



**IDENTIFICATION OF DISEASE SPECIFIC DEPENDENCY GENES
THROUGH COMPARATIVE ANALYSIS OF B-CELL DERIVED
MALIGNANCIES USING A NOVEL BIOINFORMATIC APPROACH**

A thesis submitted in part requirement for the degree of

Doctor of Philosophy

Newcastle University Centre for Cancer

Faculty of Medical Sciences

Biosciences Institute

July 2024

Hande ATASOY

Declaration

I certify that no part of the material documented in this thesis has previously been submitted for a degree or other qualification in this or any other university. I declare that this thesis represents my own unaided work, carried out by myself, under the guidance of my supervisors.

Hande ATASOY

July 2024

Acknowledgements

At the very outset, I would like to express my deepest sense of gratitude and indebtedness towards my supervisors Dr Gordon Strathdee, Dr Chris Bacon for their continuous scientific support during my PhD study along with their patience, motivation and immense source of knowledge in cancer biology. This thesis would not have been possible without the advice, discussion, guidance and support of you. I sincerely thank for all their ideas, contributions of time, and for encouraging my research work and making my PhD experience productive and stimulating. Gordon, huge thanks for the extensive amount of time and effort you dedicated to supporting my training and research.

I would like to thank my project progression assessors, Professor Ian Hickson and Dr Simon Whitehall for their encouragement, thoughtful comments and suggestions that widen my research from various perspectives. Beside my supervisors and assessors, I would like to thank Republic of Turkey-The Ministry of National Education, for providing me the opportunity to do my PhD at Newcastle University.

I would also like to thank all the past and current members of Gordon's group: Lalchununga, Chalermisin Permtermsin, and past and current member of Pual O` German Building; Emre Ozturk, Nurettin Ayvali and soul of Ahmet Yagbasan for their contribution and support during my PhD study.

I cannot conclude my acknowledgements without mentioning continuous support of my dear friends who live in Turkey. Even though we were far from each other, but I still felt close you thanks to your support. And my lovely Turkish friends who locate in Newcastle, I cannot imagine how would be living in here without you.

Last, but not least, my special thanks to my family whom I dedicate this thesis to: my daughter Bilge, my husband Gokhan and my mother Gulsen, my father Kemal and brother Hasan and his family. Thank you all for their countless sacrifices and your endless support. Bilge, you are my sunshine every day and my star every night and your love kept me strong. Gokhan, my dearest, this PhD cannot be possible without your continuous support and sacrifices. I really appreciate both of your time being patient and calm when I overwhelmed.

Abstract

Despite recent improvements in cancer treatment, many patients still relapse after initial treatment and survivors often exhibit serious long-term health effects. Therefore, more cancer specific treatments, leading to increased survival and reduced long-term toxicities are required.

Cancer cells exhibit consistent and dramatic changes in global DNA methylation patterns. Our group has recently developed a new bioinformatic approach which integrates these changes in global DNA methylation with genome-wide expression data to identify candidate subtype specific synthetic lethal genes in specific cancer subtypes. A limitation to this approach, is that any potential therapeutic targets identified are only relevant in the specific subset and not all patients with that cancer type. However, as highly similar methylation changes are shared across all B-cell derived malignancies, we hypothesised that this approach could be expanded to allow comparison between different types of B-cell malignancies to allow the identification of Disease Specific Dependency Genes (DSDG) and disease specific tumour suppressor genes (TSG), which would be functionally important in all subtypes of a particular cancer type. This analysis was able to provide proof-of-principle that the bioinformatic approach could successfully identify candidate functionally relevant genes at the whole disease level. In total, 13 candidate genes (seven DSDG and six TSG), were identified. Functional assessment of DSDG candidates did not identify any clear functional impacts of targeting DSDG by siRNA-mediated knockdown, however, this approach resulted in only partial, temporary reductions in expression and alternative approaches to allow stable knockdown will be required. Functional assessment of TSG candidates in ALL identified *SLC22A15*, as a novel negative regulator of ALL cell growth across all tested ALL genetic subtypes. These results provide initial proof-of-principle that our original approach for identifying functionally relevant genes at a subtype specific can be expanded to the whole disease level and can uncover previously unknown functionally relevant genes.

Table of Contents

Declaration	1
Acknowledgements.....	2
Abstract	3
1 INTRODUCTION	17
1.1 General View of Haematopoiesis	18
1.1.1 B-Cell Differentiation.....	20
1.2 Haematological Malignancies.....	25
1.2.1 Leukaemia Overview	25
1.2.1.1 Acute Lymphoblastic Leukaemia	26
1.2.1.1.1 BCP- ALL	27
1.2.1.1.2 Genetic Risk Factors and Cytogenetic Subclasses of BCP-ALL	28
1.2.1.2 Chronic Lymphocytic Leukaemia (CLL)	29
1.2.2 Lymphoma Overview	31
1.2.2.1 Mantle Cell Lymphoma	32
1.2.2.2 Diffuse Large Cell Lymphoma	33
1.2.2.3 Primary Central Nervous System Lymphoma	34
1.3 Epigenetics	35
1.3.1 DNA Methylation	37
1.3.2 Basic Mechanism of DNA Methylation	39
1.3.3 Locations of DNA Methylation.....	42
1.3.4 Other Epigenetic Mechanism.....	43
1.3.4.1 Histone Modifications	43
1.3.4.2 Non-Coding RNAs	47
1.3.4.3 Crosstalk of Different Epigenetic Layers	49
1.3.5 Role of DNA methylation on Haematopoiesis and Haematological Malignancies 50	
1.4 Cancer Overview	52
1.4.1 Genetics of Cancer.....	54
1.4.1.1 Tumour Suppressor Genes	56
1.4.1.2 Oncogenes	59
1.4.1.3 Non-Oncogene Addiction.....	60
1.4.1.3.1 Synthetic Lethality	61
1.4.1.3.2 Cancer Dependency Gene.....	63
1.5 Novel Bioinformatic Approach to Identify SL / Disease Specific Dependency Gene(s) 65	
1.6 Hypothesis and Aims of Project.....	66

1.6.1	Hypothesis.....	66
1.6.2	Aims of the Project	69
2	<i>MATERIALS AND METHODS</i>	70
2.1	List of Reagents.....	71
2.2	Cell Culture Methods	72
2.2.1	Cell Lines	72
2.2.2	Passaging and seeding of cells.....	73
2.2.3	Counting cells to estimate cell densities	74
2.3	Transduction of cells with lentiviral expression constructs	75
2.3.1	Production of Expression Vector (pSIN-SIEW-)	75
2.3.1.1	Recombination of target sequence into Plasmid	75
2.3.1.2	Transformation of competent bacteria	75
2.3.1.3	Extraction of Plasmid DNA	76
2.3.2	Lentiviral production using the 293T packaging cell line	77
2.3.3	Lentiviral Transduction of cell lines.....	81
2.3.4	Assessing GFP expression by flow cytometry.....	81
2.4	Cell Death Experiments.....	81
2.4.1	Assessment of Cell death by Annexin V and PI staining.....	81
2.4.2	Assessment of apoptosis by caspase activation	82
2.4.3	Assessment of apoptosis indirectly by Caspase Inhibition Assay	83
2.5	Assessment Of Cell Proliferation After Lentiviral Transduction	83
2.6	RNA Extraction and cDNA Synthesis	84
2.7	qRT-PCR.....	85
2.8	Preparation Of Cell Pellets and Protein Extraction	87
2.9	Determination of Total Protein Concentration By Pierce® Bca Protein Assay	88
2.10	SDS Page And Western Blot.....	88
2.11	siRNA Mediated Transient Knockdown Of Target Genes In Leukaemia And Lymphoma Cell Lines	90
2.11.1	Checking transfection efficiency by Flow-cytometer	91
2.11.2	. Cell Viability assessment by XTT	91
2.11.3	Assessment of caspase activation to assess apoptosis	92
2.12	Treatment Of Cell Lines with Cytotoxic Agent	93
2.13	ELISA Assays To Assess Protein Secretion.....	93
2.14	CTGF Inhibitor Antibody Experiment	96
2.15	Bioinformatic Analyses.....	96
2.15.1	DNA methylation Datasets used	96
2.15.2	Cleaning and processing methylation data	99
2.15.3	Gene expression datasets	100
2.15.4	Integration DNA methylation and gene expression data	101

2.15.5	Identification of DMR and Cancer Specific DSD/TS genes candidates	101
2.16	Statistical Analysis	104
3	<i>RESULTS OF IDENTIFICATION OF TSG AND DSDG CANDIDATES VIA NOVEL BIOINFORMATIC APPROACH IN B-CELL MALIGNANCIES</i>	105
3.1	Introduction.....	106
3.2	Aims of the Chapter	107
3.3	Results.....	107
3.3.1	ALL specific candidate genes	111
3.3.1.1	ALL specific TSG candidates.....	111
3.3.1.2	ALL specific DSDG candidates.....	113
3.3.2	MCL specific candidates.....	116
3.3.3	CLL specific candidates	119
3.3.4	PCNSL specific candidate	121
3.3.5	PCNSL & DLBCL specific candidate	122
3.3.6	Overall Results	124
3.4	Discussion.....	130
4	<i>FUNCTIONAL ASSESSMENT OF ACUTE LYMPHOBLASTIC LEUKAEMIA-SPECIFIC TUMOUR SUPPRESSOR CANDIDATES IN CELL LINE MODELS</i>	136
4.1	Introduction.....	137
4.1.1	Lentiviral system for a gene-delivery	138
4.2	Aims of The Chapter	139
4.3	Functional Assessment	139
4.3.1	Expression of TSG candidates.....	139
4.3.2	Cloning of the Candidate TSG Into the Lentiviral Expression Vector	141
4.4	Functional Validation of SLC22A15	144
4.4.1	Impact of SLC22A15 gene re-expression following transduction of ALL cell lines	144
4.4.2	Cell survival impact of re-expression of SLC22A15 on ALL cell lines.....	146
4.4.2.1	Assessment of apoptosis using binding of annexin V	146
4.4.2.2	Assessment of Caspase 3/7 activity following SLC22A15 re-expression	149
4.4.2.3	Assessment of the ability of caspase inhibition to block the induction of cell death following SLC22A15 re-expression	151
4.4.3	Assessment of the impact of re-expression of SLC22A15 on cell proliferation .	153
4.4.4	Impact of re-expression of SLC22A15 on lymphoma cell lines.....	155
4.4.5	Assessment of mRNA-expression of SLC22A15 after transduction in All and Lymphoma cell lines.....	157
4.4.6	Assessment of protein expression of SLC22A15 by Western blotting.....	159
4.5	Functional Validation of <i>THEM4</i>	163
4.5.1	Re-expression of <i>THEM4</i>	163
4.5.2	Assessing <i>THEM4</i> protein expression on ALL cell lines after transduction	166

4.6	Functional Validation of <i>TTC12</i>	167
4.6.1	Re-expression of <i>TTC12</i> in leukaemia cell lines	167
4.6.2	mRNA expression level after <i>TTC12</i> re-expression	169
4.7	Transduce All Cell Lines With pSIN-MAP9 Vector	171
4.8	Discussion.....	174
5	<i>FUNCTIONAL ASSESSMENT OF ACUTE LYMPHOBLASTIC LEUKAEMIA-SPECIFIC DISEASE SPECIFIC DEPENDENCY CANDIDATES IN CELL LINE MODELS</i>	180
5.1	Introduction.....	181
5.1.1	siRNA mediated knockdown.....	181
5.1.2	Targeting DSDG candidates as a therapeutic strategy	183
5.1.2.1	BMP2 Pathway.....	183
5.1.2.2	CTGF Pathway.....	185
5.1.2.3	NPY Pathway.....	185
5.1.2.4	Crosstalk between candidates.....	186
5.2	Aims of the Chapter	187
5.3	Results.....	187
5.3.1	Expression Profiles of DSDG Candidates in cell line models.....	187
5.3.1.1	BMP2 Expression Profile	187
5.3.1.2	CTGF Expression Profile	189
5.3.1.3	NPY Expression Profile	190
5.3.2	Functional analysis of <i>BMP2</i>	191
5.3.2.1	Assessment of secreted BMP2 protein by ELISA	191
5.3.2.2	Determination of intracellular BMP2 protein expression in cell lines by Western Blotting	193
5.3.2.3	Determination of BMP inhibitor effect to ALL and Lymphoma cell line`s growth	194
5.3.2.4	BMP2 siRNA transfection on ALL and Lymphoma Cell Lines	196
5.3.2.4.1	Transfection Efficiency	196
5.3.2.4.2	Knockdown Efficiency of <i>BMP2</i> -siRNA Transfection.....	199
5.3.2.4.3	Effect of BMP2 siRNA transfection on Cell Proliferation	199
5.3.2.4.4	Effect of <i>BMP2</i> siRNA transfection on Cell Survival	200
5.3.3	Functional analysis of <i>CTGF</i>	201
5.3.3.1	Determination of secreted CTGF protein on cell lines media by ELISA	201
5.3.3.2	CTGF siRNA transfection on ALL and Lymphoma Cell Lines	202
5.3.3.2.1	Knockdown Efficiency of CTGF-siRNA Transfection.....	202
5.3.3.2.2	Effect of <i>CTGF</i> siRNA transfection on Cell Proliferation.....	203
5.3.3.2.3	Effect of <i>CTGF</i> siRNA transfection on Cell Survival.....	204
5.3.3.3	Antibody against to CTGF	205
5.3.4	Functional assessment of <i>NPY</i>	206

5.3.4.1	Determination of secreted NPY protein in cell lines media by ELISA.....	206
5.3.4.2	Determination of intracellular NPY protein by Western Blotting	207
5.4	Discussion.....	209
5.4.1	All Candidate Expression Profile	209
5.4.1.1	mRNA expressions of Candidate Genes	209
5.4.1.2	Protein expression of Candidate Genes	210
5.4.2	siRNA mediated knockdown of candidate genes.....	211
5.4.3	Inhibitors	212
6	<i>FINAL DISCUSSION</i>	214
6.1	Overview of Study.....	215
7	<i>REFERENCES</i>	220

List of Figures

Figure 1.1. The Classical and Revised Models of Hematopoietic Hierarchy.	19
Figure 1.2 Antibody Structure.	20
Figure 1.3 B-cell Development.	22
Figure 1.4 Stages of B-cell Development and Differentiation.	24
Figure 1.5 ALL statistics in UK.	26
Figure 1.6 CLL statistics in UK.	30
Figure 1.7 Frequency of Mature B-Cell Non-Hodgkin Lymphomas.	32
Figure 1.8 Schematic overview of origin of major Non-Hodgkin Lymphoma subtypes.	34
Figure 1.9 Types of epigenetic modifications.	37
Figure 1.10 Schematic representation of DNA methylation.	37
Figure 1.11 Representative model of how DNA methylation affects gene expression.	38
Figure 1.12. Schematic representation of DNMTs family.	40
Figure 1.13 Illustration of transcription repression by cytosine methylation through MeCP2 protein.	41
Figure 1.14 Illustration of the relationship of CpG Island, CpG shore and CpG shelf regions.	43
Figure 1.15 Illustration of Structure of Nucleosome.	44
Figure 1.16 Schematic Overview Post-Translational Modifications of the Histone Tails.	45
Figure 1.17 Illustration of Epigenetic crosstalk between different mechanism.	50
Figure 1.18 Percentage of total incidence and percentage of total death of 20 cancer types over 185 countries.	53
Figure 1.19 Percentage of total incidence and percentage of total death of most common cancer types in male (A) and female (B).	54
Figure 1.20 Gatekeeper, caretaker, and landscaper mutations.	57
Figure 1.21 Knudson's two-hit hypothesis in Retinoblastoma both inherited and sporadic.	59
Figure 1.22 Schematic representation of cellular effects of oncogene and non-oncogene addiction: synthetic lethality and cancer dependency gene/pathway.	61
Figure 1.23 Schematic overview of synthetic lethality.	62
Figure 1.24 Illustration of oncogene and non-oncogene addiction through tumour development.	64
Figure 1.25 Outline of the Process That Leads to the Proposed Generation of SL Genes in ALL.	66
Figure 1.26 Average beta values of 7107 CpG sites across B-Cell Malignancies and normal B cells.	68
Figure 2.1. Schematic representation of vectors used for producing the lentiviral supernatant.	79
Figure 2.2. Overview of the Human Methylation 450K BeadChip arrays.	97
Figure 3.1. Represent examples that does not meet (A) and does meet (B) methylation and expression criteria for DSDG for ALL.	110
Figure 3.2. Methylation and Expression pattern of CCND1 gene.	110
Figure 3.3. Methylation and expression data.	113
Figure 3.4. Methylation and expression data.	116
Figure 3.5. Methylation and expression pattern of MCL DSDG candidates.	118
Figure 3.6. SOX11 expression on panel of ALL subtypes, and other Leukaemia Types.	119
Figure 3.7. Methylation and expression pattern of CLL specific DSDG candidate, SNX18.	120
Figure 3.8. Methylation and expression pattern of PCNSL TSG candidate, FSCN1.	122
Figure 3.9. Methylation and expression pattern of the DLBCL/PCNSL TSG candidate SDK2.	123
Figure 4.1. Schematic diagram of main features of expression vector.	138
Figure 4.2. Expression pattern of each candidate genes on both primary sample and model cell lines.	141
Figure 4.3. An electrophoresis results of pSIN-TEHM4/SLC22A15/MAP9/TTC12 after DNA digestion.	143
Figure 4.4. Results of Re-expression of SLC22A15 in ALL cell lines.	146
Figure 4.5. Representative flow cytometer results of Annexin V/PI and GFP in Reh cell line after transduction with psin-SLC22A15.	148
Figure 4.6. Cell death assessed by annexin V/PI staining following re-expression of SLC22A15 in Reh and PreB697 cell lines.	149
Figure 4.7. Caspase 3/7 activation after SLC22A15 re-expression.	150

Figure 4.8. Caspase 3/7 inhibition does not impact of GFP-positive cells after transduction with pSIN-SLC22A15.	152
Figure 4.9. SLC22A15 re-expression does not impact the rate of cell proliferation in ALL cell lines.	154
Figure 4.10. Impact of re-expression of SLC22A15 on Lymphoma cell lines and comparing results with ALL cell lines.	156
Figure 4.11. SLC22A15 in transduced and parental cell lines.	158
Figure 4.12. Retention of SLC22A15 protein expression following transduction.	163
Figure 4.13. Representative Flow-Cytometer Results of transduction both testing group and control group on PreB697 Cell Line post-transduction at day4 and day22.	164
Figure 4.14. Results of ALL cell lines after transduced with pSIN-SLC22A15 and p-SIN-EV	165
Figure 4.15. THEM4 protein expression following transduction with the THEM4 re-expression construct	167
Figure 4.16. Retention of GFP positive cells following transduction with the pSIN-SIEW-TTC12 vector	169
Figure 4.17. TTC12 expression at the mRNA level following transduction of the ALL cell lines	171
Figure 4.18. Examples of GFP expression in Reh and 293T cell lines at day4 after transduction with various expression constructs	173
Figure 5.1. Schematic representation of siRNA interference in a mammalian cell	182
Figure 5.2. Schematic representation domains of human BMP2	184
Figure 5.3. Representative graph of BMP2 mRNA expression on parental ALL and Lymphoma cell lines	188
Figure 5.4. Representative graph of CTGF mRNA expression on parental ALL and Lymphoma cell lines	189
Figure 5.5. Representative graph of NPY mRNA expression on parental ALL and Lymphoma cell lines	190
Figure 5.6. ELISA result of BMP2 protein expression in parental ALL and Lymphoma cell line`s media	192
Figure 5.7. Determination of intracellular BMP2 protein by Western Blotting on ALL and Lymphoma Cell lines	194
Figure 5.8. Effect of BMP inhibitor on proliferation of ALL and Lymphoma Cell Lines	195
Figure 5.9. Determination of Transfection Efficiency of ALL and Lymphoma Cell Lines	198
Figure 5.10. siRNA mediated knockdown result of BMP2	199
Figure 5.11. Cell Proliferation after BMP2 siRNA transfection in ALL and Lymphoma cell lines	200
Figure 5.12. Results of Caspase 3/7 activation after BMP2 siRNA mediated transfection on ALL and Lymphoma Groups	201
Figure 5.13. siRNA mediated knockdown result of CTGF	202
Figure 5.14. XTT results after CTGF siRNA transfection	203
Figure 5.15. Detection of apoptosis dependent cell death after CTGF siRNA mediated transfection	204
Figure 5.16. Results of anti-CTGF antibody on All and Lymphoma cell Lines	205
Figure 5.17. ELISA result of NPY protein expression in parental ALL and Lymphoma cell line`s media	207
Figure 5.18. Detection of intracellular NPY protein	208

List of Table

<i>Table 1.1 WHO Classification of Haematolymphoid Tumours.</i>	27
<i>Table 1.2 WHO Classification of Haematolymphoid Tumours.</i>	47
<i>Table 2.1 Clinical features of the ALL and Lymphoma cell lines in the present study.</i>	73
<i>Table 2.2 Culture densities of the cell lines used in this study.</i>	74
<i>Table 2.3 Master mix component and their volume for cDNA reaction.</i>	85
<i>Table 2.4qRT-PCR conditions</i>	86
<i>Table 2.5 Primers and their sequence used for qRT-PCR</i>	87
<i>Table 2.6 List of antibodies used in this project</i>	90
<i>Table 2.7 List of siRNA target gene and sequence used in this project</i>	91
<i>Table 2.8 Methylation data set used in study.</i>	99
<i>Table 2.9 Gene expression data set used in study.</i>	100
<i>Table 3.1 Data Information that was used for Bioinformatic Analysis</i>	108
<i>Table 3.2 Criteria for Identification of Candidate Disease Specific Dependency and Tumour Suppressor Genes</i>	109
<i>Table 3.3 All candidate information about function and investigation on both any cancer and leukaemia/lymphoma</i>	129
<i>Table 3.4 DSD and TS genes candidates identified in five types of B-cell malignancies.</i>	130
<i>Table 4.1 Identified candidate functional genes.</i>	137
<i>Table 4.2 Expected bans size for the EcoRI and PstI digest of pSIN-GTW vectors.</i>	142
<i>Table 5.1mRNA expression status of BMP receptors in ALL and Lymphoma cell lines from in-house unpublished data.</i>	189

Abbreviations

HSCs	Hematopoietic stem cells
BM	The bone marrow
ST-HSCs	Short-term self-renewing HSCs
LT-HSCs	Long term self-renewing HSCs
MPPs	Multipotent progenitor population
CLPs	Common lymphoid progenitors
CMFs	Common myeloid progenitors
MEPs	Megakaryocyte/erythrocyte progenitors
GMPs	Granulocyte/macrophage progenitors
IT-HSCs	Intermediate-term HSC
BCR	B cells as antigen receptors
CDRs	Complementarity determining regions
VH	Variable region of heavy chain
VL	Variable region of light chain
CH	Constant region of heavy chain
CL	Constant region of light chain
FOB	Follicular B cells
MZB	Marginal zone B cells
TdT	Terminal deoxynucleotyl transferase
CSR	Class switch recombination
SHM	Somatic hypermutation
FDC	Follicular dendritic cells
TH	T helper
ALL	Into acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
CLL	Chronic lymphocytic leukaemia
CML	Chronic myeloid leukaemia
BCP-ALL	B-cell precursor
T-ALL	T-cell precursor
WHO	World Health Organization
NOS	<i>Not otherwise specified</i>

M-CLL	CLL with exhibit a high level of IGHV somatic hypermutation
U-CLL	CLL have low or absent IGHV mutation
FL	Are Follicular Lymphoma
MZL	Marginal zone lymphomas
BL	Burkitt's lymphoma
NHL	Non-Hodgkin lymphoma
MCL	Mantle cell lymphoma
C-MCL	Classical MCL
L-NN-MCL	Leukemic non-nodal MCL
DLBCL	Diffuse large B cell lymphoma
GCB	Germinal centre B-cell-like subtype
ABC	Activated B-cell-like
MM	<i>Multiple myeloma</i>
PCNSL	Primary central nervous system lymphoma
UPR	Unfolded protein response pathway
Me	Methylation
Ac	Acetylation
Ph	Phosphorylation
DNA	Deoxyribonucleic acid
ncRNA	Noncoding RNA
5mC	5-methylcytosine
DNMTs	DNA methyltransferases
SAM	S- adenyl methionine
MBD	Methyl-CpG-binding domain
TDR	Transcriptional repression domain
UHRF	Ubiquitin-like, containing PHD and RING finger domain
AID/APOBEC	Activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme complex
TET	Ten-eleven translocation
HDACs	Histone deacetylases
HATs	Histone acetyltransferases
ATP	Adenosine triphosphate
rRNA	Ribosomal RNAs

tRNA	Transfer RNAs
snRNA	Small nuclear RNAs
snoRNA	Small nucleolar RNAs
nt	Nucleotide
MILE	The Microarray Innovations in Leukaemia
mRNA	messenger RNA
RNA	Ribonucleic acid
siRNA	Small interfering RNA
CpG	Cytosine-guanine dinucleotide
UTR	Untranslated region
HSPCs	Hematopoietic stem and progenitor cells
BRCA	Breast cancer type susceptibility protein
preBC	Precursor B cells
gcBC	Germinal centre B-cells
AFB1	Aflatoxin B1
TSGs	Tumour suppressor genes
bp	base pair
qRT-PCR	Real-Time Quantitative Reverse Transcription PCR
PCR	Polymerase chain reaction
APC	Adenomatous polyposis coli
Rb	Retinoblastoma
SL	Synthetic lethality
NHEJ	Non-homologous end joining
HR	Homologous recombination
DSDG	Disease specific dependency genes
FCS	Foetal calf serum
DMEM	Dulbecco`s Modified Eagle Medium
RPMI	Roswell Park Memorial Institute
PBS	Phosphate-buffered saline
DMSO	Dimethyl sulphoxide
kDa	kilodalton
LB	Luria-Bertani

SFFV	Spleen focus-forming virus long terminal repeat
IRES	The internal ribosome entry site
cPPT	Central polypurine tract
SFFV	Spleen focus forming virus
IRES	Internal ribosomal entry site
eGFP	Enhanced green fluorescent protein
WPRE	Woodchuck hepatitis virus post transcriptional regulatory element
LTR	Long terminal repeat
cDNA	Complementary deoxyribonucleic acid
PI	Propidium iodide
PS	Phosphatidylserine
nf-water	Nuclease-free water
RT	Reverse transcriptase
dNTP	Deoxynucleoside triphosphates
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
SDS	Sodium dodecyl sulphate
RIPA	Radioimmunoprecipitation assay buffer
BSA	Bovine serum albumin
ddH₂O	Double distilled
PVDF	polyvinylidene difluoride
PMS	N-methyl dibenzopyrazine methyl sulphate
XTT	2,3-Bis-(2-Methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide, disodium salt
ELISA	Enzyme-linked immunosorbent assays
BMP2	Bone morphogenetic protein 2
TMB	3,3',5,5'-tetramethylbenzidine
GEO	Gene Expression Omnibus
TCGA	The Cancer Genome Atlas
EGA	European Genome-phenome Archive
SSV	Subtype specific vulnerability genes
DMRs	Identifying differentially methylated regions
TSS	Transcriptional starting site
IN	Integrase

RNAi	RNA interference
RISC	RNA induced silencing complex
Ago	Argonaute-2
BMP	Bone morphogenetic proteins
CV%	Percentage coefficient variance
shRNA	Short hairpin RNA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
PARP	Poly (ADP-ribose) polymerase
GLOBOCAN	Global Cancer Observatory

1 INTRODUCTION

1.1 General View of Haematopoiesis

The haematopoietic system is composed of distinct cell types, each specialised for a unique purpose. It is a highly regenerative and adaptable tissue and its cells have different features, including very different life spans, cell morphology and function [1]. These cell types originate from hematopoietic stem cells (HSCs), which are primarily located in the bone marrow (BM) that is the key site for adult blood cell production [2]. According to their self-renew capacity, hematopoietic stem cells consist of three different cell populations: long term self-renewing HSCs (LT-HSCs), short-term self-renewing HSCs (ST-HSCs), and multipotent progenitors (without detectable self-renewal potential)[3]. LT-HSCs are quiescent at the steady-state to maintain permanent capacity for reconstitution, but when exposed to stress, they can be reactivated then enter cell cycle [4]. ST-HSCs have finite self-renewal capacity and produce multipotent progenitor population (MPPs) which have both a higher mitotic activity and increased differentiation activity, but no detectable self-renewal ability, compared to ST-HSCs and LT-HSCs [3].

Differentiation after this point occur stepwise manner. The first step begins with MPPs differentiating into either common lymphoid progenitors (CLPs) that only possess lymphoid-restricted differentiation ability or common myeloid progenitors (CMPs) that have the capacity to differentiate into both megakaryocyte/erythrocyte progenitors (MEPs) and granulocyte/macrophage progenitors (GMPs). These oligo-potent progenitors produce lineage-restricted haematopoietic cells which cannot dedifferentiate and do not possess self-renewal capacity. Finally, these differentiate to unipotent progenitors that produce a single blood cell lineage. Ultimately, this process leads to the production of mature blood cells [4-6].

The traditional hematopoietic hierarchy model is broadly in depicting the stepwise differentiation of hematopoietic stem cells (HSCs). However, the discovery of novel hematopoietic cell subsets requires revisions to this classic hierarchy model. Revised hematopoietic hierarchy models has additionally Intermediate-term HSC (IT-HSCs) population between ST-HSCs and LT-HSCs based on their self-renewal capacity [7, 8]. Moreover, in this model, the MPP population consists of MPP1, MPP2, MPP3, and MPP4 subpopulations which exhibit differences in cell-cycle distribution, immunophenotype, differentiation ability [9-11]. In this model, MPP1 is more similar ST-HSCs and even though MPP2 and MPP3 are distinct

myeloid-biased MPP subset, they work with lymphoid-primed MPP4 subpopulations to regulate cell production [9] (Figure1.1).

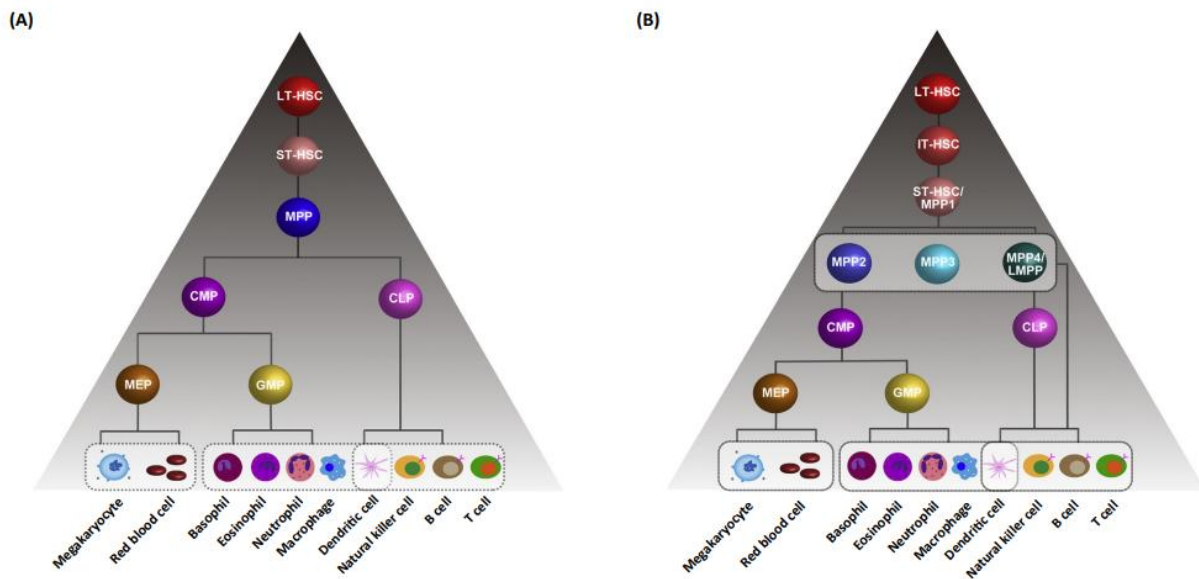


Figure 1.1. The Classical and Revised Models of Hematopoietic Hierarchy.

(A) In the classical model, LT-HSCs and ST-HSCs are both capable of differentiating into multiple cell types but occupy distinct levels in the hierarchy due to their varying self-renewal capacities. As HSCs develop into MPPs, their ability to self-renew decreases. Further downstream, MPP cells differentiate into myeloid and lymphoid lineages. CLPs can generate lymphocytes while CMPs diverge into MEP and GMP cells. (B) In the revised model, LT-HSCs, IT-HSCs, and ST-HSCs/MPP1 cells have multipotent capabilities but differ in terms of their capacity for self-renewal. HSCs differentiate different MPPs subpopulations which are MPP2, MPP3 and MPP4/LMPP at branching point. MPP2 and MPP3 cells primarily differentiate into CMPs while MPP4/LMPP cells predominantly give rise lymphoid lineages. Abbreviations: CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte/macrophage progenitor; IT-HSC, intermediate term hematopoietic stem cell; LMPP, lymphoid-primed multipotent progenitor; LT-HSC, long-term hematopoietic stem cell; MEP, megakaryocyte/erythrocyte progenitor; MPP, multipotent progenitor; ST-HSC, short-term hematopoietic stem cell. Figure is obtained from [12].

1.1.1 B-Cell Differentiation

Adaptive immunity in humans has been categorised as cellular or humoral. T lymphocytes mediate cellular responses by recognising and attacking their targets directly or indirectly with the help of other immune cells, whereas antibodies, which are produced by B lymphocyte and plasma cells, mediate humoral response [13]. Antibodies, in other words immunoglobulins, are glycosylated protein structures and they are either found on surface of B cells as antigen receptors (BCR) or are released from cells into extracellular space to bind and neutralise their target antigens. An antibody consists of two heavy protein chains and two light protein chains, held together by disulfide bonds. The N-terminus region include one variable domain of both heavy and light chain making the binding site for antigens. These variable domains include complementary-determining regions that shows sequence variance so plays role in determining specificity of antibody. On the other hand, the C terminus of both chains, called the constant region, do not involve antigen binding. Instead, it has effector roles designed to eliminate antigen. Five isotype or classes Ab have been discovered, which vary in their heavy chain constant regions: IgM, IgD, IgG, IgA, and IgE [14].(Figure1.2)

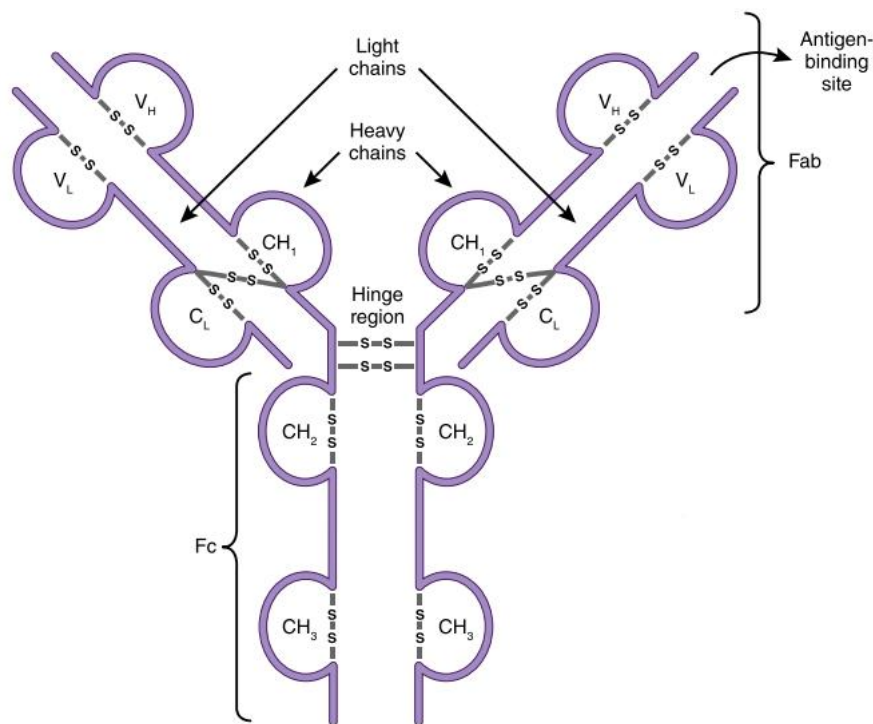


Figure 1.2 Antibody Structure.

Antibodies are consisting of two heavy chains (VH and CH) and two light chains (VL and CL). The antigen-binding fragment, Fab, is formed from one variable domain from each heavy and light chain (VH and VL). The variable domains include the complementarity determining regions (CDRs) that exhibit the high sequence diversity so that play a role in defining antibody specificity. The constant domains CH2 and CH3 of the heavy chain comprise the crystallizable fragment, Fc, which play a effector role by binding receptor on the cells. VH; Variable region of heavy chain, VL; Variable region of light chain CH; Constant region of heavy chain, CL; Constant region of light chain. Figure is adapted from [13].

Three B cells subpopulations have been distinguished based on their ontogeny and anatomic location: B1, B-2 and regulatory B cells in human. The B1 cell population, which consist of B-1a and B-1b subsets, originates from fetal liver and then persists as self-renewing populations after birth in the bone marrow. They have a key role as part of innate immune system by producing IgM to show rapid immune defence based on T-independent antigen which does not require T-cell help to trigger antibody production [13, 15]. On the other hand B-2 cells, produced from adult BM and found in secondary lymphoid organs, plays a role on adaptive immune system via differentiation into both follicular B cells (FOB) and marginal zone B cells (MZB) which are differentiate into memory B cells. Another B cell subpopulation is regulatory B cells, namely Bregs, which functions to suppress the immune response [16]. MZB are produced in the marginal sinus of the spleen and IgM is released from it to defends body from bloodborne microorganism. Like B-1 cells, they act as part of innate immunity via mediating rapid IgM responses FOB cells originate from transitional B cells in the spleen and are subsequently located in the spleen and lymph nodes. Their main role is producing long-lasting and high affinity IgG antibodies after infection [13].

B cells go through three major stages: development, maturation and function. B cell development starts in BM with an antigen independent (Ag-independent) manner. Following the differentiation of HSCs into common lymphoid progenitors (CLPs) these then further differentiate to produce distinctive intermediate B cell populations; progenitor B (pro-B) cells, precursor B (pre-B) cells, and immature B cells based on their specific cell surface marker and rearrangements of the immunoglobulin (Ig) heavy (H) and light (L) gene segments [17].

Ag-independent development in the BM starts with recombination-activating endonuclease activity of the RAG1/RAG2 complex which promotes the formation of the μ chain by allowing

V-D-J joining of Ig heavy chain gene segments. It also facilitates the transition of CLP cells into pro-B cells by promoting the expression of markers specific to the B cell lineage, such as B220, terminal deoxynucleotyl transferase (TdT), mb-1, and B29. Next, μ chains, Ig α /Ig β , and the surrogate light chains (VpreB and $\lambda 5$), which are expressed on the surface of B cells, form together to produce the pre-BCR. Thereafter, pre-BCR expression induces Ig kappa and/or Ig-lambda of L-chain rearrangement and following replacement of the surrogate light chains, then IgM is expressed on surface of immature B cell. This identifies that the precursor B lymphocyte is at the final stage of antigen-independent development and it is ready to leave BM for Ag-dependent development in the periphery [17]. Cells expressed BCRs (B cell receptor) are tested for autoreactivity in BM, before they go into peripheral to move spleen. Among this cell population, only cells which do not show robust autoreactivity, now termed transitional B cells, migrate to the spleen to complete their maturation to either follicular or marginal zone (MZ) B cells [18]. (Figure 1.3)

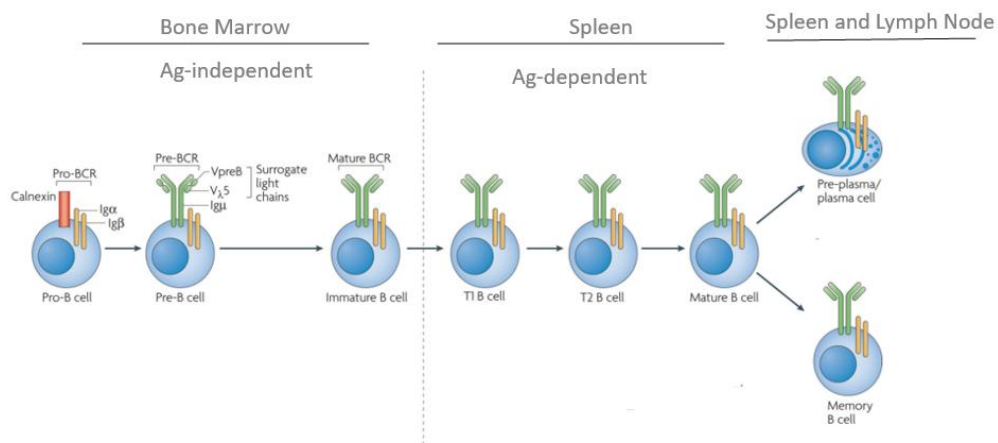


Figure 1.3 B-cell Development.

The differentiation of B cells occurs in both the BM and peripheral lymphoid tissues, such as the spleen. Within the BM, B cell development progresses through the pro-B-cell, pre-B-cell, and immature-B-cell stages. During this process, rearrangements at the immunoglobulin locus led to the generation and surface expression of the pre-B-cell receptor (pre-BCR, which is comprised of an Ig μ heavy chain and surrogate light chains (VpreB or V $\lambda 5$)) and eventually a mature B cell receptor (comprised of rearranged heavy- and light-chain genes) capable of binding antigen. At the immature stage, B cells undergo a selection process to prevent the

further development of self-reactive cells. Cells that successfully complete this checkpoint leave from the BM and enter secondary lymphoid organs (i.e., spleen) where become transitional B cells (T1 and T2), eventually maturing into either mature follicular B cells or marginal-zone B cells (not shown in the figure). Following an immune response, antigen-specific mature B cells leave germinal centre and differentiate into either plasma which secretes antibody or memory B cells. Abb, BCR; B-cell Receptor. Ag-independent; antigen independent, Ag-dependent; antigen dependent. Figures is adapted from [19].

MZ B cells responsible of initially defence against the infections are predominantly located in the spleen. When encounter with antigens, MZ B cells then rapidly develop into IgM-secreting plasma cells. On the other hand, antigen needs to be transferred to the B cell follicles of secondary lymphoid organs to enable activation of follicular B cells. Following to the antigen recognition and activation, follicular B cells have the capacity to present their antigen on MHC II to T-helper cells so that second additional activation signal provided by T-helper cell and cytokines is received. Activated follicular B cells are stimulated to differentiate into rapidly proliferating blasts that generate germinal centres (GC) [18]. Cells in the GC regions of secondary lymphoid organs express the enzyme activation-induced deaminase (AID), which deaminates cytidine residues in the VDJ and switch regions of the Ig gene, resulting in somatic hypermutation (SHM) and class switch recombination (CSR). There is a 10^6 times higher estimated mutation rate in the Ig variable (V-) regions during SHM in the GC when compared with normal somatic mutation. In addition to Ig loci, AID can target ssDNA when transcription occurs, leading to genomic damage that may result in mutations in oncogene, chromosome translocations and potentially drive the development of GC lymphomas [20]. Ig SHM, CSR and regulation of the complex network made by stroma cells, follicular dendritic cells (FDC), T cells, and the factors they express or produce give rise to the development of long lived plasma cells which have capacity to secrete class- switched immunoglobulins or memory B cells which express surface IgG, IgE or IgA [17, 18]. (Figure 1.4)

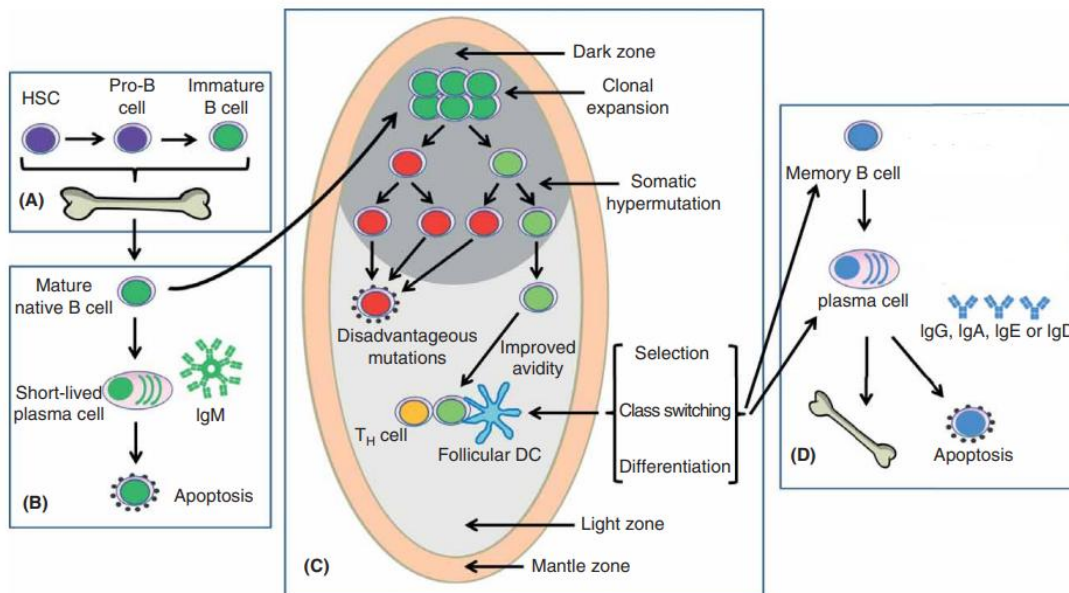


Figure 1.4 Stages of B-cell Development and Differentiation.

Stages of B-cell Development and Differentiation. (A) In the BM, HSCs are differentiating into lymphoid progenitors and B-lineage progenitor cells develop into pre-B cells, pro-B cells and immature B cells. (B) Mature naïve B cells exit the BM and encounter T cell-independent antigens in the bloodstream, prompting their differentiation into short-lived plasma cells that secrete low-affinity IgM antibodies. These IgM-producing plasma cells lack somatically mutated immunoglobulin genes and ultimately undergo apoptosis at the site of their activation. (C) Mature naïve B cells migrate into secondary lymphoid organs, where they form germinal centres. Within the dark zone of a germinal centre, the mature naïve B cells become activated, proliferating and differentiating into centroblasts. The proliferating centroblasts initiate a process known as somatic hypermutation. These centroblasts then migrate to the light zone of the germinal centre, where they transition into centrocytes. The centrocytes interact with T helper cells and follicular dendritic cells, receive survival signals to escape apoptosis. In contrast, centroblasts have unfavourable mutations undergo apoptosis. Furthermore, these centrocytes continue to proliferate and undergo selective processes. Additionally, the immunoglobulin genes of numerous centrocytes are remodelled through class switching that substitutes the originally expressed immunoglobulin heavy-chain constant region genes with those of another class (i.e., IgG, IgA, IgE or IgD). Ultimately, the selected centrocytes differentiate into either memory B cells or plasma cells and depart the germinal centre. (D) Post-germinal centre plasma cells migrate to the BM and become long lived. Figure

is obtained from [21]. BM: Bone marrow; FDCs: Follicular dendritic cells; HSCs: Haematopoietic stem cells; IgM: Immunoglobulin M; TH: T helper.

1.2 Haematological Malignancies

1.2.1 Leukaemia Overview

The term “leukaemia” is made up from two Greek words; `leukos` which means white and `hamia` which means blood. Leukaemia is used to describe various types of blood cancer that are categorized based on their morphology, immunophenotype, genetic abnormalities and clinical characteristics. Excessive white blood cells in the bloodstream of patients is indicative of leukaemia. Leukaemia clinically can be classified into four main groups based on origin of lineage; myeloid or lymphoid and onset; acute or chronic [22]. While acute leukaemia refers to excessive proliferation of immature white blood cells or blasts, chronic leukaemia indicates excessive proliferation of mature cells. In regards to origin of white blood cells, lymphoid origin refers to T or B cells, and myeloid origins refer to neutrophils, basophils, eosinophils, and monocytes lineages. In the light of this information, leukaemia is divided into acute lymphoblastic leukaemia (ALL), acute myeloid leukaemia (AML), chronic lymphocytic leukaemia (CLL) and chronic myeloid leukaemia (CML) [23].

The development of AML is the result of a transformation in progenitor myeloid or precursor myeloblast cells, causing a blockage in differentiation. This leads to an accumulation of poorly differentiated blast cells in the BM, along with disturbance in the regular production of blood cells [24]. According to the Cancer Research UK, each year around 3000 new AML patients are diagnosed in UK that counts less than 1% of all newly diagnosed cancer [25]. And around 2700 deaths related to the AML occur in UK every year and this counts for 2% of all cancer deaths in UK [26].

In respect of CML; it is defined by the unregulated proliferation of myeloid cells at various maturation stages. The translocation between the long arms of chromosome 9 and 22 (t(9;22)) is detected in nearly all patients. The t(9;22) results in the BCR::ABL fusion oncoprotein that has constitutive tyrosine kinase activity that drives uncontrolled cellular proliferation [27]. According to the Cancer Research UK, each year around 830 cases are diagnosed with CML [28] and around 230 deaths caused by CML occur in UK [29].

1.2.1.1 Acute Lymphoblastic Leukaemia

ALL is the most common childhood malignancy and is classified as B-cell precursor (BCP-ALL) produced from B lineage and T-cell precursor (T-ALL) produced from T-lineage [30]. More than 80% are B-cell precursor ALL (BCP-ALL), with the rest being T-lineage ALL [31]. Both BCP-ALL and T-ALL are characterized by multiple subtypes that are defined by structural chromosomal alterations, which are considered initiating lesions. In addition, secondary somatic alterations, such as DNA copy number alterations and point mutations also contribute to leukemogenesis. Aneuploidy and chromosomal rearrangements can lead to oncogene deregulation or the production of chimeric fusion genes. Identification of these alterations are important in terms of accurate diagnosis, risk classification, and in some cases, targeted therapy [30]. On the other hand, exposure to environmental factors like ionizing radiation, pesticides, certain solvents or viruses such as Epstein-Barr Virus can act as predisposing factors [32].

According to Cancer Research UK, ALL accounts for less than 1% of newly diagnosed cancers, with 790 new cases per year in the UK. The highest incidence occurs in individuals aged 0-4 years [33].

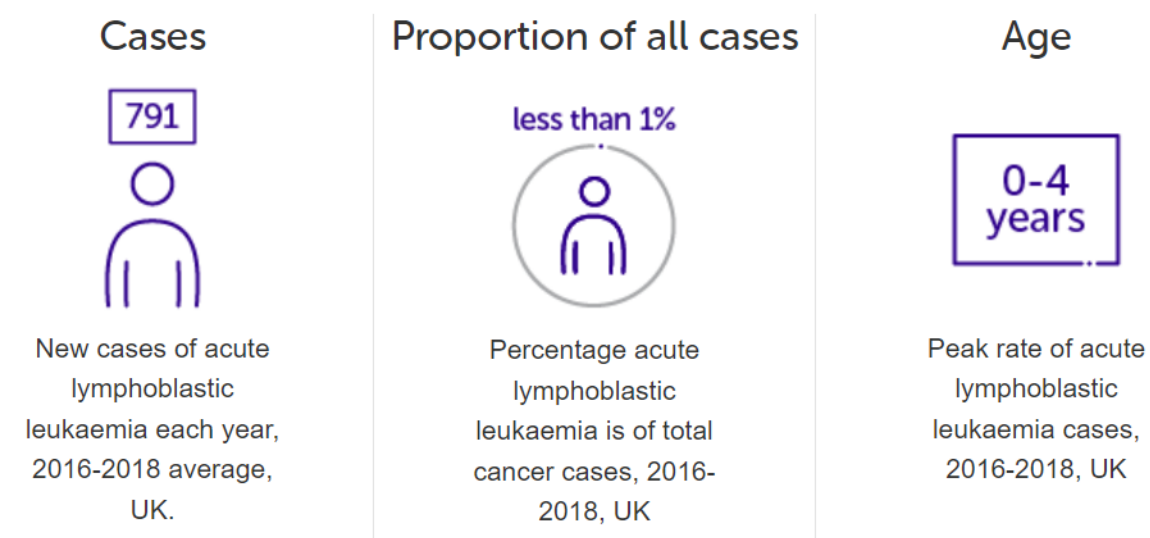


Figure 1.5 ALL statistics in UK.

Data is obtained from Cancer Research UK [33]. ALL; Acute Lymphoblastic Leukaemia

1.2.1.1.1 BCP- ALL

BCP-ALL in children generally has a favourable prognosis, with cure rates reaching around 90% for patients treated on modern risk-adjusted protocols. In spite of these advancements in treatment response, ALL remains a significant cause of cancer-related mortality in children due to relapse or treatment-related complications. Patients who survive after ALL treatment, also suffer from long-term effects of toxic chemotherapy. For these reasons, it is crucial to keep identifying patients who can benefit from less intensive therapy and to explore new targets for developing novel therapeutic agents with lower toxicity [31]. Because leukaemia has such a wide range of biological characteristics, consistent diagnostic criteria are required. The suggested categorization by the World Health Organization (WHO) aims to evaluate a broad range of available information on morphology, immunophenotype, biology, genetics, and clinical data pertaining to leukaemia [34]. According to the WHO Precursor B-cell lymphoblastic leukaemia/lymphomas are classified 11 major subgroups [35]. (Table1.1)

B-lymphoblastic leukaemia/lymphoma, NOS
B-lymphoblastic leukaemia/lymphoma with high hyperdiploidy
B-lymphoblastic leukaemia/lymphoma with hypodiploidy
B-lymphoblastic leukaemia/lymphoma with iAMP21
B-lymphoblastic leukaemia/lymphoma with BCR::ABL1 fusion
B-lymphoblastic leukaemia/lymphoma with BCR::ABL1-like features
B-lymphoblastic leukaemia/lymphoma with KMT2A rearrangement
B-lymphoblastic leukaemia/lymphoma with ETV6:: RUNX1 fusion
B-lymphoblastic leukaemia/lymphoma with ETV6::RUNX1-like features
B-lymphoblastic leukaemia/lymphoma with TCF3::PBX1 fusion
B-lymphoblastic leukaemia/lymphoma with IGH::IL3 fusion
B-lymphoblastic leukaemia/lymphoma with TCF3::HLF fusion
B-lymphoblastic leukaemia/lymphoma with other defined genetic abnormalities

Table 1.1 WHO Classification of Haematolymphoid Tumours.

WHO Classification of Haematolymphoid Tumours, 5th edition: B-cell lymphoid proliferations and lymphomas is adapted from [35]. NOS; not otherwise specified.

1.2.1.1.2 Genetic Risk Factors and Cytogenetic Subclasses of BCP-ALL

BCP-ALL is made up of various subtypes that are identified by recurring genetic abnormalities responsible for initiating the disease and are crucial for risk stratification. These genetic abnormalities are either aneuploidy which is chromosomal number alterations via gaining or losing whole chromosomes or chromosomal translocations that perturb genes through the formation of chimeric fusions or position of them near the enhancers involve hematopoietic transcription factors, epigenetic modifiers, cytokine receptors, and tyrosine kinases [30]. Thanks to genetic research, ALL subclassification has been further elucidated and inherited and somatic genetic changes in ALL development are found. Identifying these changes is important not only for diagnosis and risk-stratification purpose but also it is important to development of novel and targeted approaches [36].

Aneuploidy, which is changing in the number of chromosomes, is detected 30% of B-ALL cases by karyotype analysis [37]. B-ALL includes 3 subclasses based on numerical changes;

High hyperdiploidy is defined by gaining extra chromosomes resulting in 51 to 65 chromosomes in the leukemic cells. It constitutes around 25% of BCP-ALL and is more common in children but rare in both infants and adults. It is one of the favourable outcome subtypes of B-ALL and survival rates are above 90% [35]. The chromosomes which are gained are not random and generally include chromosomes 4, 10, 14, 21 and X. Additionally, other genetic events like mutations of the Ras pathway and epigenetic modifiers are seen in this group [38].

Another type of numerical changes B-ALL is **hypodiploidy** which is characterised by loss of at least one or more chromosomes. It is divided into three subclasses according to the total chromosome number in the cell which are near haploid with 24-31 chromosomes, low-hypodiploid with 32-39 chromosomes and high-hypodiploid with 40-43 chromosomes [35]. The incidence of hypodiploidy in adult B-ALL is 10%, while it is rare in children. Prognosis is poor and direct correlation between lower chromosome number poorer outcome is detected [39].

In regard of last subtype of numerical alteration in B-ALL is **iAMP21**. It is characterized by intrachromosomal amplification of a portion of the long arm of chromosome 21 and seen more prevalent in older children (median age 10 years) and rare occurrence in patients over 30 years old [40].

B-ALL can also be divided into four further subclasses based on translocations;

The t(12;21)(p13;q22) translocation results in the **ETV6::RUNX1** fusion and is found in approximately 25-35% of B-ALL patients that one of the highest prevalence along with high hyperdiploidy. It mostly seen in children ages between 2 and 10 and uncommon in adults (less than 3%) and infants. Cure rate with favourable prognosis is good [41].

Fusion involving the *KMT2A* gene is most common subtype of B-ALL in infants lower than 1 years old, and also shows a 15% incidence in adult ALL. It is characterised with poor prognosis in all age groups [42]. *KMT2A* (formerly known as MLL) is gene located in 11q23 and numerous partners are identified associated with translocations of *KMT2A* [36].

BCR::ABL1 fusion (also known as Philadelphia chromosome) is produced by a translocation between the *BCR* gene on chromosome 22q11.2 and the *ABL1* gene on chromosome 9q34.1. It results in expression of a constitutively active tyrosine kinase, leading to elevated downstream signalling which promotes cell growth and survival [41]. It is frequently seen in adult cases (around 25%) but is quite rare in children, with a frequency of around 2%–5% [43]. Outcome is typically poor, however this has been improved by combining chemotherapy with tyrosine kinase inhibition [36].

The *TCF3-PBX1* fusion is caused by t(1;19)(q23;p13) and is found in about 4% of BCP-ALL cases. Although initially considered high risk, the treatment outcomes for these patients have greatly improved with modern therapies [31].

1.2.1.2 Chronic Lymphocytic Leukaemia (CLL)

According to the Cancer Research UK data based on 2016-2018, approximately 3800 cases per year are diagnosed in the UK. Diagnosis is strongly age related with a peak rate between 85-89 [44]. In regard of mortality, total 976 death are associated with CLL each year. Also, the percentages of CLL-related mortality in female and male are 39% and 61% respectively. The peak rate of ages in death relation with CLL is over 90+ years [45].

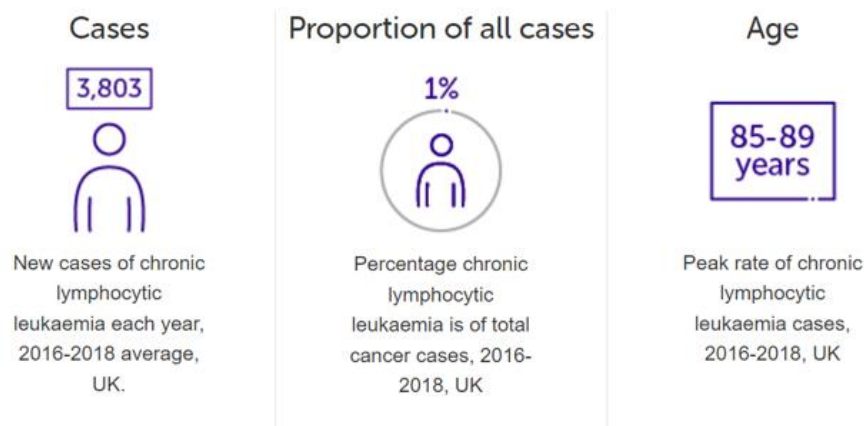


Figure 1.6 CLL statistics in UK.

Data is obtained from Cancer Research UK [44]. CLL; Chronic Lymphoblastic Leukaemia

CLL is characterized with accumulation of mature B cells in the blood, BM, lymph node and spleen [46]. Patients with CLL shows diverse clinical progress like some patients experience slow progression while others showed rapid disease course with more insensitive to therapy and overall, relatively short survival. Several clinical factors such as gender, age, Rai/Binet stage, IGHV mutational status, *ZAP70* status, and genetic abnormalities detected by FISH play a significant role in determining prognosis. Of these factors, IGHV mutational status is an important independent prognostic indicator, with IGHV unmutated cases being associated with poorer outcome[47]. CLL often begins with the loss or gain of chromosomal change, subsequently, additional mutations drive more aggressive disease. At least one of four common chromosomal alterations, del(13q), del(11q), del(17p), or trisomy 12, are detected around 80% of CLL patients [48], [49].

CLL is classified into two main subgroups depends on mutational status of IGHV which codes variable region of the immunoglobulin heavy chain [50]. CLL with exhibit a high level of IGHV somatic hypermutation (M-CLLs) are related to favourable outcome while CLL have low or absent IGHV mutation (U-CLLs) shows 98% homology to germline sequence and associated with poorer clinical outcome compared to mutated counterpart's [50], [51]. It was inferred that CLL cells which include an unmutated IGHV region are derived from a B cell in the germinal centre because they have not gone through somatic hypermutation in the IGV(D)J region. Conversely, M-CLLs originated from B cells post-germinal centre [51], [52].

1.2.2 Lymphoma Overview

Lymphoma encompasses a range of cancer neoplasms formed by lymphocytes, and may affect the lymphatic tissue, BM, or outside of the lymph nodes. According to the WHO's categorization system, over 90 distinct subtypes are identified [53]. It is initially and broadly divided as Hodgkin and Non-Hodgkin lymphoma (NHL) then stratified by cell type of origin (B-cell, T-cell or natural killer cell). Further subclassification is determined by morphology, immunophenotype, genetic and clinical features [54].

According to the analysis based on UK's Haematological Malignancy Research Network between 2004 and 2012, 5796 lymphoma patients diagnosed. Among them, Mature B-cell malignancies accounts approximately 95% while rest of identifies as subtype of T-cell malignancies. Non-Hodgkin lymphomas are more common (86% of cases), while Hodgkin lymphoma accounts for around 14% of cases.

Non-Hodgkin lymphoma is further subdivided into multiple subtypes. Diffuse large B cell lymphoma has the most frequent incidence (around 40% of total cases) with a median diagnostic age of 70 years. Around 5% of total cases comprise Mantle Cell Lymphoma with 75.5 median diagnostic age [55] (Figure 1.7). Other major subtypes are Follicular Lymphoma (FL) accounts for around 20% among the lymphomas and it is derived from GC type b-cells [56]. Most cases are linked with BCL2-IgH translocation [57]. Marginal zone lymphomas (MZLs) constitute of 10% of all NHLs. It derived from post-GC memory B-cells in the marginal zone of lymph nodes [58]. Burkitt's lymphoma (BL) is counts for around 2% of NHLs and it is aggressive type of mature B-cell NHL [56]. Most of cases shows translocations of MYC along with three immunoglobulin genes result in constitutive overexpression of MYC oncogene [35] (Figure 1.7). In regards of risk factor for lymphoma, individuals who have first- degree relatives with lymphoma have increased risk to be lymphoma [59]. In addition, infection contributes to risk of lymphoma via immunosuppression or direct transformation of lymphocyte. Tobacco, obesity and long-term pesticide exposure have also been identified as risk factors [54].

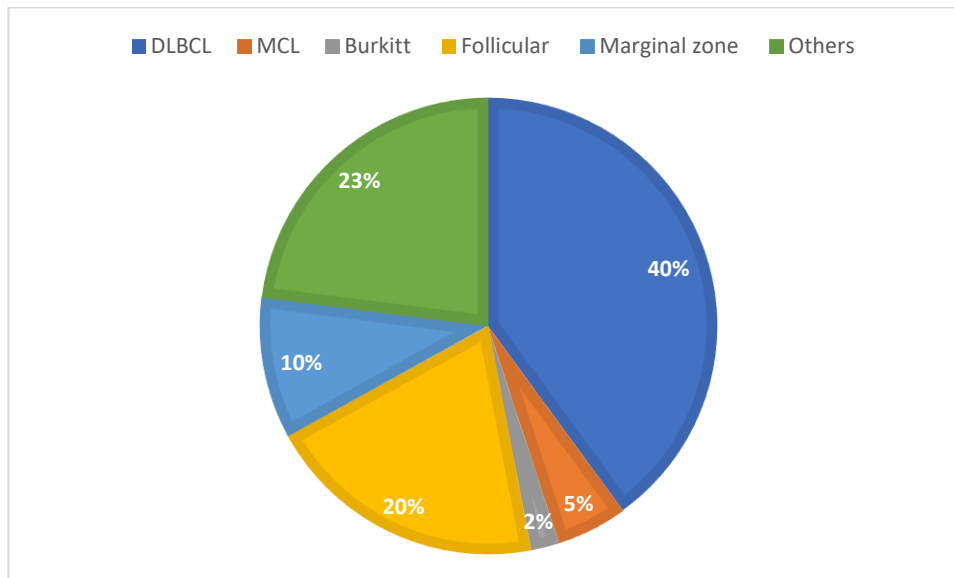


Figure 1.7 Frequency of Mature B-Cell Non-Hodgkin Lymphomas.

Figure is adapted from [56].

1.2.2.1 Mantle Cell Lymphoma

MCL as incurable type of B-cell lymphoma is characterized by t(11;14) (q13;q32) translocation leading to overexpression of cyclin D1 [60]. MCL accounts for approximately 5% of non-Hodgkin lymphomas [61]. Classification based on The 2016 WHO is divided to two subtypes show different molecular pathway resulting with distinct clinical presentations; classical MCL (C-MCL) that involves to lymph node and extra nodal site and leukemic non-nodal MCL (L-NN-MCL) that involves to BM, peripheral blood and spleen [53]. C-MCL consist of 90% of patients, who also present with expression of the transcription factor SOX11 and unmutated IGHV and are associated with an aggressive clinical progression. L-NN-MCL accounts for 10% of patients, who are SOX11 negative, mutation of CCND1 and TLR2, and somatic hypermutation of IGHV, resulting in an indolent clinical presentation [60]. C-MCL is derived from mature B cells that have not entered the germinal centre and they often develop additional abnormalities in cell cycle, DNA damage response and cell survival pathways. On the other hand, L-NN-MCL is generally indolent, but can become more aggressive with gain additional abnormalities like TP53 mutations [62]. (Figure 1.8)

1.2.2.2 Diffuse Large Cell Lymphoma

Diffuse large B cell lymphoma (DLBCL) is the most prevalent type of B-Non-Hodgkin Lymphoma (B-NHL), accounting for approximately 40% of new diagnosis in adulthood. It can be de novo or evolve from less aggressive B-NHL like FL or CLL [63]. Its name suggests large cells with nuclei that are at least twice size of lymphocytes. It can be involved both nodal and extra-nodal sites [64]. According to gene expression data, it is classified into two major groups; the germinal centre B-cell-like (GCB) subtype which arising from GC light zone B cells and activated B-cell-like (ABC) which is from later stage of GC differentiation when B cells are committed to plasmablastic differentiation. About 10-15% of cases are unclassified [65], [66]. This separation based on different stages of lymphoid differentiation is also associated with distinct tumorigenic mechanisms [67]. The median age of diagnosis for DLBCL is around the mid-60s, with about 30% of patients being over 75 years old. Research indicates that there is association between of DLBCL and inherited immunologic deficiency diseases like ataxia-telangiectasia, immune dysregulation such as autoimmune lymphoproliferative syndrome, viral such as HIV infection or environmental exposures, such as pesticides [64].

Abnormal somatic mutations are found in over 50% of DLBCL and include mutations of genes such as *PAX5* and *MYC*, which are believed to play important roles in tumour development [68]. Chromosomal translocations resulting in rearrangements of *BCL-6*, which is a transcriptional repressor expressed by germinal centre B cells to regulate GC formation, are seen in approximately 30-40% of DLBCL [69]. Further common chromosomal rearrangements underlying GCB-DLBCL target *MYC* and *BCL2*, which are also seen in Burkitt and FL, and are identified in approximately 10% and 40% of GCB-DLBCL respectively [70]. On the other hand, two primary events identified as underlying development of ABC type DLBCL; include the continuous activation of NF- κ B and also prevention of terminal differentiation into plasma cells by inactivation of PRDM1 transcription factor to trigger plasma cells differentiation [63]. (Figure 1.8)

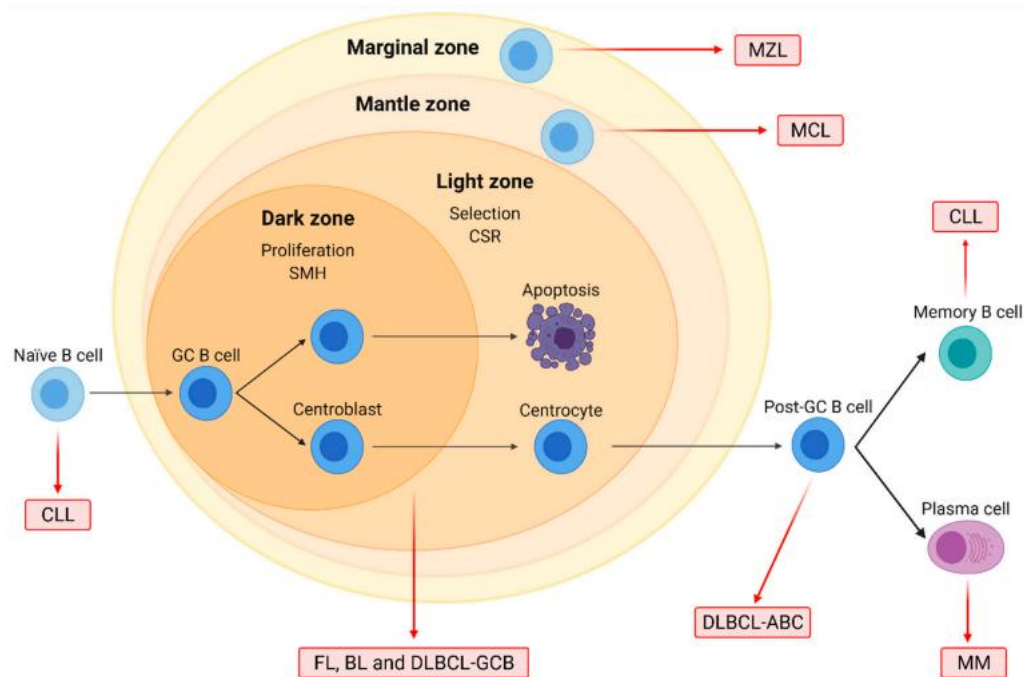


Figure 1.8 Schematic overview of origin of major Non-Hodgkin Lymphoma subtypes.

Schematic overview of origin of major Non-Hodgkin Lymphoma subtypes. Mature naïve B cells move into secondary lymphoid organ's dark zone where form germinal centre (GC) become centroblast. In the dark zone, centroblast undergo proliferation and somatic hypermutation. Germinal centre (GC) B cells serve as the normal counterparts of certain B-cell malignancies, including follicular lymphoma (FL), Burkitt lymphoma (BL), and the GC subtype of diffuse large B-cell lymphoma (GCB-DLBCL). In contrast, the activated B-cell subtype of DLBCL arises from post-GC cells. Multiple myeloma (MM) originates from terminally differentiated plasma cells. Chronic lymphocytic leukaemia (CLL) may develop from either naïve or differentiated memory B cells. Additionally, mantle cell and marginal zone lymphomas arise from B cells situated within the mantle and marginal zones of lymphoid follicles, respectively. Figure is obtained [71].

1.2.2.3 Primary Central Nervous System Lymphoma

Primary central nervous system lymphoma (PCNSL) is a rare type of extra-nodal NHL affecting brain, leptomeninges, eyes or spinal cord without widespread involvement at initial diagnosis [72]. According to the WHO classification, Primary CNS diffuse large B-cell lymphoma is classified as distinct subtype of DLBCL [35]. 90% of non-HIV-associated PCNSL are DLBCL while

the remaining cases encountered low grade lymphomas such as T-cell lymphomas or Burkitt lymphomas [73].

The only established risk factor for PCNSL is congenital or acquired immunodeficiency and PCNSL patients with HIV show significantly higher incidence compared to the general population by 3600-fold [74]. The median age of PCNSL is ranges from 53 to 57 years old in immunocompetent patients versus 31-35 years old in immunocompromised patients. With respect of gender distribution, male and female ratio is nearly same in immunocompetent patients (1.2:1.0), while there is a pronounced male bias in AIDS-associated PCNSL (male to female ratio, 7.38:1) [75].

Rubenstein et al. compared the gene expression profiles of 23 patients with PCNSL and 9 patients with lymph node DLBCL. Their analysis revealed that individual PCNSL cases could be classified into GCB, ABC, or type because most of PCNSL samples expressed either GCB marker (Bcl-6) or ABC marker (MUM1). Furthermore, PCNSLs were parted from DLBCL by elevated expression of regulators in terms of regulators of unfolded protein response (UPR) pathway, by oncogenes c-Myc and Pim-1 and by divergent apoptosis regulators [76]. Montesinos-Rongen and colleagues found that PCNSL were similar expression pattern of late germinal centre B cells and PCNSL and DLBCL are not markedly distinct from each other based on their gene expression patterns. Lastly, PCNSLs expression profile resemble of each type DLBCL by gene expression profiling analysis of 21 PCNSL comparing with brain tissue, normal B- cells and systemic DLBCL [77]. Moreover, elevated expression of multiple extracellular matrix and adhesion related pathway genes are found in PCNSL when compared to non-CNS lymphomas [78]. This difference may be linked to specific interactions between tumour cells and extracellular matrix proteins in the CNS that could explain the neuro invasion observed in PCNSL [79].

1.3 Epigenetics

Epigenetic term was coined the first time by Conrad Waddington in 1942 as combining two words; epigenesis and genetics [80]. He refers that genetic and environmental factors influence development process [81]. David Nanney suggested that environmental influences play a role in shaping epigenetic mechanisms, which in turn regulate gene expression and are maintained through cell divisions. In this way, he addressed heritable regulation of

epigenetic and its control upon gene expression [82]. Wu and Morris defined epigenetic in 2001 that epigenetic pattern of gene expression continues through cell division (mitosis and meiosis) without any alteration of DNA sequence [83]. Epigenetic marks in multicellular organisms during development could be essential for producing a wide variety of phenotypes from the same genotype and also help to adapt to respond to intrinsic and extrinsic stimuli throughout life [84]. The first epigenetic modification discovered was DNA methylation in 1940s [85] and its role in controlling gene expression was identified after approximately thirty years later [86], [87]. Since then, epigenetic studies have shifted from its initial focus on development to examine the regulatory role of epigenetic modifications in gene expression. Along with DNA methylation, other types of DNA modification like hydroxy methylation and post-translational histone modifications like acetylation, methylation, phosphorylation, ubiquitination, SUMOylation and ADPribosylation among others, have been identified so far [88]. The genetic material in mammals is packaged in the chromatin. The arrangement of chromatin to get a three-dimensional compact form is influenced by enzymatic DNA methylation and histone modifications that create epigenetic layers for controlling gene expression and cellular processes without changing the actual DNA sequence. In addition to these mechanisms, non-coding RNA has an effect on regulation of gene transcriptional activity in different physiological pathways like cell division and apoptosis. (Figure 1.9). Dysfunctional epigenetic patterns have been linked to abnormal development as well as diseases such as cancer [89].

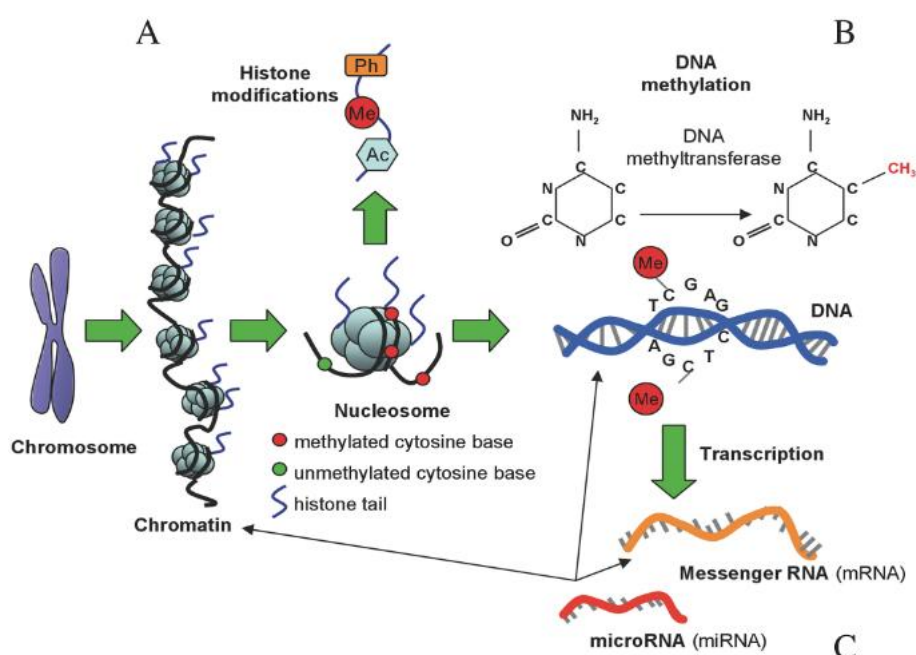


Figure 1.9 Types of epigenetic modifications.

A represents histone modifications including phosphorylation (Ph), methylation (Me) and acetylation (Ac). These modifications are involved in chromatin remodelling and transcriptional regulation. B indicates DNA methylation that adding methyl group at 5` carbon of cytosine bases and is catalyses by DNA methyltransferase enzyme to repress gene activity. C shows repressing transcriptional and translational processes by microRNAs which is a class of noncoding RNA (ncRNA). Figure is adapted from [90].

1.3.1 DNA Methylation

The most extensively researched epigenetic modification in humans is cytosine methylation. DNA methylation in mammals occurs specifically at cytosines and involves the addition of a methyl group to 5th carbon on the cytosine ring, leading to the formation of 5-methylcytosine (5mC) (Figure 1.10). DNA methylation was initially identified by Rollin Hotchkiss in 1948 with the identification of a modified cytosine (5-methylcytonise) [85], after decades with several studies that DNA methylation with other key player is major epigenetic process changing transcriptional activity in the cell [86], [87].

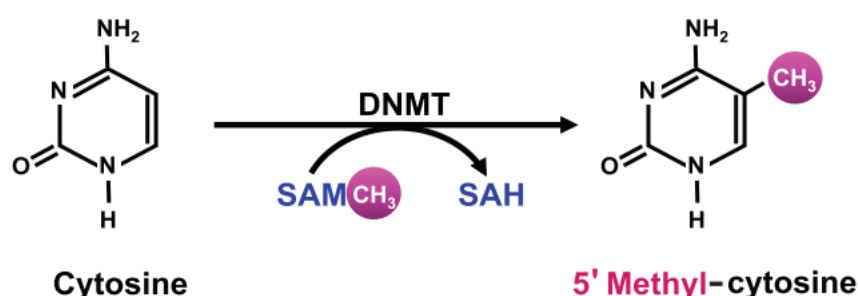


Figure 1.10 Schematic representation of DNA methylation.

DNA methyltransferases (DNMTs) catalyse addition of methyl group from S-adenyl methionine (SAM) to cytosine to form 5-methyl-cytosine. Figure is adapted from [91]

Role of DNA methylation as suppressor on gene transcription occur by three ways modifying the affinity of transcription factors for a gene promoter, affecting the binding of specific methylation recognition factors to promoters or gene bodies and changing the structure of

chromatin and the accessibility of transcription factors and other DNA-binding proteins [89](Figure 1.11).

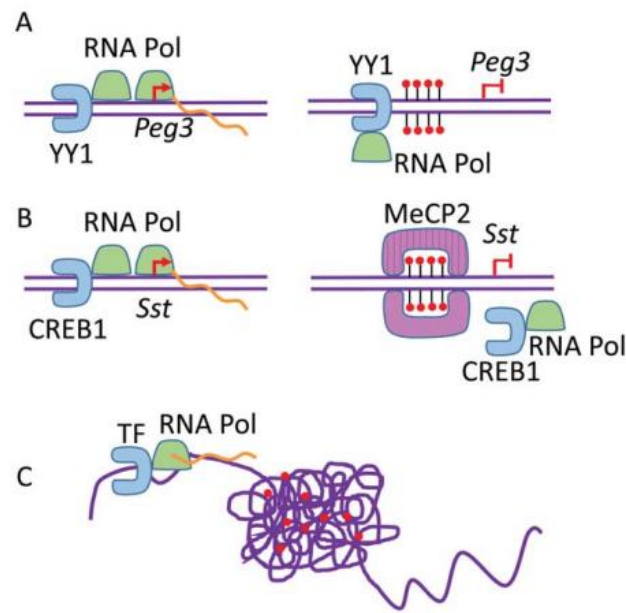


Figure 1.11 Representative model of how DNA methylation affects gene expression.

A indicates that transcription factors like YY1 are sensitive to methylated DNA and cannot initiate transcription in presence of DNA methylation. B indicates that methyl binding proteins (MBD) proteins (MeCP2 in figure) can bind to methylated DNA, thereby inhibiting transcription factor (CREB1 in the figure) access. C indicates that methylated DNA can be compacted by chromatin remodelling enzymes thereby cause restricting transcription factor binding and access to heterochromatin. Lolly with red heads (in figure A, B) and red dots (in figure C) represent methylated CpG dinucleotides; orange curves represent RNA transcripts. Figure is obtained from [89].

DNA methylation is essential for establishing genomic imprinting, whereby genes are expressed from only one of the two inherited parental alleles, as seen in X chromosome inactivation and tissue-specific gene expression [92]. In addition to, it has an important role upon preventing genome from chromosomal instability via methylation of transposable and viral elements which accounts for around 45% of mammalian genome and can cause gene distribution[84].

1.3.2 Basic Mechanism of DNA Methylation

The enzymes which are involved in regulation of biological activities due to DNA methylation can be divided into three classes: writers, erasers, and readers, writers refer to enzymes can catalyse the addition of the methyl group to cytosine. Readers are group of enzymes that recognise and bind to methyl groups to change gene expression whereas erasers can modify and remove then methyl group [92].

DNA methylation is catalysed by a family of enzymes (DNMTs) which catalyse the transfer of a methyl group from S-adenosyl methionine to cytosine residues in DNA. Five DNMT family members have been identified in mammals; DNMT1, DNMT2, DNMT3a, DNMT3b, DNMT3L [93]. DNMT1 is known as a maintenance methyltransferase and is the predominant enzyme involved in replicating the pattern of DNA methylation in newly synthesised DNA when cell divide. During DNA replication, DNMT1 localises to the replication fork where newly synthesised DNA is hemi-methylated (i.e. methylation is only present on the old, and not the newly synthesised, DNA strand) and targets the nascent DNA strand to methylate the unmethylated cytosine in the newly synthesised strand, in the context of a palindromic CpG site. In this way the methylation pattern from the original cell is copied to the two daughter cells [94], [95] (Figure 1.12-B).

Two other family members, Dnmt3a and Dnmt3b, also have important catalytic roles in establishing the cellular methylation patterns. Contrary to Dnmt1, these enzymes do not show a pronounced preference for hemi-methylated DNA and are predominantly involved in methylation of regions of the genome which were previously unmethylated. Thus, these enzymes are referred to as de novo DNMTs [96]. DNMT3a and DNMT3b are de novo DNMTs which are thought have responsibility for establish methylation pattern during embryonic development after fertilization and are highly expressed in embryonic stem cells. Both enzymes are then downregulated once cells differentiate [93] (Figure 1.12-A).

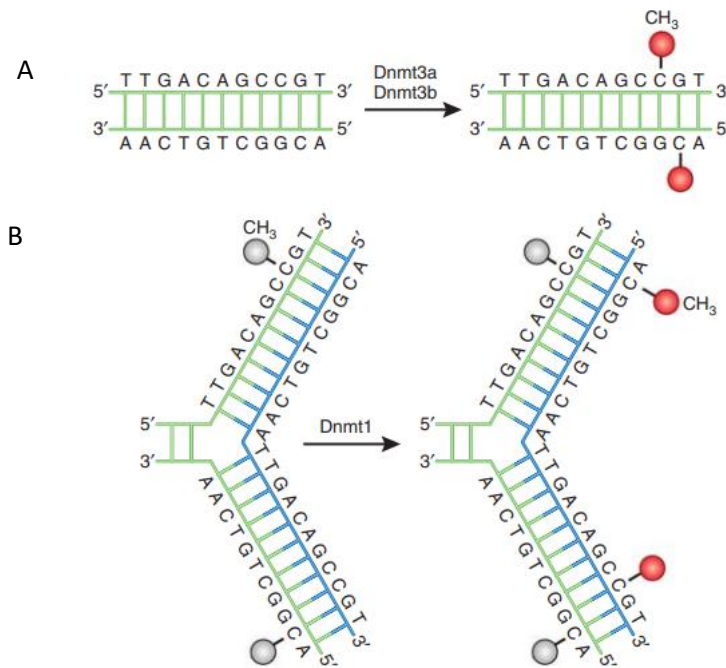


Figure 1.12. Schematic representation of DNMTs family.

A) Dnmt3a and Dnmt3b are referred to as the *de novo* DNA methyltransferases, and they are responsible for adding methyl groups (red) to previously unmethylated DNA. B) In contrast, Dnmt1 maintains the existing DNA methylation profile during DNA replication. As DNA undergoes semiconservative replication, the parental DNA strand retains its original methylation pattern (gray). Dnmt1 localizes to the replication foci and preserves the existing DNA methylation pattern by adding methyl groups to the newly synthesized daughter strand (blue). Figure is obtained from [92].

Reading of DNA methylation is directed by group of methyl binding proteins. Adding methyl group to cytosine bases causes direct suppression of gene expression by inhibiting transcription factor binding to their target sequence on gene promoter. In addition, DNA methylation allows accumulation of specific transcriptional repressor proteins to methylated DNA causing indirectly blocks transcription factor binding. One of the well-known protein family to target methylated DNA is methyl-CpG-binding domain (MBD). MBD proteins contain a conserved methyl-CpG-binding domain which enhances their affinity for single methylated CpG sites [97] and transcriptional repression domain (TDR) to bind variety of repressor complexes (review in [92]). TDR domain of MeCP2 protein, first identified member of the MBD protein family, can attach to a corepressor complex containing the transcriptional repressor mSin3A and histone deacetylases

and suppress transcription [98] (Figure 1.13). Against the other methyl binding proteins, The UHRF (ubiquitin-like, containing PHD and RING finger domain) proteins are not directly bind to DNA and suppress it, instead of initially interact with DNMT1 via its SET and RING DNA-binding domain then locate it to the hemi-methylated DNA to maintain DNA methylation especially during DNA replication [99].

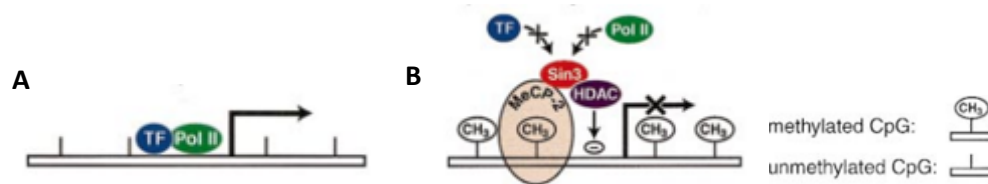


Figure 1.13 Illustration of transcription repression by cytosine methylation through MeCP2 protein.

A) Active transcription is demonstrated by the presence of unmethylated CpG sites, which enable transcription factors (TF-blue) and RNA polymerase II (green) to bind and initiate transcription. However, (B) histone deacetylase activity (HDAC-purple) can lead to a blockage at the transcription start site, preventing the binding of transcription factors and RNA polymerase II (Black arrow with cross indicates transcriptional repression). Figure is adapted from [98].

DNA demethylation is classified as either passive or active. Passive DNA demethylation takes place in dividing cells. During cell replication, Dnmt1 actively preserves DNA methylation; however, its inhibition or malfunction results in the retention of unmethylated cytosine and subsequently reduces the overall level of methylation after each cell division. Active demethylation of DNA can operate in both dividing and nondividing. This process involves enzymatic reactions to convert the 5mC through deamination by AID/APOBEC (activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme complex) or oxidation reactions by TET (ten-eleven translocation) family enzymes into a form recognized by the base excision repair pathway, which then replaces the modified base with unmodified cytosine [92].

1.3.3 Locations of DNA Methylation

The majority of DNA methylation occurs on cytosines that come before a guanine nucleotide, known as CpG sites [100]. 5mC only accounts for 1% of nucleic acids in the human genome [101]. The dinucleotide CpG has been gradually depleted from higher eukaryotic genomes during the evolution [102]. Cytosine methylation seems play an essential role in this process because most of CpG sites lost resulted from mutagenic potential of methyl cytosine to thymine through deamination [103]. The remaining CpG sites are distributed throughout the genome, being highly methylated except for CpG islands. CpG island is region where longer than 200bp with at least 50% of G+C content as a higher than rest of genome and the ratio of CpG to GpC of at least 0.6 and generally unmethylated [104]. Around 70% of gene promoters are located in the CpG island [105] and they are found unmethylated therefore they are preserved through evolution [106] and this result in interaction of transcription factors with gene promoters and allow to gene expression [94]. Because of the many transcriptions binding site in promoter have high GC content, CpG island in those regions are likely facilitating accessibility to DNA and promote transcription factor binding. In addition to CpG island in promotor region, CpG island located in chromatin is usually highly acetylated, does not have histone 1 and contain nucleosome free region. Therefore, these make it open chromatin structure may allow interaction of transcription factors to gene promoters [107]. On the other hand, methylation of CpG island is associated with suppression of gene expression result from that inhibit the binding of transcription factors, recruit repressive methyl-binding proteins. However, CpG island especially linked to gene promoters are rarely methylated [92]. CpG islands located at transcription start sites rarely associate with tissue-specific methylation patterns. Regions closely proximal to CpG islands are referred to as CpG island shores which are defined as 2kb upstream or downstream of CpG islands with lower CpG density. Methylation in these regions is more frequently associated with tissue-specific expression pattern (Fig 1.14). Methylation of CpG island shore is linked with gene suppression like methylation of CpG island [108].

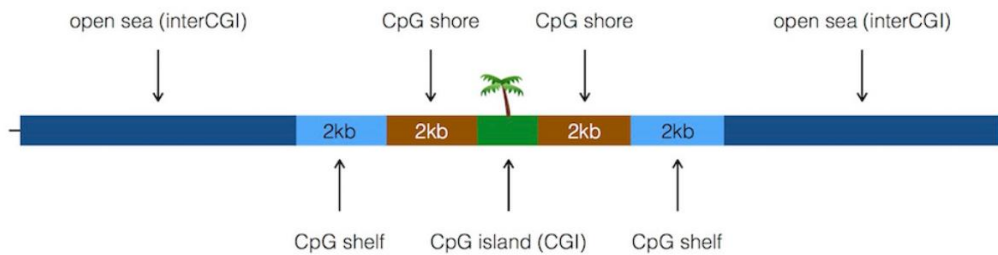


Figure 1.14 Illustration of the relationship of CpG Island, CpG shore and CpG shelf regions.

CpG island is located transcription start sites. the region is termed as CpG island shore is found 2kb upstream/downstream of the CpG island. CpG shelves are defined as another 2kb upstream/downstream of the of the CpG shores. The remaining genomic regions make up the inter-CGI annotation. Figure is obtained from [109].

1.3.4 Other Epigenetic Mechanism

1.3.4.1 Histone Modifications

Histone proteins play an essential role as part of nucleosomes, and their post-translational modifications are linked to the structure of chromatin [98]. There are five primary types of histones: H1, H2A, H2B, H3, and H4. They have a high amount of positively charged amino acids such as lysine and arginine. Nucleosomal histones are considered highly conserved proteins with nearly 100% homology in their amino acid sequences across all eukaryotic organisms. The nucleosome is the basic component of chromatin, octamer of the four core histones (H3, H4, H2A, H2B) around which 147 base pairs of DNA are wrapped – forming the nucleosome core unit [110]. These units are then linked together by histone H1 to form 30 nm fibrils (Figure 1.14). Linker histone, H1, plays a role in stabilizing chromosomes and creating more complex structures. Histones, especially their tails, are notable for the numerous and diverse modified residues they contain and with its modifications influencing the formation of heterochromatin and transcriptional activation [111].

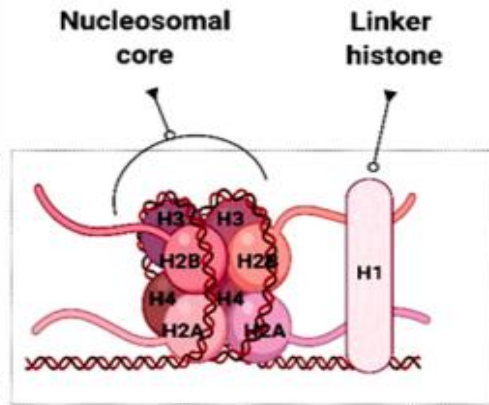


Figure 1.15 Illustration of Structure of Nucleosome.

Octamer of the four histones (H3, H4, H2A, H2B) makes nucleosome core and this unit is linked by linker histone, H1. Figure is adapted from [112].

Histone modifications serve two main functions: creating global chromatin contexts and regulating DNA-related biological processes. These alterations aid in organizing the genome into separate domains such as euchromatin and heterochromatin, facilitating accessibility of DNA for specific tasks such as transcription and replication [113]. The human genome can be divided into two distinct chromatin types based on their transcriptional activity: transcriptionally silent heterochromatin and transcriptionally active euchromatin. Heterochromatin plays a crucial role in protect chromosome ends and ensuring proper chromosome separation during mitosis. on the other hand, euchromatin has a loose configuration to allows flexibility to DNA in biological processes like transcription and typically contain high expression genes [12]. Epigenetic alterations may serve as one of the indicators for distinguishing between different types of chromatins; for instance, euchromatin often exhibits high levels of acetylation while heterochromatin tends to have low acetylation levels. A diverse range of histone modifications have been identified until now, and they are crucial for regulating chromatin state and controlling transcription processes during cell proliferation. These posttranslational modifications include methylation, acetylation, sumoylation, glycosylation, phosphorylation, poly-ADP-ribosylation, and ubiquitination [112] (Figure 1.16).

Histones can undergo simultaneous modifications at multiple sites. Each core histone forming the nucleosome can carry numerous modifications, leading to interactions between different marks. Interactions among histone modifications can take place within the same

site and across different histone tails. Consequently, a single modification on a histone does not solely determine the outcome; instead, it is the collective presence of all modifications in a nucleosome or region that determines the result.

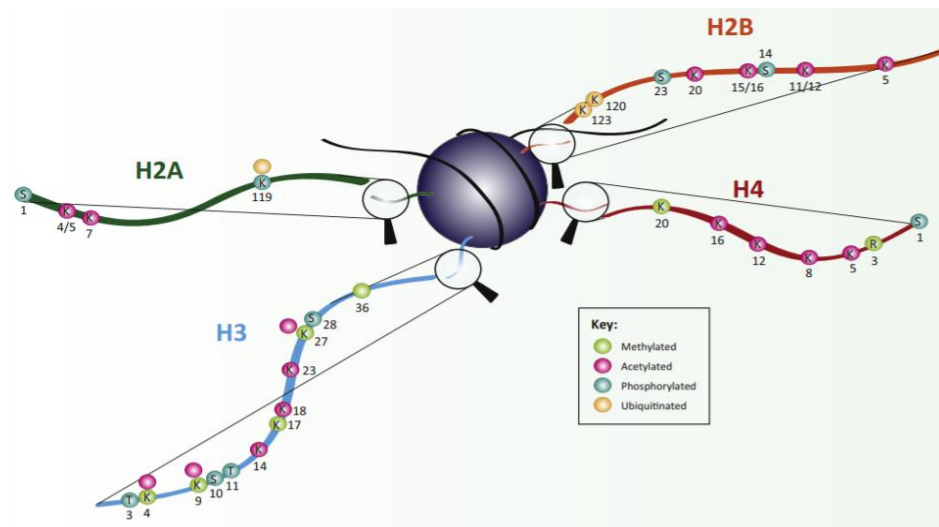


Figure 1.16 Schematic Overview Post-Translational Modifications of the Histone Tails.

The location of each modification is shown in black, and the amino acid modified at each position is also shown (K = lysine, R = arginine, S = serine, T = threonine). The modifications to each amino acid are visually represented using a colour scheme (green = methylated, pink = acetylated, turquoise = phosphorylated, beige = ubiquitinated). Figure is obtained from [114].

Acetylation of lysine is controlled by two opposing enzyme families: histone acetyltransferases (HATs) and histone deacetylases (HDACs) [115]. Acetylation refers to transfer acetyl group of acetyl-CoA to the ϵ -amino group of the internal lysine residue where near the amino termini of histone proteins by histone acetyltransferases (HATs) (Table 1.3) [116]. Adding acetyl group cause neutralization of positive charge of lysine leading disruption of the electrostatic interaction between histones and DNA subsequently decreased affinity of histone for DNA, modification/ relaxing of condense heterochromatin structure, impact gene organization then result in activate transcription process [117]. To support this hypothesis, histone acetyltransferase (HAT) activity of several transcriptional activator complexes in mammalian cells have been discovered (reviewed in [89]). On the other hand, removing acetyl group from lysine is performed by enzymes called histone deacetylases (HDACs). This results in the restoration of the positive charge of lysine, resulting in increased

histone-DNA affinity, compaction of chromatin and consequent transcriptional repression [118].

Histone methylation takes place on amino acids; arginine, lysine, and histidine [119]. Moreover, there is an additional layer of complexity in this modification; lysine can undergo mono-, di-, or tri-methylation, while arginine may be mono-, di-methylated. Differ from acetylation and phosphorylation, however, histone methylation does not change the charge of the histone protein [115]. Histone methylation is regulated by functionally two opposing group enzymes: histone methyltransferases and histone demethylases. These enzymes have a crucial role in regulating transcription and changing chromatin structure leading to coordinating gene expression, cell survival, and differentiation [120]. Lysine methyltransferases play a role in transfer methyl group from S-Adenosyl-L-methionine (SAM) to specific lysine's ϵ -amino group residues on histones which are most common in this process H3K4, H3K9, H3K27, H3K36, H3K79, and H4K20 [121], [122] (Table 1.3). Mono-, di- or tri-methyl group can be added to lysine residue but, different histone methyltransferase shows different affinity to different substrate, so each has different capacity about catalysing of different level of methylation [123]. In regards of arginine methylation, the enzymes which add methyl group to arginine residue is divided into two-classes; type-I and type-II enzymes. Both types facilitate the transfer of a methyl group from SAM to the ω -guanidino group of arginine in various substrates [118] (Table 1.3). Histone methylation has been linked to a range of cellular processes, including transcription, DNA replication, the response to DNA damage and repair, the formation of heterochromatin, and somatic cell reprogramming. Among these functions, transcriptional activation and repression have received the most attention. H3K4 trimethylation is associated with euchromatin and promotes active transcription by recruiting the RNA polymerase II complex. Conversely, H3K27 trimethylation is commonly viewed as a mark for repressing transcription [124]. On the other hand, arginine methylation of histones (H3 or H4) generally promotes transcriptional activation mediated by arginine methyltransferases (PRMTs) [125].

Phosphorylation of histones occurs predominantly at serine, threonine and tyrosine amino acids in N-terminal tail. This process is regulated by two enzymes family; kinases which use ATP for transfer phosphate group to target amino acid chain and phosphatases which remove phosphate group [126], [127]. Histone phosphorylation may act by recruiting chromosome condensation elements. Histone phosphorylation is linked with the compaction

of chromatin that takes place during mitosis and meiosis. Phosphorylation is suggested to be a crucial stage in the condensation of higher-ordered chromosomes, which plays a significant role in subsequent chromosome alignment and separation during cell division. Multiple sites undergo phosphorylation such as H3S10, H3T3, and H3T11 during mitosis or meiosis. Heavily compacted metaphase chromosomes exhibit substantial phosphorylation at all these locations [128].

Histones can undergo mono- and poly-ADP ribosylation on glutamate and arginine residues (Table 1.3). The process of histone poly-ADP-ribosylation is mediated by the enzymes from the poly-ADPribose polymerase family, while it is reversed by the enzymes belonging to the poly-ADP-ribose-glycohydrolase family. These two groups of enzymes work in coordination to regulate levels of poly-ADP ribosylated histones, which have been associated with a relatively open chromatin state [129].

Ubiquitin, a 76-amino acid polypeptide, is linked to histone lysine residue through the coordinated activity of three enzymes: E1-activating, E2-conjugating, and E3-ligating enzymes [130] (Table 1.3). Notably, two distinct sites, H2A and H2B, are found related to this process. While H2AK119ub1 contributes to gene silencing, the presence of H2BK123ub1 is critical for transcriptional initiation and elongation [131].

Sumoylation modification is a process linked to ubiquitylation and it takes place attaching small modifier molecules, similar to ubiquitin, to histone lysine through E1, E2, and E3 enzymes (Table 1.3). It has primarily been associated with suppressive function [132].

Chromatin Modifications	Residues Modified	Functions Regulated
Acetylation	K-ac	Transcription, Repair, Replication
Methylation (lysine)	K-me1, K-me2, K-me3	Transcription, Repair
Methylation (arginine)	R-me1, R-me2a, R-me2s	Transcription
Phosphorylation	S-ph, T-ph	Transcription, Repair, Condensation
Ubiquitylation	K-ub	Transcription, Repair
Sumoylation	K-su	Transcription
ADP ribosylation	E-ar	Transcription

Table 1.2 WHO Classification of Haematolymphoid Tumours

List of common histone modifications. Table is adapted from [133].

1.3.4.2 Non-Coding RNAs

More than 70% of the human genome is transcribed into non-coding RNAs (ncRNAs), making them the largest component of the genome [134]. NcRNAs can be categorized as

housekeeping (including ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs)) and the other category is regulatory ncRNAs commonly further divided based on their lengths into small ncRNAs such as small interfering RNAs (siRNAs), micro RNAs (miRNAs) and piwi-interacting RNAs (piRNAs) and long non-coding RNAs (lncRNAs) [135].

A category of small non-coding RNAs consists of 15–200 nucleotide (nt) transcripts long that play a role in regulating gene expression after transcription. Small non-coding RNAs have been found to be involved in various fundamental processes, including maintaining stemness and germline, development and differentiation, as well as silencing genes at both transcriptional and post-transcriptional levels [136].

miRNAs represent the most extensively researched category of small ncRNAs, consisting of evolutionarily conserved transcripts ranging from 17 to 25 nt. They function in regulating gene expression post-transcriptionally by either suppressing target mRNA translation or inducing mRNA degradation [137]. miRNAs exhibit complete complementarity with their mRNA targets, they can directly endonucleolytic cleavage of transcripts or suppressing their translation by partially binding on 3'UTR of a target mRNA [138].

Small interfering RNA (siRNA) is made from external origin retrotransposons and viral via processing by Dicer enzyme as a long fully complementary double strand RNA in 20-24 nucleotides length and linear form. siRNA can suppress the translation by cleavage of perfect complementary matched mRNA. It is also observed that siRNA induce heterochromatin epigenetic marks and cause direct sequence-specific transcriptional gene silencing [138], [139].

When compare siRNA to miRNA, there are some differences like dsRNA (30-100 nt length) can be transcribed or artificially introduced to the cell then Dicer process it into siRNA. Precursor of miRNA (7-100 nt length), however, produced inside the nucleus by RNA Pol II then cleaved by Drosha then transported to cytosol. In addition to, for Endonucleolytic cleavage of target mRNA by siRNA must be fully complementary matched between guide siRNA and target mRNA, in contrast partially complementary matched between guide miRNA and target mRNA will result in either degradation or translational repression of target mRNA in some cases when complementary level is high between miRNA and mRNA, endonucleolytic cleavage of mRNA [140].

In contrast to the two prior categories of small ncRNAs, piRNAs originate from a single-stranded precursor rather than a double-stranded RNA, and their production are not relied on the Dicer enzyme. It is 26–31 nt-long piRNAs and name comes from interaction to Piwi protein that play a role in epigenetic regulation [141]. The main role of piRNAs is protect genome against to transposon activity. They are involved in translational suppression of transposon mRNA degradation by identify transposon sequence then guide cleavage of transposon sequence match with 3' or 5' UTRs of mRNAs with PIWI proteins [139].

1.3.4.3 Crosstalk of Different Epigenetic Layers

DNA methylation and histone modifications are closely connected and depend on each other to preserve the epigenetic state of mammalian cells. CpG methylation can act as a signal for histone modifications. Specifically, methyl binding proteins MeCP2 and MBD not only bind methylated CpGs but also interact with enzymes involved in modifying histones such as deacetylases and lysine methyltransferases [142]. Simultaneous interaction of MBD1 with methylated CpG and SUV39H1 enables this protein to serve as a connection between CpG methylation and histone H3K9 methylation. This dual binding capability of MBD1 plays role as bridge between two distinct epigenetic level control DNA methylation and histone modification, thereby guaranteeing the silencing of targeted genes through epigenetic mechanisms. Histone methylation may also attract DNMTs. The recruitment of DNMT3a/3b to H3K9 methylated chromatin occurs through direct interaction with the heterochromatin protein HP1, which binds to methylated H3K9 via its chromodomain [143]. DNMT1 also can be recruited by methylated H3K9 through UHRF1, guaranteeing the accurate transmission of DNA methylation during mitosis [144] (Figure 1.17). This illustrates that in heterochromatin, methylated H3K9 is involved in controlling DNA methylation. In addition to, a number of miRNAs have been reported to target DNA methyltransferases or enzymes responsible of DNA demethylation, consequently influencing the state of DNA methylation [145], [146].

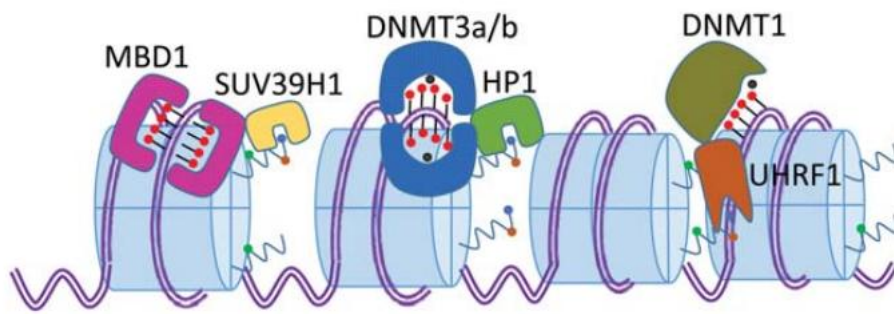


Figure 1.17 Illustration of Epigenetic crosstalk between different mechanism.

H3K9 methylation and DNA methylation are coupled processes. The MBD1 protein can identify methylated DNA and recruit the H3K9 methyltransferase SUV39H1. Similarly, the HP1 protein binds to K9-methylated histone H3 and recruits DNMT3 to allow de novo methylation to nascent DNA; UHRF1 can recognize H3K9 methylation and guide DNMT1-mediated maintenance of DNA methylation. Lolly with red heads represent methylated CpG dinucleotides; blue dots represent methylated H3K9; black dots represent S-adenosyl methionine substrate for DNMTs. Figure is obtained from [89].

1.3.5 Role of DNA methylation on Haematopoiesis and Haematological Malignancies

Hematopoietic stem and progenitor cells (HSPCs) can undergo either self-renewal by mitotic division to produce identical stem cells with their developmental potential or differentiation which leads to the production of differentiated daughter cells from same mother cells. Cellular fate, whether to undergo self-renewal or differentiation, is carefully coordinated by network of specific transcription factors, genes, signalling pathways and epigenetic regulation via changing gene expression pattern [147], [148].

Among the epigenetic changes DNA methylation, histone modifications as well as the non-coding RNAs play crucial role upon regulating B cell fate related genes at both the transcriptional and post-translational level [17].

Throughout early stages of HSC development, genes which have a role in maintaining cell number in BM via self-renewal and multipotency are in an open-access state to transcription and actively transcribed. In contrast, genes with specific functions in lineage commitment are

epigenetically suppressed. As differentiation proceeds, lineage-specific genes become active whereas genes linked to pluripotency are inactivated by epigenetic mechanisms [149].

It has been showed that DNMT1 is required for HSC differentiation from stem cells to multipotent progenitors to lineage-restricted myeloid progenitors in mice [150]. In addition, loss of DNMT3a leads to aberrant expression of self-renewal genes (Runx1 and Gata3), resulting in expansion of HSC population and blocking differentiation. Moreover, specific deletion of both DNMT3a and DNMT3b in HSC causes activation of B-catenin signalling that results in enhanced self-renewal over differentiation [151].

Hypomethylation and hypermethylation of DNA are terms used to indicate lower or higher levels of methylation compared to a standard DNA, which in the context of cancer epigenetics is normal tissue [152]. Global DNA hypomethylation is frequently observed in cancer and may be linked to chromosomal instability, transposon reactivation and loss of the genomic imprinting [153], [154]. On the other hand, it has been showed that aberrant methylation occurs in 5-10% of normally demethylated promoters in cancer. Therefore, tumour suppressor gene can be inactivated by abnormal hypermethylation and cause tumorigenesis. Moreover, this abnormal methylation effect non-coding RNA expression that contribute malignant transformation as well [155], [156]. Promoter hypermethylation appears to be as prevalent as the disruption of classic tumour-suppressor genes through mutation in human cancer and may potentially be even more common. Approximately 50% of the genes that lead to familial cancer when mutated in the germline are known to undergo silencing associated with methylation in various sporadic forms of cancer [157]. Epigenetic silencing plays a crucial role in the development of non-hereditary forms of cancer, as demonstrated by research on the BRCA1 gene. This gene was previously believed to be significant solely for familial breast cancer. However, it is now evident that 10-15% of women with the non-familial version of this disease have tumours in which BRCA1 is hypermethylated [158].

To investigate epigenetic relationship between normal B cell development and cancerous process, Kulis et al compared B-Cell cancers in comparison to their healthy cell counterparts by DNA methylome analysis. They compared different types of malignancy with normal cells at an equivalent stage of differentiation, i.e. ALL with precursor B cells (preBCs), DLBCL-GCB type with germinal centre B-cells (gcBCs) and multiple myeloma (MM) with plasma cells. Their results showed that a significant proportion of differentially methylated CpGs in tumours are

methyated through normal B cell differentiation ranging from 53% to 82% for hypermethyated sites and 29% to 84% for hypomethyated sites [159].

In addition, Norlund et al studied methylation level of 663 patients with paediatric B-ALL and 101 patients with paediatric T-ALL, and their results showed that each ALL subtype exhibited a particular methylation signature, and these signatures could be help to predict genetic subtype of paediatric ALL [160].

Genome-wide loss of methylation and site-specific gain of methylation were identified in CLL by different techniques and groups [161], [162]. Kulis et al compared subset of CLL in comparison to their healthy cell counterparts by DNA methylome analysis. They compared different subtypes of CLL (U-CLL and M-CLL) with normal cells at an equivalent stage of differentiation, i.e. U-CLL with naïve B cells, M-CLL type with memory b-cells and compared two subtypes of CLL each other. Their results showed that U-CLL were epigenetically resemble to naïve b-cells, whereas M-CLL resembled memory b-cells. Also, global hypomethylation that found gene body and enhancer region between both CLL subtypes and b-cells differentiation suggesting DNA methylation outside of promotor regions has a function [162].

.

1.4 Cancer Overview

Cancer is disease in which cells divide continuously and excessively because of the pathophysiological changes [163]. Cancer is recognised as one of the leading causes of disease-associated mortality worldwide every year. According to the Global Cancer Statistics 2020 data by Hyuna Sung et al [164], 19.3 million cancer patients were newly diagnosed and around 10 million death were recorded due to cancer in 2020 in 185 countries (GLOBOCAN). According to the data, the most common cancer types regardless of gender are breast, lung, and prostate cancer. The leading causes of cancer-related deaths are lung, liver, stomach, and breast cancer. Gender-specific trends show that lung, prostate, non-melanoma skin, and stomach cancers are most prevalent in men, while breast, lung, and cervical cancers are most prevalent in women [164] (Figure 1.19).

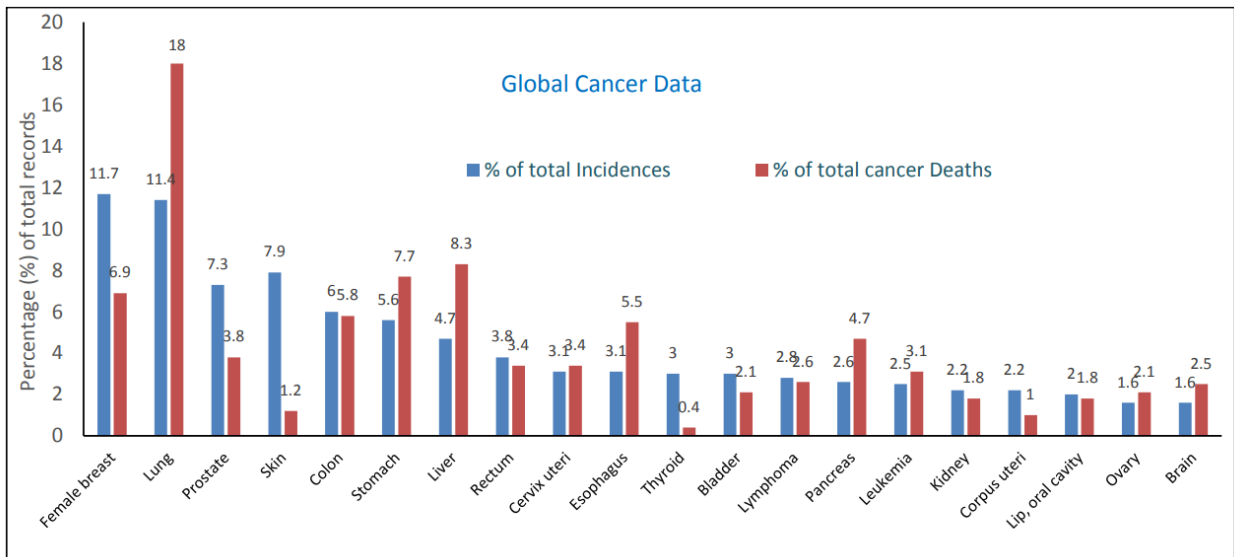
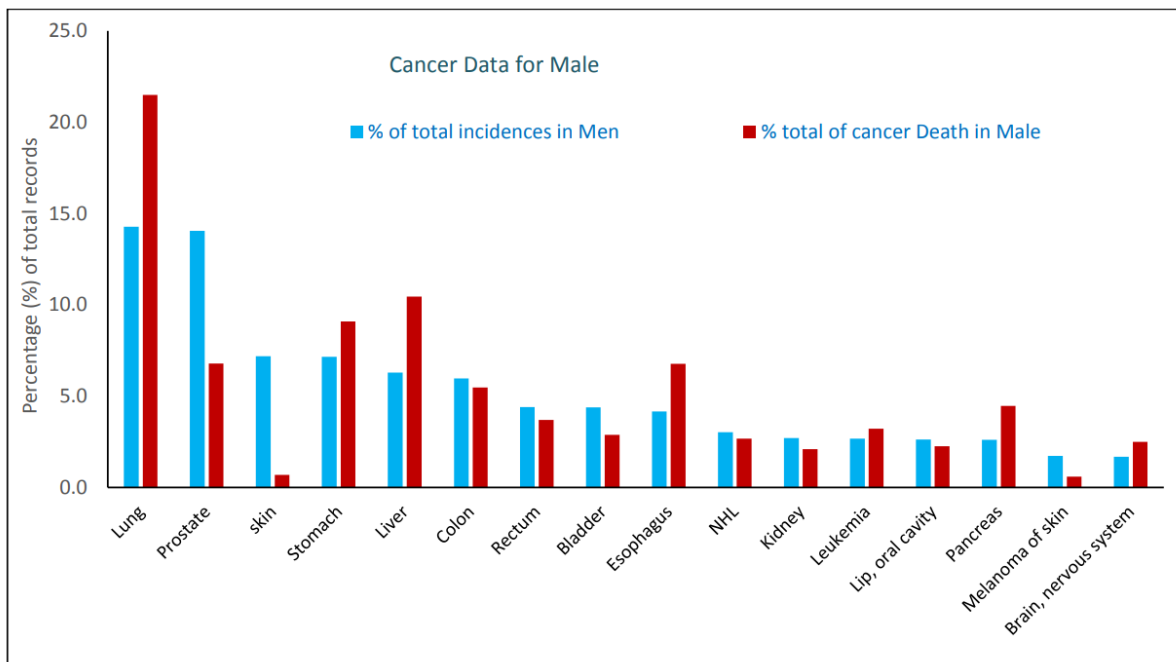


Figure 1.18 Percentage of total incidence and percentage of total death of 20 cancer types over 185 countries.

This graph was reproduced from Sung, H., et al [164].

A



B

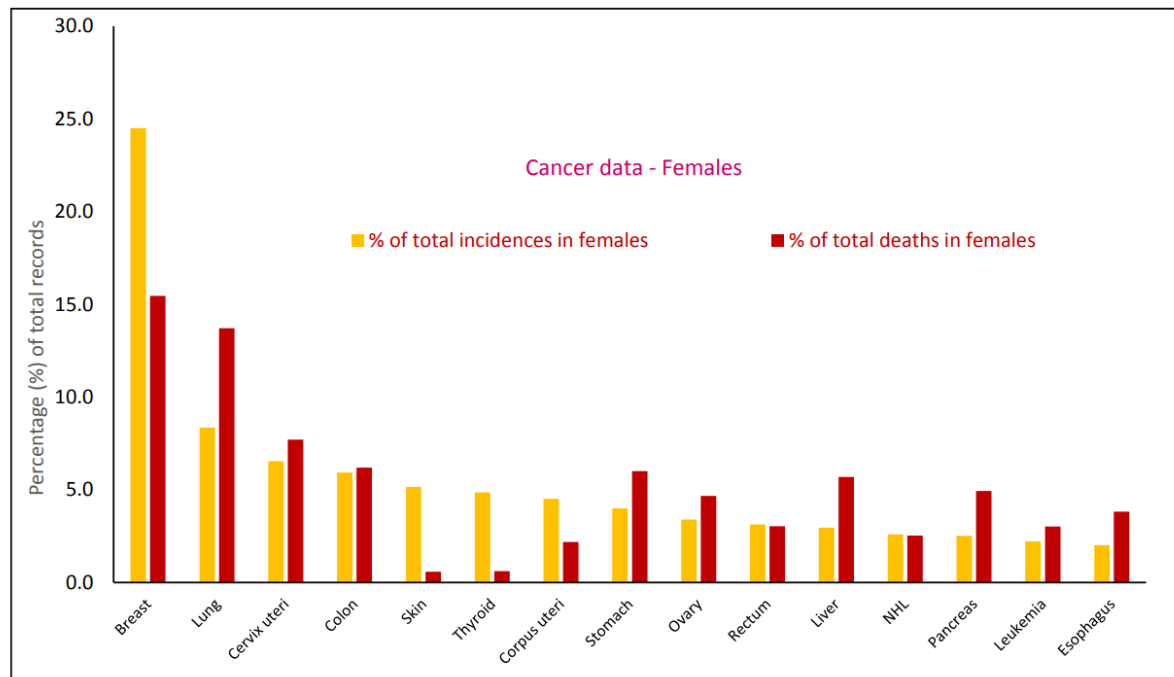


Figure 1.19 Percentage of total incidence and percentage of total death of most common cancer types in male (A) and female (B).

This graph was reproduced from Sung, H., et al [164].

1.4.1 Genetics of Cancer

A genomic role in cancer development was proposed by David von HanSEMann [165] first, then by Theodor Boveri [166] based on microscopic observation of dividing cancer cells revealed unusual chromosomal abnormalities. They proposed idea that cancers form of abnormal cell clones caused by hereditary material abnormalities. Cancer as a genetic disease arises from accumulation of mutations in genes controlling cell division, survival, invasion and other hallmarks of transformed phenotype [167]. The transformation process from normal to cancerous cell requires changes at both the genomic and epigenomic levels [168]. Family background and increasing age are can be key risk factors for many cancers type, indicating that both inherited and acquired factors which play roles in development of cancer [169], [170]. Germline mutations are congenital genetic alterations in reproductive cells derived from either sperm or egg and are typically present in every cell of the resulting offspring's body. Somatic mutations accumulate throughout an individual's lifetime as a result of internal and external factor that induce DNA damage. Detection of germline mutations is intended to determine inherited malignancies and identify high-risk families, and

detection of somatic mutation is proposed to find targeted drugs, monitor tumour loading for guided therapy, and evaluate prognosis [171].

Somatic mutations can take a number of different forms; base substitutions which change one base with another one, insertions or deletions of DNA segments, rearrangements that involve breaking and rejoining of DNA segments to other locations in the genome, copy number variations, either due to gain or loss of large chromosomal sections, or more localised gene amplification or deletion [172]. As an example, Philadelphia chromosome is produced by a translocation between the *BCR* gene on chromosome 22q11.2 and the *ABL1* gene on chromosome 9q34.1. It results in expression of a constitutively active tyrosine kinase, leading to elevated downstream signalling which promotes cell growth and survival (mentioned in section 1.2.1.1.2) [41].

The DNA sequence in normal cells is constantly exposed to mutagens originating from both intrinsic and extrinsic sources. Exposure to high levels of external mutagens such as tobacco smoke carcinogens, aflatoxins produced by fungi, and various forms of radiation like ultraviolet light can further elevate mutation rates. While most of this damage is repaired, a small portion may result in permanent mutations (170). These exposures have been linked to increased incidences of lung, liver, and skin cancer respectively. Somatic mutations in types of cancers often display specific mutational patterns that are linked to the causative mutagens (i.e., codon 249 mutation in TP53 gene was associated with exposure to aflatoxin B1 (AFB1) in hepatocellular carcinoma) [173]. Moreover, entirely new DNA sequences from external sources derived from particularly virus like human papillomavirus, Epstein-Barr virus and hepatitis B virus may have been embedded cancer cell genome and contribute pathway of some cancer types [174].

Cancer-associated genes are generally categorized into two main groups: proto-oncogenes and tumour suppressor genes (TSGs). Proto-oncogenes are commonly associated with pathways that stimulate the growth of cells. Activation of these genes through mutations or chromosomal alterations can promote cellular transformation. Mutations in proto-oncogenes generally exhibit dominant characteristics and mutant versions of proto-oncogenes are called as oncogenes [175]. On the other hand, TSGs involve several essential cellular pathways like DNA damage repair, inhibition cell division, initiation of apoptosis and repression of metastasis. Thus, loss of TSG function contributes to the initiation and promotion of cancer development [176].

1.4.1.1 Tumour Suppressor Genes

TSGs can be involve various type cellular pathway [177], including genes code intracellular proteins which responsible to control a progression from G1 to S phase of the cell cycle like pRB which controls progression from G1 to S phase [178], genes which encode either receptor or signal transducer for secreted hormones and encode signals to inhibit cell proliferation such as transforming growth factor (TGF)- β and adenomatous polyposis coli (APC) [179], genes coding for cell cycle checkpoint control proteins which normally function to arrest the cell cycle in response to unrepaired DNA damage or chromosomal abnormalities like BRCA1 and p16 [180], genes coding proteins which are responsible for regulation of cell death mechanism, such as apoptosis, like p53 [181] and genes coding proteins take apart DNA mismatch repairing pathway like DNA mismatch repair protein 2 (MSH2) [182].

TSGs have been proposed to fall into three subclasses based on the role of their protein product in tumorigenesis: gatekeeper, caretaker and landscaper [183]. The “gatekeeper” term encompasses direct inhibition of cell growth via suppressing proliferation, inducing apoptosis, or promoting differentiation [184].

The concept of the "gatekeeper" was first introduced to elucidate the function of the APC TSG, which consistently undergoes mutations early in colorectal tumorigenic pathway. Mutations of the APC are thought to be a rate-limiting step for tumour initiation because if APC is functional, mutations of other genes, like p53, are unable to induce tumorigenesis process. On the other hand, “caretaker” class of TSGs play indirect a role on inhibit growth by maintaining the accuracy of the DNA code through repair of DNA damage or prevention of genomic instability like chromosome instability. And many DNA repairs genes can be included in this class. Also, loss of function of caretaker TSGs promotes cancer development via increasing DNA mutation rate [184], [185]. TSG known as "landscaper" are expected to influence the tumour cell microenvironment, potentially through regulating extracellular matrix proteins, cell surface markers, adhesion proteins, or secreted growth/survival factors. If a 'landscaper' TSG loses its function, it could lead to abnormal functioning and facilitate adjacent epithelial cells' transformation into neoplastic cells [184] (Figure 1.20).

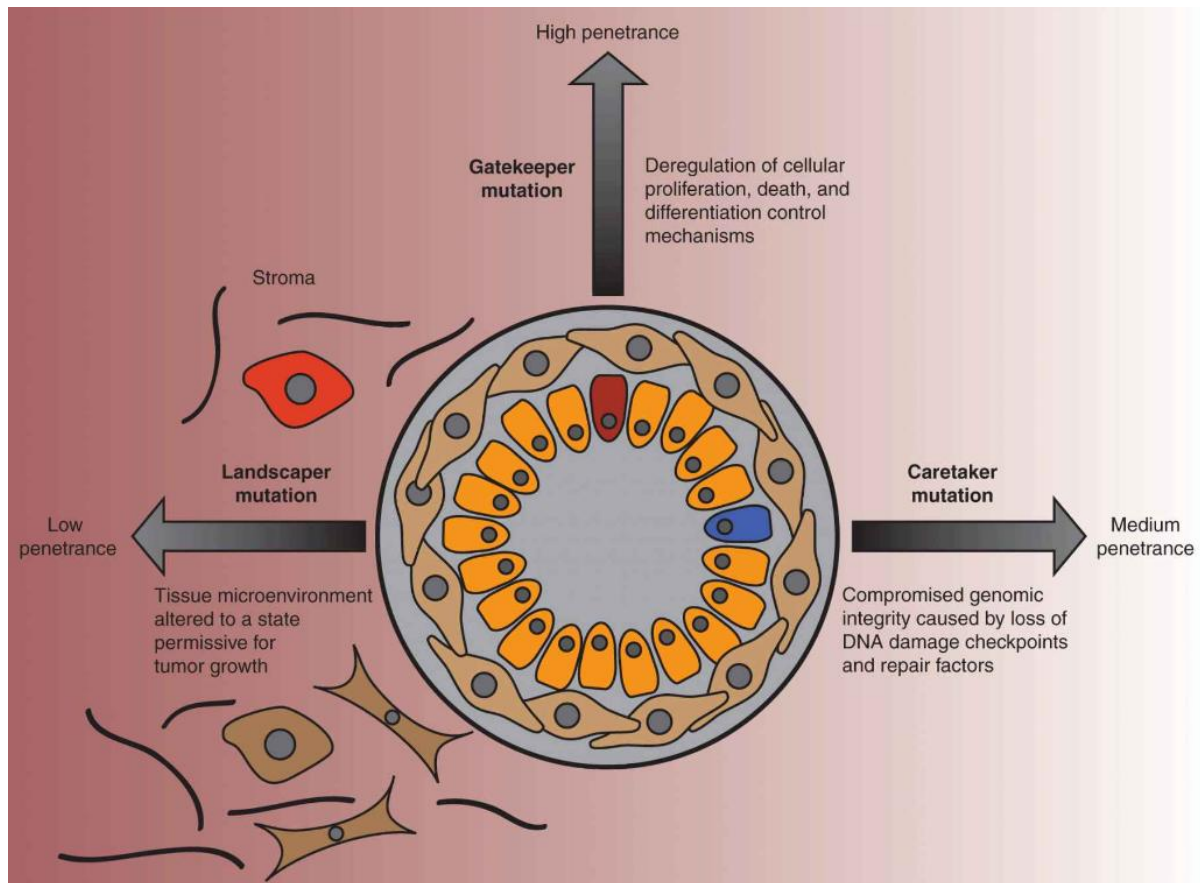


Figure 1.20 Gatekeeper, caretaker, and landscaper mutations.

Tumour-promoting genes have been categorized into gatekeepers, caretakers, and landscapers based on their roles. Gatekeeper genes produce proteins that control various cell activities such as growth, differentiation, or death. Mutations in these genes are strongly associated with the risk of cancer due to their direct influence on abnormal growth regulation. Caretaker genes encode proteins responsible for maintaining genome integrity and preventing the accumulation of mutations that could lead to oncogenic growth. Meanwhile, mutations occurring in the cells surrounding the tumour site can create a supportive microenvironment for tumorigenesis, and these associated genes are referred to as landscapers. As mutations in landscaper genes primarily play a supportive role without significant impact alone. Figure is obtained from [186].

Knudson first coined the term two-hit hypotheses as an explanation for how mutation of TSG occurred during cancer development. In this model, an inherited mutation in one copy of a TSG predispose of cancer, with a second, acquired mutation in the remaining copy occurring

during the individual's lifetime, leading to tumour initiation and progression based on study hereditary cancer Retinoblastoma (Rb) [187]. Rb is most common eye neoplasm in children and is strongly linked to mutations of the RB1 gene, which is localized on chromosome 13q14.2 led to these malignancies [188], [189]. Rb is often hereditary, around 40% of cases are linked to the inheritance of an autosomal dominant mutant of the Rb gene. During tumour development the second RB allele becomes sporadically inactivated, leading to a deficiency in the Rb protein (pRb) which impairs normal regulation of the of cell cycle and leads to uncontrolled cell proliferation [190] (Figure 1.21).

Multiple studies have conclusively demonstrated that tumours can persistently harbour mutations in one allele of a gene while the other allele is epigenetically silenced through hypermethylation, resulting in functional inactivation of the gene. Furthermore, in cases of familial cancer where one allele of a gene is mutated in the germline, the resulting tumour often retains both alleles, with hypermethylation serving as the second inactivating mechanism [191], [192], [193]. As mentioned in section 1.3.3., BRCA1 gene was previously believed to be significant solely for familial breast cancer. However, it is now evident that 10-15% of women with the non-familial version of this disease have tumours in which BRCA1 is hypermethylated [158].

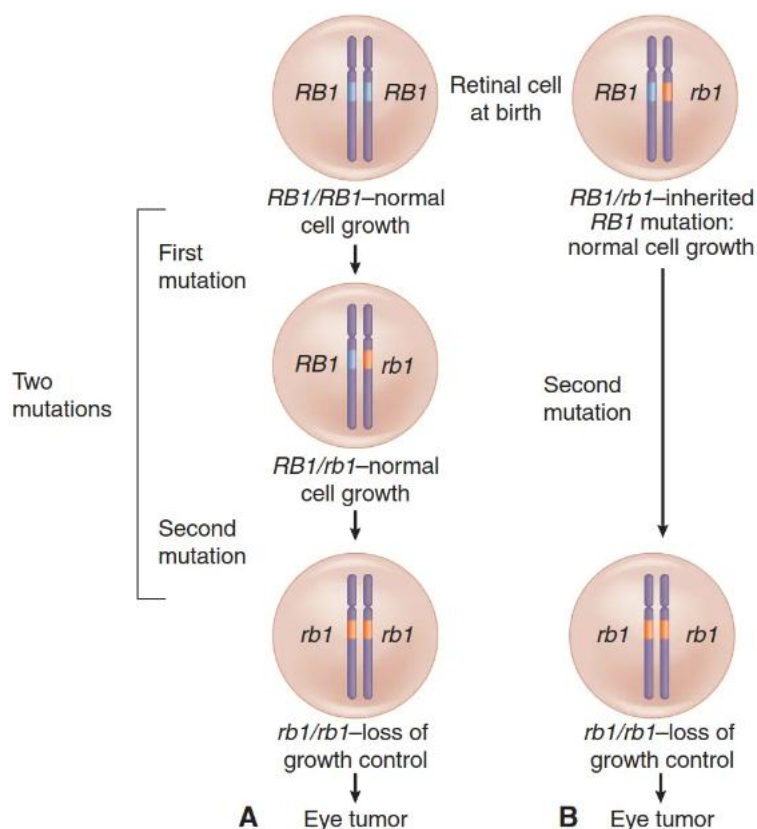


Figure 1.21 Knudson's two-hit hypothesis in Retinoblastoma both inherited and sporadic.

(A) Sporadic fraction of retinoblastoma cases; independent de novo mutations occur in the two copies of RB gene during the lifetime and cause Retinoblastoma development. (B) Hereditary retinoblastoma; individual born with germline mutation in one copy of RB gene following the second mutation (hit) result in Retinoblastoma development. RB1 and rb1 represents wild type and mutated allele respectively. Figure is obtained from [194].

1.4.1.2 Oncogenes

Proto-oncogenes in normal cells play crucial role in various cellular processes, serving as growth factor, transmitters of cellular signals and transcription factors [195]. In normal cell, expression of pro-oncogene is to allow the gene to respond various physiological signals. Depending on the metabolic requirements of cells, their expression can be at low levels or be elevated when required [196]. Proto-oncogene can become oncogene due to structural changes caused by point mutations result in substitution of single base by new one led to changing of amino acid in oncoprotein and elevate its activity such as point mutation at codon 12 in RAS oncogene, re-arrangements resulting in gene fusion or proximity to enhancer elements as an example MYC translocation in Burkitt's lymphoma or by gene amplification which incorporation of multiple copies of oncogene led to elevated production of oncoprotein [197], [198].

Activation of oncogene through these mechanisms provides growth advantages or enhanced survival of the cells which carry these alterations. Each mechanism results in either changes in the structure of the gene itself or deregulation of its expression [199]. Mutations and translocations happen at initiation of tumorigenesis or during tumour progression while gene amplification generally happen during progression [200]. Oncogene products can be classified into six main categories according to their roles: transcription factors, chromatin remodelers, growth factors, growth factor receptors, signal transducer and apoptosis regulators [201].

Vast majority of tumour cells rely on constant activation of oncogenic signalling for their survival which is concept know as oncogene addiction [202]. Identification of new therapies have taken advantage of inhibition oncogene addiction based on hypothesis that eliminate tumour cells with elevated oncogenic signalling whereas preserving normal tissue [203]. A

variety of new drug, small molecule, and monoclonal antibodies to directly impact oncogene product have been developed [201].

As an example of oncogenic addition targeting, Imatinib Mesylate is a targeted therapy commonly used for CML patients with the BCR-ABL oncogenic fusion protein and is selective inhibitor of ABL and its derivate tyrosine kinase protein [204]. Imatinib inhibits the BCR-ABL tyrosine kinase by binding to its ATP-binding site, preventing activation and downstream signalling that drives disease progression in CML [205]. While it is effective for most CML patients particularly in the chronic phase, significant proportion of patients in the advanced disease phase and a minority of chronic phase patients have refractory disease or develop resistance, leading to relapse over time [206]. While cancer treatments can initially result in tumour regression, the subsequent emergence of drug-resistant clones often leads to disease relapse. In addition to strategies targeting oncogenes can only be applied to a limited number of genes. A potential approach to overcome this limitation may be to shift the focus from only targeting oncogenes on which cancer cells depend, instead identify genes are not inherently oncogenic but tumour specifically depends on its activity [207].

1.4.1.3 Non-Oncogene Addiction

In the cell, there are buffering systems allowing cells to respond to changes in the environment. An example of these buffering system is functional redundancy, such as non-homologous genes operating in the same cellular process or in back-up pathways. Therefore, systems in the cell that maintain homeostasis ensure that important cellular processes do not depend on only single component that could be perturbed by mutation or environmental effects [208]. However, because of the disruptions to normal cellular function that occur in tumour development, the buffering capacity of cancer cells is often reduced [207].

The continued survival of cancer cells relies on numerous factors that set them apart from normal cells, not only oncogenic stimulants but also, they depend on mechanisms to cope with elevated stress levels. Even though the genes and pathways do not directly cause the development of cancer, their optimal functioning is crucial for cell survival, a phenomenon referred to as "non-oncogene addiction" [209]. For example, as most synthetic lethal partners of oncogenic driver genes are not contributing to tumorigenesis, but tumour cells might depend on constitutive function of that synthetic lethal partner gene, so synthetic

lethality can be regarded as a specific subtype of non-oncogene addiction [207]. Another example is cancer dependency genes that support tumour cells to overcome stress, which is augmented specifically in cancer cells as compared to normal cells, through by intrinsic and extrinsic mechanism. Thus, key benefit of targeting these genes and pathways is less toxic and more effective [210] (Figure 1.22).

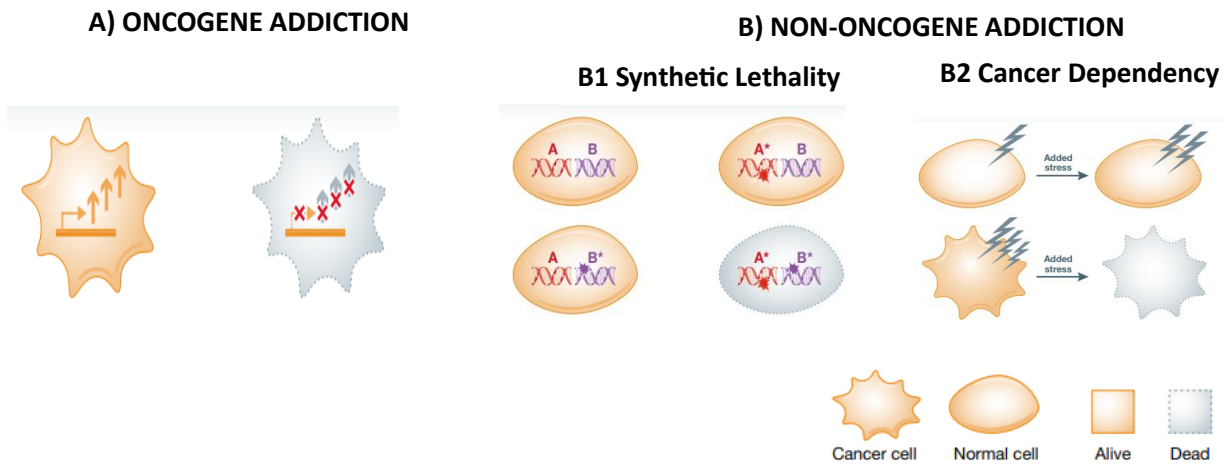


Figure 1.22 Schematic representation of cellular effects of oncogene and non-oncogene addiction: synthetic lethality and cancer dependency gene/pathway.

A) Oncogene addiction refers to the reliance of cancer cells on continuous oncogenic signalling for their survival, as depicted by the increased signalling pathways represented by the arrows. B1) In contrast, synthetic lethality occurs when the mutation of individual genes is compatible with cell viability, but the combined mutation of these genes leads to cell death. B2) Cancer Dependency Gene/Pathway describes how cancer cells harbour elevated levels of various stresses, stemming from collateral events during the tumorigenic process. These tumour cells can be specifically targeted and killed by the application of additional stress or the inhibition of specific salvage pathways, while normal cells can tolerate such perturbations. Figure is adapted from [207].

1.4.1.3.1 Synthetic Lethality

The proposed idea that non-oncogenic addiction can be used to for new therapeutic approach in cancer treatment based on phenomenon called `synthetic lethality`. Synthetic

lethality (SL) was first reported by Calvin Bridges, based on observation that when fruit flies were crossed, specific non-allelic genes caused lethality only when combined, even though the homozygous parents were perfectly viable [211]. The term "synthetic lethality" was later coined by Theodore Dobzhansky after observing a similar phenomenon in *Drosophila pseudoobscura* about two decades later [212]. Synthetic lethality refers to the situation where functional changes in two genes or their product results in cell death, whereas survival is maintained if either gene was changed on its own [213] (Figure 1.23).

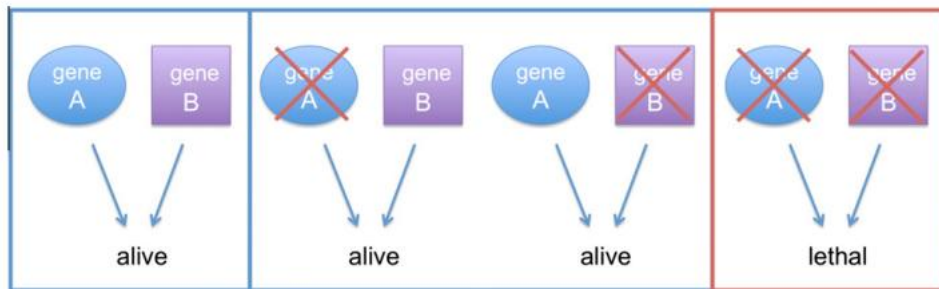


Figure 1.23 Schematic overview of synthetic lethality.

Two genes are synthetic lethal only when their simultaneous inactivation results in cellular or organismal death. In this example, deletion of either gene A or gene B does not affect viability whereas inactivation of both at the same time is lethal. Figure is produced from [208].

A synthetic lethal interaction in context of cancer is linked to a mutation specific to cancer, such that the mutant tumour cell's survival relies on the activity of synthetic lethal partner gene so it is referred as non-oncogene addiction [207]. The protein product of a gene that has a synthetic interaction with a commonly occurring tumour specific somatic mutation could serve as an effective target for anticancer drug therapy. Exploiting this type of interaction in therapy should lead to favourable treatment outcomes, where only tumour cells carrying the mutation would be impacted by the treatment. In addition, synthetic lethal interactions can broaden the range of targets for anticancer treatments by indirectly targeting non-druggable cancer mutations through identifying an associated synthetic lethal target that may be druggable [214].

The existence of various DNA damage repair mechanisms, such as homologous recombination (HR), Non-homologous end joining (NHEJ), and base excision repair, guarantees that normal cells can manage the genetic damage they experience. DNA repair genes, such as BRCA1, are frequently inactivated or mutated during tumour development.

Intriguingly, if cancerous cells no longer have one particular repair pathway functional, they appear to rely highly on the other remaining pathways for their continued existence [215]. This discovery has resulted in the development of specific treatments that target a synthetic lethal vulnerability in two distinct DNA damage repair pathways. The idea of synthetic lethality between Homologous Recombination (HR) and single strand break DNA repair is being explored for the treatment of breast and ovarian cancers with BRCA mutations [216]. Loss of function in BRCA1 or BRCA2 leads to impaired HR, leading to a critical dependence on NHEJ for the repair of double-stranded DNA breaks leading to an increase in unrepaired DNA and ultimately triggering cell death selectively in BRCA-deficient tumours [217].

1.4.1.3.2 Cancer Dependency Gene

Cancer dependency genes are not able to drive malignant transformation by themselves as they are not inherently oncogenic, but their overexpression is essential supporting oncogenic phenotype of cancer cells however, they are not required to same degree for the viability of normal cells [209]. Tumour cells frequently exhibit elevated levels of different types of stress. The process of tumorigenesis causes a reorganization of numerous cellular processes and results in heightened levels of cellular stress, including increased DNA damage and replication stress, metabolic stress, as well as proteotoxic and oxidative stress [210], [208]. Maintenance of mis activation of cellular stress in cancer cells mostly depend on overexpression of cancer dependency gene. In order to take advantage of this unique stress feature of cancer cells, two strategies have been suggested which would impact tumour cells while preserving normal cells. One of them is that inactivating stress-reducing processes is likely to critically elevate stress levels specifically in tumour cells and result in cancer cell death. The other potential approach would be to apply additional stress to selectively kill cancer cells because of their lower buffering capacity compared to normal cells [210]. The sensitivity of cancer cells to their augmented stress levels is suggested as promising for novel therapeutic target [209]. For example, intrinsic genomic instability in cancer cells triggers a chronic DNA damage response, leading to the production of altered, misfolded, or dosage-perturbed proteins. As a homeostatic mechanism to manage this 'stress phenotype', cancer cells overexpress chaperone proteins and enhance proteasome activity, to which they become 'addicted' [209], [218]. Therapeutically, proteasome inhibitors for

multiple myeloma (MM) is one of the examples of first-in-class agents developed in this regard [219].

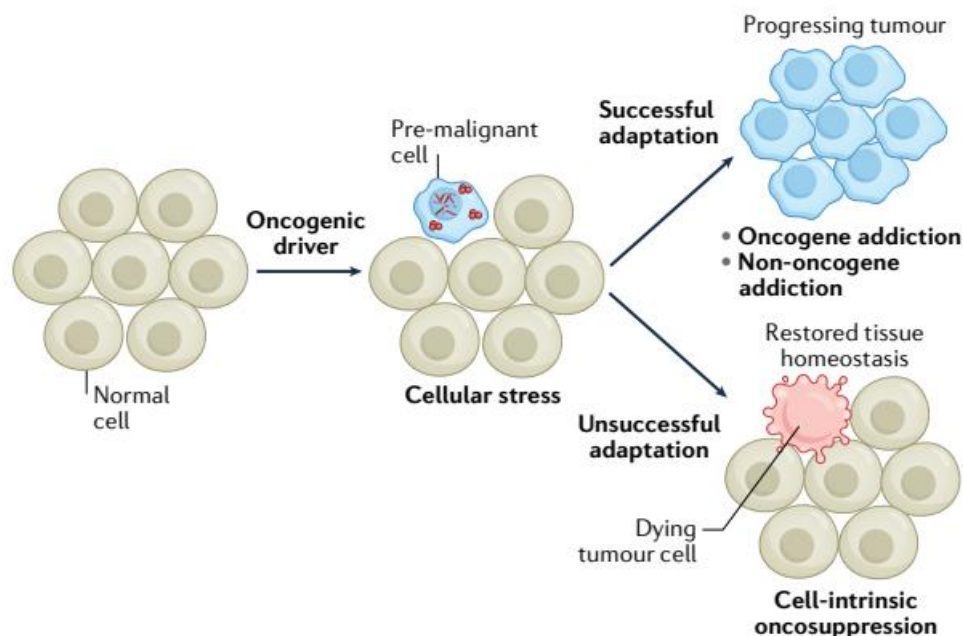


Figure 1.24 Illustration of oncogene and non-oncogene addiction through tumour development.

According to a commonly accepted theory, the process of malignant transformation is initiated by genetic or epigenetic events that give cells an advantage, typically involving increased proliferation or resistance to cell death, over their non-transformed counterparts. However, these changes often lead to heightened cellular stress due to elevated metabolic and proliferative rates. This results in the activation of protective stress-response pathways. Only malignant precursors that adapt to this cellular stress as imposed by oncogenesis are able to survive and, eventually developing into clinically detectable neoplasms as they acquire additional alterations enabling immunoevasion. Thus, tumour progression usually requires persistent oncogenic signalling (oncogene addiction) along with constant activation of non-oncogenic stress response pathways stimulated by driver events (non-oncogene addiction). Based on this conceptual framework, examples of both oncogene and non-oncogene addiction have provided numerous targets for the development of effective anticancer treatments. Figure is obtained from [220].

1.5 Novel Bioinformatic Approach to Identify SL / Disease Specific Dependency Gene(s)

Our group has recently developed a novel bioinformatic approach [221] which was originally designed to combine genome wide DNA methylation and expression data to allow the identification of synthetic lethal genes. The basic principles of how this was initially used to identify SL genes is outlined in Figure 1.19. Briefly, the early stage of cancer development includes widespread hypermethylation effecting thousands of genes. Crucially, the genes involved are highly reproducible within a specific cancer type. Clonal expansion, driven by an initiating genetic change will then result in a selective process which essentially “tests” the pool of hypermethylated genes for synthetic lethality. If any of the hypermethylated genes are synthetically lethal with a specific genetic initiating mutation, then the final cancer will be selected to have low methylation and retention of expression of this gene. In contrast, for all other genetic drivers (for which the gene is not synthetically lethal) there will be no selection against acquisition of methylation, which will result in cancers in which the gene is hypermethylated and not expressed. Thus, SL genes can be identified due to the specific failure to acquire methylation in one specific genetic subtype (Figure 1.25).

Although our pipeline was primarily designed to identify SL genes, we also hypothesised that this approach would allow us to identify subtype specific TSG. These would be loci that were not affected by the reproducible methylation. Changes in the disease as a whole and also did not become methylated in proliferating normal cells, but exhibited methylation, and loss of expression, only in one specific subtype of the cancer being studied.

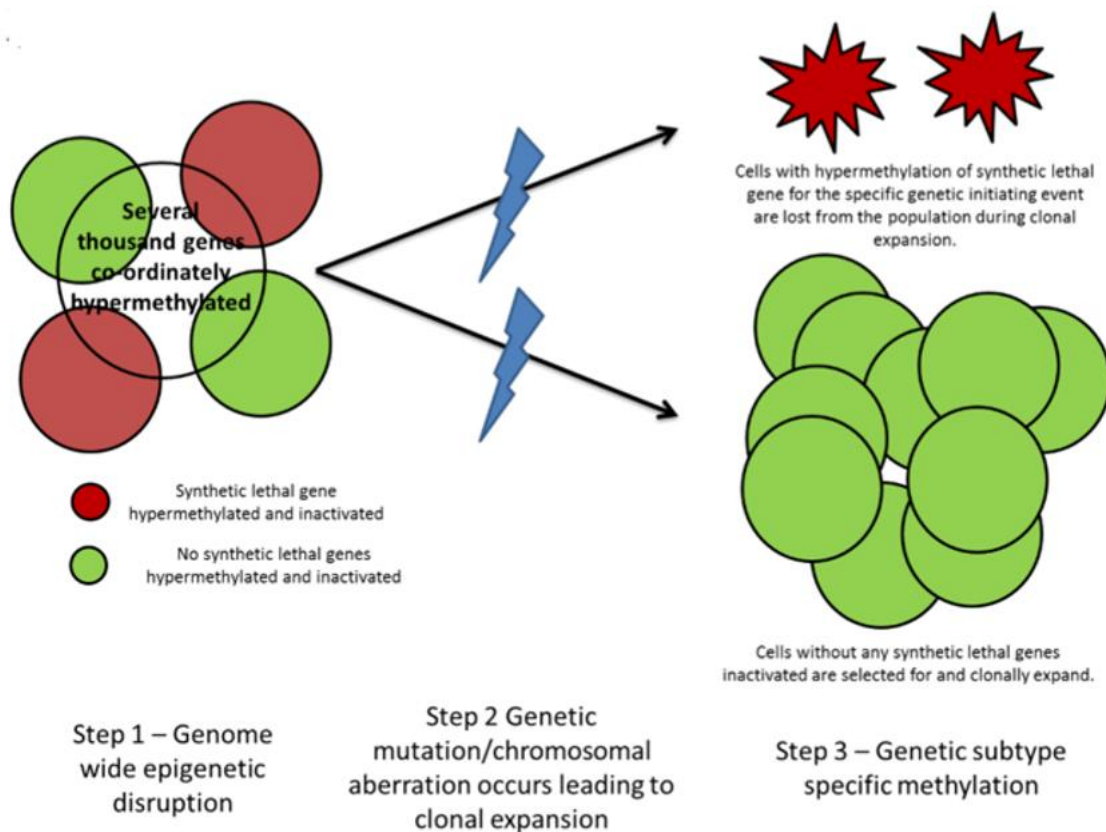


Figure 1.25 Outline of the Process That Leads to the Proposed Generation of SL Genes in ALL.

Step1- Methylation of thousands of promoters associated CpG islands occurs early in the tumour development. By chance, some cells (show in red) have hypermethylation of genes synthetically lethal with the initiating genetic change. Other cells (shows in green) lack this methylation. Step 2- Additional genetic change drives clonal expansion of epigenetically disrupted cells. Step 3- cells in which a SL gene for the specific initiating genetic events is methylated and inactivated will be lost during clonal expansion (red cells), while cells with no SL genes methylated/inactivated will clonally expand. Subsequently we can identify SL gene partners of the initiating genetic event for specific subtype of ALL based on their unusual methylation patterns (i.e., high methylation in all other subtypes, but very low methylation in the specific genetic subtype for which the gene is specifically lethal).Figure is obtained from [221].

1.6 Hypothesis and Aims of Project

1.6.1 Hypothesis

Initially the bioinformatics pipeline outlined above was developed to identify SL gene and tumour suppressor genes that were specific for a single genetic subtype in ALL. The

approach was subsequently expanded to other cancer types, as the patterns of methylation changes that underly the approach are essentially shared across all cancer types. This current project involves a further expansion of the approach to enable comparison between different cancers derived from the same normal cell type. This expansion will result in the identification of genes that are more accurately described as disease specific dependency genes (DSDG) as opposed to our original terminology of SL genes.

A limitation of the previous analysis is that SL genes can only be identified for specific subtypes within a cancer type and thus the identified targets would only be relevant for a subset of patients. This limitation could potentially be overcome if the approach could be expanded to comparing multiple cancer types and thus aim to identify novel targets that would be relevant to all patients with a particular cancer type. This may be possible when multiple cancer types derived from the same normal cell type, assuming the methylation changes across the genome would therefore be highly similar. To assess this, differential methylation in B-lymphocyte-derived cancers was compared at 7107 CpG sites previously identified as hypermethylated in all ALL subtypes [222]. All five types of B-cell malignancy, as well as normal memory B cells, showed similarly increased in methylation across this same set of CpG sites, even though they were initially identified specifically in ALL. Thus, as opposed to being disease specific, altered methylation at these sites was observed in all subsets of B-cells that had proliferated extensively (i.e. b-cell derived cancers and normal memory B-cells). In contrast, normal B-cell progenitor cells (which are in a similar differentiation state as ALL but will not have proliferated extensively) do not share the same methylation changes. Naïve B cells exhibit partial acquisition of the methylation pattern. This suggests these methylation changes shared in all B-lymphocyte-derived cancers are related to proliferation of B cells [221], [Xia M. unpublished]. It also confirms that similar methylation changes are seen across all tested B-cell malignancies (Figure 1.26).

Overall, this suggests that the high overlap in methylation patterns seen in different B-cell malignancies is similar to the high overlap between different subtypes of a single B-cell malignancy. Thus, the approach of identifying loci that specifically fail to acquire methylation could potentially be expanded to whole disease level. Candidate genes identified by this method would be DSDG as they would not be linked to a specific genetic change but would instead relate to a shared aspect of biology for the disease being assessed.

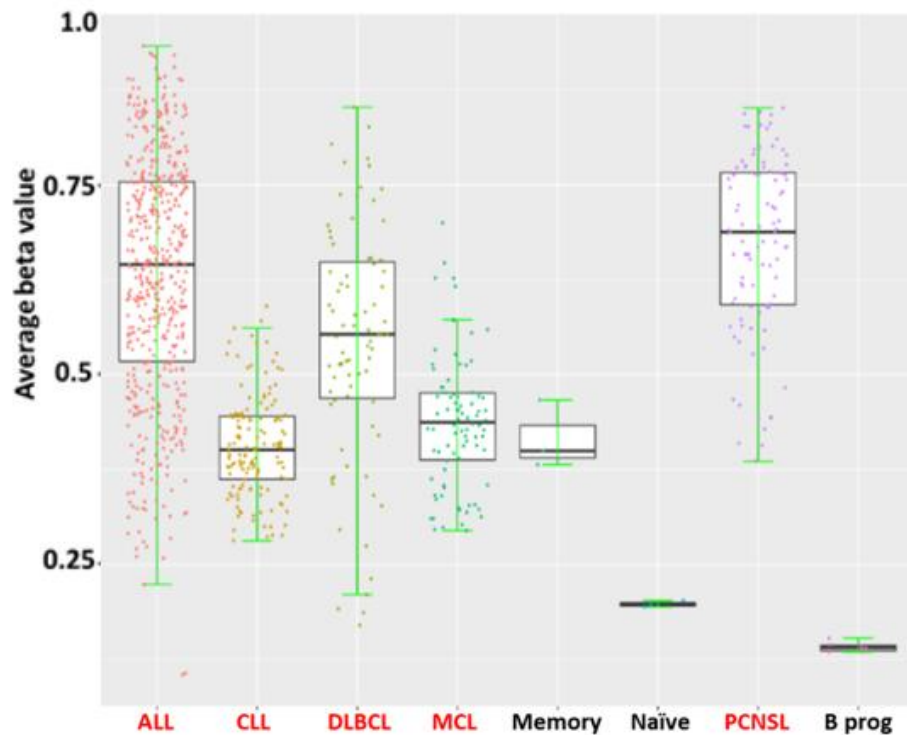


Figure 1.26 Average beta values of 7107 CpG sites across B-Cell Malignancies and normal B cells.

7107 CpG sites previously identified as hypermethylated in all ALL were assessed in across five B-cell malignancies as well as normal B-cell populations. Cell populations which have proliferated extensively, B-cell malignancies and memory B-cells, exhibited very clear hypermethylation. While cell populations which will not have previously proliferated extensively, B-cell progenitors and naïve B-cells, shows low methylation. Furthermore, the extent of methylation changes is also reflective of the extent of B-cell proliferation, with B-cell progenitor exhibiting the lowest methylation, followed by naïve B-cells and memory B-cell. Similarly, the more aggressive and proliferative B-cell malignancies such as ALL, DLBC and PCNSL shows higher methylation than the more indolent MCL and CLL. Figure is obtained from (Xia M. unpublished).

Hypothesis of project is that the similarity in methylation changes seen across all B-cell malignancies will enable us to expand our initial approach for the identification of SL and TS genes to identify DSD/TS genes at a disease wide level which would enable the identification of novel therapeutic targets that would be relevant for most or all individuals with a particular B-cell malignancy.

1.6.2 Aims of the Project

The primary aims of our project were:

- 1) To determine whether the bioinformatics pipeline designed originally for identifying subtype specific SL genes, could be used to identify DSD genes and/or TS genes relevant for all patients with a particular B-cell malignancy by using the individual B-cell malignancies as the “subtypes”.
- 2) Functional analysis of the TSG candidates identified by the approaches in aims 1 to determine if the candidate genes identified are functionally relevant in the expected disease/genetic subtypes.
- 3) Functional analysis of the DSDG candidates identified by the approaches in aims 1 to determine if the candidate genes identified are functionally relevant in the expected disease/genetic subtypes.

2 MATERIALS AND METHODS

2.1 List of Reagents

SDS Lysis buffer

- 2% SDS
- 50mM Tris-HCl (pH 6.8)
- 10mM EDTA
- 10% Glycerol

SDS loading buffer (5X)

- 250 mM Tris·HCl, pH 6.8,
- 10% SDS,
- 30% Glycerol,
- 10 mM DTT,
- 0.05% Bromophenol Blue

Running/Electrode Buffer (10X), pH 8.3

- 0.25 M Tris
- 1.92 M Glycine
- 1% SDS

Transfer buffer (10X)

- 25 mM Tris
- 192 mM glycine
- 20% methanol

TBS (10X)

- 250mM Tris
- 27mM potassium chloride
- 1.37M sodium chloride

TBS/Tween (TBST) (1X)

- 0.01M Tris-HCl pH 7.5
- 0.1M NaCl
- 0.1% Tween-20

Blocking solution

- 5% dried non-fat skimmed milk powder in TBST (X)

Laemmli Buffer (2X)

- 0.125M Tris
- 20% Glycerol
- 4% SDS
- 0.002% Bromophenol blue
- 10% β -mercaptoethanol

2.2 Cell Culture Methods

2.2.1 Cell Lines

The leukaemia and lymphoma cell lines used in this study are shown in table 2.1. In addition, the cell line of human embryonic kidney fibroblasts, 293T, which was used as a packaging cell line for production of lentiviral particle. All leukaemia and lymphoma cell lines were grown in Roswell Park Memorial Institute (RPMI) 1640 medium (Sigma, Dorset, UK) containing 10% foetal calf serum (FCS) and 1% penicillin/streptomycin in humidified incubators (Incu Safe, Sanyo, IL, USA) at 37°C with 5% CO₂. The 293T cell line was grown in DMEM (Dulbecco's Modified Eagle Medium, Sigma, Dorset, UK) containing 10% foetal calf serum (FCS) and 1% penicillin/streptomycin in humidified incubators (Incu Safe) at 37°C with 5% CO₂. Cell culture was carried out under sterile conditions in class II containment hoods. (Biomat2, Medair Technologies, MA, USA). Mycoplasma infection of cell lines was regularly checked (once every two months) for by using Mycoalert® Detection Kit (Lonza, Basel, Switzerland).

Cell Lines	Disease	*Cytogenetic Abnormalities	Expression of Fusion Gene
PreB697	BCP-ALL	t(1;19)(q23;p13.3)	TCF3-PBX1
REH	BCP-ALL	t(12;21)(p13.2;q22.1)	ETV6-RUNX1
RCH-ACV	BCP-ALL	t(1;19)(q23;p13.3)	TCF3-PBX1
SEM	BCP-ALL	t(4;11)(q21;q23.3)	KMT2A-AFF1
NALM-6	BCP-ALL	t(5;12) (q33.2;p13.2)	DUX4-rearrangement
TOLEDO	DLBCL	t(14;18)(q32;q21)	IGH rearrangement with BCL2
OCI-LY-18	DLBCL	t(14;18)(q32;q21)	IGH rearrangement with BCL2
DAUDI	Burkitt lymphoma	t(8;14)(q24;q32)	IGH rearrangement with MYC
JIYOYE	Burkitt lymphoma	t(8;14)(q24;q32)	IGH rearrangement with MYC

Table 2.1 Clinical features of the ALL and Lymphoma cell lines in the present study

*Main cytogenetic change used for diagnosis and or molecular subgrouping

2.2.2 Passaging and seeding of cells

Based on the number of cells required for an experiment, the cell lines were grown in different sized flasks: 25cm², 75cm² (Corning, Amsterdam, Netherlands). The cell lines were passaged every 2-3 days. To passage suspension cell lines the cells were collected in the 15ml falcon tube. They were centrifuged for three minutes at 125g, and the supernatant removed. The cell pellet was then re-suspended in fresh RPMI 1640 medium All cell lines were passaged three times a week, maintaining the cultures at the concentrations reported in Table 2.2.

To passage 293T cells, the cells were washed with phosphate-buffered saline (PBS) and then incubated with 1X trypsin (trypsin-EDTA (Sigma-Aldrich, USA) for approximately two minutes and the cells detached by tapping gently. To stop the action of trypsin, DMEM media containing 10% FCS was added with three times the added trypsin volume. The cell suspension was then added to a 15ml falcon tube and centrifuged for five minutes at 500g. Then supernatant was

removed, and the cells resuspended in fresh DMEM medium. The cell line was passaged three times a week, maintaining the cultures at the concentrations reported in Table 2.2.

For long term storage, cells were counted by haemocytometer to get 1 to 3 million cells. Following to determination of the cell density, cells were centrifuged at 400g for 5 minutes and supernatant was discarded. Cell pellets were resuspended in 1mL of freezing medium (FBS supplemented with 10% (v/v) dimethyl sulphoxide (DMSO, #D4540, Sigma Aldrich)) and transferred into cryovials (Invitrogen Life Technologies) prior to freezing at - 80°C using a Mr. Frosty (freezing container, Nalgene) for 24 hours. Frozen cryovials were subsequently transferred into liquid nitrogen for a long-term.

CELL LINE	MINIMUM DENSITY	MAXIMUM DENSITY
PREB697	0.5×10 ⁶ cells/ ml	2-3 x 10 ⁶ cells/ml
REH	0.5×10 ⁶ cells/ ml	3-5 x 10 ⁶ cells/ml
RCH-ACV	0.5×10 ⁶ cells/ ml	4 x 10 ⁶ cells/ml
SEM	0.2×10 ⁶ cells/ ml	3-4 x 10 ⁶ cells/ml
NALM6	0.5×10 ⁶ cells/ ml	4-6 x 10 ⁶ cells/ml
TOLEDO	0.3×10 ⁶ cells/ ml	3 x 10 ⁶ cells/ml
OCI-LY-18	0.5×10 ⁶ cells/ ml	3.5 x 10 ⁶ cells/ml
DAUDI	0.5×10 ⁶ cells/ ml	1.7 x 10 ⁶ cells/ml
JIYOYE	0.3×10 ⁶ cells/ ml	2 x 10 ⁶ cells/ml
293T	2 x10 ⁶ cells/ 75cm flask	20x10 ⁶ cells/ 75cm flask

Table 2.2 Culture densities of the cell lines used in this study

2.2.3 Counting cells to estimate cell densities

A Neubauer haemocytometer (Appleton Woods LTD, UK) was used to estimate the concentration of cell suspensions. 10µl of cell suspension was added on both grid sides of the haemocytometer. Cells were counted within two 5x5 grids (1 mm²) manually for both sides of the haemocytometer using a light microscope (Olympus, CKX53, Tokyo, Japan). The average number per grid was calculated and multiplied by 10⁴ to provide a measure of cells per ml. In case of calculation of alive cells, Trypan Blue 0.4% stain (Sigma-Aldrich, USA) was used as Trypan blue can enters only dead cells and stain their cytoplasm. At a 1:1 ratio, 10ul of cell suspension was mixed with 10ul trypan blue dye solution afterward counted by haemocytometer.

2.3 Transduction of cells with lentiviral expression constructs

2.3.1 Production of Expression Vector (pSIN-SIEW-)

2.3.1.1 Recombination of target sequence into Plasmid

In order to transfer of gene of interest from pDNOR (as entry clone) to pSIN-SIEW-GTW (expression vector), gateway technologies (Gateway™ LR Clonase™ II Enzyme Mix, Invitrogen) were used. The pSIN-SIEW-GTW vector is a modified form of pSIN-SIEW lentiviral vector (self-inactivated HIV derivated), which recombination sites, *AttR*, have been introduced to allow use of the Gateway cloning system and kindly provided by Dr Paul Sinclair, Newcastle University. pDNOR vector include target sequence of THEM4 was purchased from GeneCopoeia (GC-11844) and include target sequence of SLC22A15, TTC12 and MAP9 were purchased from Genescript (SC1200). For three of the candidates, SLC22A15, THEM4 and MAP9, there is only one known transcriptional variant. However, there are four variants for TTC12. Variant 1, which is the full-length variant, was used for this analysis. pDNOR has also sequence called AttL to be recognised by recombinase enzyme provided by kit. The sequence between the two (AttL) sites is removed and incorporated between the two destination sequences AttR in the empty pSIN-SIEW-GTW expression vector (the cloning steps based on gateway cloning system shortened in the diagram below). Briefly, destination vector at 150ng was added 2µl, entry vector at 100-300ng was added between 1-10µl and 4µl LR clonase enzyme mix and 4ul 5X LR clonase buffer was used and necessary amount of TE buffer was added to cover 20µl total volume. The LR Clonase enzyme mix was thawed on ice for two minutes and vortexed before adding to the mix. After that, the reaction was incubated at 25°C for one hour. Lastly, 2µl proteinase K solution was added to stop reaction and the mixture was incubated at 37°C for 10 minutes. pSIN-SIEW-THEM4, pSIN-SIEW-TTC12, pSIN-SIEW-MAP9 and pSIN-SIEW-SLC22A15 were generated.

AttL1-gene-AttL2	×	AttR1-ccdB-AttR2	→	AttP1-ccdB-AttP2	+	AttB1-gene-AttB2
(Donor vector)		(Empty expression vector)		(Donor vector)		(Expression vector)

2.3.1.2 Transformation of competent bacteria

Luria-Bertani (LB) (Invitrogen, Cat No:22700025) agar plates were prepared, including the ampicillin (amp) antibiotic as pSIN-SIEW vector structure has amp resistance site. To make LB agar plates, 3.2 g LB Agar (Invitrogen) was weighed and dissolved to give a final volume of

100ml distilled water and then autoclaved. Following autoclaving, the LB agar was cooled to 50°C and antibiotics (100 µg/ml ampicillin) was added. This was poured into 10cm plates and allowed to cool. Plates were stored at 4°C.

Library Efficiency® DH5α™ Competent Cells (Invitrogen, Cat No:18263012) were used for transformation. As per protocol, 100µl competent cells were thawed on the wet ice and added to chilled polypropylene tubes. DNA ligation reactions from recombination step were diluted 5-fold in TE buffer and 1µl added to the competent cells and incubated on ice for 30 minutes. Cells were the heat shocked in 42°C water bath for 45 seconds. The reactions were placed on ice for 2 minutes. Then, 0.9ml Super Optimal broth with Catabolite repression (S.O.C) medium was added to the reaction. The whole reaction was shaken at 225rpm (37°C) for 1 hour. In the final step, 200µl and 500µl of rection volumes were spread at two LB agar plates including the 100µg/ml ampicillin antibiotic and the plate was incubated overnight at 37°C.

2.3.1.3 Extraction of Plasmid DNA

Following transformation, single colonies were picked from agarose plates and used to inoculate 5ml of LB Broth Base (Invitrogen, Cat No:12780052) containing the appropriate antibiotic (100 µg/ml ampicillin) which was prepared and autoclaved previously. Cultures were then incubated overnight at 37°C in a heated shaker (225rpm). Plasmid DNA was isolated using a miniprep plasmid DNA isolation kit (QIAprep Spin Miniprep Kit, Qiagen Cat No: 27104). According to the manufacturer's instructions, the bacterial cells were harvested by centrifugation at >8000 rpm (6800 x g) in a conventional table-top microcentrifuge (Cat no: 5423 Eppendorf) for 3 min at room temperature. The harvested bacterial cells were resuspended in 250µl Resuspension buffer with RNase A by vortex until totally dissolved. 250µl Lysis buffer was added and then mixed by inverting the tube. After that 350µl precipitation buffer added and mixed by inverting tube. The suspension was centrifuged for 10 min at 13,000 rpm (~17,900 x g) in a table-top microcentrifuge. This by followed by transfer 800µl supernatant to QIAprep 2.0 Spin Column. After washing the spin column with 750µl PE Buffer, DNA was eluted by adding 50µl Buffer EB. DNA samples were quantified with a Nanodrop (Thermo Fisher Scientific Fisher) to determine the DNA concentration and the A260/A280 ratio. The DNA samples were stored at +4°C. The extraction method was used to extract plasmid DNA from expression vectors: the pSIN-THEM4 and pSIN-SLC22A15, pSIN-TTC12 and pSIN-MAP9

2.3.1.4. Restriction Endonuclease Digestion and Agarose Gel Electrophoresis

Restriction endonuclease digestion was used for confirming the presence of the expected insert in plasmids. EcoRI restriction endonuclease (New England BioLabs Cat No: R0101S) was used for confirming pSIN-THEM4, pSIN-TTC12 and pSIN-MAP9 and PstI restriction endonuclease (Thermo Fisher Scientific) Cat No: ER0611) for pSIN-SLC22A15. EcoRI digests vector at two sites and one sites on *THEM4* insert, one sites on *TTC12* insert and two sites on *MAP9* insert. PstI digests pSIN-SLC22A15 vector at three sites and also cuts four times within the expected insert. In addition, number of control vectors from previous version of vector different genes (*HOXA4* and *MSC*) cloned into were used to confirming whether both enzyme (EcoRI and PstI) digestion properly digest.

Each reaction total volume was 20 μ l, including 2 μ l 10X buffer, 1 μ l enzyme, 150-500ng of plasmid DNA and distilled water up to a final volume of 20 μ l. Digestion was conducted at 37°C for 1 hour. Meantime 0.8% agarose gel was prepared; 1.2g agarose was melted in 150ml TBE buffer. Following digestion, the 20 μ l digested products and 4 μ l Gel Red (Biotium) were mixed and loaded onto the agarose gel. 100bp DNA ladder (Cat No; G2101 Promega) and 1kb DNA ladder (Cat no:11581625, Thermo Fisher Scientific) was used as a marker. Electrophoresis was conducted at 150V approximately 45 minutes (BioRad Horizontal Electrophoresis and power supplies system) and the gel was viewed under the UV light (ChemiDoc BioRad Documentation System). To get higher concentration and total amount of plasmid DNA, bacterial colonies, containing the pSIN-SIEW-expression vector, on the agar plate were sent to MRC DNA sequencing and Services, Dundee University, for maxi prep plasmid isolation and sequencing to confirm insert matches the published coding sequence of the genes being studied. To allow long term storage of plasmids, 30% glycerol was added to the bacterial culture containing the plasmid, mixed and stored at -80°C.

2.3.2 Lentiviral production using the 293T packaging cell line

As a subfamily of retroviruses, lentiviruses are capable of delivering large amount of genetic material into host cells and incorporating it into cellular genome [223]. Lentiviral production is enabled through co-transfection of three plasmids to produce the necessary proteins to enable packaging of the expression vector. The lentiviral system used in this study belongs to the second generation of lentiviral vectors and is composed of three different

plasmids: the expression vector (pHR'- SIN-cPPT-SIEW), the packaging plasmid (pCMVΔ8.91) and a third vector (pMD2.G) expressing the envelope glycoprotein (VSV-G in this work) required to bind the receptor sites to the host cell (Figure 2.2). The first, the pCMVΔ8.91 vector encodes (ψ) packaging element which is essential for regulating the packaging process of the viral RNA genome inside the capsid (Figure 2.2-B). Also, codes necessary element for integration, reverse transcription and synthesis of capsid. The second plasmid is envelope plasmids called pMD2.G encodes envelope glycoproteins derived from vesicular stomatitis virus envelope glycoprotein (VSV-G) that enables virus cells entry to host cell (Figure 2.2-C). And last one is pSIN-SIEW which is an expression vector expressed gene of interest. pSIN-SIEW is the class of self-inactivating (SIN) vectors, in which a deletion within the 3' long terminal repeat disrupts the viral promoter, so that effectively prevent the virions from replicating themselves. Moreover, auxiliary viral genes, such as vpr, vpu, nef and vif, have been deleted from the viral genome in order to provide a degree of biosecurity [224]. The expression of the gene of interest is controlled by the spleen focus-forming virus long terminal repeat (SFFV) where is located to immediately downstream the gene of interest is the internal ribosome entry site (IRES), which allows for the concomitant transcription of the eGFP along with the gene of interest to monitor basically transduction efficiency on transduce and non-transduce mix cell populations. In addition, the vector contain additional viral elements like central polypurine tract (cPPT) which improves replication of viral genome and the woodchuck hepatitis virus post-transcriptional element (WPRE) which is increases level of the expression of transgene cloned into the lentiviral vector [225] (Figure 4A).

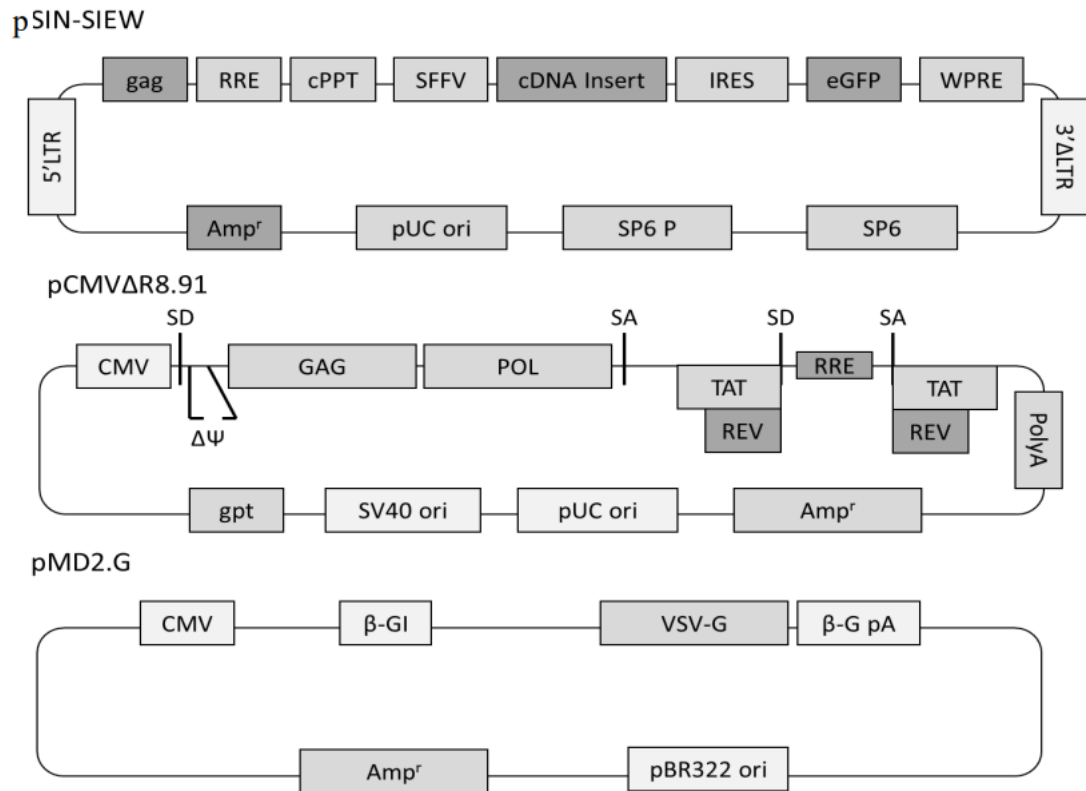


Figure 2.1. Schematic representation of vectors used for producing the lentiviral supernatant.

A)-pSIN-SIEW, B) packaging pCMVΔR8.91 vector and C) - pMD2.G vector. The following elements contribute to the structure of these vectors: GAG = group specific antigen, RRE =Rev responsive element, cPPT = central polypurine tract, SFFV=spleen focus forming virus, IRES= internal ribosomal entry site, eGFP= enhanced green fluorescent protein, WPRE=Woodchuck hepatitis virus post transcriptional regulatory element, SP6 and SP6P= transcription factor (RNA polymerase promoters), pUC ori=origin of replication for propagation in *E. coli*. AMP^r = confers resistance to the antibiotic ampicillin in *E. coli* cells. LTR= long terminal repeat, a control center for gene expression. CMV= Cytomegalovirus promoter, SD=splice donor. ΔΨ= packaging signal. Pol= DNA polymerase, SA= splice acceptor, TAT= trans-activator of transcription, REV= a trans-activating protein that is essential to the regulation of HIV1 protein expression, Poly A= poly-adenylation signal, SV40 ori= simian virus regulates viral genome replication and expression in the host cell upon infection, gpt= guaninephosphoribosyl transferase, regulating virus replication, gpt = guanine-phosphoribosyl transferase virus replication. BG-pA= beta globin polyadenylation signal, cloned in place of the 3'LTR, serves as transcriptional termination signal, VSV-G = pseudo-typed G protein of the vesicular stomatitis virus. B-GI = β-

globin intron replace the LRT for viral self-inactivating. PBR322 ori= pBR322 origin of replication in E.Coli. (Figure kindly provided by Dr Paul Sinclair, Newcastle University).

To allow production of lentiviral particles, EndoFectin™ Lenti Transfection Reagent (GeneCopoeia™) was used to transfect the packaging cell line (293T) with the lentiviral construct containing the gene of interest and the two packaging plasmids pCMVΔR8.91 and pMD2. The day prior to transfection, 293T cells were trypsinised and counted, after that 5×10^5 293T cells were seeded in a 10-cm dish to achieve 70-80% confluency at the time of transfection. DNA/Endofectin complex was prepared according to the protocol. 10µg expression vector (pSIN-THEM4, pSIN-TTC12, pSIN-MAP9, pSIN-SLC22A15), 2.5µg pCMVΔR8.91 and 7.5µg pMD2 were diluted in 200ul Opti-MEM® I (Gibco Cat no: 31985062) in sterile polypropylene tubes. 15µl Endofectin was diluted in 200ul Opti-MEM in a sterile polypropylene tube separately. Diluted EndoFectin Lenti reagent was added dropwise to the DNA solution while gently vortexing. The mixture was incubated for 20 minutes at room temperature to allow the DNA-EndoFectin complex to form. After the incubation, DNA-EndoFectin Lenti complex was added directly to plate by gently scrambling. The cells were incubated in a CO₂ incubator at 37°C overnight. 16 hours post-transfection the media was replaced with fresh DMEM medium supplemented with 2–5% heat-inactivated fetal bovine serum and penicillin/streptomycin and continue incubation in the CO₂ incubator at 37°C. Lentivirus-containing media was collected 48 hours after transfection. Virus-containing medium was transferred into a 15ml tube and centrifuged at 500×g for 10 minutes to avoid any cell debris. The medium was microfiltered using a 0.45 µm polyether sulfone filter; the supernatant was aliquoted (1000 µl/tube) and stored at -80°C.

In some cases, to get high number virus particles, six to ten transfections were carried out at the same time then after 48 hours incubation, viral preps were concentrated using the Lenti-X Concentrator (Takara Bio, Cat No: 631231), according to the manufacturer's protocol. Briefly, supernatant was collected into 50ml tube then centrifuged at 500Xg for 10 minutes. Clarified supernatant was transferred to a sterile container, with three volumes of supernatant combined with 1 volume of Lenti-X Concentrator. Following mixing by inverting, this was incubated at 4°C overnight. The next day, the sample was centrifuged at 1500g for 45 minutes at 4°C. Afterwards samples were aliquoted into microcentrifuge tubes (1ml per tube) and stored at -80.

2.3.3 Lentiviral Transduction of cell lines

2.5×10^5 of the target cells was plated per well in a 12-well plate at in 1ml fresh RPMI 1640 media, containing 5% calf foetal serum then 500 μ l lentiviral-containing media was added to each well in a total volume of 1.5ml. This was incubated at 37°C and 5% CO₂ for 16-20 hours. Untreated (parental) and mock-treated cells were used as negative controls. After incubation, at post-transduction day 1, cells were washed twice with PBS and 1.5ml fresh RPMI 1640 media added. This was then incubated in the standard growth conditions, as described above, so that infected cells could be analysed by flow cytometry at day 3-4 post infection to measure the fraction of transduced cells.

2.3.4 Assessing GFP expression by flow cytometry

Flow cytometry was used to assess the fraction of GFP positive cells following lentiviral transduction. All flow cytometry analysis was performed using the BD FACS Bioscience Canto II and BD Bioscience LSRFortessa X-20. Measurements were taken at 488/530/30 to assess GFP expression. Data was analysed using FCS Express 7 software. Cells were gated using FSC-A vs SSC-A graph to restrict subsequent analysis to viable cells and. GFP levels were typically determined at days 4, 8, 11/12, 21, 28/29 post transductions.

2.4 Cell Death Experiments

2.4.1 Assessment of Cell death by Annexin V and PI staining

Level of cell death was initially assessed by APC Annexin V and PI staining (BioLegend Cat No: 640932), consistent with manufacturer's instructions. This assay has been designed to distinguish apoptotic and necrotic cells based on annexin V and Propidium iodide (PI) staining. Annexin V binds to phosphatidylserine (PS). PS is located on intracellular layer of the healthy cells' plasma membrane, but it translocates on external layer of early apoptotic cells because of the losing membrane asymmetry on early apoptosis. Annexin V staining does not differentiate between early apoptotic and late apoptotic and/or dead (necrotic) cells. For this reason, PI was used along with annexin V. PI is cell- impermeant dye that binds DNA on late apoptotic or necrotic cells that no longer have intact membranes [226]. For the assay, cells (untreated, mock treated and treated) were washed twice with cold staining buffer (BioLegend

Cat No: 420201). Then 100,000 cells were resuspended in 100µl Annexin V binding buffer. 100µl of this cell suspension was transferred into polystyrene test tube, and 5µl of APC Annexin V and 10ul of Propidium Iodide solution were added. Following 15 minutes incubation at room temperature (25°C) in the dark, 400µl Annexin V binding buffer was added to each test tube and this was then analysed by flow cytometer on 640 670/30, 561 610/20 channel (Annexin V and PI respectively) along with GFP assessment on 488 530/30 channel. This allows three different subpopulations to be distinguished. Viable cells were negative for both dyes, annexin V positive but PI negative cells refer early apoptotic cells and both stains are positive on late apoptosis and/or dead cells. Flow cytometry was performed on the FACS Fortessa X-20 and FCS Express 7 software was used to analyse results.

2.4.2 Assessment of apoptosis by caspase activation

As a second confirmation of cell death that triggered by apoptosis induction, Caspase-Glo 3/7 Assay (Promega UK, Cat No:8091) system was used. Caspases are protease enzymes that play an important part in apoptosis. Among them, Caspase 3 and 7 have the role as executioner caspases in apoptotic pathway [227].

To assess effect of re-expression of *SLC22A15* gene on cell lines, caspase 3/7 activity was measured at post-transduction day 3, 4 and 5 according to the manufacturer protocols. Initially, cells were counted, and 10,000 cells seeded into 96 well plates in a 80µl volume. Caspase 3/7 buffer and substrate were equilibrated to room temperature, and mixed until the substrate had totally dissolved. When reagent was ready to use, 80µl reagent was added to each well to provide 1:1 ratio of Caspase-Glo® 3/7 reagent volume to sample volume. Blank wells, with only reagent and media without cells, were used as blank controls. Untreated (parental) and mock-treated cells were used as negative controls. All experiments were performed in triplicate. Plates were gently mixed using a plate shaker at 300–500rpm for 30 seconds then incubated at room temperature for 1 hour. Luminescent reading was taken using FLOUstar Omega (BMG LABTECH). The reading from blank wells were subtracted prior to calculation of final results.

2.4.3 Assessment of apoptosis indirectly by Caspase Inhibition Assay

The caspase inhibitor Ac-DEVD-CHO (Selleckchem Cat no: 7901) was used to assess the importance of caspase activity in *SLC22A15* driven cell death. The inhibitor was resuspended in Nuclease-free water (nf-water) to obtain a 20mM stock solution. This was further diluted in nf-water to prepare a 20 μ M working solution. Briefly, transduction was conducted as described above on Reh and PreB697 cell lines. On the day2 post-transduction, the working solution was prepared then added to each well which contained transduce cells with either pSIN-SLC22A15 or pSIN-EV. In addition, same amount of nf-water with inhibitor was added to each well both untreated parental cell and transduce cells (pSIN-SLC22A15 and pSIN- EV). Post-transduction at day3, 4 and 5 cells were washed with PBS prior measurement of caspase 3/7 and GFP experiments, to remove any remaining Ac-DEVD-CHO. Caspase 3/7 activation was measured with Caspase-Glo 3/7 Assay and GFP level was measured in the transduced population at post-transduction day3/4/5. The Caspase 3/7 Glo (Promega UK, Cat No:8091) kit was used according to the manufacturer's instructions (detailed in previous section). The fraction of GFP-positive cells was assessed by flow-cytometer Fortessa X-20 (BD Bioscience LSRFortessa X-20). This experiment was performed three times.

2.5 Assessment Of Cell Proliferation After Lentiviral Transduction

Staining with eBioscience™ Cell Proliferation Dye eFluor™ 450 (Invitrogen Cat No: 65-0842-85) was used to assess cell proliferation in transduced leukaemia cell lines. This fluorescent dye is able to cross cell membranes and bind covalently to amine groups on cellular proteins, thus becoming fixed within the cells. Following staining, with each cell division, the dye passes equally to each daughter cell leading to a halving of the fluorescent intensity with each round of cell division. This allows the extent of cell division to be followed for up to 7 generations. Staining was carried out according to the manufacturer's instructions., 10mM stock solution was prepared before the experiment. After lentiviral transduction on day3, 250,000 cells were washed two times with pre-warmed PBS, then resuspended in 500 μ l PBS. 20 μ M dye was prepared for each sample from the stock solution in 500 μ l volume for each sample to be stained. This was immediately applied to 500 μ l cell suspension and mixed by vortexing to get 1:1 ratio then incubated at 37°C for 10 minutes in the dark. After the incubation period, 4-5 volumes of cold complete media were added to each sample and

incubated on ice for five minutes to stop the labelling process. Afterward, cells were washed three times with complete media then cultured as normal. Samples were analysed approximately from 48 hours to 96 hours until day 17 post-transduction, after which fluorescent levels dropped too low to see detectable differences after additional rounds of cell division. FCS Express 7 software was used to analyse results.

2.6 RNA Extraction and cDNA Synthesis

To assess gene expression at the mRNA level, total RNA was extracted using a GeneJET RNA Purification Kit (Thermo Fisher Scientific, Cat No: K0731) according to the manufacturer's protocol. Briefly, cells were collected from samples and washed with PBS, then 600µl Lysis buffer was added to samples and vortexed quickly to mix, then 360µl ethanol (% 96-100) was added and mixed by pipetting. On the second step, cell lysate was added to a RNA binding column and was centrifugated for 1 min at $\geq 12000 \times g$ and this step repeated until the all-cell lysate was added to column. Then, the column was washed with wash-1 and wash-2 buffer, respectively. 50µl to 100µl water was added on the last step then centrifuged for 1 min at $\geq 12000 \times g$ to elute RNA. The extracted RNA was stored at -20°C (for use in the short term) or at -80°C if long term storage was required.

The purified RNA was quantified using the Nanodrop ND-1000 spectrophotometer (Nanodrop, Delaware, USA). About 2µg of RNA was used for cDNA synthesis using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, UK Cat No: 4368814) according to the manufacturer's protocol. 10µL of master mix and 10µL of RNA (total volume made up with RNase free water if concentration was higher than 200ng/µl) was added to a 0.2mL thin-walled PCR tubes for a total of 20µL reaction (Table 2.3). The mixture was then loaded onto a thermal cycler, using a program of 25°C (10 minutes), 37°C (120 minutes), 85°C (5 minutes) and lastly 4°C (5 minutes). Samples with no reverse transcriptase (RT) were also included in the reaction to act as a control in subsequent PCR reactions.

Component of Master Mix	Volume	
	Samples with RT	Samples without RT
10X RT Buffer	2µl	2µl
25X dNTP Mix (100 mM)	0.8µl	0.8µl
MultiScribe™ Reverse Transcriptase	1µl	-
Nuclease-free H ₂ O	4.2µl	5.2µl
10X RT Random Primers	2µl	2µl
Total per reaction	10µl	10µl

Table 2.3 Master mix component and their volume for cDNA reaction.

2.7 qRT-PCR

All qPCR reactions were carried out using Platinum SYBR Green qPCR SuperMix-UDG with ROX (Invitrogen, UK) following to the manufacturer's protocol. Total volume per reaction was the 10µl. This includes 5µl Syber Green Master mix, 2µl CDNA, 0.8µl primers mix (Eurofins Genomics, Luxembourg) and 2.2µl water to top up 10µl. Each sample was performed in triplicate in a 384- well plate (Thermo Fisher Scientific) and sealed with MicroAmp optical adhesive film (Applied Biosystem Cat No: 4311971). A water control, and no RT controls were included along with assay samples. The samples were run on a QuantStudio 7 Flex Real-Time PCR system and analysed using QuantStudio Real-Time PCR software v1.2 (both Applied Biosystems). *GAPDH* was used as reference. qRT-PCR conditions were shown in Table 2.4 and primer sequences used for running the qRT-PCR are listed in Table 2.5.

	Temperature	Time	Number of Cycle
UDG incubation	50 °C	2 minutes	1
Initial Denaturation	95 °C	10 minutes	1
Denaturation	95 °C	15 seconds	40
Anneling	60 °C	30 seconds	
Melt curve stage	95 °C	15 seconds	1
Melt curve stage	60 °C	1 minute	
Melt curve stage	95 °C	15 seconds	

Table 2.4qRT-PCR conditions

Gene	Primer Type	Sequence	*Average Ct Value in expressing leukaemia/lymphoma cell lines
THEM4	RT-FW	TCTGTGCCTGCCGCCAGTA	ND
	RT-RW	CTGGGGTTGGGGACAGAAC	
TTC12	RT-FW	GGTGTTTTTCCACAGCAGG	ND
	RT-RW	TATCACTAGCACACGCTGG	
MAP9	RT-FW	CTGATGGGTGTGAAGACATTG	28.5
	RT-RW	TGATCATCTGTGTCAAGGCTG	
SLC22A15	RT-FW	TCAGCATTTCTGTGCCTAGG	28.4
	RT-RW	ACAAGGAATGGCTGTTCAAC	
BMP2	RT-FW	GCAGCTTCCACCATGAAGA	30.4
	RT-RW	CTGAGGTGATAAACTCCTCC	
CTGF	RT-FW	GGAAGAGAACATTAAGAAGGGC	26.1
	RT-RW	CGTCGGTACATACTCCACAG	
NPY	RT-FW	CATCAACCTCATCACCAGG	26
	RT-RW	GTCTTCAAGCCGAGTTCTG	
GAPDH	RT-FW	TCAACGGATTTGGTCGTATTGG	18.7
	RT-RW	ATTTGCCATGGGTGGAATCATAT	
BMPR1B	RT-FW	CATGCCTTGTTGATAAAGGTTTCAG	34.2
	RT-RW	CCATCCTCTTTCTTGGTGCC	
ACVR1	RT-FW	GGAAGTGGCTCTGGTCTTCC	28.4
	RT-RW	ATACCTGCCTTTCCCGACAC	
ACVR2A	RT-FW	CCGAGGAAGACCAGGGAAC	28

	RT-RW	CTGATCTACCAAGTATAGCACTGA	
ACVR2B	RT-FW	CTGTGCGCCGGCTCTG	29.3
	RT-RW	CATCTAGCCAGCAGCCCTTC	
BMPR2	RT-FW	CGTTTCTGCTGTTGTAGCAC	25.6
	RT-RW	GAGACTGATGCCAAAGCAATG	
BMPR1A	RT-FW	GCCAACATCTTGGAGGAGTCG	27.2
	RT-RW	CCAGATTCTGTCCTTGAACACG	
NPYR2	RT-FW	AACTCCTAGAGGTGAACTGG	38
	RT-RW	TGGAGCAGTAGGCCAATATG	
NPYR5	RT-FW	ATCGGTAACAACCTGACCTGC	ND
	RT-RW	TCTCTGTGGCAAGTGTCTTG	

Table 2.5 Primers and their sequence used for qRT-PCR

*Average Ct value of primer set based on housekeeping gene (GAPDH). ND; not detectable in any assayed cell line.

2.8 Preparation Of Cell Pellets and Protein Extraction

After $1-3 \times 10^6$ cells were collected, they were initially centrifuged at 400xg 5 minutes after that supernatant was discarded and pellet washed and resuspended in 5ml PBS. Following the 400Xg 5 minutes centrifuge, supernatant discarded again then resuspended either with sodium dodecyl sulphate (SDS) Lysis Buffer or Radioimmunoprecipitation assay buffer (RIPA) (Sigma-Aldrich, USA Cat No: R0278) with proteinase inhibitor cocktail tablets (*Complete MINI protease inhibitor* (Roche, UK). For purification with SDS Lysis buffer, briefly, cell pellet was rinsed with PBS after harvested then resuspend in 40µl SDS Lysis buffer afterward lysate placed on heating block for 10min at 100°C. Before checking the concentration of the purified protein samples, they were sonicated on ice (at 5 mA 5 seconds repeated two times). For purification with RIPA buffer, one tablet of proteinase inhibitor was added to 10mL RIPA buffer. Following cell harvesting, samples were washed with 750µl PBS twice, then 60µl RIPA buffer (with protease inhibitors) solution was added to samples and incubated on ice 5 minutes. After sonication (as described above), samples were centrifuged at 8000xg for 10 minutes at 4°C. Protein concentrations were determined by Pierce® BCA Protein Assay Kit (Fisher Scientific UK, Leicestershire, UK) as described below. Protein samples and aliquots were stored at -20°C freezer if not used immediately.

2.9 Determination of Total Protein Concentration By Pierce® Bca Protein Assay

Estimation of total protein concentration was conducted with Pierce BCA protein Assay Kit (Thermo Fisher Scientific) based on manufacturer's protocol. The kit includes two Reagents referred to as A and B and bovine serum albumin (BSA). Initially, BSA protein was prepared as a standard in different concentration to generate a standard curve. Double distilled (ddH₂O) was used either dilute standards to get different concentrations or samples to get 1:10 dilutions. BCA working solution was made of Reagent A and Reagent B in a 50:1 ratio. 200µl of this was added for blank wells and 190µl for wells to which standards and samples would be added, in 96 well plates (Cat No: 3598, Corning CoStar, UK). Finally, 10µl aliquot of standards and samples were added to the 190µl of BCA reagent mix in quadruplicate. The 96-well plate was incubated at 37°C for 30 minutes. After incubation, absorbance at 560nm was taken with 96-well plate reader (FLUOstar Omega, BMG Labtech) and absorbance values were exported into Microsoft Excel (Microsoft Corporation, United States of America (USA)). Initially, the correlation (linear regression) between absorbance and concentration values was calculated using absorbance values of protein standards (with known concentrations). This was then utilised to determine the concentration of the protein samples.

2.10 SDS Page And Western Blot

The Laemmli method [228] was used for electrophoretic separation of proteins. 15 to 30µg of protein samples were treated with Laemmli buffer and heated at 100°C for 5 minutes to allow denaturation of proteins, with the exception of analysis of *SLC22A15* expression. No denaturation step was used for protein samples for *SLC22A15* protein expression because, due to its hydrophobic feature, *SLC22A15* protein forms large aggregates when heated and remains at the top of the gel following electrophoresis. Samples were instead incubated at room temperature for 5 minutes. Following the protein denaturation step, samples were loaded to Mini-Protean TGX precast gels (12% (456-8045), for *THEM4* and *BMP2*, 7.5% (4568023) for *SLC22A15*, Any kD (4569034) for NPY) (Bio-Rad UK) for separation according to their molecular weights. Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad, UK) system was used, and tank was filled with electrode buffer (ingredients in section 2.1). PageRuler Plus Prestained protein ladder (Thermo Fisher Scientific Cat No: 26619) was used to allow size estimation

following electrophoresis which was performed using a PowerPac Basic (Bio-Rad, UK) electrophoresis power supply. Electrophoresis was performed at 180V for 40 minutes. After electrophoresis the proteins were then transferred onto Polyvinyl difluoride transfer membranes (Immuno- Blot PDVF Membrane, Bio-Rad, UK). To activate PVDF membrane before the transfer step, it was soaked in 100% methanol for 1-2 minutes then inserted into mini-transblot cassette (Bio-Rad, UK) with sponges, filter paper, (Cat no: 1703932, Bio-Rad UK) and gel in a sandwich fashion (filter paper-gel-transfer membrane-filter paper). Finally, the sandwich cassette was placed into a transfer tank (Bio-Rad UK) then filled with transfer buffer. Transfer was performed using a PowerPac Basic (Bio-Rad Laboratories) power supply running at 100V for 30 minutes.

Following completion of protein transfer, the membrane was blocked in blocking solution which was made up from 5% dried skimmed milk with TBST (1 mM Tris-base, 5 mM NaCl, 0.01% Tween 20, TBST) for 1 hr at room temperature to prevent non-specific binding of the primary or secondary antibody. Membrane incubation with primary antibody solution (diluted in blocking solution) was performed overnight at 4°C on a roller (Table 2.6). Thereafter, membrane was washed three times with TBST solution on roller at room temperature for 10 minutes to avoid unbound antibody. Following the washing steps, the membrane was incubated with anti-mouse or anti-rabbit secondary antibody conjugated to the horseradish peroxidase (HRP) solution for one hour at room temperature on roller. Subsequently, three washing steps with TBST solution were performed. Afterward, ECL detection agent (Pierce ECL Western Blotting Substrate, Thermo Fisher Scientific Cat No: 32106) was prepared and incubated with the membrane for one to two minutes. Chemidoc Imaging system (Bio-Rad Laboratories) was used to capture chemiluminescent (corresponding to protein signal) and colorimetric (corresponding to visible protein ladder marking relative position) images of the membranes.

Membrane stripping was performed in case target proteins size were close to each other for different primary antibodies. Stripping buffer (Thermo Fisher Scientific Cat No: 21059) was added to the membrane for 15 minutes at room temperature on a roller. Following one washing with TBST, the membrane was incubated with blocking buffer for one hour at room temperature on a roller, then re-probed with another primary antibody.

To quantification SLC22A15 protein expression in leukaemia and lymphoma cell lines at day 8 post-transduction, images were captured using a GBox Chemi XL1.4 device and

GeneSys software (Syngene, UK) was used for protein quantification. Protein bands were quantified relative to the total signal captured from the reference band, which was assigned a value of 1.00.

Antibody	Host	Type	Supplier	Cat No	*Dilution
THEM4	Rabbit	Polyclonal	Novus Bio	NBP2-94437	1:1000
SLC22A15	Rabbit	Polyclonal	Proteintech	20626-1-AP	1:1000
NPY	Rabbit	Monoclonal	Cell Signaling	11976	1:500
BMP2	Mouse	Monoclonal	Rndsystem	MAB3551	1:500
B-actin	Mouse	Monoclonal	Abcam	ab49900	1:10000
Secondary Anti Mouse	Goat	Polyclonal	Dako	P0447	1:5000
Secondary Anti Rabbit	Goat	Polyclonal	Abcam	Ab6721	1:5000

Table 2.6 List of antibodies used in this project

* Dilution used for western blotting

2.11 siRNA Mediated Transient Knockdown Of Target Genes In Leukaemia And Lymphoma Cell Lines

SMARTpool *CTGF* siRNA, Individual *BMP2* siRNA were purchased from Dharmacon (L-012633-01-0010 and J-011219-06 respectively) and AlexaFluor488-labelled non-Targeting siRNA were purchased from Qiagen (AllStar Neg. siRNA, Qiagen, Cat No: 1027292 (Table2.7). And FuGENE siRNA reagent (FuGENE SI transfection reagent, Cat No:SI-1000) was used for transfection. Stock solutions of targeted and non-targeted siRNAs were initially resuspended at 20µM using nuclease free water. According to the manufacturer`s protocol, briefly, cells was counted then 50,000 cells was resuspended in 450µl of RPMI 1640 media (without penicillin and streptomycin) on 12 well-plate (Corning, Costar). Afterward, transfection reagent was prepared with mix 22µl pre-warmed Opti-MEM (Gibco, Cat No: 31985062), 3µl FuGene SI transfection reagent and 25µl siRNA (20µM stock) to make a final siRNA concentration of 100µM per well. The complex mixture was then incubated for 15 minutes to allow sufficient time for complex formation. Subsequently after, this reagent:siRNA complex mixture was added to the cells and incubated at 37°C during different time point depending on further experiments such as cell viability or apoptosis assay. Level of knockdown at the RNA level was

assessed using qRT-PCR at post-transfection day2 (detailed described at section 2.7). Each of transfection and following experiments were repeated three times on separately tests. (detailed siRNA method described in section 5).

siRNA target gene	siRNA target sequence	Company and Catolog No
BMP2- Individual seq-6	CCAGGUUGGUGAAUCAGAA	Dharmacon, J-011219-6
CTGF smart pool	ACAAUGACAUCUUUGAAUC	Dharmacon, L-012633-01-0010
	AGGAAGAUGUACGGAGACA	
	CGAUUAGACUGGACAGCUU	
	GAGAGACAUUAACUCAUUA	
Non-Targeting siRNA		Qiagen, 1027292

Table 2.7 List of siRNA target gene and sequence used in this project

2.11.1 Checking transfection efficiency by Flow-cytometer

Transfection was carried out by using non-targeting siRNA conjugated with AlexaFlour488 on different leukaemia and lymphoma cell lines. 24 hours after transfection, cells were collected then subsequently washed with PBS twice to remove any fluorophore which couldn't pass the cell membrane so that found in media efficiency. Efficiency of transfection was determined by FACS (Bioscience LSRFortessa X-20). Measurements were taken at 488/530/30 and data was analysed using FCS Express 7 software. Gating was applied to FSC-A vs SSC-A graph to determine percentage of viable cells.

2.11.2. Cell Viability assessment by XTT

Cell viability was conducted using XTT assay. This assay based on determination of live cells to investigate rate of cell proliferation and cytotoxicity after treatment. The assay contains N-methyl dibenzopyrazine methyl sulphate (PMS) and 2,3-Bis-(2-Methoxy-4-nitro-5-sulphophenyl)-2H-tetrazolium-5-carboxanilide, disodium salt (XTT) (Biotium, USA, Cat No:30007). XTT is technique that can assess the number of viable cells based on changing mitochondrial enzymes activity in live and dead cells. Viable cells have capability reduce XTT while dead cells do not. This colorimetric assay measures the metabolic activity of cells by

detecting the reduction of a yellow tetrazolium compound (3,4-tetrazolium]-bis benzene sulfonic acid hydrate or XTT) to an orange formazan dye by metabolically active cells [229]. Therefore, reduction of XTT in the solution results in a colour change, which can then be measured using a plate reader. Firstly, 3mg PMS was resuspended in 1ml of PBS and 4mg XTT resuspended in 4ml RPMI medium then working solution was prepared according to the manufacturer's protocol. Samples incubated at 37°C until day three and six post transfection of BMP2 and CTGF, pre-warmed 2.5uL activation reagent (PMS) was mixed with pre-warmed 1ml XTT solution in the dark to make working solution. After that, 50µl of this working solution was immediately added to each well. Plates were immediately covered with tinfoil, as XTT is light-sensitive, and incubated for 3-4 hours at 37°C with 5% CO₂. Post incubation, the plates were placed on plate-reader Fluostar-Omega (BMG Labtech) then shaken at 300-500 revolutions per minute (rpm) for 30 seconds to mix the dye in the solution before absorbance at 475nm wavelength was measured. The results were used to assess the effect of siRNA transfection on cell viability and readings from cells transfected with non-target siRNA was used as reference.

2.11.3 Assessment of caspase activation to assess apoptosis

To assess effect of knockdown of target gene on samples, Caspase 3/7 (Promega UK, Cat No:8091) was carried out at post transfection day 2 according to the manufacturer's instruction. Initially, cells were counted, and 10,000 cells seeded into 96 well white flat bottom plates (Cat No;3598, Corning CoStar, UK) in a 50µl volume at day2 after transfection. Caspase 3/7 buffer and substrate were equilibrated to room temperature, and mixed until the substrate had totally dissolved. When reagent was ready to use, 50µl reagent was added to each well to provide 1:1 ratio of Caspase-Glo® 3/7 reagent volume to sample volume. Blank wells, with only reagent and media without cells, were used as blank controls. Untreated (parental) and mock-treated cells were used as negative controls. All experiments were performed in triplicate. Plates were gently mixed using a plate shaker at 300–500rpm for 30 seconds then incubated at room temperature for 1 hour. Luminescent reading was taken using FLOUstar Omega (BMG LABTECH). The reading from blank wells were subtracted prior to calculation of final results.

2.12 Treatment Of Cell Lines with Cytotoxic Agent

To investigate role of BMP pathway on ALL, LDN-193189 (ab278073) was used as selective inhibitor of BMP Type1 receptors to treat cell lines at different working concentration from 0 μ M to 0.1 μ M, 0.25 μ M, 0.5 μ M and 1 μ M. Nf-water was used for dilution to make stock solution on 10mM LDN-193189. Further nf-water was used to prepare working solution immediately prior to use. Experimental design briefly, cells were counted, and 10.000 cells seeded in a 100 μ l volume into 96 well plate and each sample was triplicated thereafter seeded wells was treated with required amount of inhibitor at day 0 and were incubated at 37°C with 5% CO₂ until 48 hours and 96 hours. To assess effect of inhibitor on cell lines at different time point (48 hours and 96 hours), two different plates were prepared as explained above at day 0. To test cell viability, XTT was performed 48 and 96 hours following the end of the treatment. (XTT experiment is detailed in section 2.11.2)

2.13 ELISA Assays To Assess Protein Secretion

To assess extracellular *BMP2*, *NPY* and *CTGF* protein expression enzyme-linked immunosorbent assays (ELISA) were performed, according to manufacturer instructions.

BMP2 concentration in the cell culture supernatant was checked by Human *BMP2* ELISA Kit Abcam (ab119581) according to the manufacturer's protocol. Assay range was 31.2 pg/ml - 2000 pg/ml and is designed to quantitative measurement of Human BMP2 in bone tissue, cell culture supernatants and serum. Experimental design briefly, all cells were seeded at a density 1×10^6 cells in 2ml RPMI growing media per well in a six-well plate in and incubated at 37°C with 5% CO₂ during 2 days. On the experiment day, standards were freshly prepared according to the serial dilution principle by reconstitution of 10ng/ml stock standard with sample diluent buffer (from 1000pg/ml as highest, standard 1 to lowest concentration, standard 6 (31.2pg/l)). Meantime samples were centrifuged at 400X g at 5 minutes then supernatant collected at 1.5 ml Eppendorf tube. If required, sample diluent buffer was used to dilute samples. Samples and standards were equilibrated to room temperature prior to the experiment then 100 μ l added to 96 well plates which were precoated with BMP2 specific mouse monoclonal antibody. This was then incubated for 90 minutes at 37°C, RPMI media

alone sample served as blanks and standard diluent buffer was used as zero (control) well. Meantime, biotinylated anti-human *BMP2* antibody and Avidin-Biotin-Peroxidase Complex was diluted at 1:100 ration with antibody diluent buffer and ABC diluent buffer respectively to make it 1X concentration and working washing buffer (1X) was diluted at 1:25 ratio with distilled water. Following to the incubation period, the supernatant in each well was discard and 100µl 1X biotinylated detection monoclonal mouse antibody was added. This was incubated at 37°C for 60 minutes. Thereafter, supernatant was removed from each well then 300µl 1X washing buffer added and following the 1-minute incubation, washing buffer removed from wells and this step was repeated three times. Subsequently 100µl 1X Avidin-Biotin-Peroxidase Complex was added to each well then incubated at 37°C for 30 minutes. Afterward, unbound conjugates were washed away with 1X washing buffer this process was repeated five times and buffer stayed in 1-2 minutes each time. 90µl TMB (3,3',5,5'-tetramethylbenzidine) solution was used to visualize the HRP enzymatic reaction in which TMB was catalysed by HRP to produce a blue colour. This was incubated 37°C for 15-25 minutes in dark. Incubation time was determined according to the shades of blue colour in standard wells from highest to lowest concentration of standards. Then, 100µl of TMB Stop Solution was added into each well, the colour changed into yellow immediately. The density of yellow coloration is directly proportional to the Human *BMP2* protein concentration of the sample captured on the plate. Absorbance at 450 nm was measured at plate-reader Fluostar-Omega (BMG Labtech) within 15 minutes. Levels of protein were determined by subtracting the value from medium alone. Each assay was performed in two independent preparations. And each sample of independent experiments was duplicated.

Human NPY ELISA Kit Abbexa (abx052317) was used to determine NPY concentration in the cell culture supernatant, according to the manufacturer`s protocol. Assay range was 4.94 pg/ml - 400 pg/ml and designed to assess human NPY concentrations in serum, plasma, tissue homogenates, cell lysates, cell culture supernatants and other biological fluids. Briefly, all cells were seeded at a density 1×10^6 cells in 2ml RPMI growing media per well in a six-well plate in and incubated at 37°C with 5% CO₂ for 2 days. On the experiment day, standards were freshly prepared according to the serial dilution principle by sample diluent buffer (from 400pg/mL as highest, standard 1 to lowest concentration, standard 5 (4.94pg/ml)). Meantime samples were centrifuged at 1000X g at 20 minutes then supernatant collected at 1.5ml Eppendorf tube. If required, sample diluent buffer was used to dilute samples. Samples and

standards were equilibrated to room temperature prior to the experiment then 50µl added to 96 well plates which were precoated with NPY specific antibody. Afterward 50µl detection reagent A which was freshly prepared from 100X concentration with diluent A to 1X was added to each well then concentration incubated for 90 minutes at 37°C, RPMI media alone samples served as blanks and standard diluent buffer was used as zero (control) well. Meantime, detection reagent B was diluted at 1:100 ratio with diluent B and to make it 1X concentration and working washing buffer (1X) was diluted at 1:30 ratio with distilled water. Following the incubation time, the supernatant was removed from each well then, the plate was washed 3 times with X wash buffer then 100µl X concentration detection reagent B was aliquoted into each well then incubated 37°C for 30 minutes. Right after the incubation, supernatant was removed from each well then each well used was washed 5 times with X wash buffer. Then 90µl TMB substrate was added into each well, and incubated 37°C for 10-20 minutes in the dark. After incubation, 50µl of TMB Stop Solution was added into each well, the colour changed into yellow immediately. Absorbance at 450 nm was measured at plate-reader Fluostar-Omega (BMG Labtech) within 15 minutes. Levels of protein were determined by subtracting the value from medium alone. Each assay was performed in two independent preparations. And each sample of independent experiments was duplicated. The assay is based on competitive reaction between biotin-labelled NPY and unlabelled-NPY so there is an inverse correlation between colour development (or OD measurement) and sample concentration.

To determine extracellular CTGF protein level, Human CTGF Mini ABTS ELISA Development Kit (PeproTech, Cat No: 900-M317) and ABTS ELISA Buffer Kit (PeproTech, Cat No: 900-K00) were used by manufacturer's instructions. Assay range was 32 pg/mL - 4000 pg/ml and designed to assess human CTGF concentrations. Briefly, all cells were seeded at a density 1×10^6 cells in 2ml RPMI growing media per well in a six-well plate in and incubated at 37°C with 5% CO₂ during 2 days. Capture ab was diluted by 1:100 ration with PBS A 96-well plate was coated with capture antibody (rabbit anti-human CTGF) one day before the experiment by incubating overnight at room temperature. Next day, following the discard capture ab from each well and washing 4 times with X wash buffer, 300µl blocking buffer was added to each well and the plate was incubated for 1 hour at room temperature. During the incubation samples and standards from 4000pg/mL as highest to standard 1, 31.5pg/mL were prepared. After incubation period, each well used on the plate was washed 4 times with X

washing buffer and then 100µl sample and standards, media alone for blank and just diluent as zero (control) were aliquoted into each well. The plate was incubated at room temperature for 2 hours. Each well used on the plate was washed 4 times with X washing buffer then 100µl X concentration detection ab (biotinylated rabbit anti-human CTGF) was added to each well and incubated at room temperature for 2 hours. Afterward, the plate was washed 4 times with X washing buffer then 100µl X Avidin-HRP conjugate was added to into each well and incubated at room temperature for 30 minutes. Following the 4 times washing step, with X washing buffer, 100ul ABTS substrate solution was pipetted to each well and the plate was then incubated at room temperature for colour development. Colour development was measured at absorbance 405nm, with wavelength correction set at 650nm by plate-reader Fluostar-Omega (BMG Labtech). The density of yellow coloration represented directly proportional to the Human CTGF protein amount of sample captured in plate. Levels of protein were determined by subtracting the value from medium alone. Each assay was performed in two independent preparations. And each sample of independent experiments was duplicated.

2.14 CTGF Inhibitor Antibody Experiment

To investigate role of CTGF pathway on ALL, Pamrevlumab (FG-3019) (Selleckchem, Cat No; A2042) was used as selective inhibitor antibody against to CTGF to treat cell lines at different working amount from 0µg to 100µg/ml and 30µg/ml. Nf-water was used for dilution to make stock solution on 1mg. Further nf-water was used to prepare working solution immediately prior to use. Experimental design briefly, cells were counted, and 10.000 cells seeded in a 100µl volume into 96 well plate and each sample were triplicated thereafter seeded wells was treated with required amount of inhibitor at day 0 and were incubated at 37°C with 5% CO₂ until 48 hours and 96 hours. To assess effect of inhibitor on cell lines at different time point (48 hours and 96 hours), two different plates were prepared as explained above at day 0. To test cell viability, XTT was performed 48 and 96 hours following the end of the treatment. (XTT experiment is detailed in section 2.11.2)

2.15 Bioinformatic Analyses

2.15.1 DNA methylation Datasets used

All methylation data used in this study was derived from Infinium HumanMethylation 450K Beadchip array, which is cover 99% of RefSeq genes, 96% of CpG island from UCSC

database. The 450K chip includes two assay design. In design I, at the unmethylated CpG sites, the U-shaped bead with the tail A is matched with unmethylated CpG sites to produce a signal, however, the M-shaped bead with the tail G cannot be recognized so that there is no signal. On the other hand, at the methylated CpG sites, M-shaped beads produce the signal, but the U-shaped type does not. In design II, the bead will be added to a base instead of being divided into U-shaped or M-shaped. Furthermore, using fluorescent labels, the type of base can be determined to determine the methylation state (Figure 2.2). The quantitative measurement of DNA methylation (beta value [β]) was calculated as $\beta = M/(M+U+\alpha)$, the ratio of fluorescent signals from the methylated alleles to the sum of the signals from the methylated and methylation free alleles and $\alpha = 100$ (recommended by Illumina). Beta values ranged from 0, for completely unmethylated alleles, to 1 for completely methylated alleles [230]. Therefore, it allows us to estimate percentage of methylation differences between CpG sites under study group in comparison with all others like 0.4 β -value differences reflect 40% differences in methylation.

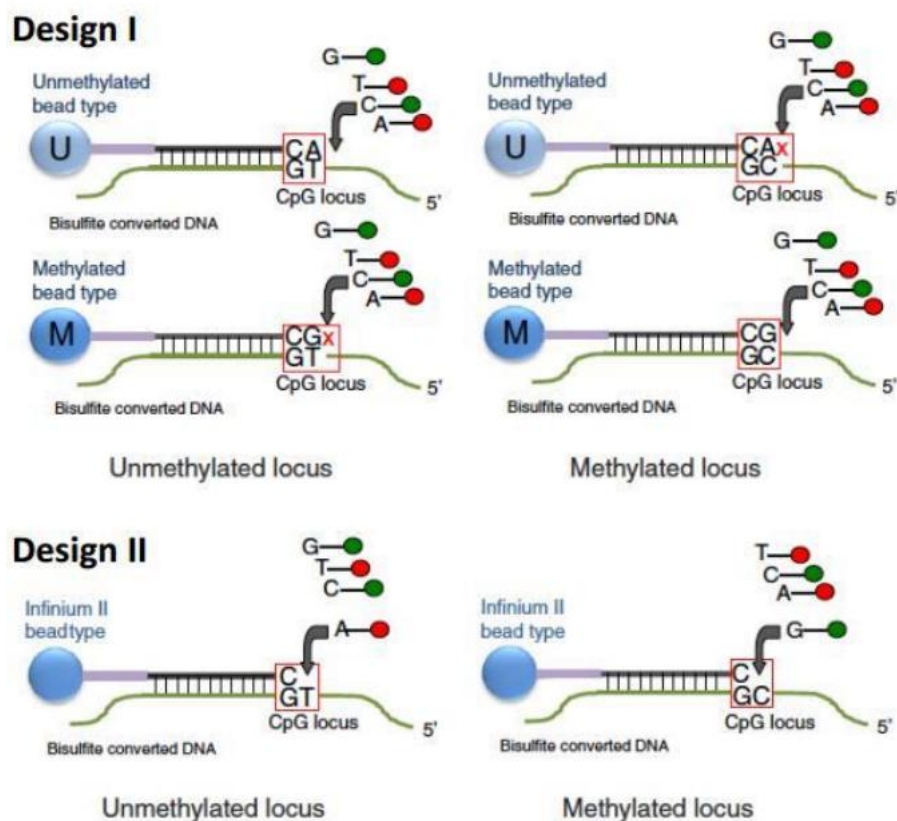


Figure 2.2. Overview of the Human Methylation 450K BeadChip arrays.

Design I: Type I probes which have two bead types (U and M), methylated site (C) and other for the unmethylated site (T) of the bisulphite modified DNA. Design II- Type II probes which have only one bead type corresponds to each target CpG site. Figure is obtained from [230].

All of methylation data used were publicly available from Gene Expression Omnibus (GEO), and The Cancer Genome Atlas (TCGA), European Genome-phenome Archive (EGA) (Table 2.8). All DNA methylation data used in this study from the Infinium HumanMethylation 450K Beadchip array datasets. The coverage on the array includes CpG sites mapping to different genomic regions such as promoter regions (TSS200 and TSS1500 representing 200bp and 1500bp upstream of transcriptional start site respectively), 5' UTR, 1st exon, gene body, 3' UTR, CpG island, Shore (N and S) 2kb flanking CpG islands and Shelf (N and S) 2kb flanking the shores.

The DNA methylation data has 8 data sets, including five types of B-lymphocyte-derived cancers, naïve B cells, memory B cells and B-cell progenitor cells According to the degree of differentiation stages, progenitor B- cell samples were divided into four different groups based on the level of CD34 protein expression and expression of the B cell marker CD19. S1 population represent predominantly Multipotent progenitors (MPP) but also containing common lymphoid progenitors (CLPs) and stem cells. S2 population represents pre-B-I cells, S3 population represent pre-B-II cells that expressed sIgM-CD19+ and S4 populations represent immature B cells that express sIgM+CD19+ [231]. Details of all the methylation data sets used in the study are listed in Table 2.8.

Type of Data	Source	No of Samples
ALL	GSE49031 [222] and GSE69229 [232]	517
CLL	EGAD00010000254 [162] and in-house dataset	187
MCL	EGAD00010001012 [233]	86
DLBCL	GSE37362 [234] and TCGA-DLBC [235]	79
PCNSL	GSE92676 [236]	95
Progenitor cell	GSE45459 [231]	22
Memory CS	EGAD00010000254 [231]	3
Memory NCS	EGAD00010000254 [231]	3
Naïve B cells	EGAD00010000254 [231]	3
Total		995

Table 2.8 Methylation data set used in study.

GSE - Gene Expression Omnibus accession numbers, EGA – European Genome-phenome Archive.

2.15.2 Cleaning and processing methylation data

The majority of methylation data used in this study were obtained as raw unprocessed file in “IDAT” format. These files represent the direct output from the methylation array as two-coloured signal intensities (red – unmethylated and green – methylated) of each probe. However, the study also includes processed files such as B-cell Progenitor cell (GSE45459) or MCL (EGAD00010001012), obtained either as beta values (β) representing the methylation level of each CpG site or methylation and unmethylation score corresponding to signal intensities. All methylation data used in this study were already processed by a previous student Lalchungnunga, normalisation and processing were performed using the R package “Minfi” [237] to filter out poorly performing probes that may have confounding effects on downstream analyses. Briefly, array background fluorescent signal bias correction was performed by ‘noob (normaleponential out-of-band)’ [238], probes that had detection p-value > 0.01 in 50% of samples were removed, cross-reactive probes [239] were also removed, as were probes from X and Y chromosomes and, probes with proximal SNPs with a minor allele frequency of 5% or greater and maximum distance (from CpG to SNP) of 2 bases. Methylation

score (β value) for the remaining probes was extracted which range from 0 to 1. Batch effect correction for the source of the materials was performed by using the removeBatchEffect function of the limma package v.3.52.1 [240].

2.15.3 Gene expression datasets

All of the samples used in this study came from a publicly accessible data source. The majority of the samples were obtained as raw unprocessed files in “CEL” format that contains probe intensity data corresponding to the level of gene expression on an Affymetrix GeneChip. The study also includes processed files such as for CLL (EGAD00010000252) containing the level of expression for all the genes present in the array. All datasets used in this study were already processed by previous student Lalchungnunga. Details of all the expression data sets used in the study are listed in Table 2.9.

Type of Data	Source	No of Samples
ALL	GSE12159 [241]	326
CLL	EGAD00010000252 [242]	143
MCL	GSE93291 [243]	122
DLBCL	GSE11318 [244]	203
PCNSL	GSE34771 [245]	34
Progenitor cell	GSE4560 [231]	31
Memory CS	GSE24759 [246]	5
Memory NCS	GSE24759 [246]	5
Naïve B-cell	GSE24759 [246]	5
Total		879

Table 2.9 Gene expression data set used in study.

GSE - Gene Expression Omnibus accession numbers, EGA – European Genome-phenome Archive.

2.15.4 Integration DNA methylation and gene expression data

In order to subsequent analysis, DNA methylation and expression data need to be integrated via R. After obtaining methylation data, they were normalised and processed by using R package called “Minfi” by previous student Lalchungnunga. Briefly, methylation B-value matrixes were calculated based on Infinium HumanMethylation450 Bead Chip then these data can be merged directly according to the name the CpG sites probe and poorly performed probes were separated. On the second stage, even though gene expression matrix data were obtained from Affymetrix Human Genome Array, these datasets correspond to different version of the array, including U219 Array, U133 Plus 2.0 Array, U95 Version 2 Array, and U133 Array. In these different versions of the array, the same gene corresponds to different probes with variable numbers. In addition, CLL data, in which its probe name is represented by Ensemble ID, is different from other datasets. Based on these, in the integration process, initially the probe name is converted to the corresponding gene name in their respective versions, and these probes which cannot correspond to the gene name are filtered. Then the average value of the probe with the same gene name is retained. Finally, the genes common these five types of cancers were kept by intersecting these datasets.

2.15.5 Identification of DMR and Cancer Specific DSD/TS genes candidates

Bioinformatic analyses to identify cancer specific DSDG/TSG in present study were undertaken using R v3.4.0 (<https://www.r-project.org/foundation>). This method was used initially to identify SL partner gene for specific genetic subtype of ALL by comparison of five common subtype of ALL [221]. The first step before identification of DMR, “cpg.annotate” package in R were used to analyse differences and annotate of single methylated site. Subsequently, these single methylated site results were used to identify cancer specific DMR candidates by DMRcate package in R [247].

To sum up bioinformatic pipeline analysis, the first step is identifying differentially methylated regions (DMRs) between the group under study (ETV6-RUNX1, in the example shown in Figure 2.3 and all other groups combined. Initially DMRs were selected by comparing average β -value for the group under study versus all other groups combined using DMRcate. The region is selected if the β -value difference is greater than 0.2 (equivalent to approximately

20% methylation difference), with a p-value of <0.0001 (step 1 in Figure 2.3). To restrict the analysis to regions more likely to impact gene expression DMRs that were located further than 20kb from any transcriptional start site are removed. As the identified DMRs can vary significantly in size, we then identify the region of biggest change within each DMR (which must include 2 or more adjacent CpG sites) (step 2 in Figure 2.3). This is done by identifying the single CpG site with the biggest change in beta value and then assessing the neighbouring upstream and downstream CpG sites to find the pair with the highest average beta value difference. This region is then potentially expanded both up and/or downstream but only if adding in the next neighbouring CpG site does not result in a reduction in the average beta value change of the region. The region of biggest change is then compared between the group under study and all other groups individually and only DMRs in which the region of biggest change is specific for the group under study (i.e. are different between this group and every other group individually) are retained. The cut-off value at this step can be varied, although to increase stringency a b-value difference of 0.4 (40% difference in methylation) with p value of less than 0.0001 is generally used (number 3 in Figure 2.3). In the last step, final candidate genes were identified by integrating with gene expression data (step 4 in Figure 2.3). Genes associated with DMRs that had met all above criteria (and had reduced methylation levels) were selected as SL candidate genes for the relevant group if the gene showed higher expression in the group of interest compared with all other groups individually. Similarly, genes associated with DMRs that had met all above criteria (and had increased methylation levels) were selected as TS candidate genes for the relevant group if the gene showed lower expression in the group of interest compared with all other groups individually. No absolute differences in the extent of differential expression were required other than being higher or lower, respectively, than all other groups. This approach was taken to account for the extensive differences in expression levels between genes such that for some genes even relatively small differences in gene expression levels may be biologically significant.

In addition to above explanation about how candidate genes generated by our bioinformatic analysis, it is worth to add that methylation differences of gene(s) between disease of interest versus all other B-cell disease understudy was looked. Genes shows high or low methylation are not take account of for analysis. Because of the methylation status of each gene is same scale from zero to one in terms of B value, we could have cut-off value when comparing disease of interest versus others like 0.4 for identifying ALL specific candidates.

However, having cut-off is not applicable for absolute expression values of gene(s). Because even small absolute expression differences between genes might have strong impact on gene function. Absolute expression differences when comparing between disease of interest and other B-cell disease is just taking forward if SLG candidate shows higher expression in absolute terms in disease of interest than all other B-cell diseases or shows lower / absent expression in just relevant disease than all other B-cell diseases.

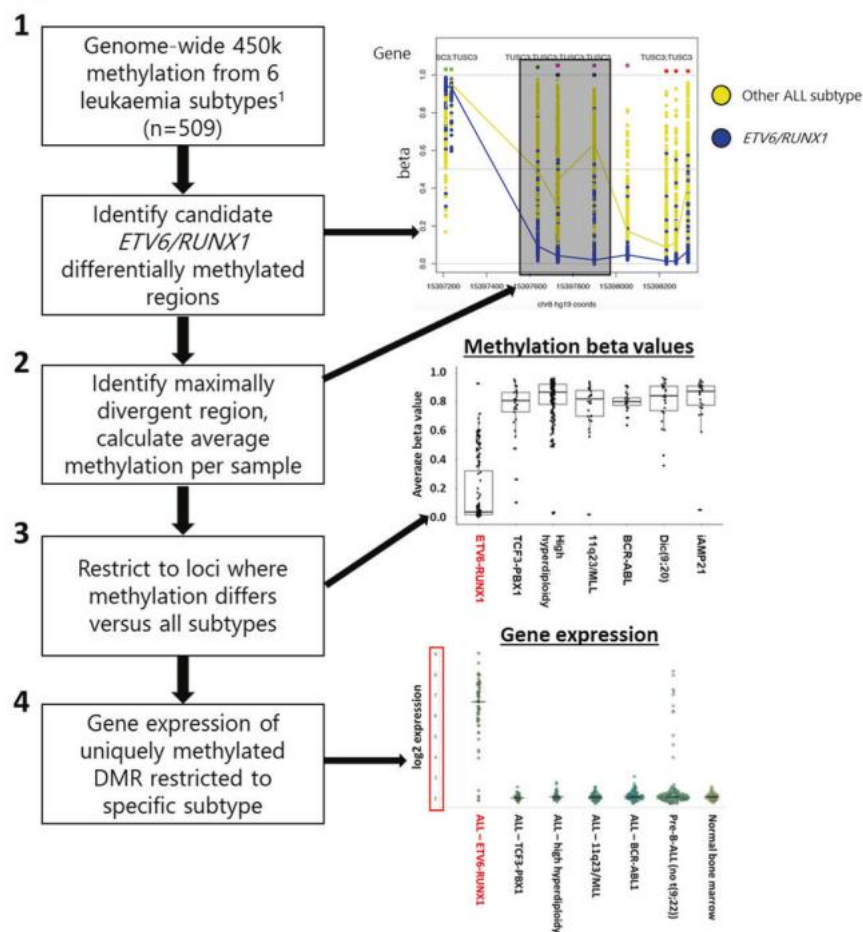


Figure 2.3. Flow-chart of Bioinformatic Pipeline for Identification SL Genes.

(1) All DMRs for a specific subtype of cancer under study (in this example *ETV6/RUNX1*) are initially selected using DMRcate. (2) The initial region is then further analysed to identify the maximally divergent region. (3) This region is then tested in comparison with all other B-cell malignancies under study and only regions that are divergent from all other types of cancer are retained as markers of potential DSD genes. (4) Gene expression data sets are used to analyse the expression of the nearest gene from loci derived from step 3. Genes in which

expression is exclusive to the cancer of interest are taken forward as DSDG candidates. Figure is obtained from [221].

2.16 Statistical Analysis

Student t-test was performed on Excel 2021 (Microsoft Corporation, USA). If not otherwise stated, statistical analysis was conducted utilising a two-tail Student's t-test. The correlation between fraction of lost of GFP and mRNA transcript level or protein level of SLC22A15 after transduction described in Section 4.4.5 and 4.5.6, was determined by linear regression using Excel 2010 (Microsoft Corporation, USA).

*3 RESULTS OF IDENTIFICATION OF TSG AND
DSDG CANDIDATES VIA NOVEL
BIOINFORMATIC APPROACH IN B-CELL
MALIGNANCIES*

3.1 Introduction

Epigenetic alterations in ALL can be divided into changes that are common across all subtypes likely due to extensive proliferation and those specific to each subtype based on biological differences induced by diverse genetic initiating events. Specific patterns of DNA methylation and gene expression might be chosen for subtype specific SL like genes, referred to as subtype specific vulnerability genes (SSV) partner gene related to the particular genetic change causing the cancer [221]. Identifying cancer SSV genes will be beneficial for novel therapeutic strategies that design for specifically kill cancer cells on the other hand, decreased toxicity in normal cells. These vulnerability genes are only required for the growth and survival of cells when partner gene is mutated or disrupted. If this other mutation is cancer-inducing, targeting the vulnerability gene should enable the selective killing of cancer cells while sparing normal tissues from toxicity.

Our group has developed novel bioinformatic approach to identify SSV gene partner in specific genetic subtype of cancer by using genome level DNA methylation and gene expression data. Briefly, among the numerous affected genes in the methylation wave dependent on B-cell proliferation, those essential for cell growth and survival due to a specific genetic initiating mutation will be strongly selected to maintain low levels of methylation (if high methylation negatively affects gene expression). This will lead to decreased methylation in that particular genetic subtype, while there will be increased methylation in all other subtypes. This differs from genes necessary for the overall survival of progenitor B-cells, which would show low methylation across all ALL subtypes and normal cells.

It is important to note that DNA methylation acts in cis and effect only adjacent gene so this method exploiting identified difference in methylation which can only affect the expression of the associated gene, enabling identification of the specific target involved in this selective process. In addition to, genome-wide methylation analysis is combined with genome-wide expression data to narrow down the analysis to sites where the changed DNA methylation is linked to a corresponding change in gene expression. In addition to favourable expression of the target gene in the specific subtype, SSV genes should show minimal or absent expression in all other subtypes.

Initially the bioinformatics pipeline outlined section 2.15.4 was developed to identify SSV gene and tumour suppressor genes that were specific for a single genetic subtype in ALL. The approach was subsequently expanded to other cancer types, as the patterns of methylation changes that underly the approach are essentially shared across all cancer types. According to Schwalbe et al, methylation change in ALL was highly predictive of changes in other B-cell malignancies and was highly predictive of mirroring methylation change in healthy memory B cells and similar methylation changes are seen across all tested B-cell malignancies [221]. High overlap in methylation patterns seen in different B-cell malignancies is similar to the high overlap between different subtypes of a single B-cell malignancy [221]. Thus, the approach of identifying loci that specifically fail to acquire methylation could potentially be expanded to whole the whole disease level. Candidate genes identified by this method would be DSDG as they would not be linked to a specific genetic change but would instead relate to a shared aspect of biology for the disease being assessed.

3.2 Aims of the Chapter

- To determine whether the previously designed bioinformatic approach for identifying subtype specific functionally relevant genes could be repurposed to identify DSD gene(s) relevant for all patients with particular B-cell malignancies by using the individual B-cell malignancies as the “subtypes”.
- To identify TS and DSD genes candidates via bioinformatic pipeline that could function as novel therapeutic target across five different types of B-cell malignancies.

3.3 Results

Bioinformatic analyses to identify DSDG for specific B-cell malignancies tested was performed using same pipeline and criteria with initial bioinformatic analysis (mentioned in section Figure 3.1). Same bioinformatic pipeline was repeat to identify candidates for each of five malignancies.

All methylation data used in the previous and current study, are derived from Illumina beadchip arrays (Illumina 450k or Methylation EPIC arrays). The data used in the analysis was primarily publicly available from Gene Expression Omnibus (GEO), and The Cancer Genome

Atlas (TCGA), European Genome-phenome Archive (EGA). DNA methylation data was integrated with publicly available gene expression data sets for all five-cancer types. Number of samples and their source for both methylation and expression were included in Table 3. 1.

Methylation of Data	Number of Data	Expression Data	Number of Data
ALL *[222], * [229]	517	ALL * [234]	326
CLL **[162]	187	CLL [235]	143
MCL ** [233]	86	MCL * [236]	121
DLBCL *, ^φ [234], [235]	79	DLBCL * [237]	203
PCNSL * [236]	95	PCNSL * [238]	34
Progenitor (S1, S2, S3, S4) * [231]	24	Progenitor (S1, S2, S3, S4) * [228], * [239]	35
Memory CS **[231]	3	Memory CS * [239]	3
Memory NCS ** [231]	3	Memory NCS * [239]	3
Naïve **[231]	3	Naïve * [239]	3
CD5 Naïve ** [231]	3	CD5 Naïve * [239]	3
CD19++ ** [231]	14	CD19++ * [239]	14

Table 3.1 Data Information that was used for Bioinformatic Analysis

* GEO, **European Genome-phenome Archive (EGA, EGAD00010001012). ^φ The Cancer Genome Atlas (TCGA)

Criteria to identify DSD/SL genes and TSG are shown in Table 3.2 (1) and example of genes which are both fulfilled the criteria for being candidate of DSDG and non-candidate are shown in Figure 3.2. In addition to, any DMR which even fulfil criteria to be DSDG candidate was removed from list, in case it has been already associated with being oncogene in any cancer such as CCDN1 gene was removed from candidate list for MCL even though it matched with candidate criteria (Figure 3.3).

	DSDG candidates	Disease Specific TSG
Methylation status Relevant vs Other Disease/Subgroups	*Hypomethylation in relevant vs Hypermethylation All other disease type	*Hypermethylation in relevant vs Hypomethylation all other disease type
Expression status Relevant vs Other Disease/Subgroups	Higher expression of relevant disease vs Low or absent expression of other disease type	Low expression in relevant vs Higher expression of other disease
Methylation status Disease type vs Healthy cells	Hypomethylation of relevant disease vs Memory B cells	Hypermethylation of disease vs B-cell subsets

Table 3.2 Criteria for Identification of Candidate Disease Specific Dependency and Tumour Suppressor Genes

* Indicates that differences in B value is greater than 0.4 in some cases 0.3 when stated

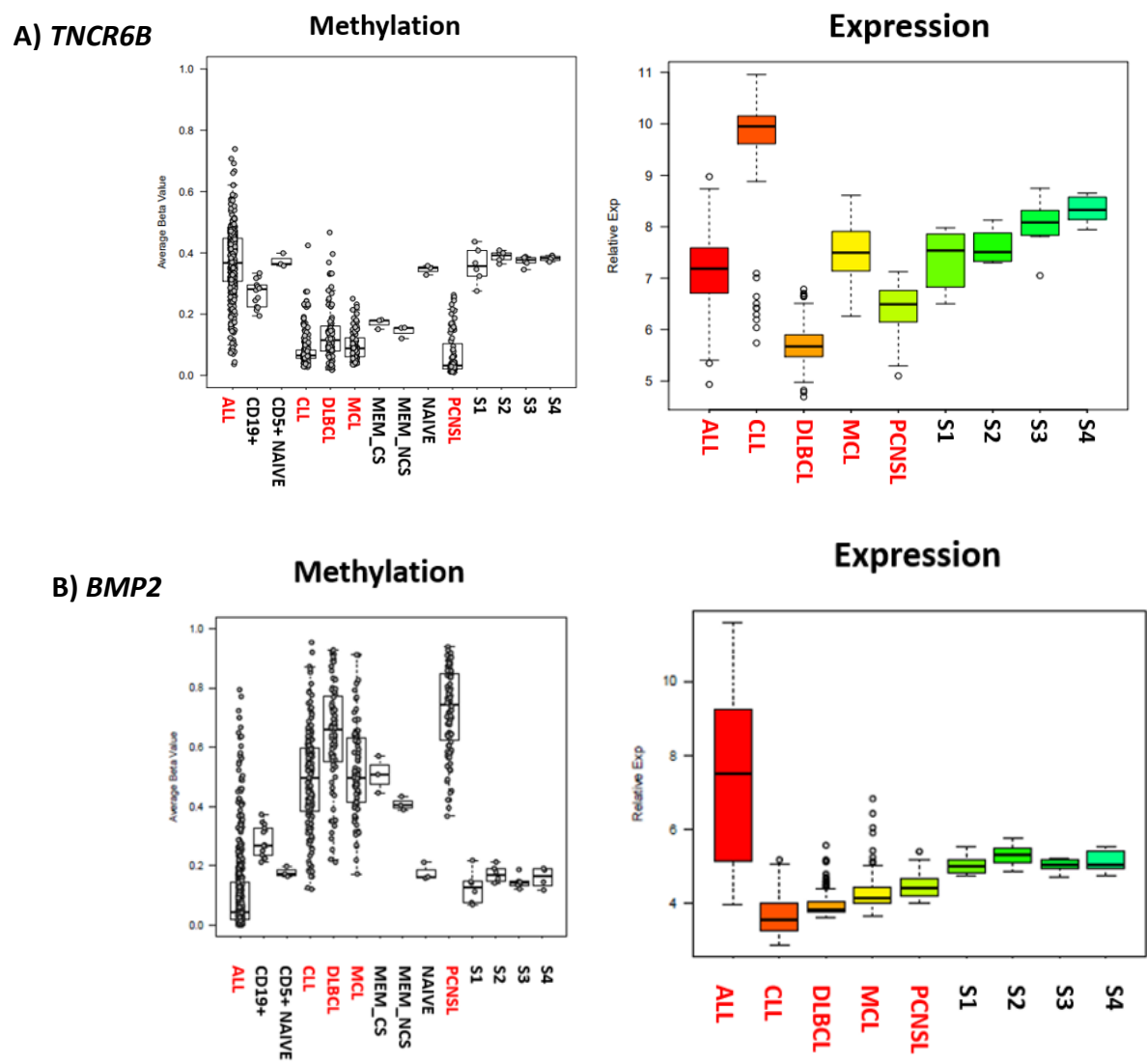


Figure 3.1. Represent examples that does not meet (A) and does meet (B) methylation and expression criteria for DSDG for ALL.

A) This gene (*TNCR6B*) would fail on the methylation expression criteria. For methylation, although different from the other diseases the difference is too small also reflects differences in methylation between B-cell progenitors and differentiated B-cells. For expression, the level is lower than CLL but similar to other diseases and thus reduced expression is not ALL specific. This would also not be a CLL-specific DSDG candidate because the low methylation is not CLL-specific. B- This gene (*BMP2*) shows specific hypo-methylation for ALL while all the other types have higher methylation. On the other hand, upregulation of candidate genes in ALL compared to all the other types of B cell-malignancies and progenitor B cells. Methylation beta values on Y-axis ranges from 0 (un-methylated) to 1 (complete methylation). B-cell progenitor cells which include S1, S2, S3 and S4 population. According to the degree of differentiation stages, progenitor B- cell samples were divided into four different groups. S1 population represent predominantly MPP (Multipotent progenitors) but also containing CLPs (common lymphoid progenitors) and stem cells. S2 population represents pre-B_I cells, S3 population represent pre-B-II cells that expressed *slgMCD19+* and S4 populations represent immature B cells that express *slgM+CD19+*.

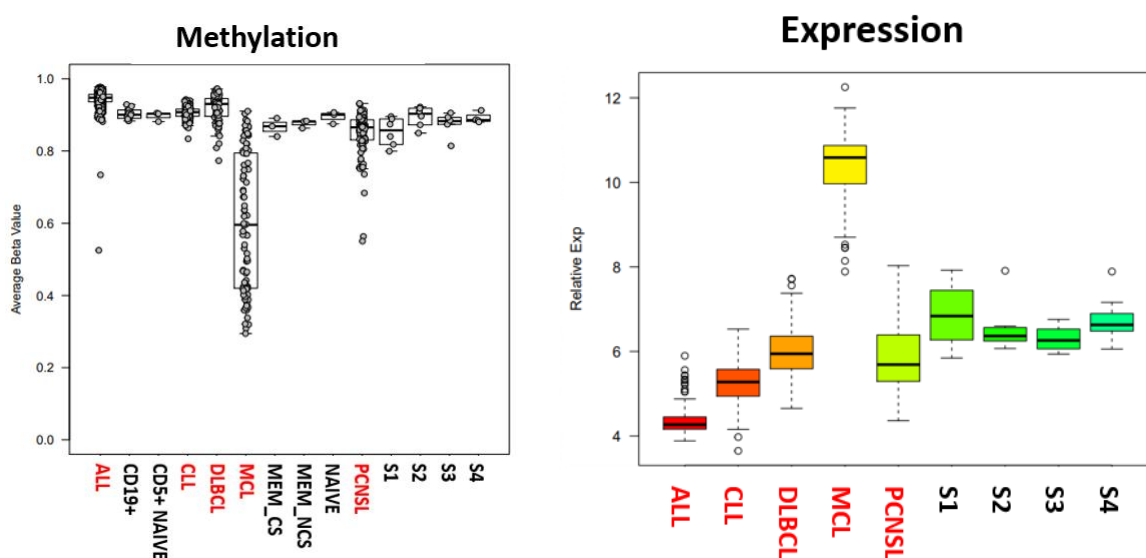


Figure 3.2. Methylation and Expression pattern of *CCND1* gene.

This gene shows specific hypo-methylation for MCL while all the other types have higher methylation. On the other hand, upregulation of candidate gene in MCL compared to all the

other types of B cell-malignancies and progenitor B cells. Methylation beta values on Y-axis ranges from 0 (un-methylated) to 1 (complete methylation).

3.3.1 ALL specific candidate genes

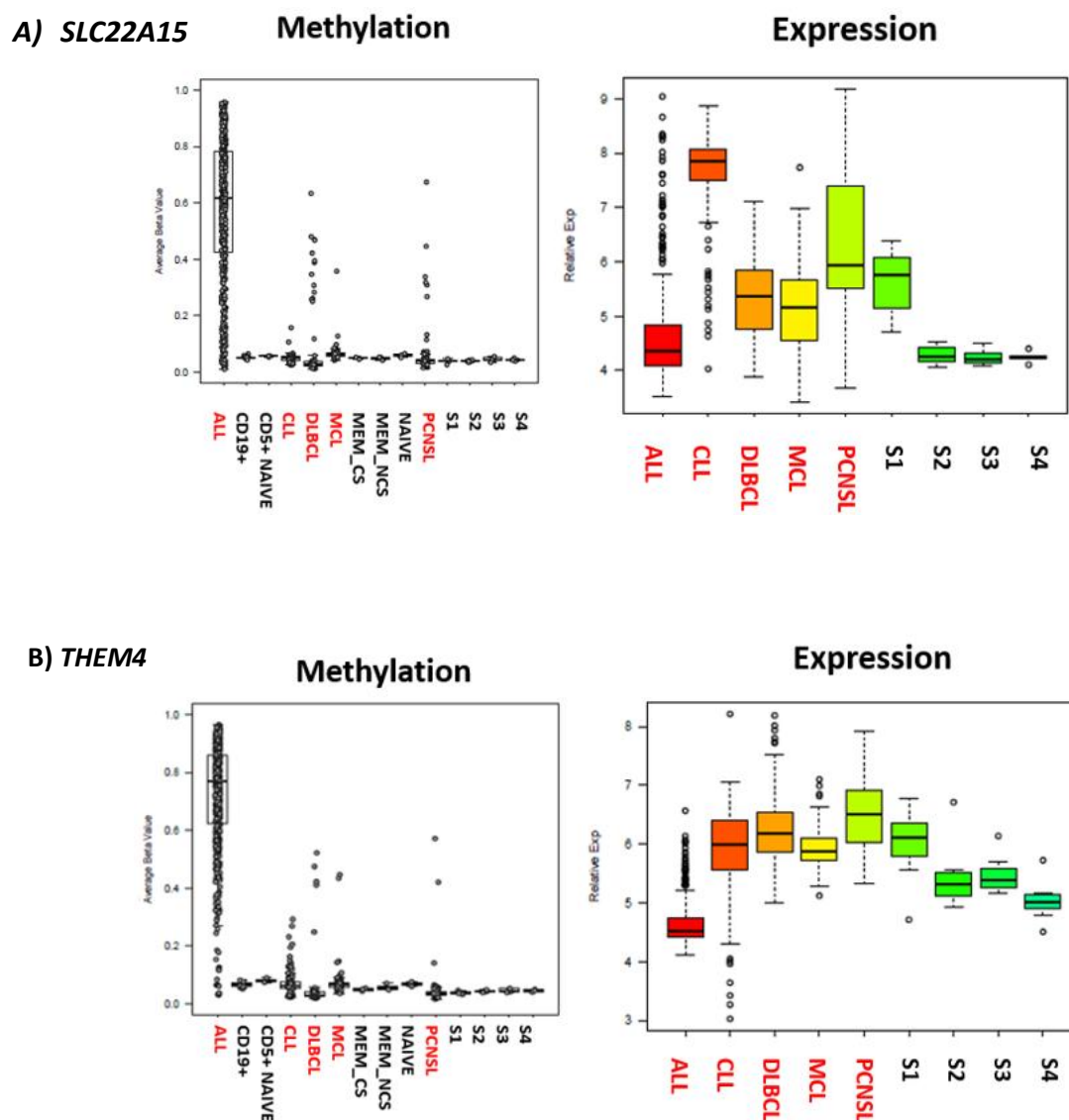
To identify ALL specific DSDG/TSG candidates, genome-wide methylation and expression data of ALL samples compared across the other four type of B-cell malignancies using the bioinformatic pipeline, with the minimum beta value differences across the DMRs set at 0.4 (approximately equivalent to a differences of 40% methylation). This analysis identified a total of 375 cancer-specific/absent DMRs. 61 of them was lack of expression value. Possible reasons might be there was no gene within the 20kb cut-off what we used for analysis and either the gene associated with this DMR was not represented in the expression array data what we used for analysis or has different name on array that could not matched up. 79 out 375 were excluded because average β value methylation differences between ALL and other individual (or all together) did not reach cut-off point 0.4 (40% differences). 227 of 375 were ruled out because methylation pattern did not correspond with expression value. Only remaining eight DMRs fulfilled criteria, four of them count for TSG and another four counted for DSDG (mentioned in section 3.3, Table 3.2).

3.3.1.1 ALL specific TSG candidates

We identified four ALL-specific TSG candidates; *SLC22A15*, *THEM4*, *TTC12*, and *MAP9*. All candidate's DMR were overlapping with the genes transcriptional start site. In addition, the DMRs of all four candidates had a high density of CpG sites, ranging from 11 to 13. Median methylation for all four candidates in ALL lies between beta values of 0.6 to 0.8 (i.e. 60 and 80% methylation). Methylation and expression pattern of all candidates seen in Figure 3.4. For all four candidates significantly, increased methylation was almost entirely restricted to ALL samples. Indeed, although the analysis had set a high minimum threshold for the difference in methylation (beta value of >0.4), the differences in the identified candidates were generally well above this cut-off. For all four candidates ALL samples exhibited lower expression than all other B-cell malignancies (as would be required to meet the expression criteria) and also exhibited low expression levels overall, compatible with the very high levels of promoter methylation identified at all four loci. For three of the four candidates the results do show both a minority of samples with lower methylation and a minority of samples with higher expression

(with the exception of *TTC12*, where almost all samples exhibit relatively low expression). However, as the DNA methylation and gene expression data sets for ALL are derived from different sample sets it was not possible to directly assess the correlation between expression and DNA methylation in the ALL samples to determine if lower methylation was associated with high expression.

Two of the candidates, *MAP9* and *SLC22A15*, have not previously been associated with B-cell derived cancer. *THEM4* and *TTC12* have also not been widely studied in ALL although they have been reported to be hypermethylated in ALL (references [248] and [249] respectively).



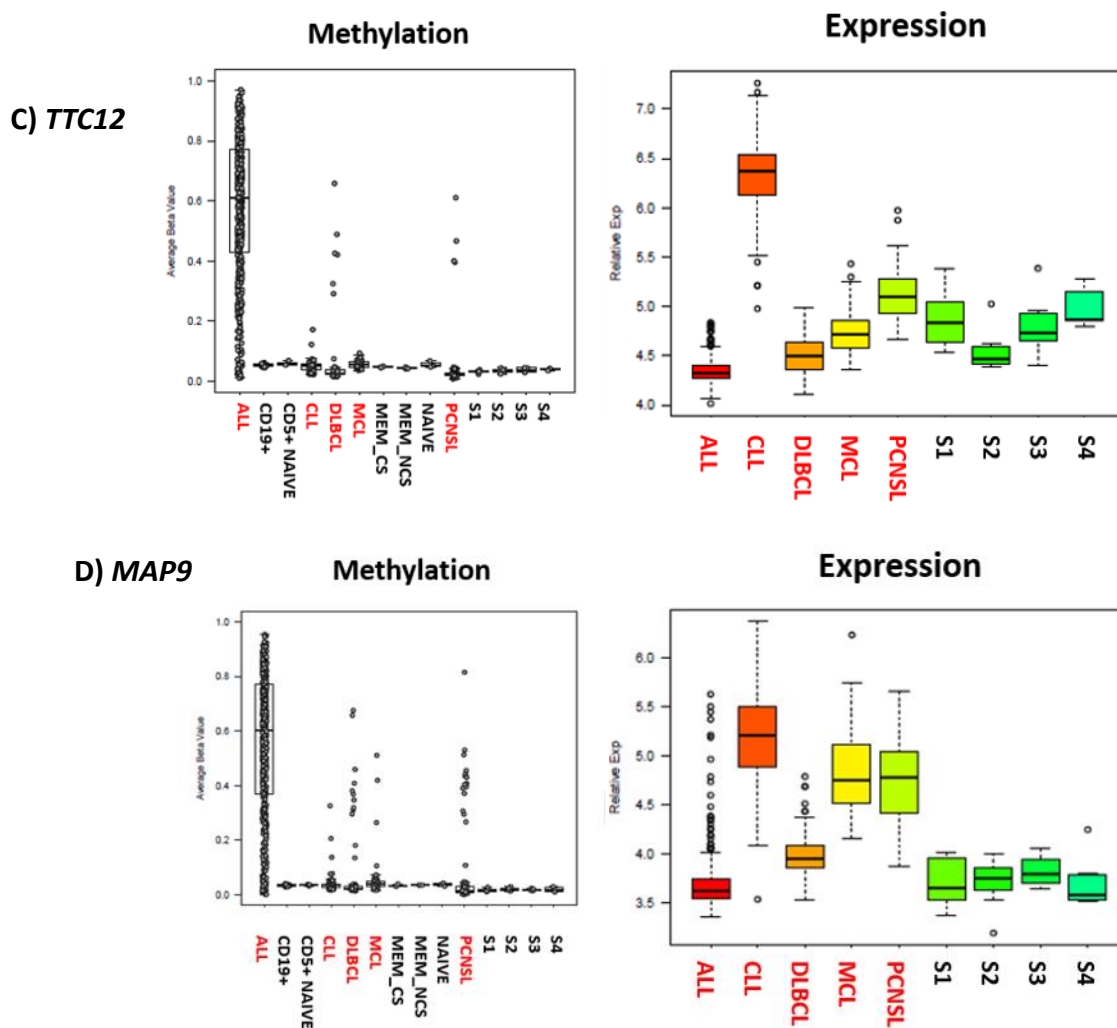


Figure 3.3. Methylation and expression data.

Methylation and expression data are shown for A) *SLC22A15*, B) *THEM4*, C) *TTC12*, D) *MAP9*. Methylation beta values on Y-axis ranges from 0 (un-methylated) to 1 (complete methylation). The candidate genes show specific hyper-methylation for ALL while all the other populations have lower methylation. Expression data shows downregulation of candidate genes in ALL compared to all the other types of B cell-malignancies.

3.3.1.2 ALL specific DSDG candidates

We identified total four DSDG candidates; *BMP2*, *CTGF*, *NPY*, and *ZNF423*. For three of four candidates, the DMR was overlapping with the TSS (*BMP2*, *CTGF*, *NPY*). However, for the other candidate (*ZNF423*), the DMR was located distal to the TSS (18461bp), although within the limit of 20kb. The number of CpG sites within the DMRs for *BMP2*, *CTGF*, *NPY*, and *ZNF423*

was 11, 14, 10, 2 respectively. Among these genes, *ZNF423* [250], [251] and *CTGF* [252] have been shown to be associated with ALL, while the other two candidates have not previously been suggested to have a functional role in B-Cell malignancies. However, elevated expression of *NPY* and *BMP2* in B-ALL samples has been reported [253], [254] respectively.

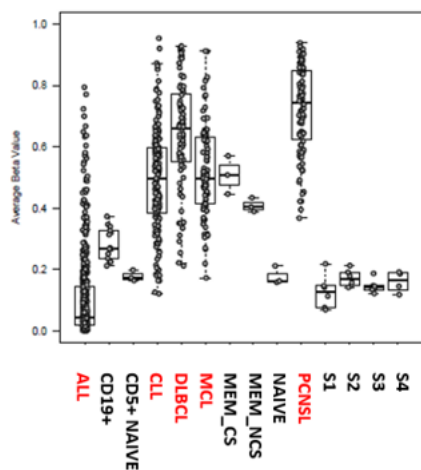
Methylation and expression levels, for each of the candidates, are illustrated in Figure 3.5. The candidates appear to split into two slightly different groups. Two of them (*BMP2*, *NPY*) are consistent with our original approach for identifying genes on the basis of specifically failing to acquire methylation in the disease of interest. For those genes methylation is low in normal progenitor cells and is then acquired in more proliferated normal B-cell subsets (e.g. naïve and memory B-cell populations). Similarly, all other B-cell malignancies also acquire methylation similar to the more proliferated normal cell populations. A potential complication with this analysis is that the methylation pattern for these two loci is also linked to differentiation status – i.e. ALL is similar to normal progenitor cells, while the mature B-cell malignancies are similar to B-memory cells. Thus, an alternative explanation could be that the differences in methylation are just a reflection of differentiation status and not an active selection against methylation (and thus for retention of expression of the candidate DSDG). However, several features argue against this interpretation. Firstly, methylation in the differentiated B-cell malignancies does not reflect their specific stage of differentiation. For example, MCL and most CLL cases are thought to be derived from a stage equivalent to naïve B-cells, but their methylation levels are much higher than naïve B-cells and similar to memory cells. Similarly, DLBCL and PCNSL are derived from a stage between naïve and memory B-cells but do not have intermediate methylation levels and instead show methylation level well above the level seen in memory B-cells. In contrast, the patterns of methylation at these genes across the mature B-cell malignancies and normal cell populations essentially perfectly reflects the pattern of proliferation related methylation seen in most CpG sites found methylated in ALL (see section 1.6 figure 1.25) and thus the methylation patterns of these genes is indistinguishable from other loci though to result as a by-product proliferation. Lastly, expression of the two genes is increased relative to normal progenitor cells as well as in comparison to other B-cell malignancies, consistent with this upregulation being specific for ALL.

The second two identified DSDG (*CTGF*, *ZNF423*) exhibit similar differences in methylation within the B-cell malignancies but show differences from *BMP2* and *NPY* in normal

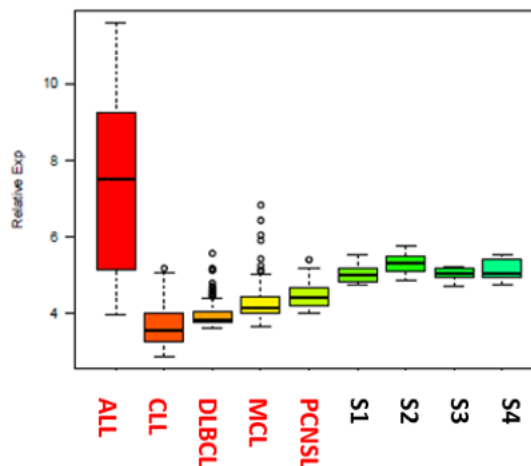
cells, in that normal cells also exhibit very high methylation at these loci. Thus, in these cases it appears that ALL exhibits specific demethylation of the loci, as opposed to retention of a low methylation level.

A) *BMP2*

Methylation

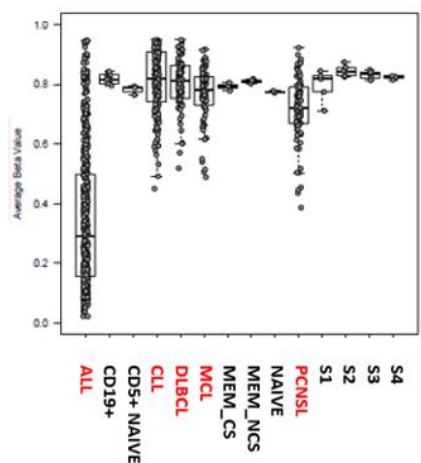


Expression

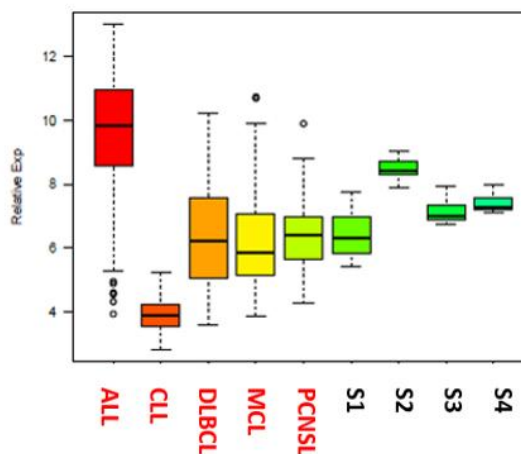


B) *CTGF*

Methylation

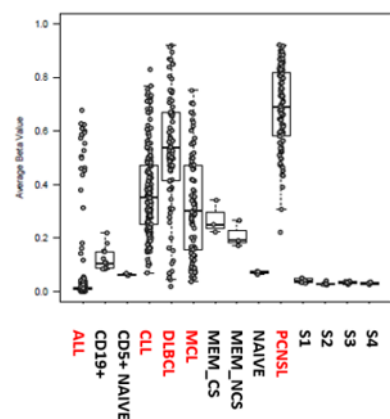


Expression

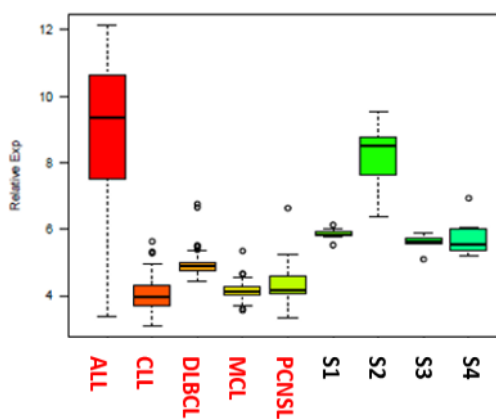


C) *NPY*

Methylation



Expression



D) ZNF423

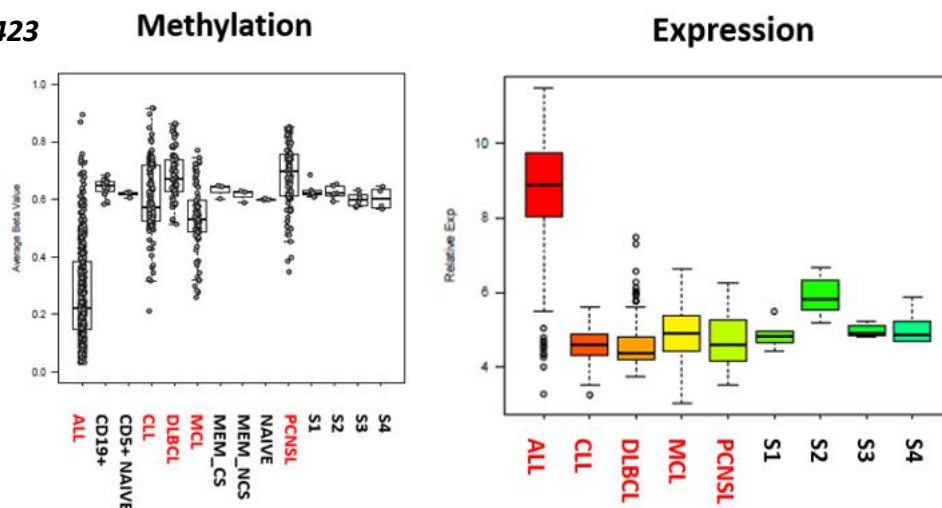


Figure 3.4. Methylation and expression data.

Methylation and expression data are shown for A) BMP2, B) CTGF, C) NPY, D) ZNF423. Methylation beta values on Y-axis ranges from 0 (un-methylated) to 1 (complete methylation). The candidate genes show specific hypo-methylation for ALL while all the other types have higher methylation. On the other hand, upregulation of candidate genes in ALL compared to all the other types of B cell-malignancies and progenitor B cells.

3.3.2 MCL specific candidates

To identify MCL specific DSDG/TSG candidates, genome-wide methylation and expression data of ALL samples compared across the other four type of B-cell malignancies using the bioinformatic pipeline, with the minimum beta value differences across the DMRs set at firstly 0.4 (approximately equivalent to a differences of 40% methylation). However, any candidate had not been identified therefore cut-off was reduced 0.3 which differs from that used in ALL to account for the broadly lower levels of acquisition of methylation changes seen in some of the other b-cell malignancies especially CLL and MCL (see section 1.6 figure 1.25). Overall, 219 cancer specific DMRs were identified in this analysis. 32 of them have lack of expression data (possible reasons mentioned in section 3.3.1). 173 out 219 were failed because lack of consistency in methylation differences between MCL and other b-cell malignancies especially similar lower levels of methylation as those seen in CLL. 12 of them did not display inverse correlation between methylation and expression value.

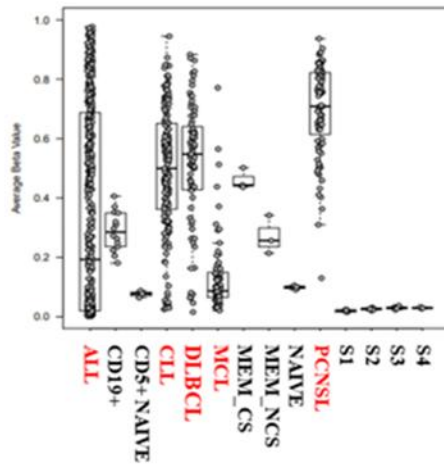
Only remaining three DMRs fulfilled criteria for being DSDG; *SOX11*/ *MFHAS1*/ *CCDN1*. However, *CCDN1* has already referred to oncogene so that it was removed from candidate list. Unlike ALL, we did not identify any TSG candidate for MCL.

The DMR of *SOX11* overlaps with its TSS, in contrast, the DMRs for *MFHAS1* are located proximal to their TSS (1636bp). In addition, the DMR's include 11, and 6 CpG sites respectively for *SOX11*, and *MFHAS1*. Median methylation for two candidates in MCL lies between beta values of 0.1 to 0.4 (i.e. 10 and 40% methylation). Methylation and expression levels, for each of the candidates, are illustrated in Figure 3.6 *SOX11* is consistent with our original approach for identifying genes on the basis of specifically failing to acquire methylation in the disease of interest. For those genes methylation is low in normal progenitor cells and is then acquired in more proliferated normal B-cell subsets (e.g. naive and memory B-cell populations). Similarly, all other B-cell malignancies also acquire methylation similar to the more proliferated normal cell populations except ALL. Methylation differences between ALL and MCL (median 0.2 and 0.1 respectively) is lower than cut-off 0.3 value. However, two ALL genetic subtype (ETV6/RUNX1 and TCF3/PBX1) is highly expressed *SOX11* gene comparing other cytogenetic subtype ALL based on Leukaemia MILE study data [255] (Figure in 3.7). *SOX11*, has been widely studied for its role in MCL and used as a diagnostic marker [256]. The gene codes a transcription factor that plays important roles in cell survival, suppression of apoptosis and induction of angiogenesis, in many cancer types [257], [258], [259]. Overall, when either methylation / expression pattern or it's role considered, *SOX11* was chosen candidate for MCL.

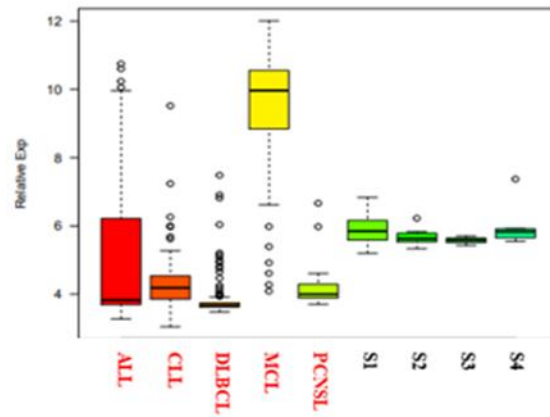
The second identified DSDG for MCL was *MFHAS1* exhibits similar differences in methylation within the B-cell malignancies even DLBCL and PCNCL show hypomethylation comparing less proliferated B-cell malignancies, CLL and ALL. In normal cells also exhibit very high methylation at these loci. Thus, in this case it appears that MCL exhibits specific demethylation of the loci, as opposed to retention of a low methylation level. In addition, *MFHAS1* has been implicated as a potential oncogene in malignant fibrous histiocytomas [260] and mutation in *MFHAS1* gene has been associated with DLBCL [261], [262].

A) SOX11

Methylation

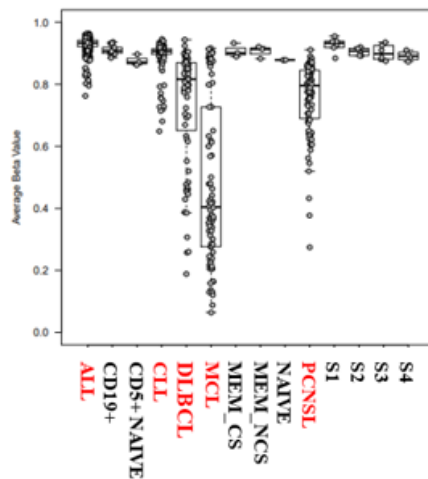


Expression



B) MFHAS1

Methylation



Expression

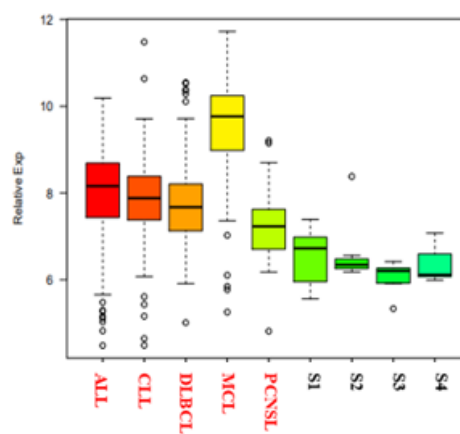


Figure 3.5. Methylation and expression pattern of MCL DSDG candidates.

Methylation and expression data are shown for A) SOX11, B) MFHAS1. Methylation beta values on Y-axis ranges from 0 (un-methylated) to 1 (complete methylation). The candidate genes show specific hypo-methylation for MCL while all the other types have higher methylation. On the other hand, upregulation of candidate genes in MCL compared to all the other types of B cell-malignancies and progenitor B cells.

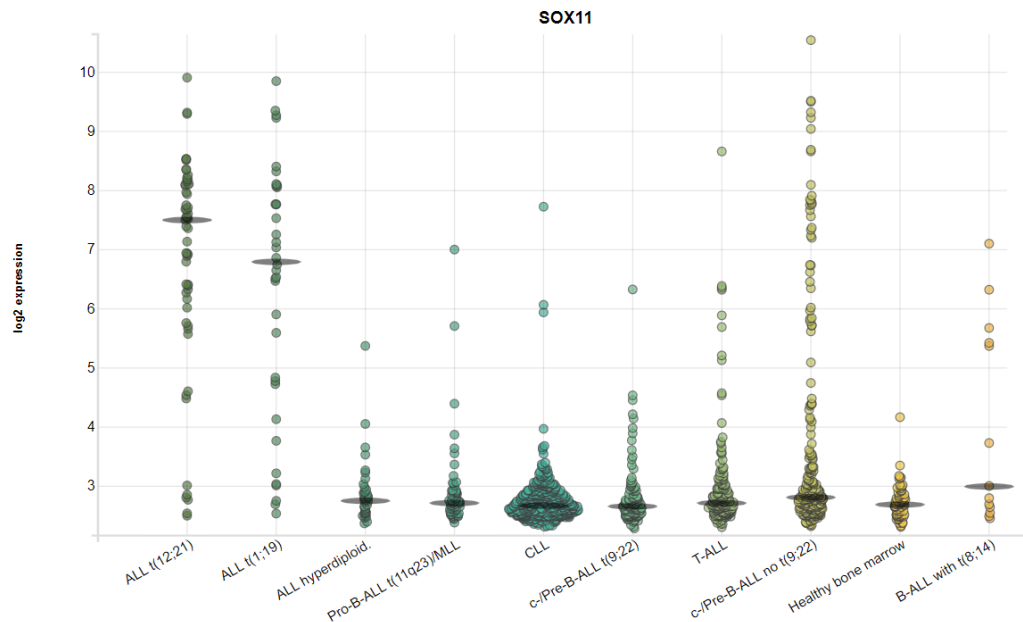


Figure 3.6. SOX11 expression on panel of ALL subtypes, and other Leukaemia Types.

SOX11 expression is specifically high in ETV6/RUNX1 and TCF3/PBX1 genetic subtypes of ALL. It's expression is limited in other subtype of ALL, CLL,, AML, T-ALL and normal healthy bone marrow cells. Figure is obtained from [255].

3.3.3 CLL specific candidates

To identify CLL specific DSDG/TSG candidates, genome-wide methylation and expression data of CLL samples compared across the other four type of B-cell malignancies using the bioinformatic pipeline, with the minimum beta value differences across the DMRs set at firstly 0.4 (approximately equivalent to a differences of 40% methylation). However, any candidate had not been identified therefore cut-off was reduced 0.3 which differs from that used in ALL to account for the broadly lower levels of acquisition of methylation changes seen in some of the other b-cell malignancies especially CLL and MCL (section 1.6, figure 1.25). As CLL has similarly lower level of methylation as those seen in MCL, the same reduced 0.3 cut-off was used. Total 684 DMRs were identified, although 120 of them has lack of expression data (possible reasons mentioned in section 3.3.1). Out 491 was failed because lack of consistency in methylation differences between CLL and other b-cell malignancies especially similar lower levels of methylation as those seen in most cases MCL. 72 of them did not display inverse correlation between methylation and expression value. Only remaining one DMR fulfilled criteria SNX18. Similarly, to MCL, we did not identify any TSG candidates for CLL. SNX18

codes membrane-associated protein that plays role in intracellular traffic and protein haemostasis in the cell [263]. There are no current studies implicating *SNX18* in haematological malignancies, although a single study identifies a potential role for *SNX18* in pancreatic cancer metastasis [264].

The DMR of *SNX18* overlaps with its TSS, and its DMR include 11 CpG sites. Median beta value of methylation of CLL candidate was approximately 0.1 (i.e. 10% methylation). Methylation and expression level of the candidate is illustrated in Figure 3.8. The DMR shows demethylation specifically in CLL across the other B-cell malignancies however, some samples in PCNSL show low levels of methylation but difference between median CLL methylation and the next lowest group, PCNSL higher than cut-off 0.3. Corresponding methylation value, highest expression of *SNX18* gene was seen in CLL compared the other B-cell malignancies and normal B-cells.

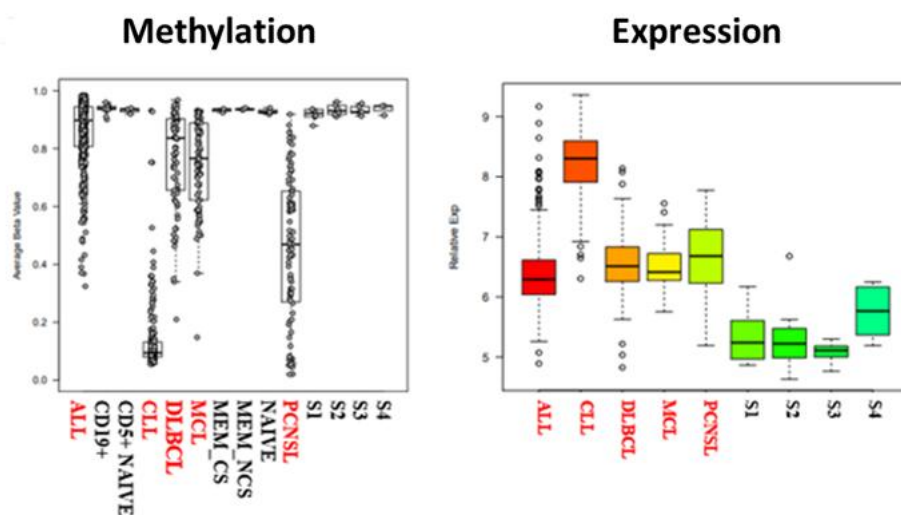


Figure 3.7. Methylation and expression pattern of CLL specific DSDG candidate, *SNX18*.

Methylation and expression data are shown for *SNX18*. Methylation beta values on Y-axis ranges from 0 (un-methylated) to 1 (complete methylation). The candidate gene show specific hypo-methylation for CLL while all the other types have higher methylation. On the other hand, upregulation of candidate gene in CLL compared to all the other types of B cell-malignancies and progenitor B cells.

3.3.4 PCNSL specific candidate

To identify PCNSL specific DSDG/TSG candidate(s), genome-wide methylation and expression data of PCNSL samples compared across the other four type of B-cell malignancies through novel bioinformatic approach using minimum beta value differences across the DMRs was set 0.4 up (approximately equivalent to a differences of 40% methylation). Total 867 DMRs were identified, although 139 of them have lack of expression data (possible reasons mentioned in section 3.3.1). 584 was failed because lack of consistency in methylation differences between PCNSL and other B-cell malignancies especially similar levels of methylation as those seen in most cases DLBCL. PCNSL is known to have a significant overlap with DLBCL, and clustering analysis performed on all the b-cell malignancy methylation data (Lalchungnunga, Atasoy manuscript in preparation) has suggested that DLBCL and PCNSL form two overlapping clusters (unlike other B-cell malignancies, which were clearly differentiated from each other. Therefore, most of DMRs do not show median methylation B-value cut-off (0.4) differences between PCNSL and DLBCL. Therefore, out of 584 did not considered as candidate for PCNSL. 143 did not display inverse correlation between methylation and expression value. Only remaining one DMR fulfilled criteria as TSG candidate, *FSNC1*. Unlike ALL, MCL and CLL, there was no DSDG candidate identified.

The DMR of *FSNC1* overlaps with its TSS, and it's DMR include 9 CpG sites. Median beta value of methylation in PSNCL was approximately 0.7 (i.e. 70% methylation). This loci specifically hyper-methylated in PCNSL samples comparing other B-cell malignancies, even though some samples of DLBCL samples exhibit similar level of methylation with PCNSL, median beta value of methylation in DLBCL was closed to fully methylated level 0. *FSNC1* overall expression in PCNSL samples exhibited lower expression than all other B-cell malignancies. *FSNC1* codes actin-binding proteins plays a role in cell migration, invasion, metastasis [265] and it already has been shown as disease marker for Reed-Sternberg cells in Hodgkin's lymphoma [266], and specific overexpression in AML [267]. Methylation and expression data for *FSNC1* are shown in Figure 3.9.

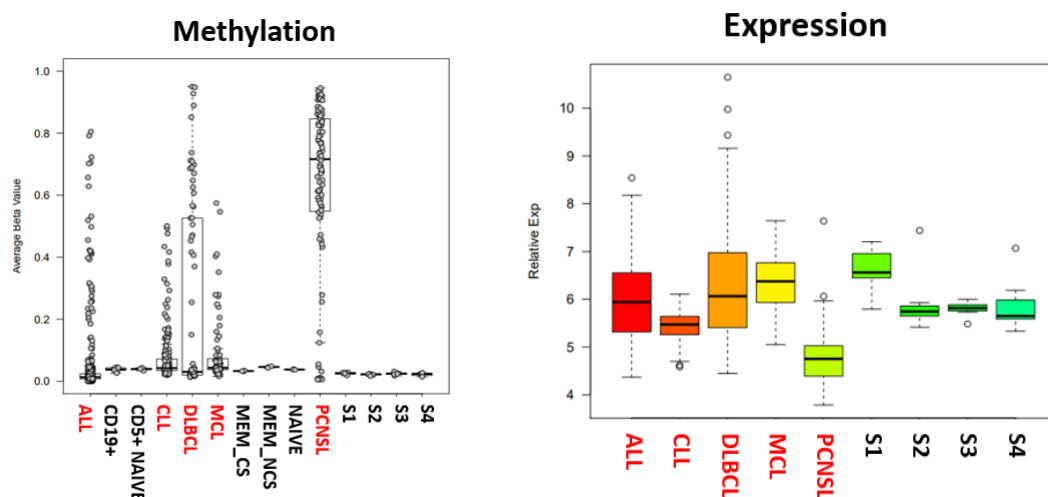


Figure 3.8. Methylation and expression pattern of PCNSL TSG candidate, FSCN1.

Methylation and expression data are shown for FSCN1. Methylation beta values on Y-axis ranges from 0 (un-methylated) to 1 (complete methylation). The candidate gene show specific hyper-methylation for PCNSL while all the other types have lower methylation. Expression data shows downregulation of candidate gene in PCNSL compared to all the other types of B cell-malignancies and progenitor B cell.

3.3.5 PCNSL & DLBCL specific candidate

Similar analysis was carried out to identify candidate DSD and TS genes in DLBCL and PCNSL individually. However, PCNSL is known to have a significant overlap with DLBCL, and clustering analysis performed on all the b-cell malignancy methylation data (Lalchununga, Atasoy manuscript in preparation) has suggested that DLBCL and PCNSL form two overlapping clusters (unlike other B-cell malignancies, which were clearly differentiated from each other). Furthermore, analysis of DLBCL as a single subtype did not identify any candidates and this analysis revealed a strong overlap between differentially methylated genes in DLBCL and PCNSL. Therefore, we ran the analysis amalgamating DLBCL and PCNSL as a single entity to identify DSDG/TSG that were relevant for both DLBCL and PCNSL. As a cut-off value, a beta-value difference of 0.4 (40% difference in methylation) was used. A total of 817 DMRs were identified, although 130 of them have lack of expression data (possible reasons mentioned in section 3.3.1). 526 out 817 was excluded because average β value methylation differences between PCNSL&DLBCL and other individual (or all together) did not reach cut-off point 0.4 (40% differences). 160 of 817 was ruled out because methylation pattern did not correspond

with expression value. Only remaining DMR fulfilled criteria as a TSG candidate for both malignancies, *SDK2*, was selected based on both the required methylation and gene expression differences.

The DMR for this candidate was located in the TSS and contains 14 CpG sites. Median beta value of methylation of the candidate was approximately 0.5 (i.e. 50% methylation) in DLBCL and approximately 0.7 in PCNSL. The candidate genes show specific hyper-methylation for DLBCL and PCNSL while all the other types have lower methylation. *SDK2* shows downregulation of candidate genes in DLBCL and PCNSL compared to all the other types of B cell-malignancies and has similar expression pattern in progenitor B cells. Methylation and expression level of the candidate is illustrated in Figure 3.10. The *SDK2* gene encodes an adhesion molecule that plays a role in synaptic connections in the retina [268]. There were no previous studies linking the genes to b-cell malignancies.

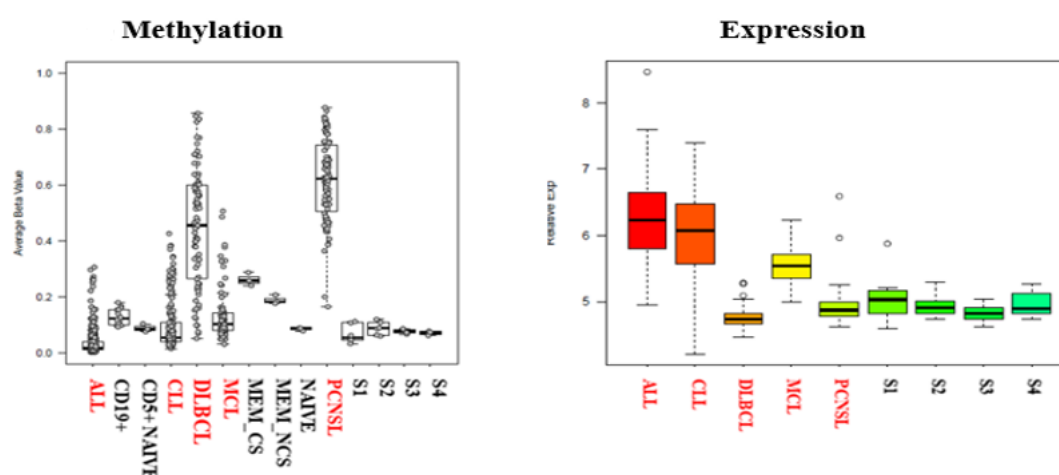


Figure 3.9. Methylation and expression pattern of the DLBCL/PCNSL TSG candidate *SDK2*.

Methylation and expression data are shown for *SDK2*. Methylation beta values on Y-axis range from 0 (un-methylated) to 1 (complete methylation). The candidate gene shows specific hyper-methylation for DLBCL and PCNSL, while all the other types have lower methylation. Expression data shows reduced expression of the candidate gene in DLBCL and PCNSL compared to all the other types of B cell-malignancies.

3.3.6 Overall Results

In the light of our criteria, this approach has been successful in identifying total 13 DSD and TS gene candidates across all the B-cell malignancies under study as shown in Table 3.1. Seven of them were DSDG and six of them were TSG candidates. Four of seven DSDG candidates for ALL and two of seven DSDG candidates for MCL and one of seven DSDG for CLL were identified. However, there was not any DSDG candidate for DLBCL and PCNSL. On the other hand, four of six TSG candidates for ALL and one of six TSG candidates for PCNSL and one of six for shared DLBCL and PCNSL were identified and there was no any TSG candidate for CLL and MCL.

Three of seven DSDG candidates across the five types of B-cell malignancies have been associated with haematological malignancies, *SOX11* [249], *MFHAS1* [261], [262] and *ZNF423* [251]. On the other hand, there are no functional studies for four of seven DSDG candidates in leukaemia/ lymphoma. However, rest of them have been associated with multiple cancers. Regards of TSG candidates, only one of six candidates has been previously linked with B-lymphocyte-derived cancers [258]. However, there were not any functional studies for rest of TSG candidates in haematological malignancies beside there were only limited functional studies in other cancers. (Table 3.3). To sum up, 5 of 13 candidates have been linked with B-lymphocyte-derived cancers before, suggesting that this method is effective to identify cancer related genes. In contrast, 8 of 13 candidates across the B- cell malignancies under study have not been associated with leukaemia/lymphoma, indicating that the method has the ability to find cancer therapeutic genes that cannot be identified in other ways.

Candidate Type	Gene Name	Function of Gene Product	Investigated in any cancer	Investigated in Leukemia/Lymphoma
DSDG for CLL	<i>SNX18</i>	A member of the sorting nexin family is involved in intracellular trafficking and protein homeostasis [263]	Pancreatic cancer metastasis [264]	-
DSDG for MCL	<i>SOX11</i>	Transcription factors, promoting cell survival, suppress apoptosis, inducing tumour angiogenesis [256]	In many cancer [257], [258], [259]	Diagnostic marker for MCL [256] DNA hypomethylation and high expression ALL cell lines [269]
DSDG for MCL	<i>MFHAS1</i>	A potential oncogene whose expression is enhanced in some malignant fibrous histiocytomas (MFH) [260]	Malignant fibrous histiocytomas [260]	Mutation on the DLBCL [261], [262]
DSDG for ALL	<i>NPY</i>	A neuropeptide that plays role in cell growth, migration, angiogenesis, and metastasis [270]	Reviewed in [270]	Elevated serum level on B-Cell Leukaemia [253]
DSDG for ALL	<i>BMP2</i>	A secreted ligand of the transforming growth factor- β (TGF- β) superfamily, plays essential roles in many cancerous processes by like angiogenesis, epithelial-mesenchymal transition (EMT), and cancer stem cells [271]	In many cancer [272], [273], [274]	Elevated expression on B-cell ALL [254]
DSDG for ALL	<i>CTGF</i>	A mitogen that is secreted by vascular endothelial cells. The encoded protein plays a role proliferation, adhesion, migration, and metastasis [252], [275], [276]	In other cancer [275], [276]	shRNA knockdown of CTGF in B-ALL cell line (Reh) cause suppression of cell growth [252]
DSDG for ALL	<i>ZNF423</i>	It is a transcription factor and play role as both activator and suppressor [250]	In other cancer [277], [278]	In ALL, inhibits transcriptional factor of B-cell differentiation (EBF1) resulting in B-cell maturation arrest [251]
TSG for ALL	<i>THEM4</i>	Carboxyl-terminal modulator protein, inhibit phosphorylation and activation of Protein kinase B (PKB) which play critical role upon cell survival and transformation [279]	In few cancers; breast cancer [280], glioblastoma [281], lung cancer [282]	Hypermethylated in B-ALL [248]
TSG for ALL	<i>MAP9</i>	Microtubule associated protein 9, a gene critical for spindle assembly and cytokinesis. Suppress cell proliferation and inducing cell cycle arrest [283]	It is downregulated in few cancers; hepatocellular carcinoma [283] breast and colorectal cancer [284]	-

TSG for ALL	<i>SLC22A15</i>	Organic ion transporters [285]	Promote epithelial-mesenchymal transition (EMT) in pancreatic cell line [286]	-
TSG for ALL	<i>TTC12</i>	Cytoplasmic protein that plays a role in the proper assembly of dynein arm complexes [287]	Loss of function in adult germ cell tumour [288]	Hypermethylated in ALL [249]
TSG for DLBCL & PCNSL	<i>SDK2</i>	Adhesion molecule [268]	Mutation in triple-breast cancer [289]	-
TSG for PCNSL	<i>FSCN1</i>	Actin-binding proteins plays a role in cell migration, invasion, metastasis [265]	Reviewed in [265]	Marker for Reed-Sternberg cells in Hodgkin's lymphoma [266] Found elevated express in AML not ALL serum samples [267]

Table 3.3 All candidate information about function and investigation on both any cancer and leukaemia/lymphoma

Candidate Type	Gene Name	Chr	Distance to TSS	DMR start	DMR ends	No of CpG
DSDG for CLL	<i>SNX18</i>	5	0	53813164	53817963	15
DSDG for MCL	<i>SOX11</i>	2	0	5831147	5834638	11
DSDG for MCL	<i>MFHAS1</i>	8	1636	8748536	8749494	6
DSDG for ALL	<i>NPY</i>	7	0	24323128	24323939	10
DSDG for ALL	<i>BMP2</i>	20	0	6748710	6751435	11
DSDG for ALL	<i>CTGF</i>	6	0	132270033	1.32E+08	14
DSDG for ALL	<i>ZNF423</i>	16	18461	49872492	49873368	2
TSG for ALL	<i>THEM4</i>	1	0	1,52E+08	1.52E+08	11
TSG for ALL	<i>MAP9</i>	4	0	1,56E+08	1.56E+08	13
TSG for ALL	<i>SLC22A15</i>	1	0	116518738	1.17E+08	12
TSG for ALL	<i>TTC12</i>	11	0	113184936	1.13E+08	13
TSG for DLBCL & PCNSL	<i>SDK2</i>	17	0	71638531	71641669	14
TSG for PNC SL	<i>FSCN1</i>	7	0	5631189	5632476	9

Table 3.4 DSD and TS genes candidates identified in five types of B-cell malignancies.

TSS refers the transcriptional starting site, Chr refers chromosome.

3.4 Discussion

Identifying genes with roles on non-oncogenic pathways which are required for cancer cell survival is one of the promising approaches to design new therapies that kill specifically tumour cells and prevent normal cell toxicity [290]. Our group present new bioinformatic approach based on integration genome wide methylation and genome wide gene expression data which was originally designed to identify molecular subtype specific SL like genes, referred to as subtype specific vulnerability genes (SSV). The idea beyond this approach based on that identifying genes which are selectively protected from acquiring methylation in the presence of specific cancer-driving mutations or pathways. This selective retention of gene expression in the presence of cancer-driving genetic events suggests that these subtype-specific vulnerability genes could be promising targets

for developing novel cancer therapies that specifically target malignant cells while minimizing toxicity to normal cells. This original approach was applied across multiple cancer types, initially focussing on genetic subtypes of ALL then medulloblastoma, and has been shown to be able to identify candidate genes in the all but one cancer subtypes assessed and was identified candidate genes in almost all subtypes assessed which had a high rate of functional confirmation (80% of candidates analysed) [221]. A limitation of the above approach is that SL genes can only be identified for specific subtypes within a cancer type and thus the identified targets would only be relevant for a subset of patients. This limitation could potentially be overcome if the approach could be expanded to comparing multiple cancer types. This may be possible for multiple cancer types derived from the same normal cell type, assuming the methylation changes across the genome would therefore be highly similar.

We applied this method to the analysis of five types of B-lymphocyte-derived cancers and identified total 13 candidates across the five types of B-cell malignancies. Seven of thirteen candidates were DSDG candidates, for ALL (*BMP2*, *CTGF*, *NPY*, *ZNF423*), CLL (*SNX18*) and MCL (*SOX11*, *MFHAS1*). No DSDG candidates were identified for DLBCL and PCNSL. Consistent with a functional role for the identified candidates, four of the genes (*CTGF*, *ZNF423*, *SOX11*, *MFHAS1*) have been previously linked to in B-cell malignancies (leukaemia/ lymphoma). The *SOX11* gene codes for a transcription factor that play critical role in tumour progresses by suppressing apoptosis and inducing tumour angiogenesis and cell survival and it has been already used as a diagnostic marker to distinguish MCL from other b-cell malignancies because it is not detected in other mature B-cell malignancies and normal lymphoid cells [256]. In addition, *MFHAS1* has been proposed as a potential oncogene, associated with malignant fibrous histiocytomas [260]. Furthermore, its mutation is linked with DLBCL [261], [262]. In terms of *CTGF*, this gene encodes a mitogen that is secreted by vascular endothelial cells. The encoded protein plays a role proliferation, adhesion, migration, and metastasis in many cancers [275], [276]. It has been found to exhibit elevated expression in B-ALL [291]. In addition, it has been shown that knock down by shRNA in an ALL cell line caused suppression of cell growth [252]. *ZNF423* gene codes transcription factor which has been implicated in both suppression and activation of tumorigenesis [250]. It has been associated with ALL

through suppression of the *EBF1* transcription factor leading to B-Cell maturation arrest [251]. For the remaining DSDG candidates, there are no functional studies regarding their potential roles in leukaemia/lymphoma malignancies, however, elevated *NPY* protein expression [253] and increased *BMP2* expression [254] have been reported in B-ALL cases.

This study also identified a number of novel tumour suppressor gene candidates. Against our initial approach that allow to identify SL gene for subtype-specific level (only detect 21 SL not any TSG candidate), we could identify total six (six of thirteen) TSG candidate across the five types of B-malignancies. Four of six TSG candidates account for ALL (*SLC22A15*, *THEM4*, *TTC12* and *MAP9*), This included one for PSNCL (*FSCN1*) and one (*SDK2*) for shared between DLBCL and PCNSL and four TSG candidates in ALL (*SLC22A15*, *THEM4*, *TTC12* and *MAP9*). Interestingly this contrasts with our original subtype specific approach which identified 21 synthetic lethal candidates in specific ALL subtypes but no TSG candidates. Of the TSG candidates identified here, *FSCN1*, has been linked to B-cell malignancies in a previous study and was suggested as a marker for Reed-Sternberg cells in Hodgkin's lymphoma [265]. Among of the other five genes, there are not any functional studies previously published in haematological cancers but *THEM4* [248] and *TTC12* [249] have been found hypermethylated in B-ALL. Although *MAP9* has not been studies any B-cell malignancies yet, it has already nameable as potential tumour suppressor gene in colorectal [292], hepatocarcinoma [293] and gastric cancer [294]. *SLC22A15*, and *SDK2* have not previously been implicated in any haematological cancers.

Additional method called “methylation mapping” was done by previous PhD student in our group by using same data set used in this project. A key aim of methylation mapping was that screening out of methylation changes that are driven by non-disease phenomena would allow identification of true cancer-related methylation changes. This study further emphasises the similarity in methylation patterns between different cancer types derived from the same normal cell type. Thus, although this study expanded the bioinformatic approach to compare whole disease together, as opposed to different subtypes of the one disease, the extent of differential methylation between the different diseases remains relatively low and thus the number of potential candidates genes identified was quite small. As will be discussed in more detail in the following chapter, a key underlying reason for this is likely to be that only a small fraction of the differences

observed between B-cell cancers and normal B-cells are actually related to disease development and the primary driver of altered methylation is likely the process of proliferation and occurs in essentially the same fashion in either transformed or normal cells which have proliferated to the same extent [221], XIA M. unpublsh)

According to the result, only a very small fraction (about 3%) of DNA methylation changes in B-cell cancers are disease related, with the overwhelming majority (97%) being driven by normal biological processes, predominantly cell proliferation. Furthermore, the low level of true disease-specific changes can potentially simplify identification of functionally relevant DNA methylation changes, allowing identification of previously unappreciated candidate drivers of cancer development. Furthermore, 15 candidate genes were identified end of the analysis (Lalchununga, Atasoy manuscript in preparation). These exhibit a highly significant overlap with the candidates identified through both methods. While, due to the differences between the approaches, each identifies some additional unique candidates, the number of candidates identified remains low, further suggesting that the number of functionally relevant methylation changes in B-cell malignancies is low, even though large number of methylation changes shared across B-cell malignancies.

The method we initially developed, focussing on identifying genes of functional relevance in a subtype specific manner, identified 21 SLG candidates with at least one candidate for each subtype have been identified for five common subtypes of ALL, however, any TSG candidate could not be identified. Similarly, in medulloblastoma, a total 15 SLG candidates were identified across the four medulloblastoma subtypes [221]. In contrast, the results presented here, identify a smaller total number of candidates, a total 13 candidate genes identified across five B-cell malignancies versus the 21 and 15 candidates identified by the original approach in single diseases. This is perhaps somewhat surprising as different diseases (such as ALL and MCL) might have been expected to exhibit greater variety in methylation patterns than different subtypes within the same disease. However, at least in this study, this did not correlate with a higher number of functionally relevant DNA methylation changes.

Our previous subtype specific analysis also found that more specifically define molecular entities generally led to identification of a higher number of candidates (e.g.

seven candidates were identified in the ETV6-RUNX1 subtype (all of which are characterised by acquisition of the same fusion protein) while only two candidates were identified in the less homogenous high hyperdiploidy subtype (characterised by acquisition of a variable number of additional chromosomes). Thus, the lower number of candidates found per disease in the current approach may be reflective that pathologically defined disease entities could be more accurately thought of a set of closely related, but molecularly distinct, diseases.

As another factor, ploidy potentially could impact methylation status of individual genes and wasn't accounted for in this analysis. This could be a particular concern in cases where genes exhibit allele specific differences in methylation, such as in the case of imprinted genes. In such cases gain or loss of a chromosome would result in changes in the level of methylation pattern at this loci. This would also be the case if allele specific differences in methylation developed by a part of cancer development. However, differences in methylation driven by stochastic gains or losses of methylation results proliferation, may not be significantly impacted. If the average rate of gain or loss of methylation across all alleles was broadly the same, then the change in average methylation is still would be similar.

Another factor that would impact DNA methylation levels is tumour cell content. Only ALL and CLL dataset used in these analyses have information about percentage of tumour cell content (>90% in both datasets) and thus it was not readily possible to incorporate this aspect into the analysis. However, this can be partially estimated by assessing the extent of methylation changes at CpG sites which were consistently methylated across all B-cell disease (see Fig. 1.6). Lower average tumour cell content would result in lower average change at these CpG sites. This will also be impacted by the rate of methylation change in each of the different B-cell malignancies. While this reduces the ability to utilise this data to estimate the impact of tumour cell content, it is the combination of both these effects which is most important to be accounted for in the analysis shown here.

Overall, the approach presented here is a promising new approach to uncover previously unidentified cancer dependency and tumour suppressor genes as new therapeutic targets with the potential to enable the development of more cancer specific

therapeutic approaches. Furthermore, using the disease level approach utilised in this study, as opposed to the subtype specific approach we previously reported, has the potential to aid in the development of therapeutic approaches in a disease that could be applicable to all patients with the disease.

*4 FUNCTIONAL ASSESSMENT OF ACUTE
LYMPHOBLASTIC LEUKAEMIA-SPECIFIC
TUMOUR SUPPRESSOR CANDIDATES IN
CELL LINE MODELS*

4.1 Introduction

Based on our bioinformatic pipeline results, we have identified TSG and DSDG candidates across the five type B-cell malignancies. The approach enabled to identified total 13 candidates, including eight candidate genes for ALL. Furthermore, a separate ongoing project in the lab used a technique referred to as `methylation mapping` (which aims to map the derivation of all methylation changes in cancer as being induced by proliferation, differentiation or cancer development) identified a similar set of 15 candidates across the five type B-cell malignancies. This included 77% of the candidate identified in chapter 3 and all but one of the eight candidates identified in ALL (Table 4.1). Thus, both approaches identified similar sets of candidates, found the highest fraction of candidates in ALL and all four ALL TSG candidates identified in chapter 3 (*SLC22A15*, *THEM4*, *TTC12*, *MAP9*) were also identified by this second approach.

Disease	Gene	Candidate type ^a	Methylation mapping ^b	Subtype Analysis ^c
ALL	<i>THEM4</i>	TSG	Yes	Yes
ALL	<i>MAP9</i>	TSG	Yes	Yes
ALL	<i>SLC22A15</i>	TSG	Yes	Yes
ALL	<i>TTC12</i>	TSG	Yes	Yes
ALL	<i>DTD1</i>	TSG	Yes	No
ALL	<i>NPY</i>	DSDG	Yes	Yes
ALL	<i>BMP2</i>	DSDG	Yes	Yes
ALL	<i>CTGF</i>	DSDG	No	Yes
ALL	<i>ZNF423</i>	DSDG	Yes	Yes
ALL	<i>SMIM3</i>	DSDG	Yes	No
ALL	<i>CMTM2</i>	DSDG	Yes	No
CLL	<i>SNX18</i>	DSDG	Yes	Yes
MCL	<i>SOX11</i>	DSDG	Yes	Yes
MCL	<i>MFHAS1</i>	DSDG	No	Yes
PCNSL	<i>FSCN1</i>	TSG	Yes	Yes
PCNSL	<i>STMN4</i>	DSDG	Yes	No
PCNSL	<i>GNG7</i>	TSG	Yes	No
PCNSL&DLCL	<i>SDK2</i>	TSG	NA ^d	Yes ^d

Table 4.1 Identified candidate functional genes

^aTSG; tumour suppressor candidate, identified as having disease specific hypermethylation and loss of expression. DSDG; disease specific dependency gene candidate, identified as having disease specific lack of acquired methylation and increased expression

^bIdentified by the methylation mapping approach

^cSee chapter 3

^dAs PCNSL and DLBCL are highly similar diseases, a combined analysis was used in chapter 3 which identified *SDK2*. However, a combined analysis was not used in the methylation mapping study.

4.1.1 Lentiviral system for a gene-delivery

Lentiviruses belong to the Retroviridae viral family and are distinguished by their utilization of viral RT (reverse transcriptase) and IN (integrase) to effectively integrate viral genetic material into the host genome. Lentiviral vectors have been established as an efficient technique for delivering specific genes into target cells [223]. Using lentiviral vectors for gene delivery has a number of benefits; it can be used in both dividing and non-dividing cells, it can infect wide-range cell types (including cell lines that are traditionally hard to transfect), it stably integrates into the host genome, resulting in long-term expression of gene of interest and has a high packaging capacity (up to 10kb) [225]. In addition, the specific expression vector to be used (pSIN-SIEW) includes an internal ribosomal entry site (IRES), downstream of the Gateway cloning site. This allows the expression of the reporter gene GFP and the gene of interest simultaneously from the same transcript, but as separate proteins (Figure 4.1). This enables the identification of successfully transduced cells using flow cytometry and facilitates monitoring of the transduced population over time by monitoring the fraction of GFP positive cells. The lentiviral vector system was used in this study to investigate the potential functional role of candidates in cell lines models.

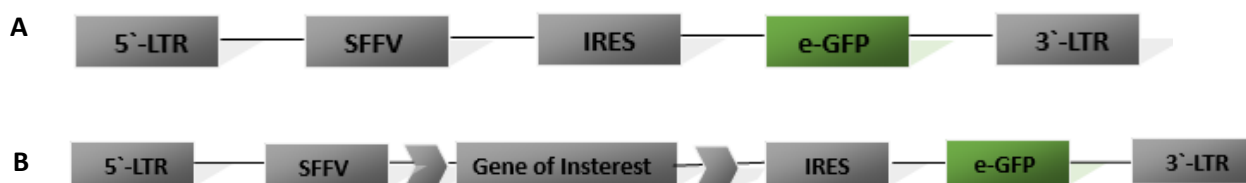


Figure 4.1. Schematic diagram of main features of expression vector.

A) Empty control vector (pSIN-SIEW), B) pSIN-SIEW- gene of interest (*SLC22A15*, *THEM4*, *TTC12* or *MAP9*). pSIN-SIEW vector with the stated gene of interest cloned at the Gateway

cloning site. SFFV: spleen focus-forming long terminal repeat (regulating the expression of the gene of interest); IRES, internal ribosome entry site (allows for the concomitant transcription of the eGFP along with the gene of interest in a bicistronic messenger RNA,); eGFP, enhanced green fluorescent protein; triangles represent the Gateway recombination signals AttB.

4.2 Aims of The Chapter

- Functional assessment of four ALL-specific TSG candidates to determine if their re-expression results in negative impacts on growth/survival of ALL cells.
- Investigation of the specificity of any identified impacts of TSG candidate re-expression in cell line derived from ALL versus other B-cell malignancies.

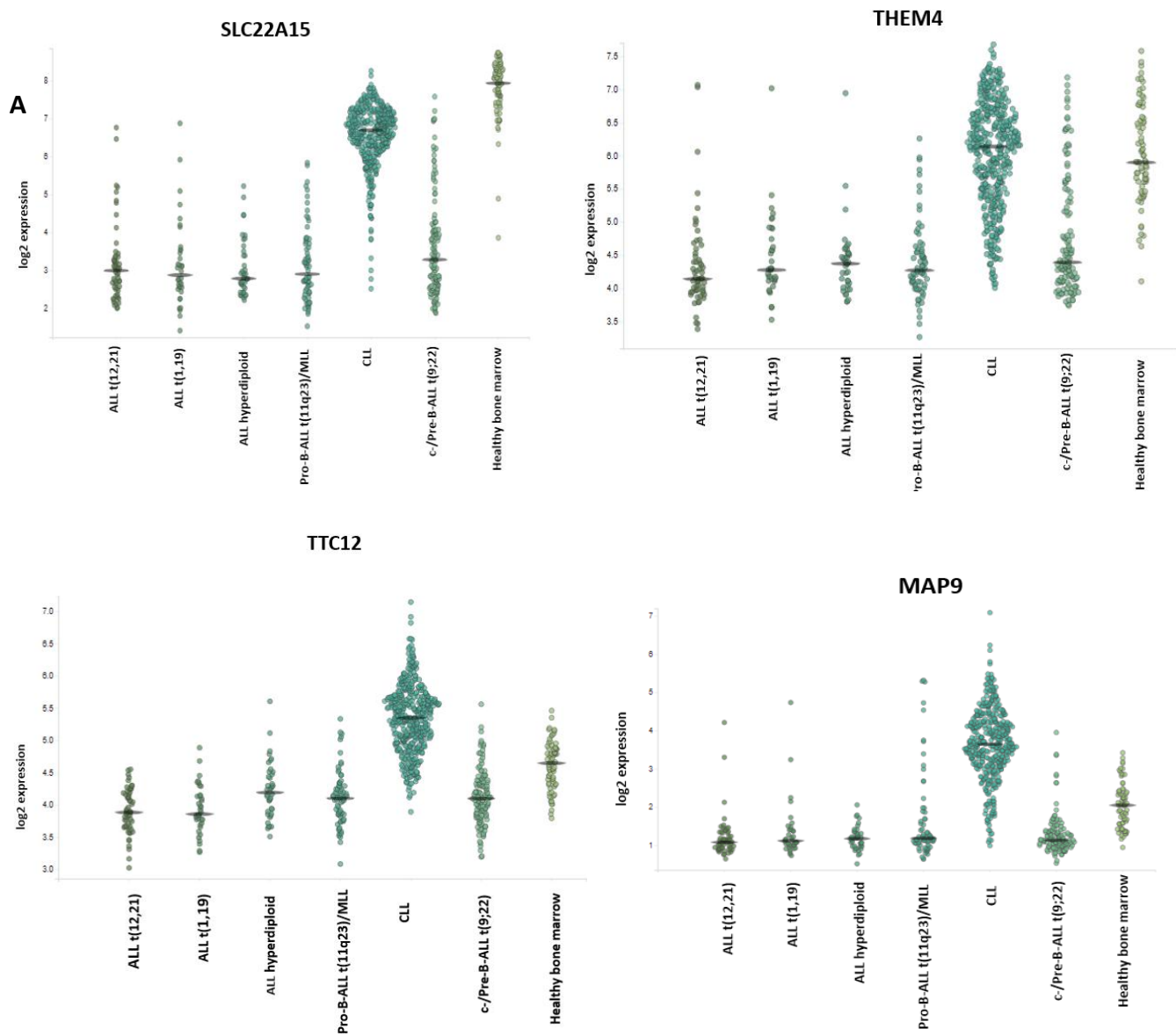
4.3 Functional Assessment

4.3.1 Expression of TSG candidates

As the TSG candidates were identified as relevant for ALL as a whole (i.e. for all/most ALL genetic subtypes) a panel of cell lines was chosen that included representatives of multiple different cytogenetic ALL subtypes. To confirm lack of expression, all candidate genes expression were confirmed across the different subtypes of ALL in primary samples (Figure 4.2A) and confirmed in the cell lines to be used by qRT-PCR, (Figure 4.2 B). The analysis of expression in primary samples confirmed the low expression in all ALL subtypes included in the dataset derived from the leukaemia MILE Study [295] with higher expression being seen in normal bone marrow and CLL. As can be seen in Figure 4.2-B, there was no detectable *THEM4* and *TTC12* transcript in any of the five ALL cell lines and there were no detectable *SLC22A15* mRNA transcript on four of five ALL cell lines (with low level expression detectable in Rch-acv (describe in section 4. 4,5)). In contrast to the other TSG candidates, *MAP9* was clearly detected at the RNA level in some of the ALL cell lines. Three of five ALL cell lines, PreB697, Rch-acv and SEM expressed

MAP9 transcript. This is also consistent with expression reported on R2 database [296].

There was no detectable MAP9 mRNA expression on NALM6 and Reh cell lines.



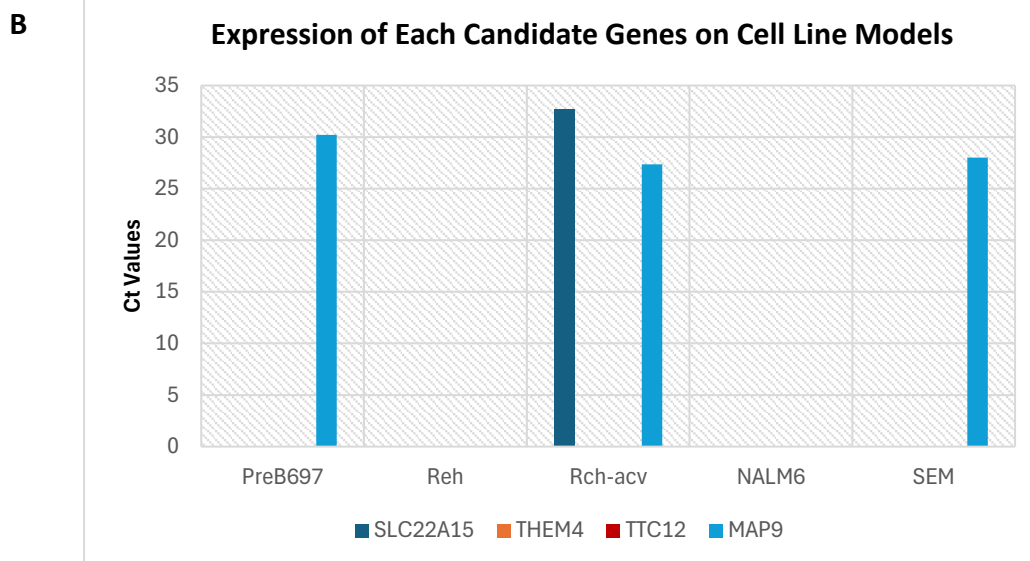


Figure 4.2. Expression pattern of each candidate genes on both primary sample and model cell lines.

A) Expression level of SLC22A15, THEM4, TTC12 and MAP9 across the different cytogenetic subtypes in primary ALL samples are lower than normal bone marrow and CLL. B) Expression in the cell line models largely corresponds with primary samples, with the exception of MAP9, where Ct values of expression of the transcript was detectable in PreB697, Rch-acv and SEM cell lines. Also, SLC22A15 expression only was detectable on Rch-acv cell line. Data shown here are representative of single qPCR experiment. Expression of the housekeeping gene GAPDH was used for normalisation.

4.3.2 Cloning of the Candidate TSG Into the Lentiviral Expression Vector

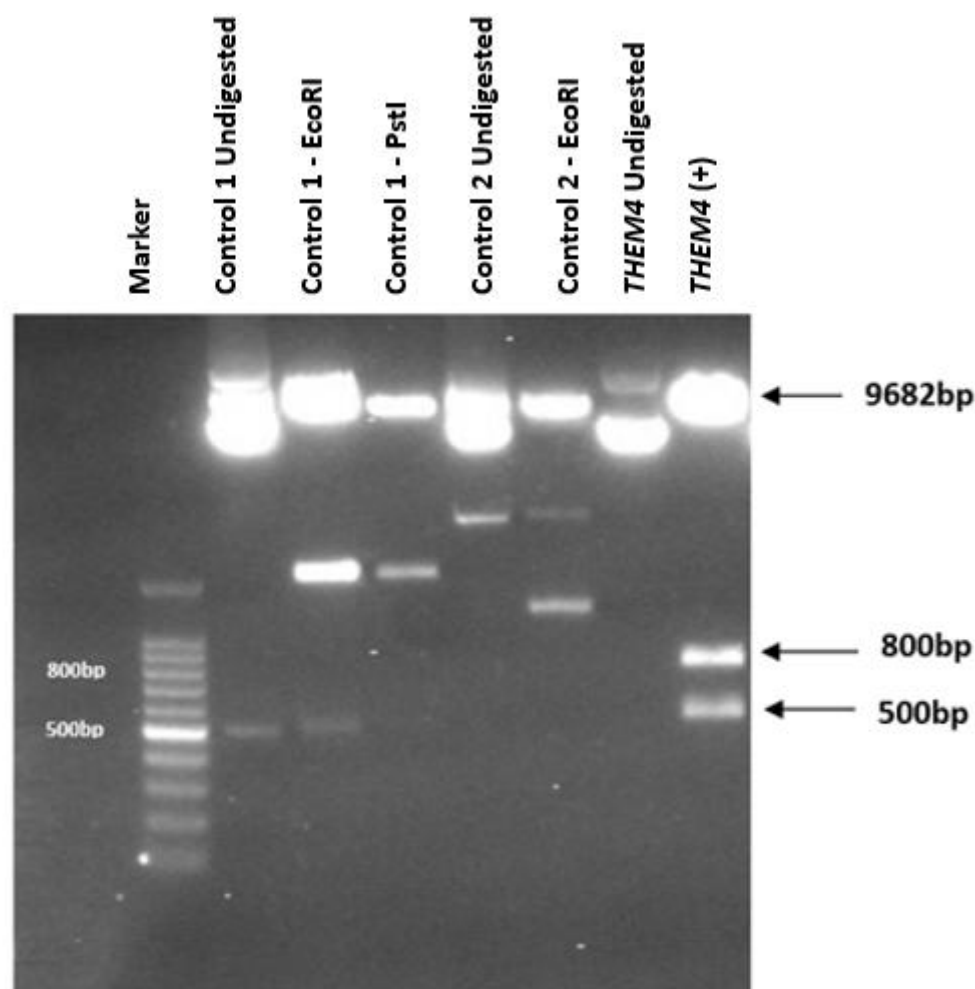
To ensure accurate long and the correct orientation of the derived THEM, SLC22A15, TTC12 and MAP9 clones, restriction digests were carried out to assess the size and orientation of the insert present in the clone. Expected band sizes for each vector is shown in Table 4.2. And the resultant fragments after digestions are shown in Figure 4.3. Based on this analysis clones could be identified for all four candidates in which the expected insert was present and also cloned in the correct orientation. Furthermore, clones with correct digestions patterns were sequenced to ensure that full sequence

matched the known genomic sequence. The sequencing data confirmed that the sequence matched the reference sequences for the entire coding regions for all candidates (NM_053055.5, NM_001318533.1, NM_001039580 and NM_018420, THEM4, TTC12, MAP9 and SLC22A15 respectively).

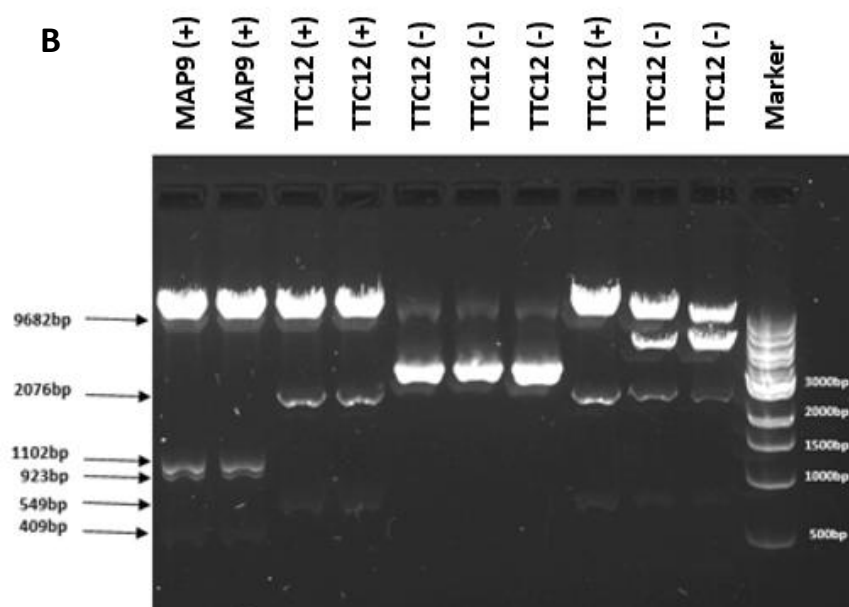
Vectors and Expected Band Size			
pSIN-THEM4 EcoRI	pSIN-TTC12 EcoRI	pSIN-MAP9 EcoRI	pSIN-SLC22A15 PstI
500bp	549bp	409bp	87bp
800bp	2076bp	923bp	167bp
9682bp	9682bp	1101bp	460bp
		9682bp	615bp
			792bp
			1328bp
			8366bp

Table 4.2 Expected bands size for the EcoRI and PstI digest of pSIN-GTW vectors.

A) THEM4



B



C

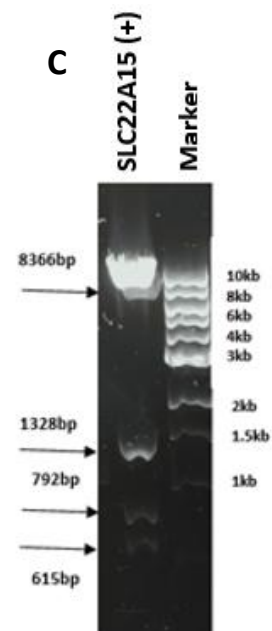


Figure 4.3. An electrophoresis results of pSIN-TEHM4/SLC22A15/MAP9/TTC12 after DNA digestion.

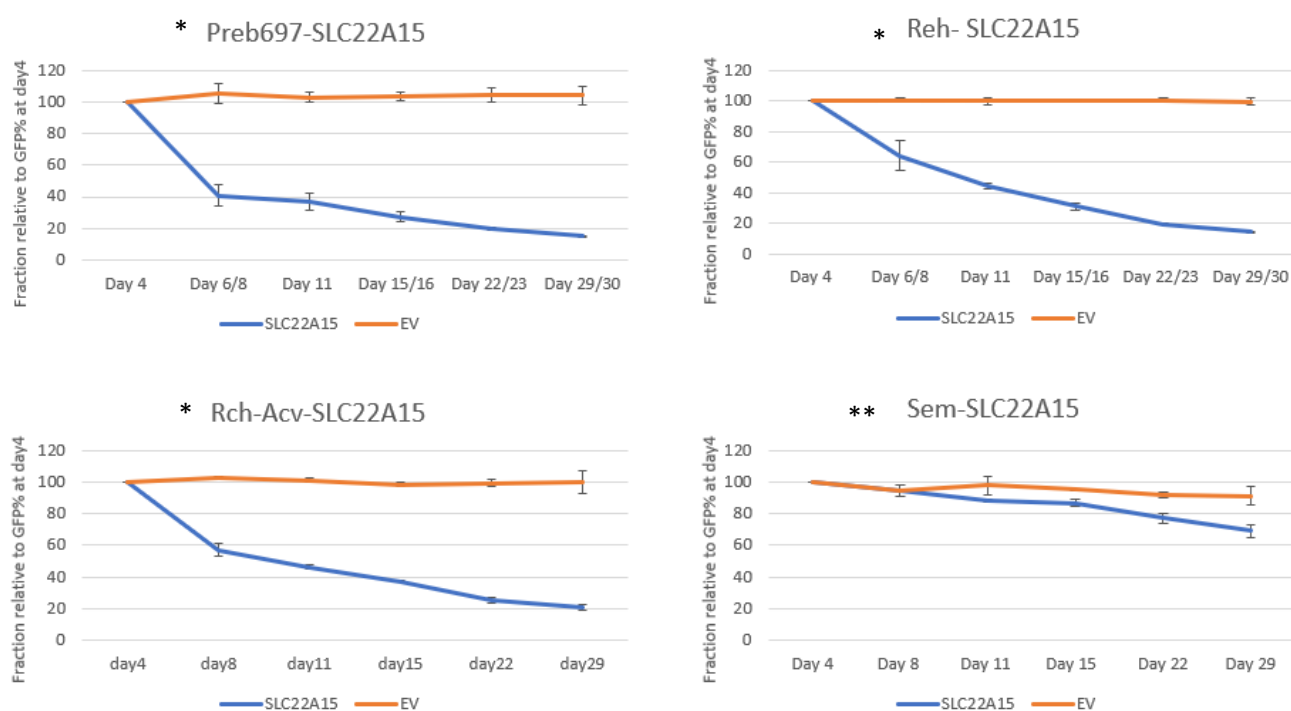
A) Anticipated band size was observed after digest with *EcoRI* for THEM4. Numbers above gel indicates sample on each line. Wells 1-5 control samples with/out digest. Well 6 was pSIN-THEM4 without digest and Well 7 was pSIN-THEM4 with *EcoRI* digest. B) Anticipated band size was observed after digest with *PstI* for MAP9 and TTC12. Well 1 and 2 were pSIN-MAP9 after digestion. Well 3, 4 and 8 were pSIN-TTC12 after digestion. Well 5,6,7,9 and 10 were not correct colonies for pSIN-TTC12. C) Anticipated band size was observed in well 1 after digest with *PstI* for pSIN-SLC22A15. 460bp, 167bp and 87bp were present although they were feint and not clear on the figure. 100bp marker was used for gel electrophoresis to visualise THEM4 anticipated bands. 1kb ladder was used as marker to visualise MAP9, TTC12 and SLC22A15. (+) indicates right clone which had predicted size of bands, (-) indicates wrong clone which had not predicted size of bands. Clone 1 indicates HOXA4, Clone 2 indicates MSC.

4.4 Functional Validation of SLC22A15

4.4.1 Impact of SLC22A15 gene re-expression following transduction of ALL cell lines

Five ALL cell lines (PreB697, Reh, SEM, Rch-acv and NALM6) were used to assess the impact of *SLC22A15* re-expression following transduction with the pSIN-SLC22A15 expression vector. All cell lines were transduced with pSIN-SLC22A15 (which expresses *SLC22A15* and GFP), and with the pSIN-EV control vector (which only expresses GFP and is referred to as “empty vector”) as a control. Any selection against cells transduced with pSIN-SLC22A15 (and thus re-expressing *SLC22A15*) could then be monitored by assessing the fraction of GFP-positive cells in the transduced cell population (using flow cytometry), i.e. if the re-expression of *SLC22A15* resulted in reduced growth or survival, GFP levels would be expected to fall in the pSIN-SLC22A15 transduced cells (while remaining stable in the control pSIN-EV transduction). The fraction of GFP positive cells was determined at days 4, 8, 11, 15, 22 and 29 after transduction. Day 4 was the first time point assessed as previous studies in the lab with the same expression vector had shown that expression typically peaks on this day, even for genes that have a detrimental effect on cell growth/survival. Roughly 80% of loss of GFP was seen on TCF-PBX cytogenetic cell lines models, PreB697 and Rch-acv, also ETV6-RUNX1 cytogenetic cell line model Reh by the

end of period assessed. There was statically significant difference between pSIN-*SLC22A15* transduce cell population and control group ($p < 0.001$). Cell lines can be categorised as more sensitive to expression of *SLC22A15* than other two ALL model cell lines. 50% of cells transduce with the *SLC22A15* expressing vector was lost from culture by post-transduction day 29 so statistically significant reduction of GFP loss was observed on NALM6 cell line. In regard od SEM cell line result as KMT2A-AFF1 genetic subtype of ALL , around 30% of pSIN-*SLC22A15* infected cell line population was lost during the time course. There were slightly significant differences between pSIN-*SLC22A15* transduce cell population and control group ($p < 0.001$). SEM and NALM6 were less sensitive cell lines to *SLC22A15* expression compared to other three ALL cell lines assessed. Overall, *SLC22A15* expression levels (assessed by fraction of GFP positive cells) significantly dropped throughout the time course in all cell lines assessed, suggesting that *SLC22A15* expression show a detrimental impact on cell growth/ survival on ALL cell lines assessed. (Figure4.4)



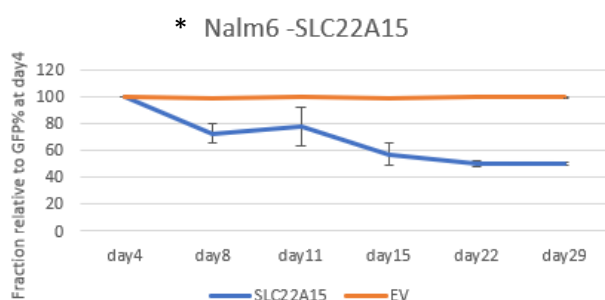


Figure 4.4. Results of Re-expression of SLC22A15 in ALL cell lines.

Five ALL cell lines (as indicated) were transduced with pSIN-SLC22A15 and separately with the empty vector control and the fraction of GFP positive cells assessed over time. Results are expressed as a fraction relative to the percentage of cells in the transduced population at day 4 post-transduction. As expected, the fraction of positive cells remains essentially constant over time in the empty vector control transductions in all five cell lines. By contrast, the fraction of GFP positive cells falls over time in the pSIN-SLC22A15 transduced populations in all five cell lines, with more rapid loss seen in three cell lines (PreB697, Reh, Rch-Acv) and a lower rate of decline seen in NALM6, and SEM. Data shown in here produced from two independent experiment results and bars represent standard error between experiments. Statistical analyses were performed with unpaired two-tailed t-test and * indicates $p < 0.001$ and ** indicates $p < 0.05$.

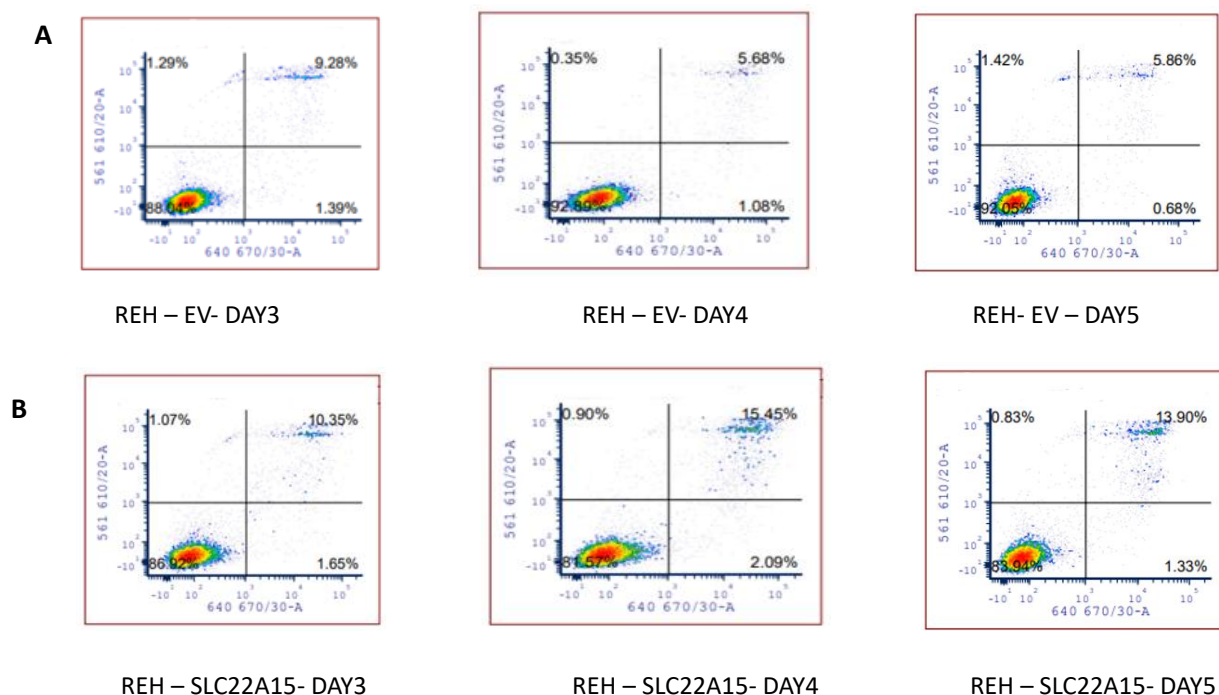
4.4.2 Cell survival impact of re-expression of SLC22A15 on ALL cell lines

4.4.2.1 Assessment of apoptosis using binding of annexin V

The above analysis indicates that re-expression of SLC22A15 had a negative impact on cell growth or survival across multiple cell lines. To begin to assess the mechanism by which this occurs staining with Annexin V/PI was used to determine if re-expression of SLC22A15 was associated with increased rates of cell death. The PreB697 and Reh cell lines were transduced with pSIN-SLC22A15 and pSIN-EV. On day 3, 4 and 5 post-transductions, cells were harvested and stained with APC-annexin V and PI kit (BioLegend Cat No: 640932). The cells were then analysed by flow cytometry to allow assessment of PI and annexin V in the whole cell population which refer both transduce and non-

transduce cells. Meantime, GFP level assessed on same day by flow-cytometer to verify GFP decreasing reflect cell death on cell population transduced with *SLC22A15* expressed vector. Based on this analysis cells were defined as non-apoptotic (Annexin V-/ PI-), early apoptosis (Annexin V+/PI-) or late apoptosis (Annexin V+/PI+) (Figure 4.5). Total cell death was calculated by adding early apoptosis cell population percentage (Annexin V+/PI-) to late apoptotic cells percentage (Annexin V+/ PI+).

The results, summarised in Figure 4.6, demonstrate that *SLC22A15* expression in both cell lines exhibited a small, non-statistically significant, increase in cell death at day3, which increased and reached statistical significance at days 4 and 5 after transduction when compared to control group that expressed just GFP (* $p < 0.05$ and ** $p < 0.005$) (Figure 4.6-A for Reh cell line and B for PreB697 cell line). In addition, as can be seen on graphs, the fraction of GFP positive cells reduced in both cells lines across the three days assessed in the *SLC22A15* re-expressing cells, while GFP expression was stable in control vector transduced populations over the course of time (Figure 4.6-A for Reh cell line and B for PreB697 cell line).



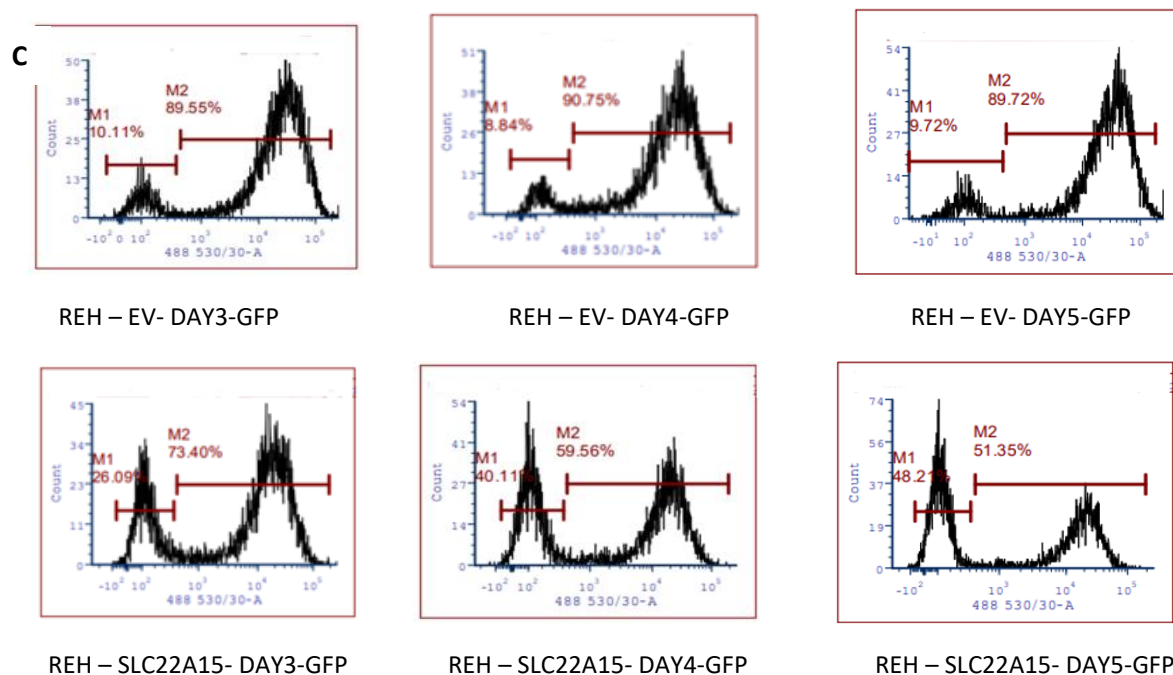


Figure 4.5. Representative flow cytometer results of Annexin V/PI and GFP in Reh cell line after transduction with psin-SLC22A15.

A and B) Annexin V and PI staining was performed at days 3, 4, 5 post-transduction with either transduction with pSIN-SLC22A15 or control (pSIN-EV) vectors. The X axis represents APC-Annexin V and Y axis represents PI staining. Intact cells are located in the lower left quadrant, while early (lower right quadrant) and late (upper right quadrant) are positive for Annexin V or double positive (Annexin V/PI), indicating early and late apoptosis, respectively. As can be seen from the profiles in the pSIN-SLC22A15 transduced Reh cell the fraction in the two right hand quadrants is increased at days 4 and 5, indicating increased cell death following SLC22A15 re-expression. In contrast, levels remain essentially stable in at days 4 and 5 in the pSIN-EV control vector transduced cells. C) Stable GFP expression was observed on control group at post-transduction day3, 4 and 5. However, the percentage of GFP of cell population expressed SLC22A15 on was 73.40 on day3 after transduction (D) and it was dropped to 51.35 at day5 (D). Histograms (with the X axis assessing GFP intensity) come from single experiments after analysis by FCS Express Program. (The M1 marker illustrates the position of non-transduced cells, while the M2 marker indicates GFP expressing cells).

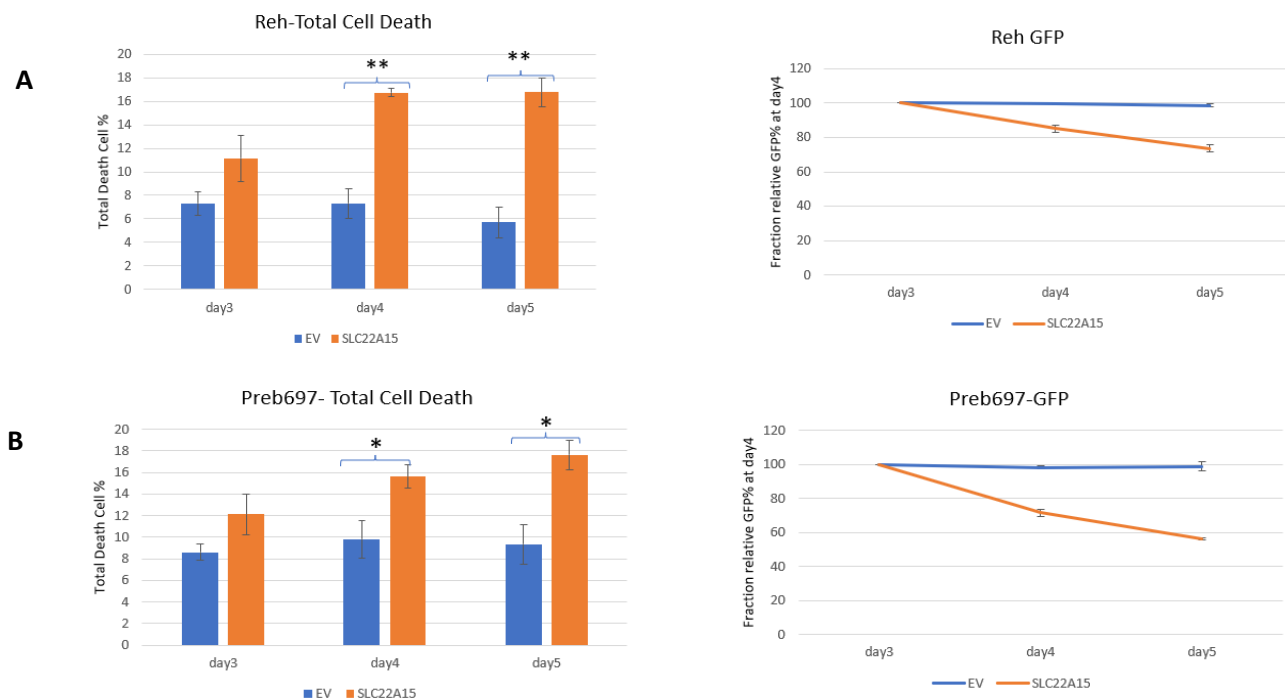


Figure 4.6. Cell death assessed by annexin V/PI staining following re-expression of SLC22A15 in Reh and PreB697 cell lines.

Binding of Annexin V/PI was used to assess cell death at days 3, 4 and 5 post-transduction with the pSIN-SLC22A15 or pSIN-EV vectors in Reh (A) and PreB697 (B) cell lines. Evidence for increased cell death is initially seen in both cell lines at day 3 and increases to statistically significant levels (* $p < 0.05$ and ** $p < 0.005$) on day 4 and 5 in both cell lines. Concomitant with the increase in cell death, both cell lines also demonstrated decreasing levels in the fraction of GFP positive cells in the pSIN-SLC22A15 transduced cell populations. Data comes from three independent experiment and bar represent standard error within experiments.

4.4.2.2 Assessment of Caspase 3/7 activity following SLC22A15 re-expression

The above results suggest that re-expression of SLC22A15 was associated with induction of cell death. To determine if this was associated with caspase activation, Caspase 3/7 activation was assessed in PreB697 and Reh cell lines at day 3, 4 and 5 following transduction with pSIN -SLC22A15 or EV control. In addition, the GFP-positive fraction in each cell population was assessed pSIN-SLC22A15 over the same time period.

This analysis demonstrated a very clear Caspase 3/7 activation triggered following pSIN-SLC22A15 transduction compared with the EV control group at days 3, 4 and 5 in both cell lines. Results closely mirrored those in the Annexin V staining experiments above. Caspase activation reached statistical significance at day3 (unlike the annexinV staining), however, there was a similar increase in the extent of caspase activation as days 4 and 5. In addition, the fraction of GFP positive cells reduced specifically in the *SLC22A15* expressing populations (Figure 4.7). Overall, these results demonstrated that re-expression of *SLC22A15* resulted a clear activation of caspase activity in a time scale that closely mirrored the increase in cell death detected by annexin-V/PI staining.

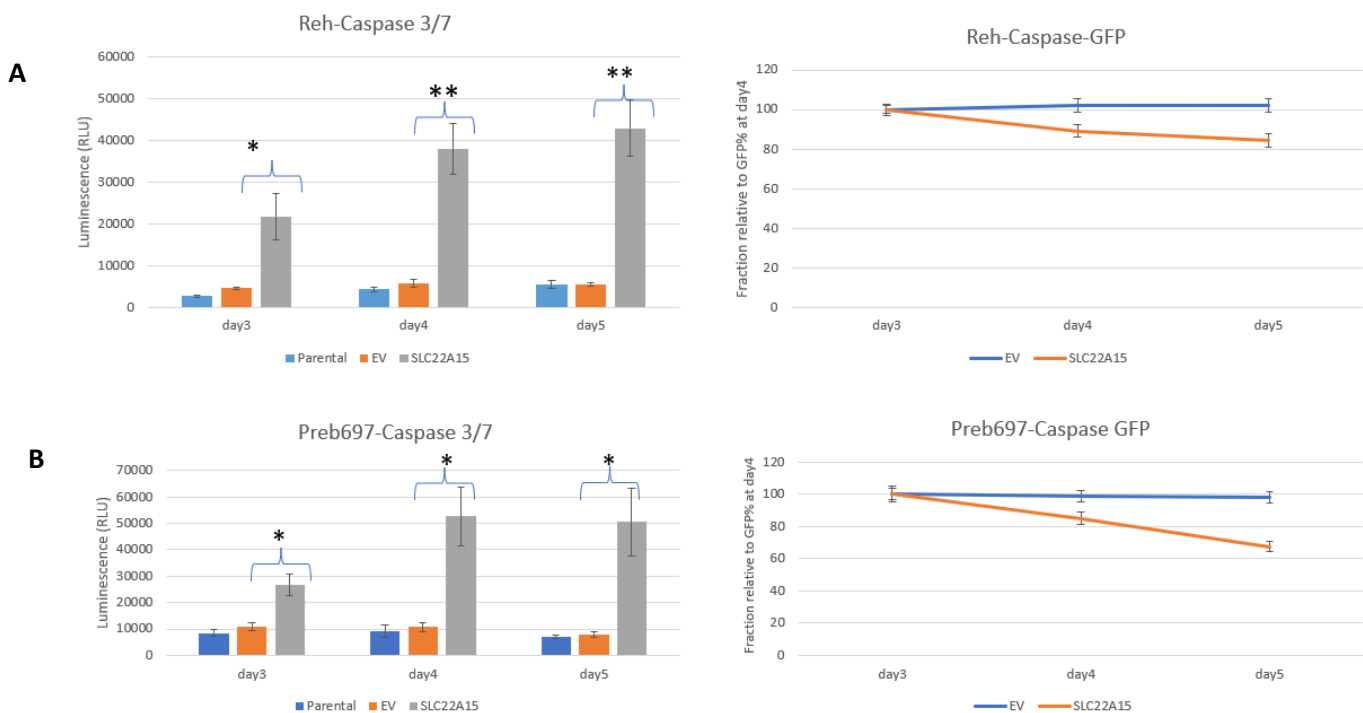


Figure 4.7. Caspase 3/7 activation after SLC22A15 re-expression.

Caspase-glo 3/7 was performed to assess caspase activation after transduction with pSIN-SLC22A15, compared with the EV control, at days 3, 4 and 5 post-transduction. Caspase 3/7 induction was detectable by day3 and reached a maximum at day4/5 after re-expression of SLC22A15 when compared to control group that expressed just GFP in Reh (left panel of A) and PreB697 (left panel of B) cell lines (* $p < 0.05$ and ** $p < 0.005$). In both cell lines the fraction of GFP-positive cells fell specifically in the pSIN-SLC22A15 transduced

populations. Data comes from three independent experiments and bars represent standard error between experiments. GFP results were normalised at day3.

4.4.2.3 Assessment of the ability of caspase inhibition to block the induction of cell death following SLC22A15 re-expression

Based on the Annexin V/PI staining, total cell death was significantly induced by day4 and day5 after re-expression of *SLC22A15* however, there was little evidence of any increase in early apoptotic cells (i.e. annexin-V+/PI-) during the experimental time period (day3, 4 and 5). The lack of an early apoptotic population may suggest a mechanism of cell death other than apoptosis, while the corresponding caspase activation is more consistent with apoptotic cell death. Therefore, we wanted to verify whether the cell death observed following *SLC22A15* re-expression was caspase dependant or not. Therefore, the combined caspase 3/7 inhibitor Ac-DEVD-CHO was used to inhibit caspase 3/7 activation after transduction, to determine if this would overcome the negative impact of *SLC22A15* expression in ALL cell lines. Transduced cell populations were treated with Ac-DEVD-CHO at post-transduction day2 (dose for Ac-DEVD-CHO was derived from previous publication [297] and as shown in figure 4.8 resulted in complete inhibition of caspase 3/7 activity). Then at day3, 4 and 5 cells were washed with PBS, to remove any remaining Ac-DEVD-CHO, and caspase 3/7 activation was measured with Caspase-Glo 3/7 Assay. In addition, GFP was measured in the transduced population to determine if caspase inhibition prevent loss of SLC22A15/GFP expressing cells. Untreated parental cells and transduced cells (both pSIN-SLC22A15 and with pSIN-EV) without inhibitor, pSIN-SLC22A15 were used as controls. Assessment of caspase activation demonstrated that Ac-DEVD-CHO treatment was a highly effective inhibitor of caspase 3/7 activation in both empty vector infected and pSIN-SLC22A15 infected cells (Figure 4.8 left upper/below) in Reh and PreB697 cell lines. Even at day 5 post transduction (72 hours after treatment with Ac-DEVD-CHO), caspase activity in both the pSIN-SLC22A15 and EV transduced populations was essentially undetectable and below the background level of untreated parental control cells (Figure 4.8). As expected, caspase 3/7 activation was readily detectable by day 3 in the cell populations which were transduced with pSIN-SLC22A15, while no increase in caspase

activity was seen in the untreated EV vector transduced cells. However, suppression of caspase 3/7 activity in treated pSIN-SLC22A15 infected cell group did not alter the rate of loss of GFP-positive cells in either cell lines. The fraction of GFP reduction in treated and untreated pSIN-SLC22A15 transduced group was indistinguishable. Therefore, these results suggest that caspase 3/7 activation is not required for induction of cell death following *SLC22A15* re-expression in the cell lines tested.

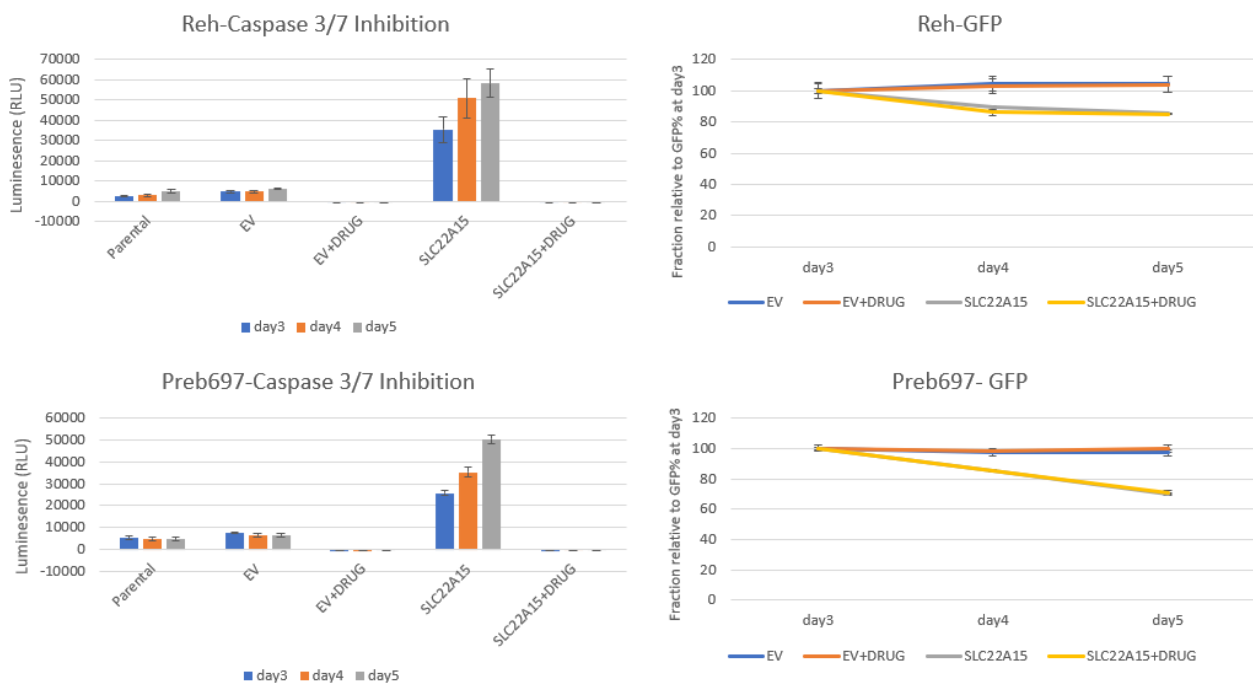


Figure 4.8. Caspase 3/7 inhibition does not impact of GFP-positive cells after transduction with pSIN-SLC22A15.

Caspase 3/7 activity was measured post-transduction at days 3, 4 and 5, either with or without treatment with the caspase inhibitor Ac-DEVD-CHO (at day2). The inhibitor treatment fully suppressed caspase activation at all time points (left hand panels) graph on Figure. Inhibition of caspase activity had no detectable impact on loss of GFP-positive cells following re-expression of *SLC22A15* (right hand panels), as rate of loss of GFP was indistinguishable between populations treated/not treated with the caspase inhibitor. Data shown here comes from three independent experiment and bars reflect standard error between experiments. GFP level was normalised at day3. Parental represents

untreated, EV (control group) represent mock treated, EV+DRUG represent mock treated with inhibitor, SLC22A15 represent testing group, SLC22A15+DRUG represents testing group with inhibitor.

4.4.3 Assessment of the impact of re-expression of SLC22A15 on cell proliferation

To assess if re-expression of *SLC22A15* in ALL cell lines impacted cell proliferation, proliferation assays were performed with the eFluor 450 dye in the Reh and PreB697 cell lines. Staining was performed at day 3 post-transduction and used to assess any changes in cell proliferation at days 5, 6, 10, 12 and 17 (Figure 4.9). As eFluor 450 dye permanently stains cells, the intensity of staining only reduces when cells divide. Thus, as can be seen in Figure 4.9, staining of the cell population steadily reduces in control cells until it reaches background levels (at day 17). Staining of both EV and pSIN-SLC22A15 transduced populations was very similar across the time course. There was very minor cell delay at day 6, 10 and 12 for Reh, at day 5, 6 and 10 for PreB697 cells compared with the control group. However, the extent of the difference was minimal and there was no evidence of a cell cycle arrested population (which would have remained at the right-hand side of the histograms). Thus, the assay indicates that re-expression of *SLC22A15* did not significantly impact the rate of proliferation or induce cell cycle arrest. This implies that the loss of SLC22A15/GFP expressing cells identified above was due to induction of cell death and was not related to cell cycle inhibition.

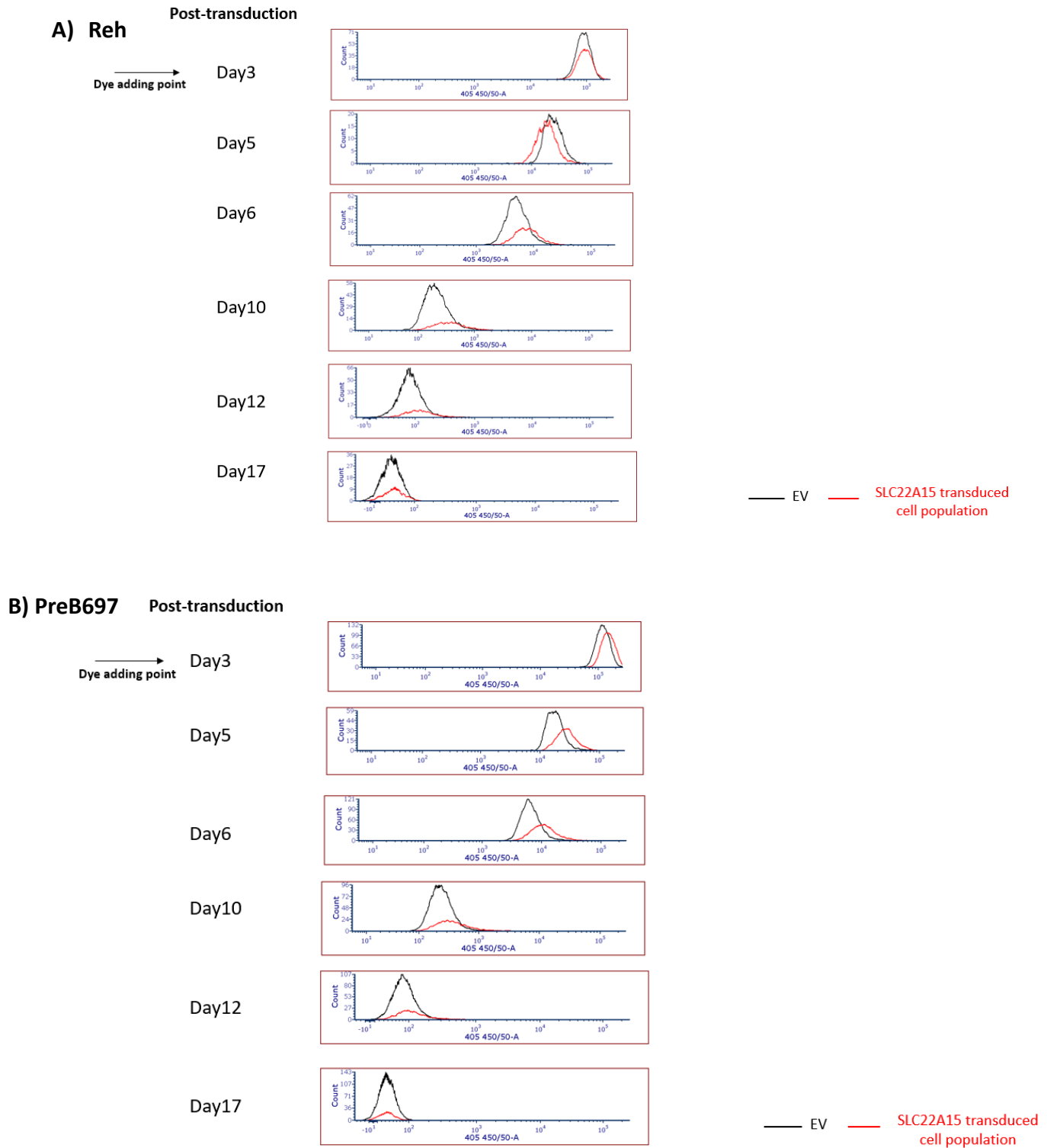


Figure 4.9. SLC22A15 re-expression does not impact the rate of cell proliferation in ALL cell lines.

e-Fluore 450 staining was carried out at day 3 post-transduction. On the same day flow cytometry was performed to record initial staining intensity levels. As can be seen on

histograms for Reh (A) and PreB697 (B) cell lines, there was no major differences in staining intensity (X axis) between SLC22A15 expressing and control populations over the time period assessed in either cell line. Represented data comes from single experiment. Black line represents control population (EV) and the red line represents SLC22A15 expressing cell populations.

4.4.4 Impact of re-expression of SLC22A15 on lymphoma cell lines

To investigate whether the impact of *SLC22A15* expression is specific for ALL or not, four lymphoma cell lines were transduced with the *SLC22A15* expressing and control vectors. The fraction of GFP positive cells was then measured at day 4, 8, 11, 15, 22 and 29 by flow-cytometry. As shown in figure 4.10, all four lymphoma cell lines demonstrated a gradual reduction in GFP levels across the time course. Two of the cell lines, Jiyoye and Oci-Ly-18, showed only comparatively minor changes in the fraction of GFP positive cells, reaching approximately 20% and 33% reductions over the time course respectively, relative to empty vector control cells, and there was statistically significant differences between pSIN-SLC22A15 transduce cell population and control group ($p < 0.05$) (Figure 4.10-A). The other two lymphoma cell lines, Toledo and Daudi, exhibited a slighter greater loss of GFP, reaching approximately 60% by day 29 and there was statically significant difference between pSIN-SLC22A15 transduce cell population and control group ($p < 0.05$).

Comparison of ALL and lymphoma cell lines (Figure 4.10B), showed that ALL cell lines generally exhibited greater sensitivity to *SLC22A15* expression, but the difference was not absolute, with 2/5 ALL cell lines exhibiting rates of loss of the GFP fraction of the population following transduction with pSIN-SLC22A15 similar to the lymphoma cell lines.

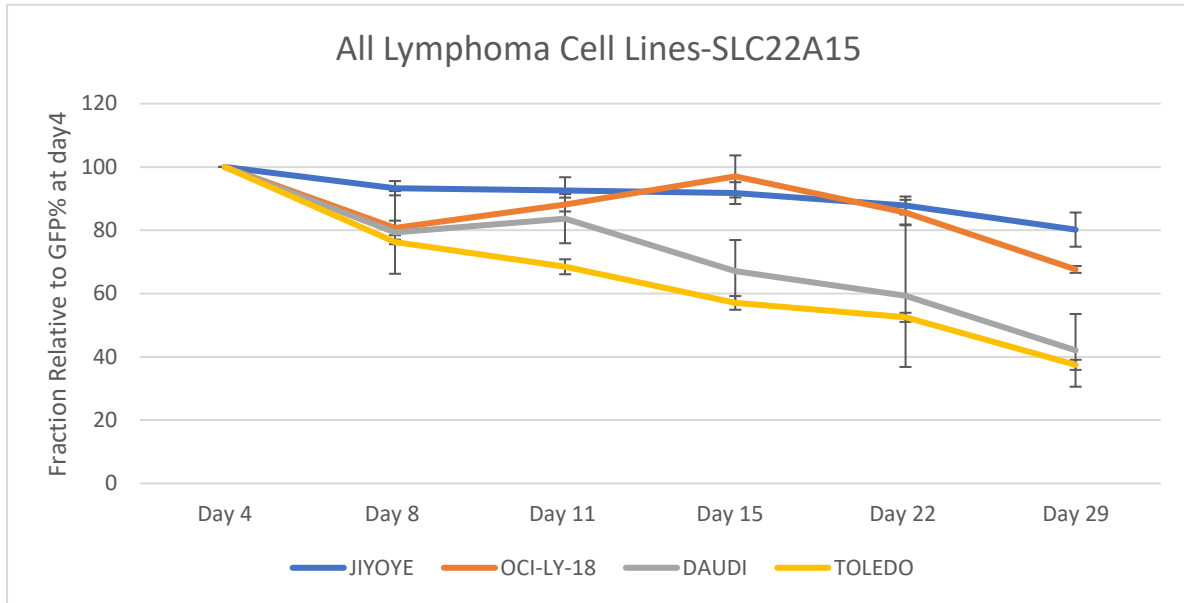
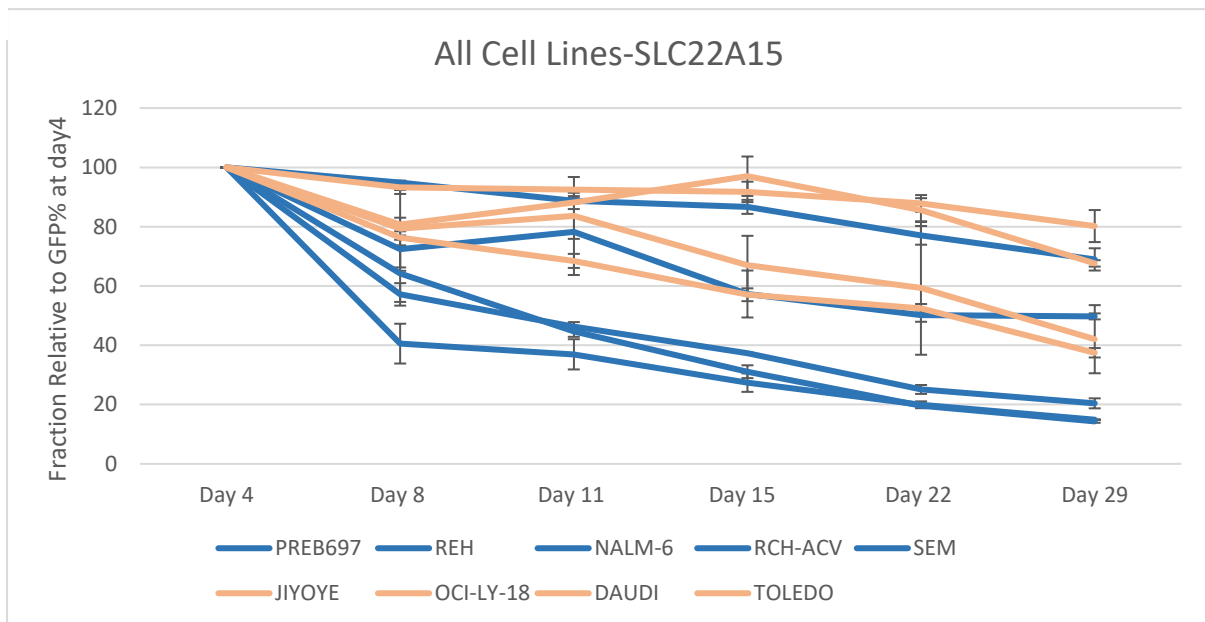
A**B**

Figure 4.10. Impact of re-expression of SLC22A15 on Lymphoma cell lines and comparing results with ALL cell lines.

(A) Lymphoma cell lines were transduced with the pSIN-SLC22A15 and control (EV) vectors then GFP assessed during the time course, at the indicated times. Re-expression of SLC22A15 resulted in a gradual loss of GFP positive cells in the lymphoma cell lines, which was clearest in Daudi and Toledo and more borderline in Jiyoye and Oci-Ly-18. (B) Comparison of the ALL and lymphoma cell lines showed that three ALL cell lines (Reh, PreB697 and Rch-acv) showed sharpest decrease, NALM6, Toledo and Daudi cell lines

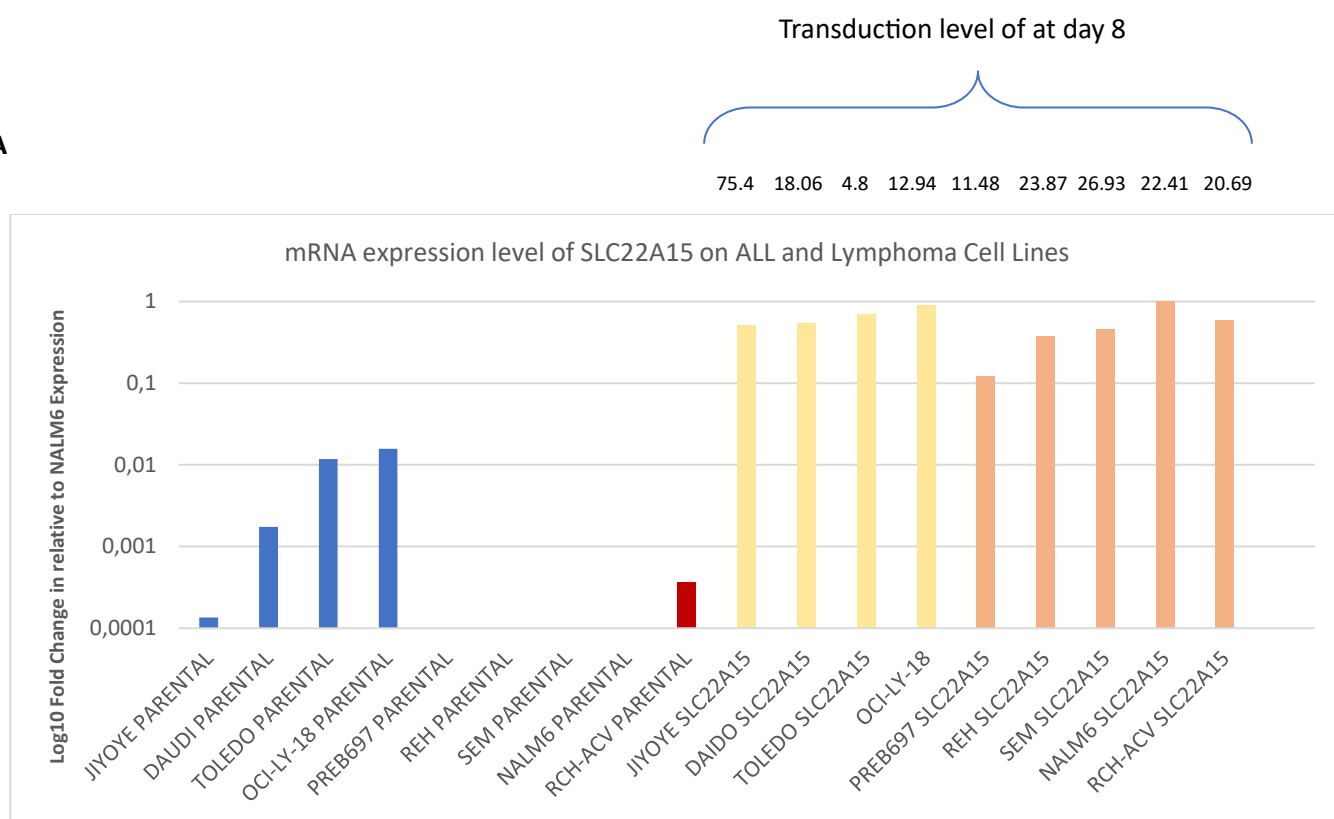
showed medium reduction and SEM, Jiyoye and Oci-Ly cell lines showed slightest decrease of the GFP expressing cell population. Data comes from two independent experiments and bar represent standard error between experiments. GFP level normalised at day4.

4.4.5 Assessment of mRNA-expression of SLC22A15 after transduction in All and Lymphoma cell lines

Responses to re-expression of *SLC22A15* varied across the cell lines. One potential explanation would be related to the level of expression achieved in each of the cell lines. Thus, mRNA expression of *SLC22A15* in parental samples and samples infected with pSIN-*SLC22A15* vector was quantified by qRT-PCR after transductions at day8. If parental (non-transduced)/EV transduced samples expressed *SLC22A15*, this value was subtracted from the total level of expression to specifically identify increased *SLC22A15* expression following transduction with the *SLC22A15* expressing lentiviral vector. Also, corrections were made to account for differences in transduction efficiency between the cell lines (based on fraction of GFP positive cells assessed on the same day as used for RNA extraction) (Figure 4.11-B). This allows comparison of expression level in the successfully transduced cells, independent of the efficiency of transduction in each cell line.

As can be seen from the raw expression analysis in Figure 4.11A, expression levels were similar in both lymphoma and ALL cell lines following transduction (the average in the lymphoma cell lines was slightly higher, although this was not statistically significant) and was clearly greater than the endogenous expression observed in several of the lymphoma cell lines. However, following correction for transduction levels, the highest expression in successfully transduced cells in the Toledo cell line which had approximately double the expression level of the next highest cell line (Oci-Ly-18). While the two highest expressing cell line were lymphoma cell lines, another lymphoma cell line, Jiyoye, was the lowest and there were no consistent differences in *SLC22A15* expression after correction for transduction levels, with lymphoma cell lines overall averaging slightly higher expression than ALL cell lines. To check correlation between level of mRNA expression and fraction of loss of GFP between day4 and day15 post-transduction, linear regression analysis was formed and there was no statically significant ($p > 0.05$ and $R^2 = 0.035$)

A



B

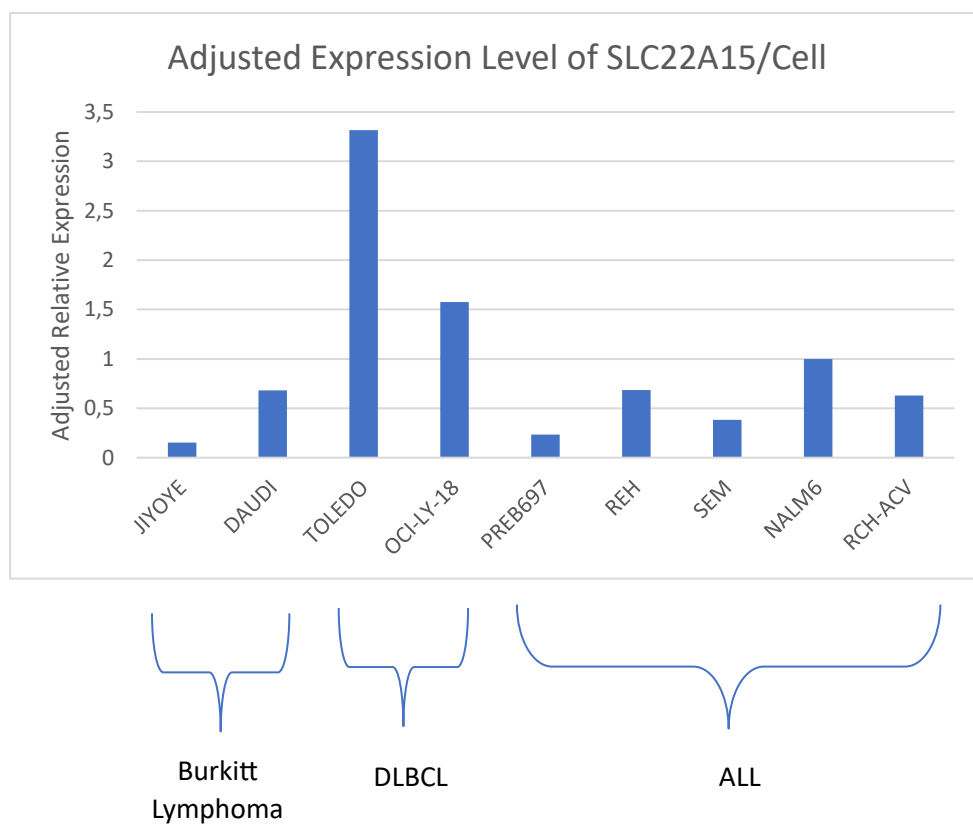


Figure 4.11. SLC22A15 in transduced and parental cell lines.

SLC22A15 transcript level was assessed by qRT-PCR post-transduction at day 8 in pSIN-*SLC22A15* transduced (labelled as *SLC22A15* in figure) and parental cell lines (labelled as parental in figure). For cell line exhibiting positive expression in the parental sample this value was subtracted from the pSIN-*SLC22A15* transduced to obtain the additional expression caused by transduction, although as parental expression levels were generally low this had minimal impact. Results are expressed as a fraction of the highest expressing cell line (NALM6) and the corresponding transfection efficiency at day 8 is shown above the graph (A). (B) The *SLC22A15* expression level adjusted for the fraction of cells successfully transduced in each cell lines, again expressed as a value relative to NALM6. Data shown here comes from single experiment. GAPDH was used as housekeeping control to normalise expression data after qRT-PCR. In figure A- blue colour represent lymphoma cell lines untreated group, red represents ALL cell lines untreated group, yellowish represents testing samples of lymphoma cell lines and pinkish colour represents testing samples of ALL group.

4.4.6 Assessment of protein expression of *SLC22A15* by Western blotting

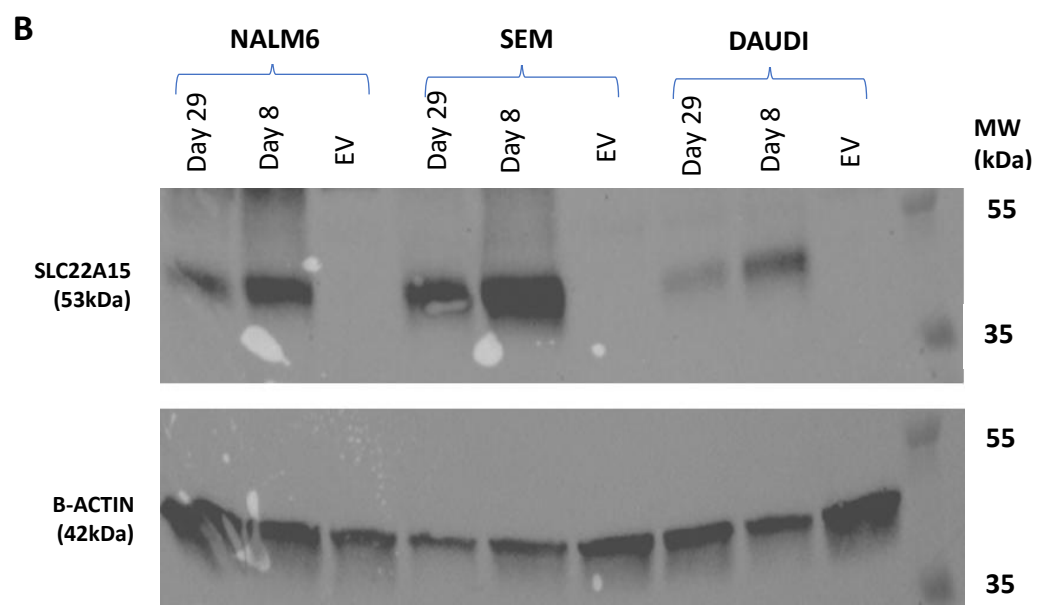
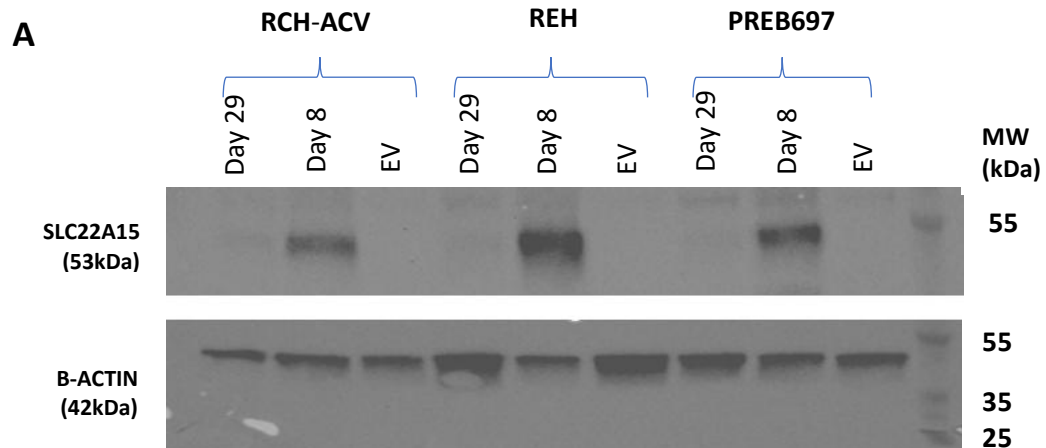
SLC22A15 re-expression at protein level on samples was assessed by western blotting. To show re-expression of *SLC22A15* and its reduction during the time course, post-transduction day 8 and day 29 samples of each cell line as well as empty vector infected samples (pSIN-EV) as control group were used for western blot experiment (Figure 4.12 – A-C). According to the results, each ALL-cell lines showed clear reduction of *SLC22A15* protein from day4 to day29. This confirms that loss of GFP corresponded with loss of *SLC22A15* protein in transduced ALL cell line (Figure 4.12 -A, B). For the three ALL cell lines that showed rapid loss of *SLC22A15*/GFP expressing cells by flow cytometry *SLC22A15* expression was found to be completely or nearly completely absent by day 29 (Figure 4.12A). Even for ALL cell lines where loss of *SLC22A15*/GFP expressing cells by flow cytometry was less rapid the loss of protein expression at day 29 is pronounced.

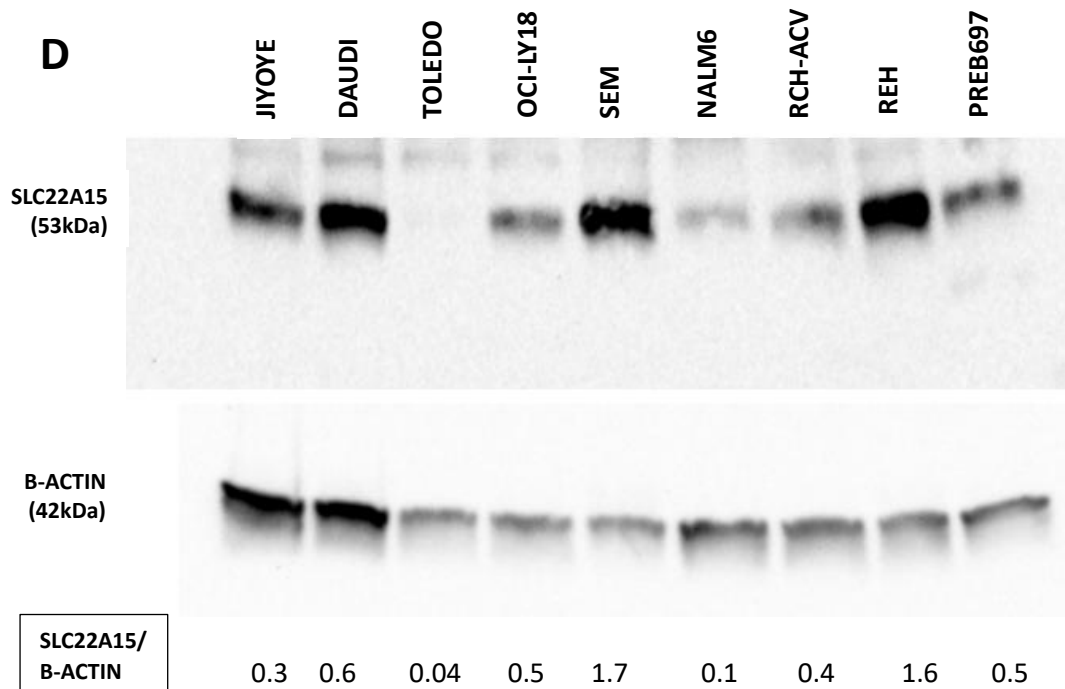
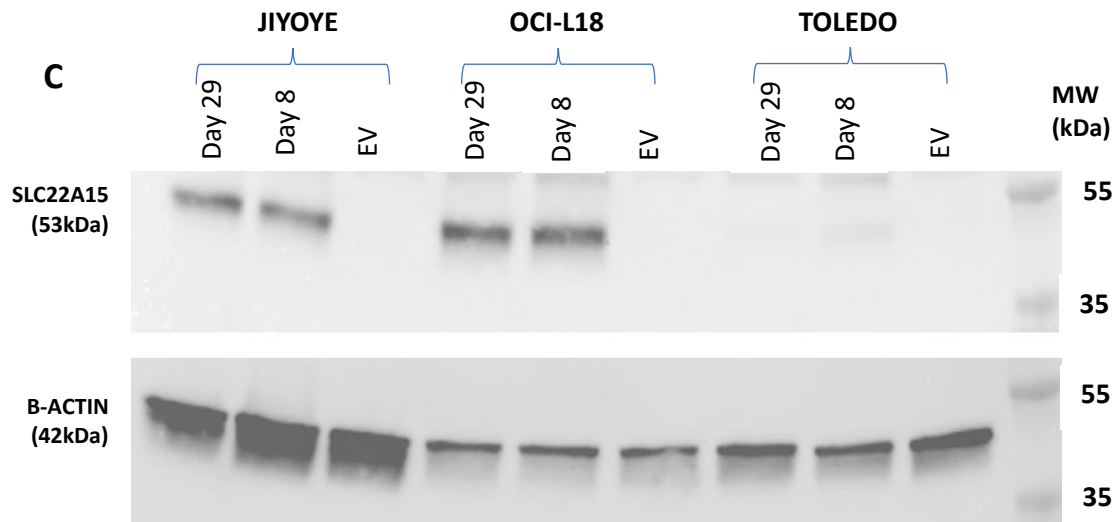
Loss of *SLC22A15* protein expression was less clear in the lymphoma cell lines. Two of the cell lines (Daudi and Toledo) showed a readily visible reduction of *SLC22A15* protein

(Figure 4.12 B-C). For Toledo the bands at both day 8 and 29 are faint due to the low transduction efficiency in this cell line. On the other hand, the two remaining lymphoma cell lines, Jiyoye and Oci-Ly-18, showed no evidence of a visible reduction of *SLC22A15* between days 8 and 29 even though approximately 20% and 30%, respectively, reduction of GFP express cells was observed at day 29 post-transduction (Figure 4.12- C).

As shown above (figure 4.11) there was no correlation between expression measured at the RNA level and the sensitivity of cell lines to *SLC22A15* re-expression. However, as RNA expression levels are not always reflective of protein expression, a similar analysis was performed to determine if protein expression in the cell lines was correlated with sensitivity to the *SLC22A15* re-expression vector. *SLC22A15* protein level of all samples at post-transduction day8 was quantified with densitometry analysis. *SLC22A15* protein was then normalised to the loading control (B-actin) (Figure 4.12-D). On the next step, protein level of each cell lines corrected to transduction level assessed by flow cytometer at same day to estimate *SLC22A15* protein expression specifically in the successfully transduced cells. The original values calculated by densitometry are shown in figure 4.12D and the values after correction for transduction efficiency are shown in Figure 4.12E.

Similar to the analysis of expression at the RNA level, protein expression level at day 8 is not clearly correlated with the impact of *SLC22A15* re-expression (as assessed either by loss of protein expression by day 29 or by loss of GFP expression in the transduced cells). Several of the cell lines with the highest levels of protein expression after correction for transduction efficiency (e.g. Oci-Ly-18, Daudi, SEM) show no or only partial loss of *SLC22A15* protein by day 29 and there was no significant correlation between protein expression level at day 8 and the rate of subsequent loss of *SLC22A15*/GFP expressing cells ($R^2= 0.0007$, $p> 0.05$).





E

Cell Lines	Corrected Protein Expression level	Loss of SLC22A15 protein
JIYOYE	0.464554286	Not observed
OCI	3.047026135	Not observed
DAUDI	3.589302558	Partial
TOLEDO	0.653447325	Partial
SEM	3.62293699	Partial
NALM6	0.478780763	Partial
REH	3.61147253	Complete
PREB	2.258951996	Complete
RCH-ACV	2.702687709	Complete

Figure 4.12. Retention of SLC22A15 protein expression following transduction.

SLC22A15 protein was assessed on all cell lines at day 8 and day29 compared with control (labelled as EV in these figures). Reduction of SLC22A15 protein between day8 and day29 after transduction was clear on all ALL cell lines along with two lymphoma cell lines (Toledo and Daudi) (A, B, C). In contrast, there were no visibly detectable reduction of SLC22A15 protein in the other two lymphoma cell lines, Jiyoye and Oci-Ly-18 between day 8 and day29 (C). There was no any SLC22A15 protein expression on control group in any cell lines (A, B, C). (D) SLC22A15 protein level was assessed on transduced cell lines at day 8 then normalized to the loading control (B-actin) by densitometric analyses. (E) To investigate if the effect of SLC22A15 expression on cell lines was related to the level of expression in transduced cells, the expression level of SLC22A15 protein was corrected for the percentage of cells transduced in each cell line. And cell lines were categorized according to the extent of loss of SLC22A15 protein (either no detectable, partial or complete). All data shown here come from single experiment. B-actin was used as loading control.

4.5 Functional Validation of THEM4

4.5.1 Re-expression of THEM4

Functional assessment of *THEM4* was initially conducted using four ALL cell lines (PreB697, Reh, SEM, and NALM6). and utilised the lentiviral system described above. All cell lines were infected with pSIN-THEM4 vector, and the pSIN-EV as a control group. Any selection against expression of *THEM4* could then be monitored by assessing the GFP expression, as described above. The fraction of GFP positive cells was determined at days day 4, 8, 11, 15, and 22 after transduction. *THEM4* expressing cells (assessed by fraction of GFP positive cells) were rapidly lost between day 4 and 22 in the PreB697 cell line (Figure 4.13A, 4.14), while the GFP positive fraction remained constant in the EV control transduced cells. In contrast, transduction of any of the other cell lines tested with the pSIN-THEM4 construct resulted in completely or near completely stable levels of GFP positive cells, which mirrored the stability seen in the EV control (Figure 4.14). This is not consistent with the identification of *THEM4* as a candidate tumour suppressor across all

ALL genetic subtypes but could be consistent with a more specific role in the TCF3-PBX1 subtype, from which PreB697 cells are derived. To address this possibility, a second TCF3-PBX1 subtype cell line, Rch-acv cell lines, was obtained from DMSZ and assessed as above. However, this also failed to replicate the rapid loss of GFP positive cells in the pSIN-THEM4 transduced cultures and instead found the stability of the post transduction GFP positive population to match that seen in the control EV transduced cells (Figure 4.14).

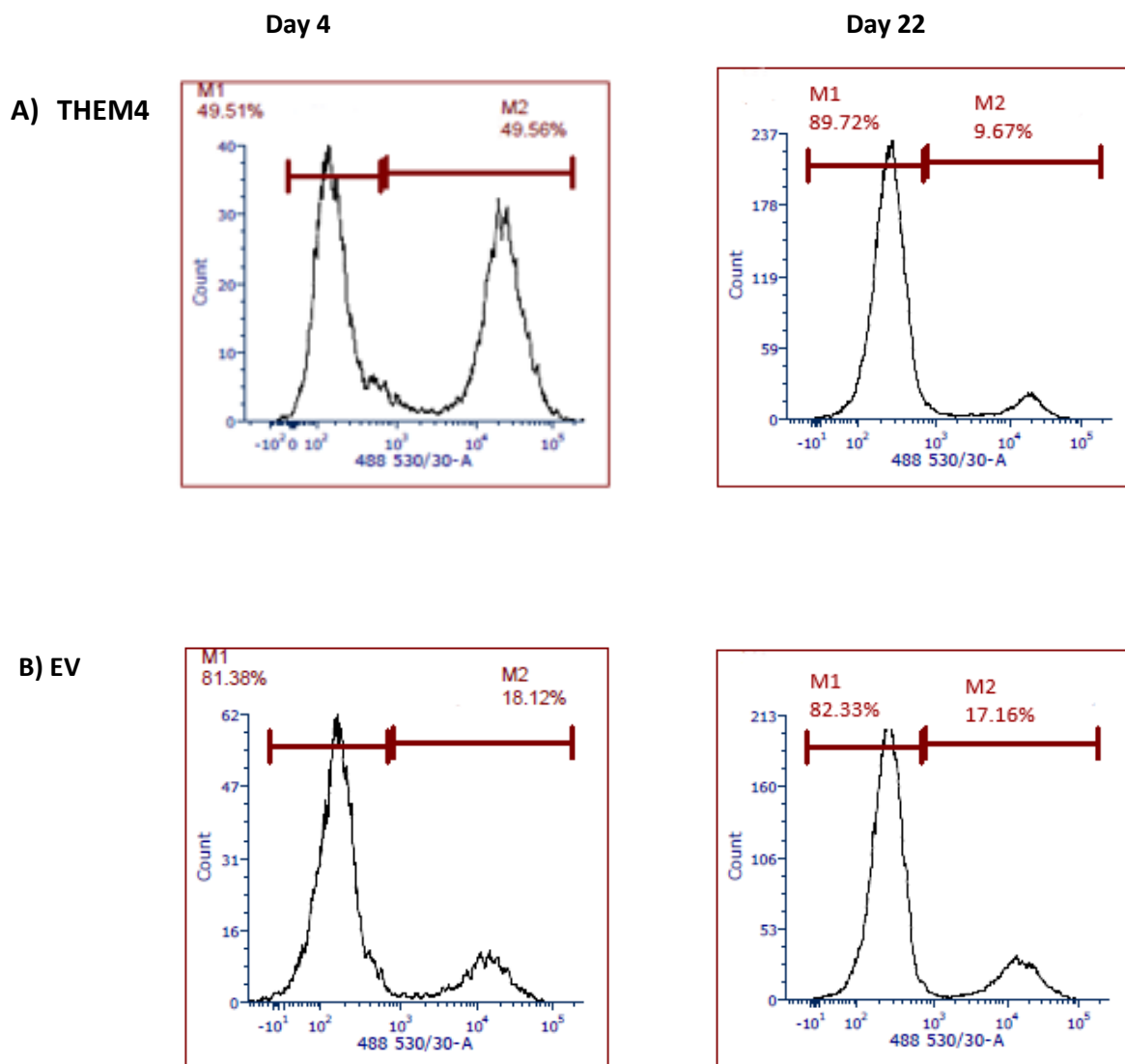


Figure 4.13. Representative Flow-Cytometer Results of transduction both testing group and control group on PreB697 Cell Line post-transduction at day4 and day22.

Stable GFP expression was observed on control group at post-transduction day4 (left on B) and day22 (right on B). However, the percentage of GFP of cell population expressed THEM4 on was 49.56% on day4 after transduction (left on A) and it was dropped to 9.67% at post-transduction day22 (right on A). Histogram data (X axis represent GFP intensity) comes from single experiment after analysis by FCS Express Program. (M1 and M2 marker were created by FCS Express program during the analysis and M1 marker represent not GFP expressed cell population while M2 marker represent GFP expressed cell populations).

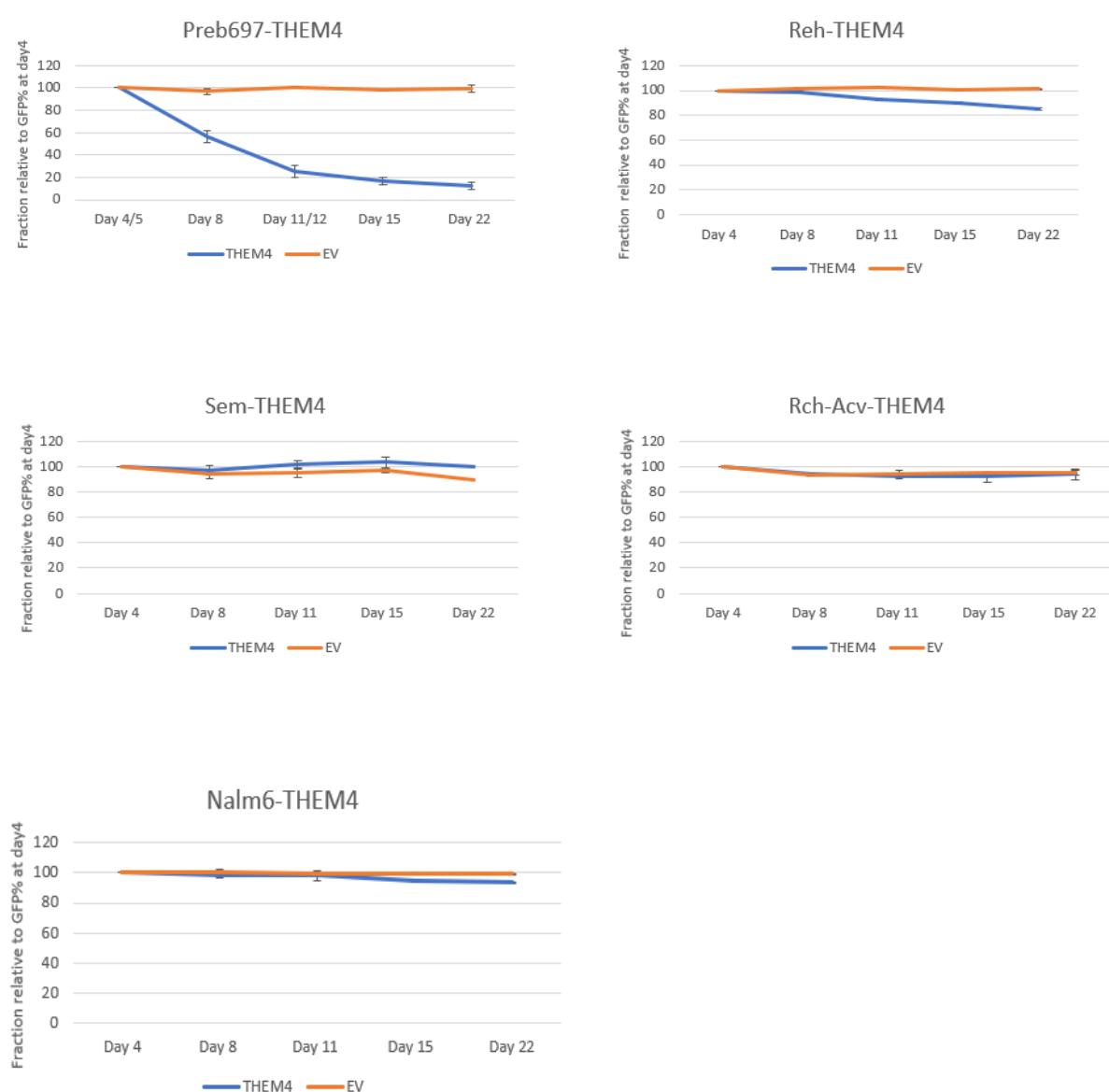


Figure 4.14. Results of ALL cell lines after transduced with pSIN-SLC22A15 and p-SIN-EV

More than 80% reduction of GFP was observed in PreB697 cell line while the control group had stable GFP expression during the time course. However, Rch-acv, SEM and NALM6 cell lines showed stable GFP expression in the both groups which were pSIN-SIEW-THEM4 vector infected and control. Even though there was small reduction of GFP expression on Reh cell line, there was no statistically significant differences comparison with control group. amount of THEM4 gene expression was retained on REH cell line by end of the time course. Data shown in here produced from two independent experiment results and bars represent standard error between experiments. Results was normalised to GFP% at day4. Blue line represent vector expressed THEM4 and GFP (called THEM in the figure, orange line represents vector express just GFP (called EV in the figure).

4.5.2 Assessing *THEM4* protein expression on ALL cell lines after transduction

To clarify whether infected cell lines expressed *THEM4* protein after transduction, western immunoblotting was performed on samples at post-transduction day4 and day22 infected with pSIN-THEM4 vector and pSIN-EV in the PreB697 cell line, in which GFP levels fall rapidly and in the NALM6 and Reh cell lines, where the GFP fraction remains largely constant in the pSIN-THEM4 transduced cells. This analysis found that *THEM4* protein was detectable at day4 in PreB697 cells, but no longer detectable at day 22. By contrast, *THEM4* protein expression was largely constant between day4 and day22 in NALM6 and Reh cells, although a small reduction in *THEM4* protein was seen at day 22 in the Reh cell line, consistent with the small reduction in GFP levels seen in Figure 4.15. This result further confirms that the negative impact of *THEM4* re-expression appears to be specific for a single tested cell line and that this is not due to other mechanisms impacting *THEM4* protein stability in the other cell lines nor is it due to higher expression levels in the one impacted cell line, as PreB697 cells exhibited clearly lower expression than Reh cells and similar levels to NALM6 cells.

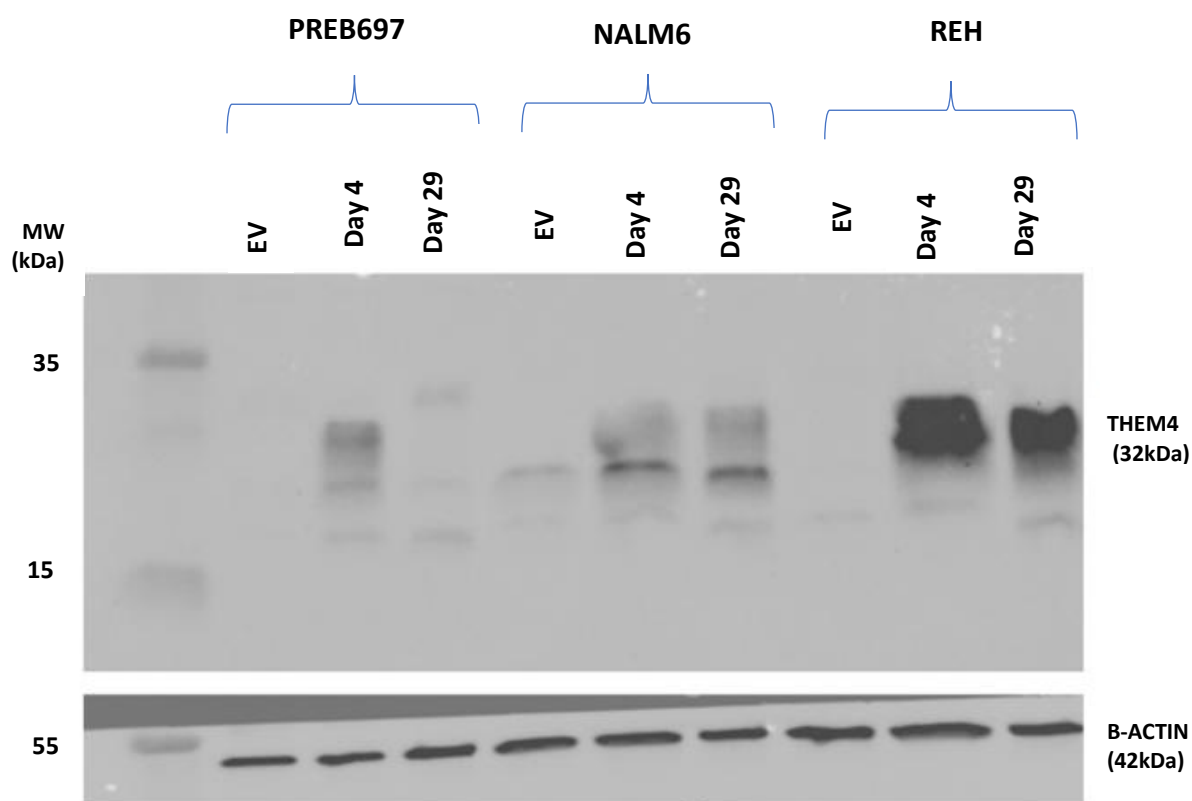


Figure 4.15. THEM4 protein expression following transduction with the THEM4 re-expression construct

Western immunoblotting was performed to show THEM4 protein expression after transduction at day4 and day22 along with control, EV transduced, group which expressed just GFP in PreB697, NALM6 and Reh cell lines (as indicated). Clear reduction of THEM4 protein was seen on PreB697 cell line between day 4 and day22 post-transduction but with little or no change in THEM4 expression in the other cell lines. B-actin was used as a loading control. Data is derived from a single western blot experiment.

4.6 Functional Validation of TTC12

4.6.1 Re-expression of TTC12 in leukaemia cell lines

Five ALL cell lines (PreB697, Reh, SEM, Rch-acv and NALM6) were used for functional assessment of TTC12 using the lentiviral system described above. All cell lines were transduced with pSIN-TTC12 vector and with the control pSIN-EV vector. The fraction of GFP positive cells was determined at days day 4, 8, 11, 15, 22 and 29 post-

transductions. As can be seen from figure 4.16, Rch-acv was only cell line showed reduction of GFP (around 30%) by the end of time course and there was statistically significant differences between control group and testing group which infected with pSIN-TTC12 vector ($p<0.05$). However, PreB697 cell line as another TCF-PBX1 subtype of ALL did not show rapid loss of GFP positive cells in pSIN-TTC12 transduce cell populations against the Rch-acv cell line. This is not consistent with the identification of *TTC12* as a candidate tumour suppressor across all ALL genetic subtypes. In addition to, GFP expression was essentially constant from day 4 to day 29 in the other ALL cell lines, Reh, SEM and NALM6 and was not readily distinguishable from the EV control. This result suggesting that *TTC12* is neither not an ALL tumour suppressor gene or that its function is not adequately modelled in the ALL cell lines (Figure 4.16).

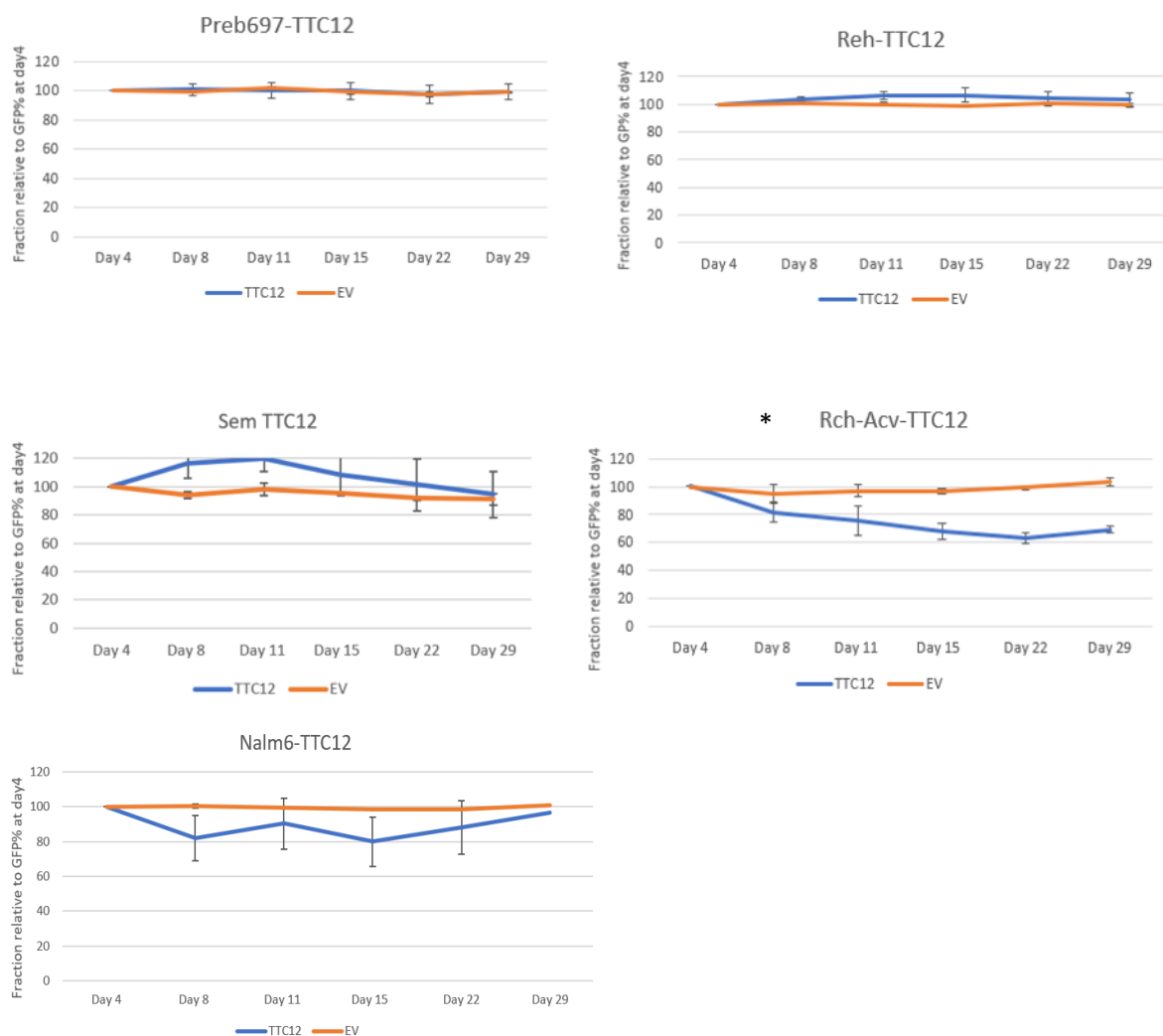
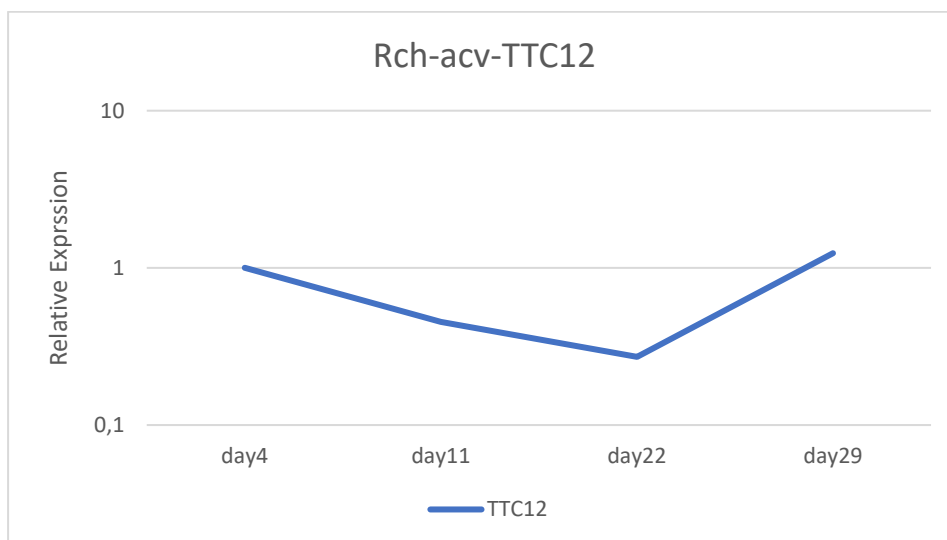
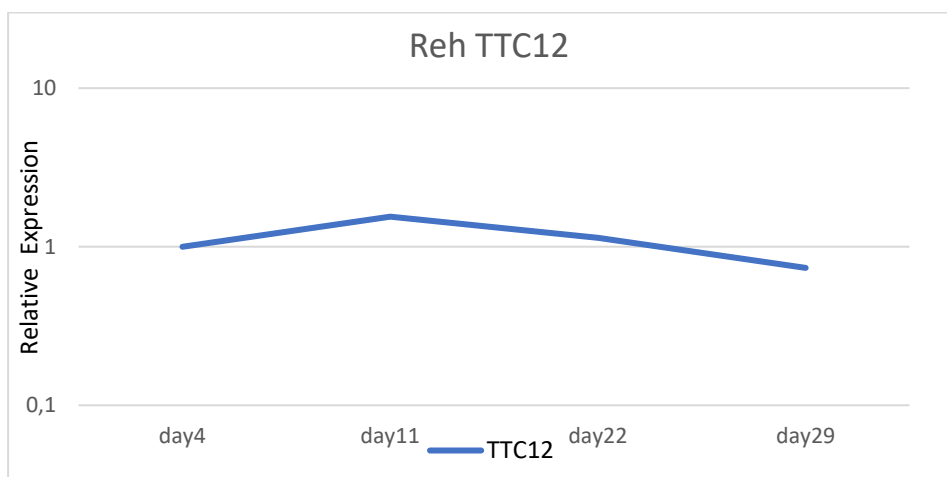
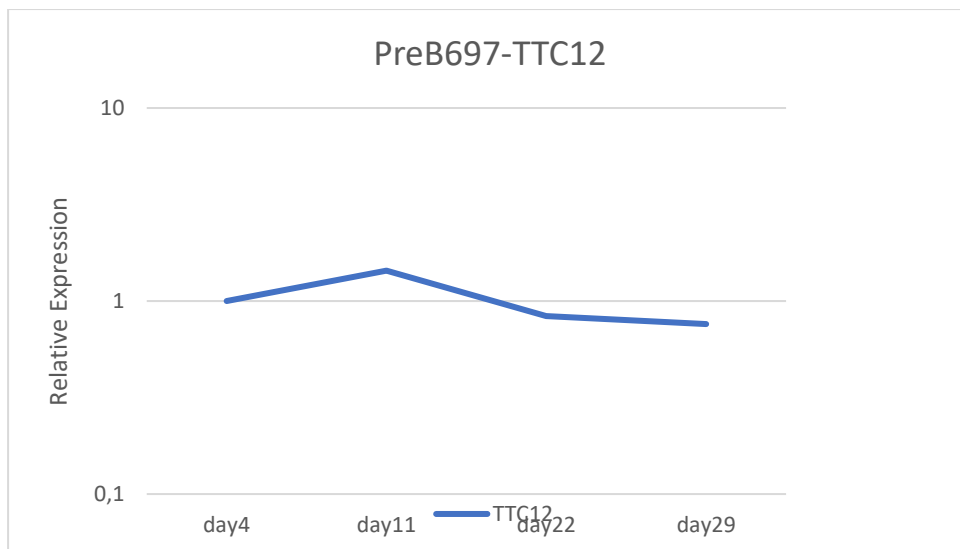


Figure 4.16. Retention of GFP positive cells following transduction with the pSIN-SIEW-TTC12 vector

*Five ALL cell lines were transduced with the TTC12 lentiviral expression vector and with the control empty vector (expressing just GFP). Cultures were monitored from day4 to day 29 after transduction. GFP level was stable during the time course in PreB697, Reh, NALM6 and SEM cell lines similar to control group. There was small reduction of GFP/TTC12 expression in the Rch-acv cell line, although the rate of loss of expression was low and the majority of transduced cells (about 70%) still retained expression at day 29. Data shown in here produced from two independent experiment results and bars represent standard error between experiments. Results was normalised to GFP% at day4. * $p < 0.05$.*

4.6.2 mRNA expression level after TTC12 re-expression

A potential explanation for the lack of impact of the TTC12 expression construct could be that it either fails to express TTC12 or that TTC12 expression falls even though GFP levels remain constant. To confirm TTC12 expression and its stability during the period assessed, qRT-PCR experiment was performed at post-transduction day4, 11, 22 and 29 on transduced cell lines PreB697, Reh, Rch-acv and NALM6. This was done using cell pellets collected at the same time points as the GFP levels were assessed by flow cytometer analyses. As can be seen from below figure 4.17, although there was fluctuation of the measured TTC12 mRNA level across the time course, TTC12 transcript expression level remained broadly similar across the time course and was not significantly lower at the end of the assessment period at day 29 than it was at day 4 in each of the cell lines (Figure 4.17). Thus, the resistance of the cell lines to transduction with the pSIN-TTC12 expression construct is not due to a lack of TTC12 mRNA expression.



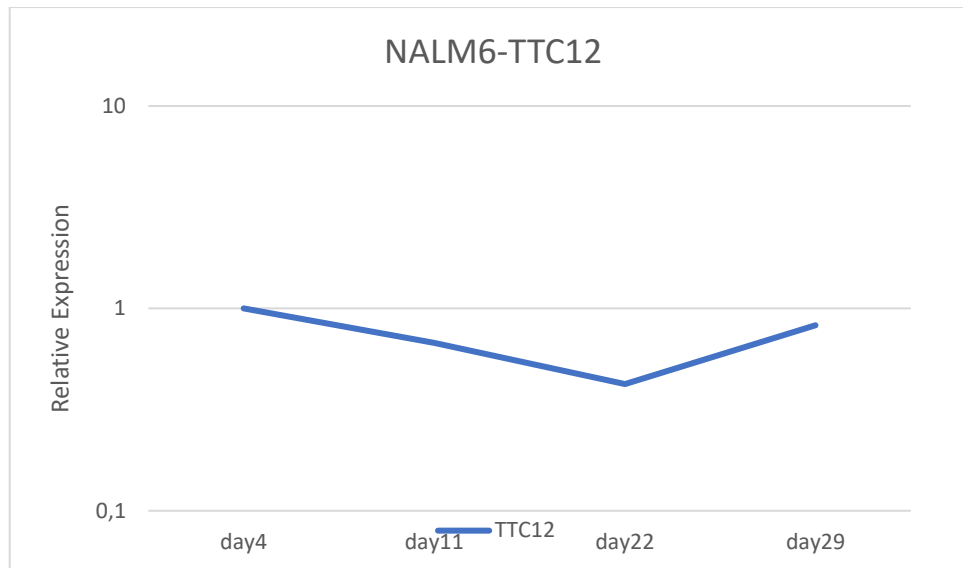


Figure 4.17. TTC12 expression at the mRNA level following transduction of the ALL cell lines

mRNA expression of TTC12 expression at post-transduction days 4, 11, 22 and 29 was assessed in the ALL cell lines. This analysis found no clear evidence of consistently reduced TTC12 expression across the study period. The results closely reflect the flow cytometry results that shows GFP changing after transduction on same period of time. Data shown here are representative of single experiment. GAPDH was used to normalised data and relative expression was calculated with at day4 on log scale.

4.7 Transduce All Cell Lines With pSIN-MAP9 Vector

Five ALL cell lines (PreB697, Reh, SEM, Rch-acv and NALM6) were used for functional assessment of *MAP9* by using lentiviral system, described above. All cell lines were transduced with pSIN-MAP9 vector and with pSIN-EV a corresponding “empty vector” The fraction of GFP positive cells was determined at days 4, 8, 11, 15, 22 post-transduction[297]. However, following transduction with the pSIN-MAP9 vector GFP expression was either undetectable or largely undetectable. An example, using the Reh cell line, is shown in Figure 4.18A. In the third panel, transduction with the pSIN-SIEW-MAP9 vector does not result in any evidence of a separate GFP positive peak as would be

expected (profiles for control empty vector and pSIN-TTC12 transduced Reh cells are shown for comparison). Similar results were obtained for all cell lines and in multiple separate transductions. While there was some evidence of a slight shift in the peak to the right, suggesting that there may be a very low level of expression, this made measuring any change in GFP levels unreliable and would not enable any definitive conclusions to be drawn, as a lack of impact of the transduced construct could be explained by the very low gene expression level.

While in this project and others in the lab, expression from the pSIN-SIEW vector can be variable depending on the cloned gene, the complete inability to observe any clear GFP expression was not something that had been noted before. A potential explanation could be post transcriptional targeting of the transcript, which could represent a second tumour suppressor mechanism leading to *MAP9* downregulation in ALL cells. If this was so, it would be expected that the lack of detectable expression would also be specific for ALL cells. To test this, the 293T cell line (derived from human embryonic kidney cells) was transduced with the pSIN-MAP9 and control vectors then assessed on day4 by flow cytometer. While the control empty vector resulted in very high expression of GFP in close to 100% of the cells, transduction with the pSIN-MAP9 vector did not result in any detectable GFP expression (Figure 4.18). As described earlier original expression construct was sequenced to confirm the correct sequence the cloning position of the insert. Furthermore, multiple separately generated viral preparations all failed to generate separate GFP positive peaks. Notably, these results are not similar to poor viral preparations that generate low levels of transduction, as this would still generate a clearly separate peak, even if the height of the peak (i.e. the fraction of transduced cells) was very low. Overall, the results suggest that cloning the *MAP9* sequence into the expression vector resulted in either an inhibition of transcription or the production of a highly unstable transcript. The results thus are not able to provide any evidence to support or refute the hypothesis that *MAP9* is an ALL specific tumour suppressor.

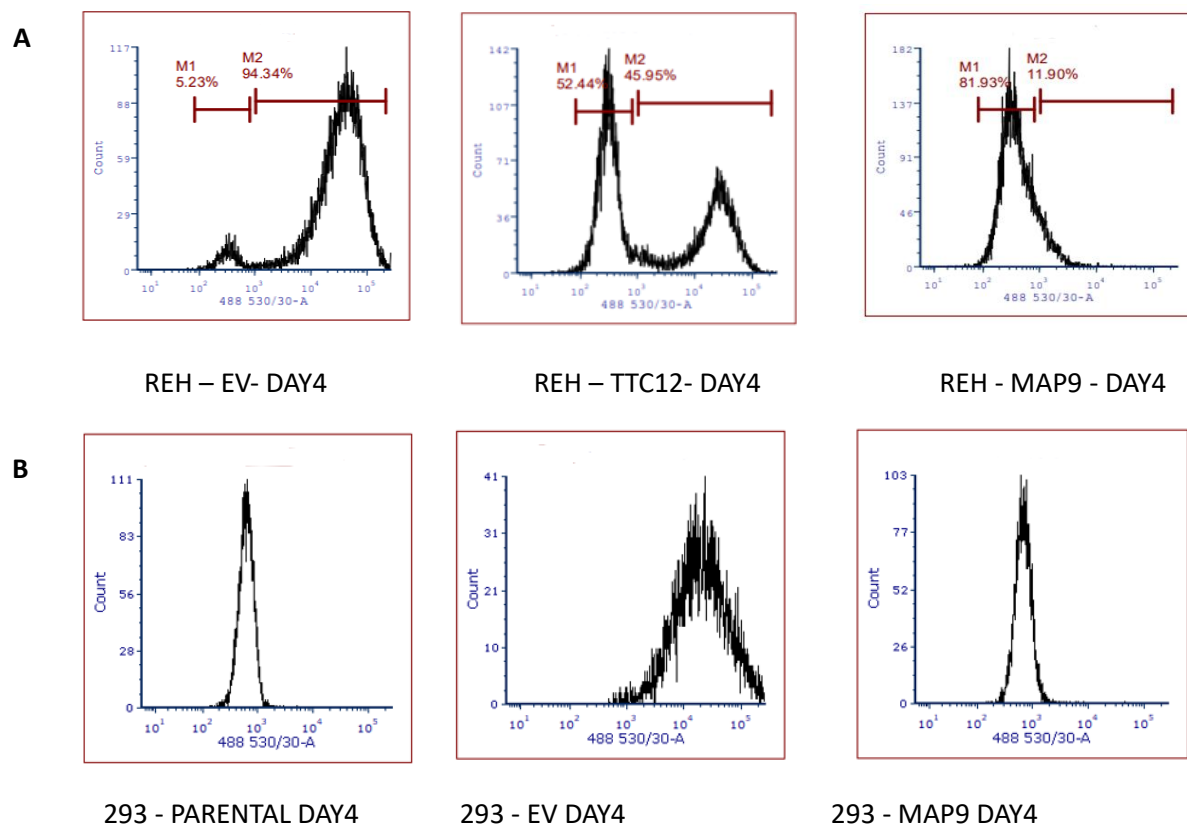


Figure 4.18. Examples of GFP expression in Reh and 293T cell lines at day4 after transduction with various expression constructs

(A), Transduction with either the EV control (middle) or TTC12 expressing construct (left) results in two clear cell populations at day4, while there was no clearly detectable separate population based on GFP expression following transduction with the pSIN-MAP construct (right). (B) For the 293T cell line parental, non-transduced cells, give a single peak as expected(left). This shifted clearly to the right following transduction with the empty vector (middle). However, no GFP expression was detectable after transduction with pSIN-SIEW-MAP9 vector at post-transduction day4 (right). Data comes from single experiment after analysis by FCS Express Program. (M1 and M2 marker were created by FCS Express program during the analysis and M1 marker represent not GFP expressed cell population while M2 marker represent GFP expressed cell populations).

4.8 Discussion

In the present study, based on novel bioinformatic approaches, we identified four TSG candidates (*SLC22A15*, *THEM4*, *TTC12* and *MAP9*) for ALL. In this chapter, we begin the initial process of functionally assessing these identified candidates. A key difference between the work reported here and our previous work in the area was to expand the approach from being cancer subtype based to covering whole diseases. Thus, a crucial element of this assessment was to identify genes that were functionally relevant across multiple genetic subtypes. Therefore, the analysis was carried out using a panel of ALL cells lined derived from multiple cytogenetic subtypes. To explore potential tumour suppressor role of candidates, re-expression of the candidates in the cell lines was achieved via transduction with a lentiviral expression construct, into which each of the genes were cloned. The constructs also co-expresses GFP from the same transcript to allow transduced cells to be readily identified by flow cytometry. Correct cloning and lack of genetic disruption of the candidates in the expression construct as confirmed by sequencing.

Members of the *SLC22A* family facilitate the transfer of various endogenous compounds and xenobiotics substances across the cell membrane [298]. One of orphan member of these family, *SLC22A15*, does not have any known substrates or inhibitors. Despite being classified as a carnitine transporter based on genetic analyses, there has been little research aimed at determining its physiological substrates or how it transports them [299]. Yee et al have identified *SLC22A15* can transport these substrates; eight zwitterions like ergothioneine, carnitine, carnosine, gabapentin, four cations such including MPP⁺, thiamine and cimetidine and additionally carnosine was identified as specific substrate of *SLC22A15* among the *SLC22A* transporter family [285]. Although there are no previous studies demonstrating a direct functional role for *SLC22A15* in carcinogenesis, the role of carnosine in cancer has been studied in multiple different cancer types as an anti-tumour metabolite [300], [301], [302]. Therefore, downregulation of carnosine uptake may be important in the functional role of *SLC22A15* in ALL. To begin to assess the potential tumour suppressor role of *SLC22A15* in ALL, the gene was re-expressed in a panel of ALL cell lines. Following re-introduction of the gene, there was a

clear decrease in GFP/*SLC22A15* positive cells across the time course (while control empty vector transduced cells demonstrated stable GFP expression levels). Importantly, while there was variation in the rate of loss of *SLC22A15* expressing cells in the different cell lines, all five ALL cell lines did exhibit a decline in *SLC22A15* positive cells across the time course, consistent with a potential functional role for *SLC22A15* in ALL that was independent of cytogenetic subtype. To investigate the mechanism by which *SLC22A15* impacts ALL cells, cell death and cell proliferation assay were performed, using two of the cell lines (PreB697 and Reh) which exhibited more rapid loss of *SLC22A15* expressing cells. This analysis determined that *SLC22A15* re-expression appear to have little or no effect on the rate of cell proliferation, but a clearly significant effect on cell survival. The co-appearance of annexin V positive staining cells with activation of caspase 3/7 (detectable from day three post-transduction) suggested that *SLC22A15* re-expression was inducing an apoptotic cell death mechanism. However, use of a caspase inhibitor that completely blocked activation of caspase 3/7 did not prevent or reduce the rate of loss of *SLC22A15* expressing cells. This suggests that either the main mechanism of cell death induced following *SLC22A15* re-expression is not apoptotic, and that activation of caspases is a by-product of the actual cell death mechanism or that blocking apoptotic cell death resulted in activation of a second cell death mechanism.

Another key aspect of the original bioinformatic approach is that the identified candidates should be specific for ALL and play no role or a lesser role in other B-cell derived cancers. Thus, to begin to assess this *SLC22A15* was also re-expressed in several mature B-cell malignant cell lines, including two Burkitt lymphoma and two DLBCL cell lines. The results of this analysis were partially consistent with ALL specificity for *SLC22A15*. While, all four cell lines showed at least some evidence of loss of *SLC22A15*/GFP expressing cells across the assay period, this was generally at a relatively low level and none of the mature B-cell cell lines displayed the rapid loss of *SLC22A15*/GFP positive cells seen in 3/5 ALL cell lines and the average loss of *SLC22A15*/GFP positive cells was significantly higher in ALL cell lines than in the lymphoma cell lines. However, these results did not identify an absolute requirement for loss of *SLC22A15* in all ALL cell lines versus lymphoma cell lines. There was clearly some evidence of selection against *SLC22A15* expression in the mature B-cell malignancy derived cell lines and two of the ALL cell lines (NALM6 and SEM) exhibited

response largely similar to those seen in the lymphoma cell lines. Further work would also be required to delineate more clearly any differences between ALL subgroups (as most are only represented by a single cell line) and whether differences in the impact of *SLC22A15* may exist between different types of mature B-cell malignancy (for example, CLL and MCL were included in the bioinformatic analysis but not represented in the cell line study). Overall, the results are consistent with a tumour suppressive like role for *SLC22A15* that is at least partially ALL specific but may also indicate that other aspects of cellular background may modify the functional impact of *SLC22A15* expression.

Quantitation of *SLC22A15* expression at the protein level was carried out by western blotting. As we expected protein level of *SLC22A15* sharply declined after between day 8 and day 29 post-transduction in three ALL cell lines, Reh, PreB697, Rch-acv and limited reductions was seen in SEM, NALM6 and two lymphoma cell lines, Daudi and Toledo. In contrast, there was no evidence of decrease of *SLC22A15* protein between day 8 and day 29 at Jiyoye and Oci-Ly-18 cell lines. Thus, reduction in GFP expression following transduction with the *SLC22A15* expressing vector was mirrored by a similar reduction in *SLC22A15* expression at protein level.

THEM4, also known carboxy-terminal modulator protein (*CTMP*), was also identified as a tumour suppressor candidate by the bioinformatic analysis and assessed in the cell line models. A number of studies have linked *THEM4* to important pathways in cancer development, including as a negative regulator of AKT [279], [281]. Maria and colleagues showed that *THEM4*, as a component of PKBa signalling, interacts directly with Akt and cause inhibits its phosphorylation at threonine 308 and serine 473serine/threonine residues in the C-terminal of PKB (protein kinase B) at the plasma membrane and leads to inhibition of AKT phosphorylation. This results in preventing inappropriate kinase activation as well as further cell growth and proliferation [279]. On the contrary, Ono and colleagues found that *THEM4* protein specifically binds to N terminal of PKB to facilitate phosphorylation so that AKT phosphorylation and anti-Initial assessment of *THEM4* in the apoptotic activity are enhanced in breast cancer cell tumorigenesis [303]. These findings suggest that *THEM4* may have different functions in different cancer types but also demonstrate that it regulates key pathways involved in cell growth and therefore could potentially have a tumour suppressor function in some cell types. However, there are no

functional studies regarding any potential role in leukaemia although it had previously been identified as hypermethylated in ETX6/RUNX1 subtype of ALL [248]. Initial assessment of in four ALL cell lines provided only partial evidence for a functional impact of *THEM4* re-expression. In one cell line (PreB697) *THEM4* re-expressing cells were rapidly lost for the population, whilst in all other cell lines *THEM4* expressing cells were largely retained at the Day 4 post-transduction levels. Testing a second TCF3-PBX1 derived cell line (as PreB697 is TCF3-PBX1 positive) also failed to show specific specificity for this genetic subtype. Thus, it remains unclear if the impact of *THEM4* re-expression in PreB697 cells is reflective of a more general role in ALL whose importance has been lost in most ALL cell lines, is reflective of a role in a subset of ALL due to other aspects of cellular background (other than cytogenetic subtype) or a cell line phenomenon in the PreB697 cells that is not relevant for primary ALL. Overall, while the limitations of the cell line model system mean that a tumour suppressor role in ALL generally cannot necessarily be ruled out, the results presented here do not provide clear evidence supporting such a role.

TTC12 gene located at 11q23 and code cytoplasmic protein that plays a role in the proper assembly of dynein arm complexes [287]. Although the function of *TTC12* has not been identified yet, *TTC12* gene has been found specifically methylated in ALL [249]. Four different protein variants (designated 1, 2, 3 and 5) are listed in the NCBI protein database. The encoded proteins are highly similar but vary slightly in lengths due to alternative splicing (711 aa, 705 aa, 680 aa and 706 aa in length, respectively). In current study, variant 1, longest variant, was used for expression of *TTC12* in cell lines. Consequently, it is hard to predict whether the different isoforms would have important functional differences that might impact their role in ALL. This would be achieved by performing qRT-PCR by primer would pick up all variants to identify which variants are expressed.

In terms of re-expression of *TTC12* four ALL cell lines did not identify any cell lines in which *TTC12*/GFP expression was rapidly or clearly lost. On the contrary, Rch-acv cell line showed reduction of *TTC12*/GFP cell population. However, assessment of expression at the RNA level further confirmed that *TTC12* mRNA expression was retained at largely the same levels at the end of the assay period as at the first assessed time point (day 4). A clear limitation to cell line based assessment of candidate genes is that there is no modelling of the microenvironment or ALL cell interactions with other cell types.

However, although a negative role in ALL development cannot be ruled out on the basis of the results, the results do not currently provide any support for an important functional role for *TTC12* in ALL.

MAP9 belongs to *MAP* family, which regulates microtubule dynamics [304]. *MAP9* has not been very widely studied to date, however, there are a number of papers which have reported potential tumour suppressor roles for *MAP9* in colorectal cancer [292], in hepatocellular carcinoma [293] and gastric carcinoma [294], No studies have yet been reported investigating the role of *MAP9* in ALL or other B-cell malignancies. The results presented here also fail to shed any clear light on the potential of a tumour suppressor role for *MAP9*, as transduction of ALL cell lines with the lentiviral expression construct failed to give any clear evidence that the cloned transcript was expressed. Multiple lines of evidence suggest this was due to a failure to obtain significant expression from the lentiviral vector as a result of the cloning of the *MAP9* sequence. Firstly, although a clear GFP expressing peak was not seen in any cell line post-transduction, some of the cell lines did exhibit a small widening of the peak compatible with very low expression levels. Notably, this is different from very low transduction efficiency (potentially due to poor quality viral preps) as this would still give a very small, but still distinct peak. Furthermore, similar results were obtained with multiple separate transfections and with multiple viral preps. While, for the other expression constructs, variability in quality of viral preps was observed, this never resulted in a complete failure to identify at least a low level of clearly transduced/GFP-expressing cells. Lastly, although such low expression of constructs based on the pSIN-SIEW vector have not been observed previously in the lab, either in this project or a number of other utilising the same lentiviral vector, it has previously been observed that cloning of different upstream sequences has significant positive or negative results on the efficiency of expression driven by the vector. Sequencing of the construct demonstrated that the gene had been appropriately cloned and that the sequence was correct and intact.

A potential hypothesis to explain the unexpectedly low expression would have been a second mechanism in ALL cells that served to suppress *MAP9* expression through destabilisation of the *MAP9* transcript, such as a miRNA. If such a mechanism did underlie the ALL specific tumour suppressor function for *MAP9* then it would be predicted that

expression in non-ALL cells would be detectable. However, transduction of a non-ALL cell line, 293T, with same lentiviral construct, resulted in a similar failure to identify a GFP positive peak post transduction (even though expression from the viral promoter is usually very high in this cell line). Thus, as no evidence could be obtained that *MAP9* had been successfully expressed in the cells, no conclusions could be drawn about the activity of the *MAP9* protein.

Overall results suggesting that of the four candidates assessed only three provided interpretable results and of this one exhibited a clear negative impact on cell survival in ALL cells that was seen across multiple genetic subtypes consistent with its identification as a pan-ALL tumour suppressor gene.

*5 FUNCTIONAL ASSESSMENT OF ACUTE
LYMPHOBLASTIC LEUKAEMIA-SPECIFIC
DISEASE SPECIFIC DEPENDENCY
CANDIDATES IN CELL LINE MODELS*

5.1 Introduction

According to the novel bioinformatic pipeline developed by our group, we have identified TSG and DSDG candidates across the five type B-cell malignancies. The approach enabled to identify total 13 candidates, including eight cancer dependency gene candidates which are four for ALL, two for MCL and one for CLL specific. Therefore, we chose to analyse functional relevance for the ALL candidates, as it has the highest number of candidates among the five types of malignancies. *BMP2*, *CTGF*, *NPY* and *ZNF423* were the identified ALL-specific candidates. Due to time limitations for laboratory work, we decided to focus on *BMP2*, *CTGF*, *NPY* for the functional analysis (as all three genes had a DMR at their TSS, while for *ZNF423* the DMR was over 18kb away from the TSS). All ALL cell lines used in this analysis were assessed for DNA methylation and gene expression of the candidate genes, which confirmed low levels of promoter associated methylation correspond with high level of expression for all candidates by a previous student in our group. To determine whether these candidates have specific roles in ALL, siRNA mediated knockdown against *BMP2* and *CTGF* was performed on two ALL cell lines and two lymphoma cell lines as a control group. Inhibitors for the BMP pathway, and *NPY* receptors and antagonistic Ab for *CTGF* were used as a second approach to assess functional roles of the candidates. Also, all three candidate genes encode proteins that are typically secreted. Therefore, ELISA assays were carried out in the untreated (parental) cell line samples to determine if secreted proteins, with the potential for autocrine activation of cell surface receptors, could be detected.

5.1.1 siRNA mediated knockdown

Targeted degradation of mRNA for individual genes provides a method for post-transcriptional silencing of gene expression. RNA interference (RNAi) technology is one of the most commonly used methods for inducing this type of gene-specific RNA degradation [305].

RNA interference involves multiple steps. Upon introduction of double-stranded RNA into cells, Dicer, a member of the RNase III family ribonuclease, first recognizes and processes

it into small interfering RNAs (siRNAs) with 21–23 base pairs [306]. In practical applications, siRNA can be artificially created and introduced directly into the cell without requiring Dicer activity. In the presence of siRNA in the cytoplasm, siRNA incorporates into protein complex known as RNA induced silencing complex (RISC) [307]. Argonaute-2 is a versatile protein found in RISC, which unwinds the siRNA and then cuts the sense strand (also known as passenger strand) [308]. The activated RISC with the antisense strand (or guide strand) of the siRNA identifies and degrades mRNA that is complementary to the antisense strand [309]. The cleavage of mRNA takes place between nucleotides 10 and 11 on the complementary antisense strand [310]. Following this process, the activated RISC complex can proceed to eliminate further mRNA targets, leading to continued gene silencing [311]. This enhanced efficacy provides a therapeutic effect for 3–7 days in rapidly dividing cells and several weeks in non-dividing cells [312] (Figure 5.1).

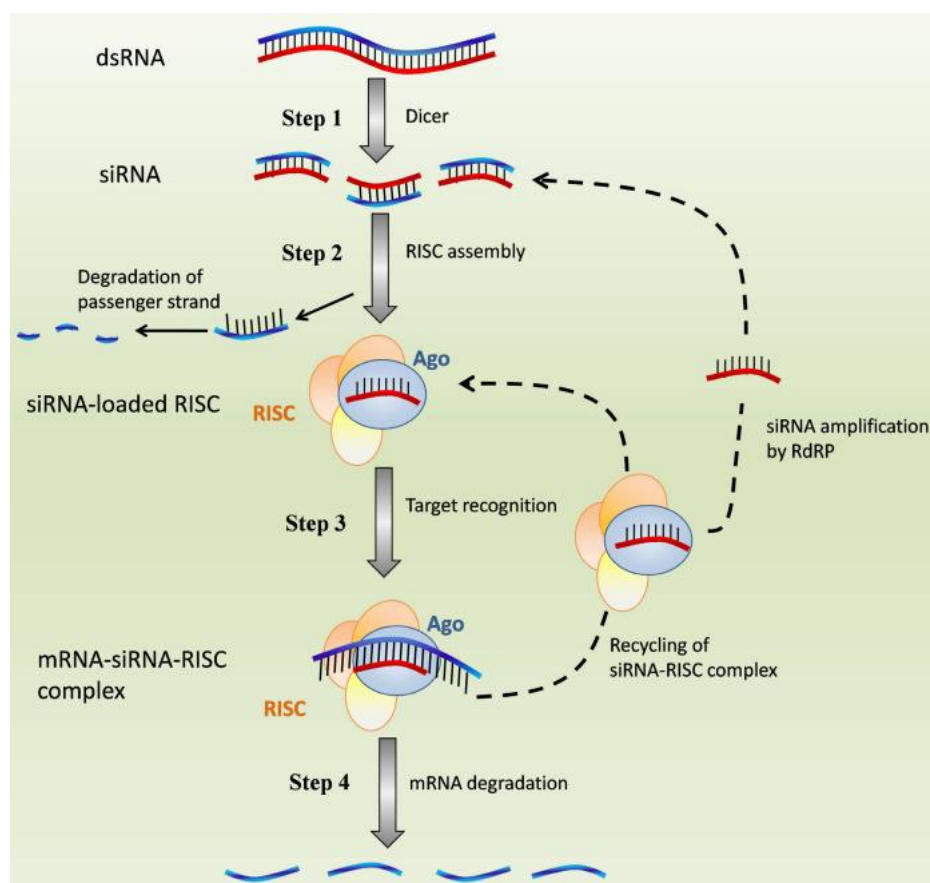


Figure 5.1. Schematic representation of siRNA interference in a mammalian cell

Step 1; Upon exposure to double stranded RNA (dsRNA), the cell triggers a response mediated by the enzyme Dicer. This results in the cleavage of dsRNA into fragments consisting of 21-23 base pairs, known as siRNA. The next step (Step 2) involves loading siRNAs into a multiprotein complex named RNA Induced Silencing Complex. Within RISC, one strand (referred to as the passenger strand) is discarded and degraded, while the guide strand remains as a template during silencing reactions. Subsequently (Step 3), the guide strand forms a functional siRNA-RISC complex by binding to Ago protein. Following this assembly, target mRNAs are identified through base pairing and captured by the siRNA-RISC complex. As a result of these interactions, mRNA degradation is initiated (Step 4); once complete, dissociation from siRNA occurs and allows for further processing of mRNA targets by releasing the siRNA-RISC complex for additional rounds of action. Figure is obtained from [313].

5.1.2 Targeting DSDG candidates as a therapeutic strategy

5.1.2.1 BMP2 Pathway

Bone morphogenetic proteins (BMP) belong to a subset of the transforming growth factor β family and consist of more than 25 protein ligands. BMPs interact with two classes of cell surface receptors known as BMPR-I and BMPR-II. The BMPR-I receptor class includes three types of receptors, including activin receptor-like kinase-2 (ALK-2 (ACVR1A), ALK-3(BMPR-IA) and ALK-6 (BMPR-IB), while the BMPR-II receptor class consists of three different types *BMPR-II*, *ACVR2A*, *ACVR2B*. Upon binding with BMPs, heterotetrametric complexes are formed, comprising two type I and two type II receptors. Each BMP receptor contains an intracellular serine/threonine kinase domain as well as an extracellular binding domain. Subsequent to the binding process, constitutively active type II receptor kinases phosphorylate the kinase domains of type I receptors that subsequently phosphorylate SMADs 1, 5, and 8 which are responsive to BMPs. These SMAD proteins can then translocate into the cell nucleus where they act as transcription factors. Phosphorylation of these particular SMADs leads to various impacts on the cells, such as controlling growth and promoting cell differentiation [314]. Urist and colleagues reported on a group of proteins that directly influenced bone development, which they named bone morphogenetic proteins [315]. Within the

transforming growth factor β family, the BMPs constitute the largest subgroup, comprising 18-20 members [316] and *BMP2* is one of the most extensively studied BMPs and has garnered significant therapeutic interest for its potential applications in bone regeneration and repair [317]. Human *BMP2* is synthesised as precursor proteins (396 amino acid (aa)) comprising three distinct regions: the N-terminal signal peptide (23 aa long) that directs the protein to the secretory pathway; the pro-domain (259 aa long), crucial for protein folding and cleavage of the secreted protein and the C terminal is mature domain (114 aa long) [318] (Figure 5.2). After synthesis of preproprotein form of *BMP2*, it enzymatically cleaved. Mature monomeric *BMP2* is 14kDa and glycosylated monomeric peptide is 18kDa [319]. *BMP2* activated through the dimerization of two 114-amino acid monomers, resulting in a protein with a molecular weight of 26kDa [320]. Additionally, extracellular protein, Furin, is required for cleave of preprotein *BMP2* to produce the mature protein [321]. Canonical *BMP2* signalling pathway is activated by through type I (*BMPR1A* and *BMPR1B*) and type II receptors (*BMPR2*, *ACTRIIA* and *ACTRIIB*) [322]. The signalling cascade is triggered when *BMP2* binds to its receptor, which possesses serine/threonine kinase activity. Transphosphorylation takes place following the binding of *BMP2* to type II receptor, which then proceeds to phosphorylate the type I receptor. The downstream consequences of the phosphorylation of type II and type I receptors are the phosphorylation of Smad transcription factors then cause upregulation of gene has role on bone formation [323]. Once secreted, the soluble forms of *BMP2* have a very short half-life, often only lasting for a few hours [324]. Half-life in circulation is shorter when it is administered systematically averaging 7-16 minutes [325].



Figure 5.2. Schematic representation domains of human *BMP2*
 Figure is adapted from [326].

To determine if inhibition of *BMP* signalling showed evidence of ALL specific effects, ALL and lymphoma cell lines were treated with LDN-193189 a selective inhibitor for which targets *ACVRLI*, *ACVR1*, *BMPR1A* and *BMPR1B* causing inhibition of BMP signalling effectors SMAD1, SMAD5 and SMAD8.

5.1.2.2 CTGF Pathway

Connective tissue growth factor is member of CCN family which has N-terminal signal peptide for protein secretion. The *CTGF* gene encodes an extracellular matrix-associated protein, full size is 36kDa with/out associated glycosylation, and 10-20kDa size was associated with secreted form of *CTGF* after cleavage of full-length protein [327]. *CTGF* has been found to associate with a range of proteins, such as integrins, bone morphogenetic proteins, transforming growth factor- β . *CTGF* has been found to plays a role in the production of extracellular matrix, as well as cellular processes like proliferation, survival, adhesion, migration, and metastasis [328], [329], [330]. *CTGF* has been widely studied and associated with tumour progression, poor prognosis in breast cancer, glioblastoma [331], [332] and an antibody directed against *CTGF* causes decreased tumour growth and metastasis in pancreatic cancer [276]. FibroGen screened a collection of human anti-*CTGF* antibodies and identified the FG-3019 antibody, which effectively bound to *CTGF* and demonstrated activity in both in vitro and in vivo disease models. [333]. Pamrevlumab, a fully recombinant human monoclonal antibody targeting connective tissue growth factor, has emerged as a promising therapeutic candidate for the treatment of idiopathic pulmonary fibrosis [334]. In present study, Pamrevlumab (FG-3019) was used as alternative method for siRNA mediated knockdown to target *CTGF* in ALL and lymphoma cell lines.

5.1.2.3 NPY Pathway

Neuropeptide Y is a 36-amino acid neuropeptide that acts as a neurotransmitter and has widespread distribution in the central nervous system. It was first discovered and isolated in 1982 by Tatenoto et al. during their examination of pig brain tissue [335]. Neuropeptide Y exerts its effects through four functionally relevant receptor subtypes: Y1, Y2, Y4, and Y5. All cloned NPY receptors are part of the G-protein-coupled receptors superfamily but have different ligand affinity profiles [336]. *NPY* role in various cancer progress such as promoting cell proliferation, migration, survival and cell differentiation has been found with growing evidence (reviewed in [337]).

5.1.2.4 Crosstalk between candidates

We have identified total four DSDG candidates for ALL. Three of four candidates encoded secreted protein. *BMP2* and *CTGF* among them have been previously linked to each other in different cell type and conditions. Chang M. et al demonstrate that the effect of *BMP2* on upregulation the expression of *CTGF* is facilitated by the ALK2/ALK3-mediated SMAD-dependent signalling pathway on human granulosa cells [338]. In addition, Takigawa and his colleague *BMP2* has the strongest affinity for *CTGF* among the osteogenic BMPs. Importantly, the combined application of *CTGF* and *BMP2* inhibited both *BMP2*-induced phosphorylation of Smad1/5/8 and the proliferation of human chondrosarcoma-derived cell line HCS-2/8 [339]. Moreover, Yan et al found that *CTGF* enhances the *BMP2* expression and phosphorylation of Smad1/5/9 pathway activity by directly interfacing with *BMP2* in periodontal ligament stem cells (PDLSCs) [340].

ZNF423 is a member of family of multi-zinc finger proteins that can function as both transcriptional activators and repressors [341]. Domain analysis of the ZNF423 protein has demonstrated that zinc fingers 2–8 comprise the DNA-binding domain for ZNF423 homodimers, while zinc fingers 9–13 form the DNA-binding domain for the ZNF423–SMAD1–SMAD4 complex, which recognizes BMP response elements. The interaction with the phosphorylated SMAD1-SMAD4 complex is facilitated by zinc fingers. ZNF423 multimerization and binding to EBF1 involve zinc fingers 20-30 in human. Consistent with its diverse functional capabilities, ZNF423 participates in various other signalling pathways, including retinoic acid signaling, Notch signaling, and the DNA damage response [342].

ZNF423 has been found to interact with a phosphorylated SMAD1-SMAD4 complex in a mutually exclusive manner upon *BMP2* stimulation [343]. ZNF423 has not been found constitutively expressed and played a role in normal haematopoiesis [341]. However, its aberrant activity has been found in B-cell malignancies [251], [344], [345]. Furthermore, Harder and colleagues showed that hypomethylation of CpG island at the genomic ZNF423 locus leads to a permissive transcriptional state of ZNF423, which is activated by *BMP2* signalling in ALL [251].

5.2 Aims of the Chapter

- Functional assessment of three ALL-specific DSDG candidates to determine if reduced expression results in negative impacts on growth/survival of ALL cells.
- Investigation of the specificity of any identified impacts of targeting the DSDG candidates in ALL versus other B-cell malignancies.

5.3 Results

5.3.1 Expression Profiles of DSDG Candidates in cell line models

Five ALL cell lines (PreB697, Reh, SEM, NALM6 and Rch-acv) and four lymphoma cell lines (Toledo, Oci-Ly-18, Daudi and Jiyoye) were assessed mRNA expression of the candidate genes. qRT-PCR was performed to verify gene of interest expression on cell lines tested. In addition to, mRNA expression of *BMP2* and *NPY* receptors were determined in same method.

5.3.1.1 BMP2 Expression Profile

Expression of DSDG candidates was determined by qRT-PCR. Three of the five ALL cell lines strongly expressed *BMP2* mRNA. Expression was present but >10-fold lower in SEM and no detectable expression was observed in the NALM6 cell line. On the other hand, only one of four lymphoma cell lines, Jiyoye, expressed *BMP2* and even in this cell line expression was far lower than the levels observed in the positive ALL cell lines. (Figure 5.3). These results correspond with methylation pattern of *BMP2* locus in four of five ALL cell lines; low methylation pattern consist with high expression in Reh, Rch-acv and SEM cell lines, and high methylated locus of *BMP2* in NALM6 cell line corresponded with lack of expression.

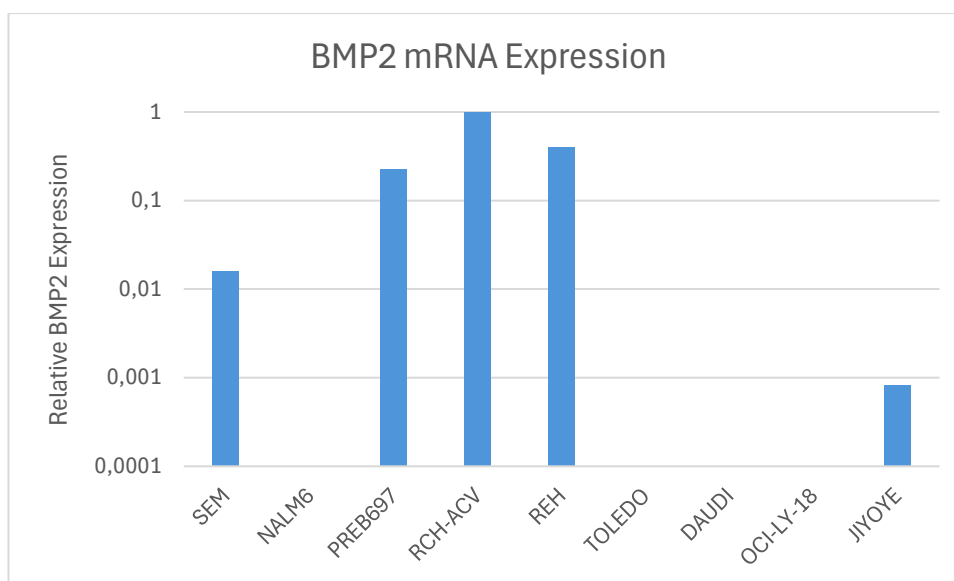


Figure 5.3. Representative graph of BMP2 mRNA expression on parental ALL and Lymphoma cell lines

qRT-PCR was performed to determine BMP2 gene expression at the mRNA level. All the genes expression were normalised against house-keeping gene GAPDH and show relative expression to the highest expressing cell line (Rch-acv). Data is derived from single experiment.

In addition to the *BMP2*, functional BMP receptors is required for BMP signalling pathway (detailed in section 5.1). In this matter, we determined BMP receptors expression on mRNA level to identify functional receptors in all cell lines by qRT-PCR. Type I BMP receptors (*BMPR1B*, *BMPR1A* and *ACVR1*) and type II BMP receptors (*BMPR2*, *ACVR2A* and *ACVR2B*) mRNA expressions are shown in Table 5.1. All ALL cell lines tested had detectable transcript of at least two type I receptors; *BMPR1A*, *ACVR1* and two type II receptor *BMPR2*, *ACVR2B*. Similar to the ALL cell line results, all lymphoma cell lines tested expressed the type I receptor *BMPR1A* mRNA and the type II receptor *BMPR2* mRNA. On the other hand, *ACVR1*, *ACVR2A* and *ACVR2B* receptors transcripts were detected on all lymphoma cell lines tested except Daudi. As a result, functional BMP type I and type II receptors present in all cell lines.

	BMPR1B	BMPR1A	ACVR1	BMPR2	ACVR2A	ACVR2B
SEM	-	+	+	+	+	+
NALM6	-	+	+	+	-	+
RCH-ACV	-	+	+	+	+	+
PREB697	-	+	+	+	+	+
REH	+	+	+	+	-	+
TOLEDO	-	+	+	+	+	+
OCI-LY-18	+	+	+	+	+	+
DAUDI	-	+	-	+	-	-
JYIOYE	+	+	+	+	+	+

Table 5.1mRNA expression status of BMP receptors in ALL and Lymphoma cell lines from in-house unpublished data.

+ indicates that transcript expressed, - indicates that undetectable on cell line. Red colour indicates Type I and black colour indicates type II receptors.

5.3.1.2 CTGF Expression Profile

As can be seen in Figure 5.4, four of five ALL cell lines expressed *CTGF* although the expression level was lower in Rch-acv. There was no detectable *CTGF* mRNA expression in three of four lymphoma cell lines and only one of lymphoma cell lines, Oci-Ly-18 showed weak expression of *CTGF*. Conversely, expression was not detectable in the lymphoma cell lines, apart from low expression observed in the Oci-ly-18 cell line.

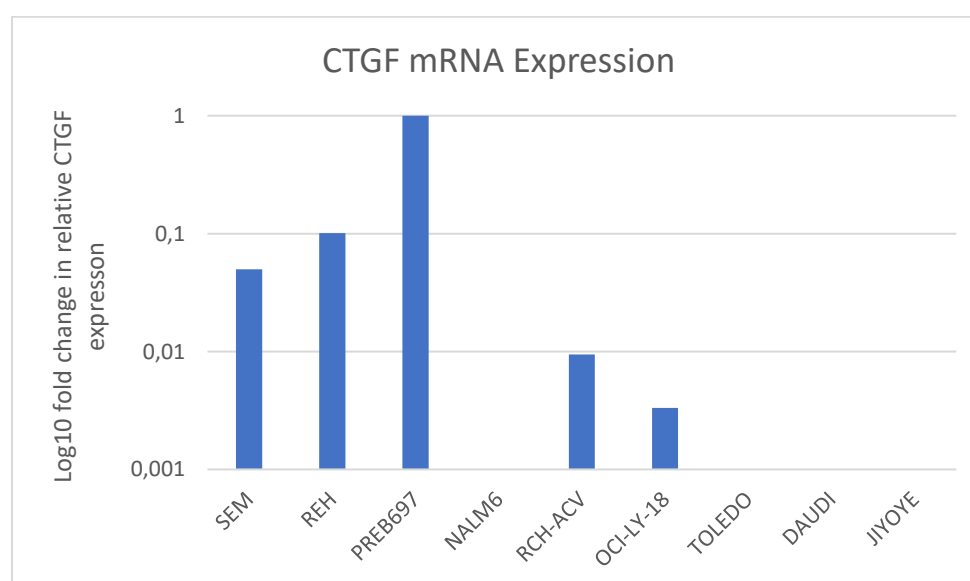


Figure 5.4.Representative graph of CTGF mRNA expression on parental ALL and Lymphoma cell lines

qRT-PCR was performed to determine CTGF gene expression. All the genes expression were normalised against house-keeping gene GAPDH and show relative expression to the highest expressing cell line (PreB697). Data is derived from single experiment.

5.3.1.3 NPY Expression Profile

As can be seen in Figure 5.5, all five ALL cell lines expressed *NPY*, although the expression level was lower in SEM in comparison to the other cell lines. There were no detectable *NPY* mRNA expression on lymphoma cell lines except Oci-Ly-18, which showed weak expression of *NPY*. To sum up, strong gene expression of the *NPY* as ALL relevant DSDG candidate based on bioinformatic analyses across B-ALL cell lines was observed against the lymphoma cell lines expressed significantly lower level on Oci-Ly-18 or non-detectable in other three cell lines tested. Expression levels of potential *NYP* receptors, *NPYR2* and *NPYR5*, was also assessed. This analysis found that only the PreB697 cell line expressed the *NPYR2* transcript (Ct mean was 38), while there was no detectable mRNA transcript of *NPYR2* in other ALL and lymphoma cell lines tested. There was no detectable *NPYR5* transcript on any cell lines tested. Data not shown in detail as only one cell line was positive for one of the assessed receptors.

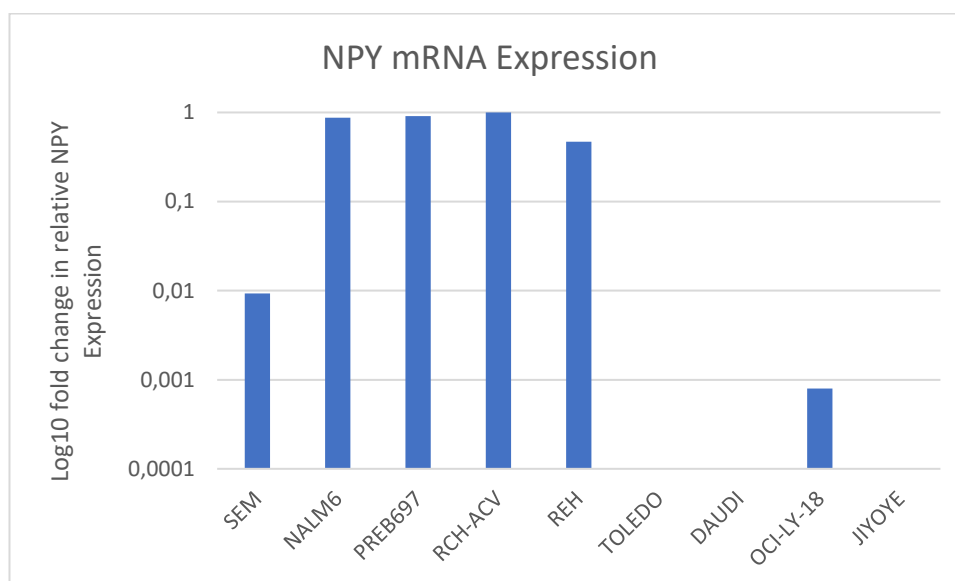


Figure 5.5. Representative graph of NPY mRNA expression on parental ALL and Lymphoma cell lines

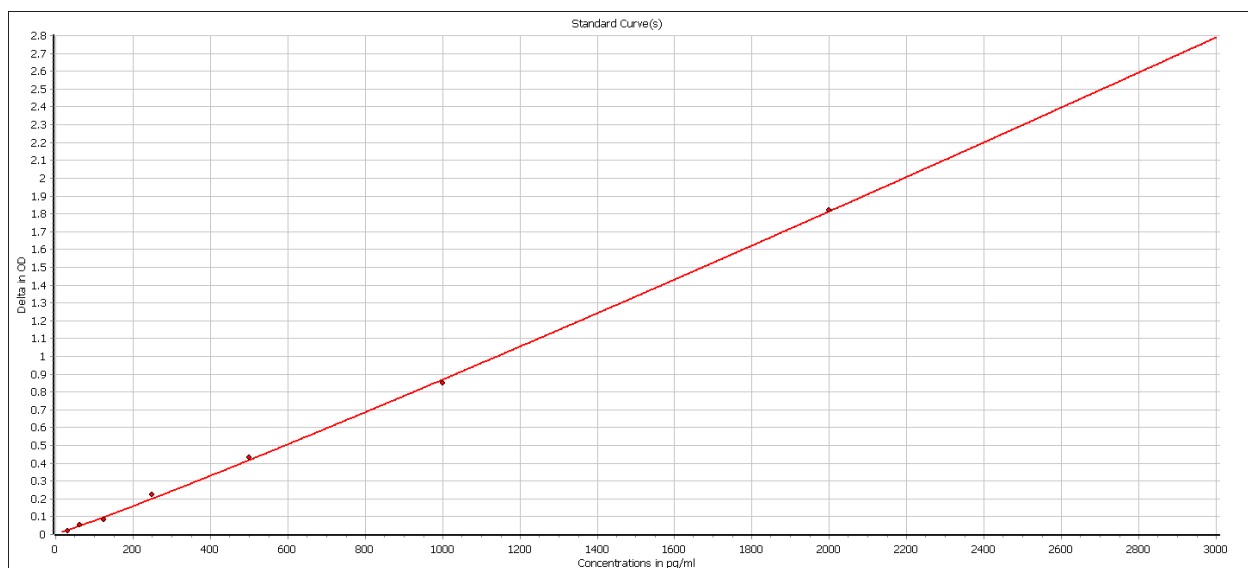
qRT-PCR was carried out to determine NPY gene expression. Gene expression was normalised against house-keeping gene GAPDH and calculated relative expression to the highest expressing cell line (Rch-acv). Data is derived from single experiment.

5.3.2 Functional analysis of *BMP2*

5.3.2.1 Assessment of secreted *BMP2* protein by ELISA

The ELISA technique was used to determine secreted *BMP2* protein on ALL and lymphoma cell lines media and was repeated on two occasions. Standard curve was plotted after each experiment based on mean absorbance of protein concentration standards. Samples were diluted on 1st ELISA experiment as kit recommendation. However, all sample were below the detectable level. Thus, the ELISA was repeated, but on this occasion the samples were used without dilution to increase sensitivity for detection. RPMI media was used to control for background noise in each experiment. In regard to second ELISA results, standard curve was plotted as explained above (Figure 5.6-A) then absolute concentration and percentage coefficient variance (CV%) of standard/samples were calculated by Omega Data Analysis program (BMG Lab Tech version 3.32) (Figure 5.6-B). As can be seen in Figure 5.6-B, absolute concentration of standards was very similar level at the kit's suggestion. On the other hand, most of the cell lines gave readings of at or below the level of control media alone. Of the five cell lines that did give readings higher than media alone, the levels were only slightly above the control and rest of four cell lines with potential low positive readings came from cell lines with no detectable *BMP2* transcript. Therefore, it is not clear whether the ELISA assay was able to detect any clear evidence of secretion of *BMP2* protein in the cell lines assessed (Figure 5.6-B).

A



B

Samples	Concentration (pg/ml) *	%CV**
STANDARD S1	2005.288	2.2
STANDARD S2	978.643	9.2
STANDARD S3	514.565	7.1
STANDARD S4	275.124	0.9
STANDARD S5	104.84	7.7
STANDARD S6	66.218	14.9
STANDARD S7	22.732	28.4
SEM	156.581	0.6
NALM6	72.776	29.3
REH	196.884	29.8
PREB697	113.171	51.9
RCH-ACV	121.955	30.7
JIYOYE	141.765	74.4
DAUDI	196.043	44.7
TOLEDO	203.663	3
OCI-LY-18	172.537	26.3
MEDIA	156.062	38.3

Figure 5.6. ELISA result of BMP2 protein expression in parental ALL and Lymphoma cell line's media

ELISA was performed to check BMP2 protein concentration in media of parental ALL and Lymphoma cell lines. A-Standard curves was plotted by mean of OD450nm reading on Y axis and concentration of standards on X axis. B- Concentrations and coefficient variant

(%) of standards and samples which are ALL and Lymphoma cell lines. Absolute concentration of standards was same as kit range and CV% was lower than 20% except lowest standard. Concentration of BMP2 protein in each cell line media were at the background level that was RPMI growing media. CV% was higher than 20% except SEM and Toledo cell lines. Data is shown derived from single experiment. (r^2 was 0.99) **picogram per ml. ** percentage of coefficient variation (%CV).

5.3.2.2 Determination of intracellular BMP2 protein expression in cell lines by Western Blotting

The ELISA results suggested that secreted BMP2 protein was low or undetectable in all cell lines. Therefore, we further wanted to check whether BMP2 protein was being produced inside the cell but not secreted. Western blotting was carried out to assess intracellular BMP2 protein expression in ALL, lymphoma cell lines along with U2OS cell line which was positive control and Toledo, Daudi and NALM6 cell lines were used as negative control based on a lack of transcript being detected by qRT-PCR. On the first attempt, there was clear detectable band on all cell lines even negative control, Toledo, NALM6 and Daudi cell lines because of the there was 100% cross reactivity human BMP4. Then we order new ab against BMP2 confirming with supplier there were less than 1% cross reactivity with other type human BMPs. Therefore, western blot was performed with second antibody against to BMP2. The predicted band size was 15-16kDa for mature monomer of BMP2 protein (against which the antibody was raised). However, no band of that size was detected in any of the cell lines, including the positive control cell line U2OS [346], (Figure 5.6). A band was observed band was around 35-36kDa size on all cell lines samples tested, although faint in U2OS cell line. However, there was no clear difference in band intensity between cell lines that were strongly positive by qRT-PCR and cell lines without detectable transcript. Thus, it does not appear that this band represents an earlier, not fully processed, BMP2 protein. On the other hand, B-actin was used as loading control and observed band was on the correct size (42kDa). Therefore, mature monomer peptide of BMP2 protein in these samples was not detected by this antibody and western blot results is not reflective of the qRT-PCR results.

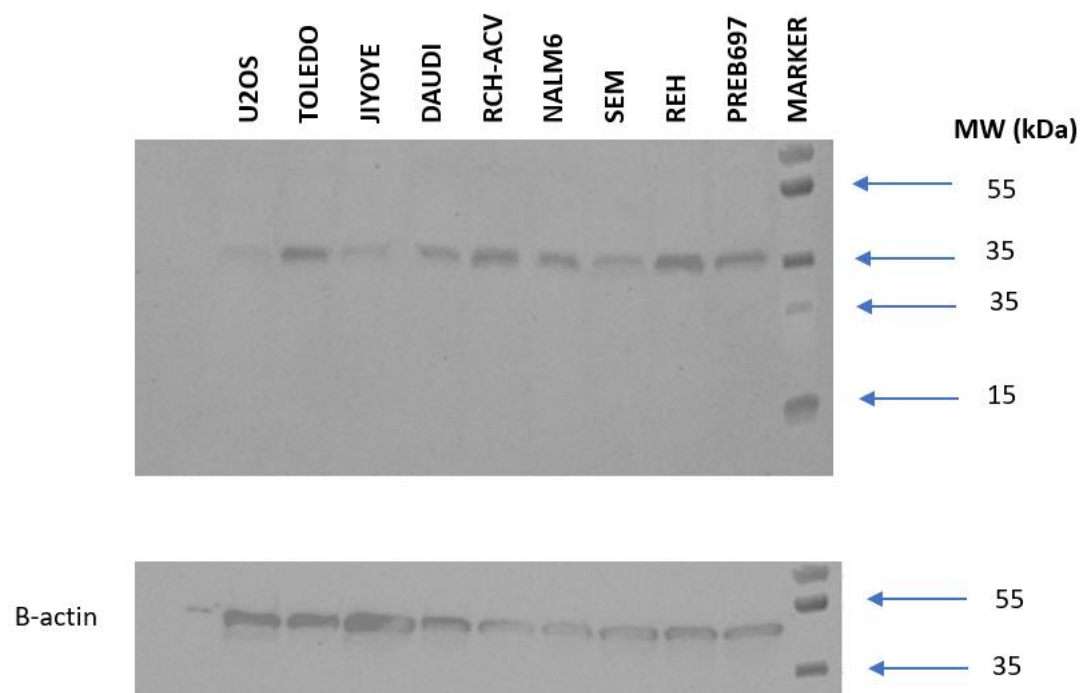


Figure 5.7. Determination of intracellular BMP2 protein by Western Blotting on ALL and Lymphoma Cell lines

Western blotting was performed to detect BMP2 protein in panel of ALL and lymphoma cell lines samples along with positive and negative control samples. Antibody detected an unknown band at 35-36kDa size but this did not match the predicted size for BMP2 and was present at similar levels in transcript positive and negative cell lines. B-actin was used as a loading control. Data is derived from a single western blot experiment. MW; molecular weight.

5.3.2.3 Determination of BMP inhibitor effect to ALL and Lymphoma cell line's growth

As the above BMP2 protein expression analysis was inconclusive, and the cell lines are all positive for BMP receptor expression, ALL except NALM6 because of lack of BMP2 expression, and lymphoma cell lines were treated with increasing doses (from 0-1 μ M) of the BMP inhibitor LDN-193189 and incubated in up to 4 days (96 hours) then assessed by XTT on both day2 (48 hours) and day4 (96 hours) to identify any impact on the numbers of viable cells. As can be seen from Figure 5.8, ALL cell lines started to exhibit reduced growth at the 0.25 μ M dose, reaching statistically significant differences comparing control

was seen at 0.5 μ M and 1 μ M concentrations on day2 and day4 comparing non-treated group ($p < 0.05$) (Figure 5.8- A and C). The four lymphoma cell lines tested showed little evidence of toxicity at 0.5 μ M but showed clear reduction in viable cells was seen after treatment with 1 μ M at day2 and day4 ($p < 0.05$) (Figure 5.8- B and D).

Comparing average reduction in growth of all ALL-cell lines at 0.5 μ M concentration at day2 and day 4 (45 and 35, respectively) versus the lymphoma cell lines at same concentration at day 2 and day4 (10 and 8, respectively), there was statistically significantly greater decrease in cell growth in the ALL cell lines at day 2 ($p < 0.05$) although this fell short of statistical significance at day 4 (Figure 5.8- E and F).

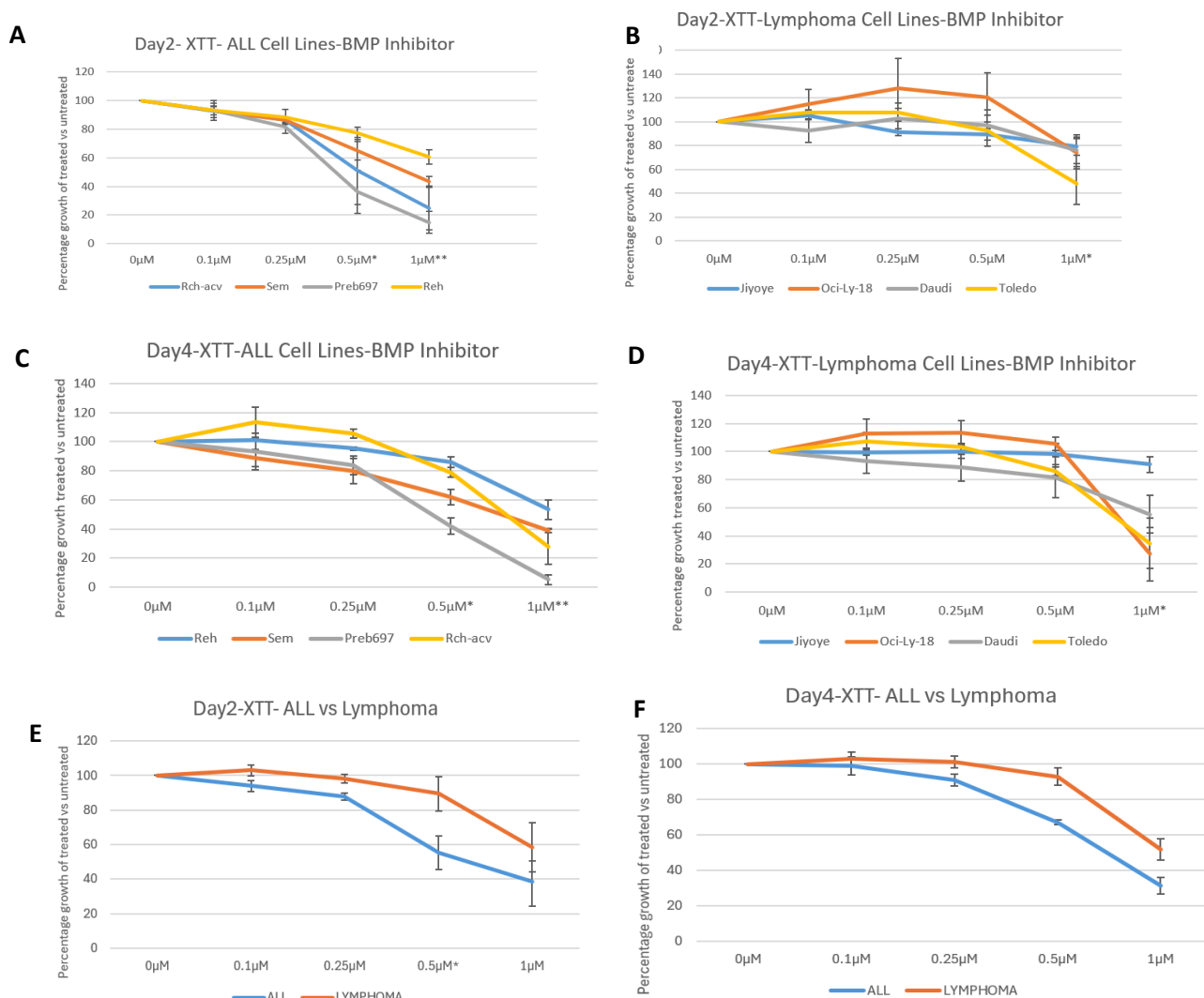


Figure 5.8. Effect of BMP inhibitor on proliferation of ALL and Lymphoma Cell Lines

Graphs show impact of LDN-193189 on proliferation of ALL cell lines (A, C) and lymphoma cell lines growth (B, D) results at day2 and day4 respectively. Growth inhibition of ALL cell

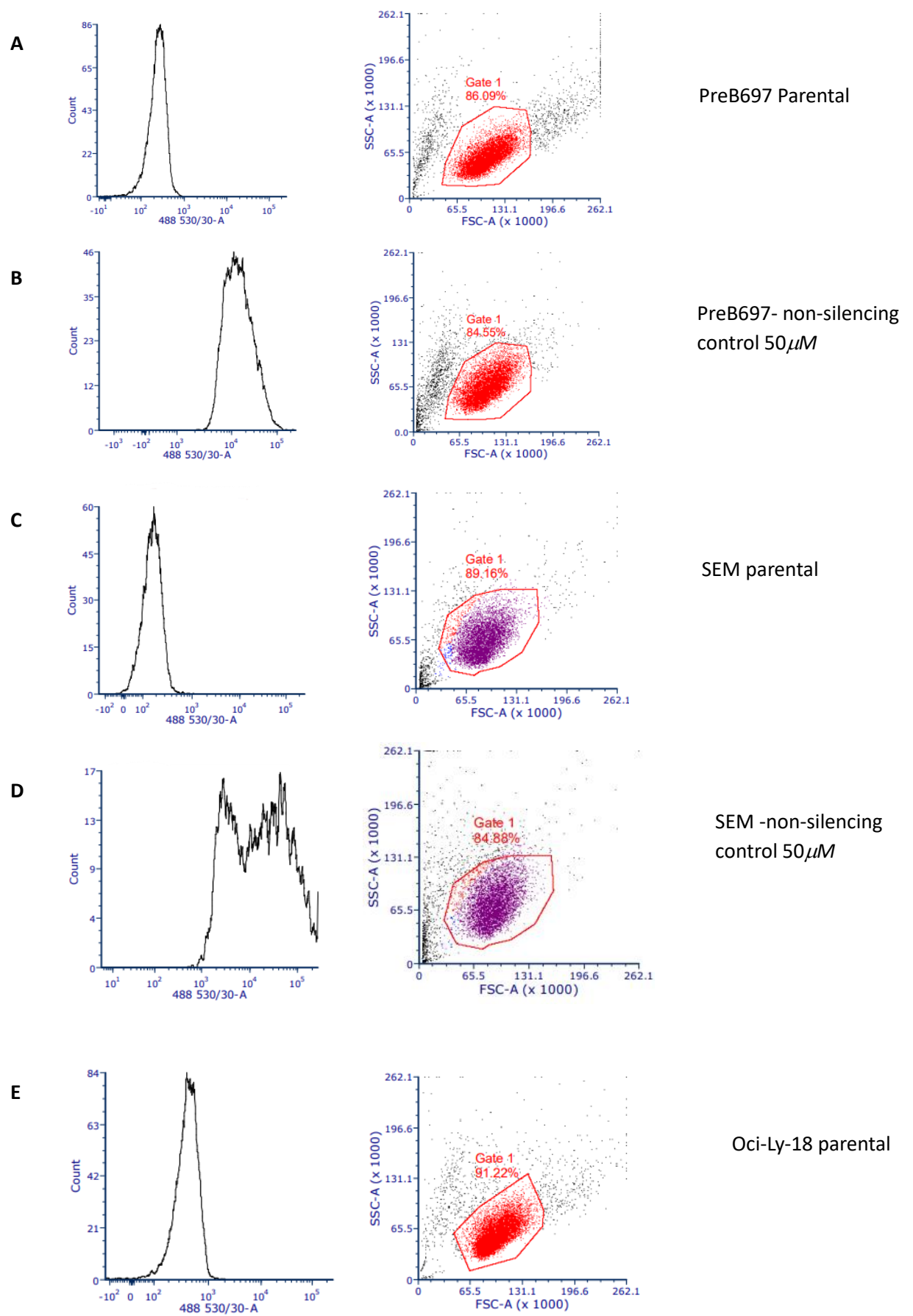
lines after exposed to BMP inhibitor at 0.5 μ M and 1 μ M concentrations either day2 or day4 showed statistically difference comparing untreated group. There was statistically significant on reduced cell growth of lymphoma cell lines at just 0.5 μ M concentrations at both day2 and day4. Average reduced cell growth of ALL cell lines at 0.5 μ M concentration was significantly differences with average reduced cell growth of lymphoma cell lines at day2 (* $p < 0.05$). Data represent average of three independent experiment and bars represent standard error between experiments. Results normalised based on untreated group (0 μ M). * $p < 0.05$, ** $p < 0.005$.

5.3.2.4 BMP2 siRNA transfection on ALL and Lymphoma Cell Lines

The above results provide some evidence of preferential sensitivity of ALL cell lines to BMP inhibition. However, the inhibitor used above is not specific for *BMP2*. Thus, to further investigate the potential functional role of *BMP2* in ALL, siRNA mediated knockdown method was attempted on ALL cell lines along with lymphoma cell lines. As described in section 5.3.1.2, PreB697 as strongly expresses *BMP2* while SEM expressed at a clearly lower level. Oci-Ly-18 and Toledo were included as negative controls. To assess whether impact of reduced expression of *BMP2* on ALL cell lines cell growth and survival, XTT and caspase 3/7 experiments performed post-transfection at 72 hr/144 hr and 48 hours respectively on both ALL and Lymphoma cell lines.

5.3.2.4.1 Transfection Efficiency

To optimise the transfection conditions cell lines were initially transfected with fluorescent non-silencing siRNA to allow assessment of transfection efficiency and cell viability by flow cytometer after 24 hours of transfection. Whole cell populations of PreB697, SEM, Toledo and Oci-ly-18 shifted right (i.e. exhibited increased fluorescence) at 50 μ M concentration when compared with the parental cell populations, indicating that the transfection was successful and transfection efficiency was close to 100% (Figure 5.9). In addition, cell lines viability in the transfected cell cultures were close to the parental cell populations (Figure 5.9).



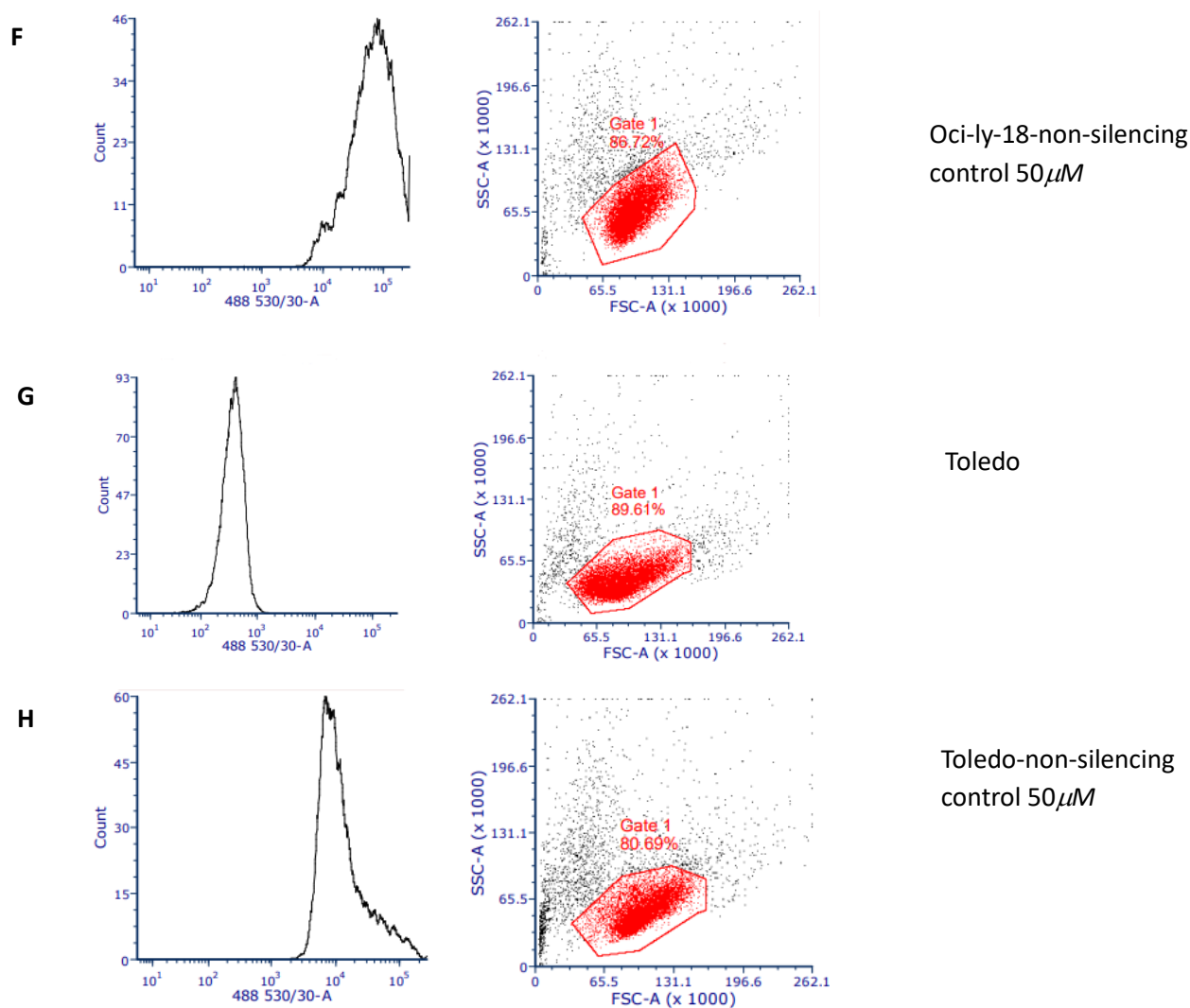


Figure 5.9. Determination of Transfection Efficiency of ALL and Lymphoma Cell Lines

PreB697 (B), SEM (D), Oci-Ly-a8 (F) and Toledo (H) were transfected with non-silencing siRNA duplex labelled with a AF488 fluorescence tag. Flow-cytometer was performed at 24 hours after transfection. In addition to, parental cell population of PreB697 (A), SEM (C), Oci-Ly-18 (E) and Toledo (G) were used to show differences of transfection efficiency and cell viability after transfection. Histograms which are located on left in the figure represent each cell population showed increased intensity and relate to entire cell population shifted right when compared to the parental cell population as a result of high transfection efficiency. Graphs which are located on right in figure based on size scatter (SSC-A in the figure) on Y axis and forward scatter (FSC-A in the figure) on X axis shows cell viability There was minimal effect of transfection reagent on cell viability in testing group when compared non-transfected parental cell populations. Data is derived from single experiment in this figure.

5.3.2.4.2 Knockdown Efficiency of *BMP2*-siRNA Transfection

Cell lines were transfected with *BMP2* specific siRNA (and a non-silencing control) using the optimised conditions defined above. qRT-PCR was carried out to assess the reduction in *BMP2* transcript after *BMP2* siRNA mediated transfection in both ALL and lymphoma cell lines at 48 hrs post-transfection. This identified only minimal evidence of knockdown, with the PreB697 cell line averaging knockdown efficiency of 20% and no detectable knockdown in the SEM, Oci-ly-18 and Toledo cell lines. Experiment was repeated three time and average presented in Figure 5.10.

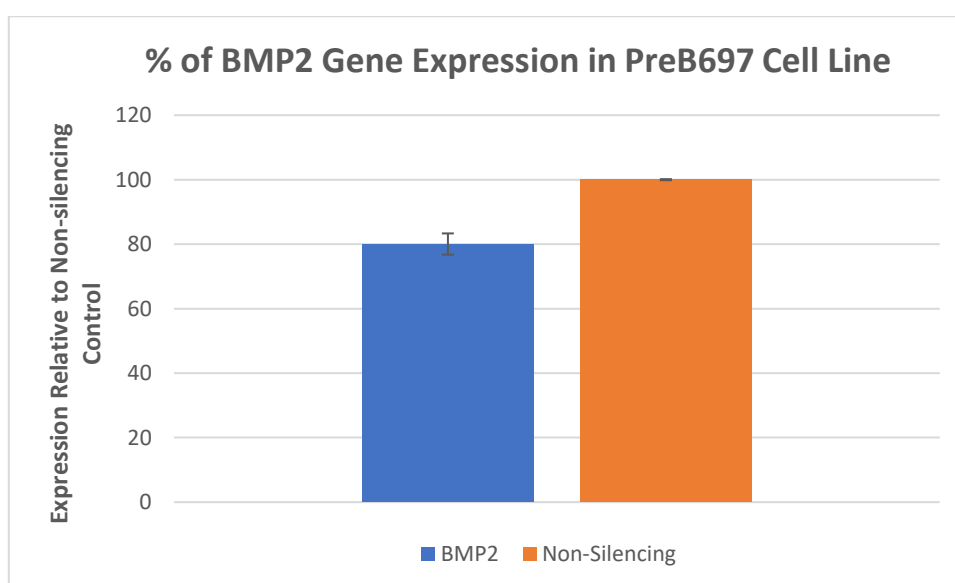


Figure 5.10. siRNA mediated knockdown result of *BMP2*

Expression level, relative to cells transfected with the non-silencing control siRNA, in preB697 cells 48 hours after transfection with BMP2 specific siRNA. Data is an average from three qRT-PCR replicates of three independent transfections. Error bars represent SEM.

5.3.2.4.3 Effect of *BMP2* siRNA transfection on Cell Proliferation

The above analysis suggests that transfection with *BMP2* siRNA had a relatively limited effect on *BMP2* expression. However, even minor reductions in expression could potentially impact cellular function. To assess if this was the case, XTT assays were performed at post-transfection 72hours and 144 hours to assess whether negative effect of *BMP2* siRNA mediated transfection or not upon cell proliferation in ALL and lymphoma

cell lines. For both groups, although there was slight decrease on cell proliferation on both 72 hours and 144 hours after *BMP2* siRNA mediated transfection when compare non-silencing control, is the difference was not statistically significant and there was no evidence of specificity for cell lines that express *BMP2* (Figure 5.11).

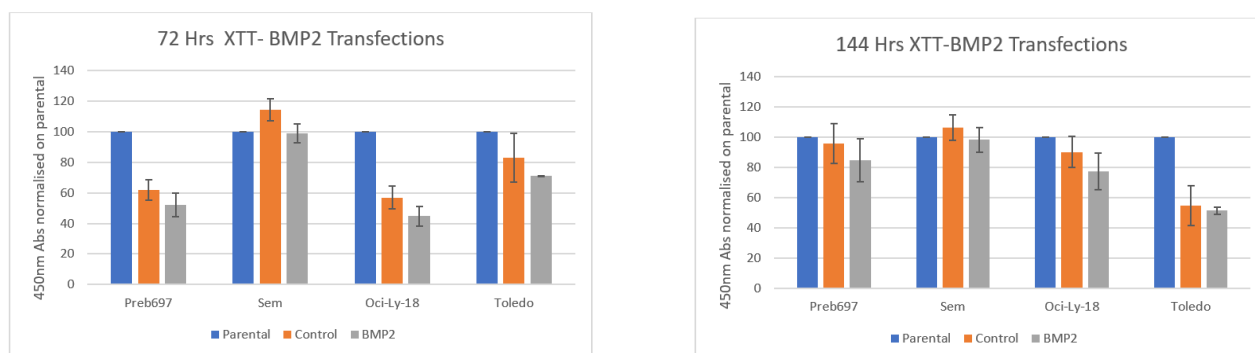


Figure 5.11. Cell Proliferation after *BMP2* siRNA transfection in ALL and Lymphoma cell lines

Cell lines were transfected with the indicated siRNA (or untransfected (parental)) and assessed for viable cells by XTT assay at 72 and 144 hours post transfection. Data comes from average of three independent experiment results and bars depict standard error between experiments. Results normalised to the values obtained with parental (non-transfected) cell lines.

5.3.2.4.4 Effect of *BMP2* siRNA transfection on Cell Survival

Caspase-glo 3/7 assay was carried out at 48hours post-transfection to determine whether the *BMP2* siRNA transfection resulted in induction of cell death and whether this was specific for ALL cell lines. As shown in Figure 5.12, SEM, there was no statistically significant increase in caspase activity ($p > 0.05$) compared with the control group (non-silencing siRNA) (Figure 5.12). Increased caspase activity is seen following transfection in the PreB697 cell lines; however, this was not significantly different between the control and *BMP2* specific siRNAs and thus is likely a non-specific effect as a result of the transfection procedure.

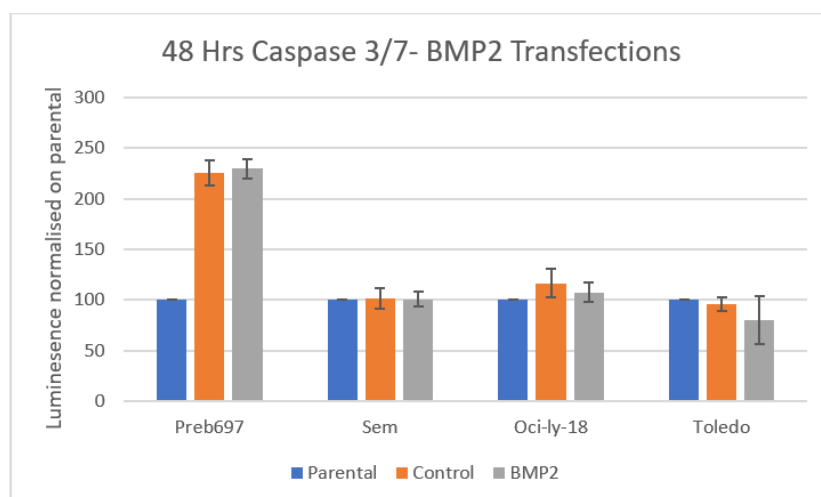


Figure 5.12. Results of Caspase 3/7 activation after BMP2 siRNA mediated transfection on ALL and Lymphoma Groups

There was no increase in caspase activity detected in any of the cell lines after transfection with the BMP2 siRNA versus the control (non-silencing). Data comes from average of three independent experiment results and bars depict standard error between experiments. Results normalised on parental.

5.3.3 Functional analysis of CTGF

5.3.3.1 Determination of secreted CTGF protein on cell lines media by ELISA

As for *BMP2*, *CTGF* is a secreted protein and could impact on cell growth via an autocrine mechanism. However, attempt to detect secreted *CTGF* (using the human *CTGF* Mini ABTS ELISA Development Kit (from PeproTech)), were not successful. The ELISA was repeated three times, however, the absolute standard concentrations based on standard curve plot was not reflective of the expected values and CV% of standards was very high because of the variation between duplicates of standard/samples. Therefore, the data obtained was not reliable (data not shown) and alternative approaches may be required to assess if ALL cells do indeed secrete CTGF.

5.3.3.2 CTGF siRNA transfection on ALL and Lymphoma Cell Lines

To investigate functional role *CTGF* expression in ALL as DSDG candidates, siRNA mediated knockdown method was attempted in two ALL and, as controls, two lymphoma cell lines. As described in section 5.3.1.2.2, PreB697 exhibited the highest expression of *CTGF* at the RNA level, while SEM expressed about 20-fold lower levels. Oci-Ly-18 expressed BMP2 mRNA transcript at a very low level and Toledo was negative. Either synthetic *CTGF* siRNA smart pool or non-silencing duplexes as a control were used for transfection. To assess impact of reduced expression of *CTGF* on ALL cell lines whether result in cell growth and/or survival, XTT and caspase 3/7 experiments performed post-transfection at 72 hr/144 hr and 48 hours respectively on both ALL and Lymphoma cell lines.

5.3.3.2.1 Knockdown Efficiency of CTGF-siRNA Transfection

qRT-PCR was carried out to assess whether reduced *CTGF* transcript after *CTGF* siRNA mediated transfection in both ALL and lymphoma cell lines, cell pellets were collected from each cell lines at 48 hrs post-transfection. For the ALL cell line, the average knockdown efficiency was relatively low for both cell lines, with 22.9% and 29.9% for PreB697 and SEM respectively. In the lymphoma cell lines, average knockdown efficiency was 30.4% for Oci-ly-18 cell line. The Toledo cell line does not express *CTGF* (Figure 5.13).

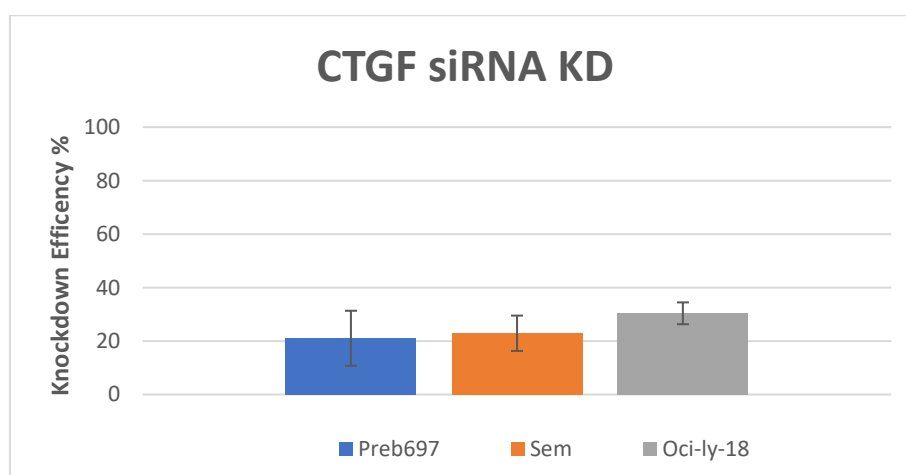


Figure 5.13. siRNA mediated knockdown result of CTGF

qRT-PCR was performed to assess KD% efficiency at post-transfection 48 hours. Average KD efficiency was 29.9% for PreB697, 22.9 % for SEM and 30.4% for Oci-Ly-18 cell lines when compared to control (non-silencing siRNA transfection Data based on qRT-PCR of three independent experiments and bars depict standard error. qRT-PCR results were normalised against to house-keeping gene GAPDH.

5.3.3.2.2 Effect of CTGF siRNA transfection on Cell Proliferation

Although, the efficiency of knockdown following siRNA transfection was low, it was consistently detectable in all *CTGF* expressing cell lines. Thus, to determine if this impacted on cell growth or survival, the XTT assay was performed at 72hours and 144 hours post-transfection. As can be seen on Figure 5.14, PreB697 and SEM cell lines exhibited a slight reduction on cell growth following *CTGF* siRNA mediated transfection, particularly at 144-hours post-transfection when compare non-silencing control. However, this difference did not reach statical significance ($p>0.05$) when compare non-silencing control. In the lymphoma cell lines, there was generally no reduction in the *CTGF* siRNA transfected cells versus the control siRNA, except a small, again non-significant, reduction in Oci-ly-18 at 72 hours post transfection.

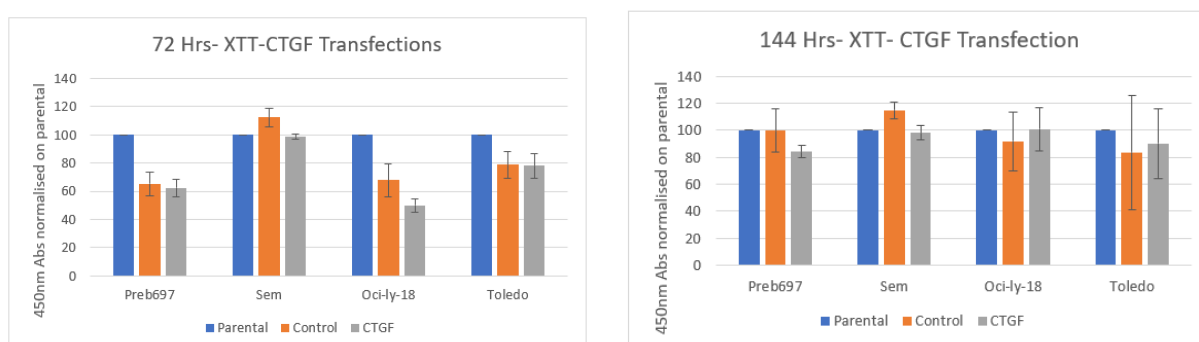


Figure 5.14. XTT results after CTGF siRNA transfection

CTGF siRNA transfection displayed slightly negative impact upon cell proliferation on ALL cell lines at 72 hours and 144 hours but there were no statistically significant differences compared with the control. Of the lymphoma cell lines, only Oci-ly-18 cell lines gave a small decrease at 72 hours, however, this was also not statically significant. Data comes from average of three

independent experiment results and bars depict standard error. Results (or absorbance at 450nm) normalised to the parental, non-transfected cells.

5.3.3.2.3 Effect of CTGF siRNA transfection on Cell Survival

Following to transfection at 48 hours, the Caspase-glo 3/7 assay was carried out to determine whether CTGF2 siRNA transfection resulted in induction of caspase activity. As can be seen in Figure 5.15, there was no clear CTGF siRNA specific increase in caspase activation in any of the cell lines. Non-specific activation of caspase activity was seen in PreB697 cells (as previously observed above following transfection with control and BMP2-specific siRNA). There was a higher level of caspase activity in this cell line following transfection with the CTGF siRNA (versus control) but this difference was not statistically significant ($p>0.05$).

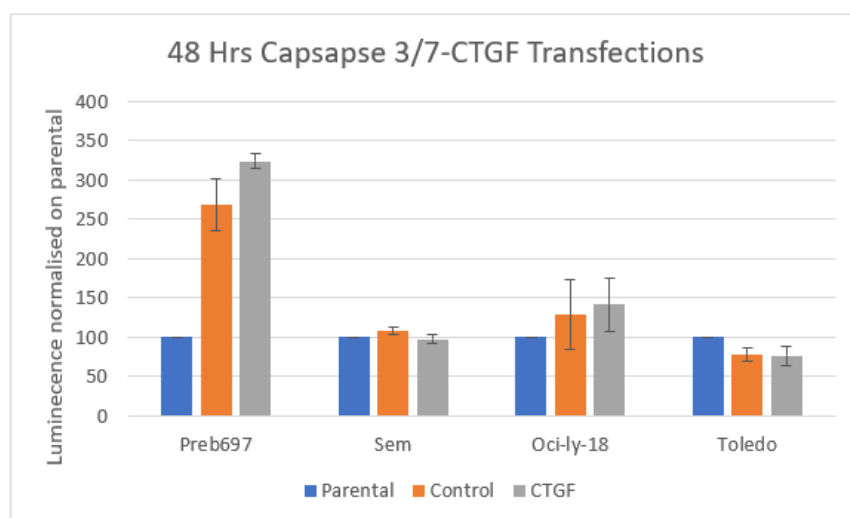


Figure 5.15. Detection of apoptosis dependent cell death after CTGF siRNA mediated transfection

Caspase-glo 3/7 assay was performed at post-transfection 48 hours showed no significant activation of caspase activity specifically associated with the CTGF siRNA transfections. There was small increase in the PreB697 cell line, but this was not statistically significant. Data is average from three independent experiment and bars represent standard error between experiments. Luminescence reading normalised to the parental cells.

5.3.3.3 Antibody against to CTGF

As mentioned in section 5.3.3.1, ELISA results were inconclusive to determine secreted *CTGF* protein by cell lines as requirement of functional autocrine signalling, therefore, we wanted to test whether secreted *CTGF* in cell lines media and if so, inhibiting extracellular *CTGF* via anti-CTGF ab to determine role of *CTGF* on ALL. Thus, ALL and lymphoma cell lines were treated with the anti-CTGF Ab (Pamrevlumab (FG-3019) at 100 μ g/ml and 30 μ g/ml to determine if this impacted cell growth/survival due to blocking of any secreted *CTGF*. The concentration of FG-3019 was selected based on a previous report [347] and [348]. After addition of the inhibitory antibody, the XTT assay was carried out at 48 hours and 96 hours post treatment. At 48hrs, there was no and/or only a slight reduction of cell growth in all cell lines and no statistically significant reductions were seen ($p > 0.05$ for all cell lines). At 96 hours, again no statistically significant decrease in viable cells was seen ($p > 0.05$, for all cell lines). In the lymphoma both cell lines showed slightly reduction of cell proliferation when treated with the 30 μ g /ml dose on both day ($p > 0.05$) (Figure 5.16). However, this wasn't consistently seen, and no reduction was seen with the higher dose of antibody. Overall, there was no clear evidence that anti-CTGF antibody impacted the growth of either ALL or lymphoma cell lines regardless of their *CTGF* expression status.

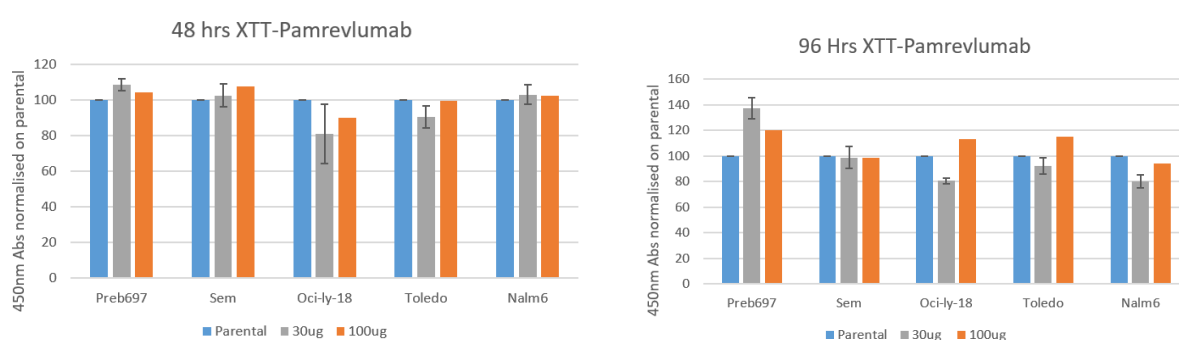


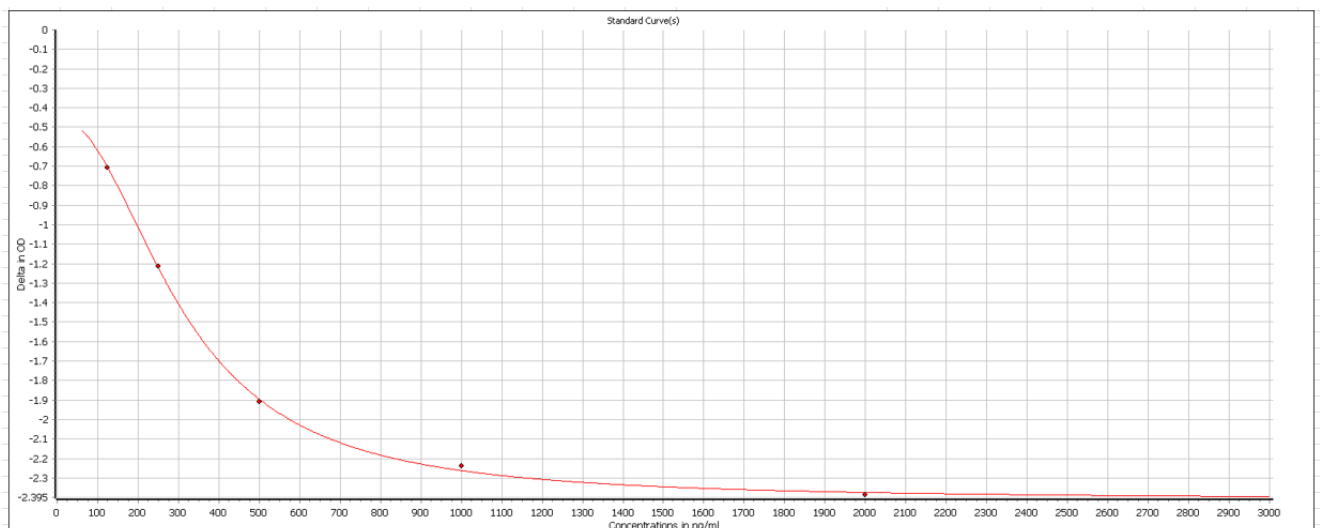
Figure 5.16. Results of anti-CTGF antibody on All and Lymphoma cell Lines

XTT assay was performed at 48 and 96-hours post-transfection. None of the cell lines exhibited consistent statistically significant reductions in viable cells at either of the doses tested. Data comes from average of two independent experiment results and bars depict standard error between experiments. Results (or absorbance at 450nm) normalised on parental.

5.3.4 Functional assessment of *NPY*

5.3.4.1 Determination of secreted *NPY* protein in cell lines media by ELISA

An ELISA assay was used to assess secreted *NPY* protein in media taken from ALL and lymphoma cell line cultures. The assay used was a competitive ELISA and consequently there was inverse correlation between OD measurements and concentration of *NPY* in the assay samples. A standard curve was plotted after each experiment based on mean absorbance on Y axis and protein concentration of standards on X axis (Figure 5.17-A). Then absolute concentration and percentage coefficient variance (CV%) of standard/samples were calculated by Omega Data Analysis program (Figure 5.17-B). The assay was repeated twice and gave similar results showing that the standards gave results close to the expected values and with relatively low coefficients of variation (CV%) in the majority of standards, although the second and lowest standard had higher CV% (31 and 45%, respectively). However, as can be seen in Figure 5.17-B, all samples gave undetectable readings, even though several of the cell lines were strongly positive in the qRT-PCR assay. Furthermore, this lack of ability to identify secreted *NPY* was not limited to ALL cell as two further positive (by qRT-PCR) cell lines were included SKNAS and GIMEN (both neuroblastoma cell lines) and these also gave undetectable reading.



Sample	Concentration (pg/ml) *	%CV**
STANDARD 1	2140,178	0
STANDARD 2	990,439	30,9
STANDARD 3	509,74	4,3
STANDARD 4	248,913	14,4
STANDARD 5	122,033	44,6
PREB697	Undetected***	n.a.
REH	Undetected***	n.a.
RCH-ACV	Undetected***	n.a.
SEM	Undetected***	n.a.
NALM6	Undetected***	n.a.
DAUDI	Undetected***	n.a.
JIYOYE	Undetected***	n.a.
TOLEDO	Undetected***	n.a.
OCI-LY-18	Undetected***	n.a.
SKNAS	Undetected***	n.a.
GIMEN	Undetected***	n.a.

Figure 5.17. ELISA result of NPY protein expression in parental ALL and Lymphoma cell line`s media

ELISA was performed to assess the presence of NPY protein in media used to grow ALL and Lymphoma cell lines. A-Standard curves was plotted by mean of OD450nm corrected to the blank reading on Y axis and concentration of standards on X axis B- Concentrations and percentage coefficient variant of standards and samples. Assessed Concentration of standards was similar to the known concentrations and percentage CV was lower than 20% except second and lowest standard. There was no detectable NPY protein in ALL, lymphoma or positive control cell lines and readings were at the background level derived from unused RPMI media. Data is shown derived from single experiment. (r^2 was 0.99) **picogram per ml. ** percentage of coefficient variation (%CV). ***Undetected refer to a reading at or below the reading given in the control media alone (i.e. fresh media never used for growing cells) wells.

5.3.4.2 Determination of intracellular NPY protein by Western Blotting

From the results described above it is clear that most of the ALL cell lines are clearly positive for NPY expression at the RNA level, but that secreted protein could not be

detected. This could be because the transcript is not translated or it is translated, but some or all of the *NPY* protein that is produced remains intracellular. Therefore, we further wanted to check whether intracellular *NPY* protein could be identified in ALL and lymphoma cell lines. The ALL cell lines SEM, PreB697 and NALM6 were used for ALL, as all three are positive, but SEM expresses about 100-fold lower levels of transcript. Three lymphoma cell lines (Daudi, Jiyoye and Toledo) were used as negative controls based on an absence of any detectable *NPY* mRNA expression (detailed in 5.3.1.2.3). Two neuroblastoma cell lines, SH-SY5Y and GIMEN were also included as additional non-ALL cell lines that were also strongly positive at the RNA level. According to the manufacturer, expected size of mature *NPY* protein was 11kDa size but as can be seen in Figure 5.18, there was no detectable *NPY* protein on any cell lines tested, even those strongly positive for the transcript. All samples (excluding Jiyoye) gave a clear positive with the B-actin loading control. A background band was seen at around 80kDa, however, this is too large to represent *NPY* and is a known background band reported by the manufacturer for this antibody. The western blot was repeated with the same cell lines however, results was identical.

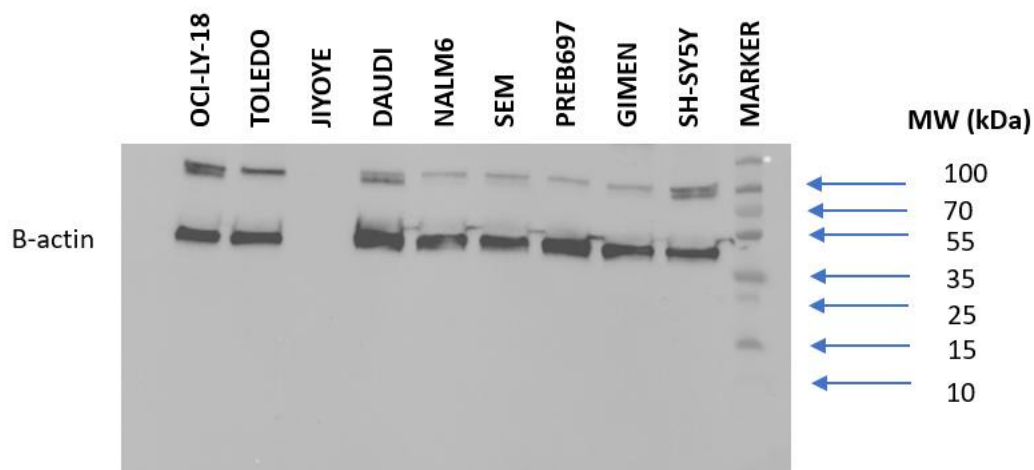


Figure 5.18. Detection of intracellular *NPY* protein

Western immunoblotting was performed to assess *NPY* protein expression in ALL and lymphoma cell lines. GIMEN and SHSY5Y cell lines were included as additional, non-ALL, cell lines that were positive at the RNA level. No mature *NPY* protein, with an expected size of 11kDa, could not detected at any of the cell lines. B-actin was used as a loading control. JIYOYE failed to exhibit any bands, even for B-actin, suggesting that there was a technical

issue with this sample, but detection of B-actin was as expected for all other cell lines. Data is derived from a single western blot experiment.

5.4 Discussion

The original bioinformatic analysis identified four DSDG candidates (*BMP2*, *CTGF*, *NPY* and *ZNF423*) predicted to be specific for ALL. The primary aim of this chapter was investigating the potential functional role of the identified DSDG candidates in ALL with the aim of identifying novel therapeutic targets that could contribute to improving the management of ALL. To represent different cytogenetic subgroups, multiple ALL cell lines were chosen for functional validation that are derived from different ALL cytogenetic subgroups. *BMP2*, *CTGF* and *NPY* genes were chosen for functional tests in this study. The role of *ZNF423* in ALL has been identified and also both DMR was located over 18kb away from the TSS and limited time in the laboratory, *ZNF423* gene was removed from functional tests.

5.4.1 All Candidate Expression Profile

5.4.1.1 mRNA expressions of Candidate Genes

As three candidate genes tested in here are secreted proteins. To understand their functional roles in ALL, we started to investigation with analyses their mRNA expression along with their receptors. In addition, ELISA was performed to determine they were secreted by ALL cell line to investigate whether they have functional autocrine pathway. Also, intracellular protein expression of candidate's genes were carried out by western blotting.

We started investigation of *BMP2* on ALL with we wanted to check *BMP2* mRNA expression, it's receptor expression at RNA level, secreted *BMP2* protein on cell lines media and intracellular *BMP2* protein expression. To identify whether functional *BMP2* signalling present on ALL cell lines, *BMP2* and its relevant mRNA expression were determined by qRT-PCR. Four of five ALL cell lines were found as expressed *BMP2* transcript (except NALM6 cell line). In addition to, as requirement of functional *BMP2*

signalling, type I and type II BMP receptors expression identified; detectable mRNA expression of *BMPR1A*, *ACVR1* and two type II receptor *BMPR2*, *ACVR2B* have been determined on all ALL cell lines. Conversely, *BMPR1A* and *BMPR1B* were type I receptors that associated with *BMP2* and *BMPR2*, *ACVR2A* and *ACVR2B* were type II receptors that associated with *BMP2* based on Phylogenetic analysis in previous reports [349] [350].

Similar to *BMP2* results, *CTGF* mRNA expression was detected in all ALL cell lines apart from NALM6. This results consistent with previous results that *CTGF* was found specifically elevated in 70-80% B-ALL samples compared with T-ALL and normal peripheral blood/hematopoietic bone marrow samples [351], [254], [352], [291]. Also, Reh, RS4;11 cell lines were found expressed *CTGF* transcript while NALM6 cell was reported as negative cell line based on *CTGF* mRNA transcription in previous report [252].

Regarding with *NPY* gene results, all of ALL cell lines showed strong positive *NPY* mRNA expression. Expression levels of potential *NYP* receptors, *NPYR2* and *NPYR5*, was also assessed. only the PreB697 cell line was found weakly expressed the *NPYR2* transcript while there was no detectable mRNA transcript of *NPYR2* in other ALL and lymphoma cell lines tested. There was no detectable *NPYR5* transcript on any cell lines tested.

5.4.1.2 Protein expression of Candidate Genes

All three of the genes that were assessed as potential DSDG were all encoding proteins that are typically expected to function following secretion from the cell. However, in none of the cases could secreted protein be identified. For one of the ELISA assays used there were significant issues with the controls and thus the results of this assay (for *CTGF*) cannot be takes as reliable. However, for the *NPY* and *BMP2* assays the control standard curves suggested the assays were functioning as expected, but still not protein could be identified. This could indicate that the proteins were either not secreted or that they were secreted at levels too low to be readily identified. In addition, although the standard curves in these two ELISA assays suggested that the assay was working well, secretion could not be identified in any cell line. There was no standard “high secretion” cell line, known to be positive by ELISA, that could be identified for either assay and so it can be excluded. In contrast to our result of *BMP2*, Dai Q. et al [353] found elevated *BMP2* protein expression detected by ELISA in B-ALL patients serum comparing control. The

reason behind these results might be short half-life of *BMP2* protein that mentioned on previous report [324]. *BMP2* and *CTGF* were previously identified as upregulated gene in primary B-ALL sample's bone marrow by microarray analysis [254]. Contrary to our results of *NPY*, one study report that plasma *NPY* level in patients with paediatric B- cell precursor ALL comparing control group by radioimmunoassay technique [253]. Potentially, this could be explained by increased stability of *NPY* protein in serum versus in vitro conditions or if the assay used was more sensitive than the assay used in this study.

Western blotting was performed to eliminate reason behind the ELISA results that might be lack of intracellular mature *BMP2* and *NPY* protein expression, however because of the limited time in the laboratory *CTGF* western blot could not be attempted. In terms of *BMP2* protein results, there were no detectable protein for *BMP2* in any cell line. The underlying reasons might be that *in vivo* extracellular Furin protease is required to obtain mature *BMP2* protein [321] and antibody may not recognise the pre-cleaved form of *BMP2*. In addition, the half-life of *BMP2* protein has been reported to be very short [319], [320], [324], therefore it may not have been detected in any cell line tested in this study due to lack of stability. Regards of *NPY* western blot, there were no expected band for mature *NPY* protein in these cell lines tested. Similarly, for *NPY*, the antibody was against the mature version may not recognised non-fully processed *NPY* if mature version of the protein is not produced by ALL cells. Alternatively, the *NPY* protein may not be stable in the ALL cell lines or mRNA of *NPY* may be blocked from translation.

5.4.2 siRNA mediated knockdown of candidate genes

To explore potential role of candidates as cancer specific gene, siRNA transfection was performed to knockdown of *BMP2* and *CTGF* then subsequently cell proliferation assay, XTT, and cell survival experiment Caspase 3/7 were carried out to identify whether knockdown of the gene have negative effect on cell growth/survival. Although high transfection efficiency, knockdown efficiency was not optimal and significant amount of candidate genes expression remained in cell lines tested. Therefore, role of *BMP2* and *CTGF* on ALL was not determined in present study. This may suggest that assessment of these genes will require a different approach. In the lab we are currently investigating the

potential for using a CRISPR/CAS9 based approach to allow more efficient assessment of the potential function of identified DSDG and synthetic lethal gene candidates.

Contrary, a previous study by Hongbo Lu and colleagues found that silencing of *CTGF* expression via shRNA in Reh and RS4;11 cell lines resulted in reduction of leukaemia cell lines' growth due to inactivation of the AKT/mTOR pathway and elevated level of cyclin-dependent kinase inhibitor, p27 [252].

5.4.3 Inhibitors

Due to the technical challenge of siRNA on suspension cell line models, we decided to perform inhibitor of candidate to assess whether they have role on ALL. As mentioned in section 5.1.2.1, there is no available selective inhibitor for *BMP2*, therefore a less selective inhibitor of BMP signalling, that targets multiple receptors, including those expressed on ALL cells, was utilised. There was statistically significant reduced growth of ALL cell lines at 0.5µM and 1µM concentrations after treated with specific inhibitor of BMP receptors compared with the untreated control. Comparison of ALL and lymphoma cell lines treated with BMP inhibitor shown statistically significant difference between growth inhibition in the ALL and lymphoma cell lines at 0.5µM dosages at day2. Briefly, these results suggest that there was significant effect of BMP pathway inhibition on specifically ALL cell line model's growth comparing lymphoma cell line models.

For the same purpose mentioned above, an antibody against to *CTGF* (Pamrevlumabor FG-3019) was used to inhibit activity of *CTGF* in cell lines. There was no evidence of growth disadvantage in the cell lines after treated with inhibitor ab against *CTGF*. These results was similar to the previous report results that FG-3019 was used to investigate role of *CTGF* inhibition on ALL [252] and pancreatic cell line models [348], FG-3019 did not impact of cell growth in vitro. To overcome, Lu et al examined FG-3019 combination with/out conventional therapeutic agent in the xenograft using patient-derived leukemic cells. Their results displayed that when conventional chemotherapy is combined with FG-3019 resulting in prolonged survival in mouse models however, FG-3019 did not show effect on progression of leukaemia in vivo [252].

Because of the non-sufficient mRNA expression of *NPY* receptors (*NPYR2* and *NPYR5*), this approach did not apply to *NPY* candidate gene.

Overall, we have identified four DGDS gene candidates across the ALL. *BMP2*, *CTGF* and *NPY* genes were chosen for functional analysis to determine their role in ALL. All candidates tested encodes secreted protein. However, cell lines models were not precisely imitate microenvironment so ELISA experiments might the failed to detect this secreted protein. In addition to, siRNA mediated transfection to knockdown of candidates was failed. Therefore, their role in ALL could not be identified in the present study. To overcome this, trying different silencing methods like short hairpin RNA (shRNA) or Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) would be worthwhile in the future to verify candidates` disease specific dependency role on ALL.

6 FINAL DISCUSSION

6.1 Overview of Study

The development of cancer is associated with the accumulation of genetic and epigenetic changes within cells resulting in uncontrolled cell growth and proliferation, [354]. Two key changes in DNA methylation pattern in cancer cells have been observed including global reductions in the in total 5-methylcytosine levels and increased methylation at promoter-associated CpG island [355]. Tumour suppressor genes may acquire hypermethylation at their promoter CpG islands, resulting in gene inactivation [152], on the other hand, some highly methylated genes and repetitive sequences can be over-expressed or activated due to hypomethylation, potentially contributing to genomic instability and cancer development [356]. Consequently, DNA methylation pattern in cancer is quite complex [357]. Even though DNA methylation is complex in cancer and can appear to be stochastic, individual types of cancer clearly display cancer type-specific DNA methylation patterns. Moreover, different cancer types which arise from same normal cells, such as leukaemia and lymphoma derived from B lymphocytes, display altered patterns of DNA methylation which are highly similar [152], suggesting that the cell of origin is a key driver of the abnormal methylation patterns seen in cancer cells.

Prior to this study, our group reported a novel bioinformatic approach which based on integration genome-wide methylation data with genome-wide expression data to identify functionally relevant genes in specific cancer subtypes. The hypothesis underlying this approach was that the vast majority of methylation changes that occur in cancer development are predictable consequences of proliferation of the specific cell type involved and the failure of these predictable changes to occur at a specific gene in a specific subtype implies that a selective pressure must exist in that subtype to prevent the acquisition of methylation of that gene and so enable retention of expression. This original approach was initially applied to ALL genetic subtypes but subsequently expanded to multiple types of haematological and solid tumours and was found to identify candidate genes in almost all subtypes assessed which had a high rate of functional confirmation (80% of candidates analysed to date). However, a limitation to this approach is that any genes identified could only be relevant in a single molecular subtype which would often represent a very small number of patients. In this study, we have aimed to build on this

initial approach by expanding the main level of analysis from a molecular subtype to the whole disease level and in this way aim to identify potential new therapeutic targets that would be applicable to larger patient groups (i.e. all patients from a single cancer type). We hypothesised that this may be possible for multiple cancer types derived from the same normal cell type as the “default” pattern of acquired methylation due to cell proliferation would be the same. Indeed, a project carried out in parallel with this study, aimed at mapping the derivation of all methylation changes occurring in B-cell derived malignancies, determined that the vast majority (88%) of methylation changes in B-cell malignancies were related to cellular proliferation and independent of disease state (with 9% of changes related to differentiation status and only 3% cancer specific) (Lalchununga, Atasoy manuscript in preparation). Therefore, these results indicating that the similarity in methylation changes seen across all B-cell malignancies will potentially enable us to expand our initial approach and identify methylation changes which are only selected for in the context of specific diseases and would be compatible with a disease specific function. In this analysis we were able to provide initial proof of principle that this disease level approach was able to identify candidate genes based on highly disease specific methylation and expression patterns. Consistent with the high degree of similarity in methylation patterns between the disease this analysis did not detect a high number of genes with disease specific methylation and expression differences, and only a total of 13 candidate DSD/TSG genes across five B-cell malignancies were identified. While the total number of DSDG candidates identified was quite low it was notable that four of the seven candidates (*SOX11* [256], *MFHSA1* [261] (for MCL) and *CTGF* [252], *ZNF423* [251] (for ALL) have been already linked B-cell malignancies, suggesting that the genes identified by this approach are likely to include those with a genuine disease specific functional role. In addition to elevated mRNA and serum expression of *BMP2* [254] and *NPY* [253] respectively, have been found in ALL samples but their functional roles, if any, in ALL is currently unknown. The final identified DSDG candidate, *SNX18*, identified in CLL, has not previously been reported to be associated with CLL or any other B-cell malignancy. This results partially suggest that novel bioinformatic approach designed by our group is effective to identify candidate genes at a disease specific level which could lead to the identification of novel therapeutic targets. However, functional tests done in present project to assess the potential role of several of the ALL-specific DSDG candidates failed

to give definitive results and further assessment of the candidates is required to get a clearer idea as to whether the candidate genes identified by this approach are likely to include genes with crucial disease specific functional roles. This could be achieved by targeting candidate DSDGs by other RNA interference techniques, such as shRNA or gene manipulation technique like CRISPR system, that can lead to more complete and stable gene inactivation. In addition to, as mentioned in chapter 3, three of four ALL specific DSDG candidate's product were proteins whose normal function is largely dependent on secretion of the protein. Further work would be needed to confirm protein secretion in the assessed cell lines models, either through more sensitive ELISA analysis or mass spectrometry approaches., If secretion can be confirmed, *in vivo* model system might be more appropriate to assess the functional role in ALL disease, as this would allow impact on the microenvironment to be taken into account..

In addition to DSDG candidates, this analyse determined that the disease level approach was also able to identify TSG candidates, although again the total number of candidates (six across five B-cell malignancies) was low. Unlike the identified DSDG candidates, most of the identified TSG candidates are novel and have not previously been clearly linked to B-cell malignancies. Interestingly, four of the six TSG candidates were identified in ALL. This contrasts with our initial subtype specific approach which identified 21 SL gene candidates but did not identify any subtype specific TSG candidates. Three of the four candidates were assessed for potential tumour suppressor function in ALL cell line models, leading to the identification of *SLC22A15* as a potential novel functionally relevant gene in ALL. These results also provide evidence that the approach trialled in this project is able to identify novel, functionally relevant, genes.

In this analysis we have focussed on the identification of candidate genes in B-cell malignancies. However, the original subtype specific approach was expanded to multiple other tumour types (medulloblastoma, neuroblastoma and hepatocellular carcinoma), raising the question about whether the disease level analysis reported here could also be expanded to other cancers. A key feature of the B-cell malignancies utilised in this project that was identified in advance was the strong similarity in methylation patterns between the diseases. Thus, prior to attempts to expand this approach to other types of cancer, such as paediatric brain tumours, it would be important to compare already existing

genome wide methylation profiles to identify sets of diseases that exhibit strong overlaps in their patterns of altered DNA methylation. As methylation profiles on relatively large sample sets are publicly available for many cancers types this analysis would be readily achievable.

Our bioinformatic approach potentially could be expanded to identify functionally relevant genes like SL partner for specific genetic lesions shared between multiple cancer. This could be done using essentially any biological or clinical aspect that allows stratification into different subgroups, such as key clinical factors, response to therapy and, in particular, specific genomic lesions. Taking advantage of specific genetic lesions that occur in multiple cancer like p53 mutation, MYC amplification, SL gene partner identified in one cancer type may be similarly targetable in other cancer types with same genetic lesion such as that Poly (ADP-ribose) polymerase (PARP) inhibitors against both breast and ovarian cancer cells which have *BRCA1/2* mutation [358].

Further work section – The functional analyses of the DSDG candidates in ALL presented here failed to clearly determine if the DSDG candidates were functionally relevant or not. The studies had a number of technical difficulties that resulted in a failure to achieve sufficient specificity of inactivating the candidate genes or their protein products. A key part of this was difficulties in achieving substantial knockdown using siRNA transfections. Therefore, trying alternative knockdown/knockout methods like shRNA and CRISPR methods may allow for more efficient but also longer-term knockdown of the candidate DSDG genes. Recently, we have initiated a project to establish a CRISPR/CAS9 based system which includes a fluorescent marker to allow rapid assessment of genes without the need for a selection step (essential for the assessment of genes predicted to be lethal when inactivated). This project is initially focussing on medulloblastoma, but a similar approach could be used to more thoroughly assess the potential of the identified DSDG candidates to be investigated as novel therapeutic targets.

Our results suggest that there might be cell death mechanisms, other than caspase dependent apoptosis, that are induced following re-expression of SLC22A15 in ALL cell line models. Therefore, investigation of other potential cell death mechanisms, would be merited. For example, autophagy could be assessed by analysis of autophagy specific

protein markers by western blot or flow cytometry identification of the specific pathways that are altered due to loss of SLC22A15 expression may allow the identification of potential targets in SLC22A15 deficient ALL cells. In addition, identification of SLC22A15 protein location within the cell membrane by immunofluorescence microscopy, may allow determination of potential interactions with other proteins to gain greater insights into the normal function of SLC22A15. In addition, this may allow clarification of whether mislocation inside the cell, such as localising to the cytoplasm and/or nucleus could occur specifically in ALL disease. Furthermore, identification of SLC22A15 specific ligands by mass spectrometry might suggested possible pathways for targeting SLC22A15 absence.

Overall, current project presented here shows that our original bioinformatic approach could be modified to allow the identification of candidate genes at a whole disease level and that this approach is able to identify previously unappreciated function genes, such as *SLC22A15*, which have functional relevance across all molecular subgroups of a specific cancer type.

7 REFERENCES

1. Rieger, M.A. and T. Schroeder, *Hematopoiesis*. Cold Spring Harb Perspect Biol, 2012. **4**(12).
2. Haas, S., A. Trumpp, and M.D. Milsom, *Causes and Consequences of Hematopoietic Stem Cell Heterogeneity*. Cell Stem Cell, 2018. **22**(5): p. 627-638.
3. Reya, T., et al., *Stem cells, cancer, and cancer stem cells*. nature, 2001. **414**(6859): p. 105-111.
4. Weissman, I.L., *Translating stem and progenitor cell biology to the clinic: barriers and opportunities*. Science, 2000. **287**(5457): p. 1442-1446.
5. Akashi, K., et al., *A clonogenic common myeloid progenitor that gives rise to all myeloid lineages*. Nature, 2000. **404**(6774): p. 193-197.
6. Kondo, M., I.L. Weissman, and K. Akashi, *Identification of clonogenic common lymphoid progenitors in mouse bone marrow*. Cell, 1997. **91**(5): p. 661-672.
7. Yamamoto, R., et al., *Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells*. Cell, 2013. **154**(5): p. 1112-1126.
8. Benveniste, P., et al., *Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential*. Cell stem cell, 2010. **6**(1): p. 48-58.
9. Pietras, E.M., et al., *Functionally distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions*. Cell stem cell, 2015. **17**(1): p. 35-46.
10. Oguro, H., L. Ding, and S.J. Morrison, *SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors*. Cell stem cell, 2013. **13**(1): p. 102-116.
11. Cabezas-Wallscheid, N., et al., *Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis*. Cell stem cell, 2014. **15**(4): p. 507-522.
12. Zhang, Y., et al., *Hematopoietic hierarchy—an updated roadmap*. Trends in cell biology, 2018. **28**(12): p. 976-986.
13. Hoffman, W., F.G. Lakkis, and G. Chalasani, *B cells, antibodies, and more*. Clinical Journal of the American Society of Nephrology, 2016. **11**(1): p. 137-154.
14. Murphy, K. and C. Weaver, *Janeway's Immunobiology*. Garland Science. 2011. The Quarterly Review of Biology, 2012. **87**: p. 266-267.
15. Montecino-Rodriguez, E. and K. Dorshkind, *B-1 B cell development in the fetus and adult*. Immunity, 2012. **36**(1): p. 13-21.
16. Wang, J.-Y., *B cells in immunity and tolerance*. Vol. 1254. 2020: Springer.
17. Bao, Y. and X. Cao, *Epigenetic control of B cell development and B-cell-related immune disorders*. Clinical reviews in allergy & immunology, 2016. **50**: p. 301-311.
18. Eibel, H., et al., *B cell biology: an overview*. Current allergy and asthma reports, 2014. **14**: p. 1-10.
19. Cambier, J.C., et al., *B-cell anergy: from transgenic models to naturally occurring anergic B cells?* Nature Reviews Immunology, 2007. **7**(8): p. 633-643.
20. Victora, G.D. and M.C. Nussenzweig, *Germinal centers*. Annual review of immunology, 2012. **30**: p. 429-457.

21. Guo, J., et al., *RNA interference for multiple myeloma therapy: targeting signal transduction pathways*. Expert Opinion on Therapeutic Targets, 2016. **20**(1): p. 107-121.
22. Arber, D.A., et al., *The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia*. Blood, The Journal of the American Society of Hematology, 2016. **127**(20): p. 2391-2405.
23. Grigoropoulos, N.F., et al., *Leukaemia update. Part 1: diagnosis and management*. Bmj, 2013. **346**.
24. Liesveld, J.L. and M.A. Lichtman, *Acute Myelogenous Leukemia*, in *Williams Hematology*, 10e, K. Kaushansky, et al., Editors. 2021, McGraw-Hill Education: New York, NY.
25. CancerResearchUK. *Acute myeloid leukaemia (AML) incidence statistics*. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml#heading-One>. 2016-2018 [cited 2024].
26. CancerResearchUK. *Acute myeloid leukaemia (AML) mortality statistics*. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-aml#heading-Zero>. 2017-2019 [cited 2024].
27. Minciacchi, V.R., R. Kumar, and D.S. Krause, *Chronic Myeloid Leukemia: A Model Disease of the Past, Present and Future*. Cells, 2021. **10**(1): p. 117.
28. CancerResearchUK. *Chronic myeloid leukaemia (CML) incidence statistics*. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-cml#heading-Zero>. 2016-2018 [cited 2024].
29. CancerResearchUK. *Chronic myeloid leukaemia (CML) mortality statistics*. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-cml#heading-Zero>. 2017-2019 [cited 2024].
30. Iacobucci, I. and C.G. Mullighan, *Genetic basis of acute lymphoblastic leukemia*. Journal of Clinical Oncology, 2017. **35**(9): p. 975.
31. Schwab, C. and C.J. Harrison, *Advances in B-cell precursor acute lymphoblastic leukemia genomics*. HemaSphere, 2018. **2**(4): p. e53.
32. Terwilliger, T. and M. Abdul-Hay, *Acute lymphoblastic leukemia: a comprehensive review and 2017 update*. Blood cancer journal, 2017. **7**(6): p. e577-e577.
33. CancerResearchUK. *Acute lymphoblastic leukaemia (ALL) statistics*. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-all#heading-Zero>. 2016-2018 [cited 2024].
34. Kęsy, J. and D. Januszkiewicz-Lewandowska, *Geny a białaczki dziecięce*. Postępy Higieny i Medycyny Doświadczalnej, 2015. **69**: p. 302-308.
35. Alaggio, R., et al., *The 5th edition of the World Health Organization classification of haematolymphoid tumours: lymphoid neoplasms*. Leukemia, 2022. **36**(7): p. 1720-1748.
36. Inaba, H. and C.G. Mullighan, *Pediatric acute lymphoblastic leukemia*. Haematologica, 2020. **105**(11): p. 2524.
37. Davis, A.S., A.J. Viera, and M.D. Mead, *Leukemia: an overview for primary care*. American family physician, 2014. **89**(9): p. 731-738.
38. Paulsson, K., et al., *The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia*. Nature genetics, 2015. **47**(6): p. 672-676.
39. Molina, O., et al., *Near-haploidy and low-hypodiploidy in B-cell acute lymphoblastic leukemia: when less is too much*. Cancers, 2021. **14**(1): p. 32.
40. Harrison, C.J., et al., *An international study of intrachromosomal amplification of chromosome 21 (iAMP21): cytogenetic characterization and outcome*. Leukemia, 2014. **28**(5): p. 1015-1021.
41. Russell, M., A.K. Ali, and K.M. Mirza, *B-Lymphoblastic Leukemia/Lymphoma: Newly Defined Entities and Subtypes by Molecular Methods*. Advances in Molecular Pathology, 2023. **6**(1): p. 39-49.

42. Hunger, S.P. and C.G. Mullighan, *Acute lymphoblastic leukemia in children*. New England Journal of Medicine, 2015. **373**(16): p. 1541-1552.
43. Roberts, K.G. and C.G. Mullighan, *The biology of B-progenitor acute lymphoblastic leukemia*. Cold Spring Harbor perspectives in medicine, 2020. **10**(7): p. a034835.
44. CancerResearchUK. *CLL incidence*-<https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-cll/incidence#heading-Zero>. 2017-2019 [cited 2024].
45. CancerResearchUK. *CLL-mortality*. <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-cll/mortality#heading-Zero>. 2017-2019 [cited 2024].
46. Hallek, M. and O. Al-Sawaf, *Chronic lymphocytic leukemia: 2022 update on diagnostic and therapeutic procedures*. American journal of hematology, 2021. **96**(12): p. 1679-1705.
47. Heerema, N.A., et al., *Prognostic significance of translocations in the presence of mutated IGHV and of cytogenetic complexity at diagnosis of chronic lymphocytic leukemia*. Haematologica, 2021. **106**(6): p. 1608.
48. Landau, D.A., et al., *Mutations driving CLL and their evolution in progression and relapse*. Nature, 2015. **526**(7574): p. 525-530.
49. Döhner, H., et al., *Genomic aberrations and survival in chronic lymphocytic leukemia*. New England Journal of Medicine, 2000. **343**(26): p. 1910-1916.
50. Zenz, T., et al., *From pathogenesis to treatment of chronic lymphocytic leukaemia*. Nature Reviews Cancer, 2010. **10**(1): p. 37-50.
51. Fais, F., et al., *Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors*. The Journal of clinical investigation, 1998. **102**(8): p. 1515-1525.
52. Stevenson, F.K., et al., *B-cell receptor signaling in chronic lymphocytic leukemia*. Blood, The Journal of the American Society of Hematology, 2011. **118**(16): p. 4313-4320.
53. Swerdlow, S.H., et al., *The 2016 revision of the World Health Organization classification of lymphoid neoplasms*. Blood, 2016. **127**(20): p. 2375-90.
54. Lewis, W.D., S. Lilly, and K.L. Jones, *Lymphoma: diagnosis and treatment*. American family physician, 2020. **101**(1): p. 34-41.
55. Smith, A., et al., *Lymphoma incidence, survival and prevalence 2004–2014: sub-type analyses from the UK's Haematological Malignancy Research Network*. British journal of cancer, 2015. **112**(9): p. 1575-1584.
56. Küppers, R., *Mechanisms of B-cell lymphoma pathogenesis*. Nature Reviews Cancer, 2005. **5**(4): p. 251-262.
57. Bachy, E., et al., *Sustained progression-free survival benefit of rituximab maintenance in patients with follicular lymphoma: long-term results of the PRIMA study*. Journal of Clinical Oncology, 2019. **37**(31): p. 2815.
58. Campo, E., et al., *The international consensus classification of mature lymphoid neoplasms: a report from the clinical advisory committee*. Blood, The Journal of the American Society of Hematology, 2022. **140**(11): p. 1229-1253.
59. Cerhan, J.R. and S.L. Slager, *Familial predisposition and genetic risk factors for lymphoma*. Blood, The Journal of the American Society of Hematology, 2015. **126**(20): p. 2265-2273.
60. Maddocks, K., *Update on mantle cell lymphoma*. Blood, The Journal of the American Society of Hematology, 2018. **132**(16): p. 1647-1656.
61. Jain, A.G., et al., *Leukemic non-nodal mantle cell lymphoma: diagnosis and treatment*. Current Treatment Options in Oncology, 2019. **20**(12): p. 85.
62. Sánchez-Beato, M., et al., *A genetic profiling guideline to support diagnosis and clinical management of lymphomas*. Clinical and Translational Oncology, 2023: p. 1-20.

63. Basso, K. and R. Dalla-Favera, *Germinal centres and B cell lymphomagenesis*. Nature Reviews Immunology, 2015. **15**(3): p. 172-184.
64. Martelli, M., et al., *Diffuse large B-cell lymphoma*. Critical reviews in oncology/hematology, 2013. **87**(2): p. 146-171.
65. Alizadeh, A.A., et al., *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling*. Nature, 2000. **403**(6769): p. 503-511.
66. Rosenwald, A., et al., *The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma*. New England Journal of Medicine, 2002. **346**(25): p. 1937-1947.
67. Sehn, L.H. and G. Salles, *Diffuse large B-cell lymphoma*. New England Journal of Medicine, 2021. **384**(9): p. 842-858.
68. Ott, G., et al., *Immunoblastic morphology but not the immunohistochemical GCB/nonGCB classifier predicts outcome in diffuse large B-cell lymphoma in the RICOVER-60 trial of the DSHNHL*. Blood, The Journal of the American Society of Hematology, 2010. **116**(23): p. 4916-4925.
69. Ye, B.H., et al., *Alterations of a zinc finger-encoding gene, BCL-6, in diffuse large-cell lymphoma*. Science, 1993. **262**(5134): p. 747-750.
70. Aukema, S.M., et al., *Double-hit B-cell lymphomas*. Blood, The Journal of the American Society of Hematology, 2011. **117**(8): p. 2319-2331.
71. Fernández-Serrano, M., et al., *Histone Modifications and Their Targeting in Lymphoid Malignancies*. International Journal of Molecular Sciences, 2022. **23**(1): p. 253.
72. Han, C.H. and T.T. Batchelor, *Primary central nervous system lymphoma*. CONTINUUM: Lifelong Learning in Neurology, 2017. **23**(6): p. 1601-1618.
73. Giannini, C., A. Dogan, and D.R. Salomão, *CNS lymphoma: a practical diagnostic approach*. Journal of Neuropathology & Experimental Neurology, 2014. **73**(6): p. 478-494.
74. Schabet, M., *Epidemiology of primary CNS lymphoma*. Journal of neuro-oncology, 1999. **43**: p. 199-201.
75. Bhagavathi, S. and J.D. Wilson, *Primary central nervous system lymphoma*. Archives of pathology & laboratory medicine, 2008. **132**(11): p. 1830-1834.
76. Rubenstein, J.L., et al., *Gene expression and angiotropism in primary CNS lymphoma*. Blood, 2006. **107**(9): p. 3716-3723.
77. Montesinos-Rongen, M., et al., *Gene expression profiling suggests primary central nervous system lymphomas to be derived from a late germinal center B cell*. Leukemia, 2008. **22**(2): p. 400-405.
78. Tun, H.W., et al., *Pathway analysis of primary central nervous system lymphoma*. Blood, The Journal of the American Society of Hematology, 2008. **111**(6): p. 3200-3210.
79. Gerstner, E.R. and T.T. Batchelor, *Primary central nervous system lymphoma*. Primary Central Nervous System Tumors: Pathogenesis and Therapy, 2011: p. 333-353.
80. Waddington, C.H., *The epigenotype*. 1942. Int J Epidemiol, 2012. **41**(1): p. 10-3.
81. Waddington, C.H., *The strategy of the genes*. 2014: Routledge.
82. Nanney, D.L., *EPIGENETIC CONTROL SYSTEMS*. Proc Natl Acad Sci U S A, 1958. **44**(7): p. 712-7.
83. Wu, C. and J.R. Morris, *Genes, genetics, and epigenetics: a correspondence*. Science, 2001. **293**(5532): p. 1103-5.
84. Portela, A. and M. Esteller, *Epigenetic modifications and human disease*. Nature biotechnology, 2010. **28**(10): p. 1057-1068.
85. Hotchkiss, R.D., *The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography*. Journal of Biological Chemistry, 1948. **175**(1): p. 315-332.
86. Compere, S.J. and R.D. Palmiter, *DNA methylation controls the inducibility of the mouse metallothionein-I gene lymphoid cells*. Cell, 1981. **25**(1): p. 233-40.

87. Holliday, R. and J.E. Pugh, *DNA modification mechanisms and gene activity during development*. Science, 1975. **187**(4173): p. 226-32.
88. Ospelt, C., *A brief history of epigenetics*. Immunology Letters, 2022. **249**: p. 1-4.
89. Zhang, G. and S. Pradhan, *Mammalian epigenetic mechanisms*. IUBMB life, 2014. **66**(4): p. 240-256.
90. Gómez-Díaz, E., et al., *Epigenetics of host–pathogen interactions: the road ahead and the road behind*. PLoS pathogens, 2012. **8**(11): p. e1003007.
91. Zakhari, S., *Alcohol metabolism and epigenetics changes*. Alcohol research: current reviews, 2013. **35**(1): p. 6.
92. Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function*. Neuropsychopharmacology, 2013. **38**(1): p. 23-38.
93. Esteller, M., *Epigenetic gene silencing in cancer: the DNA hypermethylome*. Hum Mol Genet, 2007. **16 Spec No 1**: p. R50-9.
94. Leonhardt, H., et al., *A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei*. Cell, 1992. **71**(5): p. 865-73.
95. Hermann, A., R. Goyal, and A. Jeltsch, *The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites*. J Biol Chem, 2004. **279**(46): p. 48350-9.
96. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development*. Cell, 1999. **99**(3): p. 247-257.
97. Nan, X., R.R. Meehan, and A. Bird, *Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2*. Nucleic acids research, 1993. **21**(21): p. 4886-4892.
98. Singal, R. and G.D. Ginder, *DNA methylation*. Blood, 1999. **93**(12): p. 4059-70.
99. Sharif, J., et al., *The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA*. Nature, 2007. **450**(7171): p. 908-912.
100. Bird, A.P., *DNA methylation and the frequency of CpG in animal DNA*. Nucleic acids research, 1980. **8**(7): p. 1499-1504.
101. Ehrlich, M., et al., *Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells*. Nucleic Acids Res, 1982. **10**(8): p. 2709-21.
102. Antequera, F. and A. Bird, *Number of CpG islands and genes in human and mouse*. Proc Natl Acad Sci U S A, 1993. **90**(24): p. 11995-9.
103. Tazi, J. and A. Bird, *Alternative chromatin structure at CpG islands*. Cell, 1990. **60**(6): p. 909-20.
104. Bird, A., et al., *A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA*. Cell, 1985. **40**(1): p. 91-99.
105. Saxonov, S., P. Berg, and D.L. Brutlag, *A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters*. Proceedings of the National Academy of Sciences, 2006. **103**(5): p. 1412-1417.
106. Illingworth, R.S., et al., *Orphan CpG islands identify numerous conserved promoters in the mammalian genome*. PLoS Genet, 2010. **6**(9): p. e1001134.
107. Cross, S.H. and A.P. Bird, *CpG islands and genes*. Curr Opin Genet Dev, 1995. **5**(3): p. 309-14.
108. Irizarry, R.A., et al., *The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores*. Nat Genet, 2009. **41**(2): p. 178-186.
109. Cavalcante, R.G. *annotatr: Making sense of genomic regions*. 2017-02-06 [cited 2024 December]; Available from: <https://bioconductor.riken.jp/packages/3.4/bioc/vignettes/annotatr/inst/doc/annotatr-vignette.html>.
110. Onufriev, A.V. and H. Schiessel, *The nucleosome: from structure to function through physics*. Current opinion in structural biology, 2019. **56**: p. 119-130.

111. Fyodorov, D.V., et al., *Emerging roles of linker histones in regulating chromatin structure and function*. Nature reviews Molecular cell biology, 2018. **19**(3): p. 192-206.
112. Neganova, M.E., et al. *Histone modifications in epigenetic regulation of cancer: Perspectives and achieved progress*. in *Seminars in Cancer Biology*. 2022. Elsevier.
113. Liu, C.L., et al., *Single-nucleosome mapping of histone modifications in S. cerevisiae*. PLoS biology, 2005. **3**(10): p. e328.
114. Lawrence, M., S. Daujat, and R. Schneider, *Lateral Thinking: How Histone Modifications Regulate Gene Expression*. Trends Genet, 2016. **32**(1): p. 42-56.
115. Xhemalce, B., M.A. Dawson, and A.J. Bannister, *Histone modifications*. Reviews in Cell Biology and Molecular Medicine, 2006.
116. Di Martile, M., D. Del Bufalo, and D. Trisciuglio, *The multifaceted role of lysine acetylation in cancer: prognostic biomarker and therapeutic target*. Oncotarget, 2016. **7**(34): p. 55789.
117. Narita, T., B.T. Weinert, and C. Choudhary, *Functions and mechanisms of non-histone protein acetylation*. Nature reviews Molecular cell biology, 2019. **20**(3): p. 156-174.
118. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications*. Cell research, 2011. **21**(3): p. 381-395.
119. Bedford, M.T., *Arginine methylation at a glance*. Journal of cell science, 2007. **120**(24): p. 4243-4246.
120. Kooistra, S.M. and K. Helin, *Molecular mechanisms and potential functions of histone demethylases*. Nature reviews Molecular cell biology, 2012. **13**(5): p. 297-311.
121. Greer, E.L. and Y. Shi, *Histone methylation: a dynamic mark in health, disease and inheritance*. Nature Reviews Genetics, 2012. **13**(5): p. 343-357.
122. Tian, X., et al., *Histone lysine-specific methyltransferases and demethylases in carcinogenesis: new targets for cancer therapy and prevention*. Current cancer drug targets, 2013. **13**(5): p. 558-579.
123. Bannister, A.J., R. Schneider, and T. Kouzarides, *Histone methylation: dynamic or static?* Cell, 2002. **109**(7): p. 801-806.
124. Kim, D.-H., et al., *Histone H3K27 trimethylation inhibits H3 binding and function of SET1-like H3K4 methyltransferase complexes*. Molecular and cellular biology, 2013. **33**(24): p. 4936-4946.
125. Fulton, M.D., T. Brown, and Y.G. Zheng, *Mechanisms and Inhibitors of Histone Arginine Methylation*. Chem Rec, 2018. **18**(12): p. 1792-1807.
126. Oki, M., H. Aihara, and T. Ito, *Role of histone phosphorylation in chromatin dynamics and its implications in diseases*. Chromatin and Disease, 2007: p. 323-340.
127. Hu, S., et al., *Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling*. Cell, 2009. **139**(3): p. 610-622.
128. Taylor, G.C., et al., *H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction*. Genome research, 2013. **23**(12): p. 2053-2065.
129. Hassa, P.O., et al., *Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going?* Microbiology and Molecular Biology Reviews, 2006. **70**(3): p. 789-829.
130. Hershko, A. and A. Ciechanover, *The ubiquitin system*. Annu Rev Biochem, 1998. **67**: p. 425-79.
131. Lee, J.S., et al., *Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS*. Cell, 2007. **131**(6): p. 1084-96.
132. Nathan, D., et al., *Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications*. Genes & development, 2006. **20**(8): p. 966-976.
133. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.

134. Djebali, S., et al., *Landscape of transcription in human cells*. Nature, 2012. **489**(7414): p. 101-8.
135. Bure, I.V. and M.V. Nemtsova, *Methylation and Noncoding RNAs in Gastric Cancer: Everything Is Connected*. Int J Mol Sci, 2021. **22**(11).
136. Taft, R.J., et al., *Non-coding RNAs: regulators of disease*. J Pathol, 2010. **220**(2): p. 126-39.
137. Yao, Q., Y. Chen, and X. Zhou, *The roles of microRNAs in epigenetic regulation*. Current opinion in chemical biology, 2019. **51**: p. 11-17.
138. Ferreira, H.J. and M. Esteller, *Non-coding RNAs, epigenetics, and cancer: tying it all together*. Cancer and Metastasis Reviews, 2018. **37**: p. 55-73.
139. Kaikkonen, M.U., M.T. Lam, and C.K. Glass, *Non-coding RNAs as regulators of gene expression and epigenetics*. Cardiovascular research, 2011. **90**(3): p. 430-440.
140. Lam, J.K., et al., *siRNA versus miRNA as therapeutics for gene silencing*. Molecular Therapy-Nucleic Acids, 2015. **4**.
141. Lin, H., *piRNAs in the germ line*. science, 2007. **316**(5823): p. 397-397.
142. Fujita, N., et al., *Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression*. Journal of Biological Chemistry, 2003. **278**(26): p. 24132-24138.
143. Lehnertz, B., et al., *Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin*. Current Biology, 2003. **13**(14): p. 1192-1200.
144. Rothbart, S.B., et al., *Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation*. Nature structural & molecular biology, 2012. **19**(11): p. 1155-1160.
145. Ng, E.K., et al., *MicroRNA-143 is downregulated in breast cancer and regulates DNA methyltransferases 3A in breast cancer cells*. Tumor Biology, 2014. **35**: p. 2591-2598.
146. Zhang, P., et al., *Ten-eleven translocation (Tet) and thymine DNA glycosylase (TDG), components of the demethylation pathway, are direct targets of miRNA-29a*. Biochemical and biophysical research communications, 2013. **437**(3): p. 368-373.
147. Sharma, S., et al., *Stem cell c-KIT and HOXB4 genes: critical roles and mechanisms in self-renewal, proliferation, and differentiation*. Stem cells and development, 2006. **15**(6): p. 755-778.
148. Zhou, Y., et al., *Epigenetic modifications of stem cells: a paradigm for the control of cardiac progenitor cells*. Circulation research, 2011. **109**(9): p. 1067-1081.
149. Genovese, G., et al., *Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence*. New England Journal of Medicine, 2014. **371**(26): p. 2477-2487.
150. Trowbridge, J.J., et al., *DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells*. Cell stem cell, 2009. **5**(4): p. 442-449.
151. Challen, G.A., et al., *Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells*. Cell stem cell, 2014. **15**(3): p. 350-364.
152. Ehrlich, M., *DNA methylation in cancer: too much, but also too little*. Oncogene, 2002. **21**(35): p. 5400-5413.
153. San Jose-Eneriz, E., et al., *Epigenetic regulation of cell signaling pathways in acute lymphoblastic leukemia*. Epigenomics, 2013. **5**(5): p. 525-538.
154. Baylin, S.B. and P.A. Jones, *A decade of exploring the cancer epigenome—biological and translational implications*. Nature Reviews Cancer, 2011. **11**(10): p. 726-734.
155. Timms, J.A., et al., *DNA methylation as a potential mediator of environmental risks in the development of childhood acute lymphoblastic leukemia*. Epigenomics, 2016. **8**(4): p. 519-536.
156. Dawson, M.A. and T. Kouzarides, *Cancer epigenetics: from mechanism to therapy*. cell, 2012. **150**(1): p. 12-27.

157. Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer*. Nature reviews genetics, 2002. **3**(6): p. 415-428.
158. Esteller, M., et al., *Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors*. JNCI: Journal of the National Cancer Institute, 2000. **92**(7): p. 564-569.
159. Kulis, M., et al., *Whole-genome fingerprint of the DNA methylome during human B cell differentiation*. Nature genetics, 2015. **47**(7): p. 746-756.
160. Nordlund, J., et al., *DNA methylation-based subtype prediction for pediatric acute lymphoblastic leukemia*. Clinical epigenetics, 2015. **7**: p. 1-12.
161. Wahlfors, J., et al., *Genomic hypomethylation in human chronic lymphocytic leukemia*. 1992.
162. Kulis, M., et al., *Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia*. Nature genetics, 2012. **44**(11): p. 1236-1242.
163. Knudson, A.G., *Two genetic hits (more or less) to cancer*. Nature Reviews Cancer, 2001. **1**(2): p. 157-162.
164. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA: a cancer journal for clinicians, 2021. **71**(3): p. 209-249.
165. Hansemann, D., *Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung*. Archiv für pathologische Anatomie und Physiologie und für klinische Medizin, 1890. **119**: p. 299-326.
166. Boveri, T., *Zur frage der entstehung maligner tumoren*. 1914: Fischer.
167. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. cell, 2011. **144**(5): p. 646-674.
168. Qing, T., et al., *Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden*. Nature communications, 2020. **11**(1): p. 2438.
169. White, M.C., et al., *Age and cancer risk: a potentially modifiable relationship*. American journal of preventive medicine, 2014. **46**(3): p. S7-S15.
170. Murff, H.J., D.R. Spiegel, and S. Syngal, *Does this patient have a family history of cancer?: an evidence-based analysis of the accuracy of family cancer history*. Jama, 2004. **292**(12): p. 1480-1489.
171. Gong, Y., J. Deng, and X. Wu, *Germline mutations and blood malignancy*. Oncology Reports, 2021. **45**(1): p. 49-57.
172. Stratton, M.R., P.J. Campbell, and P.A. Futreal, *The cancer genome*. Nature, 2009. **458**(7239): p. 719-724.
173. Olivier, M., et al., *TP53 mutation spectra and load: a tool for generating hypotheses on the etiology of cancer*. IARC scientific publications, 2004(157): p. 247-270.
174. Talbot, S.J. and D.H. Crawford, *Viruses and tumours—an update*. European Journal of Cancer, 2004. **40**(13): p. 1998-2005.
175. McDuff, F.K. and S.D. Turner, *Jailbreak: oncogene-induced senescence and its evasion*. Cellular signalling, 2011. **23**(1): p. 6-13.
176. Sun, W. and J. Yang, *Functional mechanisms for human tumor suppressors*. Journal of Cancer, 2010. **1**: p. 136.
177. Wang, L.-H., et al., *Loss of tumor suppressor gene function in human cancer: an overview*. Cellular Physiology and Biochemistry, 2019. **51**(6): p. 2647-2693.
178. Leiderman, Y.I., S. Kiss, and S. Mukai. *Molecular genetics of RB1—The retinoblastoma gene*. in *Seminars in ophthalmology*. 2007. Taylor & Francis.
179. Smith, A.L., T.P. Robin, and H.L. Ford, *Molecular pathways: targeting the TGF- β pathway for cancer therapy*. Clinical Cancer Research, 2012. **18**(17): p. 4514-4521.

180. Savage, K.I. and D.P. Harkin, *BRCA 1, a 'complex' protein involved in the maintenance of genomic stability*. The FEBS journal, 2015. **282**(4): p. 630-646.
181. Nayak, S.K., P.S. Panesar, and H. Kumar, *p53-Induced apoptosis and inhibitors of p53*. Current medicinal chemistry, 2009. **16**(21): p. 2627-2640.
182. Rahman, N. and R.H. Scott, *Cancer genes associated with phenotypes in monoallelic and biallelic mutation carriers: new lessons from old players*. Human molecular genetics, 2007. **16**(R1): p. R60-R66.
183. Macleod, K., *Tumor suppressor genes*. Current opinion in genetics & development, 2000. **10**(1): p. 81-93.
184. Kinzler, K.W. and B. Vogelstein, *Landscaping the cancer terrain*. Science, 1998. **280**(5366): p. 1036-1037.
185. Kinzler, K.W. and B. Vogelstein, *Lessons from hereditary colorectal cancer*. Cell, 1996. **87**(2): p. 159-170.
186. Key, O. <https://oncohemakey.com/oncogenes-and-tumor-suppressor-genes/>. [cited March 2024 March 2024].
187. Knudson Jr, A.G., *Mutation and cancer: statistical study of retinoblastoma*. Proceedings of the National Academy of Sciences, 1971. **68**(4): p. 820-823.
188. Grossniklaus, H.E., *Retinoblastoma. Fifty years of progress. The LXXI Edward Jackson memorial lecture*. American journal of ophthalmology, 2014. **158**(5): p. 875-891. e1.
189. Cassoux, N., et al., *Retinoblastoma: update on current management*. The Asia-Pacific Journal of Ophthalmology, 2017. **6**(3): p. 290-295.
190. Ramasubramanian, A. and C.L. Shields, *Retinoblastoma*. 2012: JP Medical Ltd.
191. Myöhänen, S.K., S.B. Baylin, and J.G. Herman, *Hypermethylation can selectively silence individual p16ink4A alleles in neoplasia*. Cancer research, 1998. **58**(4): p. 591-593.
192. Grady, W.M., et al., *Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer*. Nature genetics, 2000. **26**(1): p. 16-17.
193. Esteller, M., et al., *DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis*. Human molecular genetics, 2001. **10**(26): p. 3001-3007.
194. Course™, B.B.a.C.S. <https://www.ao.org/education/bcscsnippetdetail.aspx?id=1d0101bc-a400-4232-9649-gef25ed924cf>. [cited March 2024].
195. Derelanko, M.J. and M.A. Hollinger, *Handbook of toxicology*. 2001: CRC press.
196. Gutiérrez, J., et al., *V. Example 2: Estrogens, Retinoids and Cervical Cancer Development*. Molecular Oncology Principles and Recent Advances, 2012: p. 48.
197. Konopka, J.B., et al., *Cell lines and clinical isolates derived from Ph1-positive chronic myelogenous leukemia patients express c-abl proteins with a common structural alteration*. Proceedings of the National Academy of Sciences, 1985. **82**(6): p. 1810-1814.
198. 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. 1390-1393.
199. Bishop, J.M., *Molecular themes in oncogenesis*. Cell, 1991. **64**(2): p. 235-248.
200. Finger, L.R., et al., *A common mechanism of chromosomal translocation in T-and B-cell neoplasia*. Science, 1986. **234**(4779): p. 982-985.
201. Croce, C.M., *Oncogenes and cancer*. New England journal of medicine, 2008. **358**(5): p. 502-511.
202. Weinstein, I.B., *Addiction to oncogenes--the Achilles heel of cancer*. Science, 2002. **297**(5578): p. 63-64.
203. Slamon, D.J., et al., *Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2*. New England journal of medicine, 2001. **344**(11): p. 783-792.
204. Iqbal, N. and N. Iqbal, *Imatinib: a breakthrough of targeted therapy in cancer*. Chemotherapy research and practice, 2014. **2014**(1): p. 357027.

205. Marcucci, G., D. Perrotti, and M.A. Caligiuri, *Understanding the Molecular Basis of Imatinib Mesylate Therapy in Chronic Myelogenous Leukemia and the Related Mechanisms of Resistance: Commentary re: AN Mohamed et al., The Effect of Imatinib Mesylate on Patients with Philadelphia Chromosome-positive Chronic Myeloid Leukemia with Secondary Chromosomal Aberrations. Clin. Cancer Res.*, 9: 1333–1337, 2003. *Clinical cancer research*, 2003. **9**(4): p. 1248-1252.
206. Hochhaus, A. and P. La Rosée, *Imatinib therapy in chronic myelogenous leukemia: strategies to avoid and overcome resistance. Leukemia*, 2004. **18**(8): p. 1321-1331.
207. Nagel, R., E.A. Semenova, and A. Berns, *Drugging the addict: non-oncogene addiction as a target for cancer therapy. EMBO reports*, 2016. **17**(11): p. 1516-1531.
208. Nijman, S.M., *Synthetic lethality: general principles, utility and detection using genetic screens in human cells. FEBS letters*, 2011. **585**(1): p. 1-6.
209. Solimini, N.L., J. Luo, and S.J. Elledge, *Non-oncogene addiction and the stress phenotype of cancer cells. Cell*, 2007. **130**(6): p. 986-988.
210. Luo, J., N.L. Solimini, and S.J. Elledge, *Principles of cancer therapy: oncogene and non-oncogene addiction. Cell*, 2009. **136**(5): p. 823-837.
211. Bridges, C., *The origin of variation. Amer Nat*, 1922. **56**: p. 51-53.
212. Dobzhansky, T., *Genetics of natural populations. XIII. Recombination and variability in populations of Drosophila pseudoobscura. Genetics*, 1946. **31**(3): p. 269.
213. He, B., *Specific killing of BRCA2-deficient tumours with inhibitors of poly (ADP-ribose) polymerase. Nature*, 2005. **434**: p. 913-917.
214. Hartwell, L.H., et al., *Integrating genetic approaches into the discovery of anticancer drugs. Science*, 1997. **278**(5340): p. 1064-1068.
215. Konstantinopoulos, P.A., et al., *Homologous recombination deficiency: exploiting the fundamental vulnerability of ovarian cancer. Cancer discovery*, 2015. **5**(11): p. 1137-1154.
216. Farmer, H., et al., *Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature*, 2005. **434**(7035): p. 917-921.
217. Haince, J.-F., et al., *Targeting poly (ADP-ribosyl) ation: a promising approach in cancer therapy. Trends in molecular medicine*, 2005. **11**(10): p. 456-463.
218. Galluzzi, L., I. Vitale, and G. Kroemer, *Past, present, and future of molecular and cellular oncology. Frontiers in Oncology*, 2011. **1**: p. 1.
219. Moreau, P., et al., *Proteasome inhibitors in multiple myeloma: 10 years later. Blood, The Journal of the American Society of Hematology*, 2012. **120**(5): p. 947-959.
220. Petroni, G., et al., *Targeting oncogene and non-oncogene addiction to inflame the tumour microenvironment. Nature reviews Drug discovery*, 2022. **21**(6): p. 440-462.
221. Schwalbe, E.C., et al., *Integration of genome-level data to allow identification of subtype-specific vulnerability genes as novel therapeutic targets. Oncogene*, 2021. **40**(33): p. 5213-5223.
222. Nordlund, J., et al., *Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. Genome biology*, 2013. **14**: p. 1-15.
223. Bukrinsky, M.I., et al., *A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. Nature*, 1993. **365**(6447): p. 666-669.
224. Zufferey, R., et al., *Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. Journal of virology*, 1998. **72**(12): p. 9873-9880.
225. Sakuma, T., M.A. Barry, and Y. Ikeda, *Lentiviral vectors: basic to translational. Biochemical Journal*, 2012. **443**(3): p. 603-618.
226. *How does propidium iodide staining work?* [cited 2024; Available from: <https://www.aatbio.com/resources/application-notes/propidium-iodide-applications-common-issues>].
227. Sahoo, G., et al., *A review on caspases: key regulators of biological activities and apoptosis. Molecular neurobiology*, 2023. **60**(10): p. 5805-5837.

228. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4*. *nature*, 1970. **227**(5259): p. 680-685.
229. Roehm, N.W., et al., *An improved colorimetric assay for cell proliferation and viability utilizing the tetrazolium salt XTT*. *Journal of immunological methods*, 1991. **142**(2): p. 257-265.
230. Bibikova, M., et al., *High density DNA methylation array with single CpG site resolution*. *Genomics*, 2011. **98**(4): p. 288-295.
231. Lee, S.-T., et al., *A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network*. *Nucleic acids research*, 2012. **40**(22): p. 11339-11351.
232. Gabriel, A.S., et al., *Epigenetic landscape correlates with genetic subtype but does not predict outcome in childhood acute lymphoblastic leukemia*. *Epigenetics*, 2015. **10**(8): p. 717-726.
233. *BLUEPRINT DNA Methylation 450K data of mantle cell lymphoma - EGA European Genome-Phenome Archive*.
234. Asmar, F., et al., *Genome-wide profiling identifies a DNA methylation signature that associates with TET2 mutations in diffuse large B-cell lymphoma*. *Haematologica*, 2013. **98**(12): p. 1912.
235. Schmitz, R., et al., *Genetics and pathogenesis of diffuse large B-cell lymphoma*. *New England journal of medicine*, 2018. **378**(15): p. 1396-1407.
236. Nakamura, T., et al., *Genome-wide DNA methylation profiling identifies primary central nervous system lymphoma as a distinct entity different from systemic diffuse large B-cell lymphoma*. *Acta neuropathologica*, 2017. **133**: p. 321-324.
237. Aryee, M.J., et al., *Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays*. *Bioinformatics*, 2014. **30**(10): p. 1363-1369.
238. Triche Jr, T.J., et al., *Low-level processing of Illumina Infinium DNA methylation beadarrays*. *Nucleic acids research*, 2013. **41**(7): p. e90-e90.
239. Chen, Y.-a., et al., *Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray*. *Epigenetics*, 2013. **8**(2): p. 203-209.
240. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. *Nucleic acids research*, 2015. **43**(7): p. e47-e47.
241. Haferlach, T., et al., *Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group*. *Journal of clinical oncology*, 2010. **28**(15): p. 2529-2537.
242. Ferreira, P.G., et al., *Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia*. *Genome research*, 2014. **24**(2): p. 212-226.
243. Scott, D.W., et al., *New molecular assay for the proliferation signature in mantle cell lymphoma applicable to formalin-fixed paraffin-embedded biopsies*. *Journal of Clinical Oncology*, 2017. **35**(15): p. 1668-1677.
244. Lenz, G., et al., *Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways*. *Proceedings of the National Academy of Sciences*, 2008. **105**(36): p. 13520-13525.
245. Kawaguchi, A., et al., *Gene Expression Signature–Based Prognostic Risk Score in Patients with Primary Central Nervous System Lymphoma*. *Clinical Cancer Research*, 2012. **18**(20): p. 5672-5681.
246. Novershtern, N., et al., *Densely interconnected transcriptional circuits control cell states in human hematopoiesis*. *Cell*, 2011. **144**(2): p. 296-309.
247. Peters, T., et al., *The DMRcate package user's guide*. 2017.

248. Davidsson, J., et al., *The DNA methylome of pediatric acute lymphoblastic leukemia*. Human molecular genetics, 2009. **18**(21): p. 4054-4065.
249. Wattanawaraporn, R., et al., *Hypermethylation of TTC12 gene in acute lymphoblastic leukemia*. Leukemia, 2007. **21**(11): p. 2370-2373.
250. Chaiprasert, T., et al., *Roles of zinc finger protein 423 in proliferation and invasion of cholangiocarcinoma through oxidative stress*. Biomolecules, 2019. **9**(7): p. 263.
251. Harder, L., et al., *Aberrant ZNF423 impedes B cell differentiation and is linked to adverse outcome of ETV6-RUNX1 negative B precursor acute lymphoblastic leukemia*. Journal of Experimental Medicine, 2013. **210**(11): p. 2289-2304.
252. Lu, H., et al., *Targeting connective tissue growth factor (CTGF) in acute lymphoblastic leukemia preclinical models: anti-CTGF monoclonal antibody attenuates leukemia growth*. Annals of hematology, 2014. **93**: p. 485-492.
253. Kogner, P., et al., *Neuropeptide Y (NPY) synthesis in lymphoblasts and increased plasma NPY in pediatric B-cell precursor leukemia*. 1992.
254. Tesfai, Y., et al., *Interactions between acute lymphoblastic leukemia and bone marrow stromal cells influence response to therapy*. Leukemia research, 2012. **36**(3): p. 299-306.
255. *Leukemia MILE study- SOX11 expression*.
256. Zeng, W., et al., *Cyclin D1-negative blastoid mantle cell lymphoma identified by SOX11 expression*. The American journal of surgical pathology, 2012. **36**(2): p. 214-219.
257. Weigle, B., et al., *Highly specific overexpression of the transcription factor SOX11 in human malignant gliomas*. Oncology reports, 2005. **13**(1): p. 139-144.
258. Liu, D.-T., et al., *Clinical and prognostic significance of SOX11 in breast cancer*. Asian Pacific Journal of Cancer Prevention, 2014. **15**(13): p. 5483-5486.
259. Jian, J., W. Guoying, and Z. Jing, *Increased expression of sex determining region Y-box 11 (SOX11) in cutaneous malignant melanoma*. Journal of international medical research, 2013. **41**(4): p. 1221-1227.
260. Sakabe, T., et al., *Identification of a novel gene, MASL1 within an amplicon at 8p23. 1 detected in malignant fibrous histiocytomas by comparative genomic hybridization*. Cancer research, 1999. **59**(3): p. 511-515.
261. Tagawa, H., et al., *MASL1, a candidate oncogene found in amplification at 8p23. 1, is translocated in immunoblastic B-cell lymphoma cell line OCI-LY8*. Oncogene, 2004. **23**(14): p. 2576-2581.
262. Dubois, S., et al., *Next-generation sequencing in diffuse large B-cell lymphoma highlights molecular divergence and therapeutic opportunities: a LYSA study*. Clinical cancer research, 2016. **22**(12): p. 2919-2928.
263. Hanley, S.E. and K.F. Cooper, *Sorting nexins in protein homeostasis*. Cells, 2020. **10**(1): p. 17.
264. Juin, A., et al., *N-WASP control of LPAR1 trafficking establishes response to self-generated LPA gradients to promote pancreatic cancer cell metastasis*. Developmental cell, 2019. **51**(4): p. 431-445. e7.
265. Liu, H., et al., *Fascin actin-bundling protein 1 in human cancer: promising biomarker or therapeutic target?* Molecular Therapy-Oncolytics, 2021. **20**: p. 240-264.
266. Pinkus, G.S., et al., *Fascin, a sensitive new marker for Reed-Sternberg cells of hodgkin's disease. Evidence for a dendritic or B cell derivation?* The American journal of pathology, 1997. **150**(2): p. 543.
267. El Kramani, N., et al., *Clinical significance of the TNF- α receptors, TNFRSF2 and TNFRSF9, on cell migration molecules Fascin-1 and Versican in acute leukemia*. Cytokine, 2018. **111**: p. 523-529.
268. Tang, H., et al., *Architecture of cell-cell adhesion mediated by sidekicks*. Proceedings of the National Academy of Sciences, 2018. **115**(37): p. 9246-9251.

269. Grönroos, T., et al., *Clinicopathological features and prognostic value of SOX11 in childhood acute lymphoblastic leukemia*. Scientific reports, 2020. **10**(1): p. 2043.
270. Sánchez, M.L., F.D. Rodríguez, and R. Coveñas, *Neuropeptide Y peptide family and cancer: Antitumor therapeutic strategies*. International Journal of Molecular Sciences, 2023. **24**(12): p. 9962.
271. Bach, D.-H., H.J. Park, and S.K. Lee, *The dual role of bone morphogenetic proteins in cancer*. Molecular Therapy-Oncolytics, 2018. **8**: p. 1-13.
272. Langenfeld, E.M., et al., *The mature bone morphogenetic protein-2 is aberrantly expressed in non-small cell lung carcinomas and stimulates tumor growth of A549 cells*. Carcinogenesis, 2003. **24**(9): p. 1445-1454.
273. Li, C.-S., et al., *Secreted phosphoprotein 24 kD (Spp24) inhibits growth of human pancreatic cancer cells caused by BMP-2*. Biochemical and biophysical research communications, 2015. **466**(2): p. 167-172.
274. Park, Y., et al., *The bone morphogenesis protein-2 (BMP-2) is associated with progression to metastatic disease in gastric cancer*. Cancer research and treatment: official journal of Korean Cancer Association, 2008. **40**(3): p. 127-132.
275. Bennewith, K.L., et al., *The role of tumor cell–derived connective tissue growth factor (CTGF/CCN2) in pancreatic tumor growth*. Cancer research, 2009. **69**(3): p. 775-784.
276. Aikawa, T., et al., *Connective tissue growth factor–specific antibody attenuates tumor growth, metastasis, and angiogenesis in an orthotopic mouse model of pancreatic cancer*. Molecular cancer therapeutics, 2006. **5**(5): p. 1108-1116.
277. Bond, H.M., et al., *ZNF423: a new player in estrogen receptor-positive breast cancer*. Frontiers in endocrinology, 2018. **9**: p. 255.
278. Huang, S., et al., *ZNF423 is critically required for retinoic acid-induced differentiation and is a marker of neuroblastoma outcome*. Cancer cell, 2009. **15**(4): p. 328-340.
279. Maira, S.-M., et al., *Carboxyl-terminal modulator protein (CTMP), a negative regulator of PKB/Akt and v-Akt at the plasma membrane*. Science, 2001. **294**(5541): p. 374-380.
280. Lin, C.-H., et al., *Carboxyl-terminal modulator protein facilitates tumor metastasis in triple-negative breast cancer*. Cancer Gene Therapy, 2023. **30**(3): p. 404-413.
281. Knobbe, C.B., et al., *Hypermethylation and transcriptional downregulation of the carboxyl-terminal modulator protein gene in glioblastomas*. Journal of the National Cancer institute, 2004. **96**(6): p. 483-486.
282. Hwang, S.-K., et al., *Carboxyl-terminal modulator protein induces apoptosis by regulating mitochondrial function in lung cancer cells*. International journal of oncology, 2012. **40**(5): p. 1515-1524.
283. Mok, E.H.K., C.O.N. Leung, and T.K.W. Lee, *MAP9/ERCC3 signaling cascade: A new insight on understanding the chromosomal instability in hepatocellular carcinoma*. EBioMedicine, 2020. **54**.
284. Rouquier, S., et al., *Expression of the microtubule-associated protein MAP9/ASAP and its partners AURKA and PLK1 in colorectal and breast cancers*. Disease markers, 2014. **2014**(1): p. 798170.
285. Yee, S.W., et al., *Deorphaning a solute carrier 22 family member, SLC22A15, through functional genomic studies*. FASEB journal: official publication of the Federation of American Societies for Experimental Biology, 2020. **34**(12): p. 15734.
286. Nayak, D., et al., *EMT alterations in the solute carrier landscape uncover SLC22A10/A15 imposed vulnerabilities in pancreatic cancer*. Iscience, 2022. **25**(5).
287. Thomas, L., et al., *TTC12 loss-of-function mutations cause primary ciliary dyskinesia and unveil distinct dynein assembly mechanisms in motile cilia versus flagella*. The American Journal of Human Genetics, 2020. **106**(2): p. 153-169.
288. Korkola, J., et al., *In vivo differentiation and genomic evolution in adult male germ cell tumors*. Genes, Chromosomes and Cancer, 2008. **47**(1): p. 43-55.

289. Ademuyiwa, F.O., et al., *Immunogenomic profiling and pathological response results from a clinical trial of docetaxel and carboplatin in triple-negative breast cancer*. Breast cancer research and treatment, 2021. **189**(1): p. 187-202.
290. O'Neil, N.J., M.L. Bailey, and P. Hieter, *Synthetic lethality and cancer*. Nature Reviews Genetics, 2017. **18**(10): p. 613-623.
291. Sala-Torra, O., et al., *Connective tissue growth factor (CTGF) expression and outcome in adult patients with acute lymphoblastic leukemia*. Blood, 2007. **109**(7): p. 3080-3083.
292. Wang, S., et al., *MAP9 loss triggers chromosomal instability, initiates colorectal tumorigenesis, and is associated with poor survival of patients with colorectal cancer*. Clinical Cancer Research, 2020. **26**(3): p. 746-757.
293. Zhang, J., et al., *Microtubule associated protein 9 inhibits liver tumorigenesis by suppressing ERCC3*. EBioMedicine, 2020. **53**.
294. Xiao, H., et al., *Expression of MAP9 in Epstein–Barr virus-associated gastric carcinoma*. Virus Research, 2021. **293**: p. 198253.
295. Haferlach, T., et al., *Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group*. J Clin Oncol, 2010. **28**(15): p. 2529-37.
296. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity*. Nature, 2012. **483**(7391): p. 603-607.
297. Gyrd-Hansen, M., et al., *Apoptosome-independent activation of the lysosomal cell death pathway by caspase-9*. Mol Cell Biol, 2006. **26**(21): p. 7880-91.
298. Koepsell, H., *Organic cation transporters in health and disease*. Pharmacological reviews, 2020. **72**(1): p. 253-319.
299. Eraly, S.A. and S.K. Nigam, *Novel human cDNAs homologous to Drosophila Orct and mammalian carnitine transporters*. Biochemical and biophysical research communications, 2002. **297**(5): p. 1159-1166.
300. Habra, K., et al., *Anticancer actions of carnosine in cellular models of prostate cancer*. Journal of Cellular and Molecular Medicine, 2024. **28**(2): p. e18061.
301. Maugeri, S., et al., *The Anti-Cancer Activity of the Naturally Occurring Dipeptide Carnosine: Potential for Breast Cancer*. Cells, 2023. **12**(22): p. 2592.
302. Oppermann, H., et al., *Carnosine inhibits glioblastoma growth independent from PI3K/Akt/mTOR signaling*. PLoS One, 2019. **14**(6): p. e0218972.
303. Ono, H., et al., *Carboxy-terminal modulator protein induces Akt phosphorylation and activation, thereby enhancing antiapoptotic, glycogen synthetic, and glucose uptake pathways*. American Journal of Physiology-Cell Physiology, 2007. **293**(5): p. C1576-C1585.
304. Saffin, J.-M., et al., *ASAP, a human microtubule-associated protein required for bipolar spindle assembly and cytokinesis*. Proceedings of the National Academy of Sciences, 2005. **102**(32): p. 11302-11307.
305. Han, H., *RNA interference to knock down gene expression*. Disease gene identification: methods and protocols, 2018: p. 293-302.
306. Sontheimer, E.J., *Assembly and function of RNA silencing complexes*. Nature Reviews Molecular Cell Biology, 2005. **6**(2): p. 127-138.
307. Rand, T.A., et al., *Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity*. Proceedings of the National Academy of Sciences, 2004. **101**(40): p. 14385-14389.
308. Matranga, C., et al., *Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes*. Cell, 2005. **123**(4): p. 607-620.
309. Ameres, S.L., J. Martinez, and R. Schroeder, *Molecular basis for target RNA recognition and cleavage by human RISC*. Cell, 2007. **130**(1): p. 101-112.
310. Rand, T.A., et al., *Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation*. Cell, 2005. **123**(4): p. 621-629.

311. Hutvagner, G. and P.D. Zamore, *A microRNA in a multiple-turnover RNAi enzyme complex*. Science, 2002. **297**(5589): p. 2056-2060.
312. Bartlett, D.W. and M.E. Davis, *Insights into the kinetics of siRNA-mediated gene silencing from live-cell and live-animal bioluminescent imaging*. Nucleic acids research, 2006. **34**(1): p. 322-333.
313. Ullah, I., et al., *Application of RNAi technology: a novel approach to navigate abiotic stresses*. Molecular Biology Reports, 2022. **49**(11): p. 10975-10993.
314. Cuny, G.D., et al., *Structure–activity relationship study of bone morphogenetic protein (BMP) signaling inhibitors*. Bioorganic & medicinal chemistry letters, 2008. **18**(15): p. 4388-4392.
315. Urist, M.R., A. Mikulski, and A. Lietze, *Solubilized and insolubilized bone morphogenetic protein*. Proceedings of the National Academy of Sciences, 1979. **76**(4): p. 1828-1832.
316. Termaat, M., et al., *Bone morphogenetic proteins: development and clinical efficacy in the treatment of fractures and bone defects*. Jbjs, 2005. **87**(6): p. 1367-1378.
317. Carreira, A., et al., *Bone morphogenetic proteins: facts, challenges, and future perspectives*. Journal of dental research, 2014. **93**(4): p. 335-345.
318. Wozney, J.M., et al., *Novel regulators of bone formation: molecular clones and activities*. Science, 1988. **242**(4885): p. 1528-1534.
319. Sun, P.D. and D.R. Davies, *The cystine-knot growth-factor superfamily*. Annual review of biophysics and biomolecular structure, 1995. **24**: p. 269-291.
320. Scheufler, C., W. Sebal, and M. Hülsmeier, *Crystal structure of human bone morphogenetic protein-2 at 2.7 Å resolution*. Journal of molecular biology, 1999. **287**(1): p. 103-115.
321. Zylbersztejn, F., et al., *The BMP pathway: A unique tool to decode the origin and progression of leukemia*. Experimental hematology, 2018. **61**: p. 36-44.
322. Gilde, F., et al., *Stiffness-dependent cellular internalization of matrix-bound BMP-2 and its relation to Smad and non-Smad signaling*. Acta biomaterialia, 2016. **46**: p. 55-67.
323. Miyazono, K., Y. Kamiya, and M. Morikawa, *Bone morphogenetic protein receptors and signal transduction*. Journal of biochemistry, 2010. **147**(1): p. 35-51.
324. Migliorini, E., et al., *Tuning cellular responses to BMP-2 with material surfaces*. Cytokine & growth factor reviews, 2016. **27**: p. 43-54.
325. Khattar, V., et al., *Structural determinants and genetic modifications enhance BMP2 stability and extracellular secretion*. FASEB BioAdvances, 2019. **1**(3): p. 180-190.
326. Bessa, P.C., et al., *Osteoinduction in human fat-derived stem cells by recombinant human bone morphogenetic protein-2 produced in Escherichia coli*. Biotechnology letters, 2008. **30**: p. 15-21.
327. Moussad, E.E.-D.A. and D.R. Brigstock, *Connective tissue growth factor: what's in a name?* Molecular genetics and metabolism, 2000. **71**(1-2): p. 276-292.
328. Leask, A. and D.J. Abraham, *All in the CCN family: essential matricellular signaling modulators emerge from the bunker*. Journal of cell science, 2006. **119**(23): p. 4803-4810.
329. Dhar, A. and A. Ray, *The CCN family proteins in carcinogenesis*. Exp Oncol, 2010. **32**(1): p. 2-9.
330. Hall-Glenn, F. and K.M. Lyons, *Roles for CCN2 in normal physiological processes*. Cellular and Molecular Life Sciences, 2011. **68**: p. 3209-3217.
331. Kang, Y., et al., *A multigenic program mediating breast cancer metastasis to bone*. Cancer cell, 2003. **3**(6): p. 537-549.
332. Xie, D., et al., *Levels of expression of CYR61 and CTGF are prognostic for tumor progression and survival of individuals with gliomas*. Clinical Cancer Research, 2004. **10**(6): p. 2072-2081.

333. Lipson, K.E., et al. *CTGF is a central mediator of tissue remodeling and fibrosis and its inhibition can reverse the process of fibrosis*. in *Fibrogenesis & tissue repair*. 2012. Springer.
334. Sgalla, G., et al., *Pamrevlumab for the treatment of idiopathic pulmonary fibrosis*. Expert Opinion on Investigational Drugs, 2020. **29**(8): p. 771-777.
335. Tatemoto, K., M. Carlquist, and V. Mutt, *Neuropeptide Y—a novel brain peptide with structural similarities to peptide YY and pancreatic polypeptide*. Nature, 1982. **296**(5858): p. 659-660.
336. Thorsell, A. and A.A. Mathé, *Neuropeptide Y in alcohol addiction and affective disorders*. Frontiers in Endocrinology, 2017. **8**: p. 178.
337. Tilan, J. and J. Kitlinska, *Neuropeptide Y (NPY) in tumor growth and progression: Lessons learned from pediatric oncology*. Neuropeptides, 2016. **55**: p. 55-66.
338. Chang, H.-M., et al., *Connective tissue growth factor mediates bone morphogenetic protein 2-induced increase in hyaluronan production in luteinized human granulosa cells*. Reproductive Biology and Endocrinology, 2022. **20**(1): p. 65.
339. Maeda, A., et al., *CCN family 2/connective tissue growth factor modulates BMP signalling as a signal conductor, which action regulates the proliferation and differentiation of chondrocytes*. Journal of biochemistry, 2009. **145**(2): p. 207-216.
340. Yan, S., et al., *CTGF promotes the osteoblast differentiation of human periodontal ligament stem cells by positively regulating BMP2/Smad signal transduction*. BioMed Research International, 2022. **2022**(1): p. 2938015.
341. Warming, S., et al., *Evi3, a common retroviral integration site in murine B-cell lymphoma, encodes an EBF4-related Kruppel-like zinc finger protein*. Blood, The Journal of the American Society of Hematology, 2003. **101**(5): p. 1934-1940.
342. Harder, L., A.-C. Puller, and M.A. Horstmann, *ZNF423: transcriptional modulation in development and cancer*. Molecular & cellular oncology, 2014. **1**(3): p. e969655.
343. Hata, A., et al., *OAZ uses distinct DNA-and protein-binding zinc fingers in separate BMP-Smad and Olf signaling pathways*. Cell, 2000. **100**(2): p. 229-240.
344. Warming, S., et al., *Early B-cell factor-associated zinc-finger gene is a frequent target of retroviral integration in murine B-cell lymphomas*. Oncogene, 2004. **23**(15): p. 2727-2731.
345. Miyazaki, K., et al., *Enhanced expression of p210BCR/ABL and aberrant expression of Zfp423/ZNF423 induce blast crisis of chronic myelogenous leukemia*. Blood, The Journal of the American Society of Hematology, 2009. **113**(19): p. 4702-4710.
346. *BMP2 protein western blot*. [cited 2024; Available from: <https://www.thermofisher.com/antibody/product/BMP2-Antibody-Polyclonal/BS-0514R>].
347. Ohara, Y., et al., *Connective tissue growth factor-specific monoclonal antibody inhibits growth of malignant mesothelioma in an orthotopic mouse model*. Oncotarget, 2018. **9**(26): p. 18494.
348. Dornhofer, N., et al., *Connective tissue growth factor–specific monoclonal antibody therapy inhibits pancreatic tumor growth and metastasis*. Cancer research, 2006. **66**(11): p. 5816-5827.
349. Jiramongkolchai, P., P. Owens, and C.C. Hong, *Emerging roles of the bone morphogenetic protein pathway in cancer: potential therapeutic target for kinase inhibition*. Biochemical Society Transactions, 2016. **44**(4): p. 1117-1134.
350. Yadin, D., P. Knaus, and T.D. Mueller, *Structural insights into BMP receptors: Specificity, activation and inhibition*. Cytokine & growth factor reviews, 2016. **27**: p. 13-34.
351. Vorwerk, P., et al., *CTGF (IGFBP-rP2) is specifically expressed in malignant lymphoblasts of patients with acute lymphoblastic leukaemia (ALL)*. British journal of cancer, 2000. **83**(6): p. 756-760.

352. Boag, J.M., et al., *High expression of connective tissue growth factor in pre-B acute lymphoblastic leukaemia*. British journal of haematology, 2007. **138**(6): p. 740-748.
353. Dai, Q., et al., *miR-539-5p targets BMP2 to regulate Treg activation in B-cell acute lymphoblastic leukemia through TGF- β /Smads/MAPK*. Experimental Biology and Medicine, 2024. **249**.
354. Kondo, Y. and J.-P.J. Issa, *DNA methylation profiling in cancer*. Expert reviews in molecular medicine, 2010. **12**: p. e23.
355. Robertson, K.D., *DNA methylation and human disease*. Nature Reviews Genetics, 2005. **6**(8): p. 597-610.
356. Ehrlich, M., *DNA hypomethylation in cancer cells*. Epigenomics, 2009. **1**(2): p. 239-259.
357. Tirado-Magallanes, R., et al., *Whole genome DNA methylation: beyond genes silencing*. Oncotarget, 2017. **8**(3): p. 5629.
358. Audeh, M.W., et al., *Oral poly (ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial*. The lancet, 2010. **376**(9737): p. 245-251.