
Modelling *Crebbp* loss in *BCL2* driven non-Hodgkin's lymphoma

J.C. Dawes

A Thesis Submitted to Imperial College London for the Degree of Doctor of Philosophy

1st October 2015

Cancer Genomics Group

Medical Research Council, Clinical Sciences Centre

Imperial College London

Declaration of Originality

I declare that the work presented in this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Signed:

A handwritten signature in black ink, consisting of several loops and a long horizontal stroke at the bottom.

Date: 1st October 2015

Abstract

Non-Hodgkin's lymphomas (NHL) are a spectrum of hematopoietic cancer accounting for 4% of new cancer diagnoses each year. Approximately 95% of all NHL are of B-cell origin; diffuse large B-cell lymphomas (DLBCL) and follicular lymphoma (FL) accounting for 30-40% and 20% of B-NHL respectively.

Recent human mutation profiling and resequencing studies have shown that *CREBBP* and *BCL2* are frequently mutated, early events in B-NHL that are often concurrent. This thesis presents a study of their role in oncogenesis by generating a novel model of B-NHL overexpressing *BCL2* in the haematopoietic compartment combined with the conditional heterozygous loss of *Crebbp* in the germinal centre B-cells. Loss of *Crebbp* significantly accelerated *BCL2* driven lymphomagenesis. Characterisation of these lymphomas by flow cytometry demonstrated that they were phenotypically similar to mature human B-cell lymphomas of germinal centre origin.

Additionally an insertional mutagenesis screen was performed in mice sensitised to B-NHL. Somatic MoMuLV caused mutations interact with the sensitising mutations, facilitating tumourigenesis and recapitulating the stepwise and accumulative progression of human disease.

Provirus insertions were detected using a novel method of insertion site cloning called UMI-LM-PCR which amplifies the region flanking the provirus. Insertions were then mapped to the mouse genome flagging genes with a putative role in tumourigenesis. This method represents the highest throughput method to date and its applicability to Illumina sequencing permits the most comprehensive quantitative survey of subclonal mutations.

This work confirms *Crebbp* is a tumour suppressor gene and that its loss cooperates with overexpression of *BCL2* to accelerate lymphomagenesis. Through insertional mutagenesis screening of this model hundreds of putative cancer genes have been identified including *Pou2f2* and *Tfrc* which have a propensity for mutation specifically in B-cell lymphomas deficient in *Crebbp* and overexpressing *BCL2*. These candidates now need to be validated to better characterise their role in B-cell lymphoma.

Acknowledgements

Thank you to my supervisor Anthony Uren for the opportunity and his guidance; and to the MRC and Imperial for making our work possible. Thank you to my lab, Cancer Genomics: Phil, Jakub, Kata, Barbara, Hamlata and Bruce for keeping me company, especially to my lab brothers for keeping me going and keeping me laughing.

I am also grateful for all the support I have received whilst at the CSC: Ola and Emma in Central Biomedical Services, Laurence, Ivan, Kat and Harshal in the Genomics Facility and James and Thomas in the FACS Facility.

I am indebted to my former lab at the Università degli Studi di Cagliari. Prof., Vera, Gabriella and Monica, I can hear your words of advice with every experiment I do. It is as if you are standing over my shoulder again. To my undergraduate lab at University of Surrey thank you for your introduction to research and writing.

To all the fish in the fishbowl and my PhD buddies with whom I shared the last four years of my life, you are all a part of this. Thank you for every kind word, snack and uplifting cat video along the way.

To my parents, thank you for giving me the curiosity to learn and for all the advice along the way. To my sister, Antonia, thank you for being by my side for the last four years and putting up with me for the last twenty eight...To my Dario, thank you for being you, and for being my biggest supporter. I am your biggest fan.

And thank you to all the giants whose shoulders I am standing on.

Softly, softly, catchee monkey.

Cuoncio cuoncio!

Abbreviations

ABC	Activated B-cell
Aicda	Activation-induced cytidine deaminase
ALV	Avian leukaemia virus
BCL2	B-cell lymphoma 2
BCR	B-cell receptor
BL	Burkitt's lymphoma
B-NHL	B-cell non-Hodgkin's lymphoma
bp	Base pair
BP	B-cell Panel
CIS	Common insertion site
CREBBP	CREBBP binding protein
CSR	Class switch recombination
DLBCL	Diffuse large B-cell lymphoma
EP300	E1A binding protein p300
FL	Follicular lymphoma
FLV	Feline leukaemia virus
GC	Germinal centre
GCB	Germinal centre B-cell
GCP	Germinal Centre Panel
GP	General Panel
IgH	Immunoglobulin heavy chain
IgL	Immunoglobulin light chain
IM	Insertional mutagenesis
LAT3	lysine acetyltransferase 3 family
LKP	Lambda/Kappa Panel
LM-PCR	Ligation mediated PCR
mCAT1	Murine cationic amino-acid transporter
MMTV	Mouse mammary tumour virus
MoMuLV	Moloney murine leukaemia virus
NHL	Non-Hodgkin's lymphoma
RAG	recombination-activation gene
R-CHOP	Rituximab, cyclophosphamide, oncovin (vincristine), prednisolone
RTS	Rubinstein-Taybi syndrome
SHM	Somatic hypermutation
SPRI	Solid phase reversible immobilisation
UMI	Unique molecular identifier

Table of Contents

Declaration of Originality.....	i
Copyright Declaration	i
Abstract.....	ii
Acknowledgements.....	iii
Abbreviations.....	iv
Table of Contents.....	vi
List of Figures.....	x
List of Tables	xii
1. Introduction.....	1
1.1. Non-Hodgkin’s Lymphoma	1
1.1.1. Development of B-cells and B-cell lymphoma	2
1.1.2. Diffuse Large B-cell Lymphoma	7
1.1.3. Follicular Lymphoma	7
1.2. B-cell lymphoma 2, BCL2	8
1.3. CREB-binding protein, CREBBP	10
1.4. Insertional Mutagenesis	13
1.4.1. Moloney Murine Leukaemia Virus	14
1.4.2. Integration site cloning technological advancements.....	19
1.5. Murine Models of NHL	20
1.5.1. VavP-BCL2 transgenic.....	20
1.5.2. <i>Crebbp</i> flox allele.....	21
1.5.3. <i>Aicda</i> ^{Cre} knockin.....	23
1.6. Research Aims and Approach.....	24
2. Materials and Methods	25
2.1. Animal Models.....	25
2.1.1. Breeding strategy.....	25
2.1.2. Genotyping	26
2.1.2.1. <i>Crebbp</i> ^{fl}	26

2.1.2.2.	<i>Aicda</i> ^{Cre}	27
2.1.2.3.	<i>VavP-BCL2</i>	27
2.2.	Moloney Murine Leukaemia Virus	28
2.2.1.	MoMuLV production	28
2.2.2.	MoMuLV quantification	28
2.3.	Generation of MoMuLV driven lymphoma	29
2.4.	Flow Cytometry	29
2.5.	Tumour RNA and DNA extraction	31
2.6.	Spleen suspension B- and T-cell purification	31
2.7.	UMI-LM-PCR (Unique Molecular Identifier – Ligation Mediated – PCR).....	32
2.7.1.	Shear	32
2.7.2.	Blunt and A-tail	32
2.7.3.	Ligation of indices, unique adaptor	33
2.7.4.	EcoRV digest.....	34
2.7.5.	Size selection.....	34
2.7.6.	Primary PCR.....	35
2.7.7.	Secondary PCR.....	36
2.7.8.	HiSeq.....	36
2.8.	Bioinformatics Pipeline.....	39
3.	Characterisation of a novel murine model of spontaneous B-cell lymphoma.....	43
3.1.	Spontaneous lymphoma cohort.....	43
3.2.	Spontaneous lymphoma survival analysis	44
3.3.	Cytometric characterisation of the spontaneous lymphoma	48
3.3.1.	General Panel	48
3.3.2.	B-cell Panel	56
3.3.3.	Lambda/Kappa Panel	63
3.3.4.	Germinal Centre Panel	66
3.4.	Summary of the spontaneous lymphoma	70

4.	Retrovirus induced lymphoma.....	71
4.1.	Quantification of MoMuLV.....	72
4.2.	MoMuLV induced lymphoma cohort.....	74
4.3.	MoMuLV induced lymphoma cohort survival analysis.....	76
4.4.	Cytometric characterisation of retrovirus induced lymphoma.....	78
4.4.1.	General Panel.....	78
4.4.2.	B-cell Panel.....	88
4.4.3.	Lambda/Kappa Panel.....	94
4.4.4.	Germinal centre Panel.....	98
4.5.	Summary of MoMuLV induced lymphoma.....	101
5.	Retroviral insertion site cloning and analysis of MoMuLV induced lymphoma.....	103
5.1.	Improvements to library preparation and sequencing.....	104
5.1.1.	UMI-LM-PCR.....	104
5.1.2.	Sequencing.....	108
5.1.3.	Quality Control of UMI-LM-PCR.....	109
5.2.	Characterisation of inserts.....	111
5.3.	Common insertion sites.....	113
5.4.	Subclonal mutations in B-cell lymphoma.....	117
5.5.	Analysis of lymphoma dissemination.....	120
5.5.1.	Analysis of cells specific insertions within the spleen.....	122
5.6.	Genes collaborating with <i>BCL2</i> or <i>Crebbp</i> deficiency.....	124
5.6.1.	Genes that collaborate with <i>BCL2</i>	127
5.6.1.1.	<i>BCL2</i> specific mutation: <i>Pou2f2</i>	133
5.6.2.	Genes that collaborate with <i>Crebbp</i> deficiency independently of <i>BCL2</i>	135
5.6.2.1.	<i>Crebbp</i> specific mutation in the absence of the <i>BCL2</i> transgene: <i>Tfrc</i>	138
5.7.	Co-occurrence and mutual exclusivity.....	141
5.8.	Summary of insertion data.....	146

6.	Discussion and future work	147
6.1.	Overview	147
6.2.	<i>Crebbp</i> and the acceleration of <i>BCL2</i> driven lymphomagenesis	148
6.3.	Technological advancement.....	150
6.4.	Future work.....	151
6.5.	Concluding remarks	152
	References.....	153
	Appendices	165
A.1	Full list of Adaptors, MRC#854-945.....	165
A.2	Secondary PCR primers, MRC#946-1029	169

List of Figures

Figure 1.1. B-cell development and surface markers	3
Figure 1.2. B-cell remodelling and cellular origin of B-cell lymphoma	6
Figure 1.3. Retroviral insertional mutagenesis	15
Figure 2.1. Detail of UMI-LM-PCR: Example Upper Stand Adaptor MRC#850	33
Figure 2.2. Detail of UMI: EcoRV digest strategy	34
Figure 2.3. Detail of UMI-LM-PCR: Primary PCR	35
Figure 2.4. Overview of UMI-LM-PCR and HiSeq	38
Figure 2.5. Bioinformatics Pipeline	40
Figure 2.6. Analysis Pipeline	41
Figure 3.1. BCL2 non-transgenic cohort survival curves	45
Figure 3.2. BCL2 transgenic cohort survival curves	47
Figure 3.3. Spontaneous lymphoma cohort GP gating strategy	50
Figure 3.4. Spontaneous lymphoma cohort examples of the varying proportions of B and T-cells	51
Figure 3.5. Spontaneous lymphoma cohort representative CD4 and CD8a plots	52
Figure 3.6. Spontaneous lymphoma cohort GP outliers	53
Figure 3.7. Spontaneous lymphoma cohort GP overview of tumours composition	54
Figure 3.8. Spontaneous lymphoma cohort GP overview of tumours composition by genotype	55
Figure 3.9. Spontaneous lymphoma cohort BP gating strategy	57
Figure 3.10. Spontaneous lymphoma cohort characteristic immunoglobulin proportions	59
Figure 3.11. Spontaneous lymphoma cohort BP outliers	60
Figure 3.12. Spontaneous lymphoma cohort BP overview of immunoglobulin proportions	61
Figure 3.13. Spontaneous lymphoma cohort BP overview of immunoglobulin proportions by genotype	62
Figure 3.14. Spontaneous lymphoma cohort LKP gating strategy	64
Figure 3.15. Spontaneous lymphoma cohort demonstration of homogeneity of Lambda/Kappa ratios	65
Figure 3.16. Spontaneous lymphoma cohort gating strategy for the GCP	66
Figure 3.17. Spontaneous lymphoma cohort typical CD95 PNA expression	67
Figure 3.18. Spontaneous lymphoma cohort GCP overview of tumours composition	68
Figure 3.19. Spontaneous lymphoma cohort GCP overview of tumours composition by genotype	69
Figure 4.1. Quantification of MoMuLV	73
Figure 4.2. Examples of mouse pathology	75
Figure 4.3. MoMuLV induced lymphoma survival curves	77
Figure 4.4. MoMuLV induced lymphoma cohort GP gating strategy	78
Figure 4.5. MoMuLV induced lymphoma a cohort examples of the varying proportions of B and T-cells	79

Figure 4.6. MoMuLV induced lymphoma cohort with unusual B and T-cell profiles.....	80
Figure 4.7. MoMuLV induced lymphoma cohort representative CD4 and CD8a plots.....	81
Figure 4.8. MoMuLV induced lymphoma cohort abnormal CD8a and CD4 profiles	82
Figure 4.9. MoMuLV induced lymphoma cohort GP overview of tumours composition	83
Figure 4.10. MoMuLV induced lymphoma cohort GP overview of tumours composition by genotype....	85
Figure 4.11. MoMuLV induced lymphoma cohort GP overview of T-cell proportions by genotype	87
Figure 4.12. MoMuLV induced lymphoma cohort BP gating strategy.....	88
Figure 4.13. MoMuLV lymphoma cohort characteristic immunoglobulin profiles	90
Figure 4.14. MoMuLV lymphoma cohort atypical plots.....	91
Figure 4.15. MoMuLV induced lymphoma cohort BP overview of tumours composition	92
Figure 4.16. MoMuLV induced lymphoma cohort BP overview of immunoglobulin proportions by genotype.....	93
Figure 4.17. MoMuLV induced lymphoma cohort LKP gating strategy	94
Figure 4.18. MoMuLV induced lymphoma cohort demonstration of homogeneity of Lambda/Kappa ratios	96
Figure 4.19. MoMuLV induced lymphoma cohort abnormal Lambda/Kappa ratios.....	97
Figure 4.20. MoMuLV induced lymphoma cohort gating strategy for the Germinal Centre Panel.....	98
Figure 4.21. MoMuLV induced lymphoma cohort typical CD95 PNA expression.....	99
Figure 4.22. MoMuLV induced lymphomas assessment of germinal centres	100
Figure 5.1. Quality control of libraries using the Agilent bioanalyser	105
Figure 5.2. Definition of insert, reads and fragments	106
Figure 5.3. Schematic of UMI-LM-PCR.....	107
Figure 5.4. Heat maps comparing the number of insertion sites shared between mice.....	109
Figure 5.5. Reproducibility of UMI-LM-PCR	110
Figure 5.6. Entropy scoring	112
Figure 5.7. Schematic of the Gaussian Convolution	113
Figure 5.8. Histogram describing the of B-cell lymphomas.....	117
Figure 5.9. Comparison of top ten most clonal insertions in different organs of the same animal	121
Figure 5.10. Comparison of top ten most clonal insertions in whole spleen and its equivalent B or T-cells	123
Figure 5.11. Pou2f2 insertions.....	133
Figure 5.12. Tfrc insertions	138
Figure 5.13. Co-occurring and mutually exclusive insertions in the BCL2 transgenic tumours.....	142
Figure 5.14. Co-occurring and mutually exclusive insertions in the BCL2 non-transgenic tumours	143
Figure 5.15. Co-occurring and mutually exclusive insertions in the Crebbp deficient tumours	144

List of Tables

Table 2.1. Genotype Key	25
Table 2.2. Summary of antibodies used in flow cytometry	30
Table 3.1. Spontaneous lymphoma cohort	44
Table 4.1. MoMuLV induced lymphoma cohort.....	74
Table 5.1. Top 100 CIS in the screen	116
Table 5.2. Top 50 B-cell lymphoma specific CIS	119
Table 5.3. Example contingency table test comparing Notch1 insertions in the transgenic and non-transgenic BCL2 mice	124
Table 5.4. Contingency table comparing the number of inserts in the spleens of the BCL2 transgenic mice compared to the non-transgenic BCL2 mice	127
Table 5.5. Contingency table comparing the number of inserts in the lymph node of the transgenic BCL2 mice compared to the non-transgenic BCL2 mice	128
Table 5.6. Contingency table comparing the number of inserts in the thymi of the transgenic BCL2 mice compared to the non-transgenic BCL2 mice	129
Table 5.7. Contingency table comparing the number of inserts in the spleen of the CrebbpCreBCL2 mice compared to the other transgenic BCL2 mice	130
Table 5.8. Contingency table comparing the number of inserts in the lymph node of the CrebbpCreBCL2 mice compared to the other transgenic BCL2 mice.....	131
Table 5.9. Contingency table comparing the number of inserts in the thymi of the CrebbpCreBCL2 mice compared to the other transgenic BCL2 mice	132
Table 5.10. Contingency table comparing the number of inserts in the spleen of the CrebbpCre mice compared to the Crebbp & Cre & wild type mice	135
Table 5.11. Contingency table comparing the number of inserts in the lymph node of the CrebbpCre mice compared to the Crebbp & Cre & wild type mice	136
Table 5.12. Contingency table comparing the number of inserts in the thymi of the CrebbpCre mice compared to the Crebbp & Cre & wild type mice	137
Table 5.13. Example contingency table test demonstrating the mutual exclusivity of Myc and Mycn mutations.....	141

1. Introduction

1.1. Non-Hodgkin's Lymphoma

Non-Hodgkin's lymphomas (NHL) are a spectrum of hematopoietic malignancies of B, T or natural killer lymphocytes accounting for approximately 4% of new cancer diagnoses each year. In the United Kingdom NHL is the sixth most frequently diagnosed cancer and the eleventh most common cause of cancer death (*Cancer Research UK*, <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/non-hodgkin-lymphoma#heading-One>, Accessed July 2015). The vast majority of NHLs are of B-cell origin, 95%, with B-cell non-Hodgkin's lymphoma (B-NHL) being further subdivided into approximately fifteen subgroups based on architecture and morphology (described later in Figure 1.2).

Since 1978 the empiric treatment of B-NHL has been CHOP (cyclophosphamide, doxorubicin hydrochloride, vincristine (previously oncovin) and prednisolone). The addition of rituximab, a humanised monoclonal against the pan-B-lymphocyte antibody CD20 of B-lymphocytes has improved outcomes (Coiffier et al., 2002). Unfortunately R-CHOP remains non-curative in many patients and has undesirable side effects such as anaemia, nausea and hair loss, highlighting the need for novel agents and therapeutic targets (Pfreundschuh et al., 2008). Despite recent discoveries about the genetic and molecular changes of the various B-NHL subtypes and more detailed stratification in the WHO 2008 report, (reviewed in Campo et al., 2011), it is clear that new insight is needed into this constellation of diseases to allow more targeted treatments with fewer side effects and better outcomes.

In recent years genome-wide analysis of copy number alterations and exome sequencing of B-NHL genomes have shown that, like most tumours, they suffer genetic aberrations like amplifications, deletions, deregulation of methylation and non-synonymous point mutations. B-NHLs are particularly characterised by chromosomal translocations and aberrant somatic hypermutation; processes which are integral to antibody remodelling in the germinal centres (GCs) and a fundamental vulnerability of B-cell development (Basso and Dalla-Favera, 2015; Küppers, 2005; Pasqualucci et al., 2001).

1.1.1. Development of B-cells and B-cell lymphoma

The massive clonal expansion and genome remodelling that occurs during B-cell development makes it intrinsically prone to oncogenic transformation. Early B-cells development occurs in the foetal liver or the adult bone marrow and originates from hematopoietic stem cells that undergo a stepwise maturation process (Figure 1.1).

Early development is characterised by the progressive development and modification of the B-cell receptor (BCR). Immunoglobulin heavy chain (IgH) gene rearrangements (V_H , D_H and J_H) together with the V_L - J_L rearrangements of the light chain (IgL) genes segments generate a B-cell repertoire expressing antibodies capable of recognising a wide variety of antigens. Once the BCR is formed, the immature naïve B-cells are able to participate in the immune response and leave the bone marrow. If the BCR does not form correctly, or is autoreactive, the cell undergoes apoptosis (Figure 1.2). Immature B-cells, now with functional BCRs, migrate via the blood to the secondary lymphoid organs like the spleen and lymph nodes where they form follicles during the T-cell dependent immune response.

Within the follicle, B-cells undergo two further remodelling steps in the germinal centres (GC): somatic hypermutation (SHM) and class switch recombination (CSR) (Figure 1.2). Somatic hypermutation is a genetic mechanism that introduces point mutations, deletions or duplications into the variable region of the immunoglobulin genes (Küppers et al., 1993). Immunoglobulins also alter their functionality by CSR, the process of replacing the IgH chain constant region, altering the effector response but not the antigen binding domain. GC B-cells that have undergone CSR and have high affinity IgG BCRs are selected for and differentiate into plasma cells and long lived plasma cells, the effectors of adaptive immunity.

Somatic hypermutation and class switch recombination are mediated by the enzyme activation-induced deaminase (encoded by the *AICDA* gene) which is primarily expressed in the GC B-cells. AICDA deaminates cytosine residues of single stranded DNA which are exposed during transcription. The resulting mismatch is recognised and processed by error prone DNA repair strategies that can introduce somatic mutation and predispose a sequence to chromosomal translocation (Muramatsu et al., 1999).

The genome is normally protected from AICDA-dependent IgH translocations by *Atm*-, *p53*- and *p19Arf*-dependent pathways that detect and signal DNA damage and oncogenic stress (Ramiro et al., 2006). Unfortunately it is evident that these protective mechanisms can fail and as a consequence the IgH locus is involved in nearly all cancer-associated chromosomal translocations in mature B-cell lymphomas (reviewed in Küppers and Dalla-Favera, 2001).

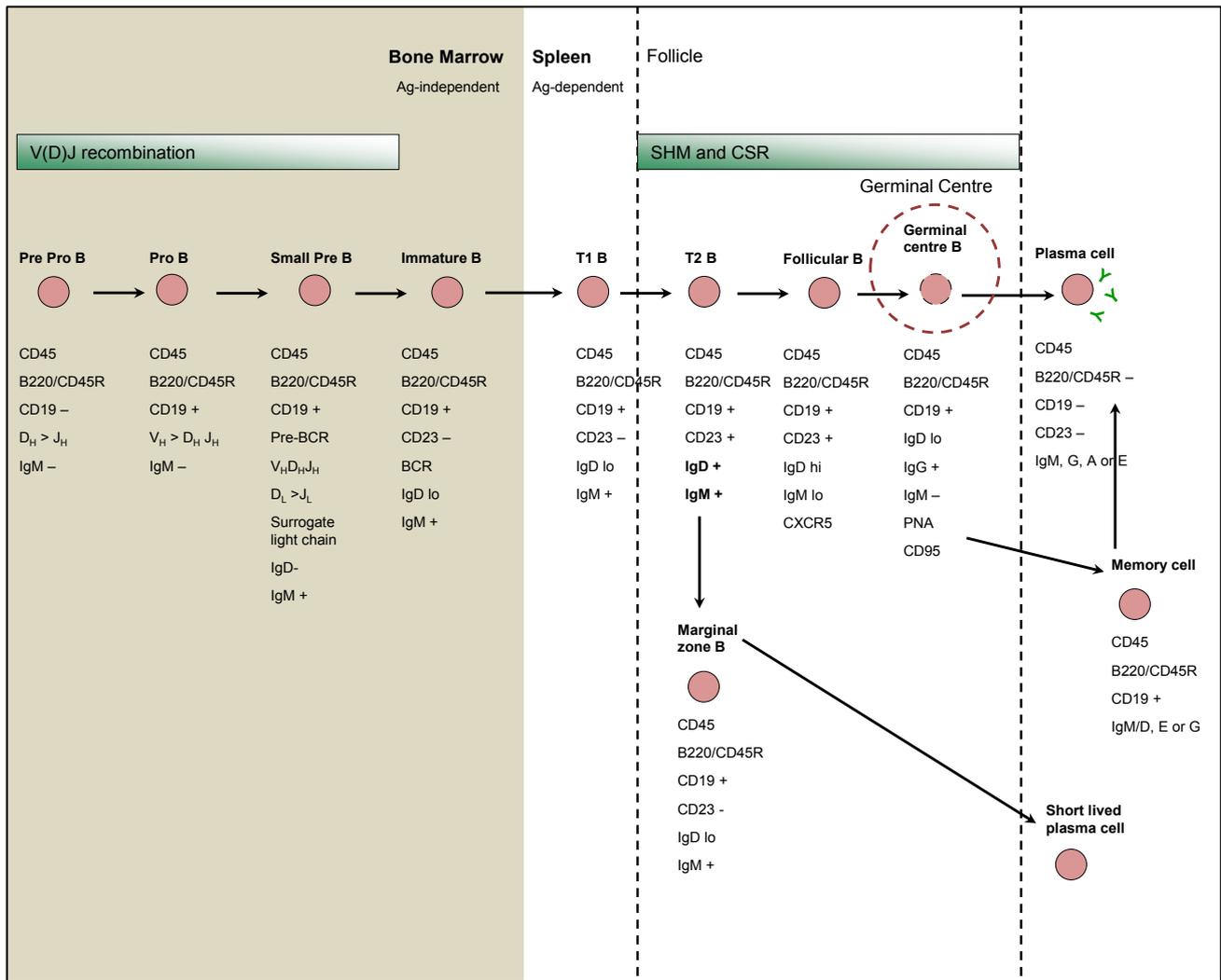


Figure 1.1. B-cell development and surface markers

B-cell differentiation is characterised by the progressive development and modification of the BCR. Immunoglobulin heavy and light chain gene rearrangements generate a B-cell repertoire expressing a variety of antibodies capable of recognising a variety of antigens. Once the heavy and light chains successfully rearrange into a functional BCR, the immature B-cell can leave the bone marrow and go, via the blood, to the secondary lymphoid organs like the spleen. Before migration, the BCR of immature cells is tested for reactivity to self-antigens: autoreactivity induces apoptosis by the BCR, while weak reactivity induces anergy (Cambier et al., 2007). When affinity for self-elements is low, the cells migrate out of the bone marrow to the secondary lymphoid organs (Niiri and Clark, 2002).

B220/CD45R is expressed from the earliest pre-pro-B-cells throughout B-cell development. CD19 is present from the pre-B-cell stage through to terminal differentiation of plasma cells. As immature B-cells differentiate their CD19 expression can increase 3-fold. Immature B-cells enter the spleen as transitional T1 B-cells which develop into T2 B-cells and then follicular B-cells or marginal zone B-cells. Marginal zone cells have low avidity BCRs and go to the marginal zone surrounding the germinal centre. T2 B-cells are the first to be IgD+ IgM+ and indicate the midpoint between the IgD- cells earlier in development, the IgD+ and the subsequently class switched IgG+, IgE+ and IgA+. GC B-cells undergo somatic hypermutation and class switch recombination to produce diverse BCRs. GC B-cells with high affinity IgG are selected for differentiation into plasma and memory cells, the effectors of the adaptive immune system. Some IgG+ cells transiently switch to IgE+ and then quickly differentiate into plasma cells (Haas and Tedder, 2005; Nojima et al., 2011; Xu, 2006) Figure adapted from Küppers, 2005 and Pieper et al., 2013.

In the T-cell dependent immune response B-cells present antigen to primed T-helper cells in the follicles of the secondary lymphoid organs and then, stimulated by signals from the T-helper cells, they clonally proliferate forming a follicle (see Figure 1.2). Some B-cells leave the extra follicular zone and become plasmablasts which differentiate into short lived plasma cells. Other B-cells rapidly proliferate and form the germinal centres where B-cells genes undergo SHM of the immunoglobulin gene and CSR to IgG, IgA and IgE producing diverse affinity antibodies (Rajewsky, 1996).

The follicle is made up of the marginal zone, mantle zone and the GC. This micro anatomical structure within secondary lymphoid organs is where mature B-cells are presented antigen by T-helper cells (Figure 1.2. blue cell) and follicular dendritic cells (purple cell) and leads to the development of antibody secreting plasma cells and memory cells.

The marginal zone is rich in B-cells and forms the first line of defence against blood borne pathogens. When these are encountered the marginal zone B-cells develop into extra-follicular plasma cells that secrete IgM which forms immune complexes with pathogens neutralising them. Human marginal zone B-cells express IgM with SHM within their variable region and they circulate in the blood indicating that they have passed through the GC unlike murine marginal zone B-cells that do not undergo SHM and remain in the spleen. The mantle zone is composed of naïve B-cells that have bypassed the GC, the majority of which are CD5+.

The GC is divided into the light and dark zones. In the dark zone cells proliferate rapidly and undergo clonal expansion to produce a variety of B-cells with different BCRs. They are also exposed to SHM of the immunoglobulin variable regions through the activity of AICDA which increases the diversity of the variable regions. Non-immunoglobulin genes like BCL6 can also be targeted for mutation by aberrant SHM in the GC which is thought to facilitate translocation (Pasqualucci et al., 2001). The cells then undergo class switch recombination of the immunoglobulin constant region and selection in the light zone of the GC. SHM and CSR aim to produce high diversity, high affinity antibodies and both these processes are mediated by AICDA. Mutations and recombination that reduce the BCR affinity for antigen are selected against and these cells undergo apoptosis.

Finally GC B-cells differentiate into plasma cells, which are the effectors of the immune system secreting antibody, or memory B-cell that leave the germinal centre providing a reservoir of quiescent antigen experienced B cells that can be reactivated upon re-exposure to antigen.

Each subgroup of B-NHL is representative of a discrete stage in B-cell differentiation and the tumours that arise from them are associated with different characteristic phenotypes and genetic lesions suggesting that different pathways of tumourigenesis were involved in their manifestation (reviewed in Küppers,

2005) (see Figure 1.2). Approximately 80% of B-NHL lymphoma, including diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL) and Burkitt's lymphoma (BL), are derived from GC B-cells as demonstrated by the presence of SHM (Stevenson et al., 2001).

Reciprocal chromosomal translocation of the immunoglobulin loci and proto-oncogenes is a classical hallmark of B-NHL whereby double strand DNA breaks put an oncogene under the control of the highly expressed immunoglobulin locus causing it's deregulated and constitutive expression; this can occur in a variety of ways and at different stages in B-cell development (reviewed in Willis and Dyer, 2000). Common translocations include BCL2 (Manolov and Manolova, 1972) BCL6 (Weiss et al., 1987) and MYC (Ladanyi et al., 1991)

By identifying the break point of a translocation it is possible to assess at what stage it occurred. Pathological chromosomal translocations with break points adjacent to J_H or D_H indicate that the translocations occurred at the pre-pro-B-cell stage in the bone marrow. Breakpoints adjacent to rearranged V(D)J genes with somatically mutated variable portions indicate that the translocation occurred later in development in the GC dark zone. Translocations are also possible during class-switch recombination in the light zone of the GC, evidenced by breakpoints in the IgH constant region.

Breakpoints in the immunoglobulin locus are formed by an interruption of normal V(D)J recombination, however it is unclear how the equivalent break forms in the translocating oncogene (reviewed in Küppers and Dalla-Favera, 2001). There is evidence that aberrant SHM of non-immunoglobulin genes like the proto-oncogene BCL6 facilitates translocations to the IgH locus with hotspots of hypermutation found adjacent to the translocation break point (Pasqualucci et al., 2001).

The IgH-BCL2 translocation, common in FL, can occur through aberrant recombination-activation gene (RAG) mediated V(D)J recombination early in B-cell development in the primary lymphoid organs (Jäger et al., 2000) and have been demonstrated to be early events in disease progression. Errors in AICDA-mediated CSR or SHM have been shown to cause Ig-MYC translocations (reviewed in Basso and Dalla-Favera, 2015).

The two most common subgroups of B-NHL are DLBCL and FL accounting for 48% and 19% of diagnosis respectively (Cancer Research UK, as above). Both of these subgroups are classified as mature B-cell neoplasms and both originate in the GC (summarised in Figure 1.2) (reviewed in Küppers, 2005). Recent mutation profiling and resequencing studies of human B-NHL patients have identified dozens of recurrently mutated genes and highlighted genetic similarities between DLBCL and FL (Morin et al., 2011; Pasqualucci et al., 2011a).

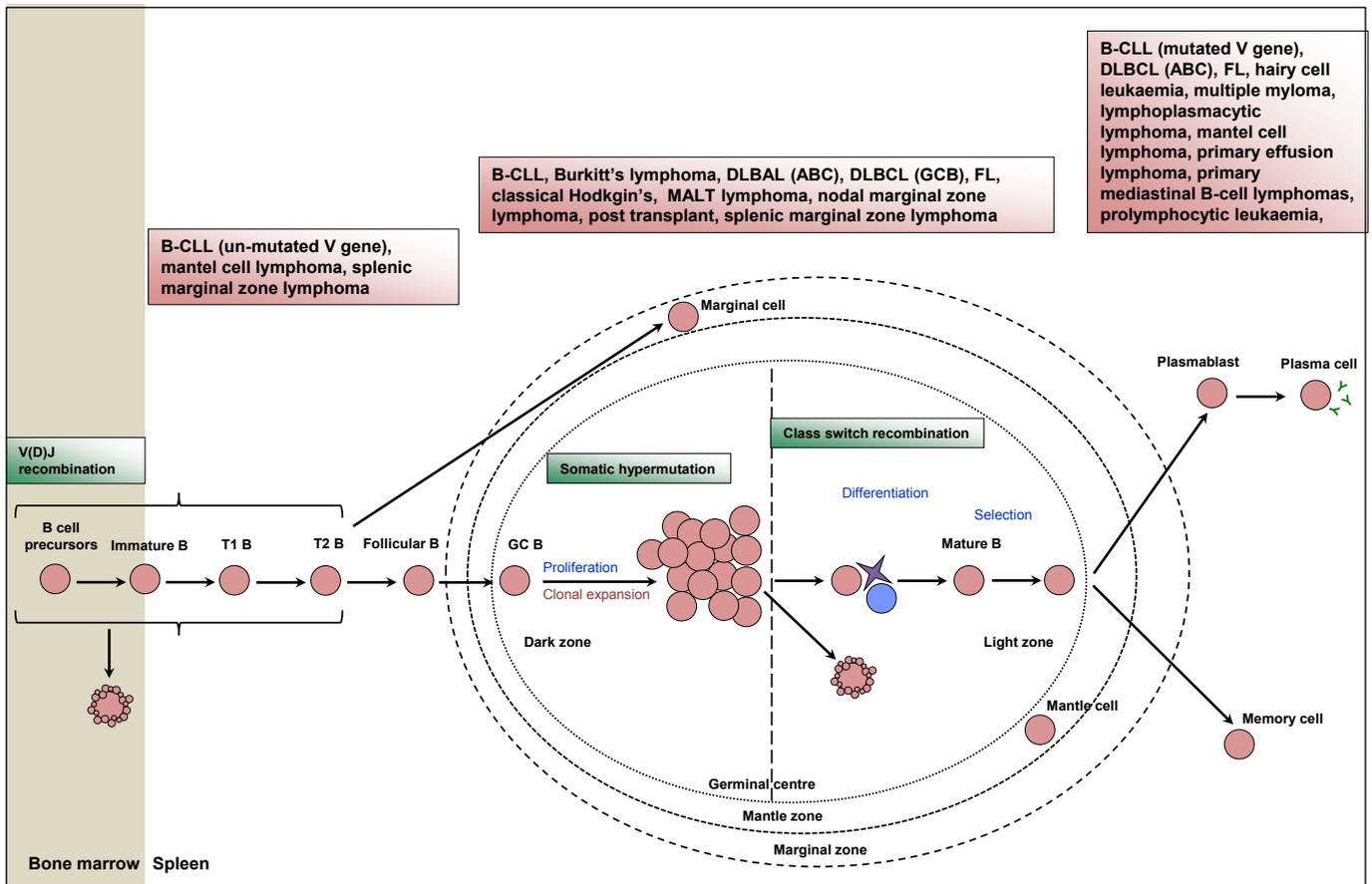


Figure 1.2. B-cell remodelling and cellular origin of B-cell lymphoma

The cellular origin of B-cell lymphomas can be deduced from the structure of the BCR, the expression pattern of surface markers, the structures where they arise, and more recently, by expression profiling. Though the exact cell type of origin or normal non-malignant counterpart is not always known for each tumour type it is apparent that the majority of B-NHL derive from GC B-cells or post GC B-cells. Oncogenic transformation is intrinsically linked to B-cell maturation processes like V(D)J recombination, SHM and CSR. Throughout maturation there are various checkpoints where abnormal cells can undergo apoptosis (cell with blebs).

B-CLL, mantle cell lymphoma, splenic marginal zone lymphoma originate from these early stages in B-cell development prior to entry in the GC Splenic marginal zone lymphoma arises from follicular or marginal zone B-cells prior to mutations of the immunoglobulin variable region (Dogan and Isaacson, 2003). Mantle cell lymphomas are predominately derived from the naïve CD5+ B-cell that make up the mantle zone. However 20-30% of cases have mutated variable regions indicating that they have passed through the GC and some also have immunoglobulin translocations indicating they are from even later stage B-cells that have undergone CSR (Küppers, 2005). B-CLL can arise throughout B-cell development with early examples similar to marginal zone B-cells (Schaffner et al., 2000), cases with mutated variable regions (suggesting they arise later) and cases with un-mutated variable genes but that have expression profiles similar to the latest stage memory B-cells (Klein et al., 2001).

Burkitt's lymphoma (BL), FL, DLBCL (GCB) and DLBCL (ABC) originate from GC derived cells. DLBCL (GCB) and DLBCL (ABC) were originally characterised as having differential gene expression profile (Alizadeh et al., 2000) but they can also be distinguished by their BCR isotype. DLBCL (ABC) expresses the primary immunoglobulin isotypes IgM or IgD whereas DLBCL (GCB) arises after CSR and expresses the secondary isotypes IgG or IgA (Ruminy et al., 2011). FL is morphologically and phenotypically similar to GC B-cells and can transform into GCB DLBCL.

B-cell prolymphocytic leukaemia and hairy-cell leukaemia originate from memory post GC B-cells and have undergone CSR Figure adapted from Küppers, 2005; Pieper et al., 2013.

1.1.2. Diffuse Large B-cell Lymphoma

DLBCL are typified by the diffuse destruction of the normal architecture of the lymph nodes or extranodal sites affected. It is particularly aggressive with a three-year survival rate of just 60% (Pfreundschuh et al. 2008). Aside from chromosomal translocations, DLBCL is particularly characterised by the inactivation of epigenetic modifiers including EP300, CREBBP and MLL2. These events only target one allele suggesting they are haploinsufficient unlike the classical “two-hit” tumour suppressor genes (reviewed in Basso and Dalla-Favera, 2015).

DLBCL has been further subdivided by expression profiling into two main groups, according to cell type of origin, germinal centre B-cell (GCB) DLBCL and the more aggressive activated B-cell (ABC) DLBCL (see Figure 1.2) (Jaffe 2009). Based on morphological similarities to the non-malignant cell type, GCB-DLBCLs were speculated to originate from the GC B-cells in light zone whereas ABC-DLBCL were thought to originate from a later stage of differentiation once B-cells were committed to plasmablastic differentiation (Alizadeh et al., 2000). More pertinent to the clinic, GCB-DLBCL is generally IgG positive whereas ABC-DLBCL is generally IgM positive (Ruminy et al., 2011).

GCB-DLBCL is known to be associated with chromosomal translocations of *MYC* (10%) and *BCL2* (40%) conferring poor prognosis (Aukema et al., 2011; Li et al., 2013). It is also known to have amplification of *c-REL* gene and of the *miR-17-92* microRNA cluster (Lenz et al., 2008).

ABC-DLBCL is characterised by translocation of *BCL6*, the constitutive activation of *NF-κB* with mutation of *TNFAIP3* and *MYD88* and a block on terminal differentiation to plasma cells by inactivation of *PRDMI* (reviewed in Basso and Dalla-Favera, 2015).

1.1.3. Follicular Lymphoma

FL is a nodal lymphoma with a follicular growth pattern. It is an indolent but incurable disease that can transform (40%) into the more aggressive DLBCL with a worse prognosis (Montoto and Fitzgibbon, 2011). Gene expression profiling of human tumour biopsies have shown that FL also has a GCB-cell-like expression signature indicating that it originates from GC B-cells making them the predominant cell type of origin of B-NHLs (Alizadeh et al., 2000).

FL is characterised by the elevated expression of *BCL2* and the inactivation of *MLL2* (reviewed by Basso and Dalla-Favera, 2015). *BCL2* activation is a prerequisite for FL however it is not sufficient to initiate disease being found also in healthy individuals. The IgH-BCL2 t(14;18) chromosomal translocation occurs early in B-cell development during primary V(D)J recombination in the bone marrow, mediated by RAG. *BCL2* increases resistance to apoptosis in the GC, facilitating the persistence of t(14;18)(q32;q21) translocated cells and therefore accumulation of other mutations.

MLL2 is a methyltransferase found to be mutated in at least 80% of FL. Its mutations also occur early in FL development and could cooperate with *BCL2* to facilitate progression (Morin et al., 2011). Chromatin modifiers are often mutated in FL including *EZH2* (27%) (Morin et al., 2010) and *CREBBP* (33%) (Kridel et al., 2012).

In recent years there has been an increased focus on sequencing the B-NHL exome identifying novel high frequency mutations (Morin et al., 2011a; Pasqualucci et al., 2011b; Lohr et al., 2012). *BCL2* and *CREBBP* have been found to be frequently commutated and have been defined as early, driving events in FL (Green et al., 2013; Morin et al., 2011).

1.2. B-cell lymphoma 2, *BCL2*

B-cell lymphoma 2 (*BCL2*) is a proto-oncogene within chromosome 18q21 that promotes B-cell survival via inhibition of apoptosis. The *BCL2* family includes both anti-apoptotic and pro-apoptotic members that form homodimers and heterodimers. *BCL2* heterodimerizes with, and limits the effects of death signals from the pro-apoptotic members. B-cells that fail to produce a functional BCR or fail to compete for presented antigen in the germinal centres will undergo apoptosis by a mechanism that is sensitive to *BCL2* levels (reviewed in Abramson and Shipp, 2005).

The *BCL2*-*IGH* t(14;18)(q32;q21) translocation is found in 90% of FL and 35% of DLBCL (most commonly GCB) (Küppers, 2005). This translocation places *BCL2* under the control of the *IGH* promoter sequence causing overexpression of *BCL2*, thus inhibiting apoptosis and promoting cell survival. Though this translocation is commonly an initiating event in development of FL it is not sufficient to initiate tumourigenesis and similar translocations can be detected in the peripheral blood and lymph nodes of healthy individuals (Limpens et al., 1995).

Mouse models have shown that the balance of pro- and anti-apoptotic family members is crucial, determining tissue homeostasis and malignancy in the hematopoietic compartment. Overexpression of the anti-apoptotic *BCL2* leads to development of follicular hyperplasia whereas mice lacking the pro-apoptotic *BCL2* family member *BAD* develop a DLBCL like disease (Ranger et al., 2003). Overexpression of *BCL2* from the pan-haematopoietic promoter *Vav-P* predisposes mice to late onset follicular lymphoma with approximately half of animals succumbing to disease by 18 months. These animals also have gross alteration of their lymphoid cells indicative of a *BCL2* specific role in the haematopoietic compartment (Egle et al., 2004).

There has been considerable effort towards drugging *BCL2*. Early attempts aimed to interfere with *BCL2* expression and thus reduce *BCL2* protein levels. Oblimersen (Genasense, Genta, NJ) is an antisense

oligonucleotide that binds to *BCL2* mRNA interfering with translation. This drug was able to induce apoptosis *in vitro* however its clinical efficacy proved limited in Phase III trials (O'Brien et al., 2009).

Subsequently researchers attempted to directly inhibit the function of the BCL2 protein. BCL2's pro-survival signals counteract BAX and BAK pro-death stimuli by directly binding them or indirectly by sequestering their activators like the BH3 domain protein BIM causing cell death (Czabotar et al., 2014).

GX-15-70 (Obatoclax) is a pan-BCL2 family inhibitor, binding BCL2, BCL_{XL}, BCL_W and MCL1, killing cells by preventing these proteins from inhibiting the death signals of BAX and BAK (Nguyen et al., 2007). After low response rates and neurological side effect in clinical trials, its development has been ceased (Trudel et al., 2007).

The most promising BH3 mimetic is currently ABT-737 which was found to kill cells in a BAX/BAK dependent manner (van Delft et al., 2006). Due to limited oral bioavailability the sister molecule ABT-263 (Navitoclax) was developed and it has shown promise in the treatment of lymphoid malignancies (Roberts et al., 2012). In a Phase I trial of ABT-263 in patients with NHLs an overall response rate of 22% was observed however dose escalation trials in CLL showed it to be thrombotic (Wilson et al., 2010) because of its strong binding to BCL_{XL} (Strasser et al., 1994).

Recently ABT-263 was re-engineered to ABT-199 which is potent, orally bioavailable, BCL2 selective and has been shown to spare platelets. Preliminary data from a NHL Phase I trial indicates an overall survival rate of 53% (Souers et al., 2013).

Although these drugs have not yet been approved for DLBCL or FL these trials indicate BCL2 is an appropriate target for therapy in B-cell lymphoproliferative disorders.

1.3. CREB-binding protein, CREBBP

CREB binding protein (CREBBP) and E1A binding protein p300 (EP300) are the only members of the lysine acetyltransferase 3 family (LAT3). They are highly conserved and ubiquitously expressed transcriptional co-activators. They interact physically or functionally with over four hundred different proteins, via binding sites unique to the LAT3 family, making them one of the most heavily connected nodes in the mammalian protein interactome (Bedford et al., 2010). CREBBP acts as a molecular scaffold facilitating the coupling of multiple transcription factors to the core transcriptional machinery activating transcription (Nakajima et al., 1997).

As a protein acetyltransferase CREBBP is able to acetylate lysine residues on histone and non-histone proteins. By acetylating the N-terminus of histone tails CREBBP, in association with SRC-1 (Yao et al., 1996) and pCAF (Yang et al., 1996) remodels chromatin, thereby activating transcription (Rando and Chang, 2009). CREBBP also acetylates non-histone proteins like the transcription factor p53 (Gu and Roeder, 1997) and the haematological differentiation factor GATA1 (Blobel et al., 1998) activating them whereas acetylation of the anti-apoptotic BCL6 inactivates it (Green et al., 2013). CREBBP acetylation of BCL6 prevents it from repressing transcription of anti-apoptotic proteins in a dose dependent manner thus increasing apoptosis. Conversely p53 needs to be acetylated to potentiate its transcriptional activity (Pasqualucci et al., 2011b).

This balance between the proto-oncogene BCL6 and the tumour suppressor gene p53 is critical in regulating DNA damage response in the GC cells during immunoglobulin remodelling (Phan and Dalla-Favera, 2004; Phan et al., 2005). ChIP assays have shown that *in vivo* BCL6 directly binds the *p53* promoter thus repressing *p53* and its target genes (Phan and Dalla-Favera, 2004). This group also demonstrated that enforced BCL6 expression suppresses the p53-mediated apoptotic response to genotoxic stress in B-cells. They engineered a B-lymphoma cell line to express a mutant form of BCL6 that was resistant to etoposide-induced ubiquitin-mediated degradation and found that, upon treatment with etoposide, wild type BCL6 expression was downregulated unlike the mutant BCL6 and that the cells were therefore more resistant to apoptosis, as demonstrated by 7-aminoactinomycin D staining.

CREBBP has also been shown to mediate p53 response to DNA damaged by ionizing irradiation. Ionization promotes phosphorylation of p53 which in turn increases its affinity for CREBBP thus increasing p53's acetylation and consequent activation (Lambert et al., 1998). Loss of CREBBP's attenuation of BCL6 and its activation of p53 could affect cellular response to other types of DNA damage such as aberrant SHM and CSR. This could increase a cell's tolerance to DNA damage in the context of diminished apoptosis and cell cycle arrest leaving the GC cells particularly vulnerable to oncogenic transformation (Pasqualucci et al., 2011b).

The central position of CREBBP and EP300 in the interactome is consistent with their requirement for normal development and nullizygous mutation of *Crebbp* or *Ep300* are embryonic lethal in mice by day 10.5. Their heterozygous loss has been found to cause the congenital developmental disorder Rubinstein-Taybi syndrome (RTS) in humans (reviewed in Bedford and Brindle, 2012). RTS is characterised by growth impairments, mental retardation, distinctive facial and skeletal abnormalities and an increased susceptibility to cancer (Cantani and Gagliesi, 1998). *CREBBP* is lost in up to half of RTS cases (Bartsch et al., 2005) and aberrations range from large, hundred kilobase deletions to deletions of single exons, translocations or inversions. The majority of the aberrations are point mutations, small deletions or duplications that may cause premature stops codons, amino acid substitutions or splicing defects. The exact molecular mechanisms of the syndrome is not known but the heterozygous mutations found in *CREBBP* indicate that deficiency of CREBBP protein function has a role (Bartholdi et al., 2007).

Kung and colleagues modelled the global heterozygous loss of *Crebbp in vivo* and found it to recapitulate human RTS with mice exhibiting developmental abnormalities and increased susceptibility to tumours (Kung et al., 2000). Aged *Crebbp*^{+/-} mice were found to have severe splenomegaly by 12-18 months with spleen mass increasing approximately 8-fold. These mice had increased numbers of myeloid and erythroid cells when compared to wild type mice without splenomegaly. This was corroborated by the increase in extramedullary haematopoiesis highlighting abnormal haematopoiesis. The *Crebbp*^{+/-} mice that developed tumours also showed propensity for loss of heterozygosity, losing the remaining intact *Crebbp* allele and retaining the mutant allele, suggesting for the first time that *Crebbp* was a tumour suppressor gene. This is consistent with the observation that *Crebbp* is targeted by viral oncogenes like adenovirus E1A and SV40 large T antigen to realise the full oncogenic potential of the virus (Kung et al., 2000).

CREBBP is rarely mutated in solid tumours but has recently been found to be mutated in B-lymphocyte pathologies suggestive of a specific role for LATs in B-NHL. Human sequencing studies have identified *CREBBP* loss in a variety of B-NHL (Kridel et al., 2012; Lohr et al., 2012; Morin et al., 2011; Pasqualucci et al., 2011b). Pasqualucci and colleagues found that 29% of patients with DLBCL had mutations in *CREBBP*. Of those cases 41.5% were GCB-DLBCL and 17% were ABC-DLBCL.

CREBBP mutations are an early event in FL which is maintained during tumour progression indicating that their loss is driving tumourigenesis and has a causal role (Green et al., 2013). Of the *CREBBP* mutations found in FL 32.6% were analogous to those in DLBCL suggesting that they share a common *CREBBP* dependent process in their pathogenesis (Pasqualucci et al., 2011b). Mutations in *CREBBP* have been found throughout the open read frame suggesting they cause loss of function however they were heterozygous indicating that loss of *CREBBP* promoted lymphoma in an haploinsufficient manner

(Chung et al., 2012). The majority of the mutations found in *CREBBP* caused the elimination or truncation of the acetyltransferase domain (Morin et al., 2011; Pasqualucci et al., 2011a) that reduced acetylation activity. In the 134 human DLCBLs sequenced by Pasqualucci, half of the mutations were predicted to cause the elimination or truncation of the LAT domain. These inactivating mutations included nonsense mutations, frameshift insertions and deletions and mutations at the splice donor or acceptor site which generate premature stop codons. The remaining mutations, inframe deletions and missense mutations, were also primarily in the LAT domain suggesting that it is specifically the loss of acetyltransferase activity that is functionally important to lymphomagenesis (Pasqualucci et al., 2011b).

Sequencing data from human studies have shown, for example, that *CREBBP* mutations frequently co-occur ($p < 0.05$) with mutations of the apoptotic regulator *BCL2* whereas they rarely occur ($0.1 < p < 0.3$) with mutations of the anti-proliferative gene *BTG1* (Morin et al. 2011). It is this kind of specific genetic interaction that can provide information regarding underlying mechanisms in oncogenic pathways.

1.4. Insertional Mutagenesis

Insertional mutagenesis (IM) is a process whereby the genome is mutated by the insertion of retroviruses or transposons. IM has been used as a research tool to perform forward genetic screens in cell lines, in the germline of a variety of model organisms and in the somatic cells of mice. One of the strengths of IM screens is the ease with which insertion mutations can be identified and mapped to the genome. The insertion sites can be detected by amplifying the regions of the genome that flank the insertion site using PCR based amplification methods (reviewed by Uren et al., 2005).

Somatic insertional mutagenesis screens in mice have primarily been applied to the study of cancer. Mouse tissues can be mutagenized through the use of replicative retroviruses or germline copies of transposons mobilized in different tissues. A subset of the mutations, created by these mutagens, give rise to a malignant transformation. Genes and loci which are repeatedly found to have insertion sites in independent tumours, more frequently than expected by chance, are called common insertion sites (CISs) and are considered to have a role in tumourigenesis (reviewed in Kool and Berns, 2009).

Oncogenic retroviruses are divided into two groups: acute and slow transforming viruses. Acute transforming retroviruses induce polyclonal tumours 2-3 weeks after infection via the expression of viral oncogenes such as v-Abl in Abelson Murine leukaemia virus (Shore et al., 2002) or v-Myb in avian myoblastosis virus (reviewed by Lipsick and Wang, 1999). Slow transforming retroviruses induce mono- or oligoclonal tumours with a longer latency of 3-12 months as a result of mutations caused by integration of the proviral retrovirus into the host genome. Examples of oncogenic slow transforming retroviruses are feline leukaemia virus (FLV), avian leukaemia virus (ALV), mouse mammary tumour virus (MMTV) and Moloney murine leukaemia virus (MoMuLV) which was used in the experiments in this thesis.

IM can also be mediated via the recently adapted sleeping beauty and piggybac transposons that mutate DNA as they hop autonomously around the genome. Though this technique offers greater capacity to target different tissues and a different spectrum of mutations, remobilization of the transposon during tumour development can cause loss of the initiating mutations and there is a higher level of background unselected non-CIS mutations (Jaap Kool & Anton Berns 2009).

In recent years the success of monoclonal antibodies, small molecules, peptide mimetic and antisense oligonucleotides that modulate specific signalling pathways or mutated genes that cause cancer has encouraged researchers to identify and comprehensively map all pathways and genes involved in cancer, even those infrequently mutated in human cancer, with the view to identifying novel targets amenable to therapy (Copeland and Jenkins, 2010). IM screens lend themselves to this unbiased forward screening of genes with a putative role in cancer. They are also complementary to human genome sequencing and copy number analysis of tumours as they have the potential to uncover genes that are rarely mutated but

more commonly subject to recurrent selection for deregulation including epigenetic modification (Albihn et al., 2010). Another strength is that by using transgenic mice already with mutations sensitising them to disease it is possible to build profiles of co-occurring or mutually exclusive mutations which highlight different approaches to targeting crucial genes (van Lohuizen et al., 1991)

Somatic retroviral mutations interact with predisposing germline or conditional mutations, facilitating tumourigenesis; it is the combination of these mutations that directs tumour presentation and course (Gabriel, 2006) recapitulating the stepwise and accumulative progression of human disease. Functional interactions between mutations can be investigated by using transgenic mouse models whereby somatic virus caused mutations cooperate with or are mutually exclusive to sensitising germline mutations (van Lohuizen et al., 1991; Stewart et al., 1993) (see Figure 1.3.C). By using mice with a sensitised background it is possible to identify mutations that co-operate with the sensitising mutations or those that are mutually exclusive, allowing the definition of tumour mutation profiles.

Genes and loci which are repeatedly found to have proviral insertion sites in independent tumours are called CISs and are considered to have a role in tumourigenesis (reviewed in Kool and Berns, 2009). The majority of the tumours are oligoclonal or polyclonal with some sharing one or more initiating, oncogenic insertions but also with additional discrete insertion suggesting that consecutive insertions occur at different stages of clonal expansion. Analysis of the relative clonality of insertions allows tracing of the lineage of tumours and defines the order in which lesions occurred (Huser et al., 2014).

1.4.1. Moloney Murine Leukaemia Virus

Moloney murine leukaemia virus (MoMuLV) is a slow-transforming retrovirus, originally isolated from a BALB/c mouse sarcoma, which upon injection into new-born mice causes polyclonal leukaemias with a median latency of four months after inoculation and with 100% incidence by nine months (Moloney, 1960). It has a linear single stranded, RNA genome consisting of 5'LTRs-gag-pol-env-3'LTR (see Figure 1.3.A) which upon integration is reversed transcribed by the *pol* protein creating a DNA provirus which can insert into the host genome.

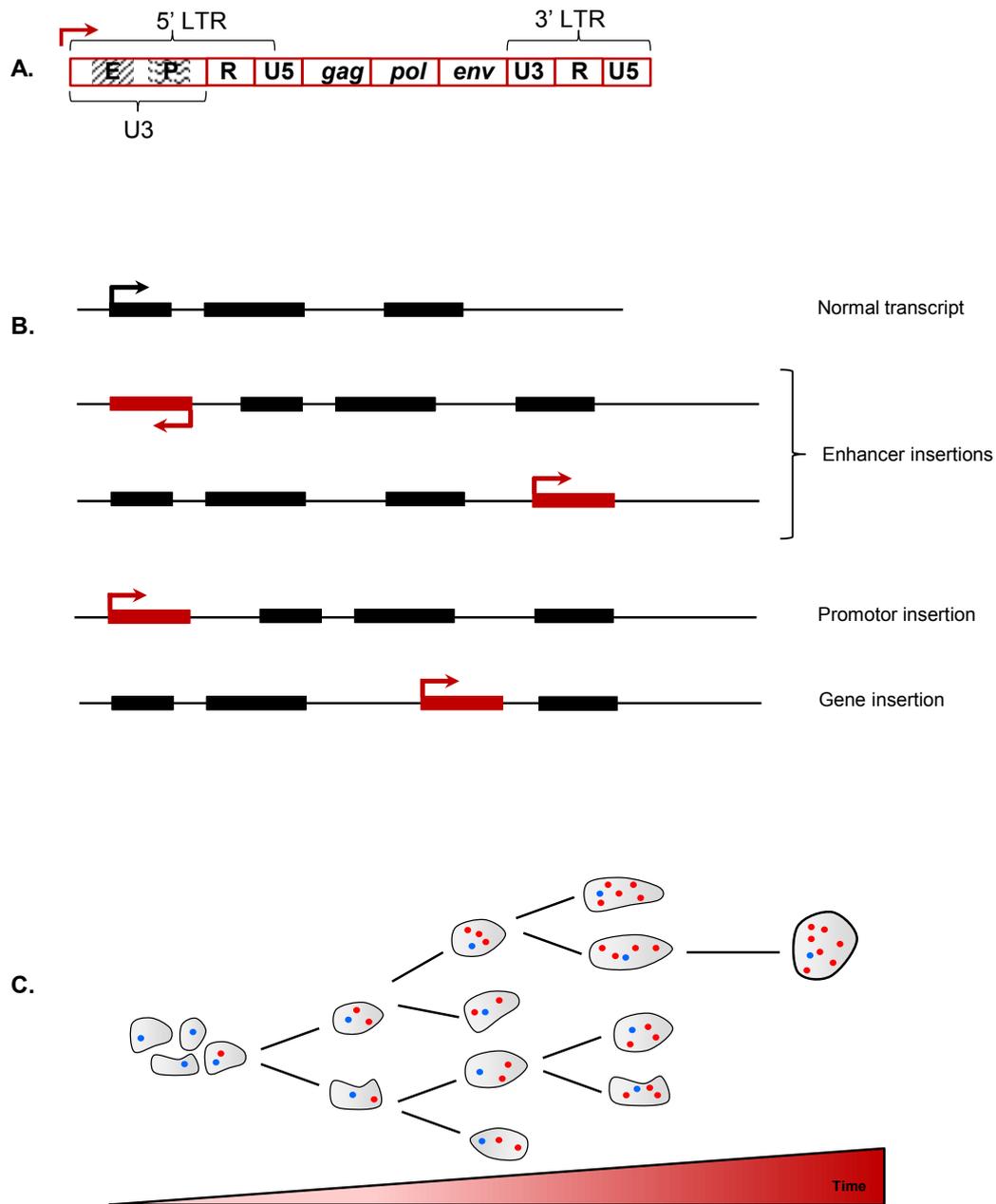


Figure 1.3. Retroviral insertional mutagenesis

A. The retrovirus genome is composed of two LTRs flanking the viral assembly genes: gag, pol and env. The LTRs are subdivided into three regions: U3, R and U5. U3 contains the enhancer and promoter portions. The enhancer contains binding sites for specific transcription factors that confer tissue specificity and the promoter is involved in recruitment of the basal transcription machinery. R contains the start and stop motifs for transcription. Once integrated the pol protein reverse transcribes the virus to produce a DNA provirus capable of inserting into the host genome.

B. Retroviruses (red) are able to induce mutations in multiple ways by integrating into the host genome, depending of the site of provirus integration, within or adjacent to a gene, the mutations can interfere with gene transcription and truncate or deregulate oncogenes or tumour suppressor genes. Insertion of the MoMuLV provirus upstream of a gene in the antisense orientation or downstream of a gene in the sense orientation cause enhancer mutations. Mutations upstream of a gene in the sense orientation often effect the gene promoter. Insertions within a gene can cause truncation of a protein causing it to have aberrant functions or to be degraded.

C. Retrovirus caused, somatic mutations (red) can interact with sensitising germline mutations (blue) which over time leads to a selective advantage, clonal expansion and tumour development mimicking the stepwise progression of human cancer.

MoMuLV is ecotropic i.e. only able to infect mouse cells where the cell surface marker (mCAT1) specific to its envelope proteins are expressed. Once the virus has integrated, the host cell will start to produce viral envelope proteins occupying the surface receptors and minimizing further infection. MoMuLV is able to circumvent this by recombining with endogenous viral sequences, thereby producing mutant envelope proteins specific to different receptors and by having an unstable envelope sequence which frequently rearranges. This permits reinfection of a cell and the accumulation of mutations (reviewed in Uren et al., 2005).

When newborn mice are infected with slow transforming retroviruses like MoMuLV, via an intra-peritoneal injection, they do not mount an immune response and consequently develop a viremia that lasts the lifetime of the animal. Successive rounds of infection and reinfection lead to an accumulation of insertion mutations in infected cells, especially in the hematopoietic compartment where the high rate of proliferation during early postnatal development makes them the preferred host cells for virus propagation (reviewed in Kool and Berns, 2009). Depending on the site of provirus integration, within or adjacent to a gene, the mutations can interfere with gene transcription and truncate or deregulate oncogenes or tumour suppressor genes. Some proviral mutations will be advantageous to the cell providing a selective advantage and over time, mutations will accumulate leading to the clonal expansion of cells each with multiple oncogenic mutations (see Figure 1.3.B).

MoMuLV predominantly causes T-cell pathologies however, this can be altered in a variety of ways. Different strains of mice have different susceptibilities with BXH-2 mice being predisposed to myeloid leukaemia whereas AKxD mice are prone to B-cell lymphoma. Transgenic mice have also been used to promote B-cell lymphoma in MoMuLV screens by recapitulating genetic aberrations common in human disease. Overexpression of the oncogene Myc in E μ -Myc transgenic mice results in 90% pre-B-cell lymphoma with mean latency of seven weeks (van Lohuizen et al., 1991).

Sequences within the virus LTRs are responsible for cell type-specific expression (Fan, 1997). Each LTR is subdivided into U3, R (containing transcription start and termination sites) and U5 portions (see Figure 1.3.B). The U3 contains enhancer and promoter regions that confer the cell specificity of viral expression. The enhancer consists of two 75bp tandem repeats and contains specific transcription factor binding sites limiting efficient transcription to cell types where those factors are present. Previous studies have shown that enhancers can be modified to effect the expression specificity and infection profile. Starkey and colleagues substituted the MoMuLV enhancer with the equivalent region from feline leukaemia virus inducing a different spectrum of integration insertions (Starkey et al., 1998). Hanecak et al. replaced the MoMuLV enhancer with that of the simian virus 40 (SV40) making it more B-cell specific and increasing latency to 17 months (Hanecak et al., 1988).

The majority of mutations in RV IM insertions are of enhancers causing increased expression of cellular genes. Mutations in enhancers tend to be upstream and antisense of the gene they modulate or downstream and sense placing the viral promoter in control of the candidate gene enhancer (Figure 1.3.B). MoMuLV is also known to have some integration site bias including preferentially integrating at transcription start sites (Wu et al., 2003) and unwound chromatin (Babaei et al., 2015).

In MoMuLV screens *c-Myc* is primarily overexpressed by enhancer mutations (Cuypers et al., 1984) and there are examples of these mutations being up to 270kb downstream of the *cMyc* promoter (Lazo et al., 1990).

Proviral insertions in the 3'UTR-encoding regions of a gene can remove regulatory or destabilising motifs in the gene mRNA by halting transcription. This can lead to an accumulation of protein-coding mRNA and therefore increase protein levels. For example, insertions in the last exon of *Pim-1* remove the motif responsible for reducing its mRNA stability allowing it to persist and ultimately increasing expression levels (Selten et al., 1985). *Pim1* strongly cooperates with *c-Myc* to promote lymphomagenesis and so these mutations frequently co-occur (van Lohuizen et al., 1991).

Provirus insertions within the promoter of a gene can produce chimeric transcript where the gene is under the control of the viral promoter causing activation and overexpression (Swain and Coffin, 1993). Proviral insertions in the sense orientation upstream or within the first noncoding exons of *Evi5* result in a viral-cellular fusion transcript leading to overexpression of *Evi5* which has been attributed to reduced terminal differentiation in haematopoietic cells and lymphoma (reviewed in Uren et al., 2005).

Insertions within tumour suppressor genes like *p53* can cause premature transcript termination leading to truncated, inactive or unstable mRNAs and reducing protein expression (Mowat et al., 1985).

Notch1 is frequently mutated in MoMuLV screens where proviral insertions cause truncation of *Notch1* to produce a constitutively active protein which has oncogenic properties and is able to activate transcription (Radtke et al., 2004).

It can be difficult to correlate insertion sites to change in gene expression. In one study of MoMuLV induced lymphoma, all the genes within 100kb of their 20 insertions sites of interest were quantified by Q-RT-PCR (Sauvageau et al., 2008). Only approximately half of the insertions had a measurable effect on gene expression by QPCR. This lack of correlation could be because the insert was only present in a minor proportion of an oligoclonal tumour DNA analysed; the deregulation required for oncogenesis might be too subtle to reliably measure; the effects might be minimal at the mRNA level but potent at the protein level (due to truncating mutations or deregulated splicing or translation); or the insertion might

alter multiple genes with varying amounts of effect in a manner similar to single nucleotide polymorphisms in a genome wide association studies (Sauvageau et al., 2008).

Previous IM screens have successfully identified novel candidate cancer genes, identified genetic interaction between genes driving tumour evolution and identified genes that confer resistance to anticancer drugs or pathogens (reviewed in Uren and Berns, 2009). The following studies identified novel candidate genes using IM genes that were subsequently found to have a role in human disease: Cuypers et al., 1984, identified the serine-threonine kinase *Pim-1* in MoMuLV induced lymphoma, van Lohuizen et al., 1991 found the transcriptional repressor *Bmi-1* in MoMuLV induced B-cell lymphoma.

IM screens have also been used to pinpoint genetic interactions between genes driving tumour evolution for example Uren et al., 2008 found *Runx1* to be preferentially mutated in $p19^{Arf^{-/-}}$ and $p53^{-/-}$ mice when compared to wild type, suggesting that mutation of *Runx1* is selected for in cells that lack the $p19^{ARF}$ - $p53$ tumour suppression pathway. In another screen *Ccnd3* mutations were found to be mutually exclusive with loss of the cyclin dependant kinase *Cdkn2A*, indicating that *Ccnd3* mutation is redundant when the inhibitor of *Ccnd3*/*Cdk4* complexes has already been lost (Kool et al., 2010). Lauchle et al., 2009 found a MEK inhibitor caused regression in *Nf1*-deficient tumours but then resistance arose in a subset of these tumours that also had *Rasgrp1* overexpression. This highlights a strength of co-occurrence data when the co-mutation of two genes makes a drug ineffective. IM screens have also been used to identify regulators of tumour suppressor genes like *Cdkn1b* which encodes *p27*. *Cdkn1b*^{+/-} mice have decreased *p27* expression but some mice in the screen had none which was found to co-occur with mutations in *Id3* suggesting *Id3* was an upstream regulator of *Cdkn1b* (reviewed in Kool and Berns, 2009).

Though recent studies have identified key aberrations and pathways in the various types of B-NHL a comprehensive review published in 2015 by Basso and Dalla-Favera concluded that this still only represents a small portion of the large number (approximately 100) of mutations found in most human tumours (Basso and Dalla-Favera, 2015). Though obviously not all of these mutations are driving oncogenesis there is need for other methods such as functional genomic screens, mouse models, IM screens and comparative genomics to scrutinise driver mutations and better characterise cancer pathways.

1.4.2. Integration site cloning technological advancements

Though IM screens have been performed for decades their feasibility and throughput has increased greatly with recent technological advances. Initial methods for integration site cloning were very labour intensive and low throughput, they involved making genomic libraries of each tumour and then screening them by colony lifting using the mutagen as a probe (Cuypers et al., 1984). This was superseded by PCR based methods that enrich for insertion sites by amplifying the region of genomic DNA adjacent to the insertion mutation. A variety of PCR methods were published but ligation mediated-PCR methods such as splinkerette-PCR became the most commonly used (Devon et al., 1995). The key advancement of splinkerette-PCR was the use of a non-complimentary hairpin shaped adaptor that facilitated enrichment of integrated sequences.

In a study of MoMuLV induced lymphoma, combining restriction enzyme fragmentation, ligation mediated PCR with shotgun subcloning and capillary sequencing, 46,197 mappable sequences were retrieved, representing more than 10,000 viral insertion sites from a cohort of 510 MoMuLV tumours (Uren et al., 2008). This approach of insertion site cloning was again improved by the use of shearing by sonication to fragment the tumour DNA into controlled length, amplifiable sequences (Koudijs et al., 2011). This minimised sequence and amplification bias caused by the fixed position of restriction enzymes recognition sites relative to virus integrations. The method also allowed the clonality of inserts to be quantified as each adaptor ligation point was now unique and the number of different sheared fragment lengths, in the initial ligation, could be used as a read out for the relative abundance of each mutation. This study also took advantage of 454 pyrosequencing which greatly increased throughput of insert identification. From the 16 MMTV tumours they retrieved 5134 mappable reads (Koudijs et al., 2011).

These improvements were consolidated when ligation mediated PCR and 454 sequencing was applied to 28 MoMuLV tumours that had already been screened using shotgun cloning and capillary sequencing. The original paper only yielded 272 insertion sites (Stewart et al., 2007) whereas Huser and colleagues obtained 12,485 unique retroviral insertion sites (Huser et al., 2014), highlighting the massive improvement in yield of IM in less than a decade. With the advent of higher-throughput next-generation sequencing the frontiers of insertional mutagenesis can be expanded even further.

1.5. Murine Models of NHL

Typically mice have been used to model human cancers with many genes (Waterston et al., 2002) and cancer pathways (Frese and Tuveson, 2007) being common to both organisms. Mouse cells however are easier to transform than human cells suggesting that they are less tumour resistant (Rangarajan and Weinberg, 2003). One of the aims of this work was to refine and advance mouse models of NHL using combinations of mutations found in recent human B-NHL studies which can be done using transgenic animal to combine multiple sensitising mutations. The following alleles were used in this study.

1.5.1. VavP-BCL2 transgenic

The *VavP-BCL2* transgenic mouse was made by Ogilvy and colleagues by placing human *BCL2* cDNA under the control of the pan-haematopoietic promoter Vav-P. They aimed to study the role of *BCL2* in the development and homeostasis of the haematopoietic compartment where *BCL2* was known to reduce apoptosis and retard cell cycle entry (Ogilvy et al., 1999a). The *vav* promoter was found to be active throughout the haematopoietic compartment and occasionally in organs such as the lung, liver and small intestine which are known to contain haematopoietic cells. Levels of *vav* expression were found to be high throughout lymphoid development from primitive pre-progenitors until lineage-committed progenitors (Ogilvy et al., 1999b).

Transgenic *VavP-BCL2* was highly expressed in all nucleated cells of the haematopoietic compartment including B- and T-lymphocytes, granulocytes, monocytes, megakaryocytes, eosinophils and nucleated erythroid cells. The resistance of the haematopoietic cells with transgenic *BCL2* to apoptosis leads to their persistence and accumulation within the compartment. When characterising the peripheral blood of these mice they found fifteen times the total lymphocytes compared to non-transgenic controls; and the total number of cells in the spleen increased six-fold causing the spleen to increase in weight from $80\pm 12\text{mg}$ to $267\pm 48\text{mg}$ in the *VavP-BCL2* transgenics.

The transgenic spleens also contained more lymphoid follicles, with enlarged germinal centre, which incidentally contained less apoptotic cells. The thymi of these mice were morphologically normal but has altered levels of T-lymphocytes with a significant decrease in $\text{CD4}^+\text{CD8a}^+$ double positive cells from 84% in the wild-types to 44% in the transgenics, compensated overall by an increase in double negative and single positives (Ogilvy et al., 1999a). This expansion of the T-cell compartment was also observed in another study of *VavP-BCL2* transgenic mice by Egle and colleagues. The authors postulated that CD4^+ T-cells were required for the development of enlarged GCs and found that when they deleted them, using antibody targeting, that the GC hyperplasia was lost (Egle et al., 2004).

These mice were highly predisposed to late onset FL with the cumulative incidence, of their three experimental strains, ranging between 37% and 50% at 18 months. These mice presented at necropsy with splenomegaly and enlarged lymph nodes, particularly the mesenteric lymph nodes. Even in young, premalignant animal, the germinal centres of the *VavP-BCL2* mice were enlarged and more frequent than in controls (Egle et al., 2004).

Immunohistochemical analysis of the GC spleens cells (PNA+) indicated that they had undergone immunoglobulin class switching as they were generally IgM negative but IgG positive (7/8 FL samples). Of the B220+ B-cells 30-50% were IgM- IgD- double negative and the majority of these expressed the Kappa light chain showing that they were mature cell that had undergone class switching. These IgM- IgD- κ + B-cells were significantly increased (50-fold) in the *VavP-BCL2* mice relative to wild type mice indicating that the GC and post GC B-cells were greatly expanded; their IgM+ IgD+ T2B-cells were also increased approximately 4.5-fold (Egle et al., 2004).

Together this indicates that overexpression of BCL2 in the haematopoietic compartment predisposes mice to mature B-cell neoplasms resembling follicular lymphoma.

1.5.2. *Crebbp* flox allele

Nullizygous loss of *Crebbp* in mice is embryonic lethal by day 10.5, whereas monoallelic inactivation of the *Crebbp* gene cause mice to develop highly penetrant, multi-lineage defects in hematopoietic differentiation which with advancing age includes an increased incidence of hematologic malignancies (Kang-Decker et al., 2004).

Zhang and colleague produced a conditional *Crebbp* loss allele by flanking exon 7 of *Crebbp* with loxP sites, *Crebbp^{fl}* which, upon recombination with a *Cre* driver causes the truncation and non-sense mediated decay of the transcript (Zhang et al., 2004). To test the conditional inactivation of *Crebbp^{fl}* the authors, who were interested in studying the role of *Crebbp* during embryonic development and adult physiology, bred *Crebbp^{fl/fl}* mice with a Ca^{2+} /calmodulin-dependent kinase II Cre recombinase mouse line and confirmed, by in situ hybridization, that loss of *Crebbp* was confined to the desired compartments of the brain.

Xu *et al.* studied the effects of *Crebbp* loss in B-cell by using the Cd19^{Cre} driver which is expressed throughout B-cell development from the pro-B stage (Xu, 2006). Southern blot analysis confirmed the successful loss of *Crebbp* in the *Crebbp^{fl/fl};CD19^{+/Cre}* mice in splenic B-cell DNA and western blots confirmed loss of the *Crebbp* protein. The western blot indicated that there was also a reduction of *Crebbp* in the *Crebbp^{fl/fl};CD19^{+/+}* relative to the *Crebbp^{+/+};CD19^{+/Cre}* suggesting perhaps that the presence of the intronic *loxP* sites alters expression and/or splicing of the transcript. These mice were generally

healthy but had increased incidence of death after one year when compared to the wild types, the causes of death were undetermined but were not lymphoma.

The *Crebbp*^{fl/fl};*CD19*^{+Cre} mice had a 50% decrease in the number of B-cells in the peripheral blood, spleen and lymph nodes when compared to their controls, *Crebbp*^{fl/fl} and *CD19*^{+Cre}. Flow cytometry analysis of splenic B-cells found reductions in the number of IgM and IgD positive B-cell classes; IgM⁺IgD⁻ which includes mature marginal zone B-cells and T1 immature B-cell; IgM⁺ IgD⁺ which comprise T2 immature B-cells and proliferating B-cells; and IgM⁻IgD⁺ which are follicular mature B-cells. This suggests that loss of *Crebbp* results in increase immunoglobulin class switching to IgM- IgD- IgG+/IgA+/IgE+.

Using microarrays Xu and colleagues tested the global gene expression of the *Crebbp*^{fl/fl};*CD19*^{+Cre} mice but found that global gene expression was largely unperturbed. However there were some genes with 2-3-fold changes in expression including downregulation of B-cell specific transcriptional regulators like *C2ta*, *Ets1* and *Relb*, and upregulation of transcriptional cofactors like *Mcm2* and *Smarca3* suggestive of a compensatory mechanism for the loss of *Crebbp*.

BCR-responsive gene expression is thought to involve transcription factors that interact with *Crebbp*. By stimulating splenic B220⁺ B-cells with anti-IgM, Xu *et al.* were able to measure BCR-dependent gene expression in the B-cells by RT-QPCR. They measured the expression of five BCR-inducible genes, with a known role in B-cell biology, and that are thought to be *Crebbp* targets: tumour necrosis factor alpha (*TNF-α*), c-Myc, Bcl2-related protein A1a, cyclin D2 (*Ccnd2*), and cyclin-dependent kinase 4 (*Cdk4*). *TNF-α* expression was reduced by approximately 68% in the *Crebbp* null B220⁺ B-cells and similar reductions were also seen in follicular B-cells (B220⁺ CD21^{med} CD23^{hi}) and marginal zone splenic B-cells (B220⁺ CD21^{hi} CD23^{lo}). *Myc* BCR-dependent expression was reduced approximately 30% in the *Crebbp* null mice. *Myc* expression is highly important in B-cells proliferation and apoptosis and even small modulations can be detrimental (de Alboran *et al.*, 2001). *Ccnd2* and A1 are NK-κB targets and both were reduced in the *Crebbp* nulls by 34% and 84% respectively. *Cdk4* expression was unaffected by *Crebbp* loss (Xu, 2006).

Due to the pleiotropic effects of *Crebbp* heterozygosity in the hematopoietic compartment, we chose to model its loss using the more restricted germinal centre B-cell specific *Aicda*^{Cre} driver rather than *CD19*^{+Cre}.

1.5.3. *Aicda*^{Cre} knockin

AICDA is under tight transcriptional regulation and is only expressed during class switch recombination and somatic hypermutation in the germinal centre B-cells. Previously *Aicda* has been shown to be expressed highly but transiently in B-cells during GC reactions and low but biologically active levels have been shown in developing B-cells (Gourzi et al., 2006; Han et al., 2007; Mao et al., 2004). *Aicda* expression has also been reported in epithelial cell and pluripotent tissue however these experiments were performed *in vitro* or with tumour tissue, so expression in non-lymphoid tissues under physiological conditions was unclear (Okazaki et al., 2007).

Robbiani and colleagues produced an *Aicda*^{Cre} knockin allele by targeting exon 1 of endogenous *Aicda* with *Cre recombinase*, putting it under the control of the endogenous *Aicda* promoter (Robbiani et al., 2008). By stimulated resting B-cells with LPS and IL4 the authors were able to induce *Aicda* expression and *Cre-recombinase* activity in antigen-experienced B-lymphocytes and Peyer's patches.

Using fate mapping, where a reporter gene is conditionally and irreversibly expressed after the expression of Cre, Rommel *et al.* demonstrated that *Aicda* is also expressed in post GC plasma cells in the bone marrow and spleen, in a fraction of T-cells in old mice and in a small fraction of non-lymphoid embryonic cells (Rommel et al., 2013).

1.6. Research Aims and Approach

The research was guided by the pretext that retroviral insertion mutations perturb the function or expression level of their proximal candidate gene contributing to lymphomagenesis in collaboration with the sensitising mutations: overexpression of *BCL2* and/or downregulation of *Crebbp*.

This led to two specific research aims:

- The first aim of the project was to experimentally demonstrate *in vivo* that loss of one copy of *Crebbp* can accelerate *BCL2* driven lymphomagenesis and characterise these tumours.
- The second aim was to identify genes mutated in *BCL2* driven lymphoma that are more or less frequently mutated in the context of *Crebbp* hemizyosity.

The approach adopted in the research was to generate a colony of mice sensitised to B-cell lymphoma by overexpressing *BCL2* and/or hemizygous loss of *Crebbp* in the haematopoietic compartment. These animals were then aged and the resultant spontaneous lymphomas were characterised by flow cytometry.

Then, insertional mutagenesis screening was performed in a second cohort of B-cell lymphoma sensitised mice with the view to identifying novel genes that promote or protect against lymphoma in the context of *BCL2* overexpression and *Crebbp* deficiency and assess cooperating or mutually exclusive relationships between them.

To support and enable this, an innovative high-throughput DNA library preparation pipeline was designed that is applicable to an Illumina sequencing platform and facilitates the deep-sequencing and quantitative measurement of retroviral insertions.

2. Materials and Methods

Described in this chapter are the main experimental methods that pertain to this thesis. Subsection 2.7 details our novel insertion site cloning method.

2.1. Animal Models

The three transgenic mice used in this study were:

- i. *Crebbp* flox mouse, *Crebbp*^{fl}, (*Crebbp*^{tm1.2Ltz}, MGI:3056564) was kindly provided by Dr B. Lutz at Johannes Gutenberg-University, Mainz (Zhang et al., 2004).
- ii. *Aicda* Cre driver mouse, *Aicda*^{Cre} (*Aicda*^{tm1(cre)Mnz/J}) Jackson laboratories:007770 (Robbiani et al., 2008).
- iii. *VavP-BCL2* mice were generated at Walter and Eliza Hall Institute of Medical Research and provided by Andreas Villunger, Medical University Innsbruck (Ogilvy et al., 1999a).

Full genotype	Name	Group #
<i>Crebbp</i> ^{+/fl} ; <i>Aicda</i> ^{+/Cre} ; <i>VavP-BCL2</i> ⁺	CrebbpCreBCL2	1
<i>Crebbp</i> ^{+/fl} ; <i>Aicda</i> ^{+/+} ; <i>VavP-BCL2</i> ⁺	CrebbpBCL2	2
<i>Crebbp</i> ^{+/+} ; <i>Aicda</i> ^{+/Cre} ; <i>VavP-BCL2</i> ⁺	CreBCL2	3
<i>Crebbp</i> ^{+/+} ; <i>Aicda</i> ^{+/+} ; <i>VavP-BCL2</i> ⁺	BCL2	4
<i>Crebbp</i> ^{+/fl} ; <i>Aicda</i> ^{+/Cre} ; <i>VavP-BCL2</i> ⁻	CrebbpCre	5
<i>Crebbp</i> ^{+/fl} ; <i>Aicda</i> ^{+/+} ; <i>VavP-BCL2</i> ⁻	Crebbp	6
<i>Crebbp</i> ^{+/+} ; <i>Aicda</i> ^{+/Cre} ; <i>VavP-BCL2</i> ⁻	Cre	7
<i>Crebbp</i> ^{+/+} ; <i>Aicda</i> ^{+/+} ; <i>VavP-BCL2</i> ⁻	Wild type	8

Table 2.1. Genotype Key

To aid ease of reading and brevity in figures the various genotypes will be referred to by the abbreviated names or numbers above.

Mice were housed in a controlled environment with day/night cycles and *ad libitum* access to food and water. All procedures were performed in accordance with the UK Home Office Animals Scientific Procedures Act 1986, (Project Licence Number 70/7353, Personal Licence Number 70/24032).

2.1.1. Breeding strategy

Breeding pairs of all combinations were crossed. Upon weaning the mice were ear notched for identification and the tissue was used for genotyping. For the survival cohort breeding continued until each predefined genotype quota was reached and then genotypes were kept blinded until after the death of the animal.

2.1.2. Genotyping

Genomic DNA was extracted using Puregene® Cell and Tissue Kit (Qiagen; 158388) as per the manufacturer's instructions. The tissue was digested for approximately four hours at 55°C with 200µl of Cell Lysis Solution and 1.5µl Proteinase K. Once the tissue was completely lysed, 80µl of Protein Precipitation Solution was added and the samples were incubated on ice for five minutes. After centrifugation of 5 minutes at 13,000 rpm the supernatant was discarded, the pellet was washed with 70% ethanol and then centrifuged again for 5 minutes at 13,000 rpm. The supernatant was then discarded and the pellet allowed to dry. The DNA was then re-dissolved in 25µl of DNA hydrating Solution and incubated at 65°C for up to an hour to allow solubilisation. Concentrations were measured by spectrophotometer and the absorbance ratio D260/D280 ratios noted.

Genotyping was performed by PCR with specific primers, fragment size confirmed by gel electrophoresis on a 2% agarose gel (w/v) with 0.05% (w/v) ethidium bromide at 100V for 45minutes, using a DNA 100bpladder (NEB; N3231S) to size the PCR products. The following protocols were used.

2.1.2.1. *Crebbp^{fl}*

PCR amplification were performed using Taq Polymerase (Life Technologies; 18038026) according to the manufacturer protocol using primers MRC#644 (Z37) 5'-CCTCTGAAGGAGAAACAAGCA-3' and MRC#645 (Z38) 5'-ACCATCATTCATCAGTGGACT-3' (Zhang et al., 2004). Genomic DNA, extracted from ear biopsies, was diluted to approximately 100ng/µl and 1µl used in a 25µl PCR reaction along with 1x PCR buffer, 0.2mM per dNTP, 2.5mM MgCl₂, 0.5µM forward and reverse primers and 2.5units Taq Polymerase. Cycling parameters were: 95°C for 3minutes, 29cycles of 95°C for 30seconds, 54°C for 30seconds and 72°C for 30seconds followed by 3minutes at 72°C for final extension. Specific amplification produced the 300bp *Crebbp^{fl}* and 230bp *Crebbp⁺* fragments.

2.1.2.2. *Aicda*^{Cre}

The *Aicda* alleles were also amplified using Taq Polymerase (Life Technologies; 18038026). The reaction mix was: 100ng DNA, 1x PCR buffer, 0.2mM per dNTP, 1.5mM MgCl₂, 0.5μM forward and reverse primers and 2.5units Taq Polymerase in a 25μl reaction volume. Cycling parameters were: 95°C for 5minutes, 26cycles of 95°C for 30seconds, 70°C for 30seconds and 72°C for 45seconds followed by 30seconds at 72°C for final extension.

The *Aicda*⁺ and *Aicda*^{Cre} alleles were genotyped separately using a common forward primer MRC#89 5'-GGACCCAACCCAGGAGGCAGATGT-3' and the reverse primers MRC#90 5'-CCTCTAAGGCTTCGCTGTTATTACCAC-3' and MRC#88 5'-CACTCGTTGCATCGA CCGGTAATG-3' respectively resulting in 484bp wild type and 283bp mutant products.

2.1.2.3. *VavP-BCL2*

To avoid false positives *VavP-BCL2* was genotyped by QPCR using Phusion Hot Start II DNA Polymerase (Life Technologies; F549L) and SYBR® Green I nucleic acid stain (Sigma-Aldrich; S9430) allowing evaluation of cycle number rather than relying on visualisation by electrophoresis. The amplification mixture contained approximately 100ng genomic DNA, 1x HF-buffer, 0.2mM per dNTP, 0.4μM forward primer, 0.4μM reverse primer, 0.1X SYBRGreen and 0.4units of Phusion in a total of 20.0μl. *VavP-BCL2* primers were MRC#98 5'-AGACATGATAAGATAACATTGATGAG-3' and MRC#99 5'-CGAAGGGGTTCTCTAGTG-3' and the internal control primers were MRC#86 (oIMR7338) 5'-CTAGGCCACAGAATTGAAAGATCT-3' and MRC#87 (oIMR7339) 5'-GTAGGTGGAAATTCTAGCATCATCC-3' (Khalaj et al., 2013). The cycling parameters were: 98°C for 30seconds followed by 45 cycles of denaturation at 98°C for 5seconds, primer annealing at 60°C for 10seconds and extension at 72°C for 10seconds, with 10minutes final extension at 72°C. Specific amplification was confirmed by comparing the cycle numbers of known negative and positive controls to the samples.

2.2. Moloney Murine Leukaemia Virus

The pNCA plasmid contains full length MoMuLV, a gift from Stephen Goff (Addgene; 17363) (Colicelli and Goff, 1988). Before use the prep was fully sequenced using primers approximately every 400bp (data not shown).

2.2.1. MoMuLV production

MoMuLV was obtained from the filtered culture supernatant of 293T cells (ATCC; CRL-3216) transfected with the pNCA plasmid. Approximately 0.3×10^6 293T cells were plated on 10cm tissue culture dishes in DMEM media with 10%FCS (Life Technologies; 10270-106), 1X PenStrep (Life Technologies; 25030024) and 1X L-glutamine (Life Technologies; 25030024). Once they reached 60% confluency the media was replaced with 9ml serum free media and the transfection mixture was added. The transfection mix contained 10.0µg pNCA and 50.0µg polyethylenimine in 1ml of serum free media and had been allowed to complex for 30minutes. After incubating for four hours, the transfection media was removed and replaced with growing media. After three days the media, which now contained MoMuLV particles, was collected, filtered (0.45µM) and frozen at -80°C until use.

2.2.2. MoMuLV quantification

To quantify the titre of MoMuLV supernatant *in vitro* it was used to infect 293mCAT1 cells. 293mCAT1 cells are a modified human 293T cells that expresses the murine cationic amino-acid transporter (mCAT1) receptors, facilitating infection by MoMuLV, and inGluc-MLV-DERSE a GaussiaLuciferase (GLuc) reporter gene (kindly provided by the lab of Dr. Alan Rein, National Cancer Institute, Frederick, MD, USA).

mCAT1s were plated at 1.5×10^4 cells per 24well plate and allowed to reach 60% confluency at which point the media was removed, replaced with 500µl of transduction mixture and incubated for four hours. The cells were treated with a serial dilution of transduction mixtures, in triplicate, with 400µl diluted MoMuLV supernatant, 8µg/ml polybrene and made up to 2.0ml with DMEM media. Five dilutions of MoMuLV were tested: 1/5, 1/50, 1/500, 1/5,000 and 1/50,000. After incubation the cells were washed and returned to growing media.

Replicating MoMuLV is able to reverse transcribe the reporter GLuc to produced luciferase which is secreted into the media. The mCAT supernatant was sampled at 2, 4, 6, 8 and 10 days post infection and then the luciferase was quantified using the BioLux® Gaussia Luciferase Flex Assay Kit (NEB; E3300) and an injector equipped luminometer. In black plates with optical bottoms (Greiner Bio-One; 655096) 20µl of mCAT supernatant was mixed with 0.5µl BioLux GLuc Flex Substrate, 8µl BioLux GLuc Flex Stabilizer and 50µl BioLux GLuc Flex Assay Buffer. After incubating in the dark for 25minutes the plates

were read using the following parameters: 50µl of injection, 35-40 seconds of delay and 2-10 seconds of integration. The luminescence measured in the supernatant is proportional to the amount of luciferase produced which is in turn proportional to the reverse transcriptase activity of MoMuLV.

2.3. Generation of MoMuLV driven lymphoma

Newborn pups were injected with 50µl of MoMuLV supernatant *intra peritoneum* within 48hours of birth. After weaning at 3weeks mice were checked triweekly for signs of disease and weighed once week over a period of 13months. When moribund the mice were culled by Schedule 1 methods and tumours isolated. Moribund was defined as displaying any of the following phenotypes: enlarged spleen or lymph nodes, 20% weight gain (after 5weeks of age) or 20% weight loss (at any age) within one week, respiratory distress or two or more of the following clinical signs: piloerection, hunched posture, reduced mobility, pallor, ocular or nasal discharge, diarrhoea.

As standard the following was collect: whole spleen, spleen suspension, whole lymph node, lymph node suspension, whole thymus, thymus suspension and bone marrow suspension. Cell suspension were made by homogenisation using the gentleMACS™ Dissociator (Miltenyi Biotec;130-093-235) and gentleMACS C Tubes (Miltenyi Biotec; 130-093-237) and then filtering, 0.45µm. Whole organs were snap frozen in liquid nitrogen and stored at -80°C whereas cells suspension were preserved in 10%DMSO/media and gradient frozen before being stored in liquid nitrogen.

Kaplan-Meier survival curves were used to assess rate of death and the Gehan-Breslow-Wilcoxon test was used to determine the statistical significance. This test was chosen because it does not require a consistent hazard ratio (which in our cohorts increases with increasing age), but does require that one group consistently have a higher risk than the other.

2.4. Flow Cytometry

Spleen suspensions were characterised using flow cytometry using the Attune NxT Acoustic Focusing Cytometer, Life Technologies with a 96 well platform. The suspensions were defrosted and washed twice in buffer, PBS-2% FCS, and then resuspended in an appropriate volume of buffer. The cells were then incubated with 2.0µg Fc block per 10⁶ cells for 15minutes (TruStain fcX™ anti-mouse CD16/32, Biolegend;101320). The samples were then incubated with the antibody panels, see Table 2.2, for 15 minutes after which they were washed and then resuspended in 400µl 1X Focusing fluid (Life technologies; 4488621). All analyses were performed using FlowJo.

Antibody	Fluorophore	Emission	Dilution factor	Catalogue #
General Panel (GP)				
CD3	AF700	720/30R	1/200	Biolegend; 197780
CD4	PerCP-Cy5.5	695/40B	1/100	Biolegend; 100434
CD5	PE	583/22YG	1/100	Biolegend; 100608
CD8a	FITC	530/30B	1/100	Biolegend; 100706
CD19	BV421	440/50V	1/100	Biolegend; 115538
CD45	BV605	603/48V	1/100	Biolegend; 103140
B-cell Panel (BP)				
B220/CD45R	BV510	512/25V	1/100	Biolegend; 103247
CD19	BV421	440/50V	1/100	Biolegend; 115538
IgD	APC-Cy7	780/60R	1/100	Biolegend; 405716
IgM	PE	583/22YG	1/100	Biolegend; 406508
IgG1	FITC	530/30B	1/100	BD Bioscience; 553443
IgG2a/2b	FITC	530/30B	1/100	BD Bioscience; 553399
IgG3	FITC	530/30B	1/100	BD Bioscience; 553403
Lambda/Kappa Panel (LKP)				
CD19	BV421	440/50V	1/100	Biolegend; 115538
CD23- B3B4	BV510	512/25V	1/100	BD Bioscience; 563200
CXCR5- L138D7	BV605	603/48V	1/400	Biolegend; 145513
Kappa	AF700	720/30R	1/100	Biolegend; 409508
Lambda	APC	660/20YG	1/100	Biolegend; 407306
Germinal Centre Panel (GC)				
CD19	BV421	440/50V	1/100	Biolegend; 115538
CD95	PECy7	780/60YG	1/200	BD Bioscience; 557653
PNA	FITC	488/10B	1/100	Vector Laboratories FL-1071

Table 2.2. Summary of antibodies used in flow cytometry

Listed, the antibodies used in the cytometric characterisation tumours, what fluorophores they were conjugated with and at what wavelength they were read.

2.5. Tumour RNA and DNA extraction

DNA extraction: AllPrep RNA/DNA 96 Kit (Qiagen; 80311) or the AllPrep RNA/DNA Mini Kit (Qiagen; 80204) as per the manufacturer's instructions. Disposable pestle and mortars were used to disrupt tissues and the QIAshredder (Qiagen; 79656) was used to homogenise tissues. The DNA was quantified using Qubit® dsDNA BR Assay Kit and fluorometer. The RNA was stored for future experiments.

2.6. Spleen suspension B- and T-cell purification

Spleen cell suspensions were defrosted and washed twice in PBS-2% FCS and then resuspended to 1.0×10^8 cells/ml. Cells were purified using the EasyPlate™ EasySep™ Magnet for 96well plates (STEMCELL TECHNOLOGIES; 18102) and their negative B-cell (19854) and T-cell (19751) selection kits as per the manufacturer's instructions and then pelleted and stored for subsequent RNA and DNA extraction.

2.7. UMI-LM-PCR (Unique Molecular Identifier – Ligation Mediated – PCR)

Tumour DNA was extracted and processed into libraries of 96 tumours using a semi-automated, high-throughput library preparation pipeline in combination with a modified Illumina sequencing protocol to identify MoMuLV integration sites. This method has not yet been published. Figure 2.4. is a detailed schematic of the method.

2.7.1. Shear

Tumour DNA was diluted to 20ng/μl and 1.1μg was sheared using the Covaris E220 Sonicator, 96 well, with the following settings: Peak Incident Power 175watts, Duty Factor 10%, Cycles per Burst 200, Treatment Time 180sec, Temperature range 0-40°C. The product size, 400bp, was confirmed using the Agilent Bioanalyser HS DNA assay.

2.7.2. Blunt and A-tail

After shearing the DNA was blunted to create 5'-phosphorylated blunt ends using NEBNext® End Repair Module (NEB; E6050L) in 77μl using 4000U of T4 polynucleotide kinase and 12U T4 DNA polymerase (components on the NEBNext® End Repair Enzyme Mix).

After blunting the DNA was cleaned using solid phase reversible immobilisation (SPRI) beads, (Agencourt AMPure XP magnetic beads (Beckman Coulter; A63882)), on the Biomek® NXP Laboratory Automation Workstation (Beckman Coulter; A31839) in 96 well format. Briefly, 77μl of blunt DNA was added to 90μl of beads, mixed and then incubated for 10minutes to allow the DNA to bind to the beads. The plate was then incubated on a 96 well magnet. After 10minutes the supernatant containing the impurities was removed and the DNA-bead pellet was washed twice with 80% ethanol. After the DNA-bead pellet was air dried for 5 minutes, it was re-suspended in 50μl of distilled water and allowed to solubilise for 2minutes. The DNA was then returned to the magnet for a further 5minutes after which 42μl of cleaned DNA was aspirated and moved into a clean 96 well plate. All further clean steps were performed in this manner but with the volumes of DNA and beads varying slightly.

To the 3'-end of blunt DNA adenosine nucleotide overhangs were added using the NEBNext® dA-Tailing Module (NEB; E6053L) according to the manufactures protocol using 3μl Klenow Fragments (3'→5' exo⁻) in a 50μl reaction.

The DNA, 50μl, was then cleaned again as described above using 90μl of beads producing 36μl of clean DNA.

2.7.4. EcoRV digest

The DNA is then digested with EcoRV-HF® (NEB; R3195L) prior to primary PCR to prevent amplification of internal virus fragments. The DNA was digested in a 60µl reaction with 1X CutSmart buffer (NEB; B7204S) and 20U EcoRV-HF for 16hours at 37°C and then the enzyme was inactivated at 65°C for 20minutes.

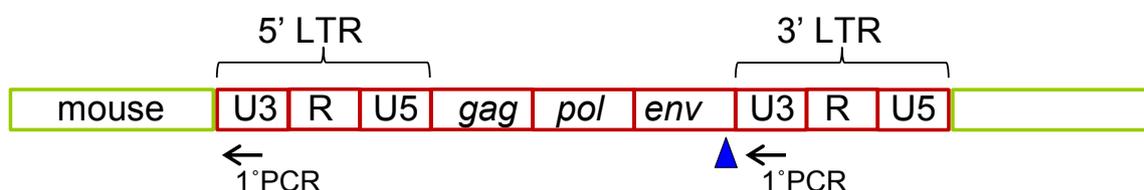


Figure 2.2. Detail of UMI-LM-PCR: EcoRV digest strategy

Sonication has removed the need to fragment the DNA using restriction enzymes however an EcoRV digest is still needed to digest the virus LTR. Digestion with EcoRV (blue triangle) prevents the primary PCR from amplifying between the LTRs. When the primary PCR LTR primer binds it amplifies the integration site in the mouse DNA (green).

The digested DNA, 60µl, was then cleaned as above (section 2.7.2) using 60µl of beads to produce 100µl of clean DNA ready for fragment size selection.

2.7.5. Size selection

A two stage size selection was performed using Agencourt AMPure XP SPRI beads to consecutively remove the long fragments and then the short. Binding to SPRI beads can be controlled by varying the polyethylene glycol concentration from the bead buffer in the mixture (Yamanaka et al., 2011). Low buffer concentrations permit binding of long fragments which are discarded; increasing the buffer concentrations allows binding of shorter fragments to the beads which are then retained leaving the shortest fragments in the supernatant to be discarded (Quail et al., 2009).

First 60µl beads are added to 100µl DNA and incubated for 10 minutes, this low concentration of buffer permits binding of the longest fragments to the beads. The DNA is then incubated on the magnet for 10 minutes and then 158µl of supernatant, containing the unbound medium and short fragments, is aspirated from the beads and retained for the second size selection. A second volume of beads, 50µl, is then added making the cumulative bead buffer volume 110µl relative to the 100µl DNA. This increase in bead buffer concentration allows the medium fragments to bind and be retained on the beads as they incubate for 10 minutes. The mixture is incubated again on the magnet for 10 minutes and then the supernatant, containing the shortest fragments, is discarded. The beads bound to the desirable medium fragments are then washed twice with 80% ethanol, dried and then re-suspended in 32µl water.

2.7.6. Primary PCR

Fragments of interest, containing MoMuLV insertions, were amplified using nested PCR with a first strand synthesis (Figure 2.3). The primary PCR was performed using MRC#216 5'-GCGTACTTAAGCTAGCTTGCCAAACCTAC-3' and MRC#415 5'-AATGATACGGCGACCACCGAGATCTACAC-3'. Initially only MRC#216, which is complementary to the 5' LTR of MoMuLV, is able to bind meaning that the first cycle of PCR enriches only fragments with inserts. From the second cycle, MRC#415 is able to bind to the newly amplified portion complementary to the adaptor overhang, only present after the first cycle of PCR.

First Strand Synthesis:

```

5' -AATGATACGGCGACCACCGAGATCTACACXXXNNNGTACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxTGAAGAGCCCACTGTAGGTTGGCAAGCTAGCTTAAGTAACGC..other adaptor-3'
                                     |
                                     MRC#216 3' -CATCCAACCGTTCGATCGAATTCATTGCG-5'

3' -AGAAAGGGATGTGCTGCGAGAAGGCTAGAxxxACTTCTGGGGTGGACATCCAACCGTTCGATCGAATTCATTGCG..other adaptor-5'

```

Subsequent amplification:

```

5' -AATGATACGGCGACCACCGAGATCTACACXXXNNNGTACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxTGAAGAGCCCACTGTAGGTTGGCAAGCTAGCTTAAGTAACGC-3'
                                     |
5' -AATGATACGGCGACCACCGAGATCTACAC-3' MRC#415                                     MRC#216 3' -CATCCAACCGTTCGATCGAATTCATTGCG-5'
|
3' -TTACTATGCCGCTGGTGGCTTAGATGTGXXXNNNCACTGACCTCAAGTCTGCACACGAGAAGGCTAGAxxxACTTCTGGGGTGGACATCCAACCGTTCGATCGAATTCATTGCG-5'
|---de novo sequence-----|

```

Figure 2.3. Detail of UMI-LM-PCR: Primary PCR

The mouse genome (light green) with adaptor ligated undergoing primary PCR with a first strand synthesis step enriching for DNA with MoMuLV (red) insertion in.

PCR was performed using Phusion Hot Start II High-Fidelity DNA Polymerase kit (Thermo Scientific, F549L) as per the manufactures protocol with SYBR® Green I (Sigma-Aldrich; S9430). To the size selected DNA, 5X HF-buffer, 0.4mM per dNTP, 1.0µM MRC#415, 1.0µM MRC#216, 0.1X SYBRGreen and 1.0unit of Phusion in a total of 50µl. The cycling parameters were 98°C for 30sec, followed by 16 cycles of 98°C for 10seconds, 66°C for 30 seconds, and 72°C for 30 seconds, and finally 72°C for 5minutes.

The primary PCR product, 50µl, was then cleaned as above (section 2.7.2), using 50µl of beads to produce 32µl of clean DNA which contained, on average was 80ng of DNA.

2.7.7. Secondary PCR

The secondary PCR further enriches PCR products with virus inserts by using a nested primer within the virus LTR. This primer also adds a second index and a sequence complementary to oligo D on the Illumina HiSeq chip to the virus LTR end of the PCR product. Twelve indices were used per 96 samples and were alternated on the sample plate so that adjacent wells never contained the same so that index analysis could resolve intra-sample cross-contamination. For the full list of secondary PCR primers see Appendix 0, page 169. The whole, clean primary PCR product was then input into the secondary PCR which was processed as the primary with MRC#415 and the appropriate indexed secondary PCR primers.

The secondary PCR product, 50µl, was then cleaned as above (section 2.7.2), using 50µl of beads to produce 100µl of clean DNA which then underwent a second size selection as in section 2.7.5 to produce the final product. The DNAs were then quantified and then 20ng of each sample was pooled to make a library of 96 samples.

2.7.8. HiSeq

The libraries were quantified for amplifiable fragments by qPCR using the KAPA Illumina SYBR Universal Lib Q. Kit (Anachem; KK4824) as per the manufacturer's instructions. This kit contains primers complementary to oligos C and D, the Illumina sequences at each end of the PCR product. Equal amounts of each of the seven libraries were pooled to produce the final library for sequencing. The pooled library was sequenced on the Illumina HiSeq, a dual-index, paired-end protocol was used, with a modified Nextera protocol (summarised in Figure 2.4). The modifications included reversing oligos C and D and modifying the number of dark cycles For more details see section 5.1. The Illumina sequencing instruments generate *.bcl files as their primary sequencing output which CASAVA demultiplexes and converts to FASTQ files (done by the CSC Genomics Facility).

UMI-LM-PCR

Example barcode #101 5'-TCTGTATTTC-3'

Upper Strand Adaptor with barcode #101: MRC#850 5'-AATGATACGGCGACCACCGAGATCTACACGAAATACAGANNNNNNNGTACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
|-oligoC-----| |5'index-||-UMI--|

Secondary PCR Primer with barcode #101: MRC#946 5'-CAAGCAGAAGACGGCATAACGAGATTTCTGTATTTCTGCTAGCTTGCCTAACCTACAGGTGG-3'
|-oligo D overhangs----||3'index-||-matches 5'LTR-----|

Annealed Adaptor

Adaptor MRC#850 5'-AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXXXXXXXXNNNNNNNGTACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'
Universal lower strand Adaptor MRC#443 Universal 3'-AGAAAGGGATGTGCTGCGAGAAGGCTAG*-5'

Ligated fragment DNA and adaptor

Adaptor #8505' -AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXXXXXXXXNNNNNNNGTACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxxxx.....TGAAAGACCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGC..... other adaptor -3'
AGAAAGGGATGTGCTGCGAGAAGGCTAGAxxxxxx.....ACTTCTGGGGTGGACATCCAACCCTTCGATCGAATTCATTGCG..... other adaptor -5'

Primary PCR

1st strand synthesis: 5'-AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXXXXXXXXNNNNNNNGTACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxxxx.....TGAAAGACCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGC..... other adaptor -3'
MRC#216 3'-CATCCAAACCGTTCGATCGAATTCATTGCG-5'
AGAAAGGGATGTGCTGCGAGAAGGCTAGAxxxxxx.....ACTTCTGGGGTGGACATCCAACCCTTCGATCGAATTCATTGCG..... other adaptor -5'

Subsequent amplification: 5'-AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXXXXXXXXNNNNNNNGTACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxxxx.....TGAAAGACCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGC-3'
MRC#415 5'-AATGATACGGCGACCACCGAGATCTACAC-3'
MRC#216 3'-CATCCAAACCGTTCGATCGAATTCATTGCG-5'
3'-TTACTATGCCGCTGGTGGCTCTAGATGTGXXXXXXXXXXXXXXXXNNNNNNNCACTGACCTCAAGTCTGCACAAGAGAAGGCTAGAxxxxxx.....ACTTCTGGGGTGGACATCCAACCCTTCGATCGAATTCATTGCG-5'
|---de novo in 1stSS-----|

Secondary PCR

5'-AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXXXXXXXXNNNNNNNGTACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxxxx.....TGAAAGACCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGC-3'
MRC#415 5'-AATGATACGGCGACCACCGAGATCTACAC-3'
3'-TTACTATGCCGCTGGTGGCTCTAGATGTGXXXXXXXXXXXXXXXXNNNNNNNCACTGACCTCAAGTCTGCACAAGAGAAGGCTAGAxxxxxx.....ACTTCTGGGGTGGACATCCAACCCTTCGATCGAATTCATTGCG-5'

Secondary PCR Product

5'-AATGATACGGCGACCACCGAGATCTACACXXXXXXXXXXXXXXXXNNNNNNNGTACTGGAGTTCAGACGTGTGCTCTTCCGATCTxxxxxx.....TGAAAGACCCACCTGTAGGTTTGGCAAGCTAGCXXXXXXXXXXATCTCGTATGCCGCTTCTGCTTG-3'
3'-TTACTATGCCGCTGGTGGCTCTAGATGTGXXXXXXXXXXXXXXXXNNNNNNNCACTGACCTCAAGTCTGCACAAGAGAAGGCTAGAxxxxxx.....ACTTCTGGGGTGGACATCCAACCCTTCGATCGXXXXXXXXXXTAGAGCATACGGCAGAGACGAAC-5'

HISEQ

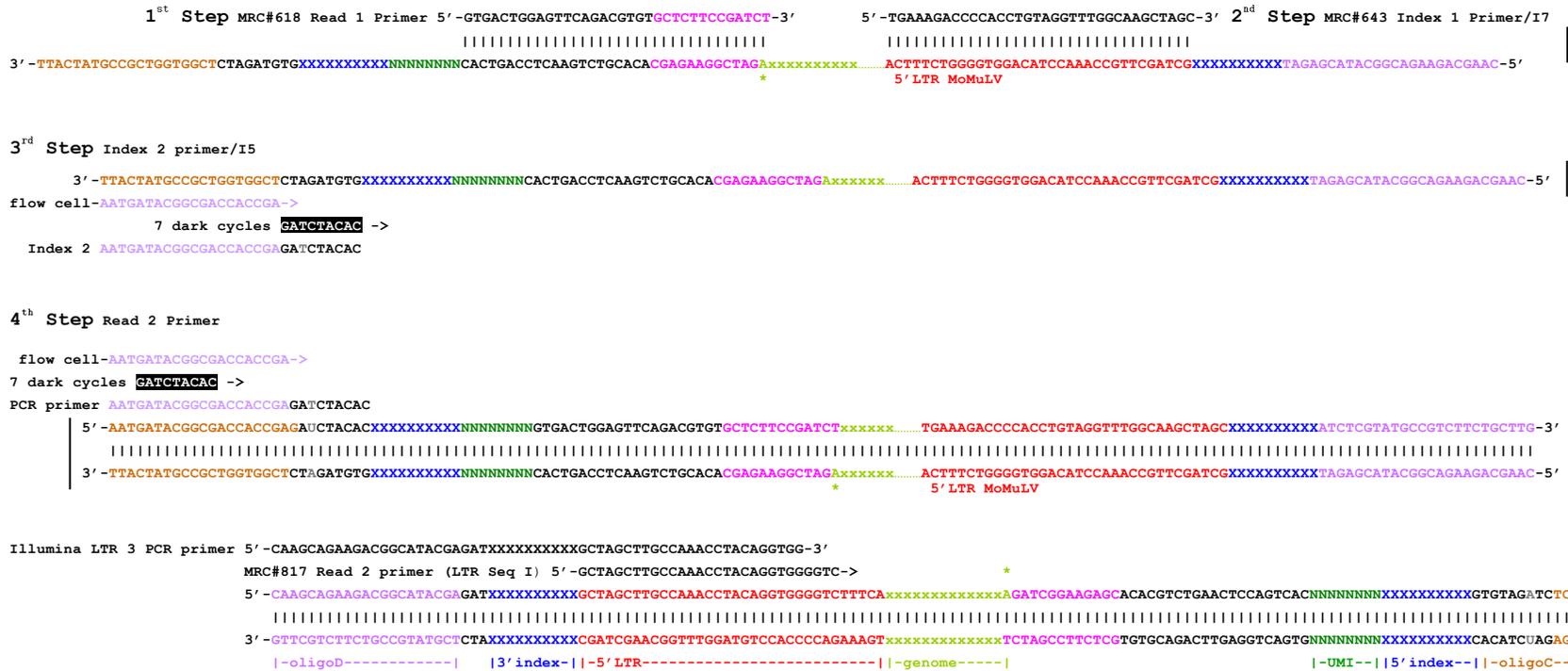


Figure 2.4. Overview of UMI-LM-PCR and HiSeq

Overview of the library preparation protocol. Ten base pair indices (blue) were adapted from 25mer indices published by Xu et al., 2009. The same 96 indices are used in the upper strand adaptor and the secondary PCR primers but in opposite orientations. The secondary PCR primer contains a common overhang which bind the sequencing flow cell. Twelve secondary PCR primers, with different indices, were used per 96 tumour library in unique combinations with the adaptor, allowing all samples to be pooled and later demultiplexed.

The adaptors are made by ligating the upper strands (96x, MRC#850-945) to the universal lower strand (MRC#433) via their complementary portion (pink) to make a hairpin shaped adaptors with a T-overhang. The adaptor contains a common portion (orange) which also binds the sequencing chip (Illumina oligo C) (orange), a 10bp index (blue) and an 8bp UMI (dark green). The adaptor is ligated to the sheared end of tumour DNA via a complementary A-overhang which was added to the randomly sheared ends by A-tailing (*).

This product then undergoes two rounds of PCR. The primary PCR first strand synthesis is exclusive to DNA fragments containing virus insertions. The adaptor specific primer (MRC#415) has no template until after the reverse MoMuLV specific primer (MRC#216) has bound and sequenced the first strand. After amplification of the first strand then both primers can bind and amplify the fragments for a further 15 cycles.

The secondary PCR is nested further improving specificity of amplification for a further 16 cycles. It also adds oligo D to the PCR products to allow annealing of the final product to the sequencing flow cell.

The library was sequenced on the Illumina HiSeq, a dual-index, paired-end protocol was used. In the first step the Read 1 Primer (MRC#618) sequences from the sheared end of the DNA (*) towards MoMuLV 5'LTR including the junction between the mouse genome and the MoMuLV insertion enabling visualisation of the exact insertion point. The second step, Index 1 Primer/ I7 (MRC#643) sequences from the MoMuLV 5'LTR through the 3' index junction. The third sequencing step occurs when oligo C is cleaved from the chip at the grey T making the Index 2/I5 primer that sequences the 5' index and the UMI after 7 dark cycles where no fluorescence is measured. In the final step the Read 2 primer (MRC#817) sequences from the 5' LTR end into the tumour DNA. Vertical lines indicate binding to the flow cell.

2.8. Bioinformatics Pipeline

The bioinformatics pipeline was designed and applied by our bioinformatician Barbara Iadarola.

The Hiseq reads were demultiplexed and aligned to the mouse genome (mm10) by their LTR position. The paired and mapped reads were made into an inserts table and inserts were merged if they were within 10bp of each other. The inserts were then filtered to remove PCR artefacts and cross-contamination between samples and then divided into three tables: 1. with inserts for each sample processed by UMI-LM-PCR, 2. grouping the inserts from DNA samples that had been repeated and 3. grouping all inserts from multiple organs from one mouse.

The inserts grouped by DNA sample were then used to calculate the entropy of a tumour and the clonality of its inserts. The clonal inserts were then used to identify CIS. CIS are identified using a Gaussian kernel convolution based method (de Ridder et al., 2006) and putative target genes for each insertion are assigned automatically using Kernel Convolution Rules Based Mapping (KC-RBM) software (de Jong et al., 2011). KC-RBM automates identification of statistically significant CIS with rule-based mapping of inserts to genes that are likely to be deregulated, based on the literature of interactions between retroviral insertions and the expression of nearby genes. The algorithm considers the quantity and orientation of the inserts near a gene.

The KC-RBM package initially identifies common insertion sites by plotting a Gaussian kernel at the site of each insertion. Overlapping kernels are summed and the resulting density distribution (convolution) is scanned to identify the peaks. The heights of these peaks are an indicator of insertion density. Each insertion is assigned to the nearest peak and all inserts assigned to the same peak are merged into a cluster. For each gene in the genome, the presence of a cluster near that gene is identified using four window sizes (upstream-sense and antisense, downstream-sense and antisense, with respect to the transcription start site). When a cluster falls within a given window, the inserts in that cluster are associated to the gene. Where more than one gene is associated with a cluster, genes are ordered by proximity to that cluster.

The output of KC-RBM program is directly saved into a local MySQL database, and automatically integrated with local copies of Ensembl, COSMIC and DrugBank.

7 libraries of 96 samples were pooled and sequenced on the Illumina HiSeq, using a dual-index, paired-end protocol.

The HiSeq generates *.bcl files as their primary output which CASAVA demultiplexes and converts to FASTQ files. One mismatch was allowed in the index (done by the CSC Genomics Facility). Three FASTQ files are produced per sample: one for lane 1, one for lane 2 and one for the UMIs. FASTQ format provides information about the quality of DNA sequencing, expressed as a score associated to each nucleotide. The bioinformatics from this step onwards was performed by Barbara Iadarola.

The two lanes were concatenated producing a unique file per sample which was then trimmed (allowing 2 mismatches) removing low quality reads and the sequence external to the indices. They were then aligned to the mouse reference genome (mm10).

The alignment outputs Sequence Alignment/Map (SAM) files which are converted into BAM files, whose indexing facilitates fast retrieval of alignments overlapping a specific region (the indices). From the BAM file relevant information can be extracted e.g. the starting and ending position of the aligned sequence, the chromosome and the strand orientation. Paired and mapped reads have both indices (5' and 3') and map to the genome.

The list of paired and mapped reads for each sample is transformed into an inserts table. The 5'LTR positions were clustered if they occurred within 10 bases of each other. The inserts were built by clustered LTR positions and each one given a unique ID, chromosome location, orientation, sample_id, minimum LTR position and maximum LTR position from the alignment.

One LTR position can have multiple reads which are subdivided into fragments based on the number of unique UMIs. Other information is then added to the inserts table e.g. the "best base" -LTR position with highest number of fragments, the median base and information about the such as mouse like mouse_id, age and genotype.

The inserts are grouped by mouse and filtered to remove contamination. LTR positions present in more than one mouse are collated in a list and then checked; any insertion present in a non-infected mouse or that occurs more than 10 mice is removed. If three or more inserts are shared by a pair of mice they are also removed. The contaminant list is then manually curated e.g. all Mycn inserts are rescued unless they are shared by two or more mice. Suspicious regions e.g. regions resembling mispriming of the PCR primers directly to the genome: chr9:3000251-3000895, chr9:3024394-3024454, chr21:4795984-4795984 , chr2: 98662236-98667333 , Chr12:3109866-3110127 are removed.

Three inserts tables are then generated: 1. built by sample (each sample is considered as unique), 2. built by DNA sample (repeats of the same sample are merged) and 3. built by mouse (all tissues from one mouse are merged together). Clonality (the relative abundance of an individual insertion in a particular sample) is then calculated within a sample, a DNA sample or a mouse. *Clonality = number of fragments/ the total number of fragments in the sample. Normalised clonality = all the clonality values : the most clonal insert within one sample (most clonal 1, less clonal 0).*

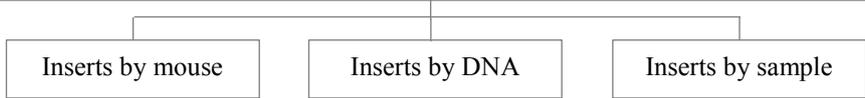


Figure 2.5. Bioinformatics Pipeline

Summary of the major steps in the bioinformatics pipeline.

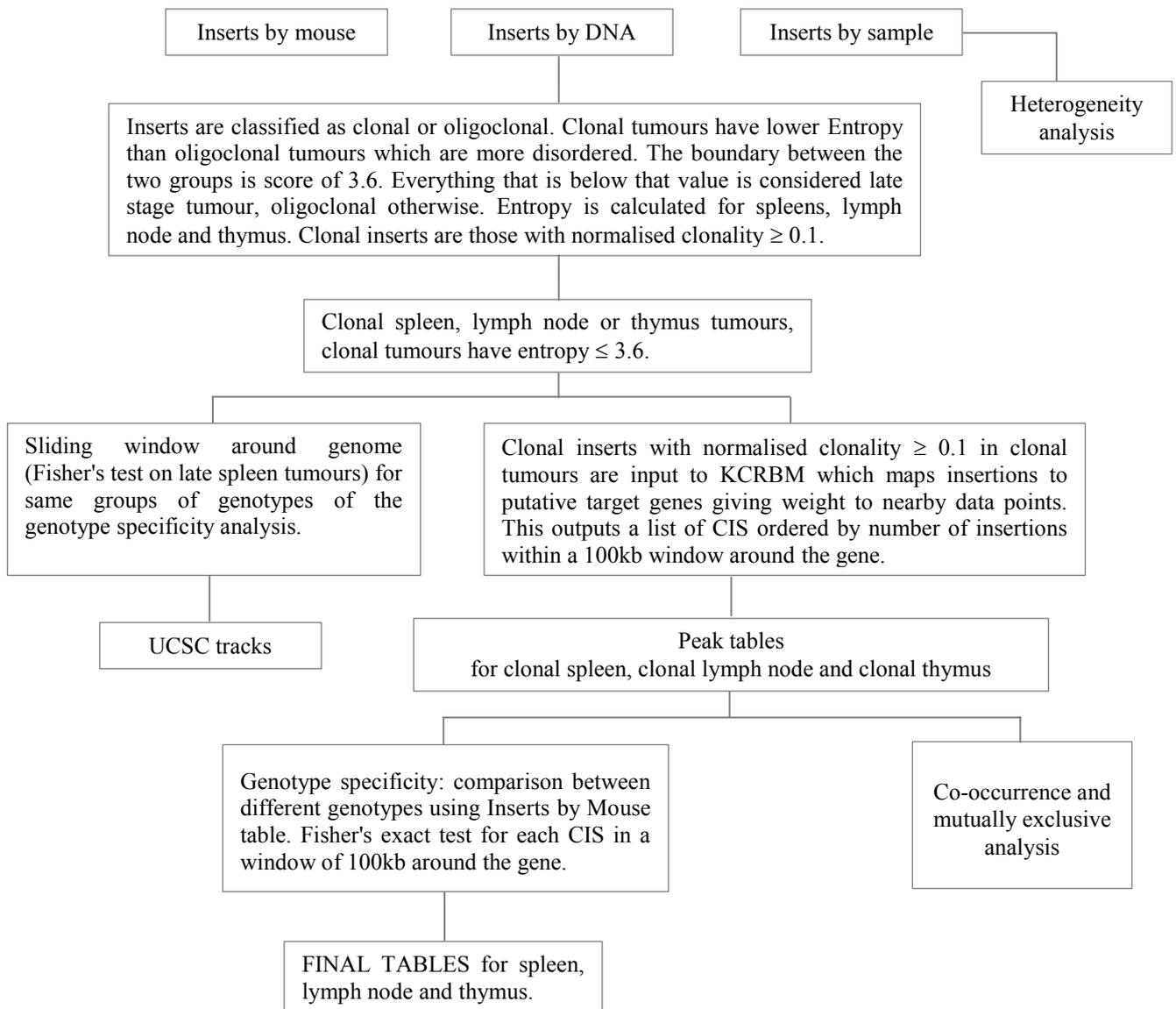


Figure 2.6. Analysis Pipeline

Summary of the analysis performed on the insertion dataset.

All inserts > spleen inserts (entropy > 3.6) > all end stage inserts (entropy ≤ 3.6) > all end stage clonal inserts (clonality ≥ 0.1)

3. Characterisation of a novel murine model of spontaneous B-cell lymphoma

In order to investigate the contribution of *Crebbp* loss in the context of *BCL2* overexpression in B-cell lymphoma we generated a novel murine model of B-NHL that recapitulates two of the earliest events in FL and DLBCL, namely overexpression of *BCL2* and hemizygous loss of *Crebbp*. To this end we combined three alleles, *VavP-BCL2*, *Crebbp flox* and *Aicda Cre* to produce a model with overexpression of *BCL2* in the haematopoietic compartment and the conditional loss of one allele of *Crebbp* in the GC B-cells.

3.1. Spontaneous lymphoma cohort

Mice were crossed to give the eight allele combinations (see Table 3.1). They were examined tri-weekly for up to 400 days at which point they were culled. Mice that exhibited external symptoms of lymphoma or became moribund were sacrificed and the presence of lymphoma was determined by necropsy. Samples of hematopoietic tissues were snap frozen and cell suspensions were cryopreserved for later analysis by flow cytometry. Animals that became moribund for reasons other than lymphoma or that reached 400 days without succumbing to disease were censored from survival analysis.

As mice aged they increasingly developed spontaneous lymphomas (38.8%). Almost half (46.7%) of the mice survived until 400 days symptom free and were censored from survival analysis. Six mice were culled when they lost 20% body weight but were found to have no enlarged lymphoid organs and therefore they were censored. Other reasons for being censored include: seizure, blocked gastrointestinal tract, prolapse and inner ear infections. The mean spleen weight of the censored mice was 0.22g whereas it was 0.91g in those classified as spontaneous lymphoma.

		Spontaneous lymphoma cohort				
		Male	Female	TOTAL	MS, days	Cause of death
Crebbp^{+fl} ; Aicda^{+Cre} ; VavP-BCL2⁺	CrebbpCreBCL2	18	23	41	297	5 symptom free, 4 censor, 32 lymphoma
Crebbp^{+fl} ; Aicda^{+/+} ; VavP-BCL2⁺	CrebbpBCL2	15	13	28	329	4 symptom free, 10 censor, 14 lymphoma
Crebbp^{+/+} ; Aicda^{+Cre} ; VavP-BCL2⁺	CreBCL2	9	3	12	>387	6 symptom free, 1 censor, 5 lymphoma
Crebbp^{+/+} ; Aicda^{+/+} ; VavP-BCL2⁺	BCL2	5	6	11	>311	3 symptom free, 4 censor, 4 lymphoma
Crebbp^{+fl} ; Aicda^{+Cre} ; VavP-BCL2⁻	CrebbpCre	19	11	30	>402	25 symptom free, 3 censor, 2 lymphoma
Crebbp^{+fl} ; Aicda^{+/+} ; VavP-BCL2⁻	Crebbp	7	3	10	>402	all symptom free
Crebbp^{+/+} ; Aicda^{+Cre} ; VavP-BCL2⁻	Cre	5	7	12	>402	11 symptom free, 1 lymphoma
Crebbp^{+/+} ; Aicda^{+/+} ; VavP-BCL2⁻	Wildtype	4	4	8	>402	7 symptom free, 1 lymphoma
TOTAL				152		

Table 3.1. Spontaneous lymphoma cohort

Summary of the spontaneous lymphoma cohort including numbers of animal of each genotype, their mean number of days to death and their cause of death. Almost half (46.7%) of the mice survived for 400 days symptom free and were censored. Other reasons for censoring include seizure, blocked gastrointestinal tract, prolapse and inner ear infections.

3.2. Spontaneous lymphoma survival analysis

Comparison of the spontaneous lymphoma cohort survival is shown in Figure 3.1. and Figure 3.2. Of the mice that survived to 400 days symptom free 74.6% lacked the *BCL2* transgene. The mean and median survival of the combined non-transgenic *BCL2* mice was 57.3 weeks and 58.4 week respectively whereas they were 39.5 weeks and 39.1 weeks in the combined *BCL2* transgenic mice.

All four groups of mice lacking the *BCL2* transgene (Crebbp, Cre and CrebbpCre, refer to Table 2.1. Genotype Key) had similar survival curves and all had median survival lengths greater than 402 days (see Figure 3.1.A). It is also important to note that loss of *Crebbp* alone in the CrebbpCre mice (red line) does not adversely affect survival (see Figure 3.1.A). This is consistent with a previous study of conditional *Crebbp* loss, in CD19+ B-cells, where the mice reached a year without succumbing to lymphoma (Xu, 2006).

Of the *BCL2* transgenic mice that died, almost all developed late onset lymphoma (93.2%). This replicates previous studies that showed that overexpression of *BCL2* causes increased incidence of lymphoma (Egle et al., 2004). Overexpression of *BCL2* decreased the median survival from > 402 day in the wild types to >311 days in the BCL2 mice ($p = 0.2004$) (see Figure 3.1.B).

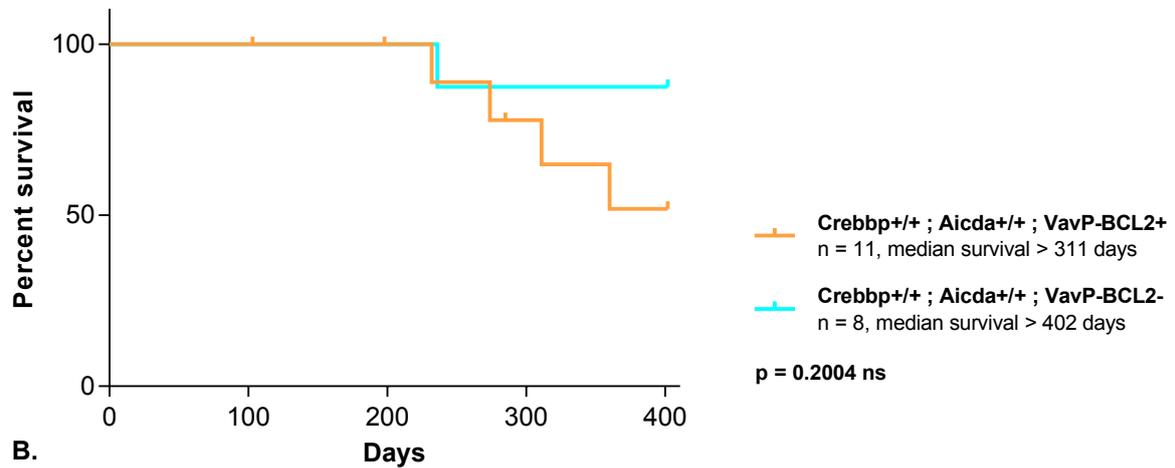
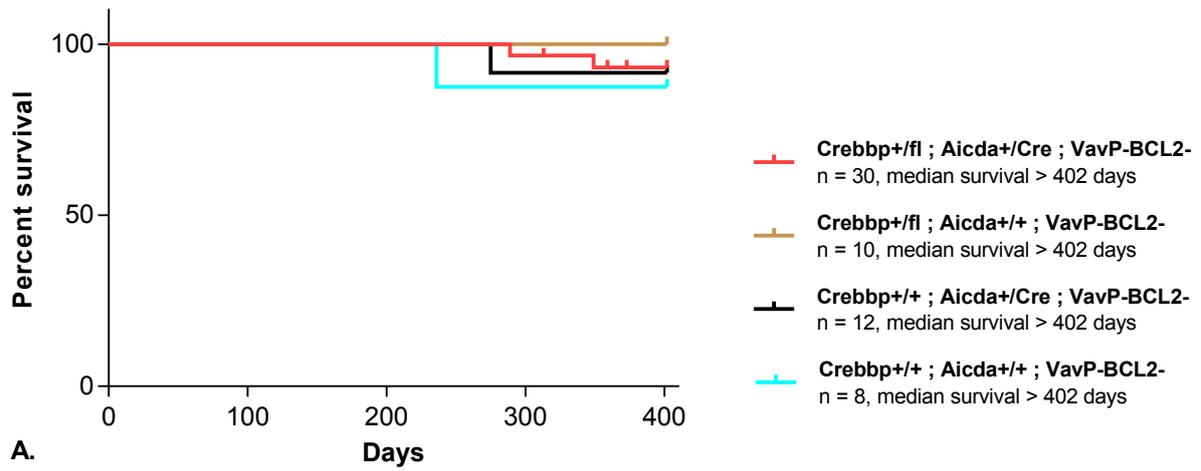


Figure 3.1. Non-transgenic BCL2 cohort survival curves

A. The non-transgenic BCL2 mice had survival similar to wild type mice. **B** Overexpression of BCL2 reduces survival but not significantly at the 400 day time point.

Loss of *Crebbp* in conjunction with overexpression of *BCL2* significantly increased speed on onset of lymphoma. The *CrebbpCreBCL2* mice died significantly faster than the *CreBCL2* (($p = 0.0344^{**}$) Figure 3.2.A) and the *BCL2* mice (($p = 0.0393^{*}$) Figure 3.2.B). There are many mechanisms by which this could be possible but they have not been investigated in this thesis. For example loss of *Crebbp* could create an environment where other mutations that cooperate with *BCL2* overexpression are tolerated and thus promote oncogenesis. By comparing the mutations that are specific to the *CrebbpCreBCL2* mice and the *BCL2* or *CreBCL2* mice examples of mutations that are only tolerated when *Crebbp* is missing will be identified and can be interrogated further. It is clear from the human sequencing studies that *Crebbp* loss occurs concurrently with *BCL2* overexpression and so this relationship warrants further investigation.

Survival of *CreBCL2* and *BCL2* did not significantly differ (Figure 3.2.C) indicating the presence of *Cre* had no influence on *BCL2* driven lymphomagenesis. Combining these two genotypes and comparing their survival to the *CrebbpCreBCL2* (Figure 3.2.D) increased the statistical power of the comparison and the significance to $p = 0.0048^{**}$.

This is the first time that the loss of *Crebbp* has been shown directly to cooperate with overexpression of *BCL2* to promote spontaneous lymphoma in mice, though they have been found to be commutated in human B-NHL (Morin et al., 2011).

It is noteworthy that the *CrebbpBCL2* mice have reduced survival relative to the *BCL2* and *CreBCL2* groups (green line Figure 3.2.E). Though this reduction is not significant when compared to the grouped *BCL2* and *CreBCL2* ($p = 0.0586$) the presence of the *Crebbp*^{+*fl*} allele even without the recombination driver *Aicda*⁺/*Cre* reduces survival to an intermediate degree. The difference between *CrebbpCreBCL2* and *CrebbpBCL2* is also not significant ($p = 0.4234$). Together this may suggest that the *Crebbp*^{+*fl*} allele is hypomorphic, perhaps because the *loxP* sites are altering transcription of *Crebbp* and causing it to have an intermediate phenotype. Expression of this allele has previously been shown, by western blot, to be lower than wild type alleles (Xu, 2006). Although the *Crebbp* floxed allele has been extensively backcrossed to a C57BL/6 background from a 129 background (Beat Lutz, personal communication) it is also theoretically possible that other linked genes flanking the original 129 allele are influencing latency.

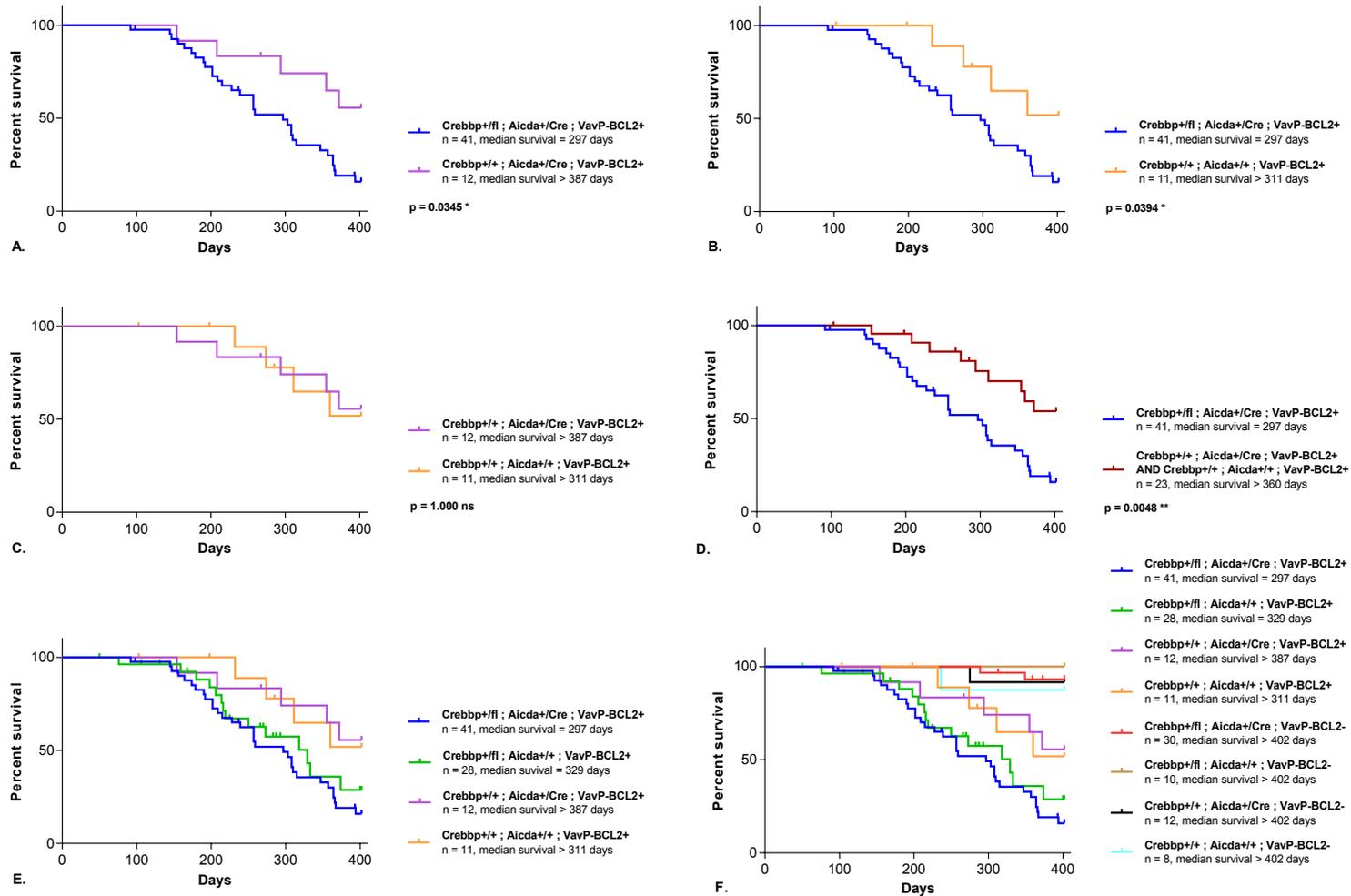


Figure 3.2. BCL2 transgenic cohort survival curves

A Loss of *Crebbp*, *CrebbpCreBCL2* (blue), significantly decreased survival relative to the *CreBCL2* (purple) ($p = 0.0345$). **B** The *CrebbpCreBCL2*, have a significantly decreased survival relative to the transgenic *BCL2* mice (orange) ($p = 0.0394$). **C** The difference between the *CreBCL2* and *BCL2* is not significant and so they can be summed as transgenic *BCL2* mice (dark red). **D** The *CrebbpCreBCL2* have very significantly decreased survival relative to the summed transgenic *BCL2*s ($p = 0.0048$). **E** Summary of the transgenic *BCL2* genotypes. The intermediated position of the *CrebbpBCL2* indicates that *Crebbp* allele is hypomorphic. **F** Overview of the spontaneous lymphoma cohort.

3.3. Cytometric characterisation of the spontaneous lymphoma

We characterised 52 cryopreserved spleen suspensions by multi-parameter flow cytometry. Of these mice, 23 had enlarged spleens, 11 had moderate spleens and 18 had no signs of disease. These samples were stained with four panels of antibodies: General Panel (GP), B-cell Panel (BP), Lamda/Kappa Panel (LKP) and Germinal Centre Panel (GCP).

FLs are characterised by high expression of CD23 (a marker of follicular B-cells) and tend to be CD5 negative. DLBCLs typically express CD19, CD20, CD22, CD79a and often have surface immunoglobulin, a small portion express CD5 (Harris et al., 1994). DLBCL (GCB) is typically IgG positive (occasionally IgA positive) whereas DLBCL (ABC) is typically IgM positive (Ruminy et al., 2011).

The majority of the spontaneous lymphomas were of B-cell origin. They were generally mature B-cell neoplasms with increased germinal centre involvement and class switching than the pre-malignant samples analysed.

3.3.1. General Panel

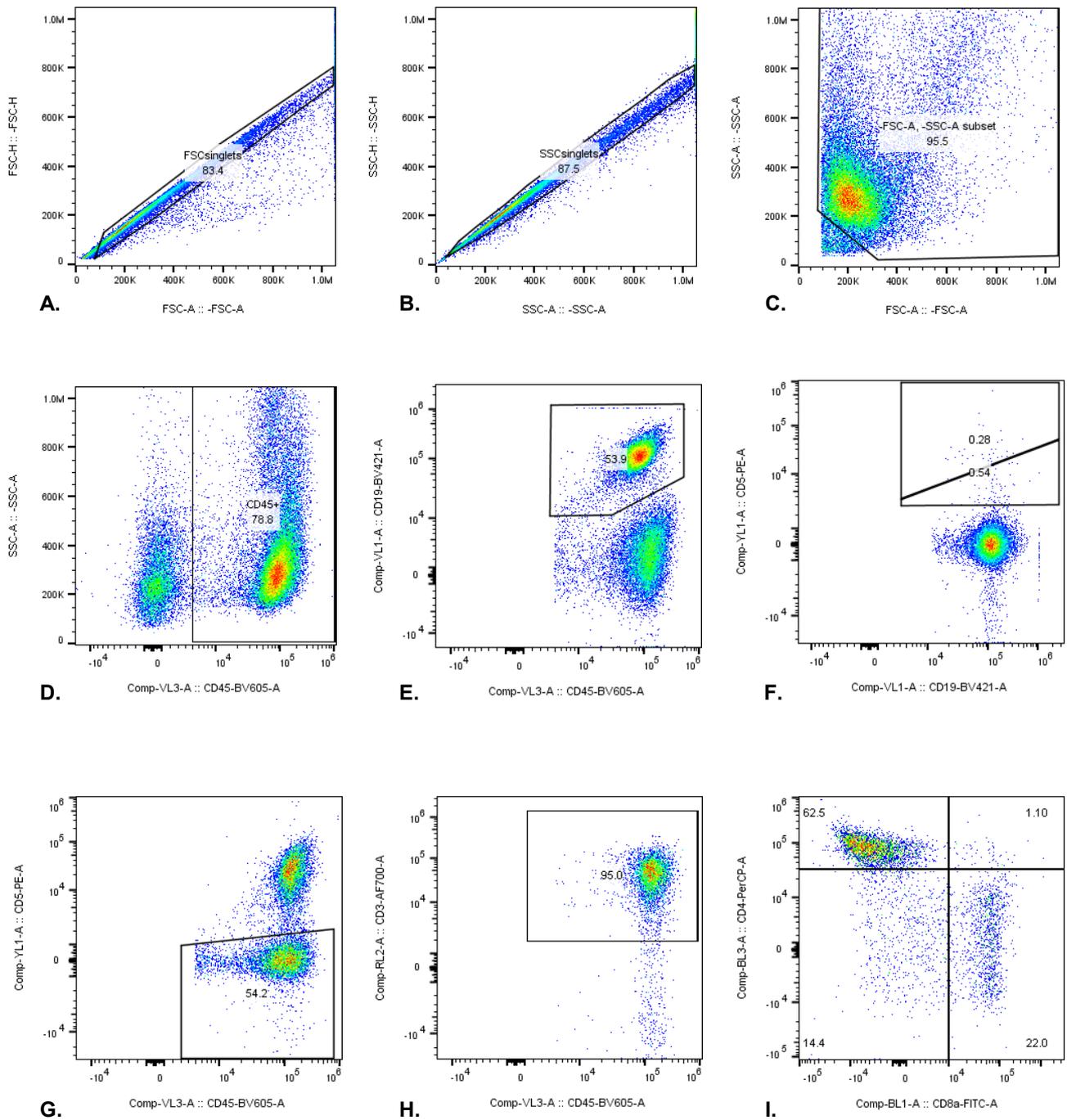
The General panel aimed to distinguish between B- and T-cell lymphoma and to identify major T-cell subtypes. CD45 is a pan lymphocyte marker and CD19 is a B-cell specific marker that is expressed from the earliest precursor cell until the plasmablast stage. It was therefore used as the marker for B-cells defined as CD45⁺ CD19⁺.

The T-cells were characterised by assessing the CD45⁺ CD19⁻ cells for CD5, CD3, CD4 and CD8a expression. CD5 is a surface glycoprotein expressed on all T-cells and a subset of B-cells (B1 and some B-cell malignancies). CD3 is the T-cell receptor. It is expressed at low levels in double negative thymocytes and expression increases in mature T-cells. Early thymic T-cells are double negative for CD4 and CD8a and then mature into single positive cells with low CD3. T-cells then become late thymic double positive before reverting to single positives with normal CD3 expression. T-helper cells are CD4⁺ whereas cytotoxic T-cells and some natural killer and dendritic cells are CD8a⁺ (reviewed in Yui and Rothenberg, 2014).

Figure 3.3. shows the gating strategy for the General Panel. Doublets were excluded by first comparing FCS-H and FSC-A (**A**) and then comparing SSC-H and SSC-A (**B**). Then the cellular debris was removed and the remainder of cells included in the analysis gate (**C**). These samples were frozen and the percentage of living and dead cells as indicated by FCS & SSC was quite variable between samples. Antibody binding in both populations was similar and so both populations were kept in the analysis.

The cells that are not lymphocytes like red blood cells and macrophages were excluded by gating for CD45+ cell (**D**) and then the B-and T-cells were separated by evaluating CD19 status (**E**). The sample in Figure 3.3.**E** contains 53.9% CD45+ CD19+ B-cells. Some B-cells and B-cell neoplasms can be CD5+ and so the CD5 expression was examined in the CD45+ CD19+ cells (**F**). Generally very few of the B-cells expressed CD5, in this example less than 1% and there was no evidence of clonal expansion of this population. As CD5 does not exclusively stain T-cells (**G**) CD3 was additionally used to define the T-cells (**H**). These CD45+ CD19- CD5+ CD3+ T-cell were then assessed for maturity by considering the relative proportions of CD4 and CD8a (**I**). Refer also to Figure 3.7 for an overview of the composition of each suspension.

The tumours had varying proportions of B-cells ranging from 4.8-80.1% however the median, 63.5%, and mean, 61.9%, indicate that these were predominantly B-cell lymphomas. Some typical CD19 plots are shown in Figure 3.4. which is arranged by increasing CD19 positivity. If the majority of the cells are of B-cell origin the tumours was classified as B-cell lymphoma.



Crebbp ^{+fl} ; Aicda ^{+/+} ; VavP-BCL2⁺

Figure 3.3. Spontaneous lymphoma cohort GP gating strategy

Doublet and debris were removed A-C and lymphocytes selected based on CD45 expression D. The B-cells were defined as CD19+ E and their CD5 levels were considered in F. The CD45+CD19- T-cells were categorised as CD5+ G and CD3+ H. These CD3+ CD5+ cells were then assessed for CD4 and CD8a expression I. This example is from tumour #5047 which was Crebbp ^{+fl} ; Aicda ^{+/+} ; VavP-BCL2⁺ and the T-cells are mainly CD8a- CD4+.

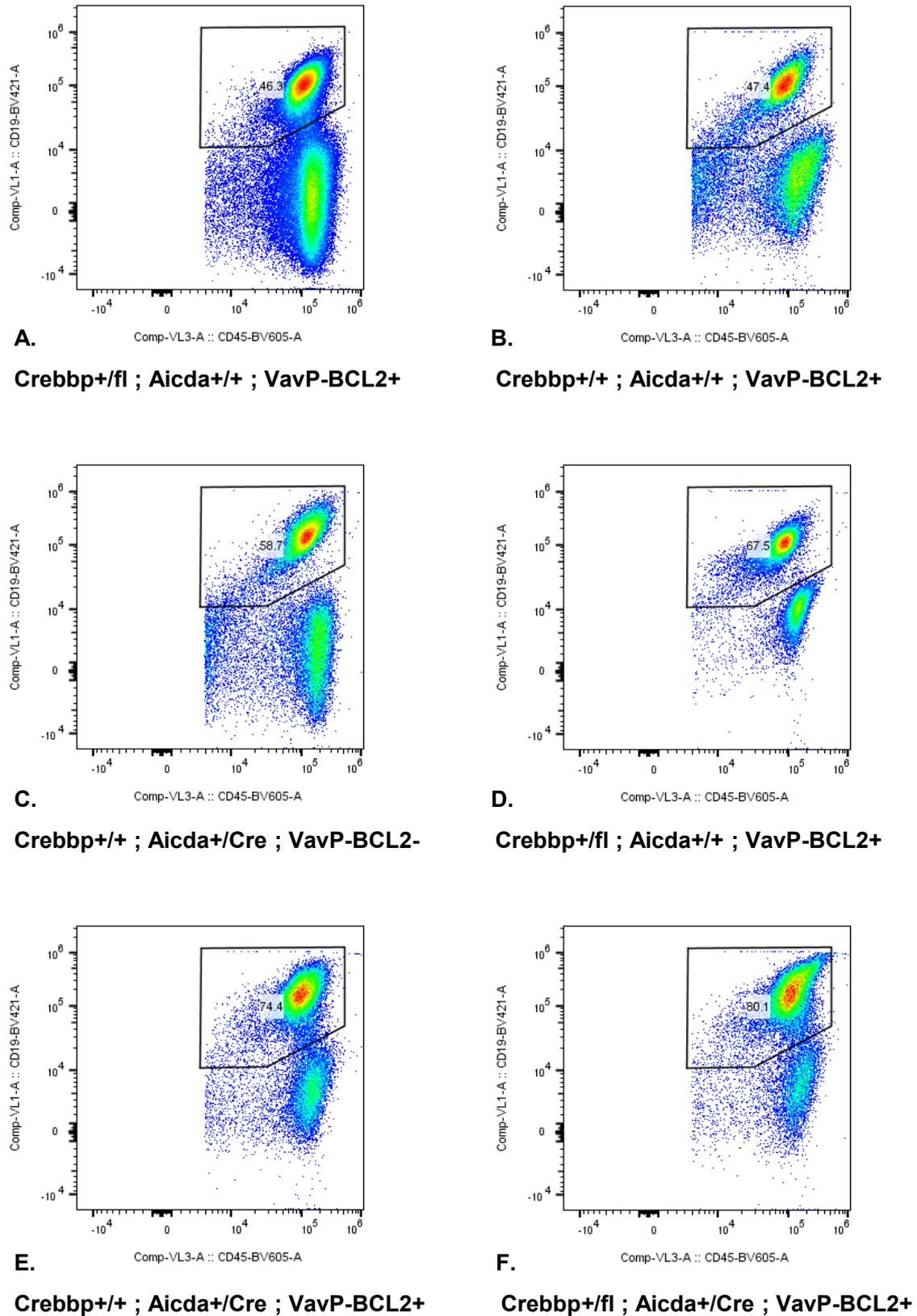


Figure 3.4. Spontaneous lymphoma cohort examples of the varying proportions of B- and T-cells

There was a wide range of B-cell composition, from 4.81%-80.1%. The above figures show some representative CD45 CD19 plots. **A** (#5196) and **B** (#5178) are examples of T-cell lymphoma with low fraction of B-cells, **C** (#5210) and **D** (#5099) show examples with average B-cell proportions and **E** (#5141) and **F** (#5140) show examples with high proportion of B-cell. **C – F** are defined B-cell lymphomas.

The T-lymphocytes (CD45+ CD19- CD3+ CD5+) broadly fell in to one of two CD4 CD8a expression profiles. Figure 3.5.A-C shows tumours with two substantial populations of single positive T-cells (CD4+ CD8a- and CD4- CD8a+) whereas the examples in Figure 3.5.D-F have predominantly CD4+CD8- cells. These single positives match a phenotype of early thymic T-cells, T-helper cells (CD4+), cytotoxic T-cells (CD8a+) or memory T-cells. The lack of early stage T-cells (double negatives or double positives) is suggestive of a mature phenotype indicating that the single positives are T-helper or cytotoxic T-cells having developed from mature double positive cells. This is highlighted in Figure 3.7 where the prevalence of single positives (CD4 + dark red, CD8+ red) rather than double negatives (yellow) or double positives is visible.

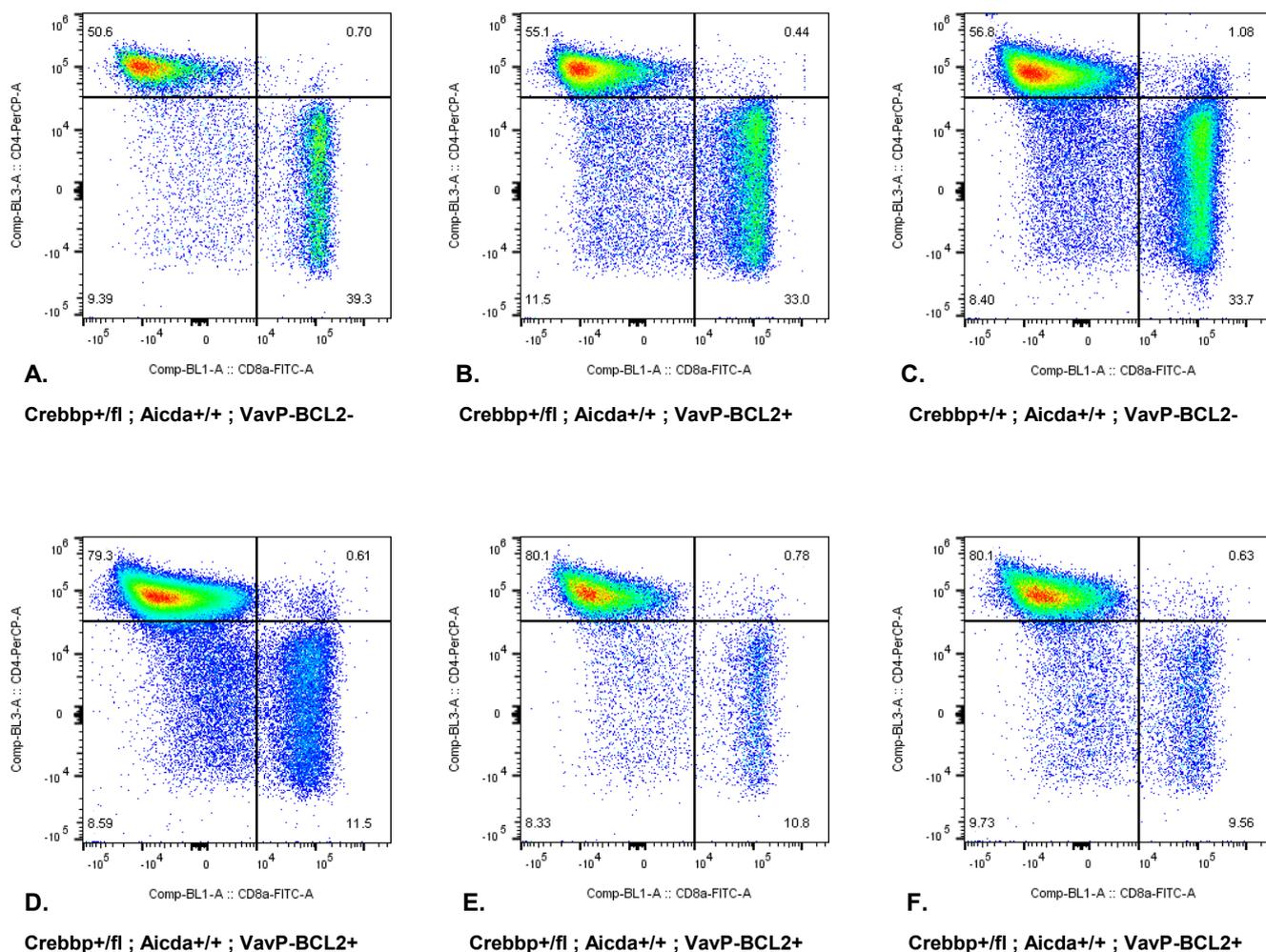
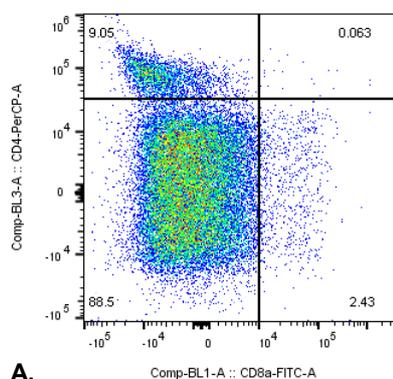


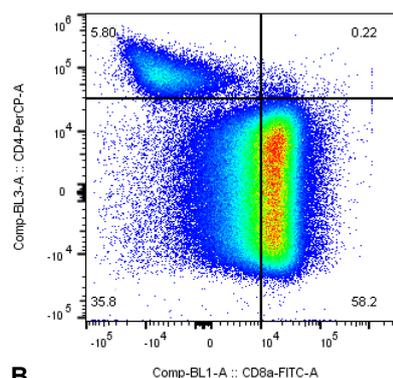
Figure 3.5. Spontaneous lymphoma cohort representative CD4 and CD8a plots

The T-cells (CD45+ CD19- CD3+ CD5+) tended to have both CD8a+ and CD4+ populations of single positive cells A-C (#5183, 5207 and 5099) with the CD4+ being larger population. A-C represent the more common profile. D-F (#5360, 5180 and 5227) had majority CD4+ single positives.

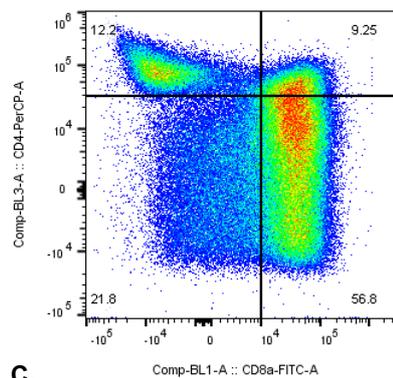
There were a few examples that deviated from these trends and are highlighted in Figure 3.6. The tumours shown in **A** had a significant proportion of double negatives (80.5%) suggesting the tumour originated from very early thymic T-cells. **B** had atypical ratio of single positives with a large proportion of CD8a+ cells (74.4%). Occasionally suspensions had an expansion of double positive T-cells like in **C**.



A.
Crebbp+*fl* ; Aicda+*Cre* ; VavP-BCL2-



B.
Crebbp+*fl* ; Aicda+*Cre* ; VavP-BCL2+



C.
Crebbp+*fl* ; Aicda+*Cre* ; VavP-BCL2-

Figure 3.6. Spontaneous lymphoma cohort GP outliers

The lymphoma had quite similar cytometry profiles however there were some exceptions. The tumour in **A** (#5184) had a majority double negatives. **B** (#5330) has an atypical ratio of single positives with CD8a+ cells predominating. **C** (#5284) had a rare double positive T-cells population.

The tumour heterogeneity can be seen in Figure 3.7. The majority of the tumours are B-cell lymphomas, approximately 85%, whose relative proportions of the T-cell subtypes is quite consistent. The bottom three tumours are T-cell lymphomas with gross expansion and skewing of their T-cell populations.

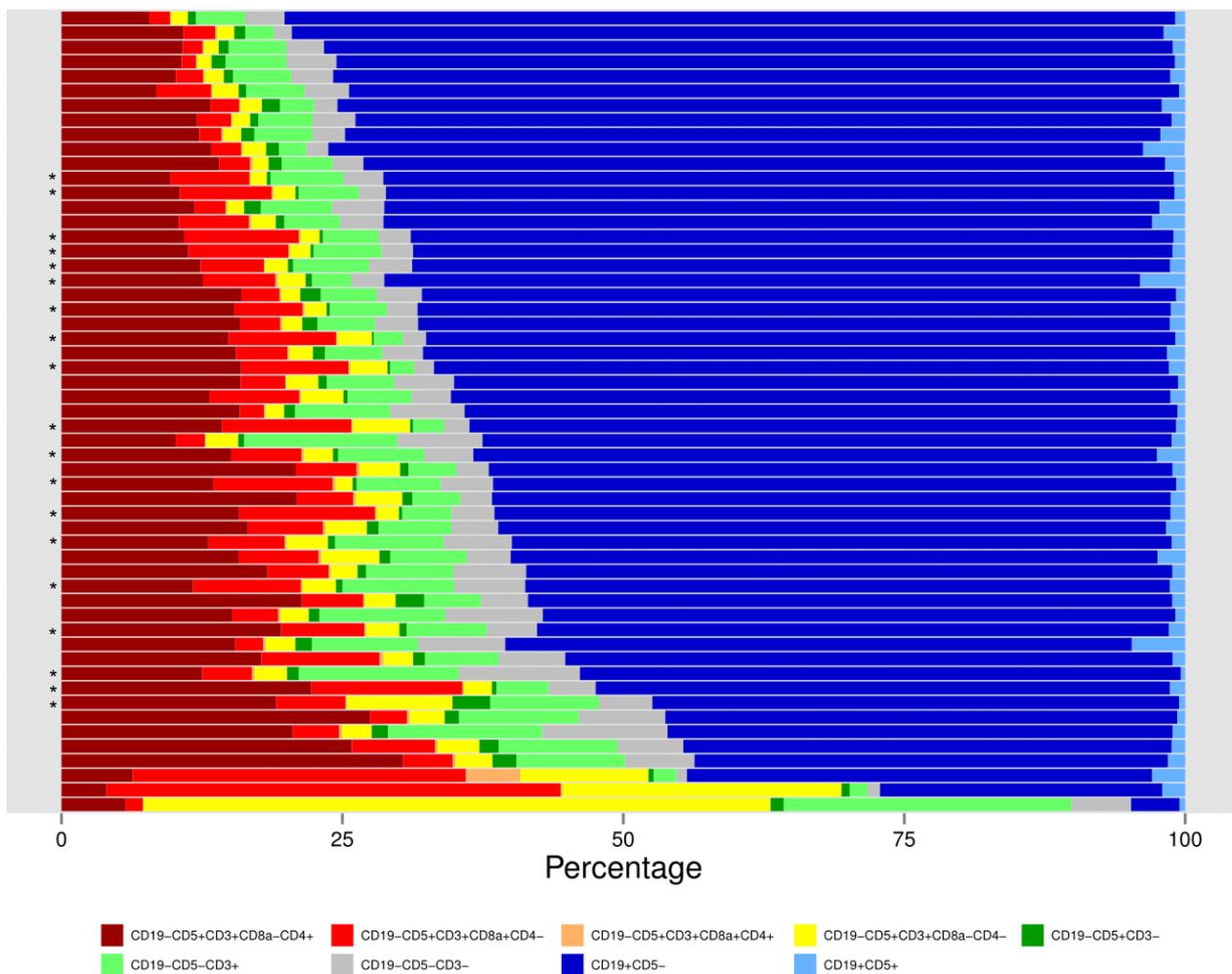


Figure 3.7. Spontaneous lymphoma cohort GP overview of tumours composition

Histogram overview of the CD45+ cells analysed in the General Panel ranked by CD19+. The majority of the tumours are B-cell lymphoma with heterogeneous T-cell composition. The B-cell lymphomas T-cell fractions follow the same trend with CD8a- CD4+ (dark red) being the major portion followed by CD8+ (red), CD3+ (light green), CD5- CD3- (grey), CD4- CD8a- (yellow), CD3- (dark green) and CD8a+ CD4+ (orange). The CD45+ CD19- CD5- CD3- (grey) population is of unknown origin.

The outliers from Figure 3.6. are recognisable as the bottom three rows in this histogram.

Rows marked * are profiles of premalignant animals that did not develop lymphoma.

The tumours were then subdivided by genotype and the B- and T-cell composition assessed in Figure 3.8. All of the genotypes were more susceptible to B-cell lymphoma rather than T-cell with majority blue population. All the genotypes follow the same general trend of CD4+ single positives (dark red) being the predominant T-cell subtype followed by CD8a single positives (red), CD3+ (light green), CD8a- CD4- double negatives (yellow), CD5+ (dark green) and very few CD8a+ CD4 double positives (orange).

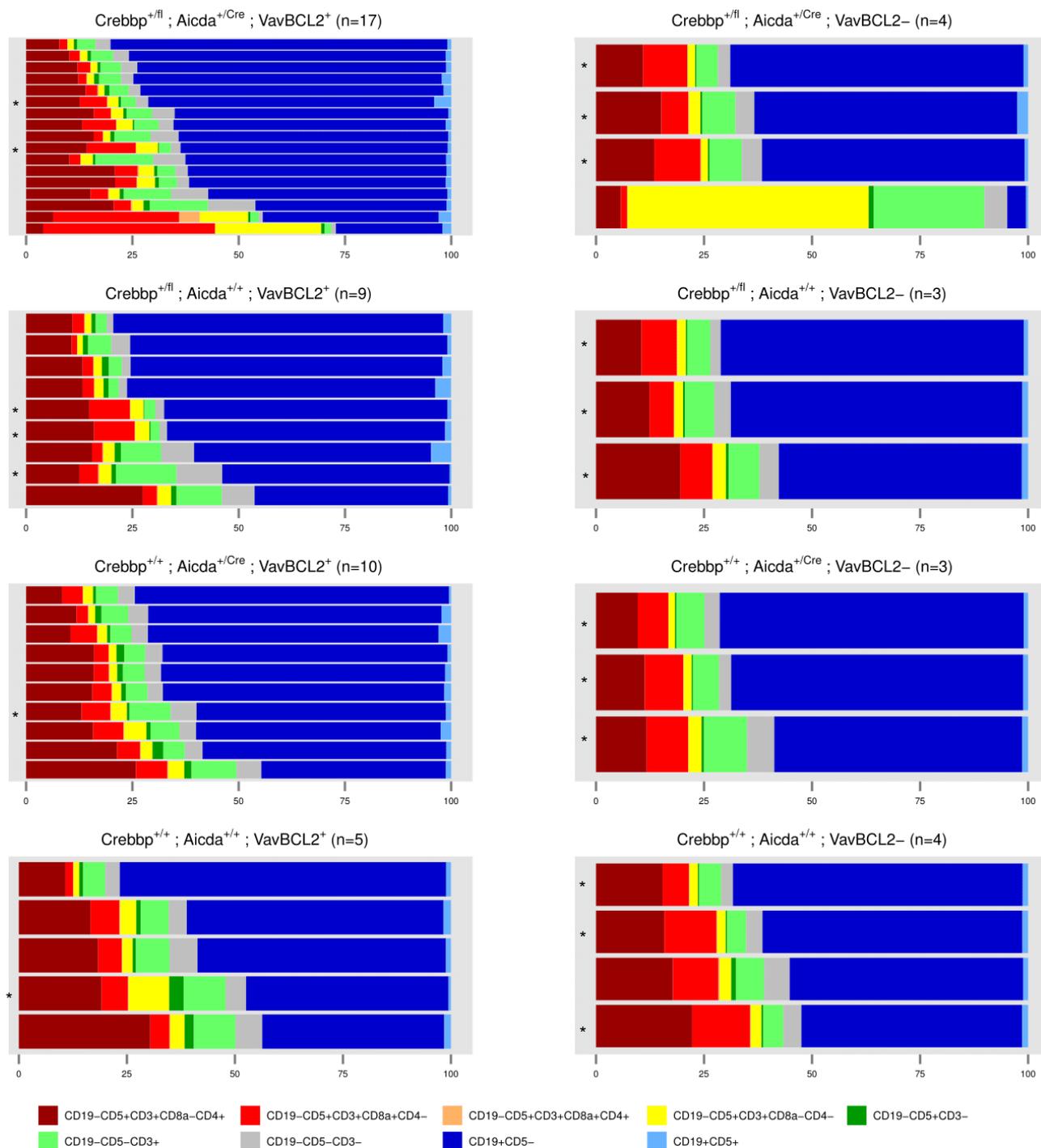


Figure 3.8. Spontaneous lymphoma cohort GP overview of tumours composition by genotype

CD45+ lymphocyte fractions grouped by genotype. The non-transgenic BCL2 genotypes (right hand side) were less susceptible to lymphoma with the majority not developing lymphoma by the end of the experiment (*). They represent the non-pathological, normal ratios of B and T-cells.

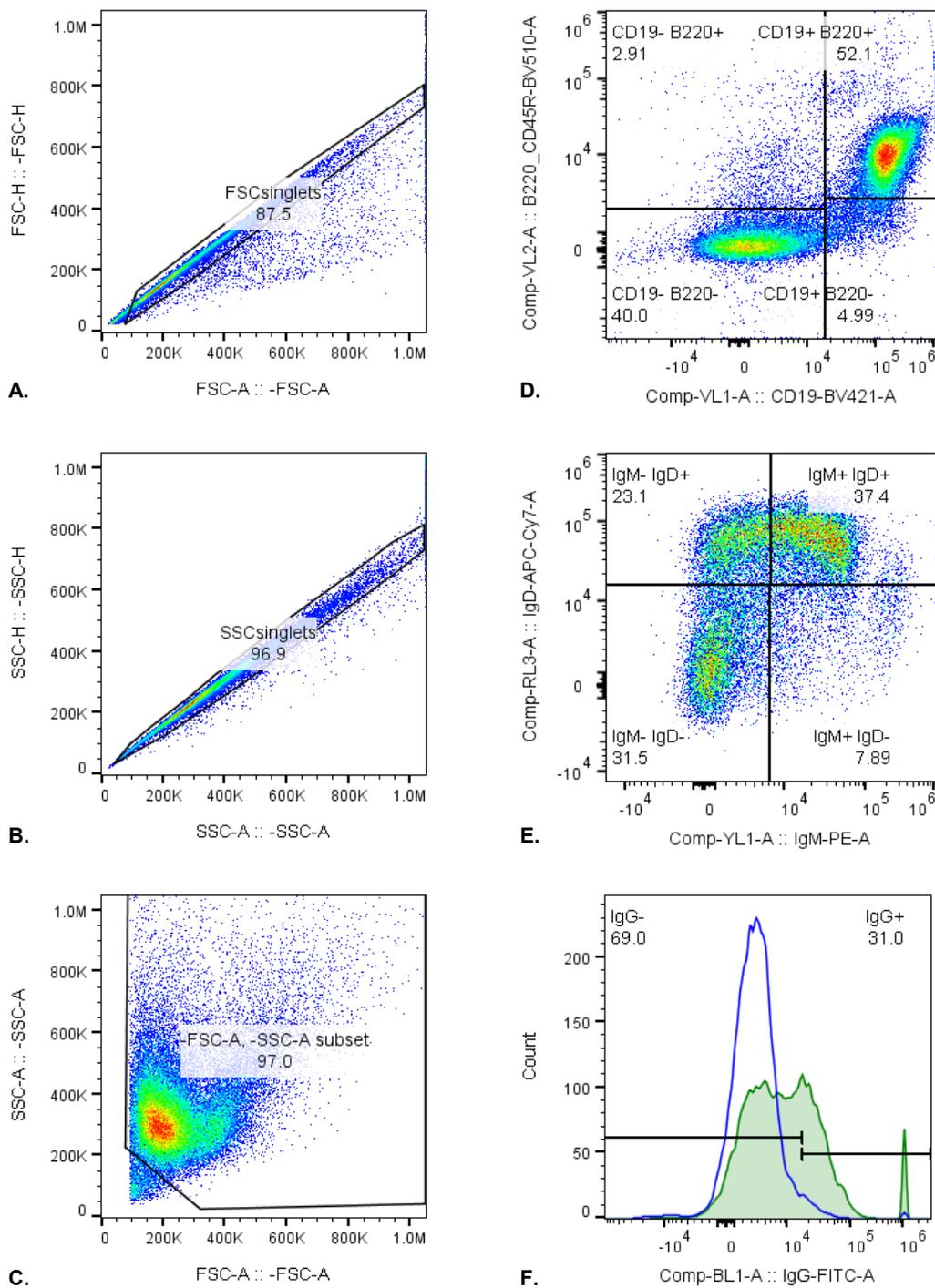
3.3.2. B-cell Panel

The B-cell panel included the pan-B-cell marker B220/CD45R which is present from the pre-pro-B-cell stage, CD19 is also a general B-cell marker present on pro-B-cells until terminal plasma cell differentiation. All four isotypes of murine immunoglobulins (IgG3, IgG2b, IgG2a and IgG1) were stained for and the expression levels summed for an IgG overview. The various combinations of the immunoglobulins can determine the stage of development of the B-cells. IgM is the first immunoglobulin to be expressed from small pre-B-cells, then IgD is expressed on T2 B-cells along with IgM for the first time. IgG expression only occurs after class switching (refer to introductory Figure 1.1, page 3).

Figure 3.9 shows the gating strategy for the B panel. Comparison of B220/CD45R and CD19 demonstrates that this spleen suspension is 2.91% B220/CD45⁺ CD19⁻ and 52.0% B-cells (B220/CD45⁺ CD19⁺). Almost half of the cells in the suspension are lymphocytes (B220/CD45⁺) (Figure 3.9.D).

Of the B-cells (B220/CD45⁺ CD19⁺) 31.5% are IgM⁻ IgD⁻, 7.9% IgM⁺ IgD⁻ (corresponding to small-pre-B-cells, immature B-cells, T1 B and marginal zone B-cells), 37.4% are IgM⁺ IgD⁺ (like T2 B-cells) and 23.1% are IgM⁻ IgD⁺ (like follicular B-cells) (E).

The IgM⁻ IgD⁻ could be infiltrating pre-pro and pro-B-cell from the bone marrow or they could be mature B-cells that have undergone class switching. The IgG levels were then compared between the combined IgM and/or IgD positive cells (blue line) and the IgM⁻ IgD⁻ double negative cells (green line). The IgM and/or IgD immunoglobulin positives do not express IgG as they have not yet undergone class switching unlike the IgM⁻ IgD⁻ IgG⁺ cells (31.0%). The IgM⁻ IgD⁻ cells that do not express IgG are probably IgA or IgE positive (F).



Crebbp fl/+ ; Aicda+/Cre ; VavP-BCL2+

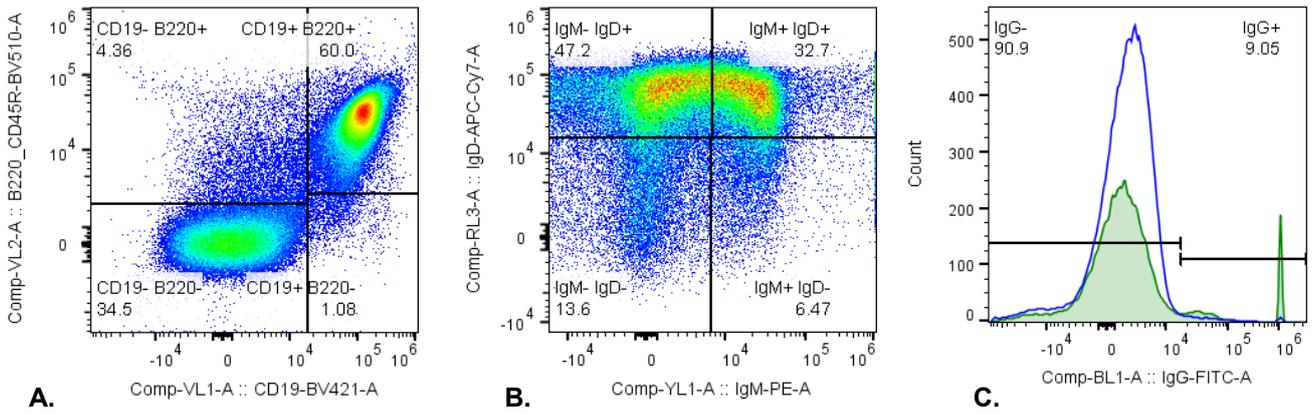
Figure 3.9. Spontaneous lymphoma cohort BP gating strategy

A-C the study population was gated (as described in section 3.3.1) and then the relative quantities of CD19 and B220 were assessed in D. E The predominant B-cell population is IgM+ IgD+ like T2 B-cells. F The IgG levels were then assessed in the immunoglobulin positive cells (IgM+ IgD- and IgM+ IgD+ and IgM- IgD+) (blue line) and the IgM- IgD-double negative cells (green line). The immunoglobulin positives do not express IgG as they have not yet undergone class switching like the double negative cells of which 31.0% are IgG+. The remaining double negative cells that do not express IgG are probably IgA and IgE positive. The gating strategy was set using tumour #5109 which is Crebbp^{fl/fl} ; Aicda^{+ / Cre} ; VavP-BCL2⁺.

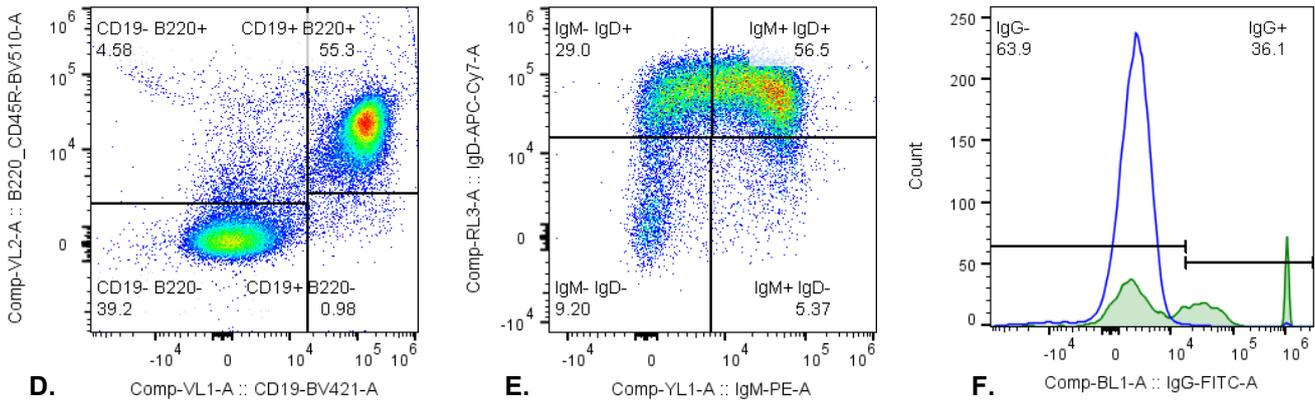
Some exemplar plots from the B-cell Panel are shown in Figure 3.10. These three suspensions have similar proportions in B-cells (**A**, **D** and **G**) however their relative proportions of IgM and IgD are quite different. **B** has 80.0% IgD⁺ cells with the IgM⁺ IgD⁺ cells corresponding to a T2 B-cell phenotype and the IgM⁻ IgD⁺ cells a follicular B-cell phenotype. This lymphoma has very few IgG⁺ class switched cells (9.05%) **C**. **E** Has more IgM⁺ IgD⁺ cells than the previous tumour and 36.1% IgM⁻ IgD⁻ IgG⁺ class switched cells **F**.

Half of **H** is double negative IgM⁻ IgD⁻, of these cells 43.9% are IgG⁺ (green line) **I** indicating that this lymphoma has undergone class switching to IgG, characteristic of a post GC derived tumour. This recapitulates the phenotype of the tumours in VavP-BCL2 mice described by Egle et al. that were IgM negative indicating that the tumours originated from cells that had undergone class switching (Egle et al., 2004).

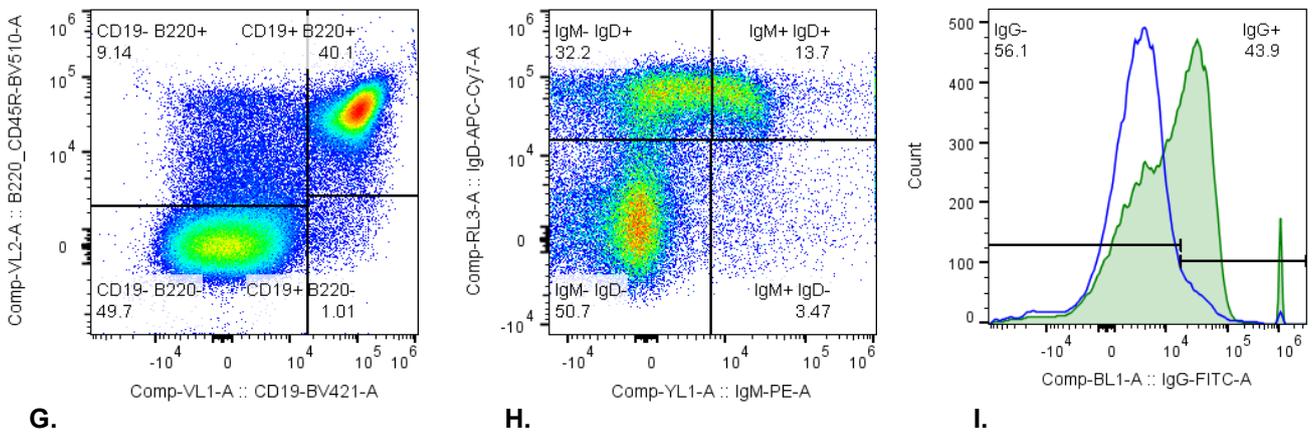
The IgM⁻ IgD⁻ IgG⁻ cells are unlikely to be very early B-cells, prior to BCR expression, which have infiltrated from the bone marrow. The LK panel demonstrates that the vast majority of cells have BCRs (refer to section 3.3.3) so it is more likely that the IgM⁻ IgD⁻ cells have undergone class switching and are secondary immunoglobulin expressing GC B-cells, plasma cells or memory cells.



A. Crebbp+*fl* ; Aicda+/+ ; VavP-BCL2+



D. Crebbp+*fl* ; Aicda+/Cre ; VavP-BCL2+

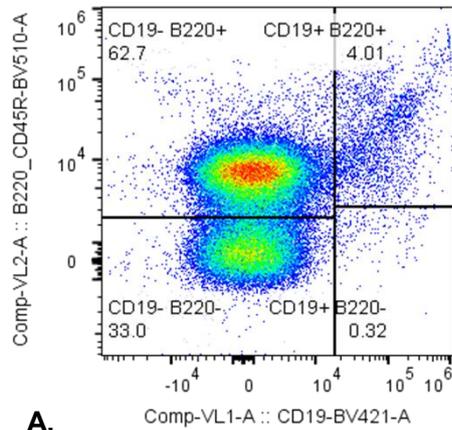


G. Crebbp+*fl* ; Aicda+/+ ; VavP-BCL2+

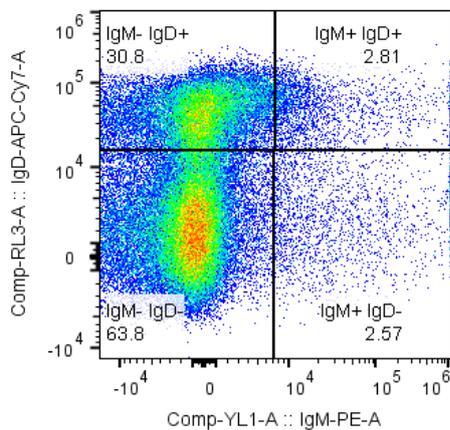
Figure 3.10. Spontaneous lymphoma cohort characteristic immunoglobulin proportions

Example of tumours with varying B- and T-cell proportions and their respective immunoglobulin fraction. The IgG expression is compare in the IgM- IgD- cells (green histogram) compared to that of the other three quadrants (blue). The B-cell tumour in **A** (#5099) is mainly IgD+ with very little IgG **C** indicating that it originates in the follicle but has not yet undergone CSR. The second example has similar IgM and IgD proportions as the first tumour but has more IgG+ cells **E** (#5142) indicating it has a higher fraction of post-C cells. A significant proportion is **H** (#5196) is double negative and 43.9% of these cells express IgG, those that don't likely express.

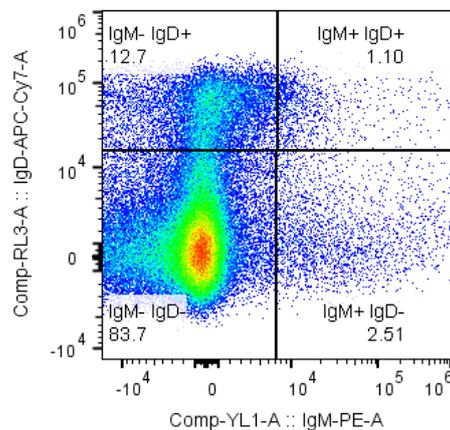
There were some samples that did not conform to the patterns outlined above as shown in Figure 3.11. The tumour in **A** had an unusual B220/CD45 CD19 plot where over half the tumour was B220+CD19- suggesting there is expansion of early stage leukaemic cells or plasma cells. **B** and **C** are composed of IgM- IgD+ cells which represent follicular like cells and class switched IgM- IgD- cells indicating there has been expansion of mature follicle derived B-cell.



A.
Crebbp+/fl ; Aicda+/Cre ; VavP-BCL2-



B.
Crebbp+/+ ; Aicda+/Cre ; VavP-BCL2+



C.
Crebbp+/+ ; Aicda+/+ ; VavP-BCL2+

Figure 3.11. Spontaneous lymphoma cohort abnormal immunoglobulin profiles

Tumour A (#5184) is predominantly B220+ CD19- indicating it originated from very early infiltrating pre-pro-B-cells or long lived plasma cells. **B** (#5191) and **C** (#5227) have high proportions of double negative cells indicating that they arise from mature class switched cells.

When looking in detail at the B-cell populations (Figure 3.12) it becomes apparent that the samples broadly fall into two groups; those with majority blue populations (IgD- IgM+, IgD+ IgM+ and IgD+ IgM-) and those with majority IgD- IgM-, class switched post germinal centre populations (pink). Alteration to the immunoglobulin populations with increased class switched immunoglobulins is correlated with disease.

The bottom third of the histogram is indicative of normal proportions of the various immunoglobulins in premalignant animals (*) where approximately 15% of the B-cells have undergone class switching. The bulk of the cells are transitional B- cells T1, IgM- IgD+ (dark blue) and T2, IgM+ IgD+ (medium blue).

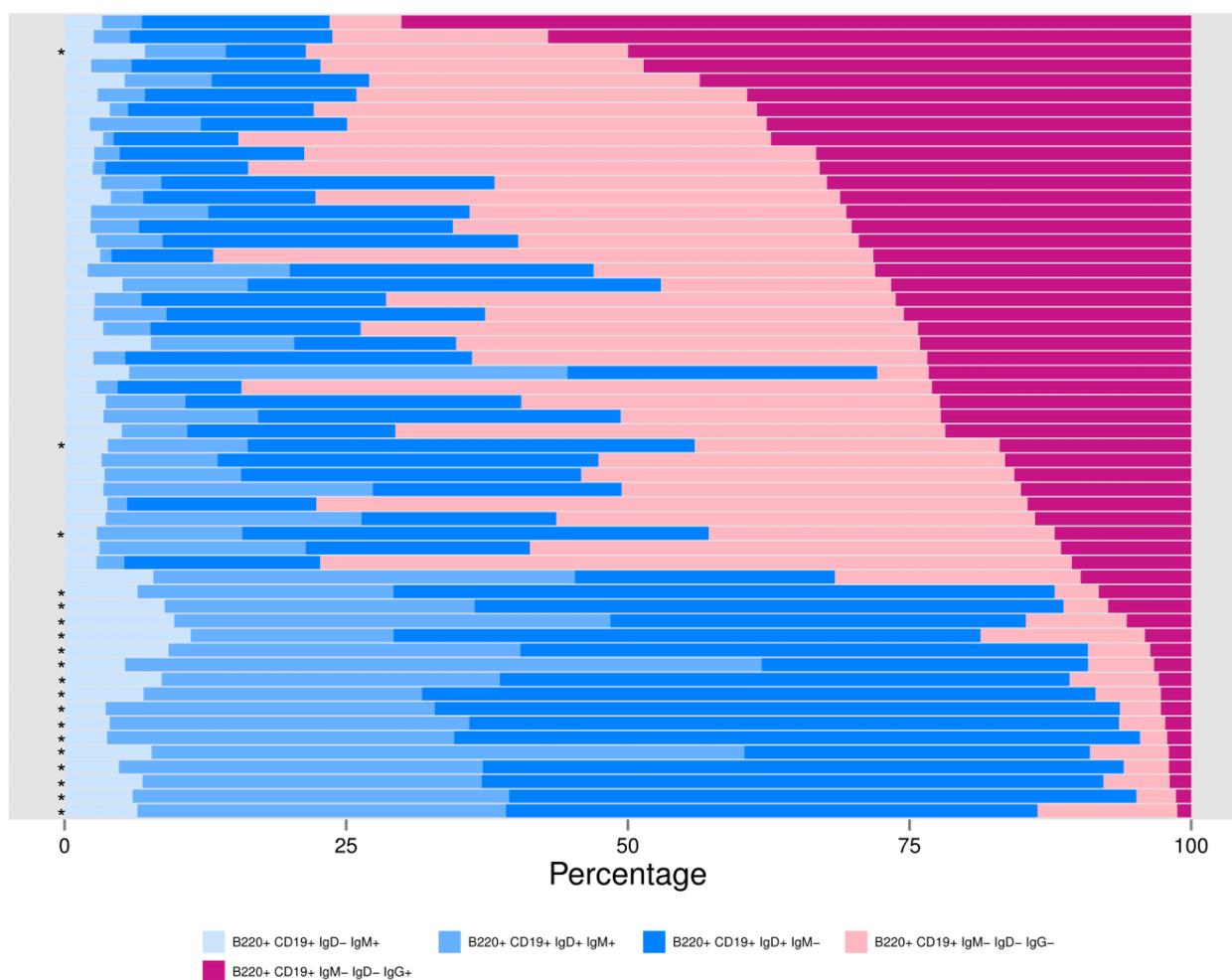


Figure 3.12. Spontaneous lymphoma cohort BP overview of immunoglobulin proportions

When the B-cell fractions were ranked by IgM- IgD- IgG+ they segregate into two groups the larger with majority IgM- IgD- IgG+ (blues) and the smaller with IgM+ and or IgD+ majority (blues); indicating that post GC lymphomas are more common in this cohort. The histogram is ranked by IgM- IgD- IgG+. Rows marked * are of premalignant animals that did not develop lymphoma.

It is evident that pre-malignant animals (*) with no signs of disease have reduced class switching (Figure 3.13). The premalignant mice with abnormally increased class switch cells are *BCL2* transgenic which is known to be a characteristic of the allele from 18 weeks of age even without malignancy (Egle et al., 2004).

The animals with non-transgenic *BCL2* mice (right hand side) that developed lymphoma demonstrate an increase in class switching suggesting that it is correlated with disease. The expansion of class switched cells suggests that their lymphomas are derived from GC or post GC cells.

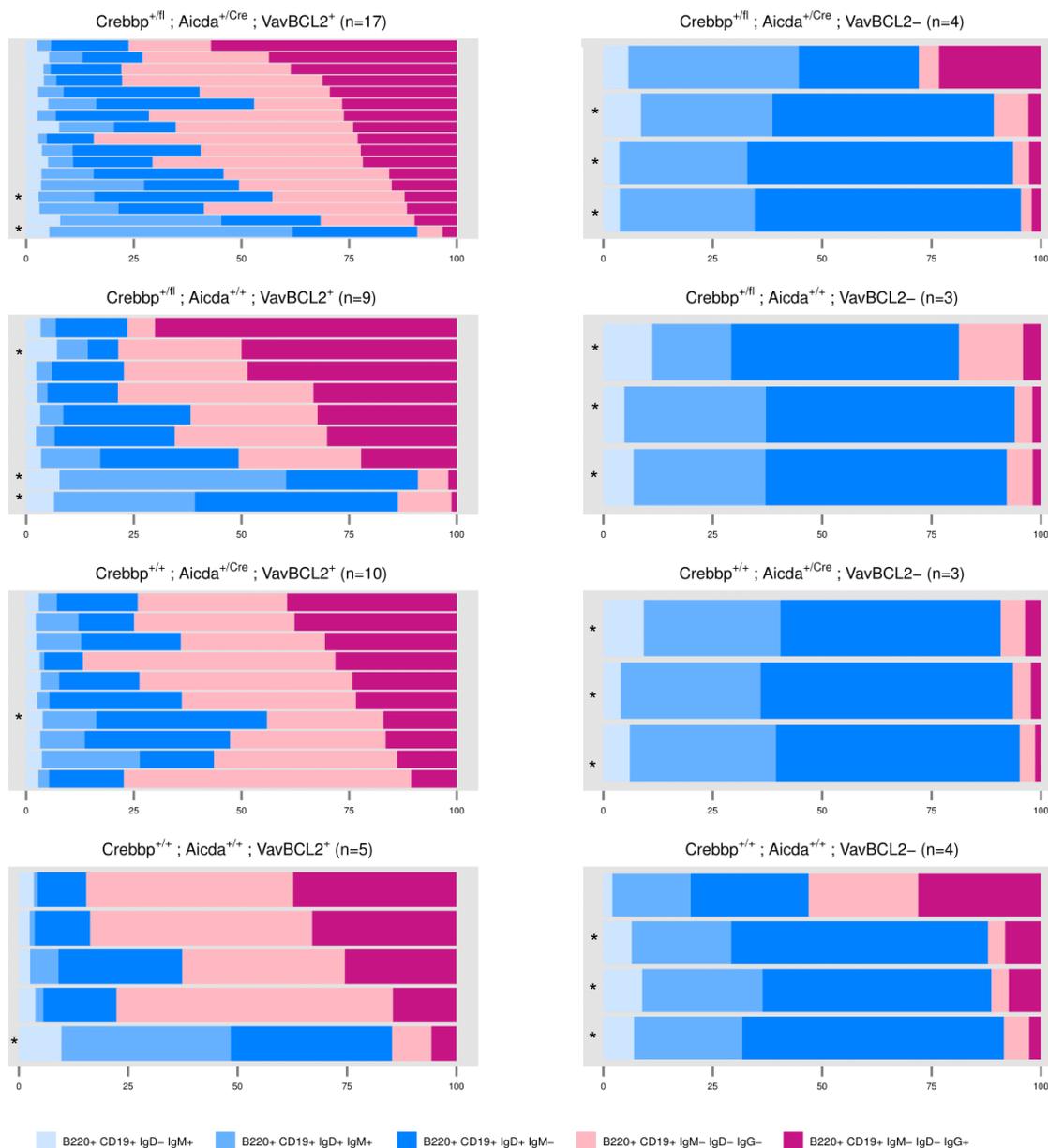


Figure 3.13. Spontaneous lymphoma cohort BP overview of immunoglobulin proportions by genotype

The *BCL2* transgenic mice (left hand column) appear to have more class switched cells (pink populations) whereas the non-transgenic *BCL2* mice (right hand column) have more pre-GC B-cells with IgM and or IgD prior to class switching. Rows marked * are of pre-malignant animals that did not develop lymphoma.

3.3.3. Lambda/Kappa Panel

The Lambda/Kappa Panel aimed to categorise the lymphomas as early or late based on whether the CD19⁺ B-lymphocytes had fully formed BCR, with Lambda and Kappa light chain expression. The BCR is fully formed by the immature B-cell stage, when the cells migrate out of the bone marrow in to the spleen. Lambda/Kappa expression is then maintained throughout maturation.

In addition CD23 and CXCR5 were measured as markers of mature B-cells and of follicular lymphoma and Burkitt's lymphoma respectively. CD23 is a follicular dendritic cell marker whose expression can be used to identify the classic follicular growth pattern of FL. Activated B-cells also express CD23 strongly until isotype switching (Kridel et al., 2012; Thorns et al., 2007). The chemokine receptor CXCR5 was first isolated from Burkitt's lymphoma and was named Burkitt lymphoma receptor 1. It is only expressed by mature recirculating B-cells and a small subset of CD4⁺ CD8⁺ T-cells. CXCR5 recruits circulating naïve B-cells to the follicles and is also thought to be involved in the dissemination of malignant haematopoietic cells (Bürkle et al., 2007; Dürig et al., 2001).

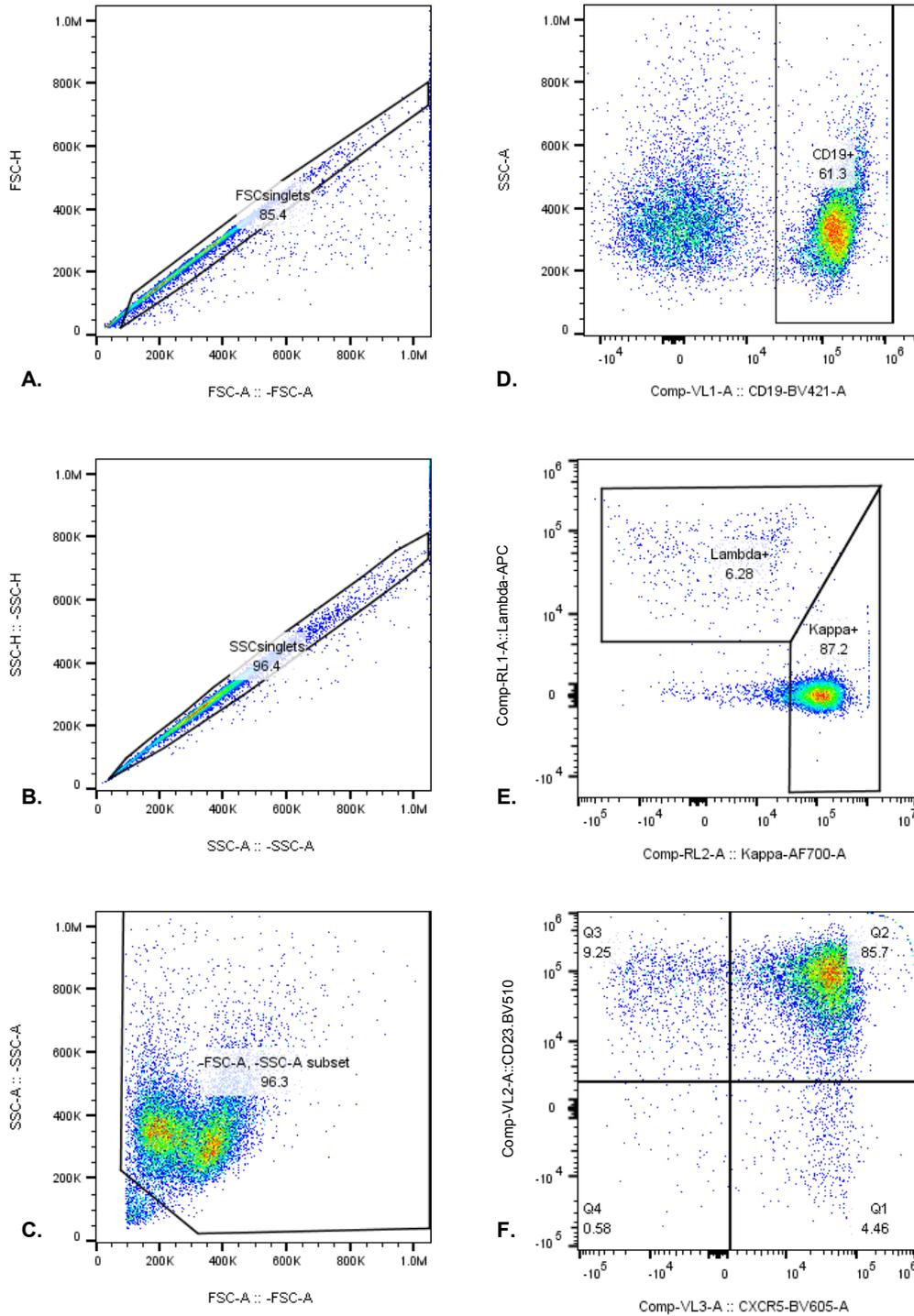
Figure 3.14. details the Lambda/Kappa gating strategy performed on a pre-malignant wild type spleen suspension. Of note, the mean Lambda/Kappa ratio in this sample and other pre-malignant spleens was approximately Lambda 1:15 Kappa.

In humans perturbation of the normal Lambda/Kappa ratio (termed light chain restriction) is indicative of disease however in mice Kappa is far more frequently expressed than Lambda making Kappa restriction simultaneously more likely to occur and more difficult to observe.

The spontaneous tumours broadly maintained the ratio set by the pre-malignant spleen analysis. Figure 3.15. illustrates some variety in the relative proportions of Lambda and Kappa positive cells, samples **D** and **H** have some degree Lambda restriction which could indicate pathology.

The presence of fully formed BCRs indicates that the cells have reached the immature B-cells stage and therefore were able to migrate to the spleen. Only B-cells within the bone marrow are light chain negative and so it would be unlikely for significant amounts of them to be found in the spleen except in the case of pre-B malignancy and infiltrating leukaemia.

The majority of the B-cells expressed CD23 and/or CXCR5 and so it was not a discriminatory parameter (data not shown).



Crebbp^{+/+} ; Aicda^{+/+} ; VavP-BCL2-

Figure 3.14. Spontaneous lymphoma cohort LKP gating strategy

A-C the study population was gated (as described in section 3.3.1) and then the CD19 expression was used to define the B-cells (CD19+) D The proportion of Lambda and Kappa in the B-cells was assessed in E (1:14) as was the CD23 and CXCR5 levels in F This spleen suspension is from a 13 week old mouse (#5357) prior to malignancy, it is Crebbp^{+/+} ; Aicda^{+/+} ; VavP-BCL2⁻.

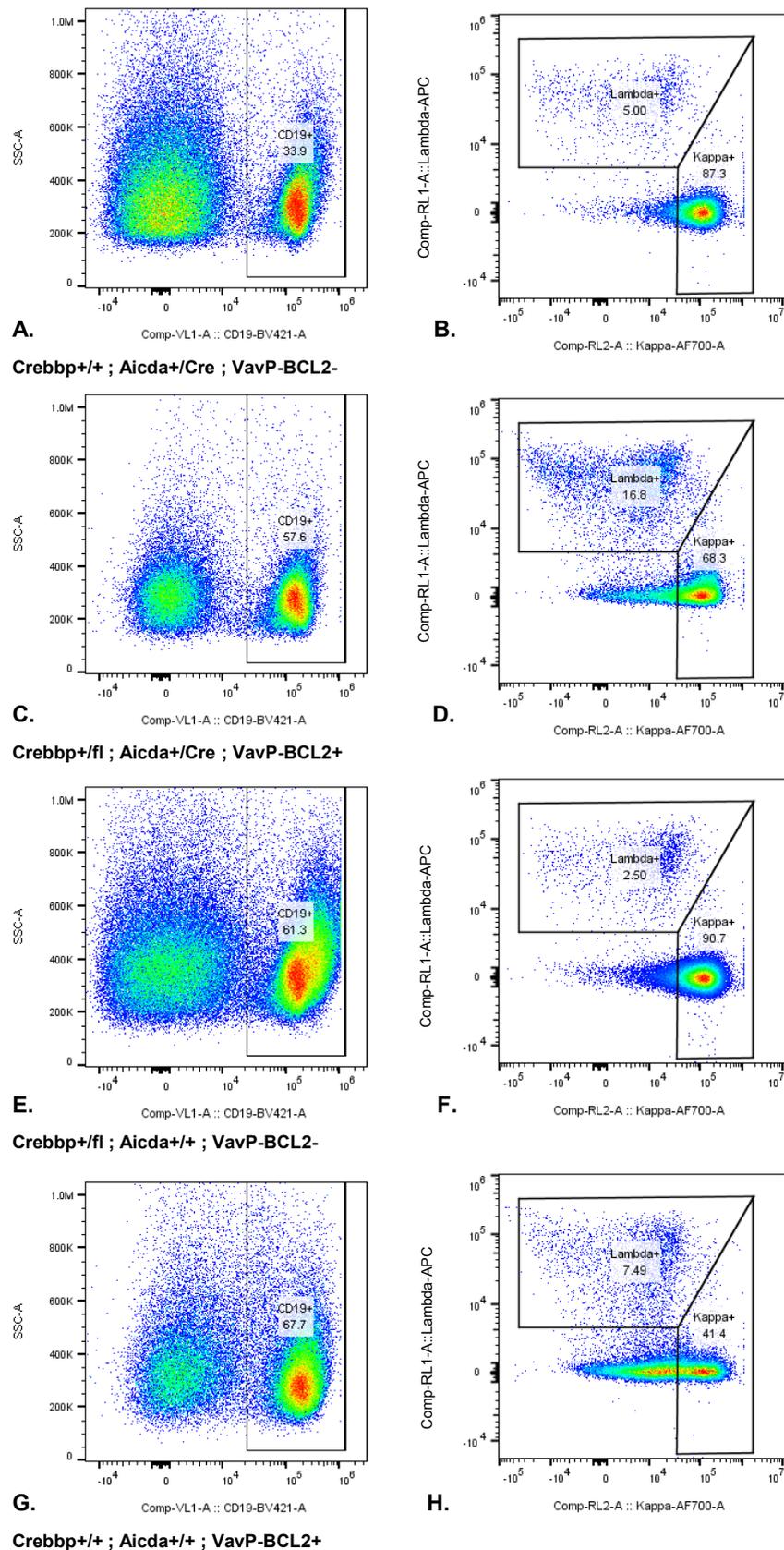


Figure 3.15. Spontaneous lymphoma cohort demonstration of homogeneity of Lambda/Kappa ratios

The left hand column shows spleen suspensions with increasing percentage of B-cells (CD19+) (A, C, E and G) and on the right the Lambda/Kappa ratios are assessed. The BCL2 transgenic genotypes, **D** 1:4 (#5142) and **H** 1:6 (#5141), show some degree Lambda restrictions whereas the non-transgenics broadly maintain a normal ratio, **B** 1:17 (#5210) and **F** 1:36 (#5214). **H** also shows evidence of an expanded population lacking surface immunoglobulin.

3.3.4. Germinal Centre Panel

In the Germinal Centre Panel CD19⁺ B-cells was assessed for CD95 and PNA levels the co-expression of which is characteristic of GC B-cells.

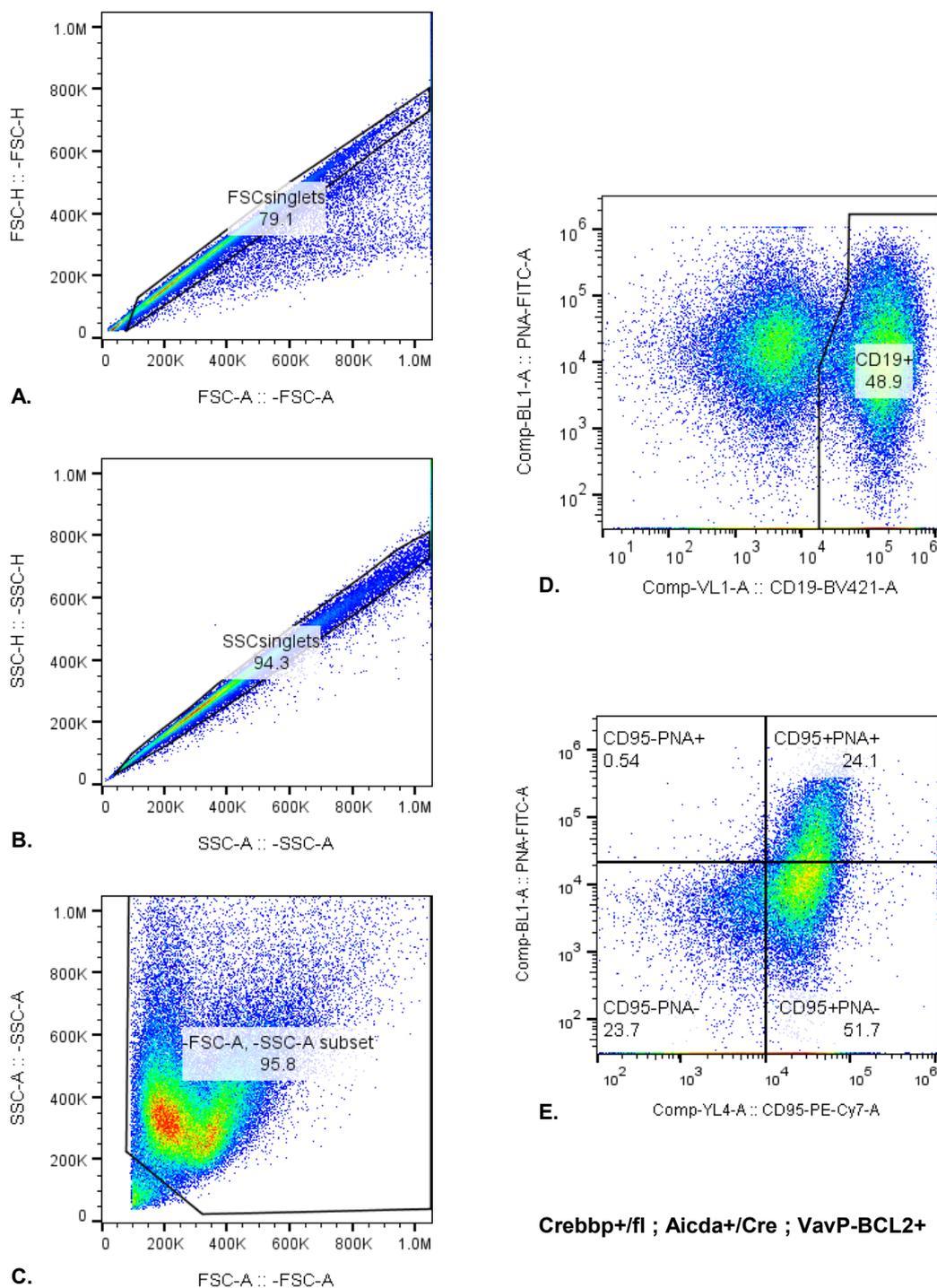


Figure 3.16. Spontaneous lymphoma cohort gating strategy for the GCP

The GC panel included CD19, CD95, PNA and CD86. **A-C** The analysis population was selected as described above in section 3.3.1. **D** The B-cells were then identified as the CD19⁺ population which was then assessed for GC B-cell proportions in **E** CD95⁺PNA⁺ cells are considered to be germinal centre B-cells, this spleen suspension contains 24.1% GC cells. This example is #5361, Crebbp^{+/fl} ; Aicda^{+/Cre} ; VavP-BCL2⁺.

In pre-malignant spleens the majority of B-cells (CD19+) are CD95- PNA- with a small percentage of active GC cells (CD95+ PNA+) as shown in Figure 3.17.A and B Increases of these minor fractions is observed in nearly all diseased animals (C-H).

Control

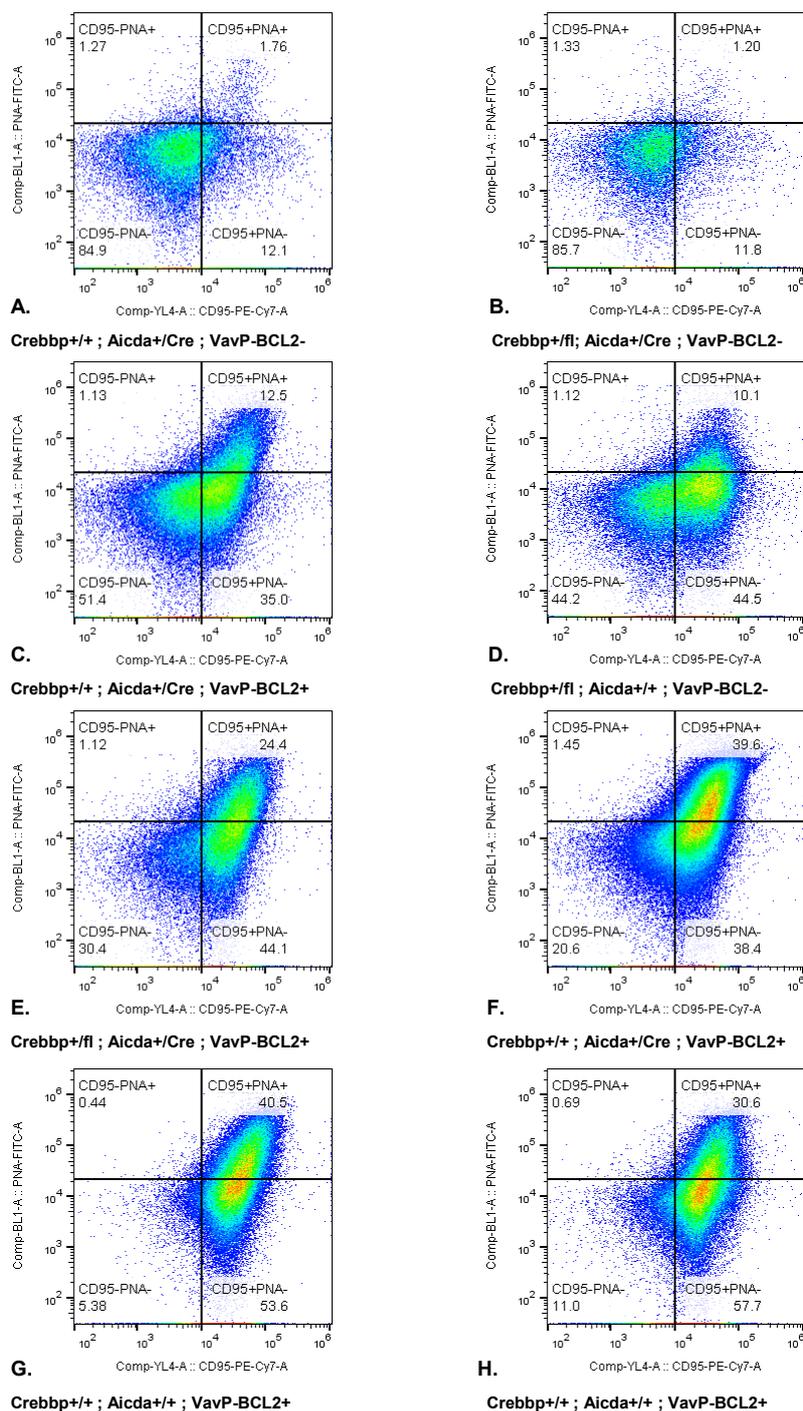


Figure 3.17. Spontaneous lymphoma cohort typical CD95 and PNA expression

The CD95 PNA staining broadly fell into four groups that with increasing double positive demonstrate activation of the splenic germinal centre. Samples **A** (#5208) and **B** (#5204) have predominantly double negative cells indicating that the spleens do not have expanded GC B-cells. These spleens were from aged mice that did not have lymphoma and represent a normal GC profile. The remaining examples are from lymphoma samples with highly disturbed CD95 and PNA **C** (#5303) and **D** (#5216) still have a significant proportion of double negatives however there are increasing CD95+ cells. Sample **C** had no signs of lymphoid hyperplasia upon necropsy and was considered pre-malignant, however the increase in CD95+ PNA+ cells is indicative of some transformation. **E** (#5330) and **F** (#5218) have three populations of double negatives, double positives and CD95+ PNA-. All the cells in **G** (#5171) and **H** (#5227) are CD95+ with less than half also binding to PNA.

The CD95 and PNA profiles segregate into two groups; those in the bottom third of Figure 3.18 are predominantly CD96- PNA- indicating that there is no significant expansion of GC B-cells, this coincides with the pre-malignant animals and can be used as an indication of normal GC B-cell proportion. The upper two third of the samples have greatly expanded CD95+ PNA- and CD95+ PNA+ populations highlighting expansion of GC-B-cells and increased proliferation. Alteration of the proportion of GC cells is strongly correlated with lymphoma.

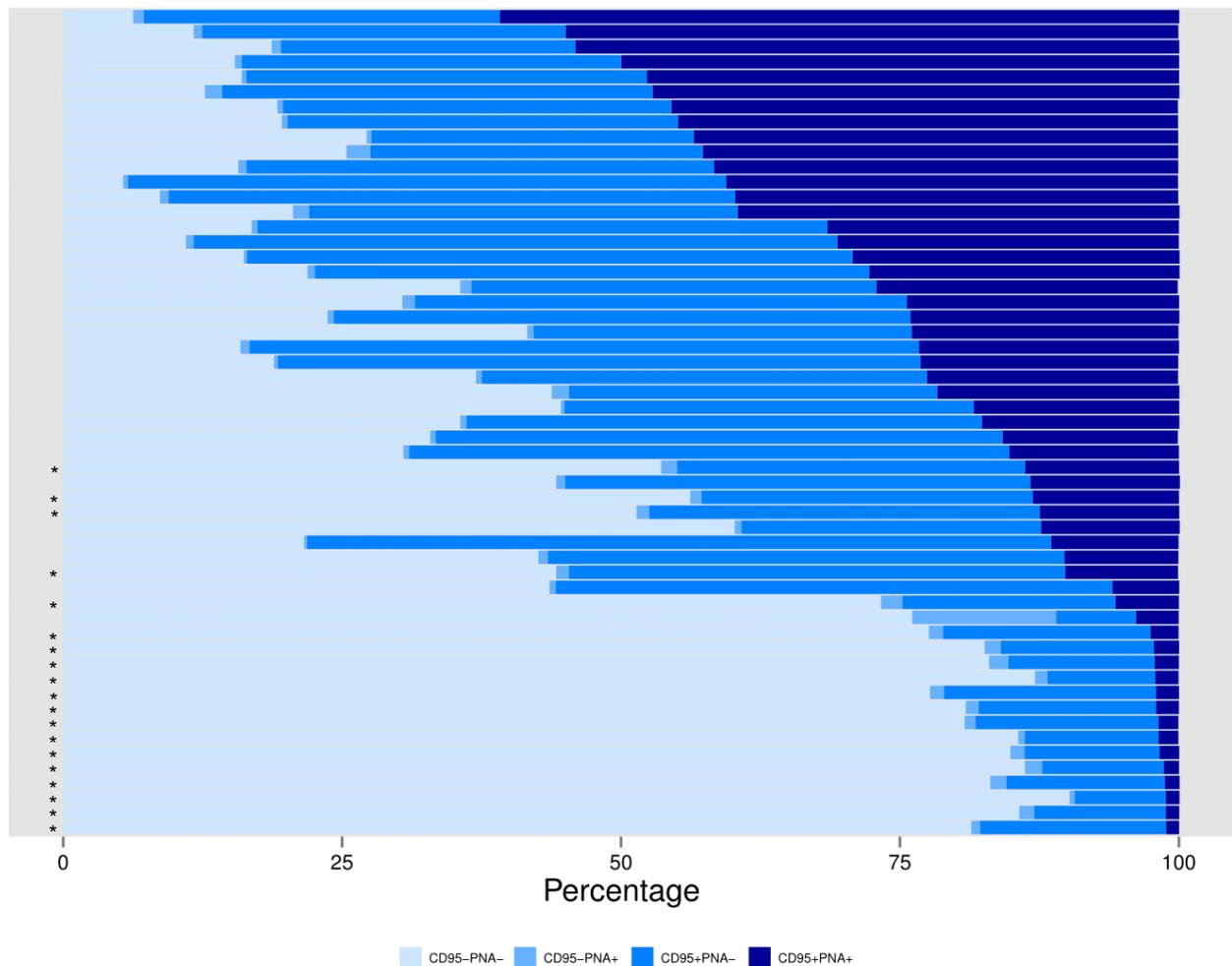


Figure 3.18. Spontaneous lymphoma cohort GCP overview of tumours composition

The histogram overview of B-cell (CD19+) fraction shows a clear delineation between the premalignant mice (bottom third) with normal C95 and PNA expression and the mice with abnormal GC activation. Rows marked * are of premalignant animals that did not develop lymphoma.

When CD95 and PNA fractions are grouped by genotype (Figure 3.19) it becomes evident that the *BCL2* transgenic spleens have more expanded GC populations. Egle and colleagues noted that by 18 weeks of age the VavP-BCL2 mice had enlarged germinal centres, compared to the E μ -BCL2 mice, which stained strongly for PNA. These PNA+ cells were IgM- indicating that they had undergone class switching (Egle et al., 2004).

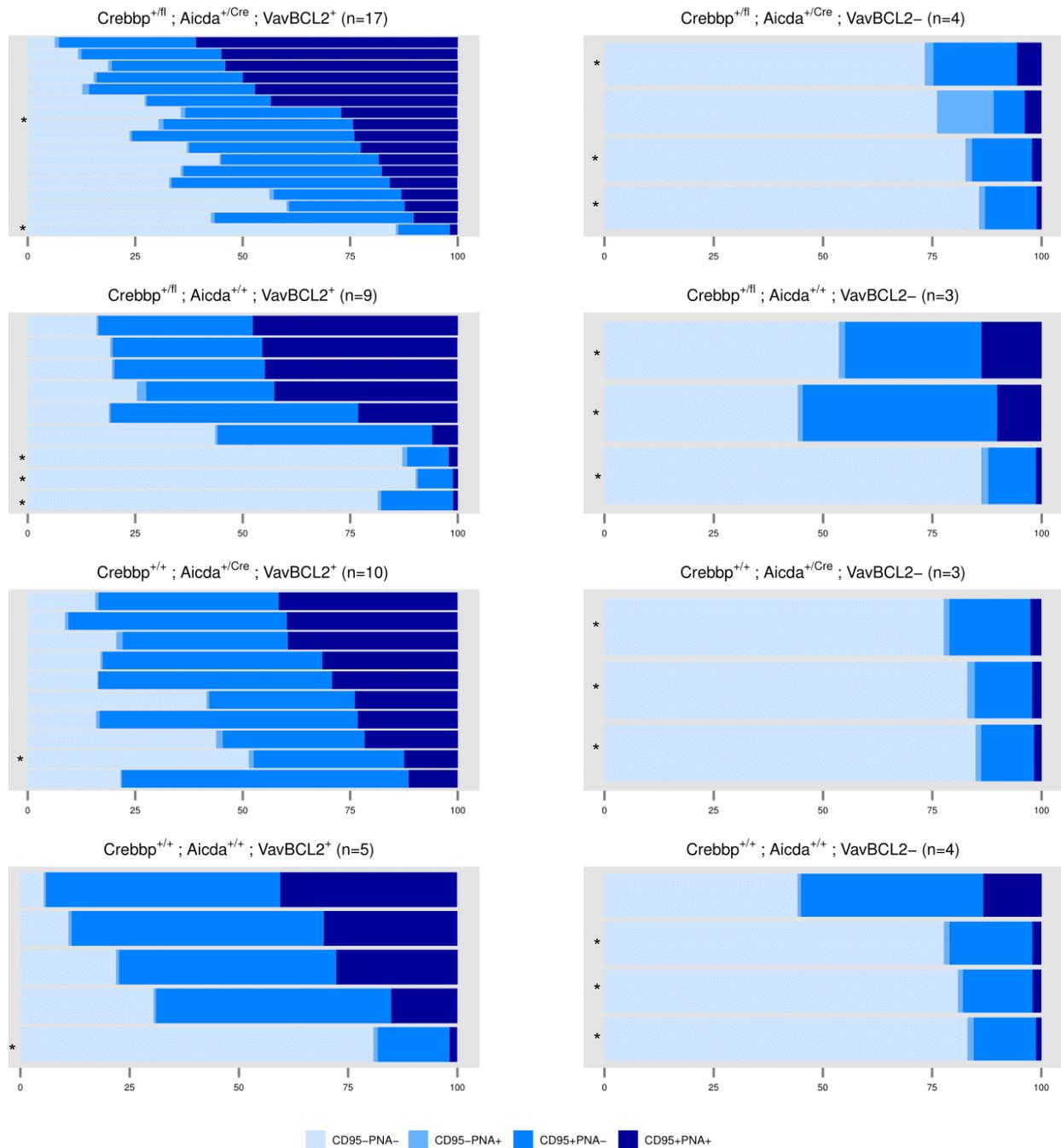


Figure 3.19. Spontaneous lymphoma cohort GCP overview of tumours composition by genotype

Division of the Germinal Centre Panel by genotype highlights that the *BCL2* overexpressing mice (left hand column) have very abnormal GC marker expression and also draws attention to the improved survival and normal GCs of mice with wild type *BCL2* levels (right hand column). Rows marked * are of premalignant animals that did not develop lymphoma.

3.4. Summary of the spontaneous lymphoma

- Conditional loss of *Crebbp* conjunction with overexpression of *BCL2* significantly accelerates lymphomagenesis corroborating human sequencing data that indicated that there are frequently co-mutated. *Crebbp* deficiency alone does not accelerate tumour onset.
- This is the first time that *Crebbp* has been shown experimentally to cooperate with *BCL2* overexpression acting as an haploinsufficient tumour suppressor gene B-cell lymphoma.
- The majority of the spontaneous tumours were B-cell lymphomas. They demonstrated expansion of the GC-B-cells, an increase in proliferative cells and increased class switching. This indicated that the lymphomas were predominantly germinal centre or post germinal centre derived.
- This novel model faithfully recreates the general immunophenotype of mature B-cell lymphomas. Further immunochartersation of the model, to demonstrate which subtype of B-NHL is most closely recapitulates, could yield a useful model for this disease

4. Retrovirus induced lymphoma

Neonatal infection of mice MoMuLV leads to accelerated lymphoma onset with 100% incidence by twelve months in wild type mice. The addition of sensitising mutations can accelerate this even further to varying degrees depending on the aberration. MoMuLV induced tumours have increased clonal complexity and lymphoid dissemination compared to spontaneous tumours (Huser et al., 2014).

MoMuLV induced somatic mutations which can cooperate with or be mutually exclusive to the sensitising germline mutations and the resulting genotype specificity of these mutations highlights genes that collaborate to drive lymphomagenesis.

In order to investigate selection for mutations that cooperate with overexpression of *BCL2* and or loss of *Crebbp*, 249 mice of the eight previously mentioned genotype combinations were infected with MoMuLV perinatally and monitored until disease onset. The resulting tumours were characterised by flow cytometry and then the mutations that drove oncogenesis are described in Chapter 5.

4.1. Quantification of MoMuLV

MoMuLV virus was produced in 293T cells and the titre of the supernatant was quantified by infecting 293mCAT1 engineered to express a reporter gene when infected with a retrovirus. These cells are human 293T cells that have been modified to express the mCAT1 receptor, allowing infection by MoMuLV. Upon infection the virus is reverse transcribed by its *pol* protein, generating a DNA provirus. This process is unique to the retroviral infection cycle and enables integration of the provirus into the host cell's genome. This intrinsic reverse transcription capability is harnessed to give a read out of virus concentration in the inGluc-MLV-DERSE reporter assay where the luciferase reporter, present in the reverse orientation, is only produced and secreted into the media if retrovirus retrotranscriptase is present. The luminescence measured in the supernatant of the GLuc expressing MCAT cells is proportional to the amount of reverse transcriptase produced and in turn the amount of MoMuLV (Aloia et al., 2013).

MoMuLV quantity increase with time after a short lag period of approximately four days whilst the virus produces reverse transcriptase and integrates. By day four the 1/5 dilution of MoMuLV induced detectable luciferase protein. Luciferase emission increases with time proportionally to the increase in MoMuLV particles (Figure 4.1.A).

Four virus batches were produced in the lab, their relative concentration ten days after infection of the mCATs is compared in Figure 4.1.B MoMuLV JD and MoMuLV JK are of similar concentration, MoMuLV PW1 is approximately 30% the concentration and MoMuLV JD and MoMuLV PW2 is 60%.

MoMuLV PW1 had been used in a previously completed screen and so this assay confirmed that MoMuLV JD, used in the IM screen in this thesis, was functional and sufficiently concentrated to cause disease.

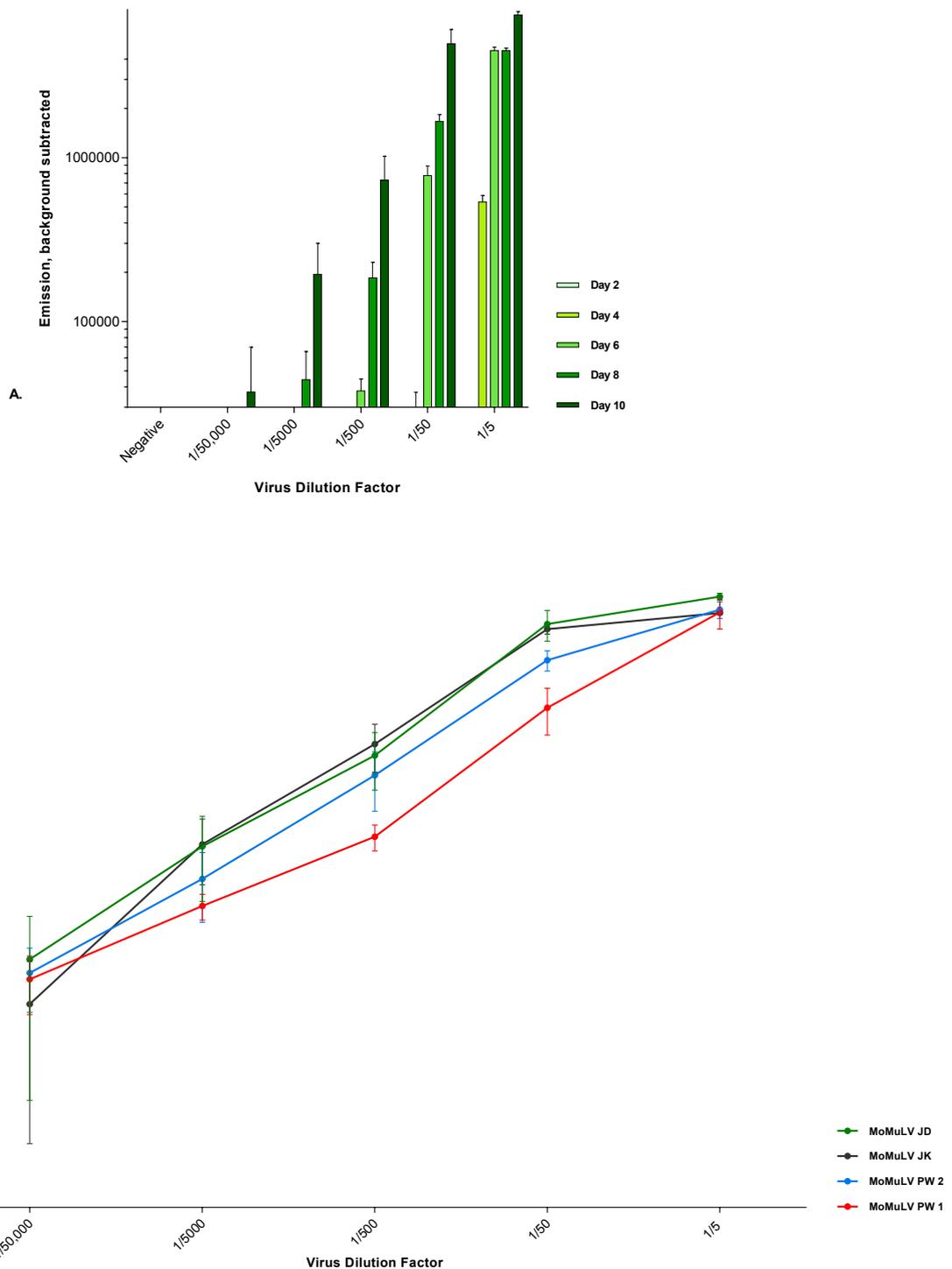


Figure 4.1. Quantification of MoMuLV

MoMuLV was quantified by serially diluting it and infecting 293mCAT1 cells, which express mCAT1 receptors and a reverse orientation GaussiaLuciferase (GLuc) reporter gene. Replicating MoMuLV reverse transcribes the reporter producing luciferase protein which is secreted exogenously. **A** Virus supernatant was collected at days 2, 4, 6, 8 and 10 days. By day 4 the virus has integrated into the cells and produced the translational machinery necessary to produce the reporter. Virus input increases with time indicating that there are more virus particles integrating in the cells. **B** The experiment was repeated with the four virus batches produced in the lab and sampled on day 10. MoMuLV JD and MoMuLV JK are of similar concentrations, MoMuLV PW 2 has 60% the activity of MoMuLV JD and MoMuLV PW 1 has only 30%. MoMuLV JD was used in the screen described in this thesis. The levelling off of the graph indicates that the system is saturated and that after 10 days the cells are approaching maximum reporter expression.

4.2. MoMuLV induced lymphoma cohort

BCL2 transgenic mice are known to be predisposed to follicular B-cell lymphoma with cumulative incidence of 37-50% by 18 months (Egle et al., 2004); inoculation with MoMuLV accelerates this.

Of the MoMuLV infected mice, all except five died of lymphoma within a year. Four mice were culled and censored for symptoms other than lymphoma (specifically seizure, blocked gastrointestinal tract and prolapse). One CrebbpCre mouse survived to 424 days without succumbing to any disease and so was culled and censored from the experiment. When the spleen DNA from this mouse was sequenced it transpired that it had no retroviral inserts so it seems likely that it was not inoculated with MoMuLV perinatally (see Chapter 5).

The vast majority of mice presented with enlarged spleens, often in conjunction with enlarged lymph nodes, which never occurred in isolation. The average spleen weight was 0.99g, i.e. greater than five times larger than wild type pre-malignant animals. Thirty two mice had enlarged thymi without spleen involvement, they tended to have non-transgenic *BCL2* genotypes. These mice had irregular breathing and deteriorated rapidly (see Figure 4.2. for example of pathology).

		MoMuLV induced lymphoma				
		Male	Female	TOTAL	MS, days	Cause of death
Crebbp^{+fl} ; Aicda^{+Cre} ; VavP-BCL2⁺	CrebbpCreBCL2	25	31	56	104	55 lymphoma, 1 censor
Crebbp^{+fl} ; Aicda^{+/+} ; VavP-BCL2⁺	CrebbpBCL2	10	18	28	107	all lymphoma
Crebbp^{+/+} ; Aicda^{+Cre} ; VavP-BCL2⁺	CreBCL2	8	13	21	91	all lymphoma
Crebbp^{+/+} ; Aicda^{+/+} ; VavP-BCL2⁺	BCL2	8	9	17	115	all lymphoma
Crebbp^{+fl} ; Aicda^{+Cre} ; VavP-BCL2⁻	CrebbpCre	28	26	54	134	1 symptom free, 1 censor, 52 lymphoma
Crebbp^{+fl} ; Aicda^{+/+} ; VavP-BCL2⁻	Crebbp	10	14	24	167	22 lymphoma, 2 censor
Crebbp^{+/+} ; Aicda^{+Cre} ; VavP-BCL2⁻	Cre	8	12	20	153	all lymphoma
Crebbp^{+/+} ; Aicda^{+/+} ; VavP-BCL2⁻	Wild type	15	14	29	110	all lymphoma
TOTAL				249		

Table 4.1. MoMuLV induced lymphoma cohort

Summary of the mice included in the retrovirus induced lymphoma cohort subdivided by genotype and gender with details of their median survival (MS) and causes of death. Only one mouse reached 400 days symptom free and was censored. Animals that died from reason other than lymphoma e.g. seizure, blocked gastrointestinal tract and prolapse, were also censored.

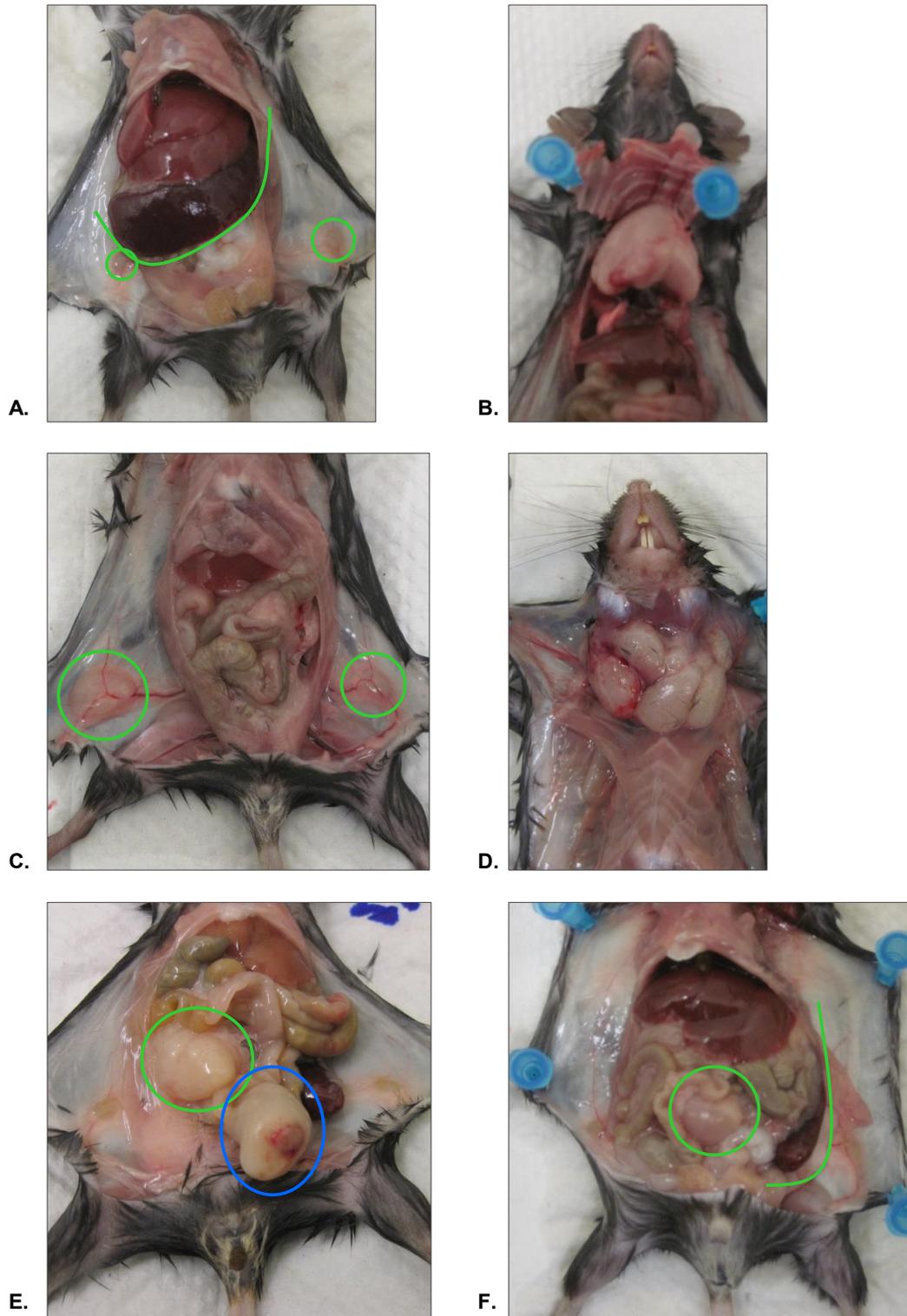


Figure 4.2. Examples of mouse pathology

The majority of the MoMuLV infected mice presented with enlarged spleens often in conjunction with enlarged lymph nodes and others had enlarged thymi. Enlarged lymph nodes were never the primary pathology. **A** #5380 had slightly enlarged inguinal lymph nodes and severe splenomegaly (spleen weight 1.9g). **B** #5332 presented with breathing difficulties and had a very enlarged thymus and a moderate spleen (0.63g). **C- F** demonstrate a variety of lymph nodes enlargements. **C** and **D** (#5240) show the enlarged inguinal lymph nodes and significant cervical lymph node enlargement in a mouse which were palpable and visible externally **D**. **E** (#5194) is an example of a uninfected mouse that had an enlarged spleen but also an enlarged mesenteric lymph node (green) and a gut-associated lymph node (blue) that had ulcerated. **F** (#5335) shows a mouse with an enlarged mesenteric lymph node and spleen and highlights that the spleens presented in various positions, high and behind the intestine rather than horizontally like in **A**.

4.3. MoMuLV induced lymphoma cohort survival analysis

Retrovirus induced lymphoma onset was rapid in all genotypes with the cohort mean and median time to death 18.7 weeks and 16.7 weeks respectively.

There is some separation between the grouped *BCL2* transgenic and non-transgenic genotypes (see Figure 4.3.A and B). When the *BCL2* transgenic groups are combined and compared to the non-transgenic *BCL2* groups there is a significant difference in survival ($p = <0.0001****$) (Figure 4.3.C). *BCL2* overexpression alone is not however, sufficient to significantly accelerate lymphoma when compared to the wild types ($p = 0.8459$) (Figure 4.3.D) this is probably due to the decrease in statistical power when analysing these smaller groups. MoMuLV acceleration of lymphoma is very rapid and larger numbers of mice need to be compared to demonstrate differences in survival.

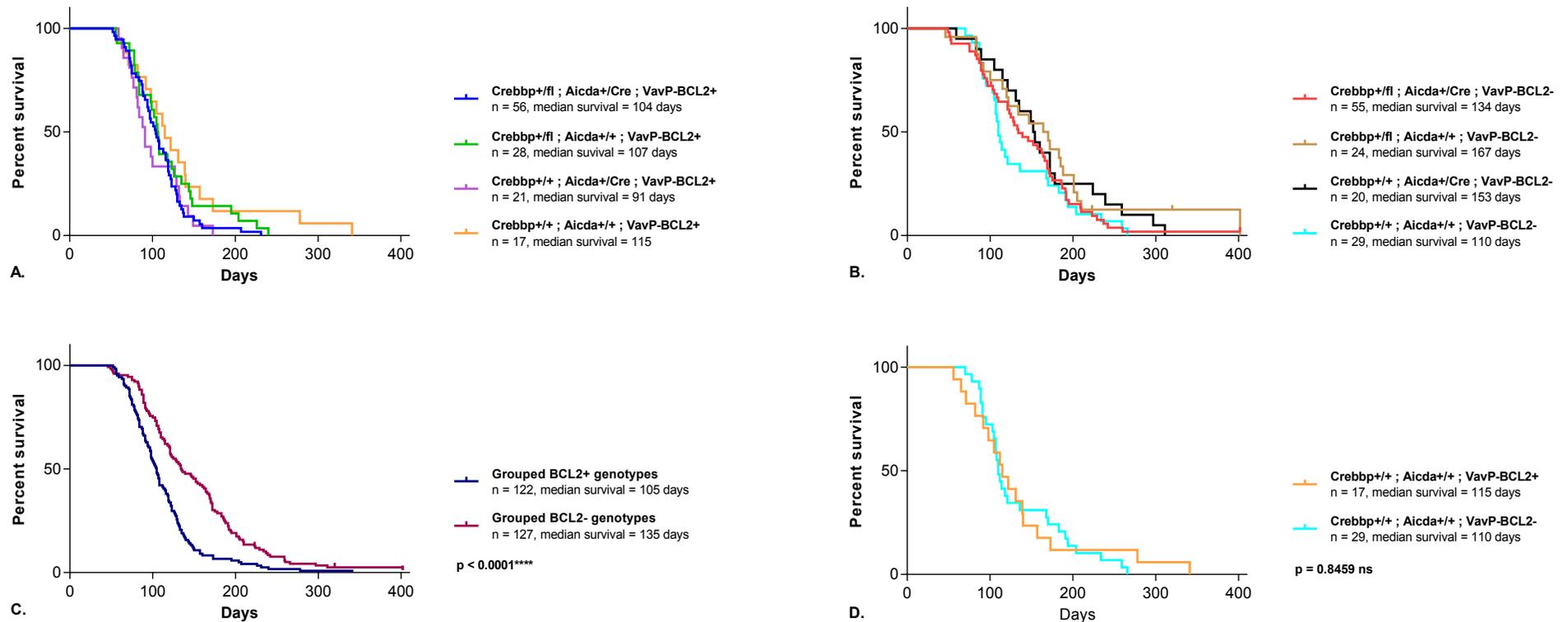


Figure 4.3. MoMuLV induced lymphoma survival curves

Comparison of the survival curves of the MoMuLV infected mice with the *BCL2* transgenic mice shown in **A** and the *BCL2* non-transgenic mice in **B**. MoMuLV greatly accelerates tumour onset, combining the *BCL2* transgenic genotypes and comparing them to the *BCL2* non-transgenic genotype increases the statistical power by increasing the sample number and shows a significant acceleration of tumour onset in the *BCL2*+ animals ($p < 0.0001$) **C** *BCL2* overexpression alone does not accelerate tumour onset relative to the wild type mice **D**, this could be a consequence of the lower sample size.

4.4. Cytometric characterisation of retrovirus induced lymphoma

The MoMuLV induced lymphomas were characterised by flow cytometry using the same antibody panels and gating strategies outlined in section 3.3. A total of 257 spleen suspensions were processed and stained with the four antibody panels. These mice had a high proportion of T-cell lymphomas though *BCL2* overexpression increased the proportion of B-cell lymphoma. The B-cell lymphomas exhibited surface IgM and/or IgD, increased immunoglobulin class switching and enrichment for GC and post GC cells indicating that they are phenotypically similar to mature B-NHLs.

4.4.1. General Panel

The General Panel was used to assess the relative proportions of B- and T-cells to categorised tumours as B or T-cell lymphomas. The T-cell populations were also assessed using the gating strategy delineated in Figure 4.4.

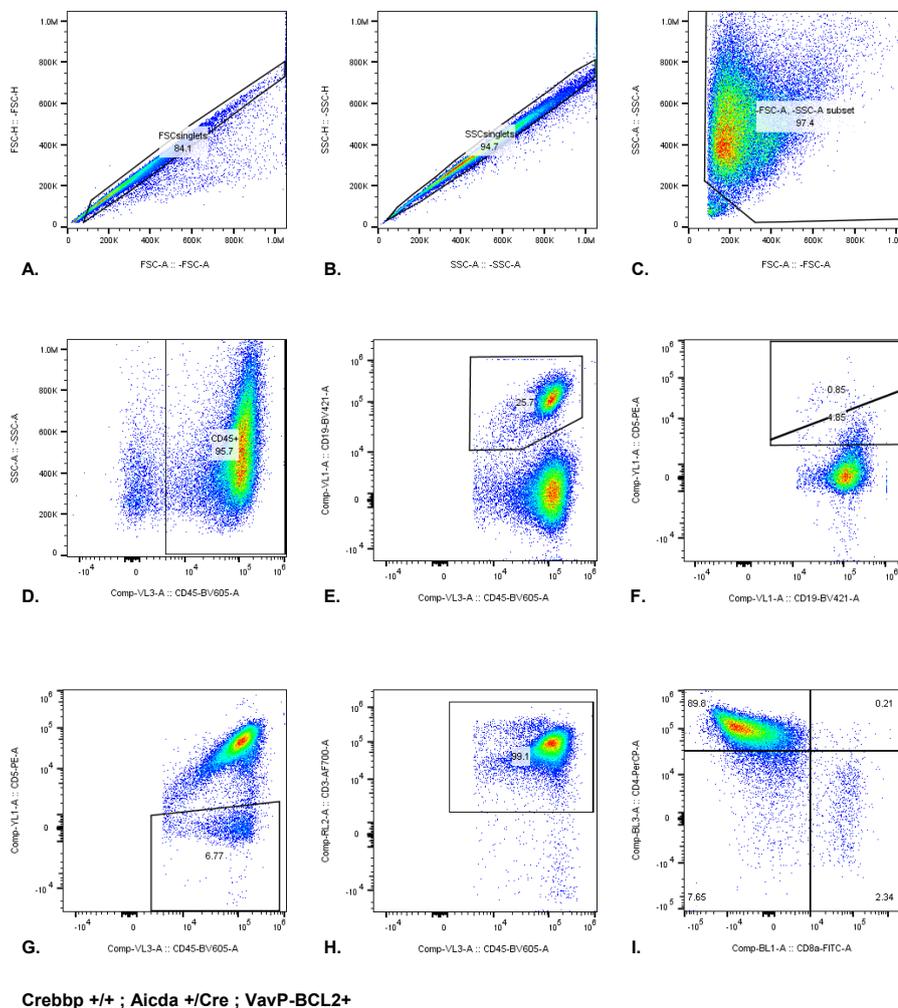


Figure 4.4. MoMuLV induced lymphoma cohort GP gating strategy

The lymphocytes of interest were captured **A-D** The B-cells were defined as CD19+ **E** and their CD5 levels were considered in **F**. The CD45+CD19- T-cells were categorised as CD5+ **G** and CD3+ **H** These CD3+ CD5+ cells were then asses for CD4 and CD8a expression **I** This example if from tumour #5025 which was Crebbp^{+/+} ; Aicda^{+/-}Cre ; VavP-BCL2⁺.

There was a wide range of B-cell proportions (2.5-90.3%) (see Figure 4.5), however the MoMuLV induced lymphomas were predominantly of T-cell origin with the median quantity of B-cells 33.3% and mean 39.2%. Refer also to Figure 4.9 for an overview of the cohort.

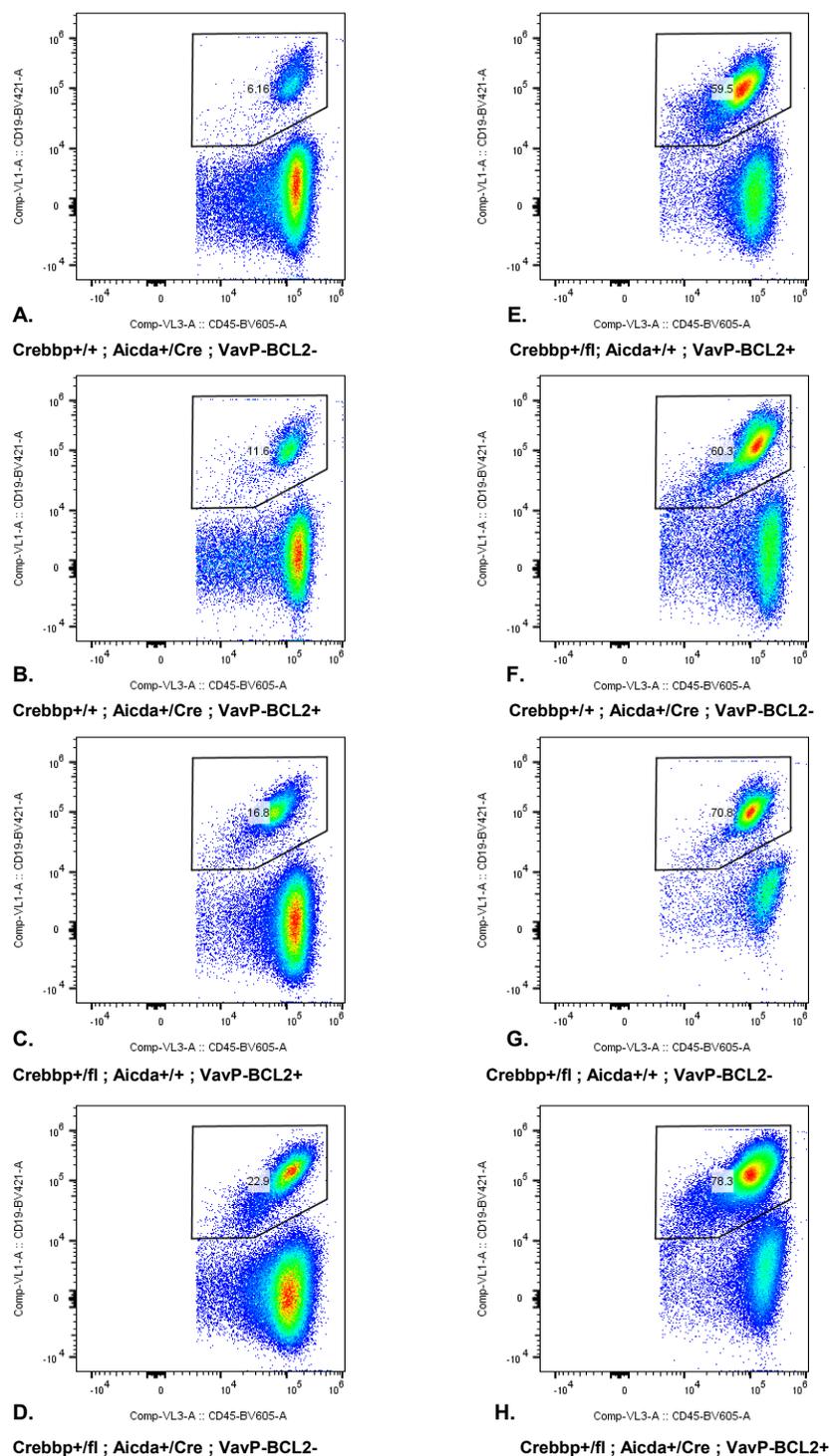


Figure 4.5. MoMuLV induced lymphoma cohort examples of the varying proportions of B- and T-cells

The majority of the tumours were of T-cell origin. **A-D** are typical examples of T-cell lymphomas with significantly less than 50% B-cells (CD45+ CD19+) (**A** #5177, **B** #5044, **C** #5012 and **D** #5004). **E-H** are B-cell lymphomas with minor T-cell fractions (CD45+ CD19+), (**E** #5017, **F** #5098, **G** #5143 and **H** #5245).

There were some tumours with unusual B220/CD45 CD19 plots (see Figure 4.6). Examples **A-C** are predominantly T-cells (CD19-) and have two distinct subgroups of CD45+ expression. **D-F** have increasing proportion of B-cells (CD19+) which also have two distinct populations based on CD45+ expression, suggestive of polyclonal disease.

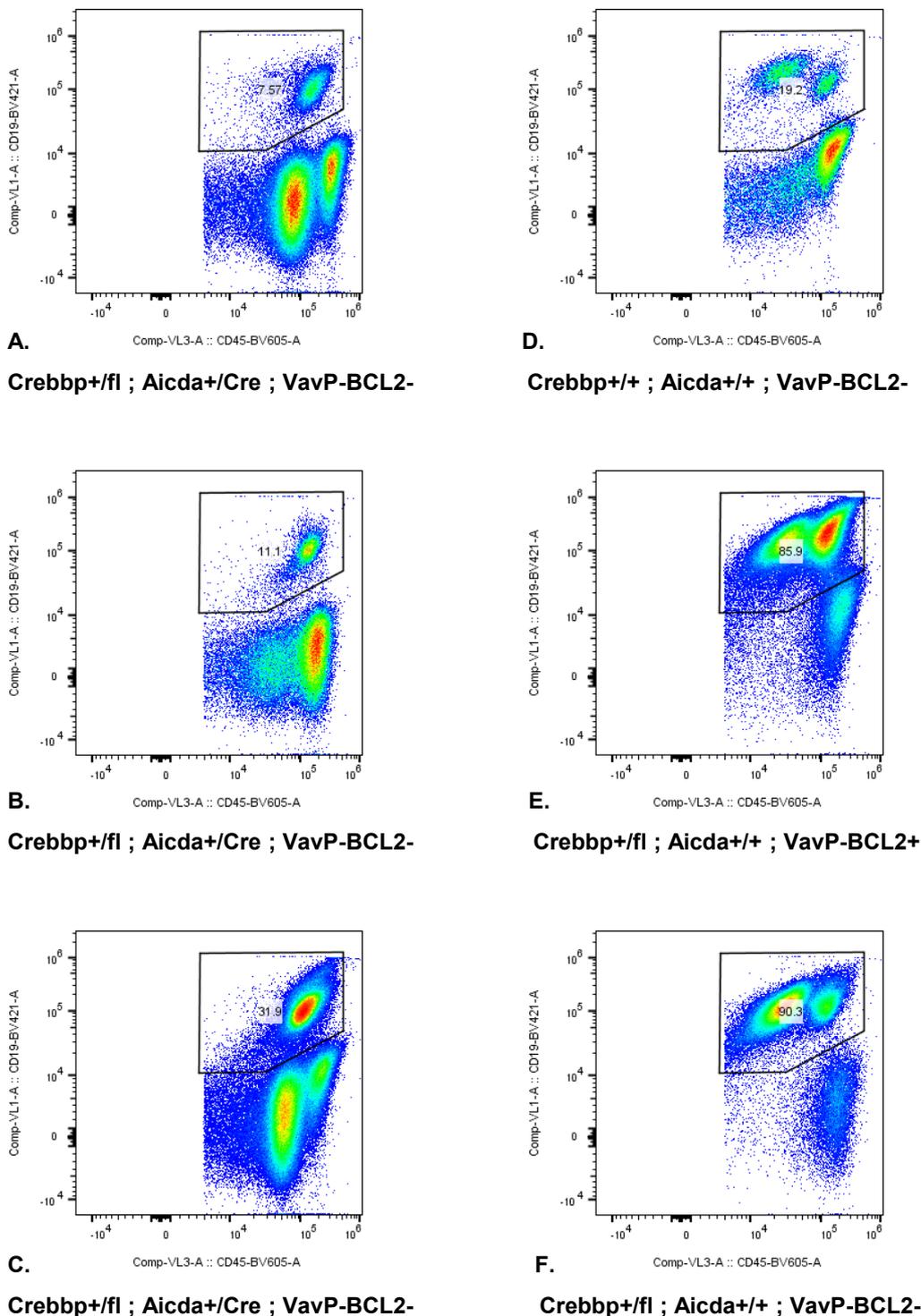


Figure 4.6. MoMuLV induced lymphoma cohort with atypical B- and T-cell profiles

Occasionally the tumours had populations with variable levels of CD45 expression. **A-C** are T-cell lymphomas with minority CD19+ populations. The T-cells exhibit two distinct subgroups of CD45 expression (**A** #5277, **B** #5267 and **C** #5338). **D-F** are B-cell lymphomas with two populations of CD45 expression (**D** #5097, **E** #5167 and **F** #5176)

The T-lymphocytes (CD45+ CD19- CD5+ CD3+) were further characterised using CD8a and CD4. The premalignant samples indicated that CD4+ and CD8a+ single positive are the major population with CD8+ cells being less numerous (Figure 3.7, page 54). The MoMuLV induced lymphomas tended to have a single population of CD4+ (Figure 4.7.A-B) or two populations of CD4+ and CD8a+ single positives (Figure 4.7.C-D). Some examples had significant proportions of double negatives usually in combination with CD4+ (E, F and G) and occasionally also with CD8a+ (H).

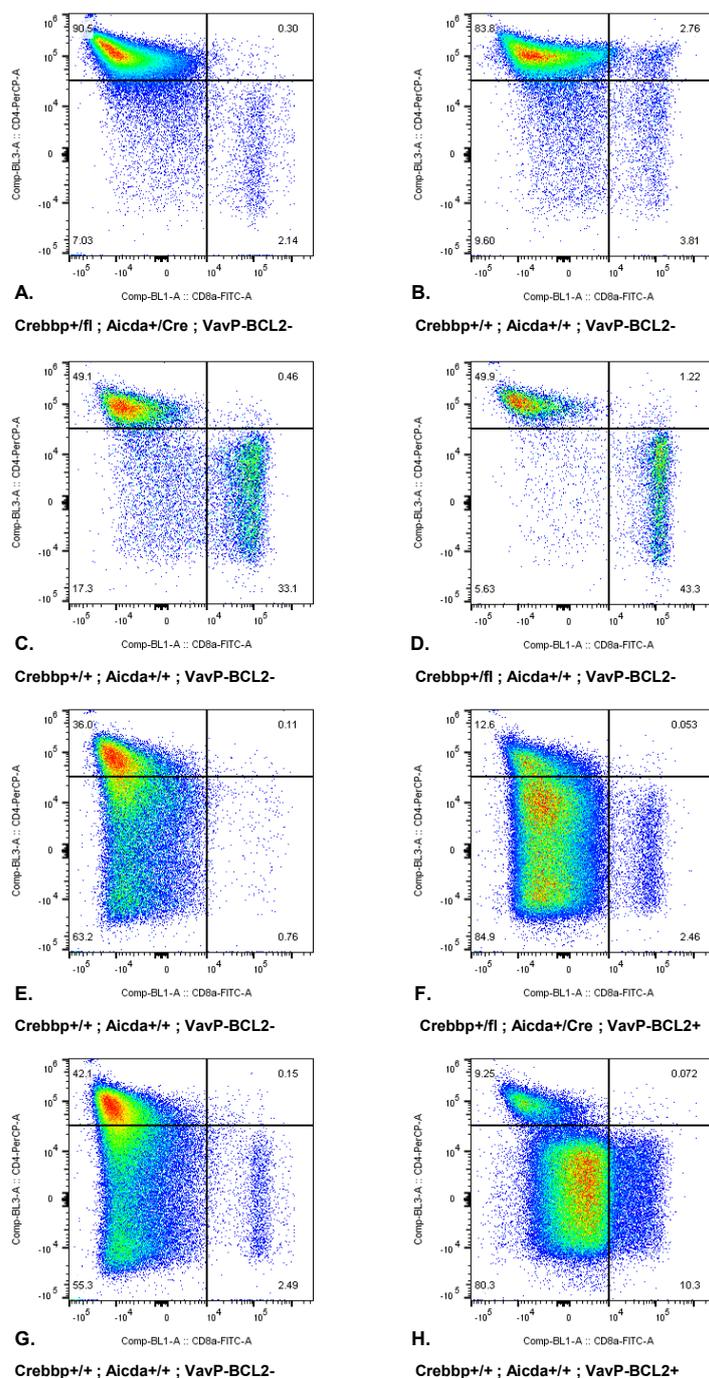


Figure 4.7. MoMuLV induced lymphoma cohort representative CD4 and CD8a plots

The majority of the T-cells (CD45+ CD19- CD5+ CD3+) were CD4+ single positive **A-B** (#5004 and #5073) or CD4+ and CD8a+ single positive **C-D** (#5056 and #5175). Less frequently there were tumours with significant double negative populations usually with a CD4+ population (**E** #5024, **F** #5007 and **G** #5016) and sometimes also with CD8a+ population (**H** #5067).

There were generally very few double positive cells but a few examples are highlighted in Figure 4.8. where **A-D** show tumours with increasing proportions of mature CD8a+ CD4+ cells. Spleen **E** also has some double positive cells but it is predominantly CD8a+ single positive and possibly of cytotoxic T-cell origin. **F** and **G** are also almost exclusively CD8a+ whereas the spleen in **H** is very heterogeneous with cell in all quadrants but mainly CD4- cells as does **I** which only has a very small discrete population of CD4+.

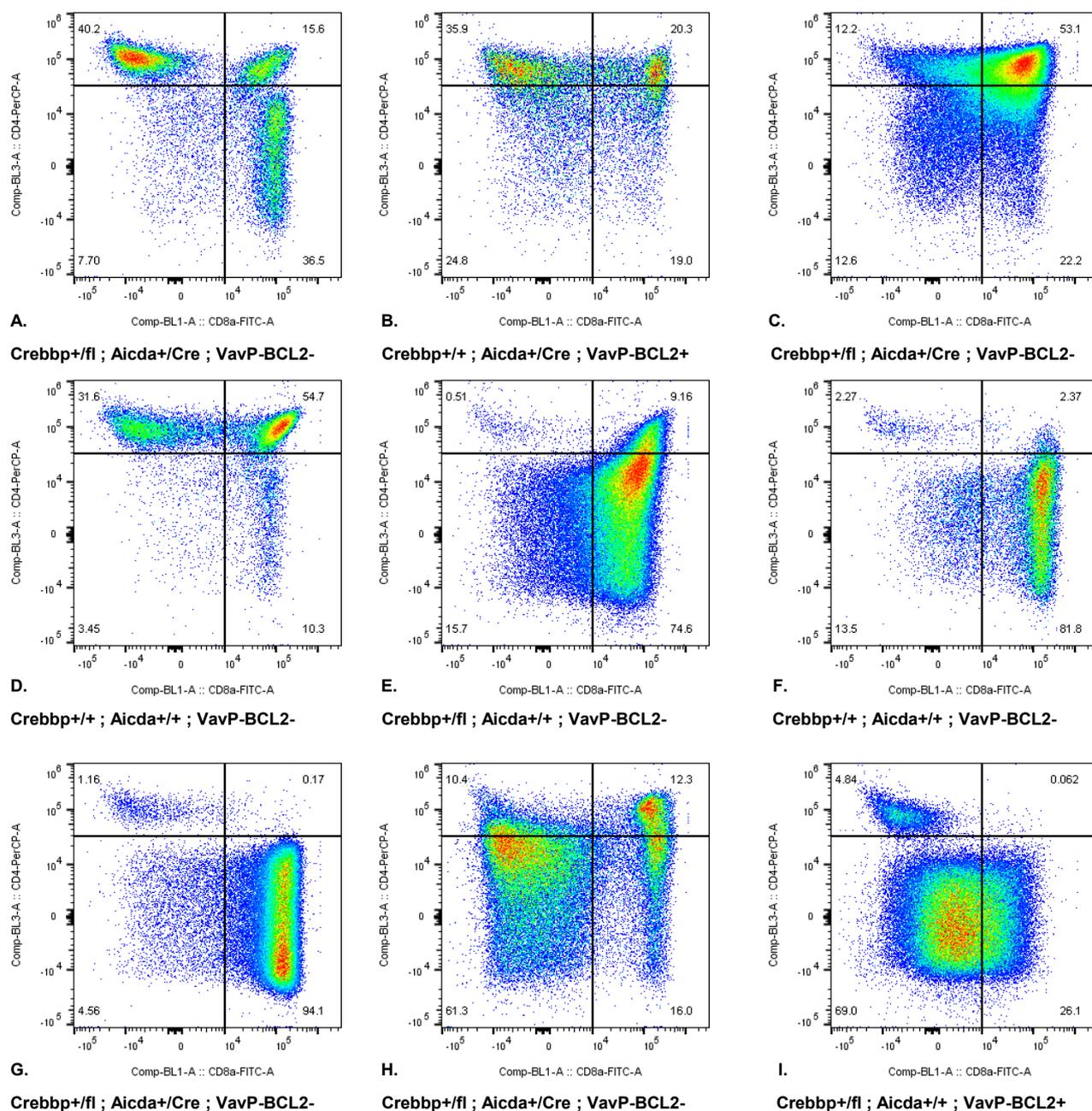


Figure 4.8. MoMuLV induced lymphoma cohort abnormal CD8a and CD4 profiles

Some tumours deviated from the majority CD4+ profiles. **A-D** have increasing CD8a+ CD4+ double positives (**A** #5030, **B** #5046, **C** #5144 and **D** #5032). **E** (#5154) also has some double positive cells but it is predominantly CD8a+ single positive. **F** (#5097) and **G** (#5347) are mainly CD8a+. The tumour in **H** (#5077) is mainly double negative cells as is **I** (#5026) which also has a small discrete population of CD4+.

Figure 4.9 below give an overview of the cellular composition of the MoMuLV induced lymphomas. The majority are T-cell lymphomas, approximately 62%, with minority CD45+ CD19+ populations in contrast to the spontaneous lymphomas which were predominantly B-cell lymphomas. Approximately one third are B-cell lymphomas whose proportions of T-cell follow a similar trend; CD5+ CD3+ CD8- CD4+ (dark red) is the largest fraction followed by CD5+ CD3+ CD8+ CD4- (red), CD5- CD3+ (light green), CD5+ CD3+ (yellow) and CD5+ CD3- (dark green).

The T-cell lymphoma have more varied T-cell proportions but tend to be predominantly single positive CD8a- CD4+ (dark red) likely arising from immature single positives or T-helper cells, or double negative CD8a- CD4- (yellow) which resemble early thymic cells.

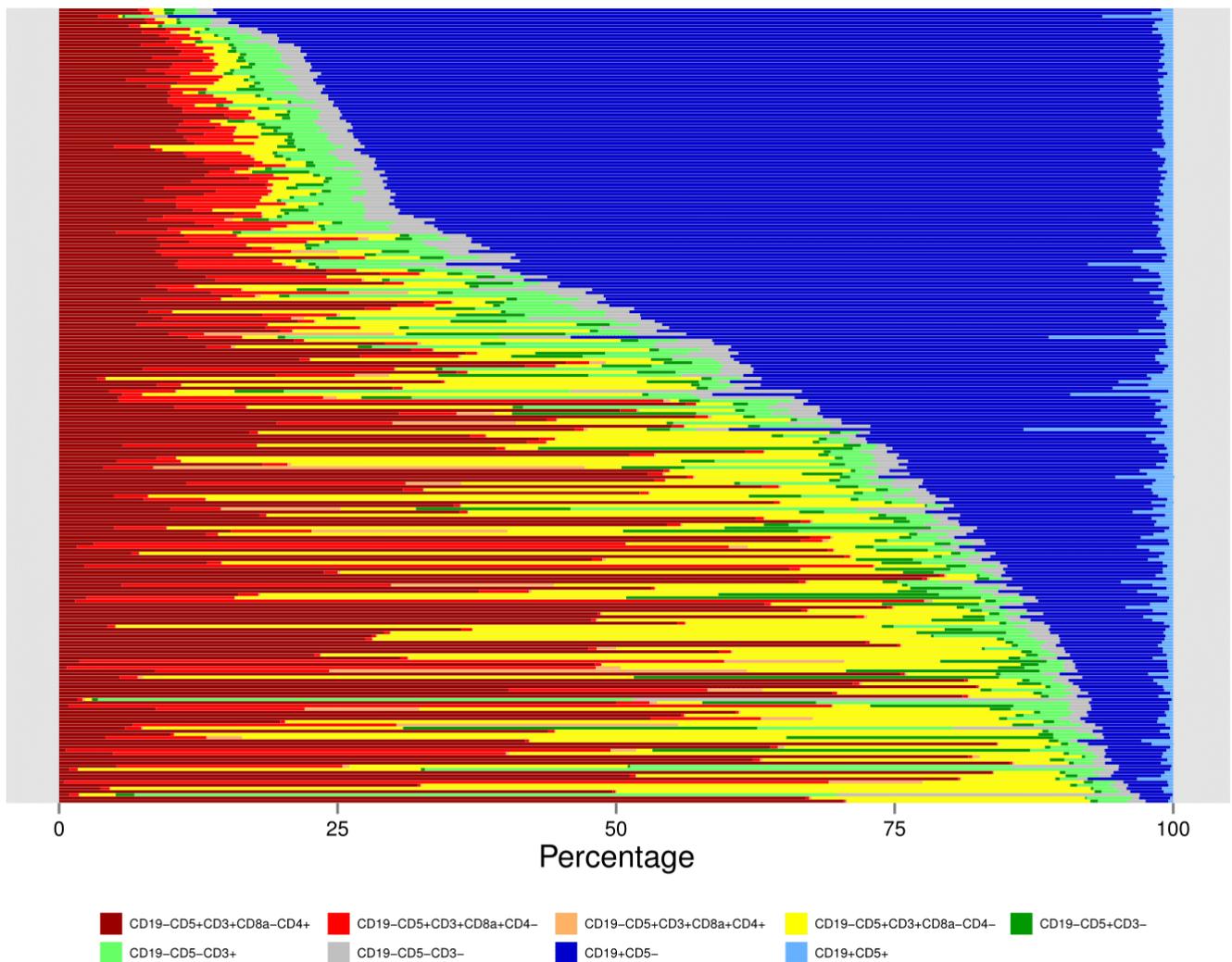


Figure 4.9. MoMuLV induced lymphoma cohort GP overview of tumours composition

Graphical representation of the cellular proportions of the CD45+ lymphocytes ranked by CD19 positivity. The majority of MoMuLV induced lymphomas are of T-cell origin (bottom two thirds).

When the data is grouped by genotype it becomes clear that the *BCL2* transgenic mice (Figure 4.10, left hand side) are more susceptible to B-cell lymphomas; though these do also occur in the non-transgenic *BCL2* samples on the right hand side but in lower proportions. The majority of the non-transgenic *BCL2* mice had T-cell lymphomas with *Crebbp* loss alone not seeming to promote B-cell lymphoma (Figure 4.10, right hand side).

The sensitising mutations of overexpression of *BCL2* in combination with loss of *Crebbp* predispose the mice to B-cell lymphoma. In these samples the T-cell populations are more similar. In the T-cell lymphomas the T-cell subtype profiles are more heterogeneous with greater degrees of expansion in the different subtypes indicating different clonal diseases.

The skew to B-cell lymphoma in the transgenic *BCL2* mice is greater than that in the *Crebbp* deficient *CrebbpCre* mice because *BCL2* is a more potent accelerator of lymphoma. Overexpression of *BCL2* accelerated spontaneous lymphoma onset relative to the wild type mice ($p = 0.2004$) (refer to Section 3.2, Figure 3.1.B) whereas loss of *Crebbp* alone was not sufficient ($p = 0.5290$).

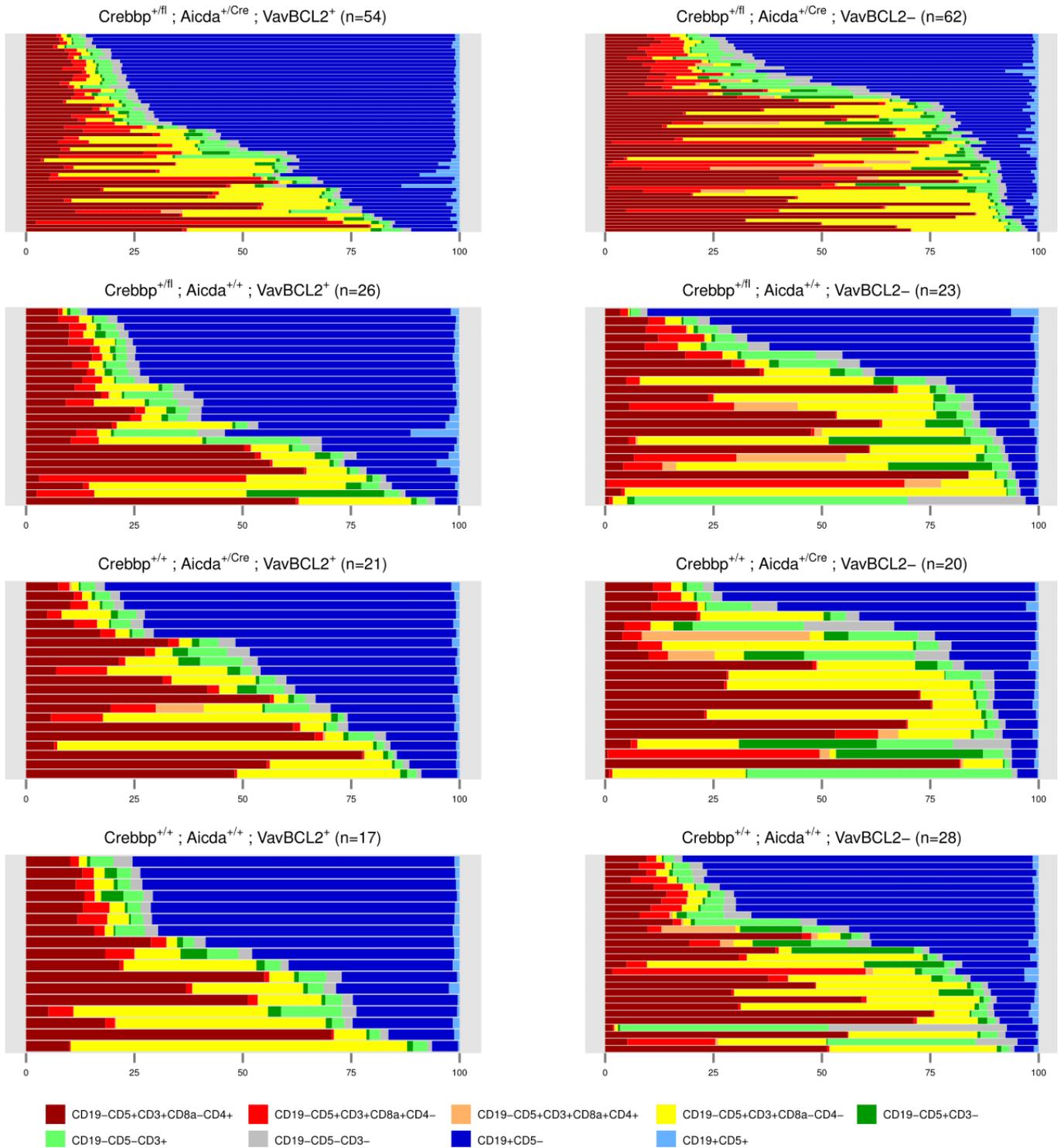


Figure 4.10. MoMuLV induced lymphoma cohort GP overview of tumours composition by genotype

Graphical representation of the cellular proportions of the CD45⁺ lymphocytes divided by genotype and ordered by CD4⁺ positivity. The *BCL2* transgenic have greater proportion of B-cells (blue populations) and are more likely to be B-cell lymphomas.

Visualisation of the T-lymphocytes in isolation emphasises their heterogeneity and considering the CD8 and CD4 status indicates the stage of maturity of the T-cells (Figure 4.11).

Generally the major proportion in these tumours is CD8⁻ CD4⁺ (dark red) indicating expansion of immature single positives or T-helper cells. Immature single positives have lower CD3 expression than late thymic T-helpers and T-cytotoxic cells. The CD45⁺ CD19⁻ cells generally showed high CD3 expression which would suggest that the CD8⁻ CD4⁺ are in fact T-helper cells (data not shown). Likewise the CD8⁺ CD4⁻ (red) are most likely T-cytotoxic cells which are rarely the major component of the tumours. The CD5⁺ CD3⁻ CD8⁻ CD4⁻ (yellow) are early thymic double negative cells, these cells are expanded in quite a few of the tumours. CD8a⁺ CD4⁺ (orange) late thymic double positives occur infrequently and almost exclusively in the non-transgenic *BCL2* tumours (right hand side).

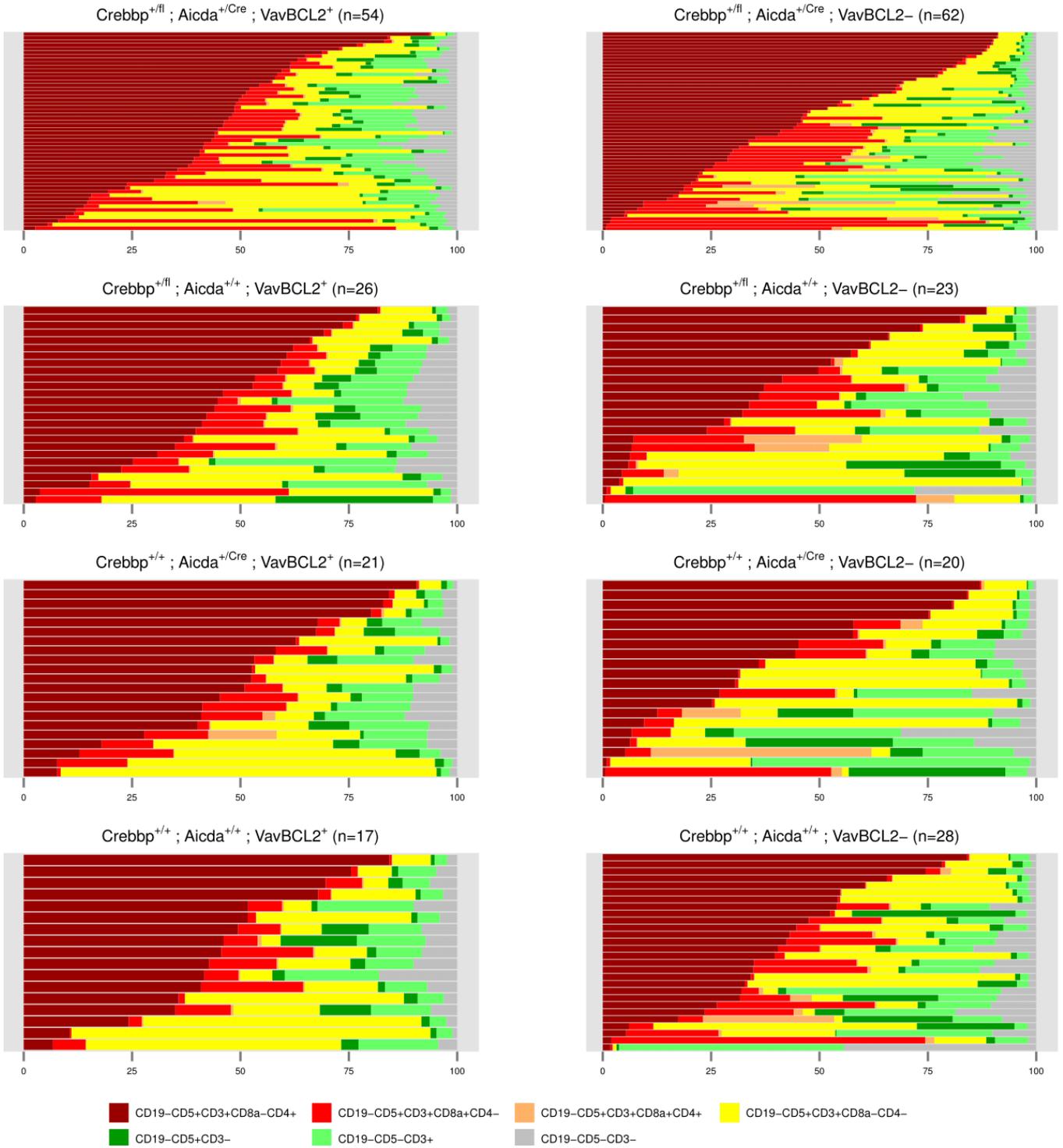
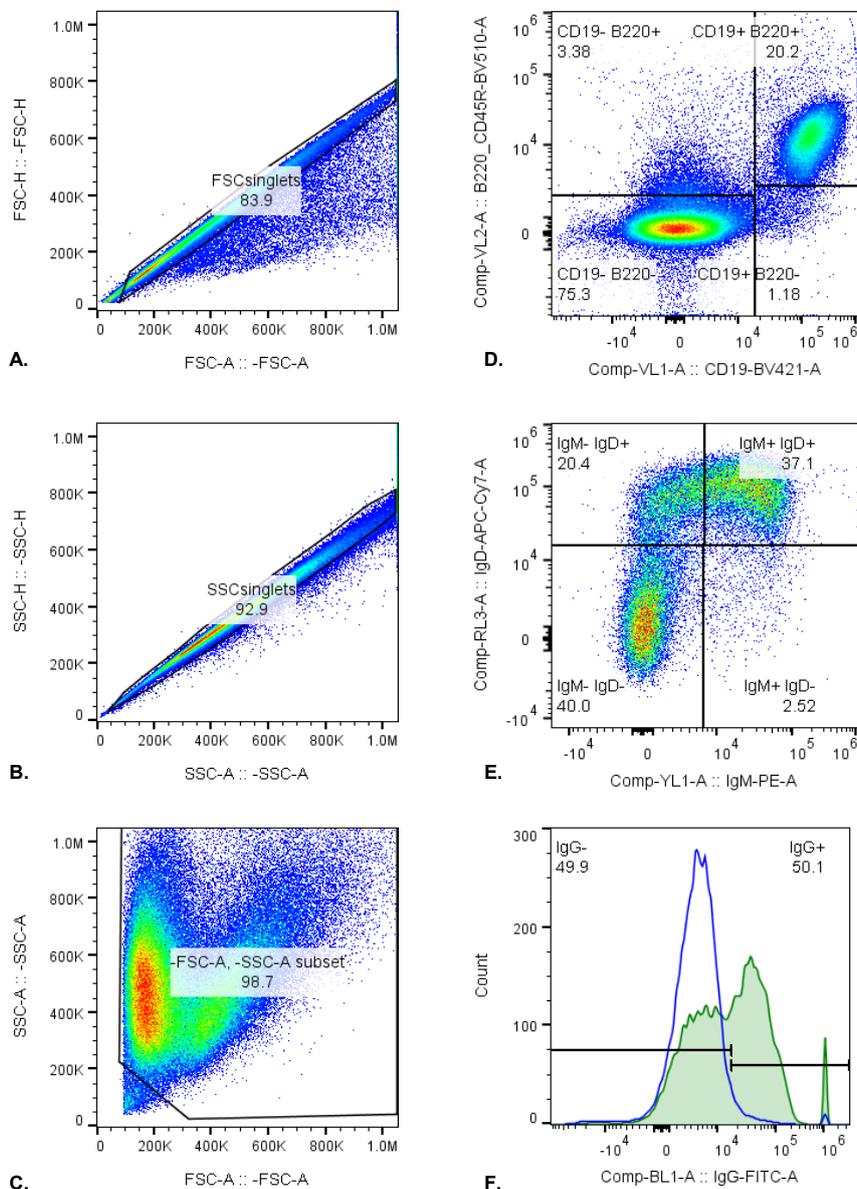


Figure 4.11. MoMuLV induced lymphoma cohort GP overview of T-cell proportions by genotype

Graphical representation of T-cell subtypes of the CD45+ CD19- T-lymphocytes divided by genotype and ordered by CD45 positivity.

4.4.2. B-cell Panel

The B-cell Panel uses immunoglobulin status to assess the stage of development of the expanded B-cells with the earliest IgM+ cells representing T1 B-cells through to the class switched GC, plasma and memory cells which will be IgM- IgD- after class switching but express IgG, IgA or IgE Figure 4.12. details the gating strategy for the B-cell Panel.



Crebbp^{+/+} ; Aicda^{+/+} ; VavP-BCL2-

Figure 4.12. MoMuLV induced lymphoma cohort BP gating strategy

A-C the study population was gated (as described in section 3.3.1) and then the relative quantities of CD19 and B220 were assessed in D This sample is predominantly CD19- B220- indicative of non-B-cell expansion. The B-cells (CD19+) from the pro-B-cell stage onwards were then assessed for IgM and IgD in E This sample is composed of IgM- IgD- class switched cells (40.0%), IgM+ IgD+ like T2 B-cells (37.1%) and IgM- IgD+ like follicular cells (20.4%) with very few IgM+ IgD- single positives indicating that there are few early T1 B-cells (2.52%). F The IgG levels were then assessed in the immunoglobulin positive cells (IgM+ IgD- and IgM+ IgD+ and IgM- IgD+) (blue line) and the IgM- IgD- double negative cells (green histogram). The immunoglobulin positive cells do not express IgG as they have not yet undergone class switching. Approximately 50% of the class switched cells express IgG and the remainder probably express IgA or IgE. The gating strategy was set using tumour #5033 which is Crebbp^{+/+} ; Aicda^{+/+} ; VavP-BCL2⁻.

Immunoglobulin expression of a tumour can indicate what cells it originates from and where in the haematopoietic compartment. IgD- IgM+ are phenotypically like T1 B-cells which originate in the spleen but have not yet encountered antigen or gone into the GC, or they can be B-cells from the marginal zone. IgD+ IgM+ cells are like T2 B-cells entering the follicle. IgM- IgD+ represent follicular B-cells which also express CD23 and CXCR5. IgM- IgD- cells can be GC B-cells (also PNA+ CD95+), plasma cells or memory cells. These cells have passed through the GC and have undergone class switching to IgG, IgA or IgE

Figure 4.13. shows some typical immunoglobulin profiles which tended to have a three quadrant profile with very few IgM+ IgD- which represent the earlier stages of B-cell development prior to entry into the follicle. The IgM- IgD- cells show differing degrees of class switching to IgG, the IgG- cells presumably being IgA or IgE positive, but generally IgG is the most common secondary isotype. GCB DLBCL tends to be IgG+ or less commonly IgA+ in contrast to ABC DLBCL which tend to be IgM+ or IgD+ (Ruminy et al., 2011).

A number of lymphomas with an atypical phenotype with abnormal CD19 B220/CD45 plots (Figure 4.14. **A-F**) or abnormal immunoglobulin plots (**G-I**) are detailed below.

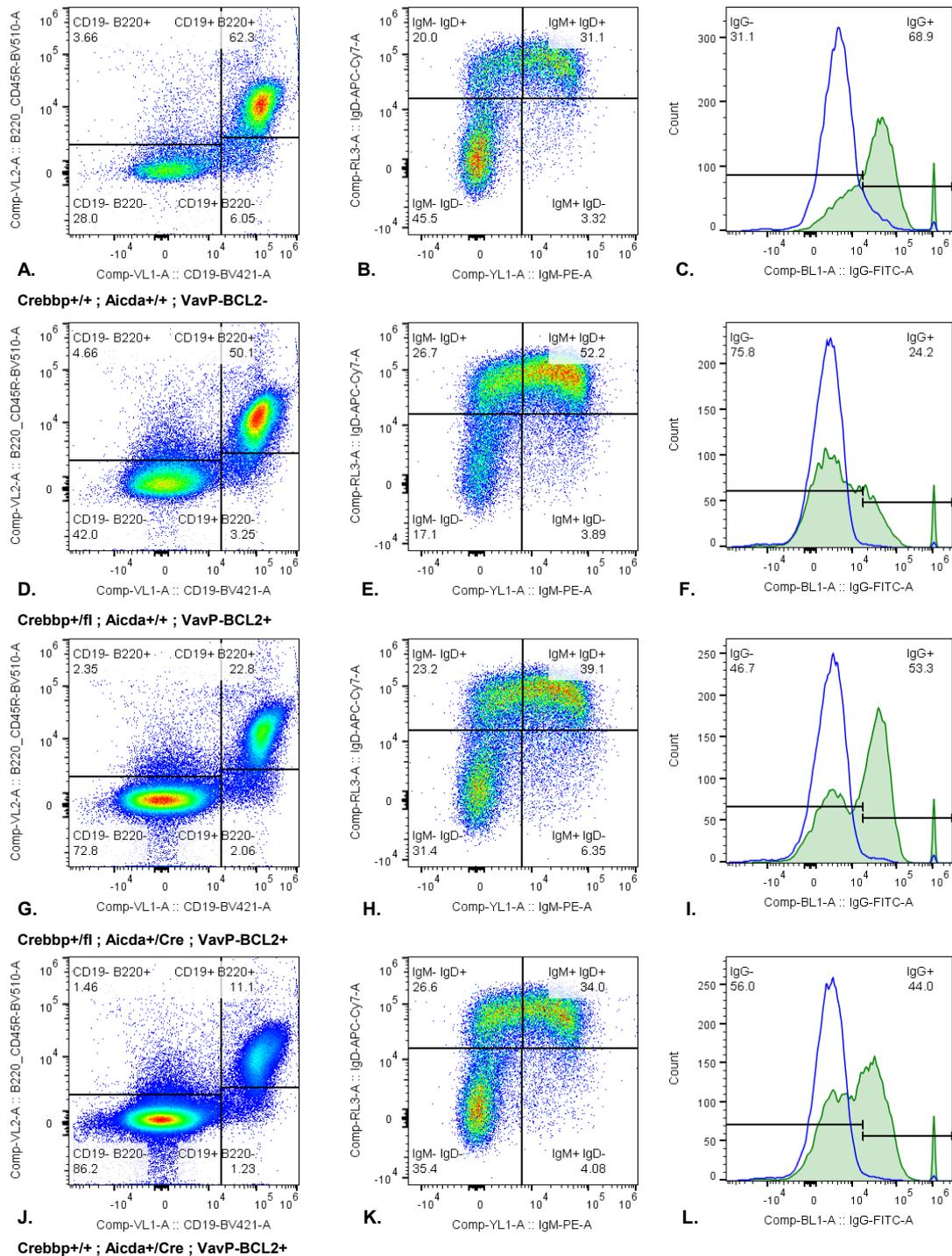


Figure 4.13. MoMuLV lymphoma cohort characteristic immunoglobulin profiles

Four examples of the relative B- and T-cell proportions of tumours (left hand column) with their respective IgM IgD plots (central column) and the IgG expression (right hand column) of their double negative B-cells (CD19+ IgM- IgD-) (green) compared to that of the immunoglobulin positive cells (blue) (includes IgM+ IgD- and IgM+ IgD+ and IgM- IgD+).

A (#5053) is an example of a B-cell lymphoma (62.3% CD19+ B220+) with predominantly double negative cells (45.5% IgM- IgD-) **B** of which 68.9% express IgG **C**.

D (#5017) has approximately equal proportions of B- and T-cells. The B-cells (CD19+) are predominantly IgD+ **E** Of the double negative cells only 24.2% of cells are IgG+ indicating that little class switching has occurred **F**.

G (#5019) is predominantly T-cells (72.8%), of the B-cells the majority are immunoglobulin positive **H** and do not express IgG however half of the IgM- IgD- cells express IgG and have undergone class switching **I**.

J (#5036) demonstrates a similar pattern to the previous example and it has intact Crebbp.

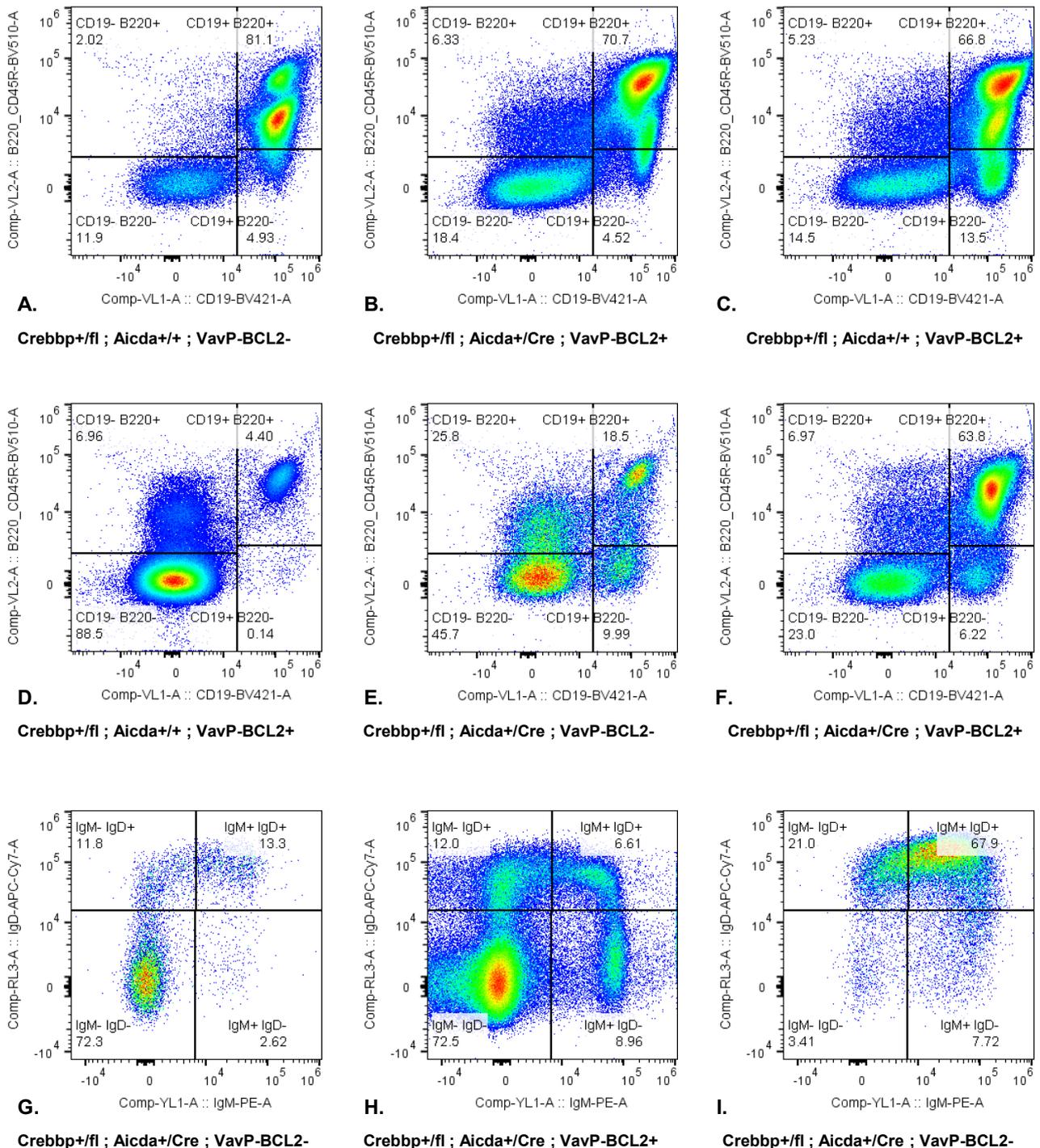


Figure 4.14. MoMuLV lymphoma cohort atypical plots

A-F are examples of tumours with unusual CD19 B220/CD45 plots (**A** #5176, **B** #5317, **C** #5167, **D** #5285, **E** #5276 and **F** #5329). **A-C** Have multiple CD19+ B220/CD45 populations with varying expression of B220/CD19. **D** has a very discrete population of CD19- B220+ cells accounting for 6.96% of the total. **E** Has many more CD19- B220+ cells (25.8%) and also a small population of CD19+ B220/CD45- cells of unknown origin as does **F**.

G-I are examples of abnormal IgM IgD plots (**G** #5034, **H** #5317 and **I** #5004). **G** and **H** are both predominantly class switched IgM- IgD-, however their profiles look very different and **H** also has a more distinct IgM+ IgD- population than usual, giving it a horseshoe shape. **I** is mainly IgD+. The majority population is IgD+ IgM+ indicating expansion of cells with a T2 B-cell like phenotype.

The majority of the MoMuLV induced tumours are of T-cell origin with expanded CD19- populations (red and orange). In these tumours there is very little class switching (pink populations) with the majority of class switching occurring in the B-cell lymphomas, lower portion of the graph (Figure 4.15).

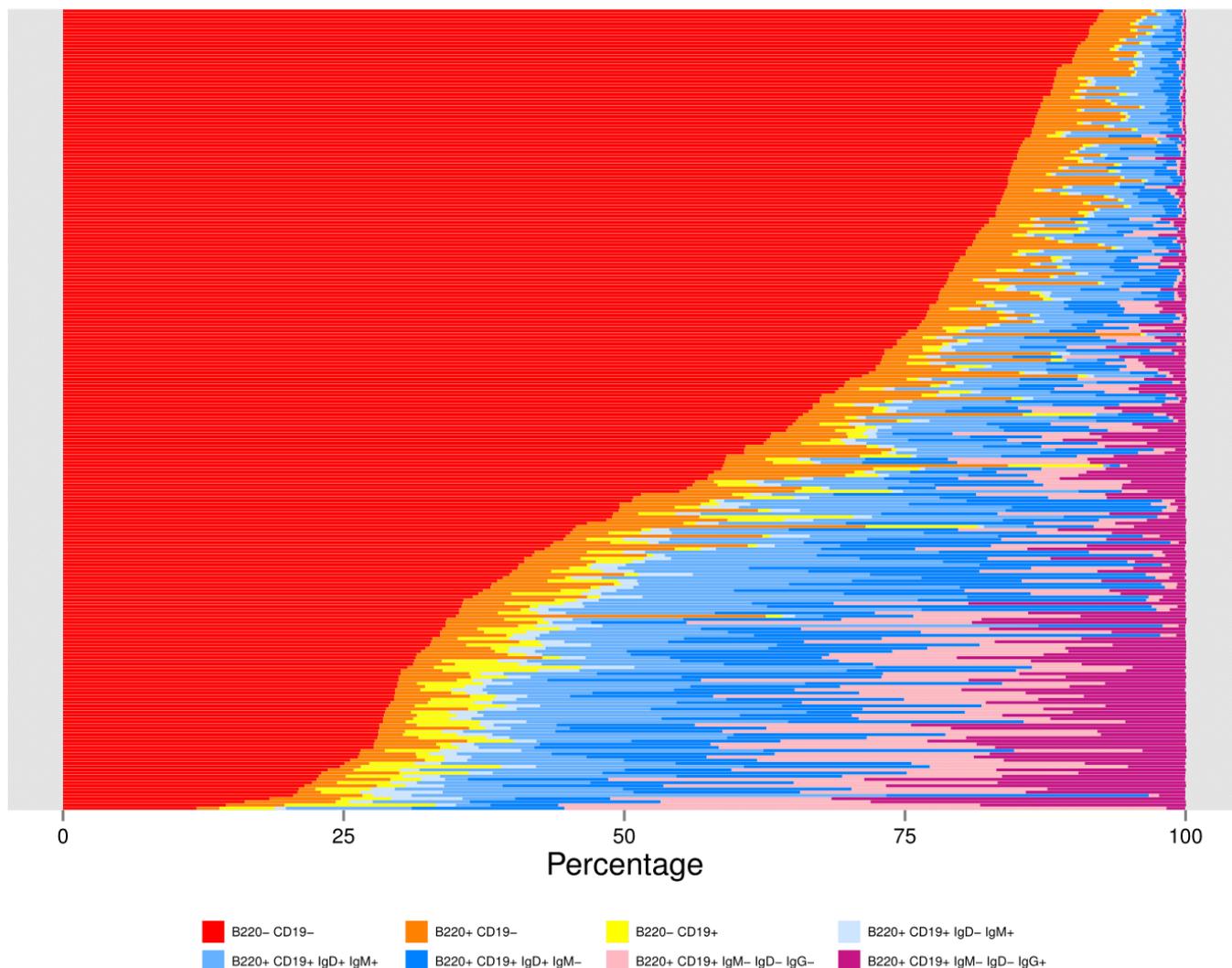


Figure 4.15. MoMuLV induced lymphoma cohort BP overview of tumours composition

Overview of the T (CD19- red and orange populations) and B-cell proportions in the tumours are their immunoglobulin fractions.

The control premalignant samples demonstrate that less than 15% of normal cells are class switched cells (Section 3.3.2., Figure 3.12) which suggest that the majority of these tumours do not have a significant proportion of class switched B-cells (pink populations) with the major immunoglobulin status being IgM+ IgD- like T2 B-cells (medium blue).

The increase in class switched cells is correlated to the increasing numbers of B-cells. The classed switched IgM- IgD- IgG+ (dark pink) and IgM- IgD- IgG- (light pink) are generally of a similar proportion where, with few exceptions, IgG is the predominant secondary isotype. Approximately two

thirds of the tumours have expansion of class switched cell indicative of hyperplasia of post GC B-cells recapitulating GCB DLBCL.

The majority of the significantly class switched cells can be attributed to the *BCL2* transgenic tumours conforming to previous work that indicated that they had increase number of germinal centres and increased class switching (Egle et al., 2004) (Figure 4.16. left hand side). Xu and colleagues indicated that *Crebbp* deficiency also increases class switching (Xu, 2006) which also seems to be the trend in this data. Increased class switching is positively correlated with disease incidence.

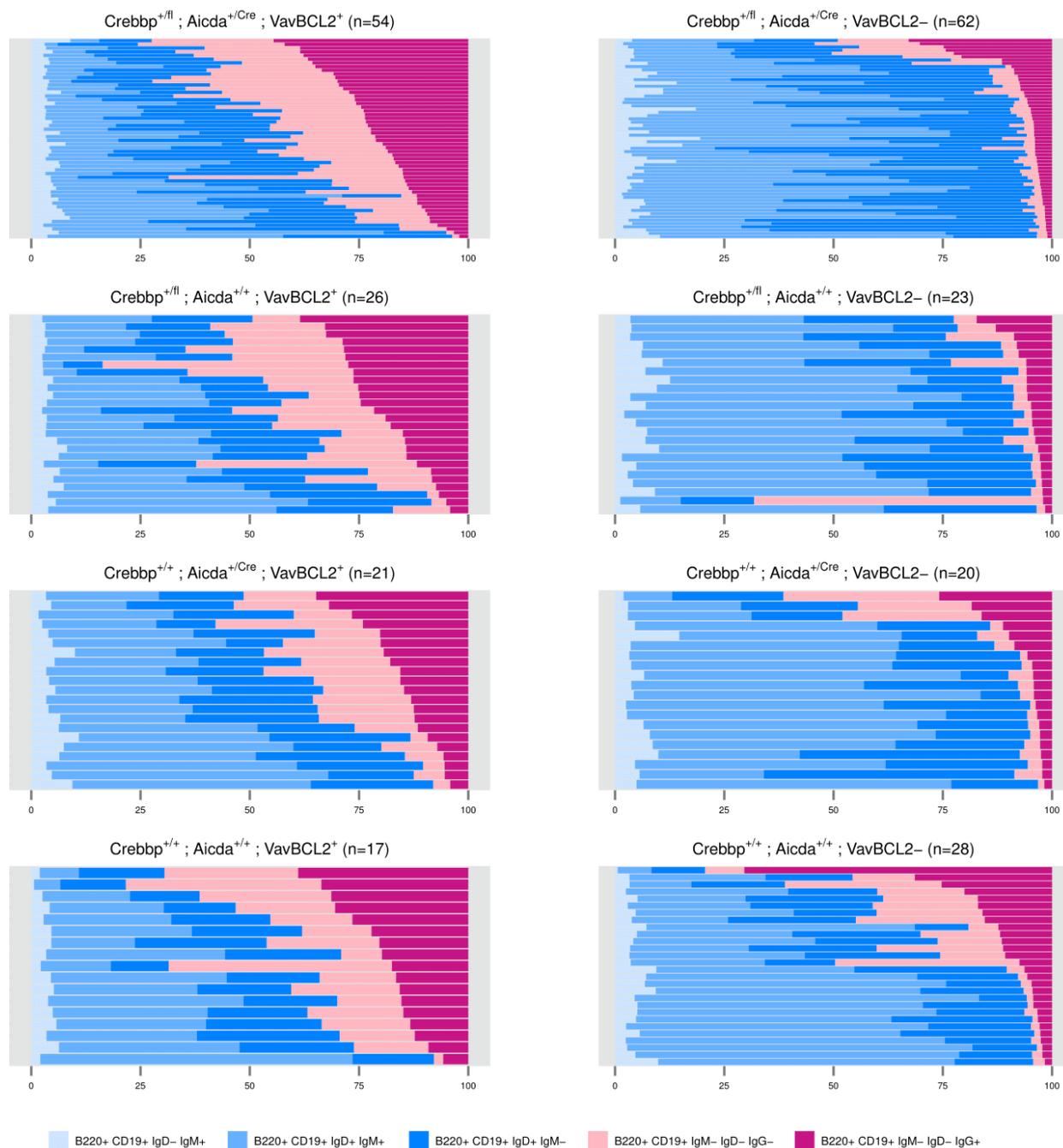


Figure 4.16. MoMuLV induced lymphoma cohort BP overview of immunoglobulin proportions by genotype

Dividing the immunoglobulin profiles by genotypes segregates the significantly class switch cells to the *BCL2* transgenic genotypes on the left hand side.

4.4.3. Lambda/Kappa Panel

The Lambda/Kappa Panel was used to distinguish between early B-cells prior to immunoglobulin light chain formation, and post immature B-cell lymphomas with altered Lambda and Kappa ratio being indicative of disease. The gating strategy is described below (Figure 4.17).

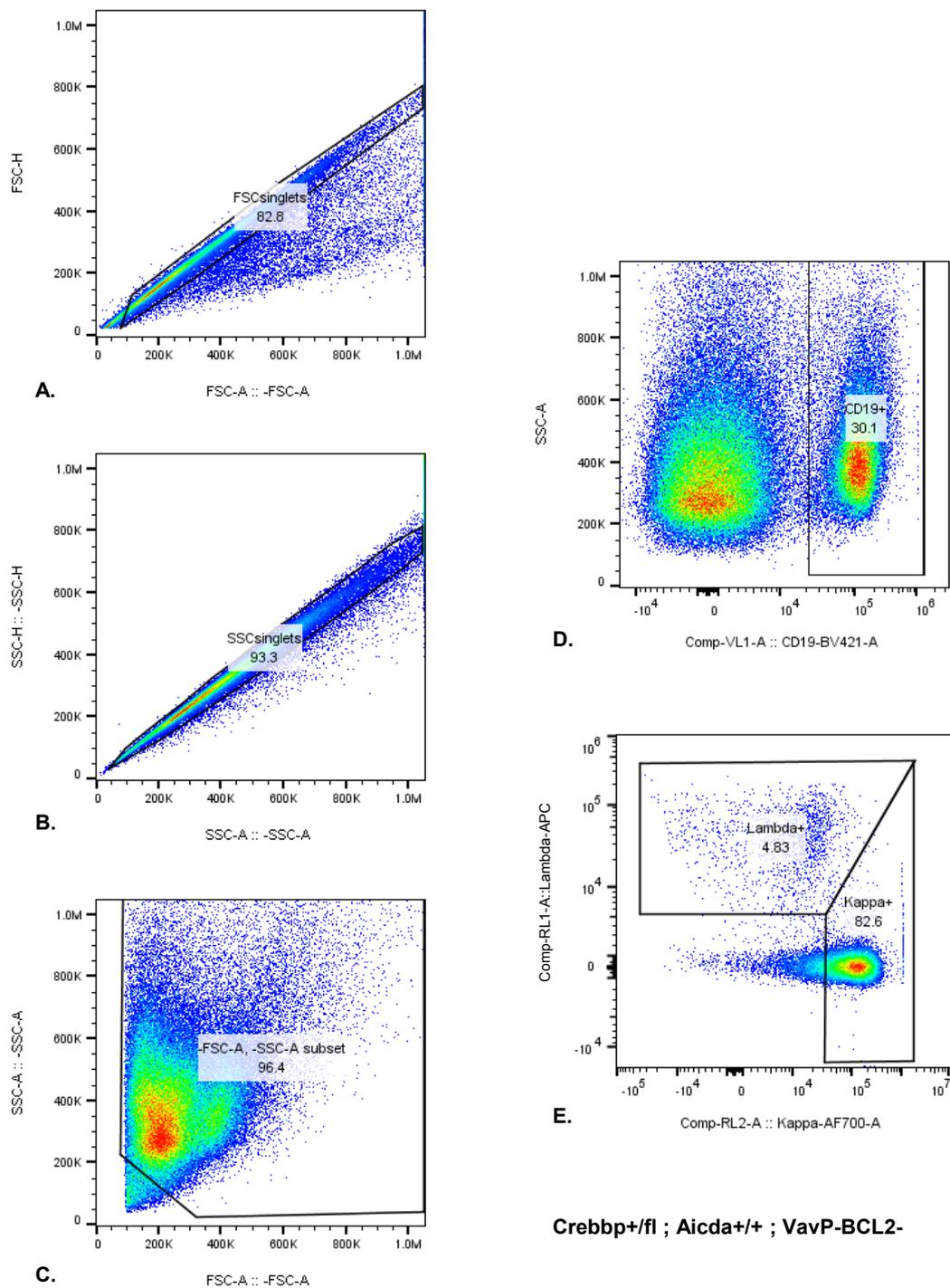


Figure 4.17. MoMuLV induced lymphoma cohort LKP gating strategy

A-C the study population was gated (as described in section 3.3.1) and then the CD19 expression was used to define the B-cells (CD19+) **D** The proportion of Lambda and Kappa in the B-cells was assessed in **E**

The vast majority of the tumours expressed Lambda and Kappa, though the ratios varied there was little evidence of significant light chain restriction. The mean Lambda/Kappa ratio was 1:10 with generally more Kappa expression than in the premalignant animals (1:15). There was some variability in the Lambda and Kappa ratio but broadly the profiles remain very similar (Figure 4.18). Figure 4.18.H was probably the most obvious example of Lambda restriction.

The majority of the B-cells were single or double positive for CD23 and CXCR5 and so it was not useful in distinguishing between T2 B-cells (CD23+) and follicular B-cells (CD23+ CXCR5) (data not shown).

There were a few examples of tumours with significant populations of Lambda Kappa double negative cells. These represent outgrowth of the very earliest stages of B-cells that have infiltrated from the bone marrow. They could also perhaps be mature cells that have non-functional BCRs that have survived in the spleen because of altered surveillance and apoptosis (Figure 4.19).

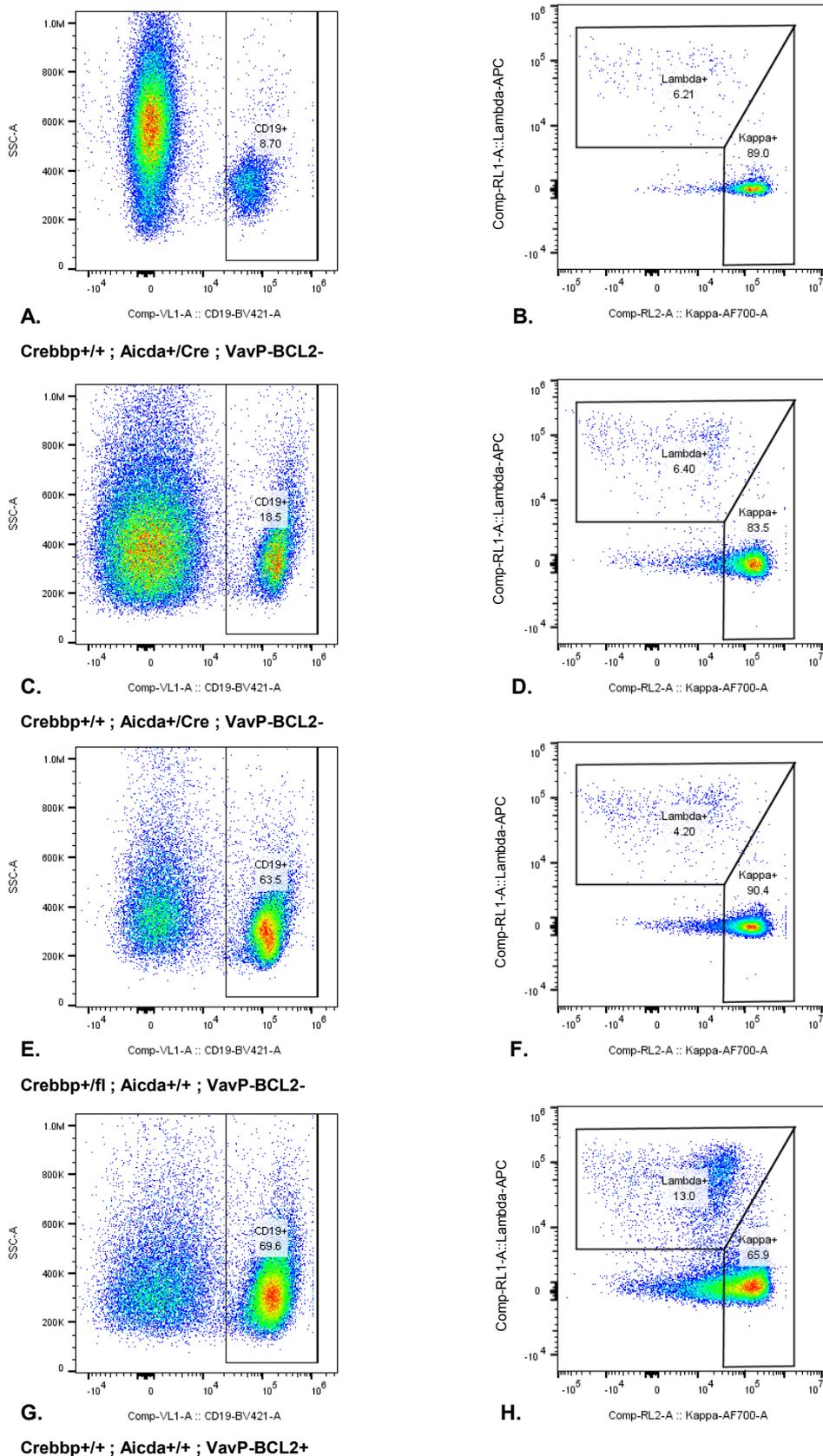


Figure 4.18. MoMuLV induced lymphoma cohort demonstration of homogeneity of Lambda/Kappa ratios

The left hand column shows spleen suspensions with increasing percentage of B-cells (CD19+) (A (#5027), C (#5091), E (#5143) and G (#5094) and on the right their Lambda/Kappa ratios are assessed. There is some variety in Lambda/Kappa ratios in the cohort B 1:14, D 1:13, F 1:22 and H 1:5 but are almost always expressed confirming that the lymphomas originate after the immature B-cell stage and they show little evidence of gross restriction. H appears to be Lambda restricted.

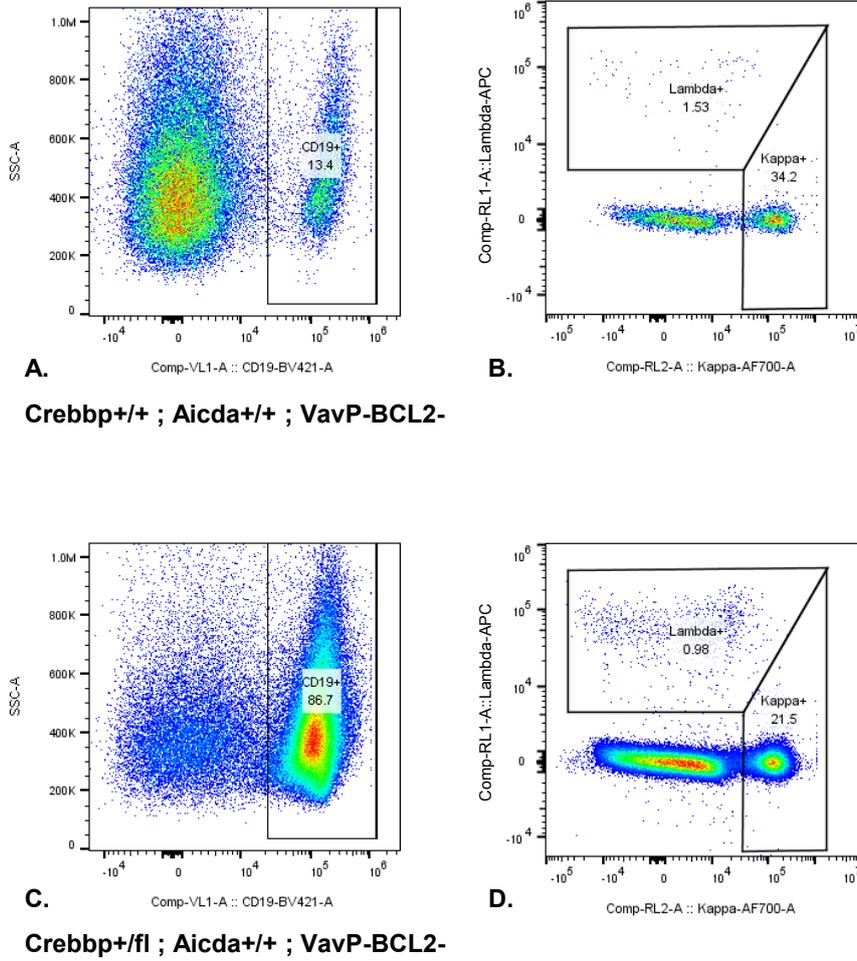


Figure 4.19. MoMuLV induced lymphoma cohort abnormal Lambda/Kappa ratios

Examples of lymphoma with majority Lambda/Kappa double negative cells. **B** (#5097) has 54% double negatives and **D** (#5176) has 79%. This indicates that these tumours are derived from pre-immature B-cells that have infiltrated the spleen or from cells that have non-functional BCRs.

4.4.4. Germinal centre Panel

By assessing CD19+ cells for CD95 and PNA expression the Germinal Centre Panel identifies expansion of germinal centre like B-cells which would suggest a germinal centre phenotype (Figure 4.20).

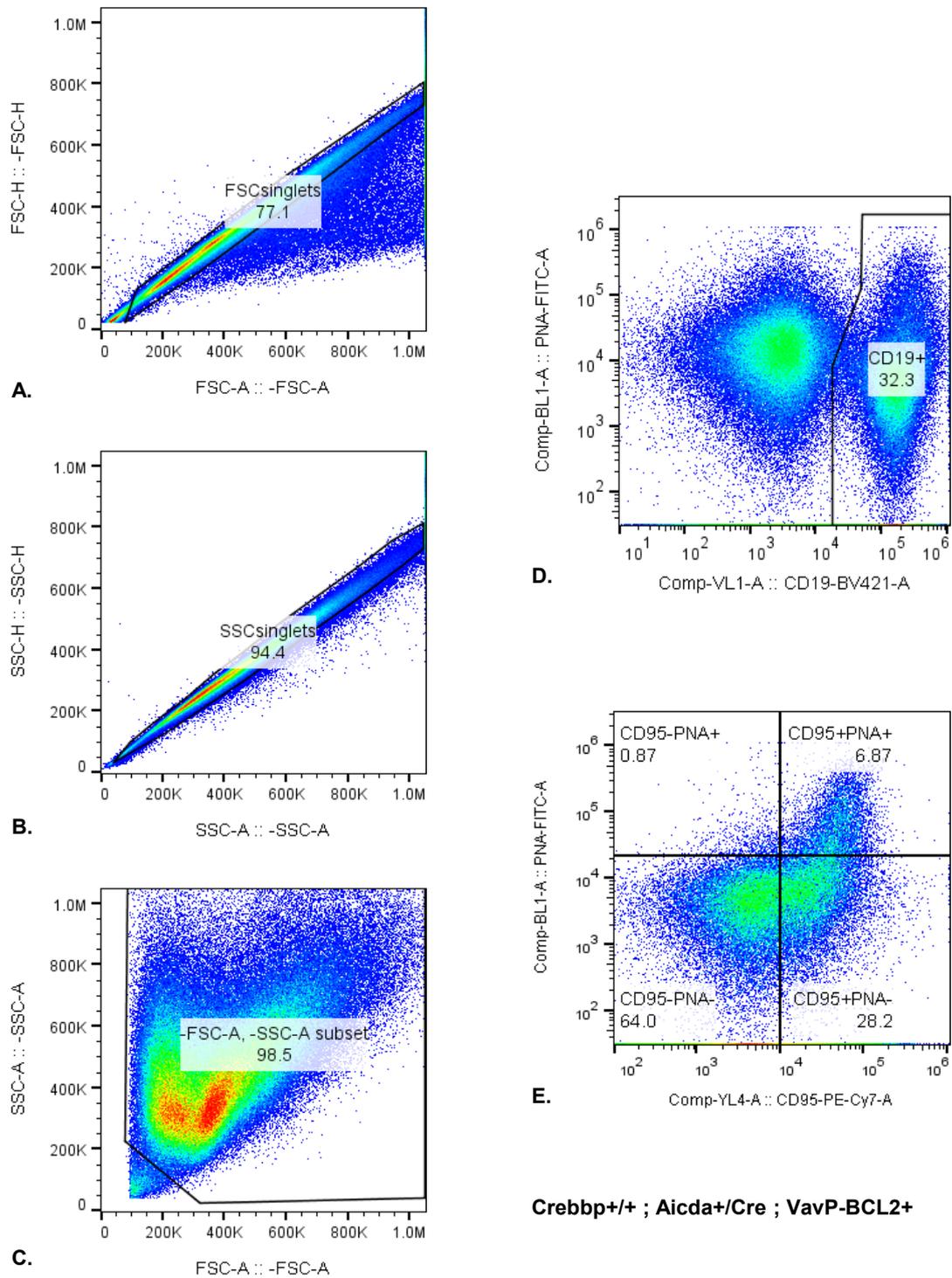


Figure 4.20. MoMuLV induced lymphoma cohort gating strategy for the GCP

A-C The analysis population was selected as described above in section 3.3.1. D The B-cells were then identified as the CD19+ population which was the assessed for GC B-cell proportions in E CD95⁺PNA⁺ cells are considered to be germinal centre B-cells. This example is #5006, Crebbp^{+/+} ; Aicda^{+/-} Cre ; VavP-BCL2⁺.

The tumours had three approximate patterns of CD95 and PNA expression. Figure 4.21. **A-C** are examples of CD95⁻ PNA⁻ samples with very few GC B-cells. **D-F** Have a small population of double positive GC B-cells. **G-I** have the mostly GC B-cells and they have a reduction in their CD95⁻ PNA⁻ cells. In all samples CD95⁻ PNA⁺ cells were rare.

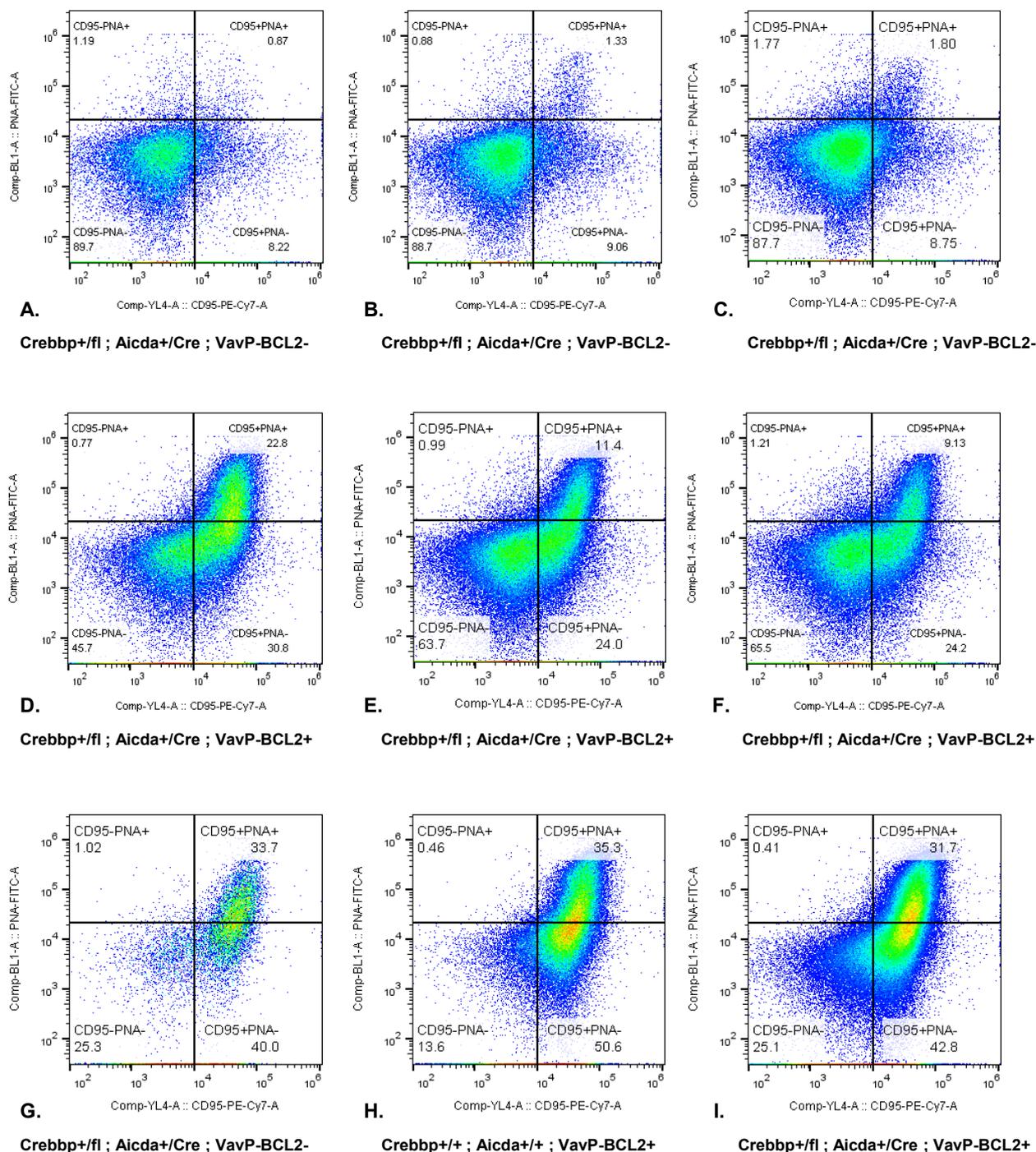


Figure 4.21. MoMuLV induced lymphoma cohort typical CD95 and PNA expression

Within the GC panel there were broadly three CD95 PNA profiles. Above triplicate examples of each of these profiles. **A** (#5279), **B** (#5281) and **C** (#5294) are examples of the CD95⁻ PNA⁻ profiles. **D** (#5254), **E** (#5258) and **F** (#5288) show the crescent shaped plots with varying proportions of cells in each quadrant with less than 2% CD95⁻ PNA⁺. **G** (#5034) **H** (#5179) and **I** (#5255) have predominantly CD95⁺ cells with CD95⁺ PNA⁺ being the largest proportion of cells.

The *BCL2* transgenic tumours have increased proportions of GC B-cells (CD95+ PNA+) and increased class switching (refer to Figure 4.16) recapitulating the pre-malignant model published by (Egle et al., 2004). These alterations are correlated with disease onset and highlight the role of GC and post GC B-cells in lymphoma.

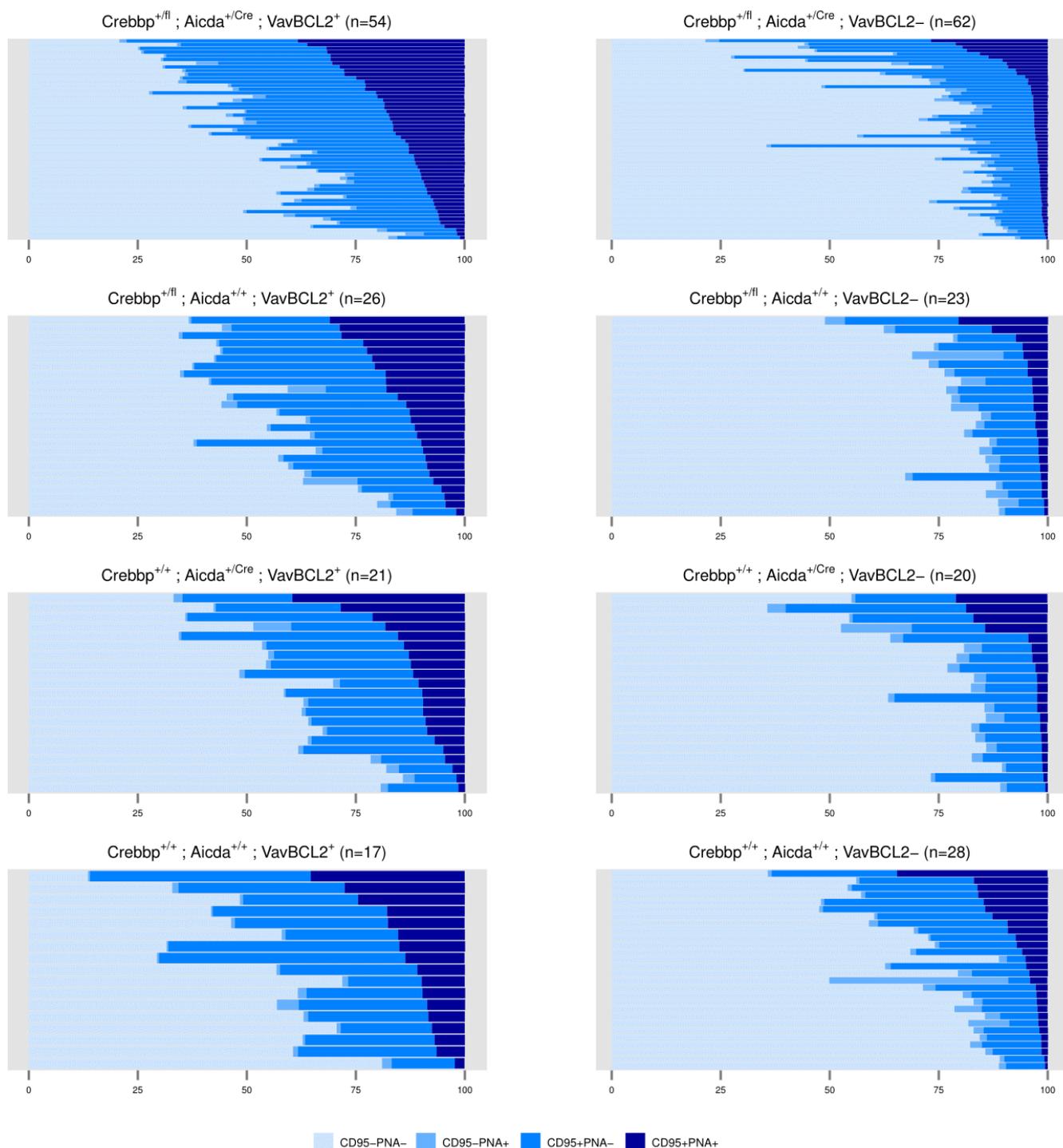


Figure 4.22. MoMuLV induced lymphomas assessment of germinal centres

Division of the Germinal Centre Panel by genotype shows the enrichment of GC B-cell (CD95+ PNA+, dark blue) in the *BCL2* transgenic samples (left hand side) relative to the *BCL2* non-transgenics (right hand side).

4.5. Summary of MoMuLV induced lymphoma

- MoMuLV rapidly accelerate onset of lymphoma.
- The majority of the MoMuLV induced lymphomas are T-cell lymphomas.
- Overexpression of *BCL2* and loss of *Crebbp* does sensitise the mice to B-cell lymphoma with increased incidence in these genotypes.
- The B-cell lymphomas demonstrated expansion of the GC-B-cells, an increase in proliferative cells and increased class switching indicating that they were predominantly germinal centre or post germinal centre derived recapitulating the general phenotype of human FL and DLBCL.

5. Retroviral insertion site cloning and analysis of MoMuLV induced lymphoma

To identify genes mutated by proviral insertions in the MoMuLV induced lymphomas genomic DNA was processed into a pooled library which was sequenced on the Illumina HiSeq using a novel protocol, UMI-LM-PCR (unique molecular identifier ligation mediated PCR).

UMI-LM-PCR was performed on 671 DNA samples, of these 627 were infected with MoMuLV and the remaining 44 were non-infected murine control DNA from the spontaneous lymphoma cohort and human DNA controls. These included the spleen DNA from all the MoMuLV induced lymphomas analysed in this thesis (248), along with the DNA from a subset of purified splenic B- and T-cells, 23 and 21 samples respectively. Additionally 72 enlarged lymph nodes and 51 thymi were processed.

The UMI-LM-PCR method was developed for the Illumina sequencing platform based on a previously published approach using sheared DNA, ligated to adaptors for use on the Roche/454 pyrosequencing platform (Koudijs et al., 2011). Modification of this protocol has greatly increase throughput and yield of IM screens and the inclusion of unique molecular identifiers (UMIs) within the adaptors has improved absolute quantification of insertion sites. It is now possible to prepare a library of 96 tumours in one week and the indexing approach described herein facilitates sequencing of up to 768 samples simultaneously with unprecedented coverage.

From the 671 samples processed 66,292,198 mapped and paired reads were output from the HiSeq which before filtering was equivalent to 627,919 insertion sites of which 398,787 pertain to the cohort of 477 lymphoma samples analysed in this thesis. After filtering and the removal of any contamination this yielded 382,528 inserts. To give this context, the most recently published MoMuLV screen yielded 12,485 retroviral insertion sites from 28 MoMuLV tumours (Huser et al., 2014).

Established oncogenes and tumour suppressors from lymphoid malignancies rank highly in the screen like *Myc* (coding mutations and amplification in FL and DLBCL), *Mycn* (amplified in a subset of CLL and particularly in CLLs that transform to DLBCL), *Ccnd3* (coding mutations and amplification observed in FL and DLBCL), *Notch1* (coding mutations observed in FL, DLBCL and T-ALL) and *Ikaros1* (mutated and or deleted in T-ALL).

This data set represents the largest number of MoMuLV tumours ever sequenced and with the deepest sequencing to date.

5.1. Improvements to library preparation and sequencing

Optimisation of the UMI-LM-PCR protocol and modifications to the Illumina sequencing protocol were conceived by my colleagues Dr. Anthony Uren and Dr. Philip Webster and it will be briefly outlined here. Unique to this thesis is the application of unique molecular identifiers (UMIs) to an IM screen to tag individual insertions sites.

Briefly genomic DNA is sheared, blunted, A-tailed and ligated to custom sequencing adaptors. A primer against the virus LTR is used for the first strand of PCR synthesis and subsequently any fragment derived from this virus LTR primer is exponentially amplified during the primary and nested secondary PCRs. The amplified library is then sequenced on an Illumina HiSeq and the inserts identified analysed (refer to Figure 5.3. page 107 for an overview).

5.1.1. UMI-LM-PCR

The use of sonication to fragment the genomic DNA prior to library construction is an improvement on fragmentation by enzymatic digest in that it generates a continuous distributions of fragments without sequence dependent biases and so does not skew PCR amplification based on the position of restriction sites.

Each library was assessed for size and quality using the 2100_High Sensitivity DNA Assay on the Agilent bioanalyser (see Figure 5.1). The bioanalyser plots demonstrate that sonication is an effective method of fragmenting DNA facilitating the unbiased shearing of DNA into uniform 400bp lengths and allowing optimal PCR amplification.

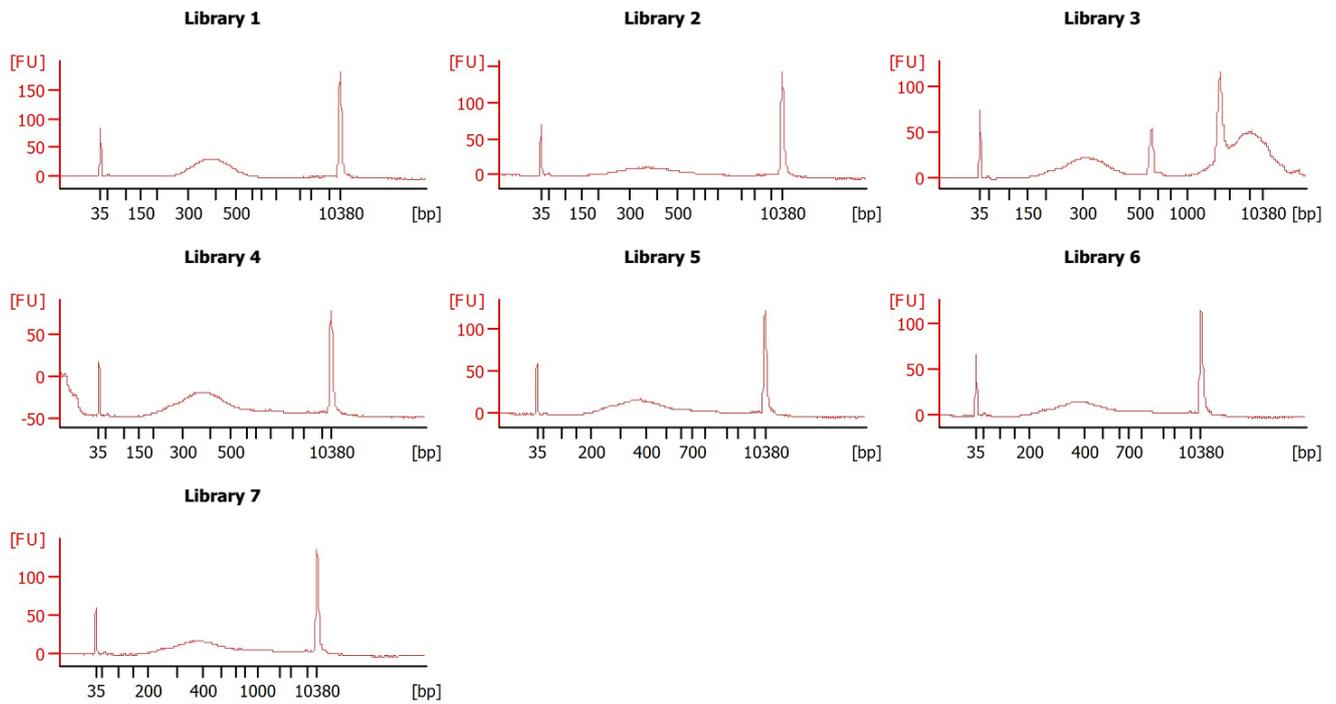


Figure 5.1. Quality control of libraries using the Agilent bioanalyser

This electrophoretic assay measures the size and quality of a DNA sample. The mean library size was 411bp, within the optimal range for sequencing and demonstrating that the shearing and size selection steps were effective. Library 3 highlights the occasional phenomenon where some un-sheared DNA remains in the sample.

The application of UMIs has made it possible to identify each individual DNA fragment enabling the removal of any contamination after ligation of the adaptors. When the reads were aligned to the genome their LTR position was noted. Reads mapped to the same LTR/genome junction with the same UMI were considered to be the same initial fragment PCR amplified multiple times. In previous methods the number of sheared fragments of different lengths was used to quantify the relative abundance of each insertion in the starting library. The same LTR position can have multiple fragments lengths originating from different samples that will have different UMIs (Figure 5.2). If a sample was cross contaminated prior to ligation of the UMIs, multiple reads with exactly the same LTR position in the same orientation, would be found in different mice. These are flagged as potential contaminants and removed in the bioinformatics pipeline

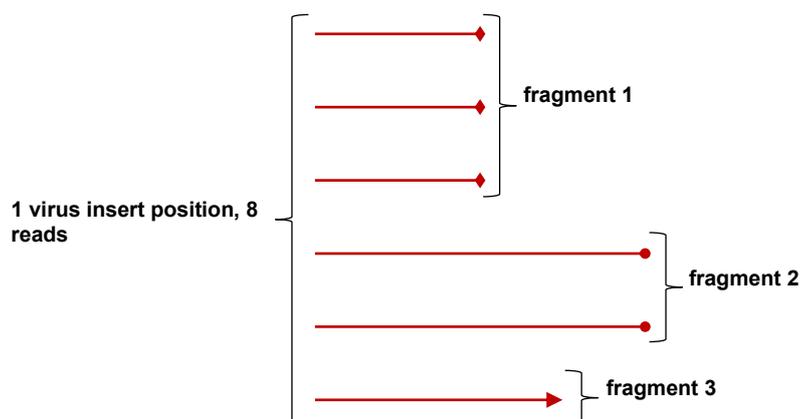


Figure 5.2. Definition of insert, reads and fragments

Paired end reads from the HiSeq are aligned to the mouse reference genome (mm10). Reads that map to within one base of each other at the LTR end (left hand side) are grouped to one insertion with the randomly sheared ends on the right hand side. The example above shows one LTR insert position with six reads. Reads with the same UMI (diamonds, circles or triangles) are assumed to be PCR amplifications of the same DNA fragment and are grouped.

Earlier LM-PCR protocols have used as many as 29 and 25 cycles for the primary and secondary PCRs respectively (Koudijs et al., 2011; Uren and Berns, 2009; Uren et al., 2008). UMI-LM-PCR uses a total of only 32 cycles, divided equally between the primary and secondary PCRs, decreasing the probability of amplification bias or PCR artefact.

Previous methods used column purification to clean the sample DNA of residual enzymes and buffers between library preparation steps and electrophoresis with gel purification to size select DNA fragments prior to PCR enrichment. In UMI-LM-PCR these processes have been substituted with solid phase reversible immobilisation (SPRI) beads that allow cleaning of the DNA when bound to the beads and size selection based on bead buffer ratios which facilitate binding of different fractions. Utilisation of SPRI beads increases throughput by allowing 96 samples to be processed simultaneously in a plate using a Beckman Biomek Robot. The reduction in manual handling of the DNA also decreases the probability of the samples becoming contaminated.

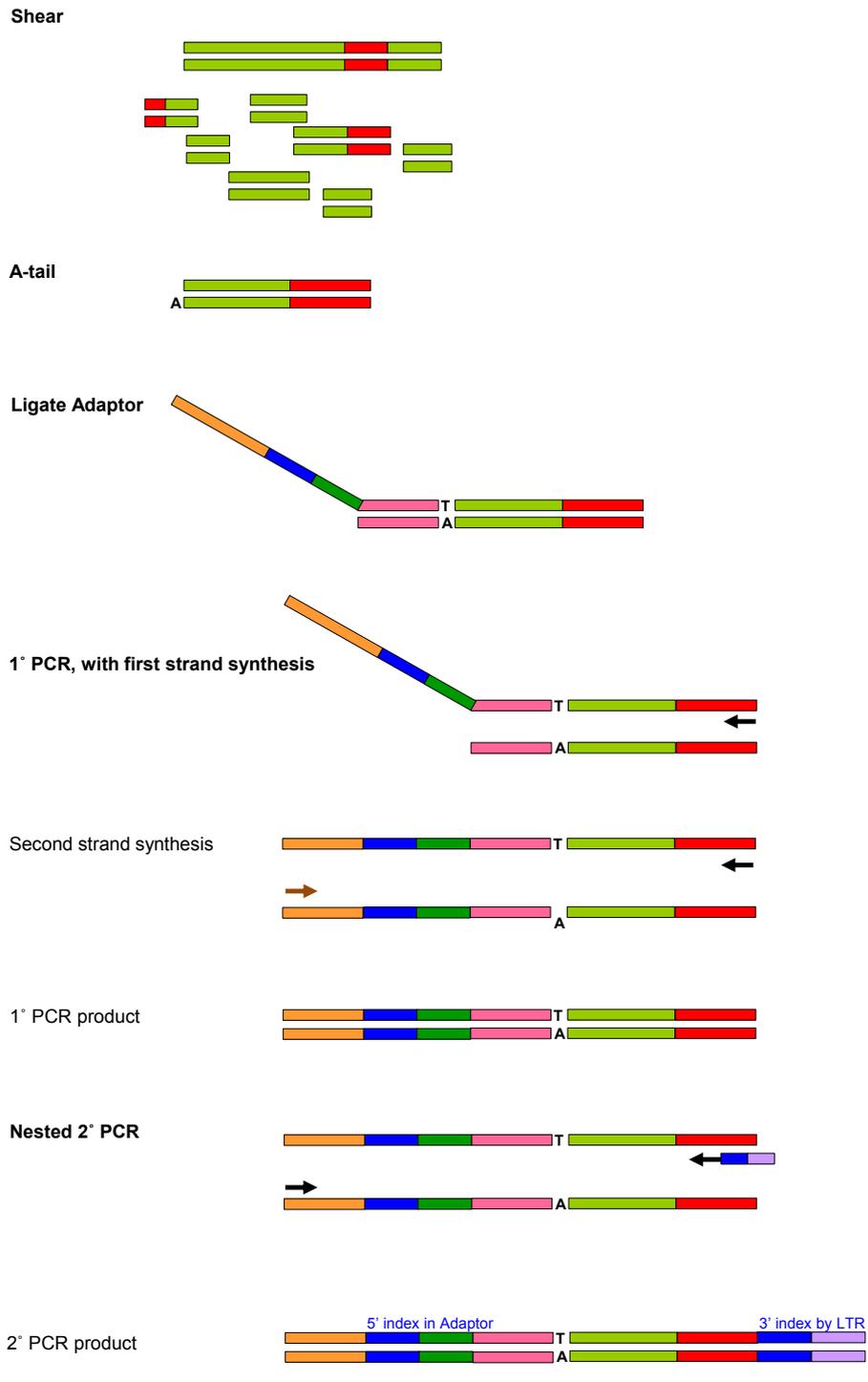


Figure 5.3. Schematic of UMI-LM-PCR

Overview of the library preparation protocol. First the mouse DNA (light green) was randomly sheared by sonication; by chance some fragments will contain MoMuLV integrations (red). The DNA was then A-tailed to provide a complementary overhang for adaptor ligation. The adaptors contain: a motif that is complementary to Illumina Oligo C (orange) and permits hybridization of the DNA to the HiSeq flow cell, a unique 10bp 5' index (blue) and an 8bp unique molecular identifier (UMI) (green). The pink portion of the adaptor is the only complementary portion of the primer leaving a non-complementary overhang containing the aforementioned motifs. During the primary PCR, first strand synthesis enriches for DNA with MoMuLV integrations. The primers were designed so that only the MoMuLV LTR (red) primer could bind and amplify the DNA in the first PCR cycle whereas the second, adaptor primer (orange) can only bind once the complementary portion of the adaptor overhang has been synthesised (orange portion of template) after the first cycle of PCR. The secondary PCR further enriches for MoMuLV inserted DNA by using a nested LTR primer further in towards the mouse DNA-MoMuLV junction. This primer also adds a second, 3' index (blue) and a portion complementary to Illumina Oligo D (lilac) for binding to the sequencing flow cell (vertical line). When the secondary PCR product is sequenced by HiSeq, Read 1 starts sequences from the 5' index end towards the chip.

5.1.2. Sequencing

The completed libraries were quantified using the KAPA Illumina SYBR Universal Lib Q. Kit, a qPCR based method that contains primers complementary to the sequences that bind the Illumina flow cell (oligo C and D) which flank the PCR product. The assay quantifies the proportion of fragments in the library that can be amplified on the flow cell and hence dictates the loading concentration. Equal quantities of each library (2nmol), which contain 96 samples, were pooled (671 samples in total) and then the Illumina HiSeq was performed by the CSC Genomics Facility, a dual-index, paired-end protocol was used.

In UMI-LM-PCR the 5', sheared end index is on the upper strand within the adaptor rather than on the lower strand as in the Illumina protocol. This inversion ensures that during first stand synthesis in the primary PCR, which only amplifies fragments with MoMuLV inserts, that all products are indexed. There is no indiscriminate amplification of product without viral integration sites, thus conserving reagents in the PCR reaction (Figure 5.3.).

The adaptor orientation was designed such that the sequencing originated from the sheared end of the DNA. This guarantees the greatest diversity of bases in the initial 10 cycles of sequencing, which is a requirement of the Illumina platform for efficient cluster recognition. To achieve this, the flow cell binding adaptor sequences "Oligo C" (orange) and "Oligo D" (purple) were exchanged so that Illumina oligo D binds the DNA template to the chip initially so that when Read 1 binds, it sequences from the sheared end of the DNA towards the LTR.

In order to increase throughput and enable pooling of 96 samples per library, Illumina indices were not used; at the time of starting the project only 24 indices were commercially available. In designing adaptors, ten base pair indices were adapted from 25mer indices published by (Xu et al., 2009) to generate 96 unique indices used in the upper strand adaptor. Furthermore a dual indexing protocol for library preparation was devised so that both ends of the PCR product were indexed. Twelve secondary PCR indices were used per 96 sample library and were distributed so that the same index was never directly adjacent to itself on a 96 well plate. This produced 768 unique index pairs facilitating pooling of up eight libraries. Each set of twelve secondary PCR indices was specific to one library of 96 samples facilitating identification of any inter-library contamination.

5.1.3. Quality Control of UMI-LM-PCR

After demultiplexing a total of 66,292,198 reads were paired and mapped. Reads corresponding to the same genome-LTR position in the same sample were merged into a single insertion giving 25,070,149 unique reads before filtering. The subset of tumours analysed in this thesis yielded 398,787 inserts of which 4% were removed after filtering (refer to Figure 5.4).

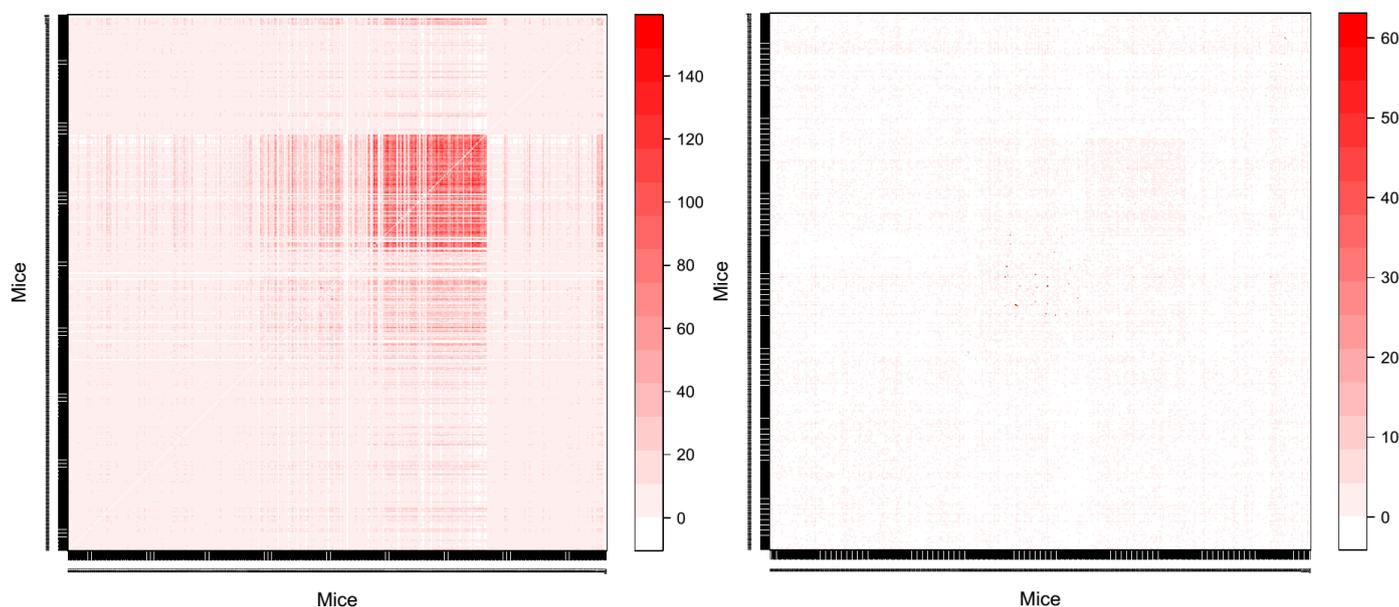


Figure 5.4. Heat maps comparing the number of insertion sites shared between mice

The inserts from each mouse were compared pairwise and the number of inserts that they had in common were plotted as a heat map where darker red indicates a greater number of identical insertions between mice. Prior to filtering (left hand side) there is an area in darker red indicating that these mice shared multiple inserts which is statistically unlikely and indicative of contamination. This degree of overlap between two mice would indicate that some DNA with those inserts was spread between multiple samples perhaps in a contaminated solution and so they should be removed. Other recurrent identical insertions may represent PCR artefacts. After application of the filtering pipeline, the majority of these shared inserts were resolved.

Human thymic DNA and non-MoMuLV infected mouse spleen DNA was included in each library to monitor levels of sample cross contamination or background insertion levels. Occasionally the control mouse DNA had inserts; these represent endogenous sequences that bound the PCR primers (in some cases these are endogenous germline retrovirus sequences) and were removed in the filtering pipeline. None of the human DNA had spurious insertion sites.

The following criteria flagged an insert as suspicious: any insert that occurred in more than 10 mice (assumed to be a PCR artefact or the result of cross contamination prior to adaptor ligation) and any three or more inserts were shared by a pair of mice.

The contaminant list was then manually curated in order to rescue a small number of loci that are known to be mutated frequently enough that repeated insertions at the same position are reasonably likely, for example all *Mycn* inserts are rescued unless their exact location was shared by two or more mice. There were also several suspicious regions containing large numbers of subclonal insertions in continuous runs,

like: chr9:3000251-3000895, chr9:3024394-3024454, chr21:4795984-4795984, chr2: 98662236-98667333, Chr12:3109866-3110127. These were presumed to be PCR artefacts and also manually removed.

The reproducibility of UMI-LM-PCR was evaluated by processing the same DNA sample multiple times in different libraries. The top ten most clonal inserts of a mouse were monitored for relative clonality in each library when they were sequenced (Figure 5.5).

Generally the relative clonality of the inserts remains the same between the different plates indicating that the method is quantitative and reproducible.

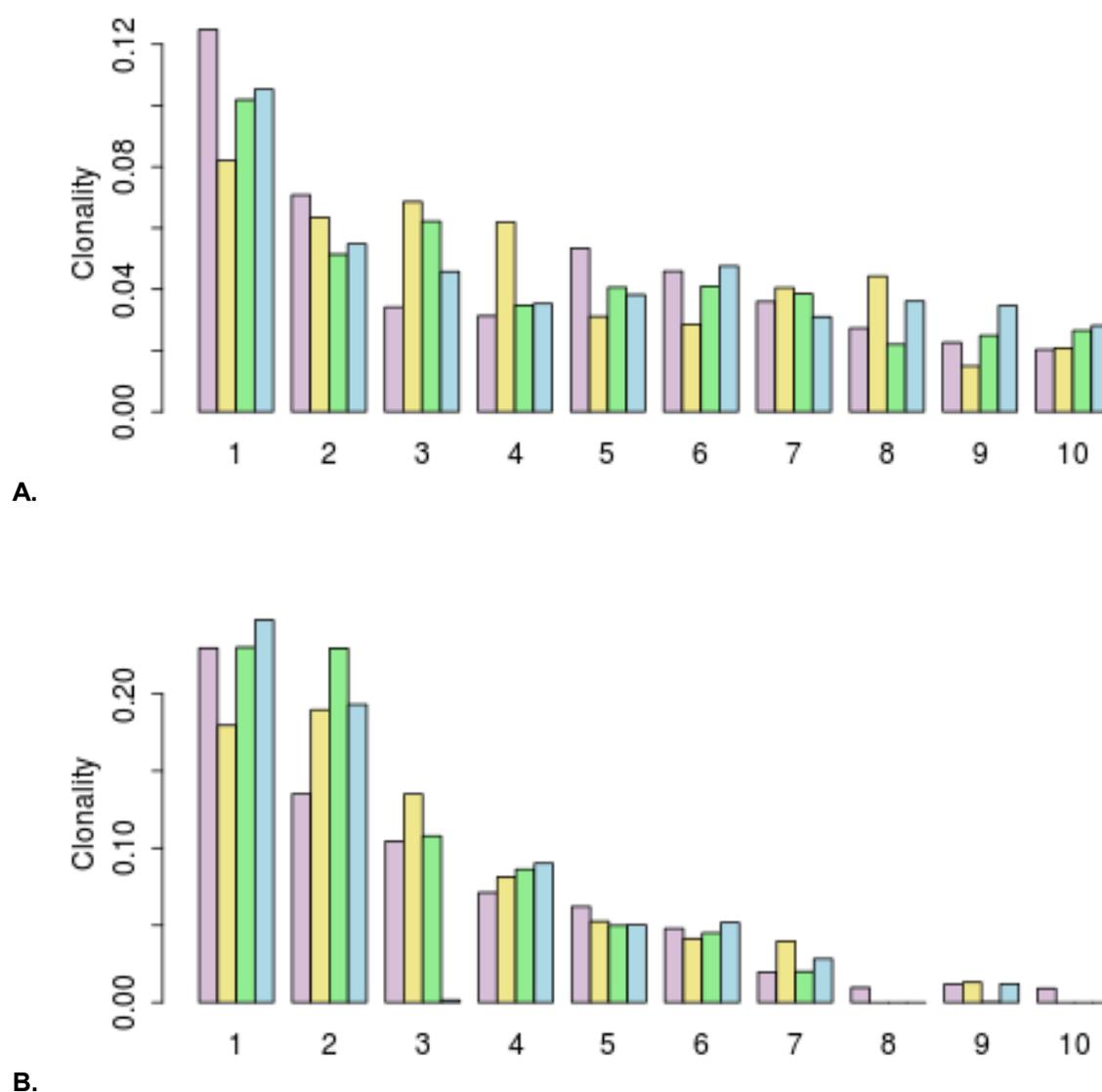


Figure 5.5. Reproducibility of UMI-LM-PCR

Spleen DNA #5165 (A) and #5275 (B) were processed in four different libraries to evaluate the reproducibility of UMI-LM-PCR. The top ten most clonal inserts from each tumour were assessed for their clonality in the different libraries. The relative clonality of the inserts broadly stays the same between each different libraries. Analysis done using the inserts by sample table so that each replicate was considered separately.

5.2. Characterisation of inserts

The bioinformatics pipeline was designed and implemented by the Cancer Genomics group's bioinformatics, Barbara Iadarola, and generates three inserts tables:

1. Inserts built "by sample" where each sample processed by UMI-LM-PCR is considered unique. This was used in the reproducibility analysis.
2. Inserts built "by DNA sample" where repeats of the same sample were merged but separating the inserts from different tissues from the same mouse. The majority of analysis were performed on this group.
3. Inserts built "by mouse" where all tissues samples from one mouse are merged together, this combines all inserts from all organs that mouse.

The samples were then assessed for clonality and entropy. **Clonality is a measure of relative abundance of an individual insert within a DNA sample.** *Clonality = number of fragments ÷ the total number of fragments in the sample.* **Entropy is the measure of disorder within the clonality of a tumour.** A sample will have high entropy if it contains many oligoclonal inserts of equal abundance and a low entropy if it has only a few clonal inserts.

Plotting the normalised clonality of the top 50 most clonal inserts in each tumour generates the plots in Figure 5.6 allowing comparison between samples with differing clonality. The end stage lymphomas have developed into full blown disease and have undergone some degree of clonal selection. These samples have small numbers of inserts with high normalised clonality with a background population of subclonal insertions. After examining these plots, the high entropy samples lacking evidence of clonal outgrowth (with entropy > 3.6 (blue, left hand side)) were separated from the end stage lymphomas with clonal outgrowth (those with entropy \geq 3.6, (red, right hand side)).

From the low entropy samples (entropy \leq 3.6) the inserts are again divided by clonality. A highly clonal insert is defined as one with a normalised clonality score \geq 0.1 which corresponds to the most vertical part of the graph. Clonal inserts occur most frequently and have undergone a greater degree of selection making it likely that they promote lymphoma.

Subsets of spleen inserts, for example, will be described using the following terms:

All inserts in the screen > all spleen inserts (entropy > 3.6) > all end stage inserts (entropy \leq 3.6) > all end stage clonal inserts (entropy < 3.6 and clonality \geq 0.1).

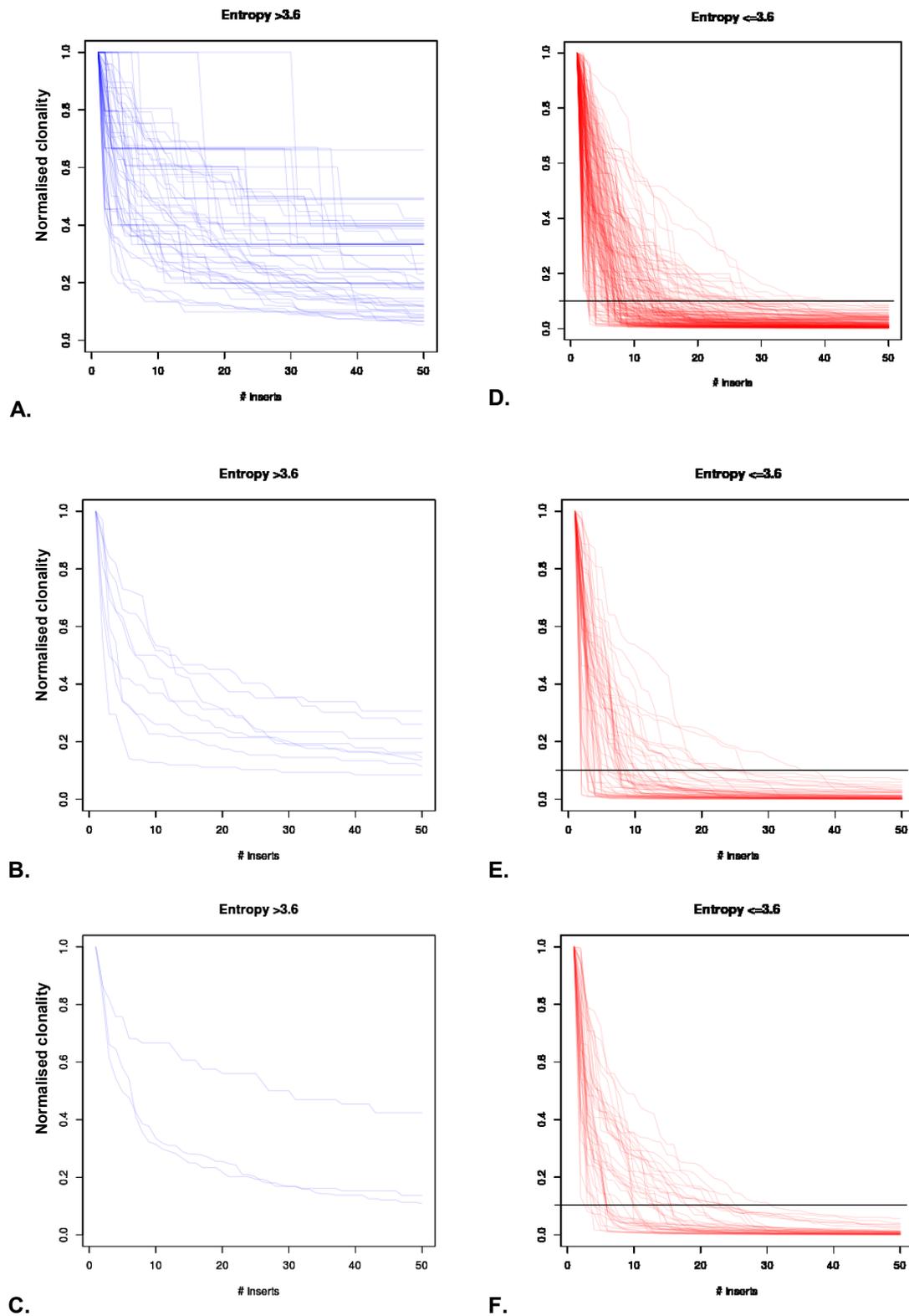


Figure 5.6. Entropy scoring

Comparison of normalised clonality of the top 50 most clonal inserts within each tissue with each line representing one sample. On the left hand side (blue), samples with high entropy > 3.6 , in the spleen (A), lymph node (B) and thymus (C) with disordered profiles. Right hand column (red) shows more selected for tumours with low entropy (≤ 3.6) in the spleen (D), lymph node (E) and thymus (F).

Tumours with low entropy have undergone a greater degree of clonal expansion.

5.3. Common insertion sites

Tumours with entropy ≤ 3.6 and inserts with clonality ≥ 0.1 are input into the KCRBM R package to identify common insertions site (CIS) i.e. regions of the genome more frequently mutated than expected by chance. The presence of recurrent insertions in different independent tumours suggests that there is positive selection for their mutation and that it promotes lymphoma. To identify regions of the genome that are mutated more frequently than expected by chance we used a method that compares the observed and expected density of insertions across the genome (de Ridder et al., 2006). This method is described in full in section 2.8, page 39.

Briefly a Gaussian kernel is plotted about each insertion. Overlapping kernels are summed giving a density distribution (convolution) and the insertion density is calculated. The density of insertions over the genome is compared to 1000 permutations of a random distribution of inserts of equal number. This yields a p-value that is adjusted using a Bonferroni correction for false discoveries in multiple testing and those smaller than 0.05 are defined as CIS (de Ridder et al., 2006). For automated assignment of insertions to genes, each insertion is assigned to the nearest peak and all inserts assigned to the same peak are merged into a cluster. The mean of each cluster is then assigned to a gene based on its position relative to nearby genes. This method was implemented using an R package described in (de Jong et al., 2011).

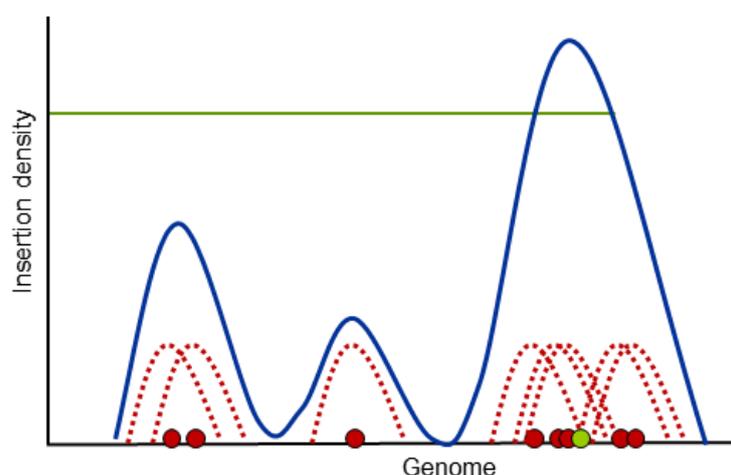


Figure 5.7. Schematic of the Gaussian Convolution

The inserts are mapped to the mouse genome (red dots). A Gaussian kernel is plotted about each insertion (red curves) and then overlapping kernels are summed to give a density convolution (blue line). This convolution is compared to 1000 random convolutions of an equal number of random insertions across the genome and assigned a p-value. If the density distribution surpasses a statistically established threshold the mean of the convolution (green dot) is defined as a CIS and considered for its role in oncogenesis.

Each gene in the genome is assessed for the presence of insertions within its vicinity using four windows sizes (upstream-sense and antisense, downstream-sense and antisense, with respect to its transcription start site). If a cluster of insertions falls within one of these windows it is assigned to that gene.

Application of KCRBM allows prediction of the most frequently inserted genes in the screen which will include oncogenes and tumour suppressor with a putative role in lymphoma. Table 5.1. lists the top 100 most frequently mutated genes in this IM screen when considering tumours with entropy ≤ 3.6 and inserts with clonality ≥ 0.1 i.e. the most selected for tumours and insertions. Almost 300 significantly mutated genes have been identified in the end stage clonal tumours all of which could have a role in lymphomagenesis.

Some common IM insertion site like *Rasgrp1* and *Gfi1* (Huser et al., 2014; Kool et al., 2010; Uren et al., 2008) and known oncogenes and tumour suppressors like *Myc*, *Notch1* and *Ikzf1* (see Table 5.1) feature in the list however there are also novel genes specific to this screen that perhaps cooperate with the unique sensitising mutations (*BCL2* overexpression and or *Crebbp* loss) or the spectrum of disease they produce.

Chromosome	Peak Position	CIS Gene Name	Insert Count 100k	P-Value
5	107737975	Evi5	78	0.00E+00
12	12933473	Mycn	67	0.00E+00
15	62020371	Myc, Pvt1	35	0.00E+00
2	26488418	Notch1	23	0.00E+00
2	117407181	Gm13982	20	0.00E+00
11	11702820	Ikzf1	18	0.00E+00
17	47527783	Taf8	16	0.00E+00
7	114175415	Copb1	14	0.00E+00
7	25134886	Grik5	11	0.00E+00
5	33685619	Slbp	10	0.00E+00
17	29493028	Mtch1, Pim1	9	0.00E+00
19	37494510	Exoc6	9	0.00E+00
4	134107056	Cd52	9	0.00E+00
1	138170299	Ptprc	9	0.00E+00
20	52743681	Mir106a	8	0.00E+00
6	127262427	Gm7308, Ccnd2	8	0.00E+00
7	129784192	NA	8	0.00E+00
12	107141424	Gm16086, Gm2800	7	0.00E+00
13	28538833	Gm12061	6	0.00E+00
5	65901808	2610307P16Rik	6	0.00E+00
7	73586721	Rhoh	6	0.00E+00
4	133785600	7SK, Chd2	6	0.00E+00
10	21155410	Arid1a	6	0.00E+00
13	30776900	Myb, Ahi1	5	0.00E+00
9	61185719	Dusp22, Exoc2	5	0.00E+00
1	156946140	NA	5	0.00E+00
20	7864509	Pik3r5	5	0.00E+00
9	32604198	NA	5	0.00E+00
4	136103985	Kcnd1	4	0.00E+00
7	127277387	Gfi1b	4	0.00E+00
4	135042203	9330198N18Rik	4	0.00E+00
11	23792442	Fli1, Ets1	4	1.61E-06
13	52864825	Ggta1	4	1.84E-06
18	54965566	Polh	4	2.50E-06
18	60850244	NA	4	2.50E-06
8	120472974	Rfx2	4	3.45E-06
2	167795306	NA	4	3.73E-06
20	13257592	Enthd1	4	5.50E-06
11	68446043	Msi2	4	8.88E-06
7	126386296	Gm8849	4	1.02E-05
3	52436290	Prkab1	4	1.20E-05
3	30062301	Card14, Cbx8	4	1.61E-05
10	19618783	Gse1	4	1.88E-05
7	144991488	Wdr66	4	2.04E-05
15	63279993	Rabep2	4	2.69E-05
8	33657494	Rpl11	4	2.93E-05
2	35498473	Gm14319	4	4.03E-05
17	46270362	NA	3	4.34E-05
2	28611869	U6	3	4.48E-05
16	91486315	Zfp438	3	5.01E-05
12	107739858	Cdc42bpg	3	5.12E-05
3	51765630	Hs1bp3, NA	3	5.62E-05
19	45598545	Gm7129	3	6.27E-05
1	121618039	Arhgef10	3	6.42E-05
8	14978748	Ptgds	3	6.56E-05
6	34905329	2610307P16Rik, Sox4	3	7.53E-05
14	25446835	Mecom	3	8.77E-05
5	113874856	Lrrc33	3	8.81E-05
2	30647654	Gsr	3	8.95E-05

14	115035193	Cxcr5	3	9.22E-05
17	56839731	NA, Gm13475	3	1.11E-04
1	171926559	Maml3	3	1.43E-04
6	119352919	Tcf7, Vdac1	3	1.60E-04
5	116031058	Smndc1	3	1.61E-04
11	119144455	Zfp608	3	1.65E-04
5	123319215	Psmc6	3	1.66E-04
13	28944896	Tcof1	3	1.68E-04
2	165954333	Rgs3	3	1.69E-04
15	95751298	NA	3	1.75E-04
8	122740001	Chsy1, Lrrk1	3	1.81E-04
15	80549304	Sla	3	1.92E-04
18	5347030	Egr2, NA	3	2.13E-04
3	107328014	Pgpep1l	3	2.14E-04
3	101673971	Mettl21d	3	2.57E-04
18	4321669	Alms1, Gm4477	3	2.68E-04
3	131034011	Celsr1	3	3.16E-04
20	103527759	Clns1a, Pak1	3	3.16E-04
11	88608412	Galc	3	3.44E-04
19	4224848	Med24	3	3.72E-04
16	32153680	Tns4	3	3.72E-04
9	44526110	Stat3	3	4.68E-04
4	62655988	NA, Atp1a1	3	4.80E-04
12	85625017	Sic16a4	3	5.56E-04
4	149683097	NA	3	5.61E-04
15	97760923	BC022687, Akt1	3	6.91E-04
15	86030099	Coro1c	3	7.73E-04
14	118909234	BC057022	3	7.74E-04
3	68737074	Mir18	3	9.65E-04
11	115478570	Atp5h	3	1.01E-03
19	6319050	Sep-09	3	1.05E-03
10	67494311	NA, Mdm1	3	1.07E-03
6	85701733	Fbxl14	3	1.11E-03
10	118087746	NA	3	1.11E-03
2	25465655	Metrn1	3	1.22E-03
7	68279346	Cbfa2t3	3	1.23E-03
4	140740755	Itgal	3	1.28E-03
6	71323412	NA	3	1.31E-03
11	4332051	Fgfr2	3	1.32E-03
7	132800191	NA	3	1.34E-03
17	52170567	Mad11l	3	1.60E-03

Table 5.1. Top 100 CIS in the screen

Listed, the top 100 most frequently inserted genes in the spleen (with entropy ≤ 3.6 and inserts with clonality ≥ 0.1) as calculated by KCRBM and ranked by the number of insertions about the gene peak in a 100kb window. They are all very significantly enriched with all $p < 0.006$.

5.4. Subclonal mutations in B-cell lymphoma

Using the flow cytometry data (section 4.4) the MoMuLV induced tumours were divided into B- or T-cell lymphomas based on their percentage of B-cells. Below are histograms highlighting the division between samples with high and low percentage of CD19 positive cells. Samples with approximately 50% or more B-cell were classified as B-cell lymphomas. This representation reinforces the observation that the spontaneous lymphomas which were predominantly B-cell lymphoma with approximately 85% incidence whereas there were only 37% in the MoMuLV derived lymphomas. (Figure 5.8).

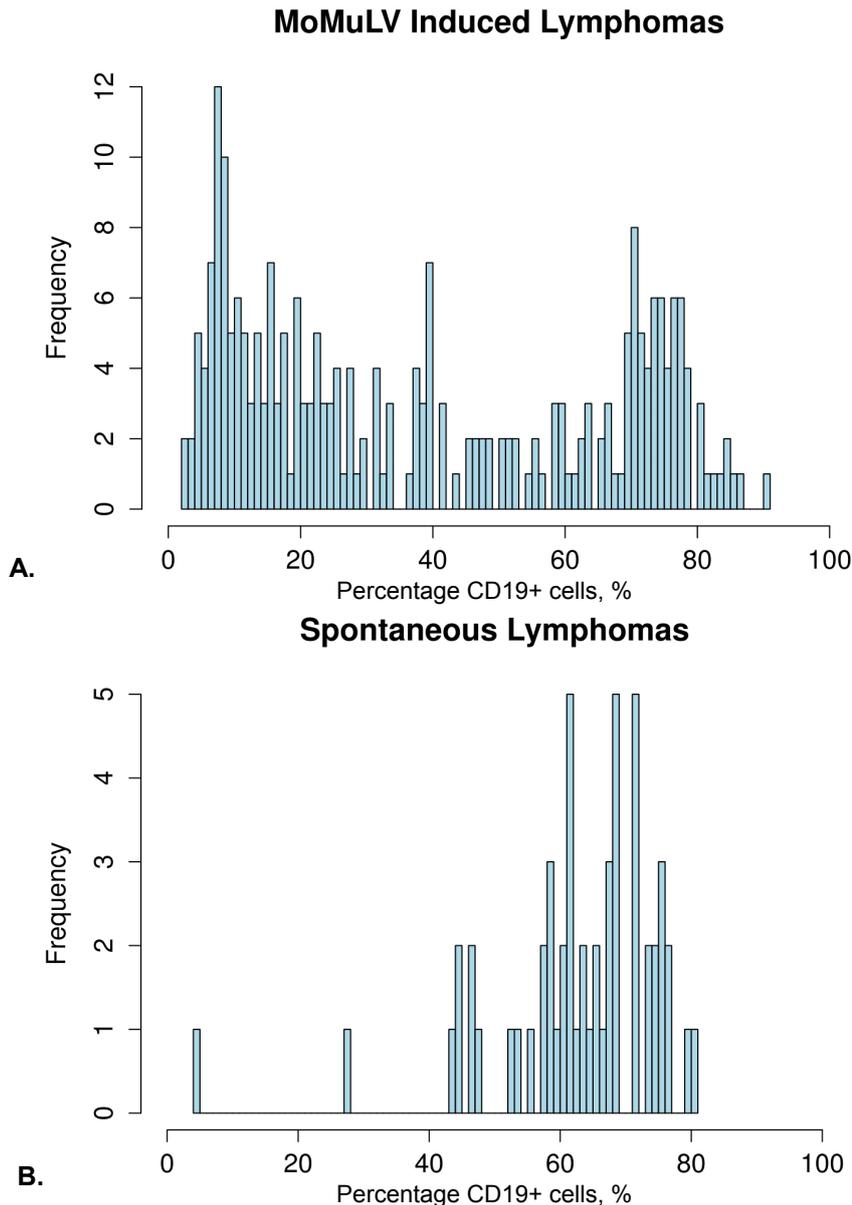


Figure 5.8. Histogram describing the of B-cell lymphomas

The percentage B220+ CD19+ B-cell was compared in the tumours with those containing more that 50% B-cell classified as B-cell lymphomas. **A** details the MoMuLV induced lymphomas which are a mixture of B-cell and T-cell origin whereas the spontaneous lymphomas **B** are predominantly B-cell in origin.

To increase statistical power, by increasing the number on inserts analysed, all end stage tumours with entropy ≤ 3.6 and all insertions (clonal ≥ 0.1 and subclonal < 0.1) were included to generate a list of B-cell lymphoma specific genes (Table 5.2). Over 1000 genes have been identified to be significantly mutated in the B-cell lymphomas.

Comparison of Table 5.1 and Table 5.2 highlights that there are differences in the most frequently mutated genes. This indicates that there is subgroup of genes whose mutation contributes directly to B-lymphomagenesis which differ from those the whole list which represents predominantly T-cell lymphoma genes as there are the major fraction of the tumours.

Notable loci that are more highly ranked in the B-cell specific subset include *Tfrc* (known to overexpressed in human B-cell lymphoma (Chitambar et al., 1983), *Pou2f2*, *Kit*, *Irf8* (mutated in B-cell lymphoma) (Zhang et al., 2013)

Chromosome	Peak Position	CIS Gene Name	Insert Counts 100kb	P-Value
9	3013254	Mir101c	141	<0.0001
7	25126195	Pou2f2	133	<0.0001
2	11629964	Il2ra	82	<0.0001
11	98489581	Zpbb2	80	<0.0001
5	75809332	AC134903.1, NA	73	<0.0001
17	47539041	Ccnd3	72	<0.0001
15	74949504	Cyp11b2, Ly6e	70	<0.0001
4	134116257	Gm7534	70	<0.0001
5	107740022	Evi5	68	<0.0001
17	35177569	Csnk2b	66	<0.0001
16	49804379	Gm15518, Cd47	59	<0.0001
3	30096434	Mecom	57	<0.0001
11	100865326	Stat5b	57	<0.0001
6	48702439	Gimap9	55	<0.0001
1	86471568	Ncl, Ptma	53	<0.0001
12	12933704	Mycn	52	<0.0001
8	120749733	7SK, RP23-80D9.1	50	<0.0001
6	129242094	AC142191.1, Cd69	48	<0.0001
12	113265845	Ighg2c, U6	47	<0.0001
8	126805170	NA	47	<0.0001
2	165942958	Gm11461	47	<0.0001
1	171645259	Cd48	47	<0.0001
11	116400543	Exoc7, Prpsap1	46	<0.0001
19	53462450	Smndc1, Dusp5	44	<0.0001
13	56157378	SNORA17	44	<0.0001
5	100614830	Coq2, Plac8	44	<0.0001
2	132036458	Rassf2, Slc23a2	44	<0.0001
15	73721742	Ptp4a3	41	<0.0001
15	80571876	Enthd1, Fam83f	41	<0.0001
2	170218211	Zfp217, Gm14270	41	<0.0001
19	5828852	Frmf8	40	<0.0001
11	23785617	Rel, Gm12061	40	<0.0001
2	26534245	Notch1, Egfl7	40	<0.0001
17	29439356	Mtch1	40	<0.0001
5	65912685	Chrna9	40	<0.0001
8	70660097	Lsm4	40	<0.0001
7	73595788	U6, Chd2	40	<0.0001
11	87744959	Supt4h1	40	<0.0001
19	4185655	Tbc1d10c	39	<0.0001
10	80673531	Btbd2	39	<0.0001
7	110955518	Mrv1	39	<0.0001
4	136073033	Gm13008	39	<0.0001
11	117309779	Sep-09	38	<0.0001
2	117398149	Gm13982	38	<0.0001
14	25506476	AC157650.1, Zmiz1	37	<0.0001
7	25665509	Bckdha	37	<0.0001
19	32263302	Sgms1	37	<0.0001
10	40469239	Cdk19	37	<0.0001
12	86865214	NA, 2310044G17Rik	37	<0.0001
1	157008031	Ralgps2, NA	37	<0.0001

Table 5.2. Top 50 B-cell lymphoma specific CIS

Listed, the top 50 most frequently inserted genes in the spleen B-cell lymphomas (with entropy ≤ 3.6) as calculated by KCRBM and ranked by the number of inserts within the 100kb window about the mean Gaussian peak position.

5.5. Analysis of lymphoma dissemination

Where multiple tissues from the same mouse were analysed, the ten most clonal inserts from each organ were compared to evaluate similarities and differences between organs. On the most simplistic level, two organs with totally different clonal insert represent two different diseases with different initiating events, whereas organs with the same inserts are likely to be the result of the same primary event and represent the same disease that has metastasised to a second site.

Figure 5.9. compares the most clonal inserts in the lymph node, spleen and thymus of the same mouse and highlights the variety of ways that disease presented. The first quadrant in the top panel (top left) is ordered by the ten most clonal inserts in the lymph node with their respective clonality in the spleen (middle quadrant) and thymus (right hand quadrant). The middle panel is ordered by clonality of the spleens inserts and the bottom panel is ordered by the clonality of the thymus inserts.

All three organs in **A** share the same most clonal inserts indicating that all organs have the same disease originating from the same primary clone. One of the most clonal insert in the spleen is noticeably absent in the thymus (and not top ten in the lymph node). This may suggest that it is a mutation that was gained after dissemination to the lymph node and spleen and that the thymus is the primary site of disease. The lymph node and spleen share all of the thymus inserts showing their common origin except for that insert.

The organs in **B** have very few clonal insertions compared to **A** and very few subclonal inserts demonstrating a high degree of selection in these tumours. The lymph node and the thymus share the same clonal disease whereas the spleen has a second totally independent clonal disease.

The primary, most clonal tumour in **C** is in the spleen. A clonal insert appears to have disseminated to the lymph node and thymus where it has generated related but less clonal diseases.

The lymph nodes and spleen in **D** share a common disease insertion profile with a second disease in the thymus that is quite distinct from the other two organs except for one common insertion that has been enriched in the thymus. This could indicate that there were two independent primary clones within the same mouse that developed into two different diseases.

In **E** the primary site of lymphoma seems to be the lymph nodes where the mutations are most clonal and seem to have had time to undergo selection increasing clonality. The spleen and thymus have similar less clonal disease that seems to originate from and share insertions with the lymph node tumour.

The spleen and lymph nodes exhibit similar disease in **F** with a second distinct disease in the thymus. The most clonal disease is in the thymus but perhaps it is the secondary sites of disease and represents a highly clonal tumour from the lymph node and spleen that had the necessary mutation to migrate.

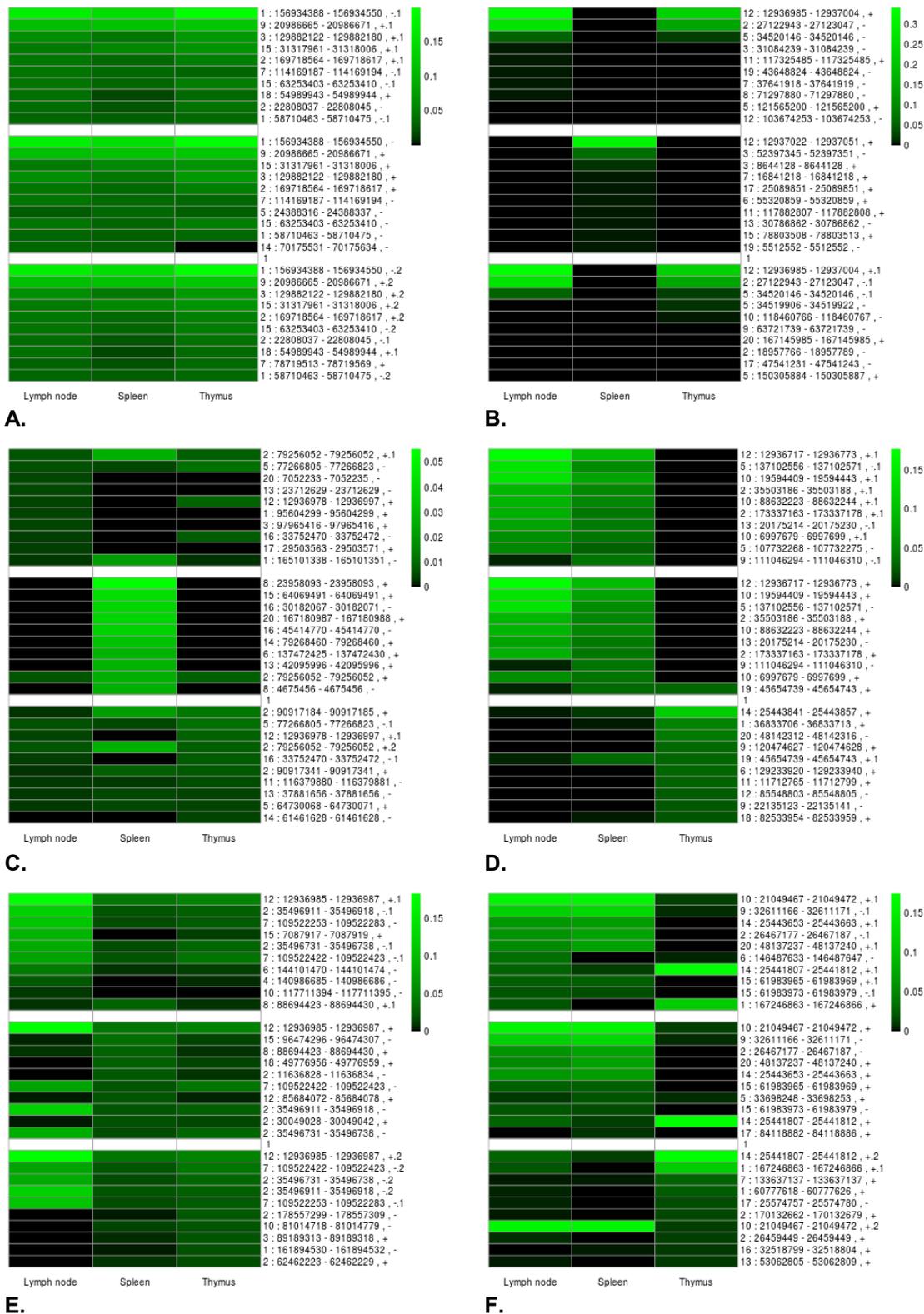


Figure 5.9. Comparison of top ten most clonal insertions in different organs of the same animal

All organs in **A** (#5154) exhibit the same disease. The spleen in **B** (#5036) has a distinct disease to that shared by the lymph nodes and thymus. The spleen inserts in **C** (#5229) are most clonal and seem to have given rise to the less clonal disease in the lymph node and thymus. The spleen and the lymph nodes share a common disease in **D** (#5157) with a distinct second disease in the thymus. The primary tumour seems to be in the lymph node in **E** (#5231) with a common less clonal disease in the spleen and thymus. The lymph node and spleen in **F** (#5265) share the same disease with a distinct more selected for tumour in the thymus. Human tumours tend to be less polyclonal than these murine lymphomas.

5.5.1. Analysis of cells specific insertions within the spleen

With the intention of investigating mutations specific to the B-cells of B-cell lymphomas or the T-cells of T-cell lymphomas tumour spleen suspensions were bead purified to produce uniform populations of cells from which the DNA was extracted and processed by UMI-LM-PCR and sequenced. Figure 5.10. includes some examples of B- or T-cell specific mutations.

A Compares the insertions found in a spleen and in its purified B-cells. They share the majority of their inserts, differing only in the less clonal ones, with some selection for clonal inserts in the B-cells.

The spleen and T-cells in **B** appear to have different diseases with a common clonal origin. This example highlights the differences in insertion profile between a whole organ and a specific subset with the T-cell relevant inserts not ranking highly in the whole spleen.

Comparison of the spleen, B- and T-cell in **C** indicates that, though they have insert in common, selection has occurred in the T-cells and this is likely a T-cell lymphoma. The importance of these T-cell specific inserts was not as evident in the whole spleen sample where the importance was diluted by the mixed population.

The samples in **D** all have totally different inserts; it is worth noting that the spleen DNA was extracted from a piece of cryopreserved spleen whereas the B and T-cells were purified from a spleen suspension, they are different portions of the same spleen and this might account for the differences in the insert profiles.

The lack of correlation between inserts in the spleen and its cellular subtypes suggested that bead purification was not sufficient to identify B- or T-cell specific mutations and so the approach was changed and the tumour specific inserts were correlated to the percentages of B- and T-cells found by flow cytometry (see section 4.4.1).

These experiments illustrate how description of organ specific mutations profiles is important when considering the site of patient biopsy and treatment recommendation. The differences observed in clonally related tumours at independent sites mimics what has been observed in human tumours and could be used to plan an effective treatment strategy with different mutation making the site amenable to different treatment. Such analysis can also give an indication of site of origin and describe which clones were able to metastasise, also relevant in treatment and potentially understanding the genetic basis of lymphoma dissemination to different environments.

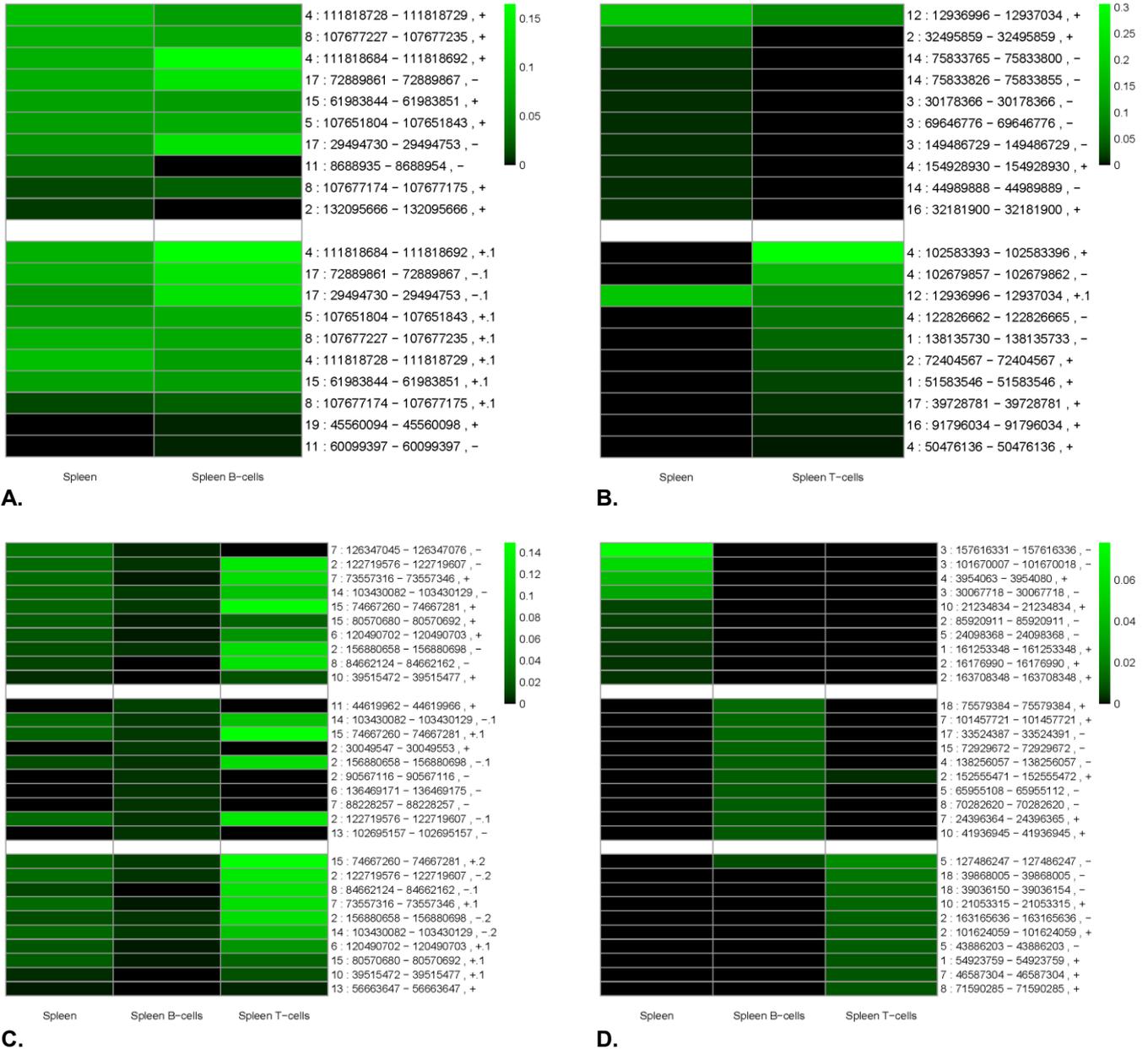


Figure 5.10. Comparison of top ten most clonal insertions in whole spleen and its equivalent B- or T-cells

The spleen and B-cells in **A** (#5021) have common mutations with evidence of selection for these mutations in the B-cells. The spleen and T-cells have different diseases in **B** (#5034) with one common insert. **C** (#5022) is an example where all samples share the same insertion but they are more clonal in the T-cells indicating a T-cell lymphoma. **D** (#5038) highlights the deficits in purification with all samples having different insertions also drawing attention to the fact that the whole spleen and purified sample, taken from the spleen suspension, may contain different portions of the tumour.

5.6. Genes collaborating with *BCL2* or *Crebbp* deficiency

After application of KCRMB to the insert table, which predicts the putative target gene of an insert, the extent to which certain insertions can be found in one genotype more than the other (genotype specificity) was assessed.

A 100kb window was taken around the mean of each Gaussian convolution peak and the inserts within the window were counted, these were defined as “IN”. This was then compared to all the other inserts in the genome outside the window for that genotype, these were defined “OUT”. Pairwise comparisons are then made between inserts within a gene window versus those outside using the Fisher’s exact test which assesses whether there is a non-random association of inserts at a given CIS to the genotypes in question. A worked example is shown below.

2 tailed p value ~ 0.000	<i>Notch1</i> inserts_IN	Other inserts_OUT	Fold enrichment
<i>BCL2</i> transgenic genotypes	33	67,506	0.31
<i>BCL2</i> non-transgenic genotypes	145	91,250	3.25

Table 5.3. Example contingency table test comparing *Notch1* insertions in the transgenic and non-transgenic *BCL2* mice

A pairwise comparison is made between the number of end stage inserts within a 100kb window of *Notch1* and the rest of the genome in the *BCL2* transgenic genotypes (*CrebbpCreBCL2*, *CrebbpBCL2*, *CreBCL2* and *BCL2*) and those with non-transgenic *BCL2* expression (*CrebbpCre*, *Crebbp*, *Cre*, wild type) demonstrating significant enrichment for mutation of *Notch1* in the non-transgenic *BCL2* tumours.

Making these comparisons allows inference about collaboration between frequently mutated genes and the sensitising genes that promote or prevent lymphoma. Examples of previously identified candidate genes, that also rank highly in this screen, with known mechanism of genotype specificity are described below.

Notch1 overexpression was first identified in haematological malignancy as part of the t(7:9)(q34;q34) translocation in patients with T-cell acute lymphoblastic leukaemia (Ellisen et al., 1991), where it is the most frequently mutated gene occurring in over 50% of cases (Weng et al., 2004). *Notch1* is also mutated in 5-20% B-NHL and it is known to promote marginal zone B-cell differentiation (Radtke et al., 2013). In this screen *Notch1* is found to be mutated more frequently in the *BCL2* non-transgenic lymphomas. The flow cytometry data demonstrated that the non-transgenic *BCL2* tumours were predominantly T-cell lymphomas. *Notch1* is known to promote T-cell growth, differentiation and survival suggesting a possible role for its mutations in these tumours.

IKZF1 and *IKZF3* (IKAROS Family Zinc Finger 1 and 3) are transcription factors involved in the determination of haematopoietic stem cell fate and lymphocyte development (John and Ward, 2011) with expression mainly restricted to lymphoid tissues including spleen and thymus (Molnár et al., 1996). Both genes rank highly in this screen.

IKZF1 deletions occur in 83.7% of B-cell acute lymphoblastic leukaemia indicative of its role as a tumour suppressor and an important driver of this disease (Mullighan et al., 2008). Mutations of *IKZF1* also occur in approximately 5% of T-cell acute lymphoblastic leukaemia where they frequently co-occur with activating *NOTCH1* mutations (Witkowski et al., 2015). Recent ChIP studies have shown that *Ikzf1* directly represses *Notch1* in thymocytes suppressing its pro-oncogenic activities (Geimer Le Lay et al., 2014). This example illustrates that co-occurrence data can indicate points of direct interaction between genes. In this screen, end stage *Ikzf1* mutations are 3-fold enriched in non-transgenic *BCL2* mice and it ranks highly ($p < 0.0001$) (Table 5.4.B) in the B-cell lymphoma specific mutations (Table 5.2). This indicates that mutation of *Ikzf1* is not tolerated in the absence of *BCL2* overexpression in B-cell lymphoma.

IKZF3 is thought to activate transcription of *BCL2* thus reducing apoptosis (Romero et al., 1999). In this screen end stage *Ikzf3* mutations are 4-fold enriched in *BCL2* transgenics (Table 5.4.B) and it ranks highly in the B-cell lymphoma specific mutations in this screen (Table 5.2, $p < 0.0001$). It is likely that *Ikzf3*'s role in these tumours is not to activate *BCL2*, which would be redundant as is already overexpressed. In this context, mutation of *Ikzf3* is tolerated and its loss can contribute to lymphomagenesis in a *BCL2* independent manner. Loss of *Ikzf3* in mice leads to the development of B-cell lymphomas (Wang et al., 1998) and it has also been found to be mutated in human FL (Okosun et al., 2014) and chronic myeloid leukaemia (Menezes et al., 2013).

The following tables, Table 5.4.-Table 5.12. are subsets of the genotype specificity tables showing the top ten most genotype specific genes, ranked by 2 tailed p-value, when considering all the inserts in that genotype (A), the end stage inserts (entropy ≤ 3.6) in animals with lymphoma (B) and the clonal end stage inserts (entropy ≤ 3.6 and clonality ≥ 0.1) that have undergone the greatest degree of selection (C). The enrichment for insertions in a particular genotype is suggestive of a lymphoma promoting phenotype in perturbing that gene, in combination with the sensitising genotype.

Table 5.4.-Table 5.6. list genes that are significantly mutated in the spleen (red tables), lymph node (green tables) and thymus (blue tables) of animals grouped by elevated transgenic *BCL2* levels compared to those with non-transgenic *BCL2*. Considering the spleen end stage inserts (Table 5.4.B) *Pou2f2* is preferentially mutated in the transgenic *BCL2* mice relative to the non-transgenic *BCL2* mice showing a 21-fold increase in insertions (2 tailed $p = 5.2 \times 10^{-53}$). Conversely, *Tfrc* is mutated 4-fold more in the non-transgenic *BCL2* genotypes (2 tailed $p = 2.2 \times 10^{-14}$).

Table 5.7.-Table 5.9. list genes that are significantly mutated in the spleen, lymph node and thymus of animal with *Crebbp* loss in an elevated *BCL2* context. Again when assessing the spleen end stage inserts (Table 5.7.B) *Pou2f2* is preferentially mutated in the *CrebbpCreBCL2* mice relative to the *BCL2* overexpressing mice showing a 2-fold increase in insertions (2 tailed $p = 1.5 \times 10^{-4}$). This suggests that *Pou2f2* collaborates with *Crebbp* loss in an elevated *BCL2* environment. In this context the only gene that is mutated more frequently in the transgenic *BCL2* groups compared to the *CrebbpCreBCL2* mice is *Sash3* whose insertions are enriched 6-fold transgenic *BCL2* group (2 tailed $p = 1.7 \times 10^3$). This could suggest that loss of *Crebbp* and insertions near *Sash3* have similar roles in lymphomagenesis or that mutation of *Sash2* is not tolerated in the *Crebbp* deficient mice.

Table 5.10.-Table 5.12. list genes that are significantly mutated in the spleen, lymph node and thymus of animals with *Crebbp* loss in a non-transgenic *BCL2* context. Considering the spleen end stage inserts in animals with non-transgenic *BCL2* (Table 5.10.B) *Tfrc* is mutated 11-fold more in the *CrebbpCre* mice compared to the non-transgenic *BCL2s* (2 tailed $p = 9.0 \times 10^{-36}$) indicating that mutations of *Tfrc* cooperates directly with *Crebbp* loss. The only gene that shows enrichment for mutations (2-fold) in the non-transgenic *BCL2* controls relative to the *CrebbpCre* mice is *Rasgrp1*, when taking into account all in inserts in the spleen (2 tailed $p = 2.8 \times 10^{-8}$) (Table 5.10.A) indicating that its mutation is not tolerated when *Crebbp* is intact or that where *Crebbp* is lacking it provides a selective advantage.

5.6.1. Genes that collaborate with *BCL2*

Chromosome	Peak Position	Gene name	Rank	P-value KCRBM	All BCL2_IN	All non BCL2_IN	All BCL2_OUT	All non BCL2_OUT	Fold enrichment BCL2	Fold enrichment non BCL2	2 tailed p-value	right handed p-value	left handed p-value
A. All spleen inserts													
7	25,134,886	Pou2f2	10	0.000	479	41	206,822	146,378	8	0	2.8E-66	1.8E-66	1.0E+0
16	36,665,016	Cd86	469	0.432	149	9	207,152	146,410	12	0	1.7E-24	1.4E-24	1.0E+0
11	11,702,820	Ikzf1	6	0.000	124	250	207,177	146,169	0	3	4.7E-23	1.0E+0	2.8E-23
2	117,407,182	Rasgrp1	5	0.000	153	265	207,148	146,154	0	2	1.6E-19	1.0E+0	1.1E-19
11	98,523,851	Ikzf3	84	0.004	259	65	207,042	146,354	3	0	5.7E-16	2.8E-16	1.0E+0
2	26,488,418	Notch1	4	0.000	172	263	207,129	146,156	0	2	1.4E-15	1.0E+0	9.2E-16
16	32,515,348	Tfrc	444	0.198	205	291	207,096	146,128	0	2	1.0E-14	1.0E+0	7.1E-15
16	19,019,411	IgL cluster	387	0.553	112	13	207,189	146,406	6	0	3.6E-14	1.9E-14	1.0E+0
6	48,711,191	Gimap1, Gimap5	518	0.710	265	86	207,036	146,333	2	0	4.9E-11	2.4E-11	1.0E+0
11	23,792,442	Rel/Bcl11a	21	0.000	191	52	207,110	146,367	3	0	6.0E-11	3.5E-11	1.0E+0
B. End stage spleen inserts													
7	25,134,886	Pou2f2	10	0.000	183	12	67,356	91,383	21	0	5.2E-53	5.2E-53	1.0E+0
16	32,515,348	Tfrc	444	0.198	33	163	67,506	91,232	0	4	2.2E-14	1.0E+0	1.4E-14
12	12,933,473	Mycn	2	0.000	107	38	67,432	91,357	4	0	3.6E-14	2.1E-14	1.0E+0
16	36,665,016	Cd86	469	0.432	47	4	67,492	91,391	16	0	9.8E-14	9.8E-14	1.0E+0
2	26,488,418	Notch1	4	0.000	33	145	67,506	91,250	0	3	1.6E-11	1.0E+0	9.7E-12
11	98,523,851	Ikzf3	84	0.004	78	26	67,461	91,369	4	0	2.2E-11	1.7E-11	1.0E+0
11	11,702,820	Ikzf1	6	0.000	25	118	67,514	91,277	0	3	2.6E-10	1.0E+0	1.6E-10
16	19,019,411	IgL cluster	387	0.553	37	7	67,502	91,388	7	0	3.1E-08	1.6E-08	1.0E+0
11	44,793,284	Ebf1	503	0.240	23	2	67,516	91,393	16	0	3.0E-07	3.0E-07	1.0E+0
1	86,443,288	Ncl	665	0.549	57	24	67,482	91,371	3	0	7.0E-07	3.6E-07	1.0E+0
C. Clonal end stage spleen inserts													
12	12933473	Mycn	2	0.000	51	15	954	1,021	4	0	3.8E-06	2.0E-06	1.0E+0
2	26488418	Notch1	4	0.000	2	22	1,003	1,014	0	11	3.4E-05	1.0E+0	2.2E-05
5	33695376	Slbp	8	0.000	0	14	1,005	1,022	0	-	1.2E-04	1.0E+0	7.2E-05
11	11702820	Ikzf1	6	0.000	1	17	1,004	1,019	0	17	1.4E-04	1.0E+0	8.7E-05
7	25134886	Pou2f2	10	0.000	11	1	994	1,035	11	0	2.9E-03	2.7E-03	1.0E+0
20	52743681	Mir106a	17	0.000	8	0	997	1,036	-	0	3.4E-03	3.4E-03	1.0E+0
2	117407182	Rasgrp1	5	0.000	3	16	1,002	1,020	0	5	4.3E-03	1.0E+0	2.6E-03
7	114175415	Copb1, Rras2	9	0.000	2	12	1,003	1,024	0	6	1.3E-02	1.0E+0	7.4E-03
5	107737975	Gfi1, Evi5	1	0.000	25	46	980	990	1	2	2.1E-02	9.9E-01	1.1E-02
7	129784192	Wgr11/Fgfr2	12	0.000	1	9	1,004	1,027	0	9	2.1E-02	1.0E+0	1.2E-02

Table 5.4. Contingency table comparing the number of inserts in the spleens of the *BCL2* transgenic mice compared to the non-transgenic *BCL2* mice

All *BCL2* = CrebbpCre*BCL2*, Crebbp*BCL2*, Cre*BCL2* and *BCL2*. Non-*BCL2* = CrebbpCre, Crebbp, Cre and wild type. *Two tailed p-values in italic are non-significant.*

Chromosome	Peak Position	Gene name	Rank	P-value KCRBM	All BCL2_IN	All non BCL2_IN	All BCL2_OUT	All non BCL2_OUT	Fold enrichment BCL2	Fold enrichment non BCL2	2 tailed p-value	right handed p-value	left handed p-value
A. All lymph nodes inserts													
7	25,128,589	Pou2f2	5	0.000	478	41	206,823	146,378	8	0	4.7E-66	2.8E-66	1.0E+00
16	36,700,990	Cd86	30	0.002	150	11	207,151	146,408	10	0	4.4E-23	3.5E-23	1.0E+00
11	11,706,012	Ikzf1	4	0.000	124	248	207,177	146,171	0	3	1.1E-22	1.0E+00	7.3E-23
2	117,412,341	Rasgrp1	19	0.000	152	263	207,149	146,156	0	2	2.8E-19	1.0E+00	1.6E-19
12	113,228,080	U6	264	0.344	194	38	207,107	146,381	4	0	4.1E-16	2.1E-16	1.0E+00
2	26,486,260	Notch1	7	0.000	172	263	207,129	146,156	0	2	1.4E-15	1.0E+00	9.2E-16
5	107,727,908	Glmn, Evi5	2	0.000	278	331	207,023	146,088	1	2	1.4E-10	1.0E+00	7.9E-11
4	134,117,909	Cd52	21	0.000	280	105	207,021	146,314	2	1	9.9E-09	5.6E-09	1.0E+00
5	122,222,213	Ccdc63	287	0.603	148	41	207,153	146,378	3	0	1.3E-08	8.5E-09	1.0E+00
12	79,437,245	Rad5111	200	0.728	58	6	207,243	146,413	7	0	2.6E-08	1.5E-08	1.0E+00
B. End stage lymph nodes inserts													
2	117,412,341	Rasgrp1	19	0.000	18	42	44,092	16,630	0	6	8.5E-12	1.0E+00	8.5E-12
11	11,706,012	Ikzf1	4	0.000	20	40	44,090	16,632	0	5	2.8E-10	1.0E+00	2.8E-10
7	114,175,447	Copb1, Rras2	6	0.000	18	33	44,092	16,639	0	5	3.1E-08	1.0E+00	3.1E-08
7	25,128,589	Pou2f2	5	0.000	119	13	43,991	16,659	3	0	1.1E-06	5.8E-07	1.0E+00
11	100,822,405	Stat5a	244	0.171	28	36	44,082	16,636	0	3	1.5E-06	1.0E+00	1.2E-06
2	26,486,260	Notch1	7	0.000	35	39	44,075	16,633	0	3	5.1E-06	1.0E+00	3.8E-06
5	107,727,908	Glmn, Evi5	2	0.000	68	56	44,042	16,616	0	2	2.8E-05	1.0E+00	1.8E-05
10	44,993,711	Prep	34	0.004	2	11	44,108	16,661	0	15	2.9E-05	1.0E+00	2.9E-05
4	135,077,916	Runx3	59	0.001	14	22	44,096	16,650	0	4	3.4E-05	1.0E+00	2.4E-05
13	91,540,514	Ssbp2	224	0.254	1	9	44,109	16,663	0	24	6.6E-05	1.0E+00	6.6E-05
C. Clonal end stage lymph nodes inserts													
12	12,939,099	Mycn	1	0.000	19	3	277	225	5	0	3.6E-03	2.5E-03	1.0E+00
11	11,706,012	Ikzf1	4	0.000	0	6	296	222	0	-	6.5E-03	1.0E+00	6.5E-03
7	114,175,447	Copb1, Rras2	6	0.000	0	6	296	222	0	-	6.5E-03	1.0E+00	6.5E-03
1	156,943,304	Ralgs2	10	0.000	0	5	296	223	0	-	1.5E-02	1.0E+00	1.5E-02
7	25,128,589	Pou2f2	5	0.000	6	0	290	228	-	0	3.8E-02	3.2E-02	1.0E+00
2	117,412,341	Rasgrp1	19	0.000	0	3	296	225	0	-	8.2E-02	1.0E+00	8.2E-02
9	32,613,890	Fli1	13	0.000	0	3	296	225	0	-	8.2E-02	1.0E+00	8.2E-02
8	120,034,743	Gse1	20	0.000	0	3	296	225	0	-	8.2E-02	1.0E+00	8.2E-02
15	63,271,075	Ighv cluster	15	0.000	0	3	296	225	0	-	8.2E-02	1.0E+00	8.2E-02
11	68,446,930	Pik3r5	17	0.000	0	3	296	225	0	-	8.2E-02	1.0E+00	8.2E-02

Table 5.5. Contingency table comparing the number of inserts in the lymph node of the transgenic BCL2 mice compared to the non-transgenic BCL2 mice

All BCL2 = CrebbpCreBCL2, CrebbpBCL2, CreBCL2 and BCL2. Non-BCL2 = CrebbpCre, Crebbp, Cre and wild type. *Two tailed p-values in italic are non-significant.*

Chromosome	Peak Position	Gene name	Rank	P-value KCRBM	All BCL2_IN	All non BCL2_IN	All BCL2_OUT	All non BCL2_OUT	Fold enrichment BCL2	Fold enrichment non BCL2	2 tailed p-value	right handed p-value	left handed p-value
A. All thymus inserts													
11	11,715,205	Ikzf1	4	0.000	109	242	207,192	146,177	0	3	2.3E-25	1.0E+00	1.8E-25
11	98,483,978	Ikzf3	211	0.107	343	76	206,958	146,343	3	0	3.7E-24	2.0E-24	1.0E+00
2	117,409,005	Rasgrp1	1	0.000	152	264	207,149	146,155	0	2	1.9E-19	1.0E+00	1.0E-19
2	26,484,694	Notch1	2	0.000	172	263	207,129	146,156	0	2	1.4E-15	1.0E+00	9.2E-16
16	32,519,650	Tfrc	40	0.000	201	291	207,100	146,128	0	2	2.5E-15	1.0E+00	1.7E-15
5	107,734,301	Evi5	3	0.000	279	332	207,022	146,087	1	2	1.5E-10	1.0E+00	7.7E-11
12	12,937,696	Mycn	13	0.000	186	61	207,115	146,358	2	0	4.4E-08	2.9E-08	1.0E+00
6	127,243,448	Gm7308	253	0.229	184	62	207,117	146,357	2	0	1.3E-07	8.0E-08	1.0E+00
17	47,524,612	Taf8	15	0.000	469	219	206,832	146,200	2	1	2.9E-07	1.4E-07	1.0E+00
10	21,209,166	Myb, Ahi1	9	0.000	179	210	207,122	146,209	1	2	6.7E-07	1.0E+00	3.7E-07
B. End stage thymus inserts													
11	11,715,205	Ikzf1	4	0.000	57	126	29,580	20,034	0	3	9.4E-15	1.0E+00	7.6E-15
2	117,409,005	Rasgrp1	1	0.000	84	133	29,553	20,027	0	2	6.6E-10	1.0E+00	4.8E-10
5	107,734,301	Evi5	3	0.000	43	88	29,594	20,072	0	3	9.4E-10	1.0E+00	6.0E-10
10	21,209,166	Myb, Ahi1	9	0.000	35	76	29,602	20,084	0	3	4.4E-09	1.0E+00	2.3E-09
3	51,757,662	Maml3	145	0.043	1	21	29,636	20,139	0	31	7.6E-08	1.0E+00	7.6E-08
2	26,484,694	Notch1	2	0.000	73	109	29,564	20,051	0	2	1.3E-07	1.0E+00	9.7E-08
2	170,159,202	Zfp217	7	0.000	25	55	29,612	20,105	0	3	4.7E-07	1.0E+00	3.0E-07
14	122,555,475	Pcca	248	0.451	0	14	29,637	20,146	0	-	3.2E-06	1.0E+00	3.2E-06
13	52,885,204	Auh	22	0.000	72	16	29,565	20,144	3	0	1.0E-05	6.4E-06	1.0E+00
16	97,458,645	Mx1	207	0.974	0	10	29,637	20,150	0	-	1.2E-04	1.0E+00	1.2E-04
C. Clonal end stage thymus inserts													
12	12,937,696	Mycn	13	0.000	4	0	134	340	-	0	6.7E-03	6.7E-03	1.0E+00
14	31,548,890	Hacl1	20	0.000	3	0	135	340	-	0	2.4E-02	2.4E-02	1.0E+00
2	26,484,694	Notch1	2	0.000	0	11	138	329	0	-	3.9E-02	1.0E+00	2.2E-02
8	114,815,638	Wwox	68	0.001	2	0	136	340	-	0	8.3E-02	8.3E-02	1.0E+00
14	25,073,504	Zmiz1	34	0.009	2	0	136	340	-	0	8.3E-02	8.3E-02	1.0E+00
1	139,643,276	U6	75	0.001	2	0	136	340	-	0	8.3E-02	8.3E-02	1.0E+00
3	68,756,835	Irf80, Il12a	55	0.002	2	0	136	340	-	0	8.3E-02	8.3E-02	1.0E+00
12	108,937,033	Wdr25	67	0.001	2	0	136	340	-	0	8.3E-02	8.3E-02	1.0E+00
5	140,283,187	Mad111	76	0.003	2	0	136	340	-	0	8.3E-02	8.3E-02	1.0E+00
3	19,622,240	Trim55	33	0.001	2	0	136	340	-	0	8.3E-02	8.3E-02	1.0E+00

Table 5.6. Contingency table comparing the number of inserts in the thymi of the transgenic BCL2 mice compared to the non-transgenic BCL2 mice

All BCL2 = CrebbpCreBCL2, CrebbpBCL2, CreBCL2 and BCL2. Non-BCL2 = CrebbpCre, Crebbp, Cre and wild type. *Two tailed p-values in italic are non-significant.*

Chromosome	Peak Position	Gene name	Rank	P-value KCRBM	CrebbpCreBCL2_IN	BCL2+_IN	CrebbpCreBCL2_OUT	BCL2+_OUT	Fold enrichment CrebbpCreBCL2	Fold enrichment BCL2+	2 tailed p-value	right handed p-value	left handed p-value
A. All spleen inserts													
7	25,134,886	Pou2f2	10	0.000	259	220	90,263	116,559	2	1	5.8E-06	3.0E-06	1.0E+00
14	25,080,495	Zmiz1	59	0.000	37	13	90,485	116,766	4	0	2.0E-05	1.3E-05	1.0E+00
5	123,319,215	Wdr66	98	0.003	83	58	90,439	116,721	2	1	3.3E-04	2.0E-04	1.0E+00
11	11,702,820	Ikzf1	6	0.000	36	88	90,486	116,691	1	2	1.0E-03	1.0E+00	5.7E-04
6	127,252,893	Ccnd2	13	0.000	102	81	90,420	116,698	2	1	1.3E-03	6.8E-04	1.0E+00
3	115,813,736	S1pr1	264	0.014	33	17	90,489	116,762	3	0	1.6E-03	1.2E-03	1.0E+00
17	46,817,821	Cnpy3	160	0.111	43	26	90,479	116,753	2	0	2.2E-03	1.4E-03	1.0E+00
7	114,175,415	Copb1	9	0.000	44	98	90,478	116,681	1	2	2.2E-03	1.0E+00	1.3E-03
9	112,213,053	Arpp21	767	0.658	40	24	90,482	116,755	2	0	3.4E-03	1.9E-03	1.0E+00
12	99,280,183	Foxn3	237	0.095	32	17	90,490	116,762	2	0	3.5E-03	1.9E-03	1.0E+00
B. End stage spleen inserts													
7	25,134,886	Pou2f2	10	0.000	114	69	32,462	34,894	2	1	1.5E-04	8.9E-05	1.0E+00
2	30,647,654	Cstad	30	0.000	24	5	32,552	34,958	5	0	2.4E-04	1.3E-04	1.0E+00
6	127,252,893	Ccnd2	13	0.000	46	19	32,530	34,944	3	0	2.7E-04	1.9E-04	1.0E+00
11	104,578,411	Itgb3	745	0.747	10	0	32,566	34,963	-	0	6.8E-04	6.8E-04	1.0E+00
11	44,793,284	Ebf1	503	0.240	19	4	32,557	34,959	5	0	1.2E-03	7.4E-04	1.0E+00
4	137,350,847	Cdc42	291	0.045	19	4	32,557	34,959	5	0	1.2E-03	7.4E-04	1.0E+00
20	48,137,287	Sash3	163	0.001	3	18	32,573	34,945	0	6	1.7E-03	1.0E+00	1.3E-03
5	123,319,215	Wdr66	98	0.003	28	10	32,548	34,953	3	0	1.8E-03	1.3E-03	1.0E+00
7	66,213,190	Chsy1, Lrrk1	74	0.002	24	8	32,552	34,955	3	0	2.4E-03	1.9E-03	1.0E+00
2	11,643,743	Il15ra	359	0.679	35	15	32,541	34,948	3	0	2.6E-03	1.5E-03	1.0E+00
C. Clonal end stage spleen inserts													
11	88,608,412	Msi2	42	0.000	4	0	495	506	-	0	<i>6.0E-02</i>	6.0E-02	1.0E+00
12	12,933,473	Mycn	2	0.000	19	32	480	474	1	2	<i>8.4E-02</i>	9.8E-01	4.7E-02
5	123,319,215	Wdr66	98	0.003	3	0	496	506	-	0	<i>1.2E-01</i>	1.2E-01	1.0E+00
8	120,472,974	Gse1	46	0.000	3	0	496	506	-	0	<i>1.2E-01</i>	1.2E-01	1.0E+00
18	4,322,812	Map3K8	52	0.000	3	0	496	506	-	0	<i>1.2E-01</i>	1.2E-01	1.0E+00
4	136,103,985	Rpl11	49	0.000	3	0	496	506	-	0	<i>1.2E-01</i>	1.2E-01	1.0E+00
7	25,134,886	Pou2f2	10	0.000	8	3	491	503	3	0	<i>1.4E-01</i>	1.1E-01	9.7E-01
2	30,647,654	Cstad	30	0.000	2	0	497	506	-	0	<i>2.5E-01</i>	2.5E-01	1.0E+00
3	115,813,736	S1pr1	264	0.014	2	0	497	506	-	0	<i>2.5E-01</i>	2.5E-01	1.0E+00
1	82,829,969	Tm4sf20	210	0.026	2	0	497	506	-	0	<i>2.5E-01</i>	2.5E-01	1.0E+00

Table 5.7. Contingency table comparing the number of inserts in the spleen of the CrebbpCreBCL2 mice compared to the other transgenic BCL2 mice

BCL2+ = CrebbpBCL2, CreBCL2 and BCL2. Two tailed p-values in italic are non-significant.

Chromosome	Peak Position	Gene name	Rank	P-value KCRBM	CrebbpCreBCL2_IN	BCL2+_IN	CrebbpCreBCL2_OUT	BCL2+_OUT	Fold enrichment CrebbpCreBCL2	Fold enrichment BCL2+	2 tailed p-value	right handed p-value	left handed p-value
A. All lymph node inserts													
7	25,128,589	Pou2f2	5	0.000	259	219	90,263	116,560	2	1	4.5E-06	2.4E-06	1.0E+00
12	99,301,855	Foxn3	239	0.649	29	12	90,493	116,767	3	0	7.2E-04	4.2E-04	1.0E+00
11	11,706,012	Ikzf1	4	0.000	36	88	90,486	116,691	1	2	1.0E-03	1.0E+00	5.7E-04
7	66,229,386	Lrrk1	39	0.004	52	34	90,470	116,745	2	1	2.2E-03	1.3E-03	1.0E+00
10	68,018,541	Rtkn2	182	0.016	15	4	90,507	116,775	5	0	2.2E-03	1.9E-03	1.0E+00
7	114,175,447	Copb1, Rras2	6	0.000	44	98	90,478	116,681	1	2	2.2E-03	1.0E+00	1.3E-03
5	122,222,213	Ccdc63	287	0.603	83	65	90,439	116,714	2	1	2.7E-03	1.6E-03	1.0E+00
19	46,463,482	Actr1a	145	0.734	52	37	90,470	116,742	2	1	5.4E-03	3.6E-03	1.0E+00
12	113,228,080	U6	264	0.344	104	90	90,418	116,689	1	1	5.8E-03	3.4E-03	1.0E+00
8	61,472,723	Cbr4	174	0.218	9	1	90,513	116,778	12	0	6.7E-03	3.5E-03	1.0E+00
B. End stage lymph node inserts													
13	51,885,434	Gadd45g	153	0.514	0	8	25,122	18,980	0	-	1.2E-03	1.0E+00	1.2E-03
16	36,700,990	Cd86	30	0.002	8	22	25,114	18,966	0	4	1.3E-03	1.0E+00	7.6E-04
16	95,749,790	Ets2	232	0.074	2	12	25,120	18,976	0	8	1.7E-03	1.0E+00	1.3E-03
2	27,122,309	Dbh	117	0.406	3	14	25,119	18,974	0	6	2.1E-03	1.0E+00	1.1E-03
6	89,047,334	Mcm2	220	0.056	0	7	25,122	18,981	0	-	2.7E-03	1.0E+00	2.7E-03
8	46,857,229	Irf2	35	0.001	25	5	25,097	18,983	4	0	2.9E-03	2.1E-03	1.0E+00
15	66,843,830	Sla	16	0.000	8	18	25,114	18,970	0	3	9.1E-03	1.0E+00	6.3E-03
1	156,415,240	Axdnd1	329	0.240	9	0	25,113	18,988	-	0	1.3E-02	6.3E-03	1.0E+00
1	171,922,707	Slamf6	338	0.147	30	9	25,092	18,979	3	0	1.4E-02	7.7E-03	1.0E+00
9	116,339,058	NA	274	0.684	8	17	25,114	18,971	0	3	1.5E-02	1.0E+00	1.0E-02
C. Clonal end stage lymph node inserts													
15	62,010,102	Myc, Pvt1	3	0.000	4	0	134	158	-	0	4.6E-02	4.6E-02	1.0E+00
9	44,535,795	Cxcr5	8	0.000	3	0	135	158	-	0	1.0E-01	1.0E-01	1.0E+00
8	46,857,229	Irf2	35	0.001	2	0	136	158	-	0	2.2E-01	2.2E-01	1.0E+00
18	5,342,938	Zfp438	24	0.001	2	0	136	158	-	0	2.2E-01	2.2E-01	1.0E+00
4	134,117,909	Cd52	21	0.000	2	0	136	158	-	0	2.2E-01	2.2E-01	1.0E+00
17	47,527,781	Taf8	9	0.000	2	0	136	158	-	0	2.2E-01	2.2E-01	1.0E+00
11	86,402,431	Med13, Vmp1	46	0.003	2	0	136	158	-	0	2.2E-01	2.2E-01	1.0E+00
8	108,902,593	Zfhx3	51	0.001	2	0	136	158	-	0	2.2E-01	2.2E-01	1.0E+00
6	34,795,891	Cald1, Stra8	29	0.000	2	0	136	158	-	0	2.2E-01	2.2E-01	1.0E+00
15	66,843,830	Sla	16	0.000	0	3	138	155	0	-	2.5E-01	1.0E+00	1.5E-01

Table 5.8. Contingency table comparing the number of inserts in the lymph node of the CrebbpCreBCL2 mice compared to the other transgenic BCL2 mice

BCL2+ = CrebbpBCL2, CreBCL2 and BCL2. Two tailed p-values in *italic* are non-significant.

Chromosome	Peak Position	Gene name	Rank	P-value KCRBM	CrebbpCreBCL2_IN	BCL2_IN	CrebbpCreBCL2_OUT	BCL2_OUT	Fold enrichment CrebbpCreBCL2	Fold enrichment BCL2	2 tailed p-value	right handed p-value	left handed p-value
A. All thymus inserts													
14	25,073,504	Zmiz1	34	0.009	36	14	90,486	116,765	3	0	7.8E-05	4.7E-05	1.0E+00
11	11,715,205	Ikzf1	4	0.000	30	79	90,492	116,700	0	2	6.6E-04	1.0E+00	3.7E-04
12	99,308,753	Foxn3	213	0.804	29	13	90,493	116,766	3	0	1.5E-03	7.9E-04	1.0E+00
6	127,243,448	Gm7308	253	0.229	102	82	90,420	116,697	2	1	1.7E-03	8.7E-04	1.0E+00
14	25,467,509	Zmiz1	19	0.000	36	86	90,486	116,693	1	2	1.8E-03	1.0E+00	9.2E-04
7	114,171,111	Copb1, Rras2	16	0.000	44	98	90,478	116,681	1	2	2.2E-03	1.0E+00	1.3E-03
1	139,643,276	U6	75	0.001	0	10	90,522	116,769	0	-	3.5E-03	1.0E+00	3.2E-03
2	30,638,549	Cstad	14	0.000	50	34	90,472	116,745	2	1	4.0E-03	2.5E-03	1.0E+00
8	106,114,894	Nfatc3	228	0.880	8	30	90,514	116,749	0	3	4.9E-03	1.0E+00	3.1E-03
12	108,937,033	Wdr25	67	0.001	4	20	90,518	116,759	0	4	7.0E-03	1.0E+00	5.2E-03
B. End stage thymus inserts													
14	25,073,504	Zmiz1	34	0.009	21	10	9,274	20,332	5	0	4.2E-05	3.4E-05	1.0E+00
15	36,435,935	NA	126	0.233	18	10	9,277	20,332	4	0	3.7E-04	3.4E-04	1.0E+00
2	26,484,694	Notch1	2	0.000	10	63	9,285	20,279	0	3	8.7E-04	1.0E+00	4.2E-04
11	11,715,205	Ikzf1	4	0.000	7	50	9,288	20,292	0	3	1.4E-03	1.0E+00	7.3E-04
8	120,492,996	Gse1	26	0.000	2	29	9,293	20,313	0	7	1.5E-03	1.0E+00	9.6E-04
2	27,123,748	Fam163b	35	0.005	2	26	9,293	20,316	0	6	3.7E-03	1.0E+00	2.5E-03
1	193,088,795	Syt14	286	0.084	11	6	9,284	20,336	4	0	6.4E-03	4.8E-03	1.0E+00
3	130,941,206	SNORA61	73	0.002	24	23	9,271	20,319	2	0	6.7E-03	3.8E-03	1.0E+00
17	64,332,165	Pja2	163	0.219	4	0	9,291	20,342	-	0	9.7E-03	9.7E-03	1.0E+00
16	32,519,650	Tfrc	40	0.000	21	82	9,274	20,260	1	2	1.9E-02	1.0E+00	8.9E-03
C. Clonal end stage thymus inserts													
14	25,073,504	Zmiz1	34	0.009	2	0	52	84	-	0	<i>1.5E-01</i>	1.5E-01	1.0E+00
18	5,343,329	Zfp438	18	0.000	2	0	52	84	-	0	<i>1.5E-01</i>	1.5E-01	1.0E+00
11	86,398,919	Med13, Vmp1	60	0.003	2	0	52	84	-	0	<i>1.5E-01</i>	1.5E-01	1.0E+00
8	114,815,638	Wwox	68	0.001	2	0	52	84	-	0	<i>1.5E-01</i>	1.5E-01	1.0E+00
15	36,435,935	NA	126	0.233	1	0	53	84	-	0	<i>3.9E-01</i>	3.9E-01	1.0E+00
1	193,088,795	Syt14	286	0.084	1	0	53	84	-	0	<i>3.9E-01</i>	3.9E-01	1.0E+00
3	130,941,206	SNORA61	73	0.002	1	0	53	84	-	0	<i>3.9E-01</i>	3.9E-01	1.0E+00
17	64,332,165	Pja2	163	0.219	1	0	53	84	-	0	<i>3.9E-01</i>	3.9E-01	1.0E+00
7	102,340,232	Stim1	221	0.305	1	0	53	84	-	0	<i>3.9E-01</i>	3.9E-01	1.0E+00
14	79,197,796	Vva8	179	0.096	1	0	53	84	-	0	<i>3.9E-01</i>	3.9E-01	1.0E+00

Table 5.9. Contingency table comparing the number of inserts in the thymi of the CrebbpCreBCL2 mice compared to the other transgenic BCL2 mice

BCL2+ = CrebbpBCL2, CreBCL2 and BCL2. Two tailed p-values in italic are non-significant.

5.6.1.1. *BCL2* specific mutation: *Pou2f2*

One of the most significant associations in the screen is *Pou2f2* mutation cooperating with both overexpression of *BCL2* and additionally with loss of *Crebbp* showing significant enrichment for inserts in these genotypes. *Pou2f2* is also the fourth most frequently mutated gene in the B-lymphoma specific dataset. This would suggest that mutation of *Pou2f2* accelerates B-cell oncogenesis in tumours overexpressing *BCL2* and deficient in *Crebbp* indicative of a shared oncogenic pathway.

By superimposing the screen mutations onto UCSC Genome Browser it is possible to visualise the exact location of the screen inserts relative to the gene transcript (Figure 5.11), (Raney et al., 2014). *Pou2f2* is significantly mutated in the first half of the gene with no strong stand specific bias. Inserts within a transcript can interfere with transcription resulting in mRNA that encodes a truncated, inactive or unstable form of the protein or perhaps an active variant of the protein. Insertion could also induce aberrant splicing events abrogating gene function.

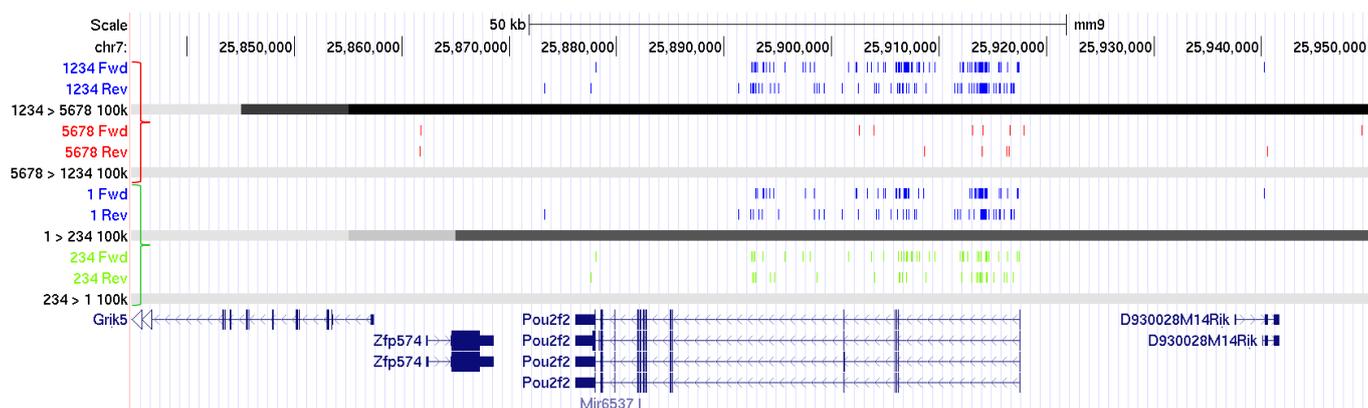


Figure 5.11. *Pou2f2* insertions

Custom tracks illustrating the screen inserts in UCSC Genome Browser. Each vertical coloured line represents one insertion at that location in each respective genotype. Fwd and Rev indicates if the mutation occurred in the sense or antisense strand respectively. The greyscales lines running horizontally indicate the p-value of those insertions with black being the most significant p-value. There are two p-values per genotype comparison representing enrichment in one or the other genotype.

Genotypes 1, 2, 3 & 4 are the transgenic *BCL2* genotypes, *CrebbpCreBCL2*, *CrebbpBCL2*, *CreBCL2* & *BCL2*. Genotypes 5, 6, 7 & 8 are the non-transgenic *BCL2* genotypes, *CrebbpCre*, *Crebbp*, *Cre* & wild type.

Pou2f2 showed a 21-fold enrichment in inserts in the spleens of *BCL2* transgenics tumours compared to the non-transgenic *BCL2* (end stage inserts, 2 tailed p-value = 5.2×10^{-53}) and a 2-fold enrichment in inserts in the *CrebbpCreBCL2* spleens compared to the transgenic *BCL2* (end stage inserts, 2 tailed p-value = 1.4×10^{-4}).

Pou2f2 (Oct2) is a POU-homeodomain family member that was first defined as a B-lymphocyte specific transcription factor present in the promoter and enhancer elements of the immunoglobulin genes (Singh et al., 1986). *Pou2f2* has been found to be expressed only in B-cells, activated T-cells, macrophages and neuronal cells (Staudt et al., 1988, He et al., 1989; Kang et al., 1992; Dunn et al., 1996). Mouse knock-out models however have shown that *Pou2f2* is not essential for early B-cell development and immunoglobulin expression though its loss causes perinatal lethality (Corcoran et al., 1993).

Pou2f2 is able to activate transcription via its C-terminus whereas the N-terminus contains an inhibitory domain capable of repressing transcription (Lillycrop and Latchman, 1992). Deletion of the C-terminus in transgenic mice recapitulates the phenotype of the complete loss of function allele demonstrating that the C-terminus is critical to its function. *Pou2f2* has a variety of alternative splice transcript which produce proteins with different properties. The predominant transcript in B-lymphocytes is *Pou2f2.1* is an activating transcript whose C-terminal activation effects, dominate the inhibitory N-terminus effects, facilitating transcription. Conversely in the neurons *Pou2f2* transcripts tend to have inhibitory functions (reviewed in Latchman, 1996).

A study in human FL cells lines compared five FL lines with the classical IgH-BCL2 t(14;18) translocation and two mature B-cell lines; they found that the *BCL2* translocated cells had greatly increased POU2F2 protein levels and that silencing of *POU2F2* using siRNA caused a reduction in *BCL2* expression. This would suggest that overexpression of *POU2F2* increases survival of follicular lymphoma-like cells by increasing *BCL2* expression (Heckman et al., 2006).

A recent exome sequencing and SNP array study found *POU2F2* was mutated in 8% of patients with FL (9/114 cases) by purifying FL cells, based on their light chain restriction to improve sensitivity. The majority of mutations were heterozygote missense mutations to the POU DNA binding domain suggesting that reduction of *Pou2f2* transcription activation facilitated oncogenesis. Mutated isoforms were tested using a luciferase based assay and confirmed reduced transactivator function of *Pou2f2* mutants (Li et al., 2014).

The conflicting evidence may be in part due to *Pou2f2*'s ability to activate or inhibit gene transcription depending on the dominating terminus and perhaps different transcripts were overexpressed in the Heckman study compared to those deregulated in the Li study.

Other than both being mutated in cancer there is no connection between *Crebbp* and *Pou2f2*. This novel interaction is promisingly associated to B-cells in this study and its known role in the haematopoietic compartment mean it warrants further investigation.

5.6.2. Genes that collaborate with *Crebbp* deficiency independently of *BCL2*

Chromosome	Peak Position	Gene name	Rank	P-value KCRBM	CrebbpCre_IN	Crebbp & Cre & WT_IN	CrebbpCre_OUT	Crebbp & Cre & WT_OUT	Fold enrichment CrebbpCre	Fold enrichment Crebbp & Cre & WT	2 tailed p-value	right handed p-value	left handed p-value
A. All spleen inserts													
16	32,515,348	Tfrc	444	0.198	212	79	59,811	86,317	4	0	5.8E-28	3.1E-28	1.0E+00
2	11,643,743	Il15ra	359	0.679	154	98	59,869	86,298	2	0	1.3E-10	9.6E-11	1.0E+00
9	32,604,198	Fli1, Ets1	24	0.000	64	23	59,959	86,373	4	0	1.0E-09	7.4E-10	1.0E+00
2	117,407,182	Rasgrp1	5	0.000	65	200	59,958	86,196	0	2	2.8E-08	1.0E+00	1.3E-08
13	52,861,424	Auh	70	0.000	45	16	59,978	86,380	4	0	2.9E-07	2.1E-07	1.0E+00
10	44,992,333	Prep	157	0.036	12	0	60,011	86,396	-	0	2.3E-05	2.3E-05	1.0E+00
12	116,661,958	Ptprn2	787	0.670	12	0	60,011	86,396	-	0	2.3E-05	2.3E-05	1.0E+00
8	10,900,038	Irs2	56	0.001	35	15	59,988	86,381	3	0	3.9E-05	3.2E-05	1.0E+00
5	116,031,058	Prkab1	44	0.000	17	4	60,006	86,392	6	0	2.4E-04	2.2E-04	1.0E+00
15	84,328,168	Parvb	212	0.045	36	20	59,987	86,376	3	0	5.6E-04	3.7E-04	1.0E+00
B. End stage spleen inserts													
16	32,515,348	Tfrc	444	0.198	144	19	37,674	53,558	11	0	9.0E-36	7.3E-36	1.0E+00
2	11,643,743	Il15ra	359	0.679	84	46	37,734	53,531	3	0	1.4E-07	7.7E-08	1.0E+00
17	45,490,320	Cdc5l	505	0.583	20	2	37,798	53,575	14	0	1.8E-06	1.8E-06	1.0E+00
12	85,625,017	Jdp2	41	0.000	32	10	37,786	53,567	5	0	5.6E-06	4.8E-06	1.0E+00
9	32,604,198	Fli1, Ets1	24	0.000	34	12	37,784	53,565	4	0	1.4E-05	7.8E-06	1.0E+00
13	52,861,424	Auh	70	0.000	31	11	37,787	53,566	4	0	2.6E-05	2.1E-05	1.0E+00
5	140,246,861	Mad11l	292	0.090	23	6	37,795	53,571	5	0	4.0E-05	3.6E-05	1.0E+00
6	34,905,331	Agbl3, Stra8	140	0.000	24	7	37,794	53,570	5	0	6.5E-05	4.9E-05	1.0E+00
4	134,116,798	Cd52	19	0.000	39	19	37,779	53,558	3	0	8.3E-05	6.2E-05	1.0E+00
5	116,031,058	Prkab1	44	0.000	15	2	37,803	53,575	11	0	9.1E-05	9.1E-05	1.0E+00
C. Clonal end stage spleen inserts													
10	21,174,618	Myb, Ahi1	15	0.000	0	7	413	616	0	-	4.6E-02	1.0E+00	2.8E-02
5	116,031,058	Prkab1	44	0.000	3	0	410	623	-	0	6.3E-02	6.3E-02	1.0E+00
20	13,248,114	Ddx3x	57	0.000	3	0	410	623	-	0	6.3E-02	6.3E-02	1.0E+00
1	138,170,299	Ptprc	14	0.000	0	6	413	617	0	-	8.7E-02	1.0E+00	4.7E-02
5	33,695,376	Slbp	8	0.000	9	5	404	618	3	0	9.6E-02	5.6E-02	9.8E-01
8	10,900,038	Irs2	56	0.001	2	0	411	623	-	0	1.6E-01	1.6E-01	1.0E+00
3	101,673,971	Atp1a1	240	0.010	2	0	411	623	-	0	1.6E-01	1.6E-01	1.0E+00
8	121,025,199	Foxf1	273	0.010	2	0	411	623	-	0	1.6E-01	1.6E-01	1.0E+00
12	81,000,758	Smoc1	208	0.047	2	0	411	623	-	0	1.6E-01	1.6E-01	1.0E+00
7	110,708,974	Ampd3	255	0.021	2	0	411	623	-	0	1.6E-01	1.6E-01	1.0E+00

Table 5.10. Contingency table comparing the number of inserts in the spleen of the *CrebbpCre* mice compared to the *Crebbp & Cre & wild type* mice

Two tailed p-values in italic are non-significant.

Chromosome	Peak Position	Gene name	Rank	P-value KCRBM	CrebbpCre_IN	Crebbp & Cre & WT_IN	CrebbpCre_OUT	Crebbp & Cre & WT_OUT	Fold enrichment CrebbpCre	Fold enrichment Crebbp & Cre & WT	2 tailed p-value	right handed p-value	left handed p-value
A. All lymph node inserts													
13	52,831,993	Auh	159	0.396	58	18	59,965	86,378	5	0	5.1E-10	4.1E-10	1.0E+00
2	117,412,341	Rasgrp1	19	0.000	63	200	59,960	86,196	0	2	8.0E-09	1.0E+00	4.6E-09
9	32,613,890	Fli1	13	0.000	58	27	59,965	86,369	3	0	6.4E-07	3.6E-07	1.0E+00
10	44,993,711	Prep	34	0.004	12	0	60,011	86,396	-	0	2.3E-05	2.3E-05	1.0E+00
9	57,972,208	Sema7a	167	0.768	40	21	59,983	86,375	3	0	1.3E-04	9.2E-05	1.0E+00
5	116,027,513	Prkab1	53	0.004	17	4	60,006	86,392	6	0	2.4E-04	2.2E-04	1.0E+00
15	84,320,908	Parvb	208	0.204	37	20	59,986	86,376	3	0	3.7E-04	2.3E-04	1.0E+00
6	52,870,301	Jazf1	161	0.802	29	14	59,994	86,382	3	0	5.4E-04	4.1E-04	1.0E+00
8	120,462,203	Gse1	283	0.383	35	20	59,988	86,376	3	0	8.6E-04	5.8E-04	1.0E+00
17	26,380,251	Hba-ps4	115	0.603	12	2	60,011	86,394	9	0	1.4E-03	7.9E-04	1.0E+00
B. End stage lymph node inserts													
1	156,943,304	Ralgps2	10	0.000	0	15	7,546	9,111	0	-	1.2E-04	1.0E+00	1.2E-04
10	44,993,711	Prep	34	0.004	11	0	7,535	9,126	-	0	1.6E-04	1.6E-04	1.0E+00
10	19,597,428	Ifngr1	97	0.498	1	16	7,545	9,110	0	13	8.4E-04	1.0E+00	5.3E-04
7	25,128,589	Pou2f2	5	0.000	11	2	7,535	9,124	7	0	4.8E-03	4.4E-03	1.0E+00
8	14,988,721	Kbtbd11	94	0.510	1	13	7,545	9,113	0	11	5.0E-03	1.0E+00	2.7E-03
5	116,027,513	Prkab1	53	0.004	9	1	7,537	9,125	11	0	7.1E-03	4.7E-03	1.0E+00
12	12,939,099	Mycn	1	0.000	2	14	7,544	9,112	0	6	9.9E-03	1.0E+00	6.2E-03
11	11,706,012	lkzf1	4	0.000	10	30	7,536	9,096	0	2	1.0E-02	1.0E+00	6.7E-03
14	30,010,676	Il17rb	121	0.299	1	10	7,545	9,116	0	8	1.6E-02	1.0E+00	1.3E-02
11	4,385,840	Hormad2	22	0.031	5	0	7,541	9,126	-	0	1.9E-02	1.9E-02	1.0E+00
C. Clonal end stage lymph node inserts i													
1	156,943,304	Ralgps2	10	0.000	0	5	102	121	0	-	6.7E-02	1.0E+00	5.0E-02
8	120,034,743	Gse1	20	0.000	3	0	99	126	-	0	<i>8.8E-02</i>	8.8E-02	1.0E+00
5	65,906,763	Chrna9	11	0.000	3	0	99	126	-	0	<i>8.8E-02</i>	8.8E-02	1.0E+00
5	116,027,513	Prkab1	53	0.004	2	0	100	126	-	0	<i>2.0E-01</i>	2.0E-01	1.0E+00
7	127,476,487	Itgal	57	0.002	2	0	100	126	-	0	<i>2.0E-01</i>	2.0E-01	1.0E+00
14	25,421,873	Zmiz1	12	0.000	2	0	100	126	-	0	<i>2.0E-01</i>	2.0E-01	1.0E+00
10	44,993,711	Prep	34	0.004	1	0	101	126	-	0	<i>4.5E-01</i>	4.5E-01	1.0E+00
11	4,385,840	Hormad2	22	0.031	1	0	101	126	-	0	<i>4.5E-01</i>	4.5E-01	1.0E+00
2	132,031,828	Rassf2	300	0.839	1	0	101	126	-	0	<i>4.5E-01</i>	4.5E-01	1.0E+00
6	146,486,718	Itpr2	321	0.102	1	0	101	126	-	0	<i>4.5E-01</i>	4.5E-01	1.0E+00

Table 5.11. Contingency table comparing the number of inserts in the lymph node of the CrebbpCre mice compared to the Crebbp & Cre & wild type mice

Two tailed p-values in italic are non-significant.

Chromosome	Peak Position	Gene name	Rank	P-value KCRBM	CrebbpCre_IN	Crebbp & Cre &WT_IN	CrebbpCre_OUT	Crebbp & Cre &WT_OUT	Fold enrichment CrebbpCre	Fold enrichment Crebbp & Cre &WT	2 tailed p-value	right handed p-value	left handed p-value
A. All thymus inserts													
16	32,519,650	Tfrc	40	0.000	212	79	59,811	86,317	4	0	5.8E-28	3.1E-28	1.0E+00
13	52,885,204	Auh	22	0.000	55	17	59,968	86,379	5	0	1.7E-09	1.1E-09	1.0E+00
2	117,409,005	Rasgrp1	1	0.000	64	200	59,959	86,196	0	2	1.3E-08	1.0E+00	8.0E-09
9	32,625,197	Fli1	21	0.000	63	33	59,960	86,363	3	0	1.2E-06	9.7E-07	1.0E+00
12	116,656,665	Ptprn2	69	0.001	12	0	60,011	86,396	-	0	2.3E-05	2.3E-05	1.0E+00
14	122,555,475	Pcca	248	0.451	14	1	60,009	86,395	20	0	3.5E-05	3.5E-05	1.0E+00
8	10,921,942	Irs2	97	0.748	29	14	59,994	86,382	3	0	5.4E-04	4.1E-04	1.0E+00
2	163,167,193	Gtsf1l	277	0.313	19	6	60,004	86,390	5	0	7.1E-04	4.1E-04	1.0E+00
8	120,492,996	Gse1	26	0.000	37	23	59,986	86,373	2	0	1.5E-03	9.8E-04	1.0E+00
12	108,937,033	Wdr25	67	0.001	15	4	60,008	86,392	5	0	1.5E-03	8.7E-04	1.0E+00
B. End stage thymus inserts													
16	32,519,650	Tfrc	40	0.000	59	49	4,074	15,978	5	0	5.3E-15	5.3E-15	1.0E+00
14	122,555,475	Pcca	248	0.451	14	0	4,119	16,027	-	0	2.3E-10	2.3E-10	1.0E+00
12	116,656,665	Ptprn2	69	0.001	10	0	4,123	16,027	-	0	1.3E-07	1.3E-07	1.0E+00
9	32,625,197	Fli1	21	0.000	16	8	4,117	16,019	8	0	1.3E-06	1.3E-06	1.0E+00
16	97,458,645	Mx1	207	0.974	8	2	4,125	16,025	16	0	9.4E-05	9.4E-05	1.0E+00
2	163,167,193	Gtsf1l	277	0.313	8	3	4,125	16,024	10	0	2.8E-04	2.8E-04	1.0E+00
19	11,540,855	Ms4a4d	99	0.394	6	2	4,127	16,025	12	0	1.4E-03	1.4E-03	1.0E+00
8	3,280,841	Insr	84	0.721	5	1	4,128	16,026	19	0	1.8E-03	1.8E-03	1.0E+00
20	52,759,298	Mir106a	149	0.503	0	28	4,133	15,999	0	-	3.4E-03	1.0E+00	1.6E-03
2	129,184,526	Slc20a1	257	0.185	6	3	4,127	16,024	8	0	3.5E-03	3.5E-03	1.0E+00
C. Clonal end stage thymus inserts													
12	116,656,665	Ptprn2	69	0.001	2	0	61	277	-	0	3.4E-02	3.4E-02	1.0E+00
6	34,971,481	Cnot4, Stra8	41	0.004	2	0	61	277	-	0	3.4E-02	3.4E-02	1.0E+00
2	29,097,299	Ttf1	36	0.006	2	0	61	277	-	0	3.4E-02	3.4E-02	1.0E+00
2	117,409,005	Rasgrp1	1	0.000	4	6	59	271	3	0	9.3E-02	9.3E-02	9.8E-01
2	30,638,549	Cstad	14	0.000	2	2	61	275	5	0	1.6E-01	1.6E-01	9.8E-01
14	122,555,475	Pcca	248	0.451	1	0	62	277	-	0	1.9E-01	1.9E-01	1.0E+00
16	97,458,645	Mx1	207	0.974	1	0	62	277	-	0	1.9E-01	1.9E-01	1.0E+00
19	11,540,855	Ms4a4d	99	0.394	1	0	62	277	-	0	1.9E-01	1.9E-01	1.0E+00
8	3,280,841	Insr	84	0.721	1	0	62	277	-	0	1.9E-01	1.9E-01	1.0E+00
2	129,184,526	Slc20a1	257	0.185	1	0	62	277	-	0	1.9E-01	1.9E-01	1.0E+00

Table 5.12. Contingency table comparing the number of inserts in the thymi of the CrebbpCre mice compared to the Crebbp & Cre & wild type mice

Two tailed p-values in italic are non-significant.

5.6.2.1. Crebbp specific mutation in the absence of the *BCL2* transgene: *Tfrc*

The novel murine model describe in this thesis facilitated the demonstration that *Crebbp* is a tumour suppressor gene in the context of *BCL2* driven lymphomas. Genotype specific analysis provides a unique opportunity to identify regulators of *Crebbp* tumours suppressor activity or downstream targets of it. If a gene is more frequently mutated in the genotypes with intact *Crebbp* it suggests that it may be an upstream regulator or downstream effector of *Crebbp* and mutations of it either deregulates *Crebbp* or attenuates its tumour suppression actions.

The most frequently inserted locus in the spleens and thymi of mice deficient in *Crebbp* is a cluster adjacent to the *Tfrc* gene; it is also the second most frequently mutated gene in the B-cell lymphoma subset. This indicates that aberration of *Tfrc* cooperates with loss of *Crebbp* specifically in B-cell lymphoma however other genes within the region may also be responsible for the observed selection.

The mutations associated with *Tfrc* are upstream of it and may upregulate expression by placing it under the influence of the strong enhancers in the MoMuLV LTRs. It is possible that *Tfrc* is a downstream target of *Crebbp* and loss of *Crebbp* leads to a decrease in *Tfrc* expression consequently upregulating, viral mutations are selected for. Other genes in the region that may also be deregulated by insertions are *Tnk2* and *Pcyf1A*

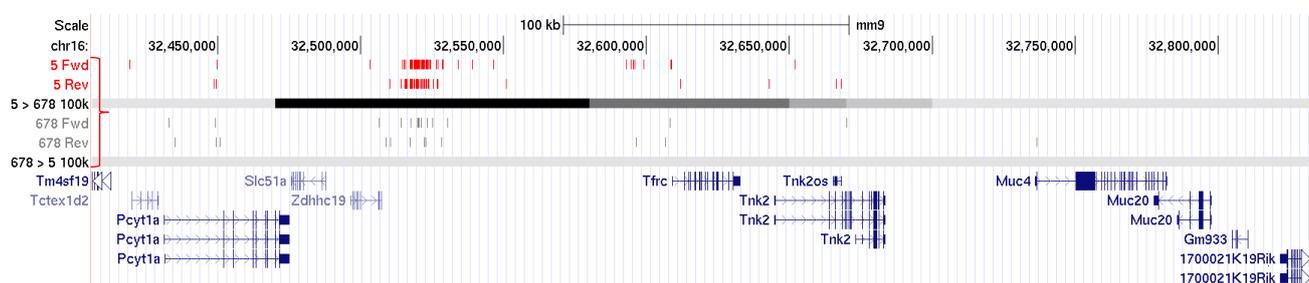


Figure 5.12. *Tfrc* insertions

Custom UCSC tracks illustrating the screen inserts. Each vertical coloured line represents one insertion at that location in each respective genotype. Fwd and Rev indicates if the mutation occurred in the sense or antisense strand respectively. The greyscales lines running horizontally indicate the p-value of those insertions with black being the most significant p-value. There are two p-values per genotype comparison representing enrichment in one or the other genotype.

Genotype 5 is *Crebbp*Cre, loss of *Crebbp*. Genotype 6, 7 & 8 are *Crebbp*, Cre and wildtype.

Tfrc end stage spleen insertions are significantly enriched, 11-fold, in the *Crebbp*Cre mice compared to the *Crebbp* & Cre & wild type mice, p-value = 9.0×10^{-36} .

Comparing the transgenic *BCL2* groups and the non-transgenic *BCL2* shows a 4-fold enrichment for mutation of *Tfrc* in the non-transgenic *BCL2* mice (which includes the *Crebbp*Cre mice) ($p = 1.4 \times 10^{-14}$)

Transferrin receptor (TFRC) is a transmembrane cell surface receptor (CD71) that regulates the uptake of iron bound to transferrins by receptor mediated endocytosis (Cheng et al., 2004). Many crucial cellular and metabolic pathways are reliant on iron including cell growth, DNA synthesis, mitogenic signalling and cell survival. There is a well-established correlation between the number of transferrin receptors and the rate of proliferation of a cell with transferrin receptor expression higher in actively proliferating cells and also consequently in cancer cells (Chitambar et al., 1983).

A recent comprehensive study has shown that activation of *Tfrc* by *c-Myc* is integral to increased proliferation in B-cell lymphoma (Donnell et al., 2006). Using an *in vitro* model of Burkitt's lymphoma, with a human B- lymphocyte cell line expressing tetracyclin repressible *c-Myc*, the researchers found that *Tfrc* expression was upregulated 7-fold by *c-Myc* overexpression and that this induction was maintained also in serum deprived cells indicating that the upregulation in *Tfrc* was not a by-product of increased proliferation. ChIP studies of the Burkitt's like cells also confirmed the direct interaction between *c-Myc* and *Tfr*.

Tfrc knock-down with siRNA *in vitro*, reduced cell proliferation and caused an accumulation of arrested cells in the G1 stage of cell cycle unable to progress to S phase. Iron deprivation with desferrioxamine had the same effect in these Burkitt's like cells indicating that reduction in iron uptake when *Tfrc* is downregulated is responsible for reduced proliferation.

Iron uptake and the transferrin receptors have already been targeted in the clinic with iron chelators being used to treat leukaemia for some time (Richardson, 2002) and the transferrin receptor mediated uptake of gallium has been used to image lymphoma in patients (Feremans et al., 1991).

Interestingly microarray study of the siRNA-treated Burkitt's like cells demonstrated that depletion of *Tfrc* alters cell cycle and pro-apoptotic genes. They found 202 downregulated transcripts involved in cell division like cyclins and DNA replication and 647 upregulated transcripts involved in p53-responsive and pro-apoptotic genes (Donnell et al., 2006).

This is particularly relevant to this screen where *Tfrc* mutation occurs less frequently in the *BCL2* overexpressing genotypes which are resistant to apoptosis. It suggests that there is an inverse relationship between *Tfrc* and apoptosis; upregulation of *Tfrc* downregulates of pro-apoptotic genes consequently decreasing apoptosis. If reduction of *Tfrc* is pro-apoptotic then it is conceivable that upregulation of *Tfrc* could be anti-apoptotic and therefore redundant in a context of *BCL2* overexpression.

They also demonstrate that *c-Myc* activated *Tfrc in vivo* using a model of murine B-cell lymphoma where animals are transplanted with p53-null bone marrow transduced with *c-MycER* retrovirus. Removal of the 4-hydroxytamoxifen treatment inactivated *c-Myc* and causes a six-fold downregulation in *Tfrc*.

Using the immortalised, non-transformed fibroblast cell line Rat1a, which is non-transplantable in nude mice, they showed that c-Myc and Tfrc cooperate *in vivo* to accelerate tumourigenesis by overexpressing *Tfrc* in Rat1a-Myc cells and causing tumours in the nude mice with significantly faster onset than the Rat1a-Myc mice.

It is possible that there is enrichment for *Tfrc* mutations in the *Crebbp* deficient mice because *Crebbp* is an upstream activator of *Tfrc*. *Crebbp* null mice are known to have reduced *Myc* expression (Xu, 2006) and the work by Donnell demonstrates that *Myc* activates *Tfrc*. Enrichment of activating mutation to *Tfrc* may compensate for the reduction of its activator in the *Crebbp* deficient mice.

Another possible mechanism of regulation that warrants further investigation is the relationship between CREBBP, GATA-1 and TFRC. GATA-1 is a transcriptional activator of *TFRC* (Shimizu et al., 2007). Reduction of Gata-1 would cause downregulation of *Tfrc* which, in a tumour, would be a disadvantage, reducing its proliferative capacity. *Crebbp* acetylation activates transcription of *Gata-1* (Kung et al., 2000) and so in the *Crebbp* deficient mice there is probably a reduction in Gata-1 activity. This could be the reason why activating mutations in *Tfrc* are enriched in the *Crebbp* deficient mice compensating for the loss of their activator.

The accessibility of *Tfrc* as a transmembrane protein makes it an attractive target for therapy. Monoclonal antibodies against *Tfrc* are used in renal cell carcinoma immunotherapy demonstrating that depriving a tumour of iron reduces its growth and proliferation (Kudoh et al., 1994). This should encourage the investigation of the role of TFRC in lymphoma.

5.7. Co-occurrence and mutual exclusivity

The top fifty most clonal inserts in the spleen DNA were compared pairwise by Fisher's exact test to measure the probability of them being mutated together (co-occurrence) or not (mutually exclusive) in different genotype contexts.

Mutations that co-occur are thought to have cooperating roles in oncogenesis in that they function in different oncogenic pathways and it is advantageous to abrogate both. Mutually exclusive genes may participate in the same oncogenic pathway and so it is redundant to mutate both. It is also possible that mutation of one gene is not tolerated if the first has already been mutated and so mutation of it is selected against.

Detailed below an example contingency table used to prepare the co-occurrence and mutual exclusivity heat maps.

2 tailed p value ~ 0.008	<i>Mycn inserts</i>	Non Mycn inserts	Fold enrichment
<i>Mice with clonal Myc insert</i>	5	8670	0.30
<i>Mice without clonal Myc insert</i>	102	58762	3.00

Table 5.13. Example contingency table test demonstrating the mutual exclusivity of Myc and Mycn mutations

When assessing the collaboration between *Myc* and *Mycn* mutation in the BCL2 transgenic mice a pairwise comparison is made between the number of end stage inserts within a 100kb window of *Mycn* and the rest of the genome in the mice bearing a clonal insertion of *Myc* and mice lacking a clonal insert of *Myc*. *Mycn* and *Myc* mutations are significantly mutually exclusive with a 0.33 reduction in mutation of *Mycn* in tumours with a clonal *Myc* mutation. Genotypes included: CrebbpCreBC2, CrebbpBCL2, CreBCL2 and BCL2.

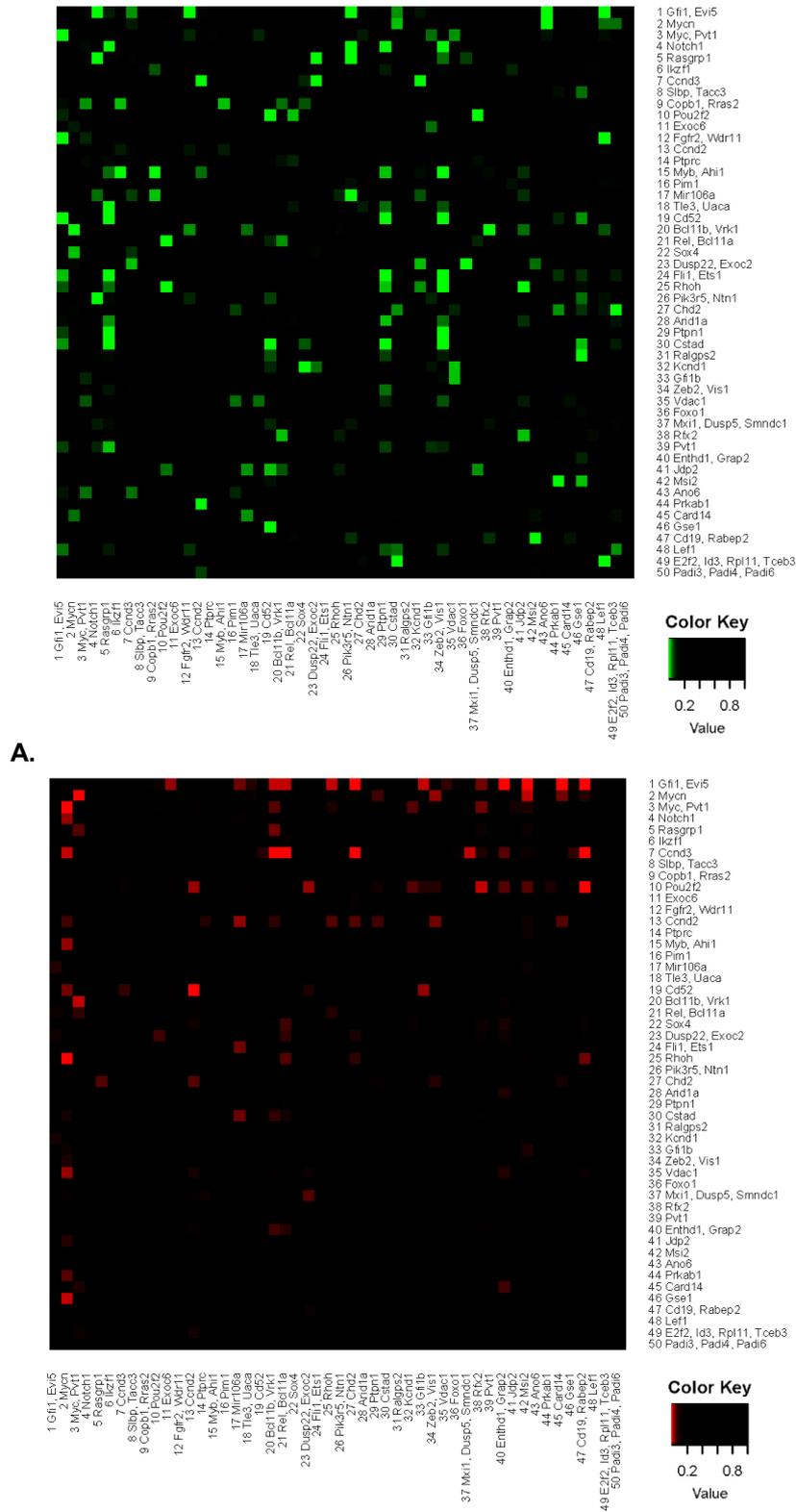


Figure 5.13. Co-occurring and mutually exclusive insertions in the *BCL2* transgenic tumours

Collaborations between the top 50 most frequently mutated loci (clonality ≥ 0.1) in the screen were compared to see if they co-occurred with or were mutually exclusive to each other in a *BCL2* overexpressing context. Genotypes included: CrebbpCreBCL2, CrebbpBCL2, CreBCL2 and BCL2.

A Shows examples of co-occurring genes. *Gfi1* most frequently occurs with *Notch1*, *12NA*, *Pik3r5*, *Gm8849* and *48NA*. This indicates that together *BCL2*, *Gfi1* and these genes cooperate in lymphomagenesis. **B** shows examples of exclusivity. *Gfi1* mutations rarely occur with *Chd2*, *Enthd*, *Msi1*, *Tbc1d16* or *Rabep2*.

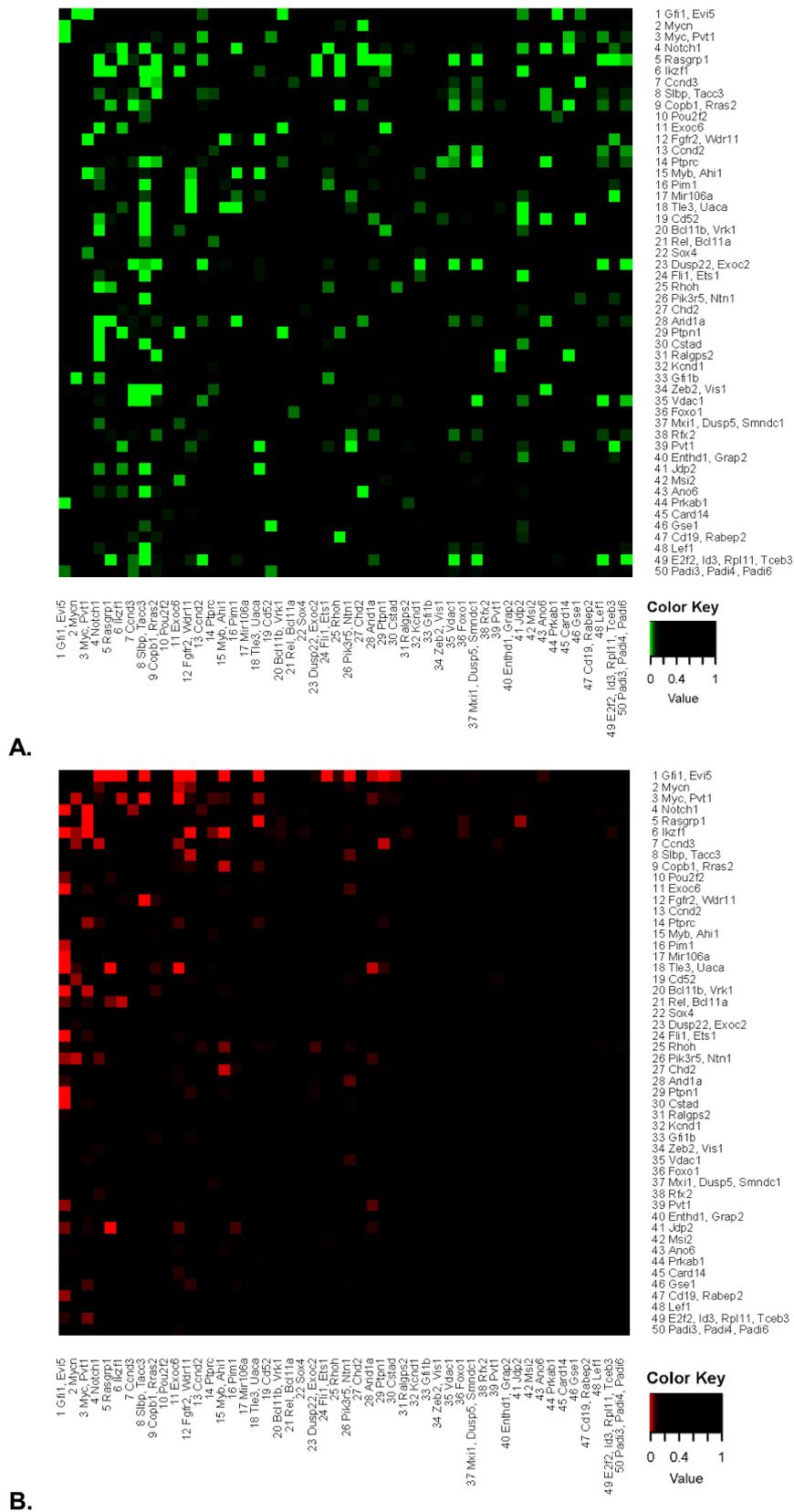


Figure 5.14. Co-occurring and mutually exclusive insertions in the non-transgenic *BCL2* tumours

Collaborations between the top 50 most frequently mutated loci (clonality >0.1) in the screen when compared in a non-transgenic *BCL2* background to see if they co-occurred with or were mutually exclusive to each other. There is more cooperation in these genotypes compared to those that overexpress *BCL2* possible because more oncogenic aberrations need to occur to cause lymphoma. Genotypes included: CrebbpCre, Crebbp, Cre and wildtype.

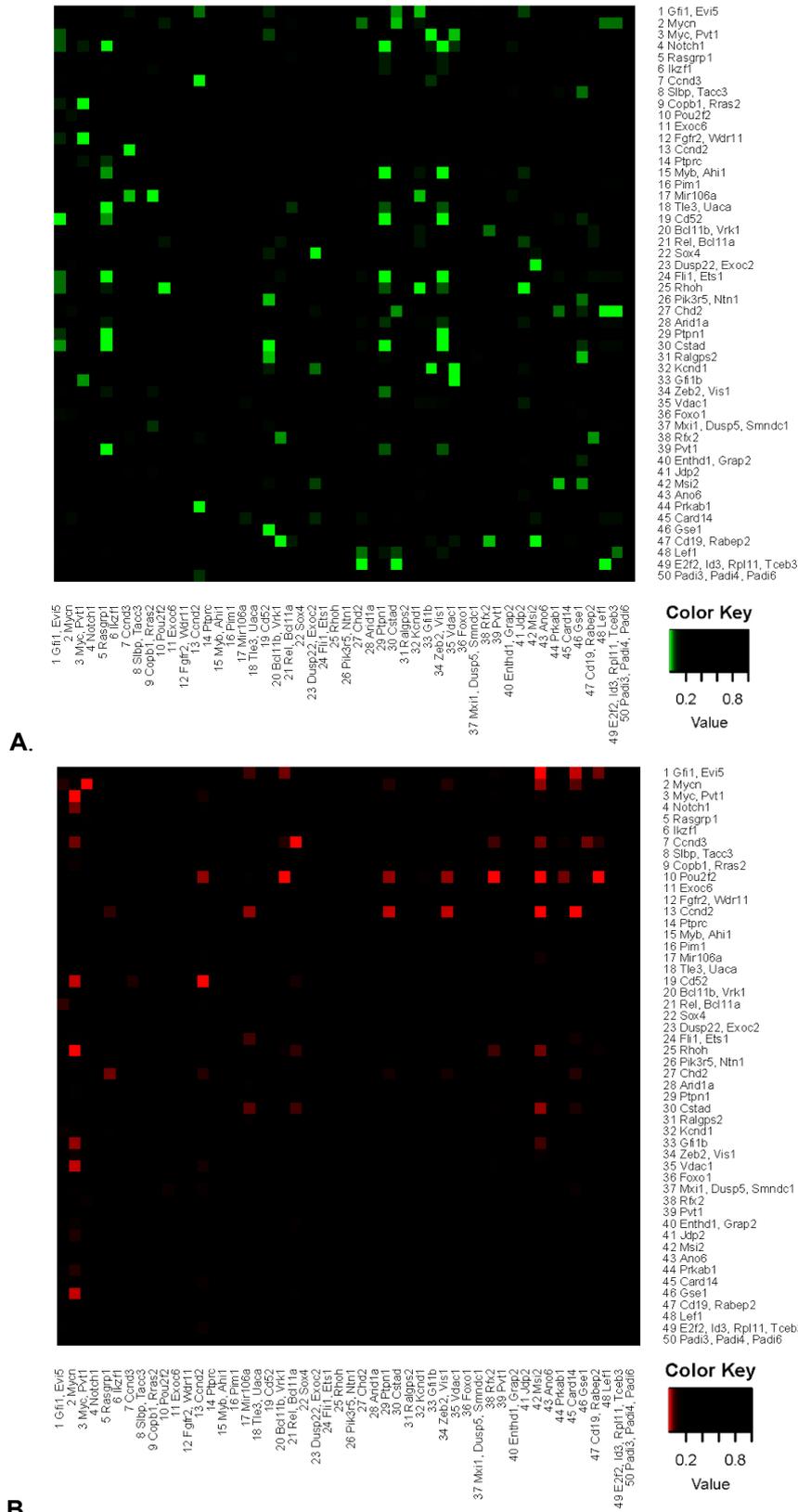


Figure 5.15. Co-occurring and mutually exclusive insertions in the *Crebbp* deficient tumours

Collaborations between the top 50 most frequently mutated loci (clonality >0.1) in the screen when compared in a *Crebbp* deficient background. In both plots there are examples of parallel mutations that are always co-mutate of mutually exclusive with a particular gene suggesting that they work in the same or different oncogenic pathway to that primary gene. For example *Pou2f2* and *Cnd2* are both mutually exclusive to *Ptpn1*, *Zeb2*, and *Msi1* but only *Pou2f2* is mutually exclusive to *Rfx2*.

Identifying genes that cooperate or are mutually exclusive to each other provides information about mechanisms of synergy in oncogenic pathways in an unbiased manner and may be useful when used to highlight new strategies for treatment that target parallel pathways. This adds an additional layer to the genotype specific data allowing the construction of mutation profiles with multiple connected nodes.

For example the oncogene *c-Myc* is frequently mutated in human lymphoma and has been shown to cooperate with *Pim1* or *Gfi1* overexpression in murine transplant experiments (van Lohuizen et al., 1989; Verbeek et al., 1991). Loss of a collaborator like *Pim-1* has been shown to delay *c-Myc* driven tumour onset (Jacobs et al., 1999; van der Lugt et al., 1995). This illustrates how when a target of interest is not druggable or has undesirable effects alternative treatment approaches might be devised by taking advantage of synthetic lethality

Analysis of mutually exclusivity has proven to be informative in patient treatment. The lung adenocarcinoma drugs gefitinib and erlotinib are used in the treatment of non-small cell lung cancers with KRAS mutations. EGFR mutations are mutually exclusive to KRAS mutations and so these drugs are of no use to patients with mutant EGFR (Pao et al., 2005).

Mouse models have also be screened to assess drug resistance. In a murine model of leukaemia, mice were infected with *Bcr-Abl1* expressing retroviruses and then treated with imatinib a Bcr-Abl1 inhibitor. These mice showed selection for *Runx1* and *Runx3* mutations which were shown to protect the cells from imatinib-mediated apoptosis making it an ineffective treatment (Miething et al., 2007).

5.8. Summary of insertion data

- A novel high-throughput DNA library preparation pipeline, UMI-LM-PCR, was designed for use on the Illumina sequencing platform that facilitates the deep-sequencing and quantitative measurement of retroviral insertions.
- Addition of a UMI to the DNA facilitates recognition of individual fragments aiding filtering and allowing the absolute quantification of inserts.
- 671 samples processed by UMI-LM-PCR yielding 66,292,198 mapped and paired reads, this represents the largest dataset with deepest sequencing to date.
- Hundreds of significantly mutated loci have been identified in the screen, including known cancer genes and novel candidates with a possible role in lymphoma.
- Hundreds of significantly mutated loci have been found to be mutated specifically in the B-cell lymphomas.
- Dissemination analysis highlighted the heterogeneity of disease within one animal which could be relevant to treatment.
- Analysis of genotype specific genes identifies relationships between genes. Specifically *Pou2f2* is downregulated in *BCL2* overexpressing tumours and *Tfr3* is activated in *Crebbp* deficient tumours.
- Co-occurrence and mutual exclusivity analysis can inform the study of how mutations cooperate in the development of lymphoma and may suggest combinatorial therapeutic approaches targeting parallel pathways.

6. Discussion and future work

6.1. Overview

Non-Hodgkin's lymphomas (NHL) are a spectrum of hematopoietic cancers accounting for 4% of new cancer diagnosis each year. The majority of NHL are derived from B-cells including, diffuse large B-cell lymphomas (DLBCL) and follicular lymphoma (FL). Both DLBCL and FL are mature B-cell neoplasms derived from germinal centre or post-germinal centre B-cells (reviewed in Basso and Dalla-Favera, 2015).

BCL2 and *CREBBP* are frequently mutated in human B-NHL and have been shown to be early events in disease progression that often occur concurrently (Morin et al., 2011). *BCL2* chromosomal translocation is an archetypal marker of B-NHL causing its overexpression and increasing a cells resistance to apoptosis (Küppers, 2005). Recently heterozygous *CREBBP* loss of function mutations were found frequently in B-NHL suggesting it was a haploinsufficient tumour suppressor gene in B-NHL (Pasqualucci et al., 2011b).

This Thesis presented a study of the role of overexpression of *BCL2* and *Crebbp* deficiency in oncogenesis achieved by modulating their expression in the haematopoietic compartment and additionally performing an insertional mutagenesis screen using MoMuLV in the resulting transgenic mice. To support and enable this, an innovative high-throughput DNA library preparation pipeline was designed that was applicable to an Illumina sequencing platform and facilitated the deep-sequencing and quantitative measurement of retroviral insertions.

6.2. *Crebbp* and the acceleration of *BCL2* driven lymphomagenesis.

This study is the first experimental model to confirm that loss of *Crebbp* is able to accelerate *BCL2* driven B-cell lymphoma where it acts as an haploinsufficient tumour suppressor gene. This is complementary to the human sequencing data that indicated that they were co-mutated in the same tumours.

The B-cell neoplasms were extensively characterised by flow cytometry and demonstrated an immunophenotype similar to mature human B-cell lymphomas with expansion of germinal centre B-cells and increased class switching. This confirms that the model generated (*Crebbp*^{+/*fl*} ; *Aicda*^{+/*Cre*} ; *VavP-BCL2*) is a faithful model of human germinal centre or post germinal centre derived lymphoma. Time point experiments using this model could yield information about early events in disease progression.

Having demonstrated that *Crebbp* is a tumour suppressor gene, analysis of retrovirus induced, genotype specific insertion analysis gives a unique opportunity to identify regulators of *Crebbp* tumour suppressor activity or downstream targets of it. By mining the insertion data we aim to identify regulators of it which could then be modulated to enhance its tumour suppressive effects or allow us to act downstream of it when *Crebbp* is lost.

Genes that are less frequently mutated when *Crebbp* is lost could function in the same pathways, meaning their mutation is redundant when *Crebbp* is already lost. If a gene is less frequently mutated in the *Crebbp* deficient mice it could also indicate that it is an upstream regulator of *Crebbp* and so it serves no purpose when *Crebbp* is already lost. It is also possible that mutation of some genes is not tolerated when *Crebbp* is lost. If a gene is more frequently mutated in the genotypes with intact *Crebbp* it suggests that they cooperate possibly by acting in different oncogenic pathways or that it is a downstream effector of *Crebbp* and mutations of it reduces the tumour suppressive effects of *Crebbp*.

CREBBP's central role in the protein-protein interactome makes it a difficult node to approach with multiple interconnected effects possible. Genotype specific mutations give the opportunity to investigate functional interactions that may not be obvious by other means.

An example of a promising genotype specific mutation is *Tfrc* which was specifically mutated in the *Crebbp* deficient tumours. The mutations of *Tfrc* are expected to upregulate it which is known to advantage tumour cells. I postulate that this upregulation of *Tfrc* is to compensate for loss of activation by Gata-1 which would be upregulated by *Crebbp* under physiological conditions.

Another important source is the information about co-occurring and mutually exclusively gene mutations. This data will provide information about which genes activate or abrogate the tumorigenic effects of

another gene. Additionally co-occurrence and mutual exclusivity analysis can inform treatment choices indicating if they will be effective or not and if they will invoke resistance.

Interrogation of clonal insertions yields information about the clonal evolution of a tumour. Clonal inserts are likely to have occurred early in tumour development and have a driving role in oncogenesis. This could provide information about alterations that could be used as early biomarkers of disease.

Together the genotype specific mutations and co-occurrence and mutually exclusivity of mutation can highlight profiles of mutations in a particular disease highlighting complementary target and identifying multiple nodes of intervention in a tumour.

6.3. Technological advancement

These discoveries have been supported by the development of UMI-LM-PCR. This novel method facilitates the high-throughput library preparation and next-generation sequencing of tumours yielding an unprecedented quantity and quality of data.

This study represents the most comprehensive insertional mutagenesis screen to date with the greatest depth of sequencing. In total 671 tumours were processed, yielding 66,292,198 mapped and paired reads. The scale of this study allows the identification of novel candidates with increased statistical power.

The application of UMIs, allowing the unique identification of each individual MoMuLV insertion point, allows direct quantification of insertions and therefore the clonality of insertions. Clonal insertions have undergone greater selection than sub clonal ones highlighting their role as drivers in lymphomagenesis. The increased number of insertions and accurate estimate of clonality facilitates the generation of dense interaction maps which can be interrogated for cancer driving pathways and important nodes of connection. Interrogation of subclonal mutations is only possible now with greater sequencing coverage and ability to quantify the insertions and may provide novel insight into the tumour microenvironment.

Though many of the major mutations in B-NHL have been described knowledge of the rare mutations or the other genes in the same pathways expands the range of potential targets as an alternative to standard chemotherapeutic regimes of lymphoma therapy.

6.4. Future work

The most important thing to do now is to validate the role of candidate genes that were particularly clonal or genotype specific in this screen.

Hairpin and overexpression vectors can be used *in vitro* to assess the effect of candidate genes in lymphoma cell line. Basic cell cycle parameters such as proliferation and apoptosis could be assessed to then prioritise candidates for validations *in vitro*. Considering the crucial importance of the germinal centre in these tumours it is fundamental to study the role of these candidates in whole system with intact immune system.

We have optimised a protocol to over express oncogenes and knock-down tumour suppressor genes *in vivo* using a retroviral vector expressing a gene open reading frame or a knock-down hairpin. Donor spleen B-cells are taken from a transgenic mouse already with sensitising mutations, cultured and then transfected with the retrovirus. After sorting the cells containing the gene of interest and transplanted into a sublethally irradiated gender matched wild type recipient via tail vein injection. These animals are then monitored for signs of disease until they develop lymphoma. Those infected with an oncogene will succumb to illness faster than the empty vector control mice as will those with a knocked-down tumour suppressor gene. The retroviral vector is tagged with a fluorescent reporter so their presence and causative role can be confirmed in eventual tumours. The use of transgenic donor cells facilitates the use of mutation combinations that are thought to cooperate in lymphomagenesis. For example, it would be interesting to transplant *Pou2f2* hairpin or overexpression vector into *Crebbp*^{+fl} ; *Aicda*^{+Cre} ; VavP-BCL2⁺ cells to resolve if its overexpression cooperates with overexpression of *BCL2* and loss of *Crebbp*.

To better characterise the model it would be useful to perform RNAseq and proteomic analysis on pre-malignant transgenic animals and compare them to the spontaneous lymphomas to compare alterations in transcript and proteins levels that might be associated with lymphomagenesis. This could also be done with the MoMuLV induced lymphoma with the view to correlation the insertion data to expression levels of mRNAs and proteins. This would define the functional role of insertions, whether is upregulates or downregulates expression and thus give more information about gene collaborations and networks.

Another area of investigation should be the interrogation of gene expression levels in the pre-malignant samples collected to assess the effects of loss of *Crebbp*. We have sorted germinal centre B-cells from mice with the genotypes of this screen for these analyses.

An IM screen has recently been integrated with chromatin conformation capture (Hi-C) to investigate the role of structural conformation in MoMuLV insertions. Usually when annotating candidates in this screen the genome is considered as linear which is too simplistic; when considering distant integration sites

might converge when coiled to affect a gene or cis inserts might actually affect a different gene when in a 3D conformation. This can result in the decreased significance of relevant genes and the incorrect assignment of inserts to a gene (Babaei et al., 2015). This phenomenon has been measured in human beta-globulin locus whose activation in K562 cells requires the formation of a ~40kb loop to put the necessary control region and activated genes in a position to activate the locus (Vakoc et al., 2005). This work also adds a caveat to mutually exclusive mutations and suggest that their lack of co-mutation might be in fact because they are affected by the same mutations rather than because there are functionally linked. Application of this approach to our data set would increase the statistical power of data and make it more comparable to the biological state.

6.5. Concluding remarks

Recent sequencing of human lymphomas have been successful in identifying driver mutations within exons (Lohr et al., 2012; Morin et al., 2011; Pasqualucci et al., 2011a). However, human populations are highly polymorphic and that, in combination with the mutator phenotype generated by many tumours, can make it difficult to subtract the background noise to evidence driver mutations. Furthermore the significance of human point mutations in cancer is derived primarily from the ratio of synonymous and non-synonymous coding mutations conversely the role of non-coding mutations is more difficult to determine. Thus the ability to only consider strongly significant mutations also means that less frequent but still important mutations are missed.

Here the complementarity of murine models is particularly useful, providing a controlled environment in which individual mutations can be interrogated. Genes highlighted by common insertion sites in insertional mutagenesis screens can prioritise which genes should be investigated in human tumours and if there is not an equivalent gene, much information can be gained from looking at mutations networks to highlight important collaborators of a gene. This research will serve as a foundation for future work which will hopefully identify novel mutations driving lymphoma and aid drug development.

References

- Abramson, J.S., and Shipp, M.A. (2005). Advances in the biology and therapy of diffuse large B-cell lymphoma: moving toward a molecularly targeted approach. *Blood* *106*, 1164–1174.
- Albihn, A., Johnsen, J.I., and Henriksson, M.A. (2010). MYC in oncogenesis and as a target for cancer therapies. *Adv. Cancer Res.* *107*, 163–224.
- De Alboran, I.M., O'Hagan, R.C., Gärtner, F., Malynn, B., Davidson, L., Rickert, R., Rajewsky, K., DePinho, R.A., and Alt, F.W. (2001). Analysis of C-MYC function in normal cells via conditional gene-targeted mutation. *Immunity* *14*, 45–55.
- Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* *403*, 503–511.
- Aloia, A.L., Duffy, L., Pak, V., Lee, K.E., Sanchez-Martinez, S., Derse, D., Heidecker, G., Cornetta, K., and Rein, A. (2013). A reporter system for replication-competent gammaretroviruses: the inGluc-MLV-DERSE assay. *Gene Ther.* *20*, 169–176.
- Aukema, S.M., Siebert, R., Schuurin, E., van Imhoff, G.W., Kluin-Nelemans, H.C., Boerma, E.-J., and Kluin, P.M. (2011). Double-hit B-cell lymphomas. *Blood* *117*, 2319–2331.
- Babaei, S., Akhtar, W., de Jong, J., Reinders, M., and de Ridder, J. (2015). 3D hotspots of recurrent retroviral insertions reveal long-range interactions with cancer genes. *Nat. Commun.* *6*, 6381.
- Bartholdi, D., Roelfsema, J.H., Papadia, F., Breuning, M.H., Niedrist, D., Hennekam, R.C., Schinzel, A., and Peters, D.J.M. (2007). Genetic heterogeneity in Rubinstein-Taybi syndrome: delineation of the phenotype of the first patients carrying mutations in EP300. *J. Med. Genet.* *44*, 327–333.
- Bartsch, O., Schmidt, S., Richter, M., Morlot, S., Seemanová, E., Wiebe, G., and Rasi, S. (2005). DNA sequencing of CREBBP demonstrates mutations in 56% of patients with Rubinstein-Taybi syndrome (RSTS) and in another patient with incomplete RSTS. *Hum. Genet.* *117*, 485–493.
- Basso, K., and Dalla-Favera, R. (2015). Germinal centres and B cell lymphomagenesis. *Nat. Rev. Immunol.* *15*, 172–184.
- Bedford, D.C., and Brindle, P.K. (2012). Is histone acetylation the most important physiological function for CBP and p300? *Aging (Albany, NY).* *4*, 247–255.
- Bedford, D.C., Kasper, L.H., Fukuyama, T., and Brindle, P.K. (2010). Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases. *Epigenetics* *5*, 9–15.
- Blobel, G.A., Nakajima, T., Eckner, R., Montminy, M., and Orkin, S.H. (1998). CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 2061–2066.
- Bürkle, A., Niedermeier, M., Schmitt-Gräff, A., Wierda, W.G., Keating, M.J., and Burger, J.A. (2007). Overexpression of the CXCR5 chemokine receptor, and its ligand, CXCL13 in B-cell chronic lymphocytic leukemia. *Blood* *110*, 3316–3325.

- Cambier, J.C., Gauld, S.B., Merrell, K.T., and Vilen, B.J. (2007). B-cell anergy: from transgenic models to naturally occurring anergic B cells? *Nat. Rev. Immunol.* 7, 633–643.
- Campo, E., Swerdlow, S.H., Harris, N.L., Pileri, S., Stein, H., and Jaffe, E.S. (2011). The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood* 117, 5019–5032.
- Cantani, A., and Gagliesi, D. Rubinstein-Taybi syndrome. Review of 732 cases and analysis of the typical traits. *Eur. Rev. Med. Pharmacol. Sci.* 2, 81–87.
- Cheng, Y., Zak, O., Aisen, P., Harrison, S.C., and Walz, T. (2004). Structure of the human transferrin receptor-transferrin complex. *Cell* 116, 565–576.
- Chitambar, C.R., Massey, E.J., and Seligman, P.A. (1983). Regulation of transferrin receptor expression on human leukemic cells during proliferation and induction of differentiation. Effects of gallium and dimethylsulfoxide. *J. Clin. Invest.* 72, 1314–1325.
- Chung, Y.R., Schatoff, E., and Abdel-Wahab, O. (2012). Epigenetic alterations in hematopoietic malignancies. *Int. J. Hematol.* 96, 413–427.
- Coiffier, B., Lepage, E., Briere, J., Herbrecht, R., Tilly, H., Bouabdallah, R., Morel, P., Van Den Neste, E., Salles, G., Gaulard, P., et al. (2002). CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N. Engl. J. Med.* 346, 235–242.
- Colicelli, J., and Goff, S.P. (1988). Sequence and spacing requirements of a retrovirus integration site. *J. Mol. Biol.* 199, 47–59.
- Copeland, N.G., and Jenkins, N.A. (2010). Harnessing transposons for cancer gene discovery. *Nat. Rev. Cancer* 10, 696–706.
- Corcoran, L.M., Karvelas, M., Nossal, G.J., Ye, Z.S., Jacks, T., and Baltimore, D. (1993). Oct-2, although not required for early B-cell development, is critical for later B-cell maturation and for postnatal survival. *Genes Dev.* 7, 570–582.
- Cuypers, H.T., Selten, G., Quint, W., Zijlstra, M., Maandag, E.R., Boelens, W., van Wezenbeek, P., Melief, C., and Berns, A. (1984). Murine leukemia virus-induced T-cell lymphomagenesis: integration of proviruses in a distinct chromosomal region. *Cell* 37, 141–150.
- Czabotar, P.E., Lessene, G., Strasser, A., and Adams, J.M. (2014). Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol.* 15, 49–63.
- Van Delft, M.F., Wei, A.H., Mason, K.D., Vandenberg, C.J., Chen, L., Czabotar, P.E., Willis, S.N., Scott, C.L., Day, C.L., Cory, S., et al. (2006). The BH3 mimetic ABT-737 targets selective Bcl-2 proteins and efficiently induces apoptosis via Bak/Bax if Mcl-1 is neutralized. *Cancer Cell* 10, 389–399.
- Devon, R.S., Porteous, D.J., and Brookes, A.J. (1995). Splinkerettes--improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res.* 23, 1644–1645.
- Dogan, A., and Isaacson, P.G. (2003). Splenic marginal zone lymphoma. *Semin. Diagn. Pathol.* 20, 121–127.

- Donnell, K. a O., Yu, D., Zeller, K.I., Kim, J., Racke, F., Dang, C. V, and Thomas-tikhonenko, A. (2006). Activation of Transferrin Receptor 1 by c-Myc Enhances Cellular Proliferation and Tumorigenesis †. *Mol. Cell. Biol.* *26*, 2373–2386.
- Dunn, T.L., Ross, I.L., and Hume, D.A. (1996). Transcription factor Oct-2 is expressed in primary murine macrophages. *Blood* *88*, 4072.
- Dürig, J., Schmücker, U., and Dührsen, U. (2001). Differential expression of chemokine receptors in B cell malignancies. *Leukemia* *15*, 752–756.
- Egle, A., Harris, A.W., Bath, M.L., O'Reilly, L., and Cory, S. (2004). VavP-Bcl2 transgenic mice develop follicular lymphoma preceded by germinal center hyperplasia. *Blood* *103*, 2276–2283.
- Ellisen, L.W., Bird, J., West, D.C., Soreng, A.L., Reynolds, T.C., Smith, S.D., and Sklar, J. (1991). TAN-1, the human homolog of the Drosophila notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell* *66*, 649–661.
- Fan, H. (1997). Leukemogenesis by Moloney murine leukemia virus: a multistep process. *Trends Microbiol.* *5*, 74–82.
- Feremans, W., Bujan, W., Neve, P., Delville, J.P., and Schandene, L. (1991). CD71 phenotype and the value of gallium imaging in lymphomas. *Am. J. Hematol.* *36*, 215–216.
- Frese, K.K., and Tuveson, D.A. (2007). Maximizing mouse cancer models. *Nat. Rev. Cancer* *7*, 645–658.
- Gabriel, S. (2006). Variation in the human genome and the inherited basis of common disease. *Semin. Oncol.* *33*, S46–S49.
- Geimer Le Lay, A.-S., Oravecz, A., Mastio, J., Jung, C., Marchal, P., Ebel, C., Dembélé, D., Jost, B., Le Gras, S., Thibault, C., et al. (2014). The tumor suppressor Ikaros shapes the repertoire of notch target genes in T cells. *Sci. Signal.* *7*, ra28.
- Gourzi, P., Leonova, T., and Papavasiliou, F.N. (2006). A role for activation-induced cytidine deaminase in the host response against a transforming retrovirus. *Immunity* *24*, 779–786.
- Green, M.R., Gentles, A.J., Nair, R. V, Irish, J.M., Kihira, S., Liu, C.L., Kela, I., Hopmans, E.S., Myklebust, J.H., Ji, H., et al. (2013). Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. *Blood* *121*, 1604–1611.
- Gu, W., and Roeder, R.G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* *90*, 595–606.
- Haas, K.M., and Tedder, T.F. (2005). Role of the CD19 and CD21/35 receptor complex in innate immunity, host defense and autoimmunity. *Adv. Exp. Med. Biol.* *560*, 125–139.
- Han, J.-H., Akira, S., Calame, K., Beutler, B., Selsing, E., and Imanishi-Kari, T. (2007). Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell and Toll-like receptors. *Immunity* *27*, 64–75.

- Hanecak, R., Pattengale, P.K., and Fan, H. (1988). Addition of substitution of simian virus 40 enhancer sequences into the Moloney murine leukemia virus (M-MuLV) long terminal repeat yields infectious M-MuLV with altered biological properties. *J. Virol.* *62*, 2427–2436.
- Harris, N.L., Jaffe, E.S., Stein, H., Banks, P.M., Chan, J.K., Cleary, M.L., Delsol, G., De Wolf-Peeters, C., Falini, B., and Gatter, K.C. (1994). A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood* *84*, 1361–1392.
- He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W., and Rosenfeld, M.G. (1989). Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* *340*, 35–41.
- Heckman, C.A., Duan, H., Garcia, P.B., and Boxer, L.M. (2006). Oct transcription factors mediate t(14;18) lymphoma cell survival by directly regulating bcl-2 expression. *Oncogene* *25*, 888–898.
- Huser, C.A., Gilroy, K.L., de Ridder, J., Kilbey, A., Borland, G., Mackay, N., Jenkins, A., Bell, M., Herzyk, P., van der Weyden, L., et al. (2014). Insertional mutagenesis and deep profiling reveals gene hierarchies and a Myc/p53-dependent bottleneck in lymphomagenesis. *PLoS Genet.* *10*, e1004167.
- Jacobs, J.J., Scheijen, B., Voncken, J.W., Kieboom, K., Berns, A., and van Lohuizen, M. (1999). Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. *Genes Dev.* *13*, 2678–2690.
- Jäger, U., Böcskő, S., Le, T., Mitterbauer, G., Bolz, I., Chott, A., Kneba, M., Mannhalter, C., and Nadel, B. (2000). Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. *Blood* *95*, 3520–3529.
- John, L.B., and Ward, A.C. (2011). The Ikaros gene family: transcriptional regulators of hematopoiesis and immunity. *Mol. Immunol.* *48*, 1272–1278.
- De Jong, J., de Ridder, J., van der Weyden, L., Sun, N., van Uitert, M., Berns, A., van Lohuizen, M., Jonkers, J., Adams, D.J., and Wessels, L.F.A. (2011). Computational identification of insertional mutagenesis targets for cancer gene discovery. *Nucleic Acids Res.* *39*, e105–e105.
- Kang, S.M., Tsang, W., Doll, S., Scherle, P., Ko, H.S., Tran, A.C., Lenardo, M.J., and Staudt, L.M. (1992). Induction of the POU domain transcription factor Oct-2 during T-cell activation by cognate antigen. *Mol. Cell. Biol.* *12*, 3149–3154.
- Kang-Decker, N., Tong, C., Boussouar, F., Baker, D.J., Xu, W., Leontovich, A. a, Taylor, W.R., Brindle, P.K., and van Deursen, J.M. a (2004). Loss of CBP causes T cell lymphomagenesis in synergy with p27Kip1 insufficiency. *Cancer Cell* *5*, 177–189.
- Khalaj, A.J., Yoon, J., Nakai, J., Winchester, Z., Moore, S.M., Yoo, T., Martinez-Torres, L., Kumar, S., Itoh, N., and Tiwari-Woodruff, S.K. (2013). Estrogen receptor (ER) β expression in oligodendrocytes is required for attenuation of clinical disease by an ER β ligand. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 19125–19130.
- Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2012). Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* *9*, 72–74.

- Klein, U., Tu, Y., Stolovitzky, G.A., Mattioli, M., Cattoretti, G., Husson, H., Freedman, A., Inghirami, G., Cro, L., Baldini, L., et al. (2001). Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J. Exp. Med.* *194*, 1625–1638.
- Kool, J., and Berns, A. (2009). High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. *Nat. Rev. Cancer* *9*, 389–399.
- Kool, J., Uren, A.G., Martins, C.P., Sie, D., de Ridder, J., Turner, G., van Uiter, M., Matentzoglou, K., Lagcher, W., Krimpenfort, P., et al. (2010). Insertional mutagenesis in mice deficient for p15Ink4b, p16Ink4a, p21Cip1, and p27Kip1 reveals cancer gene interactions and correlations with tumor phenotypes. *Cancer Res.* *70*, 520–531.
- Koudijs, M.J., Klijn, C., van der Weyden, L., Kool, J., ten Hoeve, J., Sie, D., Prasetyanti, P.R., Schut, E., Kas, S., Whipp, T., et al. (2011). High-throughput semiquantitative analysis of insertional mutations in heterogeneous tumors. *Genome Res.* *21*, 2181–2189.
- Kridel, R., Sehn, L.H., and Gascoyne, R.D. (2012). Pathogenesis of follicular lymphoma. *J. Clin. Invest.* *122*, 3424–3431.
- Kudoh, S., Stanley, J., Edinger, M.G., Tubbs, R.R., Klein, E., Bukowski, R.M., and Finke, J.H. (1994). T lymphocytes infiltrating renal cell carcinoma have a reduced expression of transferrin receptor. *Int. J. Cancer* *58*, 369–375.
- Kung, A.L., Rebel, V.I., Bronson, R.T., Ch'ng, L.E., Sieff, C.A., Livingston, D.M., and Yao, T.P. (2000). Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. *Genes Dev.* *14*, 272–277.
- Küppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. *Nat. Rev. Cancer* *5*, 251–262.
- Küppers, R., and Dalla-Favera, R. (2001). Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* *20*, 5580–5594.
- Küppers, R., Zhao, M., Hansmann, M.L., and Rajewsky, K. (1993). Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. *EMBO J.* *12*, 4955–4967.
- Ladanyi, M., Offit, K., Jhanwar, S.C., Filippa, D.A., and Chaganti, R.S. (1991). MYC rearrangement and translocations involving band 8q24 in diffuse large cell lymphomas. *Blood* *77*, 1057–1063.
- Lambert, P.F., Kashanchi, F., Radonovich, M.F., Shiekhhattar, R., and Brady, J.N. (1998). Phosphorylation of p53 serine 15 increases interaction with CBP. *J. Biol. Chem.* *273*, 33048–33053.
- Latchman, D.S. (1996). The Oct-2 transcription factor. *Int. J. Biochem. Cell Biol.* *28*, 1081–1083.
- Lauchle, J.O., Kim, D., Le, D.T., Akagi, K., Crone, M., Krisman, K., Warner, K., Bonifas, J.M., Li, Q., Coakley, K.M., et al. (2009). Response and resistance to MEK inhibition in leukaemias initiated by hyperactive Ras. *Nature* *461*, 411–414.
- Lazo, P.A., Lee, J.S., and Tschlis, P.N. (1990). Long-distance activation of the Myc protooncogene by provirus insertion in Mlvi-1 or Mlvi-4 in rat T-cell lymphomas. *Proc. Natl. Acad. Sci. U. S. A.* *87*, 170–173.

- Lenz, G., Wright, G.W., Emre, N.C.T., Kohlhammer, H., Dave, S.S., Davis, R.E., Carty, S., Lam, L.T., Shaffer, A.L., Xiao, W., et al. (2008). Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 13520–13525.
- Li, H., Kaminski, M.S., Li, Y., Yildiz, M., Ouillet, P., Jones, S., Fox, H., Jacobi, K., Saiya-Cork, K., Bixby, D., et al. (2014). Mutations in linker histone genes HIST1H1 B, C, D, and E; OCT2 (POU2F2); IRF8; and ARID1A underlying the pathogenesis of follicular lymphoma. *Blood* *123*, 1487–1498.
- Li, S., Lin, P., Young, K.H., Kanagal-Shamanna, R., Yin, C.C., and Medeiros, L.J. (2013). MYC/BCL2 double-hit high-grade B-cell lymphoma. *Adv. Anat. Pathol.* *20*, 315–326.
- Lillycrop, K.A., and Latchman, D.S. (1992). Alternative splicing of the Oct-2 transcription factor RNA is differentially regulated in neuronal cells and B cells and results in protein isoforms with opposite effects on the activity of octamer/TAATGARAT-containing promoters. *J. Biol. Chem.* *267*, 24960–24965.
- Limpens, J., Stad, R., Vos, C., de Vlaam, C., de Jong, D., van Ommen, G.J., Schuurin, E., and Kluin, P.M. (1995). Lymphoma-associated translocation t(14;18) in blood B cells of normal individuals. *Blood* *85*, 2528–2536.
- Lipsick, J.S., and Wang, D.M. (1999). Transformation by v-Myb. *Oncogene* *18*, 3047–3055.
- Lohr, J.G., Stojanov, P., Lawrence, M.S., Auclair, D., Chapuy, B., Sougnez, C., Cruz-Gordillo, P., Knoechel, B., Asmann, Y.W., Slager, S.L., et al. (2012). Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 3879–3884.
- Van Lohuizen, M., Verbeek, S., Krimpenfort, P., Domen, J., Saris, C., Radaszkiewicz, T., and Berns, A. (1989). Predisposition to lymphomagenesis in pim-1 transgenic mice: cooperation with c-myc and N-myc in murine leukemia virus-induced tumors. *Cell* *56*, 673–682.
- Van Lohuizen, M., Verbeek, S., Scheijen, B., Wientjens, E., van der Gulden, H., and Berns, A. (1991). Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell* *65*, 737–752.
- Van der Lugt, N.M., Domen, J., Verhoeven, E., Linders, K., van der Gulden, H., Allen, J., and Berns, A. (1995). Proviral tagging in E mu-myc transgenic mice lacking the Pim-1 proto-oncogene leads to compensatory activation of Pim-2. *EMBO J.* *14*, 2536–2544.
- Manolov, G., and Manolova, Y. (1972). Marker band in one chromosome 14 from Burkitt lymphomas. *Nature* *237*, 33–34.
- Mao, C., Jiang, L., Melo-Jorge, M., Puthenveetil, M., Zhang, X., Carroll, M.C., and Imanishi-Kari, T. (2004). T cell-independent somatic hypermutation in murine B cells with an immature phenotype. *Immunity* *20*, 133–144.
- Menezes, J., Salgado, R.N., Acquadro, F., Gómez-López, G., Carralero, M.C., Barroso, A., Mercadillo, F., Espinosa-Hevia, L., Talavera-Casañas, J.G., Pisano, D.G., et al. (2013). ASXL1, TP53 and IKZF3 mutations are present in the chronic phase and blast crisis of chronic myeloid leukemia. *Blood Cancer J.* *3*, e157.
- Miething, C., Grundler, R., Mugler, C., Brero, S., Hoepfl, J., Geigl, J., Speicher, M.R., Ottmann, O., Peschel, C., and Duyster, J. (2007). Retroviral insertional mutagenesis identifies RUNX genes involved in

- chronic myeloid leukemia disease persistence under imatinib treatment. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 4594–4599.
- Molnár, A., Wu, P., Largespada, D.A., Vortkamp, A., Scherer, S., Copeland, N.G., Jenkins, N.A., Bruns, G., and Georgopoulos, K. (1996). The Ikaros gene encodes a family of lymphocyte-restricted zinc finger DNA binding proteins, highly conserved in human and mouse. *J. Immunol.* *156*, 585–592.
- Moloney, J.B. (1960). Biological studies on a lymphoid-leukemia virus extracted from sarcoma 37. I. Origin and introductory investigations. *J. Natl. Cancer Inst.* *24*, 933–951.
- Montoto, S., and Fitzgibbon, J. (2011). Transformation of indolent B-cell lymphomas. *J. Clin. Oncol.* *29*, 1827–1834.
- Morin, R.D., Johnson, N.A., Severson, T.M., Mungall, A.J., An, J., Goya, R., Paul, J.E., Boyle, M., Woolcock, B.W., Kuchenbauer, F., et al. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat. Genet.* *42*, 181–185.
- Morin, R.D., Mendez-Lago, M., Mungall, A.J., Goya, R., Mungall, K.L., Corbett, R.D., Johnson, N.A., Severson, T.M., Chiu, R., Field, M., et al. (2011). Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* *476*, 298–303.
- Mowat, M., Cheng, A., Kimura, N., Bernstein, A., and Benchimol, S. (1985). Rearrangements of the cellular p53 gene in erythroleukaemic cells transformed by Friend virus. *Nature* *314*, 633–636.
- Mullighan, C.G., Miller, C.B., Radtke, I., Phillips, L.A., Dalton, J., Ma, J., White, D., Hughes, T.P., Le Beau, M.M., Pui, C.-H., et al. (2008). BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* *453*, 110–114.
- Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* *274*, 18470–18476.
- Nakajima, T., Uchida, C., Anderson, S.F., Lee, C.G., Hurwitz, J., Parvin, J.D., and Montminy, M. (1997). RNA helicase A mediates association of CBP with RNA polymerase II. *Cell* *90*, 1107–1112.
- Nguyen, M., Marcellus, R.C., Roulston, A., Watson, M., Serfass, L., Murthy Madiraju, S.R., Goulet, D., Viallet, J., Bélec, L., Billot, X., et al. (2007). Small molecule obatoclax (GX15-070) antagonizes MCL-1 and overcomes MCL-1-mediated resistance to apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 19512–19517.
- Niuro, H., and Clark, E.A. (2002). Regulation of B-cell fate by antigen-receptor signals. *Nat. Rev. Immunol.* *2*, 945–956.
- Nojima, T., Haniuda, K., Moutai, T., Matsudaira, M., Mizokawa, S., Shiratori, I., Azuma, T., and Kitamura, D. (2011). In-vitro derived germinal centre B cells differentially generate memory B or plasma cells in vivo. *Nat. Commun.* *2*, 465.
- O'Brien, S., Moore, J.O., Boyd, T.E., Larratt, L.M., Skotnicki, A.B., Koziner, B., Chanan-Khan, A.A., Seymour, J.F., Gribben, J., Itri, L.M., et al. (2009). 5-year survival in patients with relapsed or refractory chronic lymphocytic leukemia in a randomized, phase III trial of fludarabine plus cyclophosphamide with or without oblimersen. *J. Clin. Oncol.* *27*, 5208–5212.

- Ogilvy, S., Metcalf, D., Print, C.G., Bath, M.L., Harris, A.W., and Adams, J.M. (1999a). Constitutive Bcl-2 expression throughout the hematopoietic compartment affects multiple lineages and enhances progenitor cell survival. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 14943–14948.
- Ogilvy, S., Metcalf, D., Gibson, L., Bath, M.L., Harris, A.W., and Adams, J.M. (1999b). Promoter elements of *vav* drive transgene expression in vivo throughout the hematopoietic compartment. *Blood* *94*, 1855–1863.
- Okazaki, I., Kotani, A., and Honjo, T. (2007). Role of AID in tumorigenesis. *Adv. Immunol.* *94*, 245–273.
- Okosun, J., Bödör, C., Wang, J., Araf, S., Yang, C.-Y., Pan, C., Boller, S., Cittaro, D., Bozek, M., Iqbal, S., et al. (2014). Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat. Genet.* *46*, 176–181.
- Pao, W., Wang, T.Y., Riely, G.J., Miller, V.A., Pan, Q., Ladanyi, M., Zakowski, M.F., Heelan, R.T., Kris, M.G., and Varmus, H.E. (2005). KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med.* *2*, e17.
- Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Küppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* *412*, 341–346.
- Pasqualucci, L., Trifonov, V., Fabbri, G., Ma, J., Rossi, D., Chiarenza, A., Wells, V.A., Grunn, A., Messina, M., Elliot, O., et al. (2011a). Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.* *43*, 830–837.
- Pasqualucci, L., Dominguez-Sola, D., Chiarenza, A., Fabbri, G., Grunn, A., Trifonov, V., Kasper, L.H., Lerach, S., Tang, H., Ma, J., et al. (2011b). Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* *471*, 189–195.
- Pfreundschuh, M., Schubert, J., Ziepert, M., Schmits, R., Mohren, M., Lengfelder, E., Reiser, M., Nickenig, C., Clemens, M., Peter, N., et al. (2008). Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20+ B-cell lymphomas: a randomised controlled trial (RICOVER-60). *Lancet Oncol.* *9*, 105–116.
- Phan, R.T., and Dalla-Favera, R. (2004). The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* *432*, 635–639.
- Phan, R.T., Saito, M., Basso, K., Niu, H., and Dalla-Favera, R. (2005). BCL6 interacts with the transcription factor Miz-1 to suppress the cyclin-dependent kinase inhibitor p21 and cell cycle arrest in germinal center B cells. *Nat. Immunol.* *6*, 1054–1060.
- Pieper, K., Grimbacher, B., and Eibel, H. (2013). B-cell biology and development. *J. Allergy Clin. Immunol.* *131*, 959–971.
- Quail, M.A., Swerdlow, H., and Turner, D.J. (2009). Improved protocols for the illumina genome analyzer sequencing system. *Curr. Protoc. Hum. Genet. Chapter 18*, Unit 18.2.
- Radtke, F., Wilson, A., Mancini, S.J.C., and MacDonald, H.R. (2004). Notch regulation of lymphocyte development and function. *Nat. Immunol.* *5*, 247–253.

- Radtke, F., MacDonald, H.R., and Tacchini-Cottier, F. (2013). Regulation of innate and adaptive immunity by Notch. *Nat. Rev. Immunol.* *13*, 427–437.
- Rajewsky, K. (1996). Clonal selection and learning in the antibody system. *Nature* *381*, 751–758.
- Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.-T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W., et al. (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. *Nature* *440*, 105–109.
- Rando, O.J., and Chang, H.Y. (2009). Genome-wide views of chromatin structure. *Annu. Rev. Biochem.* *78*, 245–271.
- Raney, B.J., Dreszer, T.R., Barber, G.P., Clawson, H., Fujita, P.A., Wang, T., Nguyen, N., Paten, B., Zweig, A.S., Karolchik, D., et al. (2014). Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* *30*, 1003–1005.
- Rangarajan, A., and Weinberg, R.A. (2003). Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat. Rev. Cancer* *3*, 952–959.
- Ranger, A.M., Zha, J., Harada, H., Datta, S.R., Danial, N.N., Gilmore, A.P., Kutok, J.L., Le Beau, M.M., Greenberg, M.E., and Korsmeyer, S.J. (2003). Bad-deficient mice develop diffuse large B cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 9324–9329.
- Richardson, D.R. (2002). Therapeutic potential of iron chelators in cancer therapy. *Adv. Exp. Med. Biol.* *509*, 231–249.
- De Ridder, J., Uren, A., Kool, J., Reinders, M., and Wessels, L. (2006). Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. *PLoS Comput. Biol.* *2*, e166.
- Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A., et al. (2008). AID Is Required for the Chromosomal Breaks in c-myc that Lead to c-myc/IgH Translocations. *Cell* *135*, 1028–1038.
- Roberts, A.W., Seymour, J.F., Brown, J.R., Wierda, W.G., Kipps, T.J., Khaw, S.L., Carney, D.A., He, S.Z., Huang, D.C.S., Xiong, H., et al. (2012). Substantial susceptibility of chronic lymphocytic leukemia to BCL2 inhibition: results of a phase I study of navitoclax in patients with relapsed or refractory disease. *J. Clin. Oncol.* *30*, 488–496.
- Romero, F., Martínez-A, C., Camonis, J., and Rebollo, A. (1999). Aiolos transcription factor controls cell death in T cells by regulating Bcl-2 expression and its cellular localization. *EMBO J.* *18*, 3419–3430.
- Rommel, P.C., Bosque, D., Gitlin, A.D., Croft, G.F., Heintz, N., Casellas, R., Nussenzweig, M.C., Kriaucionis, S., and Robbiani, D.F. (2013). Fate mapping for activation-induced cytidine deaminase (AID) marks non-lymphoid cells during mouse development. *PLoS One* *8*, e69208.
- Ruminy, P., Etancelin, P., Couronné, L., Parmentier, F., Rainville, V., Mareschal, S., Bohers, E., Burgot, C., Cornic, M., Bertrand, P., et al. (2011). The isotype of the BCR as a surrogate for the GCB and ABC molecular subtypes in diffuse large B-cell lymphoma. *Leukemia* *25*, 681–688.
- Sauvageau, M., Miller, M., Lemieux, S., Lessard, J., Hébert, J., and Sauvageau, G. (2008). Quantitative expression profiling guided by common retroviral insertion sites reveals novel and cell type specific cancer genes in leukemia. *Blood* *111*, 790–799.

- Schaffner, C., Idler, I., Stilgenbauer, S., Döhner, H., and Lichter, P. (2000). Mantle cell lymphoma is characterized by inactivation of the ATM gene. *Proc. Natl. Acad. Sci. U. S. A.* *97*, 2773–2778.
- Selten, G., Cuypers, H.T., and Berns, A. (1985). Proviral activation of the putative oncogene Pim-1 in MuLV induced T-cell lymphomas. *EMBO J.* *4*, 1793–1798.
- Shimizu, R., Trainor, C.D., Nishikawa, K., Kobayashi, M., Ohneda, K., and Yamamoto, M. (2007). GATA-1 Self-association Controls Erythroid Development in Vivo. *J. Biol. Chem.* *282*, 15862–15871.
- Shore, S.K., Tantravahi, R. V., and Reddy, E.P. (2002). Transforming pathways activated by the v-Abl tyrosine kinase. *Oncogene* *21*, 8568–8576.
- Singh, H., Sen, R., Baltimore, D., and Sharp, P.A. (1986). A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature* *319*, 154–158.
- Souers, A.J., Levenson, J.D., Boghaert, E.R., Ackler, S.L., Catron, N.D., Chen, J., Dayton, B.D., Ding, H., Enschede, S.H., Fairbrother, W.J., et al. (2013). ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* *19*, 202–208.
- Starkey, C.R., Lobelle-Rich, P.A., Granger, S.W., Granger, S., Brightman, B.K., Fan, H., and Levy, L.S. (1998). Tumorigenic potential of a recombinant retrovirus containing sequences from Moloney murine leukemia virus and feline leukemia virus. *J. Virol.* *72*, 1078–1084.
- Staudt, L.M., Clerc, R.G., Singh, H., LeBowitz, J.H., Sharp, P.A., and Baltimore, D. (1988). Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. *Science* *241*, 577–580.
- Stevenson, F.K., Sahota, S.S., Ottensmeier, C.H., Zhu, D., Forconi, F., and Hamblin, T.J. (2001). The occurrence and significance of V gene mutations in B cell-derived human malignancy. *Adv. Cancer Res.* *83*, 81–116.
- Stewart, M., Cameron, E., Campbell, M., McFarlane, R., Toth, S., Lang, K., Onions, D., and Neil, J.C. (1993). Conditional expression and oncogenicity of c-myc linked to a CD2 gene dominant control region. *Int. J. Cancer* *53*, 1023–1030.
- Stewart, M., Mackay, N., Hanlon, L., Blyth, K., Scobie, L., Cameron, E., and Neil, J.C. (2007). Insertional mutagenesis reveals progression genes and checkpoints in MYC/Runx2 lymphomas. *Cancer Res.* *67*, 5126–5133.
- Strasser, A., Harris, A.W., Jacks, T., and Cory, S. (1994). DNA damage can induce apoptosis in proliferating lymphoid cells via p53-independent mechanisms inhibitable by Bcl-2. *Cell* *79*, 329–339.
- Swain, A., and Coffin, J.M. (1993). Influence of sequences in the long terminal repeat and flanking cell DNA on polyadenylation of retroviral transcripts. *J. Virol.* *67*, 6265–6269.
- Thorns, C., Kalies, K., Fischer, U., Höfig, K., Krokowski, M., Feller, A.C., Merz, H., and Bernd, H.-W. (2007). Significant high expression of CD23 in follicular lymphoma of the inguinal region. *Histopathology* *50*, 716–719.
- Trudel, S., Li, Z.H., Rauw, J., Tiedemann, R.E., Wen, X.Y., and Stewart, A.K. (2007). Preclinical studies of the pan-Bcl inhibitor obatoclax (GX015-070) in multiple myeloma. *Blood* *109*, 5430–5438.

- Uren, A., and Berns, A. (2009). Jump-starting cancer gene discovery. *Nat. Biotechnol.* *27*, 251–252.
- Uren, A.G., Kool, J., Berns, A., and van Lohuizen, M. (2005). Retroviral insertional mutagenesis: past, present and future. *Oncogene* *24*, 7656–7672.
- Uren, A.G., Kool, J., Matentzoglou, K., de Ridder, J., Mattison, J., van Uitert, M., Lagcher, W., Sie, D., Tanger, E., Cox, T., et al. (2008). Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks. *Cell* *133*, 727–741.
- Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell* *17*, 453–462.
- Verbeek, S., van Lohuizen, M., van der Valk, M., Domen, J., Kraal, G., and Berns, A. (1991). Mice bearing the E mu-myc and E mu-pim-1 transgenes develop pre-B-cell leukemia prenatally. *Mol. Cell Biol.* *11*, 1176–1179.
- Wang, J.H., Avitahl, N., Cariappa, A., Friedrich, C., Ikeda, T., Renold, A., Andrikopoulos, K., Liang, L., Pillai, S., Morgan, B.A., et al. (1998). Aiolos regulates B cell activation and maturation to effector state. *Immunity* *9*, 543–553.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* *420*, 520–562.
- Weiss, L.M., Warnke, R.A., Sklar, J., and Cleary, M.L. (1987). Molecular analysis of the t(14;18) chromosomal translocation in malignant lymphomas. *N. Engl. J. Med.* *317*, 1185–1189.
- Weng, A.P., Ferrando, A.A., Lee, W., Morris, J.P., Silverman, L.B., Sanchez-Irizarry, C., Blacklow, S.C., Look, A.T., and Aster, J.C. (2004). Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* *306*, 269–271.
- Willis, T.G., and Dyer, M.J. (2000). The role of immunoglobulin translocations in the pathogenesis of B-cell malignancies. *Blood* *96*, 808–822.
- Wilson, W.H., O'Connor, O.A., Czuczman, M.S., LaCasce, A.S., Gerecitano, J.F., Leonard, J.P., Tulpule, A., Dunleavy, K., Xiong, H., Chiu, Y.-L., et al. (2010). Navitoclax, a targeted high-affinity inhibitor of BCL-2, in lymphoid malignancies: a phase 1 dose-escalation study of safety, pharmacokinetics, pharmacodynamics, and antitumour activity. *Lancet. Oncol.* *11*, 1149–1159.
- Witkowski, M.T., Cimmino, L., Hu, Y., Trimarchi, T., Tagoh, H., McKenzie, M.D., Best, S.A., Tuohey, L., Willson, T.A., Nutt, S.L., et al. (2015). Activated Notch counteracts Ikaros tumor suppression in mouse and human T-cell acute lymphoblastic leukemia. *Leukemia* *29*, 1301–1311.
- Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* *300*, 1749–1751.
- Xu, W. (2006). Global transcriptional coactivators CREB-binding protein and p300 are highly essential collectively but not individually in peripheral B cells. *Blood* *107*, 4407–4416.
- Xu, Q., Schlabach, M.R., Hannon, G.J., and Elledge, S.J. (2009). Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 2289–2294.

Yamanaka, M., Inaka, K., Furubayashi, N., Matsushima, M., Takahashi, S., Tanaka, H., Sano, S., Sato, M., Kobayashi, T., and Tanaka, T. (2011). Optimization of salt concentration in PEG-based crystallization solutions. *J. Synchrotron Radiat.* *18*, 84–87.

Yang, X.J., Ogryzko, V. V., Nishikawa, J., Howard, B.H., and Nakatani, Y. (1996). A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* *382*, 319–324.

Yao, T.P., Ku, G., Zhou, N., Scully, R., and Livingston, D.M. (1996). The nuclear hormone receptor coactivator SRC-1 is a specific target of p300. *Proc. Natl. Acad. Sci. U. S. A.* *93*, 10626–10631.

Yui, M.A., and Rothenberg, E. V (2014). Developmental gene networks: a triathlon on the course to T cell identity. *Nat. Rev. Immunol.* *14*, 529–545.

Zhang, J., Grubor, V., Love, C.L., Banerjee, A., Richards, K.L., Mieczkowski, P.A., Dunphy, C., Choi, W., Au, W.Y., Srivastava, G., et al. (2013). Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci.* *110*, 1398–1403.

Zhang, Z., Hofmann, C., Casanova, E., Schütz, G., and Lutz, B. (2004). Generation of a conditional allele of the CBP gene in mouse. *Genesis* *40*, 82–89.

Appendices

A.1 Full list of Adaptors, MRC#854-945

Below the upper strand adaptor sequences (96) used in this screen. Each upper strand was ligated to MRC#443 universal lower strand to make 96 unique adaptors. Details include: the common orange portion which is equivalent to Illumina Oligo C and facilitates binding of the DNA to the sequencing chip, in blue the 10bp unique index, in green the UMI and in pink the portion that is complementary to the universal lower strand.

Adaptor Name	Upper Strand Adaptor Sequence
MRC#854 Adaptor #101	AATGATACGGCGACCACCGAGATCTACACGAAATACAGANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#855 Adaptor #102	AATGATACGGCGACCACCGAGATCTACACGTATCGGTATNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#856 Adaptor #103	AATGATACGGCGACCACCGAGATCTACACAGCAACTCCNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#857 Adaptor #104	AATGATACGGCGACCACCGAGATCTACACAAACTTCCAANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#858 Adaptor #105	AATGATACGGCGACCACCGAGATCTACACTTGAGTAACNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#859 Adaptor #106	AATGATACGGCGACCACCGAGATCTACACCACACGATCTNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#860 Adaptor #107	AATGATACGGCGACCACCGAGATCTACACCGATTTTAGTNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#861 Adaptor #108	AATGATACGGCGACCACCGAGATCTACACCTAACAGATNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#862 Adaptor #109	AATGATACGGCGACCACCGAGATCTACACGCTAACATGCNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#863 Adaptor #110	AATGATACGGCGACCACCGAGATCTACACCACAGGAAAANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#864 Adaptor #111	AATGATACGGCGACCACCGAGATCTACACAACGATATCCNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#865 Adaptor #112	AATGATACGGCGACCACCGAGATCTACACTAGGAGATGNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#866 Adaptor #113	AATGATACGGCGACCACCGAGATCTACACCGTCAGCATANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#867 Adaptor #114	AATGATACGGCGACCACCGAGATCTACACAAAGCCTACTNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#868 Adaptor #115	AATGATACGGCGACCACCGAGATCTACACGACGTCATACNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#869 Adaptor #116	AATGATACGGCGACCACCGAGATCTACACCGAATAAACANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#870 Adaptor #117	AATGATACGGCGACCACCGAGATCTACACAAACCCTCTCANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#871 Adaptor #118	AATGATACGGCGACCACCGAGATCTACACAGTCACACGCNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#872 Adaptor #119	AATGATACGGCGACCACCGAGATCTACACCCAAAGCTGANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#873 Adaptor #120	AATGATACGGCGACCACCGAGATCTACACGTACACAGTCNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#874 Adaptor #121	AATGATACGGCGACCACCGAGATCTACACTCCGACTTGNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#875 Adaptor #122	AATGATACGGCGACCACCGAGATCTACACTATGAAGTCANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#876 Adaptor #123	AATGATACGGCGACCACCGAGATCTACACCGCCTTGTGANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#877 Adaptor #124	AATGATACGGCGACCACCGAGATCTACACTTCGACGCTANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#878 Adaptor #125	AATGATACGGCGACCACCGAGATCTACACTGCCCGTCCGNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#879 Adaptor #126	AATGATACGGCGACCACCGAGATCTACACTAACAGAGANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#880 Adaptor #127	AATGATACGGCGACCACCGAGATCTACACCGTACTCTANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT

MRC#881 Adaptor #128	AATGATACGGCGACCACCGAGATCTACACTAGACTCATGNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#882 Adaptor #129	AATGATACGGCGACCACCGAGATCTACACCAGTAAACGCNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#883 Adaptor #130	AATGATACGGCGACCACCGAGATCTACACCGTTACTGACNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#884 Adaptor #131	AATGATACGGCGACCACCGAGATCTACACGCCCTTTCCTNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#885 Adaptor #132	AATGATACGGCGACCACCGAGATCTACACACACGGTCTTNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#886 Adaptor #133	AATGATACGGCGACCACCGAGATCTACACCAGATTAAGGNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#887 Adaptor #134	AATGATACGGCGACCACCGAGATCTACACATCAAGTAAGNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#888 Adaptor #135	AATGATACGGCGACCACCGAGATCTACACCATCGCTACNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#889 Adaptor #136	AATGATACGGCGACCACCGAGATCTACACGGAACATGGANNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#890 Adaptor #137	AATGATACGGCGACCACCGAGATCTACACCGACAAACAANNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#891 Adaptor #138	AATGATACGGCGACCACCGAGATCTACACATACGAAGCCNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#892 Adaptor #139	AATGATACGGCGACCACCGAGATCTACACGTGATCTAACNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#893 Adaptor #140	AATGATACGGCGACCACCGAGATCTACACACAAACAAACNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#894 Adaptor #141	AATGATACGGCGACCACCGAGATCTACACTTAGACCATCNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#895 Adaptor #142	AATGATACGGCGACCACCGAGATCTACACGCGACATACNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#896 Adaptor #143	AATGATACGGCGACCACCGAGATCTACACACATCAACCTNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#897 Adaptor #144	AATGATACGGCGACCACCGAGATCTACACTCGCTTCAAGNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#898 Adaptor #145	AATGATACGGCGACCACCGAGATCTACACTGAACATATGNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#899 Adaptor #146	AATGATACGGCGACCACCGAGATCTACACGAGTCTACTNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#900 Adaptor #147	AATGATACGGCGACCACCGAGATCTACACTTAGATACANNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#901 Adaptor #148	AATGATACGGCGACCACCGAGATCTACACTGGACCCACNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#902 Adaptor #149	AATGATACGGCGACCACCGAGATCTACACGTGCTTCGAGNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#903 Adaptor #150	AATGATACGGCGACCACCGAGATCTACACCACTAGGTANNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#904 Adaptor #151	AATGATACGGCGACCACCGAGATCTACACGCCATTGAANNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#905 Adaptor #152	AATGATACGGCGACCACCGAGATCTACACTAACCCGCGTNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#906 Adaptor #153	AATGATACGGCGACCACCGAGATCTACACCTACCGATGANNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#907 Adaptor #154	AATGATACGGCGACCACCGAGATCTACACACTTACAGTCNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#908 Adaptor #155	AATGATACGGCGACCACCGAGATCTACACAATATGCAGANNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#909 Adaptor #156	AATGATACGGCGACCACCGAGATCTACACATTACCGCGNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#910 Adaptor #157	AATGATACGGCGACCACCGAGATCTACACACCAGAGTTCNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#911 Adaptor #158	AATGATACGGCGACCACCGAGATCTACACACATGGAATCNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#912 Adaptor #159	AATGATACGGCGACCACCGAGATCTACACCCAGTCTATNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#913 Adaptor #160	AATGATACGGCGACCACCGAGATCTACACGCCATGTTNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#914 Adaptor #161	AATGATACGGCGACCACCGAGATCTACACGATGTTGCTNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MRC#915 Adaptor #162	AATGATACGGCGACCACCGAGATCTACACTTCGTAAGAGNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

MRC#916 Adaptor #163	AATGATACGGCGACCACCGAGATCTACACCCAGATACTNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#917 Adaptor #164	AATGATACGGCGACCACCGAGATCTACACCCCTTGACGNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#918 Adaptor #165	AATGATACGGCGACCACCGAGATCTACACTCGACTTACGNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#919 Adaptor #166	AATGATACGGCGACCACCGAGATCTACACATCCATGGTANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#920 Adaptor #167	AATGATACGGCGACCACCGAGATCTACACAAGTCTACGCNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#921 Adaptor #168	AATGATACGGCGACCACCGAGATCTACACATTCAGTATGNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#922 Adaptor #169	AATGATACGGCGACCACCGAGATCTACACTTGGACATACNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#923 Adaptor #170	AATGATACGGCGACCACCGAGATCTACACAGTGTGGTGNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#924 Adaptor #171	AATGATACGGCGACCACCGAGATCTACACTCTCCAGGCANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#925 Adaptor #172	AATGATACGGCGACCACCGAGATCTACACTACTCCGGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#926 Adaptor #173	AATGATACGGCGACCACCGAGATCTACACCTAAGGTCTTNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#927 Adaptor #174	AATGATACGGCGACCACCGAGATCTACACGAACCAAGCTNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#928 Adaptor #175	AATGATACGGCGACCACCGAGATCTACACATATACTGGANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#929 Adaptor #176	AATGATACGGCGACCACCGAGATCTACACGACCTTGTNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#930 Adaptor #177	AATGATACGGCGACCACCGAGATCTACACGGCTTCATAANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#931 Adaptor #178	AATGATACGGCGACCACCGAGATCTACACCAGAATAACAANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#932 Adaptor #179	AATGATACGGCGACCACCGAGATCTACACAACACACGAGNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#933 Adaptor #180	AATGATACGGCGACCACCGAGATCTACACAGCTTACTAANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#934 Adaptor #181	AATGATACGGCGACCACCGAGATCTACACCATGGTTCGGNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#935 Adaptor #182	AATGATACGGCGACCACCGAGATCTACACCACCTGTCCNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#936 Adaptor #183	AATGATACGGCGACCACCGAGATCTACACGGCATGGTANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#937 Adaptor #184	AATGATACGGCGACCACCGAGATCTACACTACGTCATCTNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#938 Adaptor #185	AATGATACGGCGACCACCGAGATCTACACTCAAATGGCCNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#939 Adaptor #186	AATGATACGGCGACCACCGAGATCTACACATTCAGCAATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#940 Adaptor #187	AATGATACGGCGACCACCGAGATCTACACTAATCCCTAANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#941 Adaptor #188	AATGATACGGCGACCACCGAGATCTACACAGATTAGGTGNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#942 Adaptor #189	AATGATACGGCGACCACCGAGATCTACACAAGTAGCTAANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#943 Adaptor #190	AATGATACGGCGACCACCGAGATCTACACACGGTCTGTNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#944 Adaptor #191	AATGATACGGCGACCACCGAGATCTACACCGATAGTCCNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#945 Adaptor #192	AATGATACGGCGACCACCGAGATCTACACTCGAACGTAANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#946 Adaptor #193	AATGATACGGCGACCACCGAGATCTACACATCATTAGCGNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#943 Adaptor #194	AATGATACGGCGACCACCGAGATCTACACTGTCCAATANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#944 Adaptor #195	AATGATACGGCGACCACCGAGATCTACACATCACGCACGNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT
MRC#945 Adaptor #196	AATGATACGGCGACCACCGAGATCTACACTACTAATCAGNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT

A.2 Secondary PCR primers, MRC#946-1029

Listed, the secondary PCR primers. In purple, Illumina Oligo D for binding to the sequencing chip, in blue the unique index and in red the portions complementary to MoMuLV.

Secondary PCR Primer Name	Secondary PCR Primer Sequence
MRC#946 secondary PCR #101	CAAGCAGAAGACGGCATAACGAGATTCTGTATTTTCGCTAGCTTGCCAAACCTACAGGTGG
MRC#947 secondary PCR #102	CAAGCAGAAGACGGCATAACGAGATATACCGATACGCTAGCTTGCCAAACCTACAGGTGG
MRC#948 secondary PCR #103	CAAGCAGAAGACGGCATAACGAGATGGGAGTTGCTGCTAGCTTGCCAAACCTACAGGTGG
MRC#949 secondary PCR #104	CAAGCAGAAGACGGCATAACGAGATTTGGAAGTTTGCTAGCTTGCCAAACCTACAGGTGG
MRC#950 secondary PCR #105	CAAGCAGAAGACGGCATAACGAGATGGTTACTCAAGCTAGCTTGCCAAACCTACAGGTGG
MRC#951 secondary PCR #106	CAAGCAGAAGACGGCATAACGAGATAGATCGTGTGGCTAGCTTGCCAAACCTACAGGTGG
MRC#952 secondary PCR #107	CAAGCAGAAGACGGCATAACGAGATACTAAAATCGGCTAGCTTGCCAAACCTACAGGTGG
MRC#953 secondary PCR #108	CAAGCAGAAGACGGCATAACGAGATATCTTGTAGGCTAGCTTGCCAAACCTACAGGTGG
MRC#954 secondary PCR #109	CAAGCAGAAGACGGCATAACGAGATGCATGTTAGCGCTAGCTTGCCAAACCTACAGGTGG
MRC#955 secondary PCR #110	CAAGCAGAAGACGGCATAACGAGATTTTTCCTGTGGCTAGCTTGCCAAACCTACAGGTGG
MRC#956 secondary PCR #111	CAAGCAGAAGACGGCATAACGAGATGGATATCGTTGCTAGCTTGCCAAACCTACAGGTGG
MRC#957 secondary PCR #112	CAAGCAGAAGACGGCATAACGAGATCAATCTCCTAGCTAGCTTGCCAAACCTACAGGTGG
MRC#958 secondary PCR #113	CAAGCAGAAGACGGCATAACGAGATTATGCTGACGGCTAGCTTGCCAAACCTACAGGTGG
MRC#959 secondary PCR #114	CAAGCAGAAGACGGCATAACGAGATAGTAGGCTTTGCTAGCTTGCCAAACCTACAGGTGG
MRC#960 secondary PCR #115	CAAGCAGAAGACGGCATAACGAGATGTATGACGTCGCTAGCTTGCCAAACCTACAGGTGG
MRC#961 secondary PCR #116	CAAGCAGAAGACGGCATAACGAGATTGTTTATTCGGCTAGCTTGCCAAACCTACAGGTGG
MRC#962 secondary PCR #117	CAAGCAGAAGACGGCATAACGAGATGAGAGGGTTTGCTAGCTTGCCAAACCTACAGGTGG
MRC#963 secondary PCR #118	CAAGCAGAAGACGGCATAACGAGATGCGTGTGACTGCTAGCTTGCCAAACCTACAGGTGG
MRC#964 secondary PCR #119	CAAGCAGAAGACGGCATAACGAGATTCAGCTTTGGGCTAGCTTGCCAAACCTACAGGTGG
MRC#965 secondary PCR #120	CAAGCAGAAGACGGCATAACGAGATGACTGTGTACGCTAGCTTGCCAAACCTACAGGTGG
MRC#966 secondary PCR #121	CAAGCAGAAGACGGCATAACGAGATCAAAGTCGGAGCTAGCTTGCCAAACCTACAGGTGG
MRC#967 secondary PCR #122	CAAGCAGAAGACGGCATAACGAGATTGACTTCATAGCTAGCTTGCCAAACCTACAGGTGG
MRC#968 secondary PCR #123	CAAGCAGAAGACGGCATAACGAGATTCACAAGGCGGCTAGCTTGCCAAACCTACAGGTGG
MRC#969 secondary PCR #124	CAAGCAGAAGACGGCATAACGAGATTAGCGTCGAAGCTAGCTTGCCAAACCTACAGGTGG
MRC#970 secondary PCR #125	CAAGCAGAAGACGGCATAACGAGATCGGACGGGCGAGCTAGCTTGCCAAACCTACAGGTGG
MRC#971 secondary PCR #126	CAAGCAGAAGACGGCATAACGAGATCTCTCTGTAGCTAGCTTGCCAAACCTACAGGTGG
MRC#972 secondary PCR #127	CAAGCAGAAGACGGCATAACGAGATTAGAGTCACGGCTAGCTTGCCAAACCTACAGGTGG
MRC#973 secondary PCR #128	CAAGCAGAAGACGGCATAACGAGATCATGAGTCTAGCTAGCTTGCCAAACCTACAGGTGG
MRC#974 secondary PCR #129	CAAGCAGAAGACGGCATAACGAGATGCGTTTACTGGCTAGCTTGCCAAACCTACAGGTGG
MRC#975 secondary PCR #130	CAAGCAGAAGACGGCATAACGAGATGTCAGTAACGGCTAGCTTGCCAAACCTACAGGTGG

MRC#976 secondary PCR #131	CAAGCAGAAGACGGCATAACGAGATAGGAAAGGGCGCTAGCTTGCCAAACCTACAGGTGG
MRC#977 secondary PCR #132	CAAGCAGAAGACGGCATAACGAGATAAGACCGTGTGCTAGCTTGCCAAACCTACAGGTGG
MRC#978 secondary PCR #133	CAAGCAGAAGACGGCATAACGAGATCCTTAATCTGGCTAGCTTGCCAAACCTACAGGTGG
MRC#979 secondary PCR #134	CAAGCAGAAGACGGCATAACGAGATCTTACTTGATGCTAGCTTGCCAAACCTACAGGTGG
MRC#980 secondary PCR #135	CAAGCAGAAGACGGCATAACGAGATGGTAGCGATGGCTAGCTTGCCAAACCTACAGGTGG
MRC#981 secondary PCR #136	CAAGCAGAAGACGGCATAACGAGATTCATGTTCCGCTAGCTTGCCAAACCTACAGGTGG
MRC#982 secondary PCR #137	CAAGCAGAAGACGGCATAACGAGATTTGTTTTCGGCTAGCTTGCCAAACCTACAGGTGG
MRC#983 secondary PCR #138	CAAGCAGAAGACGGCATAACGAGATGGCTTCGTATGCTAGCTTGCCAAACCTACAGGTGG
MRC#984 secondary PCR #139	CAAGCAGAAGACGGCATAACGAGATGTAGATCACGCTAGCTTGCCAAACCTACAGGTGG
MRC#985 secondary PCR #140	CAAGCAGAAGACGGCATAACGAGATGTTTGTGTTGCTAGCTTGCCAAACCTACAGGTGG
MRC#986 secondary PCR #141	CAAGCAGAAGACGGCATAACGAGATGATGGTCTAAGCTAGCTTGCCAAACCTACAGGTGG
MRC#987 secondary PCR #142	CAAGCAGAAGACGGCATAACGAGATGGTATGTCGCGCTAGCTTGCCAAACCTACAGGTGG
MRC#988 secondary PCR #143	CAAGCAGAAGACGGCATAACGAGATAGGTTGATGCTAGCTTGCCAAACCTACAGGTGG
MRC#989 secondary PCR #144	CAAGCAGAAGACGGCATAACGAGATCTGAAGCGAGCTAGCTTGCCAAACCTACAGGTGG
MRC#990 secondary PCR #145	CAAGCAGAAGACGGCATAACGAGATACATAGTTCAGCTAGCTTGCCAAACCTACAGGTGG
MRC#991 secondary PCR #146	CAAGCAGAAGACGGCATAACGAGATAGTAGGACTCGCTAGCTTGCCAAACCTACAGGTGG
MRC#992 secondary PCR #147	CAAGCAGAAGACGGCATAACGAGATTTGTATCTAAAGCTAGCTTGCCAAACCTACAGGTGG
MRC#993 secondary PCR #148	CAAGCAGAAGACGGCATAACGAGATGTGGGGTCCAGCTAGCTTGCCAAACCTACAGGTGG
MRC#994 secondary PCR #149	CAAGCAGAAGACGGCATAACGAGATCTCGAAGCACGCTAGCTTGCCAAACCTACAGGTGG
MRC#995 secondary PCR #150	CAAGCAGAAGACGGCATAACGAGATTACCTAGTTGGCTAGCTTGCCAAACCTACAGGTGG
MRC#996 secondary PCR #151	CAAGCAGAAGACGGCATAACGAGATTTCAATGGGCGCTAGCTTGCCAAACCTACAGGTGG
MRC#997 secondary PCR #152	CAAGCAGAAGACGGCATAACGAGATACGCGGGTTAGCTAGCTTGCCAAACCTACAGGTGG
MRC#998 secondary PCR #153	CAAGCAGAAGACGGCATAACGAGATTCATCGGTAGGCTAGCTTGCCAAACCTACAGGTGG
MRC#999 secondary PCR #154	CAAGCAGAAGACGGCATAACGAGATGACTGTAAGTGCTAGCTTGCCAAACCTACAGGTGG
MRC#1000 secondary PCR #155	CAAGCAGAAGACGGCATAACGAGATTCGATATTGCTAGCTTGCCAAACCTACAGGTGG
MRC#1001 secondary PCR #156	CAAGCAGAAGACGGCATAACGAGATCCGCGGTAATGCTAGCTTGCCAAACCTACAGGTGG
MRC#1002 secondary PCR #157	CAAGCAGAAGACGGCATAACGAGATGAACTCTGGTGCTAGCTTGCCAAACCTACAGGTGG
MRC#1003 secondary PCR #158	CAAGCAGAAGACGGCATAACGAGATGATTCATGCTAGCTTGCCAAACCTACAGGTGG
MRC#1004 secondary PCR #159	CAAGCAGAAGACGGCATAACGAGATATAGAAGTGGGCTAGCTTGCCAAACCTACAGGTGG
MRC#1005 secondary PCR #160	CAAGCAGAAGACGGCATAACGAGATAAACATAGGCGCTAGCTTGCCAAACCTACAGGTGG
MRC#1006 secondary PCR #161	CAAGCAGAAGACGGCATAACGAGATAGCAACAATCGCTAGCTTGCCAAACCTACAGGTGG
MRC#1007 secondary PCR #162	CAAGCAGAAGACGGCATAACGAGATCTTACGAAGCTAGCTTGCCAAACCTACAGGTGG
MRC#1008 secondary PCR #163	CAAGCAGAAGACGGCATAACGAGATAGTATCTGGGGCTAGCTTGCCAAACCTACAGGTGG
MRC#1009 secondary PCR #164	CAAGCAGAAGACGGCATAACGAGATCGTCAAGGGGGCTAGCTTGCCAAACCTACAGGTGG
MRC#1010 secondary PCR #165	CAAGCAGAAGACGGCATAACGAGATCGTAAGTCGAGCTAGCTTGCCAAACCTACAGGTGG

MRC#1011 secondary PCR #166	CAAGCAGAAGACGGCATAACGAGATTACCATGGATGCTAGCTTGCCAAACCTACAGGTGG
MRC#1012 secondary PCR #167	CAAGCAGAAGACGGCATAACGAGATGCGTAGACTTGCTAGCTTGCCAAACCTACAGGTGG
MRC#1013 secondary PCR #168	CAAGCAGAAGACGGCATAACGAGATCATACTGAATGCTAGCTTGCCAAACCTACAGGTGG
MRC#1014 secondary PCR #169	CAAGCAGAAGACGGCATAACGAGATGTATGTCCAAGCTAGCTTGCCAAACCTACAGGTGG
MRC#1015 secondary PCR #170	CAAGCAGAAGACGGCATAACGAGATCACCACTGCTAGCTTGCCAAACCTACAGGTGG
MRC#1016 secondary PCR #171	CAAGCAGAAGACGGCATAACGAGATTGCTGGAGAGCTAGCTTGCCAAACCTACAGGTGG
MRC#1017 secondary PCR #172	CAAGCAGAAGACGGCATAACGAGATATCCGGAGTAGCTAGCTTGCCAAACCTACAGGTGG
MRC#1018 secondary PCR #173	CAAGCAGAAGACGGCATAACGAGATAAGACCTTAGGCTAGCTTGCCAAACCTACAGGTGG
MRC#1019 secondary PCR #174	CAAGCAGAAGACGGCATAACGAGATAGCTTGGTTCGCTAGCTTGCCAAACCTACAGGTGG
MRC#1020 secondary PCR #175	CAAGCAGAAGACGGCATAACGAGATTCAGTATATGCTAGCTTGCCAAACCTACAGGTGG
MRC#1021 secondary PCR #176	CAAGCAGAAGACGGCATAACGAGATAACAAGGGTCGCTAGCTTGCCAAACCTACAGGTGG
MRC#1022 secondary PCR #177	CAAGCAGAAGACGGCATAACGAGATTTATGAAGCCGCTAGCTTGCCAAACCTACAGGTGG
MRC#1023 secondary PCR #178	CAAGCAGAAGACGGCATAACGAGATTTGTATTCTGGCTAGCTTGCCAAACCTACAGGTGG
MRC#1024 secondary PCR #179	CAAGCAGAAGACGGCATAACGAGATTCGTGTGTGCTAGCTTGCCAAACCTACAGGTGG
MRC#1025 secondary PCR #180	CAAGCAGAAGACGGCATAACGAGATTTAGTAAGCTGCTAGCTTGCCAAACCTACAGGTGG
MRC#1026 secondary PCR #181	CAAGCAGAAGACGGCATAACGAGATCCGAACCATGGCTAGCTTGCCAAACCTACAGGTGG
MRC#1027 secondary PCR #182	CAAGCAGAAGACGGCATAACGAGATGGACAGAGTGGCTAGCTTGCCAAACCTACAGGTGG
MRC#1028 secondary PCR #183	CAAGCAGAAGACGGCATAACGAGATTACCAATGCCGCTAGCTTGCCAAACCTACAGGTGG
MRC#1029 secondary PCR #184	CAAGCAGAAGACGGCATAACGAGATAGATGACGTAGCTAGCTTGCCAAACCTACAGGTGG