



Muir, Amy Lianne (2024) *Unravelling clinical heterogeneity in Philadelphia positive Acute Lymphoblastic Leukaemia (Ph+ALL)*. PhD thesis.

<https://theses.gla.ac.uk/84757/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

**UNRAVELLING CLINICAL HETEROGENEITY IN PHILADELPHIA POSITIVE ACUTE
LYMPHOBLASTIC LEUKAEMIA (PH+ALL)**

AMY LIANNE MUIR

BSc (hons)

**Submitted in fulfilment of the requirements for the degree of Doctor of
Philosophy**

October 2024

Section of Experimental Haematology

School of Cancer Sciences

College of Medical, Veterinary and Life Sciences

University of Glasgow

This thesis is dedicated to my grandparents (John & Mary Maloy and Cathy & William Muir) and my parents (Liam and Mary-Rose Muir).

“The way I see it, if you want the rainbow, you gotta put up with the rain” – Dolly Parton

Abstract

The understanding of the haematopoietic development process has increased exponentially over the past few decades through the advancement in technology in the field of haematology and immunology. By doing so, the classical two-tier model of haematopoiesis has been enhanced to acknowledge the numerous developmental stages, cell subtypes, transcriptional alterations, transcription factors and surface marker expression required for early lineage development. The result is a repertoire of immune cells with wide-ranging functions allowing for the rapid response to antigens and maintenance of long-term immunological memory. Hence, immune malfunction can have deleterious effects, as evidenced in leukaemia.

Development from early stem and progenitor cells to lineage-fixed effector cells has been studied in the context of leukaemia and has resulted in a deeper understanding of the biology of the disease and an improved outcome for patients. In both Philadelphia positive acute lymphoblastic leukaemia (Ph+ALL) and chronic myeloid leukaemia (CML), the disease driver is the constitutive activation of kinase BCR::ABL1. This kinase results from a translocation fusion event whereby the long arms of chromosomes 9 and 22 break and concurrently fuse together, producing the truncated Philadelphia chromosome. This event brings together the ABL gene on chromosome 9 and the BCR gene on chromosome 22 to form the proto-oncogene *BCR::ABL1* which is found in the Philadelphia chromosome. This gene encodes for a protein of the same name which functions as the aforementioned constitutively active tyrosine kinase. BCR::ABL1 interacts with cell cycle and apoptotic pathways producing cells which rapidly and uncontrollably proliferate and which do not respond to pro-apoptotic signals. In CML, this fusion event occurs at the apex of blood cell development in haematopoietic stem cells (HSCs), resulting in the potential for BCR::ABL1 activity to affect cells in both lymphoid and myeloid lineages. In Ph+ALL however, the cell of leukaemic origin is under question. The orthodoxy is that the t(9;22) fusion event occurs in a lymphoid progenitor as BCR::ABL1 activity is usually exclusively observed in the lymphoid lineage by the overproduction of lymphoblasts. However, recent observations by Hovorkova et al (2017) has prompted further discussion into the cell of BCR::ABL1 origin in Ph+ALL.

A subgroup of paediatric Ph+ALL patients were observed by Hovorkova et al (2017) as having disease features similar to CML and a poor response to standard Ph+ALL therapeutic strategies. These patients were detected during investigations into minimal residual disease (MRD) monitoring where rearrangement of Ig/TCR genes (a hallmark of the latter phases of B cell development) were compared to *BCR::ABL1* gene copy number. What was observed was a subgroup of patients who remained *BCR::ABL1* positive despite being Ig/TCR negative, in essence displaying an eradication of Ig/TCR positive lymphoid cells but with a maintenance of a significant number of *BCR::ABL1* positive cells. This contrasts with the current understanding of Ph+ALL where lymphoid-directed therapy would eradicate both blast cells and the leukaemic cell of origin (lymphoid progenitor) displaying concurrent negativity in both MRD methods. These patients also displayed multilineage disease involvement with large populations of myeloid cells at diagnosis, atypical in Ph+ALL where haematopoiesis is skewed toward production of lymphoid cells. The resultant theory is that this subgroup results from the *BCR::ABL1* fusion event occurring prior to lineage commitment, either in a multipotent progenitor (MPP) or an HSC, just as in CML. Hence, this Ph+ALL subtype was designated as 'CML-like Ph+ALL' by Hovorkova et al (2017).

CML exemplifies the importance of accurately targeting and eradicating *BCR::ABL1*+ HSCs for the prevention of disease progression or relapse. The treatment of CML includes the direct targeting of *BCR::ABL1* activity through a class of drugs called tyrosine kinase inhibitors (TKI) in addition to chemotherapy however, Ph+ HSCs in CML (leukaemic stem cells (LSC)) are able to evade eradication by drugs through a quiescent state with little proliferative activity. Therefore, haematopoietic stem cell transplantation (HSCT) may be utilised to fully eradicate all immune cells and replace them with healthy donor cells, thus preventing disease progression or relapse. This multi-agent approach and TKI development has allowed CML survival rates to increase from almost certain death to survival of the vast majority of patients. The findings from Hovorkova et al (2017) display that the standard Ph+ALL treatment is inadequate for eradication of *BCR::ABL1*+ cells in patients with CML-like Ph+ALL, hence patients with this subtype risk relapse or disease progression similar to CML.

To determine the leukaemic origin of CML-like Ph+ALL, we utilised paediatric Ph+ALL bone marrow samples harvested at diagnosis and post-induction therapy (PI) and CML samples in the acute lymphoid blast crisis phase (CML-LBC). The current definition of CML-like Ph+ALL is

the discordance between *BCR::ABL1* copy number and Ig/TCR gene rearrangement MRD methods, however such information was not available for all samples used in this project and therefore, CML-like Ph+ALL samples could not be identified from the outset as specified by Hovorkova et al (2017). However, examination of Ph+ALL patient samples by flow cytometry displayed a subset of samples with large myeloid populations at diagnosis, concordant with findings from Hovorkova et al (2017). Additionally, this subset of patients displayed large numbers of HSCs and few lymphoid progenitors at diagnosis, atypical to what would be expected in a standard Ph+ALL sample. Sorted progenitor cells (HSCs and MPPs) with *BCR::ABL1* fusion measured by FISH (fluorescence *in situ* hybridisation) were detected in the majority of samples investigated across the Ph+ALL cohort, suggesting that *BCR::ABL1* fusion in early stem and progenitor cells alone may be inadequate for establishment of the CML-like Ph+ALL subtype and perhaps, downstream transcriptional modifications determine multilineage involvement. RNAseq was used to investigate gene expression in the Ph+ALL and CML samples with findings displaying a subgroup of Ph+ALL samples which clustered distal to the rest of the cohort. Identification of differentially expressed genes (DEGs) in these samples showed an upregulation of genes involved in myeloid leukaemia and LSC activity. All samples in this cluster had detectable Ph+ HSC/MPPs and a higher incidence of disease progression, relapse or death compared to the rest of the cohort, disease outcomes that support the findings from Hovorkova et al (2017) which display poor response to standard Ph+ALL treatment in the CML-like Ph+ALL subtype.

Our findings postulate that CML-like Ph+ALL can be characterised by atypical haematopoietic dynamics, t(9;22) occurrence in HSC/MPPs, expression of CML-associated genes and poor response to standard Ph+ALL treatment.

Acknowledgements

I considered keeping the same acknowledgement section from the first version of my thesis, but the corrections process has given me a lot of time to think and grow, plus so many people have helped me in a variety of ways during this time and I really want to thank them.

The PhD process is never easy, but the corrections process has been a different beast altogether. Balancing thesis editing with full time employment, long commutes, ADHD, mental health issues and the propensity to burn the candle at both ends has at times been incredibly difficult but having an amazing and incredibly supportive community around me (and a lot of personal growth) has allowed me to see the PhD process through.

My PI Dr Gillian Horne is, without hyperbole, the smartest person I've ever met. Not only is her academic knowledge encyclopaedic, but her emotional intelligence is genius-level. Gillian has been there for me every step of the process, through tears, mindmaps and endless questions. Gillian truly meets me where I am, she understands that my brain sometimes works a bit 'differently' and has managed to learn exactly how to motivate me, interpret my ideas and support me in my development as a scientist- even understanding how to help when I don't know how to articulate my needs. Gillian and I are both aware that without her, I would've dropped out years ago. How she manages to be the most fiercely intelligent scientist and medic, and yet be the coolest and funniest person, I'll never know. She somehow puts up with my eccentricities and genuinely cares about everyone she meets. The above reasons (and more) are why I've refused to leave her lab and stayed on as a postdoc. Thank you Gillian, you're truly one of a kind.

I'm truly lucky to have such supportive parents who have been a rock not only during the PhD, but throughout my entire life. They've always encouraged me to pursue my passions, never forcing me to do things I don't want to (not that anyone could!). They've always done the utmost to provide for me and show me the value in hard work and honesty. Thank you for providing me with the tools to make opportunities for myself, for nurturing my tenacity, celebrating my uniqueness and always being encouraging with whatever I pursue. Most importantly, thank you for giving me such a good taste in music.

A big part of why my PhD has been an overall good experience has been everyone at the Paul O'Gorman Leukaemia Research Centre, thank you everyone. Alan has cheered me up countless times, I thoroughly enjoy his stories and chatting in our native Ayrshire dialect. Without Jen's advice and expertise, there would be no flow cytometry data, thank you so much Jen. Angela, administrator extraordinaire and such a warm-hearted person. Helen for so much advice, understanding and support for neurodivergent scientists. Thank you also to the rest of my supervisory team, professors Mhairi Copland and Chris Halsey, both fantastic examples of successful women in academic research.

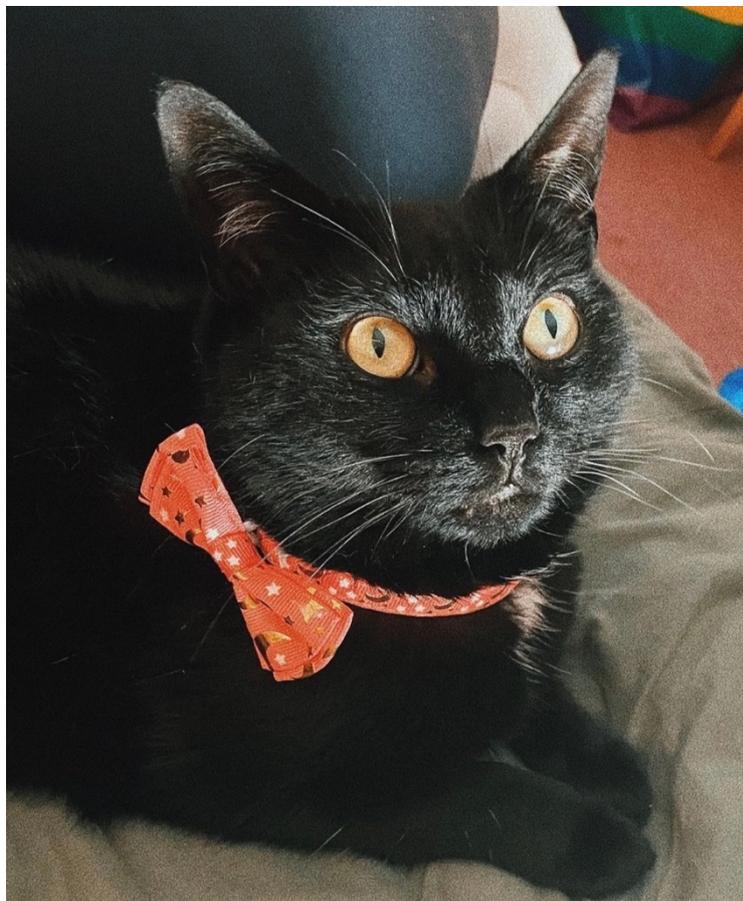
Thank you to my friends for being so supportive. Thank you, Danielle, Andy K and Chantalle, for being so patient and understanding with me, I'm so grateful to have amazing friends who I can pick up conversations with such ease, even if it's been months since our last hangout. Laura, I love that our conversations branch off like the London underground map, that we'll watch Moulin Rouge twice back-to-back (once for plot and once to sing along) and that we always end up cry laughing whenever we hang out. Shaun, I love that our friendship encompasses both silliness and depth, you're truly a fantastic scientist and a wonderful

dancer. Jodie, thank you for the encouragement and support. Rachel, I love our science chats and lab trips around Glasgow. Vivan, thank you for always being my cheerleader, if the feeling of sunshine on your face and the taste of mango was a person, it'd be you. Isla, we've been best friends in every timeline. I'm so proud to be your friend. I'm not sure if you realise how amazing you truly are but maybe now that it's down in print, you might believe it. Eden, my blood brother, thank you for always understanding me and meeting me where I am. You make me feel seen, you encourage me to be myself and I'm glad we're the same person in different fonts.

I must acknowledge the other things that have helped me through the years including the ladies in my local library (the Dick Institute), Spotify Premium and the Murder Most Irish podcast (thank you for keeping me company during the long days in the dark room!).

I also want to thank myself for putting all the hard work in. It's been a tough few years and a few occasions where I didn't think I could finish this, but I always managed to push through. Since starting my PhD in 2019 I've done so much personal growth, and this process has shown me how much inner strength I have. I'm grateful to be the person I am today, and I wouldn't change the last few years for anything (the good and the bad). Proud of you Amy.

Most importantly, I want to thank my cat, Lemon. The day I bought her (from Gumtree for £20) as a kitten and rescued her from a nasty situation, she's been my constant companion. The unconditional love and comfort that comes from animals can never be understated. Unfortunately, she can't read any of this because she's a cat and I'm not sure how much she'd care if she could. Still, love you Lem.



Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Amy Lianne Muir

Table of contents

	Page number
Abstract	iii
Acknowledgements	vi
Declaration	viii
Table of contents	ix
List of figures	xiv
List of tables	xix
List of abbreviations	xxii
Chapter 1. Introduction.	1
1.1 Introduction: Acute Lymphoblastic Leukaemia (ALL)	1
1.2.1 Introduction: ALL pathophysiology – molecular genetics and cytogenetic alterations	3
1.2.2 Introduction: ALL pathophysiology- phenotypic cellular compartments	7
1.3 Introduction: Prognostic factors and measurable residual disease	14
1.4 Introduction: Established and developing treatment strategies.	18
1.5 Introduction: Relapse/refractory disease.	22
1.6 Introduction: BCR::ABL1 fusion and cell of origin in Ph+ haematological malignancies (Ph+ALL and CML)	23
1.7 Introduction: Targeting of BCR::ABL1 by tyrosine kinase inhibitors (TKIs).	28
1.8 Introduction: The identification of a novel Ph+ALL subtype with CML-like features.	33
1.9.1 Introduction: Characterisation of CML-like Ph+ALL patients; what is known currently and how it may align with current Ph+ALL risk stratification.	38
1.9.2 Introduction: Aim1: Determine if haematopoietic dynamics differ within Ph+ALL and if differences in cell population sizes can identify distinct subgroups.	41
1.9.3 Introduction: Aim 2: Identify where during haematopoietic development BCR::ABL1 fusion occurs.	43
1.9.4 Introduction: Aim 3: Examine the heterogeneity of transcriptional profiles within Ph+ALL and determine if subtypes can be identified by differential gene expression.	44
Chapter 2. Materials and methods.	46
2.1 Materials and methods: Reagents and equipment	46

2.2 Materials and methods: Solutions	48
2.3 Materials and methods: Antibodies	50
2.4 Materials and methods: Statistical analysis	51
2.5.1 Materials and methods: Software and coding (R Studio)	52
2.5.2 Materials and methods: Software and coding (GEO2R)	56
2.5.3 Materials and methods: RNAseq (RNA sequencing) bioinformatic analysis	57
2.6.1 Materials and methods: Primary patient sample ethical considerations	60
2.6.2 Materials and methods: Primary bone marrow mononuclear cell isolation and culture - healthy patient trephine controls	61
2.6.3 Materials and methods: Primary bone marrow mononuclear cell isolation and culture - paediatric Ph+ALL bone marrow aspirate.	62
2.6.4 Materials and methods: Primary BMNC thawing (Ph+ALL & CML patient samples)	63
2.6.5 Materials and methods: Cell counting	64
2.7 Materials and methods: FACS (fluorescence-assisted cell sorting) isolation of haematopoietic progenitor cells from bone marrow samples (CML, Ph+ALL and healthy controls)	65
2.8 Materials and methods: FISH (fluorescence <i>in situ</i> hybridisation)	67
2.9.1 Materials and methods: Preservation of RNA using RNAprotect (Qiagen)	69
2.9.2 Materials and methods: RNA extraction from primary BMNC samples	70
2.10 Materials and methods: Sequencing of B cell receptor VDJ gene segments	71
Chapter 3. Results I. Utilisation of publicly available datasets to investigate the presence of Ph+ALL subtypes	73
3.1.1 Introduction: analysis of gene expression for the identification of novel Ph+ALL subtypes using publicly available datasets.	73
3.1.2 Aims.	77
3.2 Results: BCR::ABL1 promotes aberrant expression of oncogenic genes, in addition to Ph+ALL-specific metabolic activity and lymphoid differentiation arrest	78
3.3 Results: Profiling of CML and healthy Lin-progenitor cells displayed a CML-specific transcriptional signature and highlighted leukaemic-associated genes expressed similarly in CML and Ph+ALL	89

3.4 Results: HSCs in CML-CP displayed some hallmarks of normal stem cell development, shared CML-specific transcriptional hallmarks with Lin-CML samples from previous datasets and displayed expectable intra-sample variation	98
3.5: Results: Paediatric Ph+ALL patients designated as being 'good risk' based on blast cell count were transcriptionally similar to CML samples from previous datasets	105
3.6 Results: Ph+ALL patient with hyperdiploidy shared gene expression profiles with CML and 'good risk' Ph+ALL samples from previous datasets, suggesting a consistent outlying transcriptional phenotype within the Ph+ALL cohort.	112
3.7 Discussion.	120

Chapter 4. Results II. Unravelling Ph+ALL heterogeneity with FACS and FISH.

Chapter 4. Results II. Unravelling Ph+ALL heterogeneity with FACS and FISH.	124
4.1.1 Introduction: The cell of origin in Ph+ALL remains elusive.	124
4.1.2 Aims.	128
4.2.1 Results: Fluorescence-assisted cell sorting revealed a potential subgroup of Ph+ALL patients with atypical haematopoietic dynamics.	129
4.2.2 Results: HSC and MPP populations can be identified using FACS.	136
4.2.3 Results: Two Ph+ALL patients had unexpectedly large CMP populations at diagnosis.	142
4.2.4 Results: Two patients had unexpectedly large GMP populations at diagnosis.	145
4.2.5 Results: Three patients had atypically small CLP populations at diagnosis.	147
4.2.6 Results: Atypical haematopoietic population sizes were observed in a number of diagnostic and PI Ph+ALL samples	149
4.3.1 Results: BCR::ABL1 fusion was detectable by FISH in a subgroup of Ph+ALL stem and progenitor cells.	151
4.4.1 Results: Patients with atypical haematopoiesis dynamics and BCR::ABL1 fusion in stem and progenitor cells had poor clinical outcomes.	156
4.5 Discussion.	164

Chapter 5. Results III. Transcriptional investigation of a heterogeneous Ph+ALL cohort.

Chapter 5. Results III. Transcriptional investigation of a heterogeneous Ph+ALL cohort.	170
5.1.1 Introduction: Background and clinical relevance.	170
5.1.2 Aims.	173
5.1.3 Sample selection for RNAseq.	174

5.2.1 Results: CML-LBC vs Ph+ALL principle component analysis (PCA) identifies transcriptionally and clinically distinct Ph+ALL subgroups.	176
5.2.2 Results III: Gene expression profiling differentiates between bulk CML-LBC and Ph+ALL.	179
5.2.3 Processing DEGs from CML-LBC vs Ph+ALL comparison for gene set enrichment analysis (GSEA).	180
5.2.4 Results: GSEA highlighted disease-specific gene ontology terms for DEGs identified in the CML-LBC vs Ph+ALL comparison.	182
5.2.5 Results: Genes involved in circulatory system development and maintenance were differentially regulated between CML-LBC and Ph+ALL samples.	183
5.2.6 Results: Tissue development genes were differentially regulated between CML-LBC and Ph+ALL samples.	185
5.2.7 Results: Inflammatory immune response genes were upregulated in CML-LBC vs Ph+ALL bulk samples.	187
5.2.8 Results: Regulation of mRNA encoding and gene targeting genes were upregulated in CML-LBC vs Ph+ALL bulk samples.	189
5.2.9 Results: Differential gene expression analysis displayed disease-specific transcriptional profiles in CML-LBC and Ph+ALL samples.	190
5.3.1 Results: Ph+ALL samples with detectable BCR::ABL1 in stem and progenitor cells had a different transcriptional profiles to Ph+ALL samples with no BCR::ABL1 at the stem and progenitor level.	192
5.3.2 Results: Transcriptional heterogeneity exists between Ph+ALL samples with detectable BCR::ABL1 at the stem and progenitor level and Ph+ALL samples without.	194
5.3.3 Results: CNS developmental and functional genes were upregulated in Ph+ALL samples with detectable BCR::ABL1 at the stem and progenitor level.	198
5.3.4 Results: An LSC-like gene expression profile was detected in Ph+ALL samples with Ph+ stem and progenitor cells.	199
5.4.1 Results: GSEA displayed the differential regulation of few GO terms between Cluster 1 and Cluster 3.	203
5.4.2 Results: Cluster 1 and Cluster 3 shared transcriptional similarity but displayed differential regulation in microenvironmental response genes.	205
5.5.1 Results: Cluster 2 and Cluster 3 displayed markedly different overall transcriptional profiles.	206
5.5.2 Results: Cell adhesion genes were differentially regulated between Ph+ALL Clusters 2 and 3.	208

5.5.3 Results: Immune response genes were differentially regulated between Cluster 2 and Cluster 3.	210
5.5.4 Results: Cluster 2 and Cluster 3 displayed difference in the regulation of early developmental genes.	211
5.6 Results: RNAseq analysis of CML-LBC and Ph+ALL samples identified a transcriptionally and clinically distinct subgroup of Ph+ALL samples.	212
5.7 Conclusion: RNAseq analysis of CML-LBC and Ph+ALL samples identified a transcriptionally and clinically distinct subgroup of Ph+ALL samples.	214
5.7.2 Caveats and limitations.	219
Chapter 6. Concluding discussion.	220
Chapter 7. Future directions.	225
References	232

List of figures

Chapter 1. Introduction.	Page number
Figure 1.1: Surface markers of early B cell development.	8
Figure 1.2: B Cell development in bone marrow to periphery.	11
Figure 1.3: B cell lineage commitment by PAX5.	13
Figure 1.4: ALL risk stratification and treatment schema.	19
Figure 1.5: Representation of BCR::ABL1 fusion and the formation of the truncated Philadelphia chromosome.	23
Figure 1.6: Fusion transcripts arising from p190, p210 and p230 BCR::ABL1 breakpoint regions.	24
Figure 1.7: Schematic representation of BCR and ABL1 gene structures and protein domains of BCR::ABL1 fusion proteins.	25
Figure 1.8: BCR::ABL1 signalling network.	26
Figure 1.9: Schematic of BCR::ABL1 point mutation according to position.	30
Figure 1.10: Monitoring childhood ALL using BCR::ABL1 genomic breakpoints and Ig/TCR gene rearrangement identifies a discordant subgroup with CML-like biology.	34
Figure 1.11: BCR::ABL1 fusion during early stages of haematopoiesis in three disease states.	35
Figure 1.12: Updated model of haematopoiesis at steady state.	37
Figure 1.13: Sorting strategy (standard Ph+ALL).	42
Chapter 2. Materials and methods.	
Figure 2.5.1: RNAseq data analysis (R Studio), setup.	53
Figure 2.5.2: RNAseq data analysis (R Studio), define dataset.	53
Figure 2.5.3: RNAseq data analysis (R Studio), differential expression (DE).	54
Figure 2.5.4: RNAseq data analysis (R Studio), principal component analysis (PCA).	54
Figure 2.5.5: RNAseq data analysis (R Studio), differential expression analysis.	55
Figure 2.5.6: RNAseq data analysis (R Studio), Volcano plot.	55
Figure 2.5.7: RNAseq data analysis (R Studio), data export.	55
Figure 2.6.1: Schematic representation of density centrifugation separation of mononuclear cells from whole bone marrow samples.	61
Figure 2.6.2: Haemocytometer cell counting.	64

Figure 2.7.1: Fluorescence assisted cell sorting (FACS).	65
Figure 2.8.1: Schematic representation of FISH.	67
Figure 2.11: Schematic representation of ClonoSEQ Ig/TCR MRD monitoring.	72

Chapter 3. Results I. Utilisation of publicly available datasets to investigate the presence of Ph+ALL subtypes

Figure 3.1: MRD score at TP1 in the CML-like and typical Ph+ALL subgroups.	73
Figure 3.2.1: Gene expression (1) PCA plot and (2) heatmap for Ph+ cell lines (BV-173, NALM-1, SIP-B15 and TOM1), dataset GDS4175 (GSE23743). (3) DEG volcano plot Ph+ALL vs CML cell lines.	79
Figure 3.2.2: Functional enrichment analysis of genes differentially regulated in Ph+ALL cell lines (SUP-B15 and TOM1), dataset GDS4175.	84
Figure 3.2.3: Gene ontology analysis for enrichment the top 50 upregulated DEGs in Ph+cell lines (SIP-B15 and TOM1) vs CML cell lines (BV173 and NALM1), dataset GDS4175 (GSE23743).	88
Figure 3.3.1: Haematopoietic population surface marker expression for FACS.	89
Figure 3.3.2: Gene expression PCA plot (1) and heatmap (2) for for CML and healthy controls (Lin+CD34+, Lin-CD34+, Lin-CD34-), dataset GSE11675. (3) DEG volcano plot vs control cell populations.	91
Figure 3.3.3: Functional enrichment analysis of genes differentially regulated in primary Lin-CD34+ CML cells, dataset GSE11675.	94
Figure 3.3.4: Gene ontology enrichment analysis for (1) top 50 upregulated and (2) bottom 50 downregulated DEGs in CML vs healthy control cells, dataset GSE11675	96
Figure 3.4.1: Gene expression PCA plot (1) and heatmap (2) for CML and healthy control HSC cells, dataset GSE11889. (3) DEG volcano plot CML HSC vs control HSC.	99
Figure 3.4.2: Functional enrichment analysis of genes downregulated in CML HSCs, dataset GSE11889.	103
Figure 3.4.3: Gene ontology enrichment analysis for (1) top 50 upregulated and (2) bottom 50 downregulated DEGs in CML HSCs vs healthy control HSCs, dataset GSE11889.	104
Figure 3.5.1: Gene expression PCA plot (1) and heatmap (2) for untreated (Day 0) primary Ph+ALL samples, 'good' and 'poor' clinical outcomes shown, dataset GSE39335.(3) DEG volcano plot Day 0 Ph+ALL good risk vs poor risk.	106

Figure 3.5.2: Functional enrichment analysis of genes differentially regulated in untreated primary Ph+ALL samples with good or poor clinical outcomes, dataset GSE39335.	110
Figure 3.5.3: Gene ontology enrichment analysis for top 50 upregulated DEGs in Day 0 Ph+ALL good risk vs poor risk, dataset GSE39335.	111
Figure 3.6.1: Gene expression PCA plot (1) and heatmap (2) for paediatric Ph+ALL bone marrow, dataset GSE13425. (3) DEG volcano plot standard Ph+ALL vs Ph+ALL with hyperdiploidy (Patient 142).	113
Figure 3.6.2: Functional enrichment analysis of genes upregulated in primary paediatric Ph+ALL bone marrow samples, dataset GSE13425.	118
Figure 3.6.3: Gene ontology enrichment analysis for top 50 DEGs upregulated and (2) bottom 50 downregulated in standard Ph+ALL vs Ph+ALL with hyperdiploidy (Patient 142), dataset GSE13425.	119
Chapter 4. Results II. Unravelling Ph+ALL heterogeneity with FACS and FISH.	
Figure 4.1.1: Model of haematopoiesis at steady state.	124
Figure 4.1.2: Haematopoietic population surface marker expression for FACS.	125
Figure 4.1.3: Representation of BCR::ABL1 fusion during early stages of haematopoiesis in three Ph+ALL disease states.	127
Figure 4.2.1: Antibody cocktails used in FACS (1) and applied sorting strategy (Ph+ALL diagnostic, 5yo M).	129
Figure 4.2.2: Fluorescence minus one (FMO) control (surface marker and fluorophores indicated).	130
Figure 4.2.3: Post induction sample timepoint (days).	132
Figure 4.2.4: Distribution of progenitor cells (HSC, MPP, CMP, GMP and CLP) in PI Ph+ALL samples.	133
Figure 4.2.5.1: Proportion progenitor cells (HSC, MPP, CMP, GMP and CLP) in diagnostic Ph+ALL samples.	134
Figure 4.2.5.2: Proportion of progenitor cells (HSC, MPP, CMP, GMP and CLP) in PI Ph+ALL samples.	135
Figure 4.2.6: HSC population in matched Ph+ALL diagnostic and PI samples.	137
Figure 4.2.7: MPP population size in matched Ph+ALL diagnostic and PI samples.	141
Figure 4.2.8: CMP population size in matched Ph+ALL diagnostic and PI samples.	144

Figure 4.2.9: GMP population size in matched Ph+ALL diagnostic and PI samples.	146
Figure 4.2.10: CLP population size in matched Ph+ALL diagnostic and PI samples.	148
Figure 4.3.1: Representation of BCR::ABL1 detection by FISH (fluorescence <i>in situ</i> hybridisation).	151
Figure 4.3.2: Representative FISH nuclei for Ph+ leukaemias.	153
Figure 4.3.3: BCR::ABL1 positivity in haematopoietic populations (FISH), percentage of Ph+ progenitor cells.	154
Figure 4.3.4: BCR::ABL1 presence in early and late haematopoietic populations.	155
Figure 4.4.1: BCR::ABL1 positivity in patients with high-risk MRD.	157
Figure 4.4.2: BCR::ABL1 positivity in patients with cytogenetic abnormalities in addition to t(9;22).	159
Figure 4.4.3: BCR::ABL1 positivity in patients with high white cell counts at diagnosis.	160
Figure 4.4.4: BCR::ABL1 positivity in patients with reported incidences of relapse.	161
Figure 4.4.6: BCR::ABL1 positivity in patients with aberrant haematopoietic diagnostic populations measured by FACS.	162
Chapter 5. Results III. Transcriptional investigation of a heterogeneous Ph+ALL cohort.	
Figure 5.2.1: Principal component analysis (PCA) plot, CML-LBC vs Ph+ALL.	177
Figure 5.2.2: Kaplan Meier survival analysis (Ph+ALL).	178
Figure 5.2.3: Differential expression volcano plot CML-LBC vs Ph+ALL. P-value cutoff of 0.05.	179
Figure 5.2.4: Gene ontology enrichment plot (CML-LBC vs Ph+ALL)	182
Figure 5.2.5: Differentially expressed genes associated with circulatory system development with accompanying GSEA enrichment plot.	184
Figure 5.2.6: GSEA enrichment plots representing tissue development gene sets for DEGs upregulated in CML-LBC.	186
Figure 5.2.7: GSEA enrichment plots representing inflammatory immune gene sets for DEGs upregulated in CML-LBC.	188
Figure 5.2.8: GSEA enrichment plots representing mRNA encoding and gene targeting gene sets for DEGs upregulated in CML-LBC.	189
Figure 5.2.9: Gene ontology enrichment plot for genes upregulated in CML-LBC vs Ph+ALL.	191

Figure 5.3.1: Haematopoietic population surface marker expression for FACS.	192
Figure 5.3.2: Principal component analysis plot, CML-LBC (N=2) vs Ph+ALL (N=11), clusters annotated.	194
Figure 5.3.3: Principal component analysis plot, BCRABLneg (N=5) vs BCRABLpos (N=6), clusters annotated.	195
Figure 5.3.4: Differential expression volcano plot, BCRABLneg (N=5) vs BCRABLpos (N=6), clusters annotated.	196
Figure 5.3.5: Gene ontology enrichment plot (BCR::ABL1 positive HSCs/MPPs (Ph+ALL) vs BCR::ABL1 negative HSCs/MPPs (Ph+ALL)).	197
Figure 5.3.6: GSEA enrichment plots representing CNS developmental and functional gene sets for DEGs upregulated in BCRABLpos Ph+ALL.	198
Figure 5.3.7: GSEA enrichment plots representing LSC-related gene sets for DEGs upregulated in BCRABLpos Ph+ALL.	200
Figure 5.3.8: Gene ontology enrichment plot for genes upregulated in BCR::ABL1 positive HSCs/MPPs (Ph+ALL) vs BCR::ABL1 negative HSCs/MPPs (Ph+ALL).	201
Figure 5.4.1: Principal component analysis plot, Cluster 3 (N=3) vs Cluster 1 (N=2).	203
Figure 5.4.2: Differential expression analysis volcano plot, Cluster 3 (N=3) vs Cluster 1 (N=2), clusters annotated.	204
Figure 5.4.3: GSEA enrichment plots representing microenvironmental gene sets for DEGs upregulated in Cluster 3/	205
Figure 5.5.1: Principal component analysis plot, Cluster 3 (N=3) vs Cluster 2 (N=8).	206
Figure 5.5.2: Differential expression volcano plot, Cluster 3 (N=3) vs Cluster 2 (N=8), clusters annotated.	207
Figure 5.5.3: GSEA enrichment plots representing cell adhesion gene sets for DEGs upregulated in Cluster 3.	209
Figure 5.5.4: GSEA enrichment plots representing immune response gene sets for DEGs upregulated in Cluster 3.	210
Figure 5.5.5: GSEA enrichment plots representing early developmental gene sets for DEGs upregulated in Cluster 3.	211

Chapter 6. Concluding discussion.

Chapter 7. Future directions.

Figure 7.1: Schematic overview of the organisation and rearrangement of Ig and TCR genes.	228
---	-----

List of tables

Chapter 1. Introduction.	Page number
Table 1.1: ALL subtypes and characteristics.	4
Table 1.2: Characterisation of peripheral B cell subsets (human).	12
Table 1.3: Post-diagnosis, pre-therapeutic factors associated with ALL outcomes.	15
Table 1.4: Point mutation and recommended action for TKI choice.	30
 Chapter 2. Materials and methods.	
Table 2.1.1: List of reagents used for laboratory-based research.	46
Table 2.1.2: List of materials used for laboratory-based research.	47
Table 2.1.3: List of technical equipment used for laboratory-based research.	47
Table 2.2.1: Immunofluorescence reagents.	48
Table 2.2.2: FISH fixative.	48
Table 2.2.3: FISH KCl wash buffer	48
Table 2.2.4: Stock saline sodium citrate (SSC) wash buffer.	48
Table 2.2.5: FISH SSC wash buffer 1.	48
Table 2.2.6: FISH SSC wash buffer 2.	49
Table 2.2.7: Thawing solution for primary samples (DAMP).	49
Table 2.2.8: Washing solution for FACS.	49
Table 2.2.9: Human physiological growth factors for growth of primary BMNC samples.	49
Table 2.3.1: Cell surface marker expression for early haematopoietic stem and progenitor cells.	50
Table 2.3.2: Fluorochromes and wavelengths of antibodies used in FACS isolation of stem and progenitor cells from primary BMNC samples.	50

Table 2.5.1: List of software used for data generation, collection and analysis. 52

Table 2.5.2: R packages and libraries used for data analysis and visualisation. 52

Chapter 3. Results I. Utilisation of publicly available datasets to investigate the presence of Ph+ALL subtypes

Table 3.1: Gene target list. 76

Table 3.2.1: Ph+ cell line background information, dataset GDS4175. 78

Table 3.2.2: DEGs in Ph+ALL cell lines (SUP-B15 and TOM1) vs CML cell lines (BV173 and NALM-1), dataset GDS4175 (GSE23743). 82

Table 3.2.3: Bottom 50 downregulated DEGs in Ph+ cell lines (SUP-B15 and TOM1) vs CML cell lines (BV173 and NALM-1), dataset GDS4175 (GSE23743). 83

Table 3.3.1: Top 50 upregulated DEGs in CML HSCs vs healthy control cells (Lin+CD34+, Lin-CD34+, Lin-CD34-), dataset GSE11675. 92

Table 3.3.2: Bottom 50 downregulated DEGs in CML HSCs vs healthy control cells (Lin+CD34+, Lin-CD34+, Lin-CD34-), dataset GSE11675. 93

Table 3.4.1: Top 50 upregulated DEGs in CML HSCs vs healthy control HSCs, dataset GSE11889. 101

Table 3.4.2: Bottom 50 downregulated DEGs in CML HSCs vs healthy control HSCs, dataset GSE11889. 102

Table 3.5.1: Top 50 upregulated DEGs in Day 0 Ph+ALL good risk vs poor risk, dataset GSE39335. 108

Table 3.5.2: Bottom 50 downregulated DEGs in Day 0 Ph+ALL good risk vs poor risk, dataset GSE39335. 109

Table 3.6.1: Top 50 DEGs upregulated in standard Ph+ALL vs Ph+ALL with hyperdiploidy (Patient 142), dataset GSE13425. 116

Table 3.6.2: Bottom 50 DEGs downregulated in standard Ph+ALL vs Ph+ALL with hyperdiploidy (Patient 142), dataset GSE13425. 117

Table 3.7: Proposed CML-like gene expression pattern. 122

Chapter 4. Results II. Unravelling Ph+ALL heterogeneity with FACS and FISH.

Table 4.2.1: Sorted progenitor population size (percentage of live cells) in diagnostic and PI samples.	150
Table 4.4.1: Patient risk status based on clinical information provided by CellBank.	156
Table 4.4.2: Patients with high-risk MRD.	157
Table 4.4.3: Patients with cytogenetic abnormalities in addition to t(9;22).	159
Table 4.4.4: Patients with high white cell counts at diagnosis.	160
Table 4.4.5: Patients with reported incidences of relapse or death.	161
Table 4.4.6: 2way ANOVA of FISH for overall multivariate high risk vs low risk clinical factors.	163

Chapter 5. Results III. Transcriptional investigation of a heterogeneous Ph+ALL cohort.

Table 5.1.1: Proposed CML-like gene expression pattern.	172
Table 5.1.2: RNAseq sample quality information.	174
Table 5.1.3: RNAseq sample patient clinical information.	175
Table 5.6.1: Expression of proposed CML-like expression pattern in RNAseq comparisons.	212

Chapter 6. Concluding discussion.

Chapter 7. Future directions.

List of abbreviations

°C	Celsius
7AAD	7 amino actinomycin D
ABL	Abelson murine leukaemia viral oncogene homolog
ACE	Angiotensin I Converting Enzyme
ADAM	a disintegrin and metalloproteinase
ALL	acute lymphoblastic leukaemia
AIF1L	Allograft Inflammatory Factor 1 Like
AML	acute myeloid leukaemia
ANXA1	Annexin A1
AP	accelerated phase
ATCC	American Type Culture Collection BCR breakpoint cluster region protein
BCR::ABL1	breakpoint cluster region Ableson fusion gene
BM	bone marrow
BMM	bone marrow microenvironment
BMNC	bone marrow mononuclear cells
BMP	bone morphogenic protein BP blast phase
BMP4	bone morphogenic protein 4
BMPR1B	Bone morphogenetic protein receptor type-1B
BPI	Bactericidal Permeability Increasing Protein
CCyR	complete cytogenetic response
CD	cluster of differentiation
CD34-	CD34 negative
CD34+	CD34 positive
CFC	colony forming cells
CFSE	carboxyfluorescein diacetate succinimidyle ester
CH25H	Cholesterol 25-Hydroxylase
CHR	complete haematological response
CK1	casein kinase 1
CLL	chronic lymphocytic leukaemia
CLP	common lymphoid progenitors
CML	chronic myeloid leukaemia
CML-like	chronic myeloid leukaemia-like Philadelphia
Ph+ALL	chromosome positive acute lymphoblastic leukaem
CMP	common myeloid progenitors
CMR	complete molecular response
CNS	central nervous system
CO2	carbon dioxide

COL6A3	Collagen alpha-3(VI)
CP	chronic phase
CSC	cancer stem cells
CSL CBF-1-	Suppressor of Hairless/lag1
CXCL12	stromal cell-derived factor 1
CYYR1	Cysteine and tyrosine-rich 1
DADI	The DAsatinib Discontinuation study
DAPI	4',6-diamidino-2-phenylindole
DAPT	N-[N-(3,5-Difluorophenacetyl-L-alanyl)]-(S)-phenylglycine t-butyl ester
DEG	differentially expressed gene
DESTINY	DeEscalation and Stopping Treatment of Imatinib, Nilotinib or sprYcel in chronic myeloid leuk
DHH	desert hedgehog
DLL	delta-like
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DOCK1	Dedicator Of Cytokinesis 1
DPYSL3	Dihydropyrimidinase Like 3
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen E
EFNA5	Ephrin-A5
ES	enrichment score
ETP	early T cell precursor
EURO-ski	The European Stop Kinase Inhibitor trial
FACS	fluorescence activated cell sorting
FC	fold change
FEZ1	Fasciculation And Elongation Protein Zeta 1
FISH	fluorescence In Situ Hybridisation
FMO	fluorescence minus one
FSC	forward-angle scatter fucose-GlcNAc fucose-N-acetylglucosamine
FSTL4	Follistatin Like 4
GEO	gene expression omnibus
Gfi1	growth factor independence 1
GlyA	glycophorin A
GMP	granulocyte-monocyte progenitors
GMP	granulocyte-macrophage progenitor
GO	gemtuzumab ozogamicin
GO	gene ontology
GSC	gamma secretase complex

GSK3 β	glycogen synthase kinase 3 β
GSEA	gene set enrichment analysis
GSI	gamma secretase inhibitor
HBSS	hanks balanced salt solution
Hh	hedgehog
HRP	horseradish-peroxidase
HOXB9	Homeobox B9
HP	Haptoglobin
HSC	haemopoietic stem cell
HSPC	haemopoietic stem and progenitor cells
IF	immunofluorescence
IFC	integrated fluidic circuit
IHH	indian hedgehog
IS	international scale
JAG	jagged
JAK-STAT	janus kinase/signal transducers and activators of transcription
KL	Klotho
KRT8	keratin 8
LBC	lymphoid blast crisis
Lin	lineage marker
LMPP	lymphoid-primed multipotent progenitors
LN	European Leukemia Net
LOX	Lysyl Oxidase
LSCs	leukaemic stem cells
LTC-IC	long-term culture initiating cell assays
LTMCM	long-term myeloid culture medium
MAML	mastermind-like
MCyR	major cytogenetic response
MCAM	Melanoma Cell Adhesion Molecule
MDCP	monocyte-dendritic cell precursor
MEF2A	Myocyte Enhancer Factor 2A
MEP	megakaryocyte-erythroid progenitors
MFI	mean fluorescence intensity
miRNA	microRNA
MLP	mixed lymphoid progenitor
MMP8	Matrix Metalloproteinase 8
MMR	major molecular response
MNC	mononuclear cells
MPP	multipotent progenitors
MRD	minimal residual disease

Msi-1	Musashi-1
MSigDB	The Molecular Signatures Satabase
MSPC	mesenchymal stem and progenitor cells
NCBI	The National Center for Biotechnology Information
NCCN	National Comprehensive Cancer Network
NK cells	natural killer cell
NOD-SCID	non-obese diabetic severe combined immunodeficient
Padj	adjusted P value
PB	peripheral blood
PCA	principal component analyses
PCR	polymerase chain reaction
PF4	platelet factor 4
PI	propidium iodide
Ph+	Philadelphia chromosome positive
PPP1R9A	Protein Phosphatase 1 Regulatory Subunit 9A
PreB	Pre B cell
ProB	Pro B cell
PRPH2	Peripherin 2
PTCH1	Patched1
PTCH2	Patched2
PVDF	immune-Bot polucinylden difluoride
R-Jag1	recombinant jagged1
RIN	RNA integrity number
SCF	stem cell factor
RNAseq	ribonucleic acid sequencing
SDS-PAGE	sodium dodecyle sulphate-polyacrylamide gel electrophoresis
SEA	Simpler Enrichment Analysis
SFM+5GF	serum free media and 5 growth factor
SFM+LGF	low growth factor
SFM	serum free media
SHH	sonic hedgehog
SMO	smoothened
SORBS2	Sorbin And SH3 Domain Containing 2
SOX11	SRY-Box Transcription Factor 11
SRC	SCID-repopulating cells
SSC	side-angle scatter
ST18	ST18 C2H2C-Type Zinc Finger Transcription Factor
STIM	STop IMatinib
STK36	serine threonine kinase 36
TCF/LEF	T-cell factor/lymphoid enhancer factor

Tel	translocation Ets leukaemia
TFPI	Tissue Factor Pathway Inhibitor
TGF- β	transforming growth factor beta
TGF- β 1	transforming growth factor beta 1
TKI	tyrosine kinase inhibitor
TPSB2	Tryptase Beta 2
UTR	un-translated region
ZFPM2	Zinc Finger Protein, FOG Family Member 2

Chapter 1. Introduction.

1.1 Introduction: Acute Lymphoblastic Leukaemia (ALL)

Acute lymphoblastic leukaemia (ALL) is an aggressive type of leukaemia characterised by the presence of large proportions of immature lymphoblasts ('blast cells') in the peripheral blood or bone marrow. Such large populations of leukemic blast cells are the result of malignant transformation leading to uncontrolled proliferation, inhibition of differentiation and reduced responsiveness to pro-apoptotic signals (Terwilliger & Abdul-Hay., 2017). The outcome of this is global dysregulation in immune responsiveness, this is exemplified by the overproduction of dysfunctional immature lymphoblasts and the resultant suppression of normal functioning leukocytes, platelets, and erythrocytes.

In ALL, malignant transformation by chromosomal abnormalities and genetic alterations most commonly occurs in immature B cell precursors (85%) with T cell lineage ALL being observed less frequently (15%) (Raetz & Teachey., 2016). The identification of recurrent genetic alterations has allowed for the definition of ALL subtypes and the observation of subtype-specific treatment requirements. This, in combination with traditional risk-stratification clinical factors (such as white blood cell count and chemotherapy response history) has advanced management of a disease which previously had a 5-year survival rate of <10% in the 1960s to over 90% according to contemporary paediatric ALL studies today (Jeha et al., 2019 and Inaba & Mullighan., 2020). Treatment requires a multi-agent backbone for therapy including chemotherapy with the addition of vincristine, an anthracycline, corticosteroids and, where appropriate, an allogeneic stem cell transplant. Despite the successful improvement in overall survival rate, ALL remains a high-risk leukaemia with approximately 10-15% of patients being refractory to treatment or relapsing following treatment (Sidhu et al., 2023).

The incidence of ALL by age is observed as having a bimodal distribution with the highest number of new cases being reported in children aged 0-4 years, a low incidence between ages 19-55 years and a secondary peak in diagnoses at 70 years (CRUK., 2021). Cancer in children (defined as age 0-14years) and young adults (up to age 24 years) is rare, with less than 1% of all new cancer cases reported in the UK being in children and young people. However,

haematological malignancies are the most diagnosed cancer in children, with a third (31%) of cases being either leukaemia, lymphoma, myeloproliferative diseases or myelodysplastic diseases (NCRAS., 2021). In addition, the incidence rate of childhood cancer has been steadily increasing since the 1970s; in part, due to improved diagnostic criteria and broader understanding of disease pathogenesis (Smith et al., 2014). ALL is the most common cancer diagnosed in children, comprising approximately 25% of newly identified cancer diagnoses in children aged 15 years and younger (DCCPS, 2022, DuVall et al., 2022).

ALL is characterised by the uncontrolled proliferation of lymphoid progenitor cells in the bone marrow (BM), resulting in the production of leukemic lymphoid 'blast' cells. Diagnosis of ALL is established based on the presence of 20% or more lymphoid blast cells in the BM or peripheral blood, assessed via morphology, immunophenotyping (i.e. flow cytometry) and cytogenetic analysis. High numbers of leukemic blasts in the BM typically presents with symptoms of BM failure (bone pain, anaemia, thrombocytopenia, or leukopenia) (Stelljes & Marks., 2019). In addition, accumulation of poorly differentiated lymphoblasts can be observed in extramedullary sites such as eyes (ocular), kidneys (renal), bladder, liver (hepatic), central nervous system (CNS), pancreas, skin, pericardium and spleen. This accumulation of malignantly transformed progenitor cells results in hepatosplenomegaly or lymphadenopathy in 20% of patients (Jabbour et al., 2005). Though involvement in such sites is rare at presentation, blast presence is still monitored in a number of areas due to the risk of site-specific sequestered clonal cells being a route to relapse (Shahriari et al., 2020). An example of this is within the CNS, where despite only 10% of patients initially presenting with CNS symptoms, CNS-driven relapse occurs in 30% of children (Deak et al., 2021).

1.2.1 Introduction: ALL pathophysiology – molecular genetics and cytogenetic alterations

There are risk factors associated with developing ALL, these include environmental factors (such as the exposure to ionising radiation) and genetic syndromes (such as Trisomy 21) which can predispose the development of leukaemia. However, most ALL cases are observed to occur as a result of *de novo* malignant transformation in previously healthy individuals (Leuraud et al., 2015, Yokota & Kanakura., 2016).

ALL comprises a heterogeneous group of high-risk lymphoid neoplasms arising from a variety of genomic alterations (Pui et al., 2004). Not only do these genomic variations have an impact on downstream cellular activity, but they have also been a useful tool in identifying clinically distinct subtypes, as defined through the WHO 2022 lymphoid classification (Table 1.1) (Schwab et al., 2022). Though the incorporation of whole genome sequencing and other genetic approaches to the research of ALL has yielded the discovery of novel genomic lesions, there are several ALL subtypes with distinct constellations of somatic structural DNA rearrangements and sequence mutations (Lacobucci and Mullighan., 2017). These well-characterised genomic lesions perturb normal lymphoid development, expression and function of cytokine receptors, chromatin remodelling and a number of signalling pathways (i.e. various kinases and Ras).

Although diverse, such genomic alterations effect a number of mechanisms such as the aberrant expression of proto-oncogenes, an example being the pan-ALL upregulation of Bmi-1, a member of the Polycomb-group (PcG) family, involved in several biological pathways including cell-cycle and DNA damage response (Peng et al., 2017). Chromosomal translocations resulting in fusion oncogene formation is common within ALL, exemplified by the t(12;21)(p13;q22) translocation event, the most commonly diagnosed chromosomal translocation in children with ALL (Montaño et al., 2020). This results in the fusion of the transcription factors ETV6 (TEL) and RUNX1 (AML1) and the dysregulation of the PI3K/Akt/mTOR (phosphoinositide 3-kinase/Akt/mammalian target of rapamycin) pathway leading to the inhibition of apoptosis, lymphoid differentiation arrest and increased cell survival.

Subtype	Ocurrence (%)	Clinical characteristics	Description
t(9;22)(BCR::ABL1)	2-3	High risk	Presence of a translocation between chromosomes 9 and 22 [t(9;22)] which creates the Philadelphia chromosome and the <i>BCR::ABL1</i> fusion gene
t(1;19) (E2A-PBX1)	5	Low risk	Presece of translocation between chromosomes 1 and 19 resulting in the formation of the oncogenic E2A-PBX1 fusion protein
t(12;21)(TEL-AML1)	16-22	Normal ALL low risk	Presence of translocation between chromosomes 12 and 21 resulting in the formation of the oncogenic TEL-ABL1 fuson protein
t(4;11)(MLL)	5-8	Infant ALL high risk	Rearrangement of the histone lysine [K]-MethylTransferase 2A gene on chromosme 11q23 resulting in the formation of the mixed lineage leukaemia (MLL) gene
Hyperdiploid>50	25-35	Normal ALL low risk	Presence of leukaemia cells with more than 50 chromosomes
Hypodiploid<44	<7	Hig risk	Presence of leukaemia cells with fewer than 44 chromosomes
T-ALL	10-13	T-ALL moderate risk	Resulting from the activating mutations <i>NOTCH1</i> and <i>FBXW7</i>

Table 1.1: ALL subtypes and characteristics. Acute lymphoblastic leukaemia subtypes with oncogenic chromosomal aberrations or mutations, percentage frequency of overall ALL diagnoses and brief description of the leukaemia initiating event (WHO., 2022).

Recurrent genomic features, and hence ALL subtypes, are detected at different frequencies across age groups with prognostication differing by age. KMT2A (MLL) rearrangements are most frequently diagnosed in infants and represents a high-risk subtype with poor clinical outcomes. KMT2A (MLL) rearrangements, particularly the t(4;11)(q21;q23) rearrangement results in alterations to the kinase-PI3K-RAS signalling pathway; this subtype has a poor prognosis and is less frequently observed in children and adults. In contrast, the t(9;22)(q34;q11.2) translocation event which results in the formation of the constitutively active tyrosine kinase BCR::ABL1, Ph+ALL (Philadelphia chromosome positive acute lymphoblastic leukaemia) is more commonly diagnosed in adults than children with 25% of adult cases and up to 5% of childhood cases of leukaemia being Ph+ALL. These genomic alterations contribute to leukemic transformation by the dysregulation of normal cellular functions and key regulatory processes. The result being the capacity for unlimited self-renewal, subversion of normal proliferation, halting of homeostatic differentiation and promotion of apoptosis resistance. As the name ALL denotes, the cell type most affected by these dysregulations are lymphoid cells.

In addition to subtype-specific genomic alterations, a number of genetic lesions, such as *IKZF1* aberration can be observed across subtypes and used for prognostic markers (Stanulla et al., 2020). *IKZF1* alterations are present in approximately 15% of childhood ALL and when observed in adults, is associated with kinase-driven leukaemia such as Ph+ALL (Vairy and Tran., 2020). *IKZF1* encodes the transcription factor for *IKAROS*, a member of the family of zinc finger DNA-binding proteins required for haematopoietic lineage ontogeny and homeostasis. Alterations to *IKZF1* have been found to have prognostic significance in both Ph+ALL and Ph-like ALL and is, therefore, utilised as a biomarker for poor clinical outcomes. Although, in the context of newer treatments, such as blinatumomab, the prognostic significance of these mutations are unknown. Ph-like ALL, like Ph+ALL, occurs more frequently in adolescents and adults than in children, with up to 27% of young adults diagnosed with ALL between the ages of 21 and 39 years (Tran and Loh., 2016). Patients with Ph-like ALL display alterations in cytokine receptor genes and signalling pathways (commonly in the JAK-STAT and kinase pathways) similar to the downstream aberrations seen in Ph+ALL however, unlike in Ph+ALL, such patients do not display evidence of BCR::ABL1 fusion.

Similar to Ph+ALL, Ph-like ALL is a high-risk ALL subtype with a poor prognosis however, Ph-like ALL is almost three times more common than Ph+ALL and is defined by gene expression signatures (such as COG-TARGET-St Jude (Mullighan et al., 2009) and Den Boer et al., 2009) rather than the presence of chromosomal translocation and gene fusion. This subtype has a B-lineage and bears the hallmark of deletions or mutations of the lymphoid transcription factor gene *IKZF1*. The subtype is also characterised by alterations in cytokine receptor genes and signalling pathways also observed in Ph+ALL, such as the JAK-STAT kinase pathways. This genetically heterogeneous disease exhibits a variety of kinase fusions which can be targeted therapeutically, in the example of *ABL1/ABL2* gene fusion, with tyrosine kinase inhibitor drugs used to treat Ph+ leukaemia (ie imatinib/dasatinib) which will be discussed in depth later in this chapter. However, due to the variety of kinase fusions present in the Ph-like ALL cohort, a personalised treatment approach is taken and kinase fusions that can be exploited therapeutically are first identified prior to directed therapy. JAK1/2/3 inhibitors are used when alterations to *IL2RB*, *JAK2* or *TSLP* are present and TRK inhibitors when *NTRK3* alterations are observed. Hence, despite their similarities, different treatment approaches are taken between Ph+ALL and Ph-like ALL. Though both ALL subtypes share aberrations to signalling pathways

and cytokine interactions, this thesis will focus on Ph+ALL which bears the *BCR::ABL1* fusion gene.

BCR::ABL1 kinase activity has been observed to induce chromosomal and genetic instability by inducing an error-prone DNA repair system, cultivating an intracellular environment rich in reactive oxygen species (ROS) and centrosome aberration (Senapati and Sasaki., 2022). The presence of *IKZF1* alterations in both Ph+ and Ph-like ALL describes an alteration to IKAROS activity independent of *BCR::ABL1* kinase activity. The importance of identifying genomic alterations, is not only for definition of ALL subtypes which would require altered treatment strategies, but also as prognostic markers across ALL subtypes.

1.2.2 Introduction: ALL pathophysiology- phenotypic cellular compartments

Bone marrow failure observed in ALL exemplifies how detrimental the alteration of normal lymphoid development is and hence, the importance of adequate diagnosis and appropriate treatment. In 75% of ALL diagnoses, the functionally dysregulated cell lineage is B cells (Campos-Sanchez et al., 2011). Such malignant transformation disrupts normal B cell development, resulting in differentiation arrest, accumulation of non-functional progenitor B cells.

Early B cell development is a plastic but highly controlled process (Fig 1.1). B cell development from non-committed self-renewing progenitors (haematopoietic stem cells (HSC)) to mature functional plasma cells and memory B cells is dependent on both the bone marrow (BM) and lymphoid organs within the body, networks of cytokine and chemokine signalling and transcription factors (Melchers., 2015). The result of this is a plastic, yet highly conserved mechanism of establishing and maintaining a diverse repertoire of functional B cells. Initial establishment of B cell immunity occurs during early embryonic development where multipotent HSCs (mHSCs) migrate to the foetal liver wherein they develop into mature B cells before egress to and population of tissues such as lungs, epithelia and gut-associated lymphoid tissues. After these initial stages of development, B cells are continuously developed to maintain an appropriate level of adaptive immunity. B cells are produced in primary lymphoid organs from HSCs and once mature, migrate to peripheral secondary lymphoid tissues via lymphatic vessels and blood (Alberts et al., 2002). The phenotypic compartments implicated in B cell development are commonly associated with specific developmental stages, such as development from pre-B to mature B cells in the spleen.

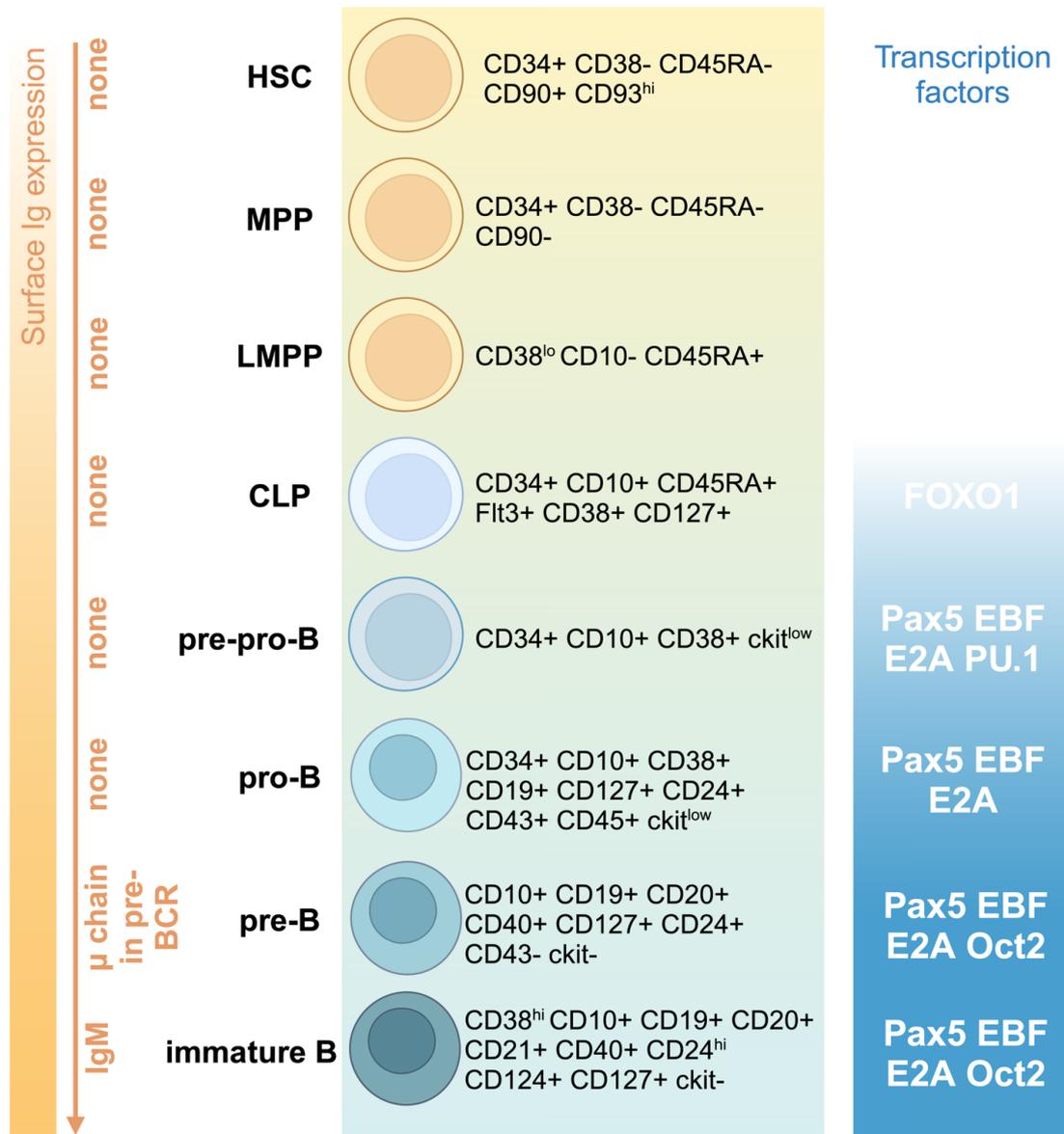


Figure 1.1: Surface markers of early B cell development. Early stages of b cell development. Ig= immunoglobulin, Pax5= paired box gene 5, EBF= early B cell factor 1, OCT2= octamer transcription factor 2, FOXO1= forkhead box protein O1. Cell surface marker and immunoglobulin expression changes during B cell development driven by presence of stage-specific transcription factors. (Adapted from Campos-Sanchez et al., 2011, Created with BioRender.com)

Emergence of mature lymphoid cells to the periphery allows lymphoid cells to exert their immunoprotective functions against specific antigenic insult. In contrast to cells such as dendritic cells (DCs), macrophages, natural killer (NK) cells, eosinophils, basophils and mast cells, lymphocytes mount a highly selective adaptive immune response, with T and B cell clones being reactive to a specific antigen. B cells produce antibodies with high levels of antigen specificity depending on antibody affinity and avidity (Rudnick and Adams., 2009). Due to some classes of antibodies being secreted and distributed via the blood, antibodies are able to exert effects over long distances. In contrast, T cells exert local effects and can be

subclassified into cytotoxic T cells and helper T cells depending on cellular function. Cytotoxic T cells directly kill cells through surface expressed or secreted cytotoxic proteins, whereas helper T cells have roles in activation of macrophages, B cells and cytotoxic T cells through secretion of cytokines and expression of membrane-bound costimulatory molecules. Though T cells act locally, their ability to migrate systemically allows T cell-mediated cell activation and cytotoxic effects to be enacted throughout the body.

B cell lineage commitment is a highly conserved multi-phasic developmental process and as a result, a number of intermediary lymphoid progenitors have been described. As previously discussed, such progenitors are restricted to lymphoid development sites with the most primitive progenitors requiring BM microenvironmental factors to support appropriate early B cell differentiation. Initial stages which lead to lymphoid lineage commitment involve the differentiation of HSCs to lymphomyeloid-primed progenitors (LMPP) which lack megakaryocyte and erythrocyte potential (Pooter et al., 2019). LMPPs express low levels of lymphoid and myeloid lineage mRNAs and are therefore still primed for multilineages with lymphoid lineage specification occurring through gene priming mediated by transcription factors PAX5, E2A-HEB and Lyl1. Normal B cell development requires a highly regulated interplay between extrinsic B cell receptor (BCR) and cytokine signalling, and intrinsic epigenetic and transcriptional programming, a multistep process with quality control checkpoints to ensure acceptable functionality (Campos-Sanchez et al., 2011). Verification of stochastic immunoglobulin gene rearrangement for the prevention of BCR autoreactivity allows precursor cells access to later lymphoid differentiation stages and prevents the establishment of autoimmune diseases (Übelhart and Jumaa., 2015). These checkpoints are multiphasic with verification of non-reactivity in the precursor-BCR (pre-BCR) resulting from heavy chain rearrangement allowing progression into light chain immunoglobulin gene rearrangement, receptor editing and the expression of self-tolerant BCR.

Understanding this physiological process has been integral in the development of measurable residual disease assessment in ALL. As previously discussed, latter stages of lymphopoiesis involve the development of antigen-specific mature effector lymphocytes which can then undergo clonal expansion in response to antigenic insult (Adams et al., 2020). The generation of antigen-specific immunoglobulins and T cell receptors involves the rearrangement of gene

segments encoding for structural elements of antigen receptors, termed variable (V), diversity (D) and joining (J) regions (Bagg., 2006). VDJ gene rearrangement occurs early during lymphoid development, resulting in a wide repertoire of antigen receptors and providing information on cell clonality- a useful diagnostic laboratory tool. In the context of leukaemia, the documentation of unique antigen receptor gene rearrangements provides a molecular fingerprint of the neoplastic clone which can then be used for prognostically relevant minimal residual disease assessment.

Progressive lineage commitment from early lymphoid progenitor (ELP) to committed B cell can be observed phenotypically through cell surface marker expression allowing B cell progenitor cells to be identified (Fig 1.2). Progressive loss of stem cell markers such as CD34 marks the gradual loss of pluripotency as cells gradually gain lymphoid-specific characteristics and move toward a fixed B cell lineage. Just as with HSCs, MPPs (multipotent progenitor cells) are produced in the BM (prenatally, in the foetal liver), MPPs are then able to differentiate into MLPs (multi-potent lymphoid progenitors) upon interaction with the appropriate differentiation signals from bone marrow stromal cells such as IL-4 and CXCR4 (Bio-Rad., 2016; Gomes et al., 2016). Early pro B cell development is then marked by the expression of recombinase activating genes (RAG) and terminal deoxynucleotidyl transferase (TdT) in addition to surrogate light chain (SLC) expression resulting from heavy chain D-J joining (Table 1.2). Similar to early pro B cells, late pro B cells also express SLC and undergo VDJ heavy chain recombination. SLC expression and heavy chain rearrangement also occurs in large preB cells, however this developmental step is marked by the expression of the μ chain in the pre-BCR and the silencing of RAG and TdT. Light chain V-J joining commences upon the development of small preB cells in addition to the expression of the μ chain on the cell surface and the re-expression of RAG, TdT but no SLC expression. Upon the formation of immature B cells, the heavy and light chains have been rearranged, and IgM and Ig $\alpha\beta$ are expressed. Immature B cells undergo egress from the BM and enter circulation as transitional B cells found most commonly in blood and secondary lymphoid organs (rarely in lymphatic vessels). Mature effector B cells can be categorised into functional subsets based on cell surface marker phenotype, associated transcription factors, cellular location and broadly characterised by the expression of functional mature BCR (Naradikian et al., 2014; Melchers., 2015).

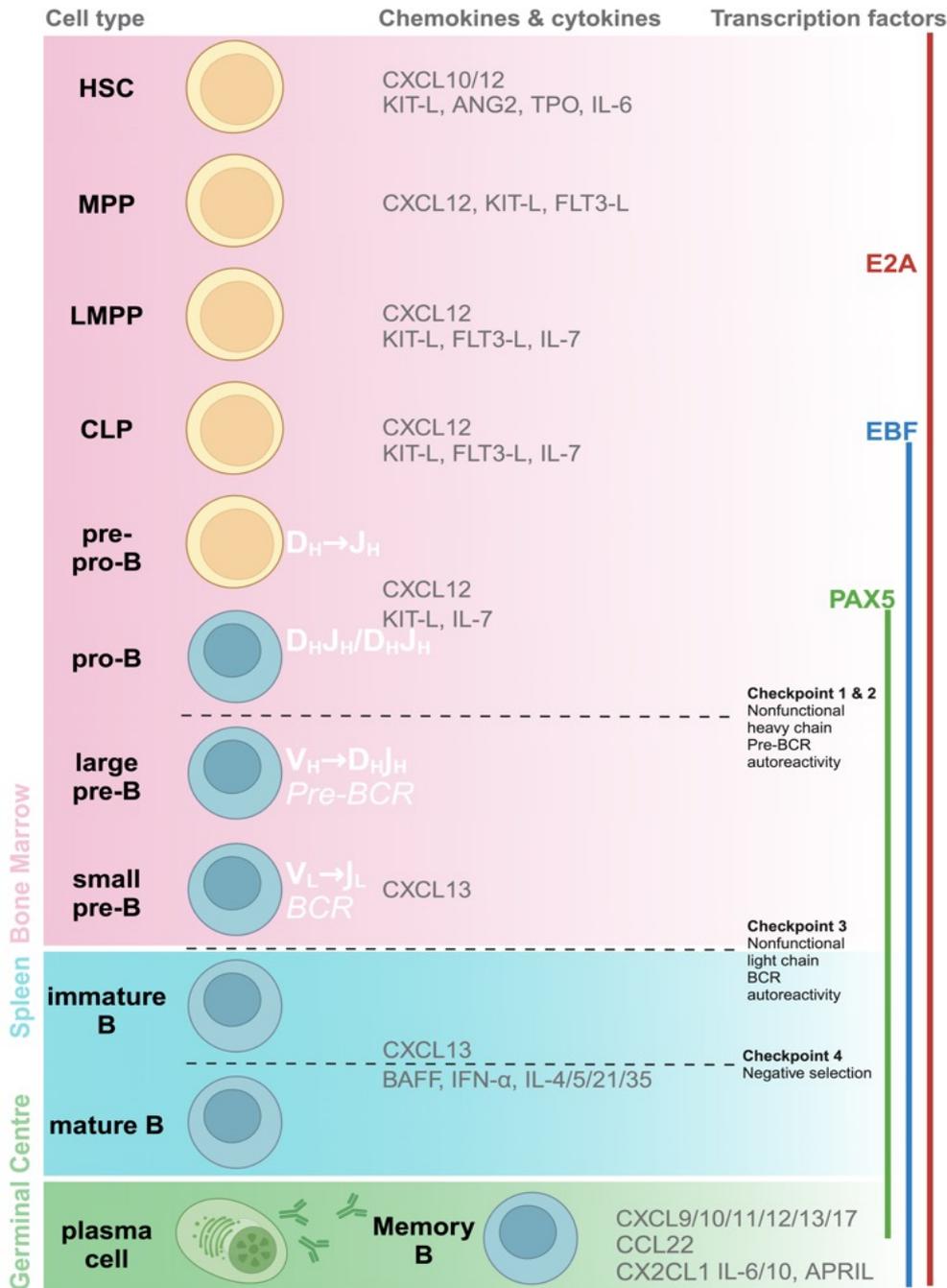


Figure 1.2: B Cell development in bone marrow to periphery. Postnatally, HSCs originate in the bone marrow (BM) and localise in specialised BM niches where progenitor cells are exposed to pro-lymphoid growth factors and cytokines. Small preB cells leave the BM for development into mature cells in the spleen. Functional effector B cells (plasma cells or memory B cells) are then active in the periphery where germinal centres may be formed. Developmental checkpoints to prevent B cell autoreactivity have been indicated.

(Adapted from Melchers., 2015, Created with BioRender.com)

Subset	Phenotype	Associated TFs	Cellular location
Transitional	CD20+ CD27- CD38hi IgM+ CD24hi BR3+	Pax5, EBF, E2A, Oct2	Migration from BM to secondary lymphoid organs
Follicular	IgMlo CD23+ CD93- CD19+ CD20+ CD21+ CD22+	Pax5	Shuttling between BM and secondary lymphoid organs
Marginal zone	IgMhi IgDlo CD1c+ CD24+ CD19+ CD20+ CD21+	Pax5, EBF, E2A, Oct2	Secondary lymphoid organ
Germinal Center	CD20+ CD38+ BR3+ IgD-	BCL6, Pax5, EBF	Secondary lymphoid organ
Plasma cells	CD20- CD38hi CD27hi CD138+ TACI+ and/or BCMA+ CD126+ CD319+ CD78+	BLIMP1, IRF4, XBP1	Long lived plasma cell in BM. Short lived plasma cell in secondary lymphoid organs
Memory B Cell	CD20+ CD38- CD27+ CD80+ CD84+ CD86+	OBF1, SPI-B	Circulating in both BM and lymphoid locations

Table 1.2: Characterisation of peripheral B cell subsets (human). TFs= transcription factors, Pax5= paired box gene 5, EBF= early B cell factor 1, OCT2= octamer transcription factor 2, BCL6= B cell lymphoma 6 protein, IRF4= interferon regulatory factor 4, XBP1= X-box binding protein 1, TACI= transmembrane activator and cyclophilin ligand interactor, BCMA= B cell maturation antigen, BR3= B lymphocyte stimulator receptor 3, Ig= immunoglobulin.
(Adapted from Melchers., 2015)

As previously discussed, B cell development is dependent on the coordinated and highly controlled action of transcription factor regulatory networks to activate B cell development programmes and silence alternative cell fates (Somasundaram et al., 2015). In ALL, malignant transformation resulting from an initial reciprocal translocation event (such as the formation of *BCR::ABL1*) induces a sequential dysregulation of such development pathways resulting in the expansion of immature non-functional progenitor cells (blast cells). Though the initial malignant transformation event varies between B-ALL subtypes as previously discussed, the alterations to essential B cell development programmes, cytokine signalling and transcription factor networks such as IKZF1, TCF3, EBF1 and PAX5 resulting in differentiation arrest is common within B-ALL as a whole (Fig 1.3).

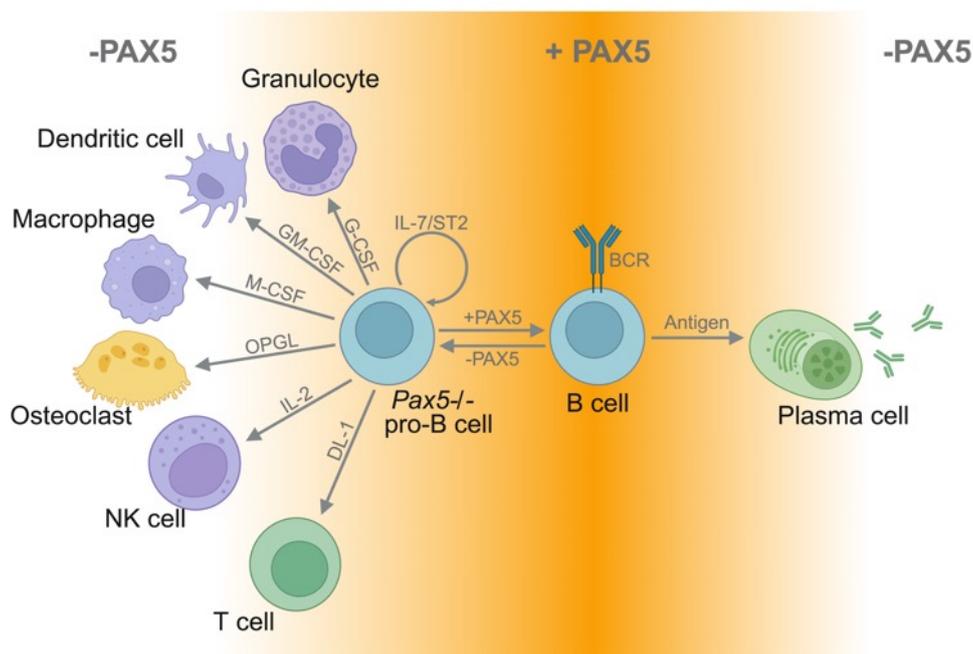


Figure 1.3: B cell lineage commitment by PAX5. Uncommitted *Pax5*^{-/-} pro-B cells are able to differentiate into a variety of myeloid and lymphoid cells in the presence of the indicated cytokines. Conditional *Pax5* deletion results in retrodifferentiation of B cells into an uncommitted progenitor cells. OPGL= osteoprotegerin ligand, ST2= stromal ST2 cells.

(Adapted from Cobaleda et al., 2007, Created with Biorender.com)

1.3 Introduction: Prognostic factors and measurable residual disease

Over the past few decades, long-term overall survival rate for paediatric ALL has greatly improved in excess of 90% (Hunger et al., 2012). This is due, in part, to advances in treatment and improved therapeutic strategies, however, a huge improvement has come from enhanced risk stratification and, importantly, the development of measurable residual disease (MRD) monitoring. Initial allocation of treatment schema at diagnosis differs from MRD risk analysis which is measured after induction treatment (Table 1.3). At diagnosis patients are assessed for determination of leukemic subtype in order to allocate the appropriate treatment; for example, the detection of the t(9;22) fusion product BCR::ABL1 leading to treatment with tyrosine kinase inhibitors, or the use of Rituximab, a monoclonal antibody targeting CD20, in those patients who are CD20 positive by flow cytometry (Marks et al., 2022). In contrast, MRD refers to the small number of cancer cells present within the body after treatment, with a positive MRD score resulting from the detection of residual cancer cells and a negative score where cancer cannot be detected. The function of MRD monitoring is two-fold; it allows clinicians to monitor the effectiveness of treatment, while also indicating which patients are likely to relapse.

Historically, traits detectable at diagnosis such as age, white cell count, immunophenotype, karyotype and additional molecular alterations have been utilised for risk allocation schemas however, with the development of MRD monitoring risk stratification has progressed to include disease-specific biological features. Detection of disease-related clinical features have also benefitted from the development and refinement of techniques to evaluate MRD, such as flow cytometry, real-time quantitative PCR (RQ-PCR) and PCR of fusion transcripts (Brüggemann et al., 2012).

Prognostic factors	Category	Prognostic impact	Potential impact on targeted therapy
Age	Paediatric	Worse outcome in infancy	NA
	Adult	Worse outcome with advancing age	
WBC at diagnosis	>30*10 ⁹ /L (B-ALL)	High WBC at diagnosis associated with poor prognosis	NA
	>100*10 ⁶ /L (T-ALL)		
Immunophenotype	CD20 expression	Conflicting prognostic data	Monoclonal antibodies
	T-ALL vs B-ALL	Early studies indicating better prognosis in T-ALL	Monoclonal antibodies, bispecific T cell engager (Nelarabine)
Cytogenetics	t(9;22)/BCR::ABL1	Poor prognosis	TKI
	t(4;11)/MLL-AF4	Poor prognosis	NA
	t(8;14)		
	Hypodiploidy (<44 chromosomes)/leukaemic cell		
	Near triploidy		
	Complex karyotype (>5 abnormalities)		
	t(1;19)	Conflicting prognostic data	NA
	High hyperdiploidy (>50 chromosomes)/leukaemic cell	Better prognosis	NA
	del(9p)		
	JAK mutations	Poor prognosis	JAK inhibitors
Specific molecular alterations	IKZF deletions/sequence mutations	Poor prognosis	
	CRLF2 overexpression	Poor prognosis- paediatric	CRLF antibodies
	ERG/BAALC expression	Conflicting prognostic data	
	NOTCH1 mutations	Conflicting prognostic data	NOTCH1 targeting

Table 1.3: Post-diagnosis, pre-therapeutic factors associated with ALL outcomes. NA= not applicable, CRLF2= cytokine receptor-like factor 2, ERG= w-ets erythroblastosis virus E26 oncogene homolog (avian), BAALC= brain and acute leukaemia cytoplasmic, IKZF= IKAROS family zinc finger, JAK= Janus kinase, TKI= tyrosine kinase inhibitors, WBC= white blood cell count.

(Adapted from Brüggemann et al., 2012)

Flow cytometry allows for the rapid detection and quantitation of aberrantly expressed antigens on leukemic cells. Over the past two decades, with the advent of multicolour flow cytometry and availability of the technology in many diagnostic laboratories, standardised MRD criteria have been developed for ALL (Chatterjee et al., 2016). This includes the aberrant expression of surface markers, such as CD34/CD19/CD20/CD21 and CD34/CD56, the combinations of which are rare in normal healthy bone marrow and hence, display the presence of developmentally arrested B cells resulting from leukemic transformation in B-ALL. This technology is dependent on the number of coloured lasers available at the diagnostic testing site and starting cell number, hence though flow cytometry MRD is largely accessible and informative, it may not be sensitive enough for the detection of very small numbers of residual leukemic cells (max 10^{-4}). Although the sensitivity is concordant to real-time quantitative polymerase chain reaction (RQ-PCR) (10^{-3} - 10^{-5}), flow cytometry-based MRD monitoring requires $>4 \times 10^6$ cells for sufficient sensitivity (Theunissen et al., 2017). Therefore, the effectiveness of flow cytometry MRD is dependent on a number of factors such as the antibody panel used, the discriminatory level of the leukaemia-associated immunophenotypes (LAIPs), cell number, multi-centre standardisation and relevant expertise for data analysis and interpretation (Tettero et al., 2022).

PCR is a commonly used diagnostic tool for the detection of target genes of low copy numbers and has been utilised outside of oncology in areas such as virology. The nucleic acid target flexibility of PCR can be evidenced in this context in the detection of viral DNA used for SARS-CoV-2/COVID-19 testing and RNA detection for retroviruses such HIV (human immunodeficiency virus). This flexibility has allowed for DNA-based real time quantitative PCR and RNA-based RT-PCR MRD analysis tools to be developed and enhanced. DNA RQ-PCR in ALL mainly targets the rearrangement of immunoglobulin (Ig) and T cell receptor (TCR) genes, but also has a common use in detecting MLL (mixed-lineage leukaemia) gene rearrangements or *SIL-TAL1* deletions. Ig/TCR gene rearrangement will be a focus in the results section of this thesis and hence will be discussed in later sections of this chapter and in further detail than MLL gene rearrangement associated with the t(4;11)(q21;q23) translocation event prevalent in infant ALL (Britten et al., 2019). As previously discussed, the rearrangement of immunoglobulin genes is a crucial step in the B cell developmental pathway, allowing for a repertoire of diverse BCRs to be established. Early T cell differentiation similarly involves

germline rearrangement of V, D and J gene segments to establish a diverse pool of non-autoreactive T cells. RQ-PCR analysis of junctional regions of rearranged Ig/TCR genes is broadly applicable for the majority of patients (>95%) and can detect patient-specific rearrangements with high specificity (one malignant cell in $<10^5$ normal cells (10^{-5})) (van der Velden and van Dongen., 2009). Gene rearrangement involves the random insertion and deletion of nucleotides at the junctional sites of V, D and J gene segments, unique to each lymphocyte, creating a 'fingerprint' which can be used as tumour-specific targets in MRD-PCR analysis.

Additional uses for PCR in MRD monitoring include RT-PCR detection of *BCR::ABL1* fusion transcripts. Quantitative co-amplification of the *BCR::ABL1* transcript in addition to an internal housekeeping gene (i.e. *GAPDH*) allows for the *BCR::ABL1* copy number to be identified from peripheral blood or BM samples. Reference standards established by WHO multicentre collaborations define *BCR::ABL1* copy number values of 10%, 1%, 0.1% and 0.01% Ph+ cells of normal cells obtained by dilution of K562 (Ph+) and HL60 (Ph-) cell lines (Cumbo et al., 2020). These standards allow for the fast and inexpensive detection of residual leukemic cells with high sensitivity. A drawback of this method is uncertain quantitation resulting from multiple *BCR::ABL1* transcripts per cell, the outcome being the overestimation of the number of Ph+ cells.

MRD and diagnostic risk stratification are, therefore, useful tools in the treatment of ALL, ensuring patients are assigned the correct treatment regimen, verifying the success of treatment and allowing for early detection of patients likely to relapse. Patient-specific differences in cell availability or *BCR::ABL1* copy number per cell displays that not all MRD methods are effective for risk stratification of each patient and therefore, the use of multiple methods are recommended for appropriate care.

1.4 Introduction: Established and developing treatment strategies.

A schematic of the treatment of paediatric ALL is listed in Figure 1.4. Patients are entered into clinical trials where available, for example ALLTogether (ClinicalTrials.gov ID: NCT03911128), for the ongoing improvement of treatment methods. Despite originating in the BM, lymphoid blast cells are able to infiltrate the cerebrospinal fluid (CSF), meningeal membranes and hence, the CNS (Künz et al., 2022). Consequently, despite the low number of patients presenting with CNS involvement at diagnosis (5-10%), it is recommended that all patients receive prophylactic intrathecal treatment and a diagnostic lumbar puncture (Del Principe et al., 2014). Multiple treatment modalities are available for prophylactic CNS therapy including intrathecal therapy (IT) with cytarabine or methotrexate, systemic therapy, and historically, cranial radiation therapy or a combination thereof (Jabbour et al., 2010). Cranial radiation therapy represents the oldest method of CNS prophylaxis, however, it comes with the risk of severe side effects such as secondary neoplasms, endocrinopathy, neurotoxicity and neurocognitive dysfunctions. An essential factor for the success of a CNS-directed chemotherapeutic agent is the ability to cross the blood brain barrier and ensure an appropriate concentration in the CSF without causing toxicity.

A therapeutic approach which has been successfully used in the refractory/relapse setting with good potential to be used first-line treatment is antibody-based therapy. The main classes of antibody-based drugs used in ALL falls into one of three categories; bi-specific T cell engagers (BiTE), antibody drug conjugate (ADC) and monoclonal antibody (mAb). Through the use of these antibodies, cell surface markers such as CD19 (blinatumomab), CD22 (inotuzumab ozogamicin) and CD20 (rituximab) can be targeted on lymphoid blast cells. In the example of blinatumomab (α -CD19), malignant cells are targeted for lysis via the recruitment of CD3 T cells. In order to effectively treat patients where CD19-targeting therapies are no longer viable due to phenotypic escape, CD22 can be targeted if present at relapse. CD22 precedes CD19 in the B cell developmental pipeline and can pose a risk of relapse if CD34+CD19-CD22+ remains after CD19-targeted treatment and the eradication of CD34+CD19+CD22- malignant cells. To best utilise these antibody-based treatments, novel approaches of chemoimmunotherapy have been investigated whereby antibodies are

administered alongside chemotherapy agents. Treating with systemic chemotherapy in combination with inotuzumab oxogamicin, blinatumomab, or rituximab has been investigated for newly diagnosed B-ALL and relapsed/refractory ALL (Rubinstein & O'Brien., 2023, Assi et al., 2017, Levato & Molica et al., 2018).

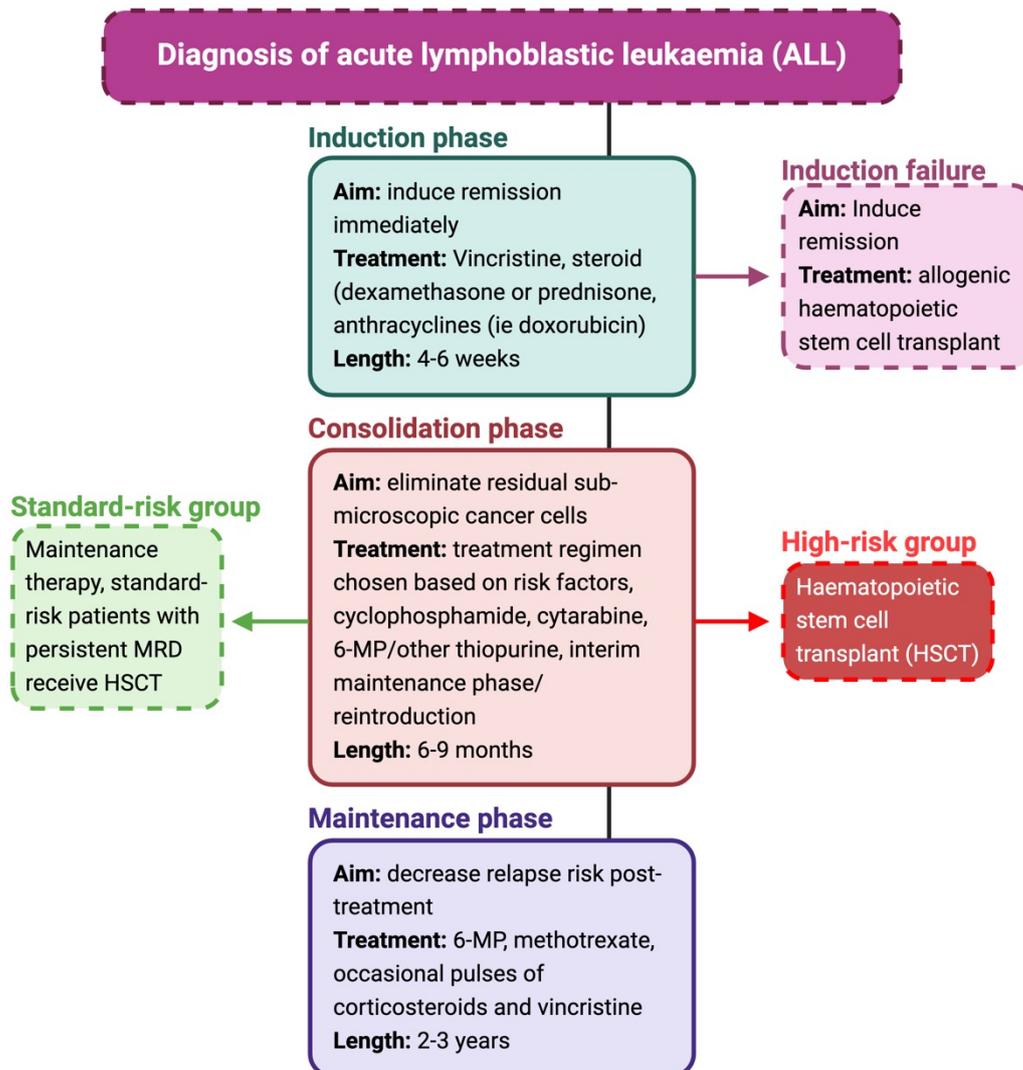


Figure 1.4: ALL risk stratification and treatment schema. After diagnosis, the treatment of patients with ALL follows three main stages; induction phase, consolidation phase and maintenance phase. Patients are stratified based on risk status with treatment regimen assigned accordingly.

(Adapted from Künz et al., 2022 and Rudin et al., 2017, Created with Biorender.com)

Another method of targeting the CD19 marker of early B cell development is chimeric antigen receptor (CAR) T cell therapy. The mechanism of this therapy involves the use of genetically modified autologous cells which target CD19+ haematological malignancies (ie ALL). CAR-T cells can be edited to express alternative co-stimulatory molecules such as CD137 (4-1BB), CD3 ζ and 19-8z CAR. These allow for the alternate activation of T cells for the destruction of malignant CD19+ lymphoid cells (Maude et al., 2018). Phase I and II clinical trials have shown excellent responses to CAR-T cell therapy for ALL with overall survival rates between 73-90% (>50% survival 18 months post-infusion (Graham et al., 2018)) but with that, a high occurrence of adverse events such as neurological events in 40% of patients and cytokine-release syndrome (CRS) in 77% (Park et al., 2018). Results from such trials found it to be imperative that patients must undergo lymphoid-depleting chemotherapy prior to CAR-T cell infusion.

Haematopoietic stem cell transplantation (HSCT) has been well-established for the treatment of ALL. HSCT involves depletion of recipient bone marrow followed by replenishment of a functional immune system with healthy HSCs. Allogenic HSCTs occur when the patient receives stem cells from a donor other than the recipient. Due to the risk of rejection, such donor samples are assessed for HLA (human leukocyte antigen) matching, either coming from an HLA-identical sibling or HLA-matched unrelated donor (MUD). HLA alleles are assessed by ELISA-based serology methods, sequence-specific primers (PCR-SSP), sequence-specific oligonucleotide probes (PCR-SSOP), sanger sequence-based typing (SBT) or next-generation sequencing (NGS) (Spellman., 2022). This allows the compatibility of donated samples to be assessed, however, if a sibling match or MUD is unavailable, umbilical cord blood (UCB), haploidentical related (haplo) or mismatched unrelated donor (MMUD) may be utilised in addition to posttransplant cyclophosphamide-based graft vs host disease (GVHD) prophylaxis to prevent rejection or development of GVHD. Autologous HSCT (α HSCT) involves using the patients' own healthy HSCs for replenishment of a healthy immune system (Snowden et al., 2018). As the donor and recipient are the same person, HLA mismatching is not an issue however, a number of studies have concluded that allogenic HSCT results in improved long-term leukaemia-free outcomes for patients with ALL (Giebel et al., 2019, and Sharma., 2018). Patients are stratified into standard or high-risk and treated either with standard chemotherapy or undergo HSCT. Though conventional therapy is utilised for standard-risk patients with an MRD level below 0.01% at the end of induction or end of consolidation, it has

been observed that patients with high-risk features, but an early MRD response also benefitted more from a standard chemotherapeutic regimen than HSCT (Ribera et al., 2021).

1.5 Introduction: Relapse/refractory disease.

Though diagnosis, treatment and monitoring of ALL has greatly improved over the last few decades, relapse and/or refractory (R/R) disease represents a high-risk disease state with poor survival outcomes and treatment-associated morbidity (DuVall et al., 2022). With the survival rate of ALL increasing to >90% for children (under the age of 15) and 80% for adults through collaborative multicentre trials, the major cause of death in ALL is now relapse with a median survival of <10-25%. The goal of R/R treatment is to therapeutically induce a remission (with undetectable MRD) and, if required, consolidate the remission with an allogenic haematopoietic stem cell transplant (allo-HCT) or CAR-T approach. In addition, methods used to reach complete response (CR) after R/R may include traditional chemotherapy and a multi-agent approach similar to that which is used in frontline therapy. However, these approaches resulted in limited success with a poor overall survival rate and high toxicity. It has been observed that poorer survival rates post R/R are associated with shorter remission times prior to relapse. Patients relapsing (CNS/BM) within 36 months of diagnosis or 6 months after completion of treatment have a 5-year survival rate of 11.5%. In contrast, patients with a late relapse (>18 months after end of treatment) have a 5-year survival rate of nearly 80%. These statistics exemplify the necessity of highly sensitive MRD monitoring to ensure the early detection of residual leukemic cells and therefore, rapid identification of patients likely to relapse (Teachey and Hunger., 2013).

1.6 Introduction: BCR::ABL1 fusion and cell of origin in Ph+ haematological malignancies (Ph+ALL and CML)

As previously discussed, ALL can be classified into subgroups based on phenotypic, cytogenetic and mutational features. This thesis focusses on ALL that originates from the reciprocal translocation event between chromosomes 9 and 22, termed the Philadelphia chromosome (Ph). It consists of a truncated chromosome 22, resulting from a reciprocal translocation, $t(9,22)(q34;q11)$, between the long arms of chromosome 9 and 22 (Fig 1.5) (Rowley, 1973). During the Philadelphia translocation event a 3' segment of the Abelson murine leukaemia viral oncogene homolog (*ABL*) gene from chromosome 9q34 is added to the 5' segment of the breakpoint cluster region protein (*BCR*) gene on chromosome 22q11 (Fig 1.6). The result is a hybrid gene named *BCR::ABL1*. This fusion gene is then transcribed into mRNA and translated to the BCR::ABL1 protein which functions as a constitutively active tyrosine kinase. The Philadelphia translocation event is detectable in the majority (95+%) of patients with chronic myeloid leukaemia (CML), in addition, 5% of children with ALL harbour the mutation, as do 15-30% of adults with ALL, and 2% of adults with de novo AML (Crews and Jamieson., 2012).

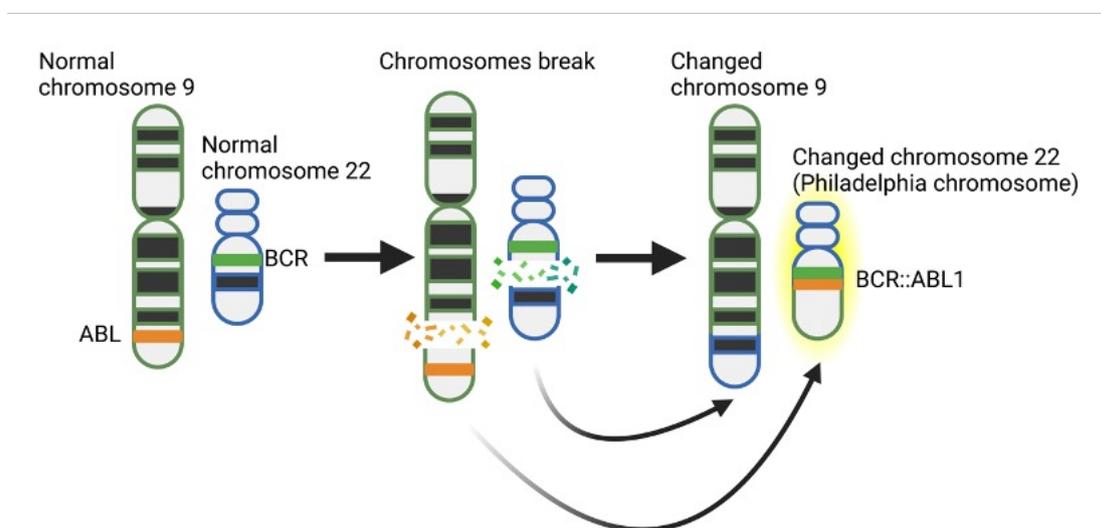


Figure 1.5: Representation of BCR::ABL1 fusion and the formation of the truncated Philadelphia chromosome. The Philadelphia chromosome is formed by the reciprocal translocation event between chromosome 9 and 22. The breakpoints at the *ABL* gene on chromosome 9 and *BCR* on chromosome 22 lead to the fusion gene *BCR::ABL1*.
(Created with Biorender.com)

Splicing at the different breakpoints on chromosome 22 (m, M and μ) results in differently sized BCR::ABL1 proteins; p190, p210 and p230, named for the different size of splice variant measured in kDa (kilodalton) (Fig 1.6). Each variant is associated with different types of leukaemia; with 70% of Ph+ALL patients having the p190 fusion (30% with p210), p210 with 90% of CML and p230 with chronic neutrophilic leukaemia (CNL)- CNL being outwith the scope of this project (Stancuioaica et al., 2019). As these statistics indicate, no one type of BCR::ABL1 splice variant is exclusively associated with a specific type of leukaemia. Hence, detection of BCR::ABL1 fusion variants in acute Ph-driven leukaemia is beneficial for diagnosis and identification of leukemic type, with the caveat that a small number of patients may possess the atypical fusion variant for that leukemic subgroup.

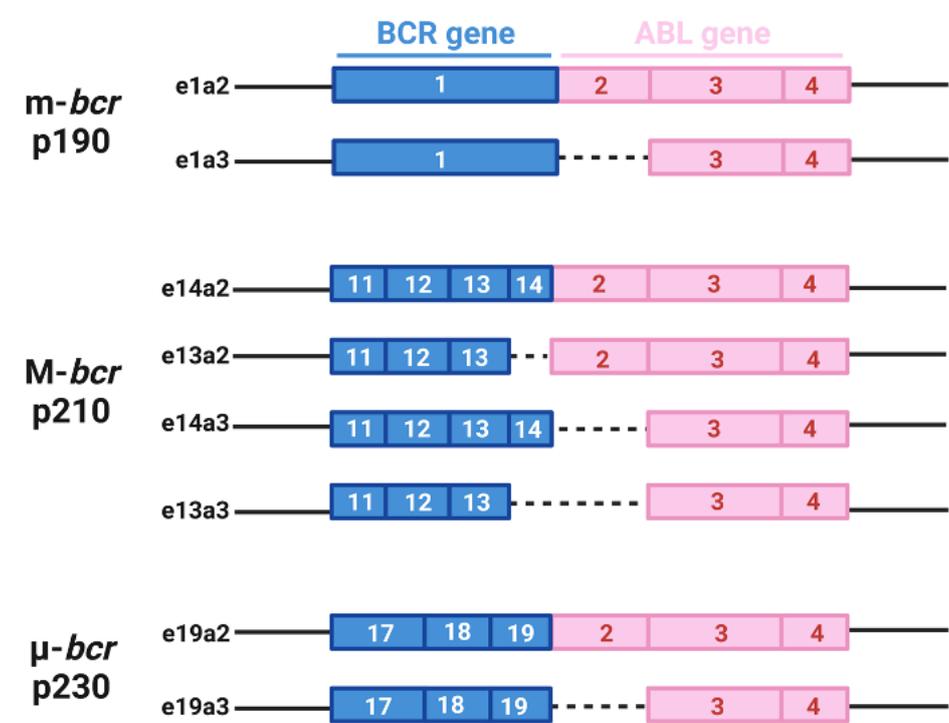


Figure 1.6: Fusion transcripts arising from p190, p210 and p230 BCR::ABL1 breakpoint regions. Alternate breakpoint regions in the BCR gene can result in at least 8 different fusion transcript variants due to alternative splicing in the ABL gene (splicing to exon 2 or 3) and M-bcr consisting of two intronic regions (intron 13 and 14).
(Created with Biorender.com)

The effect of BCR::ABL1 kinase activity can be seen in the alteration of a variety of downstream signalling pathways, displaying the numerous cell functions that are functionally impacted by BCR::ABL1. The complex signal transduction pathway activated by BCR::ABL1 kinase activity is also responsible for the chemotherapy resistance commonly observed in Ph+ALL and CML

(Amarante-Mendes et al., 2022). BCR::ABL1 activity is directed by the domains present in the fusion protein. The BCR section of the protein is involved in dimerization which allows the SH1 (SRC-homology domain 1) kinase domain of ABL to phosphorylate key tyrosine residues, enabling signal molecules to bind via their SH2 domains (Fig 1.7). Constitutive BCR::ABL1 activity is the result of the coiled-coiled domain at the N-terminus of the BCR section of the fusion protein, in addition, this domain is also responsible for oligomerisation. Other domains found within the BCR sequence includes a serine/threonine kinase (STK) domain, Ras homolog gene family/guanine nucleotide exchange factor (Rho/GEF) domain and adaptor molecule-binding domain SH2 which is able to interact with growth factor receptor-bound protein 2 (GRB2).

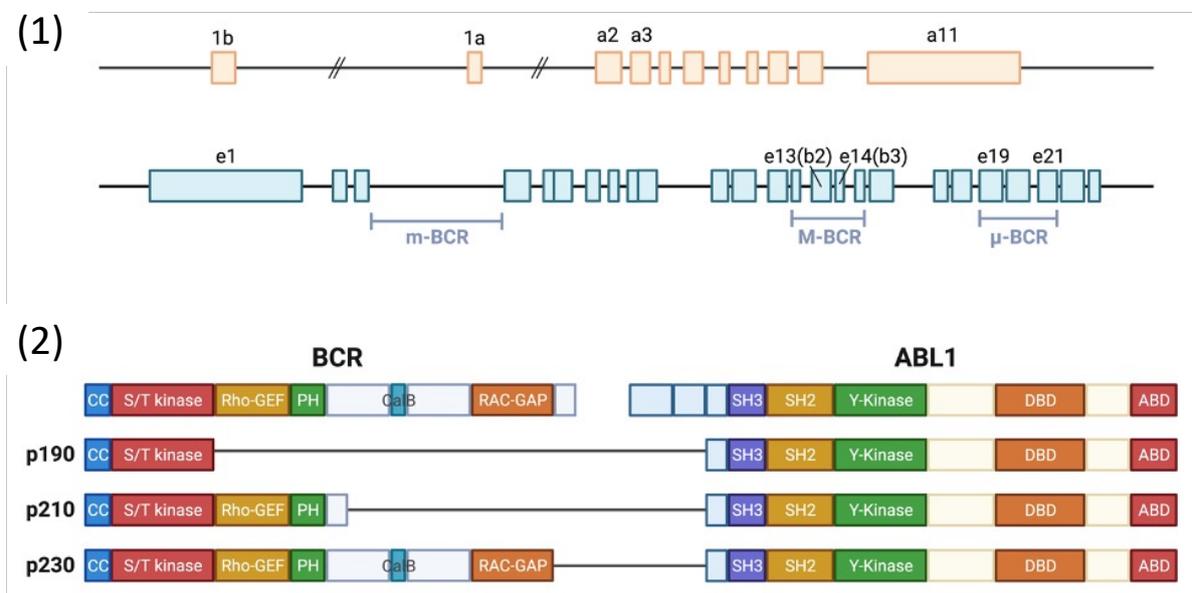


Figure 1.7: Schematic representation of *BCR* and *ABL1* gene structures and protein domains of BCR::ABL1 fusion proteins.

(1) In *ABL1*, breakpoints are distributed in the intron between exons 1b and 1a, or in the intron between exons 1a and 2. In *BCR*, most breakpoints in CML occur within the *M-BCR* region which encompasses exon 12-15. The *m-BCR* region is located in the 3' half of the first *BCR* intron. The *μ-BCR* region is located further downstream between exons 19 and 21.

(2) All three main BCR::ABL1 fusion proteins contain the common *ABL1* domains, including the SRC homology domains *SH2* and *SH3*, tyrosine kinase domain *SH1* and DNA binding domains (*DBD* and *ABD*). The fusion proteins may contain some or all of the following *BCR* domains: a coiled-coil (*CC*) oligomerisation domain, a serine/threonine kinase (*S/T-kinase*) domain, a Rho guanine nucleotide exchange factor (*Rho-GEF*) domain, a pleckstrin homology (*PH*) domain, a calcium-dependent lipid-binding domain (*Cal-B*), and a truncated *RAC-GAP* domain, depending on the *BCR* breakpoints.

(Created with Biorender.com)

The tyrosine kinase c-ABL is most commonly found in the nuclei, however, may exert actin filament binding in the cytoplasm and lamellipodial spreading at the immune synapse (Huang et al., 2008). C-ABL is conserved in the BCR::ABL1 complex, allowing its various domains to exert their functions such as the SRC-homology domains; SH2, SH3, actin transcriptome binding (AB) domain, DNA binding (DB) domain, a nuclear translocation signal (NTS) sequence, sites of phosphorylation by protein kinase C (PKC) and the highly conserved SH1 kinase domain which includes the catalytic site essential for the initiation of cellular transformation pathways, the alteration of which results in dysregulated proliferation and apoptosis resistance (Amarante-Mendes et al., 2022).

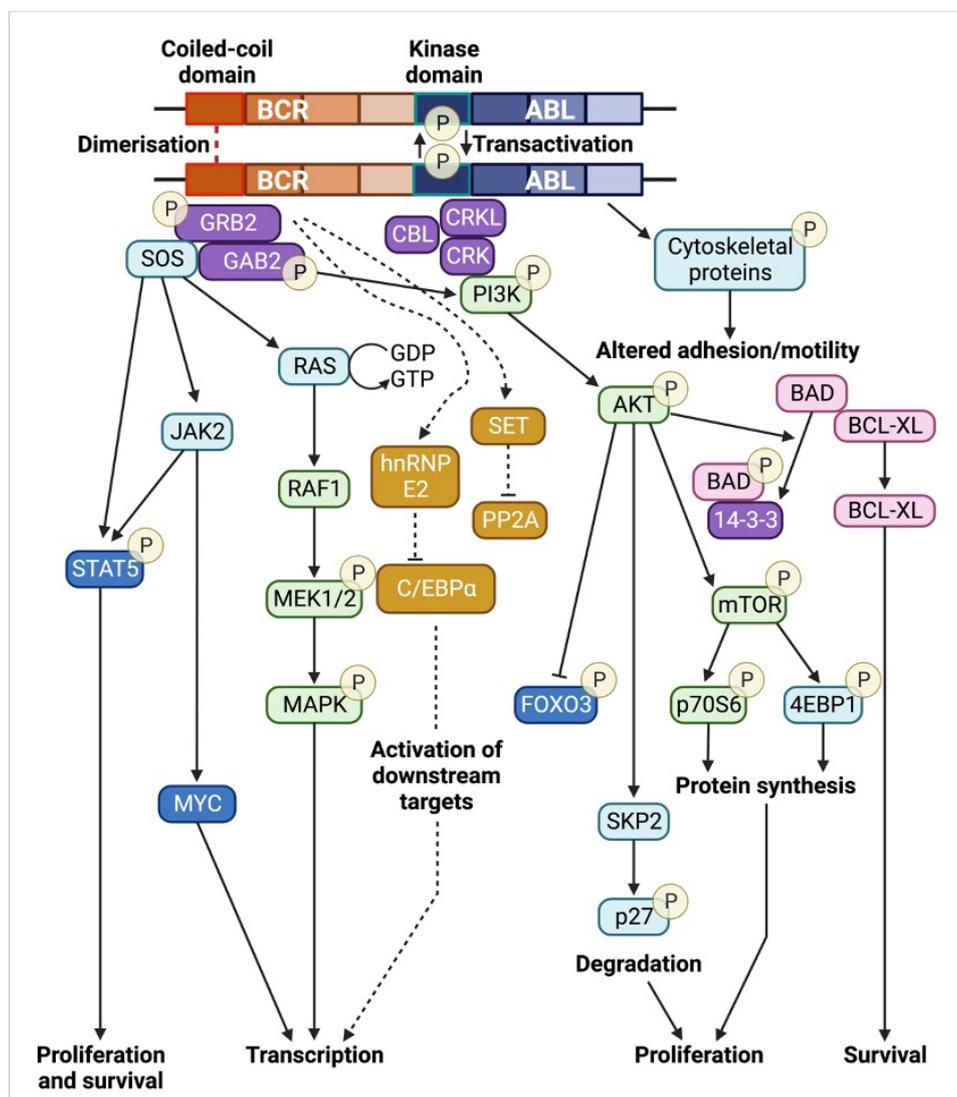


Figure 1.8: BCR::ABL1 signalling network.

Simplified schematic of downstream signalling pathways activated by BCR::ABL1. BCR::ABL1 dimerization activates the kinase domain via autophosphorylation and generates docking sites for intermediary adaptor proteins (purple). BCR::ABL1-dependent signalling activates a number of downstream signalling pathways resulting in enhanced survival, apoptosis inhibition, and perturbation of cell migration and adhesion.

Constituent transcription factors = blue, serine/threonine-specific kinases = green, and apoptosis-related proteins = red. Pathways implicated in CML LSC maintenance and BCR::ABL1-mediated disease transformation = orange.

(Adapted from O'Hare et al., 2011. Created with Biorender.com)

The number and variety of domains contained within BCR::ABL1 displays its structural complexity and hence, many proteins have been found to directly interact with BCR::ABL1 (Fig 1.8). Proteins such as GRB, SHC1, P13K and SHIP-2 have been identified as members of the BCR::ABL1 interactome through co-immunoprecipitation and mass spectrometry (Cutler et al., 2020). Further complexity is added by the recruitment of intermediate and/or effector molecules by the forementioned proteins and other interactors. By this mechanism, a plethora of signalling pathways may be activated including RAS/RAF/MAPK, P13K/AKT/mTOR and WNT/beta-catenin (Steelman et al., 2011). The effects of activation of these signalling pathways culminate in the different aspects of BCR::ABL1-induced cellular transformation. These include cell proliferation, differentiation and survival, induction of autocrine cytokines, prevention of growth factor-induced apoptosis, activation of the mTOR pathway, cytoskeleton polymerisation, FOXO-induced leukaemogenesis, drug resistance, growth factor independence and prevention of homeostatic antitumour activities such as the inactivation and degradation of BCR::ABL1 via ubiquitination by PP2A activation (Amarante-Mendes et al., 2022).

1.7 Introduction: Targeting of BCR::ABL1 by tyrosine kinase inhibitors (TKIs).

In order to directly target the tyrosine kinase activity of BCR::ABL1, a class of drugs able to directly target the constitutively activated oncoprotein was developed; tyrosine kinase inhibitors (TKIs). The clinical utility of TKI drugs was first demonstrated in chronic myeloid leukaemia, where the use and development of first, second and third generation TKIs represent a paradigm within precision medicine. Prior to the advent of TKIs, Ph+ALL carried a poor prognosis with rare durable remissions (Fielding et al., 2009). At this time, the best long-term survival rates resulted from allogenic HSCT, but were still only 40-50%. Durable remission rates at the time for patients treated with multiagent cytotoxic chemotherapy were 0-20%, thus displaying the unmet clinical need that would later be answered with TKIs and novel combination therapeutic approaches.

Early trials utilising the first-generation TKI Imatinib (UKALLXII) did so with the goal of enhancing the long-term outcomes in Ph+ALL and increasing the number of adult patients able to achieve successful HSCT (Fielding et al., 2014). The effect of imatinib was assessed by the inclusion of the TKI in frontline chemotherapeutic therapy and evidenced by an improvement in CR rates (92-82%, $P=0.004$) and 4-year overall survival rates (38-22%, $P=0.0003$). Subsequent trials investigating the addition of second generation TKIs dasatinib and nilotinib to frontline chemotherapy regimens displayed improved survival rates compared to similar treatment regimens with chemotherapeutic backbones (Kim et al., 2015, Ravandi et al., 2015 and Ravandi et al., 2016).

The improvement in clinical outcomes resulting from the addition of TKIs to chemotherapeutic treatment schema has also been confirmed in the paediatric cohort (Biondi et al., 2012, Biondi et al., 2018 & Shen et al., 2020). This phase III randomised clinical trial demonstrated an increased event-free survival rate resulting from the combination of second-generation TKI (dasatinib) with chemotherapy over first-generation imatinib (dasatinib 71% : imatinib 49% $P=0.005$). In addition, this trial demonstrated that intensive chemotherapy including dasatinib provides excellent control of central nervous system leukaemia, sparing patients from prophylactic cranial irradiation which risks toxic side effects such as secondary cancers, cognitive deficits and endocrinopathy as previously discussed (Pui et al., 2009).

Though the inclusion of TKI in treatment schema has become the gold standard in Ph+ALL treatment, tyrosine kinase inhibitors are not always a silver bullet (Leoni and Biondi., 2015). In addition to the previously discussed aberrant tyrosine kinase activity, BCR::ABL1 activity supports oxidative damage to DNA, impairment of genetic surveillance and bias toward error-prone DNA repair pathways (Senapati and Saaski., 2022). The outcome of this is genomic instability can lead to mutations in the *BCR::ABL1* gene itself. Point mutations to the *BCR::ABL1* kinase domain (KD) (thus impairing TKI binding) has been identified as a major mechanism of acquired TKI resistance and treatment failure (Pfeifer et al., 2007). Rapid onset of TKI resistance or refractory status from onset in TKI-naïve patients suggests that such mutated clones can exist prior to TKI treatment, appearing to arise after treatment due to resistant clones being selected for during treatment (Hofmann et al., 2003 and Shah et al., 2002). With improved sequencing accuracy, specific *BCR::ABL1* clones have been identified and can now be targeted by next-generation TKIs. An example of this is ABL1 kinase domain mutation, T315I, which can now be targeted by third-generation ponatinib. The addition of ponatinib to the HyperCVAD backbone (hyper-fractionated cyclophosphamide, vincristine, doxorubicin and dexamethasone) resulted in a superior event-free survival and overall survival response compared to treatment with dasatinib and HyperCVAD (Saskai et al., 2016).

TKI resistance can occur within Ph+ALL, but the data associated with it is often extrapolated from CML. For example, BCR::ABL1 dependent mechanisms of resistance can occur because of the amplification of the *BCR::ABL1* gene or the further development of point mutations. The outcomes of these can be the alteration of the *BCR::ABL1* kinase domain, rendering the TKI ineffective. Over 90 point mutations have been identified to date, however few have been biologically characterised (Azevedo et al, 2017; Cavalier et al, 2015; Gibbons et al, 2014; Khorashad et al, 2013; Shah et al, 2002; Szankasi et al, 2016). Such point mutations can be categorised into four different clusters depending on the *BCR::ABL1* domain affected by the mutation (Table 1.4 and Fig 1.9) (the P-loop, SH2 domain, SH3 domain or the activation loop) (Branford et al, 2003; Schindler et al, 2000; Soverini et al, 2011). As TKI generations have increased the variety of domains targetable therapeutically, previously untreatable point mutations can now be targeted. An example of this being the T315I mutation which is characterised by the replacement of threonine by isoleucine at ABL amino acid position 315 which was un-targetable until the development of ponatinib. With a large number and variety

of point mutations, not all mutations result in relapse or disease progression and therefore possess less clinical significance in the context of Ph+ALL and CML.

Mutation	Action
T315I	Consider ponatinib
T315A F317L/V/I/C Y253H	Consider nilotinib or bosutinib (rather than dasatinib)
F359V/C/I	Consider dasatinib or bosutinib (rather than nilotinib)
V299L	Consider nilotinib
E255K/V	Consider dasatinib
Any other mutation	Consider dasatinib, nilotinib, bosutinib, ponatinib

Table 1.4: Point mutation and recommended action for TKI choice.
(Adapted from Ai & Tu., 2014)

Additional mechanisms may also confer TKI resistance such as an increase in expression of P-glycoprotein efflux pump resulting in the transport of TKIs out of the target cell. Again, this has been highlighted within CML, rather than Ph+ALL. This was postulated in the CML-blast phase (BP) cell line K562 which demonstrated an increase in P-glycoprotein efflux pump expression when passaged with increasing doses of the TKI imatinib (Mahon et al., 2000). However, this cell line also demonstrated an amplification in *BCR::ABL1* copy number to 6x untreated levels. Hence, no discrimination between P-glycoprotein overexpression and imatinib resistance could be made from this investigation.

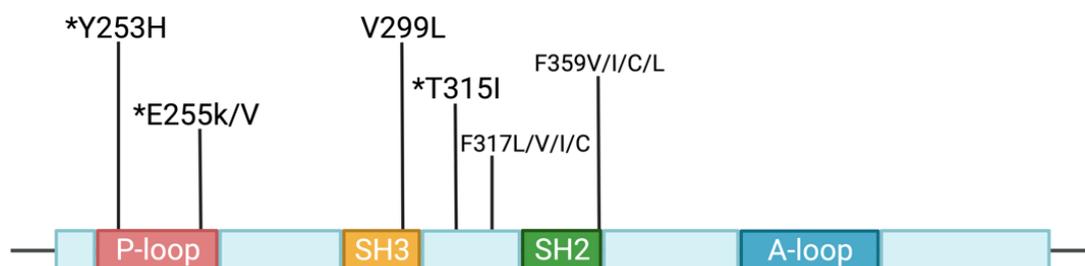


Figure 1.9: Schematic of BCR::ABL1 point mutation according to position. Schematic of point mutations associated with clinically relevant TKI choice. Asterisks represent amino acid substitutions associated with advanced disease stage.
(Created with Biorender.com)

A number of BCR::ABL1-independent mechanisms of TKI resistance have been described, displaying the importance of the alteration of intracellular signalling, cell cycle regulation and the alteration of signalling pathways by the activation of multiple non-BCR::ABL1 kinases (Li and Li., 20017). As previously discussed, the constitutively active BCR::ABL1 protein aberrantly

activates signalling pathways such as SRC kinase, RAS and JAK-STAT, leading to blast cell proliferation and apoptosis resistance (Cortez et al, 1997; Gallipoli et al, 2014; Pendergast et al, 1993). SRC family kinase activation has been demonstrated to enhance disease progression and poor response to TKIs. BCR::ABL1 is able to interact with SRC family kinases in a mutual activating manner with a conformational change to the SH2 and SH3 domains of ABL and the activation of SRC family kinases Hck, Lyn and Fyn, leading to enhanced cell proliferation, differentiation and survival (Danhauser-Riedl et al, 1996; Hu et al, 2006; Meyn et al, 2006; Stanglmaier et al, 2003). Despite no clinical examples of SRC-activating mutations in Ph+ leukaemia clinical samples or TKI-resistant cell lines, the SRC pathway is part of numerous crosstalk pathways and likely may be activated through other mechanisms (Donato et al, 2003; Ptasznik et al, 2004; Wu et al., 2008).

Pathway activation through alternative mechanisms has been demonstrated by the activation of RAS signalling by Grb2-mediated binding of the Y177 moiety in the BCR sequence, however, the specific mitogen-activated kinases activated by RAS in Ph+ALL and CML are yet to be fully elucidated (Cortez et al, 1997; Pendergast et al, 1993). Grb2 has been demonstrated to recruit Gab-2 and subsequently activate PI3K and ERK pathways, the mechanism of this activation being through the activity of plasma membrane transporter molecules such as the ABC family of transporters, ABCG2 and MDR-1 (Sattler et al., 2002). As previously discussed, the mechanism of TKI resistance conferred by such protein complexes is believed to be through the transportation of TKIs out of the target cell and therefore, a decrease in intracellular TKI concentration, however this mechanism is yet to be fully elucidated (Mahon et al., 2003).

As discussed, mechanisms causing failure to respond or resistance to TKI therapy can be BCR::ABL1-dependent or independent. Mechanisms of TKI resistance and activation of alternative signalling pathways have been an area for direct therapeutic targeting however, the phenomenon of disease persistence remains, suggesting that BCR::ABL1-independent pro-survival mechanisms are being exploited by quiescent leukemic stem cells (LSCs) in the context of CML (Bhatia et al, 2003b; Chomel et al, 2011; Deininger, 2012). Within Ph+ALL, the LSC has yet to be fully elucidated, but remains an area of research interest. Treatment of Ph+ lymphoid leukaemia (chemotherapy with TKI) is directed against the lymphoid clone; however, relapse may arise from residual Ph+ LSC populations. These LSC populations appear

to reside in HSC or MPP population (Carrol and Clair., 2018), thus it is likely that their biology resembles CML LSCs and may include TKI-resistant LSCs just as in CML. Although HSCT may eradicate LSCs, the high morbidity and mortality from this procedure mandates development of novel therapies. Key features of these TKI resistant CML LSC populations include reliance on transcriptional hubs regulating p53 and c-myc activity, epigenetic dysregulation and upregulation of autophagy in addition to a reliance on mitochondrial oxidative phosphorylation (Patel et al., 2018). Understanding these mechanisms will enable combinational therapies to be developed to target BCR::ABL dependent and independent mechanisms involved in LSC persistence making them vulnerable to repurposed drugs from a variety of classes such as antibiotics, antipsychotics, NSAIDs (non-steroidal anti-inflammatory drugs) and anti-depressants (Holyoake and Vetrie., 2017).

1.8 Introduction: The identification of a novel Ph+ALL subtype with CML-like features.

As previously discussed, ALL comprises a heterogeneous group of acute lymphoid-driven haematological malignancies. Historically, ALL was subtyped based on cellular morphology after staining and was subclassified into L1, L2 and L3 based on the presence of small-medium regular blast cells with a high nucleocytoplasmic ratio, larger pleomorphic blasts with visible nucleoli, nuclear clefts and larger cytoplasm, or blast cells with a visibly basophilic cytoplasm and cytoplasmic vacuoles respectively (Bain and Estcourt., 2013). Since then, the classification system has been updated and modernised to align with technological advancements, the increasing understanding of the heterogeneity of ALL and the discovery of distinct leukemic subtypes (Table 1.1).

The subtyping of ALL has allowed for a marked improvement in diagnosis and treatment of these haematological malignancies over the previous decades. Importantly, this has resulted in improved outcomes for patients and an increased survival rate up to 65% across subtypes with differences in survival by age (the highest mortality rate being in patients above age 50) (CRUK., 2023). However, there remains a number of patients who display atypical leukemic phenotypes. These patients may display an unpredicted response to treatment, misalignment in MRD monitoring outcomes or exhibit the involvement of leukocytes from a number of lineages in disease progression. Patients who display an atypical ALL phenotype can therefore be difficult to treat, as the understanding of disease pathology on which treatment is based is unsuitable for an unknown or undefined subtype. Hence, this issue can be seen as an unmet clinical need and necessitates further study in order to ensure correct treatment and diagnosis of patients.

In this vein, the recent identification of a group of paediatric Ph+ALL patients with atypical disease features discussed in Hovorkova et al., (2017) is at the core of this project. The study was intended to focus on comparison of different methods of MRD monitoring, directly comparing the sensitivity and accuracy of such treatment response measurements. The approach of this was to utilise the genomic breakpoint between *BCR* and *ABL1* genes for the DNA-based monitoring of MRD in comparison to standard MRD methods which are based on

immunoglobulin/T-cell receptor (Ig/TCR) gene rearrangements and *IKZF1* deletion. Broadly, there was good correlation between the newer, more sensitive, DNA based method of MRD monitoring and the standard methods used clinically. However, >20% paediatric patients of the overall Ph+ALL cohort displayed discordance between MRD methods (Fig 1.10). These patients had significantly higher levels of *BCR::ABL1* fusion than detectable Ig/TCR rearrangements or *IKZF1* deletion. As the methods of MRD monitoring discussed are intended to identify the presence of lymphoblasts after treatment, if the understanding of *BCR::ABL1* fusion occurring only in lymphoid progenitors in Ph+ALL is applicable here, such outlying patients must therefore possess cells which contain the Philadelphia chromosome outside of the lymphoid progenitor fraction and hence, do not have standard Ph+ALL.

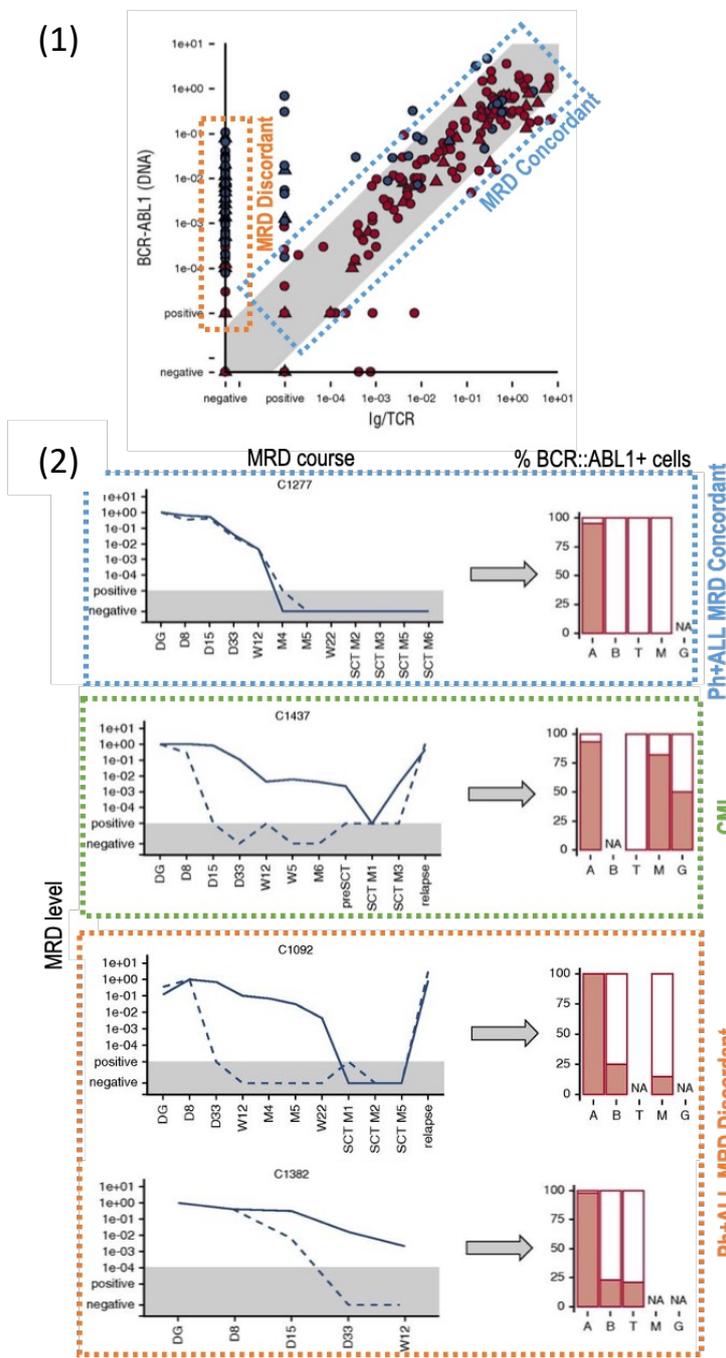


Figure 1.10: Monitoring childhood ALL using *BCR::ABL1* genomic breakpoints and Ig/TCR gene rearrangement identifies a discordant subgroup with CML-like biology.

(1) Comparison of the MRD levels in ALL patients measured by Ig/TCR vs *BCR::ABL1* transcript qualification as well as vs *BCR::ABL1* genomic breakpoint quantification in all samples and separately in selected time points during frontline treatment. Samples from patients with major *BCR::ABL1* fusion variant=triangles, and minor *BCR::ABL1*=circles. Samples from patients with concordant MRD course=red, and samples from patients with discordant MRD=blue. The light grey diagonal shape represents the area of concordance +/- log. D= day, W= week from start of treatment.

(2) Presence of *BCR::ABL1*-positive cells in haematopoietic lineages (A= ALL blasts, B= non-ALL B cells, G= granulocytes, M= myeloid cells/monocytes, T= T cells) at diagnosis in ALL patients with concordant (C1277) and discordant (C1092 & C1382) MRD courses and in CML patient (C1437). The MRD levels are shown for Ig/TCR (dashed line) and *BCR::ABL1* genomic breakpoint (full line); the grey area represents the level of sensitivity of the Ig/TCR quantification. NA= not available. (Source: Hovorkova et al., 2017)

As a result of these observations, *BCR::ABL1* presence was assessed in a variety of haematopoietic lineages. In both concordant (standard/'classical' Ph+ALL) and discordant (atypical Ph+ALL) samples, ALL blasts, non-ALL B cells, granulocytes, monocytes/myeloid cells and T cells were assessed for *BCR::ABL1* positivity by the aforementioned DNA-based method. It was observed that concordant patients displayed the t(9;22) event in ALL blast cells only, aligning with the expected standard Ph+ALL phenotype. Interestingly, the discordant patient cohort displayed *BCR::ABL1* positivity in all haematopoietic populations tested. This aligns with the MRD discordance previously observed and displays an atypical and currently undefined Ph+ALL subtype.

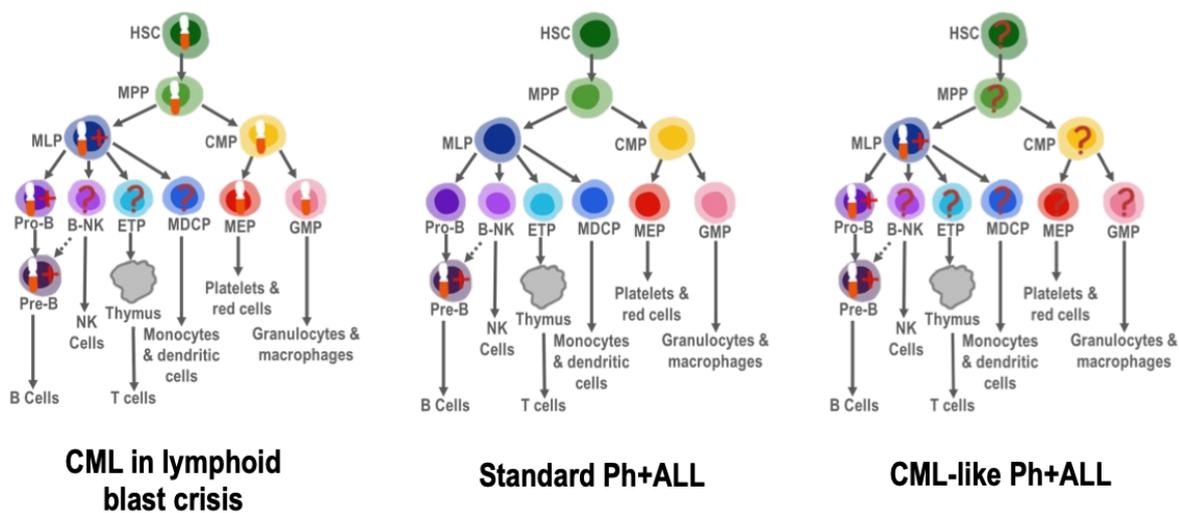


Figure 1.11: *BCR::ABL1* fusion during early stages of haematopoiesis in three disease states. Known cell of origin of *BCR::ABL1* fusion in CML (HSC). Postulated cell of *BCR::ABL1* origin in 'standard' Ph+ALL (ProB/PreB cell) and CML-like Ph+ALL (HSC or MPP). HSC= haematopoietic stem cell, MPP= multipotent progenitor, MLP= mixed lymphoid progenitor, CMP= common myeloid progenitor, MEP= megakaryocyte/erythrocyte progenitor, GMP= granulocyte-macrophage progenitor, NK= natural killer, MDCP= monocyte-dendritic cell precursor, ETP= early T cell precursor.

The outcome of these findings is the proposal of a novel Ph+ALL subtype, currently designated 'CML-like Ph+ALL', named as such due to the presence of *BCR::ABL1* in multiple haematopoietic lineages, just as in CML. CML in blast crisis can involve a number of lineages; myeloid, lymphoid or mixed lineage. This potential for multi lineage involvement arises due to the early developmental stage of the progenitor cell in which the *BCR::ABL1* fusion event occurs. The presence of Ph+ HSCs, a cell type at the apex of haematopoietic development, ensures that proceeding developing cells also contain the fusion chromosome. *BCR::ABL1* occurring in lymphoid progenitors alone results in its tyrosine kinase activity being lineage restricted.

Existence of CML-like Ph+ALL has been proposed but the subtype remains undefined (Hunger., 2017). Limited information on the potential subtype has meant that diagnostic criteria and the functional basis of the disease has not yet been established. However, the available information indicates that the BCR::ABL1 fusion event, the leukemic origin, must occur in cells prior to lineage commitment in order for the Philadelphia chromosome to be detectable in both myeloid and lymphoid lineages (Fig 1.11). Applying what is understood about CML initiation and development, the most likely cell of origin for BCR::ABL1 would either be HSCs or MPPs. Both populations are characterised as being multipotent and non-committed (Fig 1.12) (Pietras et al., 2015). While the presence and activity of LSCs has been confirmed in CML, it is unknown if Ph+ HSCs/MPPs in CML-like Ph+ALL would display the same phenotype as CML LSCs, as it would be pertinent to identify the presence of the Philadelphia chromosome in such Ph+ALL stem and progenitor cells before their potential LSC activity could be assessed (Thomas., 2012). The similarities between CML and CML-like Ph+ALL are undefined however, the identification of the same fusion event, occurring in the same cell population and yet, resulting in different diseases, may provide an interesting example of the importance of downstream signalling and transcriptional differences for establishing different disease pathologies.

The classical model of haematopoiesis has long held the dogma of a rigid developmental hierarchy with HSCs at the apex giving rise to all lineages of blood cells and possessing self-renewal capacity (Cheng et al., 2020). Investigation into single cell activity through the recent development of novel and highly sensitive technologies has challenged this model. The newer haematopoietic model takes into account the phenotypic complexity of the haematopoietic stem and progenitor cell (HSPC) populations. Examination of single cell activity reveals a heterogeneous pool of HSPCs with lineage biases. Such cells still possess self-renewal capabilities and despite being primed for differentiation into specific myeloid or lymphoid cells, are still negative for lineage markers. With this revised model in mind, BCR::ABL1 fusion in Ph+ALL or CML-like Ph+ALL may arise in a subset of lineage-primed HSPCs rather than seemingly occurring at random in a homogenous pool of HSCs. The heterogeneity of lineage involvement in CML-like Ph+ALL and CML-LBP(chronic myeloid leukaemia in lymphoid blast phase) may be the result of lineage priming, with the t(9;22) translocation event occurring in HSPC populations primed for myeloid and lymphoid populations. Therefore, the reason why

standard Ph+ALL is a lymphoid-restricted disease may be due to the formation of BCR::ABL1 in HSPCs primed for lymphoid lineages only. Rather than maintaining the older view of Ph+ALL which surmises that BCR::ABL1 fusion in Ph+ALL occurs in lymphoid progenitors and hence, only the CML-like subtype would possess Ph+ HSPCs, integrating the new model of haematopoiesis would better explain Philadelphia chromosome positivity in both standard and CML-like Ph+ALL (Fig 1.11).

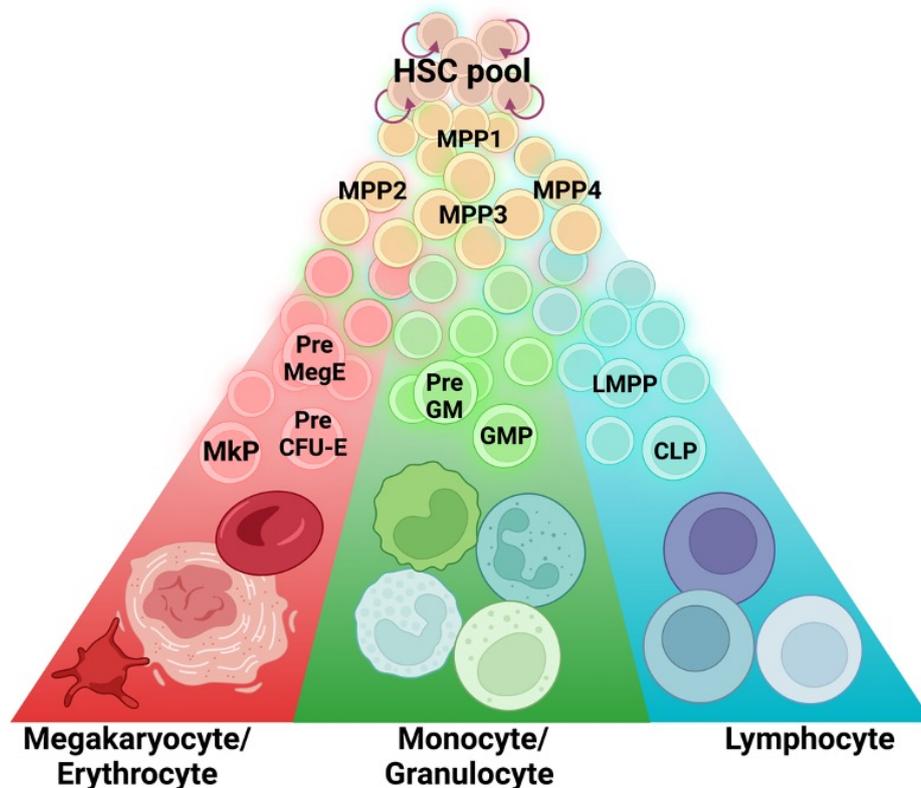


Figure 1.12: Updated model of haematopoiesis at steady state. Cell fate decisions are made during haematopoietic development and differentiation into mature cell populations. Subgroups of MPPs display lineage-specific predetermination (MPP1, 2, 3 and 4).
(Created with Biorender.com)

1.9.1 Introduction: Characterisation of CML-like Ph+ALL patients; what is known currently and how it may align with current Ph+ALL risk stratification.

Risk stratification is an essential aspect to treatment of paediatric Ph+ALL. Utilisation of disease characteristics allow for patient response to treatment to be predicted, ensuring patients who are likely to have a good clinical outcome are able to be treated with less intensive regimens and spared potentially toxic treatments, and patients with a lower probability of long-term survival given a more intensive therapy to increase their chance of cure.

In order to establish risk stratification, a number of clinical characteristics are collected including age at diagnosis, WCC at diagnosis, CNS or testicular involvement at diagnosis, occurrence of trisomy 21, sex, race and ethnicity, weight at diagnosis and during treatment. In addition, flow cytometry-based immunophenotyping can be utilised to assess the leukocyte population size. Such assessment tools may help to establish the disease phenotype at diagnosis, thus advising on the treatment protocol required for an individual patient.

Assessment of treatment response through MRD, molecular and cytogenetic responses inform on the requirement for dose escalation, de-escalation or the discontinuation of treatment. In Ph+ALL, the key times for assessment of BM MRD is day 7 of treatment, day 14, the end of induction and the end of consolidation. Peripheral blood response to steroid prophase, multiagent induction therapy, MRD before the end of induction (day 8 and 15) and any presence of leukaemia at the end of induction (induction failure) is also monitored for risk stratification.

Monitoring of treatment response allows patients to be categorised by risk status and therefore, receive appropriate treatment. However, this multi-factorial monitoring strategy also identifies patients belonging to different disease subtypes. As the identification of leukemic subtypes has been enhanced through characterisation of cellular pathway perturbation and unanticipated response to treatment, the need for rapid and accurate identification of these patients is clear (Hunger and Mullighan., 2015). Uncharacterised subtypes with no firm diagnostic criteria such as CML-like Ph+ALL represent an unmet clinical

need as these patients have been reported to respond poorly to commonly used lymphoid-directed therapies and therefore, may require an alternate therapeutic regimen (Hovorkova et al., 2017). Without specific diagnostic criteria, these patients may remain unclassified and hence, utilisation of clinical data may be beneficial in detection of those with the CML-like subtype. It is likely that these patients would have clinical characteristics atypical to the rest of the standard Ph+ALL cohort and therefore may be visible in already established disease monitoring methods. Therefore, the integration of clinical, phenotypic, and transcriptional data is essential for the characterisation of CML-like Ph+ALL.

As the proposed CML-like Ph+ALL subtype is as of yet uncharacterised, an integrated research approach must be utilised in order to fully investigate phenotypes unique to this subtype. In order to do this, immunophenotypic data, clinical information and cytogenetic analysis of progenitor cells were supplemented with gene expression analysis by RNAseq (RNA sequencing). The use of RNAseq has been revolutionary in researching complex or undefined disease states. Through the analysis of transcriptional differences between samples within a cohort, predictive modelling of disease for the identification of candidate biomarkers can be achieved (Paul et al., 2016). The emergence of this approach supports the increased utilisation of precision medicine, allowing clinicians and scientists to tailor treatment to patient subgroups according to susceptibility to specific diseases or outcomes and therapeutic response (Wang and Wang., 2023). As broad disease traits for CML-like Ph+ALL have been identified as discordance between MRD measurement techniques, poor response to traditional Ph+ALL therapy and myeloid disease involvement, RNAseq provides the opportunity to identify differential gene expression patterns which lead to such complex disease traits (Hovorkova et al., 2017). Whole transcriptome analysis has become an increasingly utilised and invaluable tool of investigating the pathology of uncharacterised diseases in an unbiased manner (Costa et al., 2013). This has already been utilised in the context of B-ALL where the genomic landscape where a multi-omic (whole genome, exome and transcriptome) approach was utilised to develop a transcriptional taxonomy and identify mutations with prognostic significance across paediatric ALL subtypes (Brady et al., 2023).

We hypothesise that a subtype of Ph+ALL exists with CML-like disease features and developmental origins arising from BCR::ABL1 fusion occurring in early stem/progenitor cells. We propose that these patients differ from the standard Ph+ALL phenotype by features including (but not limited to) an enlarged myeloid population compared to the standard Ph+ALL phenotype and will have evidence of BCR::ABL1 fusion in early stem and progenitor cells akin to CML and a transcriptional profile containing genes commonly found in both Ph+ALL and CML. We believe that due to this, these patients will require an alternative therapeutic regimen to what is currently used in Ph+ALL.

To elucidate these hypotheses, the following aims have been generated:

1. Determine if haematopoietic dynamics differ within Ph+ALL and if differences in cell population sizes can identify distinct subgroups.
2. Identify where during haematopoietic development BCR::ABL1 fusion occurs.
3. Examine the heterogeneity of transcriptional profiles within Ph+ALL and determine if subtypes can be identified by differential gene expression.

We intend to address these aims with the following objectives:

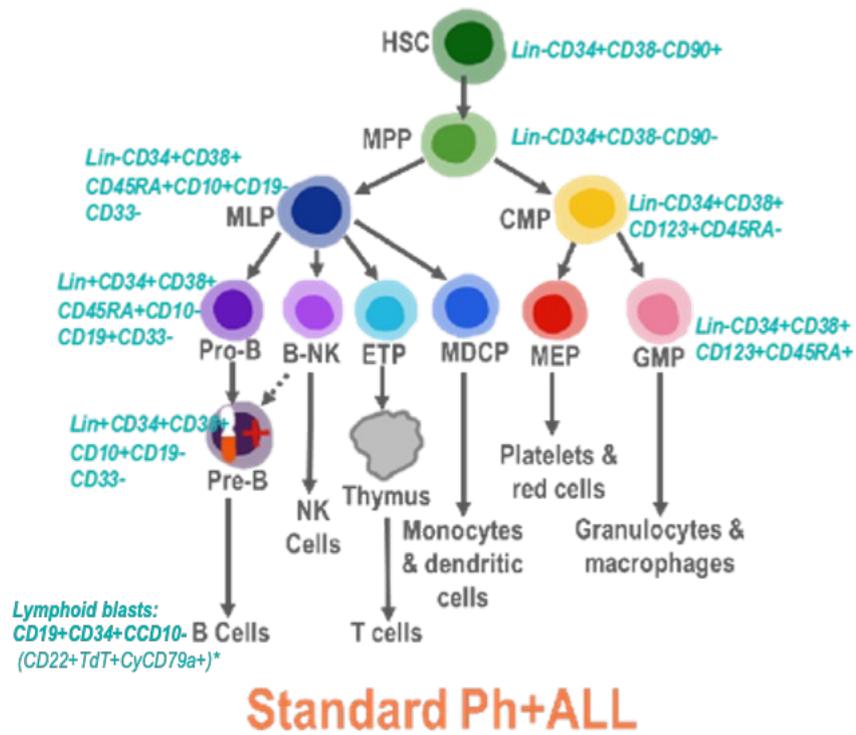
1. Assess haematopoietic progenitor population size in paediatric Ph+ALL diagnostic and post-induction bone marrow samples by FACS.
2. Examine sorted HSCs and MPPs for BCR::ABL1 fusion using FISH.
3. Investigate transcriptional profiles of Ph+ALL and CML bone marrow samples using RNAseq.

1.9.2 Introduction: Aim1: Determine if haematopoietic dynamics differ within Ph+ALL and if differences in cell population sizes can identify distinct subgroups.

As previously reported by Hovorkova et al (2017), patients identified as belonging to the CML-like subtype were observed to have an enlarged myeloid population compared to 'standard' Ph+ALL. Mature cell populations from multiple lineages (non-ALL blasts, myeloid cells, monocytes and granulocytes) were examined for BCR::ABL1 in said study however, progenitor cells and cell population size was not assessed. We hypothesise that progenitor population sizes vary within the Ph+ALL cohort and samples belonging to the CML-like subtype will have enlarged myeloid progenitor pools compared to standard Ph+ALL. We believe this is due to t(9:22) occurring at the apex of haematopoiesis, resulting in BCR::ABL1 kinase activity in multiple lineages and therefore increased proliferation of myeloid cells in addition to lymphoid cells. As discussed in section 1.2.2, lymphoid differentiation arrest occurs in ALL, producing immature lymphoid blasts, hence investigation into population dynamics of the proposed CML-like subtype would include myeloid progenitors. Based on the expansion of myeloid cells in CML-like Ph+ALL by Hovorkova et al (2017), we intend to investigate if CML-like samples possess haematopoiesis dynamics atypical to standard Ph+ALL.

In order to address this aim our objective is to sort diagnostic and post-induction Ph+ALL bone marrow samples into progenitor populations using FACS and examine population sizes across the cohort, determining if a clear subgroup can be identified (Fig 1.13). An important feature of haematopoiesis that should be remembered for the examination of differences in cell population size between diagnostic and post induction samples throughout this thesis, is its plasticity. Haematopoiesis is a dynamic developmental process with the production of different cell types varying depending on the needs of the organism. Hence, BCR::ABL1 kinase activity in multiple lineages may be observed by differencing progenitor population sizes. The purpose of sorting and collecting cell populations in addition to measurement will be elaborated on in the proceeding section.

(1)



(2)

Population	Surface Marker Expression
HSC	LIN- CD34+ CD38- CD90+ CD93+/-
MPP	LIN- CD34+ CD38- CD90-
CMP	LIN- CD34+ CD38+ CD123+ CD45RA-
GMP	LIN- CD34+ CD38+ CD123+ CD45RA+
CLP	LIN- CD34hi/+ CD38+ CD45RA+ CD10+ CD19- CD33-
ProB	LIN+ CD34+ CD38+ CD45RA+ CD10- CD19+ CD33-
PreB	LIN+ CD34+ CD38+ CD10+ CD19+ CD33-

Figure 1.13: Sorting strategy (standard Ph+ALL).

(1) Bulk diagnostic or post induction Ph+ALL samples were sorted into progenitor populations based on cell surface marker expression. The same sorting strategy was used for Ph+ALL, CML-LBC and healthy BM control samples.

(2) Cell surface marker expression profile for early haematopoietic progenitors-tabular.

*clinical markers, not used in this project (Chiaretti et al., 2014)

(Created with Biorender.com)

1.9.3 Introduction: Aim 2: Identify where during haematopoietic development BCR::ABL1 fusion occurs.

As previously discussed, findings from Hovorkova et al (2017) identified BCR::ABL1 fusion in mature cell populations from multiple lineages however, this did not represent a definite identification of Ph⁺ stem/progenitor cells or where during the haematopoietic development process t(9:22) occurred. With this in mind, we have developed the hypothesis that t(9:22) fusion occurs prior to lineage bifurcation in the CML-like subtype. This would result in BCR::ABL1 kinase activity in both lymphoid and myeloid lineages, as observed by Horakova et al (2017).

To address this aim we intend to investigate BCR::ABL1 fusion with FISH in FACS sorted stem and progenitor cells. Isolation of cells expressing HSC and MPP surface markers by FACS would allow for the investigation of cells at the apex of haematopoietic development, prior to lineage bifurcation. Due to standard Ph+ALL being a lymphoid lineage-restricted disease, the cell of t(9:22) origin is believed to be a lymphoid progenitor, hence we believe that CML-like and standard Ph+ALL samples could be differentiated from each other by the presence of BCR::ABL1 in stem and progenitor cells in the CML-like Ph+ALL samples alone. However, HSCs and MPPs have not been studied previously in Ph+ALL and BCR::ABL1 may occur prior to lineage fixing in Ph+ALL with downstream factors (such as signalling and transcriptional alterations) affecting lineage presentation of BCR::ABL1 kinase activity.

1.9.4 Introduction: Aim 3: Examine the heterogeneity of transcriptional profiles within Ph+ALL and determine if subtypes can be identified by differential gene expression.

To investigate how the same fusion event (the formation of BCR::ABL1) could occur in the same cell population (HSCs) and yet lead to two different diseases, CML-LBC and CML-like Ph+ALL we have set the objective to analyse transcriptional differences within the Ph+ALL cohort and compare to CML-LBC by RNAseq. With the leukemic initiating events being identical in both disease states, it would be logical that the phenotypic differences arose from differential transcriptional profiles. Therefore, an aim of this project is to establish how Ph+ALL, CML-LBC and CML-like Ph+ALL differ in gene expression in order to better characterise the CML-like subtype.

As discussed above, an unbiased whole transcriptome analysis approach allows for the overall gene expression profile to be investigated, allowing for detection of genes and pathways previously unanticipated to play a role in disease pathology and may potentially be utilised as biomarkers. The objective is to perform a comprehensive literature search to identify key genes involved in CML pathology, LSC activity and myeloid biology (such as *CYYR1*, *ANXA1*, *CCDC60* and *ACE*) and use these to interrogate publicly available datasets and patient cohorts. Such genes were examined in the sequenced samples for Log2Fold change and statistical significance. Utilising genes which are important in myeloid leukaemia and LSC pathology but not known to play a role in Ph+ALL allows for the identification of transcriptional elements shared between CML and the CML-like Ph+ALL subtype. Due to the known characteristics of the subtype being of a mixed clinical phenotype, it is theorised that CML-like Ph+ALL would display a mixed transcriptional phenotype, with differential expression of genes involved in both CML and Ph+ALL pathology.

As there are no specific diagnostic criteria for identifying CML-like Ph+ALL, RNAseq was performed blind on a cohort of 11 paediatric Ph+ALL patients. Sample selection criteria prioritised RNA quality but also included a range of ages, clinical outcomes and samples with or without detectable BCR::ABL1 in stem and progenitor cells. Without knowing which features align with CML-like pathology, a range of Ph+ALL samples were sequenced to create

a varied and unbiased sample cohort. In addition, to establish a CML transcriptional profile to which Ph+ALL samples could be compared, 2 CML samples were sequenced. CML is exceedingly rare in children and adolescents therefore due to sample availability, adult CML BMNC samples were utilised. To reflect the fact that Ph+ALL and CML-LBC are two distinct disease entities despite both being acute phase leukaemia with lymphoid involvement, the CML samples selected were harvested from patients in lymphoid blast crisis. This would also allow CML-like Ph+ALL to be better characterised and confirm that despite being an acute phase leukaemia with both myeloid and lymphoid disease involvement, CML-like Ph+ALL is a separate disease entity and not a misdiagnosis of CML-LBC.

The detection of CML-like subtype-specific biomarkers would represent an important tool in diagnosis of the disease and, in addition, may highlight novel therapeutic targets. While flow cytometry-based immunophenotyping is frequently used in leukaemia diagnostic practices, sorting cells for analysis of gene expression is costly in a clinical setting. Therefore, identification of clinically relevant biomarkers in a bulk BMNC population would allow for such genes to be adopted more readily into existing diagnostic practices.

Chapter 2. Materials and methods.

2.1 Materials and methods: Reagents and equipment

Reagent	Company	Technique
PBS	University of Glasgow	TC
FBS	Corning	TC
BCR-ABL DC DF probe	Abbott	FISH
Antibodies	BDBiosciences	FACS
SFM	University of Glasgow	TC
Methanol	University of Glasgow	FISH
Acetic acid	University of Glasgow	FISH
Vectashield mounting medium	2B Scientific	FISH
SSC	University of Glasgow	FISH
Agarose	University of Glasgow	PCR
TAE	University of Glasgow	PCR
SYBR safe	Thermo Fisher Scientific	PCR
RNA protect	Qiagen	RNA extraction
Microscope oil	Zeiss	FISH
Nail polish	BaryM	FISH
Trypan blue	University of Glasgow	TC
Histopaque	Sigma	TC
Growth factors	University of Glasgow	TC
Virkon	University of Glasgow	TC
ddH ₂ O	University of Glasgow	TC
RNase free water	Qiagen	RNA extraction
RNase zap	Qiagen	RNA extraction
Primers (BCR-ABL)	Thermo Fisher Scientific	PCR
Primers (housekeeping)	Thermo Fisher Scientific	PCR
Ladder (gel electrophoresis)	NEB (New England Biolabs)	PCR
KCl	University of Glasgow	FISH
Flow cytometry beads	BDBiosciences	FACS

Table 2.1.1: List of reagents used for laboratory-based research. Reagents used within the laboratory and source companies were listed and annotated by techniques used. Where reagents from multiple companies were used, all company names were listed. Where reagents were made in the laboratory, company source is listed as 'University of Glasgow'.

Material	Company	Technique
QuantiTect kit	Qiagen	cDNA synthesis
Rneasy mini kit	Qiagen	RNA extraction
Dneasy blood and tissue kit	Qiagen	DNA extraction
Syringe (20mL)	BD Plastipak	TC
Syringe filter	Greiner	TC
40uM easy cell strainer	Greiner	TC
Mortar and pestle	SLS Ltd	TC
Pasteur/dropper pipette	Greiner	TC
Pipette (1000uL)	Eppendorf	TC
Pipette (200uL)	Eppendorf	TC
Pipette (100uL)	Eppendorf	TC
Pipette (20uL)	Eppendorf	TC
Pipette (10uL)	Eppendorf	TC
Pipette (2.5uL)	Eppendorf	TC
1.5mL tubes	Eppendorf	TC
Falcon tube (50mL)	Corning	TC
Falcon tube (10mL)	Corning	TC
Stripette (25mL)	Corning	TC
Stripette (10mL)	Corning	TC
Stripette (5mL)	Corning	TC
Pipette tips (1000uL)	Thermo Fisher Scientific	TC
Pipette tips (200uL)	Thermo Fisher Scientific	TC
Pipette tips (10uL)	Thermo Fisher Scientific	TC
T25 cell culture flask	Thermo Fisher Scientific	TC
Haemocytometer	SLS Ltd	TC
Multispot slides	Shandon	FISH
Coplin jar (glass)	SLS Ltd	FISH
Glass coverslip	Fisher Scientific	FISH
FACS collection tubes	Fisher Scientific	FACS

Table 2.1.2: List of materials used for laboratory-based research. Materials used for laboratory-based research and data generation were listed and annotated by source company and technique.

Equipment	Technique
Pipette gun	TC
pH meter	TC
Vortex	TC
Mega-centrifuge	TC
Microfuge	TC
Incubator (TC)	TC
Incubator (bugs)	FISH
Fluorescence microscope (Zeiss Axio Imager M1)	FISH
FACS machine (Aria)	FACS
Flow cytometer (canto)	FACS
PCR hood	PCR
Fume hood	TC
Cell culture hood	TC
Microscope	TC
Fridge	TC
Freezer (-20C)	TC
Freezer (-80C)	TC
Liquid nitrogen tank	TC
Autoclave	TC
FlowJo dongle	FACS
Powerpack	PCR
Electrophoresis chamber	PCR
LICOR imager	PCR

Table 2.1.3: List of technical equipment used for laboratory-based research. Equipment used within the project were listed and annotated with the source company and laboratory technique.

2.2 Materials and methods: Solutions

Immunofluorescence reagents	Supplier
100% ethanol	University of Glasgow
100% methanol	University of Glasgow
Acetic acid	University of Glasgow
Vectashield mounting media	2BScientific
KCl 0.75M	University of Glasgow
BCR::ABL1 Dual Colour Dual fusion FISH probe set	Abbot Diagnostics
Poly-L-Lysine	Invitrogen
Clear nail polish	Barry M Cosmetics
SSC	Invitrogen
TWEEN	Invitrogen

Table 2.2.1: Immunofluorescence reagents. Source suppliers indicated.

FISH fixative	Volume
Acetic acid	2mL
Methanol	9mL

Table 2.2.2: FISH fixative. For use in fixing and washing of cells. Made at the POGLRC.

Hypotonic KCl solution	Volume
KCl 0.75M	1mL
dH2O	10mL

Table 2.2.3: FISH KCl wash buffer. Made at the POGLRC.

Stock SSC buffer (x20 concentration) (20x SSC)	Volume
SSC	132g
dH2O	500mL
Pure HCl	Titrate to pH 5.3

Table 2.2.4: Stock saline sodium citrate (SSC) wash buffer. Made at the POGLRC.

FISH wash buffer 1 (2xSSC 0.1% NP40)	Volume
20x SSC	100mL
dH2O	899mL
NP40	1mL
Pure HCl	Titrate to pH7

Table 2.2.5: FISH SSC wash buffer 1. 2xSSC 0.1% NP40. Made at the POGLRC.

FISH wash buffer 2 (0.4xSSC 0.3% NP40)	Volume
20xSSC	10mL
dH2O	493.5mL
NP40	1.5mL
Pure HCl	Pitrate to pH7

Table 2.2.6: FISH SSC wash buffer 2. 0.4xSSC 0.3% NP40. Made at the POGLRC.

DAMP	Volume concentration	Final concentration
PBS	418.75mL	
Tri-sodium citrate (0.155M)	53mL	8.2mM
HAS 20%	24mL	1%
Dnase I (~2500U/1mL)	2mL	10U/mL
MgCl2 (1M)	1.25mL	2.5mM

Table 2.2.7: Thawing solution for primary samples (DAMP). Made at the POGLRC.

PBS/2% FBS washing solution	Volume
PBS	1mL
FBS	49mL

Table 2.2.8: Washing solution for FACS. Made at the POGLRC.

Growth Factors	High growth factor (ng/mL)	Low growth factor (ng/mL)	Physiological growth factor (ng/mL)
IL-6	20	0.2	1
IL-7	20	0.2	-
IL-3	20	0.2	-
SCF	100	1	0.2
Flt3 Ligand	100	1	-
G-CSF	20	0.2	1
GM-CSF	-	-	0.2
LIF	-	-	0.05
MIP- α	-	-	0.2

Table 2.2.9: Human physiological growth factors for growth of primary BMNC samples. Growth factor master mix was prepared, aliquoted and stored at -80°C. Master mix was thawed for addition to sterile serum-free growth media as required.

2.3 Materials and methods: Antibodies

Population	Surface Marker Expression
HSC	LIN- CD34+ CD38- CD90+
MPP	LIN- CD34+ CD38- CD90-
CMP	LIN- CD34+ CD38+ CD123+ CD45RA-
GMP	LIN- CD34+ CD38+ CD123+ CD45RA+
CLP	LIN- CD34hi/+ CD38+ CD45RA+ CD10+ CD19- CD33-
ProB	LIN+ CD34+ CD38+ CD45RA+ CD10- CD19+ CD33-
PreB	LIN+ CD34+ CD38+ CD10+ CD19+ CD33-

Table 2.3.1: Cell surface marker expression for early haematopoietic stem and progenitor cells. Cell populations were separated by the expression of cell surface markers by FACS.

Cell Surface Marker	Fluorochrome	Wavelength Range (nm)	
CD45RA	V450	404-448	
CD90	BV510	405-510	
CD10	PerCP-Cy5.5	482-695	
CD123	PerCP-Cy5.5	482-695	
Lin	FITC	494-519	
CD33	PE	496-578	
CD38	PE-Cy7	564-785	
CD34	APC	650-660	
CD19	APC-H7	650-785	

CD3
CD14
CD16
CD19
CD20
CD56

Table 2.3.2: Fluorochromes and wavelengths of antibodies used in FACS isolation of stem and progenitor cells from primary BMNC samples. Surface marker antibodies and fluorochromes were carefully selected to prevent wavelength crossover. Surface marker cocktails were selected based on previous experiments conducted within the lab group and Horne et al., 2020.

2.4 Materials and methods: Statistical analysis

Statistical analysis of experimental FISH and FACS data was performed using GraphPad Prism 10 software. Mann-Whitney unpaired t Test was used to examine correlations between FISH scores and clinical features. An unpaired experimental design was used for the t Test. Gaussian distribution was assumed (a parametric test was used) and as an unpaired t Test, both populations were assumed to have the same standard deviation.

Multivariate analysis (Multivariate ANNOVA) analysis of the combination of different methods of risk analysis with relation to Ph positivity in HSC and MPP diagnostic and PI samples was run to determine if risk can be associated with FISH score. Multiple comparisons were used to compare cell means regardless of rows and columns and each cell mean was compared with every other cell mean.

Statistical analysis of bioinformatic data was performed using GEO2R and R Studio. In GEO2R a Benjamini and Hochberg false discovery rate was used with forced normalisation and a Log 2 Fold threshold set as 1. Volcano plots and PCA (principle component analysis) plots were used to display data distribution. In R, Log 2 Fold values were calculated to measure change in gene expression and statistical significance measured by Adjusted P value. Z scores were calculated for changes in gene expression and applied to heatmaps.

2.5.1 Materials and methods: Software and coding (R Studio)

Software	Company
Excel	Microsoft
PowerPoint	Microsoft
Word	Microsoft
RStudio	Rstudio PBC
ImageJ	ImageJ
GraphPad Prism 9	GraphPad Software, Inc.
FlowJo v10	FlowJo LLC (Becton Dickinson)
Adobe Arobat Reader	Adobe
ZEISS AxioVision Rel 4.9.1	Carl Zeiss Microscopy, LLC
BioRender	BioRender
EndNote Ver 20	EndNote

Table 2.5.1: List of software used for data generation, collection and analysis. Software and data analysis programmes were used to harvest and analyse data. Findings were presented in slides, poster and written formats. Software was used on both Mac and Windows operating systems.

Packages	amap
	apegIrn
	ashr
	BiocParallel
	clusterProfiler
	DESeq2
	digest
	DOSE
	dplyr
	EnhancedVolcano
	fastcluster
	ggfortify
	ggplot2
	limma
	MulticoreParam(2)
	org.Hs.eg.db
	RColorBrewer
	ReactomePA
	reshape2
	Rmisc
	tibble
	tidyr
	tiff
Libraries	dendsort
	genefilter
	genefilter
	ggnewscale
	ggplot2
	ggrepel
	ggridges
	ggsci
	htmlwidgets
	pathview
	pheatmap
	RColorBrewer
	VennDiagram
	viridis
	xlsx

Table 2.5.2: R packages and libraries used for data analysis and visualisation. Source: GitHub & CRAN (the Comprehensive R Archive Network).

```

# Set working directory
setwd("/Users/coplandgroup/Desktop/Amy RNAseq work space")
save.image(file = "/Users/coplandgroup/Desktop/Amy RNAseq work
space/Amy_workspace.rdata")
load("/Users/coplandgroup/Desktop/Amy RNAseq work
space/Amy_workspace.rdata")

require(ggplot2)
require("BiocParallel")
require(DESeq2)
register(MulticoreParam(2))
require(Rmisc)
require(ggfortify)
require(RColorBrewer)
require(fastcluster)
require(reshape2)
require(ama)
require(digest)
require(ReactomePA)
require(org.Hs.eg.db)
require(tibble)
require(DOSE)
require(clusterProfiler)
require(tidyr)
require(dplyr)
require(limma)
require(tiff)
require(EnhancedVolcano)
require(apeglm)
require(ashr)
library(dendsort)
library(viridis)
library("genefilter")
library(pheatmap)
library(RColorBrewer)
library(htmlwidgets)
library(ggnewscale)
library(pathview)
library(gggridges)
# library("xlsx")
library(VennDiagram)
library("ggsci")
library(genefilter)
library(ggplot2)
library(ggrepl)

```

Figure 2.5.1: RNAseq data analysis (R Studio), setup. Set local working directory and load raw data as .rdata file. Load libraries and install any required packages.

```

# Import datasets
countdata <- read.csv("countMatrix_Amy.txt", sep="\t", row.names=1)
countdata[1:5] <- NULL
countdata <- as.matrix(countdata)
exp_design <- c("4H61", "5D06", "626A", "774X", "9D02", "3683", "CML267", "CML41",
"E549", "F192", "FR0299", "FR03158803", "OR87")
colnames(countdata) <- exp_design

# Filter the genes with the low number of reads
countdata <- subset(countdata, rowSums(countdata)>=5)

#Re-order columns
col_order <- c("CML267", "CML41", "4H61", "5D06", "626A", "774X", "9D02", "3683",
"E549", "F192", "FR0299", "FR03158803", "OR87")
countdata <- countdata[,col_order]

# Define conditions for the samples
group <- factor(c(rep("CML", 2), rep("PHALL", 11)))

```

Figure 2.5.2: RNAseq data analysis (R Studio), define dataset. Import raw data dataset, define dataset as matrix, remove columns which are not required, label columns with sample ID, arrange by sample type and define sample types.

```

### Perform differential expression analysis
# Create a coldata frame for the count matrix and the conditions
coldata <- data.frame(row.names=colnames(countdata),group)
dds <- DESeqDataSetFromMatrix(countData = countdata, colData =
coldata, design = ~ group)
# dds$condition <- relevel(dds$condition, ref= "CD180_neg")
dds <- DESeq(dds)

resultsNames(dds)

```

Figure 2.5.3: RNAseq data analysis (R Studio), differential expression (DE). Perform differential gene expression analysis on data matrix using DESeq.

```

### PCA plots
## rld
rld <- rlog(dds,blind=FALSE)
Batchmat <- assay(rld)
mm <- model.matrix(~ group, colData(rld))

## Modified plotPCA from DESeq2 package. Shows the Names of the Samples (the first col of SampleTable)
and uses ggrepel pkg to plot them conveniently.
# @SA 10.02.2017
plotPCA.san <- function (object, intgroup = "condition", ntop = 500, returnData = FALSE)
{
  rv <- rowVars(assay(object))
  select <- order(rv, decreasing = TRUE)[seq_len(min(ntop,
                                                    length(rv)))]

  pca <- prcomp(t(assay(object)[select, ]))
  percentVar <- pca$sdev^2/sum(pca$sdev^2)
  if (!all(intgroup %in% names(colData(object)))) {
    stop("the argument 'intgroup' should specify columns of colData(dds)")
  }
  intgroup.df <- as.data.frame(colData(object)[, intgroup, drop = FALSE])
  group <- if (length(intgroup) > 1) {
    factor(apply(intgroup.df, 1, paste, collapse = " : "))
  }
  else {
    colData(object)[[intgroup]]
  }

  ## Select the PCAs and percentVar that you like instead of 1 and 2
  d <- data.frame(PC1 = pca$x[, 1], PC2 = pca$x[, 2], group = group,
                 intgroup.df, name = colData(rld)[,1])
  if (returnData) {
    attr(d, "percentVar") <- percentVar[1:3]
    return(d)
  }
  ggplot(data = d, aes_string(x = "PC1", y = "PC2", color = "group", label = "name")) + geom_point(size =
3) + xlab(paste0("PC1: ", round(percentVar[1] * 100), "% variance")) + ylab(paste0("PC2: ",
round(percentVar[2] * 100), "% variance")) + coord_fixed() + geom_text_repel(size=3)

}

pca_rlog <- plotPCA.san(rld,intgroup="group")

tiff(filename = "plots/PCA_plot_rlog.tif",res = 600, width=8000,height=3000,pointsize=10)
print(pca_rlog)
dev.off()

```

Figure 2.5.4: RNAseq data analysis (R Studio), principal component analysis (PCA). Visualise differential expression as PCA using DESeq2, select percentage variables based on Log2Fold DE value and save plot as .tiff file.

```

## Differential expression analysis
CMLvsPHALL <- results(dds,contrast = c("group","CML","PHALL"),
pAdjustMethod = "BH")
CMLvsPHALL <- as.data.frame(CMLvsPHALL)

# Annotate the Ensembl gene IDs to gene symbols
CMLvsPHALL$symbols <- mapIds(org.Hs.eg.db, keys =
row.names(CMLvsPHALL), column = c('SYMBOL'), keytype =
'ENSEMBL')
CMLvsPHALL <- CMLvsPHALL[,-c(1,3)]

# Remove rows with missing values
CMLvsPHALL <- CMLvsPHALL[complete.cases(CMLvsPHALL),]

# Combine rows with duplicated symbols and calculate mean of
log2FC and padj and stat
CMLvsPHALL <-
aggregate(cbind(CMLvsPHALL$log2FoldChange,CMLvsPHALL$padj,CMLvsP
HALL$stat) ~ CMLvsPHALL$symbols, FUN= mean, na.action = na.omit)
# Change colnames
colnames(CMLvsPHALL) <-
c("symbols","log2FoldChange","padj","stat")

```

Figure 2.5.5: RNAseq data analysis (R Studio), differential expression analysis. Set DE data as a data frame, change values in “symbols” column from Ensembl IDs to gene symbols, remove rows (genes) with no values, then combine gene expression data for each gene from each sample, calculate mean, Log2Fold change and adjusted P value of this DE value.

```

## Volcano plots
# For MOLM13,
vol_CMLvsPHALL <- EnhancedVolcano(CMLvsPHALL,
lab = CMLvsPHALL$symbols,
x = 'log2FoldChange',
y = 'padj',
xlim = c(-3,6),
ylim = c(0,6),
pCutoff = 0.05,
FCcutoff = 0.58,
title = 'CML vs PHALL',
subtitle = paste0('p-value cutoff
(red line) drawn ',
'at equivalent of
adjusted p=0.05'),
)

tiff(filename = "plots/volc_CMLvsPHALL.tif",res = 800,
width=10000,height=7000,pointsize=4)
vol_CMLvsPHALL
dev.off()

```

Figure 2.5.6: RNAseq data analysis (R Studio), Volcano plot. Plot DE values for each gene by Log2Fold value, define acceptable Log2Fold values and define statistical significance value for adjusted P value (displayed by red line).

```

## Save differential expressed genes as a tab delim txt file for
GSEA
write.table(CMLvsPHALL, "/Users/coplandgroup/Desktop/Amy RNAseq
work space//CMLvsPhALL4H.txt", quote=FALSE, sep="\t",
row.names=TRUE, col.names=FALSE)

```

Figure 2.5.7: RNAseq data analysis (R Studio), data export. Export DE values (including gene symbol, Log2Fold and adjusted P value) as a table. Define if table will be separated by commas or tabs (note: loading DE data into GSEA requires input table to be a tab delimited text file).

2.5.2 Materials and methods: Software and coding (GEO2R)

GEO2R was utilised to perform differential expression analysis of publicly available datasets on the Gene Expression Omnibus (GEO) repository. GEO2R uses GEOquery and Limma (Linear Models for Microarray Analysis) to perform differential expression analysis of microarray datasets. GEOquery parses GEO data into R data structures which can then be used by other packages, in this analysis, Limma was used. Limma allows statistical testing for detection of differentially expressed genes and applied multiple-testing corrections on P-values to help correct for false-positive occurrences. Analysis was performed by entering the dataset ID in the GEO accession search bar. On the sample table, groups were defined by entering the specific group name and selecting the corresponding samples to be assigned to each group. To allow gene expression comparison between sample groups, at least two groups must be defined and compared. Once sample groups were defined, the 'Options' tab was selected to define testing conditions. Benjamini & Hochberg was selected for adjustment to the P values to account false for discovery rate. Log transformation was not automatically applied to analysis as the GEO database hosts logged and unlogged data, hence the auto-detect log function was used. Force normalisation was applied for quantile normalisation to the expression data making all selected samples have identical value distribution. The adjusted P-value threshold was set at 0.05 and the Log 2 fold change threshold was set at 1. Volcano and mean-difference plot contrasts was selected to produce volcano plots which were automatically plotted by GEO2R. After running the analyses (by selecting 'Analyze'), plots could be visualised in addition to a table of 250 differentially expressed genes. This table was downloaded and exported to Excel where the gene names, gene functions, log fold changes and P values for each of the 250 DEGs could be observed. Genes were sorted from upregulated to downregulated by sorting log fold changes by largest to smallest values. This process results in a table of DEGs (differentially expressed genes) for use in downstream analyses such as gene ontology.

2.5.3 Materials and methods: RNAseq (RNA sequencing) bioinformatic analysis

Sequencing was carried out by Glasgow Polyomics using library prep approach of polyA selection for low input samples and sequencing performed at PE 2x75bp with a read depth of 30M per sample. After harvesting of RNA using the RNeasy Mini Kit (Qiagen), sample quality was assessed for all samples using Bioanalyzer. The polyA RNA libraries were prepared by Glasgow Polyomics using the TruSeq stranded mRNA kit (Illumina). Paired end sequencing was generated on NextSeq2000 that yields a read of 100bp to a depth of 30 million reads. Using FastQC and MultiQC, the quality of raw RNAseq data was verified. The Trimmomatic tool was then used to automatically remove adaptor sequences and poly G repeats. FastQC and MultiQC was performed again followed by alignment to a reference human genome. HISAT2 program was used to align the reads against a reference human genome (GCA_000001405.29, NCBI RefSeq GRCh38.p14) which created a BAM file output. MultiQC was performed again to ensure the data was of a suitable quality for downstream analyses. The number of reads were calculated for each gene using the FeatureCount function. FeatureCount was used to determine how many pairs of reads from the paired end sequencing are mapped to specific genes to improve normalisation and ensure accuracy for downstream differential gene expression analysis. Initial data processing and alignment as discussed above was performed by Glasgow Polyomics and a raw count data file was provided as a tab-delimited text file for downstream analysis performed by myself with advice from Tae-Ju Park.

RNAseq data was analysed on R (2022.07.1) using the libraries and packages listed in table 2.5.2. Raw count data was imported as a matrix, the column names changed to patient ID sample numbers, row names changed to gene IDs and genes with low read numbers (≥ 5) were filtered out. Labelling patient IDs allowed patient-specific gene expression profiles to be parsed. Labelling of gene IDs was performed to fully identify differentially expressed genes (DEGs). A PCA plot was then generated to visualise all samples and conditions in a low dimensional manner to observe sample and condition variants.

The columns were reordered, and samples defined based on features such as disease type (CMLLBC vs Ph+ALL), disease features (ie samples with Ph+ stem and progenitor cells vs those without) or retrospectively based on sample principal component analysis (PCA) clustering (ie cluster 1 vs cluster 2). This allows for specific samples or sample groups to be compared for differential gene analysis. PCA plots do not discard samples or variable characteristics but instead, reduce the number of data dimensions by constructing principal components (PCs) to elucidate patterns of expression from complex biological datasets. PCA describe variation while accounting for varied influences originating from sample characteristics (such as different disease type or age). PCA plots were generated on R studio as has been outlined in materials and methods.

Differential gene expression analysis was performed by creating a column data frame for the count matrix and conditions (sample features as discussed above) using DESeq2 (Bioconductor). Log2foldchange shrinkage was performed to generate more accurate log2foldchange estimates. This output was used to generate volcano plots to display overall differential gene expression between defined conditions. Cut-off criteria for DEGs was defined as P-value <0.05 and log2foldchange below -0.5 or above 0.5. Significant DEGs were exported as a table including log2foldchange and p value using the function write.table, saving the data as a ab delimited file which can be viewed as a .txt file or in an Excel spreadsheet.

Heatmaps were generated by selecting significant genes of interest and converting normalising and scaling raw gene counts into a normalised z-score which was used to plot the heatmap using the heatmap function in R.

Gene set enrichment analysis (GSEA) of DEGs was performed using the GSEA MSigDB (molecular signatures database) (UC San Diego, Broad Institute) software. This was done by ranking the DEGs in descending order of log2foldchange, saved as a .rnk file and uploaded to the GSEA software (version 4.2.3). Hallmark pathways were selected from the MSig data base (Gene sets database), number of permutations set to 1000 and the GSEA ran as a pre-ranked

gene list. GSEA data was exported as an .html where individual plots could then be examined. GSEA provides statistical information for examination of gene set enrichment results, in this thesis the normalised enrichment score (NES) has been utilised. GSEA utilises the enrichment score (ES), which reflects the degree to which a gene set is overexpressed at the top or bottom of a ranked gene list, and compares ES to the mean of the ESs against all permutations of the dataset. NES accounts for differences in gene set size and correlations between MSigDB gene sets and the expression dataset, allowing comparison of analysis results across gene sets.

Further analysis was performed using online gene ontology (GO) tools such as ShinyGO (Version 0.80/0.741, South Dakota State University) and GOrilla (Eden et al., 2009), and functional protein association networks generated using STRING (Version 12.0, STRING Consortium 2023©). A ranked list of statistically significant DEGs were pasted into the appropriate gene/protein name box as a single ranked list of genes, appropriate organism selected (Homo Sapiens) and the desired ontology selected.

2.6.1 Materials and methods: Primary patient sample ethical considerations

Under existing COREC approval, primary patient samples were utilised throughout this project and were sourced via the CCLG (Children’s Cancer and Leukaemia Group) CellBank (CCLGA 2019 03/304680), and the Paul O’Gorman Leukaemia Research Centre bio bank (20-WS-0066). Primary samples used for this project included paediatric Ph+ALL cryopreserved bone marrow aspirates (age 2-19 years), ‘normal’ healthy bone marrow trephines shipped in EDTA from adult patients undergoing hip replacement operations at the Queen Elizabeth University Hospital Glasgow (QEUH), adult CML cryopreserved bone marrow aspirates and fresh bone marrow aspirates harvested from an anonymised paediatric patient undergoing Ph+ALL treatment at the QEUH.

In accordance with health and safety standards, researchers received a Hepatitis B vaccine prior to working with untested paediatric bone marrow samples. Health and safety standards were maintained to the appropriate level for a biosafety category 2 laboratory (in accordance with the UK government <https://www.hse.gov.uk/biosafety/information.htm>).

2.6.2 Materials and methods: Primary bone marrow mononuclear cell isolation and culture - healthy patient trephine controls

On attainment of bone marrow trephine samples, the trephine core was removed from the EDTA sample tube and emptied into a mortar. The sample was crushed, without grinding, in 10mL sterile PBS using a pestle. 10mL sterile PBS was pipetted into the EDTA tube to rinse out any residual cells. Cells in PBS were filtered through a 40 μ M Easy Strainer into a 50mL Falcon tube, and the crushed trephine core disposed of. Filtered cells were then centrifuged for 5 minutes at 400xG, supernatant aspirated and the precipitate resuspended in 6mL sterile PBS. To 8mL room temperature Histopaque (Sigma), 6mL of cells was added carefully to prevent mixing and centrifuged for 30 minutes at 400xG with the breaks off (Fig 2.6.1). Density centrifugation resulted in a distinct white layer of leukocyte which was collected with a Pasteur pipette and centrifuged in 10mL sterile PBS for 5 minutes at 400xG twice. Cells were then counted overnight in serum-free media (SFM) supplemented with physiological growth factor +IL-7 (PGF(+IL-7)) at 37°C 5% CO₂.

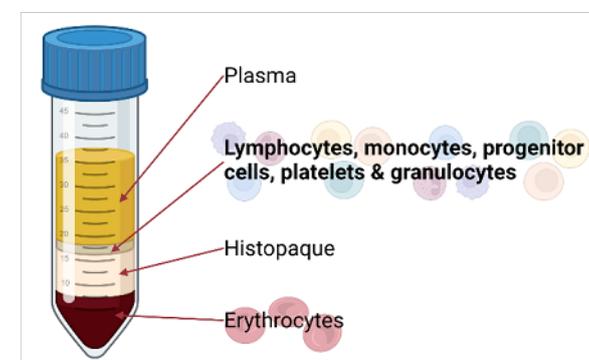


Figure 2.6.1: Schematic representation of density centrifugation separation of mononuclear cells from whole bone marrow samples. Cells were centrifuged in density gradient medium (Histopaque) until layers of cell types could be visualised. Mononuclear cells were then isolated by pipetting then cultured. *(Created with Biorender.com)*

2.6.3 Materials and methods: Primary bone marrow mononuclear cell isolation and culture - paediatric Ph+ALL bone marrow aspirate

Fluid bone marrow aspirate was decanted into a 50mL Falcon tube. 10mL sterile PBS was pipetted into the EDTA tube to rinse out any residual cells. Cells in PBS were filtered through a 40µM Easy Strainer into a 50mL Falcon tube, and the crushed trephine core disposed of. Filtered cells were then centrifuged for 5 minutes at 400xG, supernatant aspirated and the precipitate resuspended in 6mL sterile PBS. To 8mL room temperature Histopaque, 6mL of cells was added carefully to prevent mixing and centrifuged for 30 minutes at 400xG with the breaks off (Fig 2.6.1). Density centrifugation resulted in a distinct white layer of leukocyte which was collected with a Pasteur pipette and centrifuged in 10mL sterile PBS for 5 minutes at 400xG twice. Cells were then counted overnight in serum-free media (SFM) supplemented with physiological growth factor +IL-7 (PGF(+IL-7)) at 37°C 5% CO₂.

2.6.4 Materials and methods: Primary BMNC thawing (Ph+ALL & CML patient samples)

Ph+ALL and CML bone marrow leukapheresis samples were briefly stored on dry ice after removal from long-term storage in liquid nitrogen and subsequently thawed in a 37°C water bath over 3-5 minutes. 1mL of room temperature filtered DAMP (Table 2.2.7) was added to a 50mL Falcon tube, swirled to coat the inside of the tube and decanted. BMNCs were then carefully transferred to the DAMP coated Falcon tube using a Pasteur pipette. 2mL DAMP was added to the cryotube to wash out excess cells and added to the Falcon tube in a dropwise manner over 3 minutes. To the cells, an additional 8mL DAMP was added in a dropwise manner over 3 minutes and then centrifuged for 10 minutes at 250xG. Resultant supernatant was then transferred to a fresh 50mL Falcon tube and the pellet resuspended in 10mL DAMP. Both Falcon tubes were then centrifuged for 10 minutes at 250xG and pellets combined through resuspension in 10mL SFM+PGF(+IL-7). Cells were counted using a haemocytometer (Fig 2.6.2) and cultured overnight at 37°C.

2.6.5 Materials and methods: Cell counting

Cell counts were performed using an Improved Neubauer haemocytometer. Cells were pelleted by centrifuging for 5 minutes at 300xG and resuspending in 10mL SFM supplemented with PGF +IL-7. 50µL of the cell suspension was diluted 1:2 in filtered trypan blue viability dye and incubated at room temperature for 2 minutes. 10uL of the cell/trypan blue solution was pipetted onto the counting chamber of the haemocytometer and a glass coverslip affixed using liquid surface tension. Cells were then counted in the four 10^{-4} corner squares of the haemocytometer using an inverted light microscope (Fig 2.6.2). Viable cells were distinguished from non-viable cells using the trypan blue exclusion method (dead cells= blue). Cell numbers were multiplied by 10,000 and then by the dilution factor (1:2, therefore dilution factor is 2) to calculate the number of cells per millilitre and multiplied by the total volume (10mL). The percentage of viable cells were calculated by dividing the number of viable cells over the number of non-viable and then multiplying by 100.

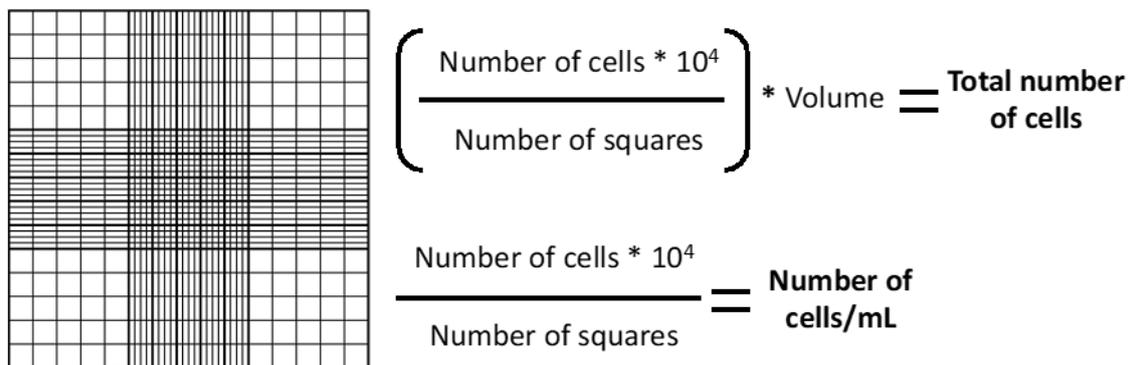


Figure 2.6.2: Haemocytometer cell counting. Haemocytometer grid with area for counting highlighted in orange. Number of cells= cell count per haemocytometer square, Volume= total volume of cell/media resuspension and Number of squares= number of haemocytometer squares counted.

2.7 Materials and methods: FACS (fluorescence-assisted cell sorting) isolation of haematopoietic progenitor cells from bone marrow samples (CML, Ph+ALL and healthy controls)

Prior to cell sorting, cells were counted using a haemocytometer and then centrifuged for 10 minutes at 250xG. The supernatant was decanted, discarded and the cells resuspended in PBS supplemented with 2% FBS. Cells were centrifuged at 250xG for 10 minutes and again washed with 2% FBS PBS, this was repeated one more time and the cells resuspended in 1×10^6 cells/ml in 2%FBS PBS. During wash steps, antibody cocktails were prepared by adding the appropriate volume of antibodies ($\sim 3\mu\text{L}$) to the appropriate labelled tube and adding 2%FBS PBS up to the volume of $100\mu\text{L}$. Antibody cocktails were vortexed thoroughly and kept in the dark until required (Table 2.3.1). From the cell suspension, $100\mu\text{L}$ was removed for the unstained control.

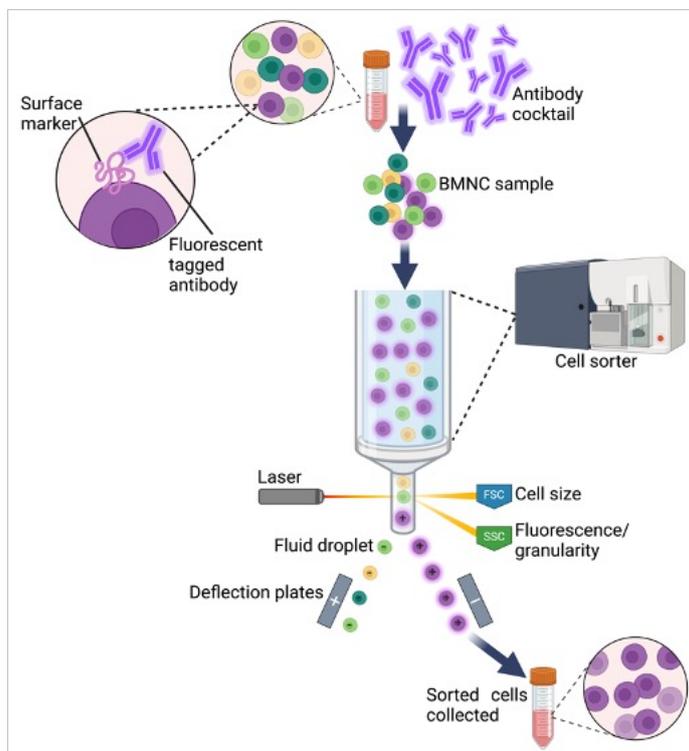


Figure 2.7.1: Fluorescence-assisted cell sorting (FACS). Mixed bone marrow mononuclear cell (BMNC) population was incubated with a fluorescently tagged antibody cocktail specific for surface markers expressed by haematopoietic progenitor cells. Cells were encapsulated into liquid droplets and tagged antibodies excited by the laser to emit light at a lower energy and longer wavelength. Forward scatter (FSC) emitted light assesses cell size and side scatter (SSC) detects fluorescence and granularity of cells. A voltage pulse applied to the liquid droplet allows the cells of interest to be deflected in an electric field between two deflection plates for collection. (Created with Biorender.com)

Antibodies for the cell surface marker cocktails were selected to ensure fluorophore compatibility (ie the peak excitation and emission spectra was different between each antibody) (Table 2.3.2). Antibody panels were designed to identify haematopoietic cell progenitor populations based on surface marker expression. Flow markers were selected based on previous FACS experiments conducted in our lab, published protocols (Vadakkemadathil et al., 2019) and clinical guidelines (Li., 2022).

Antibody cocktails were added to pelleted cells and the cells incubated in the dark at room temperature for 30 minutes. During this incubation step, flow cytometry beads were vortexed thoroughly and 1 drop added to a clean FACS tube for each single stain control and volume made up to 100 μ L with 2%FBS PBS. 1 μ L of each antibody was added to each single stain control tube and vortexed thoroughly. Antibody cocktail was washed off cells by centrifuging at 250xG for 10 minutes and resuspending in 300 μ L 2%FBS PBS. Collection tubes (1.5mL Eppendorf tube) were prepared with the addition of 2%FBS PBS. Cells were then sorted making a note of the number of cells sorted into each population (Fig 2.7.1).

2.8 Materials and methods: FISH (fluorescence *in situ* hybridisation)

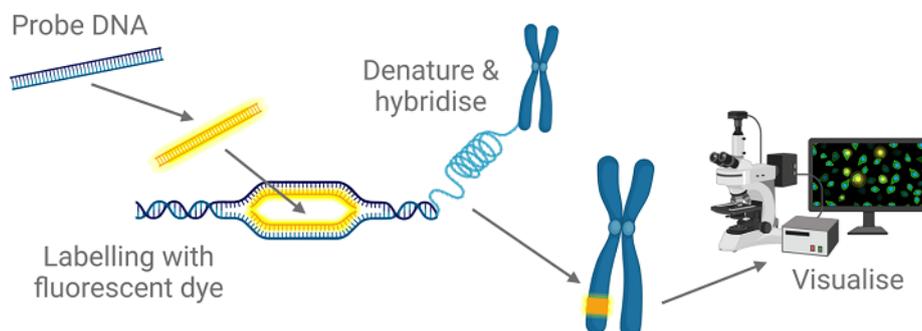


Figure 2.8.1: Schematic representation of FISH. Target cells were treated with fluorescently labelled probe which emit a different colour when hybridised (*BCR::ABL1*= yellow), allowing for the detection of the fusion gene by fluorescence microscopy. Cells which do not contain the fusion gene can be detected by identification of unbound genes (*BCR*= green, *ABL*=red). (Created with Biorender.com)

Prior to preparation of cells for fixing, fresh reagents were prepared including FISH fix (Table 2.2.1 and 2.2.2) and 0.075M KCl (Table 2.2.3). Cells were collected by FACS and centrifuged at 1000RPM for 10 minutes. Supernatant was removed by pipetting, carefully so as not to disturb the cell pellet, and then cells mixed by vortex. Cells were resuspended in 200 μ L 0.075M KCl, incubated at 37 $^{\circ}$ C for 15 minutes then centrifuged for 5 minutes at 1500RPM. Three washes were performed by supernatant aspiration, the addition of 1mL FISH fix, incubation at room temperature for 2 minutes and centrifugation for 2 minutes at 12000RPM, lastly, 1mL FISH fix was added and fixed cells stored at -20 $^{\circ}$ C.

Upon setup of the FISH slides, cells were removed from the -20 $^{\circ}$ C freezer, centrifuged for 12000RPM and then resuspended in a small volume of FISH fix. 3 μ L of cells were pipetted onto each spot of a Shandon Multispot slide and allowed to air dry. Cells were aged by incubation at 65 $^{\circ}$ C for 10 minutes before application of 1:10 FISH probe and hybridisation buffer mix prepared on the day of setup. A glass coverslip was applied to the slide and sealed using clear nail polish. The *BCR::ABL1* probe was then hybridised (Fig 2.8.1) by placing the slide on a 72 $^{\circ}$ C hot block for 5 minutes before overnight incubation in a humid 37 $^{\circ}$ C environment.

A Coplin jar containing Wash 1 (Table 2.2.5) was placed in a 75°C water bath while the Coplin jar containing Wash 2 (Table 2.2.6) was kept at room temperature on the bench. The cells were removed from the humid environment and clear nail polish sealant carefully peeled off without removing the coverslip. The glass coverslip was removed by submerging the slide in Wash 2 for 10-20 seconds until the coverslip could be easily slid off. The slide was submerged in Wash 1 then Wash 2 for 2 minutes each before mounting a glass coverslip onto the slide using 20µL DAPI (VectaShield) mounting media. BCR::ABL1 fusion events were then visualised using a fluorescence microscope (Zeiss Axio Imager M1) (Fig 2.8.1).

2.9.1 Materials and methods: Preservation of RNA using RNAProect (Qiagen)

Bulk and sorted BMNCs were stored in RNAProtect (Qiagen) by centrifugation at 300xG for 5 minutes and resuspending the pellet in RNAProtect at a 10:1 ratio (10x RNAProtect to 1x cell volume). Cell mixture was then stored at -80°C until RNA extracted.

RNA was extracted from a number of samples at the same time in contrast to extracting from one sample at a time to prevent inter-sample variability resulting from RNA extraction performed on different days. Such variability can result in batch effects and may result in inaccurate reporting of transcriptional data during RNAseq.

2.9.2 Materials and methods: RNA extraction from primary BMNC samples

RNA was extracted from bulk and sorted BMNC cells using the RNeasy Mini kit (Qiagen). Cells preserved in RNAlater (Qiagen) were thawed on ice and fresh cells were directly pelleted by centrifuging for 5 minutes at 300xG and supernatant fully aspirated. The cell pellet was loosened by flicking the tube and further disrupted by adding 350µL Buffer RLT and thoroughly vortexing. 350µL 70% ethanol was added to the homogenised lysate and mixed well by pipetting. 700µL of the sample was then transferred to a RNeasy spin column placed in a 2mL collection tube and centrifuged for 15 seconds at 8000xG. The flow-through was discarded, 700µL Buffer RW1 added to the RNeasy spin column and centrifuged at 8000xG for 15 seconds to wash the spin column. After discarding the flow-through, 500µL Buffer RPE was added to the spin column and centrifuged at 8000xG for 15 seconds and the flow-through discarded. 500µL Buffer RPE was again added to the spin column and centrifuged at 8000xG for 2 minutes. The RNeasy spin column was placed in a new 1.5mL collection tube and 30-50µL RNase-free water pipetted directly onto the spin column membrane. The sample was then centrifuged at 8000xG for 1 minute to elute the RNA. RNA was then snap frozen and stored at -80°C.

2.10 Materials and methods: Sequencing of B cell receptor VDJ gene segments

ClonoSEQ® MRD monitoring by Adaptive Biotechnologies® was utilised for sequencing of B cell receptor (BCR) gene segments in both early stem/progenitor cell populations and Pro/PreB/lymphoid cells isolated from Ph+ALL primary patient samples. ClonoSEQ® is available as an FDA-cleared *in vitro* diagnostic (IVD) test service provided by Adaptive Biotechnologies to detect minimal residual disease in bone marrow from patients with multiple myeloma or B cell acute lymphoblastic leukaemia (B-ALL) and blood or bone marrow from patients with chronic lymphocytic leukaemia (CLL). ClonoSEQ® is also available for use in other lymphoid cancers and specimen types as a CLIA-validated laboratory developed test (LDT).

Populations of progenitor and lymphoid cells were isolated and collected by FACS as previously described (Materials and methods 2.5). Cells were sorted into 1.5mL Eppendorf tubes containing 300µL 2% FBS PBS (Table 2.8). As DNA extraction is performed in-house by Adaptive Biotechnologies®, cells were immediately snap frozen upon collection and stored at -80°C until shipped on dry ice to the Adaptive Biotechnologies® laboratory (Seattle, US-WA). After gDNA extraction, quality is assessed and rearranged immune receptors are amplified using a multiplex PCR. Reaction-specific index barcode sequences for sample ID are added to the amplified receptor sequences by PCR. Sequencing libraries are prepared from barcoded amplified DNA, which are then sequenced by PCR. Sequencing libraries are prepared from barcoded amplified DNA, then sequenced by synthesis using NGS (Illumina NextSeq 500/550). Raw sequence data, once uploaded to the Adaptive analysis pipeline, are analysed by identifying the sample's sequence data using sample index sequences, processed using a proprietary Adaptive Biotechnologies® algorithm with in-line controls to remove amplification bias. Once ID assessment is concluded, the immune repertoire of the sample is checked for the presence of DNA sequences specific to a dominant clone consistent with the presence of a lymphoid malignancy. These sequences are compared against a B cell repertoire database and assigned a uniqueness value that, together with its abundance relative to the other

sequences, is used to assign the sequence to a sensitivity bin which is then used in the estimation of the reported limit of detection and limit of quantitation, required for ongoing MRD detection.

Results are delivered via a clonality ID report published on the ClonoSEQ® clinician web portal. This report provides an overview of the dominant DNA sequences identified in a patient sample, the dominant sequences typically being associated with malignancy and used as the basis for further MRD detection (no further MRD monitoring was done as part of this project).

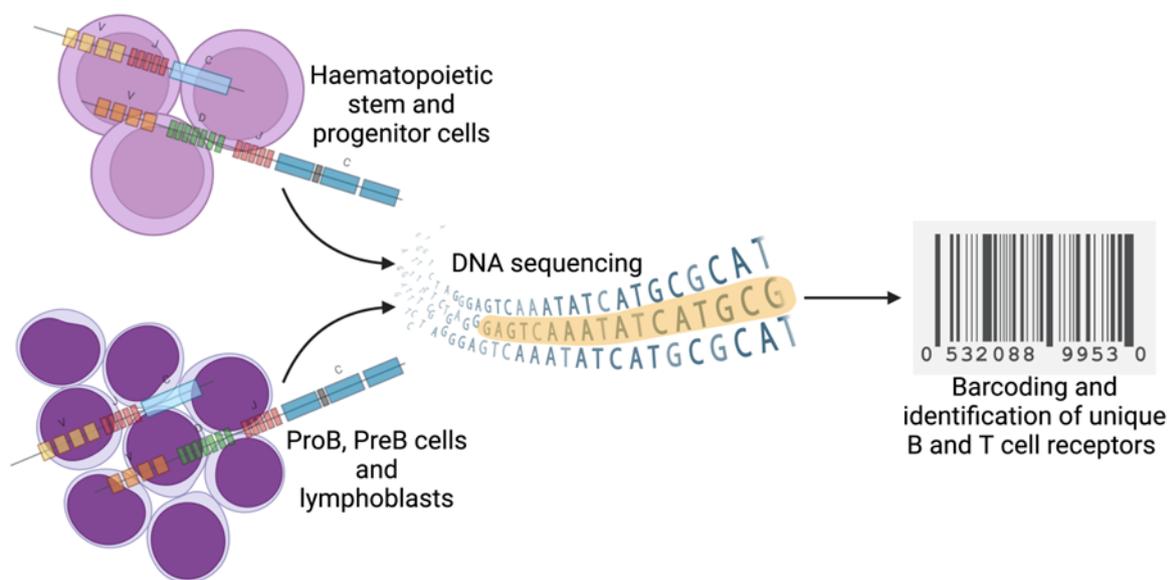


Figure 2.11: Schematic representation of ClonoSEQ® Ig/TCR MRD monitoring. Each B or T cell receptor is encoded by unique DNA sequences of 3 segments; Variable, Diversity and Joining. Such sequences are then used as barcodes which can be used to track malignant cells for MRD monitoring. ClonoSEQ® identifies and quantifies unique B and T cell receptor sequences associated with the malignant clone, acting as a direct measure of the tumour. Sequencing performed on sorted stem and progenitor cells and early lymphoid/blast cells. (Adapted from clonoseq.com/about/ Created with Biorender.com)

Chapter 3. Results I. Utilisation of publicly available datasets to investigate the presence of Ph+ALL subtypes

3.1.1 Introduction: Analysis of gene expression for the identification of novel Ph+ALL subtypes using publicly available datasets.

CML-like Ph+ALL was initially highlighted as being a potential novel subtype through clinical observations of Ph+ALL patient responses to treatment and the observation of persistent *BCR::ABL1* MRD despite loss of IgTCR MRD within this population (Hovorkova et al., 2017). This was further outlined in retrospective analysis where the prognostic significance of this sub population was highlighted, with more cases of early relapses detected in those with presumed CML-like disease (Zuna et al., 2022). Early diagnosis of this sub population is essential to enable more intensive treatment approaches with allogeneic stem cell transplantation within paediatric and adolescent patients upfront to aid better response outcomes.

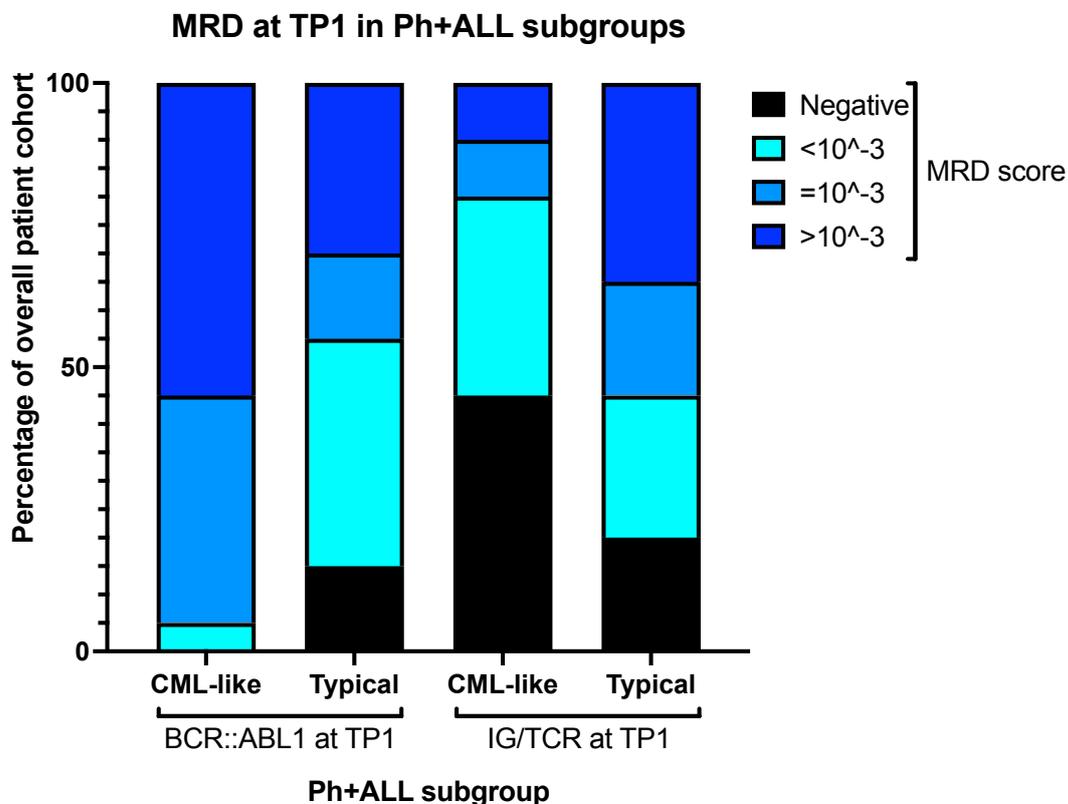


Figure 3.1: MRD score at TP1 in the CML-like and typical Ph+ALL subgroups. BCR::ABL1 and IG/TCR MRD assessment at TP1 compared in two clinically different PH+ALL subtype groups. (BCR::ABL1 $P < 0.0001$ and IG/TCR $P = 0.0006$). Adapted from Hovorkova et al., 2017.

The original study highlighted that at diagnosis, CML-like Ph+ALL patients could not be differentiated through standard diagnostic testing. They exhibited similar WCC at diagnosis, had similar FISH results and did not exhibit a propensity to express one fusion variant over another. These CML-like patients did not fall into a pattern of age or sex and only displayed themselves as outliers through MRD assessment (Fig 3.1).

There is paucity of data surrounding CML-like Ph+ALL, but it can be hypothesised that the cell of origin is key in the pathogenesis. This is in view that within previous work, *BCR::ABL1* could be identified in myeloid and lymphoid cells within patient samples identified as being CML-like Ph+ALL (Hovorkova et al., 2017). This is in contrast to the existing Ph+ALL paradigm in which *BCR::ABL1* is only found in lymphoid cells due to the initial fusion event occurring in a lymphoid progenitor. Hence, the reciprocal translocation event which gives rise to *BCR::ABL1* occurring in a progenitor cell with non-fixed lineage gives rise to a mixed-lineage disease phenotype and supports the status of CML-like Ph+ALL being a phenotypically distinct subtype.

In order to investigate the transcriptomic landscape of this population, we first utilised publicly available datasets. This was, in part, because some of the major caveats when approaching yet undefined subtypes include how to isolate affected patients from the wider population, shared phenotypic hallmarks between types of leukaemias and sample size. Ph+ALL accounts for approximately 5% of all paediatric and adolescent cases of ALL. Of these, Hovorkova et al estimates that 30% of the overall cohort will be CML-like Ph+ALL. Therefore, the likelihood of demonstrating a transcriptomic pattern in random patient sample selection is small. Hence, utilisation of publicly available datasets allowed examination of a larger Ph+ALL cohort where subtype-specific differential expression may be identified. In addition, the datasets utilised in this chapter included CML and healthy control samples. The purpose of including samples allowed for the identification of Ph+ALL specific leukaemia genes, CML genes that may share expression patterns between CML and the proposed CML-like Ph+ALL subtype and, importantly, identify genes required for leukaemia disease pathology. As the CML-like Ph+ALL subtype is, as of yet, undefined, there exists no transcriptional profile or gene set which may be used to identify patients belonging to this Ph+ALL subtype. Hence, in order to identify such patients, and eventually compile an identifying gene expression

signature, the transcriptional spectrum between CML and Ph+ALL must first be established in order to identify which genes are similarly expressed between CML-like Ph+ALL, CML and standard Ph+ALL.

The approach taken toward the investigation of the transcriptomic landscape of Ph+ALL heterogeneity was to examine publicly available datasets for the expression of phenotypically relevant genes using R packages Limma, EdgeR, Dplyr, Tidyverse and Rmarkdown. A list of genes were compiled through literature search (Table 3.1). Publications consulted included topics of Ph+ALL, CML (lymphoid blast phase), Ph+ leukaemic stem cells, HSC biology and bone marrow alteration in CML (papers used are referenced in Table 3.1 legend). Additionally, to investigate overall gene expression in the datasets, global gene expression was analysed using the GEO2R tool (NCBI). Utilising both methods allowed for genes of interest to be examined in addition to assembling a global transcriptional image.

Gene	Basic function
ADAM17	Self renewal & NOTCH signalling
AKT	Contains SH2 domains, anti-apoptotic
AMPK	Regulates energy metabolism and cell cycle in HSC
ATF4	Expansion and maintenance of HSCs in FL
BPTF	Maintenance of adult haematopoiesis, activates stemness gene expression
CCL8	Maintenance of quiescent HSC
CCND1	Cell cycle, maintenance of adult HSCs
CD79A	B lineage
CEBPA	Master regulator
CIUN	Self renewal, proliferation, differentiation and apoptosis
C-KIT	Proto-oncogene, enhances growth of myeloid and erythropoietic progenitor cells
CLCA1	Regulate stroma dependent in vitro haematopoiesis
CNR2	Regulates embryonic HSC development via PGE2
CREBBP	Tumour suppressor gene, maintenance of normal haematopoiesis
CSF1R	Essential for macrophage differentiation
CTLA2A	Required for erythropoiesis
DLL1	Self renewal pathway_2
EBF1	Early haematopoiesis endothelial-to-haematopoietic transition
EBF2	Determines osteoblastic niche required for HSC maintenance via Wnt signalling
ERG	Oncogene, required for definitive haematopoiesis and adult HSC function
ETV6	Required for megakaryocyte development
ETS1	Blocks erythroid and promotes megakaryocytic differentiation
ETS2	Proto-oncogene, erythroid to megakaryocytic phenotype switch
EZH2	Regulates normal HSC self-renewal and differentiation
FLU1	Maintenance of normal HSC homeostasis and function
FLT3	Required for normal development of HSCs, acts synergistically with leukaemia oncogenes to confer a more aggressive phenotype
FZD6	Regulates HSPC expansion and survival
GATA 1	Self renewal, erythroid gene expression and regulator of erythroid and haematopoietic cell development
GATA2	Critical regulator of normal and LSCs
GATA3	Controls expression of key lineage determining factors and cell cycle genes
GF1B	Required for multilineage development
GLUT6	Glucose uptake and utilisation, B lymphoid development
HBA1	Foetal haemoglobin subunit, erythropoiesis
HHEX	Essential regulator of embryogenesis and haematopoietic progenitor development
HIPK1	Inhibits mim-1 expression, key regulator of haematopoiesis
HK2	Expressed in myeloid progenitors
HOXA10	Maintenance of quiescent HSC, erythroid and megakaryocyte development
HOXA5	Expressed in HSCs and MPPs, erythropoiesis
HOXA9	Regulate normal haematopoiesis
HOXB4	Enhance HSC self-renewal and expansion
IGF2	Regulation of interaction between HSC and their niche, maintains balance between SC self-renewal and differentiation
IKAROS	Required during early haematopoiesis for differentiation into the three major lineages
IL-15	Associated with myeloid lineage development
ISR	Active in HSCs and facilitates their persistence
LCK	Regulator of lymphoid development
LCP1	Myeloid-specific gene
LDB1	Cofactor for haematopoietic TFs
LEF1	Multi-lineage blood reconstitution
LMO2	Active in T cell leukaemias, crucial for haematopoietic development
LYL1	Required for adult haematopoietic cell survival
MAFB	Inducer of monocyte differentiation
MAML	Maintenance and expansion of HSCs
MEK3	Proliferation, differentiation, migration and apoptosis
MCM7	Expressed in quiescent and cycling old HSCs
MCM9	Expressed in quiescent and cycling old HSCs
MEIS1	Required for HSC maintenance, erythropoiesis, megakaryopoiesis and HSC expansion
MPL	Promotes haematopoietic commitment in ESCs and establishment of definitive haematopoiesis
MPO	Selectively expressed in cells committed to granulomonocytic differentiation
MS4A4D	Haematopoietic support capacity of stroma
MYB	Oncogene, required for lymphocyte development
MYC	Oncogene, regulate HSC proliferation and differentiation
NFE2	Regulates HSC self-renewal and T cell differentiation
NOTCH1	Influences the generation of both definitive myeloid cells and lymphoid cells
NOTCH2	Early haematopoietic reconstitution
NR3C1	Regulation of LSC programmes and G1/S transition via E2F
ORC2	Transcriptional regulator that controls erythroid lineage
P13K	Mediator of cytokine signalling required for haematopoiesis regulation, overexpression associated with poor leukaemic outcome
P38	Regulation of gene and protein expression of essential haematopoietic cytokines in primary BM stromal cells
PAX5	Required for B cell development
PBX1	Maintenance of quiescent HSC
POSTN	Regulates HSC proliferation
PRDM16	Critical for HSC maintenance, supports B cell development
RAG2	Supports T cell development
RUNX1	Required for the generation of the first definitive HSCs (embryo)
SCA-1	Identifies quiescent HSCs
TAL1	Maintenance of HSC multipotency and quiescence (stage G0)
SEC23B	Involved in erythroid lineage development
SFRP2	Regeneration and maintenance of HSC pool
SMAD2	TGF-beta signalling and enhancement of erythroid differentiation
SMAD3	TGF-beta signalling and enhancement of erythroid differentiation
SMAD4	Protect HSCs against leukaemic transformation, self-renewal
SMAD5	Required for erythropoiesis
SPI1	Expressed in the monocytic and B lymphocytic lineages, proto-oncogene
SPON1	Maintenance of HSCs in quiescent state, self-renewal and cell fate decisions
STIL	Occurs in T cell leukaemias
TAL1	Regulates adult haematopoiesis
TCF3	Multi-lineage blood reconstitution
TCF7	HSC self renewal
TEK	Proliferation of primitive haematopoietic cells
TGFBI	Regulation of proliferation, quiescence and differentiation of HSCs
TLE3	Transcriptional repressor expressed in HSCs and megakaryocytes
TXNIP	Highly expressed in early HSCs and expression decreases as HSCs differentiate
YY1	Lineage differentiation and cell proliferation

Table 3.1: Gene target list. Genes of interest identified by literature search with brief description of gene function.

Publications consulted: Bernt et al., 2014; Bibi et al., 2014; Fielding et al., 2010; Giles et al., 2004; Grootens et al., 2019; Hussen et al., 2017; Kang et al., 2016; Notta et al., 2016; Olsson et al., 2015; Pane et al., 2002; Will et al., 2010; Ye et al., 2006; Ye et al., 2017 and Zhang et al., 2018.

3.1.2 Aims.

In view of the above, the aims of this chapter were to:

1. Identify transcriptomic similarities between proposed CML-like Ph+ALL and lymphoid blast phase CML and differences between standard Ph+ALL and CML
2. Identify transcriptomic differences between Ph+ driven cell activity in immature cells and Ph- immature cells
3. Identify transcriptomic differences between normal and Ph+ driven HSCs
4. Identify transcriptomic differences between good and poor responders
5. Identify possible defining genes of interest to elucidate the CML-like Ph+ALL subtype at diagnosis

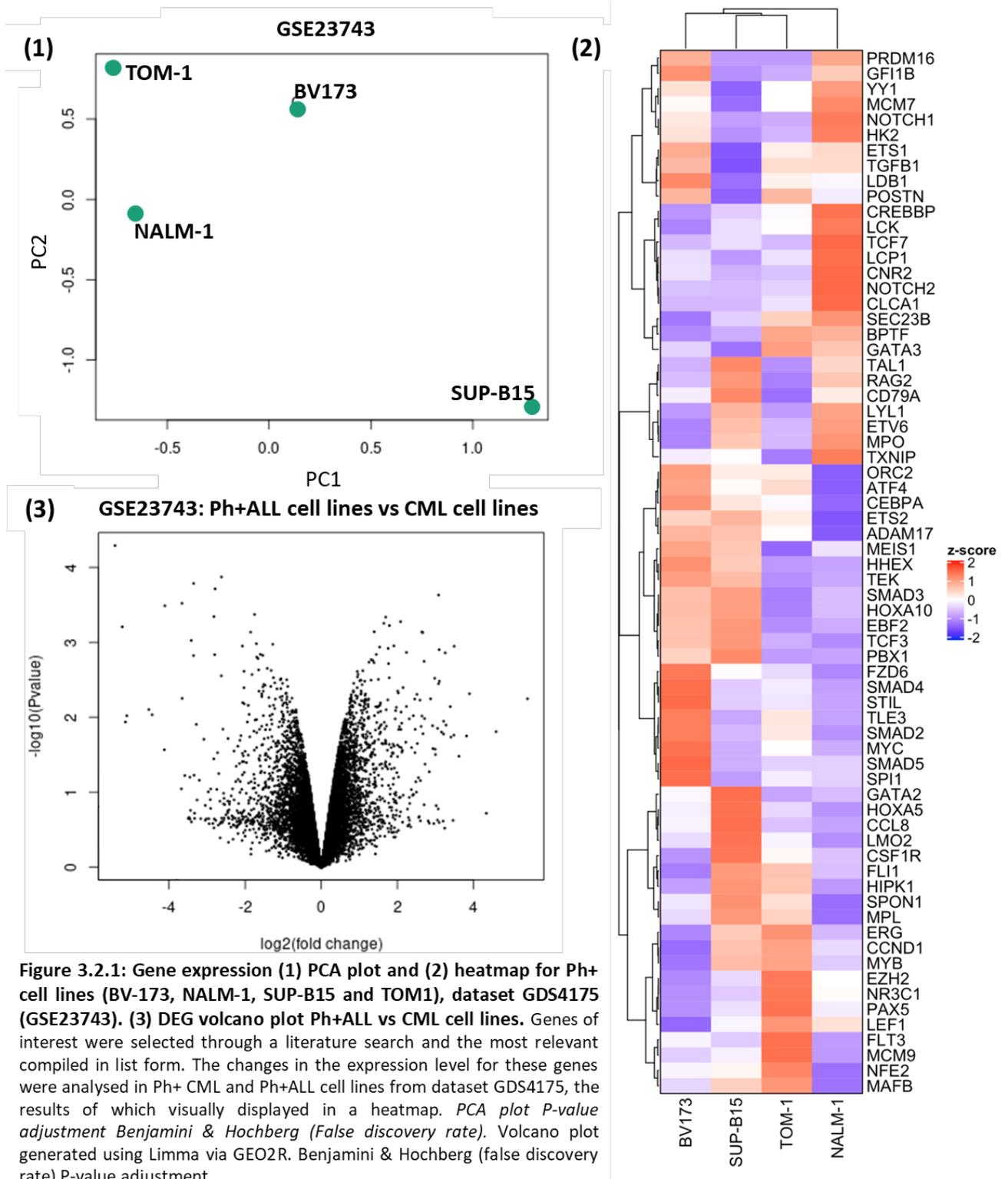
3.2 Results I: *BCR::ABL1* promotes aberrant expression of oncogenic genes, in addition to Ph+ALL-specific metabolic activity and lymphoid differentiation arrest

As has been previously discussed, the Ph chromosome is a common hallmark of both CML and Ph+ALL. However, despite sharing this fusion event, both diseases are phenotypically and clinically different. We first investigated the gene expression profiles of Ph+ cell lines to determine transcriptomic differences between the 2 disease types. This was to identify differences in the gene expression of *BCR::ABL1* occurring within different cells of origin; in an immature and mature progenitor (CML and Ph+ALL respectively).

Cell line name	Disease	BCR-ABL fusion variant	Age (years)	Sex	Additional information
BV-173	CML (BCP)	p210	45 (adult)	Male	Blast crisis
NALM-1	CML (BCP)	p210	3 (paediatric)	Female	Blast crisis
SUP-B15	ALL (BCP)	p210	9 (paediatric)	Male	Second relapse
TOM1	ALL (BCP)	p190	54 (adult)	Female	Refractory Ph+ALL

Table 3.2.1: Ph+ cell line background information, dataset GDS4175. Disease type, BCR-ABL fusion variant, age and sex of patient, disease stage and relapse status for cell lines investigated in GDS4175 (BV-173, NALM-1, SUP-B15 and TOM1).

In order to better understand the mechanisms of how these diseases differ, we first investigated gene expression in Ph+ cell lines within dataset GDS4175; this included both CML and Ph+ALL cell lines (Table 3.2.1), as well cell lines with paediatric and adult origin (Duy et al., 2011). Expression patterns observed were carried into analysis of other datasets in order to confirm observations and determine their use as potential diagnostic biomarkers. The cellular functions of relevant genes were investigated to determine functionality and how these two differing disease states may arise. Determining which genes can be attributed to age and diagnosis allows for the elucidation of the role of the Ph chromosome in cell activity and differential expression of genes between both disease states is essential for disease-specific activity. Genes with unchanged expression between disease states are deemed as being unrelated to disease-specific functions and would not be used for a biomarker panel. GDS4175 analysed 4 Ph+ cell lines, namely BV-173, NALM-1, SUP-B15, and TOM1 before and after treatment with Imatinib (Table 3.2.1). As TKI treatment and resistance is not a main focus of this project, treated samples were omitted from analysis and only untreated samples were analysed.



Gene expression in GDS4175 was profiled by array (Affymetric human genome U133A 2.0 Array) and the resultant count data uploaded to GEO (Duy et al., 2011). GDS4175 expression data was loaded into R from GEO (gene expression omnibus) and analysed using the GEO2R platform. GEO2R uses GEOquery and Limma (Linear Models for Microarray Analysis) to perform differential expression analysis of microarray datasets. GEOquery parses GEO datasets into R data structures which can then be used by other packages, in this analysis, Limma was used. Limma allows statistical testing for detection of differentially expressed genes and applied multiple-testing corrections on P-values to help correct for false-positive occurrences. Analysis was performed by entering the dataset ID in the GEO accession search bar. In the sample table, groups were defined by entering the specific group name and selecting the corresponding samples to be assigned to each group. To allow gene expression comparison between sample groups, at least two groups must be defined and compared. Once sample groups were defined, the 'Options' tab was selected to define testing conditions. Benjamini & Hochberg was selected for adjustment to the P values to account for false discovery rate. Log transformation was not automatically applied to analysis as the GEO database hosts logged and unlogged data, hence the auto-detect log function was used. Force normalisation was applied for quantile normalisation to the expression data making all selected samples have identical value distribution. The adjusted P-value threshold was set at 0.05 and the Log 2 fold change threshold was set at 1. After running the analyses (by selecting 'Analyze'), differential expression analysis of all DEGs (differentially expressed genes) was visualised as a PCA (principle component analysis) and volcano plot, heatmap and a table of 250 DEGs (Fig 3.2.1, Table 3.2.2 & 3.2.3).

We first demonstrated heterogeneity between cell lines (TOM-1, BV173, NALM-1 and SUP-B15) in GDS4175 through PCA plot (Fig 3.2.1.1). In view of the significant heterogeneity, we next utilised the previously selected gene list (Table 3.1) to generate gene expression variation within a heatmap (Fig 3.2.1.2). Heatmaps were generated in R by selecting significant genes of interest and converting normalised and scaling raw gene counts into a normalised z-score as previously described.

From the PCA plot, it could be observed that both CML cell lines (BV173 and NALM1) clustered closer (top left) than the Ph+ALL cell lines (TOM-1 and SUP-B15) (Fig 3.2.1.2). SUP-B15, in contrast, was located in the bottom right area of the PCA, suggesting an overall transcriptional difference in this cell line compared to the rest of the Ph+ cell lines in GDS4175. This displayed that despite TOM-1 and SUP-B15 being isolated from the same disease (Ph+ALL), overall gene expression between both cell lines differed. The PCA plot also displayed that BV173 and TOM-1, despite originating from two different diseases, both adult cell lines were transcriptionally similar, perhaps indicating age-related gene expression patterns shared across different diseases however more samples would need examined to verify this.

Gene	Gene function	logFC	P.Value
GPM6B	glycoprotein M6B	5.418	0.0056775
PDPN	podoplanin	3.895	0.0048667
CALD1	caldesmon 1	3.492	0.0011317
FHL1	four and a half LIM domains 1	3.39	0.0072211
CDK14	cyclin dependent kinase 14	3.257	0.001386
PECAM1	platelet and endothelial cell adhesion molecule 1	3.216	0.0113411
FHL1	four and a half LIM domains 1	3.103	0.0033058
ATP6V1B2	ATPase H+ transporting V1 subunit B2	3.097	0.0012584
LGR5	leucine rich repeat containing G protein-coupled receptor 5	3.084	0.0002337
S100A4	S100 calcium binding protein A4	3.083	0.0035792
FHL1	four and a half LIM domains 1	2.981	0.0030226
CSRP2	cysteine and glycine rich protein 2	2.874	0.0056949
GPM6B	glycoprotein M6B	2.715	0.0104514
SHOX2	short stature homeobox 2	2.701	0.0017992
NELL1	neural EGFL like 1	2.663	0.000744
TSC22D1	TSC22 domain family member 1	2.647	0.0096692
ARHGEF10	Rho guanine nucleotide exchange factor 10	2.593	0.005917
SLC35E3	solute carrier family 35 member E3	2.514	0.0049689
FAM134B	family with sequence similarity 134 member B	2.456	0.0057243
FHL1	four and a half LIM domains 1	2.419	0.0095828
CCPG1	cell cycle progression 1	2.313	0.0105249
FAM171A1	family with sequence similarity 171 member A1	2.24	0.0013677
PLEKHA5	pleckstrin homology domain containing A5	2.239	0.0071874
PLVAP	plasmalemma vesicle associated protein	2.15	0.0040212
RECK	reversion inducing cysteine rich protein with kazal motifs	2.097	0.0085311
MDM2	MDM2 proto-oncogene	2.084	0.0005305
PODXL	podocalyxin like	2.022	0.0047595
GNAI1	G protein subunit alpha i1	2.009	0.004486
LZTS1	leucine zipper tumor suppressor 1	1.973	0.0073996
ABAT	4-aminobutyrate aminotransferase	1.953	0.008205
PPP3CC	protein phosphatase 3 catalytic subunit gamma	1.943	0.0105236
MYO10	myosin X	1.909	0.0011197
DAB2	clathrin adaptor protein	1.792	0.0005987
SNRPN	small nuclear ribonucleoprotein polypeptide N	1.786	0.0029824
PPP3CC	protein phosphatase 3 catalytic subunit gamma	1.768	0.0019319
SLC7A11	solute carrier family 7 member 11	1.747	0.0058757
IGFBP7	insulin like growth factor binding protein 7	1.706	0.0079721
NMT2	N-myristoyltransferase 2	1.692	0.000462
ABAT	4-aminobutyrate aminotransferase	1.656	0.0117264
PTPN14	protein tyrosine phosphatase, non-receptor type 14	1.645	0.0005596
IKZF1	IKAROS family zinc finger 1	1.562	0.0083624
SATB1	SATB homeobox 1	1.555	0.0064535
ASAP2	ArfGAP with SH3 domain, ankyrin repeat and PH domain 2	1.534	0.0011597
SLC16A2	solute carrier family 16 member 2	1.505	0.0071539
IGFBP7	insulin like growth factor binding protein 7	1.501	0.0031341
ATRNL1	attractin like 1	1.494	0.0086483
LEPROTL1	leptin receptor overlapping transcript-like 1	1.477	0.0073581
ERG	ERG, ETS transcription factor	1.458	0.0117822
SEPHS1	selenophosphate synthetase 1	1.445	0.0014549
ARHGEF6	Rac/Cdc42 guanine nucleotide exchange factor 6	1.432	0.0072267

Table 3.2.2: DEGs in Ph+ cell lines (SUP-B15 and TOM1) vs CML cell lines (BV173 and NALM-1), dataset GDS4175 (GSE23743). DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. *LogFC*= log fold change.

Gene	Gene function	logFC	P.Value
TMEM5	transmembrane protein 5	-1.256	0.0029747
GPI	glucose-6-phosphate isomerase	-1.274	0.0010632
STEAP3	STEAP3 metalloredutase	-1.285	0.0091157
AHR	aryl hydrocarbon receptor	-1.311	0.0108041
CDC42BPA	CDC42 binding protein kinase alpha	-1.315	0.0023827
SEMA4C	semaphorin 4C	-1.325	0.0038846
TUBG1	tubulin gamma 1	-1.349	0.0095849
SDHAF3	succinate dehydrogenase complex assembly factor 3	-1.353	0.0071196
ARID3B	AT-rich interaction domain 3B	-1.377	0.005474
NAMPT	nicotinamide phosphoribosyltransferase	-1.418	0.0022315
FECH	ferrochelatase	-1.476	0.0037464
RNFT2	ring finger protein, transmembrane 2	-1.51	0.003654
DENND1C	DENN domain containing 1C	-1.52	0.0095463
TRIP10	thyroid hormone receptor interactor 10	-1.524	0.0044948
RGS16	regulator of G-protein signaling 16	-1.535	0.0039398
KLHL2	kelch like family member 2	-1.537	0.0014956
COX17	cytochrome c oxidase copper chaperone	-1.563	0.0049066
ZFP36	ZFP36 ring finger protein	-1.563	0.0053141
DDB2	damage specific DNA binding protein 2	-1.599	0.0014516
ERI2	ER1 exoribonuclease family member 2	-1.636	0.0044983
DHRS7	dehydrogenase/reductase 7	-1.645	0.0023392
RGS16	regulator of G-protein signaling 16	-1.663	0.0016457
MYO1F	myosin 1F	-1.696	0.0015761
ITPR3	inositol 1,4,5-trisphosphate receptor type 3	-1.697	0.0107703
YES1	YES proto-oncogene 1, Src family tyrosine kinase	-1.706	0.0010467
RAB29	RAB29, member RAS oncogene family	-1.738	0.0109265
AGL	amylase, alpha-1, 6-glycosidase, 4-alpha-glucanotransferase	-1.752	0.0004258
BTG1	BTG anti-proliferation factor 1	-1.819	0.0026969
BMP2	bone morphogenetic protein 2	-1.856	0.0007326
ARFIP1	ADP ribosylation factor interacting protein 1	-1.894	0.0092012
TRAF5	TNF receptor associated factor 5	-1.918	0.0069357
ZMAT3	zinc finger matrix-type 3	-1.939	0.0115969
TNNT1	troponin T1, slow skeletal type	-2.023	0.0066296
RAD50	RAD50 double strand break repair protein	-2.029	0.0024441
RAB29	RAB29, member RAS oncogene family	-2.03	0.0040453
TNF	tumor necrosis factor	-2.065	0.0064056
BTG1	BTG anti-proliferation factor 1	-2.071	0.0011251
KLF11	Kruppel like factor 11	-2.622	0.0001343
YES1	YES proto-oncogene 1, Src family tyrosine kinase	-2.8	0.0014643
CHFR	checkpoint with forkhead and ring finger domains, E3 ubiquitin protein ligase	-2.82	0.0004544
CD44	CD44 molecule (Indian blood group)	-3.353	0.0001639
CDC42BPA	CDC42 binding protein kinase alpha	-3.414	0.0009458
BHLHE40	basic helix-loop-helix family member e40	-3.656	0.005619
APOBEC3B	apolipoprotein B mRNA editing enzyme catalytic subunit 3B	-3.659	0.0003022
CD44	CD44 molecule (Indian blood group)	-4.108	0.0003259
IGK	immunoglobulin kappa locus	-4.443	0.0092992
IGKC	immunoglobulin kappa constant	-4.529	0.0079181
MIR8071-2	microRNA 8071-2	-5.109	0.0095928
RGS1	regulator of G-protein signaling 1	-5.145	0.0116062
PRPS2	phosphoribosyl pyrophosphate synthetase 2	-5.418	0.0000511

Table 3.2.3: Bottom 50 downregulated DEGs in Ph+ cell lines (SUP-B15 and TOM1) vs CML cell lines (BV173 and NALM-1), dataset GDS4175 (GSE23743). DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. *LogFC*= *log fold change*.

From the gene of interest list generated by literature search, *MPL* was found to be upregulated in Ph+ALL cell lines (SUP-B15 and TOM-1) compared to blast phase cell lines (Fig 3.2.1.2 & 3.2.2). The *MPL* gene encodes for the thrombopoietin receptor protein which promotes cell proliferation and maintenance of HSCs. In a disease context, the upregulation of this gene may be explained by the high proliferation rate of Ph+ALL blast cells in comparison to lymphoid blast crisis CML and may suggest the maintenance of LSCs in addition to HSCs, though this is yet to be elucidated in Ph+ALL. Mutations in *MPL* have been observed in Ph- myeloproliferative neoplasms, suggesting that *MPL* activity may be required within Ph+ leukaemia (Eldweny et al., 2019).

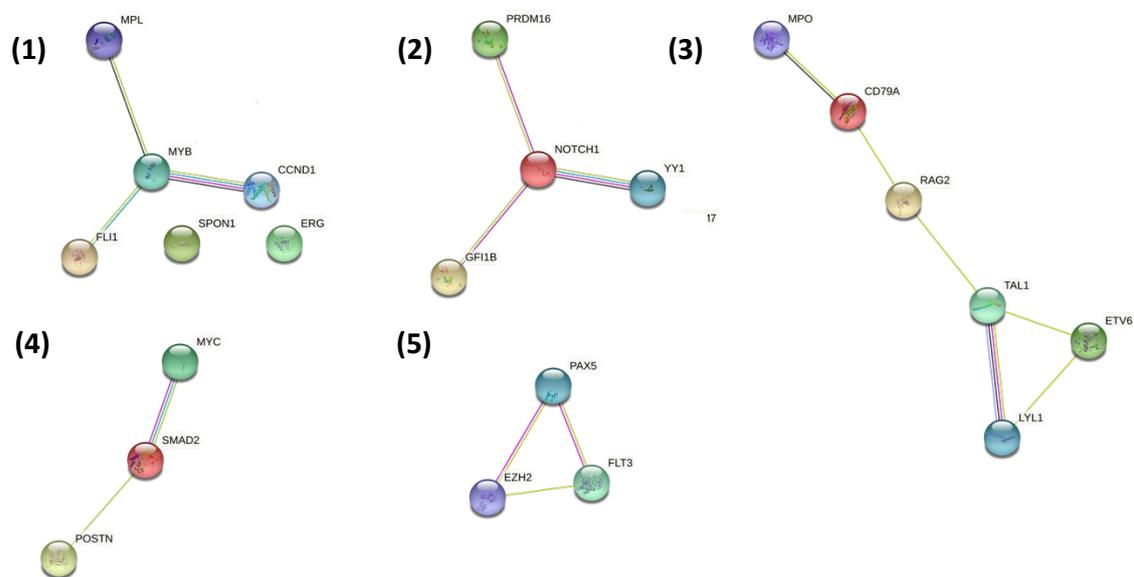


Figure 3.2.2: Functional enrichment analysis of genes differentially regulated in Ph+ALL cell lines (SUP-B15 and TOM1), dataset GDS4175. Genes of interest were selected through a literature search and the most relevant compiled in list form. The changes in the expression level for these genes were analysed in Ph+ cell lines from dataset GDS4175, the results of which visually displayed in a heatmap. From the heatmap produced, genes observed to be differentially regulated in Ph+ALL cell lines were selected, and a STRING gene-gene interaction network produced to display the functional relevance of gene expression changes in Ph+ALL cell lines. (1) Genes upregulated in Ph+ALL cell lines SUP-B15 and TOM1. (2) Genes downregulated in Ph+ALL cell lines SUP-B15 and TOM1. (3) Genes upregulated in Ph+ cell lines isolated from paediatric patients (NALM-1 and SUP-B15). (4) Genes downregulated in Ph+ cell lines isolated from paediatric patients (NALM-1 and SUP-B15). (5) Genes upregulated in a Ph+ cell line containing the p190 BCR::ABL1 fusion variant (TOM1).

HIPK1 (homeodomain-interacting protein kinase 1) is ubiquitously expressed in bone marrow however, its role in ALL has been observed through its antagonism of the master regulator of B cell development, *PAX5* (Nebral et al., 2009). *PAX5* is required for the development of B cells from Pro-B cells to committed B cell precursors, the loss of *PAX5* activity causing the arrest of B cell development. B-ALL blast cells exemplify this differentiation arrest by expression of pro- and pre-B cell surface marker expression. Upregulation of *HIPK1* expression in Ph+ALL cell lines indicates B cell developmental arrest, a feature of blast cell activity and an area to explore further.

The Ets-related gene (*ERG*) was observed to be upregulated in Ph+ALL cell lines compared to blast phase cells. Its role in the promotion and maintenance of leukaemia has previously been observed in paediatric acute megakaryocytic leukaemia and is associated with poor prognosis in adult T-cell ALL (Tsuzuki et al., 2011). The importance of *ERG* transcription factor activity in both myeloid and lymphoid leukaemia evidences similarities in functional activities between leukaemias of different lineages and suggests that Ph+ALL and CML may have shared leukaemic origins. Though the cell of *BCR::ABL1* fusion origin in CML has been identified as the HSC, the cell of origin in Ph+ALL remains elusive. Despite this, the reciprocal translocation event which results in *BCR::ABL1* fusion is, indeed, a shared leukaemic initiating event between CML and Ph+ALL.

Cyclin D1 (*CCND1*) was found to be upregulated in the Ph+ALL cell lines compared to blast phase; this observation aligns with primary ALL bone marrow expression patterns (Fernandes et al., 2018). The protein which *CCND1* encodes is essential for the transition from the G1 to the S phase of the cell cycle and hence, the dysregulation of *CCND1* has oncogenic potential. It was found to be upregulated in ALL patients, with highest expression in the presence of *BCR::ABL1* and older age. High *CCND1* levels are associated with poorer clinical outcomes and is understood to play a role in the malignant phenotype of ALL.

Increased expression of *SPON1* is associated with poorer overall survival rate in AML and was upregulated in both Ph+ALL cell lines in GDS4175 compared to blast phase cell lines (Wang et al., 2020). Its expression in solid organ tumours (such as ovarian high grade serous carcinoma) brings about enhanced cell growth and decreased apoptosis (Nagasawa et al., 2022). The involvement of *SPON1* in lymphoid, myeloid and ovarian cancers suggests that its expression is likely common in acute cancers and not associated to a specific cell lineage.

FLI1 (friend leukaemia virus integration 1), an Ets transcription factor family member, was upregulated in Ph+ALL cell lines SUPB15 and TOM1. The role of *FLI1* in AML outcomes has been well reported and increased expression levels are associated with proliferation and stromal interaction, and high *FLI1* levels are observed to be prognostically adverse (Kornblau et al., 2011). Due to its normal role in stem cell regulation and haematopoiesis, *FLI1* being

expressed similarly in acute myeloid and lymphoid leukaemia may indicate aberrant expression occurring at an early stage of haematopoiesis prior to lineage bifurcation.

MYB was upregulated in SUPB15 and TOM1. *MYB* is aberrantly expressed in leukaemia and is critical for the development of myeloid leukaemia (Nguyen et al., 2016). *MYB* encodes a transcription factor essential for the control of early haematopoiesis and hence, aberrant expression occurs prior to lineage bifurcation, resulting in its overexpression in both AML and ALL.

Genes found to be downregulated in the Ph+ALL cell lines SUPB15 and TOM1 include *NOTCH1*, the gene encoding the NOTCH1 signalling protein associated with cell fate specification and differentiation, the histone acetyltransferase *YY1*, *MCM7* which initiates genome replication by being a key component in the pre-replication complex and *GFI1B* which plays a role in the development and maturation of erythrocytes and megakaryocytes (Fig 3.2.1.2). The downregulation of *GFI1B* in a lymphoid leukaemia aligns with the lineage commitment of ALL cells, as does the downregulation of *PRDM16*, a gene which plays an important role in the pathogenesis of AML, a myeloid leukaemia.

HK2 was observed to be downregulated in Ph+ALL cell lines, its role in glucose metabolism aligning with the overall picture altered metabolomic homeostasis characteristic of Ph+ALL pathology. Hexokinases phosphorylate glucose to produce glucose-6-phosphate, the first step in the glucose metabolism pathway. CNS (central nervous system) involvement in ALL is a well-established clinical outcome and has prompted a large amount of research into the metabolic and transcriptomic alterations made by blasts in order to survive the low nutrient CNS microenvironment (Lenk et al., 2020). Previous studies have observed an upregulation in *HK2* expression in CNS-derived leukaemic cells which had adapted to oxygen-poor hypoxic conditions (Kato et al., 2017). The observations in GDS4175 seemingly contradict the literature however, both SUPB15 and TOM1 are both bone marrow derived cell lines, suggesting that ALL blasts are able to adapt to their microenvironment.

To further explore the phenotypic similarities between Ph+ALL and CML cell lines isolated from paediatric patients, the gene expression in paediatric Ph+ cell lines NALM-1 and SUP-

B15 were compared to adult cell lines BV173 (CML) and TOM1 (Ph+ALL). Both cell lines originate from paediatric patients whose ages fall into the range of ages of the BMNC samples (CellBank) utilised throughout the rest of this project (CellBank cohort range= 2- 19 years; NALM-1= 3 years; SUP-B15= 9 years). While other childhood ALL datasets are available (ie Roy et al., 2021 and Ottersbach et al., 2021), the focus of such datasets are on foetal and infant MLL-AF4 ALL, ages and disease types outwith the scope of this project. In addition, GDS4175 provided an opportunity to compare gene expression in adult and paediatric Ph+-driven leukaemia. As will be discussed in later chapters, adult CML samples were utilised in RNAseq transcriptional analysis due to the rarity of CML in the paediatric cohort. GDS4175 therefore contains gene expression data from similar samples to those used for RNAseq analyses in this project.

Significant genes found to be upregulated in paediatric cell lines compared to adult broadly fell into three functional categories: haematopoiesis and vascular function (*ETV6*, *TAL1* and *LYL1*, a paralog of *TAL1*), B cell receptor recombination and function (*RAG2* and *CD79A*) and granulocyte monocyte function (*MPO*) (Fig 3.2.1). However, as only one example was available for each category (adult vs paediatric and CML vs Ph+ALL), attribution of gene expression to age could not be made as other factors such as sex and secondary genetic abnormalities may contribute to gene expression.

The normal function of *ETV6* is that of haematopoietic maintenance and vasculogenesis, however, *ETV6* plays an important role in malignant transformation. The association between *ETV6* and ALL has been described within the context of *ETV6* forming a fusion gene with *RUNX1*, resulting in leukaemogenic kinase activity (Hock and Shimamura, 2017). As well as forming an oncogenic fusion protein with *RUNX1*, *ETV6* (also known as *TEL*) has been observed to inappropriately activate kinases through fusion with *ABL*, forming *TEL-ABL* (Voss et al., 2000). Kinases *TEL-ABL* and *BCR::ABL1* are able to similarly utilise signalling pathways (namely MAPK/Erk kinase and Akt/PKB) to negatively regulate apoptosis however, both fusion proteins have differing preferences for synthetic substrate peptides. In addition to *TEL-ABL* being observed in B-ALL and T-ALL, it has also been detected in Ph-negative CML (termed 'CML-like myeloproliferative disease') and AML (Auger et al., 2012). With similarities in kinase activity and signal transduction pathways shared between *BCR::ABL1* and *TEL-ABL* and the

capacity to initiate disease with CML-like properties, it displays the potential for development of myeloid-driven disease by *ETV6* activity.

Haematopoiesis and vasculogenesis related genes upregulated in paediatric cell lines (*ETV6*, *TAL1* and *LYL1*) may be due to the period of growth and development underway in the paediatric patients the cell lines originally arose from. The upregulation of genes associated with B cell development and receptor rearrangement (*RAG2* and *CD79A*) is notable considering the inclusion of a CML cell line. These findings display an increased level of B cell receptor rearrangement in paediatric cell lines and is supported by the observation of the highest number of plasma and memory B cells in ages 17 years and younger (Blanco et al., 2018). Pathology may also play a role in these gene expression patterns, with 89% of B cell precursor ALL displaying IGH V-D-J rearrangements (Brumpt et al., 2000).

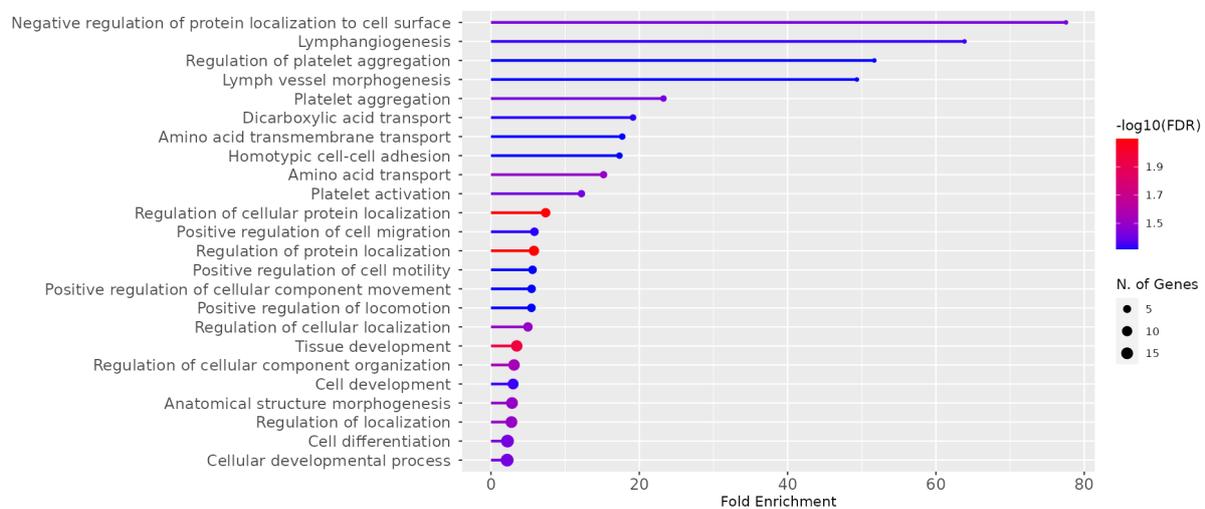
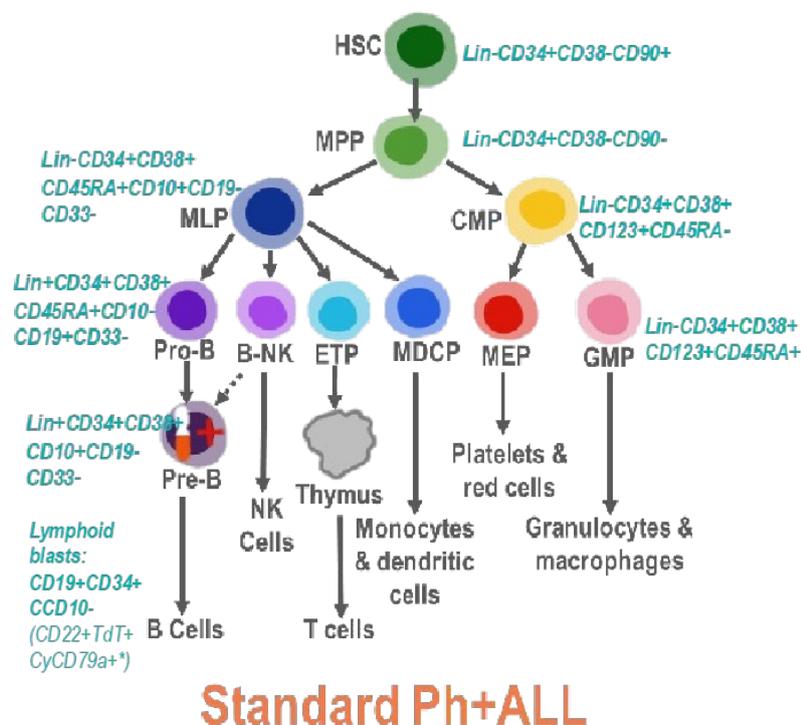


Figure 3.2.3: Gene ontology enrichment analysis for top 50 upregulated DEGs in Ph+ cell lines (SUP-B15 and TOM1) vs CML cell lines (BV173 and NALM-1), dataset GDS4175 (GSE23743). DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. $\text{LogFC} = \log \text{fold change}$. ShinyGO v0.741.

3.3 Results I: Profiling of CML and healthy Lin- progenitor cells displayed a CML-specific transcriptional signature, and highlighted leukaemic-associated genes expressed similarly in CML and Ph+ALL.

We next aimed to investigate transcriptomic differences between Ph+ immature progenitor cells and progenitors with no BCR::ABL1 fusion. To investigate this, gene expression patterns in diagnostic CML samples (Lin-CD34-, Lin-CD34+, Lin+CD34+) were compared to healthy controls using GSE11675 (Lemoli et al., 2009). Since gene expression profiles of CML-like Ph+ALL have yet to be established, it was felt imperative to understand the gene expression profiles of *BCR::ABL1* positive stem cells at different stages of maturity to allow comparison with Ph+ALL disease which typically is thought to arise in a lymphoid progenitor population (Fig 3.3.1). As CML-like Ph+ALL has been observed to possess phenotypic similarities with CML and Ph+ALL by the presence of myeloid disease involvement, gene expression profiling of this subtype may reflect a mixed transcriptional signature, exhibiting a genotype with features of both diseases. The caveat to this is that GSE11675 utilised chronic phase CML samples. There were no datasets that utilised blast phase HSCs in adequate numbers.



**clinical markers, not used in this project (Chiaretti et al., 2014)*

Figure 3.3.1: Haematopoietic population surface marker expression for FACS. Surface marker expression profile used for cell sorting based off those used clinically.

Broadly, genes related to CML-specific cellular functions were expected to be expressed similarly in primary CML patient samples belonging to dataset GSE11675 and CML cell lines in dataset GDS4175. Comparison is made with immortalised cell lines and primary progenitor cell samples. Unlike cell lines, primary samples are more heterogenous transcriptomically and proteomically.

Gene expression was profiled by array and expression data was loaded into R from GEO (gene expression omnibus) and analysed using the GEO2R platform as outlined in section 3.2. Of the genes of interest (Table 3.1), the following were upregulated in CML Lin-CD34+ compared to healthy Lin-CD34+ cells. These genes may be relevant in the investigation of heterogeneity of Ph+ALL. As LSCs have not been identified in Ph+ALL but may play a role in CML-like Ph+ALL pathology, comparison of a stem cells from patients disease with known LSC activity to healthy stem cells may highlight genes helpful for discriminating samples with LSC from samples without stem cell activity. Though identification of LSC activity would require stem cell activity assays, these genes could be used to provide an indication of LSC presence transcriptionally.

ATF4 was downregulated in CML stem cells. Its downregulation allows cancer cells to survive under amino acid deprivation, allowing leukaemic cells to evade cell death (Mesclon, 2017). *CLCA1* was downregulated in CML stem cells. It has a suggested role in pro-inflammatory response and lymphocyte-mediated lymphangiogenesis (Jordan-Williams et al., 2016). Lymphangiogenesis is the developmental process responsible for the formation of new lymphatic vessels. Under homeostatic conditions, lymphangiogenesis occurs during embryonic development and healing in response to injury or pathological injury however, this process can be dysregulated in a variety of cancers. In solid tumours, the formation of new blood vessels from pre-existing lymphatics can induce metastasis, in lymphoma this developmental process can be subverted to allow malignant cells to invade lymph nodes and lymphatic tissues through lymphatic vessels (Kadowaki et al., 2005; Stacker et al., 2014). Downregulation of genes associated with lymphocyte-mediated lymph vessel development and inflammation in CML progenitor cells demonstrates the quiescent nature of CML LSC biology, a mechanism that allows evasion of LSCs from targeting by TKI treatment. *CNR2* is

expressed predominantly on differentiated B cells, the downregulation of which may be indicative of the stem cell driven nature of disease (Gruber et al., 2021).

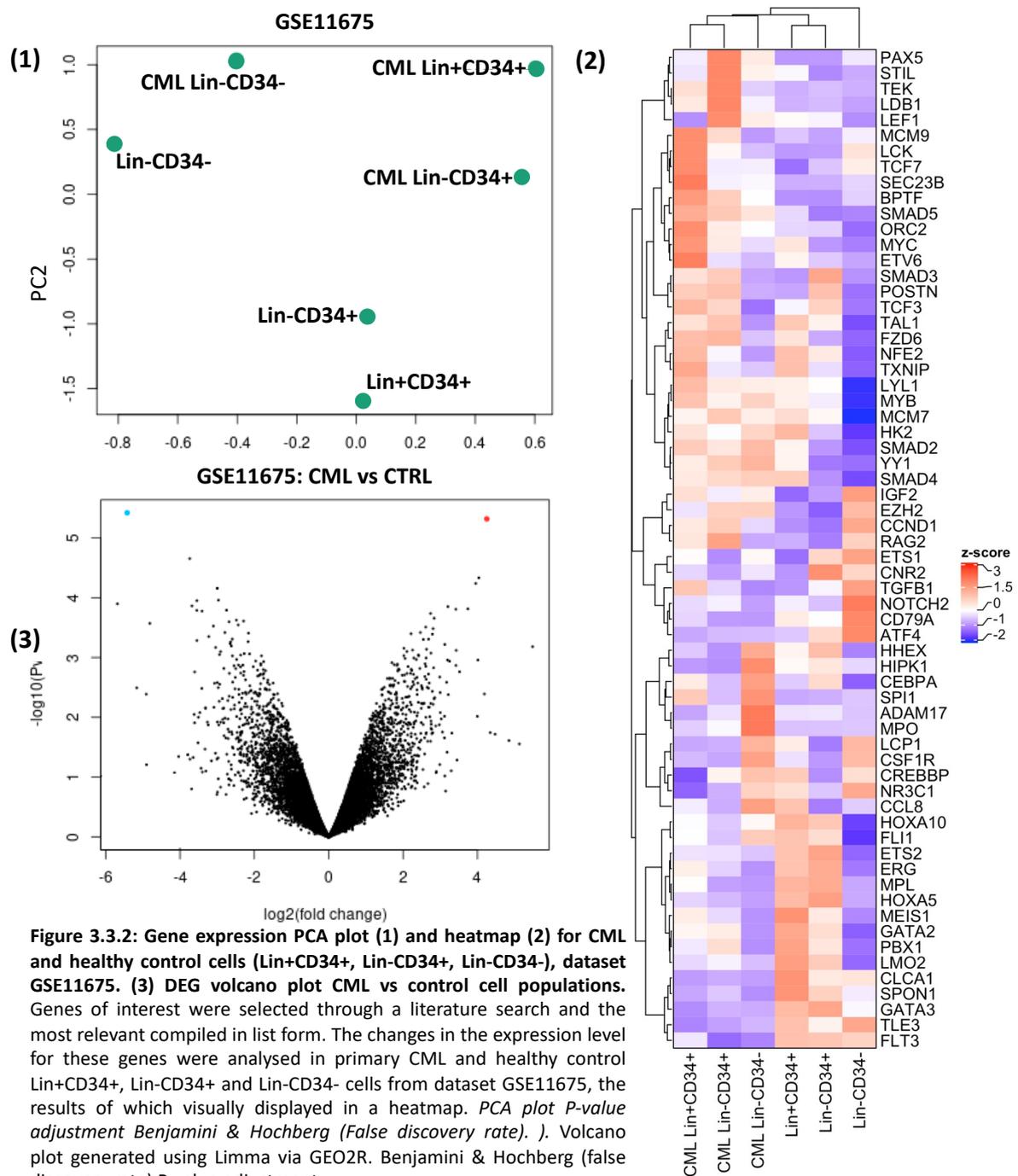


Figure 3.3.2: Gene expression PCA plot (1) and heatmap (2) for CML and healthy control cells (Lin+CD34+, Lin-CD34+, Lin-CD34-), dataset GSE11675. (3) DEG volcano plot CML vs control cell populations. Genes of interest were selected through a literature search and the most relevant compiled in list form. The changes in the expression level for these genes were analysed in primary CML and healthy control Lin+CD34+, Lin-CD34+ and Lin-CD34- cells from dataset GSE11675, the results of which visually displayed in a heatmap. PCA plot P-value adjustment Benjamini & Hochberg (False discovery rate). Volcano plot generated using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment.

Gene	Gene function	logFC	P.Value
IGFBP2	insulin like growth factor binding protein 2	5.48	0.0006587
HNRNPU	heterogeneous nuclear ribonucleoprotein U	4.25	0.00000481
HBG2	hemoglobin subunit gamma 2	4.19	0.00406875
HBBP1	hemoglobin subunit beta pseudogene 1	4.04	0.0000466
RHAG	Rh-associated glycoprotein	4.01	0.00110429
PLK1	polo like kinase 1	3.95	0.0000578
GATAD1	GATA zinc finger domain containing 1	3.74	0.00015419
ZBTB7A	zinc finger and BTB domain containing 7A	3.52	0.00173108
TGM2	transglutaminase 2	3.43	0.00075995
EPB41	erythrocyte membrane protein band 4.1	3.42	0.00015611
PTPN11	protein tyrosine phosphatase, non-receptor type 11	3.42	0.00135812
CYP7B1	cytochrome P450 family 7 subfamily B member 1	3.25	0.00115089
IL1RAP	interleukin 1 receptor accessory protein	3.22	0.00418749
RASAL2	RAS protein activator like 2	3.2	0.00110425
SNORD65	small nucleolar RNA, C/D box 65	3.17	0.00259754
FAM13C	family with sequence similarity 13 member C	3.12	0.00060397
SPC25	SPC25, NDC80 kinetochore complex component	3	0.00386192
SAP30	Sin3A associated protein 30	2.99	0.0021199
EPX	eosinophil peroxidase	2.85	0.0047076
RAB6B	RAB6B, member RAS oncogene family	2.84	0.0002203
DICER1	dicer 1, ribonuclease III	2.83	0.00490218
SELENBP1	selenium binding protein 1	2.82	0.00148835
NFATC4	nuclear factor of activated T-cells 4	2.81	0.00032916
HBD	hemoglobin subunit delta	2.81	0.00606344
ITPKA	inositol-trisphosphate 3-kinase A	2.79	0.00040871
SLC7A1	solute carrier family 7 member 1	2.79	0.00368028
STAR	steroidogenic acute regulatory protein	2.75	0.00084281
SLC26A2	solute carrier family 26 member 2	2.74	0.00018331
HBB	hemoglobin subunit beta	2.74	0.00585362
EMILIN1	elastin microfibril interfacer 1	2.7	0.00267898
MAP3K9	mitogen-activated protein kinase kinase kinase 9	2.69	0.00189069
TMPRSS11D	transmembrane protease, serine 11D	2.68	0.00049461
KIF14	kinesin family member 14	2.68	0.00326557
KIF11	kinesin family member 11	2.65	0.00314987
CCDC144A	coiled-coil domain containing 144A	2.63	0.00072856
NUP50	nucleoporin 50	2.62	0.00101789
TFRC	transferrin receptor	2.62	0.00231183
ZNF148	zinc finger protein 148	2.61	0.00031017
ATXN7	ataxin 7	2.61	0.00140922
HNRNPDL	heterogeneous nuclear ribonucleoprotein D like	2.59	0.00127421
PRG2	proteoglycan 2, pro eosinophil major basic protein	2.59	0.00207897
EAPP	E2F associated phosphoprotein	2.57	0.00582722
TFRC	transferrin receptor	2.55	0.00050006
TEAD4	TEA domain transcription factor 4	2.55	0.00583976
ARID1A	AT-rich interaction domain 1A	2.53	0.00597062
TOP2B	topoisomerase (DNA) II beta	2.52	0.00337619
C18orf25	chromosome 18 open reading frame 25	2.51	0.00322123
LGR5	leucine rich repeat containing G protein-coupled receptor 5	2.5	0.00297784
TTK	TTK protein kinase	2.49	0.00583316
TMX4	thioredoxin related transmembrane protein 4	2.46	0.00038178

Table 3.3.1: Top 50 upregulated DEGs in CML vs healthy control cells (Lin+CD34+, Lin-CD34+, Lin-CD34-), dataset GSE11675. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. *LogFC= log fold change.*

Gene	Gene function	logFC	P.Value
DST	dystonin	-2.34	0.00478222
HIST2H2AA4	histone cluster 2, H2aa4	-2.37	0.00091069
ELN	elastin	-2.37	0.00143788
HIST1H2BK	histone cluster 1, H2bk	-2.41	0.00414575
COBL	cordon-bleu WH2 repeat protein	-2.41	0.00565469
LAMA4	laminin subunit alpha 4	-2.42	0.00339214
HIST1H2BC	histone cluster 1, H2bc	-2.43	0.00028484
IL23A	interleukin 23 subunit alpha	-2.43	0.00133685
HIST1H2BL	histone cluster 1, H2bl	-2.48	0.00023658
RUNX1T1	RUNX1 translocation partner 1	-2.48	0.00107822
KCND1	potassium voltage-gated channel subfamily D member 1	-2.5	0.00152965
TAP2	transporter 2, ATP binding cassette subfamily B member	-2.51	0.00215183
KHDRBS2	KH RNA binding domain containing, signal transduction associated 2	-2.52	0.00334245
CD79A	CD79a molecule	-2.53	0.00625381
TCF20	transcription factor 20	-2.55	0.00137284
HIST1H2AE	histone cluster 1, H2ae	-2.58	0.00258009
BMX	BMX non-receptor tyrosine kinase	-2.58	0.00465029
ZNF174	zinc finger protein 174	-2.59	0.00049203
NR2F1	nuclear receptor subfamily 2 group F member 1	-2.59	0.00093826
CD53	CD53 molecule	-2.65	0.00058521
KLF4	Kruppel like factor 4	-2.68	0.00444966
ZNF165	zinc finger protein 165	-2.71	0.00524467
CYP2C8	cytochrome P450 family 2 subfamily C member 8	-2.75	0.00016154
HIST3H2BB	histone cluster 3, H2bb	-2.79	0.00038216
TRPA1	transient receptor potential cation channel subfamily A member 1	-2.82	0.00152271
ID1	inhibitor of DNA binding 1, HLH protein	-2.86	0.0048625
HCG26	HLA complex group 26 (non-protein coding)	-2.91	0.00079747
TPM2	tropomyosin 2 (beta)	-2.93	0.00279588
EMP1	epithelial membrane protein 1	-2.94	0.00089281
PPP3CA	protein phosphatase 3 catalytic subunit alpha	-2.94	0.00540462
HIST2H2BE	histone cluster 2, H2be	-2.97	0.00010966
GSTM3	glutathione S-transferase mu 3	-2.97	0.00319124
IGHM	immunoglobulin heavy constant mu	-3	0.00006974
PBX2	PBX homeobox 2	-3	0.00263057
IGHM	immunoglobulin heavy constant mu	-3.04	0.00067978
DNTT	DNA nucleotidylexotransferase	-3.07	0.00270039
HIST1H2AC	histone cluster 1, H2ac	-3.1	0.00029822
GRM5	glutamate metabotropic receptor 5	-3.23	0.00044648
AXL	AXL receptor tyrosine kinase	-3.4	0.00016792
SMARCD3	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3	-3.41	0.0032856
LIMCH1	LIM and calponin homology domains 1	-3.54	0.00011192
HIST1H1C	histone cluster 1, H1c	-3.54	0.00051918
FOLH1B	folate hydrolase 1B	-3.55	0.00016288
TMEM268	transmembrane protein 268	-3.6	0.0017453
CRTAM	cytotoxic and regulatory T-cell molecule	-3.68	0.00013734
SKIL	SKI-like proto-oncogene	-3.74	0.00002213
NPTX2	neuronal pentraxin 2	-4.81	0.0002687
MIR22	microRNA 22	-4.91	0.00409585
BLNK	B-cell linker	-5.16	0.00321616
MAPT	microtubule associated protein tau	-5.42	0.0000038
SETBP1	SET binding protein 1	-5.68	0.00012689

Table 3.3.2: Bottom 50 downregulated DEGs in CML vs healthy control cells (Lin+CD34+, Lin-CD34+, Lin-CD34-), dataset GSE11675. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. *LogFC*= log fold change.

A high expression of *ETS2* is associated with poor prognosis in AML and is upregulated in acute leukaemia with poor outcomes, therefore a high expression in Ph+ALL would be anticipated (Fu et al., 2017). *ETS2* encodes a transcription factor which regulates apoptosis and cell development genes. In addition to this, the encoded protein has been evidenced to regulate the function of telomerase and due to this pro-survival gene function, is considered a protooncogene. The observation of the downregulation of *ETS1* is in contrast to what is found in the literature where *ETS1* is significantly upregulated in CML patients via action of *BCR::ABL1* (Desterke et al., 2018).

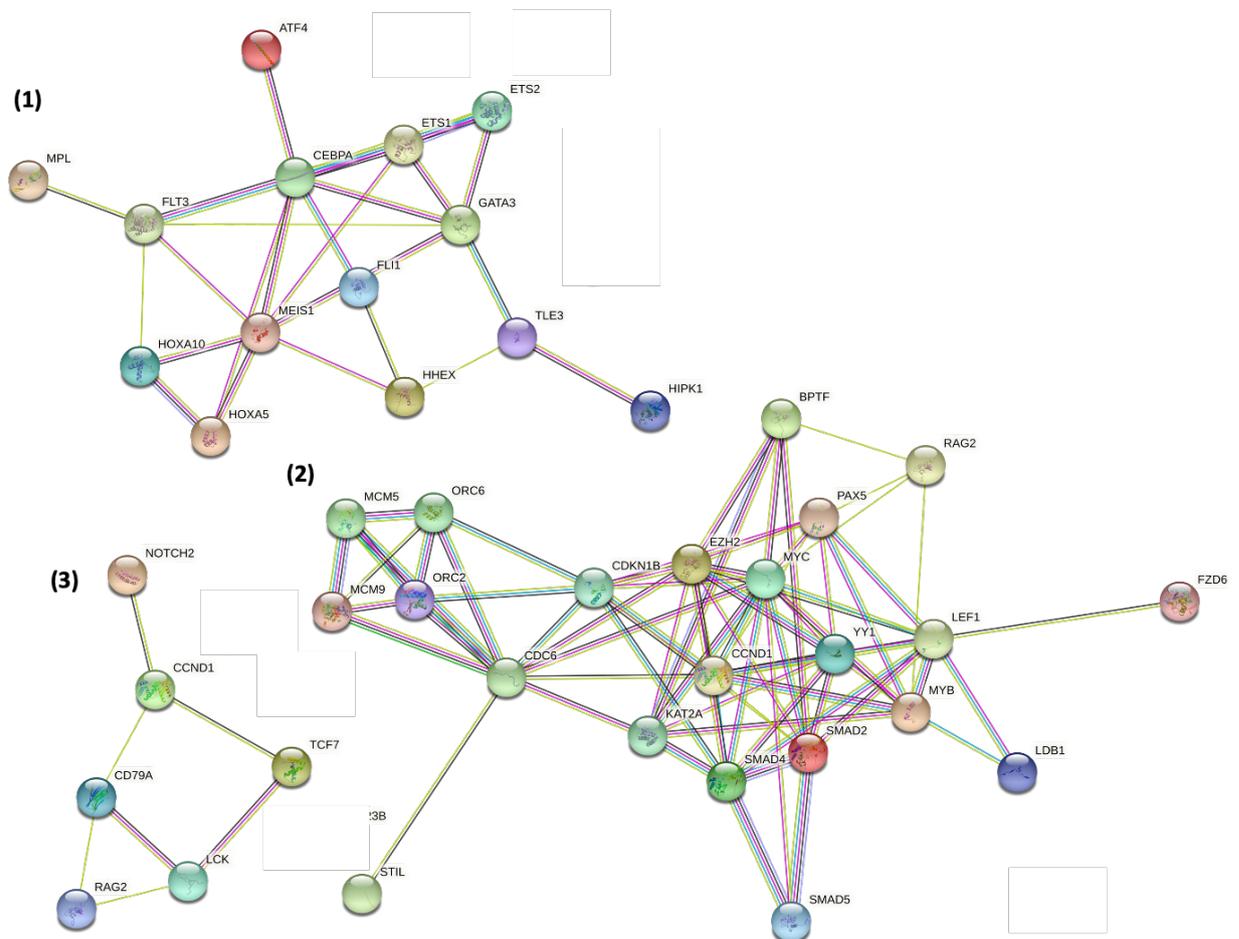


Figure 3.3.3: Functional enrichment analysis of genes differentially regulated in primary Lin-CD34+ CML cells, dataset GSE11675. Genes of interest were selected through a literature search and the most relevant compiled in list form. The changes in the expression level for these genes were analysed in Lin-CD34[±] primary CML cells from dataset GSE11675, the results of which visually displayed in a heatmap. From the heatmap produced, genes observed to be downregulated were selected and a STRING gene-gene interaction network produced to display the functional relevance of gene expression changes. (1) Genes downregulated in Lin-CD34⁺ CML cells. (2) Genes upregulated in Lin-CD34⁺ CML cells. (3) Genes downregulated in Lin-CD34⁻ CML cells.

CEBPA was downregulated in CML Lin-CD34⁺ cells, the cell population most likely to contain CML LSCs. *CEBPA* normally functions as a myeloid transcription factor though the induction of the granulocytic differentiation of myeloid progenitors by activation of myeloid-specific

genes. Downregulation of *CEBPA* in CML leads to the progression toward and development of blast crisis (Agatheeswaran and Chakraborty, 2016). This gene would be informative to follow up in Ph+ALL as it may represent the shift to acute leukaemia phenotype, and due to its involvement in myeloid disease, would be informative to investigate in myeloid-like Ph+ALL cases.

Similar to *CEBPA*, *GATA3* is involved with lymphoid and myeloid lineage commitment decisions (respectively) and is not expressed in CML CD34+ cells suggesting that such cells have a greater propensity for self-renewal than development and differentiation as in healthy HSC cells (Kronenwett et al., 2005). Overexpression of *ERG* is associated with poor outcomes in AML and T-ALL and is associated with drug resistance mechanisms. Hence, downregulation of *ERG* may reflect the chronic phenotype of CML and its upregulation associated with acute leukaemia, as is evidenced by the mesenchymal-like phenotype induced by *ERG* expression (Mochmann et al., 2013). *SPON1* is highly expressed in AML with poor overall survival rates, similar to *ETS2* stated above, may be linked to acute leukaemias, hence its downregulation in CML could be related to the early chronic disease stage at which samples were taken (Wang et al., 2020). The normal function of *FLT3* is in the regulation of haematopoiesis, with mutations in *FLT3* resulting in AML and ALL development. *FLT3* is not observed in CML (Lin et al., 2006). Lower expression of *MPL*, thrombopoietin (THPO) receptor, is associated with reduced leukaemogenic capacity and an increased sensitivity to BCR::ABL1 targeting TKIs (Zhang et al., 2014). Hence, the observation of *MPL* and *THPO* downregulation was in keeping with the reduced leukaemogenic activity of CML cells in chronic phase.

The overexpression of *HOXA10* induces progress of CML to blast crisis therefore, downregulation of *HOXA10* in these cells indicates the earlier CML stage of these sample cells and the chronic nature of cell activity (Negi et al., 2017). The downregulation of *HOXA5* suggests increased sensitivity to TKIs, indicating reliance on BCR::ABL1 kinase activity (Hikmah Elias et al., 2018). Low expression of *MEIS1* may reduce CML cell proliferation as would be observed in non-acute phases of CML (Zhao et al., 2020). *FLI1* expression is associated with poor prognosis in AML therefore supporting low proliferation rate and high survivability of chronic phase CML (Kornblau et al., 2011). The normal role of *HHEX* in haematopoietic differentiation and repression of genes in the *VEGF* signalling pathway is to inhibit myeloid

cell survival therefore, the downregulation of *HHEX* in a myeloid-mediated disease would support myeloblast proliferation and survival (Noy et al., 2012). *TLE3* gene expression is associated with myeloid disease progression to blast crisis and is associated with poor response to TKIs, thus downregulation may also be indicative of early chronic stages of CML where cells may be more sensitive to TKI treatment (Horne et al., 2017).

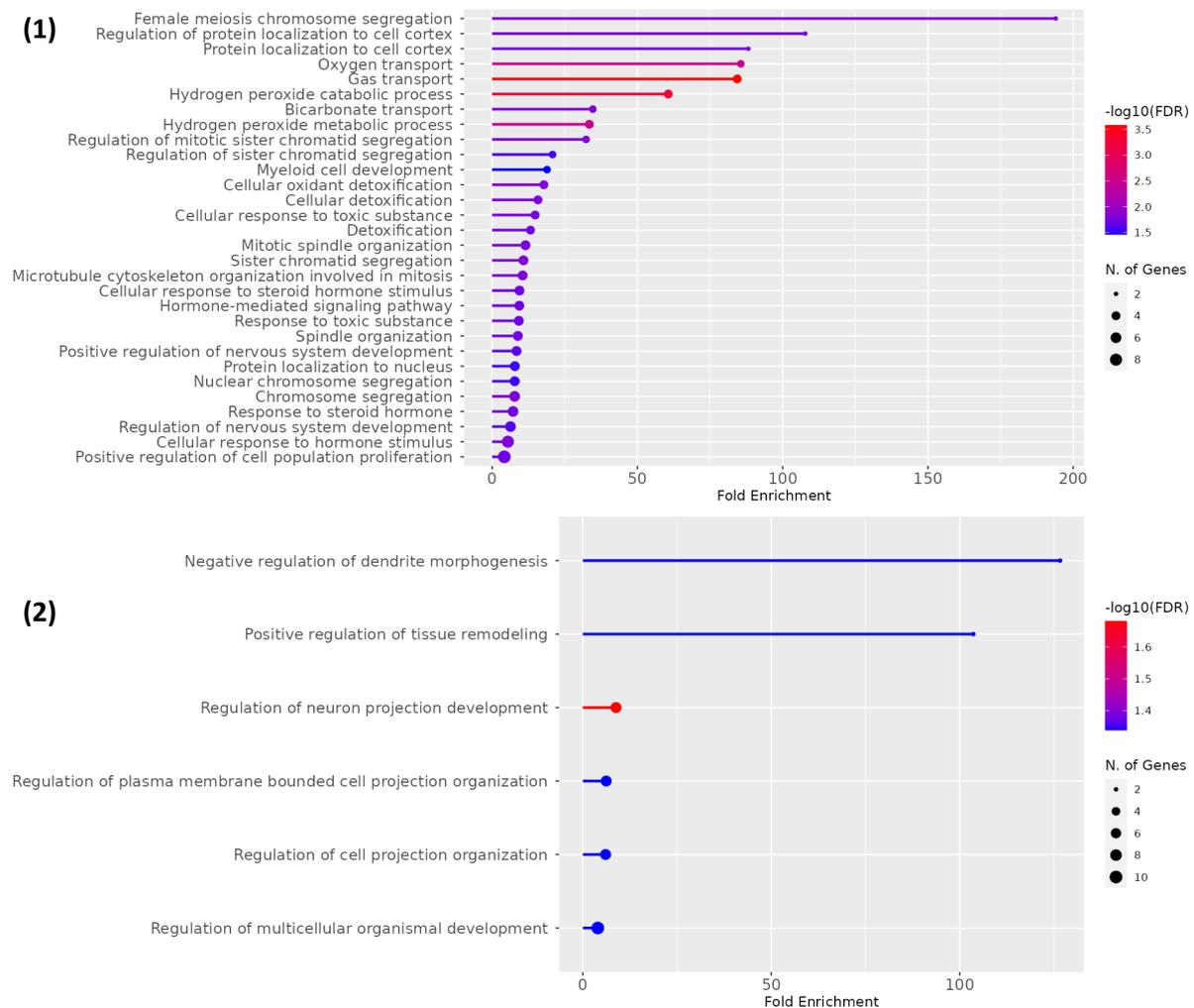


Figure 3.3.4: Gene ontology enrichment analysis for (1) top 50 upregulated and (2) bottom 50 downregulated DEGs in CML vs healthy control cells (Lin+CD34+, Lin-CD34+, Lin-CD34-), dataset GSE11675. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. $LogFC = \log$ fold change. ShinyGO v0.741.

The lymphoid related function of *HIPK1* gene has been discussed earlier and the downregulation of *HIPK1* supports the myeloid phenotype of CML cells samples for this dataset. Many of the genes observed to be downregulated in Lin-CD34+ CML cells are

involved in acute leukaemia and cell proliferation, supporting the reduced proliferation rate associated with chronic phase CML.

A number of genes were found to be upregulated in Ph+ (CML and Ph+ALL) cell lines in dataset GDS4175 and Lin-CD34+ CML LSCs in dataset GSE11675. Such genes include *EZH2*, *MCM9* and *PAX5*, their shared upregulation status in both disease types supporting involvement of such genes in BCR::ABL1-induced activity (Fig 3.3.1 and 3.3.2).

CCND1 and *MYB* were upregulated in Ph+ALL cell lines (Fig 3.3.1 and 3.3.2) and CML Lin-CD34+ cells and therefore do not suggest a disease type-specific role. In addition, primitive Lin- CD34- CML stem cells share upregulation of *FLI1*, *HIPK1* and *MYB* with Ph+ALL cell lines (Fig 3.3.2 and 3.3.3), genes associated with early, primitive stages of haematopoiesis. Putative Lin-CD34- CML stem cells display gene expression patterns distinct from traditionally defined Lin-CD34+ LSCs and display leukaemia initiating properties (Benton et al., 2015). This was demonstrated in GSE1675 by shared expression patterns of stem cell-associated early development genes such as *MCM7/9*, *MYB*, *PAX5*, *SMAD2*, *SMAD4* and *STIL*, and later effector functions such as *ATF4* and *CNR2*.

3.4 Results I: HSCs in CML-CP displayed some hallmarks of normal stem cell development, shared CML-specific transcriptional patterns with Lin- CML samples from previous datasets and displayed expectable intra-sample variation

Determining the cell of origin where the *BCR::ABL1* fusion event occurs during haematopoiesis is required for predicting the cell of origin. In addition, the cell types expressed during leukaemic-haematopoiesis often correlate with cell of origin or disease state. For example, and as would be expected in CML, CMPs and MEPs are expressed at a higher proportion than in healthy bone marrow. GSE11889 utilises gene expression profiles of CML HSCs vs normal HSCs controls. This dataset also contained expression data for other haematopoietic cell populations (CMP, GMP and MEP) for each CML patient and healthy control examined. HSC samples were selected in order to investigate differences in gene expression in healthy control HSCs and CML Ph+ HSCs, providing an opportunity to identify any CML-specific DEGs which may be beneficial in isolating CML-like samples from a wider Ph+ALL cohort in future experiments. This would then allow us to apply results from more primitive cell types to Ph+ALL cells in order to establish if there is a CML-like HSC gene expression phenotype within CML-like Ph+ALL.

The CML-like Ph+ALL subtype has been clinically observed to have a myeloid-like phenotype, with higher numbers of myeloid progenitors in this samples (observed clinically). However, it is unknown the extent to which myeloid-driven pathology affects 'classical' Ph+ALL or if myeloid involvement is common to all Ph+ALL to varying degrees (Hovorkova et al., 2019). As discussed previously, chronic phase CML has a different transcriptomic profile to blast phase disease; however Chronic phase CML HSCs were used to understand of *BCR::ABL1*-driven HSCs as the HSC within this population is better understood. The leukaemic stem cell in blast phase is not well established, particularly between myeloid and lymphoid blast phase (Copland., 2022). This may be due, in part, to the relative rarity of CML-BP in the post-TKI era however, an understanding of BP-specific transcriptional differences may benefit other acute Ph+ driven leukaemia such as Ph+ALL.

To understand the transcriptomic differences between Ph+ HSCs and normal HSCs, we first generated a PCA plot to visualise all samples in a low dimensional manner. This demonstrated a broadly heterogeneous gene expression phenotype with observable sample-specific gene expression profiles (Fig 3.4.1).

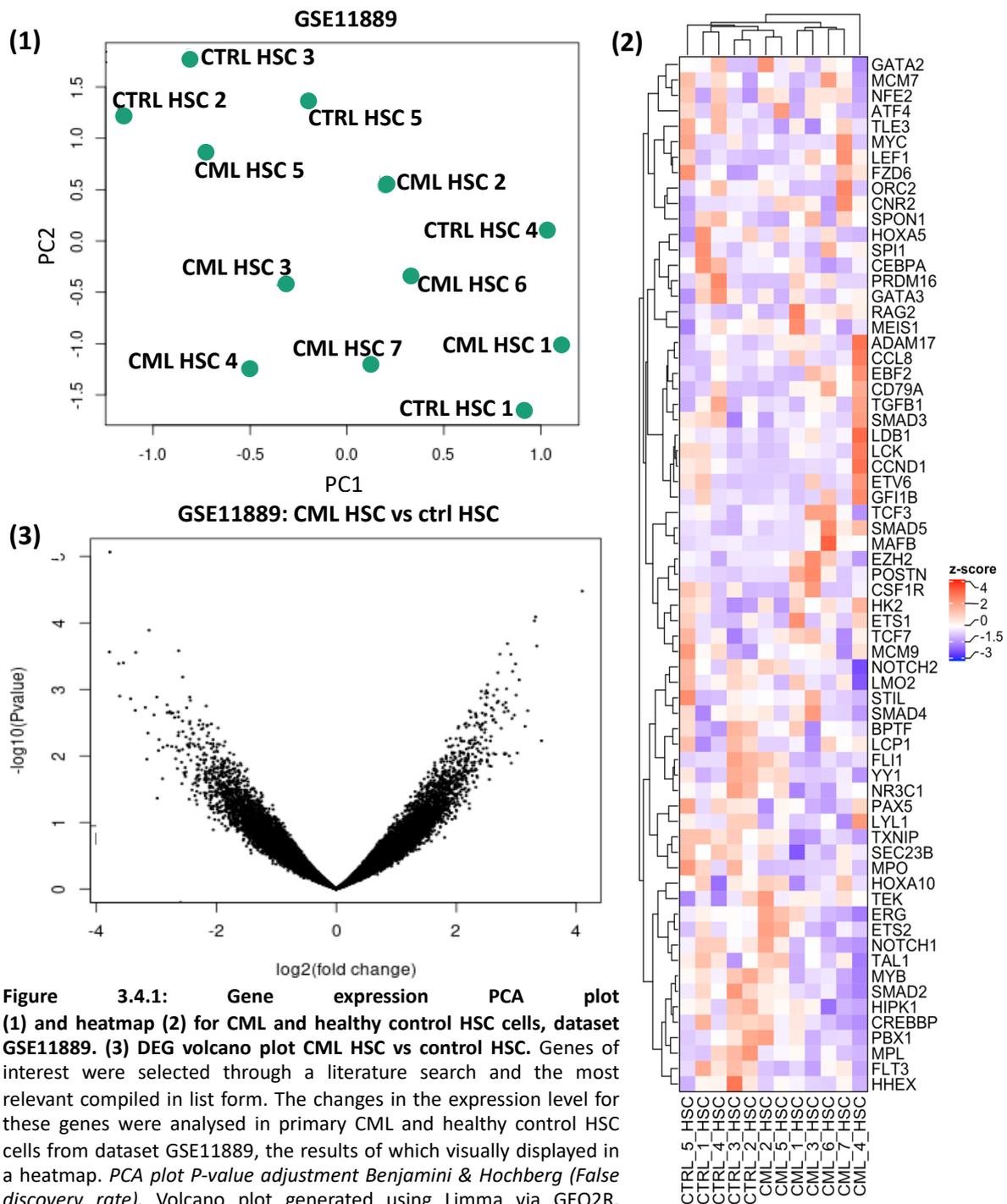


Figure 3.4.1: Gene expression PCA plot (1) and heatmap (2) for CML and healthy control HSC cells, dataset GSE11889. (3) DEG volcano plot CML HSC vs control HSC. Genes of interest were selected through a literature search and the most relevant compiled in list form. The changes in the expression level for these genes were analysed in primary CML and healthy control HSC cells from dataset GSE11889, the results of which visually displayed in a heatmap. *PCA plot P-value adjustment Benjamini & Hochberg (False discovery rate)*. Volcano plot generated using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment.

Gene expression was profiled by array and expression data was loaded into R from GEO (gene expression omnibus) and analysed using the GEO2R platform as outlined in section 3.2. We next generated heatmaps using the previously described sig genes of interest (Fig 3.4.1.2). The method is as previously described. CP is characterised by regulated proliferation of myeloid cells, the appearance of normal stem cell maturation and, comparative to acute blast crisis, the normal response of HSCs to growth factor regulators G-CSF and CM-CSF (Houshmand et al., 2019). Hence, many precursor homeostasis and cell cycle genes remain similarly regulated between CML and control samples such as *GFI1B*, *CCND1*, *LCK* and *LDB1*. A number of genes shared gene expression patterns with Lin-CD34+ CML HSCs in GSE11675 (*RAG2*, *TEK*, *CEBPA*, *ERG*, *ETS1*, *FLI1*, *FLT3*, *HHEX*, *HOXA10* and *MPL*) displaying a reliable CML phenotype to which Ph+ALL samples may be compared in order to elucidate a CML-like gene expression signature (Fig 3.4.1 and 3.4.2).

Gene	Gene function	logFC	P.Value
GAS2	growth arrest specific 2	4.1	0.0003334
HLA-DRB4	major histocompatibility complex, class II, DR beta 4	3.42	0.00592338
TAF6L	TATA-box binding protein associated factor 6 like	3.34	0.00022242
CLCA2	chloride channel accessory 2	3.32	0.00008125
PLA2G6	phospholipase A2 group VI	3.3	0.00009299
PDE4DIP	phosphodiesterase 4D interacting protein	3.19	0.00209
PPIAL4C	peptidylprolyl isomerase A like 4C	3.15	0.00358853
ISG20L2	interferon stimulated exonuclease gene 20 like 2	3.05	0.00071822
EGFL6	EGF like domain multiple 6	3.02	0.00906447
CACNA1D	calcium voltage-gated channel subunit alpha1 D	3	0.00121982
HHIPL2	HHIP like 2	2.99	0.00041363
PIEZO2	piezo type mechanosensitive ion channel component 2	2.97	0.00243667
DCX	doublecortin	2.94	0.00053581
KLHL22	kelch like family member 22	2.93	0.00274324
VCAN	versican	2.91	0.0002965
TNFRSF11B	TNF receptor superfamily member 11b	2.88	0.00999111
SENP7	SUMO1/sentrin specific peptidase 7	2.86	0.00211604
ANP32D	acidic nuclear phosphoprotein 32 family member D	2.85	0.00020523
LOC441666	zinc finger protein 91 pseudogene	2.84	0.00097856
RUNX1	runt related transcription factor 1	2.83	0.00188326
IDS	iduronate 2-sulfatase	2.83	0.00923649
CYP3A43	cytochrome P450 family 3 subfamily A member 43	2.82	0.00158036
USP49	ubiquitin specific peptidase 49	2.8	0.0021864
KLK8	kallikrein related peptidase 8	2.8	0.00920545
SYCP1	synaptonemal complex protein 1	2.8	0.00941623
GRHL2	grainyhead like transcription factor 2	2.78	0.00170405
TSNAX-DISC1	TSNAX-DISC1 readthrough (NMD candidate)	2.75	0.00384378
SMAD5	SMAD family member 5	2.73	0.00029465
EZR	ezrin	2.7	0.00150552
SFRP4	secreted frizzled related protein 4	2.69	0.00056043
GRK1	G protein-coupled receptor kinase 1	2.68	0.00100681
LOC441601	septin 7 pseudogene	2.68	0.0015853
NAV2	neuron navigator 2	2.66	0.00314396
DEDD	death effector domain containing	2.66	0.00942232
CYP4F11	cytochrome P450 family 4 subfamily F member 11	2.62	0.0012297
STAG3L4	stromal antigen 3-like 4 (pseudogene)	2.62	0.00202637
GTF2H4	general transcription factor IIF subunit 4	2.62	0.00281203
TAF6L	TATA-box binding protein associated factor 6 like	2.62	0.00919735
PSMD3	proteasome 26S subunit, non-ATPase 3	2.61	0.00291063
VCX2	variable charge, X-linked 2	2.59	0.0018844
FCN2	ficolin 2	2.59	0.00250369
PCDH7	protocadherin 7	2.58	0.00509129
RPL5	ribosomal protein L5	2.57	0.00596476
IL22RA1	interleukin 22 receptor subunit alpha 1	2.54	0.00973701
ITPR2	inositol 1,4,5-trisphosphate receptor type 2	2.52	0.00084664
TBL1X	transducin (beta)-like 1X-linked	2.52	0.00113492
MAST4	microtubule associated serine/threonine kinase family member 4	2.52	0.00921354
HRASLS2	HRAS like suppressor 2	2.51	0.0114665
MARCKS	myristoylated alanine rich protein kinase C substrate	2.5	0.00105227
BARX1	BARX homeobox 1	2.5	0.00170548

Table 3.4.1: Top 50 upregulated DEGs in CML HSCs vs healthy control HSCs, dataset GSE11889. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. *LogFC*= log fold change.

Gene	Gene function	logFC	P.Value
LTBP3	latent transforming growth factor beta binding protein 3	-2.39	0.00467912
CXCR4	C-X-C motif chemokine receptor 4	-2.41	0.00388435
SP100	SP100 nuclear antigen	-2.42	0.01163197
KBTBD11	kelch repeat and BTB domain containing 11	-2.44	0.00130438
NCOA1	nuclear receptor coactivator 1	-2.44	0.003036
PROM1	prominin 1	-2.44	0.00557933
SETD2	SET domain containing 2	-2.44	0.00643814
ULK2	unc-51 like autophagy activating kinase 2	-2.45	0.00328545
ILF3	interleukin enhancer binding factor 3	-2.46	0.00707005
DDIT4	DNA damage inducible transcript 4	-2.47	0.00703999
RRP12	ribosomal RNA processing 12 homolog	-2.47	0.00957743
ARL17B	ADP ribosylation factor like GTPase 17B	-2.48	0.00672298
SNTB1	syntrophin beta 1	-2.48	0.00687402
JOSD1	Josephin domain containing 1	-2.5	0.00188113
HCG4	HLA complex group 4 (non-protein coding)	-2.51	0.00518326
GADD45B	growth arrest and DNA damage inducible beta	-2.56	0.00065078
RHOBTB3	Rho related BTB domain containing 3	-2.56	0.00298388
KLF10	Kruppel like factor 10	-2.56	0.00429375
MYOM2	myomesin 2	-2.56	0.00798478
H1FO	H1 histone family member 0	-2.62	0.0083898
ADAM7	ADAM metalloproteinase domain 7	-2.63	0.00026263
PGK1	phosphoglycerate kinase 1	-2.63	0.00259048
MEPCE	methylphosphate capping enzyme	-2.63	0.00443598
PTGER4	prostaglandin E receptor 4	-2.64	0.00496366
POU2AF1	POU class 2 associating factor 1	-2.64	0.00830878
MALT1	MALT1 paracaspase	-2.67	0.00734646
KDM5D	lysine demethylase 5D	-2.68	0.01104284
SLC38A10	solute carrier family 38 member 10	-2.7	0.00264055
TCF4	transcription factor 4	-2.74	0.00719284
DYNC1L12	dynein cytoplasmic 1 light intermediate chain 2	-2.76	0.00229278
CXCR4	C-X-C motif chemokine receptor 4	-2.77	0.00215692
CYP3A5	cytochrome P450 family 3 subfamily A member 5	-2.82	0.00170734
LOC730101	uncharacterized LOC730101	-2.82	0.00223232
MPO	myeloperoxidase	-2.82	0.00698265
CDH2	cadherin 2	-2.89	0.00740414
SPON1	spondin 1	-2.95	0.0056781
CLC	Charcot-Leyden crystal galectin	-2.96	0.00831187
NR4A2	nuclear receptor subfamily 4 group A member 2	-2.99	0.00130517
DNTT	DNA nucleotidylexotransferase	-3.04	0.00241742
FAM105A	family with sequence similarity 105 member A	-3.12	0.0001289
ASPH	aspartate beta-hydroxylase	-3.14	0.00452852
PLAG1	PLAG1 zinc finger	-3.16	0.01119889
OXR1	oxidation resistance 1	-3.19	0.00186732
MPO	myeloperoxidase	-3.34	0.00027805
EMP1	epithelial membrane protein 1	-3.35	0.00206245
AREG	amphiregulin	-3.43	0.00138362
PDE4B	phosphodiesterase 4B	-3.55	0.00039963
ZMYND11	zinc finger MYND-type containing 11	-3.61	0.00126024
BPGM	bisphosphoglycerate mutase	-3.63	0.00040744
TCFL5	transcription factor like 5	-3.77	0.00000861
NR4A3	nuclear receptor subfamily 4 group A member 3	-3.78	0.00027404

Table 3.4.2: Bottom 50 downregulated DEGs in CML HSCs vs healthy control HSCs, dataset GSE11889. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. *LogFC= log fold change.*

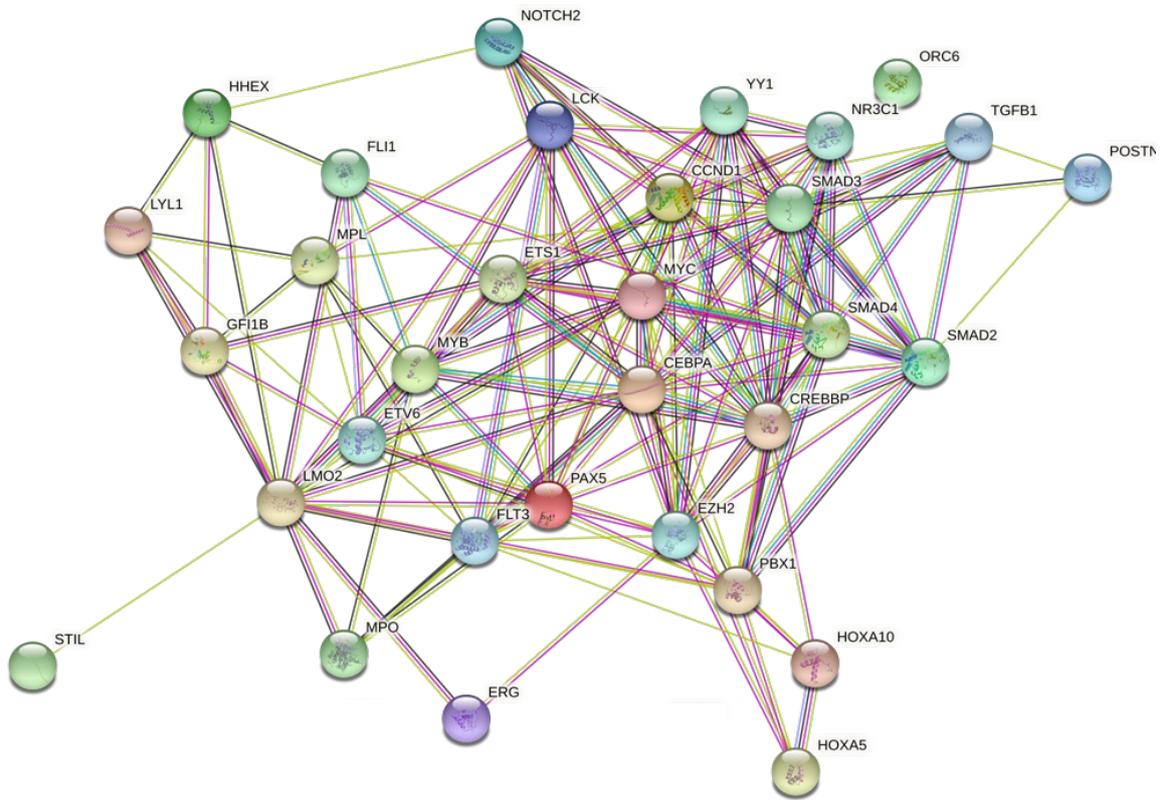


Figure 3.4.2: Functional enrichment analysis of genes downregulated in CML HSCs, dataset GSE11889. Genes of interest were selected through a literature search and the most relevant compiled in list form. The changes in the expression level for these genes were analysed in primary CML HSC cells from dataset GSE11889, the results of which visually displayed in a heatmap. From the heatmap produced, genes observed to be differentially regulated were selected and a STRING gene-gene interaction network produced to display the functional relevance of gene expression changes.



Figure 3.4.3: Gene ontology enrichment analysis for (1) top 50 upregulated and (2) bottom 50 downregulated DEGs in CML HSCs vs healthy control HSCs, dataset GSE11889. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. $\text{LogFC} = \log \text{fold change}$. ShinyGO v0.741.

3.5 Results I: Paediatric Ph+ALL patients designated as being ‘good risk’ based on blast cell count were transcriptionally similar to CML samples from previous datasets

We next questioned which genes are *BCR::ABL1*-driven and, therefore, shared in Ph+ALL and CML pathology. This would allow us to then demonstrate a ‘*BCR::ABL1*-leukaemic transcription profile’. Dataset GSE39335 was utilised to identify gene expression profiles in Ph+ALL in response to glucocorticoid (GC) treatment. As has been previously discussed, the scope of this project is not focused on the treatment of Ph+ALL and CML, therefore, the inclusion of untreated (0h) samples in this dataset has been used to identify gene expression patterns in Ph+ALL. GC treatments at timepoints 6h and 24h were omitted.

The data set used 10 paediatric patients diagnosed with Ph+ALL. These 10 samples represent all patients examined by Chen et al (2012) by microarray however, only 6 samples (Patients 184, 193, 277, 341, 205 and 241) were available as publicly available data. Patients were treated uniformly with Dexamethasone and one dose each of anthracycline, vincristine and L-Asparaginase, according to the EsphALL protocol. Samples were then analysed at day 17 and compared to the untreated (day 0) samples. As previously discussed, evaluation of treatment protocols is outwith the scope of this project, hence day 0 untreated samples were selected. Patients were designated as ‘good risk’ or ‘poor risk’ based on clinical observation of blast cell proportion in the bone marrow (<25% blasts after 8 days treatment = ‘good risk’, >25% blasts after 8 days treatment = ‘poor risk’). Poor risk samples were patient samples 205 and 241; good risk were patient samples 184, 193, 277, and 341. Retrospective designation of patients’ risk classification allows for Day 0 diagnostic samples to be investigated with the additional information of clinical outcomes, allowing for potential risk indicators to be observed at diagnosis.

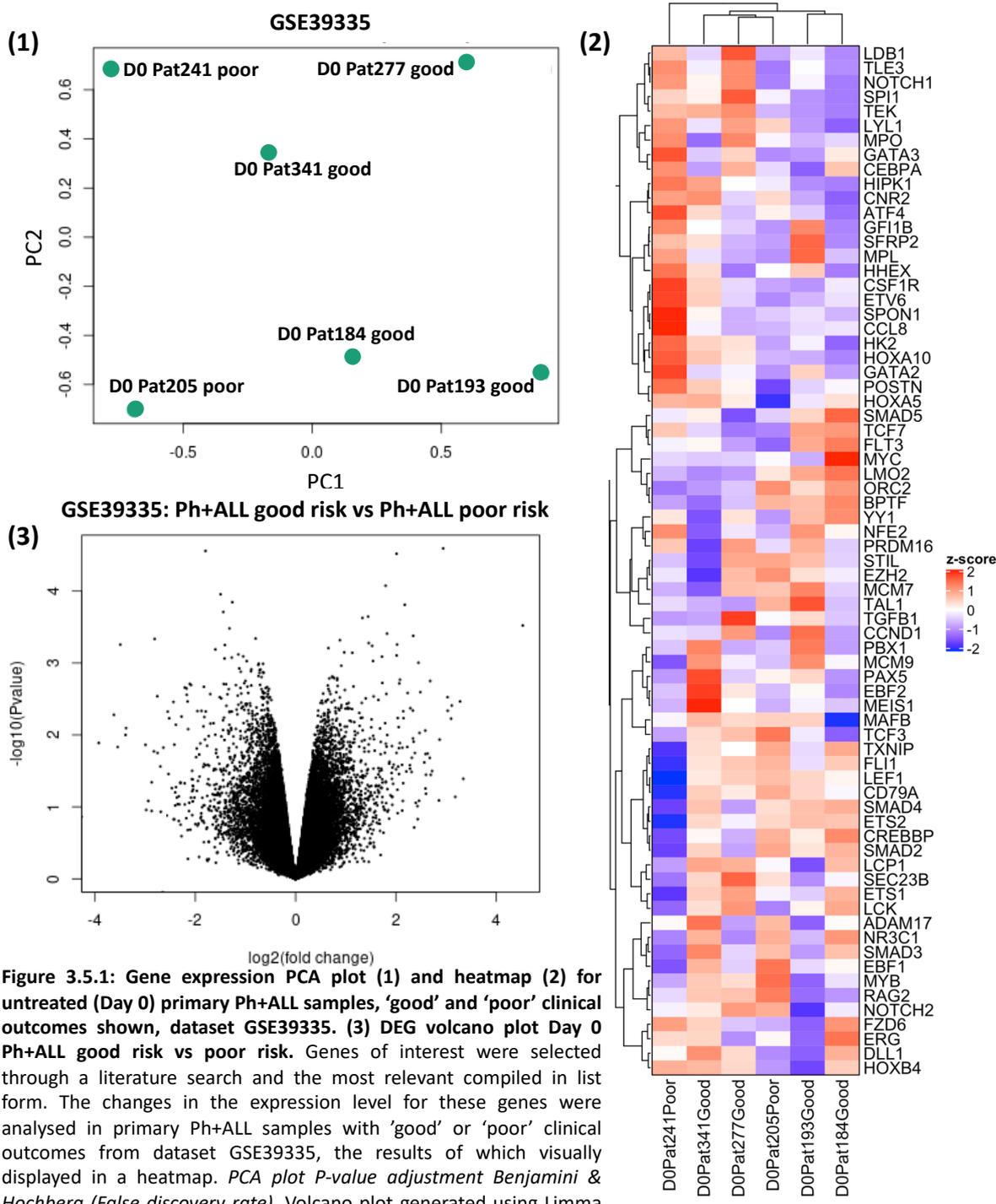


Figure 3.5.1: Gene expression PCA plot (1) and heatmap (2) for untreated (Day 0) primary Ph+ALL samples, 'good' and 'poor' clinical outcomes shown, dataset GSE39335. (3) DEG volcano plot Day 0 Ph+ALL good risk vs poor risk. Genes of interest were selected through a literature search and the most relevant compiled in list form. The changes in the expression level for these genes were analysed in primary Ph+ALL samples with 'good' or 'poor' clinical outcomes from dataset GSE39335, the results of which visually displayed in a heatmap. *PCA plot P-value adjustment Benjamini & Hochberg (False discovery rate)*. Volcano plot generated using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment.

Gene expression was profiled by array and expression data was loaded into R from GEO (gene expression omnibus) and analysed using the GEO2R platform as outlined in section 3.2. We next assessed DEG between poor and good responders (Fig 3.6.1.3). The variation in gene expression profiles between ‘poor’ and ‘good’ risk Ph+ALL is important when investigating other Ph+ALL datasets which present patient data without categorising based on disease outcome, as a number of genes display differential regulation between ‘good’ and ‘poor’ samples such as *ATF4*, *CNR2*, *LYL1* and *SMAD5*. The variation between patient subclasses means that attempting to assign a broad Ph+ALL gene expression pattern would result in subclass intricacies being neglected. This indicates that subclass gene expression phenotypes would be more adequate for capturing functional variation within Ph+ALL. With the predicted proportion of CML-like Ph+ALL patients being approximately 30% of the overall Ph+ALL cohort, it may be more pertinent to acknowledge the phenotypically different subtypes and create a number of subtype classifications (Hovorkova et al., 2017). Using selected genes, as described above, a heatmap was generated. This demonstrated that we were unable to cluster poor responders from these selected genes. This suggests that there is a heterogeneity between poor responders. We next compared the outlying poor responder to the good and poor responder cluster (Fig 3.5.1.1). This demonstrated that within our selected genes, key gene changes included upregulation in *ERG1*, *GZMA*, *SPON1*, *CCL8* and *GNLY*, and downregulated in *LEF1*, *CD79A*, *POCIB*, *SLC44A1* and *SLCC2A7* (Fig 3.5.1.2, Table 3.5.1 and 3.5.2).

Gene	Gene function	logFC	P.Value
EGR1	early growth response 1	4.532	0.003009
GZMA	granzyme A	3.282	0.0034185
GNLY	granulysin	3.116	0.0039002
COCH	cochlin	3.102	0.0059368
EGR1	early growth response 1	3.034	0.0046072
GNLY	granulysin	3.031	0.0029896
ARL4C	ADP ribosylation factor like GTPase 4C	2.945	0.0000255
ZBTB38	zinc finger and BTB domain containing 38	2.685	0.001777
BCL11B	B-cell CLL/lymphoma 11B	2.489	0.0025402
LCK	LCK proto-oncogene, Src family tyrosine kinase	2.452	0.0009901
LCK	LCK proto-oncogene, Src family tyrosine kinase	2.347	0.0004196
CD3D	CD3d molecule	2.34	0.0045862
TRDV3	T cell receptor delta variable 3	2.29	0.0042494
RASGRP1	RAS guanyl releasing protein 1	2.236	0.0026121
LY9	lymphocyte antigen 9	2.18	0.0001552
TRAT1	T cell receptor associated transmembrane adaptor 1	2.152	0.0058321
YME1L1	YME1 like 1 ATPase	2.147	0.004687
BCL11B	B-cell CLL/lymphoma 11B	2.084	0.0017032
ARL4C	ADP ribosylation factor like GTPase 4C	2.068	0.0019209
ITK	IL2 inducible T-cell kinase	2.038	0.0058232
GBP5	guanylate binding protein 5	2.025	0.0005527
FCRL3	Fc receptor like 3	2.017	0.0007855
SGPP1	sphingosine-1-phosphate phosphatase 1	2.015	0.0000302
NEXN	nexilin F-actin binding protein	1.824	0.0003902
TBCD	tubulin folding cofactor D	1.811	0.0026889
SP140	SP140 nuclear body protein	1.797	0.0000841
IL2RB	interleukin 2 receptor subunit beta	1.73	0.00189
KLRK1	killer cell lectin like receptor K1	1.723	0.0008149
ZBTB24	zinc finger and BTB domain containing 24	1.653	0.0032611
P2RY10	purinergic receptor P2Y10	1.557	0.0029127
SGPP1	sphingosine-1-phosphate phosphatase 1	1.545	0.0009371
BTG2	BTG anti-proliferation factor 2	1.528	0.0005768
SAMHD1	SAM and HD domain containing deoxynucleoside triphosphate triphosphohydrolase 1	1.516	0.0060858
P2RY10	purinergic receptor P2Y10	1.458	0.0005277
TBCD	tubulin folding cofactor D	1.447	0.0002265
IL21R	interleukin 21 receptor	1.4	0.0030886
NEXN	nexilin F-actin binding protein	1.333	0.0002361
RLIM	ring finger protein, LIM domain interacting	1.332	0.0057436
DDX28	DEAD-box helicase 28	1.321	0.0035872
SDF2L1	stromal cell derived factor 2 like 1	1.301	0.0057865
ZBTB24	zinc finger and BTB domain containing 24	1.271	0.0049388
TRAPPC2	trafficking protein particle complex 2	1.171	0.0066324
LY9	lymphocyte antigen 9	1.132	0.0007205
PCTP	phosphatidylcholine transfer protein	1.127	0.004352
CDKAL1	CDK5 regulatory subunit associated protein 1 like 1	1.107	0.0064818
ZFP36L2	ZFP36 ring finger protein like 2	1.08	0.0063383
BIRC3	baculoviral IAP repeat containing 3	1.07	0.0042676
EIF3H	eukaryotic translation initiation factor 3 subunit H	1.058	0.0035908
MAF	MAF bZIP transcription factor	1.045	0.0043219
IKZF3	IKAROS family zinc finger 3	1.036	0.0041986
SEL1L	SEL1L ERAD E3 ligase adaptor subunit	1.021	0.0036866

Table 3.5.1: Top 50 upregulated DEGs in Day 0 Ph+ALL good risk vs poor risk, dataset GSE39335. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. *LogFC*= log fold change.

Gene	Gene function	logFC	P.Value
SLC22A7	solute carrier family 22 member 7	-0.861	0.0046643
SLC44A1	solute carrier family 44 member 1	-0.862	0.0050637
POC1B	POC1 centriolar protein B	-0.884	0.0020588
GPATCH2	G-patch domain containing 2	-0.903	0.0039367
CLCN7	chloride voltage-gated channel 7	-0.91	0.0036265
MAN2A2	mannosidase alpha class 2A member 2	-0.934	0.0050096
CALN1	calneuron 1	-0.961	0.0021206
TSTD1	thiosulfate sulfurtransferase like domain containing 1	-0.968	0.0036923
SMA4	glucuronidase beta pseudogene	-0.977	0.0009403
APLP2	amyloid beta precursor like protein 2	-0.988	0.003904
MGC12916	uncharacterized protein MGC12916	-0.998	0.0036823
HRES1	HTLV-1 related endogenous sequence	-1.001	0.0017741
YOD1	YOD1 deubiquitinase	-1.026	0.0019086
FNIP2	folliculin interacting protein 2	-1.044	0.000808
MIIIP	migration and invasion inhibitory protein	-1.052	0.0040356
VAV3	vav guanine nucleotide exchange factor 3	-1.078	0.0042424
CFAP73	cilia and flagella associated protein 73	-1.094	0.0013208
ALDH8A1	aldehyde dehydrogenase 8 family member A1	-1.099	0.0022213
FCHO2	FCH domain only 2	-1.142	0.0037998
MCTP2	multiple C2 and transmembrane domain containing 2	-1.143	0.00076
RAB2A	RAB2A, member RAS oncogene family	-1.154	0.0028389
MCTP2	multiple C2 and transmembrane domain containing 2	-1.191	0.0020303
ZNF117	zinc finger protein 117	-1.263	0.0001427
SLC44A3	solute carrier family 44 member 3	-1.32	0.0003307
ZNF117	zinc finger protein 117	-1.332	0.0023041
GUSBP3	glucuronidase, beta pseudogene 3	-1.359	0.0027215
TLE4	transducin like enhancer of split 4	-1.387	0.0019718
SBF2	SET binding factor 2	-1.418	0.0008641
OSBPL5	oxysterol binding protein like 5	-1.428	0.0057435
TLE4	transducin like enhancer of split 4	-1.433	0.0019501
SBF2	SET binding factor 2	-1.441	0.000195
IKZF1	IKAROS family zinc finger 1	-1.447	0.0062447
CPEB2	cytoplasmic polyadenylation element binding protein 2	-1.483	0.003093
TAPT1	transmembrane anterior posterior transformation 1	-1.513	0.0029894
TAPT1	transmembrane anterior posterior transformation 1	-1.599	0.0006473
NAIP	NLR family apoptosis inhibitory protein	-1.702	0.0027007
NPCDR1	nasopharyngeal carcinoma, down-regulated 1	-1.793	0.0047311
ZNF117	zinc finger protein 117	-1.798	0.0000276
PDE4DIP	phosphodiesterase 4D interacting protein	-1.829	0.0021399
YBX3	Y-box binding protein 3	-1.878	0.0046295
MMRN1	multimerin 1	-1.958	0.0046575
MYEF2	myelin expression factor 2	-2.086	0.0020277
MYEF2	myelin expression factor 2	-2.111	0.0019467
MCTP2	multiple C2 and transmembrane domain containing 2	-2.174	0.0022613
CDC42EP3	CDC42 effector protein 3	-2.211	0.0021932
BAALC	brain and acute leukemia, cytoplasmic	-2.436	0.0034802
BAALC	brain and acute leukemia, cytoplasmic	-2.552	0.0047406
ADGRF1	adhesion G protein-coupled receptor F1	-2.759	0.0028982
ADGRF1	adhesion G protein-coupled receptor F1	-3.497	0.0005542
PTPRD	protein tyrosine phosphatase, receptor type D	-3.623	0.0052492

Table 3.5.2: Bottom 50 downregulated DEGs in Day 0 Ph+ALL good risk vs poor risk, dataset GSE39335. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. *LogFC= log fold change*.

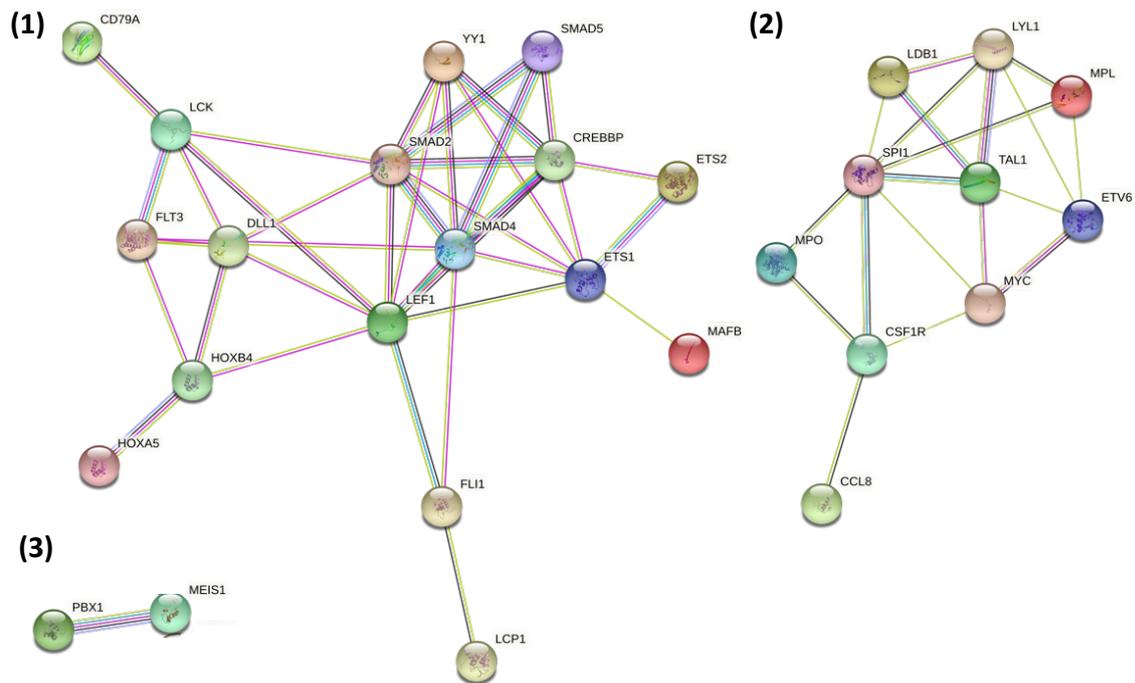


Figure 3.5.2: Functional enrichment analysis of genes differentially regulated in untreated primary Ph+ALL samples with good or poor clinical outcomes, dataset GSE39335. Genes of interest were selected through a literature search and the most relevant compiled in list form. The changes in the expression level for these genes were analysed in untreated primary Ph+ALL cells from patients with good or poor clinical outcomes from dataset GSE39335, the results of which visually displayed in a heatmap. From the heatmap produced, genes observed to be differentially were selected and a STRING gene-gene interaction network produced to display the functional relevance of gene expression changes. (1) Genes upregulated in untreated primary Ph+ALL samples with good clinical outcomes. (2) Genes downregulated in untreated primary Ph+ALL samples with good clinical outcomes. (3) Genes upregulated in untreated primary Ph+ALL samples with poor clinical outcomes.

Comparison of genes expressed in ‘poor’ Ph+ALL to CML samples from previous sections displayed no similarities (Fig 3.2- 3.5). Notably, GSE39335 only contained two samples designated as ‘poor’ outcomes and hence, these observations should only be used as indicators to display that negative Ph+ALL prognosis and CML are genetically distinct.

Conversely, samples designated by blast cell clearance as having a good clinical outcome displayed similar gene expression patterns to CML cells from datasets GSE11675 and GSE11889 (Fig 3.4 and 3.5). Clinically ‘good’ Ph+ALL samples (n=4) and CML samples displayed an upregulation in *FLI1*, *FLT3*, *SMAD2* and *SMAD4* and a downregulation of *CCL8*, *ETV6*, *LYL1*, *MPL*, *MPO*, *MYC*, *ATF4*, *CNR2* and *SPON1*. CML and ‘good’ Ph+ALL samples both similarly downregulated *LYL1*, a gene associated with ALL (McCormack et al., 2013). *FLT3* is associated with normal haematopoiesis, and similar upregulation between CML and ‘good’ Ph+ALL samples may indicate functional similarity of HSCs in these diseases. With the importance of

HSCs/LSCs in CML being previously discussed, this could indicate that such Ph+ALL samples share a CML-like gene expression and hence, may display a CML-like Ph+ALL genetic signature. This indicates that samples which may belong to the CML-like subgroup were identified as having good clinical outcomes by the single metric of blast cell clearance over 8 days with no investigation of long-term outcomes or incidence of relapse. The discordance between these results and those presented by Hovorkova et al (2017) will be elaborated in the discussion section of this chapter.

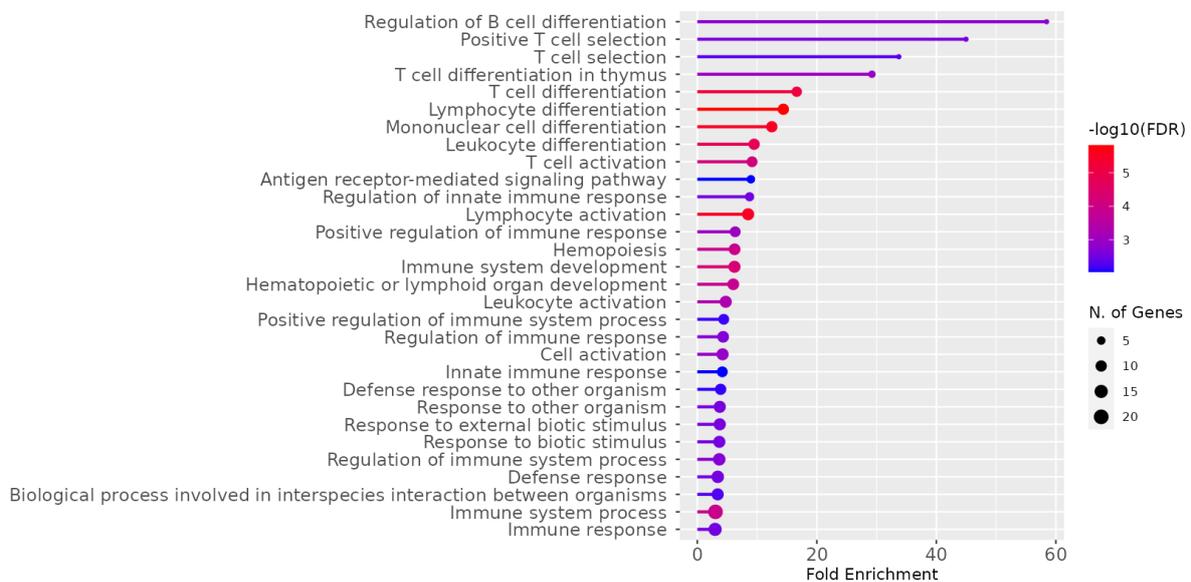


Figure 3.5.3: Gene ontology enrichment analysis for top 50 upregulated DEGs in Day 0 Ph+ALL good risk vs poor risk, dataset GSE39335. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. $\text{LogFC} = \log \text{fold change}$. ShinyGO v0.741.

3.6 Results I: Ph+ALL patient with hyperdiploidy shared gene expression profiles with CML and 'good risk' Ph+ALL samples from previous datasets, suggesting a consistent outlying transcriptional phenotype within the Ph+ALL cohort

Aligning with the analysis of dataset GSE39335, we next aimed to further elucidate shared gene expression patterns between CML and Ph+ALL by comparison of genes expressed in primary paediatric Ph+ALL bone marrow (GSE13425) to those observed in previous datasets (den Boer et al., 2009). GSE13425 allows assessment of gene expression patterns in a number of genetic and morphological subtypes of ALL. T-ALL, TEL-AML1, hyperdiploid, E2A-rearranged (EP), BCR::ABL1, MLL and Pre-B ALL paediatric bone marrow and peripheral blood were analysed for gene expression using micro-array. Ph+ALL samples were selected for further analysis. A notable inclusion in the BCR::ABL1+ samples analysed was 'Patient 142' who contained a hyperdiploidy along with a detectable Philadelphia chromosome. This provided a novel opportunity to investigate how Ph+ALL gene expression may be impacted from the gain of additional chromosomes and represents a sample which has cytogenetic features common to other ALL subtypes such as B-ALL (Chen, 2019). In contrast to the high-risk Ph+ALL, hyperdiploidy is associated with a favourable clinical outcome with a lower incidence of disease relapse post treatment (Tauro, 2003). Hence, this dataset allowed for the identification of high-risk Ph+ALL associated genes which can be detected in 'Ph+ only' and 'Ph+ and hyperdiploid' samples. Genes differentially regulated in 'Patient 142' represent hyperdiploidy associated disease phenotype, with genes which share an expression pattern across all of the samples being involved in *BCR::ABL1*-driven pathology.

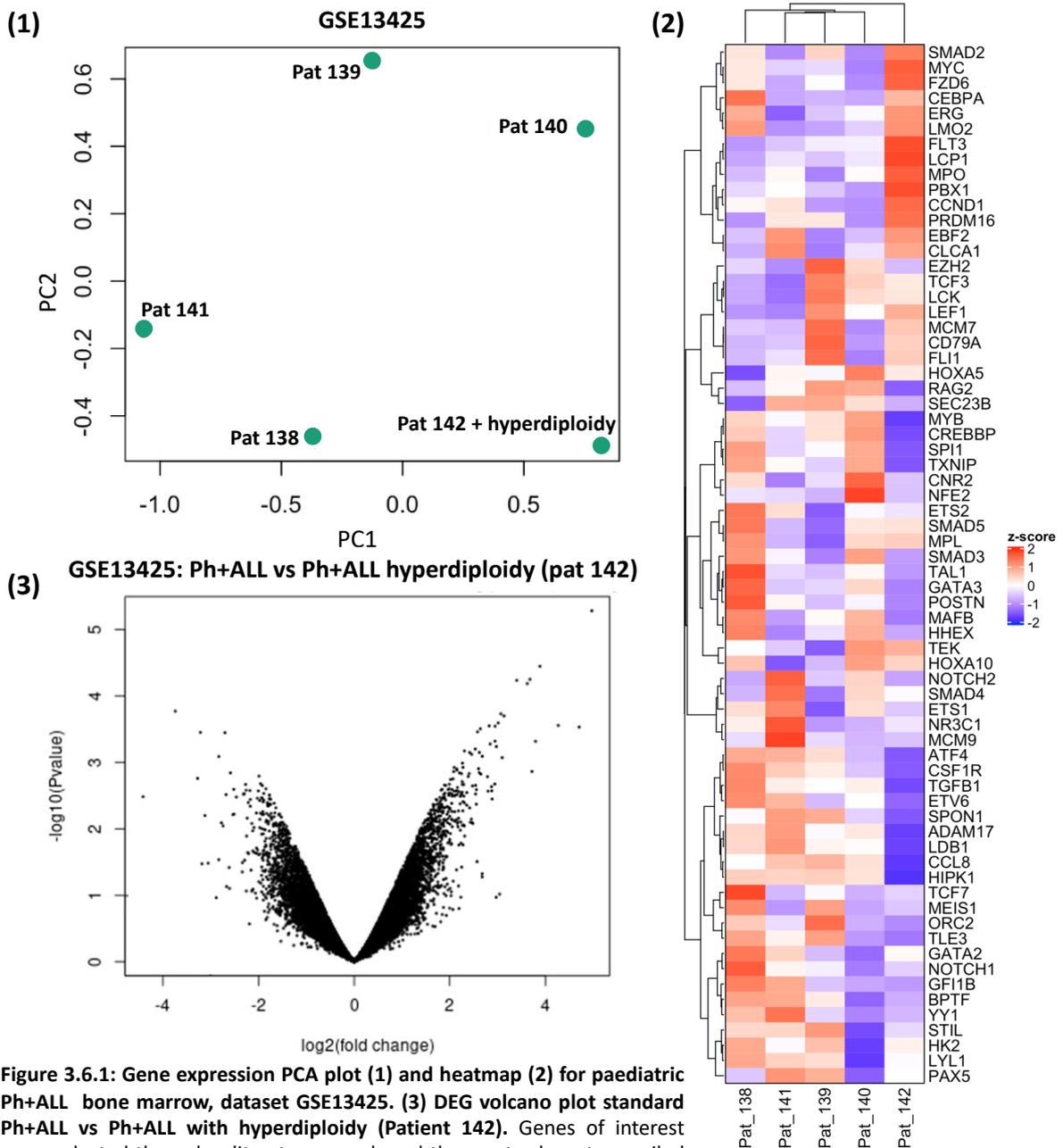


Figure 3.6.1: Gene expression PCA plot (1) and heatmap (2) for paediatric Ph+ALL bone marrow, dataset GSE13425. (3) DEG volcano plot standard Ph+ALL vs Ph+ALL with hyperdiploidy (Patient 142). Genes of interest were selected through a literature search and the most relevant compiled in list form. The changes in the expression level for these genes were analysed in primary paediatric Ph+ALL bone marrow samples from dataset GSE13425, the results of which visually displayed in a heatmap. *PCA plot P-value adjustment Benjamini & Hochberg (False discovery rate).* Volcano plot generated using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment.

Gene expression was profiled by array and expression data was loaded into R from GEO (gene expression omnibus) and analysed using the GEO2R platform as outlined in section 3.2. We first demonstrated heterogeneity between Ph+ALL samples in GSE13425 through PCA plot. In view of the significant heterogeneity, we next utilised the previously selected gene list to generate gene expression variation within a heatmap (Fig 3.10). Heatmaps were generated by selecting significant genes of interest and converting normalised and scaling raw gene counts into a normalised z-score as previously described (Fig 3.6.1.2). A notable but expected outlier was 'Pat_142' who displayed a vastly different gene expression pattern from the rest of the samples in the cohort and shared some gene expression with CML samples from datasets GSE11675 and GSE11889 (downregulation of *ATF4* and *LYL1*).

Volcano plots were generated to visualise differential gene regulation between the sample types examined in this dataset (Ph+ALL samples Pat_138-141 vs Ph+ALL hyperdiploidy Pat_142) (Fig 3.6.1.2). Differential gene expression analysis of all genes was done using Limma via GEO2R and LogFold change estimates. This DEG table was downloaded and exported to Excel where the gene names, gene functions, log fold changes and P values for each of the 250 DEGs could be observed (Table 3.6.1 and 3.6.2). Genes were sorted from upregulated to downregulated by sorting log fold changes by largest to smallest values. This process results in a table of DEGs for use in downstream analyses such as gene ontology and STRING mapping (Fig 3.6.2 & 3.6.3). After ranlong, differentially expressed genes were loaded into GSEA (MSigDB) for pathway analysis of all DEGs (significant and non-significant genes by adjusted P value). Genes of interest were determined as described previously.

'Pat_142' displayed a very different gene expression pattern from the rest of the Ph+ALL cohort (Fig 3.6.1). This sample also displayed a similar gene expression pattern to CML cells from GSE11675 and GSE11889. Such upregulated genes included *CCND1*, *FLT3*, *FZD6*, *LCP1*, *MPO*, *MYC* and *PRDM6* and downregulated genes include *CCL8*, *CREBBP*, *HIPK1*, *LDB1*, *MYB* and *TGFB1*. This evidence suggested that the outlier patient 142 may have a CML-like Ph+ALL disease phenotype and shares a similar gene expression pattern to CML progenitor cells (GSE11675 and GSE11889). Pat_142 also shared some gene expression similarities to Ph+ALL patients designated as 'good risk' in dataset GSE39335. Previously mentioned (and will be expanded in the discussion section of this chapter) is that the status of these samples as

having 'good' clinical outcomes is incongruous to what is currently understood about CML-like Ph+ALL. However, these patients were designated good risk based on blast cell clearance after 8 days of treatment with no long-term outcomes being assessed. It is unknown, therefore, if any of these patients had incidence of relapse or disease resurgence.

Gene	Gene function	logFC	P.Value
DDX3X	DEAD-box helicase 3, X-linked	4.97	0.0000522
IGHM	immunoglobulin heavy constant mu	4.71	0.00029446
IGHM	immunoglobulin heavy constant mu	4.27	0.00027876
DDX3X	DEAD-box helicase 3, X-linked	3.89	0.00003593
KLHL21	kelch like family member 21	3.8	0.00048394
PELI1	pellino E3 ubiquitin protein ligase 1	3.72	0.00137634
USP9X	ubiquitin specific peptidase 9, X-linked	3.68	0.00005632
HIPK1	homeodomain interacting protein kinase 1	3.62	0.00006527
TCF7L2	transcription factor 7 like 2	3.4	0.00005833
HNRNPL	heterogeneous nuclear ribonucleoprotein L	3.13	0.00019963
ZNF335	zinc finger protein 335	3.1	0.00085356
ITPKB	inositol-trisphosphate 3-kinase B	3.07	0.00018649
TRIM33	tripartite motif containing 33	3.01	0.00025524
RARA	retinoic acid receptor alpha	2.97	0.00057883
ZBTB43	zinc finger and BTB domain containing 43	2.95	0.00028651
MCL1	BCL2 family apoptosis regulator	2.95	0.0004811
HIP1R	huntingtin interacting protein 1 related	2.94	0.00212109
SPHK1	sphingosine kinase 1	2.89	0.00356725
ACSL1	acyl-CoA synthetase long-chain family member 1	2.88	0.00268491
NR4A3	nuclear receptor subfamily 4 group A member 3	2.87	0.00233234
DDX3X	DEAD-box helicase 3, X-linked	2.85	0.00053128
ENC1	ectodermal-neural cortex 1	2.8	0.00272152
IGHG1	immunoglobulin heavy constant gamma 1 (G1m marker)	2.79	0.00164679
PRRC2C	proline rich coiled-coil 2C	2.68	0.00081185
TCF4	transcription factor 4	2.65	0.00150918
KAT5	lysine acetyltransferase 5	2.64	0.00031309
DDX3X	DEAD-box helicase 3, X-linked	2.62	0.00071189
ERAP1	endoplasmic reticulum aminopeptidase 1	2.62	0.0014256
PRKD3	protein kinase D3	2.6	0.00124792
USP9X	ubiquitin specific peptidase 9, X-linked	2.58	0.00034862
RND1	Rho family GTPase 1	2.56	0.00258714
SRSF11	serine and arginine rich splicing factor 11	2.56	0.00426654
PDE4D	phosphodiesterase 4D	2.54	0.0022324
TRIB2	tribbles pseudokinase 2	2.54	0.00300397
PLEKHM2	pleckstrin homology and RUN domain containing M2	2.53	0.0006318
PEG10	paternally expressed 10	2.47	0.00226955
APLP2	amyloid beta precursor like protein 2	2.46	0.00075816
BANP	BTG3 associated nuclear protein	2.45	0.00719024
RASL10A	RAS like family 10 member A	2.42	0.00486528
ATP6V0C	ATPase H+ transporting V0 subunit c	2.41	0.0007145
PPP4C	protein phosphatase 4 catalytic subunit	2.41	0.00147974
CCNJ	cyclin J	2.39	0.00163478
LRP5L	LDL receptor related protein 5 like	2.36	0.00286077
SIGLEC6	sialic acid binding Ig like lectin 6	2.35	0.00442568
CDK11A/	cyclin dependent kinase 11A	2.31	0.00067581
MLXIP	MLX interacting protein	2.31	0.00407532
SETD1B	SET domain containing 1B	2.3	0.00322777
SUPT6H	SPT6 homolog, histone chaperone	2.29	0.00216426
PANK4	pantothenate kinase 4	2.28	0.00119334
ANXA11	annexin A11	2.28	0.0039065

Table 3.6.1: Top 50 DEGs upregulated in standard Ph+ALL vs Ph+ALL with hyperdiploidy (Patient 142), dataset GSE13455. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. *LogFC= log fold change.*

Gene	Gene function	logFC	P.Value
NEMP1	nuclear envelope integral membrane protein 1	-1.74	0.00678704
LCAT	lecithin-cholesterol acyltransferase	-1.75	0.0042695
GCG	glucagon	-1.76	0.00429699
FERMT1	fermitin family member 1	-1.77	0.0031039
IGLV1-44	immunoglobulin lambda variable 1-44	-1.77	0.00349097
H2AFX	H2A histone family member X	-1.77	0.00355473
EZH1	enhancer of zeste 1 polycomb repressive complex 2 subunit	-1.79	0.00444893
RAP1GAP2	RAP1 GTPase activating protein 2	-1.79	0.00628625
PLXND1	plexin D1	-1.79	0.00776081
GAS8	growth arrest specific 8	-1.8	0.0045203
EBAG9	estrogen receptor binding site associated, antigen, 9	-1.81	0.00418332
TRADD	TNFRSF1A associated via death domain	-1.81	0.00450196
RHBG	Rh family B glycoprotein (gene/pseudogene)	-1.82	0.00404596
MS4A4A	membrane spanning 4-domains A4A	-1.84	0.00742724
NMUR1	neuromedin U receptor 1	-1.85	0.00250671
MSR1	macrophage scavenger receptor 1	-1.89	0.0029489
SEL1L3	SEL1L family member 3	-1.9	0.00373341
HIBCH	3-hydroxyisobutyryl-CoA hydrolase	-1.9	0.00608658
HIST1H2BD	histone cluster 1, H2bd	-1.91	0.00393366
RAD51C	RAD51 paralog C	-1.91	0.00554041
HIST1H2BO	histone cluster 1, H2bo	-1.92	0.0022405
ZNF133	zinc finger protein 133	-1.92	0.0061625
OCRL	OCRL, inositol polyphosphate-5-phosphatase	-1.94	0.00342294
C2orf68	chromosome 2 open reading frame 68	-1.96	0.00210118
MXD3	MAX dimerization protein 3	-1.96	0.00583803
MATK	megakaryocyte-associated tyrosine kinase	-1.97	0.00721314
TOPORS-AS1	TOPORS antisense RNA 1	-1.98	0.00376637
MAGI2	membrane associated guanylate kinase, WW and PDZ domain containing 2	-1.99	0.00161632
EXO2	endo/exonuclease (5'-3'), endonuclease G-like	-2	0.00270322
RCBTB2	RCC1 and BTB domain containing protein 2	-2.03	0.00347712
AK2	adenylate kinase 2	-2.05	0.0057816
SERHL2	serine hydrolase-like 2	-2.17	0.00363141
TDP1	tyrosyl-DNA phosphodiesterase 1	-2.17	0.0046004
NAIP	NLR family apoptosis inhibitory protein	-2.22	0.00403845
SIL1	SIL1 nucleotide exchange factor	-2.24	0.00693112
CCL3L3	C-C motif chemokine ligand 3 like 3	-2.26	0.00263222
IL1B	interleukin 1 beta	-2.27	0.00215894
NGFR	nerve growth factor receptor	-2.3	0.00298696
HGF	hepatocyte growth factor	-2.32	0.00302818
SLC16A2	solute carrier family 16 member 2	-2.47	0.00333286
HIST1H1E	histone cluster 1, H1e	-2.51	0.00601028
NRXN3	neurexin 3	-2.52	0.00254509
TCL1A	T-cell leukemia/lymphoma 1A	-2.54	0.00614483
HIST1H2AE	histone cluster 1, H2ae	-2.59	0.00144775
KIR3DL2	killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 2	-2.68	0.00306132
HIST1H2BJ	histone cluster 1, H2bj	-2.7	0.0003591
HIST1H2BG	histone cluster 1, H2bg	-2.83	0.00082021
PLVAP	plasmalemma vesicle associated protein	-3.12	0.00633722
IL1B	interleukin 1 beta	-3.28	0.00174765
HSPA6	heat shock protein family A (Hsp70) member 6	-3.74	0.00016991

Table 3.6.2: Bottom 50 DEGs downregulated in standard Ph+ALL vs Ph+ALL with hyperdiploidy (Patient 142), dataset GSE13425. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. *LogFC*= *log fold change*.

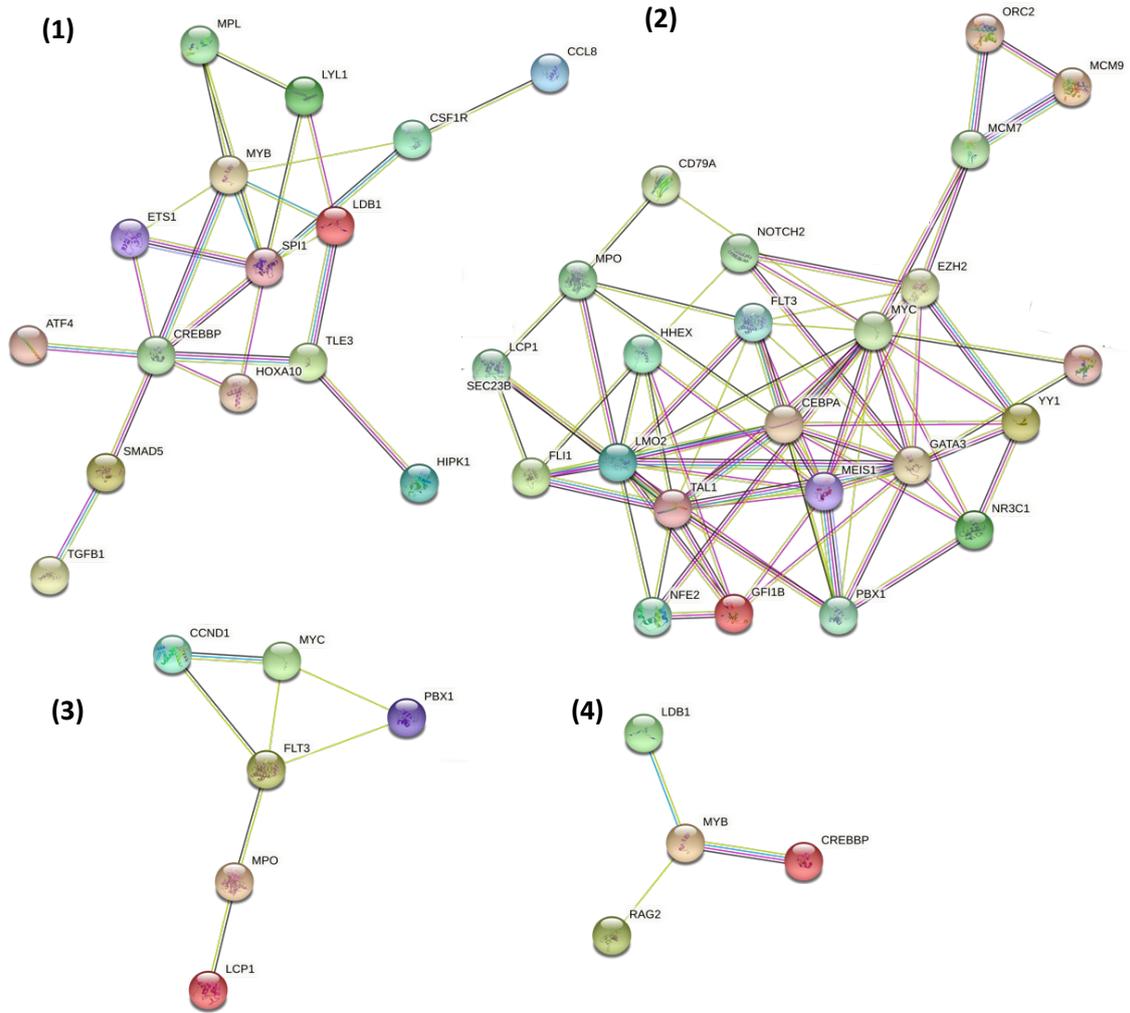


Figure 3.6.2: Functional enrichment analysis of genes upregulated in primary paediatric Ph+ALL bone marrow samples, dataset GSE13425. Genes of interest were selected through a literature search and the most relevant compiled in list form. The changes in the expression level for these genes were analysed in paediatric primary Ph+ALL bone marrow samples from dataset GSE13425, the results of which visually displayed in a heatmap. From the heatmap produced, genes observed to be upregulated were selected and a STRING gene-gene interaction network produced to display the functional relevance of gene expression changes. (1) Genes upregulated in primary paediatric Ph+ALL bone marrow samples. (2) Genes downregulated in primary paediatric Ph+ALL bone marrow samples. (3) Genes upregulated in hyperdiploidy patient (Patient 142) compared to Ph+ALL patients from the same paediatric cohort. (4) Genes downregulated in hyperdiploidy patient (Patient 142) compared to Ph+ALL patients from the same paediatric cohort.

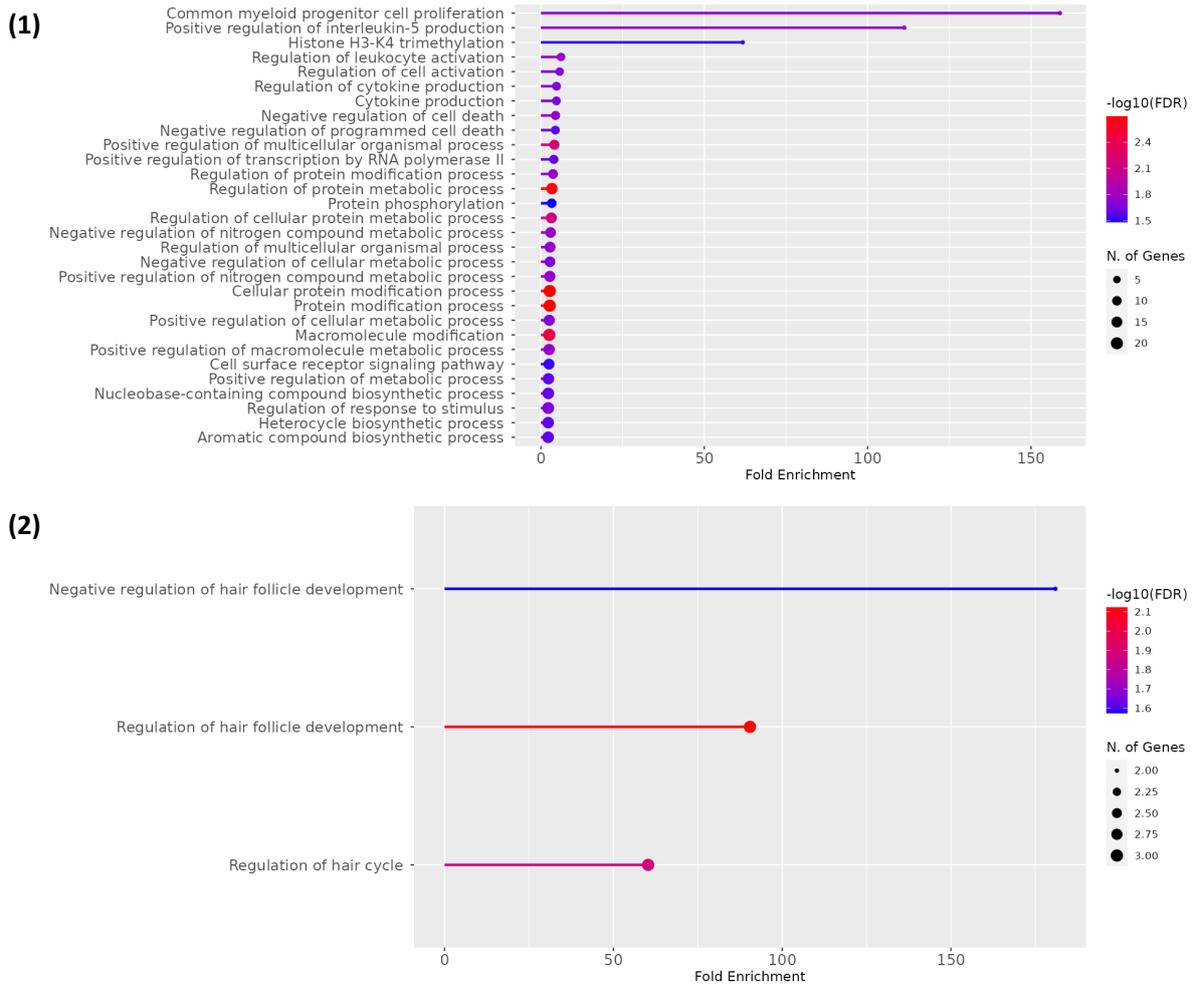


Figure 3.6.3: Gene ontology enrichment analysis for op 50 DEGs upregulated and (2) bottom 50 downregulated in standard Ph+ALL vs Ph+ALL with hyperdiploidy (Patient 142), dataset GSE13425. DEGs identified using Limma via GEO2R. Benjamini & Hochberg (false discovery rate) P-value adjustment. $LogFC = \log \text{fold change}$. ShinyGo v0.741

3.7 Discussion

The overall aim of this chapter was to investigate, through publicly available datasets, gene expression patterns in CML, Ph+ALL and normal sample with the goal to characterise potential gene signatures in the proposed CML-like Ph+ALL subtype. We did this through sequential analysis of samples to understand:

- Differences between Ph+ALL and lymphoid blast phase CML
- Differences within cell maturity of normal and Ph+-driven cells
- Differences between normal HSCs and Ph+-driven HSCs (CML)
- Differences between 'good' and 'poor' responders

Five publicly available datasets were utilised to investigate this: GDS4175, GSE11675, GSE11889, GSE39335 and GSE13425. These selected datasets encompassed gene expression in primary paediatric Ph+ALL bone marrow samples, healthy bone marrow controls, CML progenitors and chronic phase bone marrow samples, and a cell line dataset was selected to allow direct comparison of age, disease status and *BCR::ABL1* fusion variant. A clear heterogeneity was seen between primary patient samples, as well as within different cell lines through analysis of DEG and pathway analysis. This included (but was not limited to) genes involved in cell survival, cancer progression, stemness and regulation of haematopoietic cell development such as *CCND1*, *FLT3*, *FZD6*, *LCP1*, *MPO*, *MYC* and *PRDM6*, *CCL8*, *CREBBP*, *HIPK1*, *LDB1*, *MYB* and *TGFB1*. Relevant genes involved in haematopoiesis homeostasis and leukaemia initiation and development were selected based on literature searches and compiled as a list ('genes of interest').

In alignment with clinical phenotype, early progenitor (Lin- CD34+ and Lin- CD34-) CML cells and chronic phase CML HSCs (GSE11675 and GSE11889 respectively) displayed a less acute phenotype than Ph+ALL samples with a downregulation of acute leukaemia-associated genes such as *ERG*, *LYL1*, *SPON1* and *MPL*. This is likely due to the maintenance of quiescent leukaemic stem cells in CML chronic phase, prior to a switch toward an acute phenotype, increase in proliferation and disease progression. With the leukaemic cell of origin in CML

being HSCs (LSCs), this switch in cell activity is observed by an upregulation of acute leukaemic genes as listed above and downregulation of genes necessary for quiescence.

Comparison of 'good' and 'poor' clinical outcome Ph+ALL samples (GSE39335) displayed similarity between early progenitor and chronic phase CML samples (GSE11675 and GSE11889), and Ph+ALL samples designated as having good clinical outcomes. A non-exhaustive list of genes similarly upregulated in 'good' Ph+ALL and progenitor CML include *FLI1*, *FLT3*, *YY1*, *CCL8*, *LYL1*, *ATF4* and *MPL*. This contrasts with current understandings of CML-like Ph+ALL which predict a poor clinical outcome for patients belonging to this subtype. However, the classification of 'good' or 'poor' risk patients in GSE39335 was based on blast cell reduction over 8 days from commencement of treatment. Such basic metrics do not include whether patients relapsed, died, received transplants or length of survival time and does not reflect long-term response to lymphoid-directed treatment, a criterion suggested by Hovorkova et al (2017) to be important in differentiating between Ph+ALL subtypes. As lymphoid cells are sensitive to glucocorticoids, an initial reduction in lymphoblasts would be expected in either Ph+ALL subtype and would not reflect abnormal myeloid populations that may exist in CML-like Ph+ALL (Smith and Cidlowski., 2016). Hence, without longer-term clinical information, the overall clinical outcomes could not be defined as 'good' or 'poor' and though a potential CML-like transcriptional profile may have been observed in the samples designated 'good', this does not disprove the theory brought about in Hovorkova et al (2017) that patients belonging to this subtype are more likely to relapse or respond poorly long-term to lymphoid-directed treatment.

Such contrasting results to those presented by Hovorkova et al (2017) also highlight the difficulties in investigating an (as of yet) undefined disease subtype using publicly available datasets. Current CML-like Ph+ALL designation relies on discordance of Ig/TCR rearrangement and BCR::ABL1 transcript MRD methods in ALL and non-ALL cells. With the variety of MRD methods available, low likelihood of multiple MRD methods being used concurrently in the clinic and scant clinical information provided in GEO dataset, samples utilised for bioinformatic analyses therefore cannot be defined as being 'standard' or 'CML-like' from the outset of analysis. Hence, it is unknown how many (if any) samples in these datasets belong to the CML-like subtype. In addition, the transcriptional profile of CML-like Ph+ALL is

unknown, meaning that CML-like samples cannot be distinguished by interrogating datasets for specific differentially expressed genes. Despite these caveats, exploratory investigation into paediatric Ph+ALL gene expression datasets with the scope to identify any patient subsets expressing CML-related genes was conducted.

	Gene name	Function
Upregulated	ATF4	activating transctipion factor 4
	CCL8	C-C motif chemokine ligand 8
	CCND1	cyclin D1
	FLI1	Fli-1 proto-oncogene, ETS transcription factor
	FLT3	fms related tyrosine kinase 3
	FZD6	frizzled class receptor 6
	LCP1	lymphocyte cytosolic protein 1
	LYL1	basic helix-loop-helic family member
	MPL	MPL proto-oncogene, thrombopoietin receptor
	MPO	myeloperoxidase
	MYC	v-myc avian myelocytomatosis viral oncogene homolog
YY1	YY1 transcription factor 1	
Downregulated	CREBBP	CREB binding protein
	HIPK1	homeodomain interacting protein kinase
	LDB1	LIM domain binding
	MYB	MYB proto-oncogene, transcription factor

Table 3.7: Proposed CML-like gene expression pattern. Differentially expressed genes identified in CML and potential CML-like Ph+ALL samples from publicly available GEO datasets.

Taking such caveats into consideration, Ph+ALL samples with similar gene expression profiles as CML samples were compared to the Ph+ALL cohort in GSE13425. Patient 142 was initially highlighted due to the presence of a hyperdiploidy cytogenetic abnormality in addition to BCR::ABL1 fusion. Upon construction of a gene expression heatmap, Patient 142 could be clearly seen to have a different gene expression pattern in comparison to the rest of the cohort. When comparing gene expression of Patient 142 to ‘good’ Ph+ALL samples and CML samples, similar patterns could be observed such as the upregulation of *CCND1*, *FLT3*, *FZD6*, *LCP1*, *MPO*, *MYC* and *PRODM6*, and the downregulation of *CCL8*, *CREBBP*, *HIPK1*, *LDB1*, *MYB* and *TGFB1*. This congruence between outlying Ph+ALL and CML samples allowed for a list of

genes similarly expressed across such samples to be assembled (Table 3.7). These genes have been tentatively dubbed a 'CML-like' gene expression pattern which would be beneficial to examine in larger datasets and datasets which contain both CML and Ph+ALL samples which would prevent batch effects between samples.

Further follow up would be essential to confirm these findings and may be achieved by interrogating additional CML and Ph+ALL datasets and utilising clinical outcome notes to confirm if patients with CML-like Ph+ALL share a similar gene expression pattern which may be used for detection of these patients in a diagnostic setting by generation of a biomarker panel. In order to explore this further, gene expression in Ph+ALL and CML-lymphoid blast phase samples will be examined by RNAseq (Chapter 5). As a caveat of this chapter was that datasets containing a direct comparison of gene expression in CML and Ph+ALL was unavailable, comparison of a number of datasets with relevant leukaemic samples was utilised. Limitations also included the scarcity of datasets which included primary sample data which directly compared CML and Ph+ALL, hence in this chapter the closest dataset to this was one which investigated cell lines. Though helpful, cell lines represent immortalised cells to allow long term growth and may differ in metabolic and pro-survival genes from primary cells. Additionally, only one sample from each category was available (paediatric, adult, CML or Ph+ALL), meaning that more samples would be required to confer significance. Cell lines are useful models of disease but may not fully reflect the transcriptional picture of disease within patients. Additionally, publicly available datasets were sourced from different labs and projects, meaning that not only did the technology used to detect gene expression differ (ie microarray vs RNAseq) but these also represent separate experiments which cannot be directly compared due to batch effects which through technical/non-biological factors, affect variation in resultant data. Due to these datasets originating from different experiments, some data came from sorted samples (ie Lin- progenitors) and some from bulk samples, making the data too different for direct statistical comparison. Work continuing within this project aimed to follow these findings up through the utilisation of RNAseq analysis of both Ph+ALL and CML samples. This allowed for the observation of any CML-like gene expression patterns and, where available, was followed up using information on clinical outcome.

Chapter 4. Results II. Unravelling Ph+ALL heterogeneity with FACS and FISH.

4.1.1 Introduction: The cell of origin in Ph+ALL remains elusive

When investigating haematological malignancies, the dynamic process of haematopoiesis is an important concept to consider. Applying the understanding of early blood cell development to the initiation and maintenance of leukaemia has allowed for the molecular basis of diseases to be understood. A good example of this is CML, where the identification of BCR::ABL1 fusion during early haematopoiesis has not only allowed for a deeper understanding of leukaemia biology and the stem cell nature of the disease, but improved treatment by exploitation.

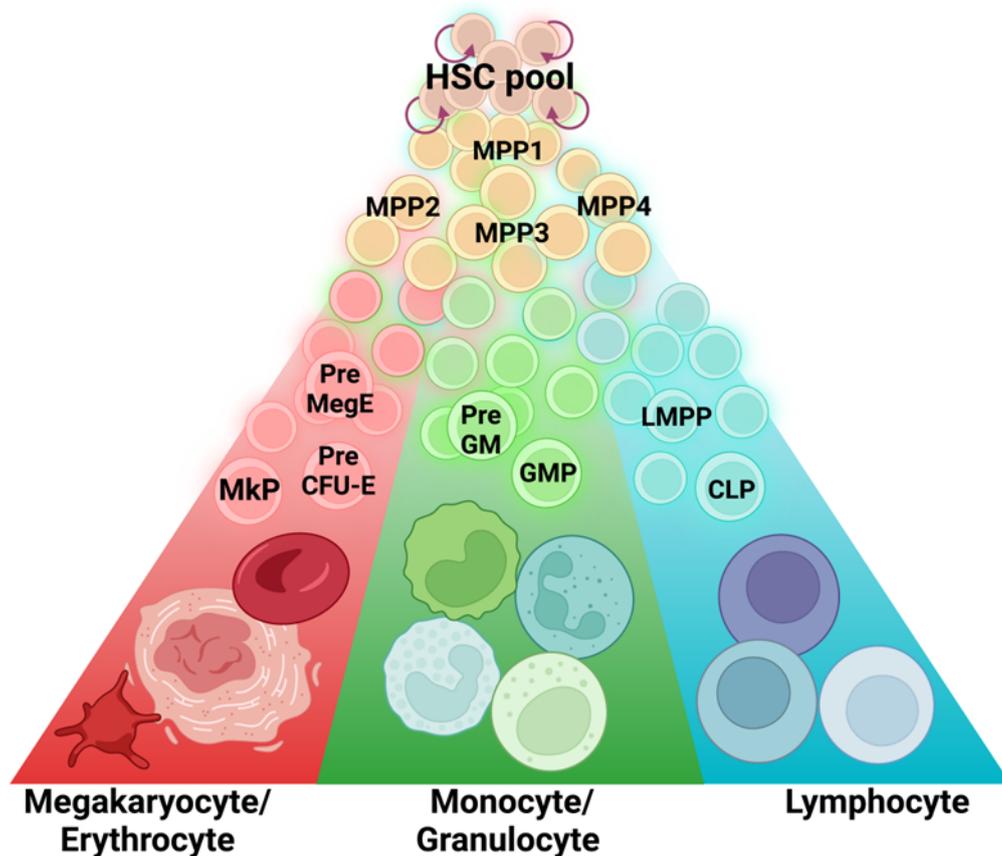
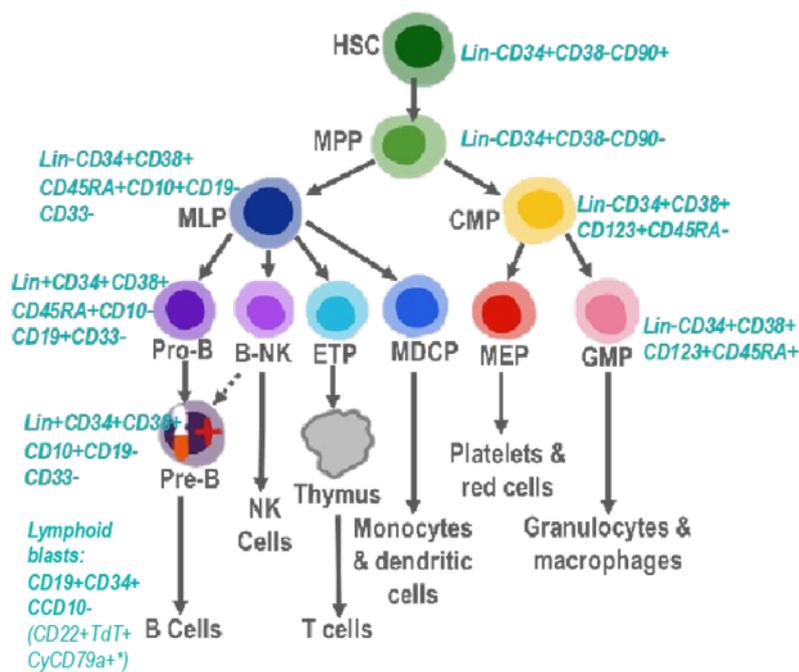


Figure 4.1.1: Model of haematopoiesis at steady state. Cell fate decisions are made during haematopoietic development and differentiation into mature cell populations. Subgroups of MPPs display lineage-specific predetermination (MPP1, 2, 3 and 4). *(created using BioRender)*

As previously discussed, haematopoiesis is the process through which mature blood cells are produced (Rodriguez-Fraticelli et al., 2018). This requires maintenance of multi- or pluripotent progenitor cells balanced with the differentiation and proliferation of mature lineage-committed cells. Previous models of haematopoiesis have displayed this process in a stepwise manner, with definitive stages of differentiation and clear groups of progenitors. However, with increasing understanding of haematological development through single cell tracking and transplantation studies, this model has been revised (Notta et al., 2016 and Sun et al., 2014). Newer understandings of haematopoiesis depict a continuous differentiation process with lineage predetermination in multilineage progenitors, such as multipotent progenitors (MPPs), but which maintain the ability to differentiate into different cell types based on organismal needs and signalling networks provided (Fig 4.1.1).



Standard Ph+ALL

**clinical markers, not used in this project (Chiaretti et al., 2014)*

Figure 4.1.2: Haematopoietic population surface marker expression for FACS. Surface marker expression profile used for cell sorting based off those used clinically.

An important feature of haematopoiesis that should be remembered when examining differences in cell population size between diagnostic and post induction samples throughout this chapter, is its plasticity. Within lineage subsets, cellular needs are met by variation of granulopoiesis, erythropoiesis and lymphopoiesis (Zaretsky et al., 2015) (Fig 4.1.2). Not only

does haematopoiesis naturally change as the organism ages (ie from embryonic to adult) but also in response to infections, malignancy and different therapeutic agents (Belyavsky et al., 2021).

As discussed previously, within CML, the BCR::ABL1 fusion originates within the stem cell compartment, namely HSC or MPP (Holyoake and Vetrie., 2016). BCR::ABL1 fusion in cells prior to lineage bifurcation results in observable BCR::ABL1 activity in cells of a number of lineages and further explains why the acute blast crisis phase of CML can be characterised by the presence of lymphoid or myeloid (or both) differentiation-arrested blast cells (Calabretta and Perrotti, 2004). Until recently, there was an orthodoxy that Ph+ALL was a distinct disease entity, with the mutation felt to arise from a progenitor B cell and believed not to occur outside of the lymphoid lineage compartment. However, recent data has challenged this binary separation (Nishiwaki et al., 2020). Expression of BCR::ABL1 in non-lymphoid lineages such as in MPO-myeloid cells indicated that the leukaemia-driving fusion event occurred in a differentiation stage prior to lymphoid lineage development (Fig 4.1.3). This was further expanded during the investigation into MRD (minimal residual disease) monitoring in paediatric Ph+ALL where a subgroup of patients were identified as harbouring BCR::ABL1 in non-ALL lymphocytes and myeloid cells (Hovorkova et al., 2017). These patients, defined as CML-like Ph+ALL, were identified as having discordant MRD due to the presence of BCR::ABL1 and absence of Ig/TcR clonal rearrangements in the non-ALL blast Ph+ cells. This data raised questions about the transcriptomic and clinical relevance of this population within Ph+ALL.

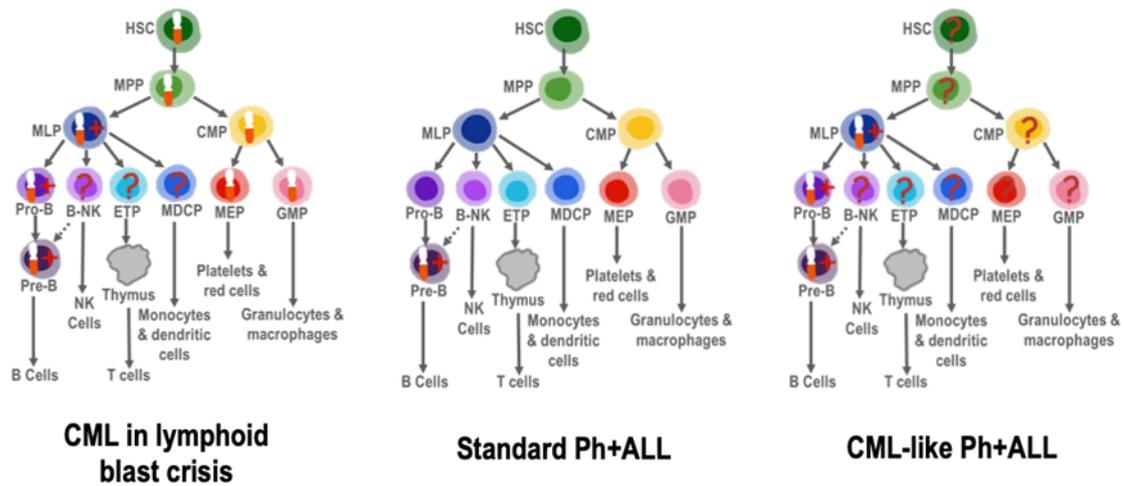


Figure 4.1.3: Representation of BCR::ABL1 fusion during early stages of haematopoiesis in three Ph+ disease states. Three disease states: chronic myeloid leukaemia in lymphoid blast crisis, standard Ph+ALL and CML-Like Ph+ALL.

HSC= *haematopoietic stem cell*
MPP= *multipotent progenitor*
MLP= *mixed lymphoid progenitor*
CML= *common myeloid progenitor*
MEP= *megakaryocyte-erythrocyte progenitor*
GMP= *granulocyte-macrophage progenitor*
NK= *natural killer cell*
MDCP= *monocyte-dendritic cell precursor*
ETP= *early T cell precursor*

Evidence that expression of BCR::ABL1 in multi-lineages can be associated with improved prognosis (Nishiwaki et al., 2020) contrasted with the finding that patients classified CML-like Ph+ALL benefitted from stem cell transplants (SCT) more than standard Ph+ALL treatment of chemotherapy alone (Hovorkova et al., 2017). These conflicting results display that the clinical outcomes of CML-like Ph+ALL is unknown and therefore alteration to standard treatment protocols may be beneficial. Bioinformatic analysis of publicly available Ph+ALL and CML datasets in the previous results chapter highlighted a potential gene expression profile that could delineate these disease populations. However, questions remain over the cell of origin for this population and if it truly represents a new disease entity within Ph+ALL. To date, this population remains poorly defined.

4.1.2 Aims

The aims of this chapter were, therefore, to:

1. Assess the heterogeneity or homogeneity of haematopoietic cell populations within paediatric and adolescent Ph+ALL samples;
2. To identify where the BCR::ABL1 fusion occurs within stem and progenitor populations of Ph+ALL samples to potentially determine the 'CML-like' patient samples;
3. To correlate outcomes with clinical characteristics.

4.2.1 Results II: Fluorescence-assisted cell sorting revealed a potential subgroup of Ph+ALL patients with atypical haematopoietic dynamics

Population	Surface Marker Expression
HSC	LIN- CD34+ CD38- CD90+ CD93+/-
MPP	LIN- CD34+ CD38- CD90-
CMP	LIN- CD34+ CD38+ CD123+ CD45RA-
GMP	LIN- CD34+ CD38+ CD123+ CD45RA+
CLP	LIN- CD34 ^{hi} /+ CD38+ CD45RA+ CD10+ CD19- CD33-
ProB	LIN+ CD34+ CD38+ CD45RA+ CD10- CD19+ CD33-
PreB	LIN+ CD34+ CD38+ CD10+ CD19+ CD33-

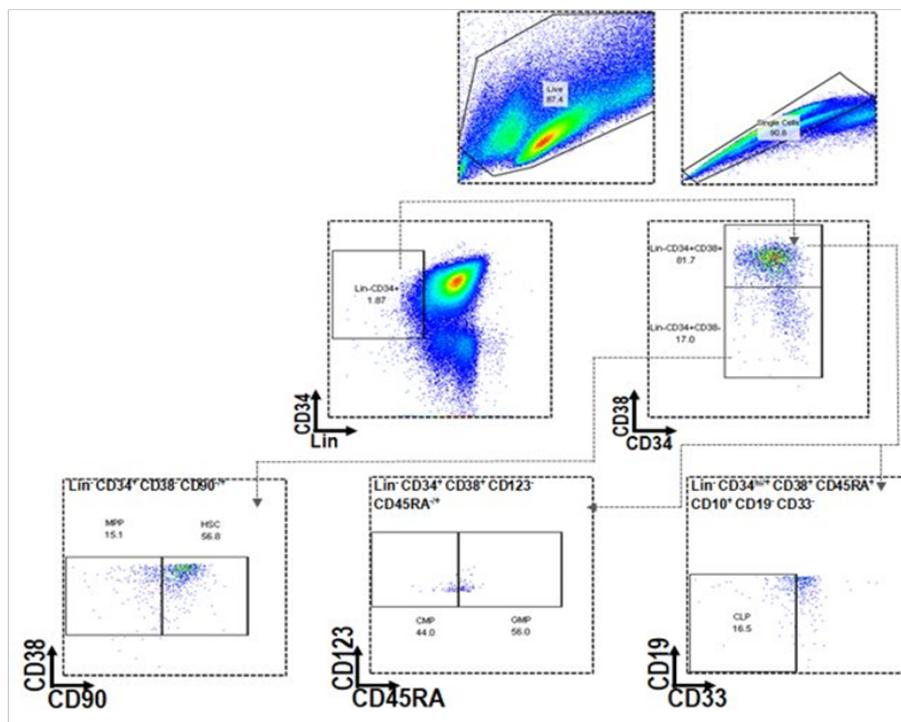


Figure 4.2.1: Antibody cocktails used in FACS (1) and applied sorting strategy (Ph+ALL diagnostic, 5yo M) (2). Haematopoietic stem and progenitor cells were isolated and collected from bulk BMNC samples using FACS. Cell populations were identified by surface marker expression. Progenitor populations were sorted from bulk BMNC. Diagnostic (untreated) Ph+ALL samples contain a high presence of lymphoid blast cells. *Data my own.*

Primary samples were sorted into stem and progenitor populations by cell surface marker expression to enable identification of BCR::ABL1 fusion events during early haematopoiesis (Fig 4.1.2, 4.1.3 & 4.2.1). Cells were strained following antibody staining to prevent clumping. Live cells were gated, and doublets removed before cells were sorted according to table 4.2.1. FMO controls were used within each sample to ensure true representation of cell populations. Within the first few samples, an aliquot of the sorted cells was run again on flow cytometry to ensure cell surface expression was comparable to the populations selected. Because of

this, the purity of the sort produced populations that were 95-99% pure for the desired cells.

An example sort, with FMO controls, is depicted in figure 4.2.2.

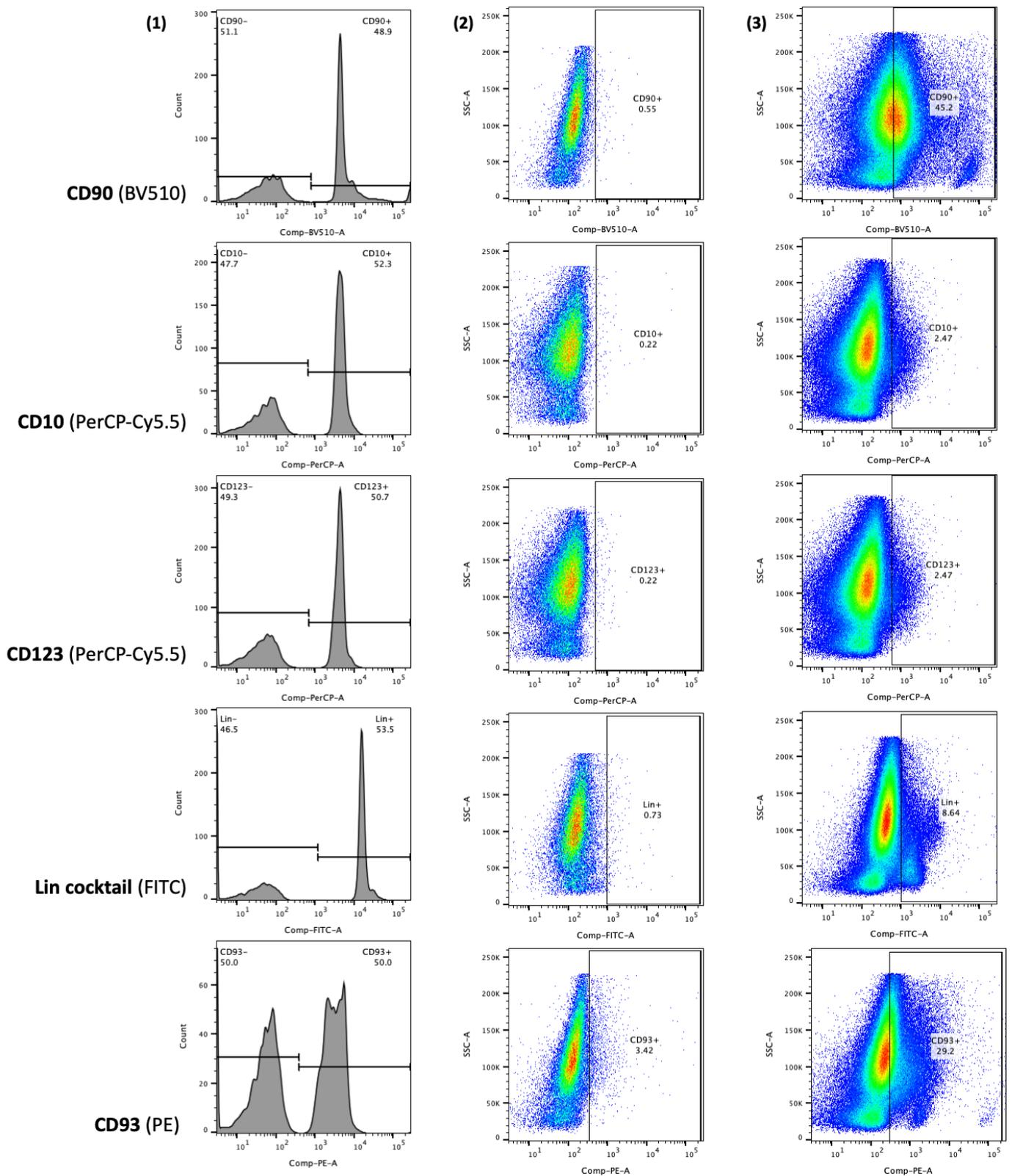


Figure 4.2.2: Fluorescence minus one (FMO) control (surface marker and fluorophores indicated). (1) Single antibody stains displaying positive fluorophore peaks using flow cytometry control and standardisation beads. (2) antibody cocktail minus one antibody, displayed on live single cells. Side scatter versus fluorophore for removed antibody. Gate cut-off determined by single stain bead control. (3) antibody cocktail with all required antibodies included to display positive signal on live single cells. (cells for all controls were from sample PALL7FPI)

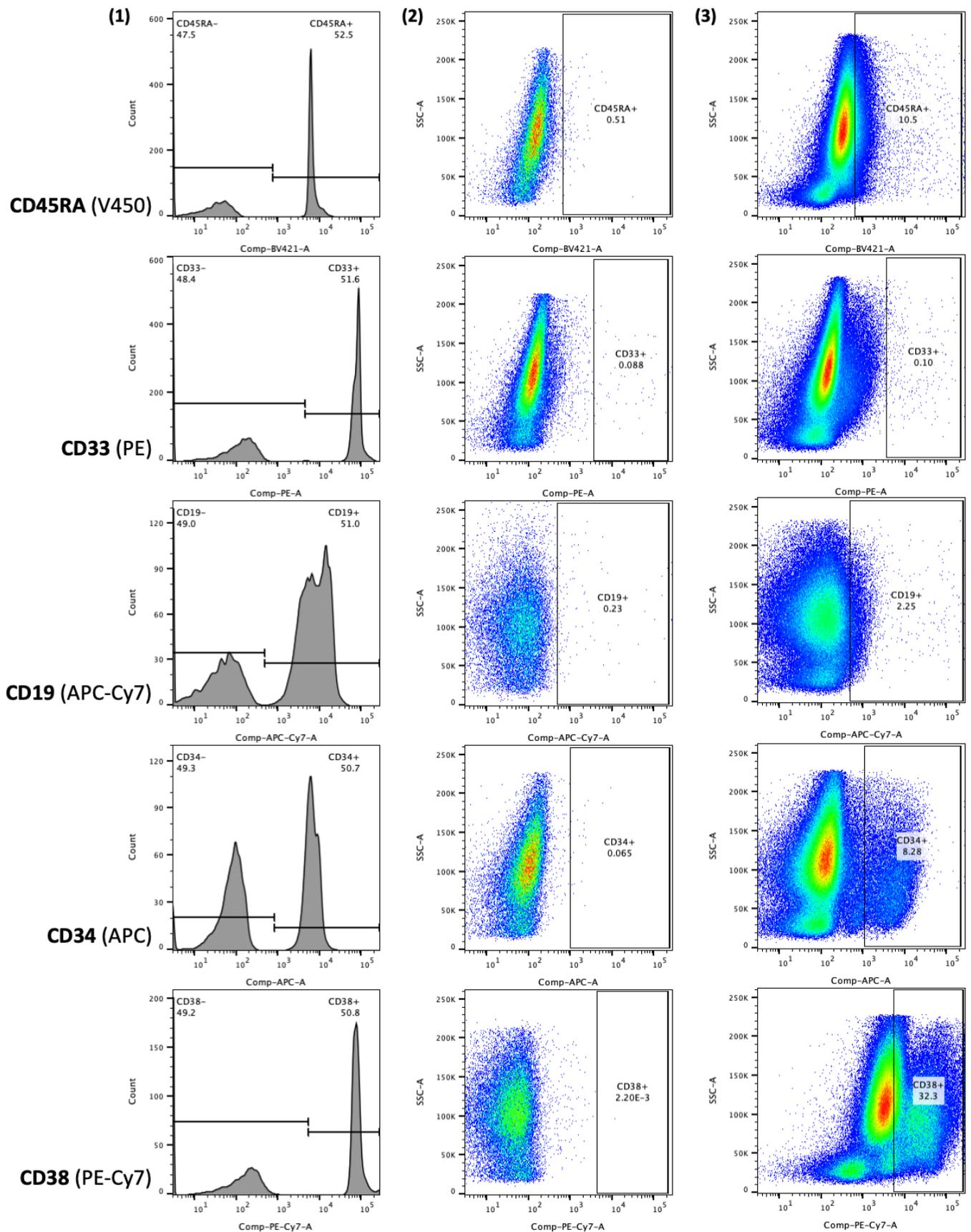


Figure 4.2.2 (contd): Fluorescence minus one (FMO) control (surface marker and fluorophore indicated). (1) Single antibody stains displaying positive fluorophore peaks using flow cytometry control and standardisation beads. (2) antibody cocktail minus one antibody, displayed on live single cells. Side scatter versus fluorophore for removed antibody. Gate cut-off determined by single stain bead control. (3) antibody cocktail with all required antibodies included to display positive signal on live single cells. (cells for all controls were from sample PALL7FPI)

23 diagnostic Ph+ALL (aged 2-19 years) and 16 post induction (PI) Ph+ bone marrow mononuclear cells (BMNC) samples were sorted into HSC, MPP, CMP GMP, and CLP populations. PI samples used in this chapter were matched to diagnostic samples and harvested at a range of timepoints (average day 70), where patients were at different treatment stages (Fig 4.2.3).

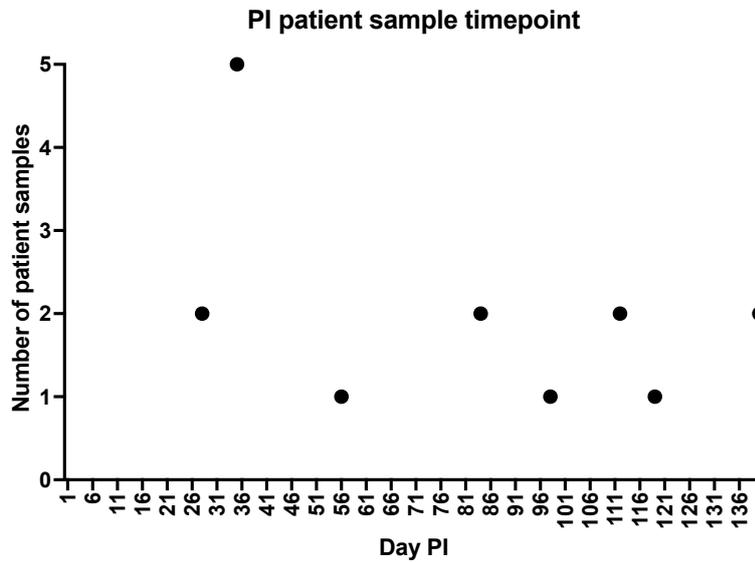


Figure 4.2.3: Post induction sample timepoint (days). Number of available patient samples and day of sampling post commencement of treatment induction.

We next assessed the heterogeneity through calculation of the proportion compared to live cells. In brief, this is calculated as the number of sorted cells for each progenitor population divided by the number of live cells measured during the sort (Fig 4.2.4). This is then expressed as a percentage. As these are human samples, live cells were felt to be an appropriate denominator. The cell populations for diagnostic and PI samples are shown in Fig 4.2.5.1 and 4.2.5.2. Upon performing FACS to isolate stem and progenitor samples, the presence of blast cells may convolute detection of lymphoid progenitors due to similarities in surface marker expression profiles (Fig 4.1.2 & 4.2.1). This data is also displayed with matched samples, displaying diagnostic and PI samples for the same patient as indicated by the colour key on the right side of the graph. Analysis of cell populations demonstrated significant interpatient heterogeneity between samples at diagnosis and post induction. At diagnosis, a Ph+ALL bone marrow sample would be expected to have large populations of CLPs and small populations of HSCs/MPPs and myeloid precursors. This is due to the skewing of haematopoiesis toward

the overproduction of lymphoid blasts, a hallmark of Ph+ALL. Hence, the expected population sizes at PI timepoints (after patients have received lymphoid-directed treatment) would include smaller populations of lymphoid progenitors such as CLPs, more similar to levels during normal haematopoiesis. Additionally, myeloid progenitors and HSPCs are expected to be larger in PI than in diagnostic samples, with the restoration of normal haematopoiesis being restored and the overproduction of lymphoblasts halted. Steroid treatment can also alter cell number, exemplified by leucocytosis/granulocytosis post glucocorticoid (ie dexamethasone) treatment (Nagakawa et al., 1998). As previously mentioned, PI samples were harvested at different timepoints during treatment (average day 70) and hence, cell numbers may vary depending on the treatment phase of PI samples. Interestingly, interpatient heterogeneity atypical of expected haematopoietic development patterns was observed in diagnostic and PI samples for a small number of patients. This will be discussed in-depth by population in later sections of this results chapter.

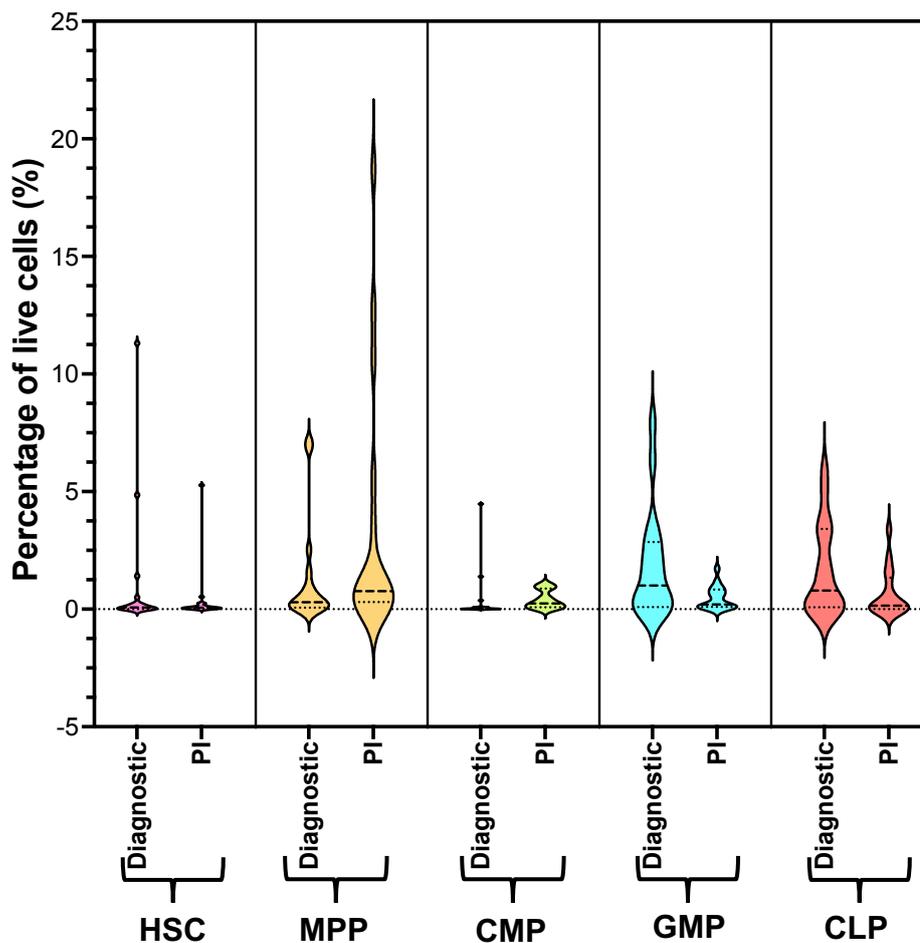


Figure 4.2.4: Distribution of progenitor cells (HSC, MPP, CMP, GMP and CLP) in PI Ph+ALL samples. Percentage of live cells for each progenitor population (HSC, MPP, CMP, GMP and CLP) identified during FACS sorting of Ph+ALL diagnostic and PI samples. Number of patients represented by each violin plot: HSC diagnostic= 17, HSC PI= 18, MPP diagnostic= 17, MPP PI= 18, CMP diagnostic= 17, CMP PI= 18, GMP diagnostic= 17, GMP PI= 18, CLP diagnostic= 16 and CLP PI= 16.

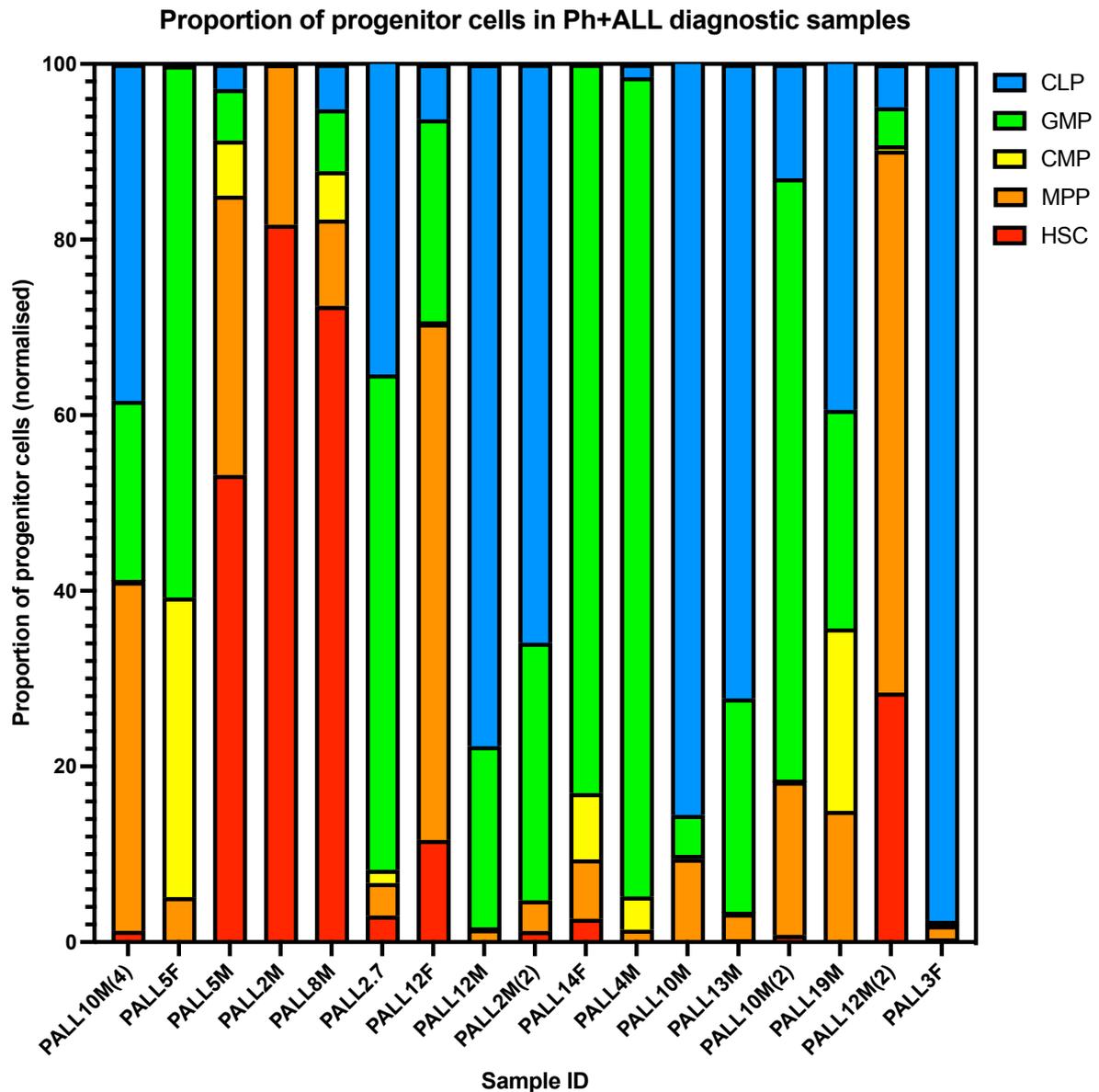


Figure 4.2.5.1: Proportion of progenitor cells (HSC, MPP, CMP, GMP and CLP) in diagnostic Ph+ALL samples. Proportion for each progenitor population was generated by normalising the percentage of live cells Lin⁻ for each population of total cells counted. This proportion was combined with the other progenitor groups in the sample to provide a total progenitor number expressed as 100%. Each progenitor group was divided by the total and multiplied by 100 to display its proportion of all progenitors measured during FACS. This calculation was done separately for each diagnostic patient sample. Each progenitor group is indicated by colour.

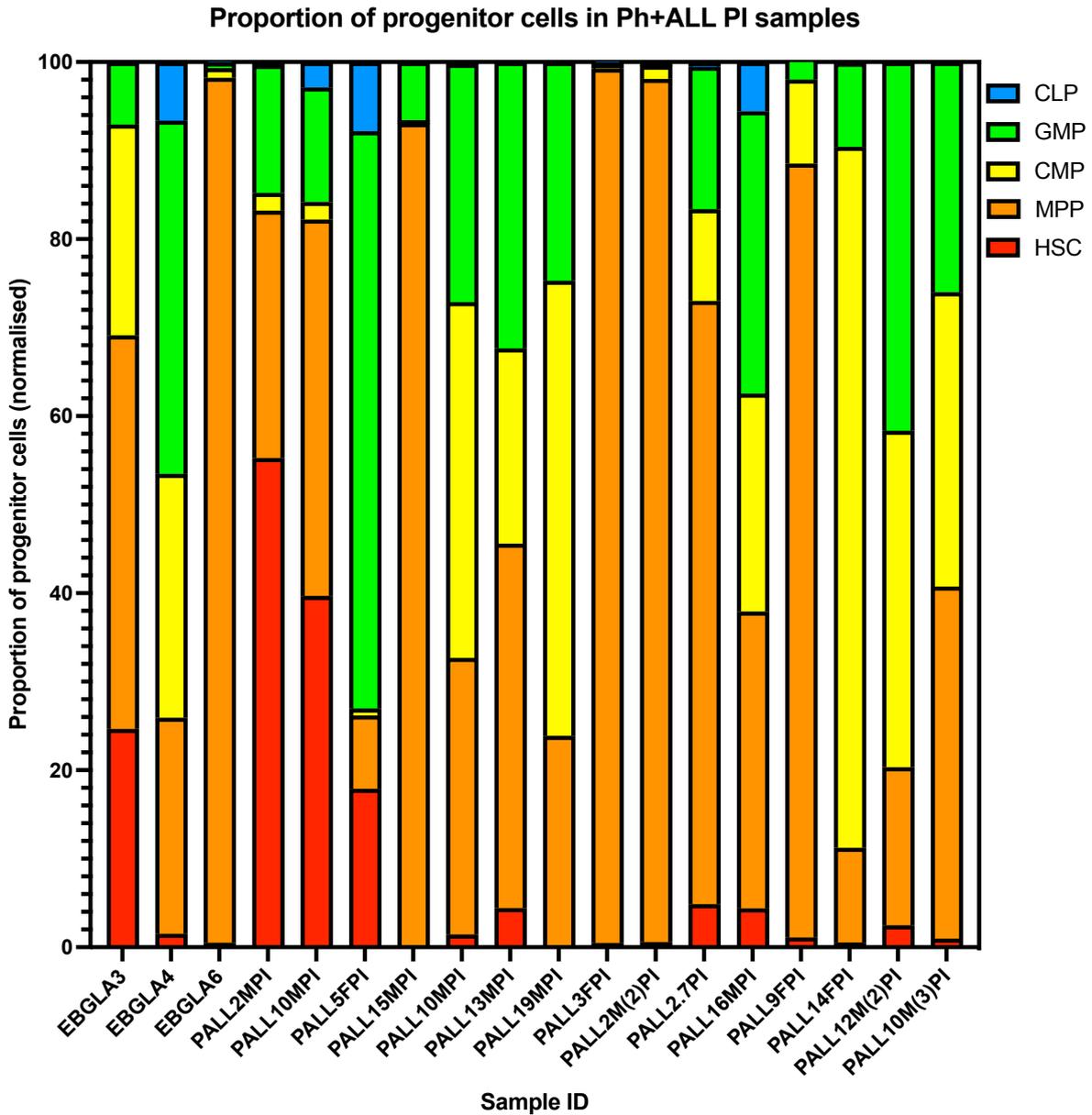


Figure 4.2.5.2: Proportion of progenitor cells (HSC, MPP, CMP, GMP and CLP) in PI Ph+ALL samples. Proportion for each progenitor population was generated by normalising the percentage of live Lin- cells for each population of total cells counted. This proportion was combined with the other progenitor groups in the sample to provide a total progenitor number expressed as 100%. Each progenitor group was divided by the total and multiplied by 100 to display its proportion of all progenitors measured during FACS. This calculation was done separately for each PI patient sample. Each progenitor group is indicated by colour.

4.2.2 Results II: HSC and MPP populations can be identified using FACS

As blood cell development in Ph+ALL is skewed toward lymphopoiesis, the HSC population size is expected to be small. The scarcity of these cells is also illustrated in healthy bone marrow in which, HSCs and MPPs represent rare populations with low frequency (0.01%) (Rossi et al., 2011). Of the approximately 10,000 HSCs in healthy BM, 1000 are predicted to contribute to haematopoiesis to maintain and restore blood cell production (Catlin et al., 2011) (Anthony and Link, 2015). In leukaemic bone marrow, wherein normal leukocyte production is corrupted, HSC homeostasis is altered resulting in the potential for a lower frequency of HSCs than in healthy bone marrow.

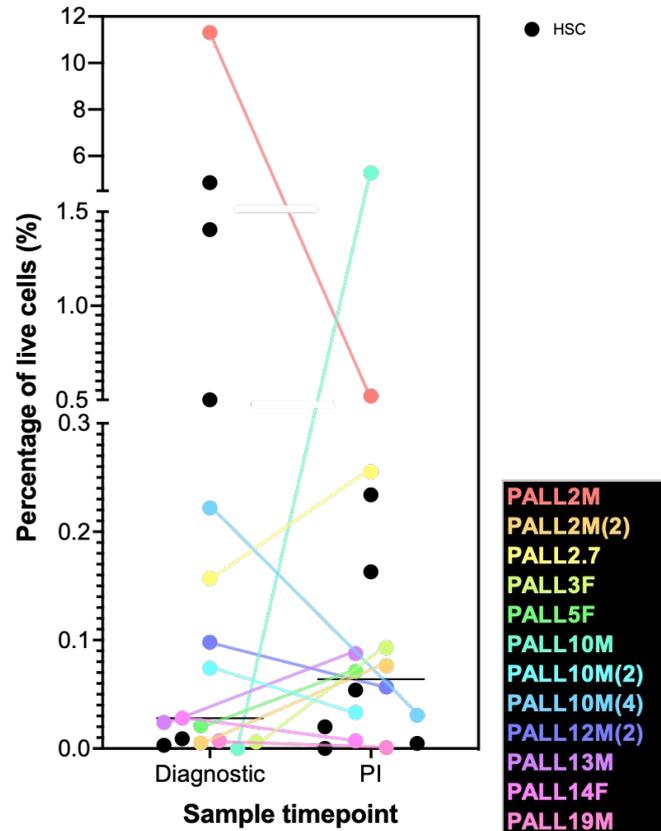
With a goal of this chapter being to collect HSCs for analysis of BCR::ABL1 fusion by FISH, the prospect of having too few cells to perform this analysis was a realistic concern.

HSCs and MPPs were sorted according to the sorting strategy in figure 4.1.2, 4.2.1 & 4.2.2 from bulk BMNC samples from Ph+ALL patients at diagnosis and post-induction therapy. As anticipated, the population size of HSCs was small in both diagnostic and PI samples (Fig 4.2.6). Paired diagnostic and PI samples were indicated by colour and connecting lines, any sample which could not be matched (either PI was not available or sequential samples were sourced from the Glasgow 8y/o male patient, previously discussed, where diagnostic sample was not available). Unmatched samples were included to display the overall variation in haematopoietic population size across the Ph+ALL cohort used in this thesis. All FACS data was reported where available, however, due to variations in starting cell number and poor recovery rate post cryopreservation, a small number of samples were too poor to identify haematopoietic progenitor populations as outlined in the sorting strategy discussed previously (n= 13 (6 diagnostic & 7PI)).

As discussed previously, the frequency of HSCs in homeostatic bone marrow is low, however with the addition of a dysregulation of the bone marrow toward the production of blast cells in Ph+ALL, this frequency is likely to be even lower. As a result, in diagnostic bone marrow the average population size of HSCs was 0.995% of total live cells and in PI samples 0.54% (Fig 4.2.6). Through PI treatment, blast cells are eradicated and hence, a smaller average

population size of HSCs than in diagnostic samples is not due to skewing of haematopoiesis to the production of lymphoblasts but is instead as a result of the treatment itself (ie steroids).

(1) HSC populations (Ph+ALL diagnostic and PI)



(2)

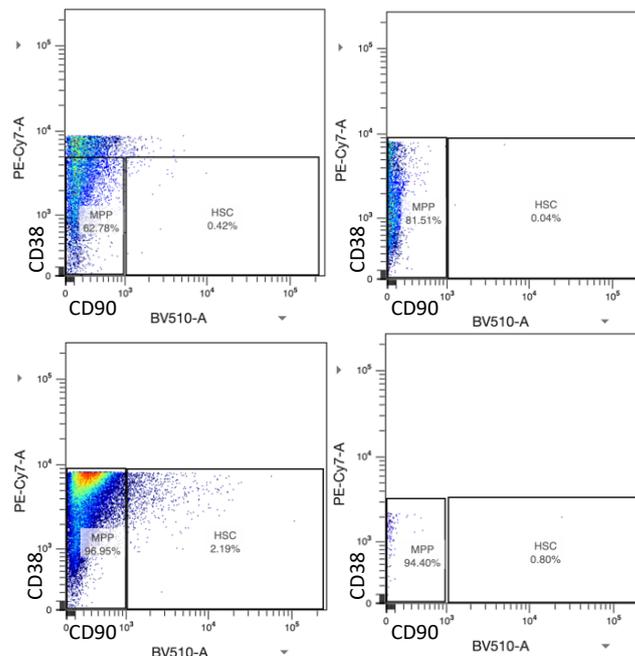


Figure 4.2.6: HSC population size in matched Ph+ALL diagnostic and PI samples . (1) Cell count normalised (expressed as a percentage of total number of cells counted during the sort) to allow for variably sized samples to be displayed on the same graph. Black dots represent samples with no matched samples available (no PI provided or post-treatment only). **(2)** Representative FACS plots displaying HSC population size in total sample (CD90/CD38). Patient samples displayed; (clockwise from top left) PALL10M, PALL19M, PALL3F & PALL2.7.

Examining specific patient samples in Fig 4.2.6 displays the variability in HSC sample size in the Ph+ALL patient cohort. For example, PALL19M (pink) displays few to no HSCs in either diagnostic or post induction sample. This may be due to issues with the cryopreservation and culture process itself which lead to the loss of HSCs or may display a naturally low number of early stem cells in this patient.

Additionally, there were patients which had very small HSC populations at diagnosis which expanded after PI. An example of this being patient PALL10M (light blue) who had 0.074% HSCs at diagnosis but 5.272% after PI. Five of the 12 patients in Fig 4.2.6 displayed an increase in HSC population size post-induction, however with 70 times increase in population size, PALL10M displays the largest expansion during treatment.

Another interesting sample is PALL2M (red) which displays the largest decrease (from 11-0.031%) in HSC population size during treatment (Fig 4.2.6). Though these numbers are small, proportionally this decrease in stem cell number is interesting and may display either a substantial response to PI therapy at day 98 post treatment (date of sample collection) or the potential for a large active population of HSCs existing at diagnosis.

A number of samples displayed little change to HSC population size over the course of treatment. An example of this being PALL2.7 (yellow) (sex of patient unavailable) which maintained a large HSC population and PALL10M(2) (turquoise) which maintained a small HSC population during treatment.

Another sample which maintained a large population of HSCs during treatment, but which was highlighted as an outlier was PALL2M (red). At diagnosis, PALL2M displayed that 11.309% of live cells captured were HSC (Fig 4.2.6). This is significantly larger than the rest of the cohort and skewed the calculation of the average population size at diagnosis. Not only is this large population unexpected for a patient at Ph+ALL diagnosis where blast cells would be expected to predominate in the sample, but also that this HSC population would be unexpectedly high for a healthy patient sample. This data suggests that there was an expansion of cells with HSC surface markers in PALL2M at diagnosis which was maintained during lymphoid-directed treatment (day 119 when sample was harvested). Without investigation of HSC activity by

functional stem cell assays, a definite cause of this unusual population size cannot be made however, it is suggestive of a different disease physiology to the rest of the Ph+ALL cohort. A possible explanation is that these are leukaemic cells which express HSC markers and are BCR::ABL1 positive by FISH, or true Ph+ HSCs, therefore this patient may have the CML-like Ph+ALL subtype.

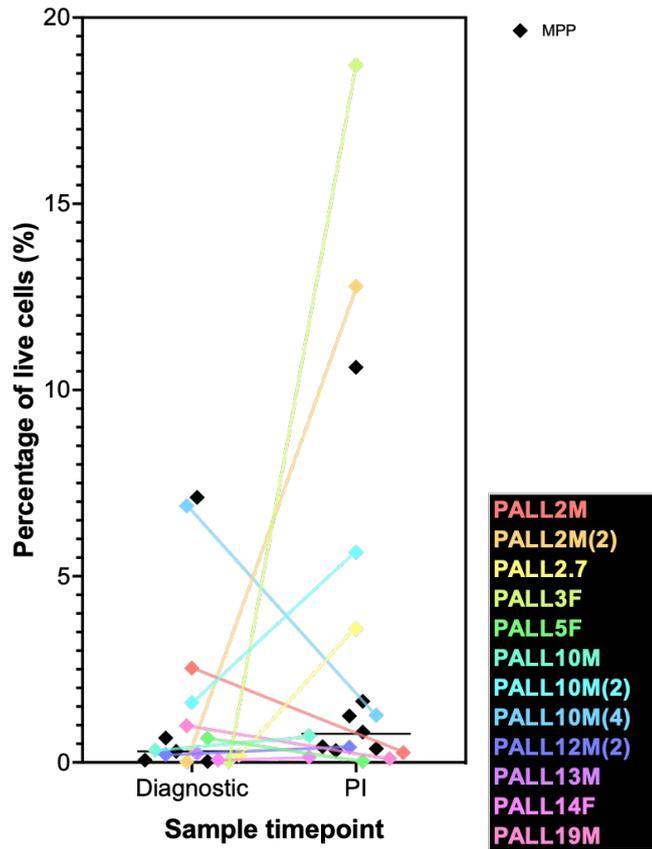
MPPs represent a non-homogenous population of early progenitor cells composed of subsets which give rise to more differentiated progenitors. Despite being early haematopoietic progenitor cells, MPPs are functionally distinct from HSCs with different cell surface marker expression patterns and intracellular signalling dynamics (Wang et al., 2021). HSCs produce a series of increasingly abundant lineage-committed cells, starting with MPPs. Observation of predetermination of cell fate decisions in subgroups of MPPs has led to the definition of lineage biased MPPs (Pietras et al., 2015). Such MPPs are able to work in tandem to adapt blood production to requirements such as homeostasis or tissue regeneration. Cell tracing has enabled detection of lineage-primed MPP subsets; MPP2 and MPP3 which are myeloid-biased, MPP4 which are lymphoid-primed and MPP1 which behave as a more metabolically active subset of HSCs (Fig 4.1.1). MPP subsets have not been further explored in this chapter as total MPP populations were collected based on cell surface marker expression. However, it is important to acknowledge that the MPPs collected by FACS are heterogeneous in lineage-biases.

As MPPs and HSCs are at the apex of the haematopoietic hierarchy and exist as non-committed progenitor cells, both populations have similar cell surface marker expression profiles (Fig 4.1.2 & 4.2.1). The marker therefore used in this chapter to distinguish such cell types and collect pure populations of HSCs and MPPs was CD90. CD90 is expressed on HSCs and has been used clinically for the accurate identification of HSCs for use in applications such as allo-HSCT. Hence, the expression profile Lin-CD34+CD38-CD90- was used to isolate pure populations of MPPs for further analysis of BCR::ABL1 fusion during early haematopoiesis. MPP population size was expectedly small in both diagnostic and PI samples. While most patients maintained a similar MPP population size during treatment, incongruent to the rest of the cohort, a number of PI samples displayed a large increase in MPP population size (such as PALL3F, PALL2M(2), PALL2.7 and PALL10M) (Fig 4.2.7). However, as with the HSC samples,

a number of patients had significant changes in MPP population size during treatment. An example of this being PALL10M(4) (mid blue) which had an atypically large number of MPPs at diagnosis (6.889% of live cells captured) which decreased to 1.226% during treatment. This near 6-fold decrease in population size mirrors that seen in the HSC sample for this patient where the number of HSCs at diagnosis were 7 times larger than at PI (Fig 4.2.6). While many samples (6/12) maintained similar MPP numbers during treatment, samples such as PALL10M(4) (mid blue) and PALL2M (red) exhibited a large decrease in population size (6-fold and 10-fold respectively). Interestingly, these were samples which exhibited HSC population sizes atypical to the rest of the Ph+ALL cohort (Fig 4.2.6).

In addition to samples which exhibited a large decrease in MPP sample size during treatment, a number of samples displayed the converse, a large increase in MPP cells (Fig 4.2.7). Samples such as PALL3F (lime green) and PALL2M(2) (orange) had some of the smallest numbers of MPPs at diagnosis (0.027 and 0.014% of live cells captured respectively) but had an observed expansion in cell number at PI (18.72 and 12.79% respectively). This expansion highlights samples going from the lowest number of MPPs at diagnosis to the highest at PI. Other samples also had increases in MPP population size (PALL10M(2) (light blue) and PALL2.7 (yellow)), however with fold changes of 693 and 913 respectively, PALL3F and PALL2M(2) displayed the largest overall change in the cohort.

(1) MPP populations (Ph+ALL diagnostic and PI)



(2)

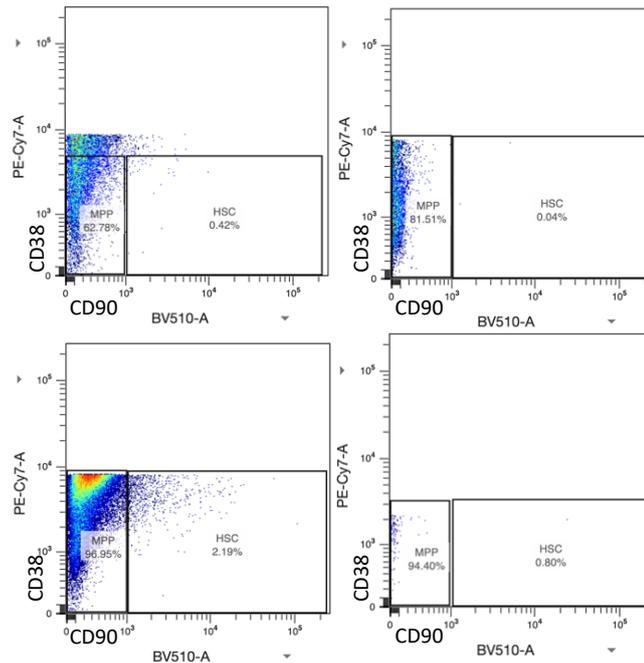


Figure 4.2.7: MPP population size in matched Ph+ALL diagnostic and PI samples . (1) Cell count normalised (expressed as a percentage of total number of cells counted during the sort) to allow for variably sized samples to be displayed on the same graph. Black dots represent samples with no matched samples available (no PI provided or post-treatment only). **(2)** Representative FACS plots displaying MPP population size in total sample (CD90/CD38). Patient samples displayed; (clockwise from top left) PALL10M, PALL19M, PALL3F & PALL2.7.

4.2.3 Results II: Two Ph+ALL patients had unexpectedly large CMP populations at diagnosis

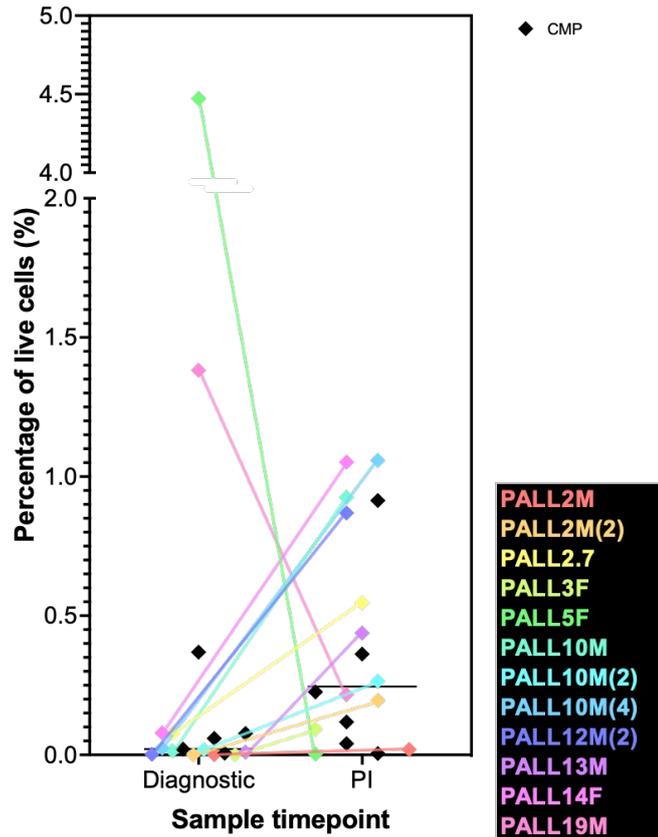
Common myeloid progenitors (CMP) result from HSC progressive commitment and are able to differentiate into either MEPs (megakaryocyte-erythrocyte progenitors) or granulocyte-monocyte progenitor (GMP) cells. CMPs represent early phase myeloid commitment during haematopoiesis (Fig 4.1.2). Monocyte commitment is a highly regulated process with transcription factors such as c-myb being an essential mediator for granulocyte/monocyte lineage decision making (Lieu and Reddy, 2012). While also functioning as progenitor cells, CMPs may play a direct role in immunosurveillance and the population kinetics of GMPs, deriving from CMPs, have been found to vary within oncogenic contexts (Pu et al., 2016). It is therefore of interest to investigate the population dynamics of myeloid progenitors in a Ph+ALL context to elucidate if the CML-like subtype displays different myeloid cell behaviour detectable at the population level by flow cytometry.

With Ph+ALL being a lymphoproliferative disease characterised by large numbers of lymphoblast cells, the size of myeloid populations at diagnosis is expected to be small. However, two samples exhibited a larger CMP population at diagnosis compared to the rest of the cohort. This was unexpected in view of the pathogenesis of classical Ph+ALL with a lymphoblast proliferation phenotype, therefore these samples may reflect CML-like Ph+ALL. As the proportion of CML-like patients in the Ph+ALL cohort was predicted to be 30% in the 2017 Hovorkova (et al) study, the small number of patients with larger than expected CMP populations at diagnosis align with these predictions.

PALL5F (green) and PALL19M (pink) can be observed as having atypical CMP population dynamics, with 4.47 and 1.382% of live cells captured being identified as CMPs respectively (Fig 4.2.8). PALL5F and PALL19M have CMP populations well above average for the rest of the cohort (average 0.023%). Not only is this atypical for the cohort studied in this chapter but also unexpected for diagnostic samples of a standard lymphoblastic leukaemia. These outlying samples suggest a different phenotype to the rest of the Ph+ALL cohort, with an unusual myeloid development programme (Fig 4.2.8). These samples may represent those discussed in Hovorkova et al, where samples designated as CML-like Ph+ALL displayed a

higher level of myeloid involvement at diagnosis, differing from the standard Ph+ALL phenotype.

(1) CMP populations (Ph+ALL diagnostic and PI)



(2)

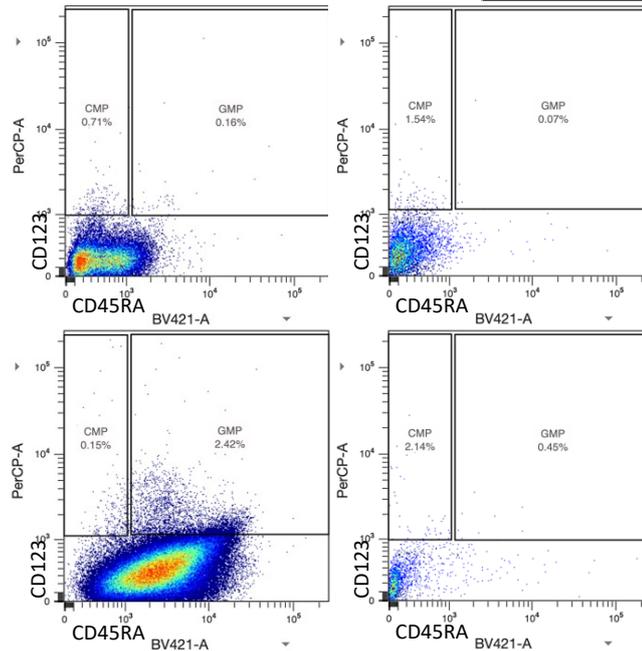


Figure 4.2.8: CMP population size in matched Ph+ALL diagnostic and PI samples . (1) Cell count normalised (expressed as a percentage of total number of cells counted during the sort) to allow for variably sized samples to be displayed on the same graph. Black dots represent samples with no matched samples available (no PI provided or post-treatment only). **(2)** Representative FACS plots displaying CMP population size in total sample (CD45RA/CD123). Patient samples displayed; (clockwise from top left) PALL10M, PALL19M, PALL3F & PALL2.7.

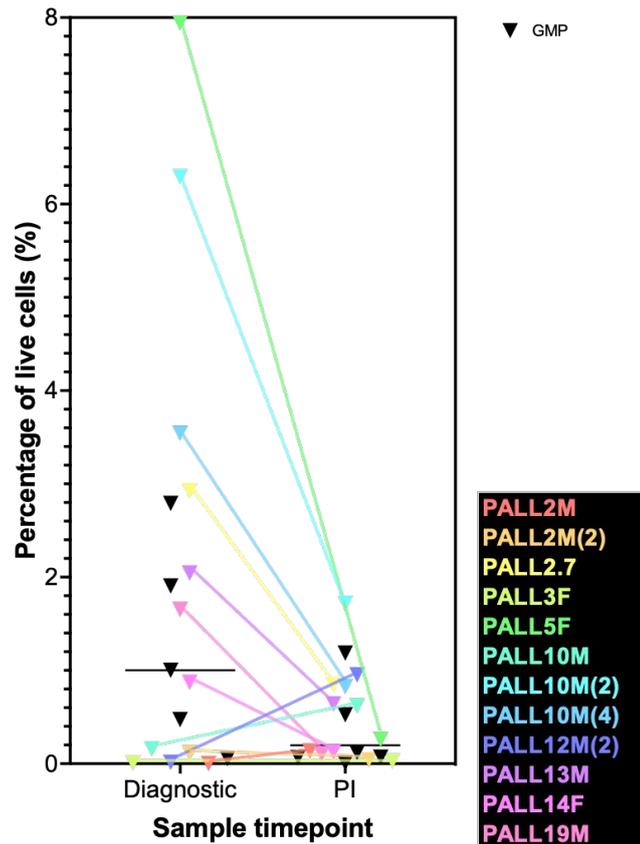
4.2.4 Results II: Two patients had unexpectedly large GMP populations at diagnosis

Similarly, Granulocyte-monocyte progenitor (GMP) cells give rise to unipotent precursor cells from which derive terminally differentiated monocytes or granulocytes (Lieu and Reddy, 2012). As previously discussed, this is a highly regulated process dependant on the activity of transcription factors (TFs) such as c-Myb, the dysregulation of which, is associated with myeloid leukaemias such as AML and CML.

GMPs are the precursor to mast cells, granulocytes and monocytes (including macrophages and monocyte-derived dendritic cells), essential components for the adequate function of the innate immune response (Zebisch et al., 2019). Additionally, GMPs are capable of inhibiting polyclonal stimuli- and alloantigen-induced T cell proliferation, displaying their importance in the modulation of various aspects of the immune response (Pu et al, 2016). Though not terminally differentiated, GMPs being at a later stage of lineage commitment than cell types discussed in previous sections of this chapter enables them to play an active role in the modulation of immune responses and therefore, can be distinguished from CMPs by the presence of the terminal differentiation marker CD45RA (Fig 4.1.2 & 4.2.1).

As with CMP cells, the number of GMPs present at diagnosis in Ph+ALL is expected to be low due to the haematopoiesis dynamic skewing toward lymphoid proliferation. It could be observed that 5 of the 12 samples had little to no GMP cells detectable in their diagnostic samples, with the other samples displaying a wide range of population sizes up to 7.94% of live cells captured being GMPs (Fig 4.2.9). Of the samples which had GMP populations over 1% of live cells captured, all have previously been highlighted as having outlying population dynamics in previously discussed cell types (HSC, MPP and CMP). The sample observed to have the largest GMP population of the cohort was PALL5F (green) with 7.94% of live cells captured expressing GMP surface markers. This observation is in line with results from the other myeloid population investigated where PALL5F also had the highest number of CMP cells (Fig 4.2.8). This supports the theory that some samples in the cohort represent Ph+ALL with a myeloid-skewed phenotype, similar to that first observed by Hovorkova et al and which had been designated as CML-like Ph+ALL (Fig 4.2.8 and 4.2.9).

(1) GMP populations (Ph+ALL diagnostic and PI)



(2)

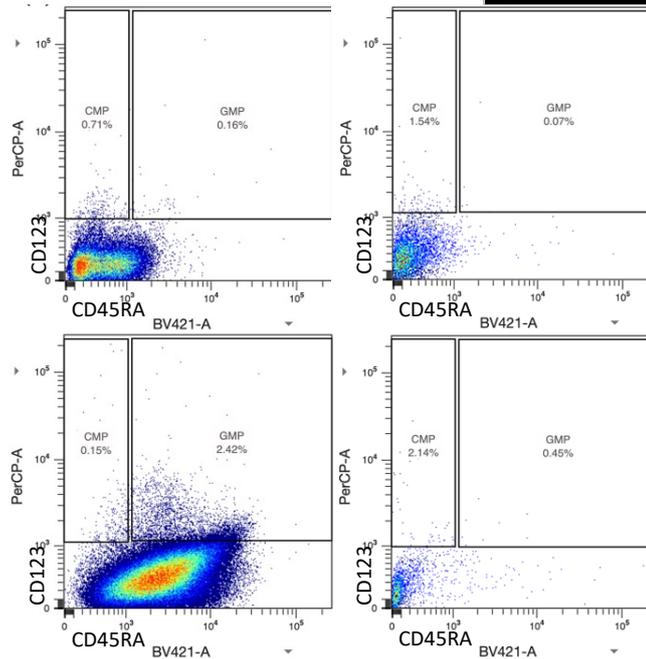


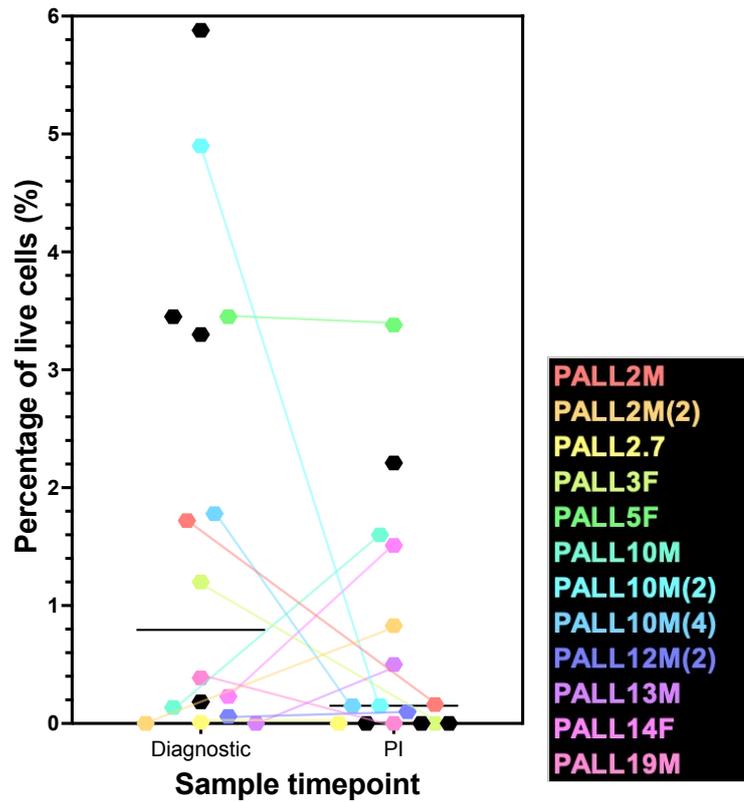
Figure 4.2.9: GMP population size in matched Ph+ALL diagnostic and PI samples . (1) Cell count normalised (expressed as a percentage of total number of cells counted during the sort) to allow for variably sized samples to be displayed on the same graph. Black dots represent samples with no matched samples available (no PI provided or post-treatment only). **(2)** Representative FACS plots displaying GMP population size in total sample (CD45RA/CD123). Patient samples displayed; (clockwise from top left) PALL10M, PALL19M, PALL3F & PALL2.7.

4.2.5 Results II: Three patients had atypically small CLP populations at diagnosis

Common lymphoid progenitors (CLPs) represent the earliest phase of lymphoid commitment during haematopoietic development. Older models of haematopoiesis classify CLPs as being at the initial lineage bifurcation step where MPPs differentiate and specialise either into a CMP or CLP (Cheng et al., 2019). Though models of haematopoiesis are constantly being revised and enhanced, this demonstrates how early in lymphoid lineage commitment CLPs reside. CLPs are able to give rise to T cells, B cells, dendritic cells (DCs) and natural killer (NK) cells, a developmental process carefully controlled by TFs and cytokines such as IL-7, SCF and TPO (Karsunky et al., 2008).

Identification and isolation of CLPs used cell markers as previously discussed, namely Lin-CD34+CD38+CD45RA+CD10+CD19-CD33- (Fig 4.1.2 & 4.2.1). As previously discussed, CLPs represent the earliest step in lymphoid lineage commitment, as treatment is aimed at eradication of lymphoid blasts, therefore, population size of lymphoid progenitors in PI samples is expected to be small. 5 of the 12 matched samples displayed a decrease in CLP number over treatment and 3 samples maintained similar CLP counts. As the majority (2/3) of samples had a small starting population size, the small PI sample size may still reflect a standard phenotype. Interestingly, 4 samples displayed an increase in CLP population size over the course of lymphoid directed therapy, a feature which may reflect immune reconstitution.

(1) CLP populations (Ph+ALL diagnostic and PI)



(2)

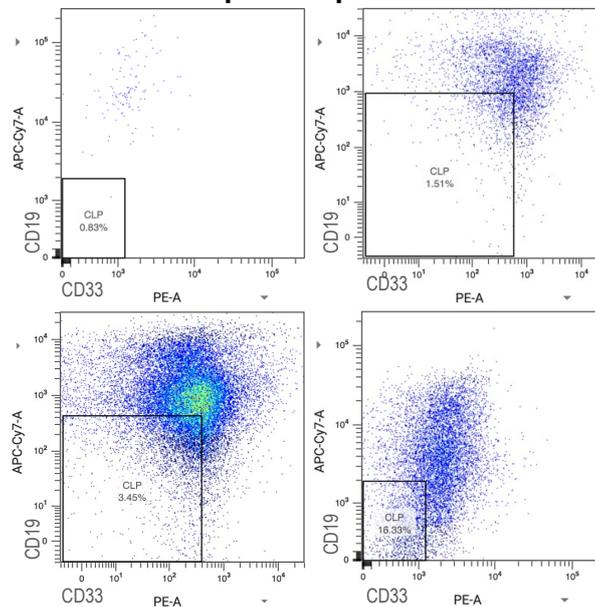


Figure 4.2.10: CLP population size in matched Ph+ALL diagnostic and PI samples . (1) Cell count normalised (expressed as a percentage of total number of cells counted during the sort) to allow for variably sized samples to be displayed on the same graph. Black dots represent samples with no matched samples available (no PI provided or post-treatment only). **(2)** Representative FACS plots displaying CLP population size in total sample (CD19/CD33). Patient samples displayed; (clockwise from top left) PALL19M, PALL8M(2), PALL9F & PALL12F.

4.2.6 Results II: Atypical haematopoietic population sizes were observed in a number of diagnostic and PI Ph+ALL samples

As discussed throughout section 4.2, a number of patients were observed to have haematopoietic progenitor population sizes atypical to the rest of the Ph+ALL cohort. The size of these populations were atypical to the standard Ph+ALL phenotype where the bone marrow is skewed toward lymphopoiesis, resulting in small populations of early haematopoietic and myeloid progenitors, and large numbers of lymphoid progenitors. Such phenotypic differences aligned with observations in Hovorkova et al (2017), where an enlarged myeloid progenitor pool was present in CML-like samples.

Matched diagnostic and PI sample data is displayed in Table 4.2.1 analysed as previously outlined in section 4.2.1. Haematopoietic progenitor population sizes are expressed as a percentage of live cells recorded for each sort, normalising the data and allowing for differences in the proportion of cell groups to be observed. 11 of the 17 diagnostic samples had small populations of HSCs and MPPs, a trait expected to be observable in standard Ph+ALL diagnostic samples where haematopoiesis is skewed toward lymphoid development. 4 of the 17 diagnostic samples had much larger HSC/MPP populations, suggesting that these samples do not represent the standard Ph+ALL phenotype. This may be due to a higher level of stem cell activity in such samples or may display an enhanced robustness of HSC/MPPs, allowing the cells to better survive the cryopreservation, culture and sorting process. Methods to assess stem cell activity will be discussed further at the end of this chapter. 5 of the 12 matched diag/PI samples displayed an increase in HSC proportions over treatment, suggesting a restoration of normal haematopoiesis due to the therapy received.

In addition to the increased HSC/MPP population size, an enlarged pool of myeloid progenitors was observed in 2 of the 17 diagnostic samples. Typical Ph+ALL presentation would include small CMP populations due to the aforementioned drive toward lymphopoiesis. The observation of an increased myeloid progenitor pool in these patients is in line with data from Hovorkova et al (2017) where larger myeloid populations were observed in the samples described as being 'CML-like'. GMP populations had a greater range than CMP (GMP range= 7.94, CMP range= 4.471) however one patient had very large

proportions of both CMP and GMP cells (PALL5F) which may represent an enhanced myeloid activity and possible ‘CML-like’ status. 8 of the 12 matched samples had an observable increase in CMPs over treatment, likely representing the establishment of normal haematopoiesis.

Due to the enhanced lymphoid proliferation characteristic of ALL, CLPs are expected to be present in large numbers in the diagnostic samples however, 8 samples displayed CLP populations of <1% of live cells measured. As all the patients in this cohort were diagnosed and treated as having Ph+ALL, these observations are likely not representative of low lymphoid activity. 2 of the 9 patients displayed an increase in population size >1% after treatment, suggesting a development normal lymphoid function or a move toward re-establishing normal haematopoiesis.

Sample ID	Timepoint	Percentage sorted progenitor population of live cells (%)				
		HSC	MPP	CMP	GMP	CLP
PALL2M	Diag	11.309	2.532	0	0.001	1.72
	PI	0.52	0.321	0.005	0.136	0.16
PALL2M(2)	Diag	0.006	0.027	0.002	0.117	0
	PI	0.076	12.786	0.019	0.104	0.83
PALL2.7	Diag	0.21	0.026	0.077	2.926	0
	PI	0.255	3.588	0.547	0.846	0
PALL3F	Diag	0.005	0.014	0.001	0.008	1.2
	PI	0.093	18.722	0.093	0.07	0
PALL5F	Diag	0.024	0.65	4.472	7.941	3.45
	PI	0.071	0.033	0.003	0.259	3.38
PALL10M	Diag	0	0.299	0.014	0.157	0.136
	PI	5.272	0.719	0.926	0.619	1.6
PALL10M(2)	Diag	0.098	1.606	0.018	6.293	4.9
	PI	0.033	5.644	0.265	1.72	0.15
PALL10M(4)	Diag	0.222	6.889	0.024	3.543	1.78
	PI	0.031	1.266	1.058	0.828	0.15
PALL12M(2)	Diag	0.157	0.192	0.01	0.015	0.06
	PI	0.057	0.408	0.869	0.952	0.1
PALL13M	Diag	0.028	0.213	0.006	2.044	0.01
	PI	0.088	0.263	0.438	0.642	0.5
PALL14F	Diag	0.074	0.062	0.079	0.872	0.23
	PI	0.007	0.142	1.052	0.141	1.51
PALL19M	Diag	0.007	0.987	1.382	1.654	0.387
	PI	0.001	0.1	0.216	0.126	0

Table 4.2.1: Sorted progenitor population size (percentage of live cells) in diagnostic and PI samples. Haematopoietic progenitor populations were sorted by FACS as previously described) and expressed as a percentage of live cells measured for each sort.

4.3.1 Results II: BCR::ABL1 fusion was detectable by FISH in a subgroup of Ph+ALL stem and progenitor cells.

The proportion of Ph+ALL patients with a mixed lineage phenotype is unknown, as is the impact of BCR::ABL1 presence in early stem and progenitor cells on disease pathology. In order to investigate this, HSC and MPP cell populations were isolated by FACS from diagnostic and PI Ph+ALL samples and BCR::ABL1 expression assessed by FISH (Fig 4.3.1). This allows for an accurate identification of BCR::ABL1 fusion during the early stages of haematopoiesis in Ph+ALL.

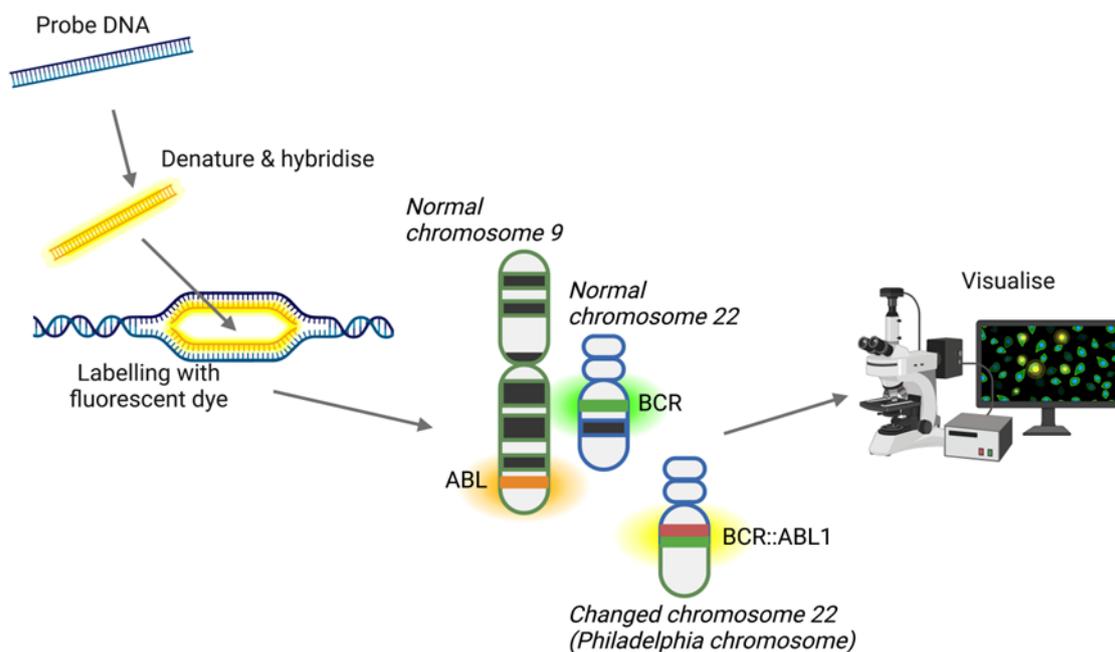


Figure 4.3.1: Representation of BCR::ABL1 detection by FISH (Fluorescence *in situ* hybridisation). FISH allows for the visualisation of BCR::ABL1 by fluorescence microscopy. Use of the dual colour dual fusion BCR::ABL1 probe (Abbott) allows for the identification of ABL on normal chromosome 9 (orange), BCR on normal chromosome 22 (green) and the fusion gene BCR::ABL1 on the Philadelphia chromosome (yellow). (Created using BioRender)

In order to verify the cell of BCR::ABL1 fusion in Ph+ALL, pure populations of haematopoietic progenitors were isolated by FACS as previously described. Applying the haematopoietic development model, where HSCs are at the apex of the hierarchy and lineage commitment increases during development, the earliest progenitors were prioritised for examination. This resulted in prioritisation of HSCs and MPPs for FISH scoring. The term 'scoring' when applied to FISH refers to the process of designating cells as Ph+ (detectable BCR::ABL1) or Ph- (BCR::ABL1 not detected) (Chase et al., 1997). This relies on the visual identification of the

fusion gene, a yellow fluorescence signal resulting from the spatial proximity of the orange ABL signal and the green BCR signal (Fig 4.3.2.1). In a healthy (Ph⁻ cell), separate orange and green signals should be observable (Fig 4.3.2.2). The Abbott probes used in this chapter allowed detection of BCR and ABL genes by detection of a 671kb length of DNA on chromosome 9 (from a point centromeric of the arginosuccinate synthase gene (ASS1) to telomeric of the of the ABL1 gene) and a 1.5Mb section of DNA on chromosome 22 (beginning with the variable segments of the IGL locus and ending 900kb telomeric to the BCR gene) (Abbott, 2023). These probes are designated SpectrumOrange ABL and SpectrumGreen BCR. Both probes are able to span the t(9;22) chromosomal breakpoints (p190, p210, and p230) for both genes and allow for the visualisation of BCR::ABL1.

K562 CML Ph⁺ cell lines were used to validate the Abbott FISH probes used (Fig 4.3.2.1). It could be observed that the nuclei size of cell lines and sorted primary patient haematopoietic cells varied, with primary cell nuclei being smaller (approx. 10 microns) (Fig 4.3.2.2). Utilising the current understanding of BCR::ABL1 presence during haematopoietic development in CML, it was theorised that BCR::ABL1 might be detectable in HSCs in CML-like Ph⁺ALL (Houshmand et al., 2019) (Fig 4.3.3). Hence, HSCs, MPPs and Lin-CD34⁺ cells were prioritised for FISH scoring in all samples with BCR::ABL1 presence in later, more committed progenitors being verified in a smaller number of patients. This was to investigate if BCR::ABL1 detected by FISH in early progenitors is maintained throughout haematopoietic development and increasing lineage commitment.

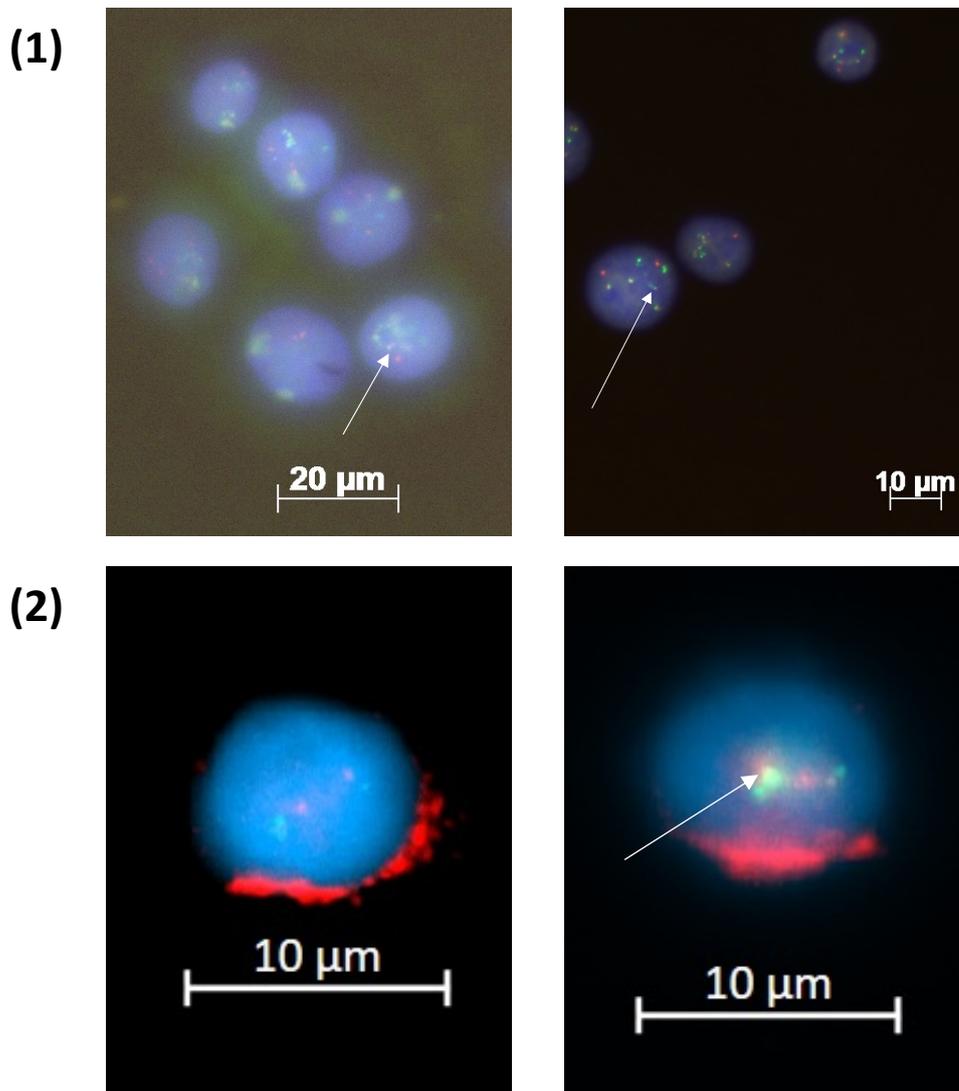


Figure 4.3.2: Representative FISH nuclei for Ph+ leukaemias (1) Nuclei of K562 CML cell line, stained for FISH. BCR-ABL indicated by white arrows. **(2)** HSCs isolated from a 5 year old male Ph+ALL patient at diagnosis, White arrow indicates BCR::ABL1, absence of white arrow to indicate no fusion event detected.

Cells were fixed and lysed as per Materials and Methods Section 2.8. Following incubation with Abbott FISH probes and washing, cells were visualised using a Zeiss Axio Imager M1 fluorescence microscope. In order to calculate the percentage of Ph+ cells, 100 cells were counted (or the maximum number of cells available) and BCR::ABL1 positive cells were calculated as a percentage of the total number of cells. Of the 34 samples analysed (diagnostic and PI), 14 failed to be reported (Fig 4.3.3). Failure of FISH analyses was due to small number of cells collected or nuclear lysis. A proposed reason for this is that samples which failed may have originated from standard Ph+ALL samples without large HSC populations. Without functional stem cell assays, it could not be determined with certainty if the HSCs isolated from the proposed CML-like samples had an increased survival advantage over the standard

Ph+ALL HSCs in terms of survivability during the harvesting, cryopreservation, culture and sorting process. As an example, sample PALL2M had the largest HSC population of the cohort (11.309% of live cells were detected as HSCs) and was able to be scored by FISH. This could mean that the cohort of samples able to be analysed using FISH may unintentionally be enriched for samples with a large or robust stem cell population. If features of CML-like Ph+ALL are an enhanced HSC pool and BCR::ABL1 presence in stem cells, the FISH data presented may over-represent samples with Ph+ stem cells. Of the total samples analysed 11/12 samples had detectable BCR::ABL1 within the HSC population and 17/23 within the MPP population.

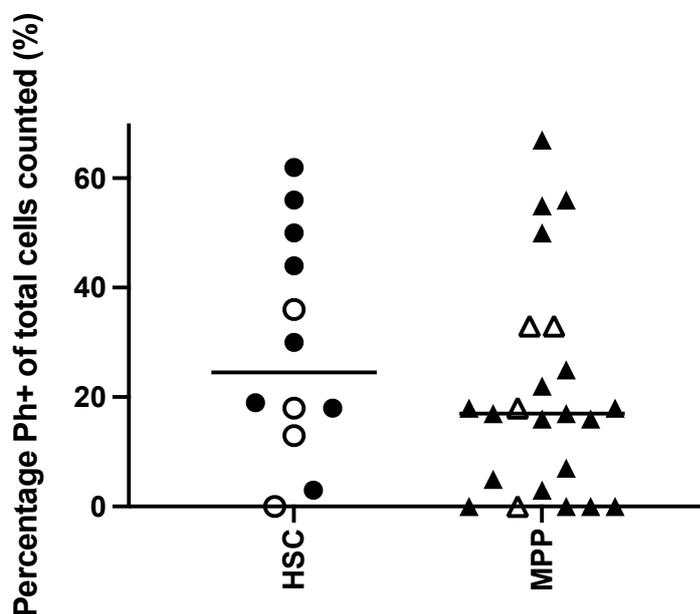


Figure 4.3.3: BCR::ABL1 positivity in haematopoietic populations (FISH), percentage of Ph+ progenitor cells. Proportion of Ph+ cells calculated by counting total number of cells in each sample, counting the number of BCR::ABL1+ cells, and expressing Ph+ cells as a percentage of total cells. (closed circle= diagnostic, open circle= PI).

As CML-like Ph+ALL is currently undefined, the presence of BCR::ABL1 at stem cell level alone is inadequate for the designation to this subtype. In order to determine if Ph+ HSC/MPPs are maintained during Ph+ALL treatment without LSC targeting, similar to CML, we assessed BCR::ABL1 in Ph+ALL diagnostic and PI samples by FISH (Fig 4.3.3). These results display the heterogeneity within the Ph+ALL cohort and additionally, the maintenance of Ph+ HSC/MPPs during treatment (Fig 4.3.3). A known mechanism of relapse in CML is the persistence of Ph+ HSC/LSCs (Mojtahedi et al., 2021). This persistence of Ph+ stem and progenitor cells post-

induction treatment could confer treatment resistance and, as a result, put such patients at risk of relapse. Further research investigating stem cell functionality would be required to determine if such Ph⁺ HSCs are, in fact, LSCs. However, the maintenance of BCR::ABL1⁺ stem and progenitor cells during and after PI treatment describes a potential mechanism for relapse and posits the requirement for adaption of therapeutic approaches to one similar to that used successfully in CML.

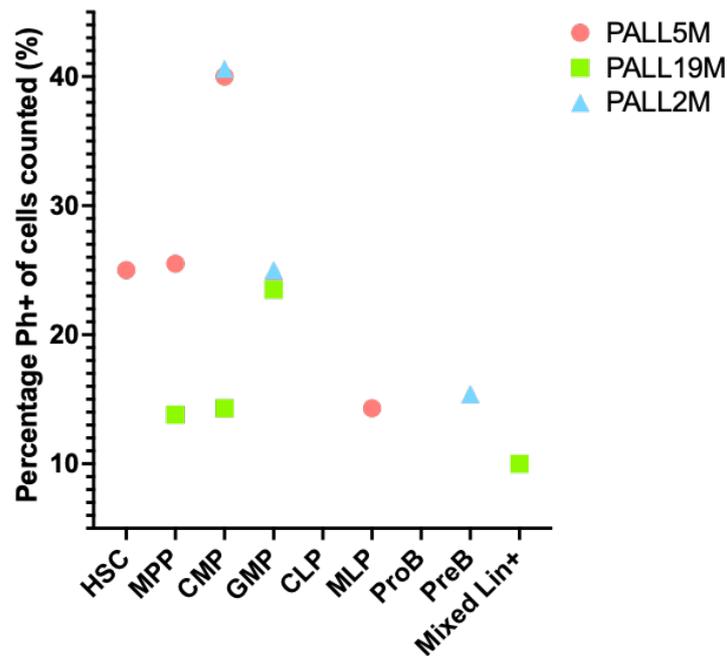


Figure 4.3.4: BCR::ABL1 presence in early and late haematopoietic populations. Percentage of Ph⁺ cells of total cells counted, three samples presented (PALL5M, PALL19M & PALL2M). Cell populations include; HSC, MPP, CMP, GMP, CLP, MLP, ProB, PreB and mixed Lin⁺ cells.

To verify that BCR::ABL1 was discoverable in cells at later haematopoietic differentiation stages than HSCs or MPPs, FISH was performed on myeloid, lymphoid and mixed Lin⁺ progenitors sorted from 3 samples (Fig 4.3.4). FISH was carried out as outlined above. Of the scorable populations, BCR::ABL1 was detectable in all progenitors. This supports the concerns that, like in CML, t(9;22) fusion at the apex of the haematopoietic hierarchy can result in BCR::ABL1 activity in multiple lineages, resulting in a mixed-lineage phenotype (as seen in CML-LBP, mixed- lineage CML-BP and CML-like Ph⁺ALL).

4.4.1 Results II: Patients with atypical haematopoiesis dynamics and BCR::ABL1 fusion in stem and progenitor cells had poor clinical outcomes.

We next used clinical information to ascertain prognostic significance of BCR::ABL1 within stem and progenitor populations at both diagnosis and PI detectable by FISH. Clinical information was provided by CellBank and CCLG (Children’s Cancer and Leukaemia Group). This provided an insight into clinical features at initial diagnosis (i.e. white blood cell count (WCC), blast percentage and immunophenotype), during treatment (i.e. end of induction Ig/TCR minimal residual disease (MRD) level and MRD group), date of PI sample harvest and additional important clinical features such as incidence of relapse/disease progression/death, transplant status/type, site of relapse and overall/event free survival (Tables 4.4.1- 4.4.5).

	Risk status			
	MRD	Cyto abnormalities	WCC	Relapse/death
PALL2M		X		
PALL2M(2)				
PALL2.7		X		
PALL3F			X	
PALL4M		X		
PALL5M			X	
PALL5F	X		X	X
PALL5F(2)		X		X
PALL7F		X		
PALL8M	X			X
PALL9F				
PALL10M			X	X
PALL10M(2)			X	X
PALL10M(4)		X		
PALL12F		X		
PALL12M		X		X
PALL12M(2)				
PALL13M				
PALL14F				
PALL15M				
PALL16M				
PALL19M				X

X Poor risk
 Good risk
 No data available

Table 4.4.1: Patient risk status based on clinical information provided by CellBank. High/poor risk indicated by ‘X’= orange, good risk= green and no data available= grey. High risk MRD was defined as having a positive EOI MRD score >0.01%. High risk cytogenetic abnormalities were defined as gene rearrangement, deletion, gain of chromosome(s), loss of chromosome(s) and ‘other abnormalities’ in addition to t(9;22). High risk white cell counts (WCC) at diagnosis were defined as >117.06 (the average WCC for the CellBank cohort). Occurrences where no data was available are indicated. High risk relapse/death was defined by the reported occurrence of relapse or death.

All clinical data received from CellBank and CCLG has been reported in tables 4.4.2- 4.4.5, however information was unavailable for a number of patients, hence the number of patients varies between tables. BCR::ABL1 transcript levels are not currently used for MRD monitoring within paediatric UK protocols and therefore this information was not available

for this comparison. The information provided allowed ‘high-risk’ patients to be identified and categorised (Table 4.4.1). ‘High-risk’ was defined using clinical information provided by CellBank, these were comprised of the presence of cytogenetic abnormalities, positive MRD score (EOI MRD>0.01%), higher WCC at diagnosis than average for the cohort and reported incidence of relapse or death. Samples displayed in tables 4.4.2- 4.4.5 represent those with available clinical data from CellBank and FISH data generated during this project. Using these metrics, 3 patients were described as high risk by MRD, 10 by cytogenetic abnormalities, 8 by high white cell counts at diagnosis and 8 by the reported occurrence of relapse or death. As this patient cohort is populated by paediatric Ph+ALL patients who were recruited from hospital sites across the UK to either ALL2003, EsPhALL, or UKALL2011 clinical trials between 2005- 2012, some clinical data was unavailable to CellBank and was therefore noted on the patient information tables.

ID	Sex	Age	EOI MRD level	MRD Group	MRD markers
PALL12M	M	12	0.00016	0.01-0.1%	Vd2-Ja29
PALL8M	M	8	0.338	>5%	VH1
PALL13M	M	13	0.0031	0.1-1.0%	VH1 VH3
PALL2M	M	2	NA	NA	Ineligible at diagnosis
PALL5F	F	5	0.011	1-5%	VH1 VH4
PALL10M	M	10	0.0011	0.1-1.0%	VH4 DD2-DD3
PALL5M	M	5	0.00021	0.01-0.1%	No results
PALL3F	F	3	0	0%	VH1 Vd2-Ja29
PALL2M(2)	M	2	9.60E-06	0-0.005%	VH3 DH4
PALL5F(2)	F	5	0.000111	0.01-0.1%	VH2 VgII-Jg1.3
PALL15M	M	15	0.000012	0-0.005%	VH3 VKIII
PALL7F	F	7	NA	NA	No results
PALL10M(2)	M	10	0.00053	0.01-0.1%	VH3 VH5
PALL19M	M	19	BCP ALL L2 morphology Ph+ with monosomy		No results

Table 4.4.2: Patients with high-risk MRD. Minimal residual disease level including markers where available (data sourced from CellBank)

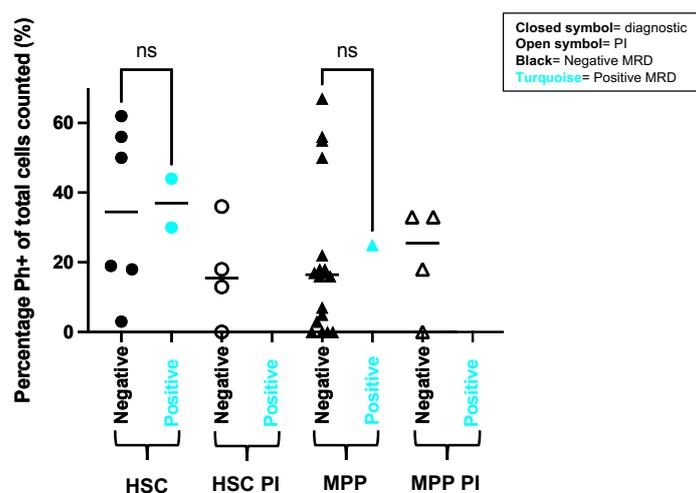


Figure 4.4.1: BCR::ABL1 positivity in patients with high-risk MRD. Patients highlighted (blue) with MRD higher than the average for the cohort (0.03 EOI MRD level).

Unpaired t Test with Gaussian distribution ‘MRD Negative vs MRD Positive’; HSC P value= 0.9035 (ns), MPP P value= 0.8383 (ns).
ns= not significant.

Patients with minimal residual disease higher than 0.01 (positive MRD in the context of IgTCR) were highlighted in the FISH data for BCR::ABL1 presence in HSC and MPPs (Table 4.4.2 and Fig 4.4.1). The FISH scores were highlighted for such high risk patients to determine if high risk MRD status aligned with BCR::ABL1 positivity in HSC and MPPs in diagnostic and post induction samples (Fig 4.4.1). Additionally, there is some evidence that MRD>0.1 is associated with a higher risk of relapse, of which only one patient had an MRD score >0.1 and displayed a FISH score of 30% (Gökbuget et al., 2021). All patients with positive MRD by IgTCR displayed Ph+ within the stem and progenitor cells. Due to small starting numbers of cells collected by FACS, the proportion of Ph+ cells in each sample may be subject to change if a larger number of cells were available. The data displayed here shows that BCR::ABL1 fusion was detectable in all samples with positive MRD but also that the majority of samples with Ph+ HSC/MPPs originated from patients with negative MRD. Hence, from this dataset, a direct link between MRD status by IgTCR and BCR::ABL1 positivity in stem cells cannot be made.

As all patients in the cohort had been diagnosed with Ph+ALL, the cytogenetic abnormality t(9;22) was present. However, patients with atypical karyotype features and cytogenetic abnormalities were highlighted by CellBank. Table 4.4.3 depicts the additional cytogenetic abnormalities present within the samples. Cytogenetic data was available for 21 patients total, of these patients, 9 had additional cytogenetic abnormalities other than BCR::ABL1 fusion. Such abnormalities broadly fell into the categories of gene rearrangement, gain of chromosomes, loss of chromosomes, gene gain and deletion, and 'other abnormal'. These were categorised by probes as alterations to CDKN2A, CRLF2, E2A, CEP2/6/7/X/Y, IGH DC/MAF, ETV6-RUNX1, TEL/AML1 and MLL (Table 4.4.3). Of these 9 samples, 7 had BCR::ABL1 FISH data available for sorted haematopoietic progenitors. Comparison of BCR::ABL1 presence in HSCs and MPPs (diagnostic and PI) samples harvested from patients with and without cytogenetic alterations in addition to t(9;22) displayed no statistical differences (Fig 4.4.2). This indicates that the presence of additional cytogenetic abnormalities is not a requirement for the occurrence of BCR::ABL1 fusion in early haematopoietic stem and progenitor cells.

ID	Probe name	Manufacturer	Interpretation	Patterns overview
PALL12M	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 2F [86%], 2R 2G OF [14%]
	p16 (CDKN2A/B) (9p21)(so)/ CEP9(sg)	Vysis	Deletion	1R 2G OF [92%], 2R 2G OF [8%]
PALL8M	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 2F [85%], 2R 2G OF [15%]
PALL13M	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 2F [88%]
PALL2M	BCR/ABL1 DC DF	Vysis	Gene rearrangement	Undefined
	CRLF2 (Xp22/Yp11) DC BAR	LRCG	Other abnormal	0R 0G 1F [10%], 0R 0G 2F [90%]
PALL5F	BCR/ABL1 DC DF	Vysis	Gene rearrangement	2R 1G 1F [78%], 1R 1G 1F [11%], 2R 2G OF [9%]
PALL10M	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 2F [98%], 2R 2G OF [2%]
PALL5M	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 2F [92%], 2R 2G OF [8%]
PALL3F	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 2F [50%], 2R 2G OF [50%]
PALL2M(2)	BCR/ABL1 DC DF	CytoCell	Gene rearrangement	1R 1G 2F [59%], 2R 2G OF [40%]
	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 3F [38%], 2R 2G OF [37%]
PALL5F(2)	TCF3 (E2A) DC BA	CytoCell	Gain of chromosome(s)	0R 0G 3F [24%], 0R 0G 2F [75%]
	CEP6	CytoCell	Gain	3R 0G OF [49%], 2R 0G OF [50%]
	CEP2 (SG)	CytoCell	Gain of chromosome(s)	0R 3G OF [59%], 0R 2G OF [41%]
	CEP 2(G) + 6(R)	CytoCell	Gain of chromosome(s)	3R 3G OF [50%], 2R 3G OF [9%], 3R 2G OF [6%], 2R 2G OF [35%]
PALL15M	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 2F [66%]
PALL7F	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 2F [49%], 2R 1G 2F [14%], 2R 2G OF [36%]
	IGH DC BAR	Vysis	Other abnormal	0R 0G 4F [6%], 0R 0G 3F [2%], 1R 0G 1F [1%], 1R 1G 1F [1%], 0R 0G 2F [86%]
PALL10M(2)	TCF3 (E2A) DC BA	CytoCell	Other abnormal	0R 0G 4F [8%], 0R 0G 3F [3%], 0R 0G 2F [84%]
	BCR/ABL1 DF+ASS	Vysis	Gene rearrangement	Undefined
PALL2.7	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 2F [66%], 1R 1G 1F [30%]
	TEL/AML1 (ETV6-RUNX1) DC DF	Vysis	Gain of chromosome(s)	2R 3G OF [85%]
	CEPX and CEPY	Vysis	Gain of chromosome(s)	0R 4G OF [92%], 1R 1G OF [8%]
PALL12F	TEL/AML1 (ETV6/RUNX1) DC DF	CytoCell	Gain	2R 4G OF [100%]
	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 3F [100%]
PALL10M(4)	BCR/ABL1 DC DF	CytoCell	Gene rearrangement	1R 1G 3F [72%], 1R 1G 2F [17%], 2R 2G OF [11%]
	MLL	CytoCell	Deletion	0R 0G 1F [70%], 0R 0G 2F [30%]
PALL9F	BCR/ABL1 DF+ASS	CytoCell	Gene rearrangement	1R 1G 2F [90%], 2R 2G OF [4%]
PALL4M	TEL/AML1 (ETV6/RUNX1) ES	Vysis	Gain	3R 2G OF [71%], 2R 2G OF [28%]
	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 3F [100%]
PALL12M(2)	IGH/MAF DC DF	Vysis	Gain	2R 3G OF [88%], 2R 2G OF [11%]
	BCR/ABL1 DC DF	CytoCell	Gene rearrangement	1R 1G 2F [53%], 1R 1G 3F [35%], 2R 2G OF [10%]
PALL14F	BCR/ABL1 DF+ASS	CytoCell	Gene rearrangement	1R 1G 1F 0S [93%], 2R 2G OF 2S [7%]
	D7S486(7q31)(so)/ CEP7(sg)	Vysis	Loss of chromosome(s)	1R 1G OF [87%], 2R 2G OF [13%]
PALL16M	BCR/ABL1 DC DF	Vysis	Gene rearrangement	1R 1G 2F [60%], 1R 1G 3F [20%], 2R 2G OF [20%]

Table 4.4.3: Patients with cytogenetic abnormalities in addition to t(9;22). Cyto-genetic abnormality interpretations including probe name. Data reported where available from CellBank..

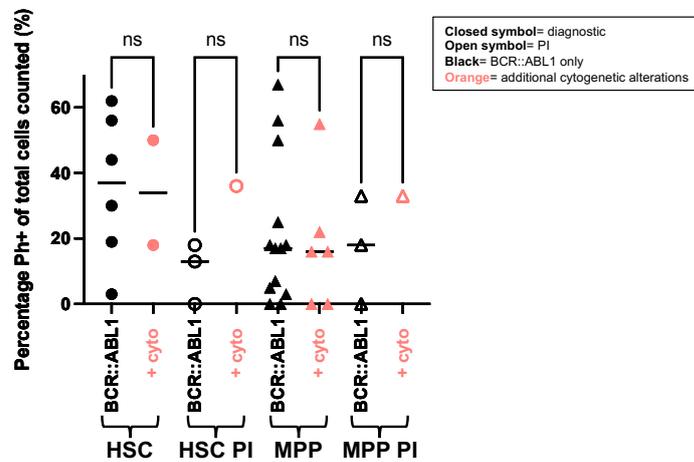


Figure 4.4.2: BCR::ABL1 positivity in patients with cytogenetic abnormalities in addition to t(9;22). Percentage of Ph+ HSC/MPP cells expressed as a percentage of total cells counted for each individual sample. Patients highlighted (orange) with cytogenetic abnormalities as reported by CellBank and CCLG. *Unpaired t Test with Gaussian distribution 'BCR::ABL1 only vs additional cytogenetic alterations'; HSC P value= 0.9310 (ns), HSC PI P value= 0.1392 (ns), MPP P value= 0.7397 (ns) and MPP PI P value= 0.4899 (ns).*
+ cyto= additional cytogenetic alterations

White blood cell count at diagnosis is a standard prognostic factor used in Ph+ALL. High WCC at diagnosis is a strong prognostic indicator for an increased risk of haematological relapse (Akahoshi et al., 2021) (Table 4.4.4 and Fig 4.4.3). Statistical significance was not detected between high white cell counts (above the cohort average of 108.7) and FISH score (Mann-Whitney unpaired t Test (Fig 4.4.3)). This suggests that high white cell count at diagnosis is not indicative of proportion of Ph+ stem and progenitor cells.

ID	Sex	Age	WCC
PALL12M	M	12	21.8
PALL8M	M	8	31.3
PALL13M	M	13	6.8
PALL2M	M	2	8.8
PALL5F	F	5	283
PALL10M	M	10	224.7
PALL5M	M	5	132
PALL3F	F	3	121
PALL2M(2)	M	2	15.3
PALL5F(2)	F	5	70.1
PALL15M	M	15	64.9
PALL7F	F	7	1.1
PALL10M(2)	M	10	541
PALL19M	M	19	U/K

Table 4.4.4: Patients with high white cell counts at diagnosis. Patient WCC at diagnosis (healthy range 4-11). Data reported where available from CellBank.

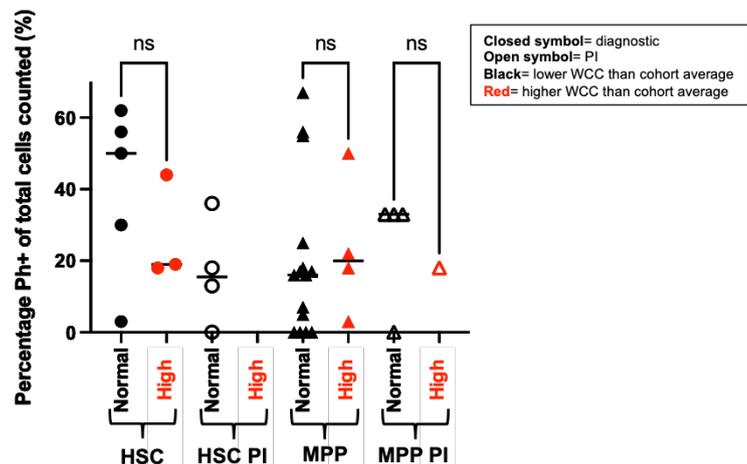


Figure 4.4.3: BCR::ABL1 positivity in patients with high white cell counts at diagnosis. Patients highlighted (red) with high WCC as reported by CellBank and CCLG. WCCs were averaged and patients with above average WCC highlighted. Unpaired t Test with Gaussian distribution 'Normal vs High WCC'; HSC P value= 0.4303 (ns), MPP P value= 0.7886 (ns) and MPP PI P value= 0.7387 (ns). Average WCC= 108.7. Normal= WCC < 108.7, High= WCC > 108.7.

Utilising WCC in addition to MRD ensures that high risk patients are detected within the cohort, even when alternate measurement methods are utilised. Comparison of WCC and MRD methods by multivariate ANOVA displayed no statistical significance and P values for each cell group were as follows; HSC MRD vs WCC $P= 0.3069$, MPP $P= 0.2062$ and MPP PI $P= 0.7729$. Comparing MRD and WCC in Ph positive and negative HSC PI samples resulted in a P value of <0.0001 however, 1+ pairwise comparisons were invalid. As previously mentioned, limited sample size may hinder the ability to detect statistical significance however, multiple methods of MRD analysis is beneficial clinically to ensure that high risk patients can be detected should another method prove inconclusive, as evidenced by the comparison of MRD methods IKZF1 deletion and *BCR::ABL1* transcript level in Hovorkova et al (2017). From this dataset, MRD and WCC alone are not able to predict Ph positivity in HSC/MPPs however, a combination of risk factors may be more beneficial for predicting t(9;22) occurrence in stem and progenitor cells.

FISH scores were then investigated in patients who had reported relapse or incidence of death (Fig 4.4.4). Of the patient samples highlighted, 7 out of the total 8 samples with incidence of relapse/death had previously been highlighted in the high risk MRD or WCC groups. This supports the efficacy of utilising MRD and WCC as indicators for clinical outcomes, as evidenced by the congruity of patients highlighted in figure 4.4.1- 4.4.4. All patients with reported incidences of relapse or death had detectable BCR::ABL1 in early haematopoietic stem and progenitor cells at diagnosis and post induction therapy. However, a number of patients had Ph+ HSC/MPPs and no incidence of death/relapse, suggesting that the presence of BCR::ABL1 in stem cells alone is inadequate for predicting death/relapse occurrence and highlighting that this is still a very heterogeneous population. Statistical analysis of FISH score in relation to likelihood of death/relapse showed no significance (Mann-Whitney unpaired t Test (Fig 4.4.4)) despite all patients with reported death occurrence having positive BCR::ABL1 FISH scores. As previously discussed, these results may be related to low sample size however, it could be surmised from this data that percentage of Ph+ HSC and MPPs is not indicative of relapse/death occurrence. These results suggest that additional factors other than BCR::ABL1 presence in haematopoietic stem and progenitor cells are causative of poor patient outcomes.

ID	Sex	Age	Relapse site	COD	OS	RFS	Transplant	TP type
PALL12M	M	12	N/A	Infection	228	228	Y	MUD
PALL8M	M	8	N/A	Other	532	532	Y	MUD
PALL13M	M	13	N/A	N/A	243	243	Y	MRD
PALL2M	M	2	N/A	N/A	4227	4227	N	N/A
PALL5F	F	5	BM	Progressive ALL	445	333	N	N/A
PALL10M	M	10	BM	Progressive disease	3416	913	N	N/A
PALL5M	M	5	N/A	N/A	295	295	N	N/A
PALL3F	F	3	N/A	N/A	3328	3328	N	N/A
PALL2M(2)	M	2	N/A	N/A	3175	3175	N	N/A
PALL5F(2)	F	5	N/A	Infection	405	405	N	N/A
PALL15M	M	15	N/A	N/A	2755	2755	N	N/A
PALL7F	F	7	N/A	N/A	2785	2785	N	N/A
PALL10M(2)	M	10	CNS	Relapse	517	384	N	N/A
PALL19M	M	19	BM	Relapse GVHD	U/K	rel 28/02/2017	21/10/2015	sib allo in CR2

Table 4.4.5: Patients with reported incidences of relapse or death. (1) Available survival data provided by CellBank. COD= cause of death, OS= overall survival (days), RFS= relapse-free survival (days), TP type= transplant type, MUD= matched unrelated donor, MRD= matched related donor, GVHD= graft versus host disease, U/K= unknown, rel= relapse, sib allo in CR2= sibling allogeneic in complete remission

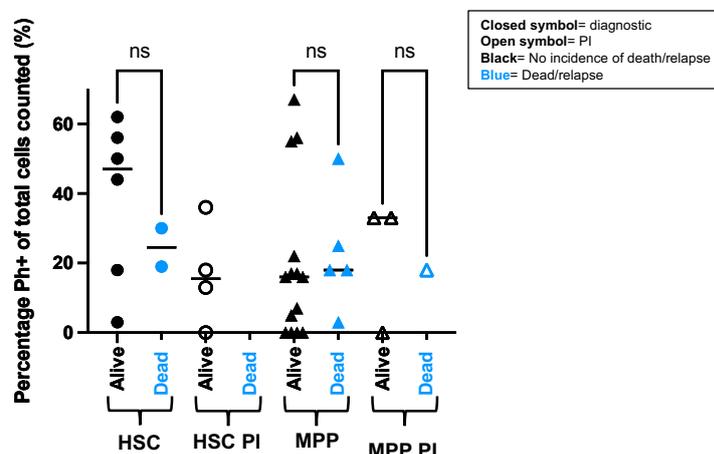


Figure 4.4.4: BCR::ABL1 positivity in patients with reported incidences of relapse or death. Patients highlighted (blue) with incidences of relapse/death as reported by CellBank and CCLG. Unpaired t Test with Gaussian distribution 'Alive vs Dead'; HSC P value= 0.4445 (ns), MPP P value= 0.7972 (ns) and MPP PI P value= 0.8725 (ns).

Clinical and molecular features used to identify high-risk patients in figures 4.4.1- 4.4.4 were those reported at diagnosis. We next sought to utilise the atypical cell populations (i.e. (large populations of myeloid progenitors and small populations of lymphoid progenitors at diagnosis) presented above to assess if these samples could be identified through BCR::ABL1 identification by FISH within the HSC and MPP population (Fig 4.4.5). HSC and MPP FISH scores for samples with aberrant haematopoietic population sizes at diagnosis were compared to samples with normal FACS population sizes. 'Aberrant' was defined as above mean of the collected population and 'Normal' was defined as below mean of the population (Fig 4.4.6). While no significance was detected, it was noted that all of the aberrant samples had positive FISH scores. This would be an interesting observation to follow up in a cohort with a larger sample pool to verify if the lack of significance is due to small sample size or if FISH scores in stem and progenitor cells are truly unrelated to samples with unusual cell population sizes.

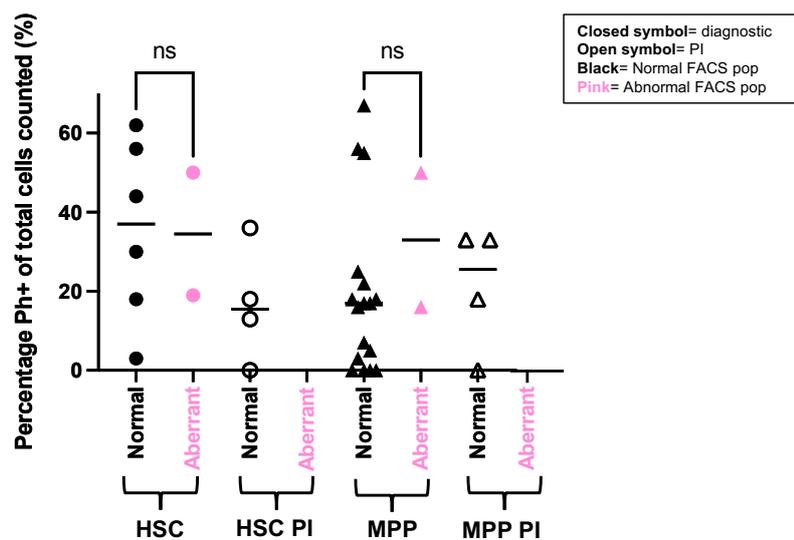


Figure 4.4.6: BCR::ABL1 positivity in patients with aberrant haematopoietic diagnostic populations measured by FACS. Mean of FACS populations were calculated for HSC & MPP diagnostic (diag) and PI. 'Normal'= below mean, 'Aberrant'= above mean.
HSC diag mean= 1.112%
HSC PI mean= 0.3878%
MPP diag mean= 1.29%
MPP PI mean= 3.283%
'Normal vs Aberrant' HSC diag P value= 0.9586 (ns) and MPP diag P value= 0.3955 (ns).

Multivariate analysis (Multivariate ANNOVA) analysis of the combination of different methods of risk analysis with relation to Ph positivity in HSC and MPP diagnostic and PI samples was run to determine if risk can be associated with FISH score. Risk features compared were the presence of cytogenetic abnormalities in addition to BCR::ABL1, high WCC at diagnosis, MRD

positivity, reported incidence of death/relapse and aberrant haematopoietic population sizes in diagnostic samples analysed by FACS (Figs 4.41- 4.4.5). No statistical significance was detected in the multivariate FISH analyses of diagnostic HSC samples (P value= 0.2108 (ns)) however, in the HSC PI samples and MPP diagnostic and PI samples, significance was found (HSC PI $P < 0.0001$ (****), MPP diagnostic $P = 0.0002$ (***) and MPP PI $P < 0.0001$ (****)) (Table 4.4.6). This suggested a significant association between high risk BCR::ABL1 FISH scores and clinical features mentioned above. As has been discussed throughout this chapter, a major limiting factor is the small sample size and failures of FISH analyses, hence, in order to bolster these findings, larger patient cohorts should be utilised.

HSC	Source of Variation	% of total variation	P value	P value summary	Significant?		
	Row Factor	20.3	0.2109	ns	No		
	Column Factor	0.2275	0.9991	ns	No		
	ANOVA table	SS (Type III)	DF	MS	F (DFn, DFd)	P value	
	Row Factor	3115	5	623.1	F (5, 30) = 1.528	P=0.2109	
	Column Factor	34.91	4	8.727	F (4, 30) = 0.02140	P=0.9991	
	Residual	12232	30	407.7			
	Data summary						
	Number of columns (Column Factor)				5		
	Number of rows (Row Factor)				6		
Number of values				40			
HSC PI	Source of Variation	% of total variation	P value	P value summary	Significant?		
	Row Factor	89.18	<0.0001	****	Yes		
	Column Factor	2.404	0.6272	ns	No		
	ANOVA table	SS (Type III)	DF	MS	F (DFn, DFd)	P value	
	Row Factor	2973	3	991	F (3, 12) = 32.97	P<0.0001	
	Column Factor	80.15	4	20.04	F (4, 12) = 0.6667	P=0.6272	
	Residual	360.7	12	30.06			
	Data summary						
	Number of columns (Column Factor)				5		
	Number of rows (Row Factor)				4		
Number of values				20			
MPP	Source of Variation	% of total variation	P value	P value summary	Significant?		
	Row Factor	43.04	0.0002	***	Yes		
	Column Factor	0.1751	0.9939	ns	No		
	ANOVA table	SS (Type III)	DF	MS	F (DFn, DFd)	P value	
	Row Factor	17202	17	1012	F (17, 74) = 3.289	P=0.0002	
	Column Factor	69.97	4	17.49	F (4, 74) = 0.05685	P=0.9939	
	Residual	22767	74	307.7			
	Data summary						
	Number of columns (Column Factor)				5		
	Number of rows (Row Factor)				18		
Number of values				96			
MPP PI	Source of Variation	% of total variation	P value	P value summary	Significant?		
	Row Factor	87.24	<0.0001	****	Yes		
	Column Factor	1.266	0.8738	ns	No		
	ANOVA table	SS (Type III)	DF	MS	F (DFn, DFd)	P value	
	Row Factor	3219	3	1073	F (3, 12) = 27.35	P<0.0001	
	Column Factor	46.73	4	11.68	F (4, 12) = 0.2978	P=0.8738	
	Residual	470.8	12	39.23			
	Data summary						
	Number of columns (Column Factor)				5		
	Number of rows (Row Factor)				4		
Number of values				20			

Table 4.4.6: 2way ANOVA of FISH scores for overall multivariate high risk vs low risk clinical factors. Ordinary Two-way ANOVA, Alpha 0.05. HSC= multivariate high risk vs low risk FISH scores in HSC diagnostic samples, HSC PI= multivariate high risk vs low risk FISH scores in HSC PI samples, MPP= multivariate high risk vs low risk FISH scores in MPP diagnostic samples, MPP PI= multivariate high risk vs low risk FISH scores in MPP PI samples.

4.5 Discussion.

Although the treatment of Ph+ALL has improved exponentially, the cell of origin remains undefined (Nishiwaki et al., 2020 and Wieduwilt., 2022). The importance of identifying the cell type in which the Philadelphia chromosome arises is essential for understanding which cells are likely to be intrinsically affected by the constitutively activated tyrosine kinase encoded by said fusion chromosome. This has been exemplified in CML where the identification and targeting of Ph+ LSCs have resulted in a clinical success story (Holyoake & Vetrie., 2017).

Despite being driven by the same fusion event, CML and classical Ph+ALL differ by which cell type predominates in disease pathology. CML is a myeloid-driven leukaemia but in the terminal blast crisis, it can be seen to have involvement by myeloid, lymphoid and mixed lineage cells, owing to its stem cell origin. In contrast, the majority of Ph+ALL cases display unchecked proliferation in lymphoid lineage cells only. Emerging reports discuss the existence of a Ph+ALL subgroup with significant myeloid disease involvement, named CML-like Ph+ALL (Hovorkova et al., 2017).

The higher frequency and reporting of mixed-lineage disease involvement in CML than in Ph+ALL is an important factor in why the cell of BCR::ABL1 origin is known in CML but remains unidentified in Ph+ALL. Referring to older models of haematopoiesis, it can be said that in order to affect cells from multiple lineages, the fusion event must occur prior to lineage bifurcation during early haematopoiesis. In newer models, this may be viewed as the fusion event occurring during the very early stages of lineage commitment, possibly in stem and progenitor cells primed for multiple lineages. Whichever model is used, it is clear that in order to understand such leukaemic initiation events, cells in the earliest stages of haematopoiesis must be examined.

As the cell of origin in Ph+ALL remains elusive, we hypothesised that the fusion event varied amongst patients and that there would be some patients where the BCR::ABL1 fusion could be identified out with a CD19+ blast population. Because of the observation by Horokova et

al in 2017, that a subpopulation of patients, termed CML-like Ph+ALL, remained *BCR::ABL1* positive by MRD analysis whilst being IgTCR negative, we hypothesised that the fusion even would arise within a stem cell compartment, namely HSCs or MPPs (Hovorkova et al., 2017). Inevitably, this would mean that Ph+ALL is heterogeneous, not solely with interpatient variability, but with the disease driver. It was therefore initially believed that that only patients with the CML-like disease would have Ph+ HSCs and MPPs, with the rest of the cohort having detectable *BCR::ABL1* fusion in lymphocyte progenitors, in line with the current understanding of 'classical' Ph+ALL. These predictions aligned with the current dogma that CML contains Ph+ stem and progenitor cells and Ph+ALL does not, with this being the reason why CML has the capacity to be a mixed lineage disease and Ph+ALL is restricted to lymphocytes.

We first approached this hypothesis by utilising a flow cytometry-based approach to identify haematopoietic stem and progenitor cells by surface marker expression and collection of such cells. Utilising the wealth of research knowledge on haematopoietic stem and progenitor cell surface marker profile, identification of HSCs, MPPs, CMPs, GMPs, CLPs, Pro and PreB cells has been reliably performed by FACS. This allowed for the isolation of rare populations of HSCs in Ph+ALL diagnostic samples with high accuracy. Though heterogeneity in population sizes was expected between samples (as cells were isolated from primary patient samples), broad trends in progenitor cell numbers were anticipated such as small HSC fractions, few myeloid progenitors and large lymphoid progenitors in diagnostic samples. Heterogeneity was indeed observed however, a small number of patients displayed progenitor population sizes contradictory to Ph+ALL cellular phenotypes and in line with results from Hovorkova et al (2017), namely, enlarged myeloid populations. Additionally, a small number of samples had large HSC populations, a surprising trait, and one which may display an enhanced survival advantage or increased drive toward the production and maintenance of HSCs or the presence of de-differentiated blast cells.

We next analysed sorted cell populations by FISH to identify those with *BCR::ABL1*. By utilising a number of cell surface markers, straining to prevent clumping, FMO controls, and validation of cell surface markers post sort, the purity of the sort was confirmed. As discussed, a number of samples failed FISH analysis due to small starting cell number or nuclear lysis. Of the

scorable samples, a high proportion of Ph⁺ HSCs and MPPs were detected. Initial predictions (based on the proposed CML-like Ph⁺ALL incidence of ~30%) were that approximately one third of samples would have BCR::ABL1 fusion in early stem and progenitor cells, based on the assumption that only CML-like samples would contain Ph⁺ HSC/MPPs. Therefore the high incidence of FISH positive cells was not anticipated and thus it was proposed that Ph⁺ALL is a highly heterogeneous disease and t(9;22) alone is inadequate for the establishment of the CML-like subtype. In addition, HSC populations were not directly investigated by Hovorkova et al (2017) and their findings displayed MRD discordance between methods, hence, unexpectedly large Ph⁺ HSC observations may be reflective of the differences between populations investigated and methods used. These observations may also be influenced by an unintended artificial enrichment of scorable cells with 'fit' HSCs better able to survive the cryopreservation process. The reason why the majority of scorable cells display t(9;22) may be that BCR::ABL1 confers a survival advantage, allowing such cells to be more robust post cryopreservation and less likely to undergo nuclear lysis than Ph negative HSC/MPPs. Hence, the FISH dataset may be enriched with fit Ph⁺ cells. The large number of samples with Ph⁺ HSC/MPPs may still reflect a heterogeneity within the Ph⁺ALL cohort in addition to BCR::ABL1-driven survival advantage.

Although this approach can ensure the purity of cell populations collected, there were a number of caveats that arose. Firstly, the cell populations, particularly HSCs and MPPs, collected were extremely small (often <100 cells). In samples such as Ph⁺ALL diagnostic BM aspirate samples, the predicted number of HSCs is low, hence collection of few HSCs in these samples were unsurprising. Additionally, washing procedures post collection and during the preparation of FISH slides can often result in the loss of cells and therefore, the number of cells available for analysis by FISH was frequently much lower than what had been in the original sample. While FISH analysis procedures used in the clinical setting aim to count 100 cells minimum in order to calculate the percentage of Ph⁺ cells, due to the conditions discussed above, this was often impossible to do. Hence, in very small samples, as many cells as possible were counted however this was not to clinical standards.

Furthermore, the use of cell surface markers alone for identification and isolation of haematopoietic cell populations has limitations. Although flow cytometry is a well-

established practice used in both research and clinical laboratories, its use in Ph+ALL can incur some issues. The foremost issue is that of blast cell de-differentiation whereby lymphoblasts aberrantly express cell surface markers more commonly associated with stem cells. This surface marker promiscuity can lead to the misidentification of blast cells as HSCs. Blast cell de-differentiation may result in a population of cells which may appear to be HSCs when phenotyped using cell surface markers however, lymphoblasts (regardless of aberrant surface marker expression) would still bear the hallmarks of having gone through latter stages of lymphocyte development, namely Ig/TCR gene rearrangement. This process ensures an adequately diverse immune repertoire of B and T lymphocytes and is a feature of lymphoid lineage fixed cells, hence, true HSCs will not display these gene rearrangements due to being uncommitted to a specific lineage. In the context of the findings from the 2017 Hovorkova et al paper, identification of BCR::ABL1 positive cells with no Ig/TCR gene rearrangement may be reflective of Ph+ HSCs discussed in this chapter which would not display hallmarks of lymphoid development. Additionally, expression of CD10 is not normally observed to be aberrantly expressed, hence the chosen cell surface markers were deemed appropriate to differentiate HSCs from lymphoblasts. Utilising phenotypic and functional assays (ie FACS and stem cell long term culture long-term culture-initiating cell (LTCIC) assays/clonogenic assays) in combination would help ensure that the cells isolated are correctly being identified as haematopoietic stem cells. This is planned within a further project within the Horne lab. Another method of verification is the use of sequencing rearranged Ig/TcR genes. This caveat was not discussed in Hovorkova et al as Ig/TCR rearrangement was not assessed in HSC populations. The goal of the Hovorkova 2017 paper was to assess discordance between MRD methodologies (Ig/TCR gene recombination and BCR::ABL1 transcript quantification). The resultant discordancy observed between methods in CML-like patients who were measured as MRD negative by Ig/TCR gene rearrangement but maintained BCR::ABL1 transcripts is, at present, the only definition available for CML-like status. As previously mentioned, clinical data was not available for all patients in this cohort, hence available MRD data and methods have been reported. Available MRD data provided by CellBank was generated by Ig/TCR rearrangement and did not include BCR::ABL1 transcript data. While utilising this MRD concordance method would be beneficial for definition of CML-like cases just as in Hovorkova et al, the available data in the CellBank cohort is variable. These limitations include (but are not limited to) the use of historical samples which pre-date MRD standardisation, multiple

and ongoing development of MRD methods and the likelihood of multiple MRD methods being used clinically. Hence, it is outwith the capabilities of this project to utilise the diagnostic method established by Hovorkova et al, which necessitates utilisation of alternative informative methods discussed in this chapter.

As there are no current diagnostic criteria for CML-like Ph+ALL, clinical information in addition to flow cytometry and FISH data were used holistically with the goal of identifying Ph+ALL subtypes with CML-like features. This revealed a larger than expected population of patients with Ph+ early stem and progenitor cells. Based off the data reported in Hovorkova et al (2017), it was postulated that only 'CML-like' patients would exhibit BCR::ABL1 fusion events in HSC/MPP populations. However, the data discussed in this chapter displayed a proportion of samples with early t(9;22) fusion, larger than the estimated 1/3 patients in the Hovorkova cohort designated 'CML-like'. Further investigation into these patients revealed that BCR::ABL1 fusion in HSC/MPPs alone, is not enough to establish a high-risk disease state, suggesting that downstream events hold more importance in CML-like Ph+ALL. Hence, Ph presence in HSC/MPPs should not be the sole diagnostic criteria for identifying CML-like patients and the large number of samples with Ph+ HSC/MPPs observed during this project, may still be congruent with the findings from Hovorkova et al (2017). In addition, Hovorkova et al assessed BCR::ABL1 presence in ALL blasts, non-ALL B cells, myeloid cells/monocytes and T cells, as BCR::ABL1 transcript levels were not investigated in HSCs, it is unknown whether similar proportions of Ph+ HSC/MPPs could be identified in both the Hovorkova and the CellBank cohort.

With such caveats in mind, promising results were generated during this investigation. It was found that the majority of Ph+ALL samples had BCR::ABL1 fusion detectable in haematopoietic stem and progenitor cells. Due to issues with sample size as previously discussed, a correlation between percentage of Ph+ cells and disease outcome could not be established. However, an interesting observation was made in that patients with high risk clinical features such as high WCC, MRD or the occurrence of relapse/death were not the only patients with BCR::ABL1 in their HSC and MPPs. Patients without such high-risk features were found to have BCR::ABL1 expression in their early haematopoietic cells, an unexpected outcome. However, all patients with above average high risk clinical features all had

BCR::ABL1 fusion at the early stem and progenitor developmental level. This suggests that BCR::ABL1 fusion alone is inadequate for establishing a high risk phenotype but its presence may be required in order for a high risk status to occur. It may also be posited that samples in which BCR::ABL1 can be detected in HSC and MPP, may have a survival advantage over those with Ph- stem and progenitor cells. It is well understood that BCR::ABL1 fusion (and resultant tyrosine kinase activity) can negatively affect downstream signalling pathways such as JAK/STAT and ERK. Hence, some of these metabolic alterations may result in a survival advantage when applied to the cryopreservation, culture and sorting process. Therefore, due to biological differences between samples, the cohort may be artificially enriched for samples containing Ph+ HSC and MPPs.

To continue investigation into the CML-like subtype, RNAseq will be used to investigate transcriptional heterogeneity within the Ph+ALL cohort (discussed in the next chapter). As there are currently no diagnostic genes for the CML-like subtype, patient samples cannot be isolated from the RNAseq data prior to data analysis. Therefore, identification of differentially expressed genes may not only highlight transcriptional differences between the proposed Ph+ALL subtypes, bringing light to mechanisms which cause these differing phenotypes, but also allow CML-like samples to be identified due to an altered transcriptional profile compared to the rest of the Ph+ALL cohort.

Despite having access to a relatively small cohort of Ph+ALL samples, and examining a rare population of cells, the investigations of this chapter has allowed for the accurate identification of BCR::ABL1 fusion in Ph+ALL HSCs. This not only represents the first observation of its kind, but also a major input in the field of Ph+ALL research.

Chapter 5. Results III. Transcriptional investigation of a heterogenous Ph+ALL cohort.

5.1.1 Introduction: Background and clinical relevance.

As previously discussed, *BCR::ABL1* in haematopoietic progenitor cells is believed to be an important aspect of the CML-like Ph+ALL subtype. Within chapter 4, we revealed a higher-than-expected proportion of Ph+ HSC/MPPs in the Ph+ALL cohort, suggesting that Ph+ALL is likely more heterogeneous than previously described. To determine if heterogeneity existed at a transcriptional level, we next sought to assess the gene expression of patient samples used in previous analysis in Chapter 4. RNAseq was used to assess transcriptional variation between the proposed Ph+ALL subtypes. RNAseq is a highly sensitive method of gene expression analysis used for high-throughput transcriptome profiling for the understanding of functional processes which underpin disease pathology (Marco-Puche et al., 2019). Utilising RNAseq to reveal the entire transcriptome of leukaemic patient samples allows for an enhanced understanding of disease pathogenesis and subtype-specific gene regulation. The clinical benefits of such investigations include the identification of new diagnostic panels (including biomarkers of disease), identification of novel therapeutic targets and increased understanding of disease mechanisms.

One aim of this chapter was to investigate the potential for transcriptional heterogeneity within the paediatric Ph+ALL cohort. Previous investigations have described the existence of CML-like Ph+ALL, represented by a group of Ph+ALL patients where *BCR::ABL1* persisted despite clearance of IgTCR MRD following induction chemotherapy (Hovorkova et al., 2017). However, disease aetiology is known. Previous data from Chapter 4 suggested that *BCR::ABL1* was able to persist after commencement of induction treatment, evidenced by the maintenance of Ph+ HSPCs in PI samples (Fig 4.4.1.2- 4.4.1.6). This data supports what is currently understood about the CML-like subtype however, gene expression features unique to this subtype are still to be elucidated. To explore this further, transcriptional profiles of Ph+ALL BMNC samples from paediatric patients used in previous experiments were examined by RNAseq. Examining overall transcriptional profiles of all samples in a cohort is an approach to identify heterogeneity in gene expression which could be used to identify disease subgroups. Primary samples will naturally display a level of transcriptional variation arising

from patient-specific differences (ie age, previous disease status and treatment history). A non-homogenous cohort of primary samples will be transcriptionally heterogeneous however, patients belonging to a distinct subgroup would predictably have a notable difference in gene expression while maintaining some transcriptional similarities with the rest of the cohort.

As genes associated with the CML-like subtype are yet to be elucidated, target genes which could be used to identify CML-like samples from a Ph+ALL population are not available. Hence, CML-like samples cannot be identified from the outset of analysis. In order to identify potential CML-like samples we propose a method of integrating the gene signature (Table 5.1.1) developed in chapter 3 with the transcriptional data generated in this chapter. We then plan to validate this with the overall RNAseq transcriptional signature of the samples and correlate findings with clinical features available for the patient samples. Finally, gene ontology tools and DEGs will be examined for any subgroups identified to determine if CML-disease specific gene expression patterns exist within the proposed CML-like samples.

Our hypothesis being that CML-like Ph+ALL is likely a heterogeneous group which shares CML disease features such as the retention of *BCR::ABL1* expression post treatment (in CML by the maintenance of LSCs), we believe that patients belonging to this subtype would have a transcriptional signature different to the standard Ph+ALL cohort, and that the gene signature could be used for diagnosis. As previously discussed in Chapter 4, HSC population size in standard Ph+ALL diagnostic samples are likely to be very small hence, HSC transcriptional signatures would be unlikely to be found in such bulk samples. Findings from Chapter 4 displayed atypically large HSC populations in a small number of samples which may represent the CML-like subtype. Upregulation of HSC/stem-related genes in CML-like Ph+ALL samples when compared to 'standard' Ph+ALL or similar regulation of such genes when compared to CML-LBC samples may suggest an increased stem cell or LSC activity, similar to disease pathology seen in CML.

	Gene name	Function
Upregulated	ATF4	activating transctipion factor 4
	CCL8	C-C motif chemokine ligand 8
	CCND1	cyclin D1
	FLI1	Fli-1 proto-oncogene, ETS transcription factor
	FLT3	fms related tyrosine kinase 3
	FZD6	frizzled class receptor 6
	LCP1	lymphocyte cytosolic protein 1
	LYL1	basic helix-loop-helic family member
	MPL	MPL proto-oncogene, thrombopoietin receptor
	MPO	myeloperoxidase
	MYC	v-myc avian myelocytomatosis viral oncogene homolog
	YY1	YY1 transcription factor 1
Downregulated	CREBBP	CREB binding protein
	HIPK1	homeodomain interacting protein kinase
	LDB1	LIM domain binding
	MYB	MYB proto-oncogene, transcription factor

Table 5.1.1: Proposed CML-like gene expression pattern. Differentially expressed genes identified in CML and potential CML-like Ph+ALL samples from publicly available GEO datasets.

5.1.2 Aims.

The aims of this chapter were:

1. To establish heterogeneity within the Ph+ALL population by investigating overall transcriptional profiles and detection of any transcriptomic subtypes that may arise;
2. To investigate genes differentially regulated between CML-LBC and Ph+ALL;
3. To correlate differentially expressed genes with clinical outcomes;
4. To determine if BCR::ABL1 presence in HSCs or MPPs (detected by FISH) results in a specific transcriptional profile;
5. To validate chapter 3 gene signature generated from publicly available datasets within primary samples.

5.1.3 Sample selection for RNAseq.

11 diagnostic Ph+ALL samples from previous experiments were selected for use in the RNAseq study. Lymphoid blast phase samples were used as a control, where cell of origin is known to be within a stem cell (or cell that has acquired stem cell properties); these samples were not age matched. As previously discussed, CML in children is exceedingly rare, hence paediatric CML LBC samples were unavailable for sequencing. RNA quality and quantity for each of the Ph+ALL and CML LBC samples sequenced is listed in table 5.1.2 with samples summarised by Glasgow Polyomics as 'Good', 'Weak but probably OK', 'Probably OK' or 'Slight degradation'. Clinical features associated with samples selected is presented in Table 5.1.3. Samples which had large amounts of RNA degradation, salt contamination or very low RNA concentrations (below 0.5ng/ μ L) were omitted based on the recommendations of Glasgow Polyomics so as not to waste money and resources running samples unlikely to yield data. Samples were processed as specified in materials and methods. Briefly, bulk samples (unsorted cells) were harvested from BMNC overnight culture (with physiological growth factor) and preserved in RNAprotect (QIAGEN) with RNA extraction being performed on multiple samples at once to reduce intra-sample variation resulting from the extraction process. Once RNA was extracted, samples were assessed by Glasgow Polyomics for RNA quality (RIN - RNA integrity number) and quantity (concentration in ng/ μ L) using a Bioanalyzer.

Sample ID	Concentration (ng/ μ L)	RIN	Comments
PALL2M	1.01	7.1	Good
PALL2M(2)	0.566	5.4	Weak but probably OK
PALL3F	0.6	6.7	Weak but probably OK
PALL4M	0.78	6.3	Probably OK
PALL5F(2)	0.874	7	Weak, slightly degraded but probably OK
PALL8M	0.722	8	Good
PALL10M	3.02	9.1	Good
PALL10M(4)	0.923	5.9	Weak but probably OK
PALL12M	0.962	7.9	Good
PALL15M	3.2	9.6	Good
PALL19M	0.654	7	Probably OK
CML-LBC1	0.633	7.1	Weak but probably OK
CML-LBC2	1.03	7.3	Slight degradation but probably OK

Table 5.1.2: RNAseq sample quality information. RNA sample quality control outputs including RNA concentration, biological concentration, RNA integrity number (RIN) and quality summary provided by Glasgow Polyomics.

Sample ID	PALL2M	PALL2M(2)	PALL3F	PALL4M	PALL5F(2)	PALL8M	PALL10M	PALL10M(4)	PALL12M	PALL15M	PALL19M
Trial	ALL2003	ALL2003	ALL2003	UKALL2011	ALL2003	ALL2003	ALL2003	UKALL2011	ALL2003	ALL2003	EsPhALL
Sex	M	M	F	M	F	M	M	M	M	M	M
Age	2	2	3	4	5	8	10	10	12	15	19
Ph+%	50 (HSC)	0	0	0	18 (HSC)	30 (HSC)	19 (HSC) 50 (MPP)	0	17 (MPP)	0	5 (MPP)
MRD Group	N/A	0-0.005%	0%	U/K	0.01-0.1%	>5%	0.1-1.0%	U/K	0.01-0.1%	0-0.005%	BCP ALL L2 morphology Ph+ with monosomy
Cause of death	N/A	N/A	N/A	U/K	Infection	Other	Progressive disease	U/K	Infection	N/A	Relapse GVHD
Relapse site	N/A	N/A	N/A	U/K	N/A	N/A	ISO BM	U/K	N/A	N/A	ISO BM
Relapse free survival (days)	4227	3175	3328	U/K	405	532	913	U/K	228	2755	rel 28/02/2017
Transplant	N/A	N/A	N/A	U/K	N/A	Y	N/A	U/K	Y	N/A	21/10/2015
Transplant type	N/A	N/A	N/A	U/K	N/A	MUD	N/A	U/K	MUD	N/A	sib allo in CR2

Table 5.1.3: RNAseq sample patient clinical information. Clinical outcomes and diagnostic features were provided for each Ph+ALL sample by CellBank. Ph%= percentage of BCR::ABL1 positive cells in stem and progenitor cells (calculated by FISH as part of this project (cell populations indicated)). N/A= information not available. U/K= information not known.

Bulk RNA was harvested as samples were thawed for sorting, due to natural heterogeneity between samples, RNA quality varied. Samples with the highest RIN and large quantities of RNA were therefore selected for sequencing and as a result, 11 Ph+ALL and 2 CML-LBC samples were selected (Table 5.1.2 & 5.1.3).

5.2.1 Results III: CML-LBC vs Ph+ALL principal component analysis (PCA) identifies transcriptionally and clinically distinct Ph+ALL subgroups.

In order to elucidate the differences in gene expression between the samples sequenced, differential expression analysis was performed. Using R Studio, datasets were imported and converted to a matrix to allow the count data to be marked as columns and desired samples for analysis to be selected. Genes with low numbers of reads (≤ 5 reads) were then filtered out. Removal of low count genes was performed as genes not expressed at a biologically meaningful level offer little insight into disease function (Bourgon et al., 2010), additionally removal of low count differentially expressed genes (DEGs) may improve sensitivity, precision and allow the mean-variance relationship in the data to be estimated with greater reliability (Law et al., 2016).

After removal of low count DEGs, the conditions of the samples were defined, in this case “CMLLBC” and “PhALL”. A ‘coldata’ data frame was then created for the count matrix and conditions specified using the DESeq and dds functions. The PCA (principal component analysis) plot was then created using the DESeq2 and ggrepel package to conveniently plot each sample by means of variance.

To first identify global differences in gene expression between samples, PCA was used (Fig 5.2.1). Three distinct populations were identified (Fig 5.2.1). Of the 11 total Ph+ALL samples sequenced, 8 formed a tight Cluster displaying significant transcriptional similarity between these samples, termed Cluster 2. The transcriptional similarity is suggestive of a similar transcriptional pathology and may infer similar disease pathogenesis. The total PCA plot displayed 3 Ph+ALL samples which were distal to the rest of the cohort in Cluster 2. These samples were designated Cluster 3. As expected, CML-LBC samples were distinct to both Ph+ALL Clusters on the PCA plot and were labelled Cluster 1.

The identification of Cluster 3 confirms transcriptional heterogeneity within the Ph+ALL cohort. The presence of a separate Ph+ALL Cluster indicates that these samples have a differing disease phenotype and biological mechanism to Cluster 2 and may represent the CML-like Ph+ALL subtype.

We next sought to determine if samples from Cluster 3 could be identified through clinical prognostic details and outcomes. Inspection of the clinical information available for samples in Cluster 3 displayed that these patients had the poorest clinical outcomes, including a higher incidence of relapse and death, and higher WCC and MRD than the rest of the cohort (Table 5.1.3). Prior to exploring DEGs within these samples, PCA analysis has displayed that the three outlying samples represent transcriptionally distinct Ph+ALL subgroup and follow up with patient data shows a clinical difference between samples. Samples previously identified as potential CML-like Ph+ALL included PALL10M, PALL19M, PALL8M, PALL2M, PALL5F(2) and PALL12M. these were highlighted due to presence of BCR::ABL1 in HSCs by FISH (PALL10M, PALL19M, PALL8M, PALL2M, PALL5F(2) and PALL12M), atypical haematopoietic progenitor population size by FACS (PALL2M, PALL10M and PALL19M) or high MRD levels (PALL8M) (Table 5.1.3).

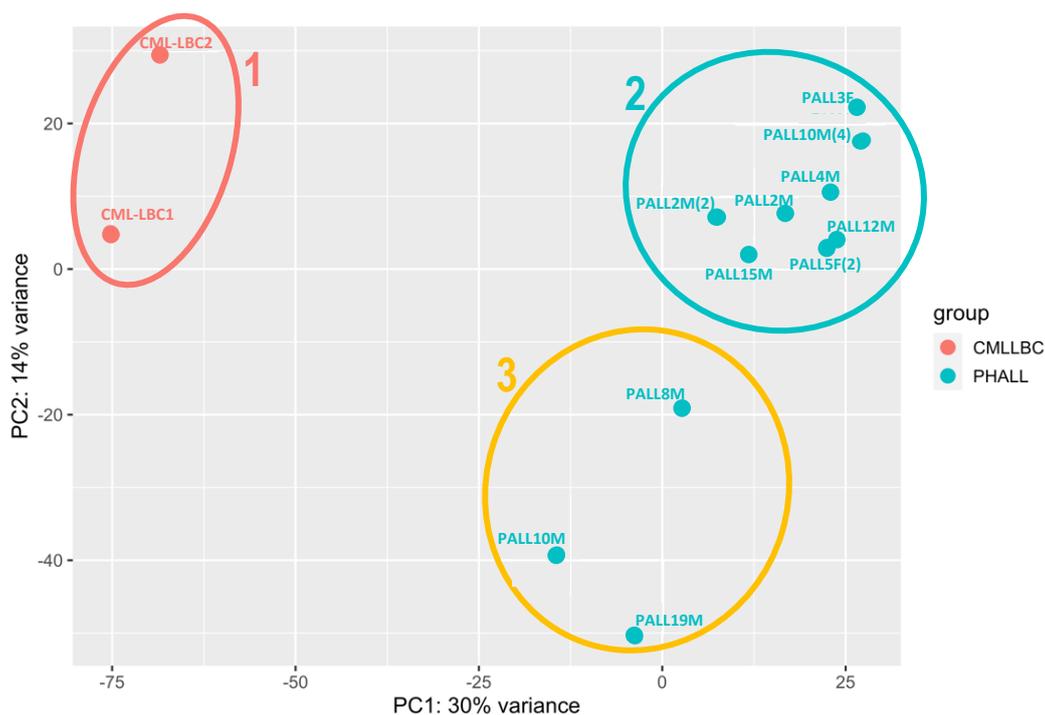
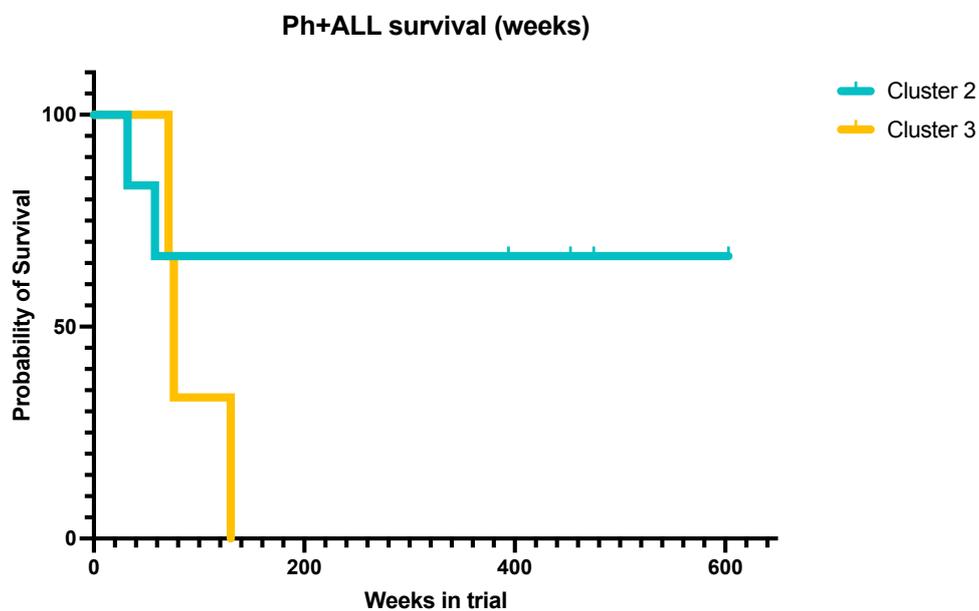


Figure 5.2.1: Principal component analysis (PCA) plot, CML-LBC vs Ph+ALL. CML-LBC (n=2) and Ph+ALL (n=11). Cluster 1 (red)= CML-LBC samples (n=2), Cluster 2 (blue)= Ph+ALL samples (n=8) and Cluster 3 (yellow)= subset of Ph+ALL samples (n=3).

A Kaplan Meier curve was plotted to demonstrate mortality rate of Ph+ALL patients of Cluster 2 and 3 over time in weeks (week 0= diagnosis) (Fig 5.2.2). This displayed that patients belonging to Cluster 3 had lower probability of survival overall. A caveat of the Ph+ALL dataset was that clinical data was not available for all samples in the dataset. Survival data was available for 14 of the 23 Ph+ALL patients in the total cohort of samples provided by CellBank. Of these 14 samples, 8 had appropriate sample quality for RNA sequencing, hence survival data was not available for samples PALL10M(4) and PALL4M (both belonging to cluster 2) (Table 5.1.2).



Sample ID	Weeks in trial	Cluster 2	Cluster 3
PALL3F	475	0	
PALL2M	603	0	
PALL2M(2)	453	0	
PALL15M	394	0	
PALL5F(2)	58	1	
PALL12M	32	1	
PALL8M	76		1
PALL10M	130		1
PALL19M	71		1

Figure 5.2.2: Kaplan Meier survival analysis (Ph+ALL). Incidence of death over time (weeks). Plot represents available data for patients in Cluster 2 (blue) & Cluster 3 (yellow) (clinical data provided by CCLG/CellBank). Samples with available information displayed in table (no available survival data for samples PALL10M(4) & PALL4M).

5.2.2 Results III: Gene expression profiling differentiates between bulk CML-LBC and Ph+ALL.

After exploring the overall transcriptional differences between CML-LBC and Ph+ALL samples in section 5.2.1, differential gene expression was interrogated (Fig 5.2.3). By comparing DEGs within the 2 populations, it was hypothesised that within a stem cell-driven leukaemia (i.e. CML-LBC), DEGs identified would be more associated with stem cell function, such as *CD74* and *MEF2A*. Whereas, within Ph+ALL, where the mutation is presumed to occur in a mature progenitor, DEGs would involve lymphoid function.

Additionally, with LSCs being a well described disease feature in CML and (as of yet) undetected in ALL, leukaemic stem cell activity-associated genes such as *HIF1A* would predictably be upregulated in CML (Soverini et al., 2021).

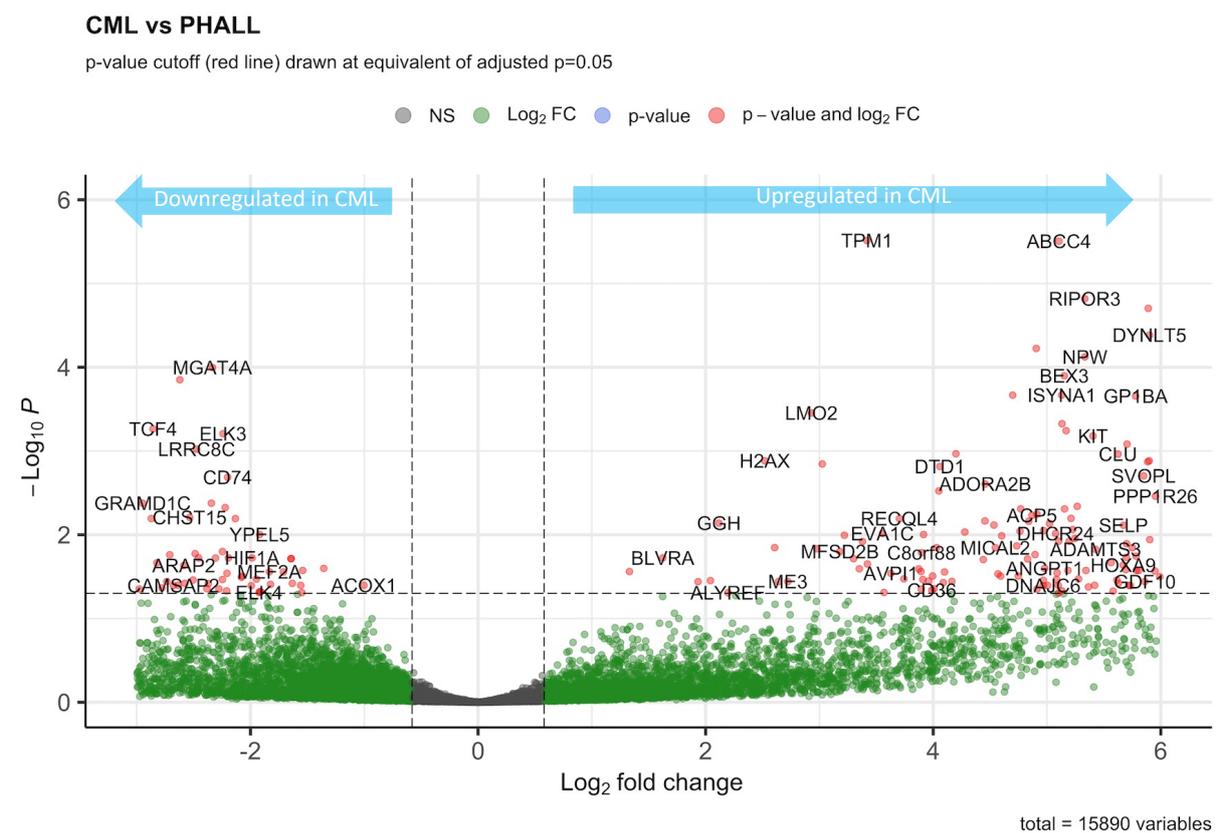


Figure 5.2.3: Differential expression volcano plot CML-LBC vs Ph+ALL. P-value cutoff of 0.05. Log₂ fold changes in adjusted P values for genes differentially regulated in CML-LBC vs Ph+ALL samples. Genes with minimal change (up or downregulated) are designated as non-significant, represented in grey and eliminated from further examination. Expression values for DEGs are displayed on the X axis by Log₂Fold change.

5.2.3 Results III: Processing DEGs from CML-LBC vs Ph+ALL comparison for gene set enrichment analysis (GSEA).

To further investigate this, gene set enrichment analysis (GSEA) was utilised. RNAseq and differential gene expression analysis yields many DEGs, however, to put this in a functional biological context, comparison of input gene set to each of the terms in gene ontology can be conducted using GSEA software. Data was prepared for GSEA by extracting statistically significant ($P_{adj} < 0.05$) genes and ordering DEGs by expression value (Log_2Fold), from highly upregulated to downregulated. This input gene set (dataset) was saved as a .rnk file for uploading to the GSEA software. Gene ontology analysis was performed by selecting terms (bins) in the gene ontology programme. Bins relevant to disease profile were selected from the Molecular Signatures Database (MSigDB), including 'immunologic signature gene sets (C7)', 'regulatory target gene sets (C3)', ontology gene sets (C5), oncogenic signature gene sets (C6) and cell type signature gene sets (C8) with a brief overview on gene set coverage and usage provided by MSigDB. Gene sets were selected based on disease relevance.

The regulatory target gene set (C3) collection was composed of 3725 individual gene sets and was selected to identify potential targets of regulation by transcription factors or microRNAs. The datasets within this collection were grouped by elements shared in non-protein encoding regions known or likely to contain regulatory elements in promoters or 3'-UTRs.

The gene ontology gene set collection (C5) was comprised of 15703 gene sets encompassing genes annotated by the same ontology term. The collection is categorised by gene sets associated with biological process, cellular component, molecular function and human phenotype ontology. This broad range of gene functions allow for a global investigation of genes highlighted through differential expression analysis.

Oncogenic signature gene sets (C6) were selected due to disease relevance in investigation of the haematological cancers in this project. This collection was comprised of 189 gene sets which represent signatures of cellular pathways commonly dysregulated in cancer and sourced from NCBI GEO and MSigDB internal profiling.

Finally, immunologic signature gene sets (C7) were selected due to the relevance of leukaemia to the function of leukocytes and control of immune system responses. As a collection of 5219 gene sets, this contained ImmuneSigDB studies of chemical and genetic immune system perturbations in human and mouse and VAX (vaccine response gene sets) curated by the Human Immunology Project Consortium (HIPC) detailing immunologic response to vaccinations in humans.

Expression values for DEGs were related to their phenotypic function and utilising GSEA deconvolutes large datasets by grouping functionally related genes. The analytical process required for GSEA has three steps:

1. Calculation of the enrichment score (ES). The ES represents the extent to which genes are over- or under-represented in the gene set;
2. Estimation of ES clinical significance. Statistical significance is calculated by a phenotypic-based permutation test to generate a null distribution for the ES. This null distribution is compared to generate the P value. This verifies the dependence of the gene set on the biologic phenotypic labels. Genes with a non-significant P value may also be included in GSEA;
3. Adjust for multiple hypothesis testing. Required when investigating a large number of genes, the enrichment scores are normalised, and a false discovery rate calculated.

Alternative methods of gene set enrichment analysis are available such as Simpler Enrichment Analysis (SEA) which assumes gene independence and uses a simpler mathematical approach to calculate t-test. However, this method may be too simple for larger RNAseq datasets wherein gene correlation has occurred, therefore the GSEA method outlined above was utilised.

5.2.4 Results III: GSEA highlighted disease-specific gene ontology terms for DEGs identified in the CML-LBC vs Ph+ALL comparison.

GSEA results were separated by contribution to enrichment results (genes with a 'yes' core enrichment value) and expression status (up or downregulated) (Fig 5.2.4). For genes positively regulated (upregulated with a positive Log2Fold score), 8/15 gene sets were upregulated. The classifications were as follows; circulatory system development, cell adhesion, tissue development, anchoring junction, defence response, proteolysis, TATAAA_TATA and AACTTT_unknown.

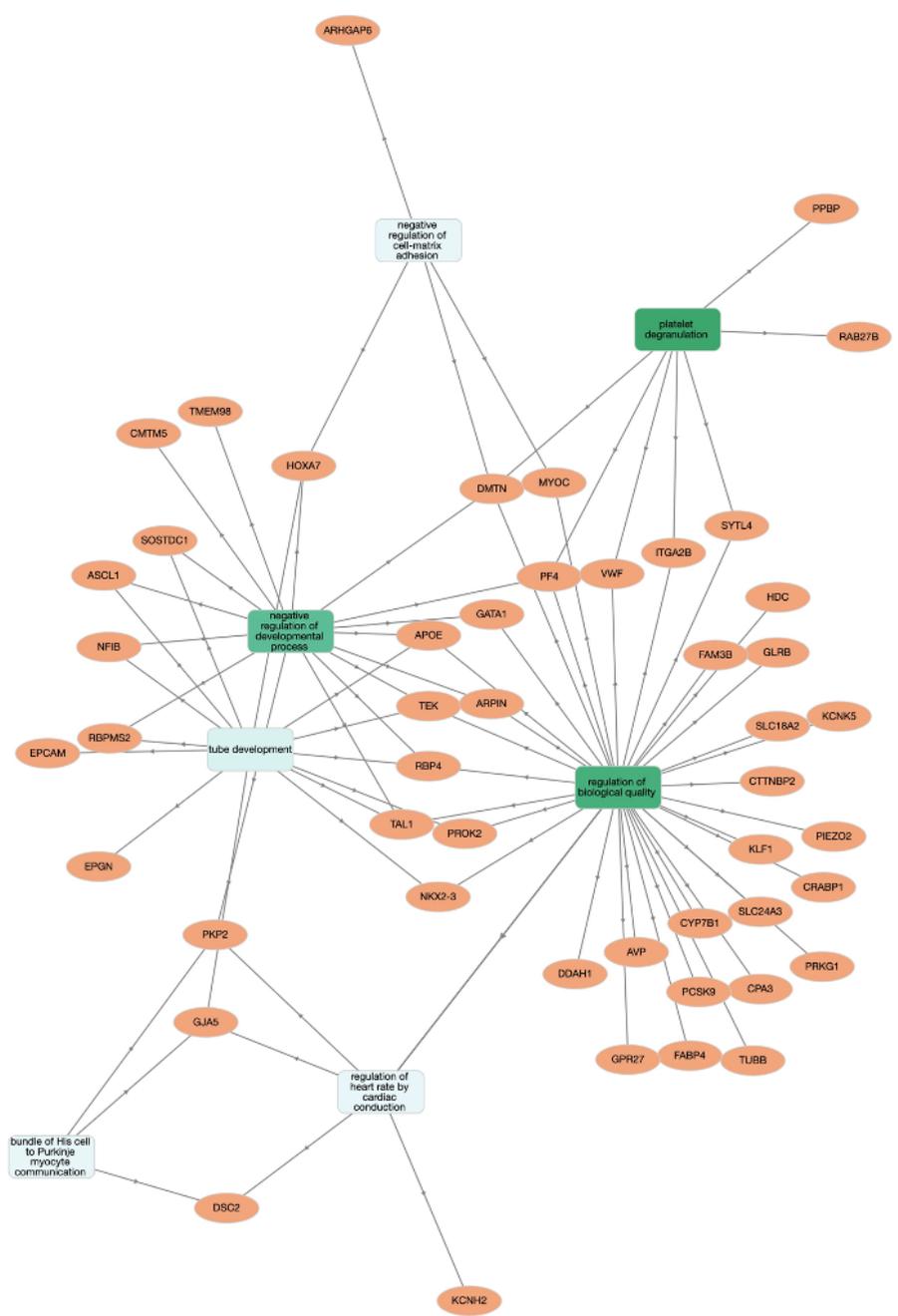


Figure 5.2.4: Gene ontology enrichment plot (CML-LBC vs Ph+ALL). Genes with positive Log2Fold changes and adjusted P value lower than 0.05 assembled into GO plot using Gonet.

5.2.5 Results III: Genes involved in circulatory system development and maintenance were differentially regulated between CML-LBC and Ph+ALL samples.

The circulatory system gene set encompasses genes involved in the progression of the circulatory system over time, from formation to maintenance of mature structure (GOC, 2022). Elucidating microenvironment alterations which impact on the movement of leukocytes may be of benefit to the understanding of disease function. With this, cell adhesion gene sets were investigated (Fig 5.2.5). Genes within this group have roles in attachment of cells to other cells or extracellular matrix by adhesion molecules. This allows for the investigation of cell-cell interaction and when combined with information in the circulatory system gene set, provides information on the trafficking and movement of leukocytes. Highlighted in both Ph+ALL and CML datasets, *ANXA1* (Annexin A1) is known to function as an anti-inflammatory phospholipid binding membrane-localised protein. The downregulation of *ANXA1* in the CML-LBC vs Ph+ALL dataset displays a loss of control of T cell activation signalling cascades, glucocorticoid-mediated innate immune responses, rearrangement of the actin cytoskeleton, cell polarisation and migration (Arcone et al., 1993, D'Acquisto et al., 2006, Ernst et al., 2004). Functionally this describes dysregulation of lymphocytes and improper response to external stimuli, a phenotype fitting in a disease model where lymphoblast formation and leukocyte function dysregulation can be observed.

NAME	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
row 0	ZFPM2	3	8.510199547	0.06527402	Yes
row 1	SORBS2	16	6.860509872	0.012603053	Yes
row 2	SNAI2	20	6.719389915	0.057204112	Yes
row 3	FZD4	23	6.418227196	0.109317586	Yes
row 4	CHI3L1	31	6.008688927	0.10175832	Yes
row 5	LOX	34	5.939679146	0.14834747	Yes
row 6	CDH2	45	5.402530193	0.10082373	Yes
row 7	PTPRM	49	5.271524906	0.12871075	Yes
row 8	FGF6	56	5.050828934	0.121083036	Yes
row 9	ACE	57	4.991694927	0.17870677	Yes
row 10	MCAM	72	3.907584906	0.069969445	No
row 11	GATA2	73	3.46874404	0.110012345	No
row 12	ANXA1	79	2.218467951	0.08067711	No
row 13	PRICKLE1	86	-4.032265186	0.06129118	No
row 14	THSD7A	103	-11.82535553	0.021977954	No

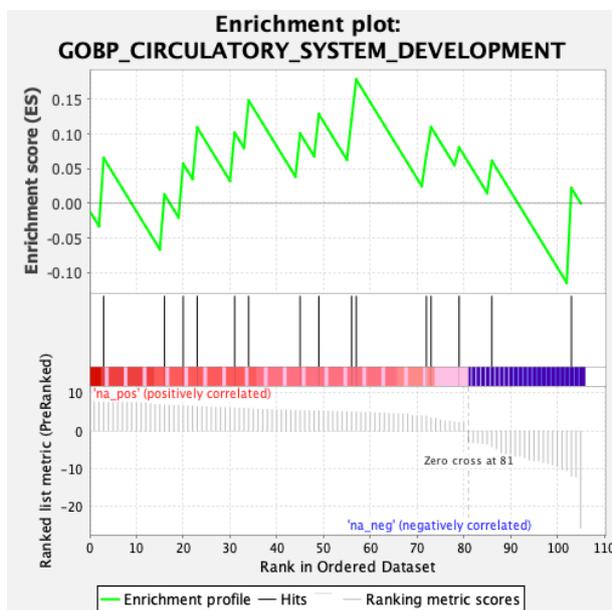


Figure 5.2.5: Differentially expressed genes associated with circulatory system development with accompanying GSEA enrichment plot. DEGs identified through differential expression analysis of CML-LBC vs Ph+ALL. NES= 0.6753.

5.2.6 Results III: Tissue development genes were differentially regulated between CML-LBC and Ph+ALL samples.

Following from investigation into circulatory and cell adhesion, genes associated with tissue development were highlighted by GSEA (Fig 5.2.6). Such genes had involvement in tissue generation from formation to mature structure (GOC, 2022). Genes which contributed most to enrichment results (genes with a 'yes' core enrichment value) were of disease relevance with *ZFPM2* as a haematopoietic transcription factor, *KRT8* as a member of the type II keratin family and responsible for maintenance of cellular structural integrity, signal transduction and cellular differentiation, the non-receptor protein kinase *SORBS2* (a member of the Abelson family) aiding in formation of the actin cytoskeleton, matrix metalloproteinase *MMP8* involved in tissue remodelling, metastasis and maintenance of inflammatory arthritis disease processes and *SNAI2*, a zinc finger transcription factor, involved in epithelial-mesenchymal transitions, displaying antiapoptotic functions and repressing E-cadherin in breast carcinoma. Taken together, these genes not only have a role in tissue remodelling and maintenance, but also in stem cell functions such as *ZFPM2* which maintains the undifferentiated state of embryonic and adult stem cells (Conta and Breitbart, 2010) and *SNAI2* which confers antiapoptotic functions and maintains stem-like features in normal and malignant cells via modulation of beta-catenin (Liu et al., 2023). The aberrant expression of *SNAI2* and *ZFPM2* are illustrative of LSC activity in CML. Additionally, *MMP8* was identified as contributing to enrichment results. *MMP8* has been identified as a marker for differential diagnosis between CML and leukaemoid reaction (LR). *MMP8* functions as a proteolytic enzyme secreted by neutrophils which plays a role in the destruction of the extracellular matrix, under pathologic CML conditions, such cells have both proliferative and functional impairment (Lin et al., 2004). Additionally, the role of *MMP8* in metastasis and inflammatory diseases such as arthritis is illustrative of leukocyte dysregulation, a hallmark of CML (Hsieh et al., 2021).

Further genes which enable cell-cell and cell-extracellular matrix interaction include *COL6A3* which encodes alpha 3 chain of type VI collagen (Jin et al., 2021), the breast cancer biomarker and actin filament bundle assembly gene *AIF1L* (Liu et al., 2018) and *MCAM*, encoding an adhesion molecule of the same name with observed roles in infiltration, apoptosis regulation and spread of ovarian cancer cells, leading to its use as a metastasis marker (Wu et al., 2012).

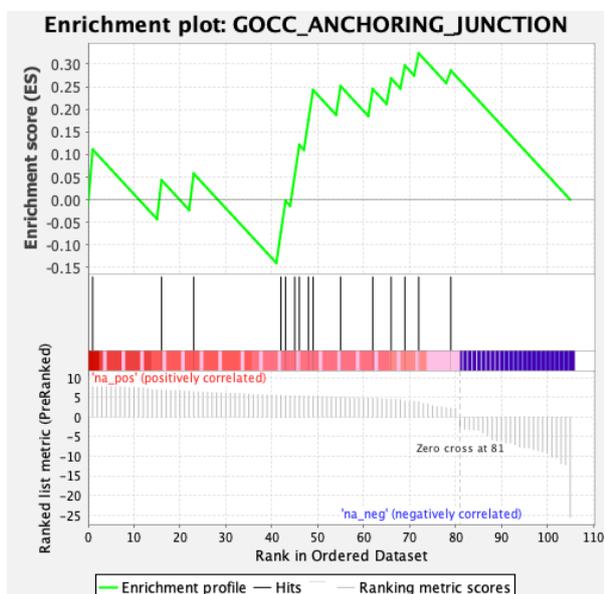
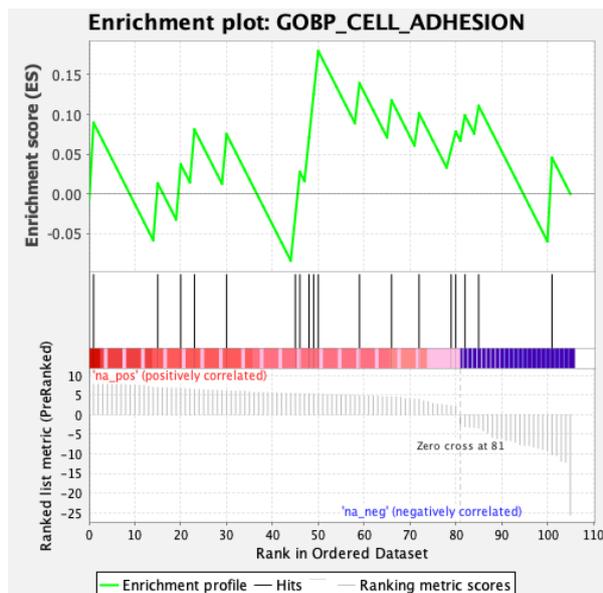
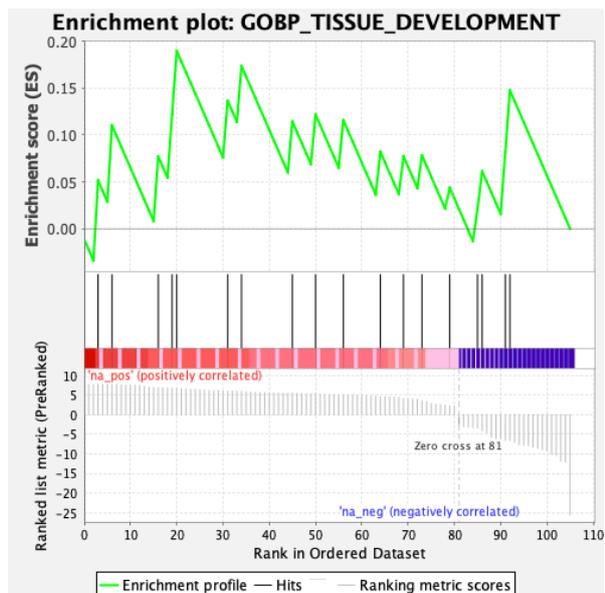


Figure 5.2.6: GSEA enrichment plots representing tissue development gene sets for DEGs upregulated in CML-LBC. DEGs identified through differential expression analysis of CML-LBC vs Ph+ALL. GOBP tissue development NES= 0.7607. GOBP cell adhesion NES= 0.7336. GOCC anchoring junction NES=1.1876.

5.2.7 Results III: Inflammatory immune response genes were upregulated in CML-LBC vs Ph+ALL bulk samples.

Dysregulation of leukocyte activity results in altered defence response, another area highlighted by GSEA (Fig 5.2.7). Genes contributing to enrichment results in this group have functions in serine protease enzyme inhibition with *SLPI* encoding a secreted inhibitor of neutrophil elastase (NE) by epithelial cells to prevent endogenous proteolysis (Sullivan et al., 2008) and *SERPINB4* which is highly expressed by tumour cells and inactivates granzyme M, a serine protease which increases tumour cell chemoresistance, colony-formation, cytokine secretion and invasiveness (Wang et al., 2015). In addition to a pro-tumorigenic function, *SERPINB4* also has a role in establishing and maintaining chronic autoinflammatory diseases such as psoriasis through the expression of the *SERPINB4*-derived autoantigen Pso p27 (Iverson et al., 2017). BPI was also indicated by GSEA, with its role in inflammatory immune response protection against gram-negative bacteria, the alterations to inflammation and tumour cell defence are clear in the genes represented in this dataset (The Human Protein Atlas, 2023).

As with altered defence response, genes which contribute to proteolysis were highlighted, again conveying dysregulation of immune response in the CML-LBC vs Ph+ALL comparison. *ST18*, a marker for poor outcomes in AML (Skou et al., 2021), modulates cytokine secretion by acting as a transcription factor regulating proapoptotic and proinflammatory gene expression in fibroblasts (Yang et al., 2008). With involvement in similar inflammatory diseases such as asthma, *TPSB2* (and *TPSD1*) encodes a mast cell tryptase resistant to all known endogenous protease inhibitors and has been identified in AML (Lamba et al., 2018). The *HP* gene encodes for haptoglobin, a plasma protein inducible by inflammation that aids in haemoglobin degradation, host antimicrobial defence response and, if mutated, can lead to the development of ahaptoglobulinaemia or hypohaptoglobulinaemia (Wang et al., 2001). Genes identified with circulation system-localised host defence functions (Wang et al., 2021) and homeostasis were *TFPI* and *ACE*, of which *TFPI* encodes a Kunitz-type protease inhibitor that regulates the formation of fibrin clots via regulation of the tissue factor dependent pathway of blood coagulation, and *ACE* in maintaining blood pressure regulation and electrolyte balance via angiotensin converting enzyme. *ACE* acts to increase blood

pressure via vasodilator bradykinin however, dysregulation of *ACE* can lead to cardiovascular disease or chronic inflammatory conditions such as psoriasis (Issa., 2022). An additional gene with immune response function highlighted by GSEA is *SPSB4* which enables ubiquitin ligase-substrate adaptor activity, regulates cell repulsive responses and its overexpression has been identified in human malignancies such as colorectal cancer (Okumura et al., 2017).

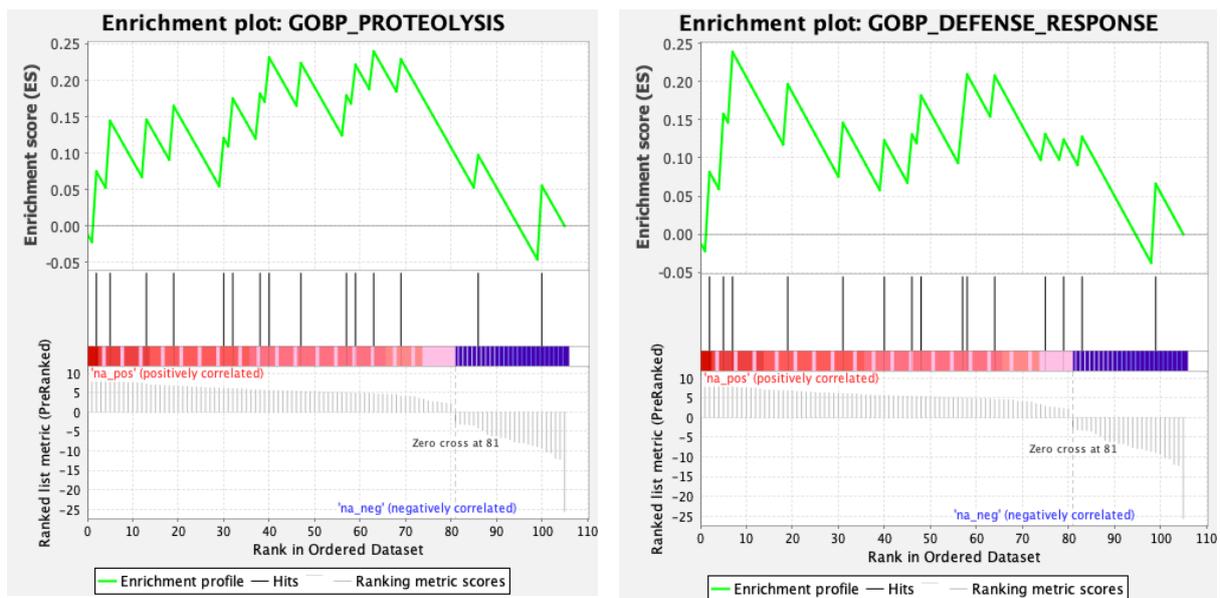


Figure 5.2.7: GSEA enrichment plots representing inflammatory immune gene sets for DEGs upregulated in CML-LBC. DEGs identified through differential expression analysis of CML-LBC vs Ph+ALL. GOBP proteolysis NES= 0.9059. GOBP defense response NES= 0.8839.

5.2.8 Results III: Regulation of mRNA encoding and gene targeting genes were upregulated in CML-LBC vs Ph+ALL bulk samples.

The TATAAA_TATA gene set contains genes which correspond to common regulatory motifs in promoters and 3' untranslated regions (3'UTRs). Genes are aligned with this gene set due to having at least one occurrence of the highly conserved motif M51 TATAAA in the regions spanning 4kb centred on their transcription starting sites (+/- 2kb) (Fig 5.2.8). Such genes likely have motifs which function as transcription factor binding sites, involvement in post-transcriptional regulation, microRNA (miRNA) encoding genes or likely miRNA binding sites. Current predictions are that at least 20% of the human genome is regulated by miRNAs, hence evolving knowledge on gene transcription regulation in disease context is of great importance to better understand gene expression and potential for therapeutic exploitation or diagnostic purposes (Xie et al., 2005).

Similarly, the AACTTT_unknown gene set represents genes which have at least one occurrence of the highly conserved motif M17 AACTTT in the regions spanning 4kb centred on their transcription starting sites (+/-2kb). Though highly conserved, this motif is not associated with any known transcription factor binding site (Xie et al., 2005). Similarly to the TATAAA_TATA gene set, genes classified under AACTTT_unknown are believed to have involvement in transcription factor binding sites, post-translational regulation, miRNA encoding and miRNA target genes.

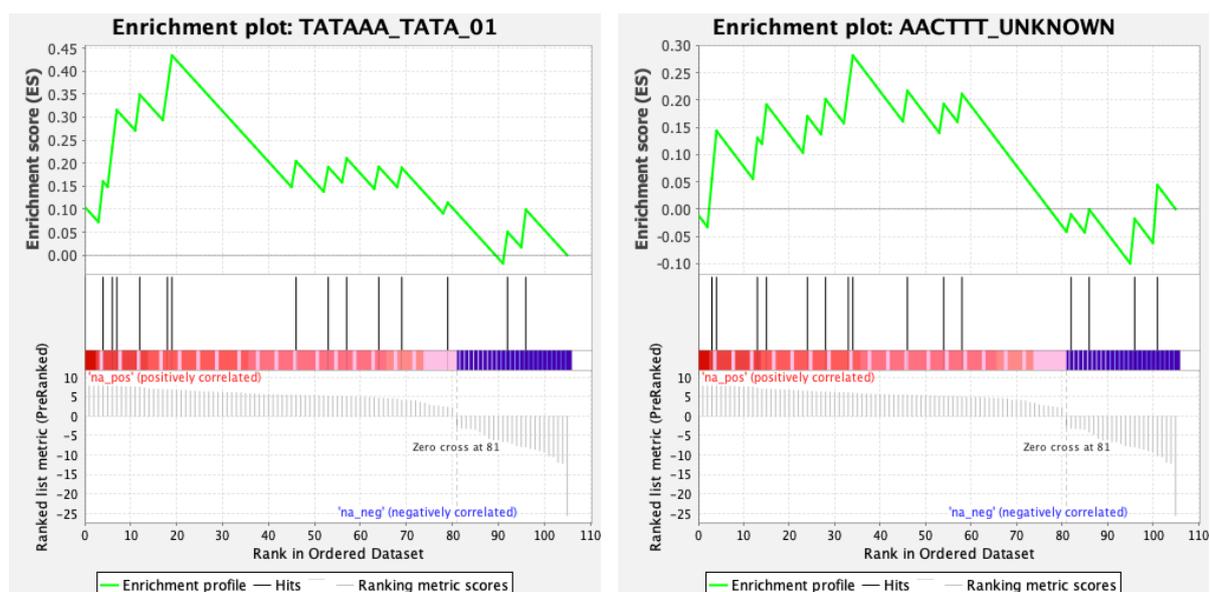


Figure 5.2.8: GSEA enrichment plots representing mRNA encoding and gene targeting gene sets for DEGs upregulated in CML-LBC. DEGs identified through differential expression analysis of CML-LBC vs Ph+ALL. TATAAA TATA 01 NES= 1.5985. AACTTT unknown NES= 1.0675.

5.2.9 Results III: Differential gene expression analysis displayed disease-specific transcriptional profiles in CML-LBC and Ph+ALL samples.

DEGs between CML-LBC (Cluster 1) and the entire Ph+ALL cohort (Cluster 2 and 3) were identified and used for gene set enrichment analysis to identify functionally relevant gene groupings and relations to known disease phenotypes (Fig 5.2.2, 5.2.8 & 5.2.9). DEGs were selected for GSEA based on statistical significance ($p < 0.05$) with the resultant gene list containing 106 genes differentially regulated between CML-LBC and Ph+ALL. Categories of gene functions highlighted by GSEA included circulatory system development, cell adhesion, tissue development, anchoring junction, defence response, proteolysis, TATAAA_TATA and AACTTT_unknown, with the latter two groups containing genes thought to play a role in mediation of mRNA encoding and gene targeting. Broadly, the functions of these DEGs were related to immune response, inflammation, cancer cell growth and maintenance, interactions between cells and the microenvironment and interactions between cells. The immune cell contexture has been well investigated in CML, as has the interaction between CML LSCs and the bone marrow microenvironment (BMM). CML LSCs are known to form a reciprocal relationship with components of the BMM whereby LSCs may be maintained and leukemogenesis boosted by the BMM and additionally, the LSCs are able to alter the BMM toward their requirements and away from homeostatic maintenance of normal healthy stem and progenitor cells (Houchmand et al., 2019). Thus, the interactions between leukaemic cells and the bone marrow microenvironment are important for establishing and maintaining CML and in order to do so, alternate transcriptional profiles must be established, as has been evidenced above. A number of genes had similar functions and could therefore be grouped into multiple ontology groups, such as those which have a role in both tissue development and maintenance, cell adhesion, anchoring junctions and defence response (Fig 5.2.5 & 5.2.6). This further emphasises the difference of cell-cell and cell-microenvironment interactions between CML-LBC and Ph+ALL samples.

A number of DEGs identified had functions in maintenance of LSCs, a known CML disease feature (Holyoake and Vetrie., 2017). Genes such as *ZFPM2* and *SNAI2* both have roles in maintenance of the undifferentiated state of LSCs, thereby conferring stem cell qualities to these transformed progenitors (Conta and Breitbart, 2010 and Liu et al., 20203). Previous

5.3.1 Results III: Ph+ALL samples with detectable BCR::ABL1 in stem and progenitor cells had a different transcriptional profiles to Ph+ALL samples with no BCR::ABL1 at the stem and progenitor level.

The existence of leukaemic stem cell activity in Ph+ALL has previously been posited, however with little in-depth investigation, the Ph+ALL LSC remains elusive (Thomas, 2012). However, as previously outlines, the presence of a CML-like population confers an alternative cell of BCR::ABL1 origin outside of the lymphoid progenitor with the fusion event likely occurring closer to the apex of haematopoiesis in an MPP or HSC population. As previously outlined in Chapter 4, stem and progenitor cells were isolated from Ph+ALL bulk BMNC samples at diagnosis and during post induction treatment and the presence of BCR::ABL1 assessed using FISH (Fig 5.3.1). BCR::ABL1 was detected in stem and progenitor cells (HSC and MPP) of 16/21 diagnostic and 9/16 PI Ph+ALL samples. These findings support the existing hypothesis that ALL is a heterogeneous leukaemic type and confirms that this heterogeneity extends to within the Ph+ALL subtype (Thomas, 2005). Additionally, findings suggest that BCR::ABL1 presence in early haematopoietic stem and progenitor cells alone is inadequate for establishment of the CML-like Ph+ALL subtype and downstream transcriptional differences are likely the origin of the novel subtype with myeloid involvement.

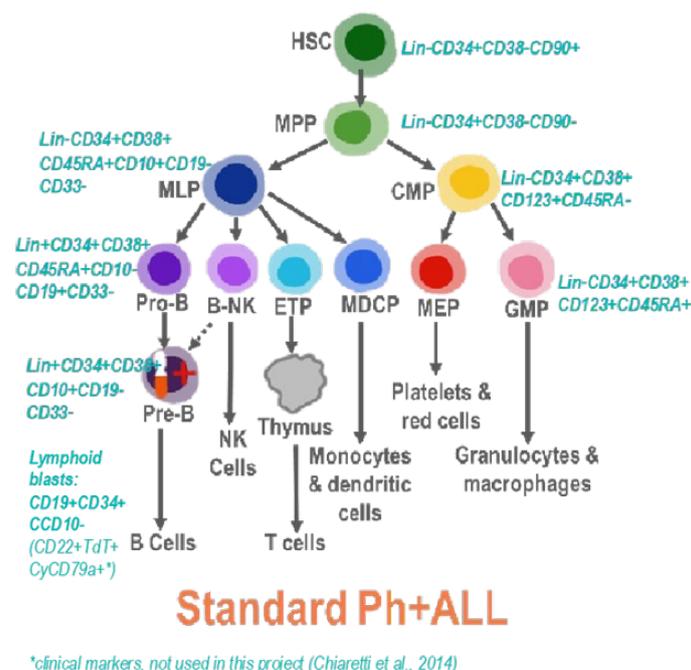


Figure 5.3.1: Haematopoietic population surface marker expression for FACS. Surface marker expression profile used for cell sorting based off those used clinically.

In order to investigate this heterogeneity further, the transcriptional profiles of Ph+ALL samples with and without detectable BCR::ABL1 in stem and progenitor cells were compared with the goal of uncovering genes differentially regulated between both groups and the potential for LSC transcriptional signature detection in the Ph+ progenitor group.

5.3.2 Results III: Transcriptional heterogeneity exists between Ph+ALL samples with detectable BCR::ABL1 at the stem and progenitor level and Ph+ALL samples without.

Again, overall transcriptional profiles of Ph+ALL samples were visualised by PCA. Investigation of Clusters 1-3 in figure 5.2.1 with the addition of BCR::ABL1 presence in Ph+ALL stem and progenitor cells displays that all samples in Cluster 3 had BCR::ABL1 identified by FISH within an HSC or MPP population. Furthermore, 3 samples in Cluster 2 had BCR::ABL1 identified by FISH within immature progenitor populations (ranging from 17-50% (Table 5.1.3)). These findings suggest there is a heterogeneity as defined by DEGs within samples that have BCR::ABL1 in the HSC and MPP populations. (Fig 5.3.2).

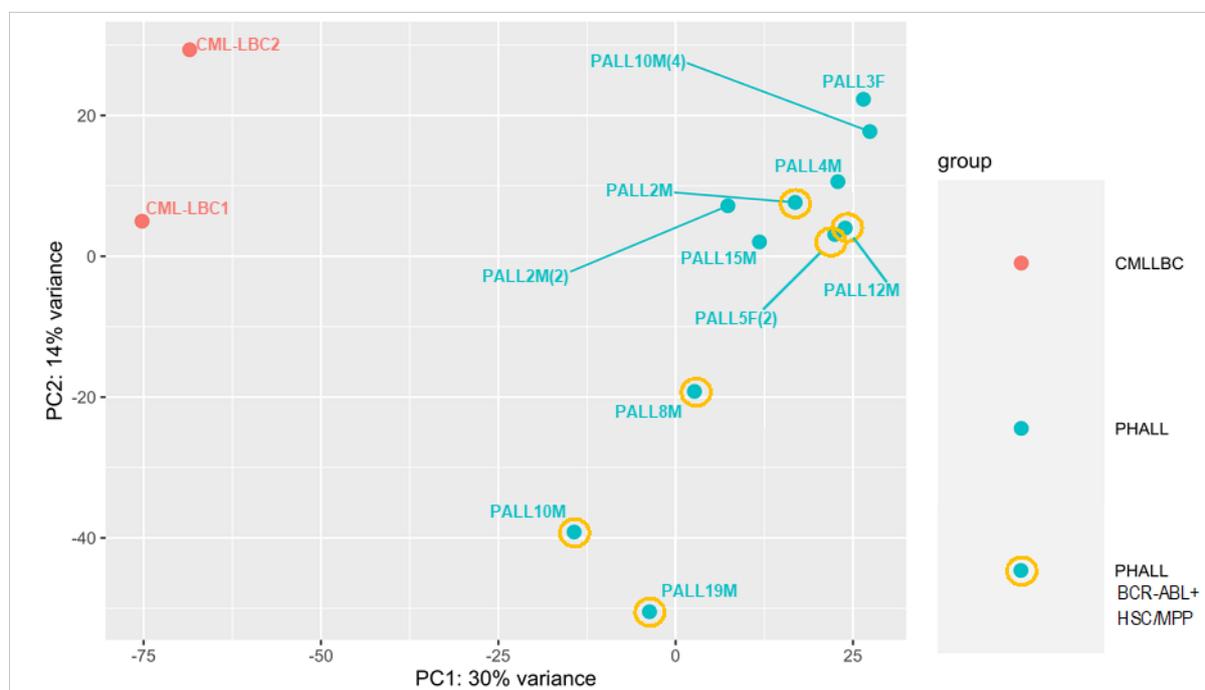


Figure 5.3.2: Principal component analysis plot, CML-LBC (N=2) vs Ph+ALL (N=11), clusters annotated. Ph+ALL samples with detectable BCR::ABL in stem and progenitor cells highlighted in yellow.

In order to further examine transcriptional similarities between the FISH positive and negative groups, a PCA plot containing Ph+ALL samples alone was generated (Fig 5.3.3). Of 11 Ph+ALL BMNC bulk samples total, 6 had detectable BCR::ABL1 in stem and progenitor cells and 5 samples had a zero FISH score (Table 5.1.3). As seen in figure 5.3.2, the overall transcriptional profile of Ph+ALL samples were heterogeneous, with tight clusters signifying transcriptional similarity not observed. However, a separation between Ph+ and Ph- samples can be seen with BCR::ABL1 negative sample locations predominating the left side of the graph and the

Ph positive samples predominating middle-right graph areas. This suggests that, despite absence of compact clusters, overall transcriptional expression differs between both groups.

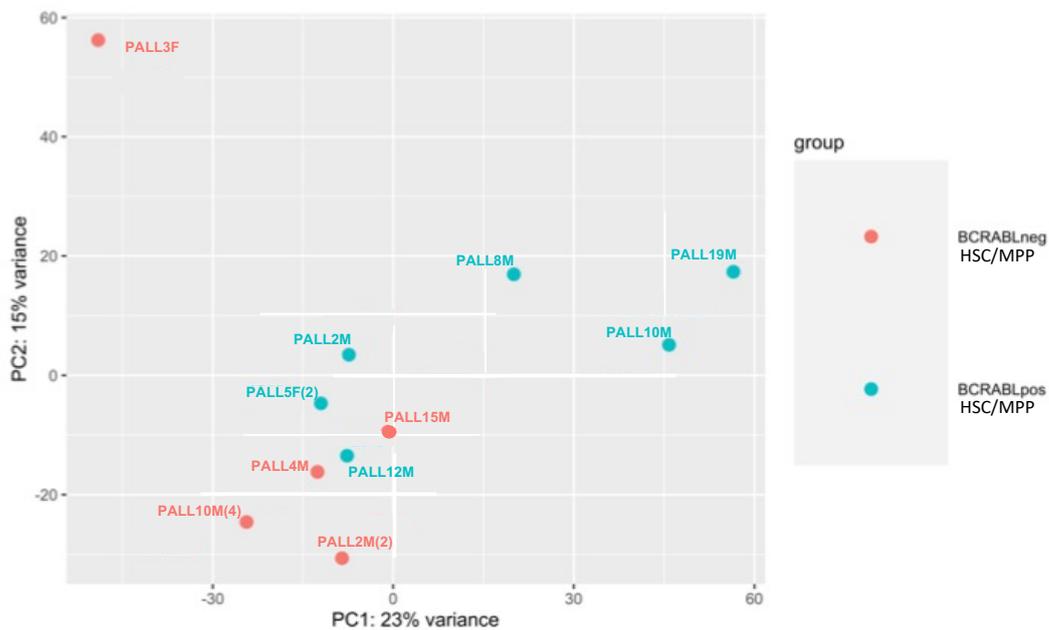


Figure 5.3.3: Principal component analysis plot, BCRABLneg (N=5) vs BCRABLpos (N=6), clusters annotated. Ph+ALL samples with detectable BCR::ABL in stem and progenitor cells coloured in blue, samples with no detectable BCR::ABL in HSC or MPP cells coloured in red.

In order to further identify transcriptional differences between Ph+ALL samples with observable BCR::ABL1 in stem and progenitor cells and those without, differential gene expression analysis was performed (Fig 5.3.4). Utilising statistical cut off ($P < 0.05$), the expression values for DEGs were expressed in Log2Fold changes.

BCRABLpos vsBCRABLneg

p-value cutoff (red line) drawn at equivalent of adjusted p=0.05

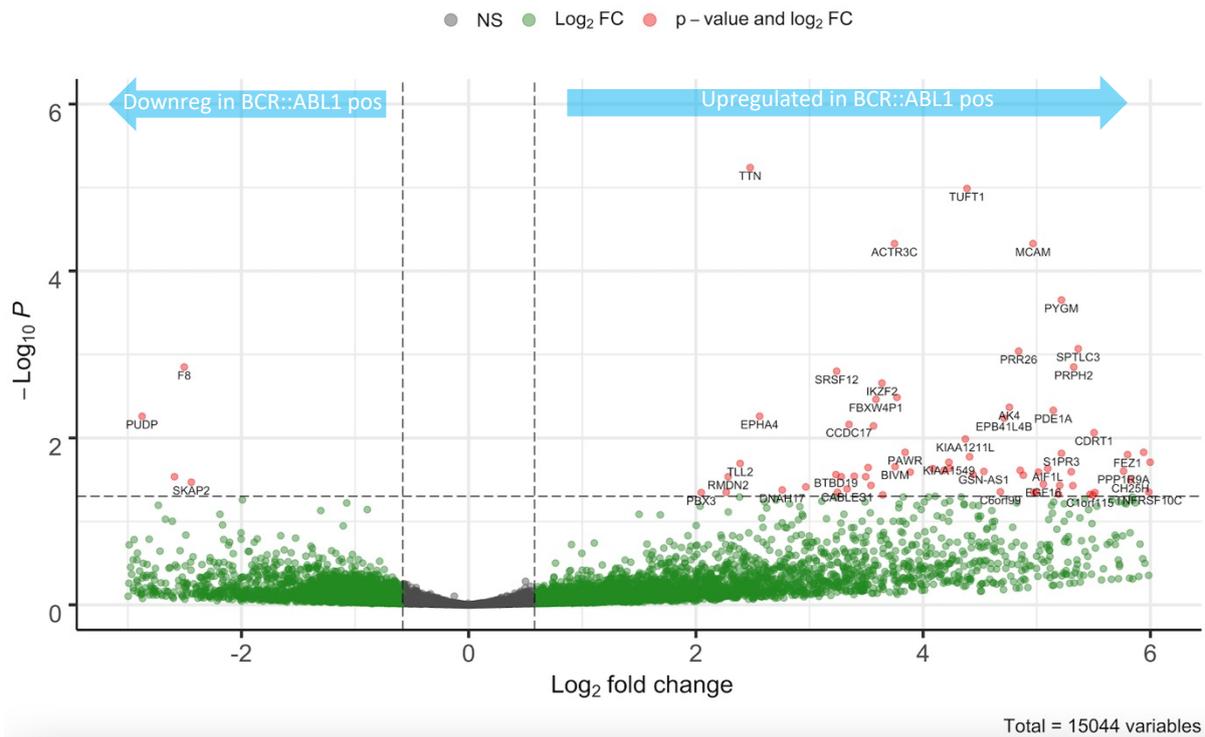


Figure 5.3.4: Differential expression volcano plot, BCRABLneg (N=5) vs BCRABLpos (N=6), clusters annotated. Log₂ fold changes in adjusted P values for genes differentially regulated in BCRABLpos vs BCRABLneg samples.

To allow for biological and disease contextualisation, DEGs were ranked based on Log₂Fold expression value (from most upregulated to most downregulated) and gene ontology analysed using GSEA as outlined previously. Broad gene sets differentially regulated between BCR::ABL1 positive vs negative Ph+ALL stem and progenitor cells were as follows; regulation of multicellular organismal development, locomotion, generation of neurons, neurogenesis, TGGAAA NFAT Q4 01, cell projection organisation, CTTTGA LEF1 Q2, cell adhesion, tissue development, AACTTT unknown and cell motility (Fig 5.3.5).

5.3.3 Results III: CNS developmental and functional genes were upregulated in Ph+ALL samples with detectable BCR::ABL1 at the stem and progenitor level.

Broad biological processes highlighted included the development and maintenance of neurons, cell trafficking and early developmental processes. Central nervous system (CNS) involvement in Ph+ALL progression and relapse has long been studied and amendments to treatment strategies made as a result of this research (Lazarus et al., 2006) (Fig 5.3.6). CNS infiltration by leukaemic cells represents oft observed route for ALL relapse but one which mechanisms remain elusive (Lenk et al., 2020). Observations of this relapse route has resulted in alterations to treatment, such as prophylactic intrathecal chemotherapy as a preventative measure rather than treating established ALL in CNS relapse (Brown, 2023). As a result, identification of genes related to CNS development and maintenance in the context of Ph+ALL and potential subtypes, is potentially of benefit for further elucidation of mechanisms of CNS-induced ALL relapse. While CNS relapse was reported for one sample (PAL10M(2)), a larger cohort of patients would allow for a better determination of CNS development related with regards to CNS relapse. DEGs associated with CNS in Ph+ALL samples with positive FISH scores were heavily involved in neuronal and nervous system development, such as *PO4F3*, *FSTL4* and *FEZ1*. In addition, a number of genes identified were integral to global early development in addition to CNS development such as *PRPH2*, *PPP1R9A*, *HEY1*, *DPYSL3* and *BMPR1B*, with *BMPR1B* activity being implicated in the persistence of LSCs in CML via Jak2 signalling (Jeanpierre et al., 2021).

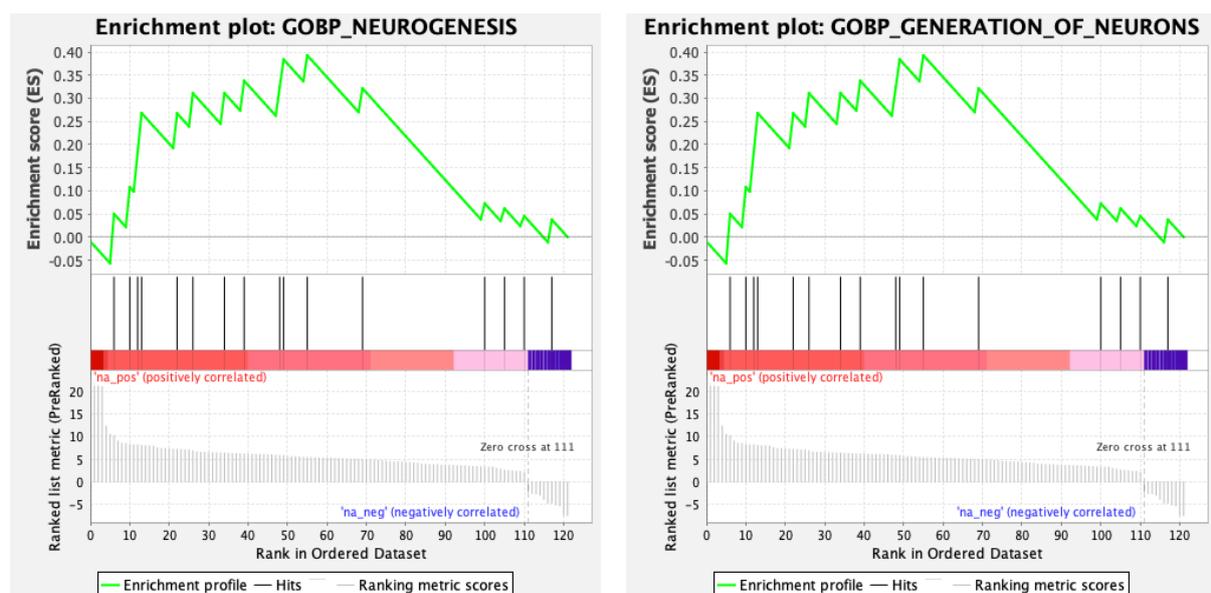


Figure 5.3.6: GSEA enrichment plots representing CNS developmental and functional gene sets for DEGs upregulated in BCRABLpos Ph+ALL. DEGs identified through differential expression analysis of BCRABLpos vs BCRABLneg Ph+ALL. GOBP neurogenesis NES= 1.2521. GOBP generation of neurons NES= 1.2546.

5.3.4 Results III: An LSC-like gene expression profile was detected in Ph+ALL samples with Ph+ stem and progenitor cells.

Alterations to the bone marrow microenvironment are essential for the maintenance of LSC activity and therefore, understanding how cells may interact with other cells or with the bone marrow niche itself, is beneficial to identifying or understanding how BCR::ABL1 presence in stem and progenitor cells may impact on disease activity. Genes associated with interactions with extracellular matrix proteins were identified, such as *SNAI2*, a transcription factor involved in generation and migration of neural crest cells and believed to repress E-cadherin transcription in breast carcinoma, and *LOX* which functions in the crosslinking of collagens and elastin (Fig 5.3.7 & 5.3.8). Genes enabling microenvironmental cellular interactions were also identified such as *CH25H* which encodes a membrane protein with functions in lipid and cholesterol metabolism via hydroxylation of hydrophobic substrates. Cell trafficking and locomotion genes were identified as DEGs in Ph+ALL samples with detectable BCR::ABL1 in stem and progenitor cells. Genes such as *CCL20* and *DOCK1* have functions in cytokinesis, immunoregulatory and inflammatory processes, with *CCL20* being implicated in enhancing tumour cell migration, proliferation and tumour microenvironment remodelling and *DOCK1* as a marker for poor outcome in AML (Kadomoto et al., 2020 and Lee et al., 2017). Contextualisation of malignancy-related functions was further supported with the identification of *HOXB9*, a transcription factor involved in cell proliferation and differentiation which has been identified as being aberrantly expressed in AML blast cells resulting in maintenance of proliferation (Heinrichs et al., 2005).

With LSC presence and activity being, as of yet, unverified in Ph+ALL, investigation of DEGs between samples with detectable BCR::ABL1 in stem and progenitor cells and those without, may elucidate stem cell-related genes which may be indicative of LSC activity. Gene set enrichment analysis highlighted DEGs with roles in early development such as *BMPRI1B*, *EFNA5*, *KL*, *PTPRD* and *SOX11*.

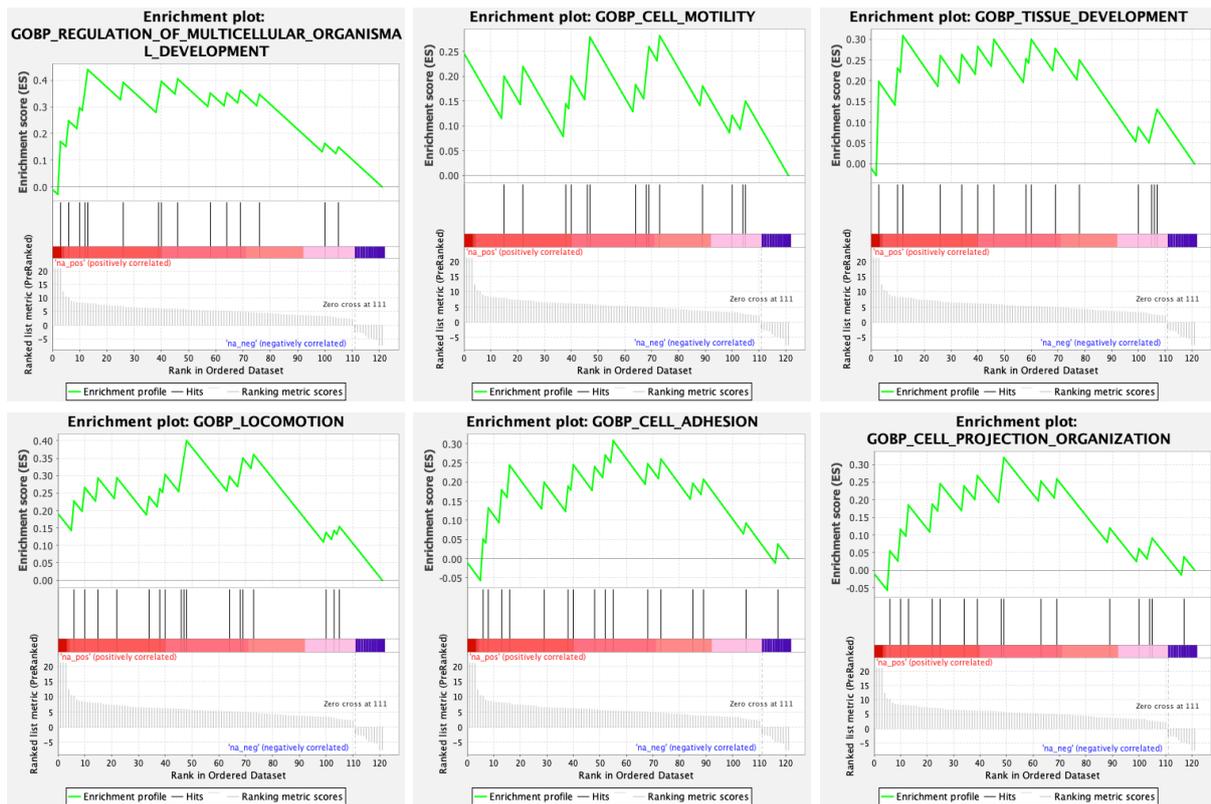


Figure 5.3.7: GSEA enrichment plots representing LSC-related gene sets for DEGs upregulated in BCRABLpos Ph+ALL. DEGs identified through differential expression analysis of BCRABLpos vs BCRABLneg Ph+ALL. GOBP regulation of multicellular organismal development NES= 1.3843. GOBP cell motility NES= 0.8727. GOBP tissue development NES= 0.9675. GOBP locomotion NES= 1.3285. GOBP cell adhesion NES= 0.9817. GOBP cell protection organisation NES= 1.0248.

SOX11 encodes a member of the SOX (SRY-related HMG-box) family of transcription factors with roles in embryonic development regulation and cell fate determination. *SOX11* activity has been implicated in diseases such as neurodevelopmental disorders, osteoarthritis and a variety of cancers, with importance in embryogenesis and being largely absent in adult tissue (Tsang et al., 2020). Of note, aberrant *SOX11* activity has been investigated in haematological malignancies such as adult *de novo* AML and mantle cell lymphoma (MCL). Overexpression of *SOX11* has been observed in adult AML, with the extent to which *SOX11* is overexpressed being highlighted as a potential AML prognostic marker (Tosic et al., 2018). Additionally, *SOX11* overexpression can be observed in cancer stem cells (CSCs) (including MCL) and is associated with aggressive behaviour in MCL (Sureda-Gómez et al., 2022). Independent of other high-risk features, *SOX11* overexpression has been evidenced to be associated with poor overall survival via the reduction in apoptosis-related genes and normal stem cell features. Such MCL CSCs displayed increased growth, survival, aberrant stemness and chemoresistance, therefore presenting itself as a potential target in MCL and displaying a critical role in CSC activity.

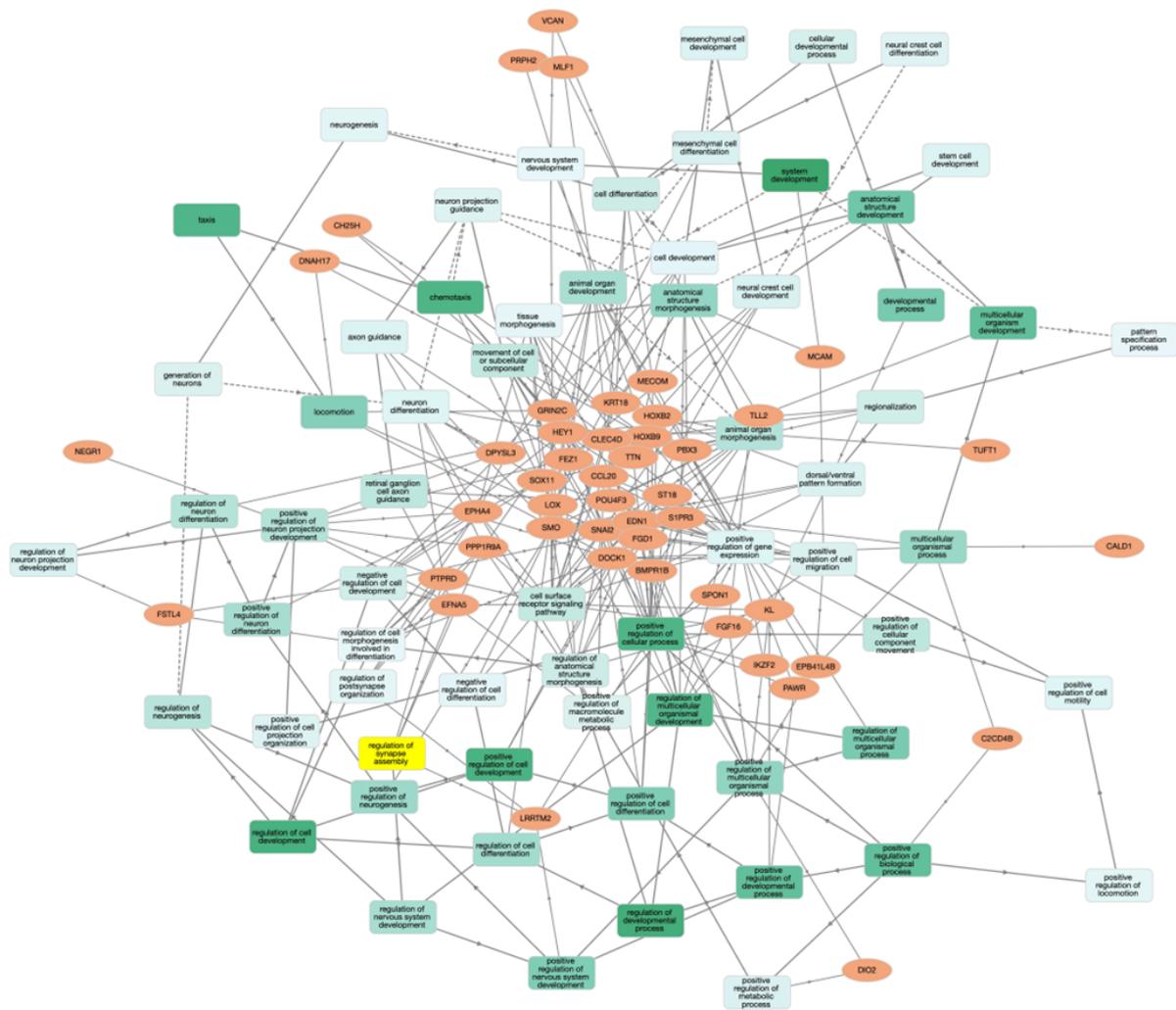


Figure 5.3.8: Gene ontology enrichment plot for genes upregulated in BCR::ABL1 positive HSCs/MPPs (Ph+ALL) vs BCR::ABL1 negative HSCs/MPPs (Ph+ALL). Upregulated genes used in GSEA analysis depicting GO enrichment networks. (Tool: GOnet)

The *BMPR1* gene encodes for a member of the bone morphogenic protein (BMP) receptor family of transmembrane serine/threonine kinases, the ligands of which, are members of the TGF-beta superfamily and which are involved in bone formation and embryogenesis. *BMPR1* has previously been identified as being upregulated in leukaemic stem cells in CML with treatment insensitive LSCs relying on the BMP pathway for survival via the BMP4 autocrine loop (Jeanpierre et al., 2021). The effect of this was evidenced by *BMPR1B*⁺ LSCs displaying a co-activated *Smad1/5/8* and *Stat3* pathway and a maintenance of proliferation status during TKI treatment. In support of previous observations of the importance of cell adhesion-related genes in samples with Ph⁺ stem and progenitor cells, *BMPR1B*⁺ cells adhering to stromal cells were the only LSCs observed to have a quiescent status, a feature of LSC activity and known mechanism for treatment evasion and disease relapse (Van Gils et al., 2021).

Ephrin-A5, encoded by *EFNA5*, is a member of the ephrin gene family, the largest subfamily of receptor protein-tyrosine kinases and have been implicated in regulating migration and positioning during early developmental events. Ephrin signalling features have been implicated in the remodelling of the LSC niche with *EphA3*, a member of the Eph receptor tyrosine kinase family, being selectively expressed on LSCs but not healthy HSCs (Slape, 2014). The importance of the bone marrow niche on maintenance of the LSC population has been demonstrated, with a bi-directional effect of LSCs remodelling the bone marrow microenvironment to support growth (Houshmand et al., 2019). This displays the importance of Eph signalling on maintenance of a leukaemic stem cell population in the bone marrow niche.

The *KL* gene encodes for a protein called Klotho, a type-1 membrane protein, related to beta-glucosidases with roles in the process of antiaging and prevention of bone loss via antioxidation, antisenesence, anti-autophagy and signalling pathways such as insulin-like growth factor and Wnt (Bian et al., 2015). While much of the research into Klotho activity relates to aging-associated diseases, dysregulation of Klotho signalling has been implicated in the loss of senescence and stem cell maintenance, both features of LSC activity.

PTPRD encodes for a member of the protein tyrosine phosphatase (PTP) family, a group of signalling molecules known to regulate cell growth, differentiation, mitotic cycle and oncogenic transformation, with *PTPRD* representing a receptor-type PTP. *PTPRD* has previously been investigated in paediatric AML, a leukaemic type with high relapse rates (30-40%) and well-documented LSC activity (Thomas and Majeti, 2017). Previous investigations of *PTPRD* in paediatric AML described a significant downregulation of *PTPRD* in AML LSCs compared to healthy HSC controls, with the dysregulation of inflammatory and immune networks in LSCs believed to be resultant from this perturbation (Depeter et al., 2020).

5.4.1 Results III: GSEA displayed the differential regulation of few GO terms between Cluster 1 and Cluster 3.

In order to investigate the differences between Ph+ALL Cluster 3 and CML Cluster 1, differential gene expression analysis was performed as described previously (Fig 5.4.1 & 5.4.2). Genes differentially expressed between Cluster 1 and Cluster 3 were analysed using gene set enrichment analysis to identify GO terms defining the transcriptional differences between the CML samples and the Ph+ALL sub cluster. Only 3 GO terms were identified representing genes differentially regulated between both Clusters: locomotion, TGGAAA NFAT and AACTTT unknown. Alteration in locomotion, cell trafficking and genes involved in cellular interaction with the microenvironment such as *CCL20*, *MCAM*, *DOCK1*, *EFNA5* and *POU4F3* indicate a difference in both cell-cell and cell-microenvironment interaction. As previously discussed, LSC-induced alterations to the bone marrow microenvironment are bi-directional with LSCs altering the BM niche to further support the maintenance of leukaemic stem cells. The differential expression of the above genes may signify a disease-specific alteration to the bone marrow between CML and the Ph+ALL subgroup.

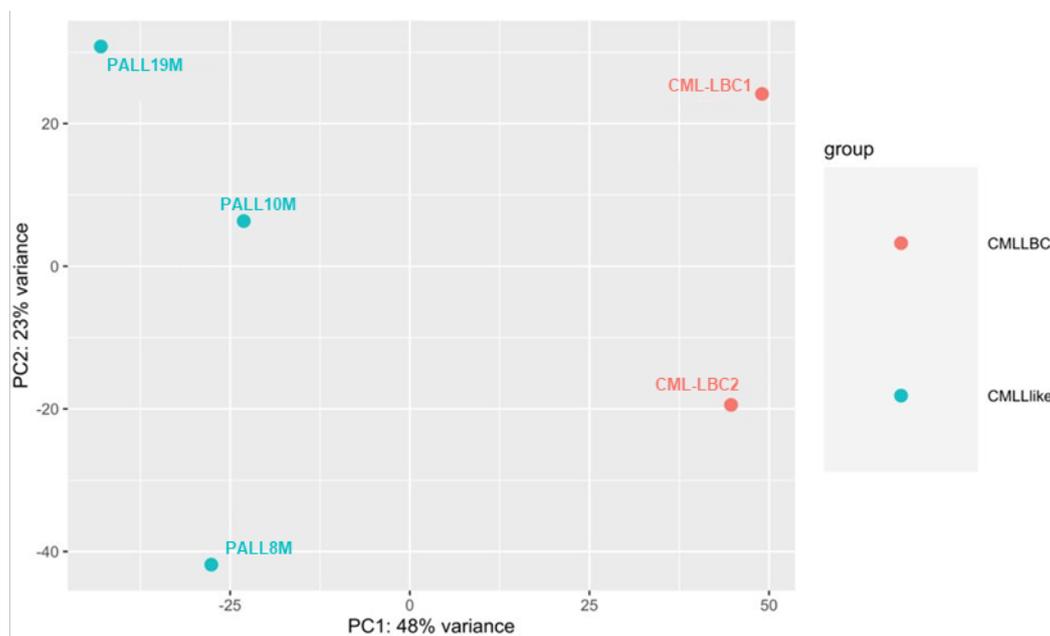


Figure 5.4.1: Principal component analysis plot, Cluster 3 (N=3) vs Cluster 1 (N=2).
Sample IDs annotated on plot.

5.4.2 Results III: Cluster 1 and Cluster 3 shared transcriptional similarity but displayed differential regulation in microenvironmental response genes.

The presence of differentially expressed genes is expected when comparing samples from two different leukaemic types, however the identification of a small number of GO terms is indicative of transcriptional similarity between Cluster 1 and Cluster 3 (Fig 5.4.3). This displays that relatively few genes are differentially regulated at a statistically significant level between these CML and Ph+ALL samples. These results indicate that this transcriptional similarity between Cluster 1 and 3 may be due to shared disease phenotypes, as would be predicted when comparing CML-LBC samples with samples designated as CML-like Ph+ALL. Proportions of Ph+ HSPCs in Cluster 2 and Cluster 3 were similar (average 28.3% and 26% respectively; range 17- 50% and 5- 50% respectively), further supporting the proposed theory that BCR::ABL1 fusion in HSPCs alone is inadequate for establishing the CML-like subtype and transcriptional variation may be an essential factor.

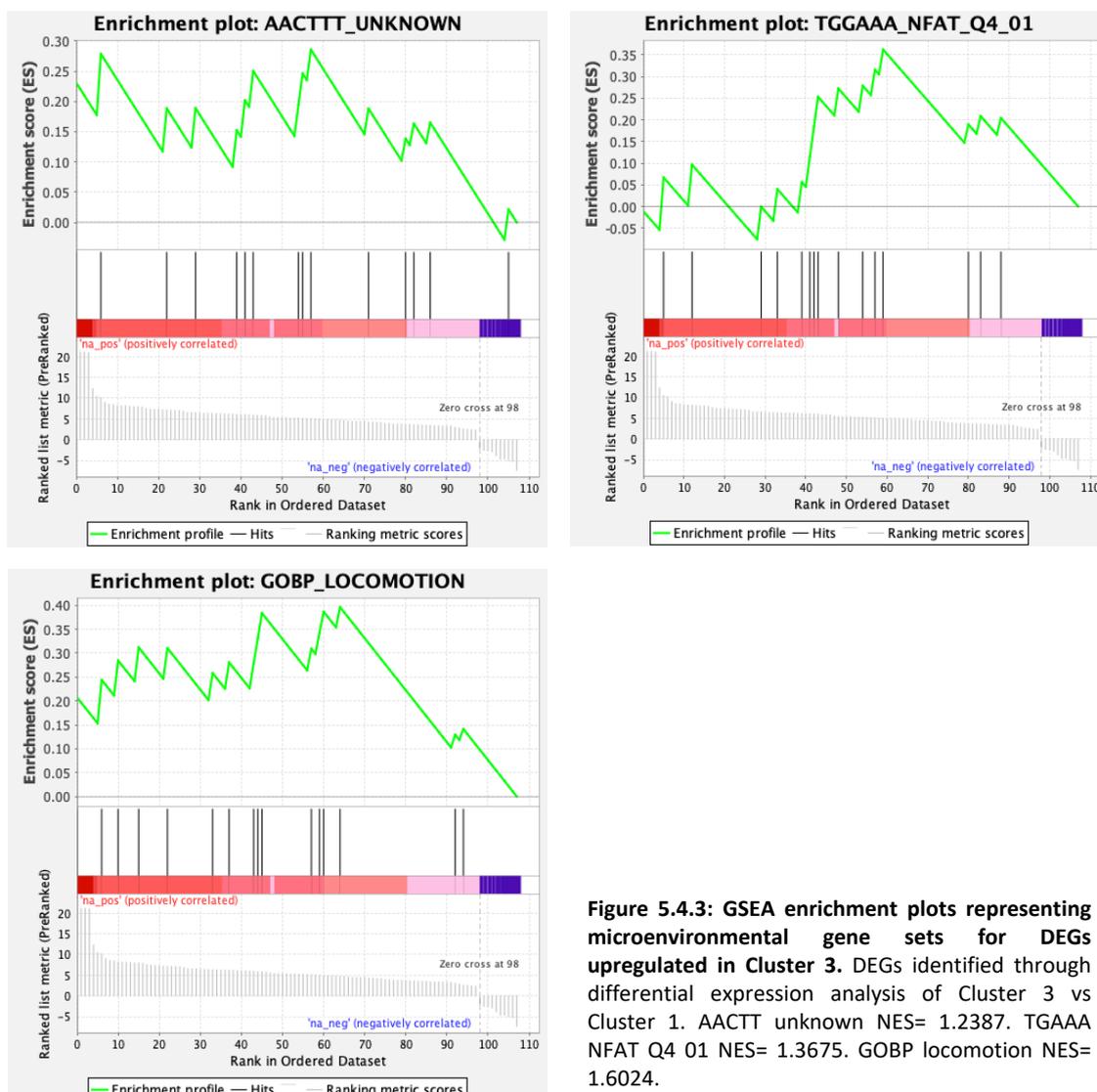


Figure 5.4.3: GSEA enrichment plots representing microenvironmental gene sets for DEGs upregulated in Cluster 3. DEGs identified through differential expression analysis of Cluster 3 vs Cluster 1. AACTT unknown NES= 1.2387. TGAAA NFAT Q4 01 NES= 1.3675. GOBP locomotion NES= 1.6024.

5.5.1 Results III: Cluster 2 and Cluster 3 displayed markedly different overall transcriptional profiles.

Identification of Cluster 3 during visualisation of overall transcriptional profiles illuminated a transcriptionally separate subgroup of Ph+ALL (Fig 5.2.1). Upon inspection of clinical outcomes, it was observed that these patients had poorer clinical outcomes than the rest of the Ph+ALL cohort with a higher incidence of ALL progression, relapse and death (Table 5.1.3 & Fig 5.2.2). Patients belonging to Cluster 3 also had distinctly different clinical features such as higher MRD, white cell count at diagnosis and BCR::ABL1 detectable in stem and progenitor cells by FISH. These patients therefore represented a subgroup of Ph+ALL samples with distinct overall transcriptional differences and clinical features.

In order to define DEGs which underpin transcriptional differences between the standard Ph+ALL samples and the newly observed subgroup, Cluster 3 and Cluster 2 were compared (Fig 5.5.1).

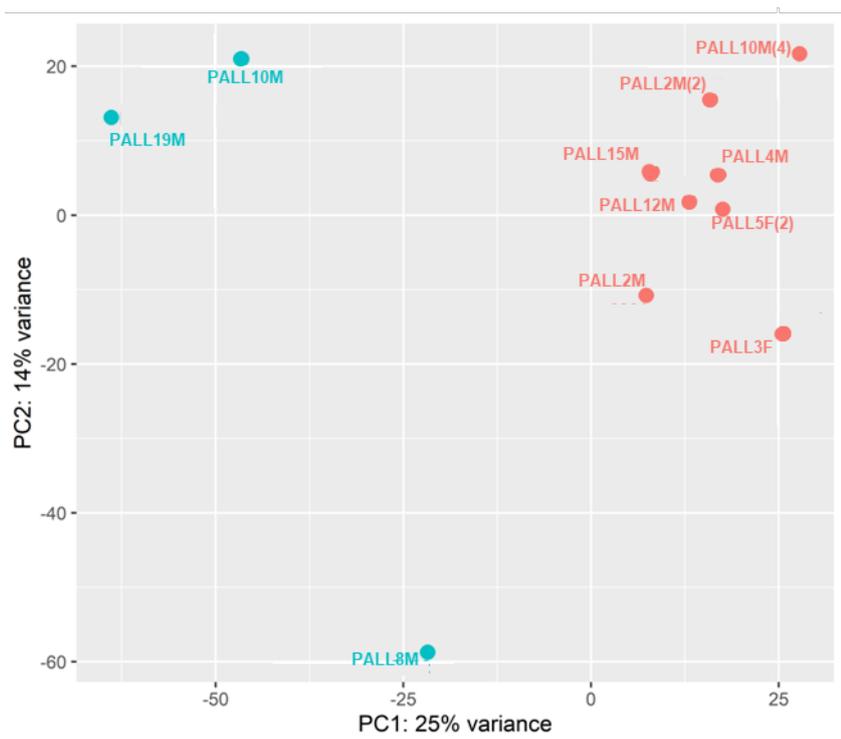


Figure 5.5.1: Principal component analysis plot, Cluster 3 (N=3) vs Cluster 2 (N=8). Sample IDs annotated on plot.

The resultant PCA plot generated to visualise overall transcriptional differences between Cluster 2 and Cluster 3 supported the initial observation of transcriptional disparity seen in Fig 5.2.1. It was observed that samples belonging to Cluster 2 highlighted in red (Fig 5.5.1) clustered tightly in the upper right quadrant of the PCA plot with 2 samples from Cluster 3

occupying the upper left quadrant. Sample PALL8M was distal to all other samples which indicated a differential overall transcriptional profile from the rest of the Ph+ALL cohort despite being designated as belonging to Cluster 3.

Genes identified as differentially regulated between Cluster 2 and Cluster 3 were visualised as a volcano plot (Fig 5.5.2). Of note, the gene XIST appears to be significantly downregulated in Cluster 3, however, this is likely due to sex differences between patients as XIST is the master regulator of X-chromosome inactivation and therefore would be highly upregulated in female patients and downregulated in males (GeneCards.org, Weizmann Institute of Science., 2024). This allowed for GSEA to be performed by isolation of statistically significant genes and ranking by Log₂Fold gene expression value. Gene ontology functions highlighted from differentially expressed genes in Cluster 3 vs Cluster 2 included anchoring junction, cell adhesion, proteolysis, response, tissue development, circulatory system development, TATAAA TATA and AACTTT unknown. The number of GO terms highlighted by GSEA displays a noteworthy difference in transcriptional expression between the Ph+ALL clusters.

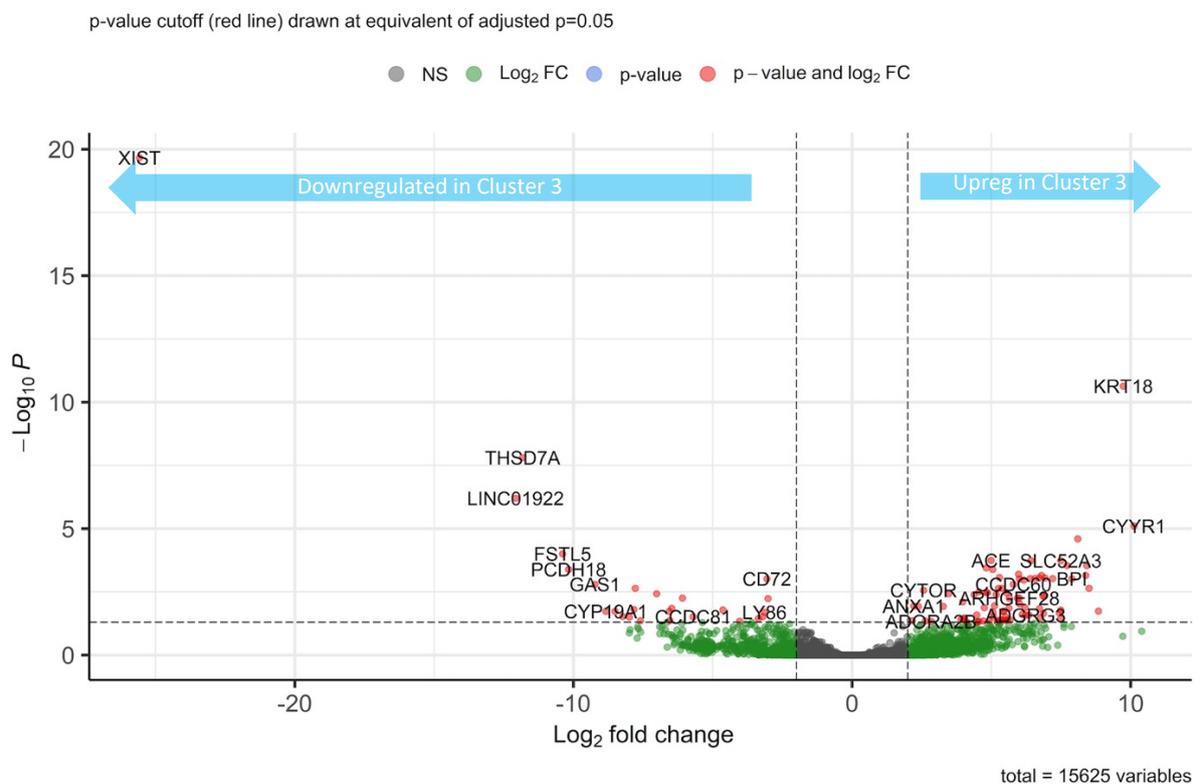


Figure 5.5.2: Differential expression volcano plot, Cluster 3 (N=3) vs Cluster 2 (N=8), clusters annotated. Log₂ fold changes in adjusted P values for genes differentially regulated in cluster 3 vs cluster 1.

5.5.2 Results III: Cell adhesion genes were differentially regulated between Ph+ALL Clusters 2 and 3.

As previously discussed, alteration in expression of genes associated with cell adhesion displays an alteration to mechanisms of cell-cell and cell-microenvironment interactions (Fig 5.5.3). Highlighted genes further supported the heterogeneity of the proposed CML-like Ph+ALL subtype that may be represented in Cluster 3. *CADM1*, the cell adhesion molecule originally identified as a tumour suppressor in lung cancer, has been highlighted as a marker for primary adult T-cell leukaemia and exerts an effect by increasing cell-cell adhesion, allowing for increased T-ALL organ infiltration (Nakahata and Morishita, 2012). Genes implicated in other acute leukaemia include *RIMS2* and *PVR* which have been identified as being differentially regulated in AML. *RIMS2* encodes a presynaptic protein which, upon interaction with *RAB3*, promotes normal neurotransmitter release and has been identified as a somatic variant in AML (Kakadia et al., 2018). *PVR* encodes a transmembrane glycoprotein belonging to the immunoglobulin superfamily, recently identified as a novel immune checkpoint which is found to be upregulated in AML and associated with poor outcomes, treatment-induced antibody blockade of *PVR* exerts anti-leukaemic effects and may be a promising treatment option in AML (Stamm et al., 2018).

In addition to differential regulation of AML-associated genes, *PTPRM* was also highlighted by GSEA and has roles in CML pathology. The protein encoded by *PTP* is a member of the protein tyrosine phosphatase family, a group of signalling molecules which regulate cell growth, differentiation, mitotic cycle and have been implicated in oncogenic transformation. *PTP* has been investigated in the context of CML due to the rise of tyrosine kinase inhibitor targeting of the tyrosine kinase BCR::ABL. *PTP* has been observed to antagonise tyrosine kinase activity and impact on TKI sensitivity. Hence, *PTP* presence in CML cells are important for the response to TKI therapy (Drube et al., 2018).

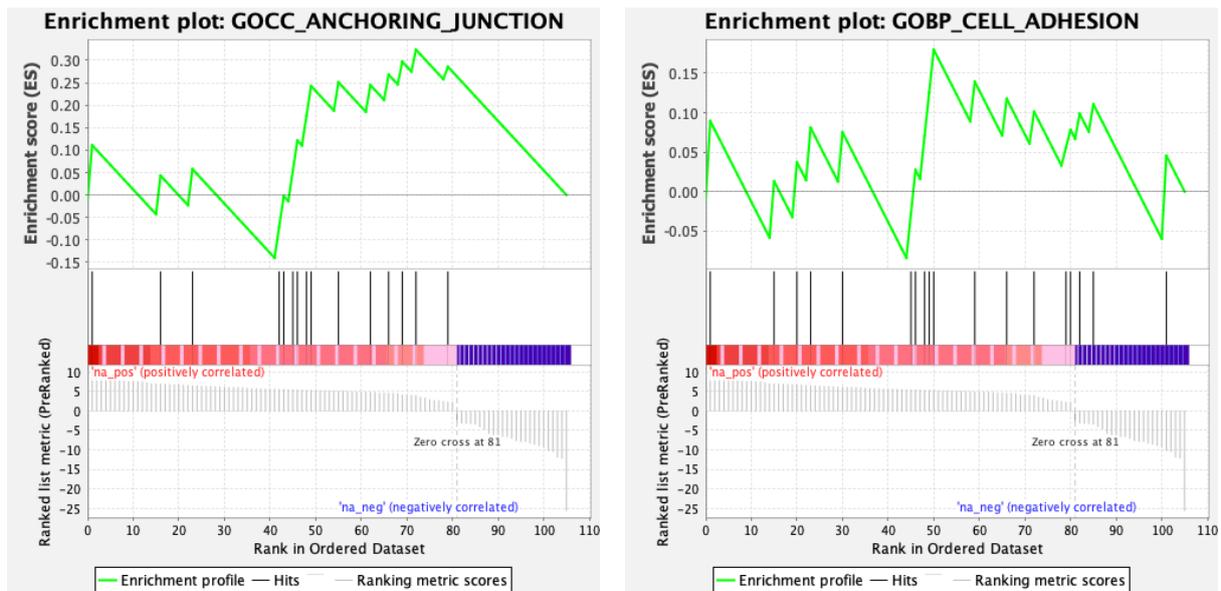


Figure 5.5.3: GSEA enrichment plots representing cell adhesion gene sets for DEGs upregulated in Cluster 3. DEGs identified through differential expression analysis of Cluster 3 vs Cluster 2. GOCC anchoring junction NES= 1.1272. GOBP cell adhesion NES= 0.6836.

5.5.3 Results III: Immune response genes were differentially regulated between Cluster 2 and Cluster 3.

Additional GO classifications identified by GSEA in Cluster 3 compared to Cluster 2 were those of immune-mediated defence response (Fig 5.5.4). Genes associated with regulation of inflammatory responses, both adaptive and inducible, were differentially regulated such as *SLPI*, *TPSB2*, *SPSB4*, *HP*, *TPSD1*, *COL6A3*, *ACE* and *TFPI*. Additionally, *SLPI*, *SPON1* and *COL6A3* demonstrate the role of extracellular matrix adaption to mediation or and epithelial protection from inflammatory responses. Other immune response genes previously implicated in multi-tissue cancer, and haematological malignancies were also identified such as *SERPINB4*, *ST18*, *MMP8* and *BPI*. This displays the importance of alterations to the immune response in both mediating inflammation and establishment of malignancy.

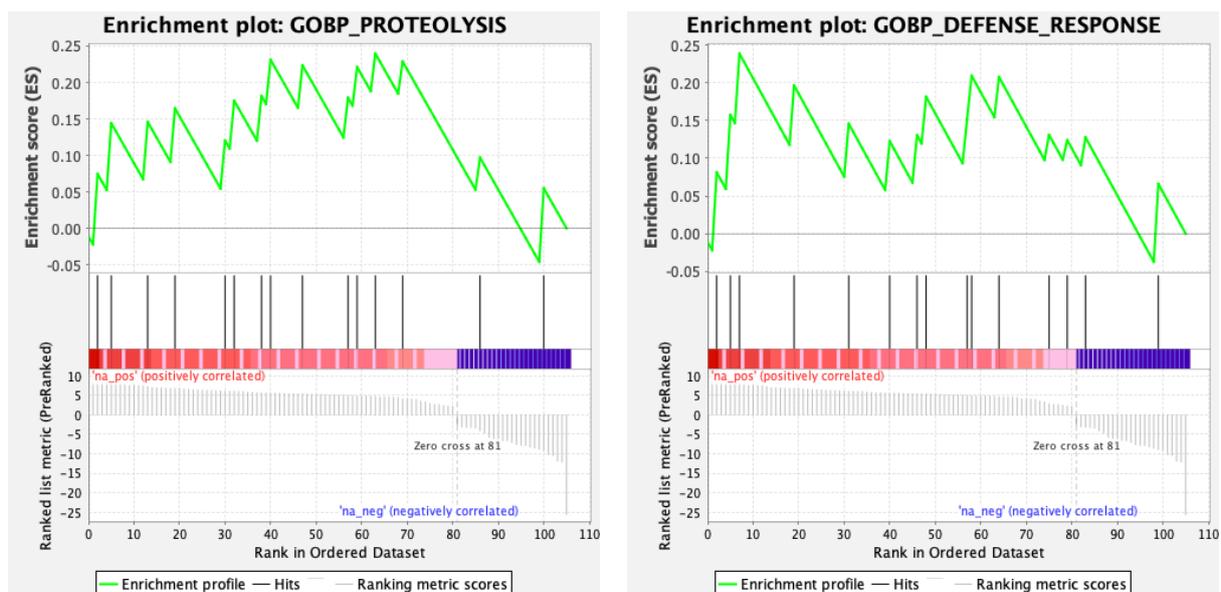


Figure 5.5.4: GSEA enrichment plots representing immune response gene sets for DEGs upregulated in Cluster 3. DEGs identified through differential expression analysis of Cluster 3 vs Cluster 2. GOBP proteolysis NES= 0.9493. GOBP defense response NES= 0.8989.

5.5.4 Results III: Cluster 2 and Cluster 3 displayed difference in the regulation of early developmental genes.

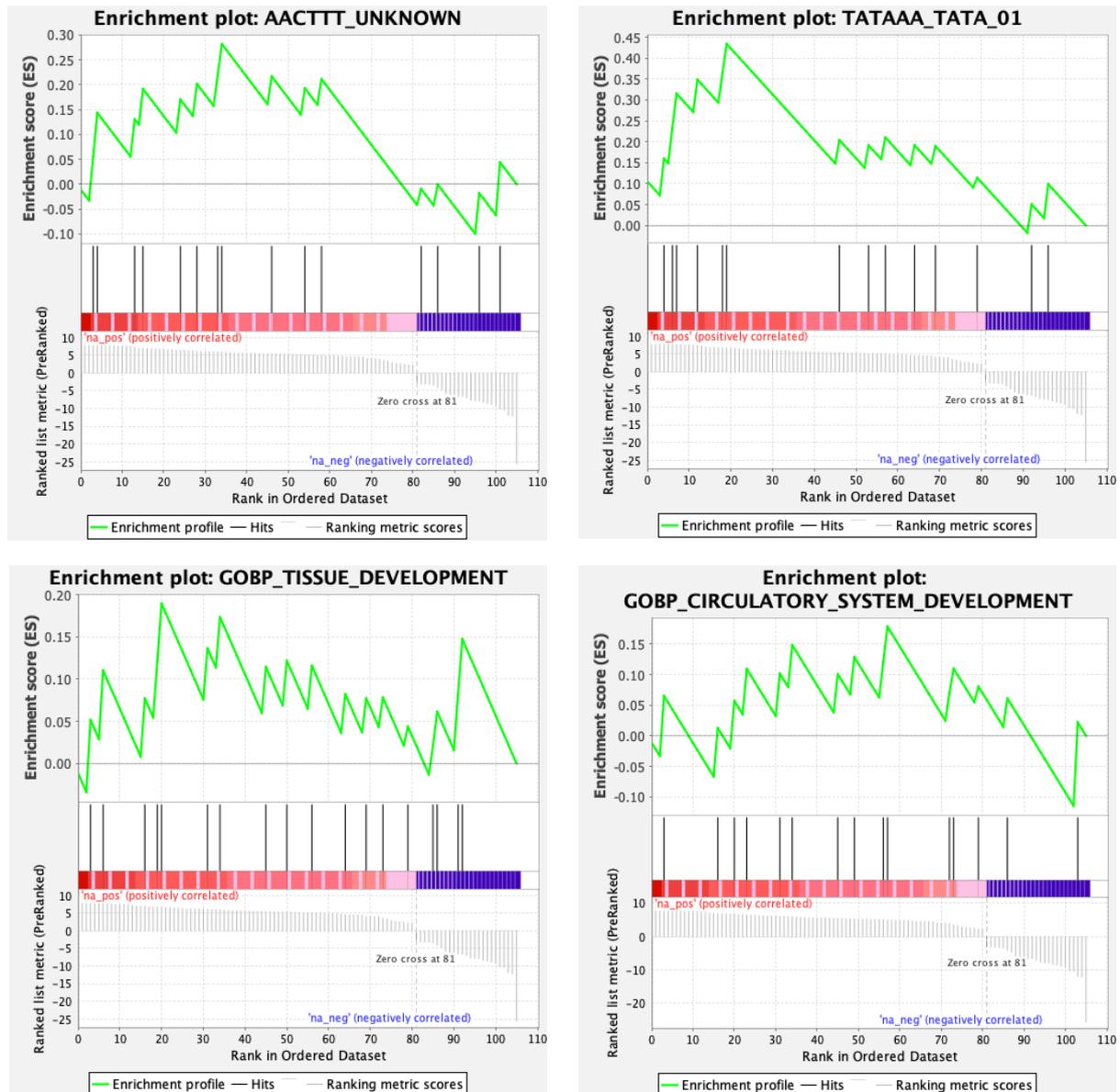


Figure 5.5.5: GSEA enrichment plots representing early developmental gene sets for DEGs upregulated in Cluster 3. DEGs identified through differential expression analysis of Cluster 3 vs Cluster 2. AACTTT unknown NES= 1.1376. TATAAA TATA 01 NES= 1.7117. GOBP tissue development NES= 0.7553. GOBP circulatory system development NES= 0.6698.

Another group of GO terms with great relevance to leukaemia initiated by early stem and progenitor cells are early development genes (Fig 5.5.5). The presence of genes commonly associated with maintenance of embryonic stem cell programmes (such as *CDH2*, *PRDM*, *LOX* and *FZD4*) were identified as differently regulated between Cluster 2 and 3 (Leszczyński et al., 2020). The upregulation of such genes in non-embryonic samples suggests an increased stem cell presence in Cluster 3.

5.6 Results III: RNAseq analysis of CML-LBC and Ph+ALL samples identified a transcriptionally and clinically distinct subgroup of Ph+ALL samples.

Overall transcriptional profiles of the CML-LBC and Ph+ALL samples investigated by RNAseq displayed an expected difference between CML and Ph+ALL. However, transcriptional profiling also revealed a smaller subgroup of Ph+ALL patients (termed 'Cluster 3') which differed transcriptionally from the rest of the cohort. Additionally, across the entire Ph+ALL cohort (Cluster 2 and Cluster 3), transcriptional heterogeneity was observed between samples which contained BCR::ABL1 positive HSCs and MPPs (measured by FISH- Chapter 4), and samples with no detectable t(9:22). DEGs and clinical considerations in samples belonging to Cluster 3 (ie incidence of death/relapse and detection of Ph+ HSC/MPPs) highlighted that these patients may represent the CML-like subtype. As previously discussed, without data from multiple MRD modalities (Ig/TCR gene rearrangement and *BCR::ABL1* copy number), CML-like Ph+ALL status cannot be confirmed using the methods presented by Hovorkova et al (2017). Hence, as there is no definitive gene panel for the CML-like subtype, patients could not be identified as CML-like from analysis outset using gene expression.

	Gene	Ph+ALLvsCML		Cluster3vsCluster1		Cluster3vsCluster2		BCRABLPOS vs BCRABLNEG	
		L2F	padj	L2F	padj	L2F	padj	L2F	padj
Upreg	ATF4	-0.5206624	0.8224058	-0.1093936	0.9534023	0.580515	0.997753	0.7535991	0.567676
	CCL8			2.262895	0.7205731	5.553026	0.3921323	4.012181	0.6660106
	CCND1	1.29047	0.8663087	1.255203	0.7053086			-1.547379	0.8033373
	FLI1	0.3295827	0.9129607	-0.1433694	0.9388738	-0.6292324	0.9978805	-0.1539163	0.9642365
	FLT3	2.588767	0.3368496	2.31354	0.07699777	-0.3817334	0.9997413	-1.464443	0.671268
	FZD6	1.435038	0.7892429	2.559063	0.05908473	1.956531	0.8185425	1.967634	0.5523988
	LCP1	0.7876063	0.4922049	0.58318888	0.5595369	-0.2961024	0.9997413	-0.27333095	0.904039
	LYL1	-1.68966	0.2942497	-1.964279	0.203308	-0.3661521	0.9997413	0.7640199	0.7585103
	MPL	-3.2891345	0.2040375	-2.8018583	0.01848699	0.7274934	0.9997413	-0.49350351	0.9939747
	MPO	-3.7959253	0.1044504	-3.5508527	0.03682876	-0.7904974	0.9997413	-1.85624938	0.6266575
	MYC	-1.26652126	0.4410453	-0.8118683	0.6387137	0.6483242	0.99974129	-0.003688694	0.9997997
YY1	0.44978787	0.7421003	0.4069389	0.6512013	-0.0785527	0.9997413	-0.05051148	0.9839762	
Downreg	CREBBP	0.3918586	0.8768832	0.3712743	0.7939281	-0.04682011	0.9997143	-0.1427875	0.9641316
	HIPK1	0.1425744	0.9617077	0.553601	0.674537	0.5854488	0.9544713	0.2716963	0.9167558
	LDB1	-1.0802818	0.6764357	-1.140708	0.2501488	-0.08666976	0.9997413	-0.06524431	0.9912882
	MYB	0.9784309	0.7743725	0.4601143	0.718795746	-0.68658085	0.9997413	-0.2465742	0.9613585

Table 5.6.1: Expression of proposed CML-like gene expression pattern in RNAseq comparisons. L2F= Log2Fold, padj= adjusted P value. Ph+ALL vs CML (PALL3F/10M(4)/4M/2M/2M(2)/15M/5M(2)/12M/10M/8M/19M compared to CMLLBC1/2), Cluster 3 vs Cluster 1 (PALL10M/8M/19M compared to CMLLBC1/2), Cluster 3 vs Cluster 2 (PALL10M/8M/19M compared to PALL3F/10M(4)/4M/2M/2M(2)/15M/5M(2)/12M) and BCRABLPOS vs BCRABLNEG (PALL2M/12M/5F(2)/8M/10M/19M compared to PALL3F/10M(4)/4M/2M(2)/15M)

However, a prospective list of genes for identification of CML-like Ph+ALL was proposed in Chapter 3 (Table 5.1.1). Such genes were selected based on expression in aberrant Ph+ALL samples, Ph+ALL samples with poor outcomes and pathology related to CML (ie LSC-associated and myeloid genes). As there are no public datasets available with clearly defined

CML-like Ph+ALL samples, this list was never intended to be a definitive diagnostic gene panel, instead proposing genes of interest that may be differentially regulated in a Ph+ALL with some cell involvement similar to CML and with poor clinical outcomes- two features of CML-like Ph+ALL posited by Hovorkova et al (2017).

Expression of these genes were analysed in the context of the sample comparisons discussed throughout this chapter (Table 5.6.1). Broadly, there was little statistical significance of gene expression throughout the comparisons tested (two genes identified with adjusted P value <0.05). Interestingly, two of the genes found to be highly upregulated and with near-significant adjusted P value were found in the Cluster 3 vs Cluster 1 analysis: FLT3 and FZD6 (L2F 2.31 and 2.56 respectively, padj 0.077 and 0.06 respectively). While FZD6 activity is associated with Wnt signalling and inhibition of oncogenic transformation (including cell proliferation and inhibition of apoptosis), FLT3 is associated with regulation of haematopoiesis including the proliferation and differentiation of haematopoietic cells in bone marrow. FLT3 mutations are associated with poor-prognosis AML, highlighting a potential transcriptomic similarity between an acute myeloid leukaemia (AML) and an acute leukaemia with myeloid and lymphoid involvement (CML-like Ph+ALL).

The gene encoding chemokine CCL8 was similarly upregulated in Cluster 3 vs Cluster 1, Cluster 3 vs Cluster 2 and BCRABLPOS vs BCRABLNEG samples (L2F 2.26, 5.55 and 4.01 respectively). The CCL8 protein functions as a chemoattractant for both lymphoid and myeloid cells, often in an inflammatory context. CCL8 has also been identified in driving breast cancer metastasis, with additional roles in diffuse large B-cell lymphoma and chemoattraction of M1 macrophages in cutaneous melanoma (Farmaki et al., 2016; Lou et al., 2022 and Yang et al., 2021). Hence, chemokine signalling may be of interest for further exploration in CML-like Ph+ALL.

5.7 Conclusion: RNAseq analysis of CML-LBC and Ph+ALL samples identified a transcriptionally and clinically distinct subgroup of Ph+ALL samples.

Overall transcriptional profiles of the CML-LBC and Ph+ALL samples investigated by RNAseq displayed an expected difference between CML and Ph+ALL. However, transcriptional profiling also revealed a subgroup of Ph+ALL samples which clustered distal to the rest of the Ph+ALL cohort. Upon investigation of clinical information provided for Ph+ALL samples, it was observed that this sub-group of Ph+ALL samples had a distinct clinical phenotype in comparison to the rest of the cohort. These samples (designated Cluster 3) had poor clinical outcomes, observed by a higher incidence of ALL progression, relapse and death (Table 5.1.3, Figure 5.2.1 & 5.2.2). These patients also had a high MRD level, higher WCC at diagnosis and all samples in Cluster 3 had detectable BCR::ABL 1 by FISH. This signified that Cluster 3 represents a transcriptionally and clinically distinct subgroup of Ph+ALL patients in this RNAseq investigation.

DEGs highlighted by GSEA in Ph+ALL samples with detectable BCR::ABL1 by FISH displayed phenotypes associated with cell trafficking, interaction with the microenvironment and early developmental processes. These DEGs displayed a phenotype related to LSC activity previously observed in CML, AML and MCL. These findings suggest an altered stem cell phenotype in Ph+ALL samples which have positive BCR::ABL1 FISH scores and represent a transcriptional alteration driven by the BCR::ABL1 translocation event in early stem and progenitor cells.

A number of DEGs uniquely upregulated in Cluster 2 throughout analyses had similar functions to genes identified in Chapter 3 from GEO datasets (Table 5.1.1). These included chemokines with pro-inflammatory functions such as *CCL8* and *CCL20*, and *FZD4/6* which is a member of the Wnt signalling pathway and plays a role in maintaining stemness. These associations display the utility of the GEO datasets investigated in Chapter 3 and indicates alterations to immune response and increased maintenance of stem cells compared to the rest of the (standard) Ph+ALL cohort. The use of such genes as biomarkers could be beneficial as high serum CCL20 is already associated with tumour progression in penile cancer, hence sensitive testing for this chemokine is already available (Mo et al., 2020). Other genes related

to stem cell maintenance in CML found to be upregulated in Cluster 2 include *ACE*, *SERPINB4*, *VCAN* and *MCAM*. The utility of these genes for use as a biomarker include detection of the surface marker CD146 (encoded by *MCAM*) by flow cytometry and measurement of ACE serum levels, as is currently used for sarcoidosis monitoring. The roles of such DEGs in the alteration of the bone marrow microenvironment to support LSC maintenance suggests that similar to CML, the BME of outlying/CML-like Ph+ALL patients may undergo structural alterations. Further exploration into useful biomarkers may therefore include genes with roles microenvironmental alteration.

Utilising FISH data generated within this project, samples with BCR::ABL1 positive stem and progenitor cells were compared to samples where this fusion event could not be detected in early haematopoiesis. Further supporting the heterogeneity of Ph+ALL, 3 samples from Cluster 2 had positive FISH scores despite being distinct from Cluster 3 transcriptionally and phenotypically. Investigation into differentially expressed genes between the BCR::ABL1 positive and negative progenitor groups revealed a number of genes associated with LSC function in CML, AML and MCL in samples with positive FISH scores in HSCs and MPPs. These results indicate the presence of more LSC-like activity in FISH positive samples. While this is not a confirmation of LSC presence in Ph+ALL or in CML-like Ph+ALL, these results display a marked difference in the transcriptional profile of early stem and progenitor cells between both sample groups.

Cluster 3 was of great interest due to its transcriptional and clinical differences from the main Ph+ALL cohort. As a result, Cluster 3 was compared to the CML Cluster 1 and Ph+ALL Cluster 2 and functions of differentially expressed genes assessed by gene set enrichment analysis. GSEA revealed few GO terms of difference between Cluster 1 and Cluster 3. These findings indicated transcriptional similarity between the CML samples and this Ph+ALL subgroup. DEGs highlighted had roles in cell interaction with the microenvironment which may indicate a disease-specific bone marrow microenvironment or cellular response. Cluster 3 was also compared to the Ph+ALL Cluster 2, with GSEA terms depicting a large number of cellular processes that were differently regulated between both Ph+ALL clusters. Biological processes which differed between Cluster 2 and 3 included immune response, cellular interaction with the microenvironment and early development genes. Not only did this display a difference in

cell activity between both Ph+ALL clusters but, importantly, genes previously implicated in CML and AML LSC activity were detected. The implications from these results are that there may be LSC activity in Cluster 3 but little in Cluster 2. As LSCs have previously been associated with CML, not Ph+ALL, this may indicate that Cluster 3 shares a mixed phenotype with the initial clinical diagnosis being Ph+ALL but with CML transcriptional signatures. Hence Cluster 3 may be representative of samples which would be designated as CML-like Ph+ALL.

In this chapter, bulk BMNC samples were used for RNAseq. Our first aim was to identify aberrances of gene expression within the heterogeneous Ph+ALL population for the identification of outlying samples with myeloid gene expression patterns with the goal of identifying CML-like Ph+ ALL samples. Bulk samples were examined instead of sorted progenitor cell groups for the potential use of subtype-specific differentially expressed genes as diagnostic biomarkers. Primary samples are examined clinically for biomarker gene expression without flow cytometry-assisted cell sorting due to sample size, financial and reagent restraints hence, biomarkers that can be identified in bulk bone marrow or peripheral blood samples are clinically relevant. Though beneficial for clinical application, there were a number of caveats to this approach. In Ph+ALL at diagnosis the proportion of bone marrow made up of lymphoblasts must be above 20% in order to be defined as leukaemic (Chiaretti et al., 2014). As a result, bulk diagnostic Ph+ALL samples contain a large number of blast cells which may skew transcription data toward a lymphoid phenotype, preventing adequate interrogation of gene expression in smaller cell populations. This issue is more relevant if examining population-specific gene expression patterns however, as discussed, the choice to examine bulk samples is more compatible with current clinical procedures. A method of preventing this transcriptional bias would be to sort samples into progenitor groups and investigate individual populations. Investigating transcriptional heterogeneity between HSC populations in Ph+ALL samples would be of interest for furthering the understanding of the effects of early stem and progenitor cell activity on the development of leukaemic subsets. With further research being conducted on HSC heterogeneity and predetermination, single cell RNAseq of Ph+ALL HSCs may be a fruitful progression of this investigation (Schroeder, 2010). This may be applied to elucidate how predetermined HSCs may lead to the development of distinct leukaemic subtypes, identification of pre-leukaemic HSCs and the importance of BCR::ABL1 fusion in HSCs predetermined for specific lineages. As the data from

Chapter 4 as shown, not all HSCs in a sample contain *BCR::ABL1* and not all samples with Ph+ stem and progenitor cells can be categorised as CML-like. Therefore, the presence of *BCR::ABL1* alone is inadequate for the establishment of CML-like Ph+ALL, with subtype-specific phenotypic differences likely arising from transcriptional differences. Hence, elucidating gene expression changes between classical and CML-like Ph+ALL would require investigation of HSC transcriptional profiles from both subgroups. Additionally, LSCs are yet to be identified in CML-like Ph+ALL, identification of LSC-associated genes (such as those used in AML (Bill et al., 2020)) would further aid understanding of this subtype and inform on best treatment strategies for such patients. Examination of aberrant gene expression and subtype-specific *BCR::ABL1* point mutations has been successful in targeting previously difficult to treat BP-CML with next generation TKIs (Hughes et al, 2006; Jabbour et al, 2006; Nicolini et al, 2006). A goal of this is to utilise gene expression data in the heterogeneous ALL to identify patients belonging to disease subtypes and alter treatment appropriately.

CML-LBC samples were selected for sequencing to establish a CML transcriptional profile which Ph+ALL samples were compared to. This is a fairly innovative investigation, with few publicly available datasets available which directly compare transcriptional profiles of CML-LBC to Ph+ALL. While there is a multitude of datasets which include sequencing data on both leukaemic types separately, accurate and significant conclusions are unable to be reached by comparing samples from separate datasets. Hence, creating this unique RNAseq dataset is not only of importance for this project, but also of great relevance to other researchers in this field. Samples of CML in lymphoid blast crisis represent a myeloid leukaemia in an acute phase with lymphoblast involvement (Ilaria, 2005). CML-like Ph+ALL shares phenotypic traits with both CML and Ph+ALL, those being an acute lymphocytic leukaemia with significant myeloid involvement (Hovorkova et al., 2017). Identifying the genes necessary for distinguishing CML-like Ph+ALL from CML-LBC would also confirm this novel subtype as being distinct from CML presenting at lymphoid blast crisis with no clinical history of chronic phase, a hypothesis which has been previously raised in this field (Hunger, 2017).

The number of GO terms and variety of gene functions highlighted as being differentially regulated between Ph+ALL Cluster 2 and Ph+ALL Cluster 3 displays a notable transcriptional difference between both Clusters. Alterations to cellular interaction with the

microenvironment, immune response and early developmental gene supports the hypothesis that Cluster 3 is representative of a distinct subset of Ph+ALL with a distinct transcriptional profile and clinical outcomes.

5.7.2 Caveats and limitations.

Caveats that arose from this sample selection include lack of BCR::ABL1 MRD clinical data, preventing CML-like patients from being identified through the same methods as used in the Hovorkova paper (2017). Additionally, there may have been unintentional bias in the samples sequenced as RNA quality could be related to CML-like status as discussed in Chapter 4 regarding CML-like samples possibly being better able to survive the cryopreservation, culture and FACS process. Finally, age difference between patients. All Ph+ALL samples sequenced were isolated from paediatric patients ranging from ages 2-19 years. However, CML-LBC samples were harvested from two adult patients aged 33 and 41 years. One reason behind the choice of these samples was the lack of availability of paediatric CML-LBC samples. Paediatric CML is rare and accounts for 2-3% of newly diagnosed leukaemia in children, the age group with the highest incidence of CML diagnosis being between 45-55 years (Pushpam and Bakhshi, 2019). Patients diagnosed with CML in lymphoid blast crisis are treated with an ALL-like induction method, highlighting the pathologic similarity between CML-LBC and ALL and clinical relevance of selecting such samples for RNAseq alongside Ph+ALL (Ilaria, 2005). Age-related transcriptional differences is a caveat to be noted, with differences to the immunologic population of bone marrow and a decline in the number of 'fit' mesenchymal stem cells (MSCs) in the bone marrow niche with age being well noted in the literature (Ganguly et al., 2017). Hence, differences in the transcriptional profile of BMNCs harvested from paediatric and adult patients may be detectable, disease relevance is satisfied by this sample selection. Future work in this area including age matched Ph+ALL and CML-LBC samples would be adequate for removing such age-related issues, however, this is outwith the scope of this project. The BMNC samples used in these transcriptional experiments were selected in order to investigate the gene expression profiles of BCR::ABL-driven leukaemia in an acute phase with distinct lymphoid cell involvement. As a result, the relation of age to disease phenotype was not explored, despite the age differences between samples (adult CML samples and paediatric Ph+ALL samples).

Chapter 6. Concluding discussion.

In order to fully examine the undefined CML-like Ph+ALL subgroup, primary paediatric Ph+ALL bone marrow samples were utilised. Samples were harvested from patients at hospital sites across the UK (for patient confidentiality, the individual location were not provided). Of a total of 24 samples, 23 were provided by CellBank via CCLG (Children's Cancer and Leukaemia Group) and one sample was sourced from the Queen Elizabeth University Hospital (Glasgow). Samples were harvested from patients ranging from age 2-19 years, with the mean age being 8.94 years. Sex was not equally weighted with 16/24 patients being male, 7/24 samples being female and one sample being unknown due to data being unavailable to CellBank. With a cohort size of 24 patients, a perfectly balanced distribution of sex is unlikely however, males are 30% more likely to be diagnosed with ALL than females which is somewhat represented by the sex distribution in the cohort (CLRUK., 2021).

All CellBank patients had diagnostic bone marrow samples available, allowing for analysis of bone marrow mononuclear cells prior to treatment. In addition to these, post induction follow-up samples were available for 16 of the 23 CellBank samples, harvested from day 28 to day 140 (average day 70) post-induction therapy commencement. The patients based in Glasgow had no diagnostic samples available however, sequential samples were available harvested during the first 3 months of treatment. In order to establish which disease elements CML-like Ph+ALL samples share with both classical Ph+ALL and CML, two control CML-LBC samples were utilised. These samples were sourced from adult CML patients (age 33 and 41 years) in lymphoid blast crisis. Adult samples were selected due to the rarity and therefore unavailability of paediatric CML samples (in LBC). To allow for accurate comparison of Ph+ALL and CML-LBC samples used in this thesis, all samples were processed in the same manner as outlined in the materials and methods section (Chapter 2).

Bulk bone marrow samples were thawed according to the protocol discussed in materials and methods (Chapter 2.6.1- 2.6.5). samples were cultured overnight in serum-free media supplemented with cytokines and growth factors at a physiological concentration (Table

2.2.9). The intention of this culture method was to allow the cells to recover from the cryopreservation process without being stimulated to proliferate or differentiate, allowing cells to be as similar to the time of harvesting as possible. After culturing overnight, cells were washed in PBS +2% FBS then incubated with the appropriate antibody cocktails (Table 2.3.1 & 2.3.2) for 1 hour to allow for Lin⁻ haematopoietic progenitor cells to be sorted and collected by FACS. Cell populations were sorted into PBS +2% FBS to allow for cells to be used for further experiments. These experiments included fixing for FISH analysis of BCR::ABL1 fusion in sorted progenitor cells, extraction of DNA for Ig/TCR gene rearrangement analysis (to be discussed further in Chapter 7) and RNAseq analysis of bulk BMNCs.

As previously discussed, the recently postulated leukaemic subtype CML-like Ph+ALL has been highlighted as an unmet clinical need which remains poorly characterised. The original study by Hovorkova et al (2017) highlighted this subgroup of Ph+ALL patients with apparent myeloid disease involvement and first designated them as 'CML-like' Ph+ALL. However, the discussions around the biological similarities between CML-LBC and Ph+ALL are not new. With clinical characteristics of CML-LBC resembling that of Ph+ALL, there has long been concern that patients with newly diagnosed an acute Ph⁺ leukaemia with lymphoid involvement could be either CML-LBC or Ph+ALL, this misdiagnosis leading to the employment of therapeutics incorrect for the disease type (Kolenova et al., 2016). The outcomes of this have the potential to be dire as paediatric Ph+ALL patients are treated with TKIs in combination with chemotherapy while an allogenic stem cell transplant is the recommended therapeutic strategy for patients with CML-LBC. Due to the activity of the constitutively active tyrosine kinase BCR::ABL1 in both diseases and therefore similar disease mechanisms, there is some crossover in treatment strategy between CML and Ph+ALL in the use of tyrosine kinase inhibitors in combination with chemotherapeutic agents. However, with the advent of personalised medicine and particularly in acute phase leukaemias, it is imperative that diagnosis is accurate, and treatment is suitable. Hence, the presence of an uncharacterised Ph+ALL subtype represents an unmet clinical need.

The group of patients highlighted by Hovorkova et al were shown to respond poorly to commonly used Ph+ALL lymphoid-directed treatment in addition to discordance in MRD monitoring techniques. Hence, not only did these patients receive unsuitable treatment, they were also difficult to measure for risk stratification. As previously discussed, to elucidate subtype-specific characteristics of CML-like Ph+ALL, standard Ph+ALL and CML-LBC samples were utilised. Using samples which share a common pathogenic lesion but different phenotypes allows for characteristics of Ph+ALL or CML to be identified in the mixed subtype.

The use of primary paediatric Ph+ALL bone marrow samples allowed for analysis of phenotypic and transcriptional differences among the Ph+ALL cohort. With the CML-like subtype being undefined at the time of writing, there were no guidelines available to allow for categorisation of these patients. Hence, it was predicted that the Ph+ALL cohort used in this thesis displayed the heterogeneity observed clinically by Hovorkova et al (2017). Prior to laboratory processing of samples, this sample variability could be observed in the clinical information provided with the cryopreserved samples from CellBank (Table 4.4.2- 4.4.5). This data displayed a variety of responses to treatment, with some patients relapsing, some reaching remission and a range of MRD scores, white cell counts at diagnosis and bone marrow blast percentages. As previously mentioned, estimates of incidence of CML-like patients within the larger Ph+ALL cohort worldwide was 20-30% (Hovorkova et al., 2017). A small sample cohort of 24 may not be representative of this however, outlying features which would align with a CML-like disease pathology was detected in a small number of patients.

Heterogeneity within the paediatric Ph+ALL cohort was evidenced throughout this thesis and was initially detected through flow cytometry-based cell sorting of haematopoietic stem and progenitor cell populations (Fig 4.2.5.1 & 4.2.5.2). Within a diagnostic Ph+ALL sample, lymphoid blast cells are expected to predominate and a low frequency of HSCs and MPPs should be present. The majority of the samples analysed showed large populations of ProB and PreB cells, two lymphoid progenitor groups likely to display contamination with lymphoblasts due to similarity in cell surface markers. However, a small number of patients displayed unexpectedly large populations of HSCs and MPPs and few lymphoid precursor cells.

These samples displayed an atypical haematopoietic dynamic for a Ph+ALL diagnostic sample and suggested an alternative disease mechanism where the main leukaemic driver may not be lymphoid cells.

One of the main aims of this thesis and question arising from Hovorkova et al (2017) was the identity of the cell of BCR::ABL1 origin in CML-like Ph+ALL. To investigate this, sorted stem and progenitor cells were analysed for BCR::ABL1 fusion by FISH as previously described in Chapters 2 and 4. It was initially predicted that only CML-like Ph+ALL samples would contain the t(9;22) fusion event due to being the proposed Ph+ALL subtype with myeloid involvement and CML-like disease features. Our investigations revealed that BCR::ABL1 could be detected in HSCs and MPPs in a number of Ph+ALL samples (Fig 4.3.3). Due to the small population sizes collected by FACS, the percentage of BCR::ABL1 positivity could not be reliably used as a predictor for disease severity, outcome or the likelihood of the sample being CML-like Ph+ALL. What was observed, however, was the majority of Ph+ALL samples having evidence of Ph+ HSCs and MPPs. Whether these cells have functions similar to CML LSCs is unknown and outwith the scope of this project, however, this represents the first identification of Ph+ HSCs in Ph+ALL to date.

As the initial theory regarding the leukaemic origins in CML-like Ph+ALL were that only samples belonging to the newly-proposed CML-like subtype would contain Ph+ HSCs, the observation of BCR::ABL1 positive stem and progenitor cells across the cohort and in samples with a variety of clinical features and haematopoietic dynamics supported the understanding that Ph+ALL is an extremely heterogeneous leukaemic subtype. Such findings display that the sample t(9;22) fusion event could occur in the same cell type (HSC) but result in three different disease states; CML, standard Ph+ALL and CML-like Ph+ALL. Therefore, the phenotypic differences between these diseases must be further downstream than the initial fusion event and likely due to transcriptional differences arising from differential gene expression. To investigate this, 11 Ph+ALL and 2 CML-LBC samples were analysed by RNAseq, the inclusion of CML samples to determine a transcriptional profile of CML in lymphoid blast crisis and confirm that the CML-like Ph+ALL subtype is not a misdiagnosis of CML-LBC. An unbiased PCA

depiction of overall transcriptional profiles displayed three distinct clusters (named 'Cluster 1-2') (Fig 5.2.1). As anticipated, the CML-LBC samples were distal to the majority of Ph+ALL samples, owing to the fact that both are different disease states. However, a separate cluster of Ph+ALL samples (Cluster 3) could be observed outside of the CML-LBC and larger Ph+ALL cluster. Upon investigation of differential gene expression between Cluster 2 (Ph+ALL) and Cluster 3 (outlying Ph+ALL), it could be observed that the outlying Ph+ALL group displayed upregulation of genes involved in CML, AML and leukaemic stem cell biology such as *RIMS2*, *PVR*, *ACE*, *PTPRM*, *PTP*, *FZD4* and *LOX*.

Transcriptional analysis of Clusters 2 and 3 displayed that Cluster 3 had an upregulation of genes associated with immunity, early development and, crucially, genes involved in myeloid leukaemia disease biology. Retrospective analysis of the samples that constitutes Cluster 3 displayed poor clinical outcomes such as relapse and death. This supports the understanding that CML-like Ph+ALL patients may benefit from haematopoietic stem cell transplant as a first-line therapy, just as in CML. Additionally, these samples had unusual population dynamics detected by flow cytometry with large populations of myeloid progenitors, fewer lymphoid progenitors and a high frequency of HSCs. Finally, all samples of Cluster 3 had BCR::ABL1 positive HSCs detected by FISH (Fig 5.3.2). Interestingly, two samples in Cluster 2 also had Ph+HSCs/MPPs detected by FISH but did not display the transcriptional, clinical or haematopoietic dynamics associated with the samples in Cluster 3. This supports the understanding that occurrence of the BCR::ABL1 fusion event in HSCs alone is not adequate for establishing the CML-like subtype and that the origins of the disease likely arise from downstream transcriptional differences. The findings of this thesis demonstrate the heterogeneity of Ph+ALL and evidences CML-like Ph+ALL as being a leukaemic subtype with distinct clinical features, haematopoietic dynamics, BCR::ABL1 fusion in primitive stem cells and a myeloid-like transcriptional profile.

Chapter 7: Future Directions

The Philadelphia chromosome (Ph+) results in the BCR::ABL1 fusion oncoprotein with enhanced tyrosine kinase activity. It is a high-risk cytogenetic abnormality found in approximately 5% of childhood and 25-30% of adult acute lymphoblastic leukaemia (ALL) and all patients with chronic myeloid leukaemia (CML). Recent data suggests that Ph+ALL can be further sub-divided into a true lymphoid leukaemia with acquisition of the BCR::ABL1 oncogene in a committed B-cell progenitor (standard Ph+ALL) and a CML-like Ph+ALL with multilineage BCR::ABL1 activity. CML-like Ph+ALL is currently defined by MRD (minimal residual disease) discrepancy between Ig/TCR and *BCR::ABL1* expression (Hovorkova et al., 2017). Relapse risk in patients with undetectable Ig/TCR MRD is higher in Ph+ALL than other ALL subtypes, supporting that some patients have biologically different disease. Therefore, understanding the CML-like Ph+ALL population is imperative to deciphering better and more effective therapeutic strategies and is a critical area of unmet clinical need; particularly identifying these patients at an early time point from diagnosis.

The results presented within this thesis have demonstrated that Ph+ALL is likely even more heterogeneous than previously anticipated. These include the identification of transcriptional heterogeneity within Ph+ALL publicly available datasets, a subgroup of patients with haematopoiesis dynamics atypical to standard Ph+ALL, the detection of *BCR::ABL1* in HSPCs by FISH in a larger number of samples than initially anticipated and finally, a small number of outlying Ph+ALL patients which share DEGs (differentially expressed genes) with CML-LBC in addition to having a high incidence of relapse compared to the rest of the Ph+ALL cohort. Hence, it is hypothesised that Ph+ALL represents a spectrum of disease rather than a single entity, owing to the heterogeneity seen both transcriptionally and clinically in disease response to treatment. Furthermore, BCR::ABL1-expressing HSC (haematopoietic stem cell) and MPP (multipotent progenitor) cells may represent the elusive leukemic stem cell (LSC) of Ph+ALL in view of treatment resistance seen within this population. Reports have demonstrated an LSC clone within leukaemia, particularly in CML and acute myeloid leukaemia, but to date, there remains limited data surrounding an LSC within ALL, with some

contradictory findings (Castor et al., 2005). Within Ph+ALL, the expression of p190 or p210 BCR::ABL1 is sufficient to cause leukaemia in transgenic mice, with B and T cell acute leukaemia more common than myeloid (Voncken et al., 1995). Some functional studies conclude that the Ph+ALL LSC is a primitive cell that is lymphoid-restricted, although the studies differ on the characterisation of what this LSC is (Cobaleda et al., 2000). The primitive CD34+CD38-CD19- cell compartment has been shown to be involved in patients with p210 BCR::ABL1 ALL but not in those with p190 BCR::ABL1 ALL (Castor et al., 2005). The p210 BCR::ABL1 transcript could also be identified in more mature CD34+CD33+ and CD34-CD33+ myeloid precursors, which was not the case for the p190 transcript. However, CD34+CD38-CD19- p210 BCR::ABL1 positive cells did not induce leukaemia in NOD/SCID mice. This contrasts with other results showing NOD/SCID engrafting leukemic cells only in the CD34+CD38- subfraction and not with CD34+CD38+ cells. If an LSC does exist, its persistence could represent a mechanism of TKI resistance and poorer outcomes, as seen with CML. Within Ph+ALL, resistance to treatment has been described. For example, Nagel et al described Ph+ALL with stem cell involvement as a mechanism of resistance to Blinatumomab. Further to this, earlier need for HSCT (haematopoietic stem cell transplant) in the CML-like Ph+ALL population has been hypothesised in view of poorer survival outcomes (Hovorkova et al., 2017).

The detection of *BCR::ABL1* in HSCs and MPPs by FISH previously reported in this thesis supports the theory that LSC activity may be present in Ph+ALL patients who responded poorly (relapse or disease resurgence) to standard Ph+ALL treatment as this reflects LSC activity in CML where *BCR::ABL1* can be detected in the stem cell fraction and where LSC-mediated treatment resistance has been reported (as discussed above). However, in addition to samples that responded poorly to treatment, Ph+ HSPCs (haematopoietic stem and progenitor cells) were also detected in patients who aligned with 'standard' Ph+ALL clinical outcomes, displayed expected haematopoiesis dynamics (by FACS) and displayed an expectedly lymphoid-biased gene expression pattern. Hence, these findings suggest that *BCR::ABL1* presence in HSCs alone is inadequate for establishing a CML-like subtype. This brings into question the importance of *BCR::ABL1* fusion in the establishment of disease state and whether downstream alterations are required to initiate a standard or CML-like Ph+ALL

phenotype. These findings may be explained, in part, by the fact that HSCs and MPPs were isolated by surface marker phenotype, not by stem cell functional assays. Surfaceome heterogeneity within tumour populations have been identified in AML and multiple myeloma, evidencing aberrant expression of surface markers within blast cell populations (Bordelau et al., 2024 and Ferguson et al., 2022). Though not yet characterised in Ph+ALL, blast cell aberrant surface marker expression may result in the misidentification of cell types by flow cytometry. Hence, HSC and MPP populations isolated by FACS in this thesis may be contaminated with blast cells aberrantly expressing stem-like surface markers, without functional assays, the true identity of these cells would be undetermined.

As previously discussed in chapter 4 of this thesis, HSC population sizes in diagnostic Ph+ALL populations are likely to be extremely small or completely diminished due to a skew toward lymphopoiesis. Hence, samples with an overall 'standard' Ph+ALL phenotype with detectable Ph+HSPCs may still have a typically small/diminished HSC population and may be displaying BCR::ABL1 positive blast cells which aberrantly express HSC-like surface markers. In addition, outlier ('CML-like') Ph+ALL samples identified by RNAseq, FACS and clinical data may have an elevated population of true HSCs which express BCR::ABL1. This theory would be more in line with our general understanding of BCR::ABL1 activity where a downstream alteration is not required for disease establishment and the point during haematopoietic development when t(9;22) arises impacts on the cell types involved in pathogenesis.

This could be elaborated by analysis of blast cell development markers in HSC and bulk cell populations in standard and 'CML-like' Ph+ALL samples where Ph+HSPCs have been detected. A feature of lymphoid blast cells which would not be present in HSCs is Ig/TCR gene rearrangement, a hallmark of lymphoid development (Fig 7.1). As HSCs have not yet undergone lineage fixing, this gene rearrangement will not have occurred and therefore identification of Ig/TCR recombination would confirm which cells are true HSCs, and which are blast cells with aberrant HSC-like surface marker expression. Analysis of Ig/TCR gene rearrangement is a minimal residual disease method used for evaluation of treatment effectiveness in patients with Ph+ALL (van der Velden & van Dongen., 2009).

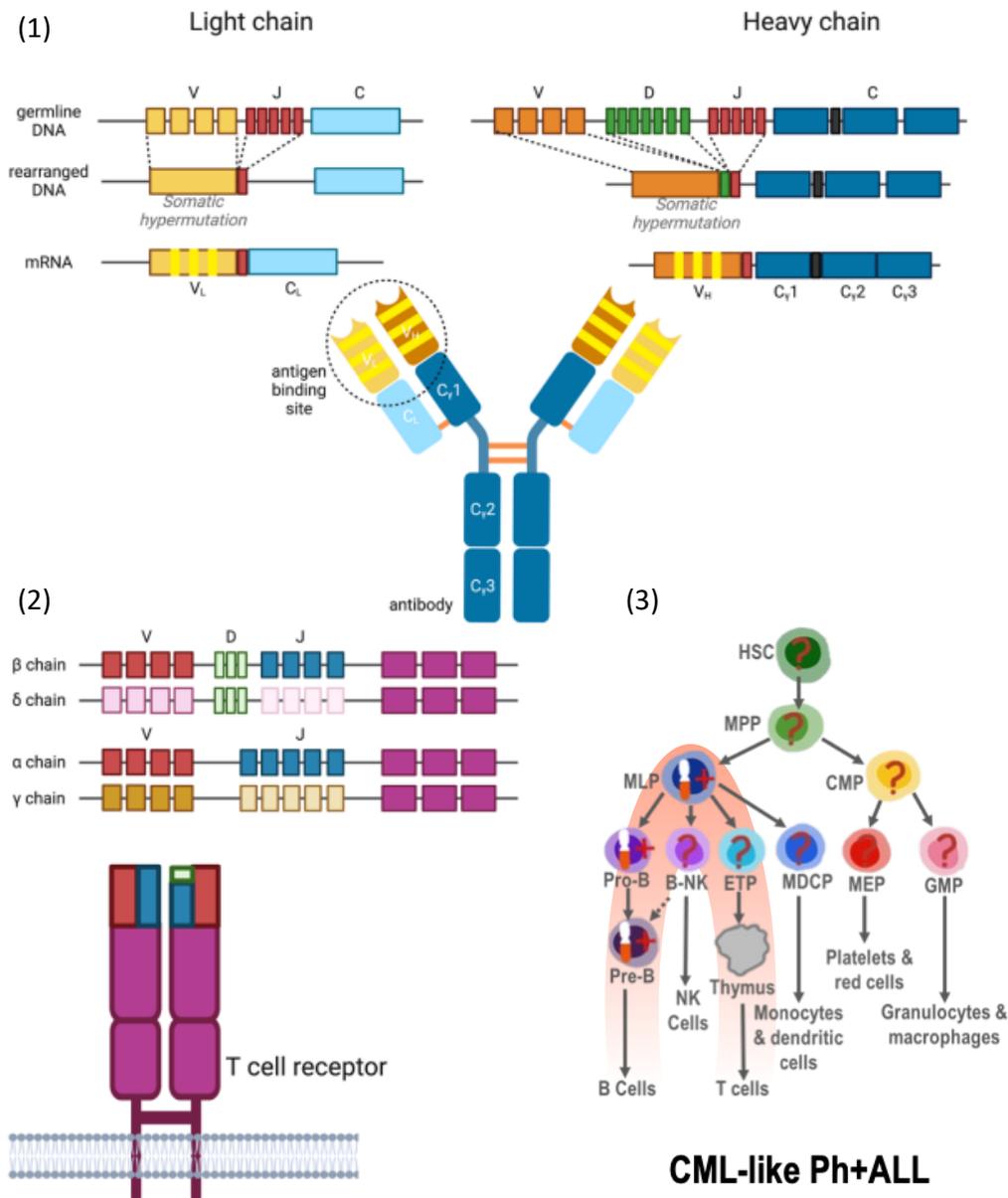


Figure 7.1: Schematic overview of the organisation and rearrangement of Ig and TCR genes.
(1) Different germline gene segments coding for the variable Ig heavy and light chains are joined by somatic V(D)J gene rearrangement, Addition or removal of nucleotides during recombination at the junctions and somatic hypermutation in the complimentary-determining regions of the VL and VH genes results in high diversity of the Ig repertoire. **(2)** T cell receptor gene rearrangement. Variable (V), joining (J) and constant regions constitute the TCR-alpha chain. VJC regions constitute the TCR-beta chain with an additional diversity (D) region. Segments from each region are recombined, with additional nucleotide additions to generate each rearranged TCR to generate T cell diversity. **(3)** Haematopoietic diagram indicating the stages of leukocyte development where Ig/TCR gene rearrangement occurs (orange background colour). Chromosome rearrangement t(9;22) presence in progenitor cells as hypothesised in the CML-like Ph+ALL subtype.

During early lymphoid development, the germline V, D and J gene segments of the immunoglobulin (Ig) and T cell receptor (TCR) genes rearrange. The outcome of this is the establishment of an expansive immune repertoire by a specific combination of V, D and J segments encoding Ig and TCR variable domains (Fig 7.1). Such variable regions have

'fingerprint'-like sequences, unique to each lymphocyte from the action of the random insertion and deletion of nucleotides at the junction sites of the V, D and J segments. Hence, this unique gene arrangement is also seen in lymphoid malignancies and allow junctional regions to be used as tumour-specific targets for MRD monitoring tracked using RT-QPCR (real-time quantitative polymerase chain reaction).

To investigate Ig/TCR gene rearrangement in the Ph+ cells sorted by HSC markers in this thesis, the ClonoSEQ® Ig/TCR MRD monitoring platform (Adaptive Biotech) could be utilised. DNA was isolated from sorted cell populations for most of the samples in this thesis (missing DNA resulted from small starting cell number and where RNA extraction or fixing for FISH was prioritised). Therefore, HSCs and bulk (unsorted) cells from Ph+ALL samples with a standard phenotype and detectable BCR::ABL1 in HSCs, Ph+ALL samples with a CML-like phenotype (as previously discussed) and standard Ph+ALL samples with no detectable BCR::ABL1 in HSPCs could be assessed for Ig/TCR gene rearrangement. This would enable verification of HSCs as being cells which are too primitive to have undergone Ig/TCR rearrangement and confirm if Ph+ HSCs from standard Ph+ALL samples are actually lymphoblast cells aberrantly expressing HSC surface markers.

Further to this, we need to determine if BCR::ABL1 positive cells are driving an LSC phenotype functionally using stem cell assays well established in the Horne lab. These include colony forming cell (CFC) assays and long-term culture initiating cell assays (LTC-IC), both of which assess the differentiation capacity of HSCs. As discussed above, characterisation of HSCs should include functional assays to assess the ability of HSCs to proliferate and differentiate into colonies in response to the appropriate cytokine stimulation. Verification of HSC functionality would then allow further investigation to determine if such Ph+ HSCs display LSC-like features such as drug resistance.

From the data discussed in this thesis, Ph+ALL is likely a heterogeneous disease, in order to heighten our understanding of this and determine if this can be observed at the single cell

level, single cell transcriptomic analysis could be an insightful approach. Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITEseq) technology can assess transcriptional alteration through RNA sequencing in addition to information (both qualitative and quantitative) on surface proteins using antibodies. This method allows a highly sensitive assessment of gene expression and proteomic analysis, with resolution of low abundance proteins (few proteins per cell) and small starting cell numbers required (as low as 500) (10x Genomics., 2018). This would allow for specific cell types to be identified by surface marker and any transcriptional heterogeneity assessed. Further exploration into transcriptional heterogeneity within the Ph+ALL cohort could be performed using long-read RNAseq. This sequencing method generates longer libraries with more accurate sequences and transcripts than libraries with greater read depth as would be yielded from the short-read methods (Pardo-Palacois et al., 2024). Though short-read RNAseq with greater read depth yields improved quantitation accuracy, long-read RNAseq allows for the detection of complex structural variants such as inversions, deletions or translocations which may not be detected with short-reads. This could also yield improved identification of novel isoforms and highly polymorphic regions. As previously discussed in chapter 1, *BCR::ABL1* activity has the downstream effect of genomic instability and therefore long-read RNAseq may be beneficial for identifying structural variants and complex rearrangements between Ph+ALL subtypes that may be beneficial for diagnosis or therapeutic targeting.

Another method of identifying intra-sample and intra-tumoral heterogeneity within the Ph+ALL cohort could be TARGET-seq. This single cell RNAseq technique genotypes somatic mutations while capturing whole transcriptomes within tumour cells. Applied to the CML-like and standard Ph+ALL cohorts, this could be utilised to identify associations between somatic mutations within lymphoblast and Ph+ HSPC cells and subsequent transcriptional phenotypes (Rodriguez-Meira et al., 2020). Concurrent single cell mutational analysis and RNA sequencing would allow for the transcriptional differences between the Ph+ALL subgroups observed in chapter 5 of this thesis to be linked to potential *BCR::ABL1* point mutations observed previously in CML (Table 1.4). As has been evidenced in CML, such mutations can be targeted by new generation TKIs such as ponatinib or dasatinib, thus preventing disease progression or relapse by focused treatment. Such *BCR::ABL1* mutations have yet to be fully explored in

Ph+ALL, hence, not only may this reveal a novel mechanism for disease progression but also generate new directed treatment schema by utilising already available drugs used in CML.

Translation of these outcomes may improve outcomes for patients with CML-like Ph+ALL, both in terms of diagnostic approaches and treatment approaches, as CML-like Ph+ALL may require stem cell- directed therapy, e.g., allogeneic stem cell transplant for cure. Further, these patients may benefit from more intensive follow-up approaches and evaluation of possible biomarkers to predict outcome and likelihood of relapse. This is imperative to ensure early recognition of treatment failure and initiation of new treatments to ensure better patient outcomes.

References

Steliarova-Foucher E, Stiller C, et al. International Classification of Childhood Cancer, third edition(link is external). *Cancer* 2005;103:1457-67.

AbuSamra, D et al., " Not just a marker: CD34 on human hematopoietic stem/progenitor cells dominates vascular selectin binding along with CD44". *Blood Adv* 2017; 1 (27): 2799–2816. doi: <https://doi.org/10.1182/bloodadvances.2017004317>

Advani, Anjali S, and Ann Marie Pendergast. "Bcr-Abl variants: biological and clinical aspects." *Leukemia research* vol. 26,8 (2002): 713-20. doi:10.1016/s0145 2126(01)00197-7

Akahoshi, Yu et al. "Newly proposed threshold and validation of white blood cell count at diagnosis for Philadelphia chromosome-positive acute lymphoblastic leukemia: risk assessment of relapse in patients with negative minimal residual disease at transplantation-a report from the Adult Acute Lymphoblastic Leukemia Working Group of the JSTCT." *Bone marrow transplantation* vol. 56,11 (2021): 2842-2848. doi:10.1038/s41409-021-01422-7

Alaggio, Rita et al. "The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms." *Leukemia* vol. 36,7 (2022): 1720-1748. doi:10.1038/s41375-022-01620-2

Alaggio, Rita et al. "The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms." *Leukemia* vol. 36,7 (2022): 1720-1748. doi:10.1038/s41375-022-01620-2

Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell*. 4th edition. New York: Garland Science; 2002. Lymphocytes and the Cellular Basis of Adaptive Immunity. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK26921/>

Arcone, R., et al. (1993). "Structural characterization of a biologically active human lipocortin 1 expressed in *Escherichia coli*." *European Journal of Biochemistry* 211(1-2): 347-355.

Aricò, M et al. "Outcome of treatment in children with Philadelphia chromosome-positive acute lymphoblastic leukemia." *The New England journal of medicine* vol. 342,14 (2000): 998-1006. doi:10.1056/NEJM200004063421402

Belyavsky, Alexander et al. "Hematopoiesis during Ontogenesis, Adult Life, and Aging." *International journal of molecular sciences* vol. 22,17 9231. 26 Aug. 2021, doi:10.3390/ijms22179231

Bennett, J M et al. "Proposals for the classification of the acute leukaemias. French-American- British (FAB) co-operative group." *British journal of haematology* vol. 33,4 (1976): 451-8. doi:10.1111/j.1365-2141.1976.tb03563.x

Bernt KM, Hunger SP. Current concepts in pediatric Philadelphia chromosome-positive acute lymphoblastic leukemia. *Front Oncol*. 2014 Mar 25;4:54. doi: 10.3389/fonc.2014.00054. PMID: 24724051; PMCID: PMC3971203.

Biondi, Andrea et al. "Imatinib after induction for treatment of children and adolescents with Philadelphia-chromosome-positive acute lymphoblastic leukaemia (EsPhALL): a randomised, open-label, intergroup study." *The Lancet. Oncology* vol. 13,9 (2012): 936-45. doi:10.1016/S1470-2045(12)70377-7

Bonifacio, Massimiliano et al. "Management of Chronic Myeloid Leukemia in Advanced Phase." *Frontiers in oncology* vol. 9 1132. 25 Oct. 2019, doi:10.3389/fonc.2019.01132

Brück, O., et al. (2018). "Immune cell contexture in the bone marrow tumor microenvironment impacts therapy response in CML." *Leukemia* 32(7): 1643-1656.

Calabretta, Bruno, and Danilo Perrotti. "The biology of CML blast crisis." *Blood* vol. 103,11 (2004): 4010-22. doi:10.1182/blood-2003-12-4111

Carrasco, Javier et al. "CD45RA on human CD8 T cells is sensitive to the time elapsed since the last antigenic stimulation." *Blood* vol. 108,9 (2006): 2897-905. doi:10.1182/blood-2005-11-007237

Casamassimi, A., et al. (2017). "Transcriptome Profiling in Human Diseases: New Advances and Perspectives." *International Journal of Molecular Sciences* 18(8): 1652.

Castor, Anders et al. "Distinct patterns of hematopoietic stem cell involvement in acute lymphoblastic leukemia." *Nature medicine* vol. 11,6 (2005): 630-7. doi:10.1038/nm1253

Charfi, C., et al. (2014). "Identification of GPM6A and GPM6B as potential new human lymphoid leukemia-associated oncogenes." *Cellular Oncology* 37(3): 179-191.

Chase, A et al. "Factors influencing the false positive and negative rates of BCR-ABL fluorescence in situ hybridization." *Genes, chromosomes & cancer* vol. 18,4 (1997): 246-53.

Chen, Z., et al. (2020). "Chronic myeloid leukemia presenting in lymphoblastic crisis, a differential diagnosis with Philadelphia-positive B-lymphoblastic leukemia." *Leukemia & Lymphoma* 61(12): 2831-2838.

Cheng, Hui et al. "New paradigms on hematopoietic stem cell differentiation." *Protein & cell* vol. 11,1 (2020): 34-44. doi:10.1007/s13238-019-0633-0

Cheng, Hui et al. "New paradigms on hematopoietic stem cell differentiation." *Protein & cell* vol. 11,1 (2020): 34-44. doi:10.1007/s13238-019-0633-0

Chiaretti, Sabina et al. "Diagnosis and subclassification of acute lymphoblastic leukemia." *Mediterranean journal of hematology and infectious diseases* vol. 6,1 e2014073. 1 Nov. 2014, doi:10.4084/MJHID.2014.073

Cilloni, Daniela, and Giuseppe Saglio. "CML: a model for targeted therapy." *Best practice & research. Clinical haematology* vol. 22,3 (2009): 285-94. doi:10.1016/j.beha.2009.04.004

Cimato, Thomas R et al. "Simultaneous measurement of human hematopoietic stem and progenitor cells in blood using multicolor flow cytometry." *Cytometry. Part B, Clinical cytometry* vol. 90,5 (2016): 415-23. doi:10.1002/cyto.b.21354

Cobaleda, C et al., A primitive hematopoietic cell is the target for the leukemic transformation in human Philadelphia-positive acute lymphoblastic leukemia. *Blood* 2000; 95 (3): 1007–1013. doi: https://doi.org/10.1182/blood.V95.3.1007.003k35_1007_1013

Conant, Joanna L, and David R Czuchlewski. "BCR-ABL1-like B-lymphoblastic leukemia/lymphoma: Review of the entity and detection methodologies." *International journal of laboratory hematology* vol. 41 Suppl 1 (2019): 126-130. doi:10.1111/ijlh.13012

Costa, Valerio et al. "RNA-Seq and human complex diseases: recent accomplishments and future perspectives." *European journal of human genetics : EJHG* vol. 21,2 (2013): 134-42. doi:10.1038/ejhg.2012.129

Cruz-Miranda, G., et al. (2019). "Long Non-Coding RNA and Acute Leukemia." *International Journal of Molecular Sciences* 20(3): 735.

Cuellar, Sandra et al. "BCR-ABL1 tyrosine kinase inhibitors for the treatment of chronic myeloid leukemia." *Journal of oncology pharmacy practice : official publication of the International Society of Oncology Pharmacy Practitioners* vol. 24,6 (2018): 433-452. doi:10.1177/1078155217710553

Dai Q, Zhang G, Yang H, Wang Y, Ye L, Peng L, Shi R, Guo S, He J, Jiang Y. Clinical features and outcome of pediatric acute lymphoblastic leukemia with low peripheral blood blast cell count at diagnosis. *Medicine (Baltimore)*. 2021 Jan 29;100(4):e24518. doi: 10.1097/MD.00000000000024518. PMID: 33530278; PMCID: PMC7850651.

Deaglio, Silvia et al. "CD38 at the junction between prognostic marker and therapeutic target." *Trends in molecular medicine* vol. 14,5 (2008): 210-8. doi:10.1016/j.molmed.2008.02.005

Druker BJ, Sawyers CL, Kantarjian H, Resta DJ, Reese SF, Ford JM, Capdeville R, Talpaz M. Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N Engl J Med*. 2001 Apr 5;344(14):1038-42. doi: 10.1056/NEJM200104053441402. Erratum in: *N Engl J Med* 2001 Jul 19;345(3):232. PMID: 11287973.

Duy C, Hurtz C, Shojaee S, Cerchietti L, Geng H, Swaminathan S, Klemm L, Kweon SM, Nahar R, Braig M, Park E, Kim YM, Hofmann WK, Herzog S, Jumaa H, Koeffler HP, Yu JJ, Heisterkamp N, Graeber TG, Wu H, Ye BH, Melnick A, Müschen M. BCL6 enables Ph+ acute lymphoblastic leukaemia cells to survive BCR-ABL1 kinase inhibition. *Nature*. 2011 May 19;473(7347):384-8. doi: 10.1038/nature09883. PMID: 21593872; PMCID: PMC3597744.

El Achi, Hanadi et al. "CD123 as a Biomarker in Hematolymphoid Malignancies: Principles of Detection and Targeted Therapies." *Cancers* vol. 12,11 3087. 23 Oct. 2020, doi:10.3390/cancers12113087

Fabian, D. K., et al. (2021). "Functional conservation in genes and pathways linking ageing and immunity." *Immunity & Ageing* 18(1).

Feng, J., et al. (2023). "Loss of bisecting GlcNAcylation on MCAM of bone marrow stroma determined pro-tumoral niche in MDS/AML." *Leukemia* 37(1): 113-121.

Fielding, Adele K et al. "Prospective outcome data on 267 unselected adult patients with Philadelphia chromosome-positive acute lymphoblastic leukemia confirms superiority of allogeneic transplantation over chemotherapy in the pre-imatinib era: results from the International ALL Trial MRC UKALLXII/ECOG2993." *Blood* vol. 113,19 (2009): 4489-96.doi:10.1182/blood-2009-01-199380

Francesca, A., et al. (2020). "Targeting of the A2A adenosine receptor counteracts immunosuppression in vivo in a mouse model of chronic lymphocytic leukemia." *Haematologica* 106(5): 1343-1353.

García-Gutiérrez, Valentín, and Juan Carlos Hernández-Boluda. "Tyrosine Kinase Inhibitors Available for Chronic Myeloid Leukemia: Efficacy and Safety." *Frontiers in oncology* vol. 9 603. 3 Jul. 2019, doi:10.3389/fonc.2019.00603

Garnache-Ottou, Francine et al. "Expression of the myeloid-associated marker CD33 is not an exclusive factor for leukemic plasmacytoid dendritic cells." *Blood* vol. 105,3 (2005): 1256-64.doi:10.1182/blood-2004-06-2416

Glatman Zaretsky, Arielle et al. "Infection-induced changes in hematopoiesis." *Journal of immunology* (Baltimore, Md. : 1950) vol. 192,1 (2014): 27-33. doi:10.4049/jimmunol.1302061

Green, T., et al. (2013). "TMEFF2 and SARDH cooperate to modulate one-carbon metabolism and invasion of prostate cancer cells." *The Prostate* 73(14): 1561-1575.

Harkness, L., et al. (2016). "CD146/MCAM defines functionality of human bone marrow stromal stem cell populations." *Stem Cell Research & Therapy* 7(1).

Horne, G & Copland, M (2017) Approaches for targeting self-renewal pathways in cancer stem cells: implications for hematological treatments, *Expert Opinion on Drug Discovery*, 12:5, 465- 474, DOI: 10.1080/17460441.2017.1303477

Houshmand, M., et al. (2019). "Bone marrow microenvironment: The guardian of leukemia stem cells." *World Journal of Stem Cells* 11(8): 476-490.

Houshmand, Mohammad et al. "Chronic myeloid leukemia stem cells." *Leukemia* vol. 33,7 (2019): 1543-1556. doi:10.1038/s41375-019-0490-0

Hovorkova, Lenka et al. "Monitoring of childhood ALL using BCR-ABL1 genomic breakpoints identifies a subgroup with CML-like biology." *Blood* vol. 129,20 (2017): 2771-2781.doi:10.1182/blood-2016-11-749978

Hsieh, Y.-C., et al. (2021). "Improving outcomes in chronic myeloid leukemia through harnessing the immunological landscape." *Leukemia* 35(5): 1229-1242.

Hu, M. C., et al. (2015). "Klotho, stem cells, and aging." *Clinical Interventions in Aging*: 1233.

Hunger, Stephen P, and Charles G Mullighan. "Redefining ALL classification: toward detecting high-risk ALL and implementing precision medicine." *Blood* vol. 125,26 (2015): 3977-87.doi:10.1182/blood-2015-02-580043

Hunger, Stephen P. "CML in blast crisis: more common than we think?." *Blood* vol. 129,20 (2017): 2713-2714. doi:10.1182/blood-2017-04-776369

Hunger, S. CML in blast crisis: more common than we think?. *Blood* 2017; 129 (20): 2713–2714. doi: <https://doi.org/10.1182/blood-2017-04-776369>

Issa, N. (2022). "Making a Case for the Anti-inflammatory Effects of ACE Inhibitors and Angiotensin II Receptor Blockers: Evidence From Randomized Controlled Trials." *Mayo Clinic Proceedings* 97(10): 1766-1768.

Janovská, P. and V. Bryja (2017). "Wnt signalling pathways in chronic lymphocytic leukaemia and B-cell lymphomas." *British Journal of Pharmacology* 174(24): 4701-4715.

Jeanpierre, S., et al. (2020). "The quiescent fraction of chronic myeloid leukemic stem cells depends on BMPR1B, Stat3 and BMP4-niche signals to persist in patients in remission." *Haematologica* 106(1): 111-122.

Jiang, D., et al. (2021). "PRICKLE1, a Wnt/PCP signaling component, is overexpressed and associated with inferior prognosis in acute myeloid leukemia." *Journal of Translational Medicine* 19(1).

Kadomoto, S., et al. (2020). "The CCL20-CCR6 Axis in Cancer Progression." *International Journal of Molecular Sciences* 21(15): 5186.

Kadowaki, I., Ichinohasama, R., Harigae, H., Ishizawa, K., Okitsu, Y., Kameoka, J. and Sasaki, T. (2005), Accelerated lymphangiogenesis in malignant lymphoma: possible role of VEGF-A and VEGF-C. *British Journal of Haematology*, 130: 869-877. <https://doi.org/10.1111/j.1365-2141.2005.05695.x>

Kakadia, P. M., et al. (2018). "Efficient identification of somatic mutations in acute myeloid leukaemia using whole exome sequencing of fingernail derived DNA as germline control." *Scientific Reports* 8(1).

Kaksonen, M., Roux, A. "Mechanisms of clathrin-mediated endocytosis". *Nat Rev Mol Cell Biol* 19, 313–326 (2018). <https://doi.org/10.1038/nrm.2017.132>

Kamoda, Yoshimasa et al. "Philadelphia Chromosome-Positive Acute Lymphoblastic Leukemia Is Separated into Two Subgroups Associated with Survival by BCR-ABL Fluorescence in situ Hybridization of Segmented Cell Nuclei: Report from a Single Institution." *Acta haematologica* vol. 136,3 (2016): 157-66. doi:10.1159/000445972

Kang, Zhi-Jie et al. "The Philadelphia chromosome in leukemogenesis." *Chinese journal of cancer* vol. 35 48. 27 May. 2016, doi:10.1186/s40880-016-0108-0

Karsunky, Holger et al. "Flk2+ common lymphoid progenitors possess equivalent differentiation potential for the B and T lineages." *Blood* vol. 111,12 (2008): 5562-70. doi:10.1182/blood-2007-11-126219

Kinstrie, Ross et al. "CD93 is expressed on chronic myeloid leukemia stem cells and identifies a quiescent population which persists after tyrosine kinase inhibitor therapy." *Leukemia* vol. 34,6 (2020): 1613-1625. doi:10.1038/s41375-019-0684-5

Kisselbach, Lynn et al. "CD90 Expression on human primary cells and elimination of contaminating fibroblasts from cell cultures." *Cytotechnology* vol. 59,1 (2009): 31-44. doi:10.1007/s10616-009-9190-3

Kolenova, Alexandra et al. "Philadelphia Chromosome-positive Acute Lymphoblastic Leukemia or Chronic Myeloid Leukemia in Lymphoid Blast Crisis." *Journal of pediatric hematology/oncology* vol. 38,6 (2016): e193-5. doi:10.1097/MPH.0000000000000582

Kolenova, Alexandra et al. "Philadelphia Chromosome-positive Acute Lymphoblastic Leukemia or Chronic Myeloid Leukemia in Lymphoid Blast Crisis." *Journal of pediatric hematology/oncology* vol. 38,6 (2016): e193-5. doi:10.1097/MPH.0000000000000582

Koo, Hong Hoe. "Philadelphia chromosome-positive acute lymphoblastic leukemia in childhood." *Korean journal of pediatrics* vol. 54,3 (2011): 106-10. doi:10.3345/kjp.2011.54.3.106

Koschmieder, Steffen, and Mirle Schemionek. "Mouse models as tools to understand and study BCR-ABL1 diseases." *American journal of blood research* vol. 1,1 (2011): 65-75.

Kumar, R. and D. S. Krause (2021). "Recent advances in understanding chronic myeloid leukemia: where do we stand?" *Faculty Reviews* 10.

Kunz, J. B., et al. (2015). "Pediatric T-cell lymphoblastic leukemia evolves into relapse by clonal selection, acquisition of mutations and promoter hypomethylation." *Haematologica* 100(11):1442-1450.

Kwabi-Addo, B., et al. (2004). "The role of fibroblast growth factors and their receptors in prostate cancer." *Endocrine-Related Cancer* 11(4): 709-724.

Lee, Ji Yoon, and Seok-Ho Hong. "Hematopoietic Stem Cells and Their Roles in Tissue Regeneration." *International journal of stem cells* vol. 13,1 (2020): 1-12. doi:10.15283/ijsc19127

Lee, S.-H., et al. (2017). "High expression of <i>dedicator of cytokinesis 1</i> (<i>DOCK1</i>) confers poor prognosis in acute myeloid leukemia." *Oncotarget* 8(42): 72250-72259.

Lemes, A et al. "p190 BCR-ABL rearrangement in chronic myeloid leukemia and acute lymphoblastic leukemia." *Cancer genetics and cytogenetics* vol. 113,1 (1999): 100-2. doi:10.1016/s0165-4608(99)00014-x

Lieu, Yen K, and E Premkumar Reddy. "Impaired adult myeloid progenitor CMP and GMP cell function in conditional c-myb-knockout mice." *Cell cycle (Georgetown, Tex.)* vol. 11,18 (2012): 3504-12. doi:10.4161/cc.21802

Liu, P., et al. (2018). "Absence of AIF1L contributes to cell migration and a poor prognosis of breast cancer." *OncoTargets and Therapy* Volume 11: 5485-5498.

Liu, X., et al. (2023). "SNAI2 Attenuated the Stem-like Phenotype by Reducing the Expansion of EPCAMhigh Cells in Cervical Cancer Cells." *International Journal of Molecular Sciences* 24(2): 1062.

Migocka-Patrzałek, M. and M. Elias (2021). "Muscle Glycogen Phosphorylase and Its Functional Partners in Health and Disease." *Cells* 10(4): 883.

Mishra, Deepshikha et al. "Role of B Cell Development Marker CD10 in Cancer Progression and Prognosis." *Molecular biology international* vol. 2016 (2016): 4328697. doi:10.1155/2016/4328697

Mojtahedi, H., et al. (2021). "Chronic myeloid leukemia stem cells: targeting therapeutic implications." *Stem Cell Research & Therapy* 12(1). Mojtahedi, Hanieh et al. "Chronic myeloid

leukemia stem cells: targeting therapeutic implications." *Stem cell research & therapy* vol. 12,1 603. 18 Dec. 2021, doi:10.1186/s13287-021-02659-1

Montaño, Adrián et al. "ETV6/RUNX1 Fusion Gene Abrogation Decreases the Oncogenicity of Tumour Cells in a Preclinical Model of Acute Lymphoblastic Leukaemia." *Cells* vol. 9,1 215. 15 Jan. 2020, doi:10.3390/cells9010215

Nagel, Inga et al. "Hematopoietic stem cell involvement in BCR-ABL1-positive ALL as a potential mechanism of resistance to blinatumomab therapy." *Blood* vol. 130,18 (2017): 2027-2031. doi:10.1182/blood-2017-05-782888235

Nakahata, S. and K. Morishita (2012). "CADM1/TSLC1 is a Novel Cell Surface Marker for Adult T-Cell Leukemia/Lymphoma." *Journal of Clinical and Experimental Hematopathology* 52(1): 17-22.

Nishiwaki, Satoshi et al. "Multi-Lineage BCR-ABL Expression in Philadelphia Chromosome- Positive Acute Lymphoblastic Leukemia Is Associated With Improved Prognosis but No Specific Molecular Features." *Frontiers in oncology* vol. 10 586567. 23 Oct. 2020, doi:10.3389/fonc.2020.586567

Nishiwaki, Satoshi et al. "Multi-Lineage BCR-ABL Expression in Philadelphia Chromosome- Positive Acute Lymphoblastic Leukemia Is Associated With Improved Prognosis but No Specific Molecular Features." *Frontiers in oncology* vol. 10 586567. 23 Oct. 2020, doi:10.3389/fonc.2020.586567

Nowell, Peter C. "Discovery of the Philadelphia chromosome: a personal perspective." *The Journal of clinical investigation* vol. 117,8 (2007): 2033-5. doi:10.1172/JCI31771

Pagliaro, L., et al. (2021). "Targeting oncogenic Notch signaling with SERCA inhibitors." *Journal of Hematology & Oncology* 14(1).

Patnaik, M. M., et al. (2018). "Biallelic inactivation of the retinoblastoma gene results in transformation of chronic myelomonocytic leukemia to a blastic plasmacytoid dendritic cell neoplasm: shared clonal origins of two aggressive neoplasms." *Blood Cancer Journal* 8(9).

Paul, Matt R et al. "Multivariate models from RNA-Seq SNVs yield candidate molecular targets for biomarker discovery: SNV-DA." *BMC genomics* vol. 17 263. 31 Mar. 2016, doi:10.1186/s12864-016-2542-4

Peng, Hong-Xia et al. "Upregulation of the proto-oncogene Bmi-1 predicts a poor prognosis in pediatric acute lymphoblastic leukemia." *BMC cancer* vol. 17,1 76. 25 Jan. 2017, doi:10.1186/s12885-017-3049-3

Pietras, Eric M et al. "Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions." *Cell stem cell* vol. 17,1 (2015): 35-46. doi:10.1016/j.stem.2015.05.003

Pietras, Eric M et al. "Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions." *Cell stem cell* vol. 17,1 (2015): 35-46. doi:10.1016/j.stem.2015.05.003

Pu, Shiming et al. "Identification of early myeloid progenitors as immunosuppressive cells." *Scientific reports* vol. 6 23115. 16 Mar. 2016, doi:10.1038/srep23115

Public Health England. Children, teenagers and young adults UK cancer statistics report 2021. Available From http://ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/cancer_in_children_teenagers_and_young_adults/(link is external), accessed March 2021.

Puckett Y, Chan O. Acute Lymphocytic Leukemia. [Updated 2023 Jun 26]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK459149/>

Pui, C.-H., Relling, M.V. & Downing, J.R. (2004) Acute Lymphoblastic Leukemia. *New England Journal of Medicine*. [Online] 350 (15), 1535–1548. Available from:doi:10.1056/nejmra023001.

Quintás-Cardama, Alfonso, and Jorge Cortes. "Molecular biology of bcr-abl1-positive chronic myeloid leukemia." *Blood* vol. 113,8 (2009): 1619-30. doi:10.1182/blood-2008-03144790

Radtke, Stefan et al. "Isolation of a Highly Purified HSC-enriched CD34+CD90+CD45RA- Cell Subset for Allogeneic Transplantation in the Nonhuman Primate Large-animal Model." *Transplantation direct* vol. 6,8 e579. 15 Jul. 2020, doi:10.1097/TXD.0000000000001029236

Reckel, S et al. "Differential signaling networks of Bcr-Abl p210 and p190 kinases in leukemia cells defined by functional proteomics." *Leukemia* vol. 31,7 (2017): 1502-1512. doi:10.1038/leu.2017.36

Rodriguez-Fraticelli, Alejo E et al. "Clonal analysis of lineage fate in native haematopoiesis." *Nature* vol. 553,7687 (2018): 212-216. doi:10.1038/nature25168

Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*. 1973 Jun 1;243(5405):290-3. doi: 10.1038/243290a0. PMID: 4126434.

Rudnick, Stephen I, and Gregory P Adams. "Affinity and avidity in antibody-based tumor targeting." *Cancer biotherapy & radiopharmaceuticals* vol. 24,2 (2009): 155-61.doi:10.1089/cbr.2009.0627

Saygin, Caner et al. "Measurable residual disease in acute lymphoblastic leukemia: methods and clinical context in adult patients." *Haematologica* vol. 107,12 2783-2793. 1 Dec. 2022, doi:10.3324/haematol.2022.280638

Schroeder, T. (2010). "Hematopoietic Stem Cell Heterogeneity: Subtypes, Not Unpredictable Behavior." *Cell Stem Cell* 6(3): 203-207.

Schultz, Kirk R et al. "Improved early event-free survival with imatinib in Philadelphia chromosome-positive acute lymphoblastic leukemia: a children's oncology group study." *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* vol. 27,31 (2009): 5175-81. doi:10.1200/JCO.2008.21.2514

Seita, Jun, and Irving L Weissman. "Hematopoietic stem cell: self-renewal versus differentiation." *Wiley interdisciplinary reviews. Systems biology and medicine* vol. 2,6 (2010): 640-53. doi:10.1002/wsbm.86

Senapati, Jayastu, and Koji Sasaki. "Chromosomal Instability in Chronic Myeloid Leukemia: Mechanistic Insights and Effects." *Cancers* vol. 14,10 2533. 21 May. 2022, doi:10.3390/cancers14102533

Smith MA, Altekruze SF, Adamson PC, et al.: Declining childhood and adolescent cancer mortality. *Cancer* 120 (16): 2497-506, 2014. [PUBMED Abstract]

Sopper, S., et al. (2017). "Reduced CD62L Expression on T Cells and Increased Soluble CD62L Levels Predict Molecular Response to Tyrosine Kinase Inhibitor Therapy in Early Chronic-Phase Chronic Myelogenous Leukemia." *Journal of Clinical Oncology* 35(2): 175-184.

Soverini, S., et al. (2019). "Treatment and monitoring of Philadelphia chromosome positive leukemia patients: recent advances and remaining challenges." *Journal of Hematology & Oncology* 12(1).

Stacker, S., Williams, S., Karnezis, T. et al. Lymphangiogenesis and lymphatic vessel remodelling in cancer. *Nat Rev Cancer* 14, 159–172 (2014). <https://doi.org/10.1038/nrc3677>

Stamm, H., et al. (2018). "Immune checkpoints PVR and PVRL2 are prognostic markers in AML and their blockade represents a new therapeutic option." *Oncogene* 37(39): 5269-5280.

Sudutan, T., et al. (2022). "Zinc finger protein 384 (<i>ZNF384</i>) impact on childhood mixed phenotype acute leukemia and B-cell precursor acute lymphoblastic leukemia." *Leukemia & Lymphoma* 63(12): 2931-2939.

Sullivan, A. L., et al. (2008). "Neutrophil elastase reduces secretion of secretory leukoproteinase inhibitor (SLPI) by lung epithelial cells: role of charge of the proteinase-inhibitor complex." *Respiratory Research* 9(1): 60.

Sun, Jianlong et al. "Clonal dynamics of native haematopoiesis." *Nature* vol. 514,7522 (2014): 322-7. doi:10.1038/nature13824

Sun, Y., et al. (2017). "SERPINB3 and B4: From biochemistry to biology." *Seminars in Cell & Developmental Biology* 62: 170-177.237

Sureda-Gómez, M., et al. (2023). "Tumorigenic role of Musashi-2 in aggressive mantle cell lymphoma." *Leukemia* 37(2): 408-421. Surveillance, Epidemiology, and End Results Program: SEER Cancer Stat Facts: Childhood

Leukemia (Ages 0–19). Bethesda, Md: National Cancer Institute, DCCPS, Surveillance Research Program. Available online. Last accessed September 7, 2022.

Szczeпаński, T et al. "Minimal residual disease in leukaemia patients." *The Lancet. Oncology* vol. 2,7 (2001): 409-17. doi:10.1016/s1470-2045(00)00418-6

Thomas, Xavier. "Philadelphia chromosome-positive leukemia stem cells in acute lymphoblastic leukemia and tyrosine kinase inhibitor therapy." *World journal of stem cells* vol. 4,6 (2012): 44-52. doi:10.4252/wjsc.v4.i6.44

Thomson RJ, Moshirfar M, Ronquillo Y. Tyrosine Kinase Inhibitors. [Updated 2023 Jul 18]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK563322/>

Tran, Thai Hoa, and Mignon L Loh. "Ph-like acute lymphoblastic leukemia." *Hematology. American Society of Hematology. Education Program* vol. 2016,1 (2016): 561-566. doi:10.1182/asheducation-2016.1.561

Vairy, Stephanie, and Thai Hoa Tran. "IKZF1 alterations in acute lymphoblastic leukemia: The good, the bad and the ugly." *Blood reviews* vol. 44 (2020): 100677. doi:10.1016/j.blre.2020.100677

Vetrie, David et al. "The leukaemia stem cell: similarities, differences and clinical prospects in CML and AML." *Nature reviews. Cancer* vol. 20,3 (2020): 158-173. doi:10.1038/s41568-019-0230-9

Voncken, J W et al. "BCR/ABL P210 and P190 cause distinct leukemia in transgenic mice." *Blood* vol. 86,12 (1995): 4603-11.

Wang, G., et al. (2020). "<p>CELSR1 Acts as an Oncogene Regulated by miR-199a-5p in Glioma</p>." *Cancer Management and Research* Volume 12: 8857-8865.

Wang, Jean Y J. "The capable ABL: what is its biological function?." *Molecular and cellular biology* vol. 34,7 (2014): 1188-97. doi:10.1128/MCB.01454-13

Wang, Kemeng et al. "CD19: a biomarker for B cell development, lymphoma diagnosis and therapy." *Experimental hematology & oncology* vol. 1,1 36. 29 Nov. 2012, doi:10.1186/2162-3619-1-36

Wang, Richard C, and Zhixiang Wang. "Precision Medicine: Disease Subtyping and Tailored Treatment." *Cancers* vol. 15,15 3837. 28 Jul. 2023, doi:10.3390/cancers15153837

Wang, Weijia et al. "Cytokine combinations for human blood stem cell expansion induce cell-type- and cytokine-specific signaling dynamics." *Blood* vol. 138,10 (2021): 847-857. doi:10.1182/blood.2020008386

Will, Britta, and Ulrich Steidl. "Multi-parameter fluorescence-activated cell sorting and analysis of stem and progenitor cells in myeloid malignancies." *Best practice & research. Clinical haematology* vol. 23,3 (2010): 391-401. doi:10.1016/j.beha.2010.06.006

Wingelhofer, Bettina et al. "Implications of STAT3 and STAT5 signaling on gene regulation and chromatin remodeling in hematopoietic cancer." *Leukemia* vol. 32,8 (2018): 1713-1726. doi:10.1038/s41375-018-0117-x

Wittwer, Nicole L et al. "High CD123 levels enhance proliferation in response to IL-3, but reduce chemotaxis by downregulating CXCR4 expression." *Blood advances* vol. 1,15 1067-1079. 20 Jun. 2017, doi:10.1182/bloodadvances.2016002931

Wu, Z., et al. (2012). "MCAM is a novel metastasis marker and regulates spreading, apoptosis and invasion of ovarian cancer cells." *Tumor Biology* 33(5): 1619-1628.

Wu, Z., et al. (2015). "HMGA2 as a potential molecular target in<i>KMT2A-AFF1</i>-positive infant acute lymphoblastic leukaemia." *British Journal of Haematology* 171(5): 818-829.