UNRAVELLING THE GENOMIC LANDSCAPE OF ACUTE LYMPHOBLASTIC LEUKAEMIA IN OLDER ADULTS

THOMAS CREASEY

A thesis submitted in part requirement for the degree of
Doctor of Philosophy from the Faculty of Medical Sciences
Newcastle University
Newcastle-upon-Tyne

Leukaemia Research Cytogenetics Group,
Translational and Clinical Research Institute

October 2020

Abstract

The objectives of this study were to characterise the primary genetic abnormalities, genomic copy number changes and mutational landscape of ALL in older adults.

The primary chromosomal abnormalities from patients aged ≥60 years recruited into the UKALL14 (n=94) and UKALL60+ (N=116) trials were first evaluated. B-cell precursor (BCP) ALL patients lacking a primary chromosomal abnormality (B-other ALL) were screened for ABL-class fusions, JAK-STAT abnormalities and other rearrangements using fluorescence *in situ* hybridisation (FISH). *CRLF2* and *ZNF384* rearrangements were detected in 17% and 7% of tested patients respectively. ABL-class fusions were notably absent.

Next, single nucleotide polymorphism (SNP) arrays were performed to identify copy number abnormalities in patients with suitable material (n=83). Deletions were detected in *IKZF1* (52%), *CDKN2A/B* (45%) and *PAX5* (39%), as well as arm level events including del(9p) (21%), monosomy 7 (10%) and gain 1q (10%). Selected novel abnormalities were then validated using a customised sequencing approach. Recurrent novel deletions were confirmed in *LEMD3*, *KDM6A* and *CXCR4*, potentially contributing to leukaemogenesis.

Separately, SNP arrays were performed on DNA from patients with low hypodiploidy or high hyperdiploidy (n=88) and machine-learning techniques were used to cluster cases based on log2 ratio data. Discrepancies between the cytogenetic-derived and SNP array-derived genetic subgroup were identified. A diagnostic classifier based on chromosomal log2 ratios was then designed using classification and regression tree analysis (CART).

Finally, the mutational landscape of ALL in older adults was characterised in selected patients using exome sequencing (n=6) and a customised sequencing panel (n=30). Pathogenic variants were identified in *TP53*, *NF1*, *JAK2* as well as members of the RAS signalling pathway, and were closely related to specific primary chromosomal abnormalities.

This project has helped characterise the landscape of genetic prognostic biomarkers in older adults with ALL, and identified novel therapeutically actionable abnormalities meriting further assessment.

Acknowledgments

I would firstly like to acknowledge my funders, specifically Bright Red and the NIHR Newcastle Biomedical Research Centre, together with the patients and staff involved in the UKALL clinical trials, without whom none of this work would have been possible.

From my very first days in the daunting world of leukaemia genetics, all members of the Leukaemia Research Cytogenetics Group have helped me, patiently answered many very basic questions and taught me how to perform all manner of experiments. I'd specifically like to thank Emilio Barretta for his hours of painstakingly teaching me the art of FISH and Amir Enshaei for his incredible help with coding, data science and statistics. Additionally, I'd like to thank James Murray for getting me through a number of library preps, Matt Bashton for all his bioinformatics input and Claire Schwab for regularly translating karyotypes into an intelligible language.

Separately, I would also like to acknowledge Claire Keeling, who has been a great friend over the last 3 years, and patiently listened to a fair amount of complaining, while keeping me well topped up with coffee throughout the day!

I could never have conducted this research without the moral support, supervision and kindness from my two supervisors, Anthony and Sarra. From the very start, Anthony's approach to supervision has been more encouraging than I could ever have hoped for. He has continually helped me steer my research during our weekly meetings, offered invaluable insights and interpretations and given me opportunities to collaborate with other leaders in the field. Sarra somehow regularly managed to find time in her hectic schedule to take me through the intricacies of molecular biology and sequencing, and guided me through designing and performing complicated experiments that would have been impossible without her expertise.

Finally, I would never have found the resilience to preserve through the more challenging periods of my PhD without the endless support from my amazing family – my parents and most importantly Nikki, our perfect daughter Eloise (who wasn't even thought of when I started this project!) and her little sister expected anytime in the next 3 weeks! I owe you all a huge debt of gratitude that I doubt I'll ever be able to repay!

Statement of work undertaken

FISH slide analysis was performed by myself with the second check performed by Emilio Barretta to ensure there was agreement in the pattern reported by two independent observers.

SNP arrays were performed by Kathryn Watts at the Newcastle Genomics Centre at the International Centre for Life, Newcastle Hospitals NHS Foundation Trust with downstream data analysis performed by myself. GISTIC significance testing of the arm-level copy number abnormalities was performed with the help of Dr Matthew Bashton.

Exome sequencing was performed by Jon Coxhead's team at the Core Genomics Facility, Newcastle University. DNA samples were provided, and all library prep and sequencing were performed by the Core Genomics Facility. Sequencing data were returned in FASTQ format and bioinformatic pipeline analysis using established variant callers was performed by Dr Matthew Bashton and John Caseman at the Bioinformatic Support Unit.

SureSelect XT2 library prep was performed by myself and libraries were then transferred to the Core Genomics Facility for sequencing. Data were returned in FASTQ format and bioinformatic analysis using HaplotypeCaller was formed by John Caseman at the Bioinformatic Support Unit.

All post bioinformatic processing of VEP output files and variant interpretation were performed by myself.

List of abbreviations

ALL	Acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
BAC	Bacterial artificial chromosome
BCP	B-cell precursor
CART	Classification and regression tree
CBS	Circular binary segmentation
CHIP	Clonal haematopoiesis of indeterminate potential
CML	Chronic myeloid leukaemia
CNA	Copy number abnormality
CNN	Copy number neutral
CNV	Copy number variation
DAPI	4’,6’-Diamidino-2-phenylindole dihydrochloride
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
ETP	Early T-precursor
FBS	Fetal bovine serum
FISH	Fluorescence *in situ* hybridisation
HeH	High hyperdiploidy
HMM	Hidden Markov model
HoTr	Low hypodiploidy/near triploidy
Ig	Immunoglobulin
IGV	Integrative genomics viewer
Indel	Insertion or deletion
LAIP	Leukaemia associated immunophenotype

LOH	Loss of heterozygosity
MDS	Myelodysplastic syndrome
MFC	Multiparametric flow cytometry
MLPA	Multiplex ligation-dependent probe amplification
MPN	Myeloproliferative neoplasm
MRD	Minimal residual disease
mRNA	Messenger RNA
NGS	Next generation sequencing
PBS	Phosphate buffered saline
PCA	Principle components analysis
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SSC	Saline-sodium citrate buffer
TCR	T-cell receptor
TKI	Tyrosine kinase inhibitor
VAF	Variant allele frequency
WCC	White cell count
WGA	Whole genome amplification
WHO	World Health Organisation

Table of Contents

Table of figures

Table of tables

Chapter 1. Introduction

1.1 Leukaemia

Leukaemia represents a heterogenous group of malignant diseases of the haematopoietic system, predominantly initiating in the bone marrow. In common with other cancers, leukaemias are genetic diseases, whereby one or more key aberrations occur within the DNA of haematopoietic cells, hijacking their normal function and leading to increased proliferation and survival (Hanahan and Weinberg, 2011).

Acute leukaemias are characterised by differentiation arrest and abnormal proliferation of a population of immature progenitors (blasts). This results in heavy bone marrow infiltration with a consequent failure of normal haematopoiesis (De Kouchkovsky and Abdul-Hay, 2016). Left untreated, patients rapidly succumb to the effects of profound immunosuppression and bone marrow failure. In comparison, chronic leukaemias usually present more indolently and are characterised by a much lower rate of proliferation and at least some retention of normal cellular function and differentiation capabilities (Jabbour and Kantarjian, 2020).

The two principal subtypes of acute leukaemia are acute myeloid leukaemia (AML) and acute lymphoblastic leukaemia (ALL). In AML, the blasts express myeloid antigens and may progress from pre-leukaemic myeloid neoplasms such as myelodysplastic syndrome (MDS) or myeloproliferative neoplasms (MPNs) (Arber et al., 2016). The incidence of AML follows the patterns of most cancers and rises with age, most likely due to the accrual of DNA damage over many years (Jan et al., 2012). In comparison ALL almost always arises de novo, with only rare cases of secondary ALL described (Rosenberg et al., 2017). Moreover, the disease incidence has a very different age-profile with a high peak in early childhood (Inaba et al., 2013), implying a distinct pathophysiology from most other malignancies.

1.2 Acute lymphoblastic leukaemia overview

ALL is a subtype of acute leukaemia affecting precursor B or T cells in the bone marrow. Clinically, the disease progresses rapidly, with leukaemic blasts usually detected in the peripheral blood and often accompanied by life-threatening anaemia, thrombocytopenia and infection. Fortunately, the management of ALL in children has been a paradigm for the success of translational medical research in improving patient outcomes. Successive large clinical trials have improved the 5-year overall survival from 30% in the 1960s to 90% in the modern era, through the adoption of better supportive care, risk stratification and intensified treatment (Inaba et al., 2013). In comparison, adult ALL has not yet seen the same degree of research or improvement in prognosis, and fewer than 50% of affected patients are currently cured of their disease (Rowe et al., 2005, Marks et al., 2019).

1.3 Epidemiology

In the UK, approximately 700-800 patients are diagnosed with ALL each year (Cancer Research UK, 2016). Three quarters of adults and 85% of children present with B-cell precursor ALL (BCP-ALL). T-cell ALL (T-ALL) is rarer and accounts for the remaining 25% and 15% of adult and paediatric patients respectively (Chiaretti and Foà, 2009).

The overall UK incidence of ALL is stable at 1.2 cases per 100,000 population per year and males are slightly more commonly affected than females with a male to female ratio of 1.4:1 (Cancer Research UK, 2016). The peak incidence is seen in early childhood with an average of 250 new cases per year in UK children aged 0-4 years. The disease is rarer in adults and becomes eclipsed by the more common AML. However, the incidence of ALL has a bimodal distribution, which increases again after the age of 60 years (figure 1.1) (Moorman et al., 2010), with approximately 120-150 new cases per year in UK adults aged over 60 years. Indeed the age-specific annual incidence in those aged 60 years and over is 0.9–1.6 per 100,000, compared to 0.4–0.6 per 100,000 in younger adults (Sive et al., 2012, Larson, 2005, Taylor et al., 1992).

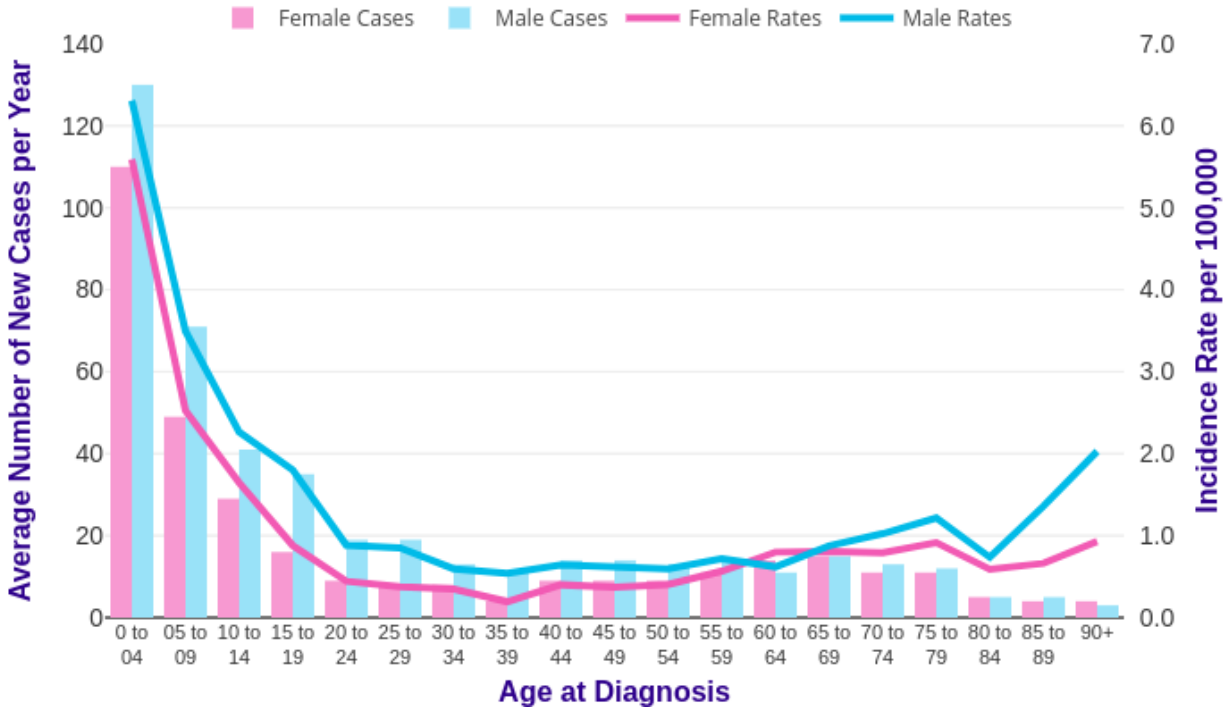


Figure 1.1. Age-specific incidence of ALL. Average number of new cases per year and age-specific incidence rates per 100,000 population split by gender, UK, 2014-2016 (taken from Cancer Research UK, cruk.org/cancerstats)

The care of older adults with ALL is an area of unmet clinical need. Although older adults comprise 20-25% of new diagnoses each year, around 60% of disease-related deaths are seen in this age group, and they are the only ALL patients not to have benefitted from the stepwise improvements in prognosis driven through successive clinical trials in children and younger adults (Dinmohamed et al., 2016).

1.4 Pathophysiology

ALL arises through sequential genetic lesions in precursor B or T cells. The mechanisms by which these aberrations arise has not been fully elucidated and is likely to differ between adults and children.

To date, the most robust evidence of an early initiating event resulting in a preleukaemic clone has been provided by the discovery that *ETV6-RUNX1* transcripts can be retrospectively detected in newborn Guthrie cards of patients presenting with ALL with t(12;21)(p13;q22)/*ETV6-RUNX1* in later childhood (Mori et al., 2002).

Moreover, both the several year latency and the observation that such transcripts can be found in up to 1% of Guthrie cards (more than 100 times the incidence of *ETV6-RUNX1+* ALL) confirms that additional co-operating secondary mutations are required for leukaemogenesis. Additionally, a high rate of concordance of specific cytogenetic ALL subtypes has been observed in monozygotic twins (Greaves et al., 2003, Ford et al., 1998). Interestingly, in monozygotic twins concordant for ALL, the same fusion gene with identical breakpoints can be found in both affected twins, unequivocally proving that preleukaemic or leukaemic cells metastasise through a shared placental circulation, rather than the concordance being related to a particularly strong germline predisposition and occurring independently in each twin.

Aside from genetic leukaemia predisposition syndromes, such as Down's syndrome or Ataxia Telangectasia, inherited susceptibility to ALL is likely to play at most a very modest role in its pathogenesis (Inaba et al., 2013). Genome-wide association studies have identified germline single nucleotide polymorphisms (SNPs) conferring only a small increased risk of ALL (Papaemmanuil et al., 2009). Common low penetrance variants in *IKZF1*, *ARID5B* and *CEPBE* were associated with odds ratios for childhood ALL of 1.69, 1.65 and 1.34 respectively.

An initiating somatic lesion or germline predisposition alone is insufficient to cause overt disease. A number of co-operating secondary events are required, often affecting genes in cell cycle regulation, B-cell development, or kinase signalling among others (Iacobucci and Mullighan, 2017). The mechanism of these events is also subject to debate. Plausibly, as with many cancers, environmental exposures are likely to play a part. Exposure to ionising radiation has been proven to have a causative role as evidenced by the increased rate of childhood ALL cases seen following the 1945 atomic bombing of Hiroshima and Nagasaki, Japan (Preston et al., 1994). Such radiation exposures are unlikely to be relevant currently, and the impact of non-ionising electromagnetic radiation in leukaemogenesis has been contentious (Inaba et al., 2013). Rather, the role of infectious exposures has proven a more plausible hypothesis for the conversion of a preleukaemic clone to overt ALL. This has been supported by observations that children exposed to common infections during infancy have a lower risk of ALL in early childhood years than those sheltered from such exposures, through a variety of common societal practices (e.g. childcare in home environment, often linked to higher socio-economic status) (Greaves et al., 1985). It is hypothesised that a subsequent delayed exposure to common infections during early childhood produces

an abnormal immune response resulting in the conversion of an existing and persistent preleukaemic clone to overt leukaemia (Greaves delayed infection hypothesis), although the specific mechanism is yet to be elucidated (Greaves, 2006).

These comprehensive aetiological studies have focussed on childhood ALL and currently very little evidence has been produced for the pathophysiology of ALL in adults and specifically older adults. In utero initiation of preleukaemic clones, germline predispositions and early childhood infections seem unlikely to be responsible for ALL occurring in later life.

1.5 Treatment and prognosis

The advances in the successful treatment of ALL in children since the 1960s has been one of the great achievements of translational and clinical research. Successive national and international collaborative clinical trials have seen the 5-year overall survival of childhood ALL increase from 35% in the first UK-wide ALL trial (UKALL1 (Medical Research Council, 1973)) to >90% in the most recently reported UKALL 2003 trial (Vora et al., 2013). Treatment of childhood ALL is now based on dose-intensive multi-agent chemotherapy accompanied by prolonged asparagine depletion and prophylactic CNS directed therapy throughout treatment. Treatment delays to allow bone marrow recovery are kept to a minimum between courses, and multiple prognostic factors are incorporated in protocols to inform either intensification or reduction of therapy at defined time points. Such is the success of multi-agent chemotherapy that allogeneic stem cell transplantation is rarely required in childhood ALL and is reserved for the relapse setting.

In recent years, younger adults with ALL have benefitted from the adoption of similar ‘paediatric-inspired’ protocols (Stock et al., 2008), although the success of these strategies has been less pronounced, and the use of stem cell transplantation in first remission with the aim of achieving long-term disease-free survival remains common.

In comparison, very little consistency exists in the treatment of older adults with ALL. Adults aged ≥60 years have frequently been excluded from clinical trials and are often not referred to specialist treatment centres (Gokbuget, 2017). Existing co-morbidities and frailty frequently preclude the use of intensive chemotherapy and/or stem cell

transplant and the use of anthracyclines and asparaginase is often associated with unacceptable morbidity and mortality. The 5-year overall survival of adults aged ≥60 years at ALL diagnosis is currently approximately 20% (Guru Murthy et al., 2015, Moorman et al., 2010). Typical treatment schedules for children and frail older patients are outlined in figure 1.2.

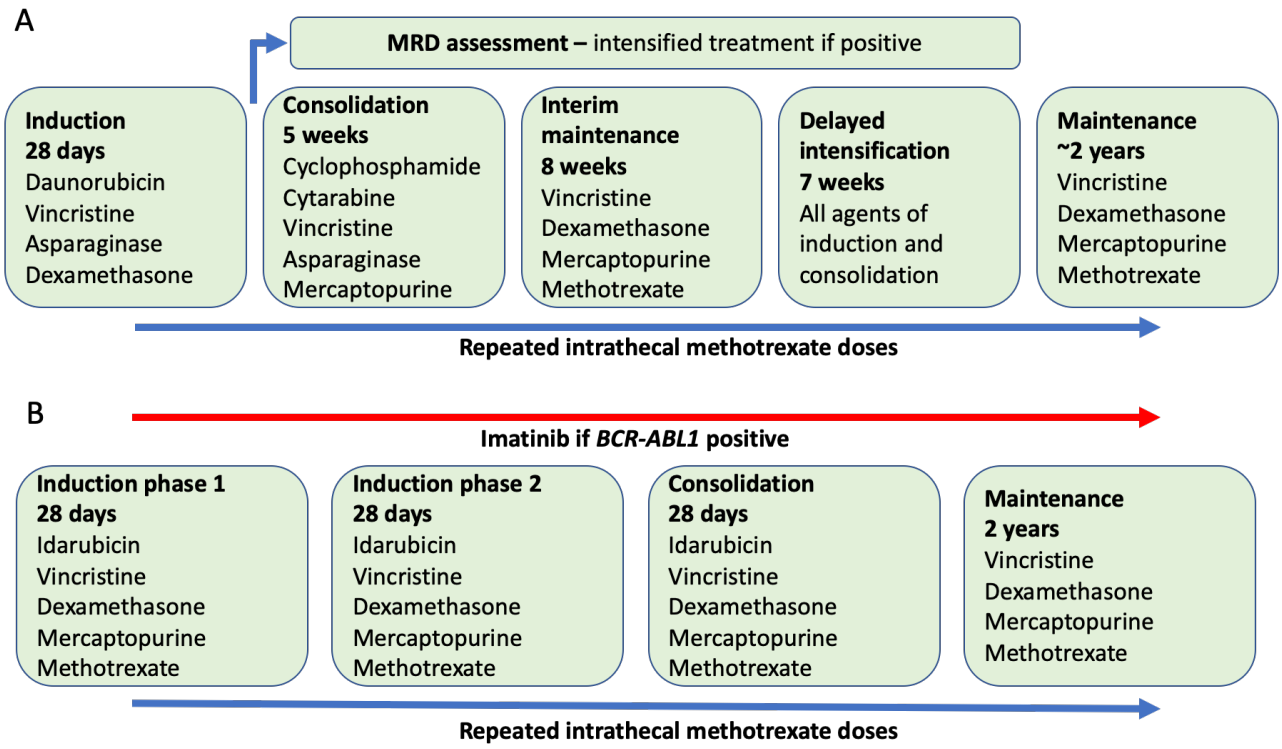


Figure 1.2. Example overview treatment schedules for paediatric patients (regimen B of UKALL2011 protocol) (A) and frail older adults (UKALL60+ trial protocol) (B). Treatment cycles are shorter in older adults and doses of anthracyclines and steroids are much lower. A number of chemotherapeutic agents are avoided due to toxicity (cyclophosphamide, cytarabine, asparaginase) and treatment is not minimal residual disease (MRD)-adapted.

1.6 Risk stratification

The vital importance of accurately identifying subgroups of ALL patients with different prognosis has been apparent for several decades and has led to intensified treatment for those with high risk features and more recently, treatment reduction for those with the best prognosis (Vora et al., 2013). A number of clinical and biological risk factors can be identified, which can inform decisions on treatment intensity and risk of relapse.

1.6.1 Sex

Historically, male sex has been associated with reduced event free survival (EFS) (Pui et al., 1999, Shuster et al., 1998). The reasons for this difference are poorly explained. Less intensive protocols from 1960s and 1970s were associated with a high risk of testicular relapse (up to 9%), although this has been <1% since the mid-1980s (Pui et al., 1999). Subsequently, persisting sex related differences in prognosis were explained by an increased rate of T cell disease in males and possible lower frequency of favourable cytogenetics (Pui et al., 1999), although modern protocols appear to have also largely abolished this residual disparity in prognosis (Pui et al., 2008, Pui et al., 2004, Vora et al., 2013). Some modern protocols still use a longer maintenance period for male patients, although this practice is not uniform. In comparison, modern protocols for adults and older adults make no distinction in the treatment of male and female patients, and a large study of more than 1500 did not show sex to be an independent predictor for overall or disease-free survival in adult ALL (Rowe et al., 2005).

1.6.2 Age and white cell count at diagnosis

In both adult and paediatric ALL population, age at diagnosis is consistently related to prognosis. Children aged ≥1 and <10 years at diagnosis have a better prognosis than those aged ≥10 years or <1 year (Chessells et al., 1998, Pui et al., 2008). Similarly, adults, aged >35-40 years have a worse prognosis than their younger counterparts (Chessells et al., 1998, Thomas et al., 2004b, Pui et al., 2008).

Along with age, white cell count (WCC) has been one of the earliest features found to be associated with prognosis, with high white cell counts at diagnosis of ALL associated with poorer outcome (Chessells et al., 1995). WCC has retained its independent prognostic value, even in the era of multiple, often co-operating and inter-linked risk features (Vaitkevičienė et al., 2011). In modern paediatric protocols, age and WCC are usually combined as an NCI risk stratification, where patients aged ≥1 and <10 years with presenting WCC <50 $x10^9$/L are designated NCI standard risk and all others NCI high risk. The basis for this can be seen from the results of the UKALLXI trial for paediatric ALL, where patients aged <10 years with a presenting WCC <50 $x10^9$/L fared significantly better than other groups (figure 1.3).

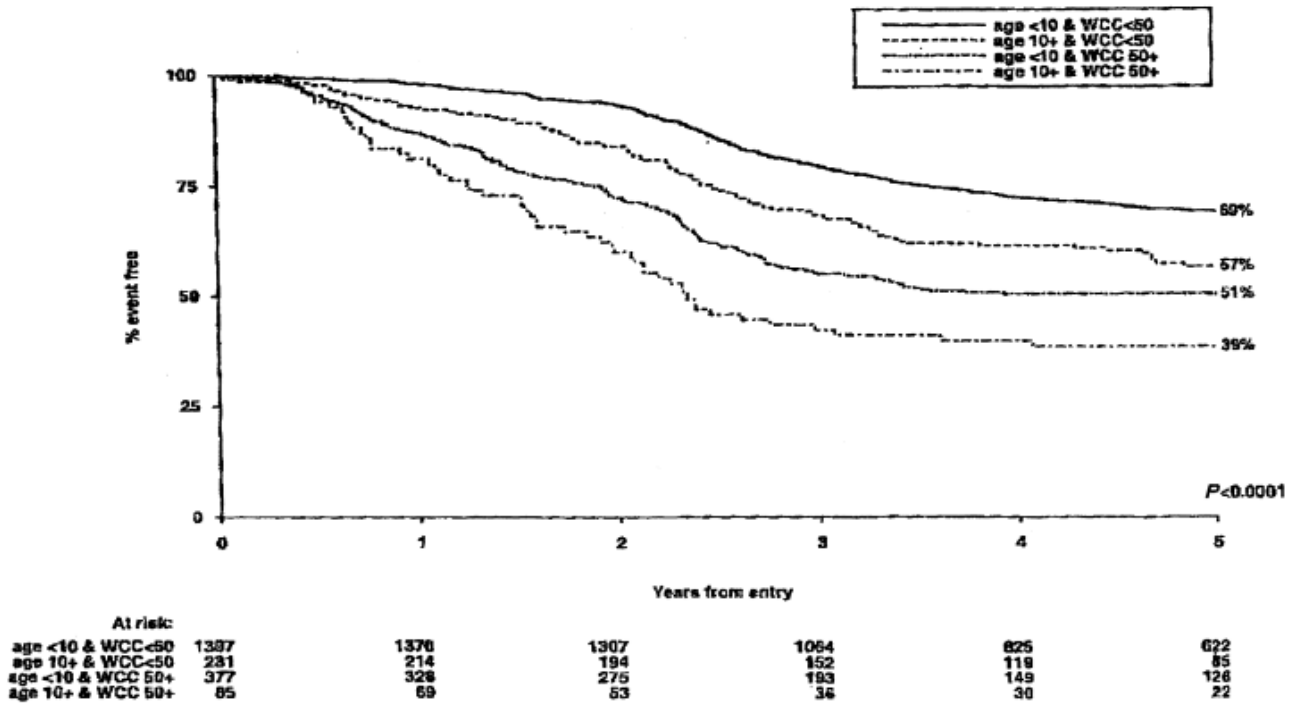


Figure 1.3. Event free survival (EFS) by age and white cell count (WCC) in UKALLXI trial. Patients aged <10 years and with WCC <50x109/L had significantly higher EFS compared with other groups. Taken from Hann et al (Hann et al., 2001).

From 1997 onwards (ALL97/99 trial), all UK clinical trials for paediatric ALL patients have incorporated NCI risk as the first separator of treatment arms with differing intensity.

Similarly, in adults, age >35-40 years and/or WCC >30x10^9/L for BCP-ALL or >100x10^9/L for T-ALL have been associated with high risk disease, and affected patients considered for allogeneic stem cell transplantation in first remission (Thomas et al., 2004b, Goldstone et al., 2008, Rowe et al., 2005). The large UKALLXII/ECOG2993 study identified age above vs below 35 years and WCC above or below 30x10^9/L and 100x10^9/L for BCP and T-cell disease respectively to be two of the strongest independent predictors of survival. This persisted in a multivariate analysis where these were analysed as continuous variables, predictably demonstrating that the poor outcome continues to worsen with advancing age beyond 35 years (Rowe et al., 2005).

However, the relationship between age and prognosis is complex and closely related to multiple other co-operating risk factors, particularly in older adults, which are both patient and therapy-related. For example, adults aged over 60 years at diagnosis often have both high risk biological features and increased rates of treatment related morbidity and mortality (Pui et al., 2008). Most patients aged under 60 years old at

diagnosis will be deemed fit for intensive therapy with curative intent. However, the situation is more complex in older adults, and historically, many such patients have been managed much more palliatively, with low expectations of durable remissions (Taylor et al., 1992, Gökbuget, 2013).

1.6.3 Cytogenetics

The study of chromosomal abnormalities in leukaemic cells has been a key feature of both diagnosis and prognosis since the discovery that the Philadelphia chromosome in chronic myeloid leukaemia (CML) arises as a result of a balanced chromosomal translocation (Rowley, 1973). Detailed karyotyping of leukaemia samples has since permitted the identification of a number of non-random chromosomal aberrations in ALL and this has led to a clear association between specific cytogenetic lesions and prognosis. Specifically, hyperdiploidy of ALL cells was first noted to be associated with a better outcome than diploid or other ploidy shifts (Secker-Walker et al., 1978). To date, the WHO recognises 9 specific recurrent genetic entities in ALL (table 1.1) (Terwilliger and Abdul-Hay, 2017, Arber et al., 2016), although others have since been described.

B-lymphoblastic leukaemia/lymphoma
B-lymphoblastic leukaemia/lymphoma, NOS
B-lymphoblastic leukaemia/lymphoma with recurrent genetic abnormalities
B-lymphoblastic leukaemia/lymphoma with t(9;22)(q34.1;q11.2);*BCR-ABL1*
B-lymphoblastic leukaemia/lymphoma with t(v;11q23.3);*KMT2A* rearranged
B-lymphoblastic leukaemia/lymphoma with t(12;21)(p13.2;q22.1);*ETV6-RUNX1*
B-lymphoblastic leukaemia/lymphoma with hyperdiploidy
B-lymphoblastic leukaemia/lymphoma with hypodiploidy
B-lymphoblastic leukaemia/lymphoma with t(5;14)(q31.1;q32.3);*IL3-IGH*
B-lymphoblastic leukaemia/lymphoma with t(1;19)(q23;p13.3);*TCF3-PBX1*
Provisional entity: B-lymphoblastic leukaemia/lymphoma, BCR-ABL1–like
Provisional entity: B-lymphoblastic leukaemia/lymphoma with iAMP21
T-lymphoblastic leukaemia/lymphoma
Provisional entity: Early T-cell precursor lymphoblastic leukaemia
Provisional entity: Natural killer (NK) cell lymphoblastic leukaemia/lymphoma

Table 1.1. 2016 revision WHO classification of lymphoblastic leukaemia/lymphoma. Adapted from Arber et al (Arber et al., 2016).

Despite the improvements in outcome achieved through treatment intensification and supportive care, cytogenetics continues to be one of the most important prognostic factors and plays a key role in treatment decisions (Moorman et al., 2007, Pullarkat et al., 2008). Cytogenetics is also closely related to age, and the frequency of poor risk abnormalities increases with age at diagnosis, with the notable exception of infant ALL which is characterised by a 70-80% frequency of the high risk t(4;11)(q21;q23)/*KMT2A-AFF1* fusion (Hilden et al., 2006).

1.6.4 Response to treatment – early marrow response and minimal residual disease assessment

The single most important recent development in the risk stratification of ALL patients has been the incorporation of response to treatment assessments in therapeutic decisions.

Early morphological assessment after 1-2 weeks of induction chemotherapy has been used to identify rapid early responders, who have a better prognosis than slow early responders (Gaynon et al., 1997). However, early morphological assessment, before haematological recovery has occurred, is challenging and error prone. Bone marrow smears are often difficult to interpret during early treatment and regenerating haematogones and lymphocytes can be mistaken for ALL blast cells and vice versa (Gupta et al., 2018). Even in optimal samples, morphological assessment is only sensitive to a blast percentage ≥5%. Although more reproducible, morphological assessments at later time points and following haematological recovery (e.g. following 4 weeks of induction treatment), are less valuable as the vast majority of patients are in morphological remission, and only patients with a dismal outcome (‘induction failure’) will be positively identified (Schrappe et al., 2012).

The development of techniques for detecting minimal residual disease (MRD) has permitted the detection of much smaller numbers of leukaemic cells, generating one of the most valuable biomarkers for risk stratification and treatment allocation (Cavé et al., 1998, Brüggemann et al., 2006). The most extensively validated and frequently used technique involves tracking leukaemia specific immunoglobulin and/or T-cell receptor rearrangements (Ig/TCR). Like all B-cells, BCP-ALL blasts undergo rearrangement of the V, D and J segments of their immunoglobulin gene through

random deletions and insertions of nucleotides at the junctional sites. T-ALL blasts undergo a similar process with their TCR. This creates a unique molecular signature which is inherited by all daughter cells produced during clonal expansion, and will therefore be specific to all leukaemic cells within a particular patient (Campana, 2010). At diagnosis, the junctional regions of the immunoglobulin or TCR gene in a bone marrow sample are sequenced to identify the patient leukaemia-specific MRD marker(s). Allele specific oligonucleotides (ASOs) can then be designed and used in a real time quantitative polymerase chain reaction (RQ-PCR) to measure residual leukaemia in follow up samples (Campana, 2010, van der Velden et al., 2007). This technique can typically detect residual ALL blasts down to a concentration of 1 in 10,000-100,000.

In patients presenting with leukaemia-specific gene fusions (such as *ETV6-RUNX1*, *BCR-ABL1*, *TCF3-PBX1*), the aberrant mRNA transcript can be used as an MRD marker (van Dongen et al., 1999). As the breakpoints cluster within well-defined genomic regions, universal primers can be used in an RQ-PCR assay without the need for time-consuming sequencing on a diagnostic sample followed by design and validation of patient specific ASOs. This MRD technique is well validated and extensively used for *BCR-ABL1+* ALL. Interestingly, although the correlation with Ig/TCR-based MRD is usually good, significant discrepancies are seen owing to the observation that the *BCR-ABL1* translocation may also be present in non-lymphoid cells (Hovorkova et al., 2017). Thus, MRD results are always viewed in the context and sensitivity of the method used.

More recently, multiparametric flow cytometry (MFC)-based MRD techniques have been developed, and make use of the observation that leukaemic cells often express aberrant cell surface markers, creating a leukaemia-associated immunophenotype, (LAIP) (Brüggemann and Kotrova, 2017). In post-remission samples, cells bearing the LAIP can therefore be detected using MFC and quantified to generate an MRD measurement.

As with other biological and clinical features, MRD kinetics are influenced by other prognostic factors. Genetic abnormalities associated with a favourable prognosis are also associated with a more rapid clearance of MRD and vice versa for high risk genetic lesions (O'Connor et al., 2018). Moreover, treatment intensity is also likely to influence the persistence of MRD, which is encountered more frequently in adult than childhood ALL (Gökbuget et al., 2018). In adults treated intensively, typically using paediatric

inspired protocols, a number of trials in *BCR-ABL1* negative patients have demonstrated that the attainment of MRD negativity is the most important prognostic factor and abrogates any survival benefit from allogeneic stem cell transplantation in first remission (Dhédin et al., 2015, Ribera et al., 2014, Bassan et al., 2020). Data, however, are currently lacking in older adults who receive the lowest intensity therapy. Interestingly, immunotherapies (such as blinatumomab) and targeted treatments (such as tyrosine kinase inhibitors for *BCR-ABL1*+ ALL) are both well tolerated and are able to produce high rates of MRD negativity (Gökbuget et al., 2018, Rousselot et al., 2016, Topp et al., 2012), even when combined with little or no chemotherapy so are particularly attractive treatment options in older and frailer patients. Indeed, one study combined the third generation TKI ponatinib with steroids alone in elderly and unfit patients with Ph+ ALL, and reported a complete molecular response in 45% of evaluable patients (Martinelli et al., 2017). Additionally, the recent D-ALBA phase 2 study used a potent chemotherapy-free combination of the bispecific T-cell engaging antibody blinatumomab and the TKI dasatinib in Philadelphia positive patients and reported a molecular response in up to 81% of patients, with an impressive overall survival of 95% at a median follow up of 18 months (Foà et al., 2020).

1.7 ALL in older adults

Approximately 130 adults over the age of 60 years are diagnosed with ALL every year in the UK (Cancer Research UK, 2016). Prior to the UKALL60+ clinical trial, there were no dedicated UK studies focused on improving outcomes through the acquisition of biological, clinical and outcome data for these patients. The UKALL XII clinical trial included 100 patients over the age of 55 years (range 55-65 years) deemed fit for intensive chemotherapy-based treatment (Sive et al., 2012). A subgroup analysis focused on the biological and clinical features of these older adults identified a significantly lower frequency of lymph node enlargement, hepatomegaly and splenomegaly compared with patients aged under 55 years old. The rate of complete remission after induction chemotherapy was 73% in older patients compared to 93% in the younger group. Treatment toxicities were much higher in older patients and the rate of death in induction was 18% compared to 4% in younger adults. Furthermore, 46% of older patients required drug dose reductions, omissions or delays in induction,

suggesting these treatments are too intensive for a significant proportion of older adults, even those deemed biologically fit. To date, the GMALL group have performed one of the largest trials of ALL therapy in older adults (Goekbuget et al., 2012). The study included 268 patients with a median age of 67 years (range 55-85). Patients were treated with a moderately intensive regime of reduced dose multi-agent chemotherapy and CNS prophylaxis. Overall, 76% of patients achieved complete remission following induction and 14% had an early death. The study confirmed that early death was associated with the ECOG status and Charlson comorbidity score before onset of ALL, highlighting the importance of a comprehensive functional assessment to determine fitness for and timing of chemotherapy.

Although the UKALL XII comparison between patients aged over and under 55 years at diagnosis did not identify significant differences in presenting white cell counts or B versus T cell disease, the population studied was aged 55-65 years so was not representative of all older ALL patients (Sive et al., 2012).

A French study also identified biological differences between ALL in older adults and other age groups (Thomas et al., 2001). The male: female ratio declined from 1.75 in younger adults to 0.97 in older adults. A moderately lower presenting WCC was also observed which in combination with lower rates of lymphadenopathy and organomegaly may suggest age-related differences in disease behaviour. B cell disease also accounts for a significantly larger proportion of ALL diagnoses in older adults (75-89%) compared with younger patients (59-66%) (Gökbuget, 2013).

1.7.1 Treatment of ALL in older patients

The clinical management of ALL in older adults requires a careful balance between optimising response whilst minimising the risk of treatment-related morbidity and mortality. The principles of treatment are similar to those adopted in younger patients, namely remission induction, followed by consolidation and maintenance, with concomitant CNS directed therapy. However, unlike younger patients, most older adults (particularly aged over 70 years) will not be suitable for allogeneic stem cell transplantation, and certain chemotherapeutic agents need to be used with greater caution due to higher toxicities, specifically anthracyclines and asparaginase. Fortunately, well tolerated targeted treatment such as tyrosine kinase inhibitors (TKIs)

are particularly beneficial in older patients presenting with *BCR-ABL1* positive disease. Indeed, such approaches have permitted significant reductions in the cytotoxic agents used, without inferior outcomes (Chalandon et al., 2015). However, similar approaches are not yet routinely available for *BCR-ABL1* negative patients, and there is little scope to intensify treatments based on the presence of high-risk disease or MRD positivity.

1.8 Cytogenetics in ALL

The hallmark of ALL is the acquisition of chromosomal abnormalities in the leukaemic blasts (Terwilliger and Abdul-Hay, 2017). To date, a spectrum of karyotypic abnormalities, some bearing a clear effect on prognosis, have been identified (Bloomfield et al., 1986, Harrison et al., 2010).

Good risk chromosomal abnormalities, such as high hyperdiploidy, deleted 9p, and t(12;21)/*ETV6-RUNX1*, have been associated with a 5-year overall survival around 60% in adults (Moorman et al., 2007). However, in comparison with childhood ALL, adult ALL has a different cytogenetic profile with increasing frequency of the poor risk chromosomal abnormalities t(9;22)(q34;q11)/*BCR-ABL1*, t(4;11)(q21;q23)/*MLL-AF4*, low hypodiploidy/near triploidy, and complex karyotype (Moorman et al., 2007). The frequency of these poor risk features increases with age (figure 1.4). Combining this with a poorer tolerance to intensive therapy, particularly in older adults, significantly impacts on survival.

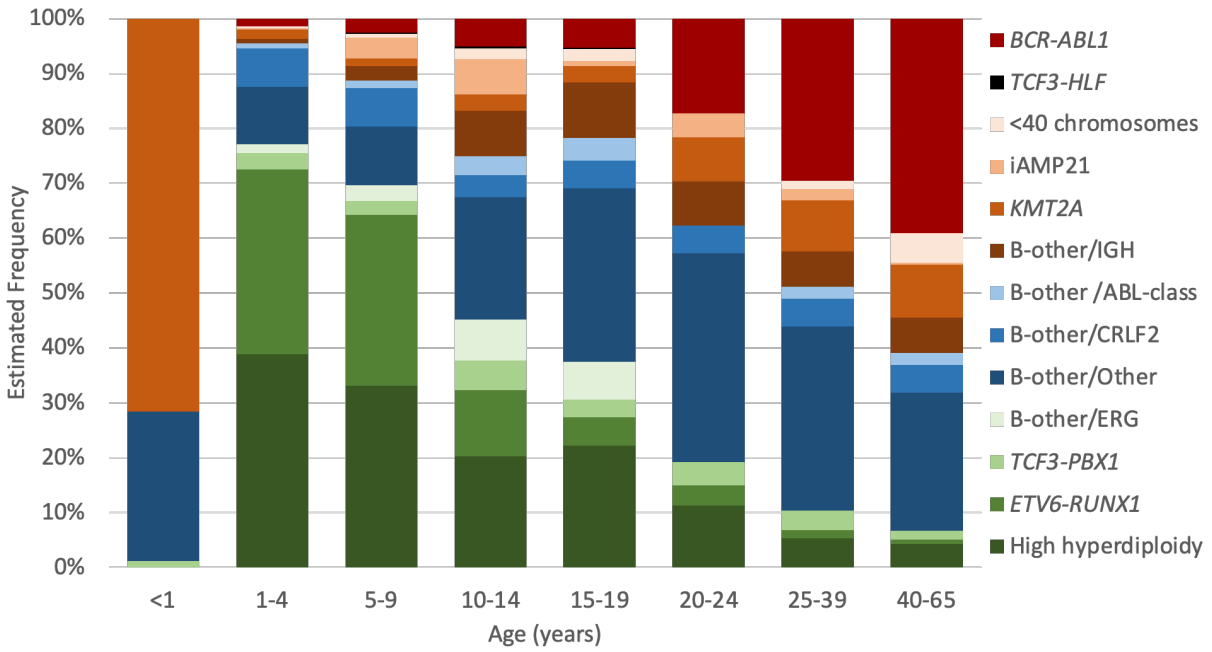


Figure 1.4. BCP-ALL cytogenetic subtypes in children, adolescents and adults aged 25-59. Red/orange regions indicate poor risk subgroups, green regions indicate good risk subgroups and blue regions indicate intermediate risk subgroups. Proportion of poor risk genetic abnormalities increases after infancy with corresponding decrease in good risk subgroups. Adapted from Moorman (Moorman, 2012).
iAMP21: intrachromosomal amplification of chromosome 21; *KMT2A*: *KMT2A* fusion; B-other/*IGH*: *IGH@* rearrangement; B-other/ABL-class: ABL-class fusion including *ABL1*, *ABL2*, *PDGFRB* or *CSF1R*; B-other/*CRLF2*: *CRLF2* rearrangement; B-other/ERG: focal ERG deletion.

Recurrent chromosomal abnormalities include large scale gains or losses to the normal diploid complement or translocations resulting in characteristic chimeric fusion genes.

1.8.1 Large scale ploidy shifts

1.8.1.1 High hyperdiploidy

High hyperdiploidy (HeH) is the most prevalent abnormality in childhood ALL and is associated with duplication of entire chromosomes (Moorman, 2016). By definition the total number of chromosomes within ALL blasts is between 51-67 and the pattern of chromosomal gains is non-random with eight chromosomes accounting for most gains (X, 4, 6, 10, 14, 17, 18, 21) (Chilton et al., 2013). Trisomies involving chromosomes 5 and 20 are also seen and have been associated with poorer risk disease. HeH is much rarer in adult ALL (figure 1.4). The UKALLXII/ECOG2993 clinical trial enrolled 1,522 adults aged 15-65 and HeH was identified as the primary cytogenetic abnormality in

around 7% of patients with successful cytogenetic analyses (Moorman et al., 2007). ALL with HeH is usually associated with a favourable prognosis (Moorman, 2016).

1.8.1.2 Low hypodiploidy

Low hypodiploid/near triploid ALL (HoTr) is rare in childhood (<2% of cases) and becomes more common with age (figure 1.4). Like HeH, the primary abnormality is characterised by gross chromosomal aneuploidy, in this case a hypodiploid clone consisting of 30-39 chromosomes. A duplication of the remaining chromosomes can also occur, resulting in a clone of cells with a near triploid karyotype (60-78 chromosomes), which can make the abnormality difficult to differentiate from HeH if only near triploid metaphases are seen on karyotypic analysis (Charrin et al., 2004). Chromosomal losses are also non-random with chromosomes 3, 7, 15, 16, 17 being preferentially deleted. Both copies of chromosome 21 appear to be consistently retained (Harrison et al., 2004). HoTr ALL is associated with a very high frequency of *TP53* mutations (over 90%), which are usually germline in paediatric cases and somatic in adults (Holmfeldt et al., 2013, Mühlbacher et al., 2014). Indeed, patients with the Li Fraumeni syndrome have a particularly high risk of developing this ALL subtype (Powell et al., 2013).

1.8.1.3 Near haploidy

Near haploid ALL (NH) is a very rare high-risk subgroup, characterised by large scale chromosomal loss in the leukaemic cells. The NH clone contains 23-29 chromosomes and like HoTr, can undergo chromosomal endoreduplication without subsequent cytokinesis, creating a "doubled-up", low hyperdiploid clone with 46 to 58 chromosomes (Ma et al., 1998, Stark et al., 2001). This, in particular, can present a diagnostic challenge in differentiating this subgroup from HeH. However, the doubled clone will present with numerous tetrasomies, and will usually lack the characteristic trisomies seen in HeH ALL. As with other genetic subtypes, chromosome 21 is always retained in a NH clone and becomes tetrasomic in a "doubled-up" clone (Harrison et al., 2004). NH-ALL occurs almost exclusively in childhood, where it accounts for 0.5% of new diagnoses and is almost unreported in adult ALL (Harrison et al., 2001, Harrison et al., 2004).

1.8.2 Gene fusions

Chromosomal rearrangements resulting in the fusion of two normally distant genes are a common feature of leukaemogenesis. If expressed, the resulting fusion gene produces a chimeric protein with altered functionality and transformative capabilities.

1.8.2.1 *ETV6-RUNX1* fusion

The t(12;21)/*ETV6-RUNX1* translocation accounts for over a quarter of childhood BCP-ALL cases but is very rare in adults (Moorman, 2016). *ETV6* contains protein dimerisation domains which are fused to the DNA binding and activating regions of *RUNX1* (Torrano et al., 2011, Golub et al., 1995). *RUNX1* forms part of the core binding factor complex which is essential for the regulation of normal haematopoiesis. Interestingly, the fusion gene usually arises antenatally and additional postnatal abnormalities (including gene deletions) are required for the development of overt disease (section 1.4) (Ford et al., 1998). In rare adult ALL patients with *ETV6-RUNX1* fusion, the disease has been associated with a low white cell count at diagnosis and comparatively favourable prognosis (Burmeister et al., 2010).

1.8.2.2 *KMT2A* fusions

The *KMT2A* (formerly *MLL*) gene on chromosome 11q23 is implicated in both myeloid and lymphoid leukaemias (Meyer et al., 2018). Translocations with a variety of partner genes destroy the normal histone methyltransferase function of *KMT2A* resulting in transcription activation (Slany, 2009). *KMT2A* translocations occur in approximately 70-80% of infant ALL and prognosis is universally poor (Brown, 2013). These abnormalities are rare in childhood ALL (1-2%) but become more common again in adults. In UKALLXII/ECOG2993, 9% of Philadelphia-negative patients (6.6% of all patients with a cytogenetic result) had a *KMT2A* translocation (Moorman et al., 2007). The majority of these (78%) had a t(4;11)(q21;q23)/*KMT2A-AFF1* rearrangement. Within this study, patients with a *KMT2A* translocation had a higher presenting white cell count but only those with t(4;11)(q21;q23)/*KMT2A-AFF1* had an inferior 5-year overall survival.

1.8.2.3 *BCR-ABL1* fusion (Philadelphia positive ALL)

Philadelphia positive (Ph+) ALL is probably the most relevant genetic subtype of adult ALL due to its high prevalence and individualised treatment. The well described t(9;22)(q34;q11.2) creates a chimeric fusion gene juxtaposing *BCR* on chromosome 22q11.2 to *ABL1* on chromosome 9q34. The resultant *BCR-ABL1* fusion gene exhibits constitutively active tyrosine kinase activity (Salesse and Verfaillie, 2002), which alone is sufficient to induce chronic myeloid leukaemia (CML). However, the development of Ph+ ALL also requires the activation of Src kinases by the bcr-abl oncoprotein (Hu et al., 2004) and has been associated with the acquisition of additional gene deletions compared with CML (Mullighan and Downing, 2009). Ph+ ALL accounts for less than 5% of childhood ALL but its prevalence rises to approximately 25% of ALL in adults diagnosed in the fourth decade of life. Some studies have shown a further rise with advancing age (Byun et al., 2017). However, this was not seen in a large cohort of Ph+ patients from multiple GMALL (German Multicenter ALL) trials (Burmeister et al., 2008) or in a population study of 349 adult ALL patients treated over 19 years in a single institution in the North East of England (Moorman et al., 2010). Historically, Ph+ ALL has consistently been associated with high risk disease and a poor prognosis. However, the advent of imatinib and the subsequent second and third generation tyrosine kinase inhibitors (TKIs), has significantly improved the outcome of Ph+ ALL. Large prospective trials have confirmed the safety and efficacy of incorporating TKIs with frontline chemotherapy with improvements in morphologic and molecular remission rates and overall survival (Wassmann et al., 2006, Thomas et al., 2004a). In the era of routine use of TKIs in the treatment of Ph+ ALL, remission rates and overall survival at 4 years have significantly improved from 82% to 92% and 22% to 38% respectively with the addition of imatinib in the UKALLXII/ECOG2993 trial (Fielding et al., 2014). More recently, incorporation of the third generation TKI ponatinib into a multi-agent chemotherapy backbone resulted in a 3 year overall survival of 76% with a high rate of complete molecular response in a phase 2 study (Jabbour et al., 2018). Additionally, a phase 3 randomised controlled trial confirmed that the incorporation of frontline TKI in Ph+ patients has permitted a reduction in the intensity of the chemotherapy backbone with no reduction in event-free or overall survival (Chalandon et al., 2015). The addition of TKIs to induction chemotherapy has also permitted more patients to undergo allogeneic stem cell transplantation due to the attainment of complete remission, which is often considered a pre-requisite for consolidation with an allogeneic transplant (Wassmann et al., 2006). Separately, one of the groups to benefit

most from this well-tolerated targeted therapy has been older adults with Ph+ ALL. A number of studies have demonstrated that TKIs administered either with steroids or low dose chemotherapy produce superior remission rates with much lower toxicity compared with standard chemotherapy schedules (Ottmann et al., 2007, Vignetti et al., 2007, Rousselot et al., 2016).

1.8.2.4 *TCF3-PBX1* fusion

A specific translocation between chromosomes 1 and 19 – t(1;19)(q23;p13) – creates a *TCF3-PBX1* fusion gene. The *TCF3* gene encodes two transcription factors (E12 and E47) that play critical roles in B-cell maturation. *PBX1* is a homeobox gene and is not usually expressed in B and T cells. *TCF3-PBX1* fusion results in unphysiological expression of this homeobox gene in lymphoid cells, leading to malignant transformation (Burmeister et al., 2010). The abnormality is seen in ALL patients of all ages, at a frequency of approximately 3-6%, and is usually associated with an intermediate prognosis, although some studies have also reported an increased risk of CNS relapse and a very poor outcome in patients who relapse (Moorman, 2016).

The majority of these abnormalities are detected using conventional karyotyping allowing enumeration of individual chromosomes and detection of structural abnormalities with a sensitivity of around 5 megabases (Mb) (Martin and Warburton, 2015). However, certain chromosomal translocation such as t(12;21)/*ETV6-RUNX1* are cytogenetically cryptic and are best detected through fluorescence *in situ* hybridisation (FISH) or RT-PCR (reverse transcription polymerase chain reaction).

1.8.3 B-other ALL

Around a quarter of adults and children with BCP-ALL lack a known primary recurrent chromosomal abnormality and are termed ‘B-other ALL’ (Moorman, 2016). Many studies have used gene expression profiling to define a group of these patients with a gene expression signature similar to *BCR-ABL1*+ ALL, now termed *BCR-ABL*-like (Den Boer et al., 2009) or Philadelphia-like (Ph-like) (Mullighan et al., 2009b) ALL (depending on studies and specific gene set). Around half of B-other ALL patients have been found to harbour a Ph-like signature, which is driven by ABL-class fusions or

JAK-STAT pathway activating lesions in over 80% of cases (Roberts et al., 2017). A range of kinase activating abnormalities have been identified, specifically ABL-class fusions caused by gene rearrangements affecting *ABL1*, *ABL2*, *PDGFRB* and *CSF1R*, or JAK-STAT pathway activating abnormalities caused by *JAK2*, *CRLF2* or rarer *EPOR* rearrangements (Roberts et al., 2014).

However, older adults have been underrepresented in many of these cohorts, and the frequency of these kinase activating lesions in older B-other ALL patients remains unclear. Interestingly, a study performed gene expression profiling across a cohort of BCP-ALL patients with older adults reasonably well represented (67 patients aged 56 to 84 years of age) (Herold et al., 2014). The authors identified a peak in the *BCR-ABL*-like signature in adolescents and young adults and a marked reduction in incidence in the older age group. This was matched with an ongoing rise in the incidence of Philadelphia-positive ALL, which peaked over 50% in the older cohort of patients.

As a group, the disease-defining gene rearrangements in B-other ALL are cytogenetically cryptic and require the use of fluorescence *in situ* hybridisation (FISH) or genomic techniques to be detected. Most abnormalities involve a specific tyrosine kinase or cytokine receptor, constitutively activated through fusion to a range of different partner genes (figure 1.5).

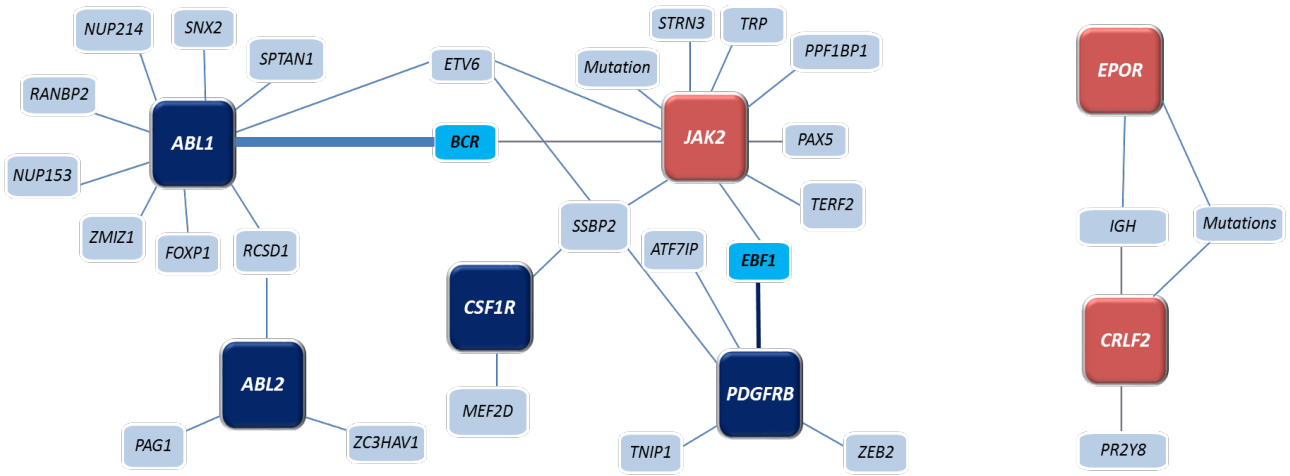


Figure 1.5. Ph-like/*BCR-ABL*-like gene rearrangements with reported partner genes. *ABL1*, *ABL2*, *CSF1R* and *PDGFRB* rearrangements represent ABL-class fusions. *JAK2*, *CRLF2* and *EPOR* rearrangements activate JAK-STAT signalling. Figure courtesy of Claire Schwab and Anthony Moorman.

1.8.3.1 ABL-class fusions

Rearrangements of *ABL1, ABL2*, *PDGFRB* or *CSF1R* result from balanced translocations, interstitial deletions or inversions and the resultant chimeric genes fuse the carboxyl terminal of the tyrosine kinase protein to the amino terminal portion of the partner gene resulting in a constitutively active tyrosine kinase (Roberts et al., 2014). Fusions with a number of different partner genes have been identified, with no evidence that specific partners impact on outcome or disease phenotype. Patients with ABL-class fusions, of which *PDGFRB* rearrangements are the commonest, often respond suboptimally to standard induction chemotherapy and are enriched in the <5% of patients with refractory disease post induction (O'Connor et al., 2017, Schwab et al., 2016). Importantly, as a result of their tyrosine kinase activity, ABL-class fusions are sensitive to imatinib and other TKIs, and early use of TKIs has been associated with an improved outcome (Tanasi et al., 2019, Moorman et al., 2020).

1.8.3.2 *CRLF2* rearrangements

CRLF2 pairs with the IL7R-alpha subunit to form a thymic stromal lymphopoietin receptor (TSLPR), which is implicated in early B-cell development. Between 15-50% of Ph/*BCR-ABL*-like patients (depending on the gene set used to define the gene expression signature) have a *CRLF2* gene rearrangement (Russell et al., 2017). The rearrangement either occurs as part of an *IGH-CRLF2* translocation or an interstitial deletion on the pseudoautosomal region (PAR) of Xp22/Yp11, juxtaposing *CRLF2* to the first non-coding exon of *P2RY8* and resulting in the expression of *P2RY8-CRLF2* fusion transcripts. These rearrangements lead to *CRLF2* overexpression. In addition, mutations activating of *JAK2* (or rarely *JAK1*) are found in approximately 50% of *CRLF2* rearranged cases (Harvey et al., 2010). The combination of *CRLF2* rearrangements and *JAK1/2* mutations result in activation of JAK-STAT signalling. *CRLF2* rearrangements are seen in greater than 50% of paediatric ALL patients with Down's syndrome and are also more common in Hispanic patients (Mullighan et al., 2009a, Harvey et al., 2010). *P2RY8* is the more common *CRLF2* rearrangement partner in children with *CRLF2* deregulated ALL, whereas *IGH-CRLF2* is seen more commonly in adults (Russell et al., 2017). *CRLF2* abnormalities have been associated with high rates of MRD persistence and poor overall outcome (Herold et al., 2017, Yoda et al., 2010). Unlike ABL-class fusions, no effective targeted treatment are

available, although the *JAK1/JAK2* inhibitor ruxolitinib has provided a plausible therapeutic option (Tasian et al., 2012) and clinical trials are underway.

1.8.3.3 *JAK2* fusions

The amino terminus of a partner gene is fused to the carboxyl terminal portion of *JAK2* resulting in activation of STAT5 and constitutive activation of JAK-STAT signalling (Schinnerl et al., 2015). These translocations or interstitial deletions are seen in approximately 5% of paediatric ALL patients and more commonly in young adults (Roberts et al., 2014). As with other JAK-STAT pathway abnormalities, *JAK2* rearranged cases have been associated with high risk disease and pre-clinical studies with ruxolitinib have produced mixed findings, both with evidence for mutations in alternative survival pathways promoting drug resistance and ruxolitinib-induced accumulation of phosphorylated JAK2 producing a sharp increase in JAK-STAT signalling on drug withdrawal (Boer et al., 2017).

1.8.3.4 *EPOR* (erythropoietin receptor) rearrangements

These are rare events, occurring in approximately 1% of BCP-ALL cases. They are usually due to truncating mutations in the *EPOR* gene arising through rearrangements with a variety of partners including the *IGH@* locus (Iacobucci et al., 2016). This results in deregulated *EPOR* expression of a truncated protein, with loss of the distal tyrosine residues, which are required for negative regulation of receptor signalling. As such, the truncated protein shows hypersensitivity to erythropoietin and increased JAK-STAT activation. The most common gene rearrangement involves insertion of a truncated *EPOR* gene, with loss of the terminal portion of exon 8, to a region distal to the *IGH@* locus. This rearrangement is cryptic by both cytogenetics and FISH, and can only be detected through sequencing or a dedicated quantitative real-time PCR assay (Iacobucci et al., 2016).

1.8.3.5 *ZNF384* translocations

Rearrangements of *ZNF384* produce a distinct subtype of ALL with a distinct immunophenotype including weak CD10 expression and aberrant expression of myeloid antigens CD13 and CD33 (Hirabayashi et al., 2017). *ZNF384* encodes a

transcription factor known to regulate promoters of the extracellular matrix genes and is rearranged with a number of different partner genes including the TET family genes (*EWSR1*, *TAF15* and *TCF3*) and *EP300* among others (Schwab and Harrison, 2018, Hirabayashi et al., 2017). These translocations have been identified in around 4% of childhood ALL patients, and these patients cluster together and separately from other genetic subtypes on gene expression analysis. The incidence of *ZNF384* translocations in adults has not yet been clearly defined but also appears to be <10% (Moorman et al., 2019, Liu et al., 2016).

1.8.3.6 *MEF2D* rearrangements

MEF2D belongs to a family of 4 myocyte enhancer factor (MEF) transcription factors and plays an important role in neuronal differentiation. *MEF2D* rearrangements with a variety of partner genes enhance its transcriptional activity resulting in activation of *HDAC9* expression and subsequent lymphoid transformation (Gu et al., 2016a). These events are reported in ~2% of childhood BCP-ALL and have been associated with an older age at diagnosis, higher white count and inferior survival (Ohki et al., 2019). The most frequent rearrangement involves a *BCL9-MEF2D* fusion. Both genes are located on 1q21.2-22 and a small cytogenetically cryptic interstitial insertion is required to produce this fusion gene. However, the mechanisms of rearrangement can be significantly more complex including insertion of the fusion into a different chromosome and can involve complex karyotypic abnormalities, making their detection more challenging (Gu et al., 2016a). *MEF2D* rearrangements result in activation of *HDAC9*, which has been shown to induce cell proliferation and has been implicated in the pathogenesis of B-cell lymphomas (Gil et al., 2016).

1.8.3.7 *IGH@* translocations

Translocations involving the immunoglobulin heavy chain (*IGH@*) locus are common primary events in mature B-cell malignancies. Typically, following a reciprocal translocation, an oncogene becomes overexpressed through close proximity to the *IGH* enhancer region. Such events are the hallmarks of Burkitt's lymphoma (t(8;14)(q24;q32)/*MYC-IGH* translocation) and follicular lymphoma (t(14;18)(q32;q21)/*IGH-BCL2* translocation) among others (Küppers, 2005). Although well described, *IGH@* translocations are rarer in ALL and are associated with different

partner genes to those described in mature B-cell malignancies (Russell et al., 2014). The consequences are similar, resulting in overexpression of a gene with anti-apoptotic or transformative properties, and contributing to leukaemogenesis. Such cases have been identified in 5% of ALL patients with a peak incidence in adolescents and young adults (Russell et al., 2014). The most frequent partner gene is *CRLF2*, which is implicated in 22% of *IGH@* translocated cases, followed by members of the CCAAT/enhancer binding protein (*CEBP*) family and *ID4* in 11% and 7% respectively (Russell et al., 2014).

More recently, *DUX4* has been identified as another *IGH@* partner gene and *IGH-DUX4* rearrangements are recurrent events, seen in ~7% of childhood BCP-ALL. *DUX4* is a small intronless gene with transcription factor activity, located within the D4Z4 repeat regions on the subtelomeric regions of 4q35 and 10q26. The abnormality is characterised by insertion of variable numbers of *DUX4* repeats within the vicinity of the *IGH@* enhancer, resulting in transcriptional activation (Tian et al., 2019). *IGH-DUX4* rearrangements are cytogenetically cryptic and challenging to detect by conventional techniques due to the small and highly repetitive nature of the inserted region, and transcriptome analysis has proven the most reliable method to date (Schwab and Harrison, 2018). However, a clear association has been identified between *IGH-DUX4* and intragenic deletions of the ETS transcription factor ERG (*ERG*) gene at 21q22.2, which are thought be exclusive to this subgroup, and occur in 50-60% of cases (Schwab and Harrison, 2018, Lilljebjörn et al., 2016). As a possible surrogate for *IGH-DUX4*, *ERG* deletions are easier to detect by conventional methods and have been associated with excellent prognosis in paediatric cohorts (Clappier et al., 2014).

1.9 T-cell ALL

The incidence of T-ALL peaks in adolescents and young adults, and is rare in older age groups (Guru Murthy et al., 2019). Recurrent chromosomal rearrangements are well established but do not impact on prognosis or risk stratification (table 1.2). The translocations are more structurally diverse than those report in B-cell disease and

usually serve to upregulate a relevant transcription factor oncogene, although rarer chimeric gene fusions are also described (Girardi et al., 2017).

Oncogene	Partner gene	Consequence
TAL1	*STIL, TCRA, TCRD*	*TAL1* upregulation
TLX1	*TCRA, TCRB, TCRD*	*TLX1* upregulation
TLX3	*TCRD, BCL11B*	*TLX3* overexpression
HOXA9/HOXA10	*TCRB*	*HOXA* gene overexpression
PICALM	*MLLT10*	*PICALM-MLLT10* fusion
SET	*NUP214*	*SET-NUP214* fusion
NKX2-1	*TCRA*	*NKX2-1* overexpression
NKX2-2	*TCRD*	*NKX2-2* overexpression
LMO1	*TCRA, TCRD*	*LMO1* overexpression
MEF2C	-	Deletion causing *MEF2C* overexpression
LYL1	*TCRB*	*LYL1* overexpression
SPI1	*BCL11B*	*SPI1* overexpression

Table 1.2. Recurrent gene rearrangements reported in T-ALL. Reported rearrangements predominantly result in overexpression of an oncogene rather than formation of chimeric fusion genes. Adapted from Girardi et al, (Girardi et al., 2017)

1.10 Copy number abnormalities in ALL

Cancer genomes harbour a range of copy number abnormalities (CNAs), characterised by losses or gains of regions of the genome. These may be random and part of the generalised genomic instability associated with malignancy (passenger lesions) or play a specific part in the cancer process, conferring transformative capabilities and survival advantage (driver lesions). Within a cohort of comparable tumour samples, driver lesions will occur at a higher frequency than passenger abnormalities, and will often be targeted to a focal gene or locus, which confers the oncogenic properties (Mermel et al., 2011).

Conversely, copy number changes also form a very common part of normal genomic variation between individuals (Mills et al., 2011). Indeed, analysis of the HapMap collection of 270 healthy individuals has identified that up to 12% of the human genome is subject to germline copy number variation (CNVs) when compared to a reference

genome, covering more genomic loci than single nucleotide polymorphisms (SNPs) (Redon et al., 2006).

Initiating primary chromosomal abnormalities alone are usually insufficient to cause ALL (Mullighan et al., 2007). A spectrum of secondary co-operating abnormalities, often characterised by deletions or gains of smaller genomic regions, contribute to leukaemogenesis and also impact on prognosis. Recurrent deletions affecting specific genes or loci are twice as common as gains (Strefford et al., 2007), and particularly affect genes encoding B-cell development and differentiation, cell cycle regulation (tumour suppressor genes) and haematopoietic control (table 1.3).

Gene/locus	Locus	ALL Subtype	Function
IKZF2	2q34	BCP*	B cell differentiation
CD200/BTLA	3q13.2	BCP	Lymphoid signalling
LEF1	4q25	T-cell	Haematopoietic control
EBF1	5q33.3	BCP	B cell differentiation
IKZF1	7p12.2	BCP	Lymphoid development
CDKN2A/B	9p21.3	BCP and T-cell	Cell cycle regulation
PAX5	9p13.2	BCP	B cell differentiation
PTEN	10q23.31	T-cell	Cell cycle regulation
RAG1/2	11p12	BCP	V(D)J-recombination
ETV6	12p13.2	BCP and T-cell	Haematopoietic control
BTG1	12q21.33	BCP	Apoptosis regulation
RB1	13q14.2	BCP and T-cell	Cell cycle regulation
IKZF3	17q12	BCP*	B cell differentiation
NF1	17q12.2	T-cell	Cell cycle regulation
TCF3	19p13.3	BCP	B cell differentiation
ERG	21q22	BCP	Haematopoietic control
PAR1‡	Xp22.3/Yp11.2	BCP	JAK/STAT activation

Table 1.3.: Recurrently deleted genes/loci in ALL identified from SNP array (Mullighan et al., 2009b, Ribera et al., 2017, Mullighan et al., 2007, Paulsson et al., 2008) and MLPA (Schwab et al., 2013) studies.
******IKZF2* **and** *IKZF3* **deletions are specifically associated with low hypodiploid and near haploid ALL respectively.**
‡Specific deletion on pseudoautosomal region 1 on Xp22.3/Yp11.2 (PAR1) results in *P2RY8-CRLF2* **fusion and** *CRLF2* **overexpression.**

In childhood ALL, specific combinations of gene deletions have a clear impact on prognosis. Paediatric ALL patients presenting with an *IKZF1* deletion co-occurring with

deletions in one or more of *CDKN2A*, *CDKN2B*, *PAX5* or PAR1 in the absence of *ERG* deletion (*IKZF1*^{plus} profile) have been found to have an event free survival of 53% compared with patients with isolated *IKZF1* deletion or intact *IKZF1*, who had an EFS of 79% and 87% respectively (Stanulla et al., 2018). In comparison, paediatric patients harbouring focal *ERG* deletions have an excellent prognosis, with an 8 year overall survival >95% (Clappier et al., 2014). These patients lack conventional chromosomal abnormalities and form part of the B-other subgroup. Interestingly over a third of affected patients had an *IKZF1* deletion, which did not impact on their prognosis, highlighting the importance of considering the interaction of CNAs rather than individual lesions in isolation. Furthermore, combining baseline cytogenetic risk groups with copy number profiles can provide a powerful prognostic biomarker in paediatric ALL populations. In one study, subsequently validated with a large international external cohort (Moorman et al., 2014, Hamadeh et al., 2019), patients could be divided into genetic good risk and genetic poor risk categories based on the combination of primary chromosomal abnormalities and copy number profiles (UKALL-CNA). This proved particularly relevant for patients with intermediate risk cytogenetics (*TCF3-PBX1* or B-other ALL), whereby a good risk UKALL-CNA profile (table 1.4) was associated with an EFS of 88% (genetic good risk), whereas those with poor risk UKALL-CNA profile had an EFS of 69% (genetic poor risk).

Classification	Risk category	Features
Cytogenetics	Good risk	HeH *ETV6-RUNX1*
	Intermediate risk	*TCF3-PBX1* B-other
	Poor risk	*BCR-ABL1* *TCF3-HLF* Near haploidy HoTr
UKALL-CNA	Good risk	No deletions Isolated deletion of *ETV6*, *PAX5* or *BTG1* *ETV6* + single deletion of *BTG1*, *CDKN2A/B* or *BTG1*
	Intermediate risk	All other CNA profiles
	Poor risk	Isolated *IKZF1*, PAR1 or *RB1* deletion Deletion of *IKZF1*/*PAX5*/*CDKN2A/B*

Table 1.4. Cytogenetic and CNA risk categories based on international study of 3239 cases of childhood BCP-ALL. Table adapted from Hamadeh et al (Hamadeh et al., 2019). Cytogenetic risk group and UKALL-CNA risk group are combined to form an overall genetic risk group. Genetic good risk consists of patients with cytogenetics good risk with any UKALL-CNA risk or cytogenetic intermediate risk with UKALL-CNA good risk. Genetic poor risk consists of patients with cytogenetic poor risk with any UKALL-CNA risk or cytogenetic intermediate risk with UKALL-CNA intermediate or poor risk.
HeH: high hyperdiploidy; B-other: BCP-ALL with no primary chromosomal abnormality identified; HoTr: low hypodiploidy/near triploidy; CNA: copy number abnormality; PAR1: pseudoautosomal region at Xp22.3/Yp11.2.

In adults, limited studies have shown specific CNAs to have an adverse effect on prognosis (e.g. deletion of *EBF1*, *IKZF1*, and *CDKN2A/B*) (Jordi et al., 2015). However, these have failed to be validated (Moorman et al., 2019) and detailed copy number analyses have not yet been performed in older patients.

1.11 Mutational landscape of ALL and relevance to relapse

Together with primary chromosomal aberrations and co-operating CNAs, ALL genomes harbour a range of smaller abnormalities including single nucleotide variants (SNVs) and small insertions and deletions (indels) that can only be detected through next generation sequencing (NGS) techniques. As with most genetic and genomic

profiling of ALL, studies have focussed on children and younger adults. Such studies have identified a low mutational burden with <10-20 non-silent mutations per case, although occasional relapse samples show a hypermutated phenotype (Mullighan, 2014). Despite the low total number of coding variants, recurrent mutations in key pathways are well described in both T and B-cell disease (tables 1.5 and 1.6).

Cellular process	Genes
Signaling	*KRAS*, *NRAS*, *NF1*, *PTPN11, FLT3, JAK1, JAK2, SH2B3, IL7R, STAT5B*
Epigenetic modification	*EZH2*, *CREBBP*, *SETD2*, *KMT2D, WHSC1, TRRAP, SETD1B, KMT2C, CTCF, KMT2A, ASXL1, ARID1B, NCOR2, CHD4, KDM6A, ARID1A, EP300, TET2, CHD8, ASXL2*
Transcription factors	*PAX5*, *IKZF1*, *ZEB2*, *ETV6*, *MGA, RUNX1, MYC*
Cell cycle regulation	*TP53*, *CDKN2A, MED12*
Others	*HERC1*, *SACS*, *USP9X*, *ASPM*

Table 1.5. Recurrent mutations identified through transcriptome sequencing of 1,223 BCP-ALL cases. Mutations data were obtained from whole exome sequencing of predominantly childhood BCP-ALL samples. Adapted from Li et al (Li et al., 2018).

Cellular process	Genes
NOTCH1 signaling	*FBXW7, NOTCH1*
Transcription factors	*BCL11, ETV6, GATA3, HOXA, LEF1, MYB, RUNX1, WT1*
Signalling	*AKT, DNM2, FLT3, JAK1, JAK3, IL7R, KRAS, NRAS, PI3KCA, PTEN, PTPN2, STAT5B*
Epigenetic factors	*DNMT3A, EED, EZH2, KDM6A, PHF6, SUZ12*
Translation and RNA stability	*CNOT3, mTOR, RPL5, RPL10, RPL22*

Table 1.6. Recurrent mutations in T-ALL. Activating *NOTCH1* mutations are the most common finding and epigenetic mutations are present in more than 50% of cases. Adapted from Girardi et al (Girardi et al., 2017).

Moreover, patterns are increasingly recognised between the mutational landscape at diagnosis and relapse. As the presumed initiating events, primary chromosomal abnormalities are almost always retained between diagnosis and relapse. Studies into relapsed childhood ALL demonstrate a similar mutational landscape at diagnosis and relapse in the majority of cases (Waanders et al., 2020). However, genes associated with drug metabolism (e.g. *NT5C2* for purine metabolism) are specifically acquired at relapse (Barz et al., 2020, Tzoneva et al., 2013). Such findings are consistent with positive selection of a clone which confers resistance to the purine analogues given

during maintenance chemotherapy, and illustrates the effects of a therapeutic selection pressure. Moreover, clear correlation has been seen between a minor subclone at diagnosis and relapse-founder clones (Ma et al., 2015, Waanders et al., 2020). Specifically, multiple mutations may be seen in RAS pathway genes at diagnosis, with a single minor subclone shown to drive disease relapse (Irving et al., 2014, Ryan et al., 2016).

Importantly, given the highly intensive and successful treatment delivered to children with ALL, detecting and quantifying these subclonal populations during remission is challenging (Ma et al., 2015). However, research into older adults receiving much less intensive therapy could provide a clearer insight into the fluctuation of subclones on treatment. In these patients, periods of remission are likely to be related to disease control rather than eradication, and understanding these mechanisms could pave the way to modifying therapy to control and suppress the disease to subclinical levels, potentially improving both quantity and quality of life for older adults with ALL. In the GMALL 05/93 trial, minimal residual disease (MRD) of 10^{-4} cells or higher was detectable in 71% of intensively treated adults with ALL at day 29 (post induction 1) and in 42% after induction 2 (Brüggemann et al., 2006). A significant proportion of follow up samples on even less intensively treated older patients are therefore likely to contain ALL cells detectable by sensitive molecular techniques and there have been no studies examining the dynamics of leukaemic clones over a long period of time and in the context of a therapeutic selection pressure. The longitudinal study of such patients would therefore provide a valuable insight into the biology and clonal dynamics of ALL in older adults, and could provide a novel understanding of clonal fluctuations on treatment. Ultimately, such data could inform the development of additional targeted therapies or re-purpose those already in existence.

1.11.1 Age-related clonal haematopoiesis

In childhood, it is clear that certain genetic abnormalities (e.g. *ETV6-RUNX1* fusion) may be present antenatally, and together with co-operating mutations acquired later in childhood, result in the development of ALL (Alpar et al., 2014). It is also well recognised that many of these primary lesions alone are insufficient to cause leukaemia. The prevalence of *ETV6-RUNX1* fusion in neonatal Guthrie spots is 100x

greater than the prevalence of ALL (Mori et al., 2002), indicating that additional lesions such as CNAs are required to cause clinical disease (section 1.4).

Much less is known about the pathogenesis of ALL in older adults, but it is reasonable to assume that it does not arise antenatally but rather via the sequential acquisition of somatic mutations during the lifetime of the patients; potentially over many years. In 2014, three large population-based studies identified a number of acquired mutations in haematopoietic cells that become more prevalent with age (table 1.7) (Jaiswal et al., 2014, Genovese et al., 2014, Xie et al., 2014). Affected individuals did not have clinical or biochemical evidence of haematological disorders and the mutations instead reflected a state whereby the pool of haematopoietic progenitors had become clonally restricted and was thus termed age-related clonal haematopoiesis. The most prevalent mutations were in the epigenetic regulators *DNMT3A*, *TET2* and *ASXL1*. These incidental findings have been identified in 10% of people over 70, and their prevalence continues to rise with advancing age (Jaiswal et al., 2014, Natarajan et al., 2018).

Gene	Function of Gene Product
DNMT3A	De novo genome-wide methylation
TET2	Conversion of methylcytosine to 5-hydroxymethylcytosine, including downstream DNA demethylation
ASXL1	Chromatin-binding protein regulating transcription through nuclear hormone receptors
TP53	Cell cycle regulation
JAK2	Protein tyrosine kinase
SF3B1	RNA splicing regulation
GNB1	Transmembrane signal transduction and modulation
CBL	Proteosomal degradation
SRSF2	Promotes RNA splicing
GNAS	Transmembrane signal transduction and modulation
PPM1D	Negatively regulates p38 MAPK
BCORL1	Interacts with histone deacetylases to repress transcription

Table 1.7. Genes with largest burden of somatic mutations in blood cells from population-based studies of individuals without haematological disease. Adapted from Natarajan et al (Natarajan et al., 2018)

Although these mutations are associated with clonal expansion and survival advantage compared to wild type cells, they are typically present at a low variant allele frequency (VAF) (mean 0.09) and most individuals carrying these mutations have no evidence of haematological disease. However, clonal haematopoiesis is associated with increased risk of AML as well as cardiovascular disease (Natarajan et al., 2018, Jaiswal et al.,

2014) and the state is now more commonly termed clonal haematopoiesis of indeterminate potential (CHIP).

Interestingly, the discovery of clonal haematopoiesis has also helped to confirm the leukaemia cell of origin in AML. *DNMT3A*-mutated AML patients were found to harbour the same mutation in their T cells, albeit at a lower VAF. This indicated that *DNMT3A* mutation had occurred very early in the disease process, probably in an HSC. The mutations were shown to result in clonal expansion of a pre-leukaemic HSC pool, from which the leukaemia arose (Shlush et al., 2014), confirming this as a disease-initiating event.

In clonal haematopoiesis, *DNMT3A* and *TET2* mutations are reported across their coding sequences and cause a loss of function effect. In comparison, the *ASXL1* variants are truncating mutations which occur in hotspot regions in exons 11 and 12 (Gelsi-Boyer et al., 2012) and are thought to result in gain of function (Natarajan et al., 2018). Regarding lymphoid malignancies, *DNMT3A* mutations have already been identified in a high frequency of patient with early-T precursor ALL (ETP-ALL) (Neumann et al., 2013) and *TET2* mutations are implicated in lymphomagenesis (Quivoron et al., 2011). However, to date, a link with ALL in older adults has not been investigated.

1.12 Aims and objectives

An ageing population together with high morbidity from the current treatment of older adults with ALL creates a disproportionate burden on healthcare resources. Conventional chemotherapeutic approaches are associated with a very high rate of complications in older patients, often leading to prolonged hospitalisations, which in turn necessitate the use of costly interventions (Sive et al., 2012, Sancho et al., 2007). Affected patients are further impacted by a decline in functional status contributing to a vicious cycle of increasing frailty and reduced fitness for treatment. Judicious use of more targeted therapy based on a comprehensive genomic characterisation and regular in-depth re-assessment of the disease's genomic profile could prolong both quality and quantity of life for these patients. The overarching aim of this research was

therefore to characterise the genetic and genomic landscape of ALL in older adults, and assess its evolution on treatment, ultimately paving the way for novel therapies.

Using cytogenetic and molecular techniques, the project was divided into the following objectives.

1. Identify primary chromosomal abnormalities in a large cohort of older adults enrolled in the UKALL14 and UKALL60+ clinical trials. Where standard diagnostic analyses had not identified a primary abnormality, extended FISH experiments were planned to characterise these B-other patients.
2. Detect and classify the secondary CNAs that drive leukaemogenesis in older adults. Using high density SNP arrays, a comprehensive profile of the genomic deletions and gains needed to be generated from older adults in the UKALL14 and UKALL60+ clinical trial cohort.
3. Optimise the use of SNP arrays to accurately classify patients with primary ploidy shift and identify discrepancies with cytogenetic analyses.
4. Identify mutations and perform serial monitoring of the clonal architecture through treatment. A small cohort of prospectively recruited older adults with ALL first needed to be defined. Sensitive NGS techniques could then be designed on diagnostic and follow up material to determine and quantify the fluctuation of leukaemia-associated mutations through treatment. The presence and role of clonal haematopoiesis mutations could also be assessed in these samples.

Chapter 2. Materials and Methods

2.1 Ethical approval

All experiments were performed on primary patient material. Patients were enrolled in the UKALL14 or UKALL60+ clinical trials and had existing REC approval in place (Study title: Genetic and functional characterisation of adult acute lymphoblastic leukaemia using patient specimens, REC reference: 16/LO/2055, IRAS project ID: 179685)

Separately, patients were recruited within the Newcastle-upon-Tyne Hospitals NHS Foundation Trust for sample biobanking at the Newcastle Biobank (Study title: Collection of samples to support the Newcastle Academic Health Partners Bioresource, REC reference: 12/NE/0395).

Informed consent was given by all patients in accordance with the Declaration of Helsinki and research ethics committee approval.

2.2 Materials

2.2.1 Manufacturers

Goods and services throughout the project were obtained from all the manufacturers listed in table 2.1. Standard laboratory equipment such as microsyringe tubes, pipettes and pipette tips are not included.

Manufacturer	Location
Abbott Molecular	Chicago, IL, USA
Affymetrix	Santa Clara, CA, USA
Agilent Technologies	Santa Clara, CA, USA
Applied Biosystems	Waltham, MA, USA
Beckman Coulter	Brea, CA, USA
BGM Labtech	Ortenberg, Germany
Cytocell	Cambridge, UK
Diagenode	Liège, Belgium
Eppendorf	Hamburg, Germany
Fisher Scientific	Loughborough, UK
IKA	Oxford, UK
Illumina	San Diego, CA, USA
Isohelix	Harrietsham, UK
Kreatech Diagnostics	Amsterdam, Netherlands
Labnet International	Edison, NJ, USA
Leica Biosystems	Nussloch, Germany
MRC Holland	Amsterdam, Netherlands
New England Biolabs	Ipswich MA, USA
Qiagen	Venlo, Netherlands
Santa Cruz Biotechnologies	Dallas, TX, USA
Sigma-Aldrich	St. Louis, MO, USA
Stem Cell Technologies	Vancouver, BC, CA
Thermo Scientific	Waltham, MA, USA
ThermoFisher Scientific	Waltham, MA, USA
Twist Bioscience	San Francisco, CA, USA
VWR International	Radnor, PA, USA
Zeiss	Oberkochen, Germany

Table 2.1. List of suppliers

2.2.2 Laboratory equipment

All equipment used to conduct the laboratory experiments throughout this project are listed in table 2.2.

Equipment	Manufacturer
Axioskop microscope	Zeiss
Bioanalyzer 2100	Agilent Technologies
Bioanalyzer chip priming station	Agilent Technologies
Bioruptor Pico	Diagenode
Bioruptor Pico microtubes (0.65ml)	Diagenode
Coplin jars	Thermo Scientific
Cover glasses (rectangular 24x50mm)	VWR International
Cover glasses (round 13mm)	VWR International
DNA LoBind tubes	Eppendorf
Fluostar Omega	BGM Labtech
Microscope slides	VWR International
MS3 vortex mixer	IKA
Mr Frosty freezing container	Thermo Scientific
Nanodrop 1000	Thermo Scientific
Nutator mixer	
SK-1S buccal swabs	Isohelix
Tapestation 4200	Agilent Technologies
Thermal Cycler 2720	Applied Biosystems
Thermobrite	Leica Biosystems
Vaccum Concentrator 5301	Eppendorf
Vortex mixer	Labnet International

Table 2.2. List of laboratory equipment required

2.2.3 Analytical software

Specialist computer software required for statistic, SNP array and bioinformatic analyses is detailed in table 2.3.

Software	Manufacturer
Bioconductor packages (various)	Bioconductor
BlueFuse Multi	Illumina
Chromosome Analysis Suite	Affymetrix
Cytovision 7.1	Leica Biosystems
GeneMarker V1.85	SoftGenetics
GenomeStudio 2.0	Illumina
Integrative Genomics Viewer	Broad Institute
Nexus copy number 10	Biodiscovery
R version 4.0	R Foundation for Statistical Computing

Table 2.3. List of computer software used

2.2.4 Chemicals and reagents

All chemicals and reagents that were not part of standardised experimental kits are listed in table 2.4.

Chemical/reagent	Manufacturer
20 x saline-sodium citrate (SSC) buffer	Fisher Scientific
4′,6-diamidino-2-phenylindole (DAPI)	Cytocell
Acetic acid (100%, glacial)	Fisher Scientific
Agencourt AMPure XP	Beckman Coulter
Distilled water	Newcastle University
Dimethyl sulfoxide (DMSO)	Sigma Aldrich
Dynabeads MyOne Streptavidin T1	Invitrogen
Ethanol (100%)	Fisher Scientific
Fetal Bovine Serum	Gibco
Hybridisation solution	Cytocell
Igepal-CA-630	Sigma Aldrich
Low TE buffer	Sigma Aldrich
Lymphoprep	Stem Cell Technologies
Methanol	Fisher Scientific
Nuclease free water	Invitrogen
Phosphate-buffered saline (PBS)	Newcastle University

Table 2.4. List of chemicals and reagents used.

2.2.5 Experimental kits

Experimental kits containing reagents and protocols for specific laboratory techniques were used and are detailed in table 2.5.

Experimental kit	Manufacturer
DNA 1000 kit	Agilent
DNA Isolation kit: DDK-50 / DDK-3	Isohelix
DNeasy Blood and Tissue Kit	Qiagen
Genomic DNA screen tape analysis kit	Agilent
High sensitivity DNA kit	Agilent
Quant-iT dsDNA Assay Kit, broad range	Invitrogen
REPLI-g Mini Kit	Qiagen
SALSA MLPA P335-IKZF1 probemix	MRC Holland
SureSelect XT2 reagent kit	Agilent

Table 2.5. List of experimental kits used.

2.2.6 Services

Specific service providers were utilised for SNP array and next generation sequencing experiments, which are conducted in specialised facilities (table 2.6).

Service	Manufacturer	Provider
Cytoscan HD arrays	Affymetrix	Newcastle Hospitals
CytoSNP 850k arrays	Illumina	Newcastle Hospitals
MLPA fragment analysis	MRC Holland	DBS Genomics
NextSeq 550 sequencing	Illumina	Newcastle University
NovoSeq sequencing	Illumina	Newcastle University
Twist Human Exome kit	Twist Bioscience	Newcastle University

Table 2.6. Internal and external providers used for array and next generation sequencing.

2.3 Patients and samples

Primary patient samples were obtained from two large UK clinical trials.

UKALL14 was a phase 3 randomised controlled trial that recruited patients with newly diagnosed ALL between 30/12/2010 and 26/07/2018. Patients were aged between 25-65 years (or 19-65 if Philadelphia positive) at diagnosis, and needed to be sufficiently fit to tolerate intensive multi-agent chemotherapy and potential allogeneic stem cell transplantation.

UKALL60+ was a phase 2 trial that recruited patients with newly diagnosed ALL between 07/12/2012 and 21/12/2018. Inclusion criteria specified that patients were aged ≥60 years at diagnosis (or ≥55 years if deemed unfit for UKALL14 treatment). A number of treatment arms were available based on the treating physician's assessment of patient fitness, ranging from intensive multi-agent high dose chemotherapy to largely outpatient based non-intensive treatment.

All patients had given written informed consent for data collection and genetic studies as specified by the trials' protocols.

Separately, newly diagnosed ALL patients at the Newcastle-upon-Tyne Hospitals NHS Foundation Trust and aged ≥60 years were approached to donate samples to the Newcastle Biobank (REC 12/NE/0395) between August 2017 and August 2019. Bone marrow samples were obtained at diagnosis and during follow up in line with clinical standard of care for the relevant treatment protocol. Additionally, buccal samples for DNA extraction were obtained using SK-1S DNA swabs (Isohelix). All patients were treated on established standard of care protocols based on the UKALL14 or the UKALL60+ clinical trials and treatment varied from intensive multi-agent chemotherapy to non-intensive predominantly oral outpatient regimes.

All genetic and genomic analyses performed on patient samples along with baseline demographic characteristics are outlined in supplementary table 1.

2.4 General molecular biology techniques

2.4.1 Isolation of mononuclear cells from patient bone marrow samples

Approximately 3-5 ml of liquid bone marrow in EDTA was obtained from the locally recruited patients at each time point of bone marrow sampling. Mononuclear cells were isolated from fresh bone marrow aspirate samples using a density gradient medium (Lymphoprep) in aseptic conditions under a laminar flow hood, sterilising all surfaces with 70% ethanol.

One part bone marrow was first diluted in 3 parts PBS in a 30 ml universal container. Two 8ml aliquots of lymphoprep medium were then introduced into separate 30ml universal containers. Half of the diluted bone marrow was then very slowly placed on top of each volume of lymphoprep, with special care to avoid disrupting the interface between the lymphoprep and the bone marrow. The samples were then centrifuged at 800 x g for 30 minutes with the centrifuge brake off. Following centrifugation, a thin buffy coat layer consisting of mononuclear cells was visible between the lymphoprep medium and the top layer of plasma. Red cells and granulocytes have a higher density than mononuclear cells and migrated through the lymphoprep medium to the bottom of the universal container during centrifugation. Using a 3ml disposable Pasteur pipette, the buffy coat layer was carefully aspirated and transferred to a fresh universal container. The mononuclear cells were re-suspended in 15ml with sterile PBS and centrifuged at 400 x g for 10 minutes. This caused the mononuclear cells to pellet at the bottom of the universal container and the supernatant was then discarded. The cells were re-suspended in 5 ml of sterile PBS and centrifuged for a further 5 minutes at 400 x g. The PBS was again discarded and the cells were then re-suspended in a solution of 10% DMSO in FBS for cryopreservation. The volume of freeze media was varied based on the pellet size, aiming for an approximate concentration of 1-5 x 10^6 cells/ml. The final cell suspension was then split into 1 ml aliquots, placed in cryovials and cooled to -80ºC in a Mr Frosty Freezing Container to achieve a rate of cooling of ~1ºC/min. The following day, the samples were transferred to a -150ºC freezer for long term storage.

2.4.2 Extraction of DNA from viable cells

The DNeasy kit for the purification of total DNA from animal blood or cells (spin-column protocol) was used to extract DNA from the mononuclear cells that had been isolated from patient bone marrow samples (section 2.4.1). Viable cells were allowed to thaw at room temperature and then centrifuged for 5 minutes at 300 x g to pellet the cells. The freeze media was removed and discarded, with care not to disturb the cell pellet. The cells were then re-suspended in 200 ul PBS and centrifuged again for 5 minutes at 300 x g to then remove and discard the supernatant containing any residual freeze media. The viable cells were re-suspended in fresh 200 ul PBS and 20 ul proteinase K was added, followed by 4 ul RNase. Next, 200 ul Buffer AL was thoroughly mixed into the sample by vortexing, before incubating at 56ºC for 10 minutes. Following this, 200 ul of 96-100% ethanol was added to the sample and mixed thoroughly by vortexing. The mixture was then carefully transferred to a DNeasy mini spin column placed in a 2 ml collection tube. The column was centrifuged at 6000 x g for 1 minute and the flow through in the collection tube was discarded. The column was placed in a fresh 2 ml collection tube and 500 ul of Buffer AW1 was added to the spin column. The mixture was then centrifuged at 6000 x g for 1 minute and the flow through discarded as previously. The spin column was placed in a fresh collection tube and 500 ul of Buffer AW2 was added, before centrifuging at 20,000 x g for 3 minutes to dry out the DNeasy membrane and discarding the flow through and collection tube. The spin column was then placed in a pre-labelled 1.5 ml Eppendorf tube. To elute the DNA, 30 ul of nuclease free water was first placed directly onto the DNeasy membrane in the spin column and allowed to incubate at room temperature for at least 1 minute. The Eppendorf tube containing the spin column was then centrifuged at 6000 x g for 1 minute. To maximise DNA yield a second elution was also performed by placing a further 30 ul nuclease free water onto the spin column in a fresh pre-labelled 1.5 ml Eppendorf tube and incubating the sample overnight at 4ºC. The following day, the spin column was centrifuged again at 6000 x g for 1 minute and the second DNA elution collected.

2.4.3 Extraction of DNA from buccal swabs

Buccal DNA samples were obtained from the locally recruited patients in accordance with the manufacturer's recommendations by rubbing the SK-1S swab against the inside of the patient's cheek for 1 minute, before placing it in the sealed tube provided.

Buccal DNA was extracted using the DDK-3/DDK-50 protocol. The swab was carefully removed from the sealed tube and placed in a fresh 1.5 ml centrifuge tube, ensuring the swab head was at the bottom of the tube. To stabilise the DNA, 500 ul of LS solution was added, followed by 20 ul of PK solution. The tube contents were mixed on a vortex mixer and at this point the DNA was stable at room temperature for at least 3.5 years. To isolate the sample DNA, the tube was first heated to 60ºC on a heating block. To increase DNA yield before discarding the swab, any remaining liquid in the swab was then extracted by briefly centrifuging the swab in a separate 1.5 ml centrifuge tube and adding the supernatant to the rest of the sample. Following this, 500 ul of CT solution was added and the sample was mixed using a vortex mixer and then centrifuged at maximum speed for 7 minutes to pellet the DNA. The supernatant was then removed and discarded, with care not to disturb the DNA pellet. A further centrifugation was then performed at 12,000 x g for 1 minute and again any remaining supernatant was removed and discarded. The pelleted DNA was then re-hydrated by adding 30 ul of nuclease free water and allowing the DNA to elute into the water for 30 minutes. The contents were then mixed on a vortex mixer and centrifuged at maximum speed for 15 minutes to pellet any undissolved debris. Following this, the supernatant, containing the extracted DNA, was transferred to a sterile 1.5 ml tube and the pellet of undissolved debris was discarded.

2.4.4 Whole genome amplification (WGA)

A minimum of 1 ug of sample DNA was required for the SureSelect XT2 library prep (section 2.7.2). Whole genome amplification (WGA) of relevant samples was therefore first required for sufficient input DNA. This was performed using the Qiagen REPLI-g mini kit and protocol. Denaturation and neutralisation buffers were prepared according to the manufacturer's protocol. Briefly ~20ng of sample DNA in 5 ul was first denatured in a gentle alkaline denaturation step by adding 5 ul of buffer D1 (denaturation buffer) and incubating at room temperature for 3 minutes. The reaction was then stopped with

the addition of 5 ul of buffer N1 (neutralisation buffer) and a master mix containing a high-fidelity polymerase enzyme (Phi 29 polymerase) was then incubated with the sample for 10-16 hours to allow whole genome amplification. Following WGA, the concentration of double stranded DNA was measured using the Quant-iT PicoGreen dsDNA Assay Kit and the FLUOstar Omega microplate reader.

2.4.5 Assessing DNA concentration and quality

DNA samples for MLPA and/or SNP array were assessed by Nanodrop only. Samples for exome sequencing and SureSelect XT2 library prep were also assessed by Quant-iT for accurate double stranded DNA (dsDNA) quantification and 4200 TapeStation System or 2100 Bioanalyzer for fragment sizing.

2.4.5.1 Nanodrop spectrophotometer

The Nanodrop 1000 spectrophotometer consists of two fibre optical cables which are located on the sample pedestal and on a mobile sampling arm. A sample is placed between the fibre optical cables and a light source and detector are used to measure absorbance at specific wavelengths (230, 260 and 280 nm), which can determine the purity and concentration of nucleic acids in the sample.

Before use, the instrument was first cleaned and primed by loading a 1 ul water sample on the lower pedestal. A blank measurement was then made with 1 ul of the relevant DNA eluent (water or TE buffer). Next, 1ul of the DNA sample to be tested was carefully placed on the Nanodrop lower pedestal and the concentration and A260/A280 ratio measured.

2.4.5.2 dsDNA concentration using Quant-iT Picogreen dsDNA reagent

Quant-iT Picogreen dye emits a fluorescent signal upon binding to dsDNA, which does not occur in the presence of RNA a single stranded DNA. The strength of fluorescence rises linearly with dsDNA concentration over a broad range so can be used for accurate dsDNA quantification. Sufficient quantities of good quality dsDNA are particularly crucial for NGS-based assays.

All reagents of the Quant-iT dsDNA Assay Kit, broad range were allowed to equilibrate at room temperature for 30 minutes and a working solution was created by diluting 1 ul Quant-iT Picogreen dsDNA reagent in 200 ul Quant-iT dsDNA BR buffer per sample to be tested. For each run, 8 standards with known DNA concentrations were included in duplicate. A flat-bottomed black 96 well plate was used and 200 ul of the diluted picogreen reagent was introduced into each well to be tested. Next, 10 ul of each standard was introduced into pre-specified wells and 1 ul of each sample DNA into other wells. The plate was sealed, mixed on a vortex mixer and briefly centrifuged to collect the liquid. The fluorescent signals emitted from each well were then measured on a FLUOstar Omega microplate reader, and a standard curve from the reference samples was constructed, which permitted quantification of dsDNA in all test samples.

2.4.5.3 Assessing DNA fragment size using Agilent 4200 TapeStation System

The Agilent 4200 TapeStation System allows automated electrophoresis for DNA fragment sizing and only requires 1 ul of sample DNA input. This was specifically used for quality control of DNA samples prior to exome sequencing. D1000 reagents were allowed to equilibrate at room temperature for 30 minutes. D1000 ScreenTape and loading tips were then introduced into the instrument. All reagents were mixed thoroughly on a vortex mixer and 3 ul of D1000 sample buffer was added into each position of a tube strip to be tested, followed by 1 ul of D1000 ladder in the A1 position of the tube strip and 1 ul of sample DNA in the other wells. The liquids were mixed at 2000 rpm on an IKA vortex mixer for 1 minute and briefly centrifuged to collect the liquid at the bottom of the tube strip wells. The samples were then ready to be loaded and analysed by automated electrophoresis on the 4200 TapeStation. A DNA integrity number (DIN) was computed by the software based on the distribution of sizes of DNA fragments in the sample. Highly degraded DNA would produce a much lower DIN than samples with well-preserved DNA. For exome sequencing, a DIN>7 was required for successful library prep.

2.4.5.4 Assessing DNA fragment size using Agilent 2100 Bioanalyzer

The 2100 Bioanalyzer performs automated gel electrophoresis and was used for fragment sizing and quantification of SureSelect XT2 libraries. DNA 1000 or High Sensitivity chips and reagents were allowed to equilibrate at room temperature for 30

minutes and mixed on a vortex mixer. Gel dye mix was first prepared by adding 25 ul of dye concentrate to the DNA gel matrix. This was mixed by vortexing for 10 seconds and then transferred to the top receptacle of a spin filter. The gel dye mix was centrifuged at 2240 x g for 15 minutes and the filter could then be discarded. The chip priming station was set to the lower position and 9 ul of gel dye mix was loaded into the specified gel priming well. The chip priming syringe plunger was set to the 1 ml mark and closed by applying a constant and even pressure, and then held in position with the securing clip. After exactly 1 minute, the securing clip was released, allowing the plunger to rise gradually to the 0.3 ml position, before raising it back to the 1 ml mark over a further 5 seconds. Next, 9 ul of gel dye mix was placed in the specified wells, followed by 5 ul of DNA marker into the ladder and all sample wells. Lastly, 1 ul of sample DNA was added to testing wells and 1 ul of DNA ladder to the pre-specified ladder well. The contents of the DNA chip were then thoroughly mixed on an IKA MS3 vortex mixer set to 2400 rpm for exactly 1 minute. The DNA chip was then loaded into the 2100 Bioanalyzer for DNA fragment analysis using the built-in software.

2.5 Fluorescence *in situ* hybridisation (FISH) to detect gene rearrangements

2.5.1 FISH introduction

FISH is based on the principle that single stranded DNA can anneal to a complementary sequence. The method was developed in the late 1980s and has permitted the targeted detection of many specific genetic aberrations (Bishop, 2010). In contrast to karyotyping, which is limited by the need for actively dividing cells to create a metaphase spread, FISH can detect abnormalities in interphase nuclei, which are usually much more abundant in primary patient samples. Specific probes are designed with sequences complementary to the DNA regions of interest. The probes are either directly or indirectly labelled to permit visualisation using fluorescence microscopy. Direct probe labelling involves integration of a fluorophore, a compound which is able to re-emit light following light excitation (Speicher and Carter, 2005). In

comparison, indirect labelling relies on the inclusion of a hapten molecule, which in turn has affinity for a secondary detection reagent (Huber et al., 2018).

FISH probes are usually created from Bacterial Artificial Chromosome (BAC) clones. These are cloned human DNA sequences, around 150-350 kb, that cover regions across the human genome, and can be amplified in bacteria (usually E. coli). A fluorochrome is integrated through nick translation. This involves creating a “nick” in the DNA sequence and extracting nucleotides through the action of polymerase I. These are then replaced with fluorescently tagged dNTPs, creating a fluorescent marker on the probe DNA sequence (Rigby et al., 1977).

Hybridisation of probe to sample DNA requires denaturation of both the probe and target DNA. Although typically, a temperature of 95°C is required for DNA denaturation, the probe hybridisation solution contains formamide, an organic solvent that allows denaturation to take place at a lower temperature. Following hybridisation, unbound or weakly bound probe then needs to be removed with a series of washes. Wash solutions, which are slightly more stringent than the probe buffer solution, are used to achieve this. A counter stain such as 4’,6’-diamidino-2-phenylindole dihydrochloride (DAPI), which binds to all DNA, is then applied to facilitate visualisation of the signals in the context of individual nuclei.

A FISH method can be used to identify rearrangements affecting *ABL1*, *ABL2*, *PDGFRB*, *CSF1R*, *CRLF2* and *JAK2*, the most common abnormalities driving the Ph/*BCR-ABL*-like signature (Schwab et al., 2016). Dual colour break-apart probes consisting of red and green fluorochromes, which hybridise to genomic loci flanking the gene of interest, are used to detect such rearrangements. Two sets of adjacent red and green signals (2 fused signals) within an interphase nucleus indicate two normal copies of an undisrupted gene. In contrast, a single fused signal and separated red and green signals indicates a typical gene rearrangement pattern, often arising through a chromosomal translocation (figure 2.1).

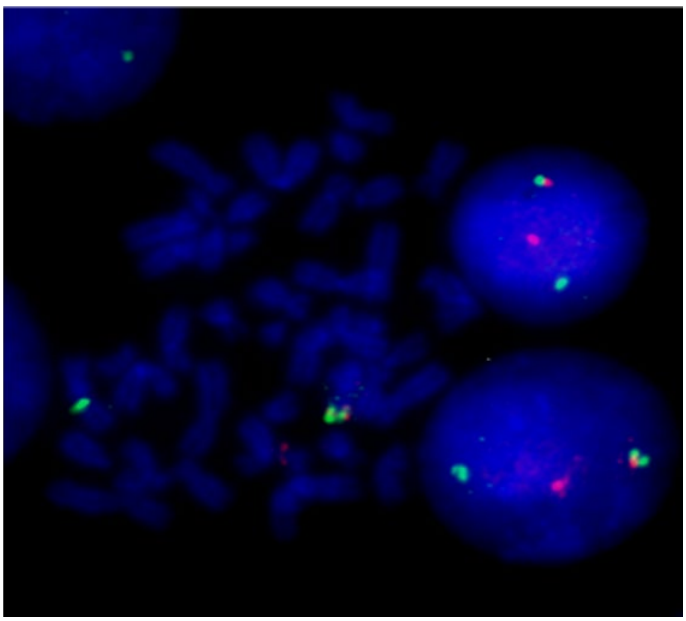


Figure 2.1. Example fluorescence *in situ* hybridisation (FISH) image of a *KMT2A* gene rearrangement using *KMT2A* break apart probe. Both nuclei demonstrate a single fused signal representing the intact *KMT2A* allele and split red and green signals representing the rearranged allele (image courtesy of Claire Schwab).

2.5.2 Patients and samples

The Leukaemia Research Cytogenetics Group (LRCG) database was searched for all patients aged 60 years and over at ALL diagnosis who were enrolled in the UKALL14 or UKALL60+ clinical trials. Patients lacking a known primary chromosomal abnormality were identified and those with available fixed cell samples were selected.

2.5.3 FISH probes

Commercially manufactured directly labelled dual colour break-apart probes were used to identify primary chromosomal rearrangements in patients hitherto lacking a recurrent cytogenetic abnormality (B-other ALL) (table 2.7).

Probe name	Company	Cytoband	Size (kb)	Position to gene	Fluorochrome colour
MEF2D	Cytocell	1q22	147	Centromeric	Red
		1q23.1	129	Telomeric	Green
ABL2	Cytocell	1q25.2	318	Centromeric	Red
		1q25.2	315	Telomeric	Green
PDGFRB/ CSF1R	Cytocell	5q33	107	Centromeric	Red
		5q33	154	Telomeric	Green
JAK2	Kreatech	9p24	550	Telomeric	Red
		9p24	525	Centromeric	Green
ABL1	Cytocell	9q34	283	Centromeric	Red
		9q34	309	Telomeric	Green
ZNF384	Cytocell	12p13.31	137	Centromeric	Red
		12p13.31	104	Telomeric	Green
IGH@	Cytocell	14q32.33	124	C-region	Red
		14q32.33	617	V-segment	Green
CRLF2	Cytocell	Xp22/Yp11	243	Centromeric	Red
		Xp22/Yp11	71, 131	Telomeric	Green

Table 2.7. Details and manufacturers of FISH probes used for characterisation of B-other ALL patients. All probes were dual colour break apart probes, with normal signal pattern consisting of two fused signals in interphase nuclei.

2.5.4 FISH method

Samples consisted of frozen fixed cells suspended in fixative solution. These were retrieved and thawed at room temperature. The samples were then centrifuged at 13,000 rpm on a bench-top centrifuge for 2 minutes to pellet the cells. The fixative solution was removed taking care not to disturb the cell pellet. Several drops of freshly prepared fixative solutions (table 2.8) were then added to each cell pellet to dilute this sufficiently to place a 3ul aliquot on a microscope slide ensuring nuclei could be readily seen without excessive overlap. The slides were air dried on a humid surface. The relevant FISH probe solution was then mixed briefly on a vortex mixer and 2ul aliquots were placed onto 13mm glass coverslips, which were subsequently applied to the fixed cell nuclei on each microscope slide. The coverslips were sealed in place with rubber solution and the slides were placed in a programmable Thermobrite probe

hybridisation system. This was heated to 75ºC for 5 minutes to denature the probe and sample DNA, then rapidly cooled to 37ºC and maintained at this temperature for a minimum of 16 hours (typically overnight) to allow probe hybridisation to their target sequences.

2.5.5 Post hybridisation washes

Following overnight hybridisation, the slides were removed from the Thermobrite system and the rubber solution was carefully peeled off. Each slide was then placed into 2xSSC solution (table 2.8) in a Coplin jar to loosen the coverslip, which was then gently removed. The slides were then soaked in Wash 1 solution heated to 72ºC for 2 minutes followed by Wash 2 solution at room temperature for a further 2 minutes (table 2.8). Separately, 9ul aliquots of DAPI were placed onto 24x50mm coverslips and these were subsequently applied to each slide after the washing procedure was completed.

Reagents	Volume	Composition	Storage
Fixative	40ml	30ml methanol 10ml acetic acid	RT
2X SSC	1 litre	900ml distilled water 100ml 20X SSC	RT
Wash 1	1 litre	980ml distilled water 20ml 20X SSC 3ml igepal-CA-630	RT
Wash 2	1 litre	900ml distilled water 100ml 20X SSC 1ml igepal-CA-630	RT

Table 2.8. Composition and storage of reagents used during FISH experiments.
SSC: saline-sodium citrate buffer; RT: room temperature.

2.5.6 FISH slide analysis

FISH signal patterns were visualised on a Zeiss Axioskop fluorescence microscope and recorded for a minimum of 50 cells by two independent observers. In the context of break apart probes, a normal pattern was represented by two fused signals and a typical pattern for a gene rearrangement consisted of split red and green signals together with a single fused signal, although loss of a green or red signal could also indicate gene rearrangements within the context of specific probes (e.g. *CRLF2*, see

chapter 3, section 3.4.2.1). Previous in-house validation of the relevant probes had confirmed a 10% cut off of abnormal nuclei to confidently report a variant signal pattern.

2.6 Detecting copy number abnormalities

Copy number abnormalities (CNAs) involve deletions or gains of genomic material in cancer cells. These can be detected either through targeted assays, focussing solely on a predefined panel of genes, or genome-wide techniques. The general principle of the detection of copy number abnormalities involves the hybridisation of probes to target regions in the genome. The abundance of each target sequence can then be measured through signal intensity or PCR amplification-based methods. Within this project, two separate methods were used for copy number analysis, namely single nucleotide polymorphism (SNP) arrays and multiplex ligation-dependent probe amplification (MLPA).

2.6.1 Single nucleotide polymorphism (SNP) arrays for the detection of CNAs

SNPs represent genomic sites of inter-individual variation and account for around 0.1% of the entire human genome (Shen et al., 2013). As well as their role in genome wide association studies (GWAS), techniques interrogating SNP loci have also been used to characterise the copy number states of sites throughout the human genome (Shen et al., 2008, Redon et al., 2006). Using subtly different chemistries, Affymetrix and Illumina SNP array platforms can be used to identify genome-wide CNAs as well as loss of heterozygosity (LOH). SNP arrays detect the intensity of a fluorescent signal released when oligonucleotide probes hybridise to their complimentary sequences. The probes are bound to an array, and each position on the array relates to specific genomic co-ordinates. Following hybridisation, the signal intensity at each position on the array is therefore related to the number of copies of a specific sequence.

2.6.1.1 Polymorphic and non-polymorphic probes

SNP array platforms consist of a variable number of non-polymorphic (copy number) probes and/or polymorphic (SNP) probes. Non-polymorphic probes cover regions of the genome with little inter-individual variation and hybridisation of target DNA fragments to probe DNA is therefore based predominantly on copy number status (Shen et al., 2008). In comparison, polymorphic probes target regions of single nucleotide polymorphisms. One of two possible alleles is present is these regions, denoted A or B. Based on the presence of a single maternal and single paternal chromosome in a normal diploid human genome, three possible allelic combinations can therefore arise at any SNP site (homozygous AA or BB or heterozygous AB alleles). As such, polymorphic SNP probes can both be used to determine the copy number state of these regions through loss of an allele (fewer target sequences binding to the array causing reduced signal intensity coupled with loss of heterozygous AB alleles), gain of an additional allele (more binding of target sequence resulting in increased signal intensity coupled with presence of AAA, AAB, ABB and BBB alleles) as well as identifying copy neutral loss of heterozygosity (CNN-LOH) (signal intensity similar to diploid background coupled with loss of heterozygous AB alleles) (figure 2.2).

The Affymetrix Cytoscan HD array consists of 1.9 million non-polymorphic (copy number) probes and 750,000 polymorphic (SNP) probes. Probes are 25 nucleotides long (25-mer probes) which are entirely or almost entirely complimentary to individual SNPs covered by the array (LaFramboise, 2009). All SNPs bind to the probes, but the binding is more efficient if the entire sequence is complimentary and a stronger signal is released. This variation in signal strength can then be used to determine the presence of an A or B allele at the relevant position in the genome.

In comparison, the Illumina CytoSNP 850k array consists exclusively of 850,000 polymorphic (SNP) probes. The probes are composed of a 50-mer oligonucleotide sequence bound to a bead. Each probe sequence is complementary to a defined region in the genome adjacent to a SNP of interest. A single base extension labelled with either a red or green fluorophore (targeting the A or B allele) is then required to produce a signal. The red or green signal produced is therefore dependent on the A or B allele at that site and the signal intensity based on the copy number status (LaFramboise, 2009).

2.6.1.2 Patients and samples

All samples from patients aged ≥60 years enrolled in the UKALL14 or UKALL60+ trials with DNA extracted from the diagnostic bone marrow sample were identified. DNA concentrations were measured on the Nanodrop spectrophotometer (section 2.4.5.1). An aliquot of 500ng of DNA was then taken from the primary sample and diluted with nuclease free water to a total volume of 10ul for a final concentration of 50ng/ul, to be used for SNP array.

2.6.1.3 SNP array protocol

SNP arrays were performed on the Affymetrix Cytoscan HD or Illumina CytoSNP 850k arrays. The SNP arrays were performed by the Core Genomics Facility at the International Centre for Life, Newcastle-upon-Tyne Hospitals NHS Foundation Trust and all subsequent array analysis was performed using Nexus Copy Number 10. Probe density of the two SNP array platforms is shown in table 2.9.

SNP array platform	SNP probes	Copy no. probes	Total
Affymetrix Cytoscan HD	750,000	1.9 million	2.67 million
Illumina CytoSNP 850k	850,000	-	850,000

Table 2.9. Details of Illumina and Affymetrix array probe characteristics. Affymetrix arrays contained both SNP and copy number probes whereas Illumina arrays consisted purely of SNP genotyping probes.

2.6.1.4 Data format and pre-processing

The raw signal intensities generated from a SNP array experiment can be influenced by many factors and normalising these values is therefore required before generating copy number and genotyping calls.

The output from Affymetrix arrays is a .CEL file, which consists of probe-level signal intensities in a text-based format (Scionti et al., 2018).

Log2 ratios are calculated using the formula:

$Log2\ ratio = Log2(sample_m / reference_m)$ where m represents each marker.

The reference log2 ratios are derived from a previously created reference model array file which includes 380 samples including 284 from the HapMap project (Scionti et al., 2018).

The output of Illumina CytoSNP 850k arrays is in the form of IDAT files, which are raw image rather than text-based files. Copy number status is derived from a logR ratio (LRR), which is similar to the log2 ratio, but is based on data from polymorphic rather than copy-number probes and is calculated using the formula:

$LRR = Log2(sample_R / reference_R)$ where R represents a combined intensity of the red and green signals.

The reference values are derived from an in-silico dataset imported in the form of an array-specific cluster file.

Signal intensities from the polymorphic SNP probes of both Affymetrix and Illumina arrays are used to determine the B-allele frequency. This complements information from the log2 ratio in identifying losses and gains, as well as characterising regions of copy number neutral loss of heterozygosity (CNN-LOH), which may arise when a genomic region is lost and then re-duplicated from the remaining homologue, potentially leading to the unmasking of recessive alleles (figure 2.2).

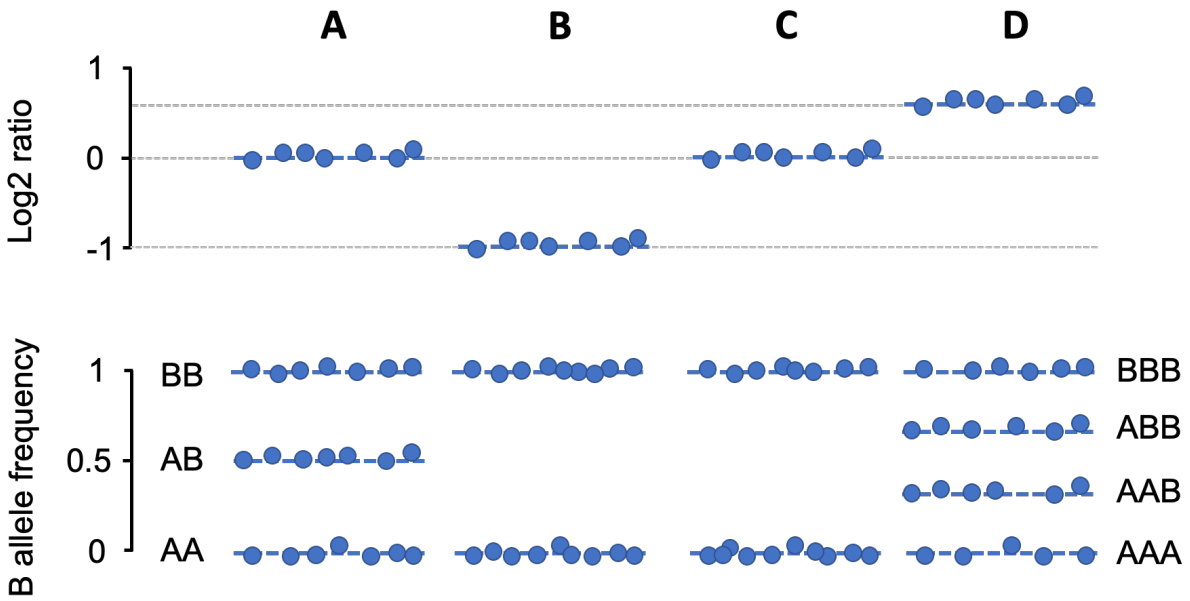


Figure 2.2. Illustration of different allelic patterns arising in SNP arrays. Each circle denotes the signal from a single probe. A represents a normal diploid pattern with log2 ratio centred around 0 (i.e. log_2 of 2/2 copy ratio) and B-allele frequency demonstrating homozygous AA and BB alleles and heterozygous AB alleles. B represents a heterozygous deletion with log2 ratio reduced to -1 (i.e. log_2 of ½ copy ratio) and B-allele frequency demonstrating loss of heterozygous AB alleles. C represents CNN-LOH with preserved log2 ratio of 0 but loss of heterozygous AB alleles. D represents single copy gain with log2 ratio increased to 0.58 (i.e. log_2 ratio of 3/2 copy ratio) and B-allele frequency demonstrating additional A or B allele.

2.6.1.5 Nexus Copy Number 10

Nexus is a desktop software package specifically designed for the analysis of copy number data. The programme is platform and manufacturer agnostic and can therefore load data from numerous different primary array and NGS platforms from which the raw copy number data has been derived. SNP array segmentation can then be performed either based on pre-defined standard platform-specific settings or customised settings to account for sample quality, background noise and the required balance between maximising detection of abnormalities and minimising false positive calls.

Raw array data from Affymetrix arrays can be loaded directly to Nexus in the form of .CEL files. In comparison, Illumina-generated IDAT files first needed to be converted into a text-based format before being loaded to Nexus. To achieve this, an Illumina-specific SNP array software package (GenomeStudio 2.0) was used, in accordance with the Nexus protocol for the analysis of Illumina arrays. The IDAT files were loaded to GenomeStudio 2.0 and then converted into text-based format by creating a Final_Report file. This was then loaded onto Nexus to visualise the data and perform copy number segmentation.

Systematic correction of the arrays was then performed. This is a recommended step in the analysis of SNP array data due to the waviness in the probe signals that can often be seen across the genome (Diskin et al., 2008). This is partly related to GC content as probes with high GC content will bind better to their target sequence, producing a higher signal intensity. As such, systematic correction was performed in Nexus using the recommended Illumina and Affymetrix correction files as provided by the manufacturer.

2.6.1.6 Copy number segmentation

Copy number analysis using SNP arrays relies on dividing the genome into segments of different copy number. Breakpoints of abrupt copy number change (from diploid) need to be identified to delineate regions of the genome that have been deleted or gained. This process of segmentation is performed computationally within SNP array analysis software and is achieved using two broad mathematical models, namely the Hidden Markov Model (HMM) and circular binary segmentation (CBS). Pure HMM-based segmentation assumes fixed integers of copy number within a sample (Seiser

and Innocenti, 2015, Fridlyand et al., 2004) and therefore relies on high tumour purity and lack of clonal heterogeneity. However, tumour samples frequently contain a mixture of tumour and normal DNA, as well as subclonal mutations within the cancer-cell population, which means that fixed integer levels of copy number state may not be apparent. In comparison, CBS models (Olshen et al., 2004) recursively partitions the genome into segments until no further statistically significant splits are possible between two adjacent segments (Tai et al., 2010). As such, copy number change is driven by observed variation between adjacent regions rather than assumption of fixed integer copy number states.

Nexus employs a hybrid segmentation algorithm termed Fast Adaptive States Segmentation Technique (FASST2). This is based on HMM-segmentation but does not assume fixed integer levels of copy number, and instead accepts a large number of potential copy number states falling between fixed integer levels.

2.6.1.7 Minimum segment size settings

Studies using SNP arrays to detect CNAs in ALL have used a number of different platforms and the minimum number of probes required for a CNA is highly dependent on array probe density. Early studies using the low density Affymetrix 100k and 250k arrays have used a minimum of 2 probes per copy number segment (Mullighan et al., 2007).

In comparison, the Affymetrix Cytoscan HD array has 2.67 million probes. Using this array, one study used a minimum threshold of 25 probes to define a copy number segment (Ribera et al., 2017). To strike a balance between the detection of novel abnormalities whilst minimising false discoveries, this analysis was performed using a minimum of 10 probes per copy number segment for Affymetrix Cytoscan HD data. Taking into account the reduced probe density of the Illumina CytoSNP 850k array, a threshold of 6 markers was used to create copy number segments from Illumina arrays. Log2 ratio thresholds and stringency settings were also adjusted with both Affymetrix and Illumina derived-data based on sample quality and degree of normal DNA contamination (chapter 4, section 4.3). All SNP array segmentation settings used are detailed in supplementary table 2.

2.6.1.8 Quality control

An QC measure is internally computed within Nexus (QC score) with lower values indicative of better quality. The QC score is based on the probe to probe variance across the genome, and aims to exclude probes where large scale variance is related to true copy number breakpoints. As such, the metric reflects short-scale variation between adjacent probes and acts as a measure of background “noise” in a SNP array profile. Based on the manufacturer’s advice, SNP arrays with QC scores <0.15-0.20 are likely to yield interpretable copy number profiles. A QC score <0.20 was therefore considered a requirement for further SNP array analysis.

2.6.2 Multiplex ligation-dependent probe amplification (MLPA) for the validation of CNAs

2.6.2.1 Patients and samples

In addition to SNP arrays, MLPA was performed on selected samples with remaining DNA to validate specific well-characterised gene deletions. All adults aged 60 years and over enrolled in the UKALL14 and UKALL60+ clinical trials were identified. Diagnostic patient samples with at least 100ng of DNA available following SNP arrays were selected for MLPA analysis.

2.6.2.2 Principle of MLPA

Multiplex ligation-dependent probe amplification (MLPA) is a targeted technique used to interrogate the copy number state of a pre-defined gene panel. The reaction involves DNA denaturation and hybridisation of test and reference probes to the denatured DNA. Each probe consists of two probe oligonucleotides. In the presence of the target sequence, a ligation reaction between the two probe oligonucleotides occurs. A PCR amplification step of all ligated probes follows and as each probe generates a uniquely sized amplification product, this can be detected by capillary electrophoresis. The relative different quantities of test and reference probes as compared to a control sample, can then be used to determine the copy number state of the specific genes under investigation.

2.6.2.3 MLPA technique using P335 kit (MRC Holland)

The P335 kit covers exons on 9 genes/loci commonly deleted in BCP-ALL, namely *EBF1* (5q33), *IKZF1* (7p12.2), *CDKN2A/B* (9p21.3), *PAX5* (9p13.2), *ETV6* (12p13.2), *BTG1* (12q21.33), *RB1* (13q14.2) and the PAR1 region (*CRLF2*, *CSF2RA*, *IL3RA* on Xp22.3/Yp11.2) and has been extensively validated in BCP-ALL (Schwab et al., 2010). Specific exons covered are detailed in table 2.10.

Cytoband	Gene	Number of probes	Exons covered
5q33	*EBF1*	4	16, 14, 10, 1
7p12.2	*IKZF1*	8	1, 2, 3, 4, 5, 6, 7, 8
9p24	*JAK2*	1	23
9p21.3	*CDKN2A*	2	5, 2A
9p21.3	*CDKN2B*	1	2
9p13.2	*PAX5*	7	10, 8, 7, 6, 5, 2, 1
12p13.2	*ETV6*	6	1, 2, 3, 5, 8
12q21.33	*BTG1*	4	1, 2
13q14.2	*RB1*	5	6, 14, 19, 24, 26
Xp22.3/Yp11.2	*SHOX*	1	NA
Xp22.3/Yp11.2	*CRLF2*	1	4
Xp22.3/Yp11.2	*CSF2RA*	1	10
Xp22.3/Yp11.2	*IL3RA*	1	1
Xp22.3/Yp11.2	*P2RY8*	1	2

Table 2.10. Details of individual exons covered by MLPA probes in IKZF1-P335 kit.

The protocol required an optimum input of 100ng of DNA in 5 ul TE buffer. Three controls (normal DNA from 2 females and 1 male) and a blank sample (TE buffer only) were prepared alongside the diagnostic DNA samples to be tested. Table 2.11 details composition of all master mixes used for the MLPA assay.

Reagent	Composition	Volume (ul)
Hybridisation master mix	MLPA buffer	1.5
	SALSA probemix	1.5
Ligase master mix	Ligase-65 buffer A	3
	Ligase-65 buffer B	3
	Nuclease free water	25
	Ligase-65	1
Polymerase master mix	SALSA PCR primers	2
	Nuclease free water	7.5
	SALSA polymerase	0.5

Table 2.11. Composition of MLPA master mixes. Quantities listed refer to volumes required per sample.

The samples were placed in a thermocycler and heated to 98ºC for 5 minutes to denature the DNA and then cooled to 25ºC. Once cooled, 3 ul of hybridisation master mix was added to each sample. The thermocycler was heated to 95ºC for 1 minute and then maintained at 60ºC for 16 hours to allow probe hybridisation to target DNA.

A ligase master mix was then prepared as shown in table 2.11. The thermocycler was set to 54ºC and 32ul of ligase master mix was added to each sample. The samples were held at 54ºC for 15 minutes to permit the ligation reaction, then heated to 98ºC for 5 minutes and finally cooled to 20ºC.

A polymerase master mix was then prepared as shown in table 2.11. Once the samples were cooled to 20ºC, 10ul of polymerase master mix was added to each sample and a PCR amplification of ligated probes was carried out for 35 cycles of 95ºC for 5 minutes, 60ºC for 30 seconds, 72ºC for 60 seconds, followed by 72ºC for 20 minutes and finally cooling to 15ºC for a further 3-4 hours. Samples were then refrigerated and sent to DBS Genomics for PCR fragment separation by capillary electrophoresis using the Applied Biosystems 3730 DNA Analyser and copy number analysis was performed using GeneMarker software. Peaks were normalised against control samples. Deletions and gains were identified when probe ratios for individual exons were <0.75 or >1.3 respectively and a CNA was called when at least 2 consecutive probes showed a consistent pattern (either deletion or gain).

2.7 Customised targeted NGS for validation of novel abnormalities

A customised NGS approach was developed to validate novel lesions. First, all genes that recurrently contained a copy number breakpoint were examined to create a list of potential target lesions for validation. Novel and highly recurrent genes or those with potential functional relevance in leukaemia were primarily targeted. To facilitate discovery of novel abnormalities, recurrent lesions with even ambiguous or subclonal copy number loss by SNP array were considered. Proximal and distal breakpoints of all deletions for validation were recorded. For each gene or locus of interest, a common region up to 200 kb was identified that spanned at least 1 breakpoint in all samples bearing the specific deletion. These regions of interest were selected to ensure at least one breakpoint (proximal or distal) of the deleted segment was covered for each of the affected genes in the relevant cases. An overview of the principles of the targeted NGS design and experimental process is shown in figure 2.3.

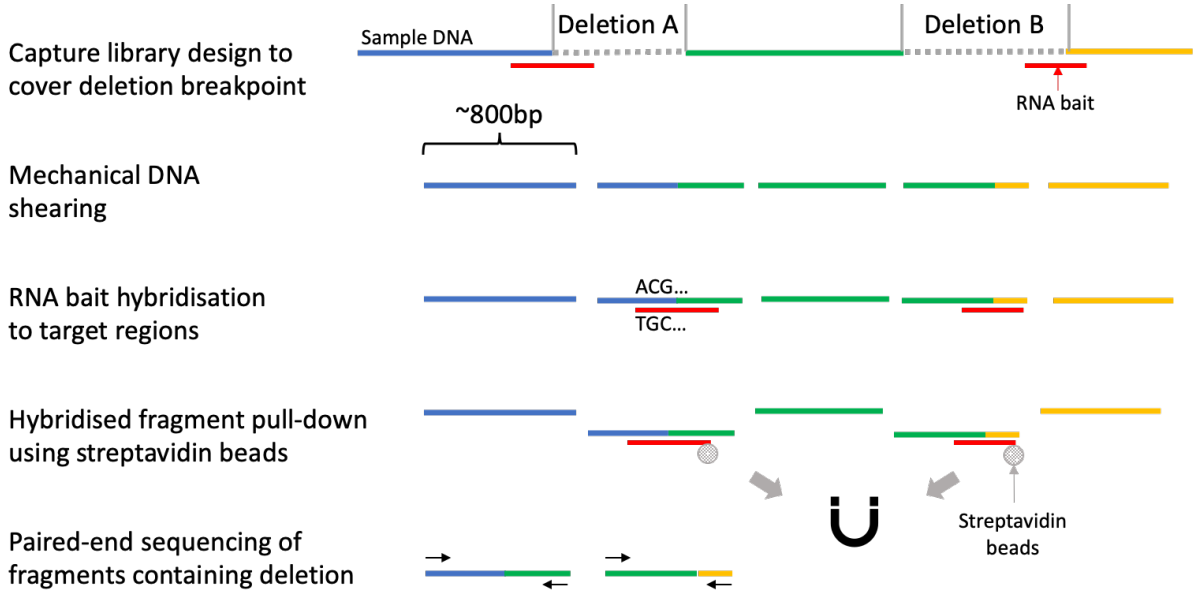


Figure 2.3. Simplified overview illustrating principle of SureSelect XT2 workflow to validate genomic deletions. Breakpoint regions of interest are first identified (either proximal or distal breakpoints). Complementary RNA baits are designed, and DNA is sheared to 800 bp fragments. During library prep, RNA baits are hybridised to their complementary regions thereby capturing fragments containing the breakpoint of interest. These fragments are selected using magnetic beads and finally sequenced using paired-end chemistry.

2.7.1 SureSelect XT2 target design

Once a common breakpoint region had been identified for each deletion of interest, the genomic co-ordinates were loaded and viewed in the SureDesign web portal (https://earray.chem.agilent.com/suredesign/) to design RNA baits complementary to the regions of interest. The coverage of all designs was viewed in the University of California Santa Cruz (UCSC) genome browser to ensure highly repetitive regions of the genome were avoided. The relevant genomic co-ordinates were then submitted to Agilent to manufacture the customised SureSelect XT2 capture library consisting of RNA baits covering the breakpoint regions of interest, together with the coding exons of several genes frequently mutated in BCP-ALL. The capture library would hybridise to and capture the regions of interest, thereby permitting subsequent sequencing of the breakpoint regions (figure 2.3). All target regions used to design the capture library can be found in supplementary tables 3 and 4.

The method is based on the principle that a deletion results in the juxtaposition of normally genomically distant loci. By capturing and then sequencing DNA fragments containing a deletion breakpoint, sequencing reads that overlap either the proximal or distal breakpoint map to non-consecutive genomic co-ordinates, resulting in a structural variant. In the case of a simple deletion, reads map to co-ordinates immediately proximal and distal to the deleted segment. Where the deletion has occurred as part of an unbalanced translocation, reads spanning the breakpoint map to different chromosomes.

To provide sufficient breadth of coverage to span both sides of breakpoints and successfully map these to the reference genome, DNA shearing was performed to yield fragments of approximately 800bp for library prep, rather than the standard 150-200bp fragments recommended by the manufacturer.

2.7.2 SureSelect XT2 library prep

Following manufacture of the customised capture library, a SureSelect XT2 library prep was then performed on all samples with at least 1ug of DNA following WGA (section 2.4.4). All incubation steps were performed in a Thermal Cycler ensuring the heated lid was kept open unless stated otherwise. Samples were first made up to 100 ul with

TE buffer and the DNA was sheared by ultrasonication using the Bioruptor Pico sonication system inputting shearing settings as follow – time on: 7 seconds; time off: 90 seconds; number of cycles: 4.

These settings aimed to shear the sample DNA into 800-1000 bp fragments, and had previously been trialled and validated by Dr Sarra Ryan and James Murray. Sample quality and fragment size were then assessed on the 2100 Bioanalyzer (Agilent) using a Bioanalyzer DNA 1000 chip, to ensure that a fragment peak was visible around 800-1000 bp (section 2.4.5.4).

Following this, an end repair reaction was performed using 50ul of sheared DNA mixed with 40ul of SureSelect End Repair Enzyme Mix and 10ul of SureSelect End Repair Nucleotide Mix. The 100ul mixture was incubated for 30 minutes at 20ºC. The DNA was then purified by mixing the 100ul sample with 180ul of magnetic AMPure XP beads and performing repeated washes with freshly prepared 70% ethanol whilst the samples were held in a Dynamag magnetic separation rack. Residual ethanol was carefully removed and finally allowed to evaporate at 37ºC. The purified DNA was then eluted from the beads in 22ul nuclease free water and the beads were discarded.

Next, dA tailing the 3’ end of the purified DNA fragments was performed by adding 20 ul of dA-Tailing Master Mix and incubating the sample at 37ºC for 30 minutes. The DNA was then ready for ligation to the pre-capture indexing adaptors.

For this, samples were held on ice and 5 ul of SureSelect Ligation Master Mix followed by 5 ul of Pre-capture Indexed Adaptor Solution was added to each sample. At this point special care was taken to ensure a different well of the Pre-capture Indexed Adaptor Solution was added to each sample without cross-contamination as this contained the unique barcode sequence to successfully de-multiplex the pooled samples during sequencing. The ligation reaction was performed by incubating the samples at 20ºC for 15 minutes. A further bead purification was performed using AMPure XP beads as described previously. Half of the indexed library was then amplified by PCR using a 25 ul of Herculase II PCR Master Mix and 1 ul of XT2 primer mix for each sample and the Thermal Cycler program described in table 2.12.

Segment	Number of cycles	Temperature	Time
1	1	98ºC	2 mins
2	5	98ºC	30 secs
		60ºC	60 secs[1]
		72ºC	2 mins[2]
3	1	72ºC	10 mins
4	1	4ºC	Hold

Table 2.12. Pre-capture PCR thermal cycler program. Annealing[1] and elongation times[2] were extended compared to manufacturer's protocol to optimise amplification of longer DNA fragments.

The amplified library was then purified using magnetic AMPure XP beads as described previously. Here, the ratio of beads: DNA was reduced to 0.7, when compared to the manufacturer's protocol, to optimise bead-binding to the longer DNA fragments in the library. Following this, the DNA fragment size and concentration were measured on the 2100 Bioanalyzer (Agilent), aiming for an average fragment size >650-700 bp.

Libraries were then pooled together by combining 15 samples, with 100ng of each sample. Using a vacuum concentrator held at 45ºC, the pools were then concentrated to a volume of 7 ul. To each of these 7 ul indexed DNA pool, 9 ul of SureSelect XT2 blocking mix was added and the samples were heated to 95ºC for 5 minutes followed by 65ºC for at least a further 5 minutes.

Next, the custom-designed capture library was prepared for hybridisation to the pooled DNA samples. Concentrated SureSelect RNase block was first diluted 1:9 with nuclease free water and 5 ul of this diluted RNase block was combined with 2 ul of the capture library. Subsequently, 37 ul of SureSelect XT2 hybridisation buffer was mixed with the RNase block/capture library solution to complete preparation of the capture library for hybridisation to the pooled DNA.

With the pooled DNA samples kept at 65ºC, 44 ul of the capture library mixture was added to these and pipette mixed rapidly to minimise evaporation, before sealing the sample wells and then incubating the DNA/capture library mixtures at 65ºC for 24 hours. During this step, the customised RNA baits in the capture library hybridised to their complementary sequences in the pooled genomic DNA samples.

The hybridised DNA fragments were then captured using streptavidin beads. Dynabeads MyOne Streptavidin T1 coated beads were first resuspended using a

vortex mixer and washed in SureSelect XT2 Binding Buffer. For each pool of DNA samples, 200ul of Binding Buffer was first mixed with 50 ul of resuspended beads and a magnetic separation device was used to isolate the beads and discard the buffer. The procedure was repeated for a total of 3 washes and the beads were re-suspended in a fresh 200 ul volume of SureSelect XT2 Binding Buffer. Meanwhile, 200 ul of SureSelect XT2 Wash 2 was heated to 65ºC for each pooled library. The pooled DNA libraries were then added to the washed Streptavidin coated beads and the samples were continually mixed on a Nutator mixer for 30 minutes at room temperature. The samples were then placed in a magnetic separation device to isolate the streptavidin-coated beads, which were now bound to the DNA fragments that had hybridised to the capture library. The supernatant was then removed and the beads were re-suspended in SureSelect XT2 Wash 1. The samples were again placed in a magnetic separation device and the cleared supernatant was discarded. The beads were then washed with the pre-warmed SureSelect XT2 Wash 2. The beads were re-suspended in 200ul SureSelect XT2 Wash 2 and incubated at 65ºC for 5 minutes before placing the samples in the magnetic separation device and discarding the supernatant. The procedure was repeated for a total of 6 washes ensuring that samples were kept on a heating block at 65ºC when not in the magnetic separation device. The beads were then fully re-suspended in 30ul nuclease-free water.

The bead-bound captured library pools were then ready for PCR amplification. Half of each pool (15 ul) was used and the remainder stored at -20ºC. For each pooled library, a PCR master mix was prepared consisting of 50 ul of LongAmp polymerase (New England Biolabs), 2 ul XT2 primer mix and 33 ul nuclease free water. This was mixed with 15 ul of the bead-bound library pool, placed in a thermal cycler and the program was started as detailed (table 2.13).

Segment	Number of cycles	Temperature	Time
1	1	95ºC	3 mins
2	8	98ºC	20 secs
		60ºC	15 secs
		65ºC	2 mins
3	1	65ºC	10 mins
4	1	4ºC	Hold

Table 2.13. Post-capture PCR thermal cycler program. Settings were modified compared to manufacturer's protocol due to use of Longamp polymerase instead of Herculase II polymerase (discussed in chapter 4, section 4.3.4).

A final magnetic bead purification was then performed using AMPure XP magnetic beads and freshly prepared 70% ethanol at room temperature as previously described. Again, the ratio of beads: sample was reduced to 0.7 to optimise bead-binding to the longer DNA fragments in the library. Each pooled library was finally eluted in 30 ul nuclease free water and the quality of the DNA for sequencing was assessed on the Bioanalyzer using the 2100 Bioanalyzer High Sensitivity DNA Assay (section 2.4.5.4). The 2 pooled libraries were then sequenced on two runs on the Illumina NextSeq 550 at the Core Genomics Facility, Newcastle University.

2.7.3 Analysis of sequencing data

Each pooled library was sequenced using a mid-output kit on the Illumina NextSeq 550 with 150bp paired end reads. Initial processing of FASTQ files was performed by the Bioinformatic Support Unit, Newcastle University. BAM files were deduplicated and re-aligned to the reference genome (hg19/GRCh37). The kit had been designed to cover the proximal and/or distal breakpoint regions of the deletions to be validated. These regions were then directly examined in the Integrative Genomics Viewer (IGV) by loading the BAM files of affected cases. In IGV, reads were coloured according to insert size. The sequencing library had been prepared for an intended insert size of ~800bp as described above. As such mate pairs were anticipated to map to regions <1kb apart. Genomic deletions result in the juxtaposition of normally distant loci, resulting in mate pairs mapping to regions further apart than the anticipated insert size. Where paired reads were >10kb apart, the reads were flagged in IGV and a breakpoint was suspected when numerous flagged reads were all present around the same genomic co-ordinate (example figure in chapter 4, figure 4.6). If the mate reads then mapped to the same distant genomic locus, a deletion was confirmed. Where the mate reads mapped to a consistent region on a different chromosome, the findings were consistent with a chromosomal translocation. Using SNP array derived deletion breakpoints, a targeted analysis was initially performed in IGV to either validate or refute the deletion under investigation. Where the deletion could not be validated in the vicinity of the SNP array derived breakpoints, the remainder of the gene was screened for structural variants. Additionally, specific genes had been included in the capture library due to their involvement in recurrent gene fusions (*ABL1*, *ABL2*, *PDGFRB*, *CSF1R*, *JAK2*,

TCF3). The entire sequence of each of these genes was screened for structural variants in IGV for all cases to identify or validate fusion genes (e.g. *BCR-ABL1*).

2.8 Identifying and tracking variants through treatment

Patients were recruited prospectively from the Newcastle-upon-Tyne Hospitals NHS Foundation Trust (section 2.3) and diagnostic bone marrow and germline DNA were isolated as described (sections 2.4.1, 2.4.2 and 2.4.3).

2.8.1 Exome sequencing

2.8.1.1 Library prep and sequencing

Samples for exome sequencing consisted of diagnostic and matched germline DNA for each patient. QC checks were performed to ensure each sample had at least 150ng dsDNA determined by Quant-iT picogreen, with DIN >7 and A260/A280 ratio of 1.75-2.04 (section 2.4.5). Library prep and sequencing were performed at the Core Genomics Facility, Newcastle University.

Library prep was performed using the Twist Human Core Exome kit. Libraries were sequenced on the Illumina NovoSeq using 100 bp paired end reads, generating on average 90x raw read depth per sample. To achieve a higher depth of sequencing for diagnostic samples with the aim of detecting low VAF abnormalities, these were put through two preparations, thereby generating an average 180x read depth. Data were returned in the form of FASTQ files.

2.8.1.2 Analysis of exome sequencing data

Raw sequencing data in the form of FASTQ files were initially processed by Dr Matthew Bashton and/or the Bioinformatics Support Unit, Newcastle University. Using the diagnostic and matched germline exomes for each case, established somatic variant calling pipelines (GATK Mutect and Mutect 2) were used to produce a list of somatic variants. Further variant annotations using Ensembl Variant Effect Predictor (VEP)

(McLaren et al., 2016) were generated to identify population level variant allele frequencies, together with reports from dbSNP and Catalogue of Somatic Mutations in Cancer (COSMIC) databases as well as the SIFT and PolyPhen variant prediction tools (Sherry et al., 2001, Tate et al., 2019, Sim et al., 2012). The resultant VEP file was used for subsequent analyses.

To generate a final list of the most likely candidate pathogenic mutations, any variants present in >1% of the general population were first excluded (MAX_AF column of VEP file). Next, the variant consequence column of the VEP file was examined and synonymous variants, intron variants, downstream gene variants, upstream gene variants, 5' UTR and 3' UTR variants were all filtered out. Subsequently, variants that were classed as both tolerated in SIFT and benign in PolyPhen were also removed. This created a final list of candidate pathogenic mutations in each diagnostic bone marrow specimen. These were examined in detail to select specific variants to track in subsequent follow up bone marrow samples.

2.8.2 Selecting mutations to design SureSelect XT HS2 kit

Single nucleotide variants (SNVs) generated through Mutect were identified and filtered as described above. Specific SNVs within each sample were then selected to design a customised capture library to target and sequence these regions to high depth in follow up samples.

A limited number of SNVs per patient (5-6) were selected for this study. Variants in genes involved in key pathways known to be deregulated in ALL were first identified and selected. Based on COSMIC data, SNVs reported as somatic variants in other cancer-types were also chosen, along with variants that appeared to be recurrently seen in the patient cohort. To identify potential low frequency variants associated with clonal haematopoiesis, all exons of *DNMT3A* and *TET2* were also included, together with exons 11 and 12 of *ASXL1*. Similarly, due to its prevalence in relapsed disease, the coding regions of *TP53* were also added to the design.

Using the SureDesign web portal (https://earray.chem.agilent.com/suredesign/), the selected SNV co-ordinates (chapter 6, table 6.5) were uploaded together with selected regions of *DNMT3A*, *TET2*, *ASXL1, TP53, KRAS, NRAS* and *PTPN11* due to their prevalence in clonal haematopoiesis or relapsed disease.

To identify variants at potentially very low VAF in remission samples, the SureSelect XT HS2 kit was chosen. This incorporates unique molecular identifiers (UMIs). Normally, the sensitivity of sequencing techniques to the detection of low VAF variants is hampered by PCR duplication. During library prep, PCR amplification of the DNA adaptor-ligated fragments is performed, generating multiple identical copies of the original DNA molecule. When more than one copy of the starting unique DNA molecule hybridises to the flow cell, this is sequenced multiple times, which generates duplicate reads. This biases sequencing coverage and if unaccounted for, can greatly exaggerate a potential artefactual variant that has arisen from sequencing error (Ebbert et al., 2016). As there is no way of telling whether identical reads have occurred by chance from 2 different DNA molecules or as a result of PCR amplification of a single unique DNA molecule, most bioinformatics tools opt to remove them from downstream analysis (de-duplication) (Ebbert et al., 2016). This limits the detection of mutations in typical NGS experiments to a VAF plateau of ~1% even with very high read depth. However, the use of UMIs is able to resolve this issue. These are short DNA sequences, which ligate to the DNA fragments in the library prior to PCR amplification. Identical reads that have occurred through PCR amplification can therefore be identified as they will possess the same UMIs, as opposed to those that have arisen from different starting DNA molecules. This permits the detection of mutations with a VAF <0.1%.

The target designs were submitted to Agilent to manufacture customised RNA baits to capture the single nucleotide sequences in an attempt to identify the variants present at diagnosis in the follow up samples taken on treatment. Unfortunately, due to the COVID-19 pandemic, it was not possible to subsequently proceed to library prep at this stage (discussed in chapter 6, section 6.4.7 and chapter 7, section 7.5).

Chapter 3. Genetic characterisation of older adults with B-other ALL

3.1 Introduction

Primary chromosomal abnormalities are one of the hallmarks of ALL and greatly influence treatment decisions and prognosis. Well defined genetic subgroups include large scale ploidy shifts resulting in high hyperdiploidy (51-67 chromosomes), low hypodiploidy (30-39 and/or 60-78 chromosomes) or near haploidy (23-29 and/or 46-58 chromosomes) and chromosomal translocations resulting in oncogenic gene fusions, specifically *BCR-ABL1*, *ETV6-RUNX1*, *TCF3-PBX1*, *TCF3-HLF* or *KMT2A*-v (chapter 1, section 1.8). The relative frequencies of these recurrent primary abnormalities have been extensively studied in children and young adults but are poorly described in older patients (Schultz et al., 2007, Gökbuget, 2013).

Moreover, a proportion of patients do not harbour a cytogenetically visible disease-defining lesion. In 2009, two publications described a subgroup of patients with a gene expression profile similar to Philadelphia positive ALL (Philadelphia-like or *BCR-ABL1*-like ALL) (Mullighan et al., 2009b, Den Boer et al., 2009). Subsequently, whole genome and transcriptome sequencing studies have identified a variety of kinase-activating gene rearrangements responsible for leukaemic transformation, detectable in up to 91% of such cases (Roberts et al., 2012, Roberts et al., 2014). The majority of patients studied were children and young adults, although a subsequent study did include some older adults (Roberts et al., 2017).

Approximately 50% of Ph-like patients have high *CRLF2* expression on their leukaemic blasts, which is driven by cytogenetically-cryptic *IGH-CRLF2* or *P2RY8-CRLF2* rearrangements. Other recurrent gene rearrangements include ABL-class fusions (affecting *ABL1*, *ABL2*, *PDGFRB* or *CSF1R*) in 9-13% of cases and the JAK-STAT pathway activating rearrangements of *JAK2* or *EPOR* in 7-10% and 3-6% of patients respectively. Importantly, age-related differences in the distribution of these

abnormalities have also been noted, with ABL-class fusions seen more commonly in childhood Ph-like ALL patients and *JAK2* or *EPOR* rearrangements more frequent in adult cases.

More recently, recurrent oncogenic fusions affecting *ZNF384* and *MEF2D* have also been identified (Hirabayashi et al., 2017, Gu et al., 2016a), and result in distinct gene expression profiles, separate from the Ph-like signature. These rearrangements are reported in approximately 2-6% of BCP-ALL cases, and have also predominantly been characterised in paediatric cohorts.

3.2 Aims and objectives

The principle aim of this work package was to define the primary genetic abnormalities in older adults with ALL. To achieve this, the following objectives needed to be addressed.

1. Discern a large clinical trial cohort of older adults with ALL
2. Categorise primary chromosomal abnormalities based on diagnostic analyses performed in regional cytogenetic centres
3. Identify all BCP-ALL patients lacking a primary chromosomal abnormality (B-other ALL)
4. Perform extended fluorescence *in situ* hybridisation (FISH) studies to identify cytogenetically cryptic rearrangements in B-other patients

3.3 Methods

The Leukaemia Research Cytogenetic Group (LRCG) database was searched for adults with ALL aged ≥60 years at diagnosis enrolled in the UKALL14 or UKALL60+ clinical trials. All karyotypes and primary chromosomal abnormalities that had been identified by the diagnostic cytogenetic laboratories were first recorded and their individual frequencies calculated. Next, all cases lacking a primary chromosomal

abnormality, hereafter termed B-other ALL, were identified. Specifically, this included cases with normal, failed or complex karyotypes or those with non-specific chromosomal abnormalities. Karyotypes required the analysis of at least 20 normal metaphase cells to be classed as normal. Failed karyotypes were defined as those where no clonal abnormality was detected and fewer than 20 metaphases were present (Medeiros et al., 2014). Additionally, failed karyotypes were required to have had *BCR-ABL1* and *KMT2A* fusions excluded by fluorescence *in situ* hybridisation (FISH) to be classed as B-other ALL. Complex karyotypes were defined by the presence of at least 5 unrelated chromosomal abnormalities in the absence of another primary abnormality (Moorman et al., 2007). All B-other cases with available fixed cells samples were specifically identified for the investigation of cytogenetically cryptic abnormalities.

FISH permits rapid identification of *ABL1*, *ABL2*, *PDGFRB/CSF1R*, *JAK2*, *CRLF2*, *IGH@* and *ZNF384* rearrangements (chapter 2, 2.5). Dual colour breakapart probes consist of red and green fluorochromes that flank either side of a gene of interest. Due to their genomic proximity, the fluorochromes generate two fused signals under fluorescence microscopy when neither copy of the gene is disrupted. However, a gene rearrangement results in a variant signal pattern, based on the type of structural variation that has occurred. Typically, *ABL1*, *ABL2*, *JAK2*, *IGH@* and *ZNF384* rearrangements occur through chromosomal translocations. In the simplest scenario, when visualised by fluorescence microscopy, this produces clear separation in the red and green signals of the affected allele, as the telomeric portion of the gene of interest has translocated to a different chromosome. In comparison, *PDGFRB* rearrangements most commonly occur through an interstitial deletion on 5q33, resulting in *EBF1-PDGFRB* fusion. In this situation, the telomeric (green) signal of the affected *PDGFRB* allele is lost, leaving a single fused signal signifying the undisrupted allele and a single red signal representing deletion of the telomeric portion of the affected allele, and subsequent juxtaposition of exon 11 of *PDGFRB* and exon 15 of *EBF1* (Schwab et al., 2016). *EBF1-PDGFRB* fusion is then confirmed through reciprocal demonstration of deletion of the centromeric signal of an *EBF1* break apart probe. *CRLF2* rearrangements can occur either through chromosomal translocations or interstitial deletions of PAR1, and can therefore generate either variant signal pattern. *MEF2D* rearrangements can be more challenging to detect by FISH. These most commonly occur as a result of a small interstitial insertion at 1q21.22-22, resulting in *MEF2D-BCL9* fusion. However, due to the proximity of these two genes, only a subtle

separation of the red and green signals is seen when this type of rearrangement has occurred.

In comparison *EPOR* rearrangements most commonly involve insertion of a truncated *EPOR* locus to a region distal to the *IGH@* enhancer, which cannot be detected by FISH (Iacobucci et al., 2016).

Dual colour break-apart probes were therefore used to detect rearrangements in *ABL1*, *ABL2*, *PDGFRB/CSF1R*, *CRLF2*, *IGH@*, *ZNF384* and *MEF2D* (chapter 2, section 2.5.3). Rearrangements in these genes are usually considered to be primary abnormalities in the majority of cases, hence there is little overlap between these events. With the notable exception of *IGH@*, cases were therefore not screened for further abnormalities in these genes if one had already been detected. In contrast to the other gene rearrangements described, *IGH@* translocations result in over-expression of the partner gene, which is responsible for the leukaemic properties. Hence, the biological significance of these events usually rests with the translocation partner (most commonly *CRLF2*), so these were not considered mutually exclusive of the other gene rearrangements.

3.4 Results

3.4.1 Established primary chromosomal abnormalities

A total of 210 patients were identified from UKALL14 (n=94) and UKALL60+ (n=116). Median patient age was 64 years (range 60-83) and 24% (n=50) were over 70 years old at diagnosis. Male: female ratio was 1:1. Overall, 200 patients (95%) had cytogenetic results available from diagnosis, or had T-cell disease. Baseline patient demographic and cytogenetic results are summarised in table 3.1.

	BCR-ABL1	*TCF3-PBX1*	*KMT2A*	HeH	HoTr	T-cell	B-other	No data	Total
N (%)	55 (28%)	3 (2%)	12 (6%)	2 (1%)	28 (14%)	11 (6%)	89 (45%)	10 (5%)	210 (100%)
Male	40%	33%	33%	100%	43%	64%	60%	40%	50%
Median age (yrs)	64	64	64	64	64	64	65	63	64

Table 3.1. Genetic subgroups of 200 adults aged ≥60 years recruited to UKALL14 or UKALL60+. *KMT2A* included t(4;11)(q21;q23)/*KMT2A-AFF1* (n=8); t(11;19)(q23;p13.3)/*KMT2A/MLLT1* (n=2) and t(1;11)(p32;q23) (n=1);
***KMT2A*: *KMT2A* fusion with any partner; HeH: high hyperdiploidy (51-67 chromosomes); HoTr: low hypodiploidy or near triploidy (30-39 or 60-78 chromosomes); T-cell: T-cell ALL; B-other: BCP-ALL with no primary chromosomal abnormality identified.**

The distribution of individual genetic abnormalities by age is shown in figure 3.1. Trial recruitment decreased with age, although this may have been confounded by the fact that patients aged 60-64 years could be recruited to UKALL14 or UKALL60+ so it is possible that more centres had a trial available for these patients. This may partly explain why patients aged 60-64 years are over-represented in the cohort.

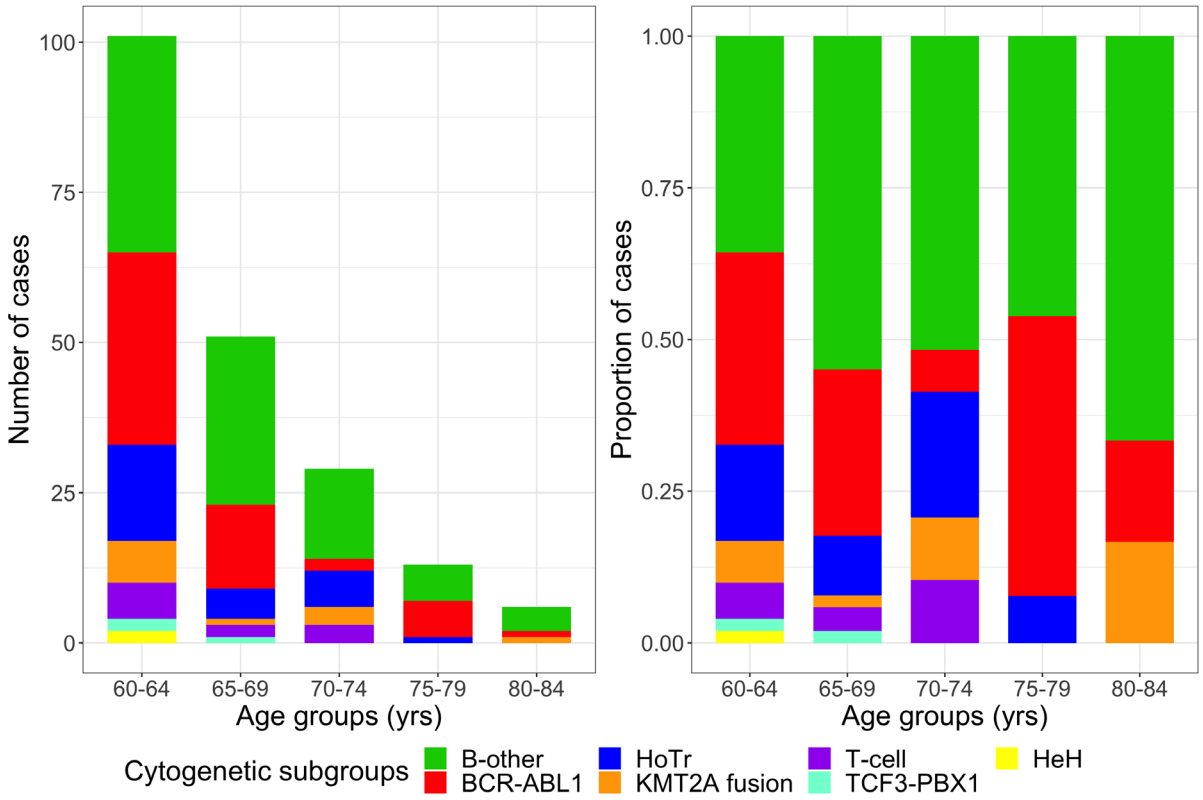


Figure 3.1. Distribution of primary chromosomal abnormalities by age groups across 200 adults aged ≥60 years (10 cases lacking sufficient cytogenetic data excluded). Left panel represents cytogenetic subgroups across age groups and right panel illustrates proportion of cases within each individual age groups.
HoTr: low hypodiploidy/near triploidy; HeH: high hyperdiploidy; T-cell: T-cell ALL; B-other: BCP-ALL with no primary chromosomal abnormality identified.

Overall, no association was seen between the individual 5 year age groups and specific cytogenetic subgroups (Pearson's Chi-squared, X^2 = 23.894, p = 0.47). To further investigate any association between *BCR-ABL1* positivity and age, the cohort was split into a younger half (median 63 years of age, range 60-64) and an older half (median 69 years of age, range 65-83). Of BCP-ALL cases with sufficient cytogenetic data (n=189), these contained 34% (32/95) and 24% (23/94) of *BCR-ABL1* positive cases respectively. Using Fisher Exact Test, there was no significant difference in the proportion of *BCR-ABL1* positive cases between these two groups (p=0.2).

3.4.2 Gene rearrangements in B-other ALL patients

B-other ALL patients were defined as those with normal or failed karyotypes, or other non-recurrent chromosomal abnormalities, including complex karyotypes, in the absence of any known risk-stratifying primary cytogenetic abnormality. Cytogenetic results for B-other patients is shown in figure 3.2.

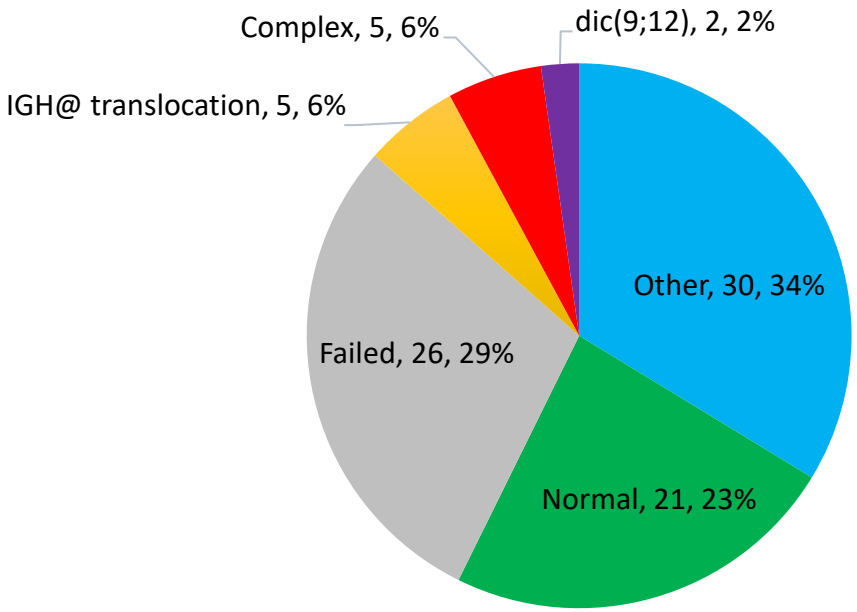


Figure 3.2. Cytogenetic details of 89 B-other ALL patients with number of cases and percentage of all B-other displayed.
Failed: failed karyotype (no cytogenetic abnormality and <20 normal metaphases available); Normal: normal karyotype (≥20 normal metaphases); dic(9;12): dicentric translocation between chromosomes 9 and 12; Other: non-recurrent or non-specific chromosomal abnormality (e.g. del(9p)) seen; Complex: ≥5 unrelated chromosomal abnormalities present on karyotype.

In total, fixed cell samples were available for 73% (65/89) of the B-other patients.

3.4.2.1 *CRLF2* rearrangements

CRLF2 break apart FISH testing was successfully performed in 74% (48/65) of the B-other samples with fixed cells. FISH was attempted in a further three cases but did not yield interpretable results and in the remaining cases, insufficient numbers of fixed cells were seen to perform FISH studies (n=12) or another primary gene rearrangement had already been identified (n=2).

CRLF2 rearrangements were identified in 17% (8/48) of successfully screened B-other ALL patients. Patient details and FISH results are displayed in table 3.2. The principal *CRLF2* rearrangement partners were *IGH@* on 14q32.33 (n=5) or *P2RY8* (n=2) on Xp22.33/Yp11.3. Importantly, these two rearrangements result in distinct FISH signal patterns. *IGH-CRLF2* occurs following a chromosomal translocation and both *IGH@* and *CRLF2* show a typical rearrangement pattern consisting of a single fused signal and split red and green signals. In comparison, *P2RY8-CRLF2* occurs following an interstitial deletion on Xp22.33/Yp11.3, resulting in a lost red signal (telomeric to *CRLF2*) on FISH.

Patient ID	Age	Sex	Karyotype	Partner gene	Abn. cells
25130	62	F	46,XX[20]	*IGH@*	12%
25246	64	M	46,XY,der(19)t(1;19)(q12;p13.3)[2]/ 46,idem,t(5;18)(q33:q23)[8]/46,XY[10]	*P2RY8*	24%
25371	60	F	46,XX[20]	*IGH@*	76%
28039*	76	M	45,XY,-7,-15,-17,+mar[10]/46,XY[1]	*P2RY8*	57%
28235	65	F	46,XX,der(19)t(1;19)(q12;q13)[4]/ 46,XX[6]	*IGH@*	30%
30102	67	F	46,XX,del(9)(p1p2)[9]/46,XX[1]	*IGH@*	62%
30297	64	F	46,XX,t(X;9)(p22;p13),der(15)t(1;15) (q21;p13)[cp8]/46,XX[4]	Unknown†	30%
30299	74	F	Not known	*IGH@*	17%

Table 3.2. *CRLF2* rearrangements. *Patient 28039 also had a separate *IGH@* rearrangement but the signal pattern was consistent with *P2RY8-CRLF2* fusion. †Patient 30297 had a *CRLF2* rearrangement pattern consistent with a chromosomal translocation (see karyotype and FISH signal pattern). However, *IGH@* FISH was normal. Based on the karyotype, *PAX5* was also tested but found to be normal.
M: male; F: female; Abn. cells: percentage of nuclei showing gene rearrangement on FISH slide.

IGH@ was the most common *CRLF2* rearrangement partner, present in 63% (5/8) of cases. In one case, the *CRLF2* rearrangement partner could not be identified, although the karyotype suggested this was located on 9p13.

3.4.2.2 *IGH@* rearrangements

Guided by karyotype abnormalities, regional cytogenetic centres had already performed *IGH@* breakapart FISH in 7 of the B-other ALL patients, and rearrangements had been detected in 5 cases (figure 3.2).

Of the remaining cases with available fixed cells, *IGH@* breakapart FISH was attempted in 77% (49/64) and successful in 46 cases. In total, *IGH@* translocations were present in 26% (14/53) of patient samples tested. Median patient age was 65 years and 57% (8/14) were male. Five cases accounted for patients with *IGH-CRLF2* translocations as detailed above (table 3.2). The remaining 9 cases involved other rearrangement partners as detailed in table 3.3. Interestingly, the vast majority (8/9) of these *IGH@* rearrangements with non-*CRLF2* partners were seen in male patients.

Patient ID	Age	Sex	Karyotype	Partner gene	Abn. cells
25552	61	M	46,XY[20]	Not tested	24%
25894	63	M	44,XY,-8,-13,der(14)t(8;14)(q11;q32)[13] /44,idem,add(19)(p13.3)[4]/46,XY[3]	*CEBPD**	91%
25907	70	M	47,XY,+X,?t(18;22)(q11;q21)[1]/46,XY[3]	Unknown	93%
27181	65	M	46,XY,inv(14)(q11q32)[2]/46,XY[18]	*CEBPE*	43%
27833	73	F	Failed	*BCL2*	94%
28039†	76	M	45,XY,-7,-15,-17,+mar[10]/46,XY[1]	Unknown	54%
29808	66	M	46,XY,add(2)(q?37),add(8)(q13),add(14) (q32) [10]/46,XY[7]	Unknown	91%
30487	60	M	47,XY,+X,t(14;19)(q32;q13)[8]/46,XY[2]	*CEBPA**	71%
25451†	63	M	46,XY[20]	Unknown	19%

Table 3.3. *IGH@* translocations (*IGH@-CRLF2* cases not shown).
***Partner genes based on karyotype and evidence from literature (Akasaka et al., 2006, Chapiro et al., 2006).**
†These patients had another primary abnormality detected by FISH (tables 3.2 and 3.4) and *IGH@* translocations were therefore likely secondary events. All cases with unknown *IGH@* partner had *CRLF2*, *CEBPA*, *CEBPB* and *CEBPE* partners excluded by FISH.
M: male; F: female; Abn. cells: percentage of nuclei showing gene rearrangement on FISH slide.

3.4.2.3 *ZNF384* rearrangements

ZNF384 FISH was successfully performed in 65% (40/65) of the B-other ALL samples with available fixed cells. *ZNF384* rearrangements were identified in 8% (3/40) of these tested patients (table 3.4). All patients were male, with a median age of 63 years.

Patient ID	Age	Sex	Karyotype	Partner gene	Abn. cells
25235	63	M	Failed	Unknown*	34%
25451	63	M	46,XY[20]	*EP300*	35%
30085	67	M	46,XY[20]	Unknown*	22%

Table 3.4. *ZNF384* rearranged cases. In patients 28235 and 30085, the rearrangement partner was not identified. *FISH for translocations involving *EP300* and *TCF3* were normal in both cases.
M: male; Abn. cells: percentage of nuclei showing gene rearrangement on FISH slide.

3.4.2.4 *MEF2D* rearrangements

FISH using the *MEF2D* breakapart probe was attempted in 60% (39/65) of the B-other cohort with fixed cells available. A single *MEF2D* rearranged case was identified in the cohort (3%, 1/39) (table 3.5).

Patient ID	Age	Sex	Karyotype	Partner gene	Abn. cells
25267	63	F	45~47,XX,+1,dic(1;17)(p32;q25),inc[cp3]	Unknown	24%

Table 3.5. *MEF2D* rearranged case. Partner gene could not be established due to insufficient material.
F: female; Abn. cells: percentage of nuclei showing gene rearrangement on FISH slide.

3.4.2.5 *PDGFRB*/*CSF1R*, *ABL1*, *ABL2*, *JAK2* rearrangements

Across the B-other cohort tested, no variant *ABL1* (0/83), *PDGFRB* (0/56), *JAK2* (0/53) or *ABL2* (0/52) were identified.

ABL1 translocations had either been excluded based on a normal *BCR-ABL1* FISH result from the diagnostic analyses performed in the regional cytogenetic centres (n=71) or by dedicated *ABL1* break apart FISH (n=12) performed as part of the extended FISH screening in this study.

3.4.3 Clinical and demographic characteristics

Median patient age of *CRLF2*, *IGH@* and *ZNF384* rearranged patients was 64.5, 65 and 63 years respectively. The majority of patients with *CRLF2* rearrangements were female (75%, 6/8), whereas 89% (8/9) of non-*CRLF2* *IGH@* and all (3/3) *ZNF384* rearranged patients were male. Further clinical and outcome data are detailed in table 3.6.

Patient ID	Trial	Abnormality	WCC (x10^9/L)	Outcome
25130	UKALL14	*CRLF2-r*	33.6	Died after 1 month
25246	UKALL14	*CRLF2-r*	6.3	Died within 1 month
25371	UKALL14	*CRLF2-r*	47.7	Alive >5 years
28039	UKALL60	*CRLF2-r*	Not known	Died after 9 months
28235	UKALL60	*CRLF2-r*	5.3	Relapsed and died after 2 years
30102	UKALL60	*CRLF2-r*	Not known	Relapsed and died after 5 months
30297	UKALL60	*CRLF2-r*	Not known	Alive >2 years
30299	UKALL60	*CRLF2-r*	Not known	Died after 4 months
25552	UKALL14	*IGH@-r*	2.9	Died after 4 months
25894	UKALL60	*IGH@-r*	0.8	Alive >5 years
25907	UKALL60	*IGH@-r*	2	Died after 1 year
27181	UKALL14	*IGH@-r*	1.2	Died after 3 months
27833	UKALL60	*IGH@-r*	14.5	Died after 2 years
29808	UKALL60	*IGH@-r*	Not known	Relapsed and died after 1 year
30487	UKALL14	*IGH@-r*	11.7	Alive after 1 year
25235	UKALL14	*ZNF384-r*	3.5	Alive >5 years
25451	UKALL14	*ZNF384-r*	34.2	Relapsed and died >5 years
30085	UKALL60	*ZNF384-r*	Not known	Alive >2 years
25267	UKALL14	*MEF2D-r*	1.4	Alive >5 years

Table 3.6. White cell count (WCC) at diagnosis and outcome for all B-other patients with gene rearrangement detected (n=19).
WCC: white cell count; *CRLF2*-r: *CRLF2* rearrangement; *IGH*@-r: *IGH*@ rearrangement (excluding *CRLF2* partner); *ZNF384*-r: *ZNF384* rearrangement; *MEF2D*-r: *MEF2D* rearrangement.

3.4.4 Summary of B-other ALL screening

In total, gene rearrangements were identified in 21% (19/89) of the B-other patient cohort as detailed in figure 3.3, demonstrating genetic landscape following standard and extended genetic analyses.

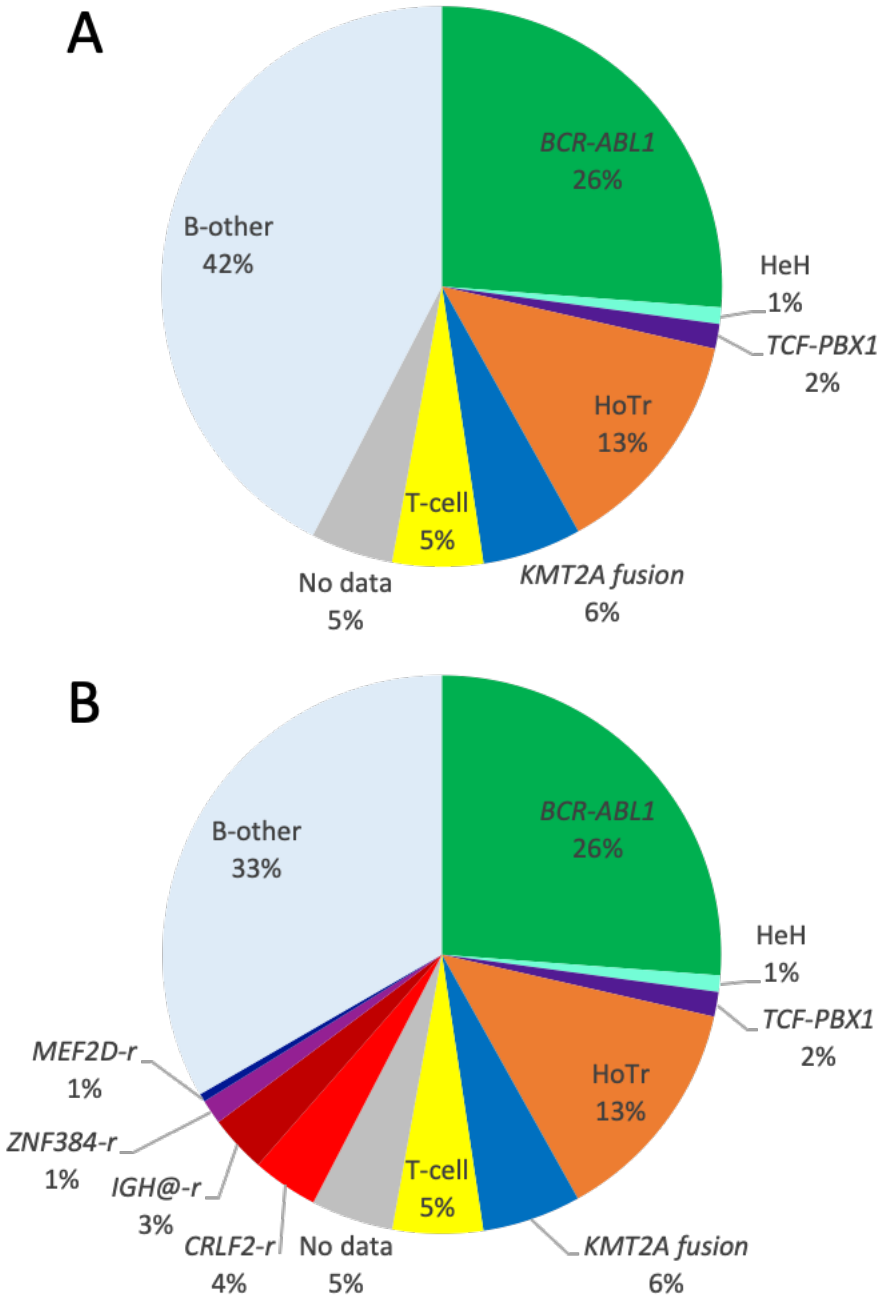


Figure 3.3. Genetic landscape of ALL in 210 older adults enrolled in UKALL14 and UKALL60+ clinical trials. Abnormalities were detected following routine cytogenetic analyses at trial entry (A) and following extended FISH screening for cytogenetically cryptic lesions (B).
HeH: high hyperdiploidy; HoTr: low hypodiploidy/near triploidy; B-other: BCP-ALL with no primary chromosomal abnormality identified; T-cell: T-cell ALL. *CRLF2*-r: *CRLF2* rearrangement; *IGH@*-r: *IGH@* rearrangement (excluding *CRLF2* partner); *ZNF384*-r: *ZNF384* rearrangement; *MEF2D*-r: *MEF2D* rearrangement.

To account for the variation in the number of cases successfully tested for each abnormality, adjusted frequencies were then calculated (table 3.7). Most of the samples without cytogenetic data had diagnostic immunophenotyping performed, permitting the detection of T-ALL as well as RT-PCR for *BCR-ABL1*. The denominator

for *IGH@*, *CRLF2*, *ZNF384* and *MEF2D* rearrangements was the sum of B-other cases successfully screened and all cases belonging to other genetic subgroups, as the presence of one primary genetic abnormality is generally considered sufficient evidence to rule out another, with the exception of occasional *IGH@* translocations (Jeffries et al., 2014).

Abnormality	Cases detected/ Total screened	Adjusted frequency
BCR-ABL1	55/207	27%
HoTr	28/200	14%
KMT2A fusion	12/203	6%
T-ALL	11/206	5%
IGH@-r	9/164	5%
CRLF2-r	8/159	5%
ZNF384-r	3/151	2%
MEF2D-r	1/150	<1%

Table 3.7. Adjusted frequencies of individual genetic abnormalities, based on variation in the number of cases successfully screened for each abnormality.
HoTr: low hypodiploidy/near triploidy; *IGH@*-r: *IGH@* rearrangement; *CRLF2*-r: *CRLF2* rearrangement; *ZNF384*-r: *ZNF384* rearrangement; *MEF2D*-r: *MEF2D* rearrangement.

3.5 Discussion

In total, 200 older adults with ALL were genetically characterised. Consistent with published data, the most prevalent primary abnormality was *BCR-ABL1*, which was seen in 26% of patients (29% of the BCP-ALL patients without missing data). To date, there have been some conflicting data regarding the frequency of *BCR-ABL1* positive disease with advancing age. One of the largest studies of over 2500 BCP-ALL patient samples clearly reported a trajectory of an increasing proportion of *BCR-ABL1* positive cases up to the age of 44 years. Thereafter, this plateaued at 42-44% of all BCP-ALL cases (Burmeister et al., 2008). In contrast, smaller studies demonstrated a persistent increase in the proportion of *BCR-ABL1* positive patients, to ~50% in older adults (Herold et al., 2014, Byun et al., 2017). In the present analysis, there was no evidence of an increasing proportion of *BCR-ABL1* ALL with advancing age in patients 60 years of age and above at diagnosis.

Interestingly, the second most frequent primary cytogenetic abnormality was low hypodiploidy/near triploidy, observed in almost 15% of patients. This contrasts sharply with children and adults aged <60 years, where it is encountered in <1% (Raimondi et al., 2003, Pui et al., 1990) and 2-4% of cases respectively (Moorman et al., 2007, Charrin et al., 2004, Moorman et al., 2010). Low hypodiploidy/near triploidy constitutes a high risk subgroup of BCP-ALL with a very poor prognosis and 5-year survival <20% in adults (Moorman et al., 2007). The overrepresentation of such high-risk genetic subtypes adds to the challenge of managing older adults with ALL as the default recourse to intensified risk-adapted treatment will often not be tolerated in older adults. Indeed, many protocols for older adults are based on an initial assessment of maximum tolerated treatment to optimise depth and duration of remission, meaning the concept of subsequent treatment intensification based on response is unsuitable for managing these patients.

Although they have no single defining genetic lesion, complex karyotypes have also consistently been associated with poor prognosis in both AML and ALL (Byrd et al., 2002, Mrózek et al., 2009). In AML, complex karyotypes are found in 10-12% of de novo diagnoses, with a frequency that correlates with advancing age, rising to around 20% in adults aged 60 years and over at diagnosis (Mrózek, 2008). A large analysis of chromosomal abnormalities in adults aged between 25-65 years identified complex karyotype in 5% of BCP-ALL cases (Moorman et al., 2007). In comparison, in this analysis of older adults with ALL, this entity was seen in just 3% (5/199) of BCP-ALL cases, indicating that unlike in AML, the frequency of complex karyotype does not increase with advancing age.

T-cell ALL was also rare in this unselected cohort of older adults, seen in only 5% of cases. This confirms that after a peak in adolescents and young adults, where it accounts for around 25% of new diagnoses, the incidence of T-cell ALL continues to reduce with advancing age (Ferrando et al., 2002, Marks et al., 2009). As a group, older adults with ALL seem to have one of the lowest frequencies of T-cell disease.

Regarding the more recently described cytogenetically cryptic abnormalities, a large US study has identified Ph-like ALL in 24% in older adults (Roberts et al., 2017), although this was seen in <10% of similar patients in a German study (Herold et al., 2014). Most studies agree the subtype peaks in young adults aged 21-39 years with estimated frequencies ranging between 25-42% of BCP-ALL cases in this age group (Roberts et al., 2017, Jain et al., 2017b, Herold et al., 2014). Interestingly the lower

estimates are usually obtained from Northern European studies, where patients of Hispanic ancestry, who are known to have higher rates of Ph-like ALL, are underrepresented compared with US derived datasets (Jain et al., 2017b). Importantly, Ph-like ALL is defined primarily by its characteristic gene expression profile, which has not been assessed in this study. However, identifying the principal kinase-activating gene rearrangements that produce the Ph-like signature provides a useful comparison between studies and in this large UK cohort of older adults, one of the main findings is the complete absence of ABL-class fusions. This contrasts with ALL in children and younger adults, where ABL-class fusions have been reported in 3-5% and 2-3% of cases respectively (Tasian et al., 2017, Roberts et al., 2014, Roberts et al., 2017).

To date, the US study provides the largest cohort with detailed transcriptomic profiling of older adults with these abnormalities (Roberts et al., 2017). Among 798 patients, 150 older adults (aged 60-86 years) were included. The study only comprised adults with BCP-ALL and excluded those with a primary ploidy shift. When restricted to similar patients, a trend towards a lower frequency of *CRLF2* rearrangements in this UK dataset was observed (table 3.8). This may be related to ethnic differences between UK and US patients as *CRLF2* rearranged ALL is significantly associated with Hispanic/Latino ancestry (Harvey et al., 2010).

Gene rearrangements	This study	US study (Roberts et al., 2017)	p-value
CRLF2	7% (8/118)	14% (21/150)	0.07
PDGFRB	0% (0/126)	<1% (1/150)	1
ABL1	0% (0/153)	0% (0/150)	1
ABL2	0% (0/122)	0% (0/150)	1
JAK2	0% (0/123)	1% (2/150)	0.50
EPOR	Not tested	<1% (1/150)	NA

Table 3.8. Comparison of frequency of kinase-activating rearrangements between UK and US cohorts in older adults. Data from Roberts et al restricted to adults >60 years at diagnosis. Denominators are based on BCP-ALL patients without primary ploidy shift. No statistically significant difference in the frequency of individual rearrangements was seen between patients aged over 60 years in the UK and US cohorts (Fisher's Exact test), although there was a trend towards lower incidence of *CRLF2* abnormalities in UK patients.

IGH@ translocations were the most frequent abnormality identified in B-other ALL patients and were present in 26% (14/53) of tested patients. As discussed previously, *IGH@* translocations are not considered a single genetic subgroup. Proximity to the *IGH@* enhancer overexpresses oncogenes, and as such the *IGH@* partner is

responsible for leukaemic transformation rather than the *IGH@* locus itself. Consistent with larger studies (Russell et al., 2014), the most common *IGH@* translocation partner was *CRLF2*, present in 36% (5/14) of cases. *CEBP* family genes were implicated in a further 21% (3/14) of *IGH@* translocations and a *BCL2* translocation was identified in a single case.

Fewer patients were tested for *ZNF384* and *MEF2D* rearrangements as many samples had already been depleted following all prior FISH experiments. However, *ZNF384* translocations were still detected in 8% of B-other cases (2% of all cases), making this the second most prevalent of the primary B-other rearrangement genes, with a frequency similar to that seen in younger adult and paediatric cohorts (3% and 4% of BCP ALL patients respectively) (Moorman et al., 2019, Hirabayashi et al., 2017). Interestingly, these patients seemed to have a more favourable outcome compared to their B-other counterparts. Although the number of patients is small, two are still alive at 2 and 5 years of follow up respectively and the patient who has died relapsed more than 7 years after initial diagnosis. In comparison, only 2/8 *CRLF2* rearranged patients and 2/9 *IGH@* rearranged (excluding *IGH@-CRLF2*) patients are still alive.

The percentage of abnormal cells identified by FISH varied considerably between samples. The gene rearrangements screened are typically considered to be primary leukaemogenic events, with the exception of *IGH@* translocations which sometimes co-occur with other abnormalities (Jeffries et al., 2014). It is therefore more likely that lower percentages of rearranged cells represented normal cell contamination rather than subclonal populations, although this could not be confirmed.

Overall, 22% (19/87) of older adults with B-other ALL had a gene rearrangement identified. A significant proportion of B-other patients therefore still remain uncharacterised following cytogenetic and FISH studies. The targeted nature of FISH experiments only permits very limited discovery of other relevant “off-target” abnormalities and it is possible that novel gene fusions are present in this cohort of older adults with B-other ALL. Additionally, the primary abnormalities may be smaller genomic lesions such as mutations and microdeletions, which are not captured by cytogenetic analyses and techniques with a much higher resolution may be required to elucidate this further.

Chapter 4. Copy number abnormalities in older adults with ALL

4.1 Introduction

Somatic abnormalities are randomly acquired during cell division. In cancer, these can range from a single base substitution to aberrations affecting entire chromosomes. Copy number abnormalities (CNAs) refer to the loss or gain of genetic material during oncogenesis, and constitutes one of the most frequently encountered classes of somatic variation in cancerous cells. Not all CNAs produce a functional impact on the cancerous cell. Driver lesions are those that are beneficial to cell survival and malignant transformation, and are positively acquired during Darwinian natural selection. In comparison, passenger abnormalities arise non-specifically during the malignant process, and are either neutral or only weakly deleterious (Mermel et al., 2011). Due to their biological importance, when a cohort of cancer samples is examined, a specific driver abnormality will be encountered more frequently than a specific passenger abnormality (Beroukhim et al., 2010). Although CNAs can contain large numbers of genes, the selective advantage of driver alterations is often mediated through one or a small number of candidate oncogenes or tumour suppressor genes within the regions of copy number change (Mermel et al., 2011).

CNAs are common secondary events in ALL. They often co-operate with primary chromosomal abnormalities to drive leukaemogeneis (Mullighan et al., 2007). Specific CNAs can be associated with genetic subtypes. For example, *BCR-ABL1*+ and *ETV6-RUNX1*+ ALL are associated with a high frequency of *IKZF1* and *ETV6* deletions respectively (Mullighan et al., 2008a). Importantly, particular combinations of CNAs have an impact on prognosis. Two validated copy number profiles, namely *IKZF1*^{plus} (Stanulla et al., 2018) and UKALL-CNA (Moorman et al., 2014, Hamadeh et al., 2019) have been developed as prognostic biomarkers in childhood ALL.

IKZF1 on 7p12.2 encodes a zinc-finger DNA-binding transcription factor that is associated with chromatin remodelling and is required for the development of all lymphoid lineages (Wang et al., 1996). As a result of alternative splicing, different IKZF1 isoforms exist. These share a common C-terminal domain, which contains two zinc-finger motifs. These are required for dimerization and interaction with other proteins and are present in all isoforms of the protein (IK1-IK8). However, some isoforms do not contain the three or more N-terminal zinc finger motifs that are required for DNA-binding, and hence may function as dominant negative isoforms (Marke et al., 2018). *IKZF1* is essential for the development of all lymphoid lineages, as well as other haematopoietic lineages to a lesser extent (Marke et al., 2018) and mice with homozygous *IKZF1* deletions fail to develop any lymphoid tissue (Georgopoulos et al., 1994). *IKZF1* is frequently disrupted in ALL, most commonly through a range of different deletions, which are seen in around 15% of childhood and 40% of adult cases (Boer et al., 2016). One copy of *IKZF1* can be lost through a focal deletion affecting the entire gene or large deletions of 7p. Additionally, recurrent small intragenic deletions also result in loss of function. Exons 4-6 encode the N-terminal DNA binding domain of the protein and intragenic deletions affecting exons 4-7 are recurrent events in ALL and produce a dominant negative isoform (IK6), lacking all DNA-binding zinc finger motifs (Boer et al., 2016). *IKZF1* deletions were the first CNAs found to have an adverse impact on prognosis in ALL (Mullighan et al., 2009b), and have since been defined as one of the principle components in the prognostic risk scores mentioned above (*IKZF1*^{plus} and UKALL-CNA).

In ALL, focal CNAs also frequently target genes that are involved in cell cycle regulation and have prominent tumour suppressor activity. *CDKN2A* and *CDKN2B* are adjacent genes on 9p21.3 that generate multiple isoforms. One of their principal functions is the inhibition of CDK kinases, thereby regulating cell cycle G1 progression. *CDKN2A* also encodes an alternate open reading frame (ARF) transcript, that functions to stabilise p53 through the sequestration of MDM2 activity (Zhao et al., 2016). *CDKN2A* and *CDKN2B* are targeted by deletions in many cancer, and frequent homozygous loss is observed in ALL, often occurring through a large 9p deletion affecting one allele and a more focal isolated *CDKN2A/B* deletion on the other allele (Sulong et al., 2009). *RB1* on 13q14.2 is another prominent tumour suppressor gene that acts as a negative regulator of the cell cycle and is recurrently deleted in numerous cancers including ALL (Mullighan et al., 2007). Similarly, *PTEN* is a tumour suppressor implicated in

numerous cancers and is specifically deleted in T-cell rather than BCP-ALL (Mendes et al., 2016).

ETV6 encodes a transcription factor that plays an essential role in normal haematopoiesis (Wang et al., 1998). It is one of the most frequently implicated genes in the pathogenesis of childhood ALL, most commonly through *ETV6-RUNX* fusion (chapter 1, section 1.8.2.1). In this genetic subtype, the wild type *ETV6* allele is often deleted, which is thought to be one of the important secondary events transforming the pre-leukaemic *ETV6-RUNX1+* clone into overt leukaemia (Cavé et al., 1997). Heterozygous *ETV6* loss is seen in other genetic subtypes, consistent with a broader role in the pathogenesis of ALL. Specific to BCP-ALL, deletions are additionally observed in genes involved in B-cell development, most commonly *PAX5, EBF1* and *TCF3* as well as *BTG1*, which has a range of functions including tumour suppressor activity (Mullighan et al., 2007, van Galen et al., 2010).

CNAs can vary from small intragenic deletions to abnormalities affecting entire chromosomes. However, the size of somatic CNAs is not random. A landmark study across multiple tumour types demonstrated that these were either very small (focal) or the length of a whole chromosome or chromosome arm (arm level) (Beroukhim et al., 2010). As such, these abnormalities need to be viewed in the context of their size and the detection method. CNAs can be identified through targeted techniques such as FISH and MLPA or genome-wide techniques such as SNP arrays and whole genome sequencing. Using MLPA, gene deletions that have arisen focally compared with those where a whole chromosome has been lost, will often not be differentiated. In comparison, this detail is appreciable through SNP arrays analysis and several studies have highlighted the important role of SNP arrays in identifying clinically relevant prognostic and therapeutic targets in ALL patients (Wang et al., 2016, Baughn et al., 2015).

One of the earliest large studies to report CNAs in ALL identified a mean of 6.46 (range 0-39) CNAs per case in a cohort of 242 paediatric ALL patients (Mullighan et al., 2007). The SNP arrays used interrogated around 350,000 loci and most recurrent CNAs were focal events <1Mb in size. Of these, deletions outnumbered gains 2:1. Recurrent segments of copy number loss had a mean size of 166 kb (range 4 – 889 kb) and 44% (24/54) of recurrent deletions affected single genes. Another study using the same platform in a subgroup of 221 high risk childhood BCP-ALL patients identified a mean of 8.36 CNAs per patient (Mullighan et al., 2009b). Similar studies of adolescent and

adult patients have reported a median of 6-8 CNAs per sample (Okamoto et al., 2010, Paulsson et al., 2008). Adults were found to have a similar pattern of focal microdeletions to paediatric patients, although one study additionally identified a higher frequency of 17p losses (Okamoto et al., 2010).

Importantly, much of the literature on CNAs in ALL has been derived from comparatively low density SNP arrays by current standards (Iacobucci et al., 2013). The use of newer high-density arrays has permitted the detection of smaller and smaller abnormalities, albeit with a risk of increasing the rate of false positive calls (Bernardini et al., 2010). Using the Affymetrix Cytoscan HD array, a more recent study identified a mean of 12.3 CNAs in adult ALL patients at diagnosis, comprising 9.6 deletions, 2.3 duplications and 0.4 regions of CNN-LOH per sample (Ribera et al., 2017). To minimise false positive calls, the authors applied a minimum segment size of 25 markers for CNAs and 20 Mb for CNN-LOH, so smaller events could not be excluded. Another study analysed SNP arrays across multiple institutions and multiple array platforms (Baughn et al., 2015). This combined analysis was performed using the Nexus software, which is able to derive copy number and B-allele frequency data from a wide range of array platforms (chapter 2, section 2.6.1.5). The authors reported on average 2-6 CNAs per case in various cytogenetic subgroups of paediatric ALL including *ETV6-RUNX1+*, high hyperdiploidy and normal karyotypes. The significance of setting appropriate platform-specific thresholds for calling CNAs was also highlighted. In the study, a minimum of 25 probes was used for Affymetrix Cytoscan HD arrays and 15 probes for Illumina Infinium 850k microarray, reflecting the importance of appreciating differences in array probe density. Additionally, log2 ratio thresholds were adjusted based on the percentage of leukaemic cells in the primary sample to optimise detection of relevant CNAs even in samples with lower blast percentages.

Guidelines from the American College of Medical Genetics and Genomics (ACMG) have suggested that a minimum sample tumour content of 25% is required before embarking on SNP array analysis, although this is likely to vary between platforms and internal laboratory validation processes (Cooley et al., 2013, Schoumans et al., 2016). Similarly, even in pure samples, CNAs may be difficult to detect if present at subclonal levels (Puiggros et al., 2013). Moreover, to date, the principle of identifying genetic mosaicism using SNP arrays has principally been explored and validated in the context

of constitutional genetic disorders, rather than in the detection of subclonal somatic CNAs (Biesecker and Spinner, 2013).

Overall, the existing SNP array literature therefore confirms the importance of integrating information on sample quality, clonality, constitutional copy number variation and probe density when analysing SNP arrays to optimise the detection of clinically relevant aberrations.

4.2 Aims and objectives

The overarching aim of this work package was to produce a comprehensive copy number profile of ALL in older adults. To achieve this, the following objectives needed to be addressed:

1. Create a SNP array cohort broadly representative of older adults with ALL
2. Optimise SNP array segmentation to facilitate discovery of novel lesions while minimising false positive calls
3. Perform MLPA to validate SNP array segmentation in specific genes
4. Identify potential novel driver genes affected by CNAs in the cohort
5. Design a customised targeted sequencing panel to validate novel driver abnormalities.

4.3 Methods

4.3.1 SNP array segmentation settings

The SNP array cohort incorporated all patients aged ≥60 years enrolled in the UKALL14 or UKALL60+ trials with DNA extracted from the diagnostic bone marrow sample. Demographic and cytogenetic information for all included patients can be found in supplementary table 1.

Bone marrow DNA samples were sent to the Northern Genetics Centre, Newcastle Hospitals NHS Foundation Trust, and SNP arrays were performed using the Illumina 850k CytoSNP or Affymetrix Cytoscan HD array (chapter 2 section 2.6.1.3). Raw signal intensity files (CEL and IDAT files) were transferred back and used for all subsequent analyses. These were loaded into Nexus Copy Number 10 (chapter 2 sections 2.6.1.5 – 2.6.1.8) and prepared for copy number segmentation.

Accurate generation of copy number segments is based on three primary considerations:

1. *The minimum positive and negative log2 ratio deflection required.*
   Nexus copy number uses a segmentation algorithm based on the Hidden Markov Model (HMM). However, in comparison to the classic HMM, integer levels of copy number are not assumed. Instead, either default or customised log2 ratio thresholds are defined, and if a group of probes falls above or below these boundaries, a copy number segment may be generated. The degree of deviation of the probe log2 ratio from the sample baseline in a CNA segment is based on a number of factors, including sample purity, the proportion of tumour DNA affected (intratumoural heterogeneity) and the SNP array platform that has been used to generate the raw data. A judgment therefore needs to be made to define the degree of positive or negative deflection in probe log2 ratio required to generate a segment of copy number gain or loss respectively. Based on assessments of sample purity and objective measures of quality, these thresholds needed to be adjusted to generate meaningful data, as described below.

2. *The minimum defined size of a copy number segment.*
   Probe density varies both across the genome and between different SNP array platforms. This needs to be taken into account when defining minimum copy number segment sizes to ensure that small intragenic lesions can be detected whilst minimising false positive calls if a very small number of probes are used to generate a copy number segment.

3. *The permitted scattering of probes in the log2 space within a copy number segment (significance threshold).*

Copy number data are inherently noisy. Probe signal intensity is influenced by a number of factors including DNA quality, GC content and the proportion of the genome affected by copy number variation. Additionally, genuine copy number changes may represent germline copy number variation (CNV), unrelated to the leukaemia genome. As such, quality control measures need to be taken into account during segmentation to ensure false positive and false negative calls are kept to a minimum. Once loaded to Nexus, a SNP array Quality score is generated. This is computed through ordering by magnitude the difference between signals from adjacent probes and then removing 1.5% of probes at each end of the variance spectrum. The smaller this final quality metric, the better quality the sample as adjacent probes will be producing a similar signal intensity. Conversely, higher Quality scores are associated with a greater spread of probes in the log2 space and more noisy data. To reduce the rate of false positive calls, a more stringent significance threshold can be applied to reduce the probability of random fluctuations in probe intensity generating a copy number segment. Adjusting the significance threshold controls the number of copy number segments generated through segmentation of the genome, with smaller significance threshold values being more stringent and generating fewer segments. In general, samples from higher density arrays and those with poorer (higher) quality scores require smaller, more stringent significance thresholds to minimise false positive copy number calls.

Based on sample quality and SNP array platform, Nexus segmentation settings were customised to account for different array types and varying levels of noise across the cohort. Minimum numbers of 10 probes for Affymetrix arrays and 6 probes for Illumina arrays were used for genome segmentation.

The significance thresholds used for genome segmentation in this study are shown in figure 4.1, based on SNP array type and quality score. Affymetrix SNP arrays had poorer quality scores than Illumina arrays and significance thresholds were set to more stringent levels with increasing (worsening) quality scores. Benchmark significance thresholds and log2 ratio deflections for segmentation were $1x10^{-12}$ and ± 0.08 respectively for Illumina arrays and $1x10^{-20}$ and ± 0.15 for Affymetrix arrays. All customised SNP array segmentation settings used for each sample can be found in supplementary table 2.

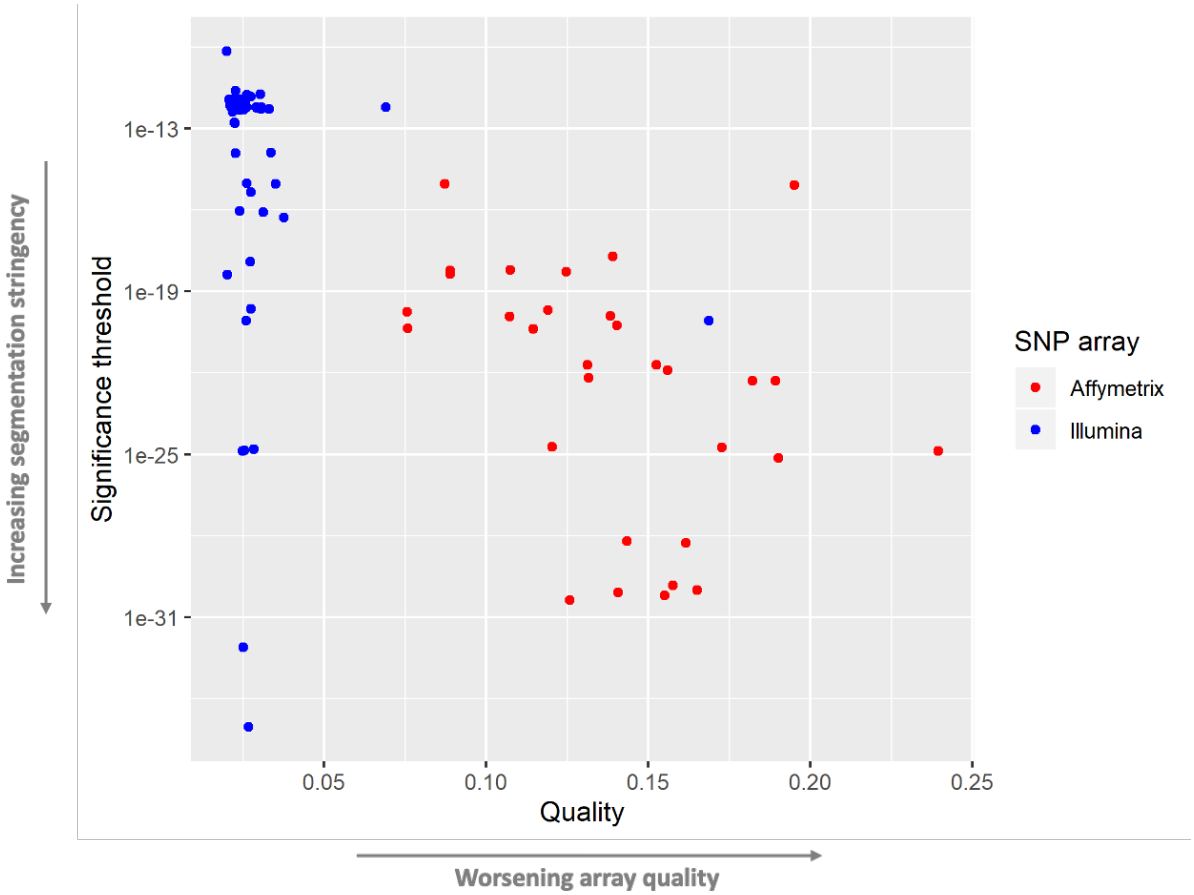


Figure 4.1. Significance thresholds used for SNP array segmentation based on array type and quality score. Samples with higher (worse) quality scores were segmented using lower (more stringent) significance thresholds to minimise the false positive call rate.

4.3.2 SNP array analysis

Following copy number segmentation, SNP arrays were analysed both visually and computationally. Initially, large abnormalities affecting whole chromosome arms were identified and described by standard visual analysis of the SNP array traces. Next, CNAs in eight well-described driver genes, namely *EBF1*, *IKZF1*, *CDKN2A*, *CDKN2B*, *PAX5*, *ETV6*, *BTG1* and *RB1*, were scrutinised case-by-case and recorded. Finally, a computational SNP array analysis method was developed using the Nexus output to identify potential novel driver abnormalities within the cohort. Often, driver lesions targeted by CNAs are either present as the only disrupted gene in a copy number segment or have breakpoints within the relevant gene (e.g. *IKZF1* exon 4-7 deletion). Therefore, to detect recurrent focal CNAs, including novel abnormalities, genes at the breakpoints of deletions were specifically extracted using a customised automated pipeline, which was created and executed in R (figure 4.2). The input file was a readily

available results file exported from Nexus, whereby each aberration within the entire cohort was listed, along with all genes present within the copy number segment boundaries.

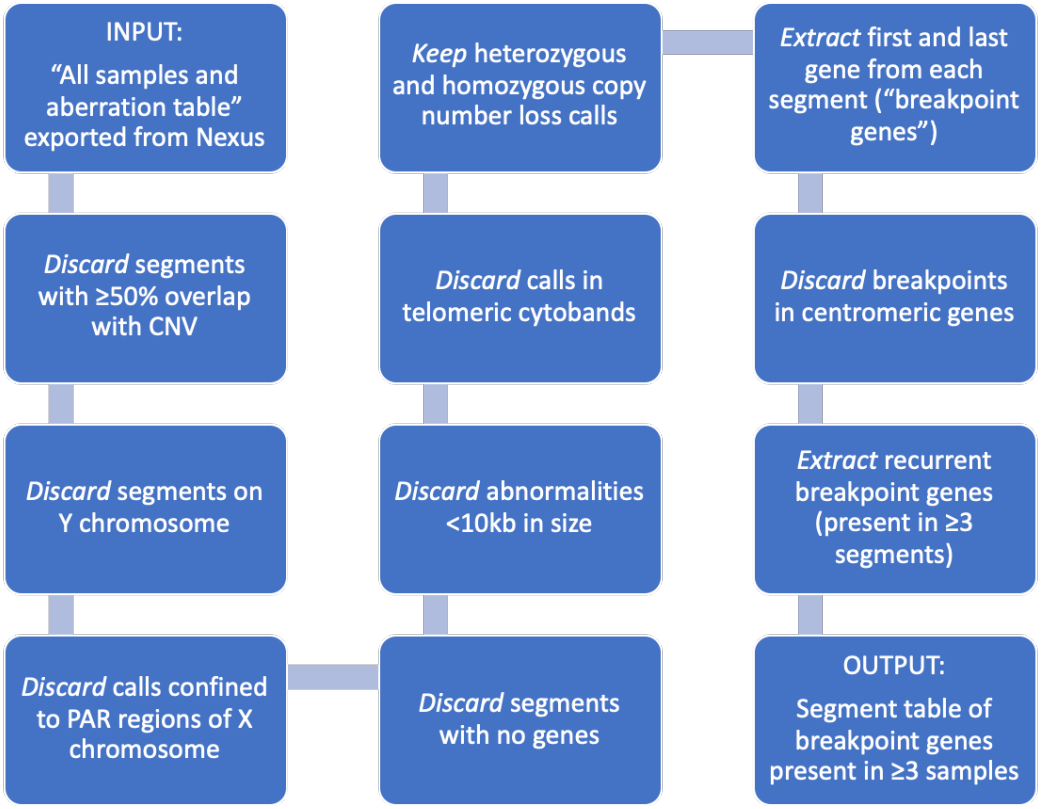


Figure 4.2. Pipeline to extract candidate driver genes from complete SNP array dataset. Individual segment information from Nexus included percentage overlap with regions of known CNV reported in the Toronto Database of Genomic Variants (MacDonald et al., 2014), which permitted the exclusion of calls with ≥50% overlap with CNVs (step 2 in pipeline). Due to sex-related inconsistencies in segmentation, all calls on chromosome Y and the pseudoautosomal regions (PAR) of chromosome X were filtered out (steps 3 and 4). Additionally, all abnormalities present in the telomeric cytobands of each chromosome were also filtered out due to inherently noisy data on these regions (step 7). Breakpoint genes in centromeric genes were discarded as these were automatically present in any arm-level or whole-chromosome abnormalities, hence did not specifically reflect the presence of driver genes (step 10).
CNV: copy number variation; PAR: pseudoautosomal regions.

4.3.3 MLPA validation

To ensure segmentation had been optimised, copy number abnormalities in 8 genes/regions were validated against MLPA, using the IKZF1-P335 kit (chapter 2, section 2.6.2). This comprised probes covering *EBF1*, *IKZF1*, *CDKN2A/B*, *PAX5*, *ETV6*, *BTG1*, *RB1* genes together with the PAR1 region of Xp22.3/Yp11.2. MLPA results were analysed using GeneMarker software (chapter 2, section 2.6.2.3).

Deletions or gains were called when probe ratios of <0.75 or >1.3 respectively were present in 2 consecutive probes in a gene.

4.3.4 SureSelect XT2 NGS validation and protocol optimisation

Whole genome amplification (WGA) and SureSelect XT2 library prep were performed as described to validate biologically relevant and potential novel CNAs from the SNP array analysis (chapter 2 sections 2.4.4 and 2.7).

Following WGA, double stranded DNA (dsDNA) concentrations were measured using the Quant-iT Picogreen assay and FLUOstar Omega microplate reader (chapter 2, section 2.4). Of the 28 initial validation samples, 24 had sufficient amounts of dsDNA following WGA to proceed to library prep.

For each case 1-2ug of sample DNA (depending on availability) was diluted in TE buffer to a volume of 100ul. The DNA was sheared by sonication using the Bioruptor Pico with a target fragment size of 800bp. Following shearing DNA fragment sizes were assessed on the Bioanalyzer (chapter 2, section 2.4.5.4) to ensure fragment peaks were present around 800 bp (figure 4.3).

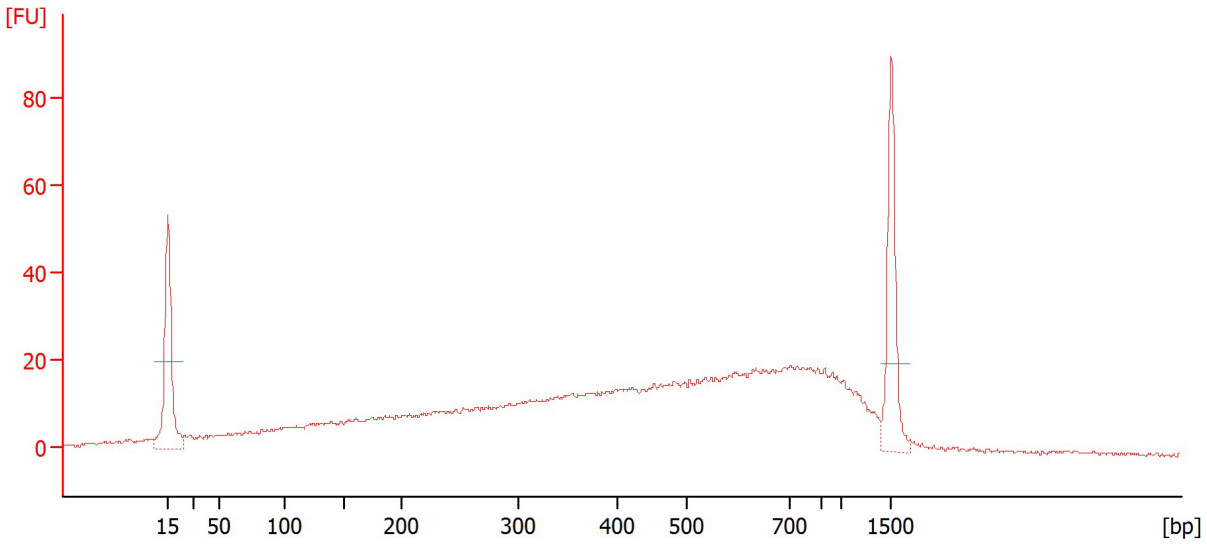


Figure 4.3. Example Bioanalyzer electropherogram following DNA shearing showing DNA fragment peak ~800bp (patient 24813). DNA fragment size shown on x-axis and fluorescence intensity on y-axis.

Electropherograms of the input material confirmed suitable DNA shearing with an acceptable fragment size in all cases.

However, following the PCR amplification stage of library prep using the Herculase II DNA polymerase (chapter 2, section 2.7.2), DNA samples showed suboptimal amplification and an over-representation of shorter DNA fragments (see figure 4.4).

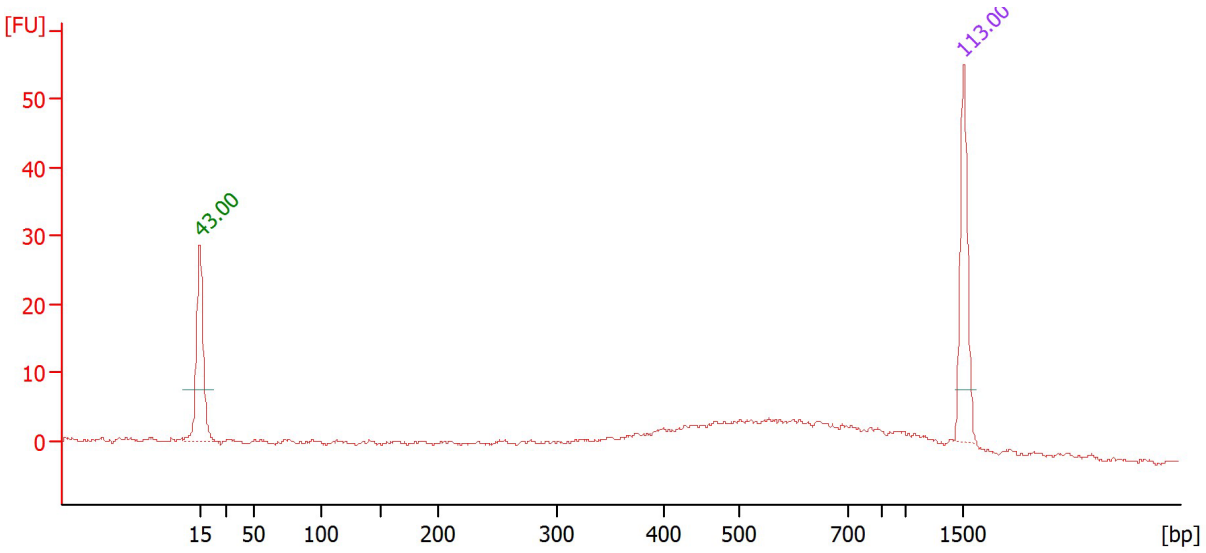


Figure 4.4. Example Bioanalyzer electropherogram following PCR amplification stage of library prep using Herculase II DNA polymerase (patient 25247). Trace demonstrates very low peak at ~550bp indicating low DNA concentration and small fragment size. DNA fragment size shown on x-axis and fluorescence intensity on y-axis.

The PCR amplification was therefore trialled using the Longamp Taq polymerase enzyme, which is optimised for amplification of longer DNA fragments. This resulted in much better amplification of the DNA samples at the PCR amplification stage (figure 4.5) and the Longamp enzyme was therefore used for all subsequent samples, as well as the final PCR amplification of the library prep.

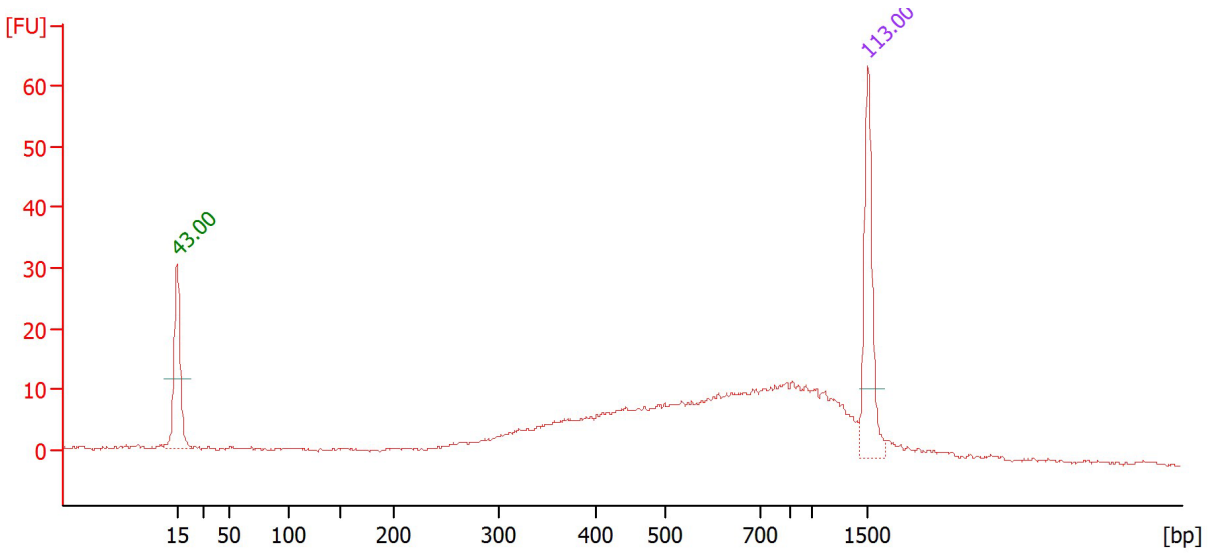


Figure 4.5. Example Bioanalyzer electropherogram following PCR amplification stage of library prep using Longamp polymerase (patient 25247). Trace demonstrates clear peak at ~800bp indicating sufficient DNA concentration and large fragment size. DNA fragment size shown on x-axis and fluorescence intensity on y-axis.

To account for the alteration in the polymerase enzyme used, the thermal cycler settings were also adjusted to optimise amplification of longer DNA fragments (table 4.1).

Segment	Number of cycles	Temperature	Time
1	1	95ºC	3 mins
2	5	98ºC	20 secs
		60ºC	15 secs
		65ºC	2 mins
3	1	65ºC	10 mins
4	1	4ºC	Hold

Table 4.1. Pre-capture PCR thermal cycler program using Longamp Taq polymerase in place of Herculase II DNA polymerase.

In total, 23/24 samples were successfully put through the full library prep. One sample (27181) had to be excluded due to poor PCR amplification. Thus, the final NGS validation cohort comprised 23 samples as detailed in table 4.2.

Patient ID	Genetic subtype	Post WGA concentration (ng/ul)	Library prep
25267	B-other	103.8	Successful
26614	B-other	27.4	Successful
28335	B-other	29.7	Successful
24890	B-other	94.6	Successful
25082	*BCR-ABL1*	39.7	Successful
25208	*BCR-ABL1*	97.4	Successful
25247	*BCR-ABL1*	100.1	Successful
28057	*BCR-ABL1*	45.8	Successful
28182	*BCR-ABL1*	20.5	Successful
28350	*BCR-ABL1*	95	Successful
28670	*BCR-ABL1*	81.1	Successful
26660	*BCR-ABL1*	93.2	Successful
25346	*BCR-ABL1*	-26.7	Insufficient DNA
25793	*BCR-ABL1*	-21.3	Insufficient DNA
25967	Complex	99.6	Successful
25437	HoTr	91.6	Successful
29407	HoTr	94.7	Successful
27833	*IGH-BCL2*	16.4	Successful
27181	*IGH-CEBPE*	36.1	Failed amplification
25130	*IGH-CRLF2*	63.8	Successful
25552	*IGH@-r*	64.4	Successful
28581	*KMT2A-AFF1*	47.9	Successful
29908	*KMT2A-AFF1*	12.1	Insufficient DNA
25100	*KMT2A-v*	74.7	Successful
27389	*KMT2A-v*	4.4	Insufficient DNA
27642	T-ALL	76.7	Successful
24813	Unknown	85.3	Successful
25451	*ZNF384-r*	40.1	Successful

Table 4.2. DNA concentrations (in 50ul) following whole genome amplification and subsequent outcome following SureSelect XT2 library prep. Negative values represent <5ng/ul.
WGA: whole genome amplification; B-other: BCP-ALL with no primary chromosomal abnormality identified; HoTr: low hypodiploidy/near triploidy; Complex: 5 or more unrelated chromosomal abnormalities on karyotype; *IGH@-r*: *IGH@* rearrangement; *KMT2A*-v: *KMT2A* fusion with non-*AFF1* partner; *ZNF384*-r: *ZNF384* rearrangement.

All samples successfully put through the library prep (n=23) were then sequenced on the NextSeq 550 using 150 bp paired end-chemistry and deletion breakpoints were directly visualised in IGV (chapter 2, section 2.6.4). Individual deletions could then be validated when concordant reads flagged a structural variant as shown in figure 4.6.

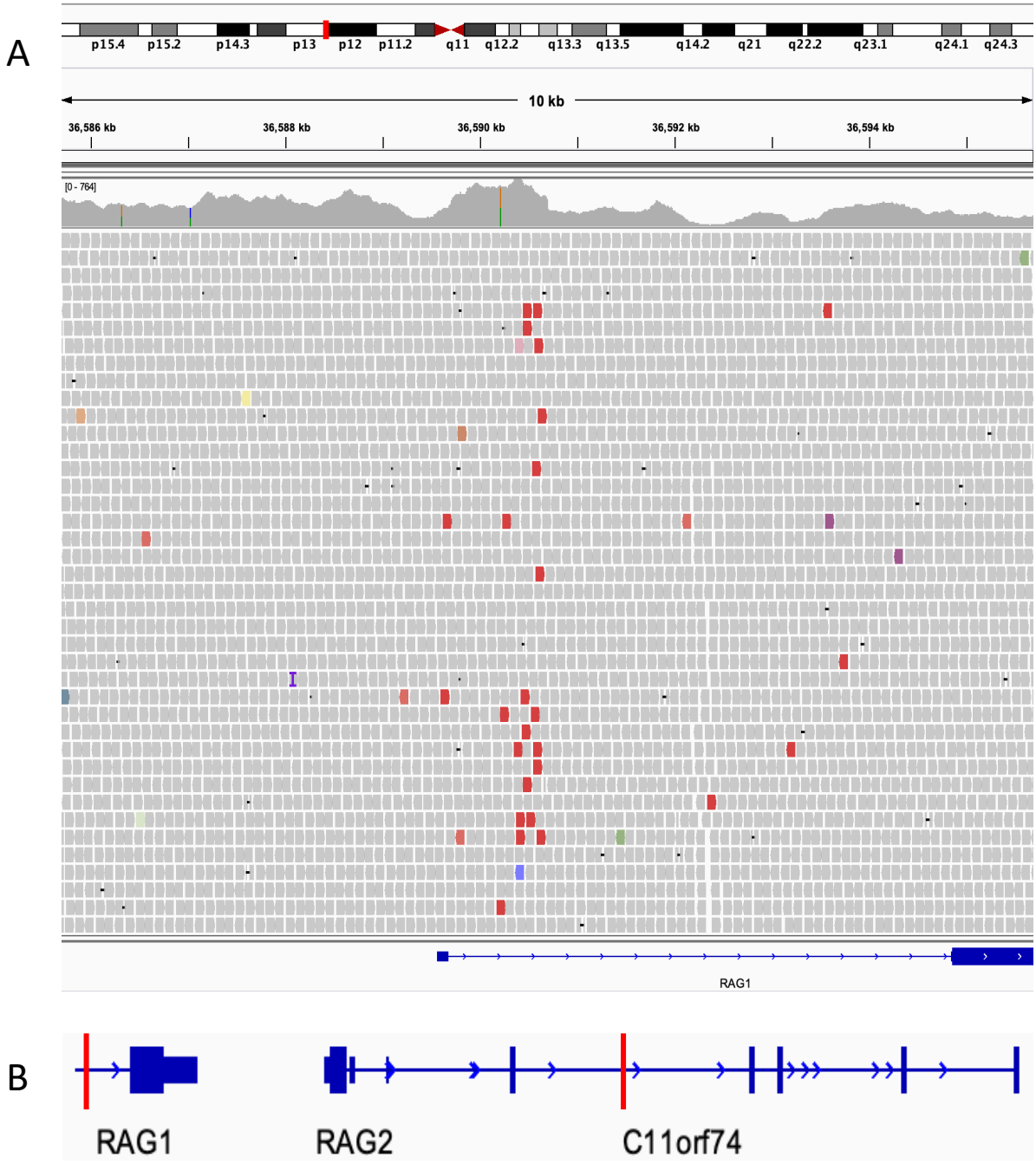


Figure 4.6. Example of validated breakpoint in *RAG1* viewed in IGV (patient 25130). Each individual segment represents a single sequencing read. Red flagged reads indicate mate pairs >10kb apart. Multiple flagged reads within same genomic region (A). Mate pairs all mapped to region in *C11orf74* indicating deletion involving exon 2 of *RAG1* and entirety of *RAG2*. Approximate breakpoints shown in B (red lines).

4.4 Results

4.4.1 Patient cohort

SNP arrays were performed on 83 patient samples using Illumina CytoSNP 850k (n=52) and Affymetrix Cytoscan HD (n=31) arrays. Median patient age was 64 years (range 60-83) and 57% were female. A range of genetic subgroups were included, broadly representative of ALL in older adults (figure 4.7). Patient samples were obtained from the UKALL14 (n=51) and UKALL60+ (n=32) clinical trials. Detailed patient characteristics and genetic data can be found in supplementary table 1.

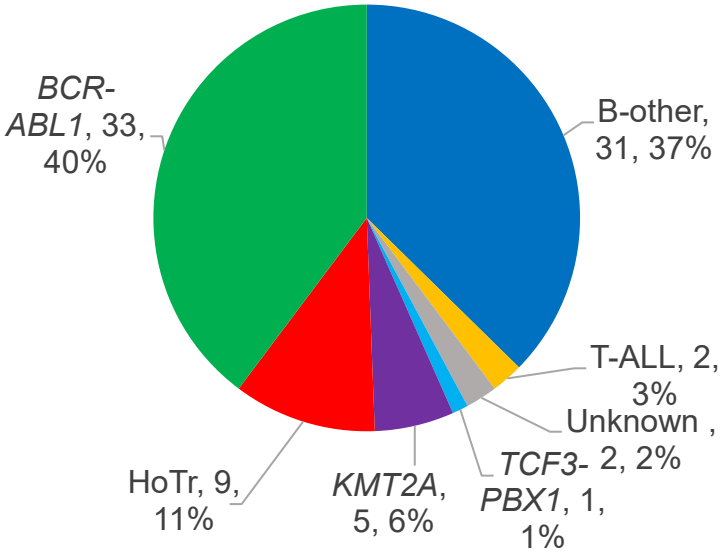


Figure 4.7. Representation of genetic subtypes in the complete SNP array cohort (n=83). HoTr: low hypodiploidy/near triploidy; *KMT2A*: *KMT2A* fusion with any partner; B-other: BCP-ALL with no primary chromosomal abnormality identified including failed, normal and complex karyotypes.

4.4.2 Copy number profile

4.4.2.1 Total CNAs per patient sample

Using the customised settings described in section 4.3.1 and supplementary table 2, the total number of CNAs generated was assessed per sample and compared between array types. Overall, Affymetrix SNP arrays generated significantly more copy number calls than Illumina arrays (figure 4.8). This was most likely due to a combination of

increased probe density and increased false positive call rate relating to poorer quality scores, despite increasing the stringency of significance thresholds (figure 4.1 and supplementary table 2).

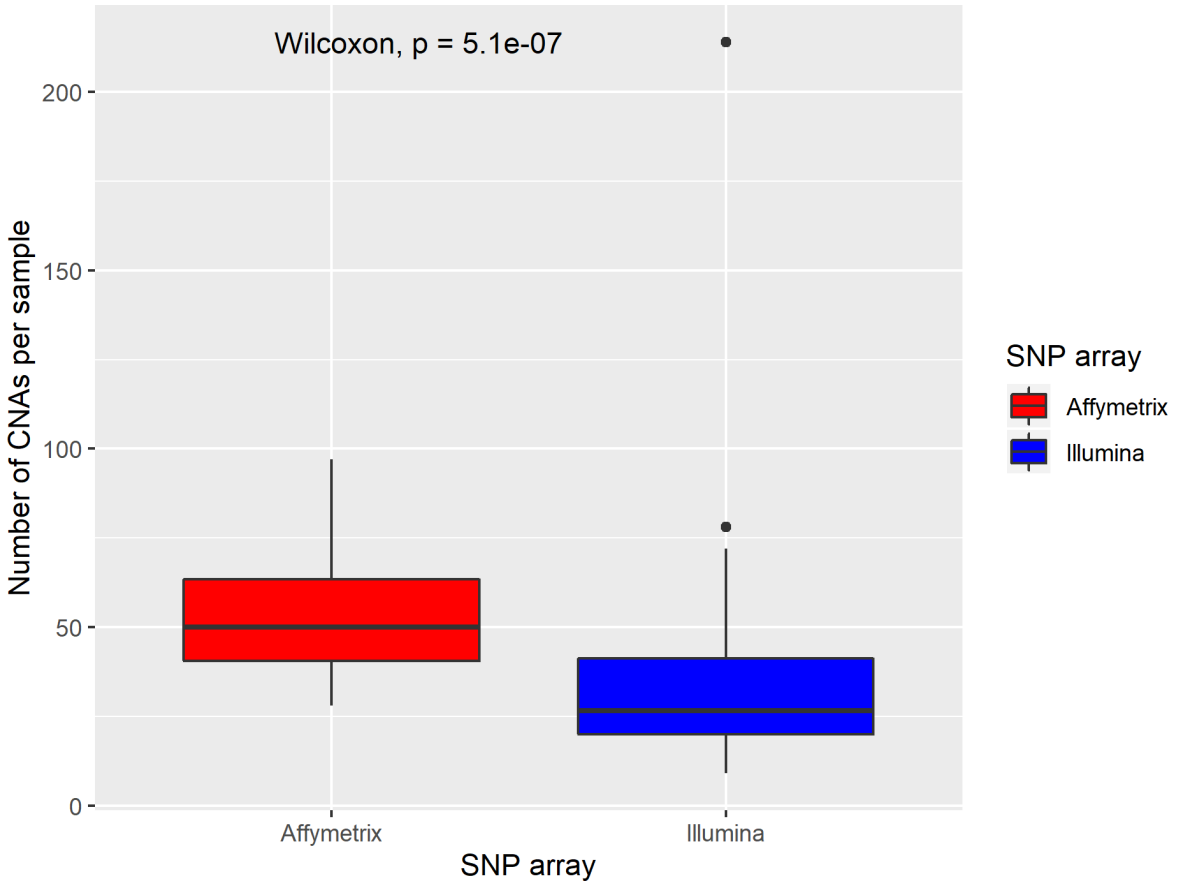


Figure 4.8. Number of CNA segments by SNP array platform. Due to higher probe density, Affymetrix Cytoscan HD array samples had on average more CNAs than Illumina 850k array samples (before germline copy number variants (CNVs) removed).
CNA: copy number abnormality.

Accepting this inevitable degree of discrepancy between Affymetrix and Illumina arrays, the segmentation settings selected in this analysis were thought to achieve the best balance between permitting the detection of novel copy number abnormalities at potentially subclonal levels, and minimising false discoveries.

Based on the customised segmentation thresholds, 4,604 individual copy number calls were made across the cohort of 83 patients. Post processing filtering was performed as detailed in figure 4.2. A total of 1,975 had at least 50% overlap with regions of known constitutional CNV in the Toronto Database of Genomic Variants (DGV) (MacDonald et al., 2014) and were removed from further analysis. Importantly, the *CDKN2A* and *CDKN2B* genes are within a reported CNV region in DGV so were safeguarded from this filtering step. Next, due to inherent noise and to remove inconsistencies between

male and female patients, calls on chromosome Y (n=213) as well as the pseudoautosomal regions of chromosome X (Xp22.33, Xq21.31 and Xq28) (n=106) were also discarded. However, as the interstitial deletion leading to *P2RY8-CRLF2* fusion occurs in Xp22.33, all SNP arrays were specifically visually assessed for this abnormality, and no cases were identified. Finally, CNA segments that did not include any genes (n=292) and those smaller than 10kb were filtered out (n=39).

This automated sequential filtering process left a total of 1,979 calls of potential interest across the 83 patient samples. Of these, 48% (n=942) were heterozygous deletions, 1% (n=26) were homozygous deletions, 12% (n=243) were one copy gains, 24% (n=469) were regions of LOH and 15% (n=299) classed as allelic imbalance. Deletions were more frequent than gains in all cytogenetic subgroups apart from high hyperdiploidy (figure 4.9).

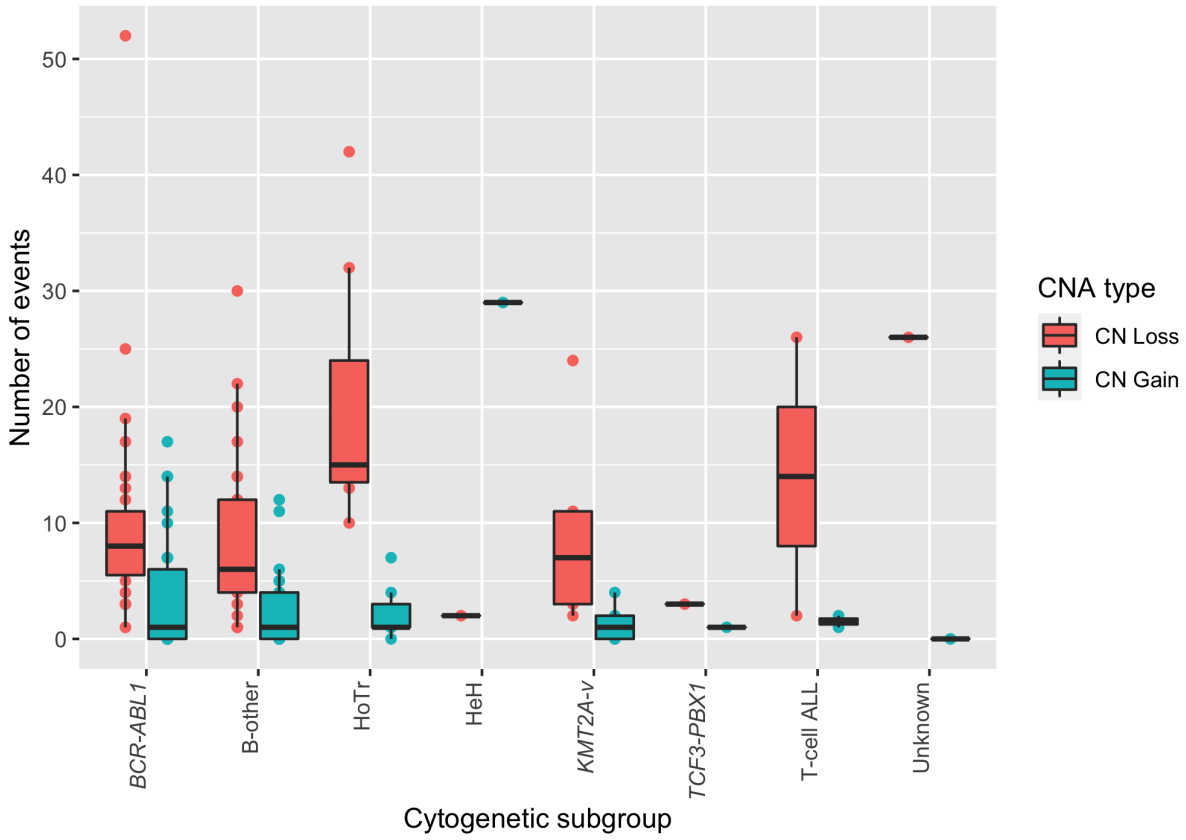


Figure 4.9. Boxplot displaying number of deletions and gains across cytogenetic subtypes in 81 older adults with ALL (2 low hypodiploid cases (28644 and 25437) excluded due to poor SNP array segmentation). B-other: BCP-ALL with no primary chromosomal abnormality identified; HoTr: low hypodiploidy/near triploidy; HeH: high hyperdiploidy; *KMT2A*-v: *KMT2A* fusion with any partner; CNA: copy number abnormality; CN Loss: copy number loss; CN Gain: copy number gain.

4.4.2.2 Chromosome arm level copy number abnormalities

The frequency of arm level events is shown in table 4.3. Arm-level CNAs were identified through direct visualisation of SNP arrays in Nexus and were included when they affected either an entire chromosome or the majority of a chromosome arm.

Abnormality	% of all cases (n)	Subgroups represented (n)
del(9p)	21% (15)	*BCR-ABL1* (11), B-other (4)
gain of Ph*	12% (9)	*BCR-ABL1* (9)
gain 1q	10% (7)	B-other (4), *BCR-ABL1* (2), *TCF3-PBX1* (1)
-7	10% (7)	*BCR-ABL1* (5), B-other (2)
del(7p)	8% (6)	*BCR-ABL1* (4), B-other (2)
gain 21q	8% (6)	*BCR-ABL1* (3), B-other (3)
del(17p)	7% (5)	B-other (4), *BCR-ABL1* (1)
del(12p)	5% (4)	B-other (2), *BCR-ABL1* (1), T-ALL (1)
gain 14q	5% (4)	*BCR-ABL1* (4)
+5	4% (3)	B-other (2), *BCR-ABL1* (1)
-9	4% (3)	*BCR-ABL1* (3)
+2	3% (2)	*BCR-ABL1* (2)
+4	3% (2)	*BCR-ABL1* (2)
+6	3% (2)	*BCR-ABL1* (1), MLL (1)
gain 8q	3% (2)	*BCR-ABL1* (1), B-other (1)
+11	3% (2)	B-other (1), *BCR-ABL1* (1)
del(13q)	3% (2)	*BCR-ABL1* (2)

Table 4.3. Number of whole chromosome and arm-level abnormalities across 73 SNP arrays. Low hypodiploid (n=9) and high hyperdiploid (n=1) cases excluded. Four *BCR-ABL1*+ cases had multiple chromosomal gains, which is a recognised secondary abnormality in this subgroup and accounts for the majority of whole chromosome gains observed.
***Consistent with extra copy of Philadelphia chromosome in *BCR-ABL1*+ cases.**
B-other: BCP-ALL with no primary chromosomal abnormality identified.

Next, significance testing was performed using GISTIC (Genomic Identification of Significant Targets in Cancer) 2.0 (Mermel et al., 2011). This is a computational method that initially models the rate of background copy number alteration and separates abnormalities by length into focal or arm-level events. Based on the background rate of CNAs in the cohort, p-values are then calculated to determine the significance of individual abnormalities. Although, the method relies on regions of constitutional CNV having been excluded prior to the analysis, these do not occur on

a whole chromosome or arm-level scale without clinical manifestation. The GISTIC 2.0 analysis was performed on the full cohort of 83 SNP arrays and generated false discovery rates (q-values) for each arm level abnormalities, with q-values <0.05 deemed to be statistically significant (table 4.4).

CNA type	Arm	Frequency	q-value
Deletion	9p	22%	1.30E-14
	7p	19%	3.04E-11
	7q	16%	5.73E-08
	17p	14%	5.03E-06
	9q	12%	0.000226
	16q*	12%	0.000758
	3p*	10%	0.0162
	16p*	10%	0.0203
	17q*	8%	0.0289
	15q*	8%	0.0456
	3q*	8%	0.0474
Gain	21q	12%	0.00645
	1q	10%	0.017

Table 4.4. Gistic 2.0 output for arm level copy number abnormalities and false discovery rates (q-values) across all 83 SNP arrays. Only arm level CNAs with significant q-value <0.05 shown.
***Specifically observed in low hypodiploid cases.**
CNA: copy number abnormality.

These data confirm that several arm level events occur recurrently in older adults with ALL. Deletions occurred more frequently than gains and 9p losses were seen in around 20% of cases. Deletions of chromosome 3, 15 and 16 were recurrent events but isolated to low hypodiploid patients so were not representative of the overall cohort.

4.4.2.3 Abnormalities in known driver genes

As mentioned previously, several driver genes are recurrently affected by CNAs in ALL, and were specifically identified from the segmented copy number data in Nexus (table 4.5).

Gene	% cases with deletion (n)	Heterozygous deletions	Homozygous deletions
IKZF1	52% (43)	41	2
CDKN2A	45% (37)	18	19
CDKN2B	45% (37)	21	16
PAX5	39% (32)	32	0
RB1	22% (18)	17	1
ETV6	21% (17)	16	1
EBF1	19% (16)	15	1
BTG1	12% (10)	9	1

Table 4.5. Deletions affecting known driver genes targeted by CNAs in ALL. Deletions reported if present in any coding part of gene (focal and arm-level events).

To examine whether specific events interacted with *BCR-ABL1* status, the frequency of each gene deletion was considered separately in *BCR-ABL1*+ and *BCR-ABL1* negative patients and Fisher's Exact Test was performed (table 4.6). *IKZF1* and *CDKN2A/B* were the most commonly deleted genes in *BCR-ABL1* positive and *BCR-ABL1* negative cases respectively.

Gene deletion	*BCR-ABL1+* % cases (n=33)	*BCR-ABL1-* % cases (n=50)	Fisher's Exact Test (p-value)
EBF1	18% (6)	20% (10)	1
IKZF1	70% (23)	40% (20)	**0.01**
CDKN2A/B	39% (13)	48% (24)	0.50
PAX5	45% (15)	34% (17)	0.36
ETV6	6% (2)	30% (15)	**0.01**
BTG1	6% (2)	16% (8)	0.30
RB1	9% (3)	30% (15)	**0.03**

Table 4.6. Frequency of individual deletions in known driver genes split by *BCR-ABL1* status. Significant differences identified between rate of *IKZF1*, *ETV6* and *RB1* deletions between *BCR-ABL1* positive *(BCR-ABL+)* and *BCR-ABL1* negative *(BCR-ABL1-)* cases.

A significantly higher frequency of *IKZF1* deletion was observed in *BCR-ABL1* positive compared to *BCR-ABL1* negative patients. Conversely, *ETV6* and *RB1* deletions had a higher prevalence in *BCR-ABL1* negative patients. The frequency of other individual abnormalities was not significantly impacted by *BCR-ABL1* status in this cohort.

The combined profile of these deletions was then examined within each patient, and subgrouped into the recurrent patterns observed (figure 4.10) and subsequently by *BCR-ABL1* status (figure 4.11). *CDKN2A* and *CDKN2B* deletions were combined as a single event, consistent with previous studies (Schwab et al., 2013, Moorman et al., 2012). In total, 22% (n=18) of patients had no deletions, 20% (n=17) had one deletion, 19% (n=16) had two deletions, 22% (n=18) had 3 deletions and 17% (n=14) had four or more gene deletions. *IKZF1* deletions in particular co-occurred with other gene deletions much more commonly than in isolation.

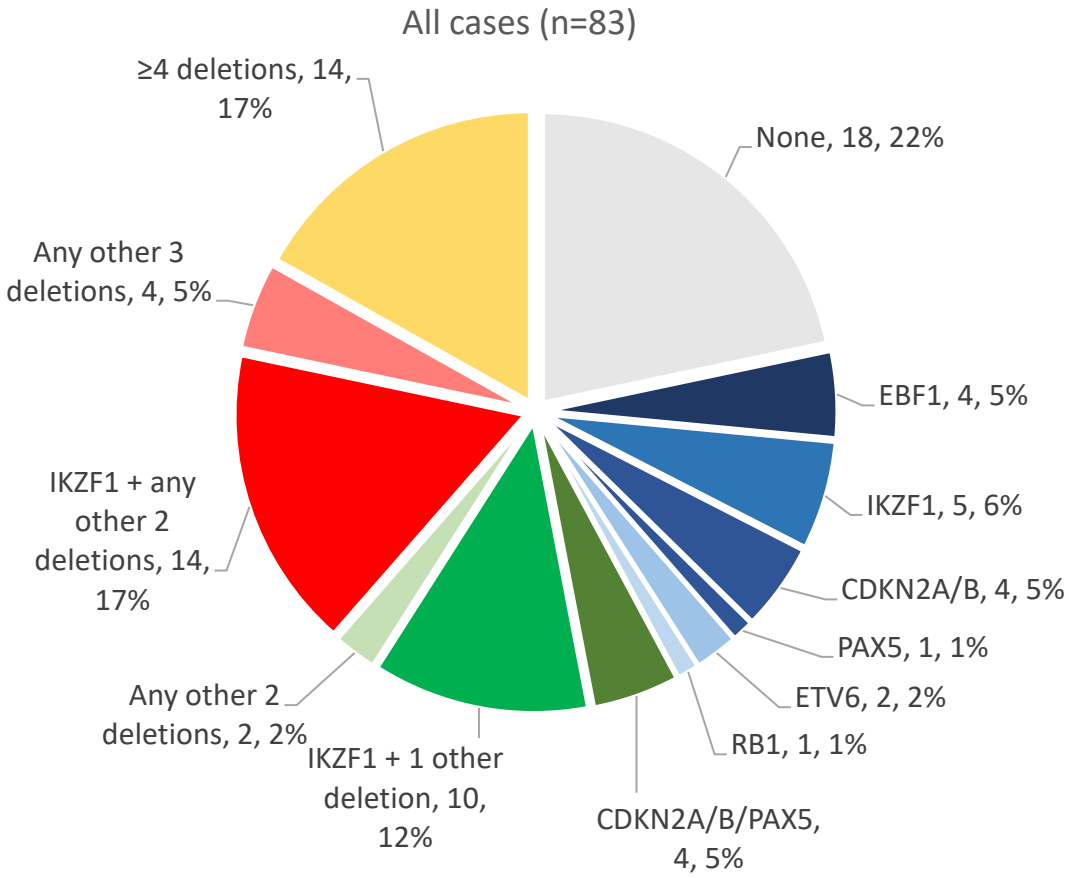


Figure 4.10. Patterns of gene deletions observed across full cohort of 83 SNP arrays. Combinations of deletions grouped by patterns observed. *CDKN2A/B/PAX5* combined deletion commonly represents del(9p). Cases with single gene deletions shown in blue segments, cases with 2 deletions in green segments and cases with 3 deletions in red segments.

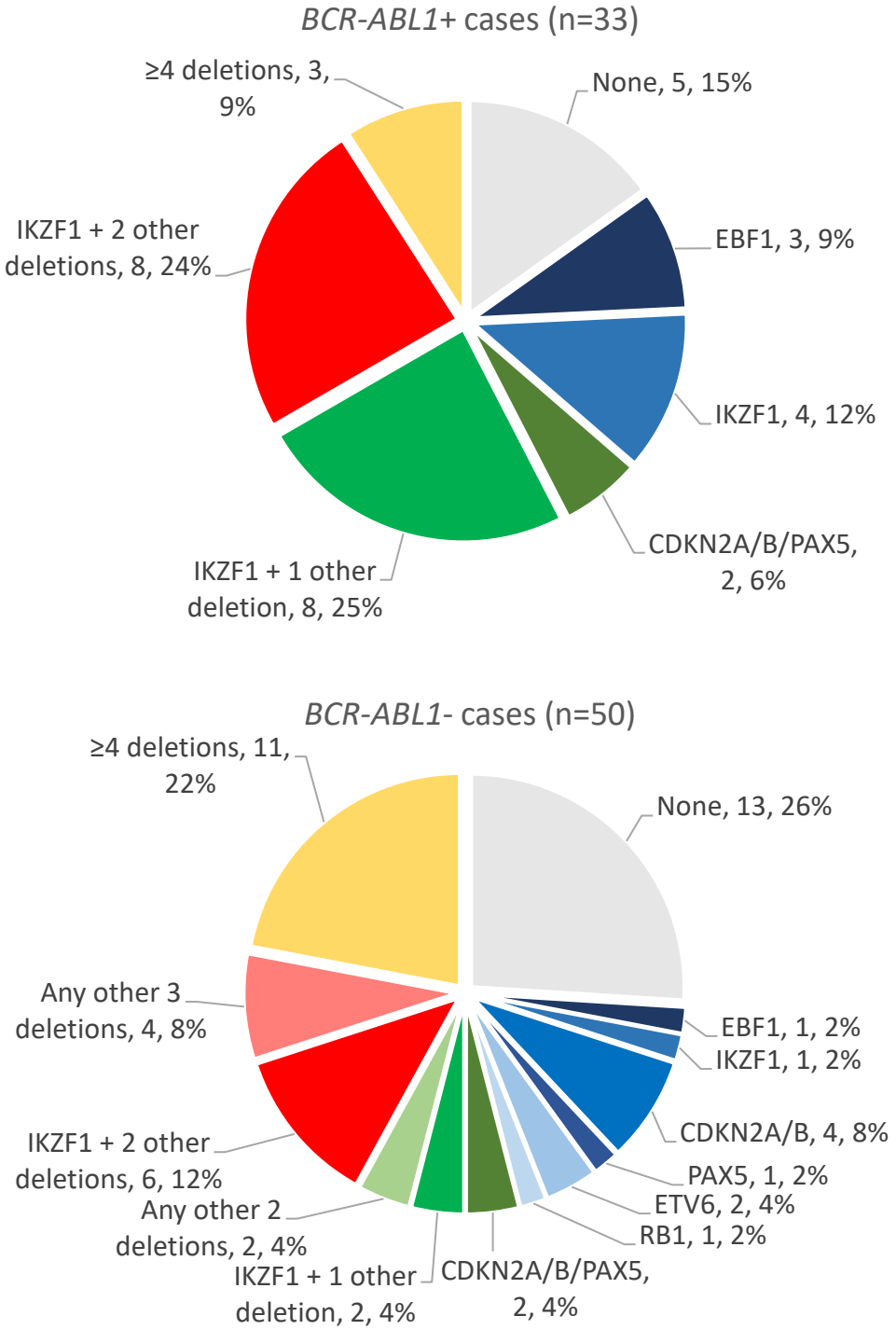


Figure 4.11. Patterns of gene deletions divided by *BCR-ABL1* status. Cases with single gene deletions shown in blue segments, cases with 2 deletions in green segments and cases with 3 deletions in red segments.
***BCR-ABL1+*: *BCR-ABL1* positive; *BCR-ABL1-*: *BCR-ABL1* negative.**

4.4.2.4 Genome-wide identification of driver abnormalities

Following the analysis of arm-level abnormalities and the specific known driver genes described above, the customised pipeline outlined in figure 4.2 was applied to extract

all genes that were recurrently present at the breakpoints of segments of copy number loss (figure 4.12).

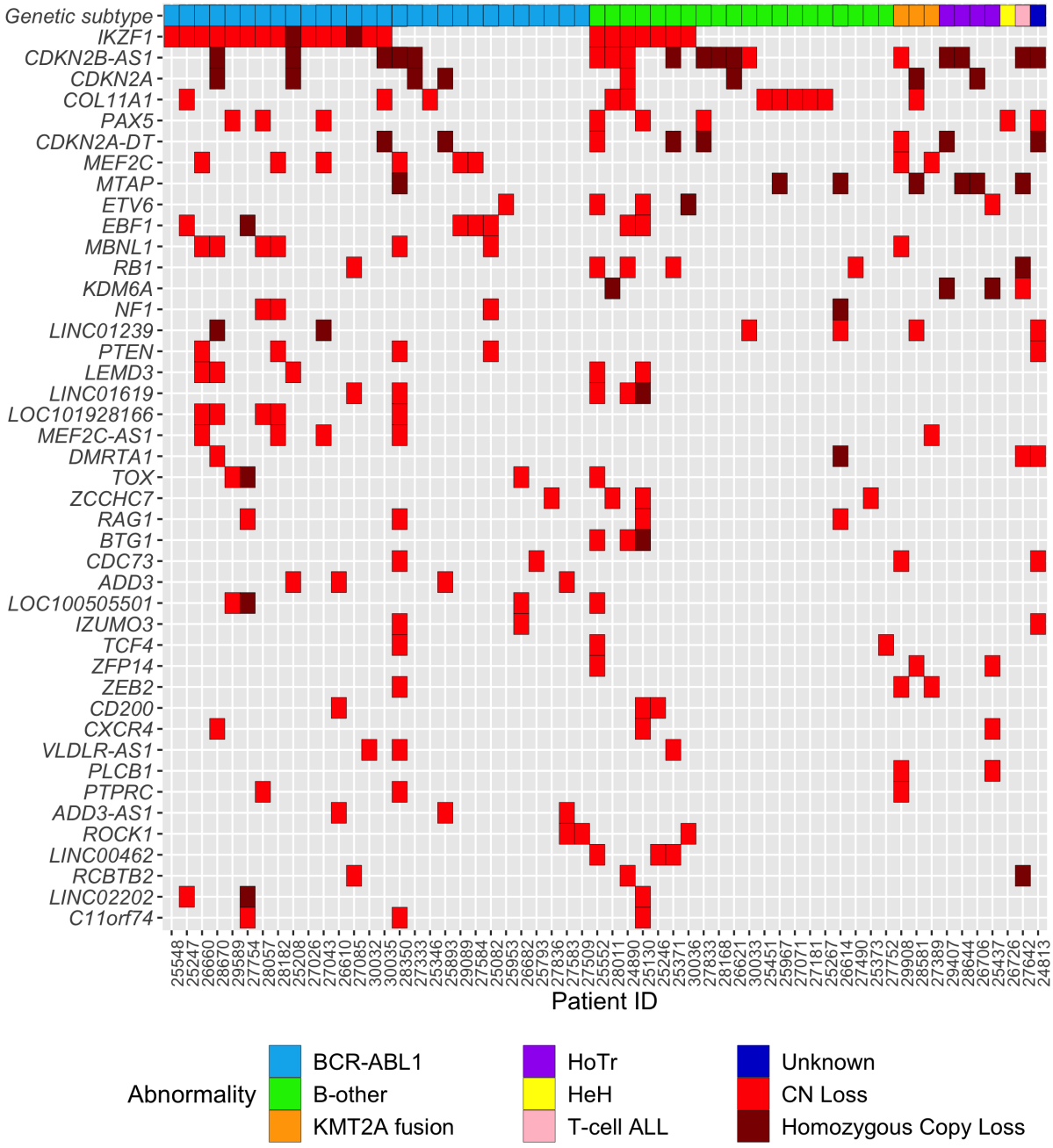


Figure 4.12. Recurrent genes within breakpoints of deletions (shown if seen ≥3 times in cohort) ordered by frequency and primary genetic abnormality. Cases are not included if the breakpoints fell outside the specified gene.
HoTr: low hypodiploidy/near triploidy; HeH: high hyperdiploidy; B-other: BCP-ALL with no primary chromosomal abnormality identified. CN Loss: copy number loss.

Reassuringly, this method still extracted all the well-established driver abnormalities detailed above, albeit only where they were present as focal deletion, at or adjacent to a copy number breakpoint, rather than as an arm-level event. Indeed, *IKZF1* remained the most frequent gene deletion, and was seen as a focal event in 27% (22/83) of

samples. Thus, these accounted for 51% of all *IKZF1* deletions, with the remaining 49% (n=19) resulting from larger deletions of 7p. Breakpoints were also present in *CDKN2A/B* in 24% (20/83) of cases, *PAX5* in 10% (8/83) of cases, *EBF1* in 8% (7/83) of cases, *RB1* in 7% (6/83) of cases, *ETV6* in 6% (5/83) of cases, and *BTG1* in 4% (3/83) of cases.

Less well described and potential novel driver genes within breakpoints of deletions included *COL11A1* in 13% (11/83) and *MEF2C* in 10% (8/83) of samples. Additionally, *MBNL1* deletions were present in 8% (7/83) of patients, all but one in *BCR-ABL1+* cases. Small intragenic deletions in *LEMD3* were identified in 6% (5/83) of cases. *PTEN*, *NF1* and *KDM6A* are known tumour suppressor genes and were targeted by small intragenic deletions in 6% (5/83), 5% (4/83) and 5% (4/83) of cases respectively. *RAG1* and *CD200* deletions are described in BCP-ALL and were seen in 5% (4/83) and 4% (3/83) of cases respectively. Similarly, *ADD3* and *TOX* deletions were each seen in 5% (4/83) of cases, predominantly *BCR-ABL1+* patients.

4.4.3 MLPA validation

Of the 83 patients in the SNP array cohort, MLPA was performed on 67 cases. The MLPA-IKZF1-P335 kit contains probes covering *EBF1*, *IKZF1*, *CDKN2A*, *CDKN2B*, *PAX5*, *ETV6*, *BTG1*, *RB1* and PAR1. No PAR1 deletions were detected either by MLPA or SNP array so this abnormality was not considered further. The frequency of individual CNAs in the other genes by MLPA is shown in figure 4.13.

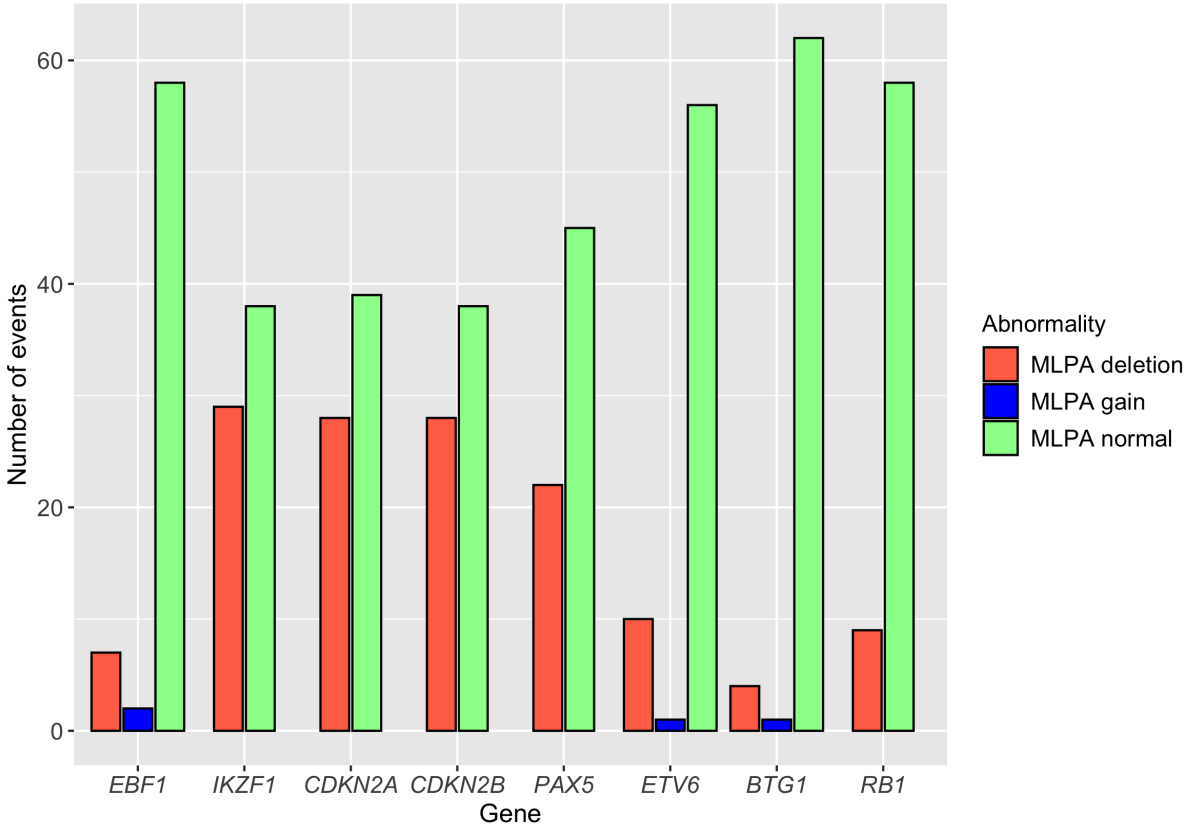


Figure 4.13. CNAs detected by multiplex ligation-dependent probe amplification (MLPA) in 67 patient samples. Deletions were present in *IKZF1* in 43% (n=29), *CDKN2A* in 42% (n=28), *CDKN2B* in 42% (n=28), *PAX5* in 33% (n=22), *ETV6* in 15% (n=10), *RB1* in 13% (n=9), *EBF1* in 10% (n=7) and *BTG1* in 6% (n=4). No PAR1 deletions were detected (data not shown).

Next, the MLPA results were compared to the SNP array-generated copy number abnormality calls in these same genes using the customised Nexus segmentation settings described previously (section 4.3.1).

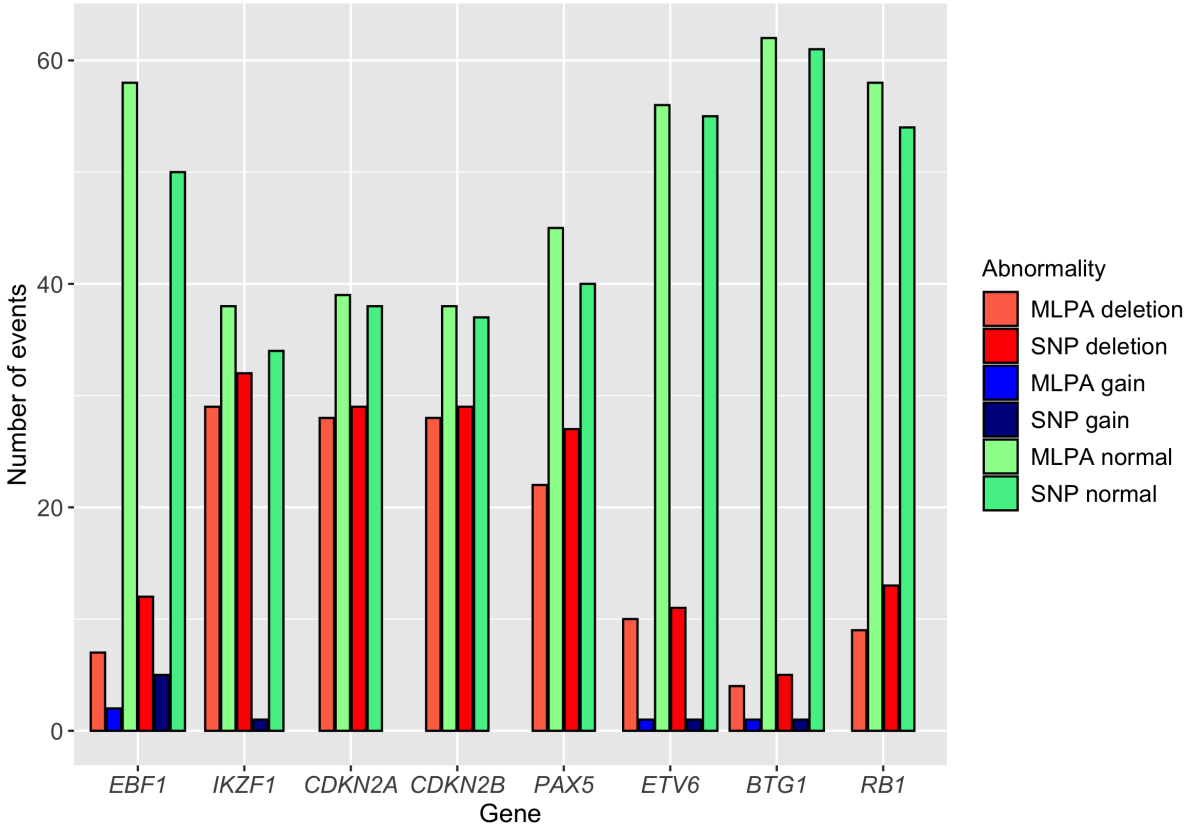


Figure 4.14. Comparison between CNAs detected by multiplex ligation-dependent probe amplification (MLPA) and SNP array in 67 patient samples. CNAs were called by SNP array more frequently than by MLPA across all 8 genes.

In total, 535 copy number calls of these genes were made by SNP array and MLPA. Slightly more CNA calls were made by SNP array than by MLPA (figure 4.14) across all 8 genes. Overall, 94% (501/535) were concordant, and the remaining 6% (34/535) were discordant. Of these, 14 related to low hypodiploid/near triploid samples. MLPA is known to be unsuitable in these samples (Schwab et al., 2010), so these were not considered further, leaving 20 discrepancies. Further details of discordant calls are shown in table 4.7.

Patient ID	Cytogenetics	Gene	MLPA result	SNP array result	Probable reason for discrepancy
25082	*BCR-ABL1*	*EBF1*	normal	**deletion**	Low VAF or blast %
26971	B-other	*EBF1*	normal	**gain**	Low VAF or blast %
28350	*BCR-ABL1*	*EBF1*	normal	**gain**	Low VAF or blast %
25246	B-other	*IKZF1*	normal	**deletion**	Low VAF or blast %
28581	*KMT2A* fusion	*CDKN2B*	normal	**deletion**	Low VAF or blast %
24813	Not done	*PAX5*	normal	**deletion**	Low VAF or blast %
25130	B-other	*PAX5*	normal	**deletion**	Low VAF or blast %
27509	*BCR-ABL1*	*PAX5*	normal	**deletion**	Low VAF or blast %
25246	B-other	*RB1*	normal	**deletion**	Low VAF or blast %
24890	B-other	*EBF1*	normal	**deletion**	MLPA probe location
27754	*BCR-ABL1*	*EBF1*	normal	**deletion**	MLPA probe location
29089	*BCR-ABL1*	*EBF1*	normal	**deletion**	MLPA probe location
24890	B-other	*CDKN2A*	normal	**deletion**	MLPA probe location
25552	B-other	*ETV6*	normal	**deletion**	MLPA probe location
27026*	*BCR-ABL1*	*EBF1*	normal	**gain**	Aneuploidy
27026*	*BCR-ABL1*	*ETV6*	normal	**gain**	Aneuploidy
27026*	*BCR-ABL1*	*BTG1*	normal	**gain**	Aneuploidy
26610†	*BCR-ABL1*	*IKZF1*	**deletion**	**gain**	Deletion and gain
27836	*BCR-ABL1*	*CDKN2A*	**deletion**	normal	Uncertain
27836	*BCR-ABL1*	*CDKN2B*	**deletion**	normal	Uncertain

Table 4.7. Details of 20 discordant SNP array-MLPA copy number calls. Majority of discrepancies were due to a CNA called by SNP array not detected by MLPA. *Case 27026 had *BCR-ABL1* and significant aneuploidy with 67-73 chromosomes on diagnostic karyotype. †Case 26610 had both gain and deletion in *IKZF1* on SNP array.
B-other: BCP-ALL with no primary chromosomal abnormality identified; VAF: variant allele frequency

Of the 20 discordant calls, the majority (90%, 18) were related to an abnormality being called by SNP array that was not identified by MLPA. Nine calls related to CNAs that did not reach the limit of detection for MLPA, although some reduction in the MLPA probe ratio could be appreciated in the majority of cases (data not shown). This indicated that using the customised Nexus segmentation settings as described above, the SNP array appeared to have a higher sensitivity for detecting abnormalities at lower variant allele fractions (VAF) – either due to subclonality or admixed non-leukaemic DNA – than MLPA. In five cases, the location of the MLPA probes meant the

abnormality could not be detected using this method as the CNA did not encompass the required minimum of 2 probes. Three of these related to *EBF1*, which only has MLPA probes covering exons 1, 10, 14 and 16 leaving a substantial gap between exons 1 and 10.

SNP array log2 ratio and B-allele frequency traces were carefully examined for all other discordant cases, where the discrepancy could not be attributed to large scale ploidy shift, and the SNP-array derived call could be confirmed visually in all cases. A single case (27836) harboured a *CDKN2A* and *CDKN2B* deletion on MLPA, which could not be confirmed by SNP array, either visually or through the segmentation algorithm, and this discrepancy therefore remains unexplained.

Overall, the SNP array and MLPA results were concordant in the vast majority of cases. The discordant cases were related to a specific cytogenetic subtype (low hypodiploidy), where large scale ploidy shift and co-existence of duplicated clones often hamper the use of molecular techniques. Where CNAs were detected by SNP array but not by MLPA, visual inspection confirmed the SNP array-derived call in all cases, thereby validating the customised segmentation settings and most likely permitting the detection of potentially subclonal abnormalities.

4.4.4 SureSelect XT2 library prep

The SureSelect XT2 capture library was designed to cover 30 of the genomic regions of interest from the SNP array analysis (partly established from figure 4.12). These had been selected to validate as many novel gene deletions as possible in cases with remaining diagnostic material. Well-established driver genes were also included as internal controls to ensure these could also be adequately detected. In total, the RNA bait library was 2.85 MB and encompassed the regions detailed in table 4.8.

Cytoband	Gene	Genomic co-ordinates covering ≥1 breakpoint in all affected cases	Number of cases affected
7p12.2	*IKZF1*	chr7:50,343,669 - 50,472,809*	22
1p21.1	*COL11A1*	chr1:103,342,013 - 103,574,062*	12
9p13.2	*PAX5*	chr9:36,833,262 - 37,034,486*	10
9q34.12	*ABL1*	chr9:133,589,258 - 133,763,072*	9
5q14.3	*MEF2C*	chr5:88,013,965 - 88,199,932*	8
3q25.1-3q25.2	*MBNL1*	chr3:151,880,890 - 151,998,320	7
12p13.2	*ETV6*	chr12:11,802,778 - 12,048,346*	6
10q23.31	*PTEN*	chr10:89,622,860 - 89,731,697*	6
1q31.2	*CDC73*	chr1:193,071,243 - 193,132,293	5
17q11.2	*NF1*	chr17:29,421,935 - 29,709,144*	5
12q14.3	*LEMD3*	chr12:65,569,801 - 65,589,942	5
Xp11.3	*KDM6A*	chrX:44,732,411 - 44,971,867*	4
5q13.2	*NIPBL*	chr5:36,876,851 - 37,066,525*	4
11p12	*RAG1*	chr11:36,532,249 - 36,614,716*	4
6p22.1	*PGBD1*	chr6:28,310,000 - 28,333,000	3
18q11.1	*ROCK1*	chr18:18,526,785 - 18,708,000	3
2q22.3	*ZEB2*	chr2:145,092,000 - 145,180,000	3
9p24.2	*VLDLR*	chr9:2,230,000 - 2,480,000	3
18q21.2	*TCF4*	chr18:52,889,552 - 53,332,028*	3
2q22.1	*CXCR4*	chr2:136,860,000 - 136,886,000	2
12q12	*ARID2*	chr12:46,110,000 - 46,301,833	2
13q12.2	*FLT3*	chr13:28,577,401 - 28,674,739*	2
13q14.11	*DGKH*	chr13:42,614,162 - 42,830,726*	1
19p13.3	*TCF3*	chr19:1,609,279 - 1,652,614*	1
5q32	*CSF1R*	chr5:149,432,844 - 149,492,945*	0
1q25.2	*ABL2*	chr1:179,068,452 - 179,198,829*	0
Xp22.3/Yp11.2	PAR1	chrX:1,321,616 - 1,397,693*	0
5q32	*PDGFRB*	chr5:149,493,390 - 149,535,445*	0
9p24.1	*JAK2*	chr9:4,985,023 - 5,128,193*	0
1q22	*MEF2D*	chr1:156,433,503 - 156,470,644*	0

Table 4.8. Targets for validation using SureSelect XT2 custom target enrichment kit. *Indicates whole gene covered. PAR1 refers to the pseudoautosomal region on Xp22.3/Yp11.2, in this case the genomic region between *CRLF2* and *CSF2RA*. Breakpoints in this region may indicate *CRLF2* rearrangements.

In total, 28 patient samples harbouring a combination of these abnormalities, were identified. As detailed in section 4.3.4, 23 of these were successfully put through library prep and formed the NGS validation cohort (table 4.9).

Patient	Genetics	Gene deletions for validation
25267	B-other	*COL11A1*
26614	B-other	*RAG1, NF1*
28335	B-other	N/A
24890	B-other	*IKZF1, COL11A1*
25082	*BCR-ABL1*	*NIPBL, CDC73, NF1, MBNL1, PTEN*
25208	*BCR-ABL1*	*ABL1, IKZF1, LEMD3*
25247	*BCR-ABL1*	*IKZF1, COL11A1*
28057	*BCR-ABL1*	*PAX5, IKZF1, NF1, MBNL1*
28182	*BCR-ABL1*	*PGBD1, NIPBL, MEF2C, IKZF1, NF1, MBNL1, PTEN*
28350	*BCR-ABL1*	*ZEB2, VLDLR, RAG1, TCF4, MEF2C, CDC73, MBNL1, PTEN*
28670	*BCR-ABL1*	*CXCR4, PAX5, IKZF1, MBNL1, LEMD3*
26660	*BCR-ABL1*	*NIPBL, MEF2C, IKZF1, NF1, MBNL1, LEMD3, PTEN*
25967	Complex	*COL11A1*
25437	HoTr	*ETV6, KDM6A*
29407	HoTr	*KDM6A*
27833	*IGH-BCL2*	*PAX5*
25130	*IGH-CRLF2*	*CXCR4, RAG1, ETV6, PAX5, IKZF1, LEMD3*
25552	*IGH@*-r	*PGBD1, TCF4, ETV6, PAX5, IKZF1, LEMD3, COL11A1*
28581	*KMT2A-AFF1*	*COL11A1*
25100	*KMT2A-v*	*PTEN*
27642	T-ALL	*KDM6A*
24813	Unknown	*ARID2, CDC73, PTEN*
25451	*ZNF384-r*	*COL11A1*

Table 4.9. NGS validation cohort. CNAs for validation in all patient samples successfully put through SureSelect XT2 library prep and targeted NGS.
B-other: BCP-ALL with no primary chromosomal abnormality identified; Complex: 5 or more unrelated chromosomal abnormalities on karyotype; HoTr: low hypodiploidy/near triploidy; *IGH@*-r: *IGH@* rearrangement; *KMT2A*-v: *KMT2A* fusion with non-*AFF1* partner; *ZNF384*-r: *ZNF384* rearrangement.

4.4.5 Targeted Next Generation Sequencing overview

Quality control thresholds were passed for both sequencing runs with the percentage of bases ≥Q30 at 92.11% and 92.46% respectively. Sequencing depth of coverage for all samples can be found in supplementary table 5.

An overview of the validation results is shown in figure 4.15. Translocations as well as the deletions, could be detected, and are discussed separately in section 4.4.8.

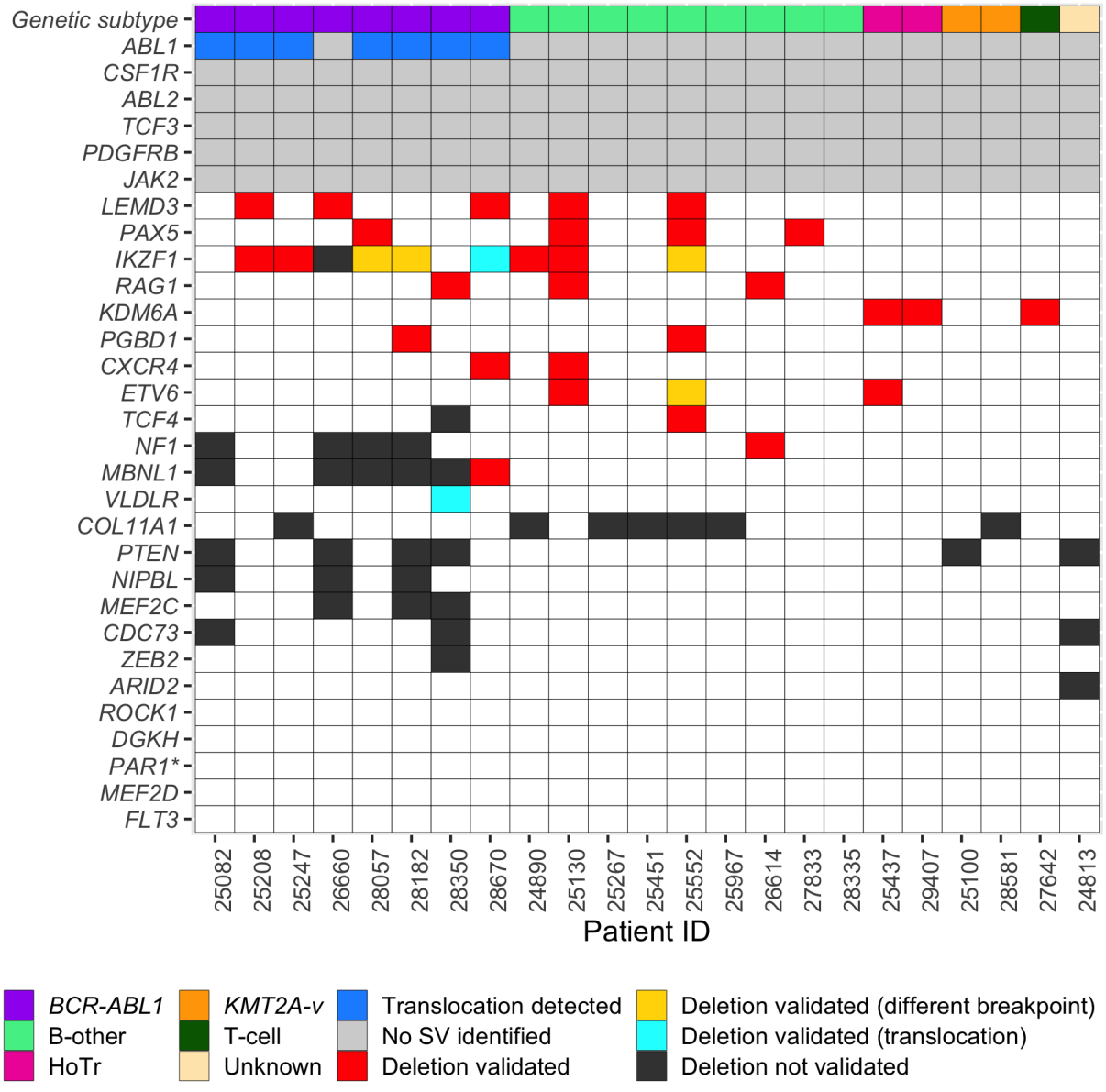


Figure 4.15. Validation of gene deletions and rearrangements. As well as the individual patient relevant CNAs detailed in table 4.9, all cases were examined for structural variants in *ABL1*, *CSF1R*, *ABL2*, *TCF3*, *PDGFRB* and *JAK2* as these are known recurrent primary abnormalities in ALL.
PAR1*: Pseudoautosomal region on Xp22.3/Yp11.2; B-other: BCP-ALL with no primary chromosomal abnormality identified; HoTr: low hypodiploidy/near triploidy; *KMT2A*-v: *KMT2A* fusion with any partner; T-cell: T-cell ALL; SV: structural variant.

In total, 49% (34/69) deletions were validated. All deletions affecting *LEMD3* (n=5), *PAX5* (n=4), *RAG1* (n=3), *KDM6A* (n=3), *PGBD1* (n=2), *CXCR4* (n=2) and *ETV6* (n=3) and 8 of the 9 *IKZF1* deletions were validated. In comparison, none of the suspected deletions in *COL11A1* (n=7), *PTEN* (n=6), *NIPBL* (n=3), *MEF2C* (n=3), *CDC73* (n=3), *ZEB2* (n=1) and *ARID2* (n=1) could be validated in this study.

4.4.6 Validated focal deletions by NGS

4.4.6.1 *IKZF1* deletions

In total, 9 cases included in the NGS validation had *IKZF1* deletions by SNP array and MLPA (table 4.10).

Patient ID	*IKZF1* deleted exons (SNP)	*IKZF1* deleted exons (MLPA)	*IKZF1* deleted exons (NGS)	Concordance
25247	ex. 4-7	ex. 4-7	ex. 4-7	Complete
24890	ex. 4-7	ex. 4-7	ex. 4-7	Complete
25130	ex. 2-8	ex. 2-8	ex. 2-8	Complete
25208	ex. 2-7	ex. 2-8	ex. 2-7	Minor discrp.
28057	ex. 1-7	ex. 4-7	ex. 4-7	Minor discrp.
28182	ex. 2-8	ex. 4-7	ex. 4-7	Minor discrp.
25552	ex. 2-7	ex. 4-7	ex. 4-7	Minor discrp.
26660	ex. 4-8	ex. 4-8	Not detected	Major discrp.
28670	ex. 2-7	ex. 2-7	Translocation	Major discrp.

Table 4.10. Details of *IKZF1* deletions in nine patients assessed by SNP array, MLPA and NGS with level of concordance/discrepancy in deletion breakpoints.
Minor discrp: minor discrepancy between SNP array, MLPA and NGS where deletion detected by all methods with some discrepancy in breakpoints; Major discrp: major discrepancy between SNP array, MLPA and NGS where deletion missed by one method or different event detected.

In 3/9 cases complete concordance was seen between SNP array, MLPA and NGS.

A further 3 cases showed minor discrepancies, where the SNP array derived deletion breakpoints differed slightly from the MLPA and NGS result (28057, 28182 and 25552). This could usually be explained by subtle discrepancies in SNP array segmentation either side of the highly repetitive region between exons 3 and 4. In one case (25208), there was a very minor discrepancy between the MLPA deletion compared with SNP array and NGS (single exon difference), although the reason for this was not clear.

In the remaining two cases (26660, 28670), more significant discrepancies were seen. One of these (26660) had an *IKZF1* deletion identified by SNP array and MLPA, which could not be detected in IGV. This sample also had a *BCR-ABL1* translocation which was not apparent in this NGS validation. Both SNP array and MLPA had yielded clear evidence of *IKZF1* loss and the bone marrow blast percentage was 84% (data not shown), rendering the possibility of discrepancy from low VAF less likely. Other considerations include uneven amplification during WGA resulting in under-representation of the relevant fragments or a sample mislabelling issue although the latter is less likely due to the successful confirmation of a *LEMD3* deletion in this case (figure 4.15). The final case (28670) had a deletion detected by SNP array and MLPA. However, in IGV, this abnormality was found to be a more complex unbalanced translocation, which could only be detected by an NGS approach (discussed later, section 4.4.8.2).

4.4.6.2 *LEMD3* deletions

Focal deletions in *LEMD3* on 12q14.3 were discovered in 6% (5/83) of the SNP array cohort. After *IKZF1*, this deletion was the most prevalent CNA in the validation cohort and was confirmed in all relevant cases (n=5). Primary chromosomal abnormalities were known in 4 of the five patients (table 4.11). Patient 25552 had an *IGH@* translocation but the partner had not been identified as fixed cells were not available. However, *CRLF2* is the most common *IGH@* rearrangement partner so it is plausible this case could also harbour *IGH-CRLF2*. Another observation was that all five cases had *IKZF1* deletions.

Patient ID	Sex (M/F)	Age (yrs)	Genetic subgroup	WCC at diagnosis ($x10^9$/L)
25208	M	62	*BCR-ABL1*	205.4
25130	F	62	*IGH-CRLF2*	33.6
28670	F	61	*BCR-ABL1*	1.6
26660	F	62	*BCR-ABL1*	18.2
25552	M	61	*IGH@* translocation	2.9

Table 4.11. Clinical and demographic details of cases with *LEMD3* deletions. Case 25552 did not have material for *IGH@* partner testing.
M: male; F: female; WCC: white cell count

A typical example of a focal *LEMD3* deletion is shown in figure 4.16.

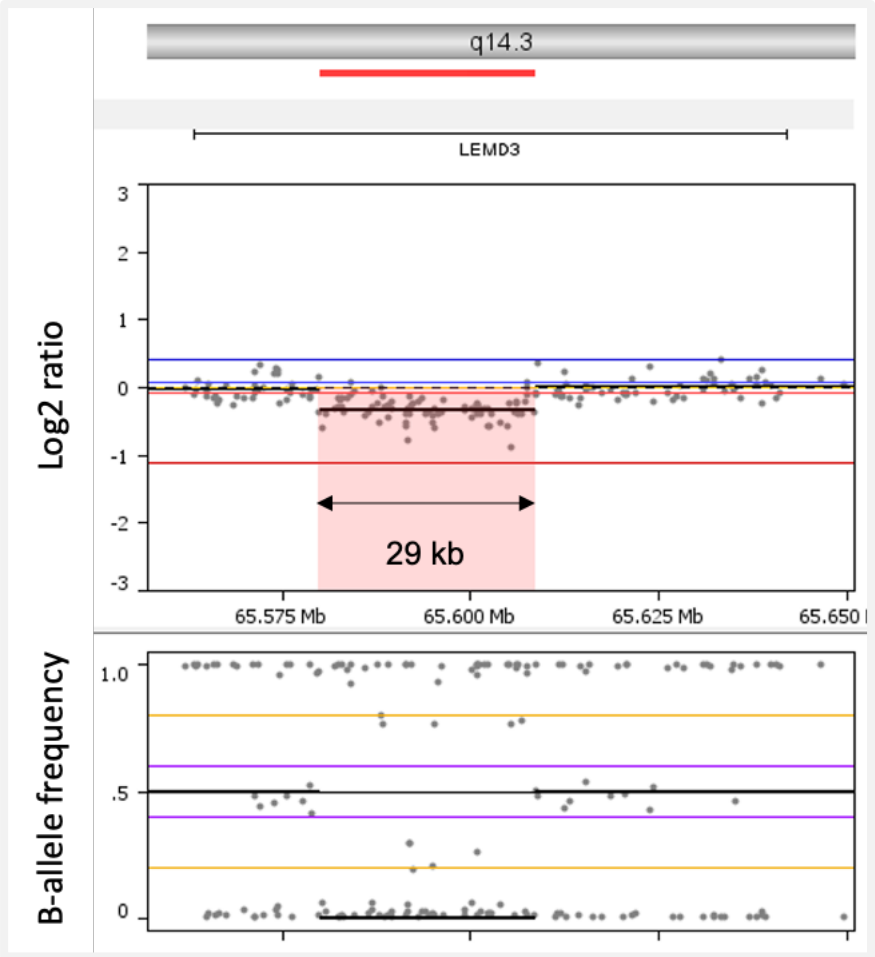


Figure 4.16.: Example focal *LEMD3* deletion as visualised in Nexus (patient 26660). Clear focal decrease in log2 ratio accompanied by loss of heterozygous AB alleles can be seen.

The deletion breakpoints matched very closely between SNP array and NGS, when visualised in IGV (figure 4.17), and the deletion was therefore clearly validated.

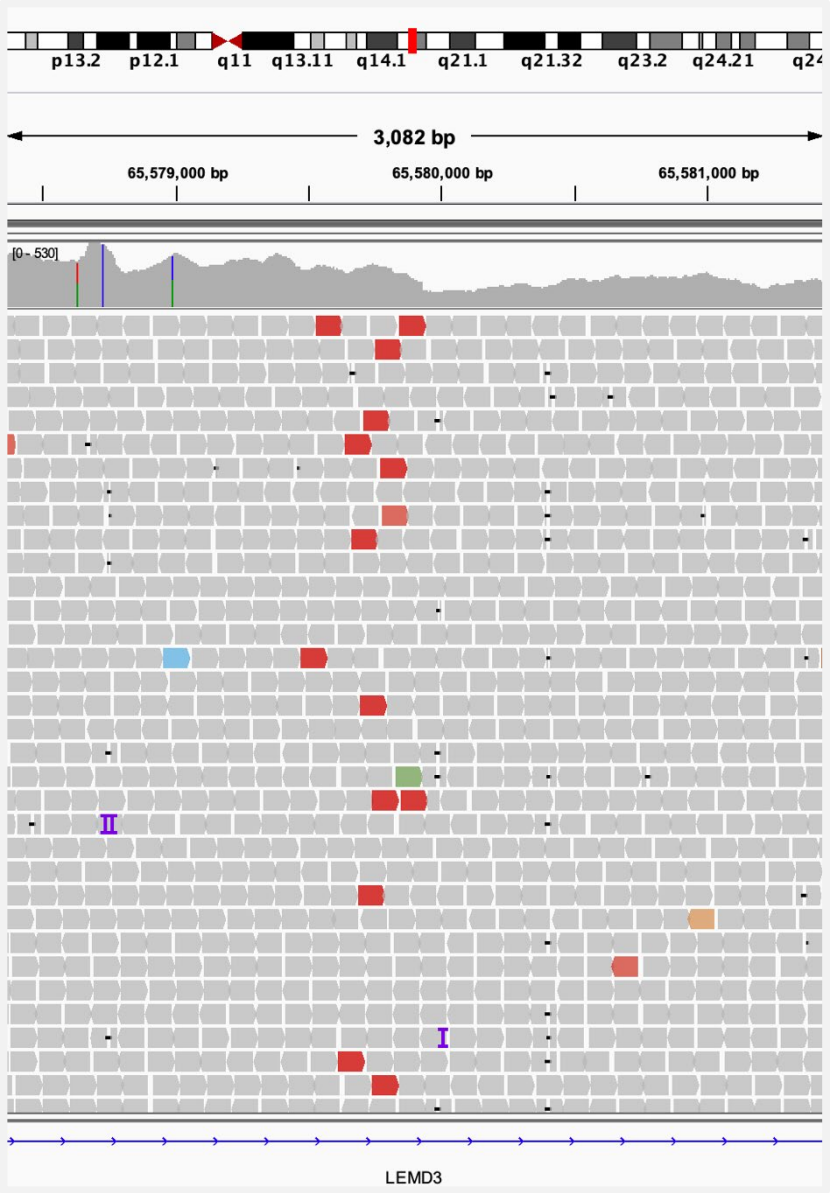


Figure 4.17. Example proximal deletion breakpoint in IGV (patient 25208) demonstrating reads with mate pairs mapping 21 kb away, consistent with ~21 kb deletion in *LEMD3*.

Patient ID	Deleted segment	Size of deletion (bp)	Deleted exons
25208	chr12:65,579,942-65,602,114	22,172	None
25130	chr12:65,579,942-65,597,922	17,980	None
28670	chr12:65,579,942-65,591,462	11,520	None
26660	chr12:65,579,942-65,608,678	28,736	ex. 2
25552	chr12:65,579,801-65,611,980	32,179	ex. 2-3

Table 4.12. Details of all *LEMD3* deletions in SNP array cohort, demonstrating highly consistent proximal breakpoint within intron 2.

The *LEMD3* deletions were focal intragenic events, ranging from 11-32 kb in size, and had highly consistent proximal breakpoints in intron 2. In three cases, the deletion was restricted to intron 2 and in the other two cases, at least one exon was also involved (table 4.12). Although the consequence of these particular intronic deletions cannot be predicted, copy number variation outside coding regions has been shown to impact splicing and gene expression (Rigau et al., 2019).

4.4.6.3 *RAG1* deletions

Focal deletions affecting *RAG1* on 11p12 were identified in 5% (4/83) of cases by SNP array. These were successfully validated in the three cases included in the NGS cohort (25130, 26614 and 28350).

Affected patients had primary chromosomal abnormalities as detailed in table 4.13. No fixed cells were available for patient 26614 for screening of B-other gene rearrangements so these abnormalities could not be excluded.

Patient ID	Sex (M/F)	Age (yrs)	Genetic subgroup	WCC at diagnosis ($x10^9$/L)
25130	F	62	*IGH-CRLF2*	33.6
26614	M	75	B-other	8
27754	F	63	*BCR-ABL1*	93.1
28350	F	62	*BCR-ABL1*	11.4

Table 4.13. Demographic and clinical details of cases with focal *RAG1* deletions.
M: male; F: female; WCC: white cell count; B-other: BCP-ALL with no primary chromosomal abnormality detected.

The focal *RAG1* deletions ranged from 14-56 kb and always affected exon 2. Three of the deletions also included the entirety of *RAG2* and the first 2 exons of *C11orf74* (figure 4.18 and table 4.14).

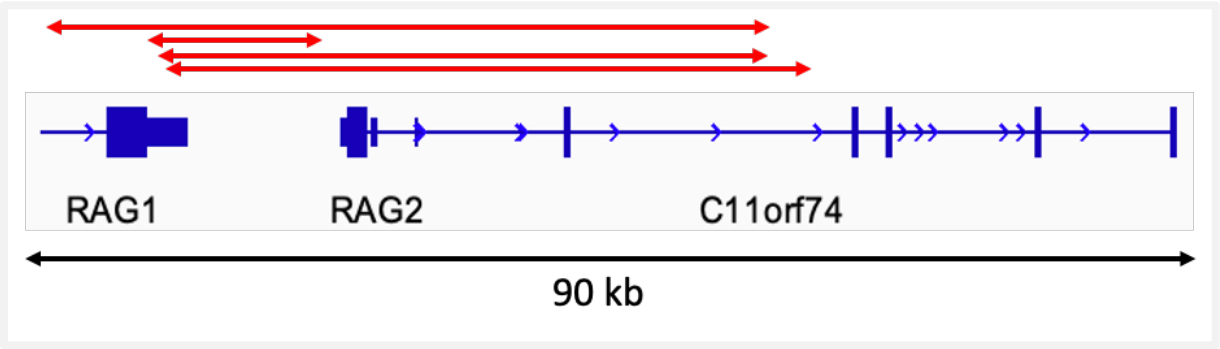


Figure 4.18. Approximate deletion breakpoints in *RAG1* and neighbouring genes in all four cases containing *RAG1* deletion by SNP array (25130, 26614, 27754 and 28350).

Patient ID	Deleted segment	Size of deletion (bp)	Genes affected
25130	chr11:36,590,053-36,642,721	52,668	*RAG1, RAG2, C11orf74*
26614	chr11:36,598,277-36,613,167	14,890	*RAG1*
27754	chr11:36,598,795-36,642,721	43,926	*RAG1, RAG2, C11orf74*
28350	chr11:36,599,564-36,656,180	56,616	*RAG1, RAG2, C11orf74*

Table 4.14. Details of deleted segments affecting *RAG1* and neighbouring genes.

4.4.6.4 *KDM6A* deletions

In total focal deletions involving *KDM6A* on Xp11.3 were discovered in 5% (4/83) of cases in the SNP array cohort. These were successfully validated in the three patients included in the NGS cohort (29407, 25437 and 27642). All events were focal intragenic deletions ranging from 56-316 kb in size.

Patient ID	Sex (M/F)	Age (yrs)	Genetic subgroup	WCC at diagnosis ($x10^9$/L)
28011	M	61	B-other	3.5
29407	F	60	HoTr	2.9
25437	F	64	HoTr	1.4
27642	F	72	T-ALL	Not known

Table 4.15. Demographic and clinical details of cases with focal *KDM6A* deletions.
M: male; F: female; WCC: white cell count; B-other: BCP-ALL with no primary chromosomal abnormality detected; HoTr: low hypodiploidy/near triploidy.

Two of the four patients had underlying low hypodiploid/near triploid ALL, and the white cell counts at diagnosis were noted to be low, although this is associated with the HoTr

phenotype. Three of the four patients had deletions affecting either both or their only *KDM6A* alleles (table 4.16).

Patient ID	Deleted segment	Size of deletion (bp)	Type of deletion	Exons deleted
28011	chrX:44,810,083-44,867,059	56,967	Hemizygous	ex. 3-4
29407	chrX:44,778,209-44,905,069	126,860	Homozygous	ex. 3-8
25437	chrX:44,775,342-44,885,557	110,215	Homozygous	ex. 3-6
27642	chrX:44,860,967-45,176,870	315,903	Heterozygous	ex. 5-29

Table 4.16. Details of deleted segments affecting *KDM6A*. SNP array from 29407 had heterozygous loss of chromosome X and focal homozygous deletion in *KDM6A*. SNP array from 25437 had a nested biallelic deletion in *KDM6A* with retained disomic complement of X chromosomes.

The affected male patient had a deletion affecting the only *KDM6A* allele as *KDM6A* is present on the X-chromosome. Interestingly, biallelic *KDM6A* deletions were seen in the two female patients with low hypodiploid ALL, albeit by two different mechanisms. By cytogenetics and SNP array, patient 29407's leukaemic blasts had lost one copy of chromosome X. SNP array and NGS confirmed the remaining homologue had a focal deletion within *KDM6A*. In comparison, in patient 25437, both copies of chromosome X had focal but subtly distinct intragenic *KDM6A* microdeletions (figure 4.19).

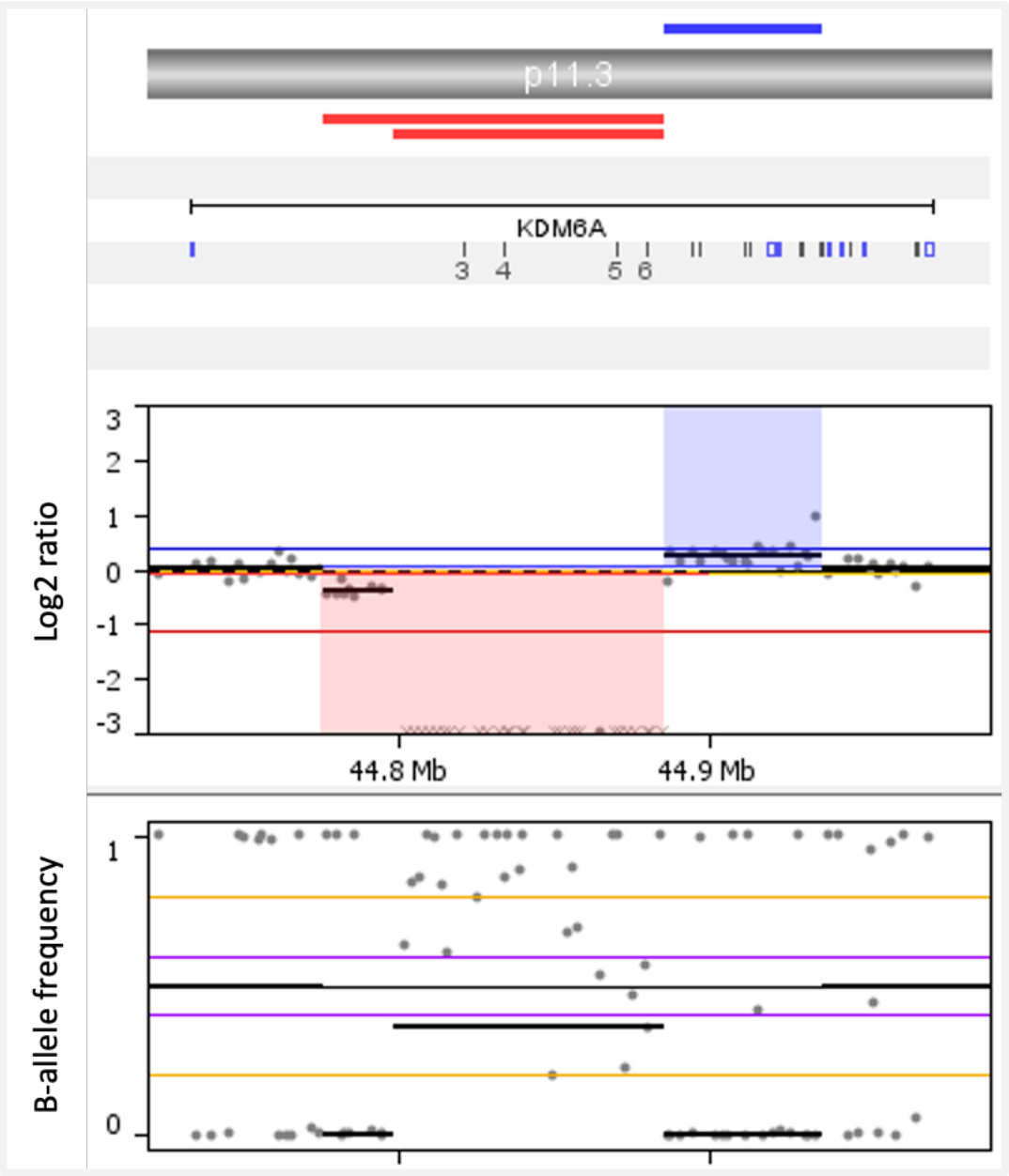


Figure 4.19. Homozygous *KDM6A* deletion in case 25437, demonstrating two slightly distinct *KDM6A* deletions measuring 110kb and 87kb respectively and resulting in biallelic loss of exons 3-6. Small gain noted following segment of homozygous deletion. Both deletions were validated by NGS in IGV.

4.4.6.5 6p22.1 deletions

Deletions encompassing three genes on 6p22.1 – *PGBD1, ZSCAN31 and ZKSCAN3* – were identified in 2% (2/83) of the SNP array cohort and were both validated by NGS. Further details of affected patients are shown in table 4.17.

Patient ID	Sex (M/F)	Age (yrs)	Genetic subgroup	WCC at diagnosis ($x10^9$/L)
25552	M	61	*IGH@-r*	2.9
28182	F	60	*BCR-ABL1*	118.3

Table 4.17. Clinical and demographic details of cases with focal deletions on 6p22.1.
M: male; F: female; WCC: white cell count; *IGH*@-r: *IGH*@ rearrangement.

Interestingly, the deletions had identical breakpoints by SNP array (table 4.18). A common feature in both cases was also the presence of 1q gain.

Patient ID	Deleted segment	Size of deletion (bp)	Genes affected
25552	chr6:28,246,968-28,323,426	76,459	*PGBD1, ZSCAN31, ZKSCAN3*
28182	chr6:28,246,968-28,323,426	76,459	*PGBD1, ZSCAN31, ZKSCAN3*

Table 4.18. Details of deleted 6p22.1 segments and affected genes

4.4.6.6 *CXCR4* deletions

Deletions affecting *CXCR4* on 2q22.1 were seen in 2% (2/83) of the SNP array cohort and were confirmed by NGS in both patients. Further details of affected patients are shown in table 4.19.

Patient ID	Sex (M/F)	Age (yrs)	Genetic subgroup	WCC at diagnosis ($x10^9$/L)
28670	F	61	*BCR-ABL1*	1.6
25130	F	62	*IGH-CRLF2*	33.6

Table 4.19. Demographic and clinical details of cases with focal *CXCR4* deletion.
M: male; F: female; WCC: white cell count.

Both deletions had very similar breakpoints (table 4.20), resulting in loss of exon 1 of the gene only, with no other neighbouring genes affected.

Patient ID	Deleted segment	Size of deletion (bp)	Deleted exons
28670	chr2:136,874,225-137,067,525	193,301	ex. 1
25130	chr2:136,874,992-137,067,525	192,534	ex. 1

Table 4.20. Details of *CXCR4* deletions from SNP array data, validated by NGS in IGV.

4.4.6.7 Deletions that were not validated

None of the *COL11A1* (n=7), *PTEN* (n=6), *NIPBL* (n=3), *MEF2C* (n=3), *CDC73* (n=3), *ZEB2* (n=1) or *ARID2* (n=1) deletions were validated by the targeted NGS analysis. Although these genes were included in the validation experiment due to their biological significance in malignancy (e.g. *PTEN*) or their high frequency of suspected copy number change in the SNP array cohort (e.g. *COL11A1*), the deleted segments were ambiguous by standard visual analysis of the relevant SNP arrays. Similarly, 1/2 *TCF4*, 1/5 *NF1* and 1/6 *MBNL1* deletions were confirmed by NGS. As with the other unvalidated CNAs, these were very equivocal findings on visual assessment of the relevant SNP arrays so were therefore likely to represent false positive calls.

4.4.7 Complete cohort copy number profile

The summary copy number profile of the complete cohort is outlined in figure 4.20. Any deletions included in the NGS experiment that could not be validated have been excluded from these data.

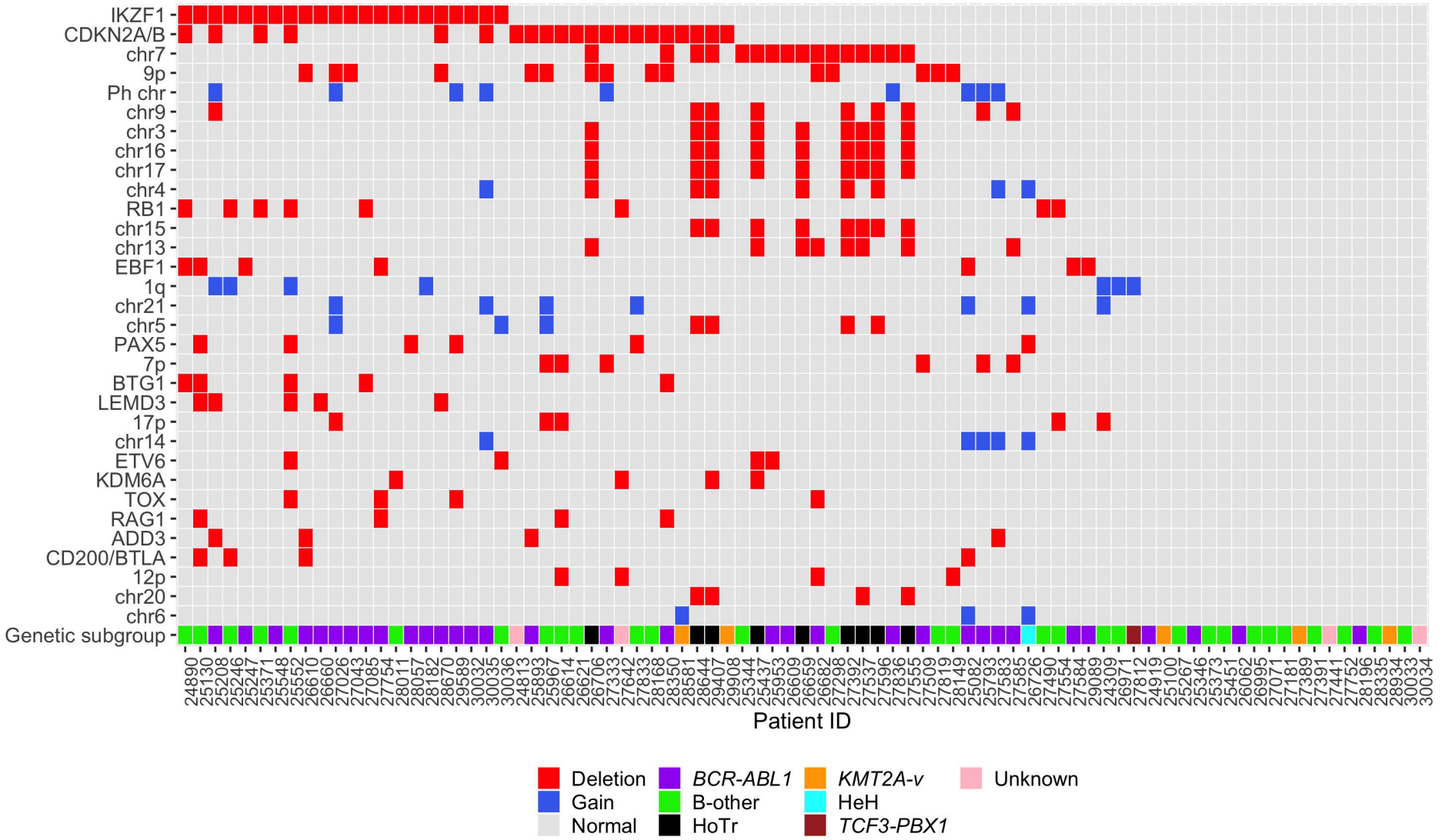


Figure 4.20. Copy number profile of 83 older adults with ALL. Focal and arm-level events shown. Only abnormalities that were detected in at least 3 cases are displayed.
Ph chr: Philadelphia chromosome; B-other: BCP-ALL with no primary chromosomal abnormality identified; HoTr: low hypodiploidy/near triploidy; *KMT2A*-v: *KMT2A* fusion with any partner; HeH: high hyperdiploidy

4.4.8 Translocations

Paired end sequencing also permits the discovery of chromosomal translocations when reads cover at least one of the translocation breakpoints. Mate pairs would usually be expected to map to genomic regions as far apart as the insert size (~1 kb in this experiment). However, when they map to regions on different chromosomes, a chromosomal translocation is suspected, and can be confirmed when significant concordance is seen in multiple reads.

4.4.8.1 *ABL1* translocations

Consistent with their primary genetic abnormality, clear breakpoints in *ABL1* were identified in 7/8 of the *BCR-ABL1+* cases. Mate pairs mapped to *BCR* thereby confirming the *BCR-ABL1* translocation. However, translocation breakpoints could not be identified in either *ABL1* or *BCR* in case 26660, although the reasons for this are not clear (discussed in section 4.4.6.1). Approximate breakpoints for all the other *BCR-ABL1* cases are shown in figure 4.21.

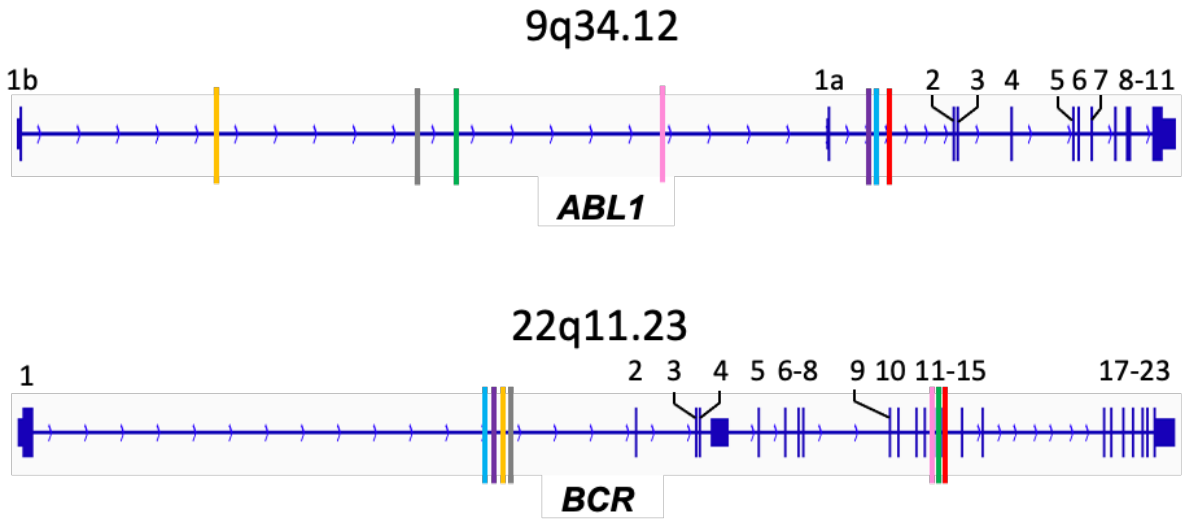


Figure 4.21. Approximate *ABL1* and *BCR* breakpoints in 7/8 *BCR-ABL1+* cases. Each coloured line represents breakpoint in *ABL1* and *BCR* for a specific case. Numbers represent individual exons.

The *BCR* breakpoints were clustered in two regions, specifically intron 1 in four cases (25082, 28182, 28350, 28670) and intron 14 in three cases (25208, 25247, 28057). These are consistent with the minor (p190) and major (p210) breakpoint cluster regions respectively. *ABL1* breakpoints were found between exons 1b and 1a (cases 25082,

25208, 28057, 28670) and 1a and 2 (25247, 28182, 28350), similar to previous studies (Score et al., 2010).

4.4.8.2 *IKZF1* translocation

An unusual *IKZF1* gene rearrangement was identified in the sample from patient 28670. The SNP array demonstrated a deletion affecting the majority of *IKZF1*. However, the telomeric portion of the gene seemed to harbour a homozygous deletion whereas the centromeric portion showed a more typical pattern of heterozygous loss (figure 4.22A). When analysed in IGV, several breakpoints were seen in *IKZF1* suggestive of a more complex structural variant, involving a pericentromeric region on chromosome 2, as well as two unmapped regions (figure 4.22B).

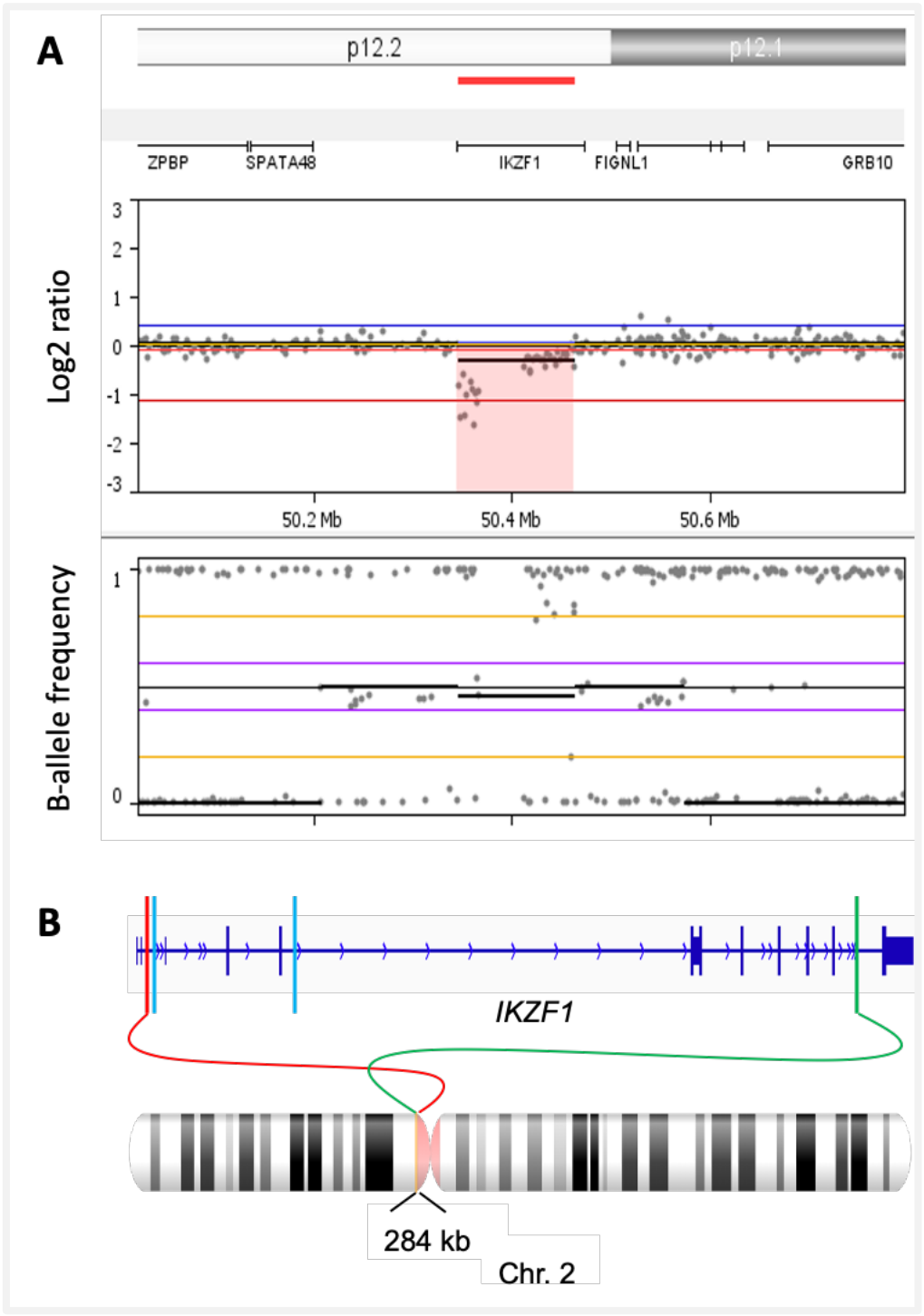


Figure 4.22. (A) SNP array demonstrating *IKZF1* deletion, which appears homozygous in telomeric and heterozygous in centromeric portion of gene. (B) *IKZF1* breakpoints identified in IGV, and mapping of mate pairs to pericentromeric region of chromosome 2 (red and green), together with additional breakpoints where mate pairs could not be mapped (blue). Findings are consistent with a complex rearrangement of *IKZF1*.

4.4.8.3 Translocation t(5;9)(q21.3;p24.2)

A translocation was also identified in patient 28350. Recurrent focal deletions had been noted between two genes on 9p24 – *VLDLR* and *SMARCA2* (figure 4.23) – and the

common breakpoint region was therefore included in the custom-designed SureSelect XT2 kit.

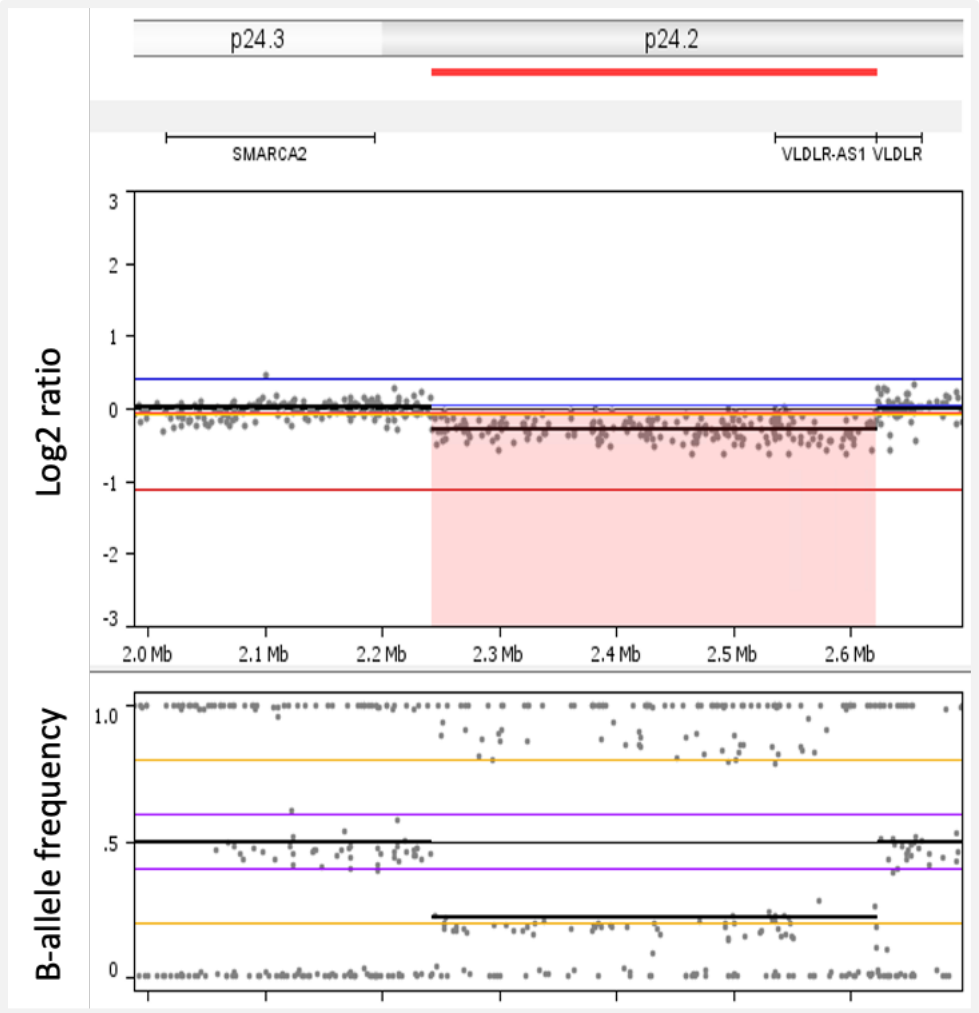


Figure 4.23. Deletion between *SMARCA2* and *VLDLR* (case 28350). In total, focal deletions in this region were seen in 4/83 cases by SNP array.

In the only case harbouring the event that could be included in the library prep (28350), the deletion was found to be part of an unbalanced translocation. However, no breakpoints were identified within genes so the consequences of this abnormality remain uncertain. *SMARCA2* is a member of the SWI/SNF chromatin remodelling complex, which is deregulated in a number of sold organ malignancies. *SMARCA2* mutations have been identified in several cases of AML with monosomy 7 (Eisfeld et al., 2017) and a translocation involving *SMARCA2*, resulting in a fusion gene, has been discovered in a single patient with MDS (Coccaro et al., 2018).

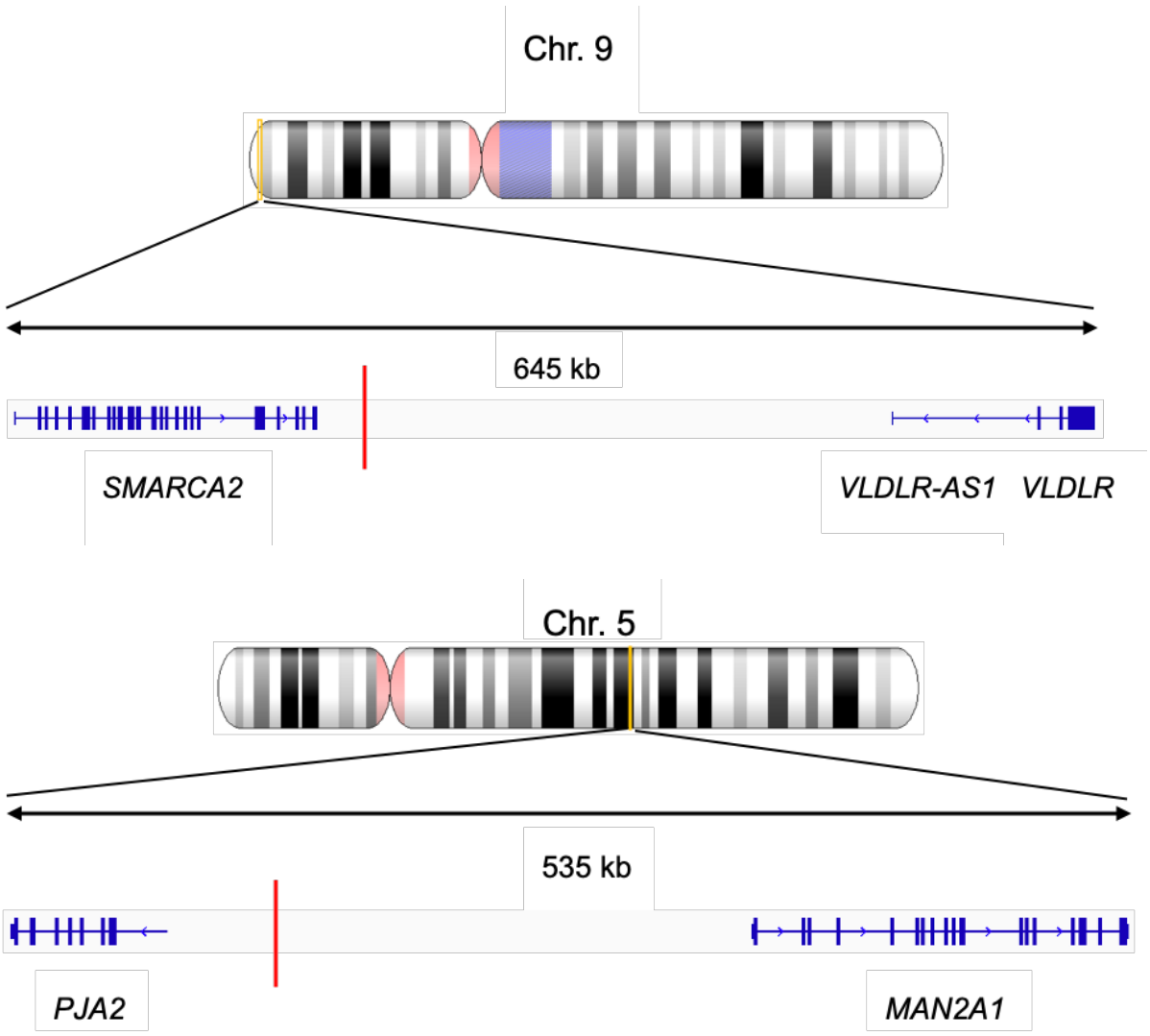


Figure 4.24. Approximate breakpoints (red lines) and neighbouring genes of *t(5;9)(q21.3;p24.2)* translocation in patient 28350. Genes closest to breakpoints were *SMARCA2* (~50kb from 9p24.2 breakpoint) and *PJA2* (~60kb from 5q21.3 breakpoint) on chromosomes 9 and 5 respectively. As the breakpoints were not in known genes or regulatory regions, the consequences of this abnormality could not be predicted.

4.5 Discussion

This study involved an in-depth copy number analysis of ALL genomes in 83 older adults, constituting one of the largest genomic analyses of ALL patients in this age group to date. Using carefully tailored SNP array segmentation settings in Nexus, consistent and biologically plausible results were seen across the cohort. A delicate balance needed to be achieved between the discovery of subtle and potentially novel abnormalities and minimising false positive calls. An initial validation between SNP

array and MLPA-derived CNAs in 8 genes confirmed the accuracy of the segmentation settings with 94% concordance between the two techniques. Discordant results were related to increased sensitivity of the customised SNP array segmentation to identify subclonal CNAs, suggesting that this achieved detection below 20% clonality, which is the limit of MLPA (Schwab et al., 2010).

Based on the validated SNP array segmentation and using the filtering steps described to exclude germline variants and minimise false positives, a mean of 10.1 deletions (range 1-51) and 3.1 gain (range 0-29) were seen per patient sample. These findings are similar to other copy number studies in adult ALL (Ribera et al., 2017) and slightly higher than those from childhood ALL cohorts (mean 6-8 CNAs per patient), although many of the latter were performed using older, lower density arrays making such comparisons difficult (Mullighan et al., 2007, Mullighan et al., 2009b).

CNAs were separated into arm/chromosome-level events and focal abnormalities. Of the former, deletions of 9p were the most frequent abnormality, seen in over 20% of cases. Monosomy 7, 7p deletion and 1q gain were also recurrent events, present in at least 10% of non-low hypodiploid cases. Focal abnormalities were considered separately, and analysis was focussed in identifying candidate “driver” genes.

4.5.1 Deletions in established driver genes

Recurrent deletions impacting known driver genes were identified affecting *IKZF1* in 52%, *CDKN2A/B* in 45%, *PAX5* in 39%, *RB1* in 22%, *ETV6* in 21%, *EBF1* in 19% and *BTG1* in 12% of patient samples. No PAR1 deletions, resulting in *P2RY8-CRLF2* fusion were seen, despite one patient having the abnormality detected by FISH (patient 25246). FISH testing had identified <25% rearranged cells in the sample, so this may have evaded detection by SNP array in this case. Overall, the frequency of deletions affecting many of these genes is higher than that observed in childhood and younger adult cohorts (table 4.21). It is noteworthy however, that *ERG* deletions were completely absent in this study. These are present in 3-4% of childhood BCP-ALL patients and are associated with an excellent prognosis, despite frequent co-existing *IKZF1* deletions (Clappier et al., 2014).

Age group	*IKZF1*	*CDKN2A/ CDKN2B*	*PAX5*	*RB1*	*ETV6*	*EBF1*	*BTG1*	*PAR1*
Older adults[1]	52%	45%	39%	22%	21%	19%	12%	0%
Young adults[2]	39%	37%	22%	7%	8%	3%	11%	1%
Paediatric[3]	14%	28%	19%	6%	22%	2%	6%	4%

Table 4.21. Comparison of specific deletions in established genes/regions in ALL patients in 3 age cohorts. [1]This study. [2]Analysis of 653 adults <65 years (Moorman et al., 2019). [3]Analysis of 1427 children with BCP-ALL (Schwab et al., 2013).

The high frequency of *IKZF1* deletions in this study is partly explained by the high proportion of *BCR-ABL1*+ patients. However, even when these are excluded, *IKZF1* deletions were still seen in 40% (20/50) of cases. Additionally, when low hypodiploid cases (n=9), which all had *IKZF1* deletions by default (due to monosomy 7), were also omitted, the findings still suggested double the *IKZF1* deletion rate compared to childhood cohorts (27% vs 14%).

IKZF1, *CKDN2A*, *CDKN2B* and *PAX5* deletions were the most frequent events in this cohort of 83 older adults. These genes are of particular interest due to their role in informing prognosis, as defined by the *IKZF1*^{plus} profile. This is based on the presence of *IKZF1* deletion co-occurring with deletions in *CKDN2A*, *CDKN2B*, *PAX5*, or PAR1, in the absence of an *ERG* deletion, and has been associated with a reduced event free survival of just 53% in paediatric ALL (Stanulla et al., 2018). This high-risk copy number profile, validated in childhood ALL, was identified in 36% (29/81) of the BCP-ALL samples in this cohort, compared with just 6% of paediatric cases in the original study. The *IKZF1*^{plus} profile correlated with primary cytogenetic abnormalities and was seen in *BCR-ABL1*+ (n=14), B-other (n=8) and low hypodiploid (n=7) patients.

To facilitate the detection of potentially novel abnormalities, genes at the breakpoints of all focal deletion segments were specifically scrutinised. This was based on the observation that “driver” CNAs are usually as small as is required to produce the survival benefit or transformative properties to the leukaemic cells, with the important exclusion of arm-level or whole chromosomal events. Hence, CNA breakpoints are recurrently observed within “driver” genes. This method appropriately captured the most common focal abnormalities (*IKZF1*, *CDKN2A/B*, *PAX5* deletions, etc), together with a number of potentially novel lesions, worthy of further investigation.

4.5.2 *LEMD3* deletions

LEMD3 deletions were recurrently observed, being present in 6% (5/83) of patient samples as a focal intragenic event, and subsequently validated by targeted NGS.

LEMD3 is a protein coding gene located on 12q14.3, containing 13 exons and measuring 79 kb. The gene encodes a nuclear envelope protein that regulates the bone morphogenic protein (BMP) and transforming growth factor β (TGF-β) pathways. TGF beta superfamily ligands, including BMPs, activate Smad proteins. These enter the nucleus and regulate gene transcription (figure 4.25). Activation of TGF-β signalling can produce both oncogenic and tumour suppressor activity in different circumstances and the pathway is known to be deregulated in a number of malignancies (Ikushima and Miyazono, 2010). Although pathogenic *LEMD3* mutations are the hallmark of Buschke-Ollendorff syndrome, a rare connective tissue disorder (Hellemans et al., 2004), a significant role in cancer has, to date, not been reported. The significance of these deletions therefore remains uncertain, particularly as 3/5 events did not include exons. However, interestingly, infrequent *LEMD3* mutations were identified in childhood BCP-ALL cases in a large pan-cancer analysis of paediatric tumours (Ma et al., 2018). A focal *LEMD3* deletion, present at both diagnosis and relapse, was also identified in a single patient in a study of 61 paediatric patients with relapsed ALL (Mullighan et al., 2008b).

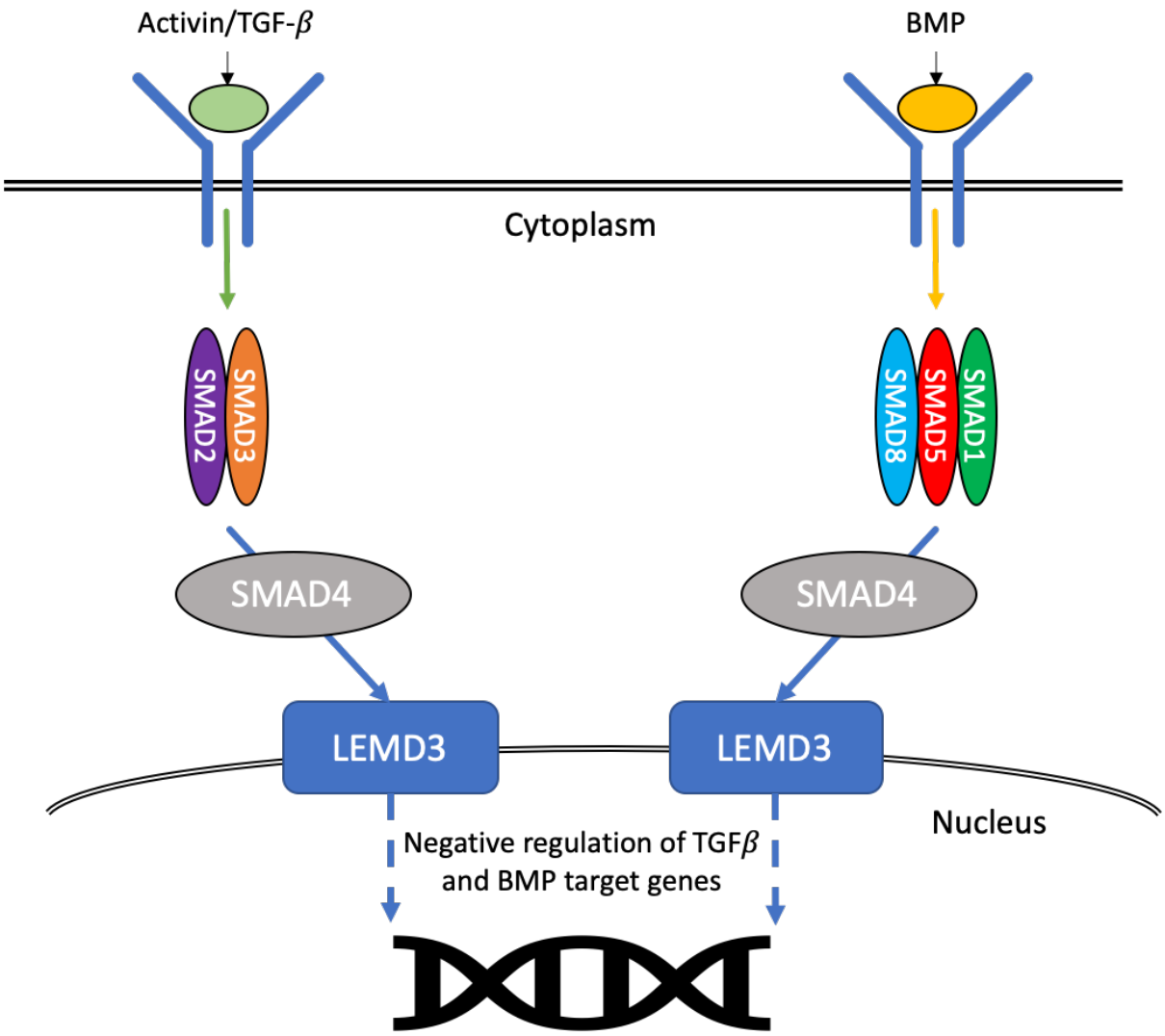


Figure 4.25. Role of LEMD3 (MAN1) in regulation of TGF-β and BMP signalling pathways. LEMD3 is a nuclear envelope protein, that downregulates transcription of TGF-β and BMP target genes. TGF-β: transforming growth factor beta; BMP: bone morphogenic protein.

In solid organ malignancies, opposing effects of increased TGF-β signalling are described depending on the stage of the disease. In early stages, TGF-β has a tumour suppressive effect by inhibiting cell cycle progression and proliferation, whereas in late stages, tumour promoting effects have been shown, by promoting tumour invasiveness and metastasis (Syed, 2016). TGF-β signalling also plays a role in normal haematopoiesis and haematological malignancies. Regulation of normal haematopoiesis is reliant of TGF-β signalling as an anti-proliferative and differentiation signal for haematopoietic progenitors, preventing cell cycle entry. Certain disease specific oncoproteins, such as RUNX1-ETO and PML-RARA in AML with t(8;21) and acute promyelocytic leukaemia (APL) respectively, produce anti-apoptotic effects and

cell cycle entry by repressing SMAD proteins and thereby reducing TGF-β signalling (Dong and Blobe, 2006). In these contexts, the TGF-β pathway appears to play a role in tumour suppressor function, with a disease-related reduction in TGF-β signalling producing the observed anti-apoptotic leukaemic properties. In comparison, loss of function genomic *LEMD3* abnormalities would be predicted to induce TGF-β signalling as *LEMD3* acts as a negative regulator of transcription of TGF-β target genes. Interestingly, certain haematological malignancies, specifically those associated with reactive bone marrow fibrosis (e.g. hairy cell leukaemia) have shown increased TGF-β signalling, which is postulated to be the cause of the fibrotic reaction (Shehata et al., 2004). At presentation, ALL may also be associated with reactive bone marrow fibrosis and although this could plausibly be induced by increased TGF-β signalling through *LEMD3* loss of function aberrations, the frequency of reactive fibrosis in ALL (up to 70% in childhood ALL) is far higher than the *LEMD3* abnormalities described (Wallis and Reid, 1989). As TGF-β signalling has a complex role in malignancy, including both pro- and anti-apoptotic properties, further work will be required to determine the consequences, if any, of the focal *LEMD3* deletions.

4.5.3 *KDM6A* deletions

KDM6A deletions were present in 5% (4/83) of the cohort and were biallelic events in 50% of affected cases. *KDM6A* (also known as *UTX*) plays a key role in epigenetic regulation. In eukaryotic cells, DNA is organised into nucleosomes, each of which comprises 147bp of DNA wound around an octamer of histone proteins. The N-terminal tails of the histone proteins can be subjected to post-translational modification, which then influences transcriptional regulation. This tightly regulated process can produce an effect on cell proliferation, survival and apoptosis. *KDM6A* encodes a histone H3 lysine 27 (H3K27) demethylase enzyme and a component of the mixed lineage leukaemia (MLL) 2/3 complexes that promote histone 3 lysine 4 (H3K4) methylation (Lee et al., 2007). H3K27 methylation correlates with genome silencing and repression of transcription, therefore, through H3K27 demethylation, KDM6A facilitates a transcription permissive chromatin state. H3K4 methylation is also a mark of open and actively transcribed chromatin (van Haaften et al., 2009, Ler et al., 2017). Through these functions, *KDM6A* antagonises EZH2/polycomb repressive complex-2 (PRC2) activity, which acts as an H3K27 methyltransferase and represses

transcription of its target genes. Inactivating *KDM6A* mutations therefore result in increased EZH2/PRC2 activity, and a sustained repression of the EZH2/PRC2 regulated genes. This aberrant EZH2 over-expression has been identified in a number of tumours and has been linked to increased tumour growth and metastatic potential (Gan et al., 2018).

KDM6A is located on Xp11.3 and is one of only a few genes that escape X inactivation in females (Dunford et al., 2017). To preserve physiological gene dosage, female eukaryotic cells inactivate one copy of their X chromosomes. However, a small number of tumour suppressor genes, including *KDM6A*, escape this inactivation and have been termed “escape from X-inactivation tumour suppressors” (EXITS). Hence, *KDM6A* has been described as a gender-specific tumour suppressor gene, whereby males, being hemizygous for *KDM6A*, can experience complete loss of *KDM6A* tumour suppressor function through abnormalities affecting the only allele. In comparison, females are relatively protected as biallelic inactivation is required (Greif et al., 2018). Indeed, this observation has been made in T-ALL patients whereby *KDM6A* mutations were exclusively observed in male patients and postulated to partly explain the male prevalence of this subtype (Van der Meulen et al., 2015).

KDM6A mutations have been identified in several tumour types, including up to 10% of myeloma cases (van Haaften et al., 2009). Inactivating *KDM6A* mutations have been reported, albeit infrequently in ALL (Li et al., 2018, Xiao et al., 2016, Mar et al., 2012). Additionally, these alterations appear to confer a survival advantage to leukaemic cells in AML, and have been specifically associated with mechanisms of cytarabine resistance (Greif et al., 2018, Stief et al., 2020), and a similar enrichment in relapsed disease has also been observed in ALL (Mar et al., 2014). To date, focal somatic CNAs, rather than sequence mutations, have only been described in very occasional ALL patients, specifically adults with relapsed disease (Ribera et al., 2017). In comparison, *KDM6A* is the second most frequently mutated gene in urothelial bladder cancer and studies have focussed on exploiting this therapeutically (Ler et al., 2017). As *KDM6A* inactivation produces uncontrolled EZH2 overexpression, inhibition of EZH2 has provided an indirect therapeutic target, which has been shown to reduce tumour size in *KDM6A*-mutated mouse and cell line models of urothelial bladder cancer. This could also provide a conceivable therapeutic avenue in *KDM6A*-deleted ALL patients and would be particularly valuable as half the affected cases had low hypodiploidy, representing a very high-risk subgroup with extremely poor prognosis.

4.5.4 *RAG1* deletions

RAG1 deletions were identified in 5% (4/83) of all cases, and were validated in all three cases included in the NGS cohort.

RAG1 is located on 11p12 and encodes the catalytic component of the RAG1/2 recombinase complex, which mediates DNA cleavage during immunoglobulin V(D)J recombination. The RAG complex preferentially binds to highly conserved recombination signal sequences (RSS), which lie adjacent to the V, D and J sequences, where RAG1 catalyses a double strand DNA break in the first stage of V(D)J recombination (Schatz and Ji, 2011). Specifically, RAG endonucleases bind DNA at RSS sites, which consist of a highly conserved heptamer (CACAGTG) followed by a 12bp or 23bp spacer sequence and a less well conserved nonamer (ACAAAAACC). Aberrant RAG activity has been identified as the key mechanism of focal deletions, including *IKZF1*, in *ETV6-RUNX1* ALL, whereby RSS-like sequences can be detected at deletion breakpoints (Papaemmanuil et al., 2014). *RAG1/2* deletions are described in BCP-ALL (Lilljebjörn et al., 2010, Mullighan et al., 2008b), and have been found to arise through this aberrant activity in the RAG endonucleases themselves (Papaemmanuil et al., 2014). Unlike other recurrent deletions, their consequences and prognostic impact are less well defined. Mouse models have shown that loss of both *CDKN2A* and *RAG1* results in a higher rate of transformation to BCP-ALL than when either gene is knocked down alone, indicating a potential leukaemogenic effect, although further data are lacking (Hauer et al., 2011).

4.5.5 *CXCR4* deletions

CXCR4 is a small 3.8 kb gene located at 2q22.1 encoding a chemokine receptor that is expressed on the surface of haematopoietic stem cells. CXCR4 is specific for stromal cell-derived factor-1 (CXCL12/SDF-1), which is secreted by bone marrow stromal cells. This interaction retains the *CXCR4* expressing cells in the bone marrow microenvironment. CXCR4 is also expressed to varying degrees on leukaemic blasts and high levels of CXCR4 expression have been shown to negatively impact treatment outcomes in both AML and ALL (Spoo et al., 2006, van den Berk et al., 2014) through a mechanism described as cell adhesion-mediated drug resistance (Damiano et al., 1999). Conversely, mouse models pharmacologically blocking CXCR4 expression,

thereby releasing leukaemic cells from the bone marrow niche, have demonstrated increased mobilisation of ALL blasts into peripheral blood (Welschinger et al., 2013) and associated increased chemo-responsiveness (Randhawa et al., 2016).

Germline mutations in *CXCR4* are recognised and produce an immunodeficiency syndrome – warts, hypogammaglobulinaemia, infections and myelokathesis (WHIM) syndrome – resulting from truncating gain of function mutations, which retain neutrophils in the bone marrow (Hernandez et al., 2003). Similar activating somatic mutations are observed in 25% of patients with Waldenstrom's macroglobulinaemia, a low grade lymphoproliferative disorder, and are associated with more aggressive disease (Treon et al., 2014).

Interestingly, focal *CXCR4* deletions are not described and hence represent a novel finding in ALL. Although most gene deletions result in reduced or absent protein functionality, the impact of these partial *CXCR4* deletions remains uncertain, particularly in the context of numerous gain of function mutations from truncation of the CXCR4 C-terminal (Poulain et al., 2016). In this study, the *CXCR4* deletions did not appear to have a consistent impact on the peripheral blood blast count at presentation (table 4.19), although the number of affected patients was very small (n=2).

4.5.6 6p22.1 deletions

Small 76kb deletions were identified in 2/83 patients, and encompassed 3 genes – *PGBD1*, *ZSCAN31* and *ZKSCAN3*. *PGBD1* is expressed in brain tissue although its function is not known. *ZSCAN31* and *ZKSCAN3* encode zinc finger transcription factors. *ZKSCAN3* has been found to modulate *Cyclin D2* expression in a myeloma cell line model (Yang et al., 2011). However, the gene was usually over-expressed through copy number gains rather than deleted. *ZKSCAN3* had additionally been identified as a master regulator of autophagy (Chauhan et al., 2013). Autophagy is a cellular process, through which damaged or defective cellular components are eliminated or recycled by lysosomes (Saftig and Klumperman, 2009). *ZKSCAN3* encodes a protein that functions as a repressor of autophagy and lysosome biogenesis and silencing of the gene has been shown to induce these processes (Chauhan et al., 2013). Overall, these three genes did not have a convincing biological role in ALL and were only present in a very small number of patients.

4.5.7 Conclusion

Copy number abnormalities were highly prevalent in older adults with ALL. Deletions affecting established recurrent driver genes were more frequent in older adults than in both younger adult and paediatric patients, with the notable exception of *ERG* and *ETV6* deletions, the latter being particularly strongly associated with *ETV6-RUNX1* fusion. Of the novel abnormalities, *KDM6A* deletions were the most biologically interesting and potentially therapeutically actionable discovery. Importantly, these were noted in patients with high risk disease, who are likely to derive the most benefit from novel treatment approaches and further screening and functional analysis of these lesions would be warranted.

Chapter 5. SNP array profiling of low hypodiploid acute lymphoblastic leukaemia

5.1 Introduction

In BCP-ALL, large non-random ploidy shifts define distinct primary genetic entities, such as high hyperdiploidy (51-67 chromosomes), near-haploidy (23-29 chromosomes) and low hypodiploidy (30-39 chromosomes) of the leukemic cells. High hyperdiploid clones typically have trisomies of chromosomes 4, 6, 10, 14, 17, 18, X and tetrasomy 21 (Heerema et al., 2007, Paulsson and Johansson, 2009, Paulsson et al., 2010). In low hypodiploidy, chromosomes, 3, 7, 15, 16, 17 are usually lost with the sex chromosomes (X and Y) and chromosome 21 always retained (Harrison et al., 2004). Low hypodiploid cells often additionally undergo chromosomal endo-reduplication without subsequent cytokinesis creating leukemic blasts with a near triploid karyotype of 60-78 chromosomes (Charrin et al., 2004, Safavi and Paulsson, 2017). This “doubling-up” phenomenon has been described in 40-65% of low hypodiploid patients and is also seen in 65% of near-haploid patients (Safavi and Paulsson, 2017, Charrin et al., 2004, Harrison et al., 2004, Pui et al., 2019). Low hypodiploidy is present in <1% childhood ALL and 3-4% of adult ALL overall; but increases in frequency with age and has been reported in >10% of patients over 60 years (see chapter 3) (Safavi and Paulsson, 2017, Moorman et al., 2010). Patients with low hypodiploidy have a very poor outcome compared to their age-matched counterparts: 5-year overall survival 0-20% in adults (Moorman et al., 2007, Safavi and Paulsson, 2017) and 35-50% in children (Safavi and Paulsson, 2017, Nachman et al., 2007, Pui et al., 2019). In comparison, high hyperdiploidy is the commonest genetic subtype of childhood BCP-ALL, seen in up to 35% of new diagnoses and was one of the first subgroups to be associated with a favourable prognosis (Lampert, 1967, Secker-Walker et al., 1978). With the use of modern protocols, childhood high

hyperdiploid ALL is now associated with an excellent 5 year overall survival >90% (Paulsson and Johansson, 2009). High hyperdiploidy is rarer in adults (5-10%) and while it appears to have a favourable outcome, the evidence is less robust with 5-year overall survival rates ranging from 33-88% (Brandwein et al., 2014, Pullarkat et al., 2008, Issa et al., 2017, Chilton et al., 2013).

Ninety percent of low hypodiploid ALL cases are additionally and very specifically associated with pathogenic *TP53* mutations (Holmfeldt et al., 2013, Mühlbacher et al., 2014). In childhood cohorts, almost half of such patients have also been shown to harbour these mutations in non-leukemic cells and childhood low hypodiploid ALL is therefore considered a manifestation of the Li Fraumeni Syndrome (Malkin, 2011). Adult patients with low hypodiploid ALL also exhibit a high frequency of *TP53* mutations, although these appear to be somatic in the majority of cases (Mühlbacher et al., 2014).

The accurate and timely detection of low hypodiploidy at the diagnosis of ALL is critical for the optimal management of patients. However, the large-scale chromosomal loss and the “doubling up” phenomenon pose significant diagnostic challenges. Although the pattern of chromosomal loss is non-random, significant variation is observed between patients, and the working definition of low hypodiploidy, and indeed high hyperdiploidy, is based on modal chromosome number rather than the loss of specific chromosomes. Thus, genome-wide techniques, such as cytogenetics and single nucleotide polymorphism (SNP) arrays, are preferred to focal testing, such as fluorescence *in situ* hybridisation (FISH). In cases where the near-triploid sub-clone is larger than the low hypodiploid clone, cytogenetic analysis frequently only detects the near-triploid sub-clone, which is referred to as “masked low hypodiploidy”. Importantly, the near-triploid clone, which will have between 60 and 78 chromosomes, can be mistaken for a high hyperdiploid clone due to the overlap in modal chromosome numbers between the subgroups, potentially resulting in a drastic impact on risk stratification and treatment protocol allocation (Carroll et al., 2019, Charrin et al., 2004).

5.2 Aims and objectives

Although the genetic subtypes of ALL characterised by large scale ploidy shifts are defined cytogenetically, in large part using the modal number of chromosomes in the leukaemic cells, SNP arrays provide complementary information, such as variation in copy number state and loss of heterozygosity, which help categorise ploidy subgroups. This chapter will focus on examining the use of SNP arrays in the accurate diagnosis of low hypodiploidy and high hyperdiploidy in ALL by addressing the following aims.

1. Defining the cytogenetic abnormalities in a cohort of adults and children with low hypodiploidy (including masked cases), and high hyperdiploidy.
2. Analysing SNP arrays on a large cohort of patients to identify discrepancies between the cytogenetics-derived and SNP-array derived subgroup.
3. Using TP53 sequencing to resolve discrepant cases.
4. Using hierarchical clustering and machine-learning techniques to develop a diagnostic classifier to accurately determine the ploidy subgroup using SNP array data in the absence of cytogenetics.

5.3 Methods

5.3.1 Patients and samples

The Leukaemia Research Cytogenetics Group (LRCG) database was first searched for all adults over 60 years of age at diagnosis enrolled in the UKALL14 and UKALL60+ trials. All patients who had been assigned a genetic subgroup of low hypodiploidy/near triploidy (HoTr) were then identified and SNP arrays were performed on all those with available DNA from diagnostic bone marrow samples. To create a comparison cohort of high hyperdiploid ALL samples, younger adult and paediatric ALL samples were also included due to the rarity of this subtype in older patients. A small number of younger patients with HoTr were also entered to optimise the size of the cohort.

5.3.2 SNP array analysis

Raw IDAT and CEL files were loaded onto Nexus as described previously (chapter 2 sections 2.6.1.4 – 2.6.1.5). Visual SNP array analysis was limited to the documentation of patterns consistent with specific ploidy groups by focussing on whole chromosomes rather than more focal abnormalities. Identifying loss of heterozygosity in multiple chromosomes at the lower copy number state (i.e. the lower log2 ratio) (hereafter termed LOH-LCN pattern) was considered consistent with HoTr. This pattern arises through an initial step of chromosomal loss (producing LOH and reduced log2 ratio). Even if the chromosomal complement then undergoes endoreduplication, chromosomes exhibiting LOH still have a lower log2 ratio than those with preserved heterozygosity. In comparison, a pattern of retained heterozygous SNPs in chromosomes at the lower copy state, combined with additional maternal and/or paternal homologues in chromosomes at the higher copy number states (indicative of chromosomal gains) (hereafter termed HET-CNG pattern) was considered consistent with HeH. An overview of these patterns is shown in figure 5.1.

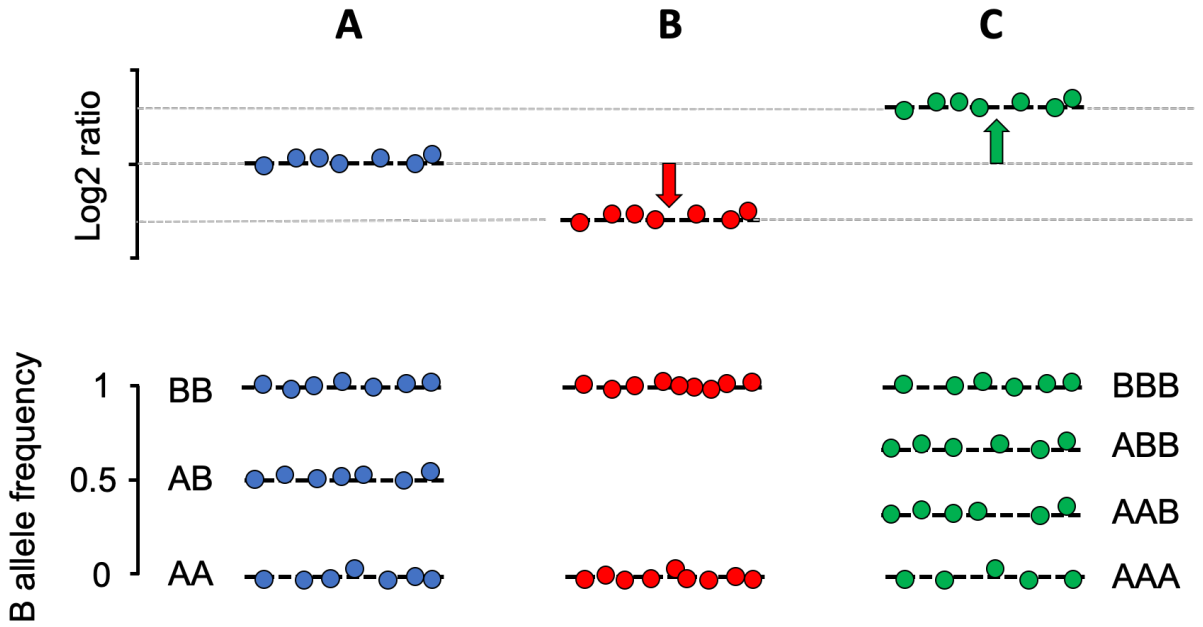


Figure 5.1. Overview of characteristic HoTr and HeH chromosomal patterns on SNP array. A demonstrates features of typical disomic chromosome with preserved heterozygous AB alleles. B demonstrates pattern arising through chromosomal loss. Log2 ratio reduces relative to heterodisomy (A) and heterozygous AB alleles are lost (LOH-LCN). Importantly, the pattern is constant even in the event of endoreduplication whereby A becomes tetrasomic and B becomes disomic with LOH. C demonstrates pattern arising through single copy gain. Log2 ratio rises relative to heterodisomy and additional allelic pattern seen on B-allele frequency due to gain of a single maternal or paternal homologue (HET-CNG).

Importantly, although HeH can be associated with whole chromosomal LOH, this does not occur as a result of initial chromosomal loss so affected chromosomes will show a similar log2 ratio to preserved heterodisomies. Using this approach, all cases were re-classified based on these SNP array patterns. Where neither pattern was apparent (including where the SNP array appeared largely normal), the case was classified as “inconclusive’.

To derive chromosome-level numerical values, without assumption of the modal chromosome number or ploidy status of each case, whole chromosome segments were manually created in Nexus and their log2 ratios extracted for further analyses

5.3.3 Creation of whole chromosome copy number segments in Nexus

Nexus segmentation algorithms were kept to default settings. Subsequently, new segments were manually created across the entire length of individual chromosomes. These overrode any existing segments. The log2 ratios of these new whole chromosome segments were automatically computed by Nexus and were exported into a single table incorporating all cases. To account for sex-related disparity in the log2 ratios of X and Y chromosomes, these were removed from all further downstream analysis.

The dataframe containing all whole chromosomal log2 ratios was then read into R (R Core Team, 2019) for subsequent clustering and classification analyses.

5.3.4 Per-sample standardisation of log2 ratios

The scale of positive and negative deflection in the log2 ratios of copy number segments is influenced by several factors. For example, CNAs in a pure sample will produce much greater positive or negative log2 ratio deflections than would be seen in a sample heavily contaminated with non-leukaemic DNA. Similarly, due to their different chemistries, Illumina and Affymetrix arrays produce different deflection in log2 ratios. To account for these variations between samples, whole chromosomal log2 ratios needed to be standardised to a consistent scale. Therefore, using R-package

BBmisc, the whole chromosomal log2 ratios within each sample were standardised to a mean of 0 and standard deviation of 1. These standardised whole chromosomal log2 ratios were then used for all subsequent analyses.

5.3.5 Principle components analysis (PCA) and unsupervised hierarchical clustering

A principle components analysis (PCA) involves reducing the dimensionality of a dataset by creating new uncorrelated variables, based upon the primary variables, in this case whole chromosome log2 ratios (Jolliffe and Cadima, 2016). Using R-package prcomp, a PCA was performed using the standardised whole chromosomal log2 ratios as input variables and visualised using R-package factoextra. Unsupervised hierarchical clustering was similarly performed using R-package ComplexHeatmap (Gu et al., 2016b). Individual cases were labelled according to their reported cytogenetic subgroup, which highlighted any cases that clustered unexpectedly. The SNP array patterns based on the visual SNP array analysis were then used to ascertain whether such cases had been incorrectly classified by cytogenetic analyses.

5.3.6 *TP53* sequencing

To add further confirmation where cases had a SNP array pattern of HoTr but had been cytogenetically classed as HeH, coding exons of *TP53* were sequenced in selected samples. This was achieved by including all exons of *TP53* in the custom-designed SureSelect XT2 target enrichment kit (Agilent, Santa Clara, CA, USA) (chapter 2, section 2.7 and supplementary table 4).

5.3.7 Classification and regression tree (CART) analysis

CART analysis is a supervised machine learning technique, whereby a dataset consisting of different classes is partitioned according to a set of variables. The variables that produce the clearest separation of the different classes are chosen by the model in a sequential manner to eventually obtain nodes that give the best

separation of the input classes. This method can therefore be used effectively to produce an algorithm (decision tree) to classify a new case based on the same variables.

To use whole chromosome log2 ratios to aid the accurate classification of ploidy status (HoTr, HeH or non-ploidy), a CART analysis was performed and a decision tree classifier was created. Initially, all HoTr and HeH cases in the cohort were re-classified based on the most probable genetic subgroup using all available data from cytogenetics, SNP arrays and *TP53* status. Where the SNP profile showed the clear LOH-LCN pattern described above, the case was categorised as HoTr, even if this conflicted with the initial cytogenetic subgroup. Similarly, if the SNP array demonstrated the HET-CNG pattern consistent with chromosomal gains, the case was classified as HeH. To create a non-ploidy group for the CART analysis, all SNP arrays lacking a major ploidy shift from the CNA analysis of older adults detailed in chapter 4 were added to this study. Whole chromosomal log2 ratios were created and standardised as detailed above.

Following any re-classification of HoTr or HeH discrepancies, the CART analysis was performed using R-package rpart (see supplementary table 8 for input dataset), and a decision tree was created based on the standardised whole chromosomal log2 ratios.

The performance of the CART model was then assessed using internal 10-fold cross-validation in R-package caret. This involved internal random partitioning of the dataset into 10 subsets. Each subset was held out in turn and the remainder of the dataset was used to train the decision tree. The accuracy of the decision tree in classifying cases in the held out subset was then assessed. As each subset was held out in turn, this generated 10 independent validations and the mean accuracy of the model was reported.

5.3.8 External validation of decision tree classifier

A cohort of 29 Affymetrix Cytoscan HD arrays was provided by the Children's Cancer Research Institute (CCRI) (Vienna, Austria) to perform external validation of the decision tree classifier. These comprised the following genetic subgroups: HeH (n=7), HoTr (n=7), near haploidy (n=8), *ETV6-RUNX1* (n=2), *TCF3-PBX1* (n=1), *KMT2A-AFF1* (n=1), B-other (n=3). Whole chromosomal log2 ratios were extracted by

collaborators at the CCRI and sent blinded to provide an external testing set for the classifier. The whole chromosomal log2 ratios were standardised using R-package BBmisc as detailed above and the resulting values were used to place each case in one of the decision tree's nodes. The genetic subgroups for each case were then unblinded by the CCRI group and compared to the output from the classifier.

5.4 Results

5.4.1 Patient demographics and cytogenetic analyses

The patient cohort comprised 88 patients with HoTr or HeH who were initially identified by cytogenetics (n=71) or SNP array (n=17). Cytogenetic results were available for 95% (n=84) of patients. Patient characteristics and detailed cytogenetic data can be found in supplementary tables 6 and 7. Centralised review of karyotype data and assignment of cytogenetic subgroups had been undertaken at the LRCG as previously described (Harrison et al., 2001). At the start of the study, 40 patients had been assigned a genetic subgroup of HeH and 48 patients HoTr (figure 5.1). The median age of the patients was 51 years (range 7-87) for the HoTr subgroup and 7 years (range 1-58) for the HeH subgroup; reflecting the disparate age-specific frequencies of the two subtypes (figure 5.2).

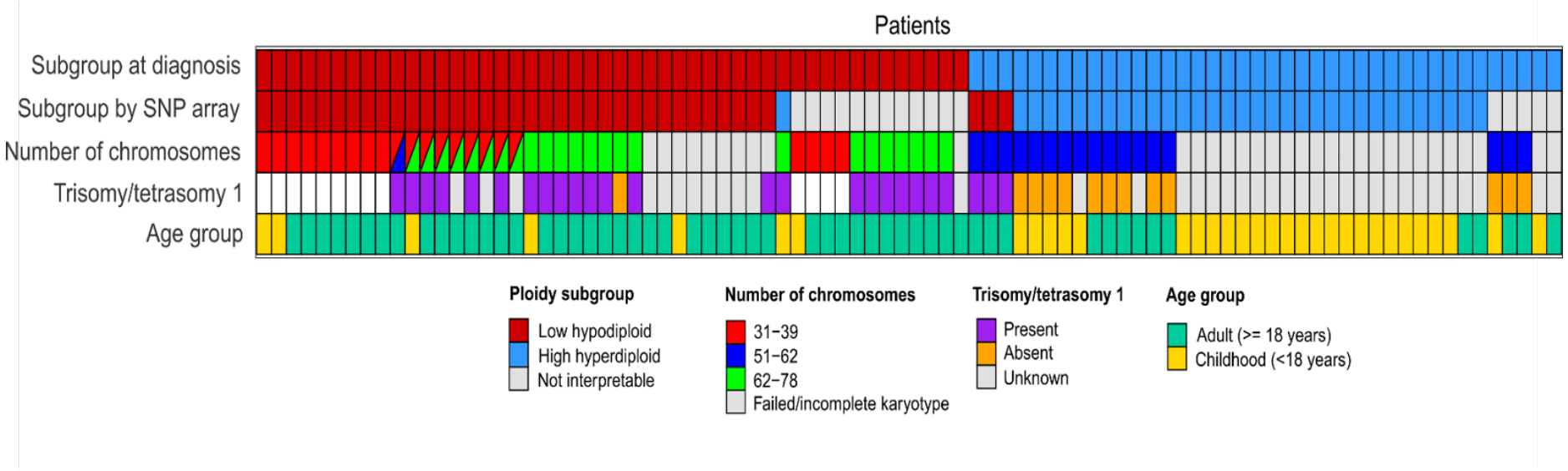


Figure 5.2. Patient demographics and cytogenetic characteristics. Patient samples were obtained from patients enrolled in UKALL14 (n=40), UKALL2011 (n=11), UKALL60+ (n=6), UKALLXII (n=6) and UKALL2003 (n=3) clinical trials as well as local non-trial cases (n=22).

All patients with complete cytogenetic analyses were first examined to document the specific whole chromosomal gains and losses that had occurred from the baseline diploid state. The chromosomal gains and losses reflected distinct patterns which are characteristic of these genetic entities (figures 5.3 and 5.4). Near triploidy is thought to arise through endoreduplication of a low hypodiploid clone, which would be expected to generate disomies and tetrasomies. It is therefore noteworthy that frequent trisomic chromosomes were also present in the near triploid clones (figure 5.4), suggesting that the process of endoreduplication is imperfect and is likely to be more complex than pure doubling of the chromosomal complement.

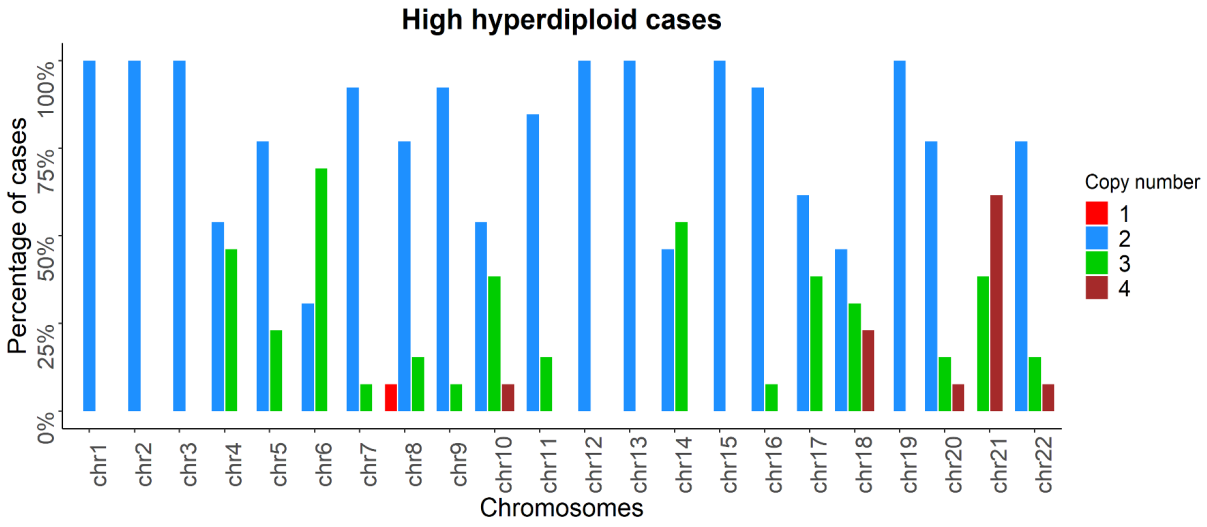


Figure 5.3. Summary of copy number state of individual chromosomes across 13 patients with high hyperdiploid ALL and a complete karyotype (3 cases with SNP arrays suggestive of masked low hypodiploidy excluded) demonstrating gains of chromosomes 4, 6, 10, 14, 17 and 21 in many cases.

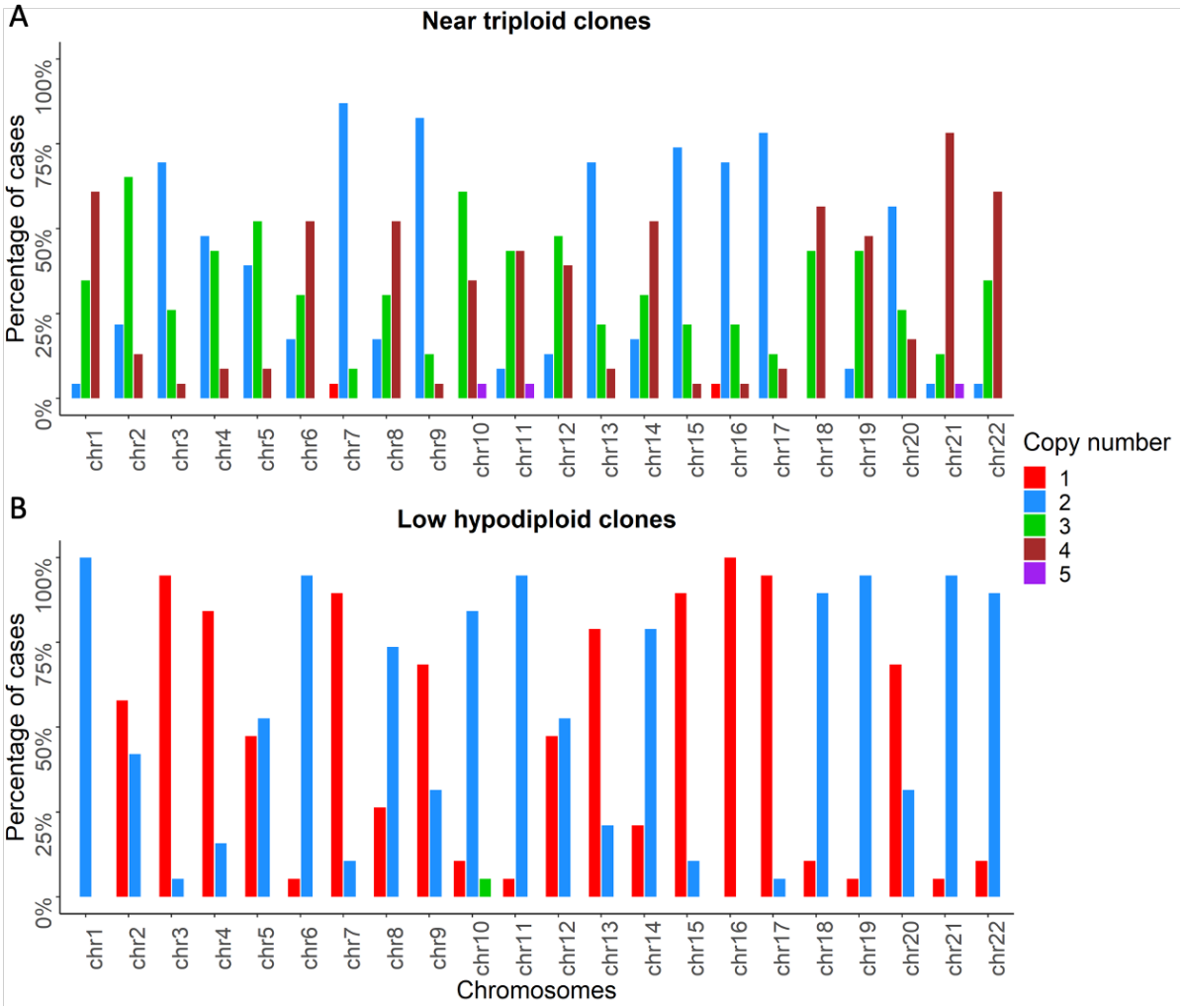
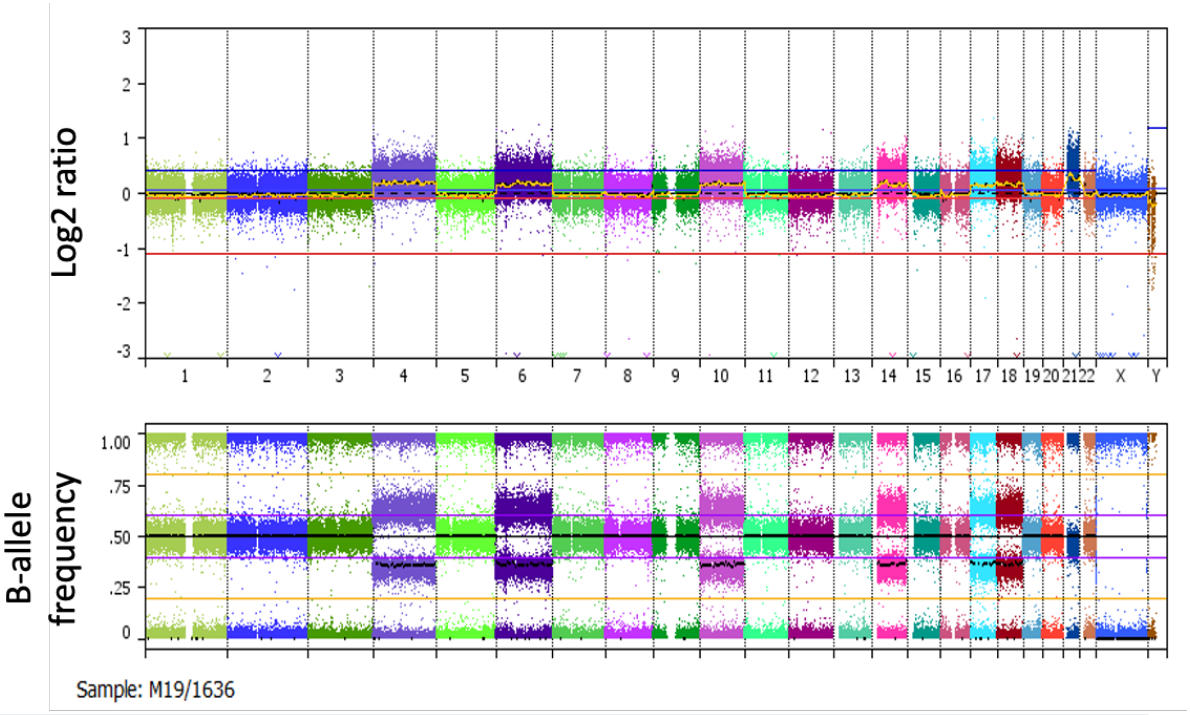


Figure 5.4. Copy number status of individual chromosomes as reported in karyotypes containing near triploid (A) and low hypodiploid (B) clones. Cases with discrepant cytogenetic and SNP array findings are not shown (n=2). Near triploid clones demonstrate distinct pattern from high hyperdiploidy with gains of chromosomes 1, 2, 4, 5, 10, 11, 12, 18, 19, 21 and 22 predominating. Disomic chromosomes in near triploid clones (A) reflect monosomic chromosomes in low hypodiploid clones (B).

5.4.2 SNP array analysis

Initially, a standard visual analysis of the SNP array profiles was performed for all 88 cases in Nexus by examining log2 ratio and B-allele frequency traces. In 32 cases the HET-CNG pattern was observed and cytogenetic analyses were consistent with the classic high hyperdiploidy profile previously described with trisomies of chromosomes 4, 6, 10, 14, 17, 18 and tetrasomy of chromosome 21 dominating (figures 5.3 and 5.5).



Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Log2 ratio	-0.04	-0.04	-0.04	0.17	-0.04	0.14	-0.03	-0.04	-0.03	0.15	-0.03	-0.04	-0.04	0.14	-0.05	-0.03	0.13	0.15	0.00	-0.03	0.26	-0.03
LOH (Y/N)	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Figure 5.5. Representative SNP array whole genome view of typical high hyperdiploid ALL case. Log2 ratio and B-allele frequency traces and table detailing whole chromosomal log2 ratios for each chromosome 1-22 and whether B-allele frequency indicates loss of heterozygosity (LOH) (Y/N). The case demonstrates classic pattern of high hyperdiploidy where chromosomes with the lowest log2 ratio possess a normal disomic component of SNPs on B-allele frequency (BB, AB and AA alleles) and chromosomes with higher log2 ratio possess a trisomic pattern of SNPs (BBB, ABB, AAB and AAA alleles) on B-allele frequency.

The SNP profile of a further 35 cases exhibited the LOH-LCN pattern, characteristic of HoTr. Importantly, this pattern was seen regardless of whether the karyotype showed a low hypodiploid clone, near triploid clone or both (figures 5.4 and 5.6). This was consistent with the LOH occurring as a result of primary chromosomal loss; typically affecting chromosomes 3, 7, 15, 16, 17 (figure 5.4). Among these 35 cases, cytogenetic analysis revealed a low hypodiploid, near-triploid or both clones in 9, 9 and 8 cases respectively; or analysis had failed (n=9) (figure 5.2).

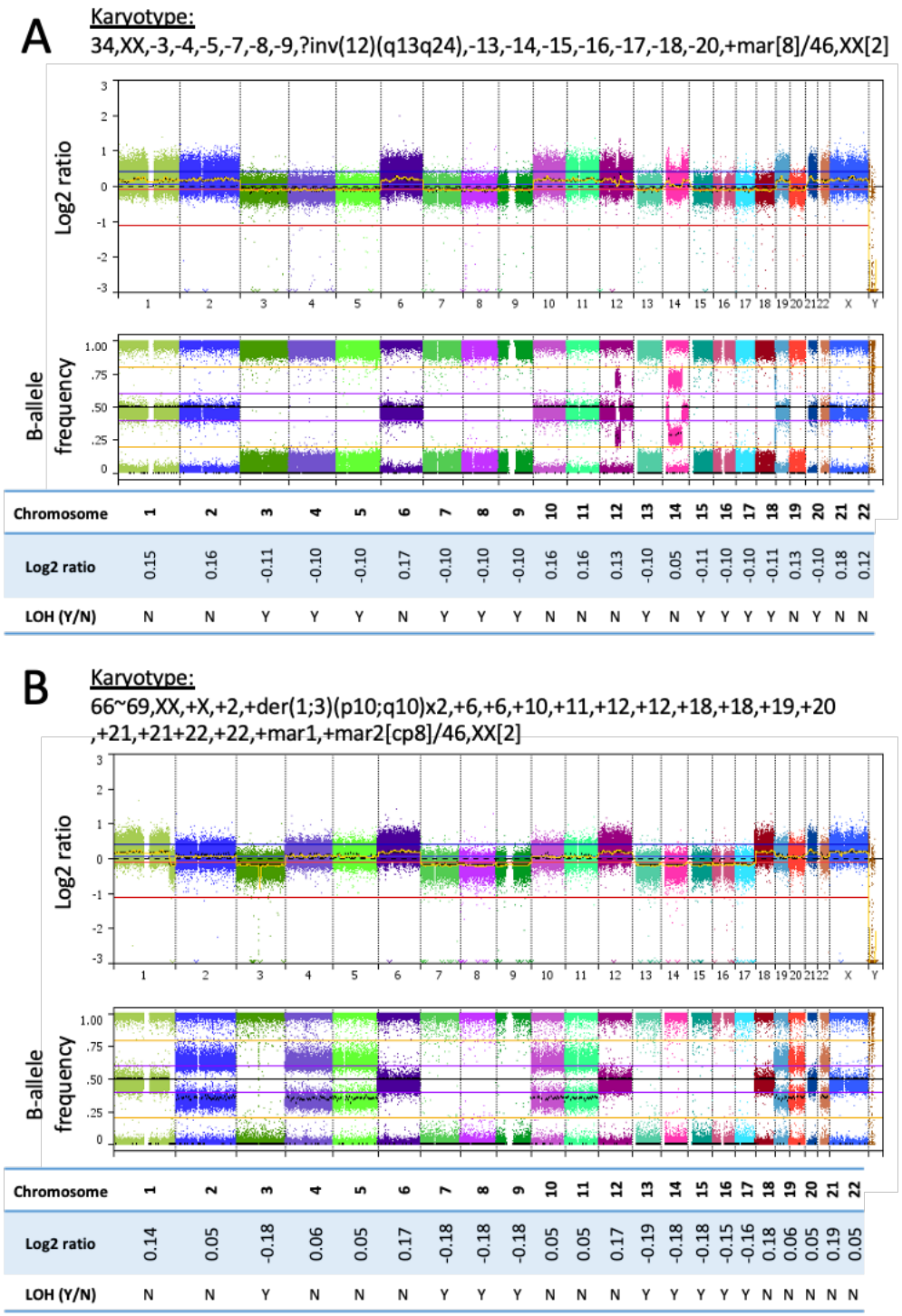
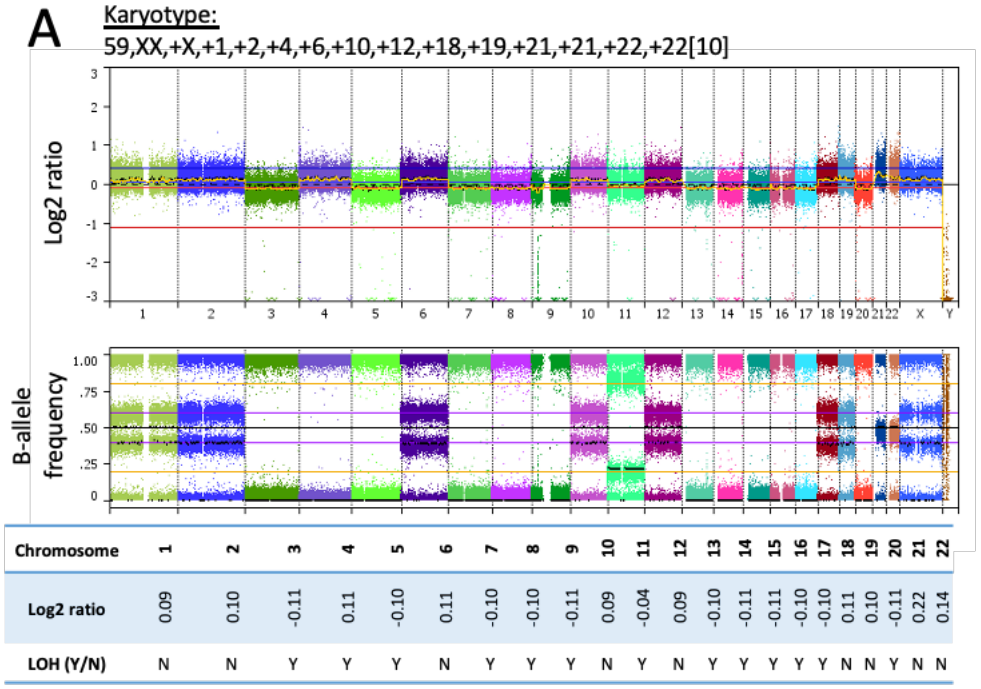
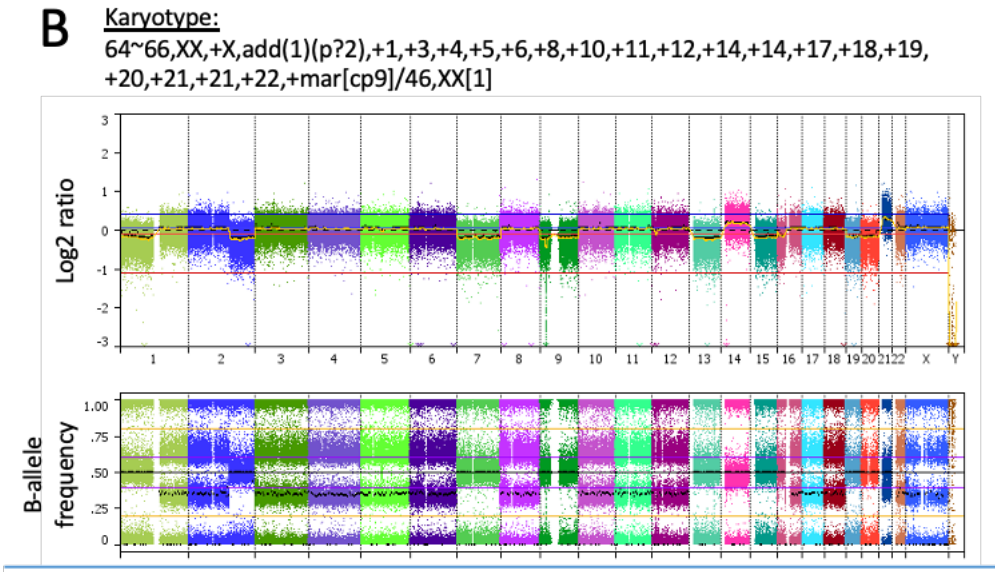


Figure 5.6. Whole genome view of SNP arrays and whole chromosome log2 ratios. (A) Example of a low hypodiploid case where only a low hypodiploid clone was detected on karyotype. Reduced log2 ratios are seen in chromosomes with complete loss of heterozygosity (LOH) on B-allele frequency trace and elevated log2 ratios in chromosomes with preserved disomic pattern of SNPs on B-allele frequency. (B) Example case where only near triploid clone was detected on karyotype. SNP array demonstrating complete LOH of chromosomes with reduced log2 ratios consistent with copy number neutral LOH (CNN-LOH). Some other chromosomes show a trisomic complement of SNPs suggestive of inexact doubling of a low hypodiploid clone.

Interpretation of the SNP profile for the remaining 21 cases led to a conclusion that contradicted cytogenetic analysis (n=4) or was inconclusive (n=17). Three cases (29491, 27478, 26910) were identified that had initially been classified as HeH but whose SNP array profile was consistent with HoTr (figure 5.7A). Crucially all three of these cases had a modal chromosome number below the theoretical threshold of 60 chromosomes for duplicated low hypodiploidy. In addition, a further case (27058) was initially classed as HoTr on the basis of the cytogenetic result but had a SNP array suggestive of HeH ALL (figure 5.7B). The inconclusive SNP array profiles (n=17) had a near-normal SNP array profile likely due to low leukemic DNA content and/or normal cell contamination.



Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Log2 ratio	0.09	0.10	-0.11	0.11	-0.10	0.11	-0.10	-0.10	-0.11	0.09	-0.04	0.09	-0.10	-0.11	-0.11	-0.10	-0.10	0.11	0.10	-0.11	0.22	0.14
LOH (Y/N)	N	N	Y	Y	Y	N	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	N	N	Y	N	N



Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Log2 ratio	-0.06	-0.03	0.05	0.04	0.05	0.04	-0.18	0.05	-0.18	0.05	0.05	0.04	-0.19	0.17	-0.17	0.00	0.05	0.05	-0.13	-0.15	0.27	0.06
LOH (Y/N)	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Figure 5.7. Two cases with discrepant cytogenetic and SNP array results.
A. Patient 27478 cytogenetically classified as high hyperdiploidy (HeH). However, SNP array demonstrates widespread loss of heterozygosity at the lower copy number state (LOH-LCN), suggestive of low hypodiploidy/near triploidy (HoTr).
B. Patient 27058 cytogenetically classified as HoTr. However, SNP array demonstrates preserved heterozygosity in chromosomes at lower copy number state, therefore excluding loss of chromosomes. Although the pattern of chromosome gains is not typical, features are otherwise consistent with HeH.

5.4.3 Clustering and classification of cases using SNP array data

The whole chromosomal log2 ratios were extracted from Nexus and can be found in supplementary table 7. The values were then standardized using R-package BBmisc as previously described.

PCA and unsupervised hierarchical clustering of standardised whole chromosomal log2 ratios demonstrated clear separation of HoTr and HeH cases (figure 5.8).

However, four cases classified by cytogenetics as HeH clustered with HoTr cases. Reassuringly, three of these four cases (26910, 27478 and 29491) also exhibited the LOH-LCN pattern by standard visual SNP array analysis, demonstrating concordance between the unsupervised clustering analyses and the standard visual SNP analysis. The fourth case (24805) had an inconclusive SNP result following visual inspection with a largely normal profile, most likely due to low tumour DNA content in the diagnostic sample. Similarly, two cases cytogenetically classified as HoTr clustered with HeH samples. Of these, one (27058) demonstrated the HET-CNG pattern by SNP array analysis (figure 5.7B). Interestingly, the other case (28893) had an unusual karyotype, which was neither suggestive of HeH or HoTr, and an inconclusive SNP array. To investigate another primary genetic abnormality, FISH was performed for B-other gene rearrangements as described previously (chapter 2, section 2.5) and an *IGH-CRLF2* gene rearrangement was discovered. As this primary genetic abnormality is distinct from both HeH and HoTr, this finding is likely to explain the unusual karyotype and clustering result.

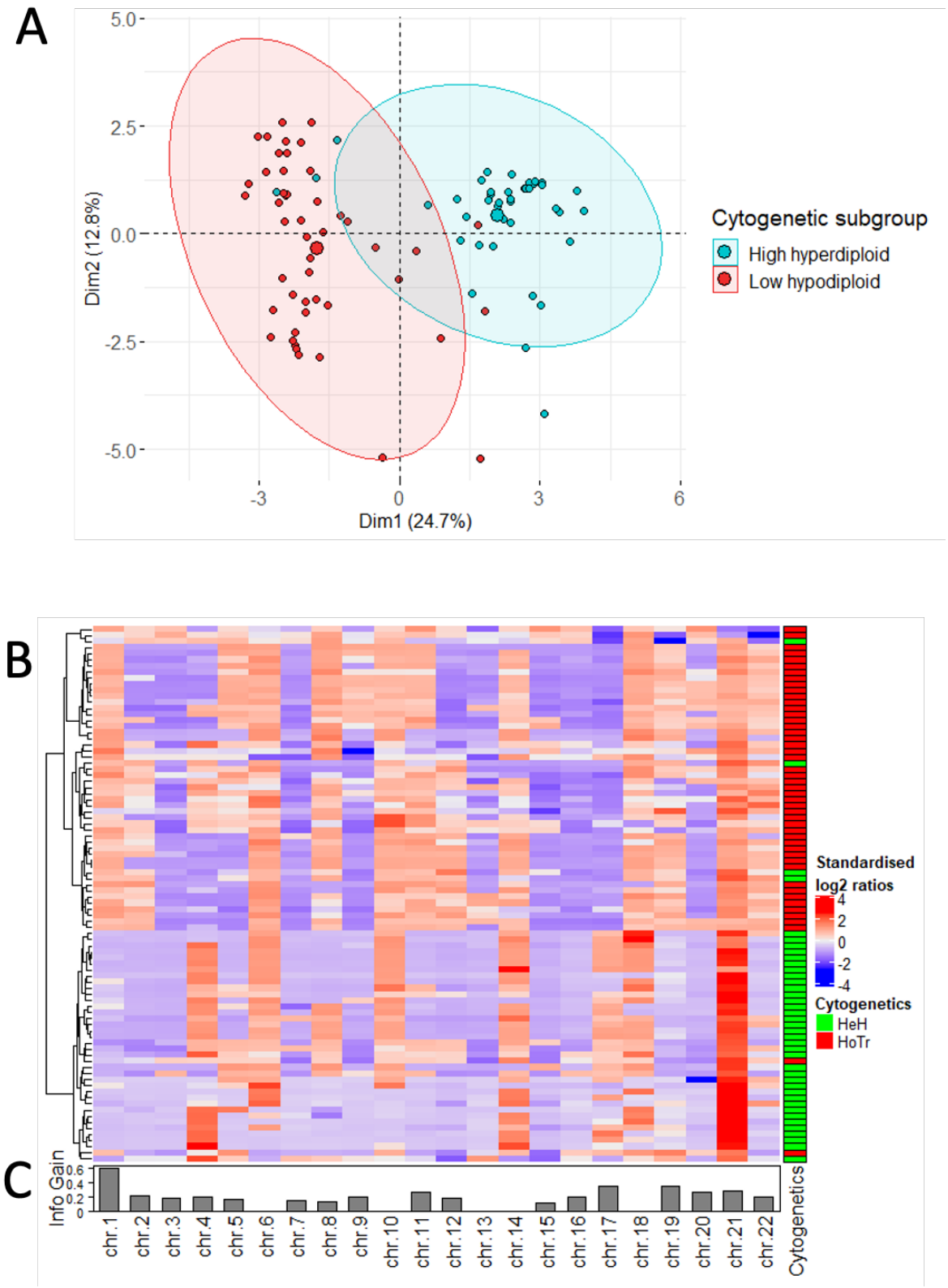


Figure 5.8. Unsupervised clustering of cases by standardised whole chromosome log2 ratios. Principal component analysis (A) and unsupervised hierarchical clustering as a heatmap (B) demonstrate separation of high hyperdiploid (HeH) and low hypodiploid (HoTr) clusters with information contributed by each chromosome (information gain) displayed as a bar chart underneath (C). Cases within the incorrect cluster based on initial cytogenetic classification are detailed in table 5.1.

The six cases that clustered discordantly based on their cytogenetic classification were examined in further detail (table 5.1). As mentioned, the three patients cytogenetically classified as HeH with SNP array profiles and clustering results consistent with HoTr (26910, 27478 and 29491) all had fewer than 60 chromosomes by karyotypes, which is the theoretical limit for duplicated low hypodiploidy (i.e. duplication of 30 chromosomes).

Patient ID	Karyotype	Subgroup by …		
		Cyto	SNP array analysis	SNP array cluster
26910	54~56,XY,+1,add(2)(q3)x2,+3,add(3)(q2),+5, +6,?del(6)(q?2),+10,+11,+14,+?16,+18, +2mar,inc[cp8]	HeH	HoTr	HoTr
27478	59,XX,+X,+1,+2,+4,+6,+10,+12,+18,+19,+21, +21,+22,+22[10]	HeH	HoTr	HoTr
29491	58~59,XY,+?X,+1,+2,+6,add(8)(q2)x2,+10, +11,+12,+12,+14,idic(15)(p1),+18, add(18)(p1),+19,+21,+21,+22,+mar,inc[cp10]	HeH	HoTr	HoTr
24805	53,XX,+5,+6,+10,+11,+20,+21,i(21)(q10),+22 [8]	HeH	Inc	HoTr
27058	64~66,XX,+X,add(1)(p?2),+1,+3,+4,+5,+6,+8, +10,+11,+12,+14,+14,+17,+18,+19,+20,+21, +21,+22,+mar[cp9]	HoTr	HeH	HeH
28893	75~80,XY,+X,+Y,+Y,+Y,+1,+1,+2,+2,+3, +4,+5,+5,+6,+7,+8,+9,+10,+11,+12,+13, +14,+14,+15,+15,+16,+16,+17,+17,+18, +18,+19,+19,+20,+20,+21,+21,+22,+22[cp4]	HoTr	Inc	HeH

Table 5.1. Cytogenetic, SNP array and standardised whole chromosomal log2 ratio clustering results of all cases that clustered discordantly from their cytogenetic subgroup.
Cyto: cytogenetics; HeH: high hyperdiploidy, HoTr: low hypodiploidy/near triploidy, Inc: inconclusive SNP array pattern.

5.4.4 *TP53* mutations

For additional evidence that the three patients classified cytogenetically as high hyperdiploidy were in fact masked low hypodiploidy (26910, 27478 and 29491), coding regions of *TP53* were sequenced by including the samples in the SureSelect XT2 library prep described previously (chapter 2, section 2.7 and supplementary table 4). This confirmed that all three patient samples harboured pathogenic *TP53* mutations (table 5.1). The *TP53* variants identified are reported in the Catalogue of Somatic Mutations in Cancer (COSMIC) (Tate et al., 2018) and were missense mutations

affecting the DNA binding domain (*TP53* p.P151S and *TP53* p.R282W) and a nonsense mutation in the C-domain (*TP53* p.K305*). Additionally, one of the two cases classified cytogenetically as duplicated low hypodiploidy that clustered with high hyperdiploid cases (28893) had a *JAK2* mutation identified, and did not harbour any pathogenic *TP53* variants. This finding was consistent with the discovery of the *IGH-CRLF2* rearrangement, which is strongly associated with *JAK2* abnormalities, and confirms this case represented a separate genetic entity (Mullighan et al., 2009a).

Demographic, clinical and outcome data as well as *TP53* status of the six cases that clustered discordantly from their cytogenetic subgroup are shown in table 5.2.

Patient ID	Age (yrs)	Sex	WCC (x10^9/L)	Subgroup by:		Mutations	Outcome
				Cyto	SNP array cluster		
26910	43	M	31.4	HeH	HoTr	*TP53* p.P151S	Died in CR1 within 1 yr
27478	58	F	8.43	HeH	HoTr	*TP53* p.R282W	Relapsed and died within 2 yrs
29491	51	M	9.5	HeH	HoTr	*TP53* p.K305*	CR1 (4 months)
24805	46	F	58.34	HeH	HoTr	Not done	Died in CR1 within 1 yr
27058	7	F	121.9	HoTr	HeH	Not done	CR1 (2 years)
28893	27	M	31.7	HoTr	HeH	*JAK2* p.T875N	CR1 (1 year)

Table 5.2. Clinical, demographic and outcome data of cases that clustered discrepantly from initial cytogenetic subgroup. Relevant mutations also shown. Case 28893 did not harbour a *TP53* mutation.
M: male; F: female; WCC: white cell count; Cyto: cytogenetics; HeH: high hyperdiploidy; HoTr: low hypodiploidy/near triploidy; CR1: first complete remission

5.4.5 Development of ploidy classifier using whole chromosome log2 ratios

The clustering analyses demonstrated that the distribution of whole chromosome log2 ratios in the HeH and HoTr cohorts was non-random, reflecting the patterns of chromosomal gains and losses in the two subgroups (figure 5.7B). The contribution of each whole chromosome log2 ratio to the unsupervised hierarchical clustering analysis was also non-random with chromosome 1 being the most informative discriminator (figure 5.7C).

To explore whether whole chromosome log2 ratios could be used to develop a ploidy classifier, CART analysis was used to identify the most informative chromosomes for a decision tree. To maximise its diagnostic power, the decision tree was built using the 88 low hypodiploid and high hyperdiploid cases, as well as an additional 72 cases broadly representative of adult ALL (supplementary table 6). Prior to running the CART analysis, four of the discrepant cases (26910, 27478, 29491 and 27058) were re-classified according to the SNP array, clustering and *TP53* results, as these provided overwhelming evidence that the cytogenetic-derived subgroup had been inaccurate. Similarly, the *IGH-CRLF2* rearranged case (28893) was re-classified into the non-ploidy subgroup as the underlying primary genetic lesion was clearly distinct from both HoTr and HeH. Thus, the final CART analysis cohort comprised 160 cases, specifically HoTr (n=50), HeH (n=41) (including three cases with *BCR-ABL1* and high hyperdiploidy) and non-ploidy (n=69).

Using the complete dataset (n=160), CART analysis identified a decision tree based on the standardised log2 ratios of chromosomes 1, 7 and 14 (figure 5.9). Using these variables, cases could be delineated into one of four terminal nodes: one each for HoTr and HeH and two for the non-ploidy cases. The majority of HoTr cases (47/50, 94%) were correctly placed into the HoTr group, while 3 cases were placed into non-ploidy groups. Similarly, the majority of HeH cases (33/41, 80%) were correctly assigned to the HeH node.

The model was then validated internally using 10-fold cross validation and delivered an overall average accuracy of 79% (95% confidence interval 72%-85%) across all three ploidy classes. Positive predictive values were 91%, 68% and 77% for HoTr, HeH and non-ploidy cases respectively and negative predictive values were 94%, 92% and 81% for HoTr, HeH and non-ploidy cases respectively.

Overall, eleven cases called as HoTr or HeH by cytogenetics and/or SNP array were incorrectly assigned by the decision tree (figure 5.8). There was no overlap between these 11 misclassified cases and the 4 cases re-classified by the SNP array analysis described above (table 1). Review of the karyotypes revealed atypical chromosome gain/loss or complex structural rearrangements in the majority.

Importantly, for clinical diagnostic practice, chromosome 1 was a very powerful discriminator between HoTr and HeH cases, and accurately segregated 97% (88/91) of cases with a ploidy shift. Taken together these data show that if cytogenetic analysis, FISH or DNA index identifies a hyperdiploid clone, the standardised log2 ratio

of chromosome 1 (<> 0.28) can extremely reliably discriminate between high hyperdiploidy and duplicated low hypodiploidy.

5.4.6 External validation of the decision tree classifier using Vienna cohort

Using the external Vienna validation cohort, the classifier correctly placed 100% (7/7) of the HoTr cases in the HoTr node. Additionally, 86% (6/7) of the HeH cases were correctly classified. One HeH case was placed in a non-ploidy group due to the absence of a chromosome 14 gain. The near haploid cases were split amongst the HeH (n=5), and non-ploidy nodes (n=3). Such cases were not included in the primary cohort so did not have a dedicated node, and these data confirm that as expected they have a SNP array profile distinct from HoTr. Unblinded cytogenetic details of the validation cases can be found in supplementary table 9.

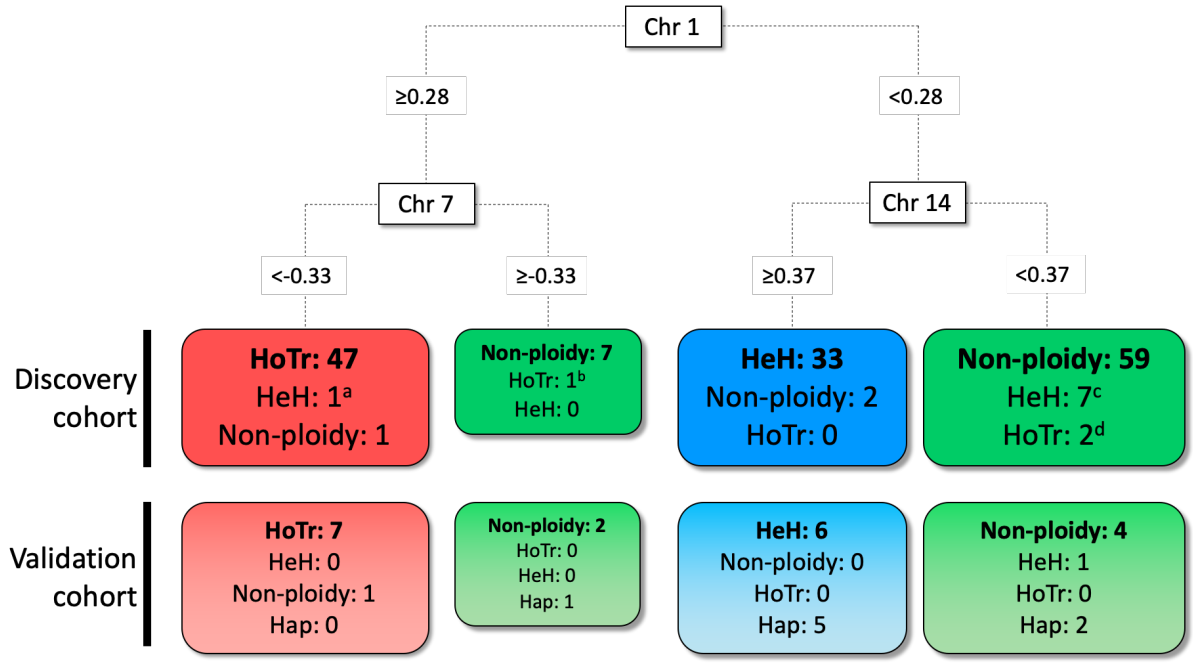


Figure 5.9. Decision tree for assigning cases to a genetic ploidy subgroup using standardized whole chromosome log2 ratios of chromosome 1, 7 and 14. SNP arrays with standardized log2 ratios for chromosome 1 ≥0.28 and chromosome 7 <-0.33 had a 94% probability of being HoTr cases. Cases with a standardized log2 ratios <0.28 for chromosome 1 and ≥0.37 for chromosome 14 had 94% probability of being HeH. Cases where the log2 ratio was <0.28 for chromosome 1 and <0.27 for chromosome 14, had an 87% probability of the absence of major ploidy shift. Importantly, these three scenarios accounted for 95% of the patients in the dataset. A total of 11 cases called by cytogenetics and/or SNP array as having ploidy shifts were incorrectly assigned by the decision tree: (a) This patient had high hyperdiploidy and t(9;22)/*BCR-ABL1*, which is recognised to have a different pattern of chromosomal gains from primary high hyperdiploidy (Chilton et al., 2013) ; (b) Although this patient had a low hypodiploid karyotype with -7 there was unbalanced translocation between the long arms of chromosome 6 and 7; (c) 4/7 cases failed cytogenetics while none of the remaining 3 cases had a +14; (d) Karyotypes had been classed as HoTr by cytogenetics but SNP array analysis was inconclusive.
Chr: chromosome; HoTr: low hypodiploidy/near triploidy; HeH: high hyperdiploidy; Hap: near haploidy; Non-ploidy: absence of primary large scale ploidy shift.

5.5 Discussion

As discussed in chapter 3, HoTr is one of the commonest subgroups in older adults with ALL, and has been poorly studied to date, largely due to its rarity in other age groups. Although younger patients were also included in this study to permit a more robust clustering analysis, there is no evidence that fundamental biological differences exist between older and younger patients with HoTr.

The results of this study provide compelling evidence of the importance of SNP arrays in the detection of clinically relevant ploidy groups at diagnosis of BCP-ALL. The most important finding is that modal chromosome number alone should not be used to

distinguish between high hyperdiploidy and masked low hypodiploidy. These data demonstrate that duplicated low hypodiploidy can give rise to karyotypes with significantly fewer than the theoretical and working lower limit of 60 chromosomes (Charrin et al., 2004, Moorman et al., 2007). Three cases with 50-60 chromosomes have been identified that are consistent with masked low hypodiploid ALL based on SNP array features (figure 5.7A) and the presence of pathogenic *TP53* mutations (table 5.1).

It is important to note that these cases did not show direct cytogenetic evidence of a low hypodiploid clone and that CNN-LOH is well described in high hyperdiploidy (Paulsson et al., 2010), where it may occur as a result of chromosomal mis-segregation during mitosis (Makishima and Maciejewski, 2011). However, several features on the SNP arrays were highly reminiscent of the HoTr pattern in all three cases. Firstly, the LOH observed was much more extensive than reported in HeH cases and affected the typical chromosomes that are lost in low hypodiploid ALL. Additionally, LOH was consistently seen in chromosomes at the lower copy number state, and those with preserved heterozygosity always had higher log2 ratios, suggesting that affected chromosomes underwent an initial heterozygous deletion followed by subsequent duplication of the remaining homologue.

Secondly, the hierarchical clustering was unsupervised and was independent of any observed loss of heterozygosity. Based solely on whole chromosomal log2 ratios, the three relevant cases clearly clustered with other HoTr samples, irrespective of the pattern and extent of the readily apparent LOH, thereby demonstrating that whole chromosomal log2 ratios alone provide a characteristic signature for low hypodiploid ALL.

Cytogenetically, all three cases had trisomy of chromosome 1. In addition, CART analysis revealed that the standardised whole chromosome log2 ratio of chromosome 1 was the most discriminating factor for distinguishing between the two ploidy groups. This strongly suggests that the presence of trisomy 1 in an otherwise high hyperdiploid karyotype should prompt further analysis with a SNP array and *TP53* mutation analysis to exclude masked low hypodiploidy. Interestingly, these cases were all adults aged over 40 years at diagnosis and accounted for 3/14 (21%) of the cytogenetically defined high hyperdiploid adult patients in this cohort. This suggests that high hyperdiploidy may be even rarer than previously thought in this age group, and that misclassified

cases of masked low hypodiploidy may account for the poorly defined prognostic value of hyperdiploidy in adults.

As SNP array analysis is performed on a fixed amount of DNA rather than a fixed number of cells, exact multiples of cellular DNA content all result in the same pattern on the microarray. For example, exact tetraploidy is indistinguishable from diploidy. Although some studies have developed methods for normalisation of aneuploid genomes, these are based on the assumption of a single aneuploid tumour population at a fairly constant ploidy level (Pounds et al., 2009, Van Loo et al., 2010). Resolving the copy number state of individual chromosomes in samples that potentially contain at least three different ploidy populations (low hypodiploid, near triploid and normal diploid DNA) is likely to be challenging. However, within each sample, relative over or under-representation of genomic loci can be inferred based on positive or negative deflections in the log2 ratio respectively. In this analysis, this principle has been used to manually derive log2 ratios for entire chromosomes. Although re-centering of log2 ratios to presumed diploid regions is often performed in aneuploid samples (either manually or computationally within SNP array analysis software), the individual sample standardisation described (to mean whole chromosome log2 ratio = 0 and SD = 1) negates any assumption of copy number state and therefore permits an unbiased analysis with no prior knowledge of the sample ploidy status. Each individual chromosome's copy number status is considered relative to the others in the sample rather than as an absolute value. Indeed, defining diploid regions is often flawed in samples with drastically disparate ploidy levels where the relative proportion of different clones is unknown.

In many cases, a descriptive analysis of SNP array patterns is sufficient to distinguish HoTr from HeH with confidence. However, these data have highlighted that a significant proportion of samples (20% of cases with a major ploidy shift in this cohort) may yield an inconclusive result. It is not clear why DNA from these 17 samples did not produce clear SNP array profiles but in three cases the DNA had been extracted from fixed cell suspension and in a further nine cases samples had been stored for >2 years prior to performing SNP arrays, potentially leading to noisy profiles. Alternatively, near normal SNP array patterns are often encountered when the leukaemic DNA content is low, although this did not seem to be the case with the majority of samples in the cohort (supplementary table 7). However, importantly, these "real-world" analysis issues did not hamper the reliability of the decision tree classifier, which was still able

to accurately delineate cases lacking a clear diagnostic SNP array profile by standard visual analysis as differences in whole chromosomal log2 ratios are present, even if not readily apparent by standard visual SNP array interpretation. Chromosome 1 is consistently relatively over-represented in HoTr compared with HeH samples and is the most discriminatory predictor to differentiate these ploidy groups. Moreover, in the absence of cytogenetics, log2 ratios of key chromosomes (1, 7 and 14) can offer valuable information to resolve the ploidy status of a sample, even when visual interpretation of the SNP array is inconclusive. Current SNP array visualisation software (e.g. Nexus copy number or Affymetrix Chromosome Analysis Suite among others) can be used to create whole chromosome segments and extract log2 ratios to support genetic risk stratification in diagnostic genetic laboratories.

This analysis is limited by case selection only on the basis of the availability of DNA, meaning that the low hypodiploid subgroup is composed predominately of adult ALL cases whereas the high hyperdiploid subgroup comprises predominately childhood ALL. However, this is a true reflection of the underlying epidemiology of these ploidy subgroups and the large number of low hypodiploid cases compiled for this study permitted a very robust clustering analysis.

In conclusion, cytogenetics alone cannot always be used to accurately distinguish between HoTr and HeH. Approximately half of all HoTr cases are masked and present with a near-triploid karyotype. Using modal chromosome number to define a lower limit is flawed, as demonstrated by cases with as few 54 chromosomes, well within the typical range for HeH. Using SNP arrays helps to detect both low hypodiploidy and high hyperdiploidy in cases with normal or failed karyotypes and can usually distinguish between these two ploidy groups. In addition, the ploidy classifier uses the log2 ratios of whole chromosomes to assist SNP array interpretation, and is of particular value in situations where low blast percentage and contaminating non-leukemic DNA may hamper clear SNP array evaluation. Virtually all childhood and most adult ALL treatment protocols will assign patients with HoTr and HeH ALL to very different treatment schedules. Hence the accurate detection of these two ploidy groups is vital to optimal patients' management.

Chapter 6. Mutational landscape of ALL in older adults

6.1 Introduction

Mutations are the hallmark of cancer development. In ALL, initiating abnormalities often involve novel fusion genes or large-scale ploidy shifts, but these events alone are usually insufficient to generate leukaemia. Secondary abnormalities include copy number abnormalities or somatic mutations, and these drive transformation of a preleukaemic clone into clinically overt disease.

Mutations refer to small changes in the genetic sequence and take different forms, with different functional consequences. The most common type of mutation involves a single nucleotide variant (SNV), which occurs when a single base in the genome is substituted with a different base. Due to redundancy in codon to amino acid mapping, many SNVs do not alter the relevant amino acid (synonymous variants). However, when a base substitution results in a changed amino acid, this disrupts the protein structure (missense mutation) or can code a premature stop codon (nonsense mutation), resulting in a truncated protein. Mutations may also involve small insertions or deletions of up to 50 bp, termed indels. These events have the potential to produce very significant disruption to the protein when they produce a shift in the reading frame (frameshift mutations).

During oncogenesis, cells acquire certain mutations that confer survival or growth advantage by Darwinian natural selection (driver mutations). In contrast, passenger mutations are neutral events which occur through mutagenic exposure or genome instability, and although these can also be passed along subsequent generations, they do not confer a specific survival advantage to the neoplastic clone (Greenman et al., 2007).

In ALL, the leukaemic genome generally harbours fewer mutations than most non-haematological tumours (Greenman et al., 2007). Indeed, a study incorporating NGS

(predominantly exome sequencing) of 203 ALL patients identified a median of 11 (range 0-88) and 9 (range 0-45) non-silent mutations in adult and paediatric cases respectively (Liu et al., 2016). On average, ALL genomes have one of the lowest mutation rates, of all cancers, shown to be 0.57 mutations per Mb of DNA in one study, compared to 18.54 mutations per Mb of DNA in melanoma (Greenman et al., 2007). Despite their low mutation rate, certain pathways are recurrently disrupted in ALL genomes. The RAS signalling pathway consists of a number of functionally related RAS proteins that bind to and activate RAF proteins (Downward, 2003). In turn, RAF phosphorylates ERK1/2, which then activate transcription factors permitting cell cycle progression through the G1 phase. Activating point mutation in RAS proteins have been recognised for many years and are one of the most frequent targets of somatic mutation in oncogenesis, present in around 20% of human tumours (Bos, 1989, Downward, 2003). These increase cell proliferation, contributing to the malignant phenotype, and constitute some of the most frequent mutations in childhood ALL samples (Case et al., 2008, Irving et al., 2014). Indeed these aberrations are particularly prevalent in certain subgroups, and have been identified in the majority of ALL patients with either high hyperdiploidy (Paulsson et al., 2015) or intrachromosomal amplification of chromosome 21 (Ryan et al., 2016), as well as being enriched in the relapse setting (Ding et al., 2017, Irving et al., 2014). As with the vast majority of translational research into ALL, published analyses of ALL genomes are heavily biased towards paediatric patients. To date, very few such studies have included adult patients, let alone older adults (Roberts and Mullighan, 2015). One large NGS study included 92 adults, analysed predominantly by exome sequencing (Liu et al., 2016). The authors reported a similar pattern of mutations to the childhood cohort with RAS pathway genes predominating, and also noted a very slight increase in the number of mutations with advancing age. Mutations in epigenetic modifiers were also seen more frequently in adult patients. As discussed in chapter 5, perhaps the closest interaction between a genetic subtype of ALL and a pathogenic mutation is that seen between low hypodiploidy and *TP53* variants (Holmfeldt et al., 2013, Mühlbacher et al., 2014). This genetic entity is seen more frequently with advancing age, and is therefore of particular relevance in the genomic landscape of ALL in older adults.

The pattern of somatic mutations in cancer is non-random and detailed genomic analysis of large datasets has identified specific mutational signatures. These differ due to the distinct mutational processes underlying oncogenesis in different tumours. Mutational signatures are characterised both by the pattern of base substitutions seen

in a sample and the genomic sequence context within which these arise. As such, the six possible substitutions – C>A, C>G, C>T, T>A, T>C, T>G (referred to by the pyrimidine of the mutated base pair) – can arise in the context of any of the four possible bases immediately 5' and 3' to the mutated base pair. This generates 96 possible mutation types, which take into account both the base substitution and the bases immediately adjacent. Mutational signatures are characterised by prominence or absence of these specific mutation types, reflecting an underlying biological process. Using these, 30 COSMIC mutational signatures (version 2) have been described and validated (Tate et al., 2019). For example, signature 1 is characterised by a prominence of C>T substitutions at NpCpG trinucleotide sites, which occurs as a result of spontaneous deamination of 5-methylcytosine. This process is particularly associated with ageing and is positively correlated with increasing patient age at cancer diagnosis. In comparison, signature 2 arises through C>T and C>G mutations at TpCpN trinucleotides, a process which results from over-activity of the APOBEC-family of cytidine deaminases (Alexandrov et al., 2013).

6.2 Aims

To date, most studies have focussed on elucidating the genomic landscape of paediatric ALL. As the frequency of the specific primary genetic abnormalities is very different in younger and older patients, it is likely that the co-operating mutational profile will also be distinct. This chapter will focus on examining the smaller genomic aberrations in a cohort of older adults with ALL with the following aims.

1. Identifying mutations in coding regions of the ALL genome using exome sequencing in a small cohort of older adults with ALL.
2. Identifying the predominant mutational signatures underpinning ALL in older adults.
3. Unravelling the spectrum of mutations in known leukaemia genes and their co-operation with the primary genetic subtypes of older adults with ALL using targeted sequencing.
4. Screening for background clonal haematopoiesis mutations in ALL in older adults.

5. Selecting and tracking variants identified at diagnosis in follow up samples to evaluate clonal persistence and evolution in older adults on treatment

6.3 Methods

6.3.1 Serial bone marrow sampling cohort

A cohort of six locally recruited patients were consented for biobanking of excess bone marrow tissue taken at routine time points during treatment. To maximise patient numbers, patients aged 50 years and over were approached for sample donation. A summary of recruited patient characteristics and treatment protocol is shown in table 6.1.

Patient ID	Age (yrs)	Sex (M/F)	Genetic subgroup	Treatment type
29779	87	M	Not done	Palliative (vincristine + dexamethasone)
29780	55	F	B-other	UKALL14 (chemo only, no allo SCT)
29854	75	F	*BCR-ABL1*	UKALL60+ (Ph +ve arm)
30142	66	F	B-other	UKALL60+ (intensive → non-intensive arm)
30643	52	M	*IGH-CRLF2*	UKALL14 (chemo only, no allo SCT)
31044	63	F	HoTr	UKALL60+ (intensive → non-intensive arm)

Table 6.1. Demographic, genetic characteristics and treatment details of the six patients recruited for sample biobanking. Patients 30142 and 31044 were initially treated on the intensive arm of the UKALL60+ protocol but were de-escalated to non-intensive treatment after induction phase 1 due to treatment related toxicity. No patients underwent allogeneic stem cell transplantation.
M: male; F: female; B-other: BCP-ALL with no primary chromosomal identified; HoTr: low hypodiploidy/near triploidy; allo SCT: allogeneic stem cell transplant.

Serial samples were collected at several time points for each recruited patient (figure 6.1). All sample time points were dictated by routine clinical care, when the patient underwent bone marrow aspiration according to standard of care assessments or due to clinical concern for relapse.

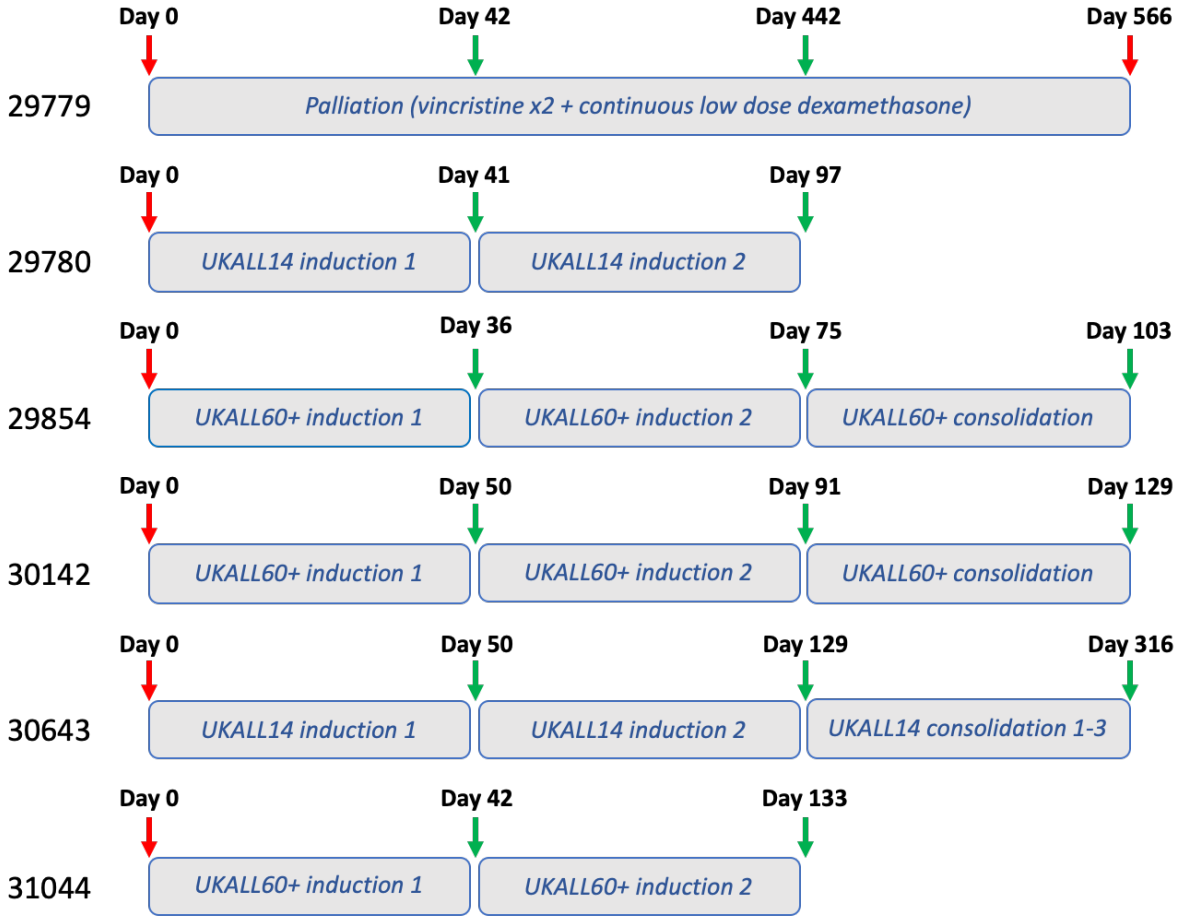


Figure 6.1. Patient progress through treatment and sample time points. Arrows indicate timepoints of sample collection (red = diagnosis/relapse, green = morphological remission). Patient 29779 had peripheral blood taken at all time points. All other patients underwent bone marrow aspiration at all time points. In addition to these samples, all patients had constitutional DNA isolated using a buccal swab.

6.3.2 Patient samples and sequencing

6.3.2.1 Exome sequencing cohort

After bone marrow or peripheral blood sample collection in the patients outlined above (table 6.1 and figure 6.1), mononuclear cells were isolated on the same day using a Lymphoprep density gradient medium (chapter 2, section 2.4.1). Viable mononuclear cells were then stored at -150°C for DNA extraction at a later date (chapter 2, section 2.4.2). Buccal swabs were obtained from each patient and constitutional DNA was extracted for use as a germline control (chapter 2, section 2.4.3). Details of all samples selected for exome sequencing are shown in table 6.2.

Patient ID	Sample type	Volume (ul)	dsDNA conc. (ng/ul)	Nanodrop conc. (ng/ul)	Sample purity (260/280)	Sequencing depth
29779	Diagnosis	15	211.6	226.2	1.87	180x
29779	Germline	30	ND	48.4	1.80	90x
29779	Relapse	20	ND	238.5	1.88	180x
29780	Diagnosis	15	136.6	154.4	1.85	180x
29780	Germline	40	52.4	83.2	1.85	90x
29854	Diagnosis	15	66.4	103.8	1.80	180x
29854	Germline	40	ND	40.0	1.95	90x
30142	Diagnosis	15	197.2	256.7	1.90	180x
30142	Germline	45	ND	38.1	1.84	90x
30643	Diagnosis	30	ND	46.3	1.67	180x
30643	Germline	15	ND	40.3	1.96	90x
31044	Diagnosis	40	ND	43.7	2.16	180x
31044	Germline	20	ND	38.7	1.82	90x

Table 6.2. Details of samples submitted for exome sequencing. For patients 29779, 29780, 29854 and 30142, buccal samples yielded insufficient high quality DNA to meet the requirements for exome library prep so remission samples were used as matched germline alternatives. For patients 30643 and 31044, buccal swabs yielded sufficient high quality DNA to proceed to exome library prep.
dsDNA conc: double stranded DNA concentration; Nanodrop conc: Nanodrop concentration; ND: not done.

Each sample was sent to the Core Genomics Facility (International Centre for Life, Newcastle University) for exome sequencing on the NovoSeq instrument. Each tumour sample (diagnosis or relapse, where applicable) was put through library prep twice and therefore sequenced as two separate samples to generate double the standard read depth of 90x to facilitate detection of lower variant allele frequency (VAF) clones. Sequencing data were returned in the form of FASTQ files and initial bioinformatic analysis of raw reads was performed by Dr Matthew Bashton using the GATK Mutect somatic variant caller (patients 29779, 29780, 29854, 30142 and 30643) and the Bioinformatic Support Unit (Newcastle University) using the GATK Mutect 2 variant caller (all patients).

6.3.2.2 Targeted sequencing cohort

Separately, a targeted NGS sequencing cohort was created from the patient samples included in the SureSelect XT2 library prep sequenced on the NextSeq instrument (chapter 4, section 4.4.4). Together with the deletion breakpoints used for the validation experiment, a number of leukaemia-associated genes were included in the

capture library, and thus constituted a targeted sequencing panel that could be applied to the discovery of pathogenic mutations in these samples (table 6.3 and supplementary tables 3 and 4).

Gene	Region captured	Gene	Region captured
ABL1	whole gene	*DNMT3A*	all exons
ABL2	whole gene	*FOXO1*	all exons
ARID2	whole gene	*IKZF2*	all exons
CSF1R	whole gene	*IKZF3*	all exons
DGKH	whole gene	*IL7R*	all exons
ETV6	whole gene	*JAK1*	all exons
FLT3	whole gene	*JAK3*	all exons
IKZF1	whole gene	*KMT2C*	all exons
JAK2	whole gene	*KRAS*	all exons
KDM6A	whole gene	*NOTCH1*	all exons
MEF2C	whole gene	*NR3C1*	all exons
MEF2D	whole gene	*NRAS*	all exons
NF1	whole gene	*NT5C2*	all exons
PAX5	whole gene	*PTPN11*	all exons
PTEN	whole gene	*RB1*	all exons
PDGFRB	whole gene	*RUNX1*	all exons
RAG1	whole gene	*TET2*	all exons
TCF3	whole gene	*TOX*	all exons
TCF4	whole gene	*SH2B3*	all exons
ASXL1	all exons	*TFDP3*	all exons
ATM	all exons	*TP53*	all exons
CREBBP	all exons	*ZFHX3*	all exons

Table 6.3. Custom 44 gene panel included in SureSelect XT2 capture library (chapter 4). Only known cancer/leukaemia genes shown. Full capture library shown in supplementary tables 3 and 4.

In total 30 patient samples were included in the targeted sequencing cohort, and consisted of the samples from the deletion validation experiment (n=23) and additional samples with ploidy abnormalities (n=7).

6.3.3 Exome sequencing variant calling

Mutect, a variant caller developed by the Broad Institute, was specifically developed to identify somatic point mutations in tumours, and has been extensively validated in this capacity (Cibulskis et al., 2013). The method identifies sites in the genome where an allele differs from the reference. These are compared to the matched germline sample and are flagged if variation is identified. Pre and post processing corrections are performed whereby low quality reads are excluded and noise corrections are implemented according to baseline somatic mutation rates.

Mutect2 is a more recently developed somatic variant caller, able to identify both SNVs and indels in tumour DNA (Benjamin et al., 2019). The method uses a system whereby active sites of possible somatic variation are identified through haplotypes that vary from the reference. Variation that is also present in the matched germline sample is discarded at an early stage, and after various stages of probabilistic scoring, candidate somatic variants are selected.

The variant call file (VCF) output from Mutect or Mutect 2 was then annotated using Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). Where available, this gathered existing information on each variant, including population frequencies, pathogenic scoring and known associations with cancer. *In silico* functional predictions of variants was obtained through the SIFT (Sim et al., 2012) and Polyphen2 (Adzhubei et al., 2010) databases.

6.3.4 Filtering pipeline

A customised filtering pipeline was then designed and executed in R to extract candidate pathogenic variants from the VEP output (figure 6.2). Briefly, likely benign variants present in at least 1% of the general population, obtained from either the gnomAD (Karczewski et al., 2020) or ExAC (Karczewski et al., 2017) databases, were first excluded. Annotation from SIFT and Polyphen2 was then used to filter out variants that were predicted to be both tolerated and benign. The final filtering step then involved removing intronic variants, upstream or downstream gene variants, or those in the 5’ or 3’ untranslated regions (UTRs). Finally, the file was converted into a mutational annotation format (MAF), a similar tab-delimited text file to VEP output with

specific column headings, and containing all samples from a project in a single file. Variant allele frequency (VAF) data for each call was also incorporated into the final MAF file from the VCF output of Mutect. Bioconductor packages maftools (Mayakonda et al., 2018) and ensembldb (Rainer et al., 2019) were used in R to visualise the data and conduct further analyses.

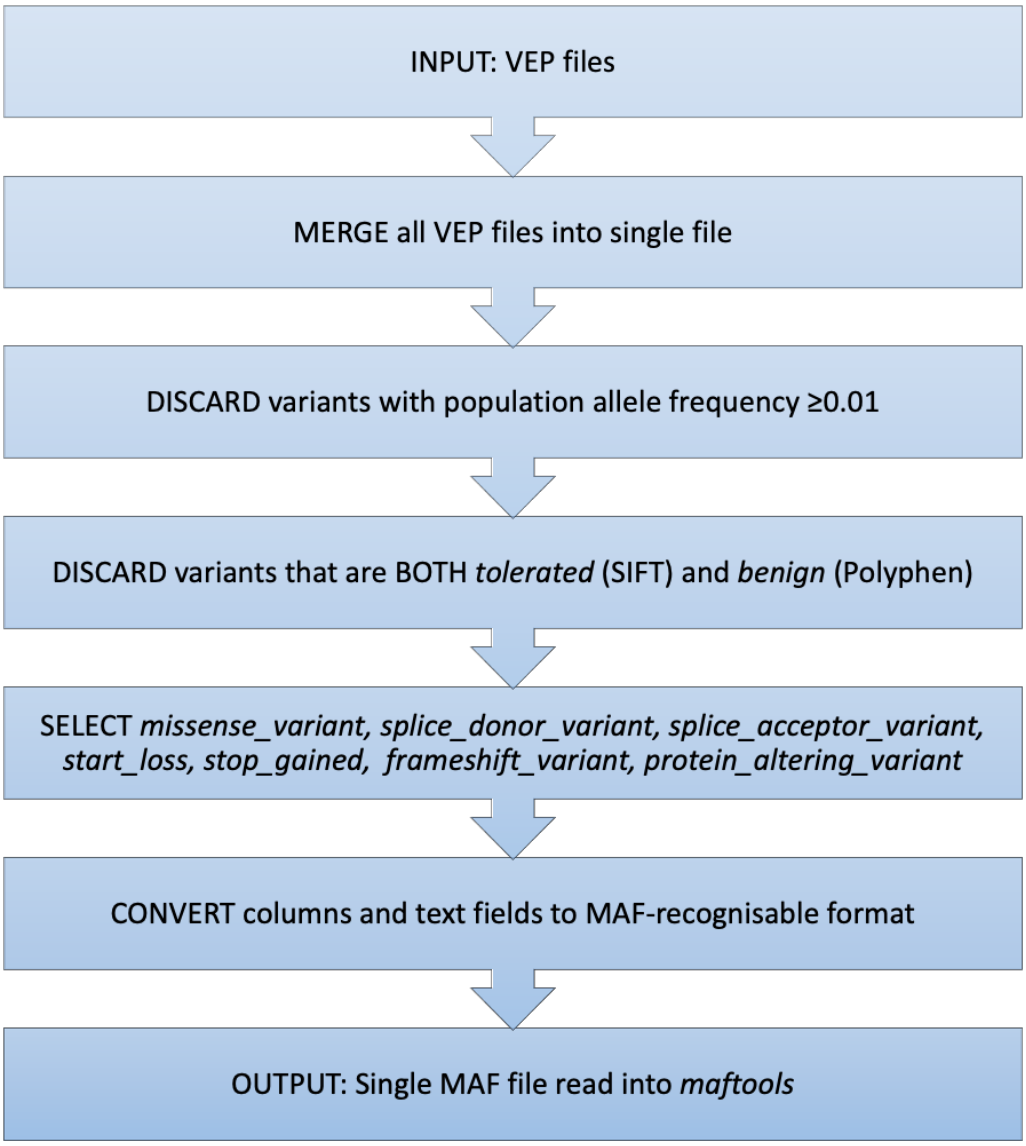


Figure 6.2. Overview of filtering pipeline to extract candidate pathogenic variants, coded and executed in R.

6.3.5 Generating mutational signatures

The Mutect output was then used to identify specific mutational signatures that were enriched in the patient cohort. The pre-filtered output was used to maximise the

number of mutational events to generate the mutational signature. As synonymous and intronic variants are as relevant as pathogenic coding variants in mutational signatures, the raw VEP output was directly converted into MAF format. Next, a trinucleotide matrix of all mutations was created by obtaining the 5' and 3' base immediately adjacent to each variant using R-packages maftools and BSgenome.Hsapiens.UCSC.hg19 as the reference genome. Once this trinucleotide sequence was available for each mutation, samples were compared to the validated COSMIC mutational signatures using the extractSignatures function in R-package maftools.

6.3.6 Targeted sequencing variant calling

The targeted sequencing cohort patient samples did not have matched germline material so variant calling using Mutect and Mutect2 was not appropriate. Instead, raw sequencing reads were processed by the Bioinformatic Support Unit using the GATK HaplotypeCaller (Poplin et al., 2017). This is a germline variant caller that uses a similar method of generating calls as Mutect2 and is able to identify both SNVs and indels. Regions of variation from the reference are first flagged as "active regions". Based on statistical modelling, the most probable haplotypes in the region of variation are then identified, and the likely genotype at each variant site is then defined. Because the software is intended for the detection of germline variation, each case only requires a single sample. If tumour DNA is used, germline variation in the sample cannot be excluded directly. However, using a targeted panel of 44 leukaemia-related genes, the potential for germline variants was already more limited. The same filtering pipeline described in figure 6.2 was executed in R to extract potential pathogenic variants from the VEP output. Additionally, all variants with existing dbSNP identifiers were scrutinised in the dbSNP database (Sherry et al., 2001) and were dismissed as potential germline variants if no clinical significance was reported. All remaining mutations were classed as candidate pathogenic variants, although without germline DNA analysis, their somatic nature could not be confirmed.

VEP outputs for SNVs and indels were processed in the same way. Additional filtering was then applied through visualisation of all post-processing calls in IGV. During this variant calling process, it was noted that frequent deletions were called by HaplotypeCaller in *ABL2*. However, when all the deduplicated and re-aligned BAM files were inspected in IGV, this region showed numerous mismatched bases in all samples

and was therefore likely to represent an artefact. A decision was made to exclude these *ABL2* calls from the downstream analysis.

6.4 Results

6.4.1 Exome sequencing Mutect results

Patients 29779, 29780, 29854, 30142 and 30643 were included in the Mutect analysis, which was performed before patient 31044 was recruited.

From these five patients, six VEP files were processed (patient 29779 had both diagnosis and relapse samples). Across these six patient samples, 966 variants had been called through Mutect. Following the post processing filtering outlined in figure 6.2, 116 candidate pathogenic mutations were flagged, equating to a median of 17 (range 9-37) pathogenic SNVs per sample (figure 6.3).

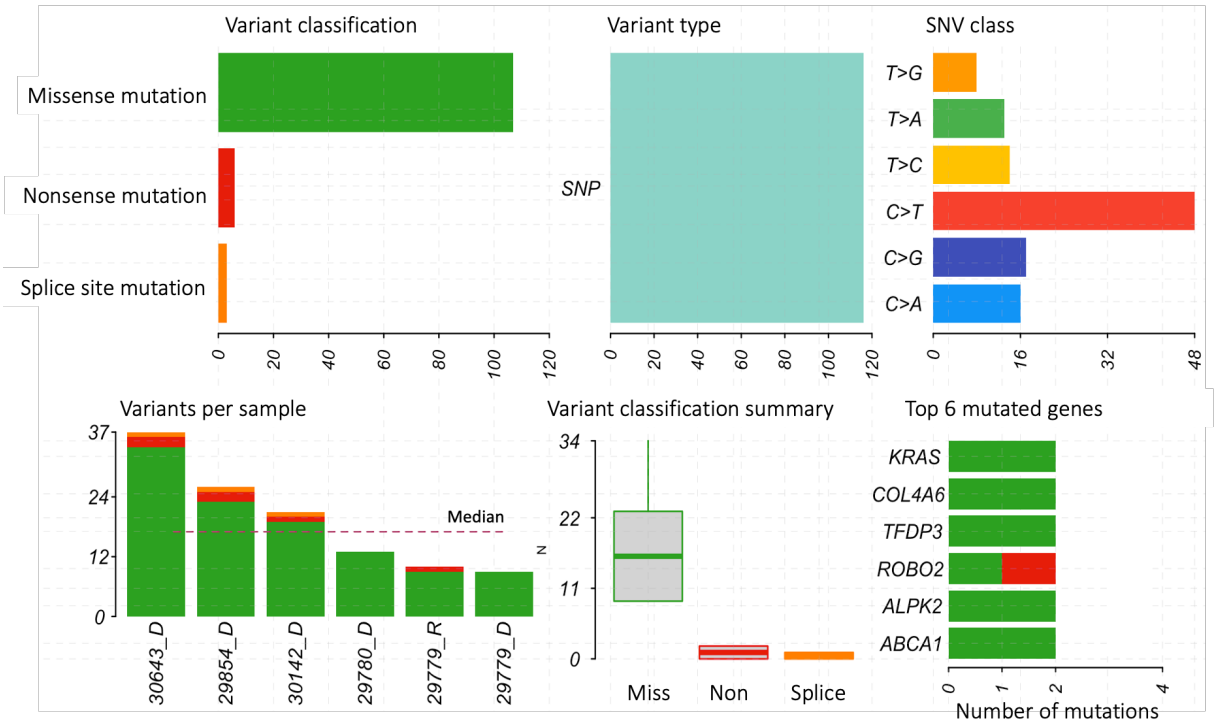


Figure 6.3. Summary of SNVs from Mutect output and downstream filtering from 6 patient samples.
SNV/SNP: single nucleotide variant; D: diagnostic sample; R: relapse sample; Miss: missense mutation; Non: nonsense mutation; Splice: splice site mutation.

The vast majority of SNVs were missense variants and C>T transitions accounted for 41% of events (figure 6.3). There was no association between the number of variants and age at diagnosis and only 6 genes were recurrently mutated. Two separate *KRAS* mutations were identified, both in the same sample (30643). Mutations in *COL4A6*, *ALPK2* and *ABCA1* were present in both diagnostic and relapse samples from patient 29779. Identical *TFDP3* mutations were present in diagnostic samples from patients 29779 and 30643. *ROBO2* was affected by missense and nonsense mutations in patients 30142 and 29854 respectively.

Two of the 6 samples had pathogenic variants in genes known to be recurrently mutated in ALL (figure 6.4).

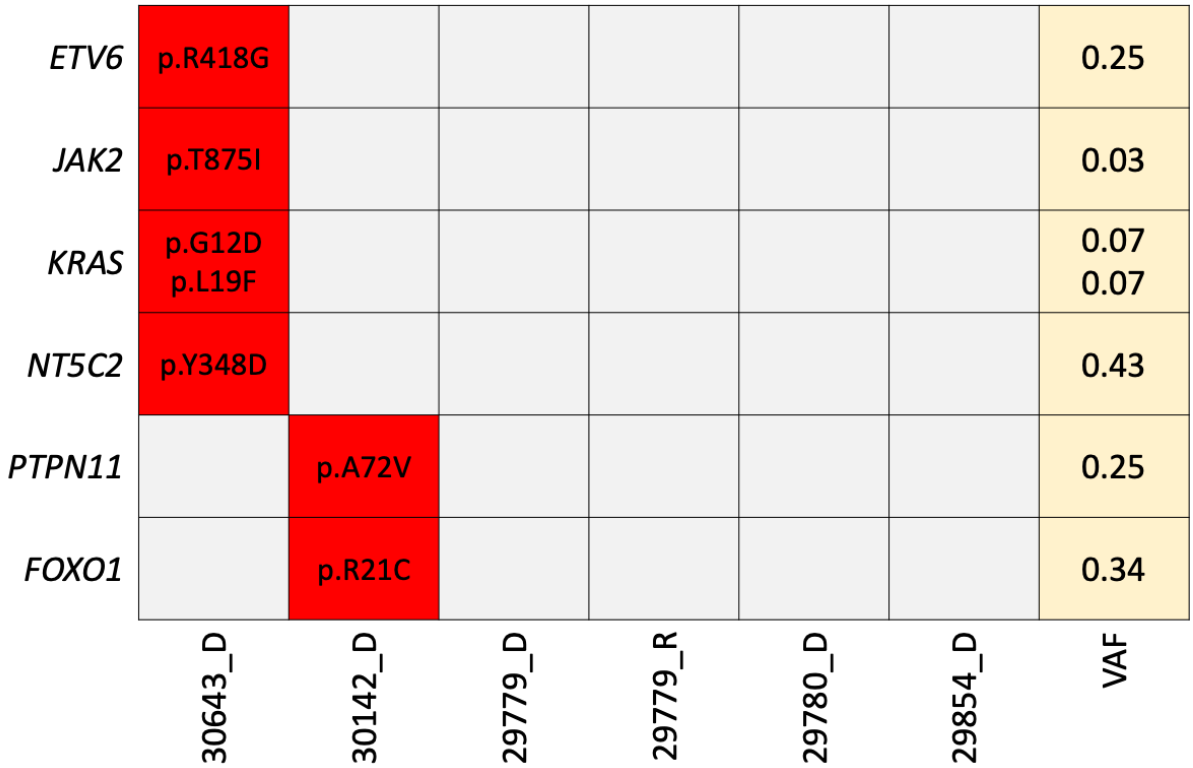


Figure 6.4. Mutations identified in known ALL-related genes (red) following Mutect variant caller and downstream filtering.
D: diagnostic sample; R: relapse sample; VAF: variant allele frequency.

6.4.2 Exome sequencing Mutect2 results

Mutect2 was run on all seven tumour-germline pairs (including one relapse-germline pair) by the Bioinformatic Support Unit (Newcastle University), and as with Mutect variant calling, an annotated VEP file was produced for downstream filtering and variant interpretation. A similar filtering pipeline as shown previously (figure 6.2) was

used to identify candidate pathogenic mutations in each sample and R-package maftools was used for subsequent analysis.

Across these seven patient samples, 5,115 variants were called by Mutect2 consisting of 2,403 SNVs, 1,083 insertions, 962 deletions, 667 substitutions. The high rate of insertions and deletions from the Mutect2 output was unusual. When visualised directly in IGV, many of these were thought to be artefact rather than genuine abnormalities. Most had very few supporting reads and were seen towards the ends of the sequencing reads. Furthermore, significant batch effect was noted. Almost all indels were detected in samples 29779 (diagnosis), 29780 (diagnosis), 30142 (diagnosis) and 30643 (diagnosis), which were all part of the first sequencing batch. Very few were seen in 29779 (relapse) and 31044 (diagnosis), which were sequenced on different runs at a later date. As such, a decision was made to exclude all the indels from downstream analysis. In comparison, the SNV calls appeared genuine in IGV and importantly, this still allowed comparison between Mutect and Mutect2 derived data. Following the subsequent filtering process outlined above (figure 6.2), 262 candidate pathogenic mutations were selected, equating to a median of 39 (range 12-79) variants per sample (figure 6.5).

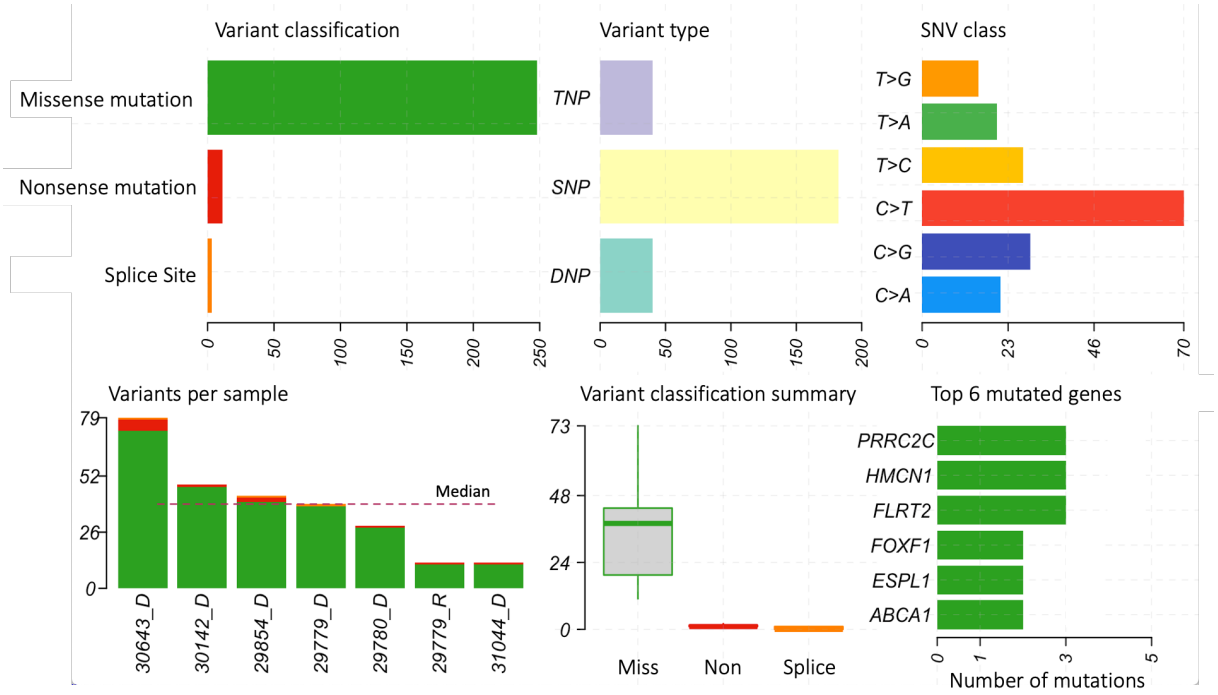


Figure 6.5. Summary of variants from Mutect2 output and downstream filtering from 7 patient samples
SNV/SNP: single nucleotide variant; DNP: dinucleotide variant, TNP: trinucleotide variant; Miss: missense mutation; Non: nonsense mutation, Splice: splice site mutation; D: diagnostic sample; R: relapse sample; Miss: missense mutation; Non: nonsense mutation; Splice: splice site mutation.

As with the Mutect data, missense variants predominated and C>T transitions accounted for 38% (70/182) of base substitutions. Mutations in known ALL related genes were also identified through Mutect2 (figure 6.6). In patient 30643's diagnostic sample, the same *ETV6*, *KRAS* and *NT5C2* variants were identified by both variant callers, but a *JAK2* variant was only identified through Mutect. In comparison, patient 30142 had variants in *PTPN11* and *FOXO1* identified by both variant callers and a *PDGFRB* variant identified only by Mutect2. Exomes from patient 31044 were only processed through Mutect2. Here, a pathogenic *TP53* mutation was detected, consistent with the underlying primary genetic abnormality of low hypodiploidy (Holmfeldt et al., 2013), as well as an *NRAS* mutation.

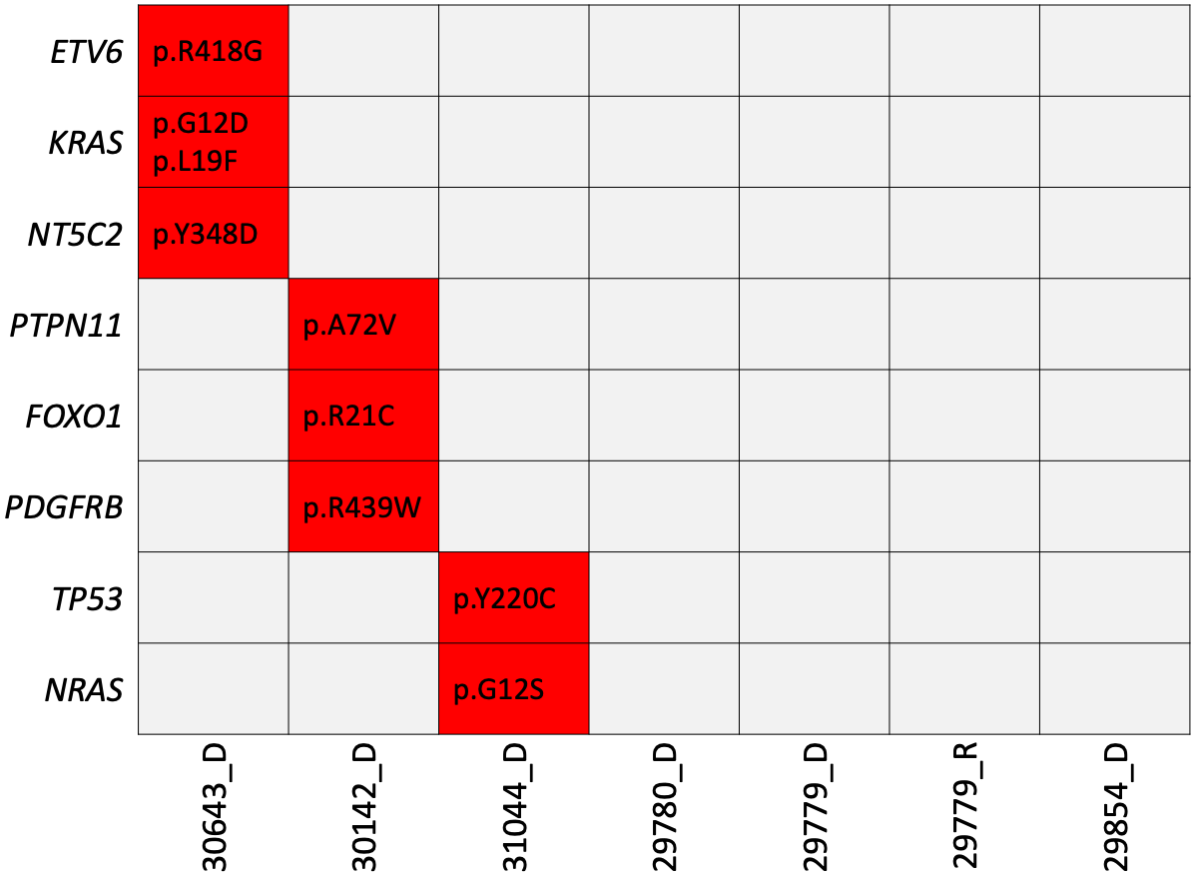


Figure 6.6. Mutations identified in known ALL-related genes (red) following Mutect2 variant caller and downstream filtering. Variant allele frequency (VAF) data was not available from the Mutect2 output.
D: diagnostic sample; R: relapse sample.

6.4.3 Mutect vs Mutect2 concordance

To determine the concordance in variant calling between Mutect and Mutect2, SNVs were examined in isolation. All indels, dinucleotide and trinucleotide substitution calls

by Mutect2 and variants in patient sample 31044 (only analysed by Mutect2) were excluded to permit a direct comparison. Following variant calling and post-processing filtering as outlined above, SNV calls by Mutect (n=116) and Mutect2 (n=171) were examined. In total, 77 concordant SNVs were identified by both variant callers, leaving 39 calls unique to Mutect and 94 calls unique to Mutect2. As noted above (figures 6.4 and 6.6), the majority of SNVs in known leukaemia driver genes were identified by both methods.

6.4.4 Mutational signatures

Mutational signatures were successfully extracted from the unfiltered Mutect output using R-package maftools. The two most prominent mutational signatures across the exome cohort were identified as COSMIC signatures 1 and 25 (figure 6.7).

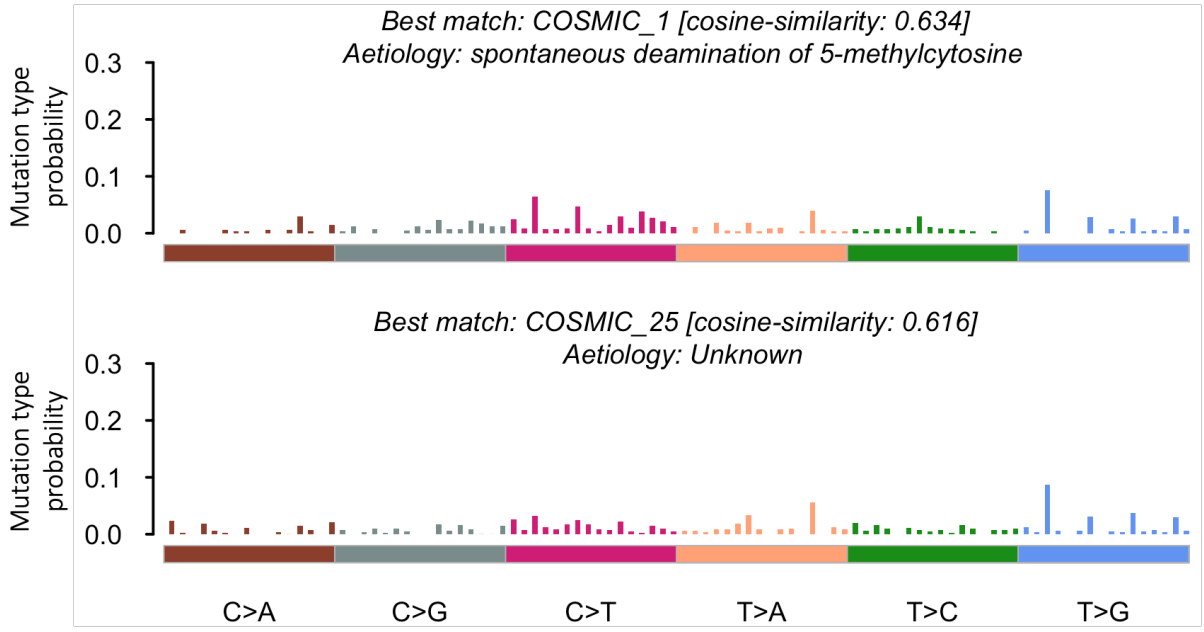


Figure 6.7. Predominant COSMIC mutational signatures in exome cohort analysed using Mutect. Mutational signatures extracted using extractSignatures function in R-package maftools.

COSMIC signature 1 is one of the signatures characterised by a prominence of C>T transitions. These occur specifically at NpCpG trinucleotide sites due to spontaneous deamination of 5-methylcytosine, which generates thymine. This leads to T-G mismatches, which may then fail to be repaired before DNA replication. This process is correlated with the age at cancer diagnosis, and is consistent with the observation that many of the observed mutations will have occurred prior to the onset of

malignancy, at a constant rate during post-natal development (Alexandrov et al., 2015). In comparison, COSMIC signature 25 has unknown aetiology and has been described in Hodgkin's lymphoma cell lines, rather than primary patient material (Tate et al., 2019, Alexandrov et al., 2015). Other mutational signatures present most prominently alongside signature 25 included COSMIC signature 5 (cosine similarity 0.60), which is found in all cancer samples and has unknown aetiology and COSMIC signature 3 (cosine similarity 0.59), which is associated with defective double stranded DNA repair by homologous recombination (Tate et al., 2019).

6.4.5 Targeted sequencing results

The targeted NGS results were analysed separately using HaplotypeCaller variant calling in the 30 patient targeted sequencing cohort. Following post-processing filtering of VEP files using the R pipeline outlined in figure 6.2, 61 SNVs and 17 indels were identified across the 30 patient samples. Apart from the artefactual *ABL2* variants described in section 6.3.6, all other calls were confirmed visually in IGV. Specific variants in 14 of the genes were filtered out due to dbsnp identifiers that were not associated with any clinical significance. The final candidate pathogenic mutations discovered in the targeted sequencing cohort consisted of 34 SNVs and 14 indels and are shown in figure 6.8 along with the primary genetic abnormalities in relevant patients. Detailed mutational data can be found in supplementary table 11. In total, at least one gene in the custom-designed NGS panel was mutated in 73% (22/30) of patients.

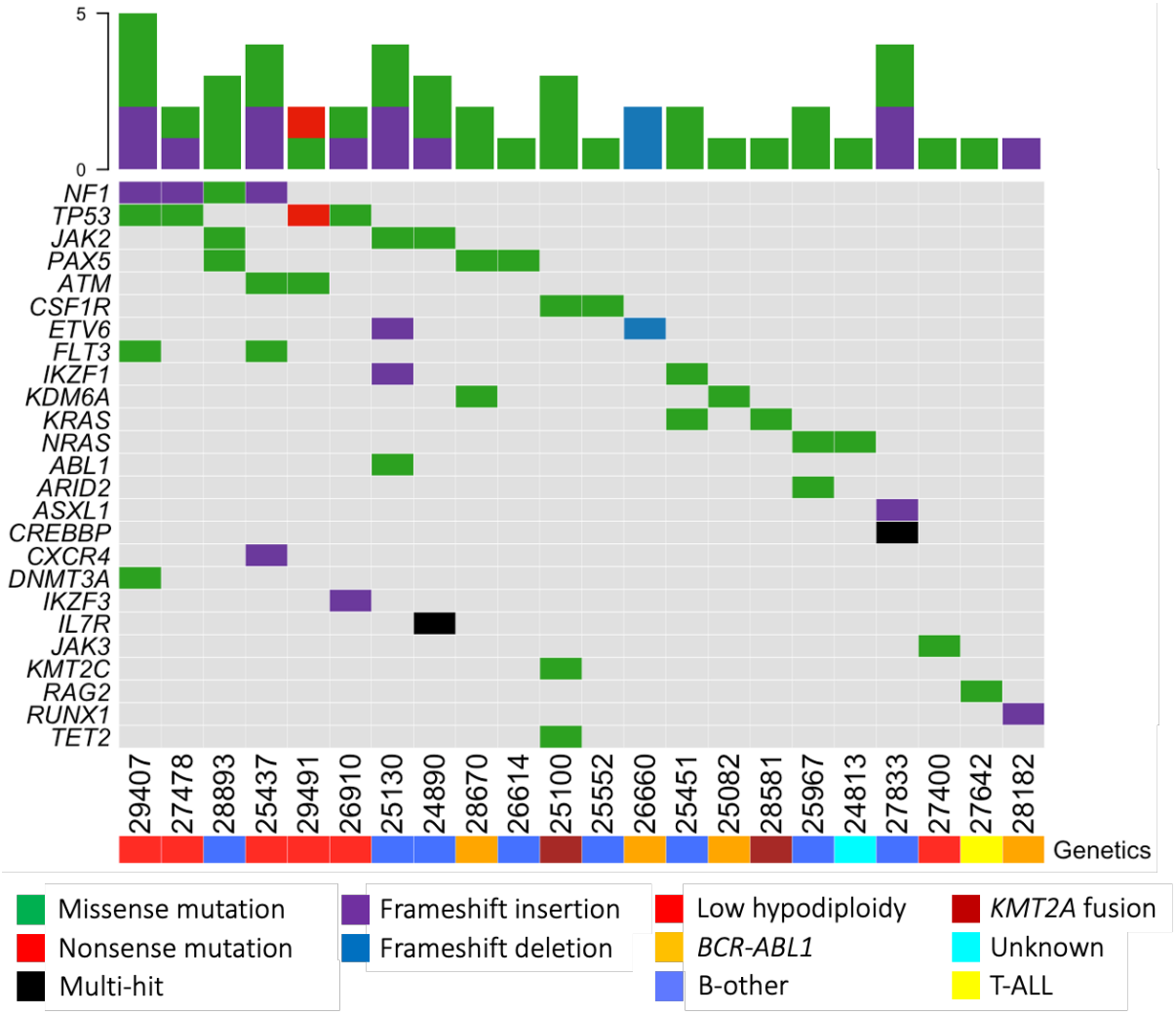


Figure 6.8. Oncoplot of all mutations identified by targeted sequencing in 22/30 patient samples using GATK HaplotypeCaller and customised filtering pipeline (created using R-package maftools). 8/30 cases (not shown) had no pathogenic mutations in any of the 44 genes covered by the targeted panel.
Multi-hit: single nucleotide variant (SNV) and indel in same gene; B-other: BCP-ALL with no identified primary chromosomal abnormality.

6.4.6 Combined exome and targeted NGS cohort

The combined NGS cohort (table 6.4) comprised a total of 36 patients, with diagnostic samples analysed by the targeted sequencing panel (n=30) and exome sequencing (n=6).

Patient ID	Age (yrs)	Sex (M/F)	Genetic subgroup	Sequencing technique
29854	75	F	*BCR-ABL1*	Exome
25082	62	F	*BCR-ABL1*	Targeted
25208	62	M	*BCR-ABL1*	Targeted
25247	64	M	*BCR-ABL1*	Targeted
28057	64	F	*BCR-ABL1*	Targeted
28182	60	F	*BCR-ABL1*	Targeted
28350	62	F	*BCR-ABL1*	Targeted
28670	61	F	*BCR-ABL1*	Targeted
26660	62	F	*BCR-ABL1*	Targeted
31044	63	F	HoTr	Exome
25437	64	F	HoTr	Targeted
29407	60	F	HoTr	Targeted
24401	54	F	HoTr	Targeted
27400	46	F	HoTr	Targeted
29491	51	M	HoTr	Targeted
26910	43	M	HoTr	Targeted
27478	58	F	HoTr	Targeted
28581	65	F	*KMT2A-AFF1*	Targeted
25100	63	F	*KMT2A-v*	Targeted
27642	72	F	T-ALL	Targeted
30643	52	M	*IGH-CRLF2*	Exome
25130	62	F	*IGH-CRLF2*	Targeted
28893	27	M	*IGH-CRLF2*	Targeted
27833	73	F	*IGH-BCL2*	Targeted
27181	65	M	*IGH-CEBPE*	Targeted
25552	61	M	*IGH*-other	Targeted
25451	63	M	*EP300-ZNF384*	Targeted
25967	60	M	Complex	Targeted
29780	55	F	B-other	Exome
30142	66	F	B-other	Exome
25267	63	F	B-other	Targeted
26614	75	M	B-other	Targeted
28335	63	M	B-other	Targeted
24890	65	M	B-other	Targeted
29779	87	M	Unknown	Exome
24813	62	F	Unknown	Targeted

Table 6.4. Combined cohort of patients with diagnostic samples analysed by NGS techniques. M: male; F: female; HoTr: low hypodiploidy/near triploidy; *KMT2A*-v: *KMT2A* fusion with non-*AFF1* partner; Complex: 5 or more unrelated chromosomal abnormality on karyotype; B-other: BCP-ALL with no identified primary chromosomal abnormality.

Median age at diagnosis for the combined cohort was 62 years (range 27-87) and 61% (n=22) of patients were female. The majority of patients (78%, n=28) were older adults aged 60 years and over at diagnosis. In total, eight younger adults were also included. Most of these (5/8) were over 50 years old, and had been included to optimise patient numbers and produce a more robust analysis. Additionally, three younger patients aged <50 years at diagnosis were also included due to unusual karyotypes which were suspicious for masked low hypodiploidy, with the aim of screening for *TP53* variants (chapter 5, section 5.4.4). The combined cohort consisted of patients with *BCR-ABL1* (n=9), low hypodiploidy (n=8), *KMT2A* fusions (n=2), T-ALL (n=1), unknown (n=2). Additionally, fourteen B-other patients were included; *IGH-CRLF2* (n=3), other *IGH@* translocations (n=3), *EP300-ZNF384* (n=1), complex karyotype (n=1), and B-other unspecified (n=6).

6.4.6.1 *TP53* mutations

TP53 was the most commonly mutated gene in this cohort with variants detected in 14% (5/36) of patients from the targeted NGS analysis (n=4) and exome sequencing analysis (n=1). These were all in low hypodiploid cases, which were relatively over-represented in the patient cohort (8/36 patients). All *TP53* variants were recognised pathogenic mutations, reported in the COSMIC database (Tate et al., 2019), and affected the DNA binding domain (*TP53* p.P151S, p.Y220C and p.R282W) and the C-domain (*TP53* p.K305*) (figure 6.9). All affected cases had additionally lost one copy of chromosome 17, which therefore led to biallelic *TP53* inactivation.

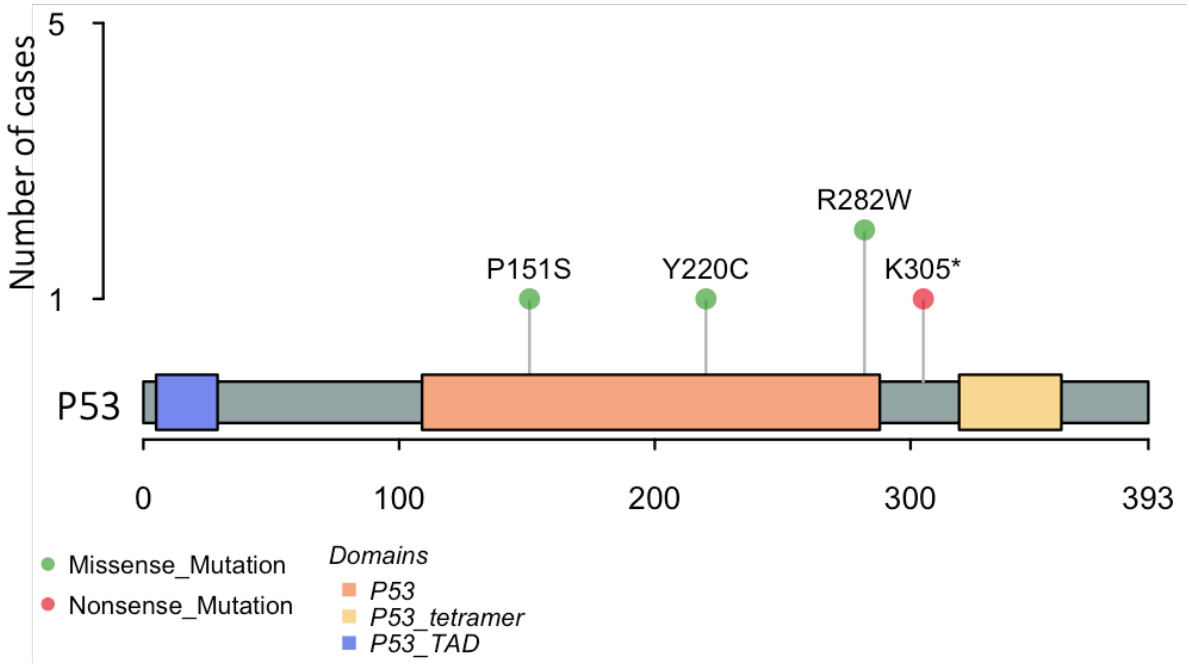


Figure 6.9. Protein domain and alteration plot detailing all *TP53* mutations from targeted (n=4) and exome (n=1) sequencing (created using R-packages ensembldb and maftools). *TP53* p.P151S, p.Y220C and p.K305* were seen in single patients and *TP53* p.R282W was present in two cases.

6.4.6.2 RAS pathway mutations

Several components of the RAS pathway were recurrently mutated in the cohort (figure 6.10). *KRAS* was the most frequently affected member of the RAS pathway, with four mutations identified in 8% (3/36) of cases. Two distinct variants were seen in the diagnostic sample from patient 30643. These were not present on the same sequencing reads when examined in IGV, implying they most likely affected either different alleles or separate subclones. Overall, the *KRAS* mutations were present in two B-other patients (*EP300-ZNF384* and *IGH-CRLF2*) and one t(4;11)(q21;q23)/*KMT2A-AFF1* patient. *NRAS* was also mutated in 3/36 patients (one HoTr, one B-other (complex karyotype) and one unknown). Two identical *FLT3* variants were also found, both in HoTr patients. Apart from patient 30643, all RAS pathway mutations were in different patients.

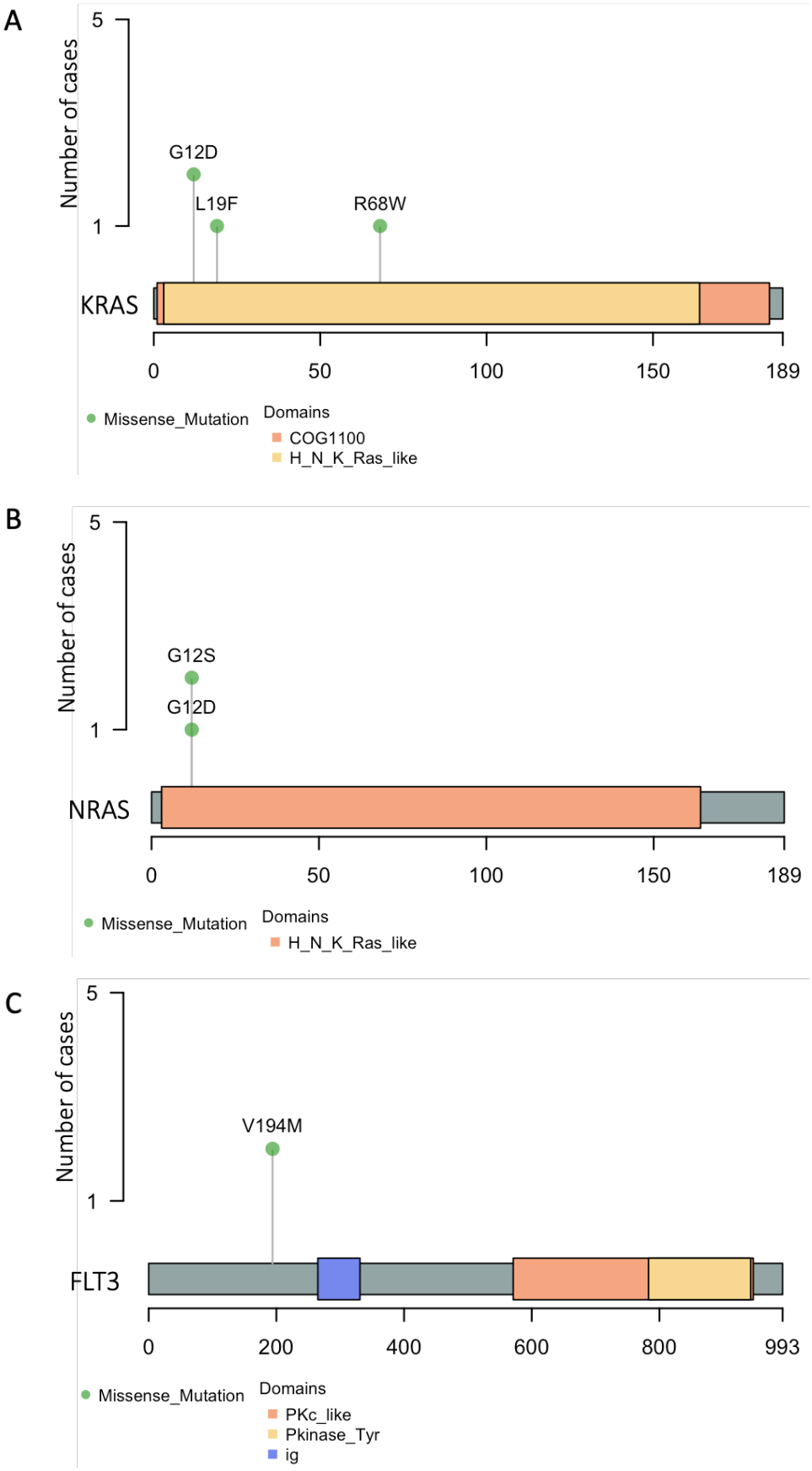


Figure 6.10. Protein domain and alteration plots of RAS pathway genes mutated more than once across all 36 patients (created using R-packages ensembldb and maftools).

Two additional mutations in RAS pathway genes were also discovered; *PTPN11* p.A72V and *NF1* p.R1375H, both in B-other ALL patients (30142 and 28893).

In total, 11 RAS pathway mutations were identified in the cohort of 36 patients. All *KRAS*, *NRAS*, *PTPN11* and *FLT3* variants were confirmed somatic mutations in COSMIC, apart from *KRAS* p.R68W, which was a novel variant. The other *KRAS*, *NRAS* and *PTPN11* mutations are well curated pathogenic mutations from multiple cancer types (Tate et al., 2019). In comparison, *FLT3* p.V194M has been shown to be a tolerated passenger mutation in AML (Fröhling et al., 2007) and *NF1* p.R1375H is a variant of unknown significance in the ClinVar database (Landrum et al., 2014).

6.4.6.3 *JAK2* mutations

JAK2 mutations were discovered in 11% (4/36) patients (figure 6.11). The variants were exclusively seen in B-other patients, specifically three cases with *IGH-CRLF2* rearrangements (30643, 28893, 25130) and one where a primary chromosomal abnormality was not identified (24890). Unfortunately, *CRLF2* break apart FISH could not be performed in this case as no fixed cells were available. *JAK2* mutations are well described in BCP-ALL and are associated with the *BCR-ABL1*-like profile (Mullighan et al., 2009c), specifically patients with *CRLF2* overexpression (Harvey et al., 2010). The p.R683 amino acid residue is the most frequent site of mutation in affected cases and is located in the pseudokinase (JH2) domain. Such mutations disrupt the negative auto-regulation of JAK2 activity, leading to constitutive activation, and increased JAK-STAT signalling (Steeghs et al., 2017).

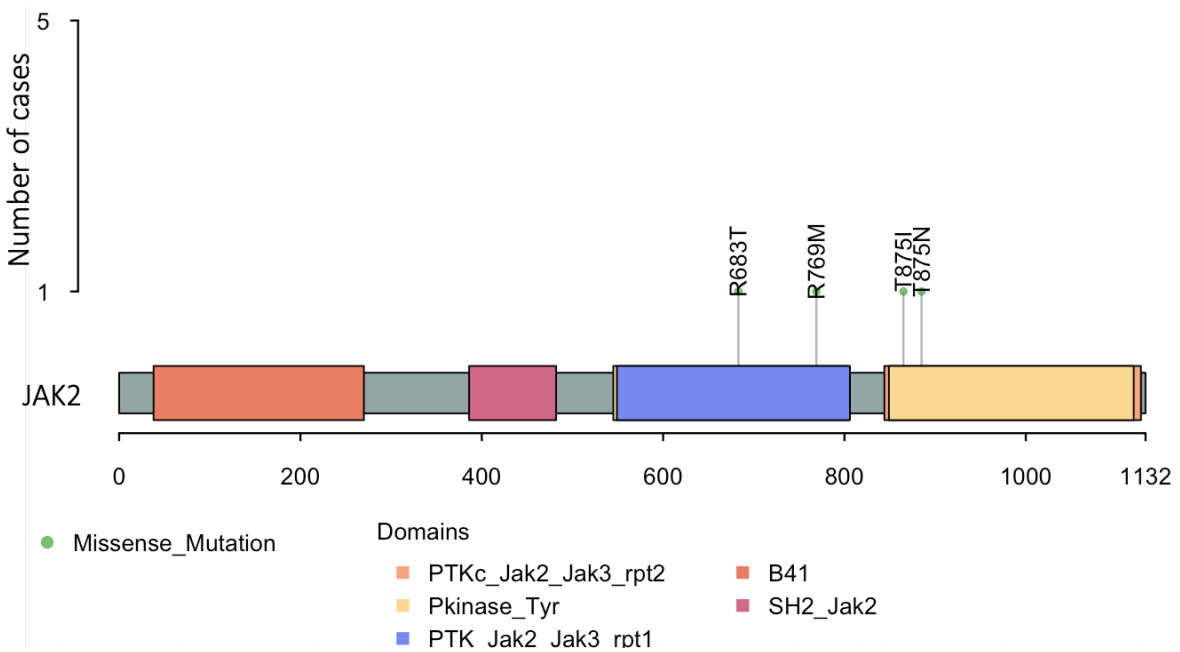


Figure 6.11. Protein domain plot detailing *JAK2* mutations in four patients (created using R-packages ensembldb and maftools).

The *JAK2* p.R683T and p.T875N are both reported pathogenic mutations in the COSMIC database and have been confirmed in BCP-ALL cohorts, whereas *JAK2* p.R769M and p.875I are novel somatic variants.

6.4.6.4 *PAX5* mutations

The *PAX5* variants (figure 6.12) were identified in 3/36 patients – two B-other and one *BCR-ABL1*+ patient.

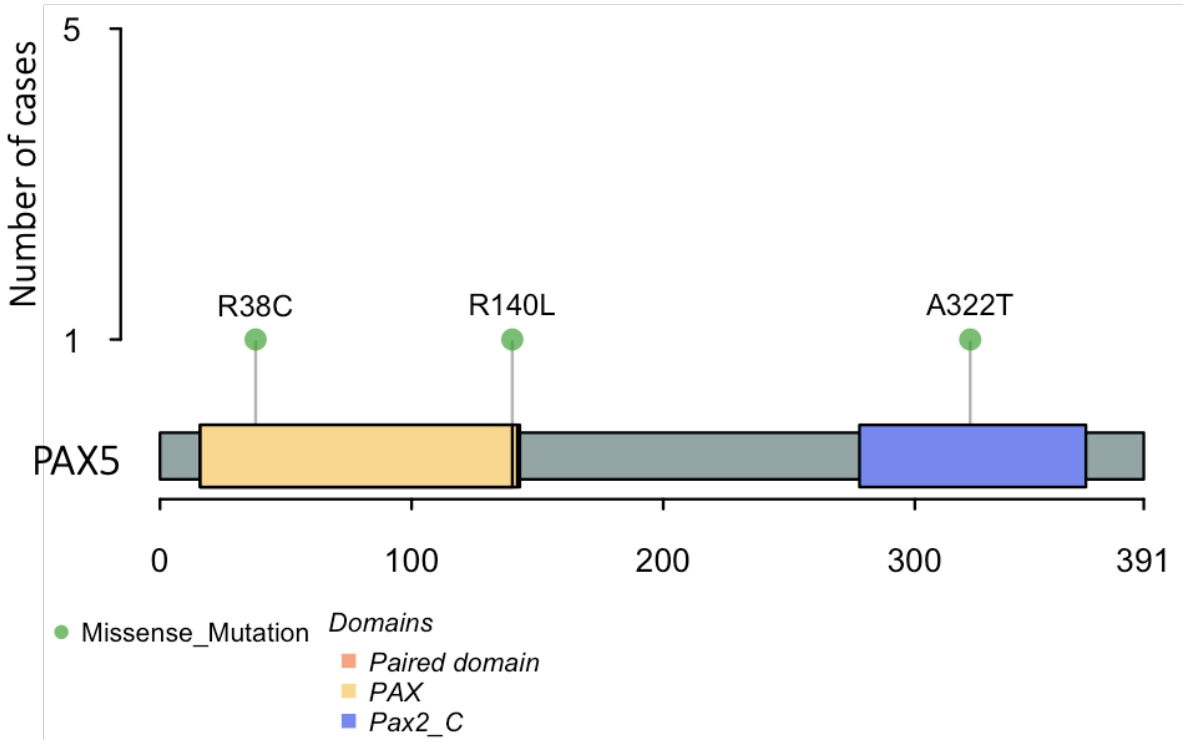


Figure 6.12. Protein domain plot detailing *PAX5* mutations in three patients (created using R-packages ensembldb and maftools).

PAX5 p.R38C is confirmed as a pathogenic somatic mutation in the COSMIC database. Both the *PAX5* p.R38C and p.R140L variants have been previously identified in a large transcriptional analysis of BCP-ALL (Li et al., 2018). In comparison *PAX5* p.A322T is present in the dbSNP database (Sherry et al., 2001) and although reported in ClinVar (Landrum et al., 2014), no conclusive evidence of pathogenicity or lack thereof is provided.

6.4.6.5 Insertions and deletions

Patient 25130 was found to harbour a 1bp insertion in *ETV6* (NM_001987.5 c.1243-1244insT) as well as a 4bp insertion in *IKZF1* (NM_006060.6 c.185-186insTTCC). Both these events were seen in the vast majority of reads in IGV. By SNP array, this case had both *IKZF1* and *ETV6* deletions present, confirming that the small frameshift insertions identified affected the only remaining allele, most likely resulting in profound loss of function.

Several frameshift insertions in *NF1* were also discovered, specifically a 7bp insertion in patient 25437 (NM_001042492.3 c.7989-7990ins GGATAAG), a 1bp insertion in

patient 27478 (NM_001042492.3 c.7292-7293insG) and an 8bp insertion (NM_001042492.3 c.1977-1978ins GTCTCCGT) in patient 29407. All three patients had low hypodiploid ALL, where one copy of chromosome 17 had been deleted. As such these mutations affected the remaining *NF1* allele, as is observed with *TP53* mutations in this subtype. A further patient (26910), also with low hypodiploid ALL, had a 1bp frameshift insertion in *IKZF3* (NM_012481.5 c.266-267insT), which is also located on chromosome 17, and therefore resulted in biallelic inactivation.

6.4.6.6 Mutations associated with clonal haematopoiesis

DNMT3A, *TET2* and *ASXL1* are the most frequently mutated variants in clonal haematopoiesis and were therefore included in the targeted sequencing panel. In total, mutations in these genes were only detected in 8% (3/36) of patients (figure 6.8). Missense mutations in *DNMT3A* (*DNMT3A* p.L653V) and *TET2* (*TET2* p.R1572W) were identified as well as a 1bp insertion in *ASXL1* (NM_015338.6 c.1926-1927insG). The *TET2* variant was present in COSMIC and has previously been reported in chronic myelomonocytic leukaemia (Kohlmann et al., 2010).

6.4.6.7 Genes associated with *BCR-ABL1*-like ALL

Two pathogenic *CSF1R* variants were identified (*CSF1R* p.G413S and p.V32G in *KMT2A* rearranged and *IGH@* rearranged patients respectively). Both mutations are reported in the COSMIC database and have previously been identified in myelofibrosis and/or myeloma, but not in ALL. Interestingly, an *ABL1* mutation (*ABL1* p.P829L) was identified in an *IGH-CRLF2* +ve patients. This variant is also present in the COSMIC database, although has only been reported in a small number of non-haematopoietic tumours. Finally, a *PDGFRB* mutation (*PDGFRB* p.R439W) was identified in the Mutect2 analysis of a B-other ALL case. The variant was predicted to be deleterious and possibly damaging in the SIFT and PolyPhen databases but has not been reported in the literature.

6.4.6.8 Other mutations

In total, four *ETV6* variants were discovered (3 indels and 1 SNV). The SNV (*ETV6* p.R418G) was identified in an *IGH-CRLF2*+ patient (figures 6.4 and 6.6), and is a

known germline variant associated with inherited thrombocytopenia and a predisposition to ALL (Noetzli et al., 2015). Interestingly, this mutation was identified in both the Mutect and Mutect2 analyses, supporting a somatic rather than germline event in patient 30643. The same patient also displayed an *NT5C2* p.Y348D variant, potentially conferring resistance to purine analogues (Tzoneva et al., 2013).

KDM6A mutations were discovered in two *BCR-ABL1+* patients in the targeted NGS cohort (figure 6.8). The *KDM6A* p.Y215H and p.K987Q variants are not reported in the literature but the SIFT and Polyphen *in silico* prediction tools described deleterious and probably damaging consequences respectively.

A somatic *FOXO1* p.R21C mutation was identified in patient 30142 in the exome sequencing cohort. This patient did not have a primary genetic abnormality identified using standard genetic analyses, and interestingly *FOXO1* variants, including p.R21C, are recurrent events in diffuse large B-cell lymphoma, where they are postulated to play a critical role in pathogenesis. *FOXO1* p.R21C affects the N-terminal domain of the protein resulting in loss of phosphorylation with associated nuclear retention and increased transcriptional activity (Trinh et al., 2013).

6.4.7 Clonal tracking of variants through ALL treatment

Diagnostic (n=6), follow up (n=15) and relapse (n=1) samples (figure 6.1) were obtained for the six patients in the exome sequencing and clonal tracking work package. Based on findings from the exome sequencing analysis, the mutations detailed in table 6.5 were selected to design the SureSelect XT HS2 capture library. In addition to these regions, all exons of *TP53*, exons 2 and 3 of *NRAS* and *KRAS* and exons 3, 8 and 13 of *PTPN11* were also included due to their prevalence in driving relapse (Irving et al., 2014). Indeed, it is conceivable that such events could be present in subclones at very low allele frequencies in diagnostic samples, evading detection by exome sequencing, but identifiable using this highly sensitive technique with molecular barcoding. Also, due their role in clonal haematopoiesis, all exons of *DNMT3A* and *TET2* together with exons 11 and 12 of *ASXL1* were included in the capture library (supplementary table 10).

Sample	Gene	Location	Consequence	Amino acid change
29779	*ABCA1*	chr9:107582289	Missense variant	V/M
29779	*ALPK2*	chr18:56247149	Missense variant	D/Y
29779	*NFKBID*	chr19:36380898	Missense variant	E/G
29779	*COL4A6*	chrX:107417755	Missense variant	P/R
29779	*TFDP3*	chrX:132351898	Missense variant	C/W
29779_rel	*HDAC1*	chr1:32790095	Missense variant	D/G
29779_rel	*EXT1*	chr8:119122811	Missense variant	D/N
29779_rel	*ABCA1*	chr9:107582289	Missense variant	V/M
29779_rel	*ALPK2*	chr18:56247149	Missense variant	D/Y
29779_rel	*NAA10*	chrX:153199405	Missense variant	L/Q
29780	*HMCN1*	chr1:186114957	Missense variant	R/Q
29780	*ACTN2*	chr1:236908040	Missense variant	R/H
29780	*COL5A2*	chr2:189916926	Missense variant	A/V
29780	*PTPRD*	chr9:8436635	Missense variant	R/T
29780	*PTCH1*	chr9:98238416	Missense variant	R/H
29854	*USH2A*	chr1:216062120	Missense variant	P/L
29854	*ROBO2*	chr3:77526570	Stop gained	R/*
29854	*DYNC1I1*	chr7:95665015	Missense variant	V/M
29854	*KMT2C*	chr7:151848600	Missense variant	C/Y
29854	*PIEZO1*	chr16:88786335	Stop gained	W/*
30142	*RIT1*	chr1:155874263	Missense variant	M/V
30142	*DNAH7*	chr2:196729067	Missense variant	R/C
30142	*ROBO2*	chr3:77626709	Missense variant	V/F
30142	*PTPN11*	chr12:112888199	Missense variant	A/V
30142	*FOXO1*	chr13:41240289	Missense variant	R/C
30643	*JAK2*	chr9:5089726	Missense variant	T/I
30643	*NT5C2*	chr10:104853013	Missense variant	Y/D
30643	*ETV6*	chr12:12038959	Missense variant	R/G
30643	*KRAS*	chr12:25398262	Missense variant	L/F
30643	*KRAS*	chr12:25398284	Missense variant	G/D
30643	*TFDP3*	chrX:132351898	Missense variant	C/W
31044	*NRAS*	chr1:115258748	Missense variant	G/S
31044	*TP53*	chr17:7578190	Missense variant	Y/C

Table 6.5. Mutations identified from exome sequencing selected for clonal tracking capture library. Unless stated otherwise events were detected in diagnostic patient samples.
rel: relapse sample

Unfortunately, prior to commencing the SureSelect XT HS2 library prep, the COVID-19 pandemic resulted in the suspension of all laboratory activity so this experiment could not be completed in the required timeframe.

6.5 Discussion

In total, 36 patient samples had NGS analyses performed, including targeted panel sequencing (n=30) and exome sequencing (n=6). A customised filtering pipeline was designed to screen the annotated VEP files generated by the GATK Mutect, Mutect2 and HaplotypeCaller variant callers.

Mutect and Mutect2 variant calling produced biologically plausible SNVs that could be readily confirmed in IGV. Unfortunately, indels generated by Mutect2 frequently appeared artefactual when visualised so were excluded from further analysis. The exact reason for the indel anomalies could not be elucidated in this study, although a very strong batch effect was noted, whereby almost all were detected in 4/6 samples, which were all sequenced on the same run. Indel artefact has been noted previously with Mutect2 in The Cancer Genome Atlas (TCGA) database, although chiefly on whole genome amplified samples (Buckley et al., 2017).

Of the SNVs detected by Mutect, 66% (77/116) were also detected by Mutect2 and of the SNVs detected by Mutect2, 45% (77/171) were also detected by Mutect. Mutect2 has the potential to detect variants at lower VAFs than most other comparable variant callers (Xu, 2018, Wang et al., 2019), which may in part explain the discrepancy and larger number of calls generated.

The most common mutations were C>T transitions, which accounted for around 40% of base substitutions in candidate driver genes. Based on the trinucleotide context of all identified mutations, the two predominant mutational signatures were COSMIC signature 1, associated with ageing, and COSMIC signature 25, of unknown aetiology. The cosine similarity index was around 0.6 for both signatures, indicating some similarities, but not a strong match. Relatedly, exome, unlike whole genome sequencing, does restrict the number of events that can be used for mutational signature profiling so these data only permitted a limited analysis of mutational signatures.

The targeted analysis confirmed SNVs (n=34) were more common than indels (n=14) in leukaemia-related genes. Two cases were affected by both an SNV and indel in the same gene ("Multi-hit") (figure 6.8 and supplementary table 11). The affected genes were *IL7R* (case 24890) and *CREBBP* (case 27833). The *CREBBP* variants were seen in the same sequencing reads and were therefore present on the same allele. The *IL7R* variants were too far apart to verify whether they were present on different alleles.

The combined exome and targeted sequencing cohort consisted of 36 patients, which represented the major genetic subgroups of adult ALL, permitting a relatively comprehensive analysis of the genomic landscape of ALL in older adults. *TP53* and *NF1* were the most frequently mutated genes, and were almost exclusively seen in low hypodiploid cases. Interestingly, *NF1* seemed particularly susceptible to indel mutations in this subtype. A single additional *NF1* variant was identified in an *IGH-CRLF2* patient, although the diagnostic karyotype also showed large scale ploidy shift. Both genes are tumour suppressors and are present on chromosome 17, which almost always becomes monosomic in low hypodiploidy, so these events will very likely have produced a profound impact on protein function. Given that these mutations occur specifically in HoTr ALL and that this entity was slightly over-represented in this NGS cohort (19% vs 14% of cases in the unselected clinical trial population of older adults in chapter 3), it is difficult to be certain of the true incidence of *TP53* and *NF1* mutations in older adults with ALL.

Pathogenic mutations in components of the RAS signalling pathway were present in 31% (11/36) of patients and affected *NF1* (n=4), *KRAS* (n=3), *NRAS* (n=3) and *PTPN11* (n=1). The variants were identified in patients with a range of primary abnormalities, specifically low hypodiploidy (n=4), *IGH-CRLF2* (n=2), *EP300-ZNF384* (n=1), *KMT2A-AFF1* (n=1), B-other (n=1) and unknown (n=1). Interestingly, RAS pathway mutations were not seen in *BCR-ABL1*+ patients. The majority of the variants detected are known to activate RAS signalling (Vatansever et al., 2019, Wang et al., 2013, Niihori et al., 2005), which leads to uncontrolled cellular proliferation and malignant transformation (Downward, 2003).

JAK2 mutations were very strongly associated with *IGH-CRLF2* and were seen in all three patients with this gene rearrangement, as well as one additional patient without an identified primary chromosomal abnormality. The co-occurrence of *CRLF2* overexpression and activating *JAK2* mutations is known to result in constitutive JAK-STAT signalling, which is a feature of this genetic subtype (Mullighan et al., 2009a).

The only gene recurrently mutated in *BCR-ABL1+* patients in this cohort was *KDM6A*, which was disrupted in two cases (28670 and 25082). Interestingly, both cases had multiple additional chromosomal abnormalities visible on karyotype. No *BCR-ABL1* kinase domain mutations were identified, which are known to confer resistance to imatinib (Soverini et al., 2006).

Interestingly, a mutation in *NT5C2* was seen at a high VAF in the diagnostic sample for patient 30643 (figure 6.4). This gene is commonly implicated in the relapse setting, whereby mutations confer resistance to thiopurines, which play a major role in the management of ALL in all age groups, throughout the treatment schedule. Unlike many other relapse-founder mutations, these events are not reported at diagnosis, purely emerging through therapy following exposure to thiopurines (Dieck and Ferrando, 2019). The clinical significance of identifying pathogenic *NT5C2* mutations in diagnostic samples is not clear but could conceivably signpost the need for increased monitoring for the emergency of treatment resistant clones. The variant identified (*NT5C2* p.Y348D) is not reported in the literature, although is predicted as deleterious and probably damaging to protein function in the SIFT and PolyPhen databases respectively.

Overall, the mutational profile of older adults with ALL correlated with the primary chromosomal abnormalities and largely mirrored the aberrations seen in younger patients. As the exome sequencing cohort was small and the targeted panel was based on known leukaemia genes, the potential for discovery of new lesions was fairly limited. Interestingly, the three genes most commonly mutated in clonal haematopoiesis (*DNMT3A*, *TET2* and *ASXL1*) were only mutated in 8% of patients, which is very similar to the background rate in age-matched controls (Jaiswal et al., 2014). This suggests that clonal haematopoiesis is unlikely to play a significant role in the pathogenesis of ALL in older adults.

Chapter 7. Discussion

ALL in older adults is an area of unmet clinical need with very poor prognosis and high treatment related morbidity. To date, the genetic and genomic characterisation of these patients has been very limited, mainly focussing on the frequency of *BCR-ABL1* positive disease. This study provides the largest genetic characterisation to date of older adults with ALL. Median patient age was 64 years, and a quarter of patients were over 70 years old.

7.1 Genetic and genomic landscape of ALL in older adults

The landscape of primary chromosomal lesions and secondary copy number abnormalities is clearly distinct from that observed in children and younger adults. The incidence of T-ALL peaks at 20-29 years of age and then rapidly drops thereafter (Marks et al., 2009). This effect seems to continue in later life, with a reported T-cell disease frequency of 14% in ALL patients aged 55-65 years (Sive et al., 2012), dropping further to only 5% in this study.

Overall, 26% of patients in this study were *BCR-ABL1* positive. There was no evidence of increasing frequency of Philadelphia positivity with advancing age over the age of 60 years, which supports the findings from the large German study where this was seen to plateau after 45 years of age (Burmeister et al., 2008). In comparison, this study is the first to demonstrate that low hypodiploidy/near triploidy becomes more frequent with advancing age, whereby it is encountered in <2% of childhood patients, 4-9% of adults aged 25-60 (Safavi and Paulsson, 2017, Moorman et al., 2019) rising to around 15% of adults aged 60 years and over (chapter 3). Other high risk cytogenetic subgroups, specifically *KMT2A* fusions and complex karyotypes were present in 6% and 3% of patients respectively, indicating a frequency very similar to that seen in younger adults (Moorman et al., 2019, Moorman et al., 2007). High hyperdiploidy was very rare in this patient cohort (<1%) and *ETV6-RUNX1* remains unreported in older patients. The cytogenetically cryptic primary chromosomal

abnormalities *CRLF2*, *ZNF384* and *MEF2D* rearrangements were identified in 5%, 2% and <1% of the complete patient cohort respectively.

Copy number abnormalities in key genes recurrently disrupted in ALL were discovered in the majority of patients. *IKZF1* loss was present in over half of all cases tested by SNP array, affecting 70% of *BCR-ABL1* positive and 40% of *BCR-ABL1* negative patients. The high rate of *IKZF1* loss in *BCR-ABL1*+ ALL is consistent with much of the published literature (Mullighan et al., 2008a, Fedullo et al., 2019). However, the frequency of *IKZF1* deletion in Philadelphia negative patients was double that reported in younger adults (Moorman et al., 2012). This discrepancy is at least in part driven by the increased frequency of low hypodiploidy in older adults. Deletions in other key driver genes in BCP-ALL (*CDKN2A/B*, *PAX5*, *RB1*, *ETV6* and *EBF1*) were also encountered more frequently than in younger adults (Moorman et al., 2019).

Interestingly, *KDM6A* deletions were discovered in 5% (4/83) of patients who underwent SNP array analysis, specifically low hypodiploid (n=2), B-other (n=1) and T-ALL (n=1) patients (chapter 4). Inactivating *KDM6A* mutations were also seen in 6% (2/36) of patients who had NGS analyses, exclusively in *BCR-ABL1* positive patients (chapter 6). *KDM6A* is well recognised as a tumour suppressor gene, involved in epigenetic regulation through repression of PRC2/EZH2 activity. For the first time, *KDM6A* disruption has been demonstrated in a significant proportion of ALL patients (~10% of screened patients combining SNP array and NGS analyses). Overall, this therefore represents a proportion of patients who may respond to EZH2 inhibition, particularly if *KDM6A* inactivation is a truncal event in leukaemogenesis of affected cases.

RAS pathway activating mutations are recurrent abnormalities in many malignancies. These have been found to be prevalent in paediatric BCP-ALL and are particularly enriched in certain subgroups such as high hyperdiploidy (Jerchel et al., 2018, Paulsson et al., 2015). In this study, of the 36 patient samples analysed by NGS techniques, RAS pathway mutations were identified in one third (11/36). Childhood ALL studies have reported frequencies of RAS pathway mutations ranging 30-45% (Case et al., 2008, Irving et al., 2014, Jerchel et al., 2018). In such cohorts, the frequency of RAS mutations is partly driven by the prevalence of high hyperdiploidy, which is particularly strongly associated with these abnormalities (Paulsson et al., 2015). To date, comparable genomic studies in adult ALL are lacking. However, this study (chapter 6) has demonstrated that the high rate of RAS pathway mutations

continues in older adults with ALL, where they were identified in a range of genetic subgroups with the notable exception of *BCR-ABL1* positive disease.

7.2 Prognostic biomarkers and druggable lesions

Chromosomal abnormalities associated with a poor prognosis are over-represented in older adults and this is accompanied by a steep decrease in those associated with a favourable outcome. The most important prognostic and predictive biomarker in older adults is the presence of *BCR-ABL1* translocation. Although *BCR-ABL1* positivity is usually associated with high risk disease, and considered an indication for allografting in first complete remission, Philadelphia positive older adults fare better than their Philadelphia negative counterparts (Gökbuget, 2013, Ribera et al., 2016). The advent of TKIs has permitted these patients to achieve complete remission with a significantly reduced burden of cytotoxic chemotherapy and its associated toxicity (Ottmann et al., 2007, Rousselot et al., 2016). Indeed, recent combinations of TKI with immunotherapy-based treatment have yielded promising chemotherapy-free options for such patients (Foà et al., 2020), thereby further reducing the treatment related morbidity and mortality, which needs to be at the forefront of all advances in the management of older adults with ALL. Although such approaches produce excellent rates of complete and even molecular remissions, the rate of sustained disease-free survival in the absence of allogeneic stem cell transplantation is still not clear. Philadelphia positive ALL may retain its high eventual relapse rate, although this needs to be viewed in the context of remaining life-years and death from other causes.

The copy number status of 8 key genes/regions provides another prognostic biomarker, which is well validated in paediatric BCP-ALL. The high risk *IKZF1*^{plus} copy number profile was identified in over a third of patients in this study (chapter 4), although its prognostic impact in older adults still needs to be elucidated. Moreover, no focal *ERG* deletions were identified. The latter are associated with favourable outcome (Clappier et al., 2014) and specifically co-occur with *DUX4* rearrangements, which represent a more recently identified primary genetic abnormality (Zhang et al., 2016). Overall, these data confirm that all genetic biomarkers typically associated with a favourable outcome, namely *ETV6-RUNX1* fusion, high hyperdiploidy and *ERG*

deletions are exceedingly rare in older adults with ALL, contributing to the challenges of treating this patient population. Additionally, as outcome data from older adult ALL trials mature, it may be that the important genetic biomarkers are distinct from those applied to younger patient groups. As mentioned previously, the presence of *BCR-ABL1* fusion improves outcome in older adults. The prognosis of older adults may therefore be primarily driven by whether or not a targetable genetic lesion is present, rather than underlying chemoresponsiveness and early attainment of MRD negativity. Over the last few years, the advent of bispecific antibodies and immunoconjugates has further improved the tolerability and effects of treatments. Currently, these are principally used in the relapse setting but, as with TKIs, there is good evidence that use of such therapies permits dramatic reduction in the burden of cytotoxic chemotherapy whilst retaining effective responses. A study of the anti-CD22 immunoconjugate inotuzumab in combination with low intensity chemotherapy in older adults with ALL resulted in a 2-year progression free survival of 59% (Kantarjian et al., 2018) and a 'chemotherapy-free' schedule of dasatinib and blinatumomab has shown particular promise in Philadelphia positive disease (Foà et al., 2020), paving the way for antibody-small molecule inhibitor combinations in ALL.

Discovering gene rearrangements in genetically uncharacterised patients has the potential of identifying targetable lesions, as well as improving prognostication. *CRLF2* overexpression usually co-occurs with *JAK2* mutations and leads to JAK-STAT pathway activation. The JAK inhibitor ruxolitinib has shown activity in patients with deregulated JAK-STAT signalling (Ding et al., 2018, Jain et al., 2017a) and a phase 2 clinical trial of ruxolitinib is in progress for children with *CRLF2* rearranged or *JAK* pathway mutated ALL (INCB-18424-269). JAK inhibition may therefore provide a therapeutic avenue for older patient with these abnormalities.

The finding of focal biallelic *KDM6A* deletions provides another potential druggable target, although was only seen in a small number of cases. Importantly, as a transcriptional regulator, loss of KDM6A function results in EZH2 overactivity, which is important in the pathogenesis of several cancers. As such, the EZH2 inhibitor tazemetostat is being investigated for the treatment of advanced urothelial cancer with *KDM6A* loss of function (Ler et al., 2017), and is already FDA approved for relapsed-refractory follicular lymphoma with activating *EZH2* mutations. Future work is needed to identify whether such compounds are effective in the treatment of ALL with *KDM6A* abnormalities.

Several groups have demonstrated that leukaemic cells bearing clonal RAS mutations are sensitive to MEK inhibitors such as selumetinib (Ryan et al., 2016, Jerchel et al., 2018), thereby providing a further avenue for non-chemotherapeutic targeted treatment options in frailer patients. Given that RAS mutations were identified in high risk subgroups, often considered incurable without stem cell transplantation, such as low hypodiploidy and *KMT2A*-rearranged ALL, MEK inhibitors may be an attractive option in older patients with these abnormalities. Relatedly, a high degree of synergism has been identified between selumetinib and glucocorticoids (Matheson et al., 2019). In turn, this has led to a phase I/II clinical trial investigating the combination of selumetinib and dexamethasone in children and adults with relapsed ALL and RAS pathway mutations (SeluDex, ISRCTN92323261).

Even though certain cryptic gene rearrangements were identified in B-other ALL samples, an even greater number of patients remain genetically uncharacterised, posing an ongoing challenge for accurate prognostication and management. Techniques such as whole genome and transcriptome sequencing will likely be crucial for future discovery of novel gene rearrangements, although do require suitable material (often including matched germline samples) for their successful use. As such methods become more widely available and routinely integrated into clinical trials, more comprehensive characterisation of the genetic drivers of leukaemogenesis will become possible. Indeed, a large study combining genetic, genomic and transcriptomic analyses of BCP-ALL samples has identified 24 subtypes (Gu et al., 2019). Nine of these were defined by integration of gene expression profile, karyotype and/or mutations, highlighting that any single technique is not sufficient to identify all subgroups. Such novel subgroups may account for the B-other patients in chapter 3 that remain genetically uncharacterised using karyotype and FISH analyses.

7.3 Novel approach to the diagnosis of low hypodiploidy

Low hypodiploidy is one of the genetic entities associated with the poorest overall survival in all age groups (Pui et al., 1987, Charrin et al., 2004), and is the second most prevalent chromosomal abnormality in older adults (chapter 3). With the aim of shedding further light on this genetic entity, and specifically its distinction from high

hyperdiploidy, SNP arrays were performed on a large cohort of samples with ploidy shift. Unexpectedly, this revealed that low hypodiploid ALL is even more common than currently appreciated. A number of patients classified as high hyperdiploidy actually had masked low hypodiploidy, confirming that modal chromosome number alone is an inaccurate segregator of the two ploidy groups. The rate of misclassification was particularly high in older adults, and crucially highlight that low hypodiploidy may be even more prevalent than is currently reported in the ALL cytogenetic literature. Importantly, characteristic fluctuations in log2 ratios on a chromosomal level have facilitated the development of a diagnostic classifier (chapter 5), which could aid future accurate classification of such cases. This analysis has been based on the largest SNP array cohort of low hypodiploid/near triploid samples reported to date, and is particularly relevant to future prognostication of adult ALL patients. Such diagnostic issues can produce a significant impact on therapeutic decisions and these findings should in particular prompt caution with the categorisation of high hyperdiploidy in adult patients.

This analysis also raised interesting biological observations into the relationship between low hypodiploidy and near triploidy. Masked low hypodiploidy is thought to arise through duplication of a low hypodiploid clone, producing cells with a near triploid karyotype. Such karyotypes would typically be expected to harbour disomies and tetrasomies, having arisen through chromosomal doubling. Interestingly, a number of otherwise characteristic masked low hypodiploid cases by SNP array had widespread trisomies. Although this calls into question the underlying mechanism through which near triploidy occurs, the presence of pathogenic *TP53* mutations and the unsupervised clustering results support low hypodiploid and near triploid cases being biologically related. Matched germline samples were not available so the origin of the *TP53* mutations could not confirmed, although previous studies suggest these are somatic in adults with low hypodiploid ALL (Mühlbacher et al., 2014). As well as *TP53* variants, other mutations were discovered in low hypodiploid samples. Affected genes were frequently also located on chromosome 17, namely *NF1* (n=4) and *IKZF3* (n=1), supporting the hypothesis that chromosomal loss unmasks recessive alleles (Safavi and Paulsson, 2017).

7.4 Limitations

As a project performed entirely on primary patient material, experiments were highly dependent on the availability of suitable samples. FISH analyses required fixed cell samples that had been obtained from diagnostic cytogenetic laboratories throughout the UK. In a number of cases, such samples were either not available or had insufficient material to perform all required FISH tests.

Similarly, SNP arrays were limited by availability of DNA extracted from the bone marrow aspirate at ALL diagnosis. A number of these samples had already been used for standard-of-care procedures such as MRD marker identification, and were often already depleted. As such, not all potential novel CNAs could be included in the subsequent SureSelect XT2 validation cohort.

SNP arrays were performed on two different platforms, with significant differences in probe density. Although they were analysed using the same software, platform-specific adjustments of segmentation settings had to be made to try to minimise bias. Despite this, there was still a small difference between the average number of calls generated through Affymetrix and Illumina arrays that is unlikely to be biologically driven. Indeed, the Affymetrix arrays had poorer quality scores and may have produced a higher rate of false positive calls. It is therefore possible that the Affymetrix array performed less well with poorer quality material, although operator and protocol familiarity issues could have also played a part. For greater consistency, the study would have ideally been performed using a single SNP array platform.

Next generation sequencing techniques that cover large amounts of the human genome such as exome or whole genome sequencing, are reliant on the provision of matched germline samples to exclude benign constitutional variants. In this project, germline DNA was extracted from buccal cells. Unfortunately, the quality of this DNA was highly variable. Despite repeated buccal sampling, a sufficient amount of suitable quality germline DNA was only obtained in two out of six patients in the exome cohort. Fortunately, remission bone marrow DNA was available to provide an acceptable matched germline alternative in the other patients. Given the depth of coverage (90x) used for sequencing these remission samples, it is unlikely that low level leukaemic clones will have been detected, and therefore incorrectly flagged as constitutional variants. However, it is possible that other somatic mutations (e.g. variants associated

with clonal haematopoiesis) could have been filtered out if they were detected both in remission and diagnostic DNA. Indeed, clonal haematopoiesis is known to frequently persist in remission in patients treated for AML (Tanaka et al., 2019).

Throughout the majority of this project, older adults have been defined as those aged 60 years and over at diagnosis of ALL, matching cut-offs used regularly in clinical trials and the majority of the existing literature (Gökbuget, 2013, Thomas et al., 2001). All patients included in chapters 3 and 4 were therefore selected specifically from this age group. However, chapter 6 required prospective collection of diagnostic and matched germline samples from locally recruited patients. To maximise enrolment for biobanking, patients aged 50 years and over were approached. In total, two out of the six patients in the exome sequencing cohort were therefore aged 50-60 years, with the remaining four patients aged over 60 years at diagnosis. Similarly, chapter 5 detailed SNP array-based signatures of genetic ploidy groups, with the primary objective of accurately identifying masked low hypodiploid cases, which is particularly important in older adults due to its prevalence. To obtain sufficient cases for a robust analysis, younger adults and children were also included, principally to create the high hyperdiploid comparison cohort, which would not have been possible with older adults alone.

7.5 COVID-19 impact on this project

The COVID-19 pandemic had a large impact on the work detailed in chapter 6 of this thesis. The aims of this work package had centred around using NGS techniques to track leukaemia-associated mutations through treatment. Patient recruitment and biobanking of samples had commenced in 2017 and a total of 22 samples had been obtained (6 diagnostic, 16 follow up). Following analysis of somatic mutations from exome sequencing, a customised sequencing panel using molecular barcoding was designed and manufactured to track and quantify these events in the remission samples. Unfortunately, immediately prior to commencing the library prep workflow, all laboratory work was suspended due to the COVID-19 pandemic.

Chapter 6 was therefore modified to focus on the mutational landscape of ALL in older adults. Together with the exome samples, the patient samples used in the NGS

validation of CNAs from chapter 4 had mutational analyses performed by virtue of the technique used. Due to the small number of RNA baits needed to cover exons alone, a number of genes could be included in this experiment, together with the larger regions required for validation of copy number breakpoints. The resulting gene panel covered most of the recurrently mutated genes in ALL (supplementary tables 3 and 4) and therefore permitted a broader mutational characterisation of ALL in older adults than had originally been intended, as an alternative to the planned clonal tracking experiment.

7.6 Future work

Overall, 37% (68/184) of older adults with BCP-ALL in the UKALL14 and UKALL60+ clinical trial cohort remain uncharacterised in terms of primary genetic abnormalities. Such cases are likely to benefit from the use of non-targeted techniques for the discovery of fusion genes or initiating mutations such as whole genome or whole transcriptome sequencing.

Several novel focal gene deletions were discovered and validated. In particular, the *KDM6A* deletions provide a plausible driver abnormality meriting further investigation. In the first instance, this would need to be functionally assessed, for example, by performing gene knockdown experiments in a cell line model, including evaluation of *EZH2* expression. The therapeutic effects of EZH2 inhibitors could then also be assessed.

With outcome data, it will be important to correlate genetic subgroups and CNAs with prognosis. In particular, the prognostic value of *BCR-ABL1* positivity compared to other subgroups, in patients who are frequently unsuitable for intensive treatments, will provide valuable insight into the benefit of targeted disease-modifying therapies in older patients. However, as the patients were treated in two separate trials, one of which (UKALL60+) offered four separate arms with different treatment intensities, a number of additional confounders will need to be taken into account to produce accurate subgroup-specific survival analyses.

The SNP array characterisation of low hypodiploid ALL identified masked cases with unexpectedly low modal chromosome numbers that had been incorrectly classified as

high hyperdiploidy. This raises interesting uncertainties around the presumed duplication process and the formation of such near triploid karyotypes, particularly where trisomies rather than tetrasomies predominate. Addressing these issues would likely benefit from single cell genomic analyses. Specific comparisons will need to be made between the three main phenotypes of the low hypodiploid subgroup, namely patients presenting with a low hypodiploid clone alone, those with both low hypodiploid and near triploid clones and those with only a near triploid clone.

Part of this work has already been presented in abstract format at the 61st ASH Annual Meeting (Creasey et al., 2019a, Creasey et al., 2019b), and granted an ASH Abstract Achievement Award. The work detailed in chapter 5 is also currently under consideration for publication.

7.7 Final conclusion

With an ageing population, the prevalence of older adults with ALL will increase. Optimising treatment for these patients poses a unique set of challenges, which are less regularly encountered in other age groups, namely that the existing approaches and aims of treatment, are often unsuitable for older individuals. Shifting the focus from traditional disease eradication through high intensity therapy to disease control through targeted therapy is of particular importance in these patients, and has already greatly benefited *BCR-ABL1* positive patients. Pursuing other targets for personalised therapy based on the genetic and genomic aberrations encountered, will ultimately lead to a reduction in treatment-related morbidity and mortality and prolong survival. To this end, a greater understanding of the mechanisms of ALL leukaemogenesis in older adults will be required. Age-related clonal haematopoiesis does not appear to feature strongly in the pre-leukaemic background and the pre-natal lesions of childhood ALL are implausible initiating abnormalities. There is a clear association between specific high-risk genetic features and advancing age at diagnosis, and molecular characterisation will have the potential to transform front-line treatment in a group of patients likely to derive the greatest benefit from novel approaches.

References

ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nature methods,* 7**,** 248-249.

AKASAKA, T., BALASAS, T., RUSSELL, L. J., SUGIMOTO, K.-J., MAJID, A., WALEWSKA, R., KARRAN, E. L., BROWN, D. G., CAIN, K., HARDER, L., GESK, S., MARTIN-SUBERO, J. I., ATHERTON, M. G., BRÜGGEMANN, M., CALASANZ, M. A. J., DAVIES, T., HAAS, O. A., HAGEMEIJER, A., KEMPSKI, H., LESSARD, M., LILLINGTON, D. M., MOORE, S., NGUYEN-KHAC, F., RADFORD-WEISS, I., SCHOCH, C., STRUSKI, S. P., TALLEY, P., WELHAM, M. J., WORLEY, H., STREFFORD, J. C., HARRISON, C. J., SIEBERT, R. & DYER, M. J. S. 2006. Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). *Blood,* 109**,** 3451-3461.

ALEXANDROV, L. B., JONES, P. H., WEDGE, D. C., SALE, J. E., CAMPBELL, P. J., NIK-ZAINAL, S. & STRATTON, M. R. 2015. Clock-like mutational processes in human somatic cells. *Nature Genetics,* 47**,** 1402-1407.

ALEXANDROV, L. B., NIK-ZAINAL, S., WEDGE, D. C., APARICIO, S. A. J. R., BEHJATI, S., BIANKIN, A. V., BIGNELL, G. R., BOLLI, N., BORG, A., BØRRESEN-DALE, A.-L., BOYAULT, S., BURKHARDT, B., BUTLER, A. P., CALDAS, C., DAVIES, H. R., DESMEDT, C., EILS, R., EYFJÖRD, J. E., FOEKENS, J. A., GREAVES, M., HOSODA, F., HUTTER, B., ILICIC, T., IMBEAUD, S., IMIELINSKI, M., JÄGER, N., JONES, D. T. W., JONES, D., KNAPPSKOG, S., KOOL, M., LAKHANI, S. R., LÓPEZ-OTÍN, C., MARTIN, S., MUNSHI, N. C., NAKAMURA, H., NORTHCOTT, P. A., PAJIC, M., PAPAEMMANUIL, E., PARADISO, A., PEARSON, J. V., PUENTE, X. S., RAINE, K., RAMAKRISHNA, M., RICHARDSON, A. L., RICHTER, J., ROSENSTIEL, P., SCHLESNER, M., SCHUMACHER, T. N., SPAN, P. N., TEAGUE, J. W., TOTOKI, Y., TUTT, A. N. J., VALDÉS-MAS, R., VAN BUUREN, M. M., VAN 'T VEER, L., VINCENT-SALOMON, A., WADDELL, N., YATES, L. R., ZUCMAN-ROSSI, J., ANDREW FUTREAL, P., MCDERMOTT, U., LICHTER, P., MEYERSON, M., GRIMMOND, S. M., SIEBERT, R., CAMPO, E., SHIBATA, T., PFISTER, S. M., CAMPBELL, P. J., STRATTON, M. R., AUSTRALIAN PANCREATIC CANCER GENOME, I., CONSORTIUM, I. B. C., CONSORTIUM, I. M.-S. & PEDBRAIN, I. 2013. Signatures of mutational processes in human cancer. *Nature,* 500**,** 415-421.

ALPAR, D., WREN, D., ERMINI, L., MANSUR, M. B., VAN DELFT, F. W., BATEMAN, C. M., TITLEY, I., KEARNEY, L., SZCZEPANSKI, T., GONZALEZ, D., FORD, A. M., POTTER, N. E. & GREAVES, M. 2014. Clonal origins of ETV6-RUNX1+ acute lymphoblastic leukemia: studies in monozygotic twins. *Leukemia,* 29**,** 839.

ARBER, D. A., ORAZI, A., HASSERJIAN, R., THIELE, J., BOROWITZ, M. J., LE BEAU, M. M., BLOOMFIELD, C. D., CAZZOLA, M. & VARDIMAN, J. W. 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood,* 127**,** 2391-2405.

BARZ, M. J., HOF, J., GROENEVELD-KRENTZ, S., LOH, J. W., SZYMANSKY, A., ASTRAHANTSEFF, K., VON STACKELBERG, A., KHIABANIAN, H., FERRANDO, A. A., ECKERT, C. & KIRSCHNER-SCHWABE, R. 2020. Subclonal NT5C2 mutations are associated with poor outcomes after relapse of pediatric acute lymphoblastic leukemia. *Blood,* 135**,** 921-933.

BASSAN, R., PAVONI, C., INTERMESOLI, T., SPINELLI, O., TOSI, M., AUDISIO, E., MARMONT, F., CATTANEO, C., BORLENGHI, E., CORTELAZZO, S., CAVATTONI, I., FUMAGALLI, M., MATTEI, D., ROMANI, C., CORTELEZZI, A., FRACCHIOLLA, N., CICERI, F., BERNARDI, M., SCATTOLIN, A. M., DEPAOLI, L., MASCIULLI, A., OLDANI, E. & RAMBALDI, A. 2020. Updated risk-oriented strategy for acute lymphoblastic leukemia in adult patients 18–65 years: NILG ALL 10/07. *Blood Cancer Journal,* 10**,** 119.

BAUGHN, L. B., BIEGEL, J. A., SOUTH, S. T., SMOLAREK, T. A., VOLKERT, S., CARROLL, A. J., HEEREMA, N. A., RABIN, K. R., ZWEIDLER-MCKAY, P. A., LOH, M. & HIRSCH, B. 2015. Integration of cytogenomic data for furthering the characterization of pediatric B-cell acute lymphoblastic leukemia: a multi-institution, multi-platform microarray study. *Cancer genetics,* 208**,** 1-18.

BENJAMIN, D., SATO, T., CIBULSKIS, K., GETZ, G., STEWART, C. & LICHTENSTEIN, L. 2019. Calling Somatic SNVs and Indels with Mutect2. *bioRxiv***,** 861054.

BERNARDINI, L., ALESI, V., LODDO, S., NOVELLI, A., BOTTILLO, I., BATTAGLIA, A., DIGILIO, M. C., ZAMPINO, G., ERTEL, A., FORTINA, P., SURREY, S. & DALLAPICCOLA, B. 2010. High-resolution SNP arrays in mental retardation diagnostics: how much do we gain? *European journal of human genetics : EJHG,* 18**,** 178-185.

BEROUKHIM, R., MERMEL, C. H., PORTER, D., WEI, G., RAYCHAUDHURI, S., DONOVAN, J., BARRETINA, J., BOEHM, J. S., DOBSON, J., URASHIMA, M., MC HENRY, K. T., PINCHBACK, R. M., LIGON, A. H., CHO, Y.-J., HAERY, L., GREULICH, H., REICH, M., WINCKLER, W., LAWRENCE, M. S., WEIR, B. A., TANAKA, K. E., CHIANG, D. Y., BASS, A. J., LOO, A., HOFFMAN, C., PRENSNER, J., LIEFELD, T., GAO, Q., YECIES, D., SIGNORETTI, S., MAHER, E., KAYE, F. J., SASAKI, H., TEPPER, J. E., FLETCHER, J. A., TABERNERO, J., BASELGA, J., TSAO, M.-S., DEMICHELIS, F., RUBIN, M. A., JANNE, P. A., DALY, M. J., NUCERA, C., LEVINE, R. L., EBERT, B. L., GABRIEL, S., RUSTGI, A. K., ANTONESCU, C. R., LADANYI, M., LETAI, A., GARRAWAY, L. A., LODA, M., BEER, D. G., TRUE, L. D., OKAMOTO, A., POMEROY, S. L., SINGER, S., GOLUB, T. R., LANDER, E. S., GETZ, G., SELLERS, W. R. & MEYERSON, M. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature,* 463**,** 899.

BIESECKER, L. G. & SPINNER, N. B. 2013. A genomic view of mosaicism and human disease. *Nature Reviews Genetics,* 14**,** 307-320.

BISHOP, R. 2010. Applications of fluorescence in situ hybridization (FISH) in detecting genetic aberrations of medical significance. *Bioscience Horizons: The International Journal of Student Research,* 3**,** 85-95.

BLOOMFIELD, C. D., GOLDMAN, A. I., ALIMENA, G., BERGER, R., BORGSTROM, G. H., BRANDT, L., CATOVSKY, D., DE LA CHAPELLE, A., DEWALD, G. W. & GARSON, O. M. 1986. Chromosomal abnormalities identify high-risk and low-risk patients with acute lymphoblastic leukemia. *Blood,* 67**,** 415-420.

BOER, J. M., STEEGHS, E. M. P., MARCHANTE, J. R. M., BOEREE, A., BEAUDOIN, J. J., BERNA BEVERLOO, H., KUIPER, R. P., ESCHERICH, G., VAN DER VELDEN, V. H. J., VAN DER SCHOOT, C. E., DE GROOT-KRUSEMAN, H. A., PIETERS, R. & DEN BOER, M. L. 2017. Tyrosine kinase fusion genes in pediatric BCR-ABL1-like acute lymphoblastic leukemia. *Oncotarget,* 8**,** 4618-4628.

BOER, J. M., VAN DER VEER, A., RIZOPOULOS, D., FIOCCO, M., SONNEVELD, E., DE GROOT-KRUSEMAN, H. A., KUIPER, R. P., HOOGERBRUGGE, P., HORSTMANN, M., ZALIOVA, M., PALMI, C., TRKA, J., FRONKOVA, E., EMERENCIANO, M., DO SOCORRO POMBO-DE-OLIVEIRA, M., MLYNARSKI, W., SZCZEPANSKI, T., NEBRAL, K., ATTARBASCHI, A., VENN, N., SUTTON, R., SCHWAB, C. J., ENSHAEI, A., VORA, A., STANULLA, M., SCHRAPPE, M., CAZZANIGA, G., CONTER, V., ZIMMERMANN, M., MOORMAN, A. V., PIETERS, R. & DEN BOER, M. L. 2016. Prognostic value of rare IKZF1 deletion in childhood B-cell precursor acute lymphoblastic leukemia: an international collaborative study. *Leukemia,* 30**,** 32-38.

BOS, J. L. 1989. *ras* oncogenes in human cancer: a review. *Cancer Research,* 49**,** 4682.

BRANDWEIN, J. M., ATENAFU, E. G., SCHUH, A. C., YEE, K. W. L., SCHIMMER, A. D., GUPTA, V. & MINDEN, M. D. 2014. Predictors of outcome in adults with BCR-ABL negative acute lymphoblastic leukemia treated with a pediatric-based regimen. *Leukemia Research,* 38**,** 532-536.

BROWN, P. 2013. Treatment of infant leukemias: challenge and promise. *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program,* 2013**,** 596-600.

BRÜGGEMANN, M. & KOTROVA, M. 2017. Minimal residual disease in adult ALL: technical aspects and implications for correct clinical interpretation. *Blood advances,* 1**,** 2456-2466.

BRÜGGEMANN, M., RAFF, T., FLOHR, T., GÖKBUGET, N., NAKAO, M., DROESE, J., LÜSCHEN, S., POTT, C., RITGEN, M., SCHEURING, U., HORST, H.-A., THIEL, E., HOELZER, D., BARTRAM, C. R. & KNEBA, M. 2006. Clinical significance of minimal residual disease quantification in adult patients with standard-risk acute lymphoblastic leukemia. *Blood,* 107**,** 1116-1123.

BUCKLEY, A. R., STANDISH, K. A., BHUTANI, K., IDEKER, T., LASKEN, R. S., CARTER, H., HARISMENDY, O. & SCHORK, N. J. 2017. Pan-cancer analysis reveals technical artifacts in TCGA germline variant calls. *BMC genomics,* 18**,** 458-458.

BURMEISTER, T., GÖKBUGET, N., SCHWARTZ, S., FISCHER, L., HUBERT, D., SINDRAM, A., HOELZER, D. & THIEL, E. 2010. Clinical features and prognostic implications of TCF3-PBX1 and ETV6-RUNX1 in adult acute lymphoblastic leukemia. *Haematologica,* 95**,** 241-246.

BURMEISTER, T., SCHWARTZ, S., BARTRAM, C. R., GÖKBUGET, N., HOELZER, D. & THIEL, E. 2008. Patients' age and BCR-ABL frequency in adult B-precursor ALL: a retrospective analysis from the GMALL study group. *Blood,* 112**,** 918-919.

BYRD, J. C., MRÓZEK, K., DODGE, R. K., CARROLL, A. J., EDWARDS, C. G., ARTHUR, D. C., PETTENATI, M. J., PATIL, S. R., RAO, K. W., WATSON, M. S., KODURU, P. R. K., MOORE, J. O., STONE, R. M., MAYER, R. J.,

FELDMAN, E. J., DAVEY, F. R., SCHIFFER, C. A., LARSON, R. A. & BLOOMFIELD, C. D. 2002. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461): Presented in part at the 43rd annual meeting of the American Society of Hematology, Orlando, FL, December 10, 2001, and published in abstract form.59. *Blood,* 100**,** 4325-4336.

BYUN, J. M., KOH, Y., SHIN, D.-Y., KIM, I., YOON, S.-S., LEE, J.-O., BANG, S.-M., KIM, K. H., JUNG, S.-H., LEE, W. S., PARK, Y., JANG, J. H., HAN, J. J., YHIM, H.-Y., KIM, D. S., LEE, Y. J., LEE, H., CHOI, Y.-S., LEE, S. & KOREAN ADULT ALL WORKING PARTY, K. S. O. H. 2017. BCR-ABL translocation as a favorable prognostic factor in elderly patients with acute lymphoblastic leukemia in the era of potent tyrosine kinase inhibitors. *Haematologica,* 102**,** e187-e190.

CAMPANA, D. 2010. Minimal Residual Disease in Acute Lymphoblastic Leukemia. *Hematology,* 2010**,** 7-12.

CANCER RESEARCH UK. 2016. *Acute lymphoblastic leukaemia (ALL) incidence statistics* [Online]. Cancer Research UK. Available: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia-all/incidence#heading-One [Accessed 02/01/18 2018].

CARROLL, A. J., SHAGO, M., MIKHAIL, F. M., RAIMONDI, S. C., HIRSCH, B. A., LOH, M. L., RAETZ, E. A., BOROWITZ, M. J., WOOD, B. L., MALONEY, K. W., MATTANO, L. A., LARSEN, E. C., GASTIER-FOSTER, J., STONEROCK, E., ELL, D., KAHWASH, S., DEVIDAS, M., HARVEY, R. C., CHEN, I. M. L., WILLMAN, C. L., HUNGER, S. P., WINICK, N. J., CARROLL, W. L., RAO, K. W. & HEEREMA, N. A. 2019. Masked hypodiploidy: Hypodiploid acute lymphoblastic leukemia (ALL) mimicking hyperdiploid ALL in children: A report from the Children's Oncology Group. *Cancer Genetics,* 238**,** 62-68.

CASE, M., MATHESON, E., MINTO, L., HASSAN, R., HARRISON, C. J., BOWN, N., BAILEY, S., VORMOOR, J., HALL, A. G. & IRVING, J. A. E. 2008. Mutation of Genes Affecting the RAS Pathway Is Common in Childhood Acute Lymphoblastic Leukemia. *Cancer Research,* 68**,** 6803.

CAVÉ, H., CACHEUX, V., RAYNAUD, S., BRUNIE, G., BAKKUS, M., COCHAUX, P., PREUDHOMME, C., LAÏ, J. L., VILMER, E. & GRANDCHAMP, B. 1997. ETV6 is the target of chromosome 12p deletions in t(12;21) childhood acute lymphocytic leukemia. *Leukemia,* 11**,** 1459-64.

CAVÉ, H., VAN DER WERFF TEN BOSCH, J., SUCIU, S., GUIDAL, C., WATERKEYN, C., OTTEN, J., BAKKUS, M., THIELEMANS, K., GRANDCHAMP, B., VILMER, E., NELKEN, B., FOURNIER, M., BOUTARD, P., LEBRUN, E., MÉCHINAUD, F., GARAND, R., ROBERT, A., DASTUGUE, N., PLOUVIER, E., RACADOT, E., FERSTER, A., GYSELINCK, J., FENNETEAU, O., DUVAL, M., SOLBU, G. & MANEL, A.-M. 1998. Clinical Significance of Minimal Residual Disease in Childhood Acute Lymphoblastic Leukemia. *New England Journal of Medicine,* 339**,** 591-598.

CHALANDON, Y., THOMAS, X., HAYETTE, S., CAYUELA, J.-M., ABBAL, C., HUGUET, F., RAFFOUX, E., LEGUAY, T., ROUSSELOT, P., LEPRETRE, S., ESCOFFRE-BARBE, M., MAURY, S., BERTHON, C., TAVERNIER, E., LAMBERT, J.-F., LAFAGE-POCHITALOFF, M., LHÉRITIER, V., CHEVRET, S., IFRAH, N., DOMBRET, H. & FOR THE GROUP FOR RESEARCH ON ADULT

ACUTE LYMPHOBLASTIC, L. 2015. Randomized study of reduced-intensity chemotherapy combined with imatinib in adults with Ph-positive acute lymphoblastic leukemia. *Blood,* 125**,** 3711-3719.

CHAPIRO, E., RUSSELL, L., RADFORD-WEISS, I., BASTARD, C., LESSARD, M., STRUSKI, S., CAVE, H., FERT-FERRER, S., BARIN, C., MAAREK, O., DELLA-VALLE, V., STREFFORD, J. C., BERGER, R., HARRISON, C. J., BERNARD, O. A., NGUYEN-KHAC, F. & THE GROUPE FRANCOPHONE DE CYTOGÉNÉTIQUE, H. M. 2006. Overexpression of CEBPA resulting from the translocation t(14;19)(q32;q13) of human precursor B acute lymphoblastic leukemia. *Blood,* 108**,** 3560-3563.

CHARRIN, C., THOMAS, X., FFRENCH, M., LE, Q.-H., ANDRIEUX, J., MOZZICONACCI, M.-J., LAÏ, J.-L., BILHOU-NABERA, C., MICHAUX, L., BERNHEIM, A., BASTARD, C., MOSSAFA, H., PEROT, C., MAAREK, O., BOUCHEIX, C., LHERITIER, V., DELANNOY, A., FIÈRE, D. & DASTUGUE, N. 2004. A report from the LALA-94 and LALA-SA groups on hypodiploidy with 30 to 39 chromosomes and near-triploidy: 2 possible expressions of a sole entity conferring poor prognosis in adult acute lymphoblastic leukemia (ALL). *Blood,* 104**,** 2444-2451.

CHAUHAN, S., GOODWIN, JINESH G., CHAUHAN, S., MANYAM, G., WANG, J., KAMAT, ASHISH M. & BOYD, DOUGLAS D. 2013. ZKSCAN3 Is a Master Transcriptional Repressor of Autophagy. *Molecular Cell,* 50**,** 16-28.

CHESSELLS, J. M., HALL, E., PRENTICE, H. G., DURRANT, J., BAILEY, C. C. & RICHARDS, S. M. 1998. The impact of age on outcome in lymphoblastic leukaemia; MRC UKALL X and XA compared: a report from the MRC Paediatric and Adult Working Parties. *Leukemia,* 12**,** 463-473.

CHESSELLS, J. M., RICHARDS, S. M., BAILEY, C. C., LILLEYMAN, J. S. & EDEN, O. B. 1995. Gender and treatment outcome in childhood lymphoblastic leukaemia: report from the MRC UKALL trials*. *British Journal of Haematology,* 89**,** 364-372.

CHIARETTI, S. & FOÀ, R. 2009. T-cell acute lymphoblastic leukemia. *Haematologica,* 94**,** 160-162.

CHILTON, L., BUCK, G., HARRISON, C. J., KETTERLING, R. P., ROWE, J. M., TALLMAN, M. S., GOLDSTONE, A. H., FIELDING, A. K. & MOORMAN, A. V. 2013. High hyperdiploidy among adolescents and adults with acute lymphoblastic leukaemia (ALL): cytogenetic features, clinical characteristics and outcome. *Leukemia,* 28**,** 1511.

CIBULSKIS, K., LAWRENCE, M. S., CARTER, S. L., SIVACHENKO, A., JAFFE, D., SOUGNEZ, C., GABRIEL, S., MEYERSON, M., LANDER, E. S. & GETZ, G. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology,* 31**,** 213.

CLAPPIER, E., AUCLERC, M. F., RAPION, J., BAKKUS, M., CAYE, A., KHEMIRI, A., GIROUX, C., HERNANDEZ, L., KABONGO, E., SAVOLA, S., LEBLANC, T., YAKOUBEN, K., PLAT, G., COSTA, V., FERSTER, A., GIRARD, S., FENNETEAU, O., CAYUELA, J. M., SIGAUX, F., DASTUGUE, N., SUCIU, S., BENOIT, Y., BERTRAND, Y., SOULIER, J. & CAVÉ, H. 2014. An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. *Leukemia,* 28**,** 70-77.

COCCARO, N., BRUNETTI, C., TOTA, G., PIERRI, C. L., ANELLI, L., ZAGARIA, A., CASIERI, P., IMPERA, L., MINERVINI, C. F., MINERVINI, A., CUMBO, C., RICCO, A., CARLUCCIO, P., ORSINI, P., SPECCHIA, G. & ALBANO, F. 2018. A novel t(3;9)(q21.2; p24.3) associated with SMARCA2 and ZNF148 genes rearrangement in myelodysplastic syndrome. *Leukemia & Lymphoma,* 59**,** 996-999.

COOLEY, L. D., LEBO, M., LI, M. M., SLOVAK, M. L., WOLFF, D. J. & COMMITTEE, A. W. G. O. T. A. C. O. M. G. A. G. L. Q. A. 2013. American College of Medical Genetics and Genomics technical standards and guidelines: microarray analysis for chromosome abnormalities in neoplastic disorders. *Genetics in Medicine,* 15**,** 484-494.

CREASEY, T., BARRETTA, E., KIRKWOOD, A. A., SCHWAB, C., PATRICK, P., CLIFTON-HADLEY, L., WRENCH, B., MENNE, T. F., MCMILLAN, A. K., MORLEY, N. J., SNOWDEN, J. A., HARRISON, C. J., LEONGAMORNLERT, D. A., PAPAEMMANUIL, E., ROWNTREE, C. J., MARKS, D. I., FIELDING, A. K. & MOORMAN, A. V. 2019a. Genetic and Genomic Characterisation of Older Adults with Acute Lymphoblastic Leukemia Treated on the UKALL14 and UKALL60+ Clinical Trials. *Blood,* 134**,** 2746-2746.

CREASEY, T., ENSHAEI, A., WATTS, K., CUTHBERT, G., SCHWAB, C., HARRISON, C. J., VORA, A., FIELDING, A. K. & MOORMAN, A. V. 2019b. Single Nucleotide Polymorphism Array-Based Signature of Genetic Ploidy Groups in Acute Lymphoblastic Leukemia. *Blood,* 134**,** 1473-1473.

DAMIANO, J. S., CRESS, A. E., HAZLEHURST, L. A., SHTIL, A. A. & DALTON, W. S. 1999. Cell adhesion mediated drug resistance (CAM-DR): role of integrins and resistance to apoptosis in human myeloma cell lines. *Blood,* 93**,** 1658-1667.

DE KOUCHKOVSKY, I. & ABDUL-HAY, M. 2016. Acute myeloid leukemia: a comprehensive review and 2016 update. *Blood Cancer Journal,* 6**,** e441-e441.

DEN BOER, M. L., VAN SLEGTENHORST, M., DE MENEZES, R. X., CHEOK, M. H., BUIJS-GLADDINES, J. G., PETERS, S. T., VAN ZUTVEN, L. J. C. M., BEVERLOO, H. B., VAN DER SPEK, P. J., ESCHERICH, G., HORSTMANN, M. A., JANKA-SCHAUB, G. E., KAMPS, W. A., EVANS, W. E. & PIETERS, R. 2009. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *The Lancet Oncology,* 10**,** 125-134.

DHÉDIN, N., HUYNH, A., MAURY, S., TABRIZI, R., BELDJORD, K., ASNAFI, V., THOMAS, X., CHEVALLIER, P., NGUYEN, S., COITEUX, V., BOURHIS, J.-H., HICHRI, Y., ESCOFFRE-BARBE, M., REMAN, O., GRAUX, C., CHALANDON, Y., BLAISE, D., SCHANZ, U., LHÉRITIER, V., CAHN, J.-Y., DOMBRET, H., IFRAH, N. & ON BEHALF OF THE, G. G. 2015. Role of allogeneic stem cell transplantation in adult patients with Ph-negative acute lymphoblastic leukemia. *Blood,* 125**,** 2486-2496.

DIECK, C. L. & FERRANDO, A. 2019. Genetics and mechanisms of NT5C2-driven chemotherapy resistance in relapsed ALL. *Blood,* 133**,** 2263-2268.

DING, L.-W., SUN, Q.-Y., TAN, K.-T., CHIEN, W., THIPPESWAMY, A. M., ENG JUH YEOH, A., KAWAMATA, N., NAGATA, Y., XIAO, J.-F., LOH, X.-Y., LIN, D.-C., GARG, M., JIANG, Y.-Y., XU, L., LIM, S.-L., LIU, L.-Z., MADAN, V., SANADA, M., FERNÁNDEZ, L. T., PREETHI, H., LILL, M., KANTARJIAN, H. M., KORNBLAU, S. M., MIYANO, S., LIANG, D.-C., OGAWA, S., SHIH, L.-Y.,

YANG, H. & KOEFFLER, H. P. 2017. Mutational Landscape of Pediatric Acute Lymphoblastic Leukemia. *Cancer Research,* 77**,** 390.

DING, Y. Y., STERN, J. W., JUBELIRER, T. F., WERTHEIM, G. B., LIN, F., CHANG, F., GU, Z., MULLIGHAN, C. G., LI, Y., HARVEY, R. C., CHEN, I. M., WILLMAN, C. L., HUNGER, S. P., LI, M. M. & TASIAN, S. K. 2018. Clinical efficacy of ruxolitinib and chemotherapy in a child with Philadelphia chromosome-like acute lymphoblastic leukemia with GOLGA5-JAK2 fusion and induction failure. *Haematologica,* 103**,** e427-e431.

DINMOHAMED, A. G., SZABÓ, A., VAN DER MARK, M., VISSER, O., SONNEVELD, P., CORNELISSEN, J. J., JONGEN-LAVRENCIC, M. & RIJNEVELD, A. W. 2016. Improved survival in adult patients with acute lymphoblastic leukemia in the Netherlands: a population-based study on treatment, trial participation and survival. *Leukemia,* 30**,** 310-317.

DISKIN, S. J., LI, M., HOU, C., YANG, S., GLESSNER, J., HAKONARSON, H., BUCAN, M., MARIS, J. M. & WANG, K. 2008. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic acids research,* 36**,** e126-e126.

DONG, M. & BLOBE, G. C. 2006. Role of transforming growth factor-beta in hematologic malignancies. *Blood,* 107**,** 4589-4596.

DOWNWARD, J. 2003. Targeting RAS signalling pathways in cancer therapy. *Nature Reviews Cancer,* 3**,** 11-22.

DUNFORD, A., WEINSTOCK, D. M., SAVOVA, V., SCHUMACHER, S. E., CLEARY, J. P., YODA, A., SULLIVAN, T. J., HESS, J. M., GIMELBRANT, A. A., BEROUKHIM, R., LAWRENCE, M. S., GETZ, G. & LANE, A. A. 2017. Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. *Nature genetics,* 49**,** 10-16.

EBBERT, M. T. W., WADSWORTH, M. E., STALEY, L. A., HOYT, K. L., PICKETT, B., MILLER, J., DUCE, J., ALZHEIMER'S DISEASE NEUROIMAGING, I., KAUWE, J. S. K. & RIDGE, P. G. 2016. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC bioinformatics,* 17 Suppl 7**,** 239-239.

EISFELD, A.-K., KOHLSCHMIDT, J., MRÓZEK, K., VOLINIA, S., BLACHLY, J. S., NICOLET, D., OAKES, C., KROLL, K., ORWICK, S., CARROLL, A. J., STONE, R. M., BYRD, J. C., DE LA CHAPELLE, A. & BLOOMFIELD, C. D. 2017. Mutational Landscape and Gene Expression Patterns in Adult Acute Myeloid Leukemias with Monosomy 7 as a Sole Abnormality. *Cancer research,* 77**,** 207-218.

FEDULLO, A. L., MESSINA, M., ELIA, L., PICIOCCHI, A., GIANFELICI, V., LAURETTI, A., SODDU, S., PUZZOLO, M. C., MINOTTI, C., FERRARA, F., MARTINO, B., CHIUSOLO, P., CALAFIORE, V., PAOLINI, S., VIGNETTI, M., VITALE, A., GUARINI, A., FOÀ, R. & CHIARETTI, S. 2019. Prognostic implications of additional genomic lesions in adult Philadelphia chromosome-positive acute lymphoblastic leukemia. *Haematologica,* 104**,** 312-318.

FERRANDO, A. A., NEUBERG, D. S., STAUNTON, J., LOH, M. L., HUARD, C., RAIMONDI, S. C., BEHM, F. G., PUI, C.-H., DOWNING, J. R., GILLILAND, D. G., LANDER, E. S., GOLUB, T. R. & LOOK, A. T. 2002. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell,* 1**,** 75-87.

FIELDING, A. K., ROWE, J. M., BUCK, G., FORONI, L., GERRARD, G., LITZOW, M. R., LAZARUS, H., LUGER, S. M., MARKS, D. I., MCMILLAN, A. K., MOORMAN, A. V., PATEL, B., PAIETTA, E., TALLMAN, M. S. & GOLDSTONE, A. H. 2014. UKALLXII/ECOG2993: addition of imatinib to a standard treatment regimen enhances long-term outcomes in Philadelphia positive acute lymphoblastic leukemia. *Blood,* 123**,** 843-850.

FORD, A. M., BENNETT, C. A., PRICE, C. M., BRUIN, M. C., VAN WERING, E. R. & GREAVES, M. 1998. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 4584-4588.

FOÀ, R., BASSAN, R., VITALE, A., ELIA, L., PICIOCCHI, A., PUZZOLO, M.-C., CANICHELLA, M., VIERO, P., FERRARA, F., LUNGHI, M., FABBIANO, F., BONIFACIO, M., FRACCHIOLLA, N., DI BARTOLOMEO, P., MANCINO, A., DE PROPRIS, M.-S., VIGNETTI, M., GUARINI, A., RAMBALDI, A. & CHIARETTI, S. 2020. Dasatinib–Blinatumomab for Ph-Positive Acute Lymphoblastic Leukemia in Adults. *New England Journal of Medicine,* 383**,** 1613-1623.

FRIDLYAND, J., SNIJDERS, A. M., PINKEL, D., ALBERTSON, D. G. & JAIN, A. N. 2004. Hidden Markov models approach to the analysis of array CGH data. *Journal of Multivariate Analysis,* 90**,** 132-153.

FRÖHLING, S., SCHOLL, C., LEVINE, R. L., LORIAUX, M., BOGGON, T. J., BERNARD, O. A., BERGER, R., DÖHNER, H., DÖHNER, K., EBERT, B. L., TECKIE, S., GOLUB, T. R., JIANG, J., SCHITTENHELM, M. M., LEE, B. H., GRIFFIN, J. D., STONE, R. M., HEINRICH, M. C., DEININGER, M. W., DRUKER, B. J. & GILLILAND, D. G. 2007. Identification of Driver and Passenger Mutations of FLT3 by High-Throughput DNA Sequence Analysis and Functional Assessment of Candidate Alleles. *Cancer Cell,* 12**,** 501-513.

GAN, L., YANG, Y., LI, Q., FENG, Y., LIU, T. & GUO, W. 2018. Epigenetic regulation of cancer progression by EZH2: from biological insights to therapeutic potential. *Biomarker Research,* 6**,** 10.

GAYNON, P. S., DESAI, A. A., BOSTROM, B. C., HUTCHINSON, R. J., LANGE, B. J., NACHMAN, J. B., REAMAN, G. H., SATHER, H. N., STEINHERZ, P. G., TRIGG, M. E., TUBERGEN, D. G. & UCKUN, F. M. 1997. Early response to therapy and outcome in childhood acute lymphoblastic leukemia. *Cancer,* 80**,** 1717-1726.

GELSI-BOYER, V., BRECQUEVILLE, M., DEVILLIER, R., MURATI, A., MOZZICONACCI, M.-J. & BIRNBAUM, D. 2012. Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. *Journal of hematology & oncology,* 5**,** 12-12.

GENOVESE, G., KÄHLER, A. K., HANDSAKER, R. E., LINDBERG, J., ROSE, S. A., BAKHOUM, S. F., CHAMBERT, K., MICK, E., NEALE, B. M., FROMER, M., PURCELL, S. M., SVANTESSON, O., LANDÉN, M., HÖGLUND, M., LEHMANN, S., GABRIEL, S. B., MORAN, J. L., LANDER, E. S., SULLIVAN, P. F., SKLAR, P., GRÖNBERG, H., HULTMAN, C. M. & MCCARROLL, S. A. 2014. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine,* 371**,** 2477-2487.

GEORGOPOULOS, K., BIGBY, M., WANG, J.-H., MOLNAR, A., WU, P., WINANDY, S. & SHARPE, A. 1994. The ikaros gene is required for the development of all lymphoid lineages. *Cell,* 79**,** 143-156.

GIL, V. S., BHAGAT, G., HOWELL, L., ZHANG, J., KIM, C. H., STENGEL, S., VEGA, F., ZELENT, A. & PETRIE, K. 2016. Deregulated expression of HDAC9 in B cells promotes development of lymphoproliferative disease and lymphoma in mice. *Disease Models & Mechanisms,* 9**,** 1483.

GIRARDI, T., VICENTE, C., COOLS, J. & DE KEERSMAECKER, K. 2017. The genetics and molecular biology of T-ALL. *Blood,* 129**,** 1113-1123.

GOEKBUGET, N., BECK, J., BRUEGGEMANN, M., BURMEISTER, T., BUSS, E. C., FRICKHOFEN, N., HUETTMANN, A., MORGNER, A., REICHLE, A., SCHMIDT-WOLF, I., SCHWARTZ, S., SERVE, H., SPRIEWALD, B. M., STARCK, M., STELLJES, M., VIARDOT, A., WENDELIN, K. & HOELZER, D. 2012. Moderate Intensive Chemotherapy Including CNS-Prophylaxis with Liposomal Cytarabine Is Feasible and effective in Older Patients with Ph-Negative Acute Lymphoblastic Leukemia (ALL): Results of a Prospective Trial From the German Multicenter Study Group for Adult ALL (GMALL). *Blood,* 120**,** 1493-1493.

GOKBUGET, N. 2017. Treatment of Older Patients with Acute Lymphoblastic Leukaemia. *Drugs & Aging,* 18**,** 18.

GOLDSTONE, A. H., RICHARDS, S. M., LAZARUS, H. M., TALLMAN, M. S., BUCK, G., FIELDING, A. K., BURNETT, A. K., CHOPRA, R., WIERNIK, P. H., FORONI, L., PAIETTA, E., LITZOW, M. R., MARKS, D. I., DURRANT, J., MCMILLAN, A., FRANKLIN, I. M., LUGER, S., CIOBANU, N. & ROWE, J. M. 2008. In adults with standard-risk acute lymphoblastic leukemia, the greatest benefit is achieved from a matched sibling allogeneic transplantation in first complete remission, and an autologous transplantation is less effective than conventional consolidation/maintenance chemotherapy in all patients: final results of the International ALL Trial (MRC UKALL XII/ECOG E2993). *Blood,* 111**,** 1827-1833.

GOLUB, T. R., BARKER, G. F., BOHLANDER, S. K., HIEBERT, S. W., WARD, D. C., BRAY-WARD, P., MORGAN, E., RAIMONDI, S. C., ROWLEY, J. D. & GILLILAND, D. G. 1995. Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences,* 92**,** 4917.

GREAVES, M. 2006. Infection, immune responses and the aetiology of childhood leukaemia. *Nature Reviews Cancer,* 6**,** 193-203.

GREAVES, M. F., MAIA, A. T., WIEMELS, J. L. & FORD, A. M. 2003. Leukemia in twins: lessons in natural history. *Blood,* 102**,** 2321-2333.

GREAVES, M. F., PEGRAM, S. M. & CHAN, L. C. 1985. Collaborative group study of the epidemiology of acute lymphoblastic leukaemia subtypes: Background and first report. *Leukemia Research,* 9**,** 715-733.

GREENMAN, C., STEPHENS, P., SMITH, R., DALGLIESH, G. L., HUNTER, C., BIGNELL, G., DAVIES, H., TEAGUE, J., BUTLER, A., STEVENS, C., EDKINS, S., O'MEARA, S., VASTRIK, I., SCHMIDT, E. E., AVIS, T., BARTHORPE, S., BHAMRA, G., BUCK, G., CHOUDHURY, B., CLEMENTS, J., COLE, J., DICKS, E., FORBES, S., GRAY, K., HALLIDAY, K., HARRISON, R., HILLS, K., HINTON, J., JENKINSON, A., JONES, D., MENZIES, A., MIRONENKO, T., PERRY, J., RAINE, K., RICHARDSON, D., SHEPHERD, R., SMALL, A., TOFTS, C., VARIAN, J., WEBB, T., WEST, S., WIDAA, S., YATES, A., CAHILL, D. P., LOUIS, D. N., GOLDSTRAW, P., NICHOLSON, A. G., BRASSEUR, F.,

LOOIJENGA, L., WEBER, B. L., CHIEW, Y.-E., DEFAZIO, A., GREAVES, M. F., GREEN, A. R., CAMPBELL, P., BIRNEY, E., EASTON, D. F., CHENEVIX-TRENCH, G., TAN, M.-H., KHOO, S. K., TEH, B. T., YUEN, S. T., LEUNG, S. Y., WOOSTER, R., FUTREAL, P. A. & STRATTON, M. R. 2007. Patterns of somatic mutation in human cancer genomes. *Nature,* 446**,** 153-158.

GREIF, P. A., HARTMANN, L., VOSBERG, S., STIEF, S. M., MATTES, R., HELLMANN, I., METZELER, K. H., HEROLD, T., BAMOPOULOS, S. A., KERBS, P., JURINOVIC, V., SCHUMACHER, D., PASTORE, F., BRÄUNDL, K., ZELLMEIER, E., KSIENZYK, B., KONSTANDIN, N. P., SCHNEIDER, S., GRAF, A., KREBS, S., BLUM, H., NEUMANN, M., BALDUS, C. D., BOHLANDER, S. K., WOLF, S., GÖRLICH, D., BERDEL, W. E., WÖRMANN, B. J., HIDDEMANN, W. & SPIEKERMANN, K. 2018. Evolution of Cytogenetically Normal Acute Myeloid Leukemia During Therapy and Relapse: An Exome Sequencing Study of 50 Patients. *Clinical Cancer Research,* 24**,** 1716.

GU, Z., CHURCHMAN, M., ROBERTS, K., LI, Y., LIU, Y., HARVEY, R. C., MCCASTLAIN, K., RESHMI, S. C., PAYNE-TURNER, D., IACOBUCCI, I., SHAO, Y., CHEN, I. M., VALENTINE, M., PEI, D., MUNGALL, K. L., MUNGALL, A. J., MA, Y., MOORE, R., MARRA, M., STONEROCK, E., GASTIER-FOSTER, J. M., DEVIDAS, M., DAI, Y., WOOD, B., BOROWITZ, M., LARSEN, E. E., MALONEY, K., MATTANO JR, L. A., ANGIOLILLO, A., SALZER, W. L., BURKE, M. J., GIANNI, F., SPINELLI, O., RADICH, J. P., MINDEN, M. D., MOORMAN, A. V., PATEL, B., FIELDING, A. K., ROWE, J. M., LUGER, S. M., BHATIA, R., ALDOSS, I., FORMAN, S. J., KOHLSCHMIDT, J., MRÓZEK, K., MARCUCCI, G., BLOOMFIELD, C. D., STOCK, W., KORNBLAU, S., KANTARJIAN, H. M., KONOPLEVA, M., PAIETTA, E., WILLMAN, C. L., L. LOH, M., P. HUNGER, S. & MULLIGHAN, C. G. 2016a. Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. *Nature Communications,* 7**,** 13331.

GU, Z., CHURCHMAN, M. L., ROBERTS, K. G., MOORE, I., ZHOU, X., NAKITANDWE, J., HAGIWARA, K., PELLETIER, S., GINGRAS, S., BERNS, H., PAYNE-TURNER, D., HILL, A., IACOBUCCI, I., SHI, L., POUNDS, S., CHENG, C., PEI, D., QU, C., NEWMAN, S., DEVIDAS, M., DAI, Y., RESHMI, S. C., GASTIER-FOSTER, J., RAETZ, E. A., BOROWITZ, M. J., WOOD, B. L., CARROLL, W. L., ZWEIDLER-MCKAY, P. A., RABIN, K. R., MATTANO, L. A., MALONEY, K. W., RAMBALDI, A., SPINELLI, O., RADICH, J. P., MINDEN, M. D., ROWE, J. M., LUGER, S., LITZOW, M. R., TALLMAN, M. S., RACEVSKIS, J., ZHANG, Y., BHATIA, R., KOHLSCHMIDT, J., MRÓZEK, K., BLOOMFIELD, C. D., STOCK, W., KORNBLAU, S., KANTARJIAN, H. M., KONOPLEVA, M., EVANS, W. E., JEHA, S., PUI, C.-H., YANG, J., PAIETTA, E., DOWNING, J. R., RELLING, M. V., ZHANG, J., LOH, M. L., HUNGER, S. P. & MULLIGHAN, C. G. 2019. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nature genetics,* 51**,** 296-307.

GU, Z., EILS, R. & SCHLESNER, M. 2016b. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics,* 32**,** 2847-2849.

GUPTA, S., DEVIDAS, M., LOH, M. L., RAETZ, E. A., CHEN, S., WANG, C., BROWN, P., CARROLL, A. J., HEEREMA, N. A., GASTIER-FOSTER, J. M., DUNSMORE, K. P., LARSEN, E. C., MALONEY, K. W., MATTANO, L. A., JR., WINTER, S. S., WINICK, N. J., CARROLL, W. L., HUNGER, S. P., BOROWITZ, M. J. & WOOD, B. L. 2018. Flow-cytometric vs. morphologic assessment of

remission in childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group (COG). *Leukemia,* 32**,** 1370-1379.

GURU MURTHY, G. S., PONDAIAH, S. K., ABEDIN, S. & ATALLAH, E. 2019. Incidence and survival of T-cell acute lymphoblastic leukemia in the United States. *Leukemia & Lymphoma,* 60**,** 1171-1178.

GURU MURTHY, G. S., VENKITACHALAM, R. & MEHTA, P. 2015. Trends in survival outcomes of B-lineage acute lymphoblastic leukemia in elderly patients: analysis of Surveillance, Epidemiology, and End Results database. *Leukemia & Lymphoma,* 56**,** 2296-2300.

GÖKBUGET, N. 2013. How I treat older patients with ALL. *Blood,* 122**,** 1366-1375.

GÖKBUGET, N., DOMBRET, H., BONIFACIO, M., REICHLE, A., GRAUX, C., FAUL, C., DIEDRICH, H., TOPP, M. S., BRÜGGEMANN, M., HORST, H.-A., HAVELANGE, V., STIEGLMAIER, J., WESSELS, H., HADDAD, V., BENJAMIN, J. E., ZUGMAIER, G., NAGORSEN, D. & BARGOU, R. C. 2018. Blinatumomab for minimal residual disease in adults with B-cell precursor acute lymphoblastic leukemia. *Blood,* 131**,** 1522-1531.

HAMADEH, L., ENSHAEI, A., SCHWAB, C., ALONSO, C. N., ATTARBASCHI, A., BARBANY, G., DEN BOER, M. L., BOER, J. M., BRAUN, M., DALLA POZZA, L., ELITZUR, S., EMERENCIANO, M., FECHINA, L., FELICE, M. S., FRONKOVA, E., HALTRICH, I., HEYMAN, M. M., HORIBE, K., IMAMURA, T., JEISON, M., KOVÁCS, G., KUIPER, R. P., MLYNARSKI, W., NEBRAL, K., IVANOV ÖFVERHOLM, I., PASTORCZAK, A., PIETERS, R., PIKO, H., POMBO-DE-OLIVEIRA, M. S., RUBIO, P., STREHL, S., STARY, J., SUTTON, R., TRKA, J., TSAUR, G., VENN, N., VORA, A., YANO, M., HARRISON, C. J., MOORMAN, A. V. & INTERNATIONAL, B. F. M. S. G. 2019. Validation of the United Kingdom copy-number alteration classifier in 3239 children with B-cell precursor ALL. *Blood advances,* 3**,** 148-157.

HANAHAN, D. & WEINBERG, ROBERT A. 2011. Hallmarks of Cancer: The Next Generation. *Cell,* 144**,** 646-674.

HANN, I., VORA, A., HARRISON, G., HARRISON, C., EDEN, O., HILL, F., GIBSON, B., RICHARDS, S. & THE, U. K. M. R. C. S. W. P. O. C. L. 2001. Determinants of outcome after intensified therapy of childhood lymphoblastic leukaemia: results from Medical Research Council United Kingdom acute lymphoblastic leukaemia XI protocol. *British Journal of Haematology,* 113**,** 103-114.

HARRISON, C. J., HAAS, O., HARBOTT, J., BIONDI, A., STANULLA, M., TRKA, J. & IZRAELI, S. 2010. Detection of prognostically relevant genetic abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: recommendations from the Biology and Diagnosis Committee of the International Berlin-Frankfürt-Münster study group. *British Journal of Haematology,* 151**,** 132-142.

HARRISON, C. J., MARTINEAU, M. & SECKER-WALKER, L. M. 2001. The Leukaemia Research Fund/United Kingdom Cancer Cytogenetics Group Karyotype Database in acute lymphoblastic leukaemia: a valuable resource for patient management. *British Journal of Haematology,* 113**,** 3-10.

HARRISON, C. J., MOORMAN, A. V., BROADFIELD, Z. J., CHEUNG, K. L., HARRIS, R. L., REZA JALALI, G., ROBINSON, H. M., BARBER, K. E., RICHARDS, S. M., MITCHELL, C. D., EDEN, T. O. B., HANN, I. M., HILL, F. G. H., KINSEY, S. E., GIBSON, B. E. S., LILLEYMAN, J., VORA, A., GOLDSTONE, A. H., FRANKLIN, I. M., DURRANT, J., MARTINEAU, M., CHILDHOOD & PARTIES,

A. L. W. 2004. Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia. *British Journal of Haematology,* 125**,** 552-559.

HARVEY, R. C., MULLIGHAN, C. G., CHEN, I. M., WHARTON, W., MIKHAIL, F. M., CARROLL, A. J., KANG, H., LIU, W., DOBBIN, K. K., SMITH, M. A., CARROLL, W. L., DEVIDAS, M., BOWMAN, W. P., CAMITTA, B. M., REAMAN, G. H., HUNGER, S. P., DOWNING, J. R. & WILLMAN, C. L. 2010. Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia. *Blood,* 115**,** 5312-5321.

HAUER, J., MULLIGHAN, C., MORILLON, E., WANG, G., BRUNEAU, J., BROUSSE, N., LELORC'H, M., ROMANA, S., BOUDIL, A., TIEDAU, D., KRACKER, S., BUSHMANN, F. D., BORKHARDT, A., FISCHER, A., HACEIN-BEY-ABINA, S. & CAVAZZANA-CALVO, M. 2011. Loss of p19Arf in a Rag1(-/-) B-cell precursor population initiates acute B-lymphoblastic leukemia. *Blood,* 118**,** 544-553.

HEEREMA, N. A., RAIMONDI, S. C., ANDERSON, J. R., BIEGEL, J., CAMITTA, B. M., COOLEY, L. D., GAYNON, P. S., HIRSCH, B., MAGENIS, R. E., MCGAVRAN, L., PATIL, S., PETTENATI, M. J., PULLEN, J., RAO, K., ROULSTON, D., SCHNEIDER, N. R., SHUSTER, J. J., SANGER, W., SUTCLIFFE, M. J., VAN TUINEN, P., WATSON, M. S. & CARROLL, A. J. 2007. Specific extra chromosomes occur in a modal number dependent pattern in pediatric acute lymphoblastic leukemia. *Genes, Chromosomes and Cancer,* 46**,** 684-693.

HELLEMANS, J., PREOBRAZHENSKA, O., WILLAERT, A., DEBEER, P., VERDONK, P. C. M., COSTA, T., JANSSENS, K., MENTEN, B., ROY, N. V., VERMEULEN, S. J. T., SAVARIRAYAN, R., HUL, W. V., VANHOENACKER, F., HUYLEBROECK, D., PAEPE, A. D., NAEYAERT, J.-M., VANDESOMPELE, J., SPELEMAN, F., VERSCHUEREN, K., COUCKE, P. J. & MORTIER, G. R. 2004. Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. *Nature Genetics,* 36**,** 1213-1218.

HERNANDEZ, P. A., GORLIN, R. J., LUKENS, J. N., TANIUCHI, S., BOHINJEC, J., FRANCOIS, F., KLOTMAN, M. E. & DIAZ, G. A. 2003. Mutations in the chemokine receptor gene CXCR4 are associated with WHIM syndrome, a combined immunodeficiency disease. *Nature Genetics,* 34**,** 70-74.

HEROLD, T., BALDUS, C. D. & GÖKBUGET, N. 2014. Ph-like Acute Lymphoblastic Leukemia in Older Adults. *New England Journal of Medicine,* 371**,** 2235-2235.

HEROLD, T., SCHNEIDER, S., METZELER, K. H., NEUMANN, M., HARTMANN, L., ROBERTS, K. G., KONSTANDIN, N. P., GREIF, P. A., BRÄUNDL, K., KSIENZYK, B., HUK, N., SCHNEIDER, I., ZELLMEIER, E., JURINOVIC, V., MANSMANN, U., HIDDEMANN, W., MULLIGHAN, C. G., BOHLANDER, S. K., SPIEKERMANN, K., HOELZER, D., BRÜGGEMANN, M., BALDUS, C. D., DREYLING, M. & GÖKBUGET, N. 2017. Adults with Philadelphia chromosome-like acute lymphoblastic leukemia frequently have IGH-CRLF2 and JAK2 mutations, persistence of minimal residual disease and poor prognosis. *Haematologica,* 102**,** 130-138.

HILDEN, J. M., DINNDORF, P. A., MEERBAUM, S. O., SATHER, H., VILLALUNA, D., HEEREMA, N. A., MCGLENNEN, R., SMITH, F. O., WOODS, W. G., SALZER, W. L., JOHNSTONE, H. S., DREYER, Z. & REAMAN, G. H. 2006. Analysis of prognostic factors of acute lymphoblastic leukemia in infants: report on CCG 1953 from the Children's Oncology Group. *Blood,* 108**,** 441-451.

HIRABAYASHI, S., OHKI, K., NAKABAYASHI, K., ICHIKAWA, H., MOMOZAWA, Y., OKAMURA, K., YAGUCHI, A., TERADA, K., SAITO, Y., YOSHIMI, A., OGATA-KAWATA, H., SAKAMOTO, H., KATO, M., FUJIMURA, J., HINO, M., KINOSHITA, A., KAKUDA, H., KUROSAWA, H., KATO, K., KAJIWARA, R., MORIWAKI, K., MORIMOTO, T., NAKAMURA, K., NOGUCHI, Y., OSUMI, T., SAKASHITA, K., TAKITA, J., YUZA, Y., MATSUDA, K., YOSHIDA, T., MATSUMOTO, K., HATA, K., KUBO, M., MATSUBARA, Y., FUKUSHIMA, T., KOH, K., MANABE, A., OHARA, A. & KIYOKAWA, N. 2017. *ZNF384*-related fusion genes define a subgroup of childhood B-cell precursor acute lymphoblastic leukemia with a characteristic immunotype. *Haematologica,* 102**,** 118-129.

HOLMFELDT, L., WEI, L., DIAZ-FLORES, E., WALSH, M., ZHANG, J., DING, L., PAYNE-TURNER, D., CHURCHMAN, M., ANDERSSON, A., CHEN, S.-C., MCCASTLAIN, K., BECKSFORT, J., MA, J., WU, G., PATEL, S. N., HEATLEY, S. L., PHILLIPS, L. A., SONG, G., EASTON, J., PARKER, M., CHEN, X., RUSCH, M., BOGGS, K., VADODARIA, B., HEDLUND, E., DRENBERG, C., BAKER, S., PEI, D., CHENG, C., HUETHER, R., LU, C., FULTON, R. S., FULTON, L. L., TABIB, Y., DOOLING, D. J., OCHOA, K., MINDEN, M., LEWIS, I. D., TO, L. B., MARLTON, P., ROBERTS, A. W., RACA, G., STOCK, W., NEALE, G., DREXLER, H. G., DICKINS, R. A., ELLISON, D. W., SHURTLEFF, S. A., PUI, C.-H., RIBEIRO, R. C., DEVIDAS, M., CARROLL, A. J., HEEREMA, N. A., WOOD, B., BOROWITZ, M. J., GASTIER-FOSTER, J. M., RAIMONDI, S. C., MARDIS, E. R., WILSON, R. K., DOWNING, J. R., HUNGER, S. P., LOH, M. L. & MULLIGHAN, C. G. 2013. The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nature genetics,* 45**,** 242-252.

HOVORKOVA, L., ZALIOVA, M., VENN, N. C., BLECKMANN, K., TRKOVA, M., POTUCKOVA, E., VASKOVA, M., LINHARTOVA, J., MACHOVA POLAKOVA, K., FRONKOVA, E., MUSKOVIC, W., GILES, J. E., SHAW, P. J., CARIO, G., SUTTON, R., STARY, J., TRKA, J. & ZUNA, J. 2017. Monitoring of childhood ALL using BCR-ABL1 genomic breakpoints identifies a subgroup with CML-like biology. *Blood,* 129**,** 2771-2781.

HU, Y., LIU, Y., PELLETIER, S., BUCHDUNGER, E., WARMUTH, M., FABBRO, D., HALLEK, M., VAN ETTEN, R. A. & LI, S. 2004. Requirement of Src kinases Lyn, Hck and Fgr for BCR-ABL1-induced B-lymphoblastic leukemia but not chronic myeloid leukemia. *Nature Genetics,* 36**,** 453.

HUBER, D., VOITH VON VOITHENBERG, L. & KAIGALA, G. V. 2018. Fluorescence in situ hybridization (FISH): History, limitations and what to expect from micro-scale FISH? *Micro and Nano Engineering,* 1**,** 15-24.

IACOBUCCI, I., LI, Y., ROBERTS, K. G., DOBSON, S. M., KIM, J. C., PAYNE-TURNER, D., HARVEY, R. C., VALENTINE, M., MCCASTLAIN, K., EASTON, J., YERGEAU, D., JANKE, L. J., SHAO, Y., CHEN, I. M. L., RUSCH, M., ZANDI, S., KORNBLAU, S. M., KONOPLEVA, M., JABBOUR, E., PAIETTA, E. M., ROWE, J. M., PUI, C.-H., GASTIER-FOSTER, J., GU, Z., RESHMI, S., LOH, M. L., RACEVSKIS, J., TALLMAN, M. S., WIERNIK, P. H., LITZOW, M. R., WILLMAN, C. L., MCPHERSON, J. D., DOWNING, J. R., ZHANG, J., DICK, J. E., HUNGER, S. P. & MULLIGHAN, C. G. 2016. Truncating erythropoietin receptor rearrangements in acute lymphoblastic leukemia. *Cancer cell,* 29**,** 186-200.

IACOBUCCI, I., LONETTI, A., PAPAYANNIDIS, C. & MARTINELLI, G. 2013. Use of Single Nucleotide Polymorphism Array Technology to Improve the Identification

of Chromosomal Lesions in Leukemia. *Current Cancer Drug Targets,* 13**,** 791-810.

IACOBUCCI, I. & MULLIGHAN, C. G. 2017. Genetic Basis of Acute Lymphoblastic Leukemia. *Journal of Clinical Oncology,* 35**,** 975-983.

IKUSHIMA, H. & MIYAZONO, K. 2010. TGFβ signalling: a complex web in cancer progression. *Nature Reviews Cancer,* 10**,** 415-424.

INABA, H., GREAVES, M. & MULLIGHAN, C. G. 2013. Acute lymphoblastic leukaemia. *Lancet (London, England),* 381**,** 1943-1955.

IRVING, J., MATHESON, E., MINTO, L., BLAIR, H., CASE, M., HALSEY, C., SWIDENBANK, I., PONTHAN, F., KIRSCHNER-SCHWABE, R., GROENEVELD-KRENTZ, S., HOF, J., ALLAN, J., HARRISON, C., VORMOOR, J., VON STACKELBERG, A. & ECKERT, C. 2014. Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. *Blood,* 124**,** 3420-3430.

ISSA, G. C., KANTARJIAN, H. M., YIN, C. C., QIAO, W., RAVANDI, F., THOMAS, D., SHORT, N. J., SASAKI, K., GARCIA-MANERO, G., KADIA, T. M., CORTES, J. E., DAVER, N., BORTHAKUR, G., JAIN, N., KONOPLEVA, M., KHOURI, I., KEBRIAEI, P., CHAMPLIN, R. E., PIERCE, S., O'BRIEN, S. M. & JABBOUR, E. 2017. Prognostic impact of pretreatment cytogenetics in adult Philadelphia chromosome–negative acute lymphoblastic leukemia in the era of minimal residual disease. *Cancer,* 123**,** 459-467.

JABBOUR, E. & KANTARJIAN, H. 2020. Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. *American Journal of Hematology,* 95**,** 691-709.

JABBOUR, E., SHORT, N. J., RAVANDI, F., HUANG, X., DAVER, N., DINARDO, C. D., KONOPLEVA, M., PEMMARAJU, N., WIERDA, W., GARCIA-MANERO, G., SASAKI, K., CORTES, J., GARRIS, R., KHOURY, J. D., JORGENSEN, J., JAIN, N., ALVAREZ, J., O'BRIEN, S. & KANTARJIAN, H. 2018. Combination of hyper-CVAD with ponatinib as first-line therapy for patients with Philadelphia chromosome-positive acute lymphoblastic leukaemia: long-term follow-up of a single-centre, phase 2 study. *The Lancet Haematology,* 5**,** e618-e627.

JAIN, N., JABBOUR, E. J., MCKAY, P. Z., RAVANDI, F., TAKAHASHI, K., KADIA, T., WIERDA, W. G., RYTTING, M. E., NUNEZ, C., PATEL, K., LU, X., TANG, G., KONOPLEV, S. N., WANG, S. A., HAN, L., THAKRAL, B., ROBERTS, K. G., DESHMUKH, A., GARRIS, R., JORGENSEN, J. L., CAVAZOS, A., VERSTOVSEK, S., GARCIA-MANERO, G., O'BRIEN, S. M., CORTES, J. E., MULLIGHAN, C. G., KANTARJIAN, H. M. & KONOPLEVA, M. 2017a. Ruxolitinib or Dasatinib in Combination with Chemotherapy for Patients with Relapsed/Refractory Philadelphia (Ph)-like Acute Lymphoblastic Leukemia: A Phase I-II Trial. *Blood,* 130**,** 1322-1322.

JAIN, N., ROBERTS, K. G., JABBOUR, E., PATEL, K., ETEROVIC, A. K., CHEN, K., ZWEIDLER-MCKAY, P., LU, X., FAWCETT, G., WANG, S. A., KONOPLEV, S., HARVEY, R. C., CHEN, I. M., PAYNE-TURNER, D., VALENTINE, M., THOMAS, D., GARCIA-MANERO, G., RAVANDI, F., CORTES, J., KORNBLAU, S., O'BRIEN, S., PIERCE, S., JORGENSEN, J., SHAW, K. R. M., WILLMAN, C. L., MULLIGHAN, C. G., KANTARJIAN, H. & KONOPLEVA, M. 2017b. Ph-like acute lymphoblastic leukemia: a high-risk subtype in adults. *Blood,* 129**,** 572-581.

JAISWAL, S., FONTANILLAS, P., FLANNICK, J., MANNING, A., GRAUMAN, P. V., MAR, B. G., LINDSLEY, R. C., MERMEL, C. H., BURTT, N., CHAVEZ, A., HIGGINS, J. M., MOLTCHANOV, V., KUO, F. C., KLUK, M. J., HENDERSON, B., KINNUNEN, L., KOISTINEN, H. A., LADENVALL, C., GETZ, G., CORREA, A., BANAHAN, B. F., GABRIEL, S., KATHIRESAN, S., STRINGHAM, H. M., MCCARTHY, M. I., BOEHNKE, M., TUOMILEHTO, J., HAIMAN, C., GROOP, L., ATZMON, G., WILSON, J. G., NEUBERG, D., ALTSHULER, D. & EBERT, B. L. 2014. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine,* 371**,** 2488-2498.

JAN, M., SNYDER, T. M., CORCES-ZIMMERMAN, M. R., VYAS, P., WEISSMAN, I. L., QUAKE, S. R. & MAJETI, R. 2012. Clonal Evolution of Preleukemic Hematopoietic Stem Cells Precedes Human Acute Myeloid Leukemia. *Science Translational Medicine,* 4**,** 149ra118.

JEFFRIES, S. J., JONES, L., HARRISON, C. J. & RUSSELL, L. J. 2014. IGH@ translocations co-exist with other primary rearrangements in B-cell precursor acute lymphoblastic leukemia. *Haematologica,* 99**,** 1334-1342.

JERCHEL, I. S., HOOGKAMER, A. Q., ARIËS, I. M., STEEGHS, E. M. P., BOER, J. M., BESSELINK, N. J. M., BOEREE, A., VAN DE VEN, C., DE GROOT-KRUSEMAN, H. A., DE HAAS, V., HORSTMANN, M. A., ESCHERICH, G., ZWAAN, C. M., CUPPEN, E., KOUDIJS, M. J., PIETERS, R. & DEN BOER, M. L. 2018. RAS pathway mutations as a predictive biomarker for treatment adaptation in pediatric B-cell precursor acute lymphoblastic leukemia. *Leukemia,* 32**,** 931-940.

JOLLIFFE, I. T. & CADIMA, J. 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 374**,** 20150202.

JORDI, R., MIREIA, M., LURDES, Z., PAU, M., INÉS, G. S., MARTA, P., JOSEP, S., RAMON, G., JOSEP, N., MAR, T., JOAQUIN, M. L., JESÚS-MARÍA, H. R., JOSÉ, G. C., PERE, B., LOURDES, E., EULÀLIA, G., FRANCESC, S., FUENSANTA, M., EVARIST, F. & JOSEP-MARIA, R. 2015. Prognostic significance of copy number alterations in adolescent and adult patients with precursor B acute lymphoblastic leukemia enrolled in PETHEMA protocols. *Cancer,* 121**,** 3809-3817.

KANTARJIAN, H., RAVANDI, F., SHORT, N. J., HUANG, X., JAIN, N., SASAKI, K., DAVER, N., PEMMARAJU, N., KHOURY, J. D., JORGENSEN, J., ALVARADO, Y., KONOPLEVA, M., GARCIA-MANERO, G., KADIA, T., YILMAZ, M., BORTAKHUR, G., BURGER, J., KORNBLAU, S., WIERDA, W., DINARDO, C., FERRAJOLI, A., JACOB, J., GARRIS, R., O'BRIEN, S. & JABBOUR, E. 2018. Inotuzumab ozogamicin in combination with low-intensity chemotherapy for older patients with Philadelphia chromosome-negative acute lymphoblastic leukaemia: a single-arm, phase 2 study. *The Lancet Oncology,* 19**,** 240-248.

KARCZEWSKI, K. J., FRANCIOLI, L. C., TIAO, G., CUMMINGS, B. B., ALFÖLDI, J., WANG, Q., COLLINS, R. L., LARICCHIA, K. M., GANNA, A., BIRNBAUM, D. P., GAUTHIER, L. D., BRAND, H., SOLOMONSON, M., WATTS, N. A., RHODES, D., SINGER-BERK, M., ENGLAND, E. M., SEABY, E. G., KOSMICKI, J. A., WALTERS, R. K., TASHMAN, K., FARJOUN, Y., BANKS, E., POTERBA, T., WANG, A., SEED, C., WHIFFIN, N., CHONG, J. X., SAMOCHA, K. E., PIERCE-HOFFMAN, E., ZAPPALA, Z., O'DONNELL-LURIA, A. H., MINIKEL, E. V., WEISBURD, B., LEK, M., WARE, J. S., VITTAL, C., ARMEAN,

I. M., BERGELSON, L., CIBULSKIS, K., CONNOLLY, K. M., COVARRUBIAS, M., DONNELLY, S., FERRIERA, S., GABRIEL, S., GENTRY, J., GUPTA, N., JEANDET, T., KAPLAN, D., LLANWARNE, C., MUNSHI, R., NOVOD, S., PETRILLO, N., ROAZEN, D., RUANO-RUBIO, V., SALTZMAN, A., SCHLEICHER, M., SOTO, J., TIBBETTS, K., TOLONEN, C., WADE, G., TALKOWSKI, M. E., AGUILAR SALINAS, C. A., AHMAD, T., ALBERT, C. M., ARDISSINO, D., ATZMON, G., BARNARD, J., BEAUGERIE, L., BENJAMIN, E. J., BOEHNKE, M., BONNYCASTLE, L. L., BOTTINGER, E. P., BOWDEN, D. W., BOWN, M. J., CHAMBERS, J. C., CHAN, J. C., CHASMAN, D., CHO, J., CHUNG, M. K., COHEN, B., CORREA, A., DABELEA, D., DALY, M. J., DARBAR, D., DUGGIRALA, R., DUPUIS, J., ELLINOR, P. T., ELOSUA, R., ERDMANN, J., ESKO, T., FÄRKKILÄ, M., FLOREZ, J., FRANKE, A., GETZ, G., GLASER, B., GLATT, S. J., GOLDSTEIN, D., GONZALEZ, C., GROOP, L., et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature,* 581**,** 434-443.

KARCZEWSKI, K. J., WEISBURD, B., THOMAS, B., SOLOMONSON, M., RUDERFER, D. M., KAVANAGH, D., HAMAMSY, T., LEK, M., SAMOCHA, K. E., CUMMINGS, B. B., BIRNBAUM, D., THE EXOME AGGREGATION CONSORTIUM, DALY, M. J. & MACARTHUR, D. G. 2017. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic acids research,* 45**,** D840-D845.

KOHLMANN, A., GROSSMANN, V., KLEIN, H.-U., SCHINDELA, S., WEISS, T., KAZAK, B., DICKER, F., SCHNITTGER, S., DUGAS, M., KERN, W., HAFERLACH, C. & HAFERLACH, T. 2010. Next-Generation Sequencing Technology Reveals a Characteristic Pattern of Molecular Mutations in 72.8% of Chronic Myelomonocytic Leukemia by Detecting Frequent Alterations in TET2, CBL, RAS, and RUNX1. *Journal of Clinical Oncology,* 28**,** 3858-3865.

KÜPPERS, R. 2005. Mechanisms of B-cell lymphoma pathogenesis. *Nature Reviews Cancer,* 5**,** 251-262.

LAFRAMBOISE, T. 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Research,* 37**,** 4181-4193.

LAMPERT, F. 1967. Cellulärer DNS-Gehalt und Chromosomenzahl bei der akuten Leukämie im Kindesalter und ihre Bedeutung für Chemotherapie und Prognose. *Klinische Wochenschrift,* 45**,** 763-768.

LANDRUM, M. J., LEE, J. M., RILEY, G. R., JANG, W., RUBINSTEIN, W. S., CHURCH, D. M. & MAGLOTT, D. R. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic acids research,* 42**,** D980-D985.

LARSON, R. A. 2005. Acute Lymphoblastic Leukemia: Older Patients and Newer Drugs. *Hematology,* 2005**,** 131-136.

LEE, M. G., VILLA, R., TROJER, P., NORMAN, J., YAN, K.-P., REINBERG, D., DI CROCE, L. & SHIEKHATTAR, R. 2007. Demethylation of H3K27 Regulates Polycomb Recruitment and H2A Ubiquitination. *Science,* 318**,** 447.

LER, L. D., GHOSH, S., CHAI, X., THIKE, A. A., HENG, H. L., SIEW, E. Y., DEY, S., KOH, L. K., LIM, J. Q., LIM, W. K., MYINT, S. S., LOH, J. L., ONG, P., SAM, X. X., HUANG, D., LIM, T., TAN, P. H., NAGARAJAN, S., CHENG, C. W. S., HO, H., NG, L. G., YUEN, J., LIN, P.-H., CHUANG, C.-K., CHANG, Y.-H., WENG,

W.-H., ROZEN, S. G., TAN, P., CREASY, C. L., PANG, S.-T., MCCABE, M. T., POON, S. L. & TEH, B. T. 2017. Loss of tumor suppressor KDM6A amplifies PRC2-regulated transcriptional repression in bladder cancer and can be targeted through inhibition of EZH2. *Science Translational Medicine,* 9**,** eaai8312.

LI, J.-F., DAI, Y.-T., LILLJEBJÖRN, H., SHEN, S.-H., CUI, B.-W., BAI, L., LIU, Y.-F., QIAN, M.-X., KUBOTA, Y., KIYOI, H., MATSUMURA, I., MIYAZAKI, Y., OLSSON, L., TAN, A. M., ARIFFIN, H., CHEN, J., TAKITA, J., YASUDA, T., MANO, H., JOHANSSON, B., YANG, J. J., YEOH, A. E.-J., HAYAKAWA, F., CHEN, Z., PUI, C.-H., FIORETOS, T., CHEN, S.-J. & HUANG, J.-Y. 2018. Transcriptional landscape of B cell precursor acute lymphoblastic leukemia based on an international study of 1,223 cases. *Proceedings of the National Academy of Sciences,* 115**,** E11711.

LILLJEBJÖRN, H., HENNINGSSON, R., HYRENIUS-WITTSTEN, A., OLSSON, L., ORSMARK-PIETRAS, C., VON PALFFY, S., ASKMYR, M., RISSLER, M., SCHRAPPE, M., CARIO, G., CASTOR, A., PRONK, C. J. H., BEHRENDTZ, M., MITELMAN, F., JOHANSSON, B., PAULSSON, K., ANDERSSON, A. K., FONTES, M. & FIORETOS, T. 2016. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in paediatric B-cell precursor acute lymphoblastic leukaemia. *Nature communications,* 7**,** 11790-11790.

LILLJEBJÖRN, H., SONESON, C., ANDERSSON, A., HELDRUP, J., BEHRENDTZ, M., KAWAMATA, N., OGAWA, S., KOEFFLER, H. P., MITELMAN, F., JOHANSSON, B., FONTES, M. & FIORETOS, T. 2010. The correlation pattern of acquired copy number changes in 164 ETV6/RUNX1-positive childhood acute lymphoblastic leukemias. *Human Molecular Genetics,* 19**,** 3150-3158.

LIU, Y.-F., WANG, B.-Y., ZHANG, W.-N., HUANG, J.-Y., LI, B.-S., ZHANG, M., JIANG, L., LI, J.-F., WANG, M.-J., DAI, Y.-J., ZHANG, Z.-G., WANG, Q., KONG, J., CHEN, B., ZHU, Y.-M., WENG, X.-Q., SHEN, Z.-X., LI, J.-M., WANG, J., YAN, X.-J., LI, Y., LIANG, Y.-M., LIU, L., CHEN, X.-Q., ZHANG, W.-G., YAN, J.-S., HU, J.-D., SHEN, S.-H., CHEN, J., GU, L.-J., PEI, D., LI, Y., WU, G., ZHOU, X., REN, R.-B., CHENG, C., YANG, J. J., WANG, K.-K., WANG, S.-Y., ZHANG, J., MI, J.-Q., PUI, C.-H., TANG, J.-Y., CHEN, Z. & CHEN, S.-J. 2016. Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia. *EBioMedicine,* 8**,** 173-183.

MA, S. K., CHAN, G. C. F., WAN, T. S. K., LAM, C. K., HA, S. Y., LAU, Y. L. & CHAN, L. C. 1998. Near-haploid common acute lymphoblastic leukaemia of childhood with a second hyperdiploid line: a DNA ploidy and fluorescence in-situ hybridization study. *British Journal of Haematology,* 103**,** 750-755.

MA, X., EDMONSON, M., YERGEAU, D., MUZNY, D. M., HAMPTON, O. A., RUSCH, M., SONG, G., EASTON, J., HARVEY, R. C., WHEELER, D. A., MA, J., DODDAPANENI, H., VADODARIA, B., WU, G., NAGAHAWATTE, P., CARROLL, W. L., CHEN, I. M., GASTIER-FOSTER, J. M., RELLING, M. V., SMITH, M. A., DEVIDAS, M., AUVIL, J. M. G., DOWNING, J. R., LOH, M. L., WILLMAN, C. L., GERHARD, D. S., MULLIGHAN, C. G., HUNGER, S. P. & ZHANG, J. 2015. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nature Communications,* 6**,** 6604.

MA, X., LIU, Y., LIU, Y., ALEXANDROV, L. B., EDMONSON, M. N., GAWAD, C., ZHOU, X., LI, Y., RUSCH, M. C., EASTON, J., HUETHER, R., GONZALEZ-PENA, V., WILKINSON, M. R., HERMIDA, L. C., DAVIS, S., SIOSON, E.,

POUNDS, S., CAO, X., RIES, R. E., WANG, Z., CHEN, X., DONG, L., DISKIN, S. J., SMITH, M. A., GUIDRY AUVIL, J. M., MELTZER, P. S., LAU, C. C., PERLMAN, E. J., MARIS, J. M., MESHINCHI, S., HUNGER, S. P., GERHARD, D. S. & ZHANG, J. 2018. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature,* 555**,** 371-376.

MACDONALD, J. R., ZIMAN, R., YUEN, R. K. C., FEUK, L. & SCHERER, S. W. 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic acids research,* 42**,** D986-D992.

MAKISHIMA, H. & MACIEJEWSKI, J. P. 2011. Pathogenesis and consequences of uniparental disomy in cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research,* 17**,** 3913-3923.

MALKIN, D. 2011. Li-Fraumeni Syndrome. *Genes & Cancer,* 2**,** 475-484.

MAR, B. G., BULLINGER, L., BASU, E., SCHLIS, K., SILVERMAN, L. B., DÖHNER, K. & ARMSTRONG, S. A. 2012. Sequencing histone-modifying enzymes identifies UTX mutations in acute lymphoblastic leukemia. *Leukemia,* 26**,** 1881-1883.

MAR, B. G., BULLINGER, L. B., MCLEAN, K. M., GRAUMAN, P. V., HARRIS, M. H., STEVENSON, K., NEUBERG, D. S., SINHA, A. U., SALLAN, S. E., SILVERMAN, L. B., KUNG, A. L., LO NIGRO, L., EBERT, B. L. & ARMSTRONG, S. A. 2014. Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. *Nature Communications,* 5**,** 3469.

MARKE, R., VAN LEEUWEN, F. N. & SCHEIJEN, B. 2018. The many faces of IKZF1 in B-cell precursor acute lymphoblastic leukemia. *Haematologica,* 103**,** 565-574.

MARKS, D. I., KIRKWOOD, A. A., ROWNTREE, C. J., AGUIAR, M., BAILEY, K. E., BEATON, B., CLIFTON-HADLEY, L., LAWRIE, E., LEE, S., MCMILLAN, A. K., MOORMAN, A. V., MITCHELL, R. J., PATRICK, P., MENNE, T. F., SMITH, P., WRENCH, B., ALAPI, K. Z. & FIELDING, A. K. 2019. First Analysis of the UKALL14 Phase 3 Randomised Trial to Determine If the Addition of Rituximab to Standard Induction Chemotherapy Improves EFS in Adults with Precursor B-ALL (CRUK/09/006). *Blood,* 134**,** 739-739.

MARKS, D. I., PAIETTA, E. M., MOORMAN, A. V., RICHARDS, S. M., BUCK, G., DEWALD, G., FERRANDO, A., FIELDING, A. K., GOLDSTONE, A. H., KETTERLING, R. P., LITZOW, M. R., LUGER, S. M., MCMILLAN, A. K., MANSOUR, M. R., ROWE, J. M., TALLMAN, M. S. & LAZARUS, H. M. 2009. T-cell acute lymphoblastic leukemia in adults: clinical features, immunophenotype, cytogenetics, and outcome from the large randomized prospective trial (UKALL XII/ECOG 2993). *Blood,* 114**,** 5136-5145.

MARTIN, C. L. & WARBURTON, D. 2015. Detection of Chromosomal Aberrations in Clinical Practice: From Karyotype to Genome Sequence. *Annual Review of Genomics and Human Genetics,* 16**,** 309-326.

MARTINELLI, G., PICIOCCHI, A., PAPAYANNIDIS, C., PAOLINI, S., ROBUSTELLI, V., SOVERINI, S., TERRAGNA, C., LEMOLI, R. M., GUOLO, F., DI BARTOLOMEO, P., LUNGHI, M., DE FABRITIIS, P., CANDONI, A., SELLERI, C., SIMONETTI, F., BOCCHIA, M., VITALE, A., FRISON, L., TEDESCHI, A., CUNEO, A., BONIFACIO, M., FALINI, B., D'ARDIA, S., TRAPPOLINI, S., TOSI, P., GALIENI, P., FABBIANO, F., ABBENANTE, M. C., MARCONI, G., SARTOR,

C., CAVO, M., FOÀ, R., FAZI, P., VIGNETTI, M. & BACCARANI, M. 2017. First Report of the Gimema LAL1811 Phase II Prospective Study of the Combination of Steroids with Ponatinib As Frontline Therapy of Elderly or Unfit Patients with Philadelphia Chromosome-Positive Acute Lymphoblastic Leukemia. *Blood,* 130**,** 99-99.

MATHESON, E. C., THOMAS, H., CASE, M., BLAIR, H., JACKSON, R. K., MASIC, D., VEAL, G., HALSEY, C., NEWELL, D. R., VORMOOR, J. & IRVING, J. A. E. 2019. Glucocorticoids and selumetinib are highly synergistic in RAS pathway-mutated childhood acute lymphoblastic leukemia through upregulation of BIM. *Haematologica,* 104**,** 1804-1811.

MAYAKONDA, A., LIN, D.-C., ASSENOV, Y., PLASS, C. & KOEFFLER, H. P. 2018. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Research*.

MCLAREN, W., GIL, L., HUNT, S. E., RIAT, H. S., RITCHIE, G. R. S., THORMANN, A., FLICEK, P. & CUNNINGHAM, F. 2016. The Ensembl Variant Effect Predictor. *Genome Biology,* 17**,** 122.

MEDEIROS, B. C., OTHUS, M., ESTEY, E. H., FANG, M. & APPELBAUM, F. R. 2014. Unsuccessful diagnostic cytogenetic analysis is a poor prognostic feature in acute myeloid leukaemia. *British journal of haematology,* 164**,** 245-250.

MEDICAL RESEARCH COUNCIL 1973. Treatment of acute lymphoblastic leukaemia: effect of "prophylactic" therapy against central nervous system leukaemia. Report ot the Medical Research Council by the Leukaemia Committee and the Working Party on Leukaemia in Childhood. *British medical journal,* 2**,** 381-384.

MENDES, R. D., CANTÉ-BARRETT, K., PIETERS, R. & MEIJERINK, J. P. P. 2016. The relevance of PTEN-AKT in relation to NOTCH1-directed treatment strategies in T-cell acute lymphoblastic leukemia. *Haematologica,* 101**,** 1010-1017.

MERMEL, C. H., SCHUMACHER, S. E., HILL, B., MEYERSON, M. L., BEROUKHIM, R. & GETZ, G. 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biology,* 12**,** R41.

MEYER, C., BURMEISTER, T., GRÖGER, D., TSAUR, G., FECHINA, L., RENNEVILLE, A., SUTTON, R., VENN, N. C., EMERENCIANO, M., POMBO-DE-OLIVEIRA, M. S., BARBIERI BLUNCK, C., ALMEIDA LOPES, B., ZUNA, J., TRKA, J., BALLERINI, P., LAPILLONNE, H., DE BRAEKELEER, M., CAZZANIGA, G., CORRAL ABASCAL, L., VAN DER VELDEN, V. H. J., DELABESSE, E., PARK, T. S., OH, S. H., SILVA, M. L. M., LUND-AHO, T., JUVONEN, V., MOORE, A. S., HEIDENREICH, O., VORMOOR, J., ZERKALENKOVA, E., OLSHANSKAYA, Y., BUENO, C., MENENDEZ, P., TEIGLER-SCHLEGEL, A., ZUR STADT, U., LENTES, J., GÖHRING, G., KUSTANOVICH, A., ALEINIKOVA, O., SCHÄFER, B. W., KUBETZKO, S., MADSEN, H. O., GRUHN, B., DUARTE, X., GAMEIRO, P., LIPPERT, E., BIDET, A., CAYUELA, J. M., CLAPPIER, E., ALONSO, C. N., ZWAAN, C. M., VAN DEN HEUVEL-EIBRINK, M. M., IZRAELI, S., TRAKHTENBROT, L., ARCHER, P., HANCOCK, J., MÖRICKE, A., ALTEN, J., SCHRAPPE, M., STANULLA, M., STREHL, S., ATTARBASCHI, A., DWORZAK, M., HAAS, O. A., PANZER-GRÜMAYER, R., SEDÉK, L., SZCZEPAŃSKI, T., CAYE, A., SUAREZ, L., CAVÉ, H. & MARSCHALEK, R. 2018. The MLL recombinome of acute leukemias in 2017. *Leukemia,* 32**,** 273-284.

MILLS, R. E., WALTER, K., STEWART, C., HANDSAKER, R. E., CHEN, K., ALKAN, C., ABYZOV, A., YOON, S. C., YE, K., CHEETHAM, R. K., CHINWALLA, A., CONRAD, D. F., FU, Y., GRUBERT, F., HAJIRASOULIHA, I., HORMOZDIARI, F., IAKOUCHEVA, L. M., IQBAL, Z., KANG, S., KIDD, J. M., KONKEL, M. K., KORN, J., KHURANA, E., KURAL, D., LAM, H. Y. K., LENG, J., LI, R., LI, Y., LIN, C.-Y., LUO, R., MU, X. J., NEMESH, J., PECKHAM, H. E., RAUSCH, T., SCALLY, A., SHI, X., STROMBERG, M. P., STÜTZ, A. M., URBAN, A. E., WALKER, J. A., WU, J., ZHANG, Y., ZHANG, Z. D., BATZER, M. A., DING, L., MARTH, G. T., MCVEAN, G., SEBAT, J., SNYDER, M., WANG, J., YE, K., EICHLER, E. E., GERSTEIN, M. B., HURLES, M. E., LEE, C., MCCARROLL, S. A., KORBEL, J. O. & GENOMES, P. 2011. Mapping copy number variation by population-scale genome sequencing. *Nature,* 470**,** 59-65.

MOORMAN, A. V. 2012. The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia. *Blood Reviews,* 26**,** 123-135.

MOORMAN, A. V. 2016. New and emerging prognostic and predictive genetic biomarkers in B-cell precursor acute lymphoblastic leukemia. *Haematologica,* 101**,** 407-416.

MOORMAN, A. V., BUTLER, E., BARRETTA, E., KIRKWOOD, A. A., SCHWAB, C., CREASEY, T., LEONGAMORNLERT, D. A., PAPAEMMANUIL, E., PATRICK, P., CLIFTON-HADLEY, L., WRENCH, B., MENNE, T. F., MCMILLAN, A. K., HARRISON, C. J., ROWNTREE, C. J., MARKS, D. I. & FIELDING, A. K. 2019. Prognostic Impact of Chromosomal Abnormalities and Copy Number Alterations Among Adults with B-Cell Precursor Acute Lymphoblastic Leukaemia Treated on UKALL14. *Blood,* 134**,** 288-288.

MOORMAN, A. V., CHILTON, L., WILKINSON, J., ENSOR, H. M., BOWN, N. & PROCTOR, S. J. 2010. A population-based cytogenetic study of adults with acute lymphoblastic leukemia. *Blood,* 115**,** 206-214.

MOORMAN, A. V., ENSHAEI, A., SCHWAB, C., WADE, R., CHILTON, L., ELLIOTT, A., RICHARDSON, S., HANCOCK, J., KINSEY, S. E., MITCHELL, C. D., GOULDEN, N., VORA, A. & HARRISON, C. J. 2014. A novel integrated cytogenetic and genomic classification refines risk stratification in pediatric acute lymphoblastic leukemia. *Blood,* 124**,** 1434-1444.

MOORMAN, A. V., HARRISON, C. J., BUCK, G. A. N., RICHARDS, S. M., SECKER-WALKER, L. M., MARTINEAU, M., VANCE, G. H., CHERRY, A. M., HIGGINS, R. R., FIELDING, A. K., FORONI, L., PAIETTA, E., TALLMAN, M. S., LITZOW, M. R., WIERNIK, P. H., ROWE, J. M., GOLDSTONE, A. H. & DEWALD, G. W. 2007. Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. *Blood,* 109**,** 3189-3197.

MOORMAN, A. V., SCHWAB, C., ENSOR, H. M., RUSSELL, L. J., MORRISON, H., JONES, L., MASIC, D., PATEL, B., ROWE, J. M., TALLMAN, M., GOLDSTONE, A. H., FIELDING, A. K. & HARRISON, C. J. 2012. IGH@ translocations, CRLF2 deregulation, and microdeletions in adolescents and adults with acute lymphoblastic leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* 30**,** 3100-3108.

MOORMAN, A. V., SCHWAB, C., WINTERMAN, E., HANCOCK, J., CASTLETON, A., CUMMINS, M., GIBSON, B., GOULDEN, N., KEARNS, P., JAMES, B.,

KIRKWOOD, A. A., LANCASTER, D., MADI, M., MCMILLAN, A., MOTWANI, J., NORTON, A., O'MARCAIGH, A., PATRICK, K., BHATNAGAR, N., QURESHI, A., RICHARDSON, D., STOKLEY, S., TAYLOR, G., VAN DELFT, F. W., MOPPETT, J., HARRISON, C. J., SAMARASINGHE, S. & VORA, A. 2020. Adjuvant tyrosine kinase inhibitor therapy improves outcome for children and adolescents with acute lymphoblastic leukaemia who have an ABL-class fusion. *British Journal of Haematology,* n/a.

MORI, H., COLMAN, S. M., XIAO, Z., FORD, A. M., HEALY, L. E., DONALDSON, C., HOWS, J. M., NAVARRETE, C. & GREAVES, M. 2002. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 8242-8247.

MRÓZEK, K. 2008. Cytogenetic, molecular genetic, and clinical characteristics of acute myeloid leukemia with a complex karyotype. *Seminars in oncology,* 35**,** 365-377.

MRÓZEK, K., HARPER, D. P. & APLAN, P. D. 2009. Cytogenetics and molecular genetics of acute lymphoblastic leukemia. *Hematology/oncology clinics of North America,* 23**,** 991-v.

MULLIGHAN, C. G. 2014. The genomic landscape of acute lymphoblastic leukemia in children and young adults. *Hematology,* 2014**,** 174-180.

MULLIGHAN, C. G., COLLINS-UNDERWOOD, J. R., PHILLIPS, L. A. A., LOUDIN, M. G., LIU, W., ZHANG, J., MA, J., COUSTAN-SMITH, E., HARVEY, R. C., WILLMAN, C. L., MIKHAIL, F. M., MEYER, J., CARROLL, A. J., WILLIAMS, R. T., CHENG, J., HEEREMA, N. A., BASSO, G., PESSION, A., PUI, C.-H., RAIMONDI, S. C., HUNGER, S. P., DOWNING, J. R., CARROLL, W. L. & RABIN, K. R. 2009a. Rearrangement of CRLF2 in B-progenitor– and Down syndrome–associated acute lymphoblastic leukemia. *Nature Genetics,* 41**,** 1243-1246.

MULLIGHAN, C. G. & DOWNING, J. R. 2009. Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions. *Leukemia,* 23**,** 1209.

MULLIGHAN, C. G., GOORHA, S., RADTKE, I., MILLER, C. B., COUSTAN-SMITH, E., DALTON, J. D., GIRTMAN, K., MATHEW, S., MA, J., POUNDS, S. B., SU, X., PUI, C.-H., RELLING, M. V., EVANS, W. E., SHURTLEFF, S. A. & DOWNING, J. R. 2007. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature,* 446**,** 758.

MULLIGHAN, C. G., MILLER, C. B., RADTKE, I., PHILLIPS, L. A., DALTON, J., MA, J., WHITE, D., HUGHES, T. P., LE BEAU, M. M., PUI, C.-H., RELLING, M. V., SHURTLEFF, S. A. & DOWNING, J. R. 2008a. BCR–ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature,* 453**,** 110-114.

MULLIGHAN, C. G., PHILLIPS, L. A., SU, X., MA, J., MILLER, C. B., SHURTLEFF, S. A. & DOWNING, J. R. 2008b. Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science (New York, N.Y.),* 322**,** 1377-1380.

MULLIGHAN, C. G., SU, X., ZHANG, J., RADTKE, I., PHILLIPS, L. A. A., MILLER, C. B., MA, J., LIU, W., CHENG, C., SCHULMAN, B. A., HARVEY, R. C., CHEN, I. M., CLIFFORD, R. J., CARROLL, W. L., REAMAN, G., BOWMAN, W. P., DEVIDAS, M., GERHARD, D. S., YANG, W., RELLING, M. V., SHURTLEFF, S. A., CAMPANA, D., BOROWITZ, M. J., PUI, C.-H., SMITH, M., HUNGER, S. P.,

WILLMAN, C. L., DOWNING, J. R. & THE CHILDREN'S ONCOLOGY, G. 2009b. Deletion of IKZF1 and Prognosis in Acute Lymphoblastic Leukemia. *New England Journal of Medicine,* 360**,** 470-480.

MULLIGHAN, C. G., ZHANG, J., HARVEY, R. C., COLLINS-UNDERWOOD, J. R., SCHULMAN, B. A., PHILLIPS, L. A., TASIAN, S. K., LOH, M. L., SU, X., LIU, W., DEVIDAS, M., ATLAS, S. R., CHEN, I. M., CLIFFORD, R. J., GERHARD, D. S., CARROLL, W. L., REAMAN, G. H., SMITH, M., DOWNING, J. R., HUNGER, S. P. & WILLMAN, C. L. 2009c. JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences,* 106**,** 9414.

MÜHLBACHER, V., ZENGER, M., SCHNITTGER, S., WEISSMANN, S., KUNZE, F., KOHLMANN, A., BELLOS, F., KERN, W., HAFERLACH, T. & HAFERLACH, C. 2014. Acute lymphoblastic leukemia with low hypodiploid/near triploid karyotype is a specific clinical entity and exhibits a very high TP53 mutation frequency of 93%. *Genes, Chromosomes and Cancer,* 53**,** 524-536.

NACHMAN, J. B., HEEREMA, N. A., SATHER, H., CAMITTA, B., FORESTIER, E., HARRISON, C. J., DASTUGUE, N., SCHRAPPE, M., PUI, C.-H., BASSO, G., SILVERMAN, L. B. & JANKA-SCHAUB, G. E. 2007. Outcome of treatment in children with hypodiploid acute lymphoblastic leukemia. *Blood,* 110**,** 1112-1115.

NATARAJAN, P., JAISWAL, S. & KATHIRESAN, S. 2018. Clonal Hematopoiesis. *Circulation: Genomic and Precision Medicine,* 11**,** e001926.

NEUMANN, M., HEESCH, S., SCHLEE, C., SCHWARTZ, S., GÖKBUGET, N., HOELZER, D., KONSTANDIN, N. P., KSIENZYK, B., VOSBERG, S., GRAF, A., KREBS, S., BLUM, H., RAFF, T., BRÜGGEMANN, M., HOFMANN, W.-K., HECHT, J., BOHLANDER, S. K., GREIF, P. A. & BALDUS, C. D. 2013. Whole-exome sequencing in adult ETP-ALL reveals a high rate of DNMT3A mutations. *Blood,* 121**,** 4749-4752.

NIIHORI, T., AOKI, Y., OHASHI, H., KUROSAWA, K., KONDOH, T., ISHIKIRIYAMA, S., KAWAME, H., KAMASAKI, H., YAMANAKA, T., TAKADA, F., NISHIO, K., SAKURAI, M., TAMAI, H., NAGASHIMA, T., SUZUKI, Y., KURE, S., FUJII, K., IMAIZUMI, M. & MATSUBARA, Y. 2005. Functional analysis of PTPN11/SHP-2 mutants identified in Noonan syndrome and childhood leukemia. *Journal of Human Genetics,* 50**,** 192-202.

NOETZLI, L., LO, R. W., LEE-SHERICK, A. B., CALLAGHAN, M., NORIS, P., SAVOIA, A., RAJPURKAR, M., JONES, K., GOWAN, K., BALDUINI, C., PECCI, A., GNAN, C., DE ROCCO, D., DOUBEK, M., LI, L., LU, L., LEUNG, R., LANDOLT-MARTICORENA, C., HUNGER, S., HELLER, P., GUTIERREZ-HARTMANN, A., XIAYUAN, L., PLUTHERO, F. G., ROWLEY, J. W., WEYRICH, A. S., KAHR, W. H. A., PORTER, C. C. & DI PAOLA, J. 2015. Germline mutations in ETV6 are associated with thrombocytopenia, red cell macrocytosis and predisposition to lymphoblastic leukemia. *Nature genetics,* 47**,** 535-538.

OHKI, K., KIYOKAWA, N., SAITO, Y., HIRABAYASHI, S., NAKABAYASHI, K., ICHIKAWA, H., MOMOZAWA, Y., OKAMURA, K., YOSHIMI, A., OGATA-KAWATA, H., SAKAMOTO, H., KATO, M., FUKUSHIMA, K., HASEGAWA, D., FUKUSHIMA, H., IMAI, M., KAJIWARA, R., KOIKE, T., KOMORI, I., MATSUI, A., MORI, M., MORIWAKI, K., NOGUCHI, Y., PARK, M.-J., UEDA, T., YAMAMOTO, S., MATSUDA, K., YOSHIDA, T., MATSUMOTO, K., HATA, K., KUBO, M., MATSUBARA, Y., TAKAHASHI, H., FUKUSHIMA, T., HAYASHI, Y., KOH, K., MANABE, A., OHARA, A. & TOKYO CHILDREN’S CANCER STUDY,

G. 2019. Clinical and molecular characteristics of MEF2D fusion-positive B-cell precursor acute lymphoblastic leukemia in childhood, including a novel translocation resulting in MEF2D-HNRNPH1 gene fusion. *Haematologica,* 104**,** 128-137.

OKAMOTO, R., OGAWA, S., NOWAK, D., KAWAMATA, N., AKAGI, T., KATO, M., SANADA, M., WEISS, T., HAFERLACH, C., DUGAS, M., RUCKERT, C., HAFERLACH, T. & PHILLIP KOEFFLER, H. 2010. Genomic profiling of adult acute lymphoblastic leukemia by single nucleotide polymorphism oligonucleotide microarray and comparison to pediatric acute lymphoblastic leukemia. *Haematologica,* 95**,** 1481-1488.

OLSHEN, A. B., VENKATRAMAN, E. S., LUCITO, R. & WIGLER, M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics,* 5**,** 557-572.

OTTMANN, O. G., WASSMANN, B., PFEIFER, H., GIAGOUNIDIS, A., STELLJES, M., DÜHRSEN, U., SCHMALZING, M., WUNDERLE, L., BINCKEBANCK, A. & HOELZER, D. 2007. Imatinib compared with chemotherapy as front-line treatment of elderly patients with Philadelphia chromosome-positive acute lymphoblastic leukemia (Ph+ALL). *Cancer,* 109**,** 2068-2076.

O'CONNOR, D., ENSHAEI, A., BARTRAM, J., HANCOCK, J., HARRISON, C. J., HOUGH, R., SAMARASINGHE, S., SCHWAB, C., VORA, A., WADE, R., MOPPETT, J., MOORMAN, A. V. & GOULDEN, N. 2018. Genotype-Specific Minimal Residual Disease Interpretation Improves Stratification in Pediatric Acute Lymphoblastic Leukemia. *Journal of Clinical Oncology,* 36**,** 34-43.

O'CONNOR, D., MOORMAN, A. V., WADE, R., HANCOCK, J., TAN, R. M. R., BARTRAM, J., MOPPETT, J., SCHWAB, C., PATRICK, K., HARRISON, C. J., HOUGH, R., GOULDEN, N., VORA, A. & SAMARASINGHE, S. 2017. Use of Minimal Residual Disease Assessment to Redefine Induction Failure in Pediatric Acute Lymphoblastic Leukemia. *Journal of Clinical Oncology,* 35**,** 660-667.

PAPAEMMANUIL, E., HOSKING, F. J., VIJAYAKRISHNAN, J., PRICE, A., OLVER, B., SHERIDAN, E., KINSEY, S. E., LIGHTFOOT, T., ROMAN, E., IRVING, J. A. E., ALLAN, J. M., TOMLINSON, I. P., TAYLOR, M., GREAVES, M. & HOULSTON, R. S. 2009. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nature genetics,* 41**,** 1006-1010.

PAPAEMMANUIL, E., RAPADO, I., LI, Y., POTTER, N. E., WEDGE, D. C., TUBIO, J., ALEXANDROV, L. B., VAN LOO, P., COOKE, S. L., MARSHALL, J., MARTINCORENA, I., HINTON, J., GUNDEM, G., VAN DELFT, F. W., NIK-ZAINAL, S., JONES, D. R., RAMAKRISHNA, M., TITLEY, I., STEBBINGS, L., LEROY, C., MENZIES, A., GAMBLE, J., ROBINSON, B., MUDIE, L., RAINE, K., O'MEARA, S., TEAGUE, J. W., BUTLER, A. P., CAZZANIGA, G., BIONDI, A., ZUNA, J., KEMPSKI, H., MUSCHEN, M., FORD, A. M., STRATTON, M. R., GREAVES, M. & CAMPBELL, P. J. 2014. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nature genetics,* 46**,** 116-125.

PAULSSON, K., CAZIER, J.-B., MACDOUGALL, F., STEVENS, J., STASEVICH, I., VRCELJ, N., CHAPLIN, T., LILLINGTON, D. M., LISTER, T. A. & YOUNG, B. D. 2008. Microdeletions are a general feature of adult and adolescent acute

lymphoblastic leukemia: Unexpected similarities with pediatric disease. *Proceedings of the National Academy of Sciences,* 105**,** 6708-6713.

PAULSSON, K., FORESTIER, E., LILLJEBJÖRN, H., HELDRUP, J., BEHRENDTZ, M., YOUNG, B. D. & JOHANSSON, B. 2010. Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences,* 107**,** 21719-21724.

PAULSSON, K. & JOHANSSON, B. 2009. High hyperdiploid childhood acute lymphoblastic leukemia. *Genes, Chromosomes and Cancer,* 48**,** 637-660.

PAULSSON, K., LILLJEBJÖRN, H., BILOGLAV, A., OLSSON, L., RISSLER, M., CASTOR, A., BARBANY, G., FOGELSTRAND, L., NORDGREN, A., SJÖGREN, H., FIORETOS, T. & JOHANSSON, B. 2015. The genomic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Nature Genetics,* 47**,** 672-676.

POPLIN, R., RUANO-RUBIO, V., DEPRISTO, M. A., FENNELL, T. J., CARNEIRO, M. O., VAN DER AUWERA, G. A., KLING, D. E., GAUTHIER, L. D., LEVY-MOONSHINE, A., ROAZEN, D., SHAKIR, K., THIBAULT, J., CHANDRAN, S., WHELAN, C., LEK, M., GABRIEL, S., DALY, M. J., NEALE, B., MACARTHUR, D. G. & BANKS, E. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv***,** 201178.

POULAIN, S., ROUMIER, C., VENET-CAILLAULT, A., FIGEAC, M., HERBAUX, C., MAROT, G., DOYE, E., BERTRAND, E., GEFFROY, S., LEPRETRE, F., NIBOUREL, O., DECAMBRON, A., BOYLE, E. M., RENNEVILLE, A., TRICOT, S., DAUDIGNON, A., QUESNEL, B., DUTHILLEUL, P., PREUDHOMME, C. & LELEU, X. 2016. Genomic Landscape of CXCR4 Mutations in Waldenström Macroglobulinemia. *Clinical Cancer Research,* 22**,** 1480.

POUNDS, S., CHENG, C., MULLIGHAN, C., RAIMONDI, S. C., SHURTLEFF, S. & DOWNING, J. R. 2009. Reference alignment of SNP microarray signals for copy number analysis of tumors. *Bioinformatics (Oxford, England),* 25**,** 315-321.

POWELL, B. C., JIANG, L., MUZNY, D. M., TREVIÑO, L. R., DREYER, Z. E., STRONG, L. C., WHEELER, D. A., GIBBS, R. A. & PLON, S. E. 2013. Identification of TP53 as an acute lymphocytic leukemia susceptibility gene through exome sequencing. *Pediatric blood & cancer,* 60**,** E1-E3.

PRESTON, D. L., KUSUMI, S., TOMONAGA, M., IZUMI, S., RON, E., KURAMOTO, A., KAMADA, N., DOHY, H., MATSUI, T., NONAKA, H., THOMPSON, D. E., SODA, M. & MABUCHI, K. 1994. Cancer Incidence in Atomic Bomb Survivors. Part III: Leukemia, Lymphoma and Multiple Myeloma, 1950-1987. *Radiation Research,* 137**,** S68-S97.

PUI, C.-H., BOYETT, J. M., RELLING, M. V., HARRISON, P. L., RIVERA, G. K., BEHM, F. G., SANDLUND, J. T., RIBEIRO, R. C., RUBNITZ, J. E., GAJJAR, A. & EVANS, W. E. 1999. Sex Differences in Prognosis for Children With Acute Lymphoblastic Leukemia. *Journal of Clinical Oncology,* 17**,** 818-818.

PUI, C.-H., REBORA, P., SCHRAPPE, M., ATTARBASCHI, A., BARUCHEL, A., BASSO, G., CAVÉ, H., ELITZUR, S., KOH, K., LIU, H.-C., PAULSSON, K., PIETERS, R., SILVERMAN, L. B., STARY, J., VORA, A., YEOH, A., HARRISON, C. J., VALSECCHI, M. G. & GROUP, O. B. O. T. P. D. L. C. A. W. 2019. Outcome of Children With Hypodiploid Acute Lymphoblastic Leukemia: A Retrospective Multinational Study. *Journal of Clinical Oncology,* 37**,** 770-779.

PUI, C.-H., ROBISON, L. L. & LOOK, A. T. 2008. Acute lymphoblastic leukaemia. *The Lancet,* 371**,** 1030-1043.

PUI, C.-H., SANDLUND, J. T., PEI, D., CAMPANA, D., RIVERA, G. K., RIBEIRO, R. C., RUBNITZ, J. E., RAZZOUK, B. I., HOWARD, S. C., HUDSON, M. M., CHENG, C., KUN, L. E., RAIMONDI, S. C., BEHM, F. G., DOWNING, J. R., RELLING, M. V. & EVANS, W. E. 2004. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital. *Blood,* 104**,** 2690-2696.

PUI, C. H., CARROLL, A. J., RAIMONDI, S. C., LAND, V. J., CRIST, W. M., SHUSTER, J. J., WILLIAMS, D. L., PULLEN, D. J., BOROWITZ, M. J., BEHM, F. G. & ET AL. 1990. Clinical presentation, karyotypic characterization, and treatment outcome of childhood acute lymphoblastic leukemia with a near-haploid or hypodiploid less than 45 line. *Blood,* 75**,** 1170-7.

PUI, C. H., WILLIAMS, D. L., RAIMONDI, S. C., RIVERA, G. K., LOOK, A. T., DODGE, R. K., GEORGE, S. L., BEHM, F. G., CRIST, W. M. & MURPHY, S. B. 1987. Hypodiploidy is associated with a poor prognosis in childhood acute lymphoblastic leukemia. *Blood,* 70**,** 247-53.

PUIGGROS, A., PUIGDECANET, E., SALIDO, M., FERRER, A., ABELLA, E., GIMENO, E., NONELL, L., HERRANZ, M. J., GALVÁN, A. B., RODRÍGUEZ-RIVERA, M., MELERO, C., PAIRET, S., BELLOSILLO, B., SERRANO, S., FLORENSA, L., SOLÉ, F. & ESPINET, B. 2013. Genomic arrays in chronic lymphocytic leukemia routine clinical practice: are we ready to substitute conventional cytogenetics and fluorescence in situ hybridization techniques? *Leukemia & Lymphoma,* 54**,** 986-995.

PULLARKAT, V., SLOVAK, M. L., KOPECKY, K. J., FORMAN, S. J. & APPELBAUM, F. R. 2008. Impact of cytogenetics on the outcome of adult acute lymphoblastic leukemia: results of Southwest Oncology Group 9400 study. *Blood,* 111**,** 2563-2572.

QUIVORON, C., COURONNÉ, L., DELLA VALLE, V., LOPEZ, C. K., PLO, I., WAGNER-BALLON, O., DO CRUZEIRO, M., DELHOMMEAU, F., ARNULF, B., STERN, M. H., GODLEY, L., OPOLON, P., TILLY, H., SOLARY, E., DUFFOURD, Y., DESSEN, P., MERLE-BERAL, H., NGUYEN-KHAC, F., FONTENAY, M., VAINCHENKER, W., BASTARD, C., MERCHER, T. & BERNARD, O. A. 2011. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell,* 20**,** 25-38.

R CORE TEAM 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

RAIMONDI, S. C., ZHOU, Y., MATHEW, S., SHURTLEFF, S. A., SANDLUND, J. T., RIVERA, G. K., BEHM, F. G. & PUI, C.-H. 2003. Reassessment of the prognostic significance of hypodiploidy in pediatric patients with acute lymphoblastic leukemia. *Cancer,* 98**,** 2715-2722.

RAINER, J., GATTO, L. & WEICHENBERGER, C. X. 2019. ensembldb: an R package to create and use Ensembl-based annotation resources. *Bioinformatics,* 35**,** 3151-3153.

RANDHAWA, S., CHO, B. S., GHOSH, D., SIVINA, M., KOEHRER, S., MÜSCHEN, M., PELED, A., DAVIS, R. E., KONOPLEVA, M. & BURGER, J. A. 2016. Effects of pharmacological and genetic disruption of CXCR4 chemokine receptor

function in B-cell acute lymphoblastic leukaemia. *British journal of haematology,* 174**,** 425-436.

REDON, R., ISHIKAWA, S., FITCH, K. R., FEUK, L., PERRY, G. H., ANDREWS, T. D., FIEGLER, H., SHAPERO, M. H., CARSON, A. R., CHEN, W., CHO, E. K., DALLAIRE, S., FREEMAN, J. L., GONZÁLEZ, J. R., GRATACÒS, M., HUANG, J., KALAITZOPOULOS, D., KOMURA, D., MACDONALD, J. R., MARSHALL, C. R., MEI, R., MONTGOMERY, L., NISHIMURA, K., OKAMURA, K., SHEN, F., SOMERVILLE, M. J., TCHINDA, J., VALSESIA, A., WOODWARK, C., YANG, F., ZHANG, J., ZERJAL, T., ZHANG, J., ARMENGOL, L., CONRAD, D. F., ESTIVILL, X., TYLER-SMITH, C., CARTER, N. P., ABURATANI, H., LEE, C., JONES, K. W., SCHERER, S. W. & HURLES, M. E. 2006. Global variation in copy number in the human genome. *Nature,* 444**,** 444-454.

RIBERA, J., ZAMORA, L., MORGADES, M., MALLO, M., SOLANES, N., BATLLE, M., VIVES, S., GRANADA, I., JUNCÀ, J., MALINVERNI, R., GENESCÀ, E., GUÀRDIA, R., MERCADAL, S., ESCODA, L., MARTINEZ-LOPEZ, J., TORMO, M., ESTEVE, J., PRATCORONA, M., MARTINEZ-LOSADA, C., SOLÉ, F., FELIU, E., RIBERA, J.-M., FOR THE SPANISH, P. G. & THE SPANISH SOCIETY OF, H. 2017. Copy number profiling of adult relapsed B-cell precursor acute lymphoblastic leukemia reveals potential leukemia progression mechanisms. *Genes, Chromosomes and Cancer,* 56**,** 810-820.

RIBERA, J.-M., GARCÍA, O., ORIOL, A., GIL, C., MONTESINOS, P., BERNAL, T., GONZÁLEZ-CAMPOS, J., LAVILLA, E., RIBERA, J., BRUNET, S., MARTÍNEZ, M.-P., TORMO, M., GENESCÀ, E., BARBA, P., SARRÀ, J., MONTESERÍN, M.-C., SORIA, B., COLORADO, M., CLADERA, A., GARCÍA-GUIÑÓN, A., CALBACHO, M., SERRANO, A., ORTÍN, X., PEDREÑO, M., AMIGO, M.-L., ESCODA, L. & FELIU, E. 2016. Feasibility and results of subtype-oriented protocols in older adults and fit elderly patients with acute lymphoblastic leukemia: Results of three prospective parallel trials from the PETHEMA group. *Leukemia Research,* 41**,** 12-20.

RIBERA, J.-M., ORIOL, A., MORGADES, M., MONTESINOS, P., SARRÀ, J., GONZÁLEZ-CAMPOS, J., BRUNET, S., TORMO, M., FERNÁNDEZ-ABELLÁN, P., GUÀRDIA, R., BERNAL, M.-T., ESTEVE, J., BARBA, P., MORENO, M.-J., BERMÚDEZ, A., CLADERA, A., ESCODA, L., GARCÍA-BOYERO, R., DEL POTRO, E., BERGUA, J., AMIGO, M.-L., GRANDE, C., RABUÑAL, M.-J., HERNÁNDEZ-RIVAS, J.-M. & FELIU, E. 2014. Treatment of High-Risk Philadelphia Chromosome–Negative Acute Lymphoblastic Leukemia in Adolescents and Adults According to Early Cytologic Response and Minimal Residual Disease After Consolidation Assessed by Flow Cytometry: Final Results of the PETHEMA ALL-AR-03 Trial. *Journal of Clinical Oncology,* 32**,** 1595-1604.

RIGAU, M., JUAN, D., VALENCIA, A. & RICO, D. 2019. Intronic CNVs and gene expression variation in human populations. *PLOS Genetics,* 15**,** e1007902.

RIGBY, P. W. J., DIECKMANN, M., RHODES, C. & BERG, P. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *Journal of Molecular Biology,* 113**,** 237-251.

ROBERTS, K. G., LI, Y., PAYNE-TURNER, D., HARVEY, R. C., YANG, Y.-L., PEI, D., MCCASTLAIN, K., DING, L., LU, C., SONG, G., MA, J., BECKSFORT, J., RUSCH, M., CHEN, S.-C., EASTON, J., CHENG, J., BOGGS, K., SANTIAGO-MORALES, N., IACOBUCCI, I., FULTON, R. S., WEN, J., VALENTINE, M.,

CHENG, C., PAUGH, S. W., DEVIDAS, M., CHEN, I. M., RESHMI, S., SMITH, A., HEDLUND, E., GUPTA, P., NAGAHAWATTE, P., WU, G., CHEN, X., YERGEAU, D., VADODARIA, B., MULDER, H., WINICK, N. J., LARSEN, E. C., CARROLL, W. L., HEEREMA, N. A., CARROLL, A. J., GRAYSON, G., TASIAN, S. K., MOORE, A. S., KELLER, F., FREI-JONES, M., WHITLOCK, J. A., RAETZ, E. A., WHITE, D. L., HUGHES, T. P., GUIDRY AUVIL, J. M., SMITH, M. A., MARCUCCI, G., BLOOMFIELD, C. D., MRÓZEK, K., KOHLSCHMIDT, J., STOCK, W., KORNBLAU, S. M., KONOPLEVA, M., PAIETTA, E., PUI, C.-H., JEHA, S., RELLING, M. V., EVANS, W. E., GERHARD, D. S., GASTIER-FOSTER, J. M., MARDIS, E., WILSON, R. K., LOH, M. L., DOWNING, J. R., HUNGER, S. P., WILLMAN, C. L., ZHANG, J. & MULLIGHAN, C. G. 2014. Targetable Kinase-Activating Lesions in Ph-like Acute Lymphoblastic Leukemia. *New England Journal of Medicine,* 371**,** 1005-1015.

ROBERTS, K. G., MORIN, R. D., ZHANG, J., HIRST, M., ZHAO, Y., SU, X., CHEN, S.-C., PAYNE-TURNER, D., CHURCHMAN, M. L., HARVEY, R. C., CHEN, X., KASAP, C., YAN, C., BECKSFORT, J., FINNEY, R. P., TEACHEY, D. T., MAUDE, S. L., TSE, K., MOORE, R., JONES, S., MUNGALL, K., BIROL, I., EDMONSON, M. N., HU, Y., BUETOW, K. E., CHEN, I. M., CARROLL, W. L., WEI, L., MA, J., KLEPPE, M., LEVINE, R. L., GARCIA-MANERO, G., LARSEN, E., SHAH, N. P., DEVIDAS, M., REAMAN, G., SMITH, M., PAUGH, S. W., EVANS, W. E., GRUPP, S. A., JEHA, S., PUI, C.-H., GERHARD, D. S., DOWNING, J. R., WILLMAN, C. L., LOH, M., HUNGER, S. P., MARRA, M. A. & MULLIGHAN, C. G. 2012. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer cell,* 22**,** 153-166.

ROBERTS, K. G. & MULLIGHAN, C. G. 2015. Genomics in acute lymphoblastic leukaemia: insights and treatment implications. *Nature Reviews Clinical Oncology,* 12**,** 344.

ROBERTS, K. G., ZHAOHUI, G., DEBBIE, P.-T., KELLY, M., RICHARD, C. H., CHEN, I. M., DEQING, P., ILARIA, I., MARCUS, V., STANLEY, B. P., LEI, S., YONGJIN, L., JINGHUI, Z., CHENG, C., ALESSANDRO, R., MANUELA, T., ORIETTA, S., JERALD, P. R., MARK, D. M., JACOB, M. R., SELINA, L., MARK, R. L., MARTIN, S. T., PETER, H. W., RAVI, B., IBRAHIM, A., JESSICA, K., KRZYSZTOF, M., GUIDO, M., CLARA, D. B., WENDY, S., STEPHEN, K., HAGOP, M. K., MARINA, K., ELISABETH, P., CHERYL, L. W. & CHARLES, G. M. 2017. High Frequency and Poor Outcome of Philadelphia Chromosome–Like Acute Lymphoblastic Leukemia in Adults. *Journal of Clinical Oncology,* 35**,** 394-401.

ROSENBERG, A. S., BRUNSON, A., PAULUS, J. K., TUSCANO, J., WUN, T., KEEGAN, T. H. M. & JONAS, B. A. 2017. Secondary acute lymphoblastic leukemia is a distinct clinical entity with prognostic significance. *Blood Cancer Journal,* 7**,** e605-e605.

ROUSSELOT, P., COUDÉ, M. M., GOKBUGET, N., GAMBACORTI PASSERINI, C., HAYETTE, S., CAYUELA, J.-M., HUGUET, F., LEGUAY, T., CHEVALLIER, P., SALANOUBAT, C., BONMATI, C., ALEXIS, M., HUNAULT, M., GLAISNER, S., AGAPE, P., BERTHOU, C., JOURDAN, E., FERNANDES, J., SUTTON, L., BANOS, A., REMAN, O., LIOURE, B., THOMAS, X., IFRAH, N., LAFAGE-POCHITALOFF, M., BORNAND, A., MORISSET, L., ROBIN, V., PFEIFER, H., DELANNOY, A., RIBERA, J., BASSAN, R., DELORD, M., HOELZER, D., DOMBRET, H., OTTMANN, O. G. & EUROPEAN WORKING GROUP ON

ADULT, A. L. L. G. 2016. Dasatinib and low-intensity chemotherapy in elderly patients with Philadelphia chromosome-positive ALL. *Blood,* 128**,** 774-782.

ROWE, J. M., BUCK, G., BURNETT, A. K., CHOPRA, R., WIERNIK, P. H., RICHARDS, S. M., LAZARUS, H. M., FRANKLIN, I. M., LITZOW, M. R., CIOBANU, N., PRENTICE, H. G., DURRANT, J., TALLMAN, M. S., GOLDSTONE, A. H., ECOG, F. & PARTY, T. M. N. A. L. W. 2005. Induction therapy for adults with acute lymphoblastic leukemia: results of more than 1500 patients from the international ALL trial: MRC UKALL XII/ECOG E2993. *Blood,* 106**,** 3760-3767.

ROWLEY, J. D. 1973. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature,* 243**,** 290-293.

RUSSELL, L., J., AMIR, E., LISA, J., AMY, E., DINO, M., HELEN, B., KARL, S. L., ADELE, K. F., ANTHONY, H. G., NICHOLAS, G., CHRISTOPHER, D. M., RACHEL, W., AJAY, V., ANTHONY, V. M. & CHRISTINE, J. H. 2014. IGH@ Translocations Are Prevalent in Teenagers and Young Adults With Acute Lymphoblastic Leukemia and Are Associated With a Poor Outcome. *Journal of Clinical Oncology,* 32**,** 1453-1462.

RUSSELL, L. J., JONES, L., ENSHAEI, A., TONIN, S., RYAN, S. L., ESWARAN, J., NAKJANG, S., PAPAEMMANUIL, E., TUBIO, J. M. C., FIELDING, A. K., VORA, A., CAMPBELL, P. J., MOORMAN, A. V. & HARRISON, C. J. 2017. Characterisation of the genomic landscape of CRLF2-rearranged acute lymphoblastic leukemia. *Genes, Chromosomes & Cancer,* 56**,** 363-372.

RYAN, S. L., MATHESON, E., GROSSMANN, V., SINCLAIR, P., BASHTON, M., SCHWAB, C., TOWERS, W., PARTINGTON, M., ELLIOTT, A., MINTO, L., RICHARDSON, S., RAHMAN, T., KEAVNEY, B., SKINNER, R., BOWN, N., HAFERLACH, T., VANDENBERGHE, P., HAFERLACH, C., SANTIBANEZ-KOREF, M., MOORMAN, A. V., KOHLMANN, A., IRVING, J. A. E. & HARRISON, C. J. 2016. The role of the RAS pathway in iAMP21-ALL. *Leukemia,* 30**,** 1824.

SAFAVI, S. & PAULSSON, K. 2017. Near-haploid and low-hypodiploid acute lymphoblastic leukemia: two distinct subtypes with consistently poor prognosis. *Blood,* 129**,** 420-423.

SAFTIG, P. & KLUMPERMAN, J. 2009. Lysosome biogenesis and lysosomal membrane proteins: trafficking meets function. *Nature Reviews Molecular Cell Biology,* 10**,** 623-635.

SALESSE, S. & VERFAILLIE, C. M. 2002. BCR/ABL: from molecular mechanisms of leukemia induction to treatment of chronic myelogenous leukemia. *Oncogene,* 21**,** 8547.

SANCHO, J.-M., RIBERA, J.-M., XICOY, B., MORGADES, M., ORIOL, A., TORMO, M., POTRO, E. D., DEBÉN, G., ABELLA, E., BETHENCOURT, C., ORTÍN, X., BRUNET, S., ORTEGA-RIVAS, F., NOVO, A., LÓPEZ, R., HERNÁNDEZ-RIVAS, J.-M., SANZ, M.-A., FELIU, E. & ON BEHALF OF THE PETHEMA GROUP, B. S. 2007. Results of the PETHEMA ALL-96 trial in elderly patients with Philadelphia chromosome-negative acute lymphoblastic leukemia. *European Journal of Haematology,* 78**,** 102-110.

SCHATZ, D. G. & JI, Y. 2011. Recombination centres and the orchestration of V(D)J recombination. *Nature Reviews Immunology,* 11**,** 251-263.

SCHINNERL, D., FORTSCHEGGER, K., KAUER, M., MARCHANTE, J. R. M., KOFLER, R., DEN BOER, M. L. & STREHL, S. 2015. The role of the Janus-faced transcription factor PAX5-JAK2 in acute lymphoblastic leukemia. *Blood,* 125**,** 1282-1291.

SCHOUMANS, J., SUELA, J., HASTINGS, R., MUEHLEMATTER, D., RACK, K., VAN DEN BERG, E., BERNA BEVERLOO, H. & STEVENS-KROEF, M. 2016. Guidelines for genomic array analysis in acquired haematological neoplastic disorders. *Genes, Chromosomes and Cancer,* 55**,** 480-491.

SCHRAPPE, M., HUNGER, S. P., PUI, C.-H., SAHA, V., GAYNON, P. S., BARUCHEL, A., CONTER, V., OTTEN, J., OHARA, A., VERSLUYS, A. B., ESCHERICH, G., HEYMAN, M., SILVERMAN, L. B., HORIBE, K., MANN, G., CAMITTA, B. M., HARBOTT, J., RIEHM, H., RICHARDS, S., DEVIDAS, M. & ZIMMERMANN, M. 2012. Outcomes after Induction Failure in Childhood Acute Lymphoblastic Leukemia. *New England Journal of Medicine,* 366**,** 1371-1381.

SCHULTZ, K. R., PULLEN, D. J., SATHER, H. N., SHUSTER, J. J., DEVIDAS, M., BOROWITZ, M. J., CARROLL, A. J., HEEREMA, N. A., RUBNITZ, J. E., LOH, M. L., RAETZ, E. A., WINICK, N. J., HUNGER, S. P., CARROLL, W. L., GAYNON, P. S. & CAMITTA, B. M. 2007. Risk- and response-based classification of childhood B-precursor acute lymphoblastic leukemia: a combined analysis of prognostic markers from the Pediatric Oncology Group (POG) and Children's Cancer Group (CCG). *Blood,* 109**,** 926-935.

SCHWAB, C. & HARRISON, C. J. 2018. Advances in B-cell Precursor Acute Lymphoblastic Leukemia Genomics. *HemaSphere,* 2**,** e53.

SCHWAB, C., RYAN, S. L., CHILTON, L., ELLIOTT, A., MURRAY, J., RICHARDSON, S., WRAGG, C., MOPPETT, J., CUMMINS, M., TUNSTALL, O., PARKER, C. A., SAHA, V., GOULDEN, N., VORA, A., MOORMAN, A. V. & HARRISON, C. J. 2016. *EBF1-PDGFRB* fusion in pediatric B-cell precursor acute lymphoblastic leukemia (BCP-ALL): genetic profile and clinical implications. *Blood,* 127**,** 2214-2218.

SCHWAB, C. J., CHILTON, L., MORRISON, H., JONES, L., AL-SHEHHI, H., ERHORN, A., RUSSELL, L. J., MOORMAN, A. V. & HARRISON, C. J. 2013. Genes commonly deleted in childhood B-cell precursor acute lymphoblastic leukemia: association with cytogenetics and clinical features. *Haematologica,* 98**,** 1081-1088.

SCHWAB, C. J., JONES, L. R., MORRISON, H., RYAN, S. L., YIGITTOP, H., SCHOUTEN, J. P. & HARRISON, C. J. 2010. Evaluation of multiplex ligation-dependent probe amplification as a method for the detection of copy number abnormalities in B-cell precursor acute lymphoblastic leukemia. *Genes, Chromosomes and Cancer,* 49**,** 1104-1113.

SCIONTI, F., DI MARTINO, M. T., PENSABENE, L., BRUNI, V. & CONCOLINO, D. 2018. The Cytoscan HD Array in the Diagnosis of Neurodevelopmental Disorders. *High-throughput,* 7**,** 28.

SCORE, J., CALASANZ, M. J., OTTMAN, O., PANE, F., YEH, R. F., SOBRINHO-SIMÕES, M. A., KREIL, S., WARD, D., HIDALGO-CURTIS, C., MELO, J. V., WIEMELS, J., NADEL, B., CROSS, N. C. P. & GRAND, F. H. 2010. Analysis of genomic breakpoints in p190 and p210 BCR–ABL indicate distinct mechanisms of formation. *Leukemia,* 24**,** 1742-1750.

SECKER-WALKER, L. M., LAWLER, S. D. & HARDISTY, R. M. 1978. Prognostic implications of chromosomal findings in acute lymphoblastic leukaemia at diagnosis. *British medical journal,* 2**,** 1529-1530.

SEISER, E. L. & INNOCENTI, F. 2015. Hidden Markov Model-Based CNV Detection Algorithms for Illumina Genotyping Microarrays. *Cancer informatics,* 13**,** 77-83.

SHEHATA, M., SCHWARZMEIER, J. D., HILGARTH, M., HUBMANN, R., DUECHLER, M. & GISSLINGER, H. 2004. TGF-beta1 induces bone marrow reticulin fibrosis in hairy cell leukemia. *The Journal of clinical investigation,* 113**,** 676-685.

SHEN, F., HUANG, J., FITCH, K. R., TRUONG, V. B., KIRBY, A., CHEN, W., ZHANG, J., LIU, G., MCCARROLL, S. A., JONES, K. W. & SHAPERO, M. H. 2008. Improved detection of global copy number variation using high density, non-polymorphic oligonucleotide probes. *BMC genetics,* 9**,** 27-27.

SHEN, H., LI, J., ZHANG, J., XU, C., JIANG, Y., WU, Z., ZHAO, F., LIAO, L., CHEN, J., LIN, Y., TIAN, Q., PAPASIAN, C. J. & DENG, H.-W. 2013. Comprehensive Characterization of Human Genome Variation by High Coverage Whole-Genome Sequencing of Forty Four Caucasians. *PLOS ONE,* 8**,** e59494.

SHERRY, S. T., WARD, M. H., KHOLODOV, M., BAKER, J., PHAN, L., SMIGIELSKI, E. M. & SIROTKIN, K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic acids research,* 29**,** 308-311.

SHLUSH, L. I., ZANDI, S., MITCHELL, A., CHEN, W. C., BRANDWEIN, J. M., GUPTA, V., KENNEDY, J. A., SCHIMMER, A. D., SCHUH, A. C., YEE, K. W., MCLEOD, J. L., DOEDENS, M., MEDEIROS, J. J. F., MARKE, R., KIM, H. J., LEE, K., MCPHERSON, J. D., HUDSON, T. J., PAN-LEUKEMIA GENE PANEL CONSORTIUM, T. H., BROWN, A. M. K., YOUSIF, F., TRINH, Q. M., STEIN, L. D., MINDEN, M. D., WANG, J. C. Y. & DICK, J. E. 2014. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature,* 506**,** 328-333.

SHUSTER, J. J., WACKER, P., PULLEN, J., HUMBERT, J., LAND, V. J., JR, D. H. M., LAUER, S., LOOK, A. T., BOROWITZ, M. J., CARROLL, A. J. & CAMITTA, B. 1998. Prognostic significance of sex in childhood B-precursor acute lymphoblastic leukemia: a Pediatric Oncology Group Study. *Journal of Clinical Oncology,* 16**,** 2854-2863.

SIM, N.-L., KUMAR, P., HU, J., HENIKOFF, S., SCHNEIDER, G. & NG, P. C. 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research,* 40**,** W452-W457.

SIVE, J. I., BUCK, G., FIELDING, A., LAZARUS, H. M., LITZOW, M. R., LUGER, S., MARKS, D. I., MCMILLAN, A., MOORMAN, A. V., RICHARDS, S. M., ROWE, J. M., TALLMAN, M. S. & GOLDSTONE, A. H. 2012. Outcomes in older adults with acute lymphoblastic leukaemia (ALL): results from the international MRC UKALL XII/ECOG2993 trial. *British Journal of Haematology,* 157**,** 463-471.

SLANY, R. K. 2009. The molecular biology of mixed lineage leukemia. *Haematologica,* 94**,** 984-993.

SOVERINI, S., COLAROSSI, S., GNANI, A., ROSTI, G., CASTAGNETTI, F., POERIO, A., IACOBUCCI, I., AMABILE, M., ABRUZZESE, E., ORLANDI, E., RADAELLI, F., CICCONE, F., TIRIBELLI, M., DI LORENZO, R., CARACCIOLO, C., IZZO, B., PANE, F., SAGLIO, G., BACCARANI, M. & MARTINELLI, G. 2006. Contribution of ABL Kinase Domain Mutations to Imatinib Resistance in

Different Subsets of Philadelphia-Positive Patients: By the GIMEMA Working Party on Chronic Myeloid Leukemia. *Clinical Cancer Research,* 12**,** 7374.

SPEICHER, M. R. & CARTER, N. P. 2005. The new cytogenetics: blurring the boundaries with molecular biology. *Nature Reviews Genetics,* 6**,** 782-792.

SPOO, A. C., LÜBBERT, M., WIERDA, W. G. & BURGER, J. A. 2006. CXCR4 is a prognostic marker in acute myelogenous leukemia. *Blood,* 109**,** 786-791.

STANULLA, M., DAGDAN, E., ZALIOVA, M., MÖRICKE, A., PALMI, C., CAZZANIGA, G., ECKERT, C., KRONNIE, G. T., BOURQUIN, J.-P., BORNHAUSER, B., KOEHLER, R., BARTRAM, C. R., LUDWIG, W.-D., BLECKMANN, K., GROENEVELD-KRENTZ, S., SCHEWE, D., JUNK, S. V., HINZE, L., KLEIN, N., KRATZ, C. P., BIONDI, A., BORKHARDT, A., KULOZIK, A., MUCKENTHALER, M. U., BASSO, G., VALSECCHI, M. G., IZRAELI, S., PETERSEN, B.-S., FRANKE, A., DÖRGE, P., STEINEMANN, D., HAAS, O. A., PANZER-GRÜMAYER, R., CAVÉ, H., HOULSTON, R. S., CARIO, G., SCHRAPPE, M., ZIMMERMANN, M., CONSORTIUM, F. T. T. & GROUP, T. I. B. S. 2018. IKZF1plus Defines a New Minimal Residual Disease–Dependent Very-Poor Prognostic Profile in Pediatric B-Cell Precursor Acute Lymphoblastic Leukemia. *Journal of Clinical Oncology,* 36**,** 1240-1249.

STARK, B., JEISON, M., GOBUZOV, R., KRUG, H., GLASER-GABAY, L., LURIA, D., EL-HASID, R., HARUSH, M. B., AVRAHAMI, G., FISHER, S., STEIN, J., ZAIZOV, R. & YANIV, I. 2001. Near haploid childhood acute lymphoblastic leukemia masked by hyperdiploid line: detection by fluorescence in situ hybridization. *Cancer Genet Cytogenet,* 128**,** 108-13.

STEEGHS, E. M. P., JERCHEL, I. S., DE GOFFAU-NOBEL, W., HOOGKAMER, A. Q., BOER, J. M., BOEREE, A., VAN DE VEN, C., KOUDIJS, M. J., BESSELINK, N. J. M., DE GROOT-KRUSEMAN, H. A., ZWAAN, C. M., HORSTMANN, M. A., PIETERS, R. & DEN BOER, M. L. 2017. JAK2 aberrations in childhood B-cell precursor acute lymphoblastic leukemia. *Oncotarget,* 8**,** 89923-89938.

STIEF, S. M., HANNEFORTH, A.-L., WESER, S., MATTES, R., CARLET, M., LIU, W.-H., BARTOSCHEK, M. D., DOMÍNGUEZ MORENO, H., OETTLE, M., KEMPF, J., VICK, B., KSIENZYK, B., TIZAZU, B., ROTHENBERG-THURLEY, M., QUENTMEIER, H., HIDDEMANN, W., VOSBERG, S., GREIF, P. A., METZELER, K. H., SCHOTTA, G., BULTMANN, S., JEREMIAS, I., LEONHARDT, H. & SPIEKERMANN, K. 2020. Loss of KDM6A confers drug resistance in acute myeloid leukemia. *Leukemia,* 34**,** 50-62.

STOCK, W., LA, M., SANFORD, B., BLOOMFIELD, C. D., VARDIMAN, J. W., GAYNON, P., LARSON, R. A. & NACHMAN, J. 2008. What determines the outcomes for adolescents and young adults with acute lymphoblastic leukemia treated on cooperative group protocols? A comparison of Children's Cancer Group and Cancer and Leukemia Group B studies. *Blood,* 112**,** 1646-1654.

STREFFORD, J. C., WORLEY, H., BARBER, K., WRIGHT, S., STEWART, A. R. M., ROBINSON, H. M., BETTNEY, G., VAN DELFT, F. W., ATHERTON, M. G., DAVIES, T., GRIFFITHS, M., HING, S., ROSS, F. M., TALLEY, P., SAHA, V., MOORMAN, A. V. & HARRISON, C. J. 2007. Genome complexity in acute lymphoblastic leukemia is revealed by array-based comparative genomic hybridization. *Oncogene,* 26**,** 4306-4318.

SULONG, S., MOORMAN, A. V., IRVING, J. A. E., STREFFORD, J. C., KONN, Z. J., CASE, M. C., MINTO, L., BARBER, K. E., PARKER, H., WRIGHT, S. L.,

STEWART, A. R. M., BAILEY, S., BOWN, N. P., HALL, A. G. & HARRISON, C. J. 2009. A comprehensive analysis of the CDKN2A gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups. *Blood,* 113**,** 100-107.

SYED, V. 2016. TGF-β Signaling in Cancer. *Journal of Cellular Biochemistry,* 117**,** 1279-1287.

TAI, Y. C., KVALE, M. N. & WITTE, J. S. 2010. Segmentation and estimation for SNP microarrays: a Bayesian multiple change-point approach. *Biometrics,* 66**,** 675-683.

TANAKA, T., MORITA, K., WANG, F., LITTLE, L., GUMBS, C., MATTHEWS, J., DINARDO, C. D., KADIA, T. M., RAVANDI, F., KONOPLEVA, M. Y., KANTARJIAN, H. M., CHAMPLIN, R. E., GARCIA-MANERO, G., FUTREAL, P. A. & TAKAHASHI, K. 2019. Clonal Dynamics and Clinical Implications of Post-Remission Clonal Hematopoiesis in Acute Myeloid Leukemia (AML). *Blood,* 134**,** 17-17.

TANASI, I., BA, I., SIRVENT, N., BRAUN, T., CUCCUINI, W., BALLERINI, P., DUPLOYEZ, N., TANGUY-SCHMIDT, A., TAMBURINI, J., MAURY, S., DORÉ, E., HIMBERLIN, C., DUCLOS, C., CHEVALLIER, P., ROUSSELOT, P., BONIFACIO, M., CAVÉ, H., BARUCHEL, A., DOMBRET, H., SOULIER, J., LANDMAN-PARKER, J., BOISSEL, N. & CLAPPIER, E. 2019. Efficacy of tyrosine kinase inhibitors in Ph-like acute lymphoblastic leukemia harboring ABL-class rearrangements. *Blood,* 134**,** 1351-1355.

TASIAN, S. K., DORAL, M. Y., BOROWITZ, M. J., WOOD, B. L., CHEN, I.-M., HARVEY, R. C., GASTIER-FOSTER, J. M., WILLMAN, C. L., HUNGER, S. P., MULLIGHAN, C. G. & LOH, M. L. 2012. Aberrant STAT5 and PI3K/mTOR pathway signaling occurs in human CRLF2-rearranged B-precursor acute lymphoblastic leukemia. *Blood,* 120**,** 833-842.

TASIAN, S. K., LOH, M. L. & HUNGER, S. P. 2017. Philadelphia chromosome-like acute lymphoblastic leukemia. *Blood,* 130**,** 2064-2072.

TATE, J. G., BAMFORD, S., JUBB, H. C., SONDKA, Z., BEARE, D. M., BINDAL, N., BOUTSELAKIS, H., COLE, C. G., CREATORE, C., DAWSON, E., FISH, P., HARSHA, B., HATHAWAY, C., JUPE, S. C., KOK, C. Y., NOBLE, K., PONTING, L., RAMSHAW, C. C., RYE, C. E., SPEEDY, H. E., STEFANCSIK, R., THOMPSON, S. L., WANG, S., WARD, S., CAMPBELL, P. J. & FORBES, S. A. 2018. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research,* 47**,** D941-D947.

TATE, J. G., BAMFORD, S., JUBB, H. C., SONDKA, Z., BEARE, D. M., BINDAL, N., BOUTSELAKIS, H., COLE, C. G., CREATORE, C., DAWSON, E., FISH, P., HARSHA, B., HATHAWAY, C., JUPE, S. C., KOK, C. Y., NOBLE, K., PONTING, L., RAMSHAW, C. C., RYE, C. E., SPEEDY, H. E., STEFANCSIK, R., THOMPSON, S. L., WANG, S., WARD, S., CAMPBELL, P. J. & FORBES, S. A. 2019. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic acids research,* 47**,** D941-D947.

TAYLOR, P. R., REID, M. M., BOWN, N., HAMILTON, P. J. & PROCTOR, S. J. 1992. Acute lymphoblastic leukemia in patients aged 60 years and over: a population-based study of incidence and outcome. *Blood,* 80**,** 1813-1817.

TERWILLIGER, T. & ABDUL-HAY, M. 2017. Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer Journal,* 7**,** e577-e577.

THOMAS, D. A., FADERL, S., CORTES, J., O'BRIEN, S., GILES, F. J., KORNBLAU, S. M., GARCIA-MANERO, G., KEATING, M. J., ANDREEFF, M., JEHA, S., BERAN, M., VERSTOVSEK, S., PIERCE, S., LETVAK, L., SALVADO, A., CHAMPLIN, R., TALPAZ, M. & KANTARJIAN, H. 2004a. Treatment of Philadelphia chromosome–positive acute lymphocytic leukemia with hyper-CVAD and imatinib mesylate. *Blood,* 103**,** 4396-4407.

THOMAS, X., BOIRON, J.-M., HUGUET, F., DOMBRET, H., BRADSTOCK, K., VEY, N., KOVACSOVICS, T., DELANNOY, A., FEGUEUX, N., FENAUX, P., STAMATOULLAS, A., VERNANT, J.-P., TOURNILHAC, O., BUZYN, A., REMAN, O., CHARRIN, C., BOUCHEIX, C., GABERT, J., LHÉRITIER, V. & FIERE, D. 2004b. Outcome of Treatment in Adults With Acute Lymphoblastic Leukemia: Analysis of the LALA-94 Trial. *Journal of Clinical Oncology,* 22**,** 4075-4086.

THOMAS, X., OLTEANU, N., CHARRIN, C., LHÉRITIER, V., MAGAUD, J.-P. & FIERE, D. 2001. Acute lymphoblastic leukemia in the elderly: The Edouard Herriot hospital experience. *American Journal of Hematology,* 67**,** 73-83.

TIAN, L., SHAO, Y., NANCE, S., DANG, J., XU, B., MA, X., LI, Y., JU, B., DONG, L., NEWMAN, S., ZHOU, X., SCHREINER, P., TSENG, E., HON, T., ASHBY, M., LI, C., EASTON, J., GRUBER, T. A. & ZHANG, J. 2019. Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia. *Nature Communications,* 10**,** 2789.

TOPP, M. S., GÖKBUGET, N., ZUGMAIER, G., DEGENHARD, E., GOEBELER, M.-E., KLINGER, M., NEUMANN, S. A., HORST, H. A., RAFF, T., VIARDOT, A., STELLJES, M., SCHAICH, M., KÖHNE-VOLLAND, R., BRÜGGEMANN, M., OTTMANN, O. G., BURMEISTER, T., BAEUERLE, P. A., NAGORSEN, D., SCHMIDT, M., EINSELE, H., RIETHMÜLLER, G., KNEBA, M., HOELZER, D., KUFER, P. & BARGOU, R. C. 2012. Long-term follow-up of hematologic relapse-free survival in a phase 2 study of blinatumomab in patients with MRD in B-lineage ALL. *Blood,* 120**,** 5185-5187.

TORRANO, V., PROCTER, J., CARDUS, P., GREAVES, M. & FORD, A. M. 2011. ETV6-RUNX1 promotes survival of early B lineage progenitor cells via a dysregulated erythropoietin receptor. *Blood,* 118**,** 4910-4918.

TREON, S. P., CAO, Y., XU, L., YANG, G., LIU, X. & HUNTER, Z. R. 2014. Somatic mutations in MYD88 and CXCR4 are determinants of clinical presentation and overall survival in Waldenström macroglobulinemia. *Blood,* 123**,** 2791-2796.

TRINH, D. L., SCOTT, D. W., MORIN, R. D., MENDEZ-LAGO, M., AN, J., JONES, S. J. M., MUNGALL, A. J., ZHAO, Y., SCHEIN, J., STEIDL, C., CONNORS, J. M., GASCOYNE, R. D. & MARRA, M. A. 2013. Analysis of FOXO1 mutations in diffuse large B-cell lymphoma. *Blood,* 121**,** 3666-3674.

TZONEVA, G., PEREZ-GARCIA, A., CARPENTER, Z., KHIABANIAN, H., TOSELLO, V., ALLEGRETTA, M., PAIETTA, E., RACEVSKIS, J., ROWE, J. M., TALLMAN, M. S., PAGANIN, M., BASSO, G., HOF, J., KIRSCHNER-SCHWABE, R., PALOMERO, T., RABADAN, R. & FERRANDO, A. 2013. Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. *Nature medicine,* 19**,** 368-371.

VAITKEVIČIENĖ, G., FORESTIER, E., HELLEBOSTAD, M., HEYMAN, M., JONSSON, O. G., LÄHTEENMÄKI, P. M., ROSTHOEJ, S., SÖDERHÄLL, S., SCHMIEGELOW, K., HAEMATOLOGY, O. B. O. T. N. S. O. P. & ONCOLOGY 2011. High white blood cell count at diagnosis of childhood acute lymphoblastic leukaemia: biological background and prognostic impact. Results from the NOPHO ALL-92 and ALL-2000 studies. *European Journal of Haematology,* 86**,** 38-46.

VAN DEN BERK, L. C. J., VAN DER VEER, A., WILLEMSE, M. E., THEEUWES, M. J. G. A., LUIJENDIJK, M. W., TONG, W. H., VAN DER SLUIS, I. M., PIETERS, R. & DEN BOER, M. L. 2014. Disturbed CXCR4/CXCL12 axis in paediatric precursor B-cell acute lymphoblastic leukaemia. *British Journal of Haematology,* 166**,** 240-249.

VAN DER MEULEN, J., SANGHVI, V., MAVRAKIS, K., DURINCK, K., FANG, F., MATTHIJSSENS, F., RONDOU, P., ROSEN, M., PIETERS, T., VANDENBERGHE, P., DELABESSE, E., LAMMENS, T., DE MOERLOOSE, B., MENTEN, B., VAN ROY, N., VERHASSELT, B., POPPE, B., BENOIT, Y., TAGHON, T., MELNICK, A. M., SPELEMAN, F., WENDEL, H.-G. & VAN VLIERBERGHE, P. 2015. The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. *Blood,* 125**,** 13-21.

VAN DER VELDEN, V. H. J., CAZZANIGA, G., SCHRAUDER, A., HANCOCK, J., BADER, P., PANZER-GRUMAYER, E. R., FLOHR, T., SUTTON, R., CAVE, H., MADSEN, H. O., CAYUELA, J. M., TRKA, J., ECKERT, C., FORONI, L., ZUR STADT, U., BELDJORD, K., RAFF, T., VAN DER SCHOOT, C. E., VAN DONGEN, J. J. M. & ON BEHALF OF THE EUROPEAN STUDY GROUP ON, M. R. D. D. I. A. L. L. 2007. Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. *Leukemia,* 21**,** 604-611.

VAN DONGEN, J. J. M., MACINTYRE, E. A., GABERT, J. A., DELABESSE, E., ROSSI, V., SAGLIO, G., GOTTARDI, E., RAMBALDI, A., DOTTI, G., GRIESINGER, F., PARREIRA, A., GAMEIRO, P., DIÁZ, M. G., MALEC, M., LANGERAK, A. W., SAN MIGUEL, J. F. & BIONDI, A. 1999. Standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease. *Leukemia,* 13**,** 1901-1928.

VAN GALEN, J. C., KUIPER, R. P., VAN EMST, L., LEVERS, M., TIJCHON, E., SCHEIJEN, B., WAANDERS, E., VAN REIJMERSDAL, S. V., GILISSEN, C., VAN KESSEL, A. G., HOOGERBRUGGE, P. M. & VAN LEEUWEN, F. N. 2010. BTG1 regulates glucocorticoid receptor autoinduction in acute lymphoblastic leukemia. *Blood,* 115**,** 4810-4819.

VAN HAAFTEN, G., DALGLIESH, G. L., DAVIES, H., CHEN, L., BIGNELL, G., GREENMAN, C., EDKINS, S., HARDY, C., O'MEARA, S., TEAGUE, J., BUTLER, A., HINTON, J., LATIMER, C., ANDREWS, J., BARTHORPE, S., BEARE, D., BUCK, G., CAMPBELL, P. J., COLE, J., FORBES, S., JIA, M., JONES, D., KOK, C. Y., LEROY, C., LIN, M.-L., MCBRIDE, D. J., MADDISON, M., MAQUIRE, S., MCLAY, K., MENZIES, A., MIRONENKO, T., MULDERRIG, L., MUDIE, L., PLEASANCE, E., SHEPHERD, R., SMITH, R., STEBBINGS, L., STEPHENS, P., TANG, G., TARPEY, P. S., TURNER, R., TURRELL, K., VARIAN, J., WEST, S., WIDAA, S., WRAY, P., COLLINS, V. P., ICHIMURA, K., LAW, S., WONG, J., YUEN, S. T., LEUNG, S. Y., TONON, G., DEPINHO, R. A., TAI, Y.-T., ANDERSON, K. C., KAHNOSKI, R. J., MASSIE, A., KHOO, S.

K., TEH, B. T., STRATTON, M. R. & FUTREAL, P. A. 2009. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nature Genetics,* 41**,** 521-523.

VAN LOO, P., NORDGARD, S. H., LINGJÆRDE, O. C., RUSSNES, H. G., RYE, I. H., SUN, W., WEIGMAN, V. J., MARYNEN, P., ZETTERBERG, A., NAUME, B., PEROU, C. M., BØRRESEN-DALE, A.-L. & KRISTENSEN, V. N. 2010. Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences,* 107**,** 16910-16915.

VATANSEVER, S., ERMAN, B. & GÜMÜŞ, Z. H. 2019. Oncogenic G12D mutation alters local conformations and dynamics of K-Ras. *Scientific Reports,* 9**,** 11730.

VIGNETTI, M., FAZI, P., CIMINO, G., MARTINELLI, G., DI RAIMONDO, F., FERRARA, F., MELONI, G., AMBROSETTI, A., QUARTA, G., PAGANO, L., REGE-CAMBRIN, G., ELIA, L., BERTIERI, R., ANNINO, L., FOÀ, R., BACCARANI, M. & MANDELLI, F. 2007. Imatinib plus steroids induces complete remissions and prolonged survival in elderly Philadelphia chromosome–positive patients with acute lymphoblastic leukemia without additional chemotherapy: results of the Gruppo Italiano Malattie Ematologiche dell'Adulto (GIMEMA) LAL0201-B protocol. *Blood,* 109**,** 3676-3678.

VORA, A., GOULDEN, N., WADE, R., MITCHELL, C., HANCOCK, J., HOUGH, R., ROWNTREE, C. & RICHARDS, S. 2013. Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): a randomised controlled trial. *The Lancet Oncology,* 14**,** 199-209.

WAANDERS, E., GU, Z., DOBSON, S. M., ANTIĆ, Ž., CRAWFORD, J. C., MA, X., EDMONSON, M. N., PAYNE-TURNER, D., VAN DE VORST, M., JONGMANS, M. C. J., MCGUIRE, I., ZHOU, X., WANG, J., SHI, L., POUNDS, S., PEI, D., CHENG, C., SONG, G., FAN, Y., SHAO, Y., RUSCH, M., MCCASTLAIN, K., YU, J., VAN BOXTEL, R., BLOKZIJL, F., IACOBUCCI, I., ROBERTS, K. G., WEN, J., WU, G., MA, J., EASTON, J., NEALE, G., OLSEN, S. R., NICHOLS, K. E., PUI, C.-H., ZHANG, J., EVANS, W. E., RELLING, M. V., YANG, J. J., THOMAS, P. G., DICK, J. E., KUIPER, R. P. & MULLIGHAN, C. G. 2020. Mutational Landscape and Patterns of Clonal Evolution in Relapsed Pediatric Acute Lymphoblastic Leukemia. *Blood Cancer Discovery,* 1**,** 96.

WALLIS, J. P. & REID, M. M. 1989. Bone marrow fibrosis in childhood acute lymphoblastic leukaemia. *J Clin Pathol,* 42**,** 1253–1254.

WANG, J., KONG, G., LIU, Y., DU, J., CHANG, Y.-I., TEY, S. R., ZHANG, X., RANHEIM, E. A., SABA-EL-LEIL, M. K., MELOCHE, S., DAMNERNSAWAD, A., ZHANG, J. & ZHANG, J. 2013. NrasG12D/+ promotes leukemogenesis by aberrantly regulating hematopoietic stem cell functions. *Blood,* 121**,** 5203-5207.

WANG, J.-H., NICHOGIANNOPOULOU, A., WU, L., SUN, L., SHARPE, A. H., BIGBY, M. & GEORGOPOULOS, K. 1996. Selective Defects in the Development of the Fetal and Adult Lymphoid System in Mice with an Ikaros Null Mutation. *Immunity,* 5**,** 537-549.

WANG, L. C., SWAT, W., FUJIWARA, Y., DAVIDSON, L., VISVADER, J., KUO, F., ALT, F. W., GILLILAND, D. G., GOLUB, T. R. & ORKIN, S. H. 1998. The TEL/ETV6 gene is required specifically for hematopoiesis in the bone marrow. *Genes & development,* 12**,** 2392-2402.

WANG, Q., KOTOULA, V., HSU, P.-C., PAPADOPOULOU, K., HO, J. W. K., FOUNTZILAS, G. & GIANNOULATOU, E. 2019. Comparison of somatic variant detection algorithms using Ion Torrent targeted deep sequencing data. *BMC Medical Genomics,* 12**,** 181.

WANG, Y., MILLER, S., ROULSTON, D., BIXBY, D. & SHAO, L. 2016. Genome-Wide Single-Nucleotide Polymorphism Array Analysis Improves Prognostication of Acute Lymphoblastic Leukemia/Lymphoma. *The Journal of Molecular Diagnostics,* 18**,** 595-603.

WASSMANN, B., PFEIFER, H., GOEKBUGET, N., BEELEN, D. W., BECK, J., STELLJES, M., BORNHÄUSER, M., REICHLE, A., PERZ, J., HAAS, R., GANSER, A., SCHMID, M., KANZ, L., LENZ, G., KAUFMANN, M., BINCKEBANCK, A., BRÜCK, P., REUTZEL, R., GSCHAIDMEIER, H., SCHWARTZ, S., HOELZER, D. & OTTMANN, O. G. 2006. Alternating versus concurrent schedules of imatinib and chemotherapy as front-line therapy for Philadelphia-positive acute lymphoblastic leukemia (Ph+ALL). *Blood,* 108**,** 1469-1477.

WELSCHINGER, R., LIEDTKE, F., BASNETT, J., DELA PENA, A., JUAREZ, J. G., BRADSTOCK, K. F. & BENDALL, L. J. 2013. Plerixafor (AMD3100) induces prolonged mobilization of acute lymphoblastic leukemia cells and increases the proportion of cycling cells in the blood in mice. *Experimental Hematology,* 41**,** 293-302.e1.

XIAO, H., WANG, L.-M., LUO, Y., LAI, X., LI, C., SHI, J., TAN, Y., FU, S., WANG, Y., ZHU, N., HE, J., ZHENG, W., YU, X., CAI, Z. & HUANG, H. 2016. Mutations in epigenetic regulators are involved in acute lymphoblastic leukemia relapse following allogeneic hematopoietic stem cell transplantation. *Oncotarget,* 7**,** 2696-2708.

XIE, M., LU, C., WANG, J., MCLELLAN, M. D., JOHNSON, K. J., WENDL, M. C., MCMICHAEL, J. F., SCHMIDT, H. K., YELLAPANTULA, V., MILLER, C. A., OZENBERGER, B. A., WELCH, J. S., LINK, D. C., WALTER, M. J., MARDIS, E. R., DIPERSIO, J. F., CHEN, F., WILSON, R. K., LEY, T. J. & DING, L. 2014. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature medicine,* 20**,** 1472-1478.

XU, C. 2018. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Computational and Structural Biotechnology Journal,* 16**,** 15-24.

YANG, L., WANG, H., KORNBLAU, S. M., GRABER, D. A., ZHANG, N., MATTHEWS, J. A., WANG, M., WEBER, D. M., THOMAS, S. K., SHAH, J. J., ZHANG, L., LU, G., ZHAO, M., MUDDASANI, R., YOO, S. Y., BAGGERLY, K. A. & ORLOWSKI, R. Z. 2011. Evidence of a role for the novel zinc-finger transcription factor ZKSCAN3 in modulating Cyclin D2 expression in multiple myeloma. *Oncogene,* 30**,** 1329-1340.

YODA, A., YODA, Y., CHIARETTI, S., BAR-NATAN, M., MANI, K., RODIG, S. J., WEST, N., XIAO, Y., BROWN, J. R., MITSIADES, C., SATTLER, M., KUTOK, J. L., DEANGELO, D. J., WADLEIGH, M., PICIOCCHI, A., DAL CIN, P., BRADNER, J. E., GRIFFIN, J. D., ANDERSON, K. C., STONE, R. M., RITZ, J., FOÀ, R., ASTER, J. C., FRANK, D. A. & WEINSTOCK, D. M. 2010. Functional screening identifies CRLF2 in precursor B-cell acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences,* 107**,** 252.

ZHANG, J., MCCASTLAIN, K., YOSHIHARA, H., XU, B., CHANG, Y., CHURCHMAN, M. L., WU, G., LI, Y., WEI, L., IACOBUCCI, I., LIU, Y., QU, C., WEN, J., EDMONSON, M., PAYNE-TURNER, D., KAUFMANN, K. B., TAKAYANAGI, S.-I., WIENHOLDS, E., WAANDERS, E., NTZIACHRISTOS, P., BAKOGIANNI, S., WANG, J., AIFANTIS, I., ROBERTS, K. G., MA, J., SONG, G., EASTON, J., MULDER, H. L., CHEN, X., NEWMAN, S., MA, X., RUSCH, M., GUPTA, P., BOGGS, K., VADODARIA, B., DALTON, J., LIU, Y., VALENTINE, M. L., DING, L., LU, C., FULTON, R. S., FULTON, L., TABIB, Y., OCHOA, K., DEVIDAS, M., PEI, D., CHENG, C., YANG, J., EVANS, W. E., RELLING, M. V., PUI, C.-H., JEHA, S., HARVEY, R. C., CHEN, I. M. L., WILLMAN, C. L., MARCUCCI, G., BLOOMFIELD, C. D., KOHLSCHMIDT, J., MRÓZEK, K., PAIETTA, E., TALLMAN, M. S., STOCK, W., FOSTER, M. C., RACEVSKIS, J., ROWE, J. M., LUGER, S., KORNBLAU, S. M., SHURTLEFF, S. A., RAIMONDI, S. C., MARDIS, E. R., WILSON, R. K., DICK, J. E., HUNGER, S. P., LOH, M. L., DOWNING, J. R., MULLIGHAN, C. G. & ST. JUDE CHILDREN'S RESEARCH HOSPITAL–WASHINGTON UNIVERSITY PEDIATRIC CANCER GENOME, P. 2016. Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nature genetics,* 48**,** 1481-1489.

ZHAO, R., CHOI, B. Y., LEE, M.-H., BODE, A. M. & DONG, Z. 2016. Implications of Genetic and Epigenetic Alterations of CDKN2A/(p16INK4a) in Cancer. *EBioMedicine,* 8**,** 30-39.

Supplementary tables

Supplementary table 1

Patient demographics and details of genetic/genomic analyses conducted for all older adults included in chapters 3, 4 and 6. 'Extended FISH' refers to FISH experiments used to identify gene rearrangements in B-other ALL patients (chapter 3).

Patient	Age (years)	Sex	Trial	Genetic subgroup	Cytogenetics	Extended FISH	SNP arrays	MLPA	Targeted NGS	Exome
24309	63	Male	UKALL14	B-other	✓	✓	✓	✓		
24813	62	Female	UKALL14	No data			✓	✓	✓	
24890	65	Male	UKALL14	B-other	✓		✓	✓	✓	
24919	64	Female	UKALL14	*BCR-ABL1*	✓		✓	✓		
24983	60	Female	UKALL14	*BCR-ABL1*	✓					
25082	62	Female	UKALL14	*BCR-ABL1*	✓		✓	✓	✓	
25100	63	Female	UKALL14	*KMT2A*-r	✓		✓		✓	
25101	63	Female	UKALL14	B-other	✓	✓				
25102	63	Female	UKALL14	HoTr	✓					
25123	60	Male	UKALL14	HoTr	✓					
25130	62	Female	UKALL14	*CRLF2*-r	✓	✓	✓	✓	✓	
25208	62	Male	UKALL14	*BCR-ABL1*	✓		✓	✓	✓	
25235	63	Male	UKALL14	*ZNF384*-r	✓	✓				
25237	63	Female	UKALL14	No data				✓		
25246	64	Male	UKALL14	*CRLF2*-r	✓	✓	✓	✓		
25247	64	Male	UKALL14	*BCR-ABL1*	✓		✓	✓	✓	
25267	63	Female	UKALL14	*MEF2D*-r	✓	✓	✓	✓	✓	
25293	63	Male	UKALL14	B-other	✓					
25344	61	Female	UKALL14	B-other	✓	✓	✓	✓		
25346	64	Male	UKALL14	*BCR-ABL1*	✓		✓	✓	✓	
25371	60	Female	UKALL14	*CRLF2*-r	✓	✓	✓	✓		
25373	65	Male	UKALL14	B-other	✓		✓	✓		
25415	64	Female	UKALL14	B-other	✓	✓				
25426	64	Female	UKALL14	B-other	✓	✓				
25437	64	Female	UKALL14	HoTr	✓		✓		✓	
25451	63	Male	UKALL14	*ZNF384*-r	✓	✓	✓	✓	✓	

Patient	Age (years)	Sex	Trial	Genetic subgroup	Cytogenetics	Extended FISH	SNP arrays	MLPA	Targeted NGS	Exome
25491	63	Male	UKALL14	*BCR-ABL1*	✓					
25548	60	Female	UKALL14	*BCR-ABL1*	✓		✓	✓		
25552	61	Male	UKALL14	*IGH*@-r	✓		✓	✓	✓	
25685	62	Male	UKALL14	No data						
25688	62	Male	UKALL14	T-cell	✓					
25694	60	Male	UKALL14	*KMT2A*-r	✓					
25695	63	Male	UKALL14	HoTr	✓					
25709	62	Female	UKALL14	*KMT2A*-r	✓			✓		
25793	67	Female	UKALL60	*BCR-ABL1*	✓		✓		✓	
25794	65	Female	UKALL60	HoTr	✓					
25842	64	Male	UKALL14	HoTr	✓					
25893	78	Female	UKALL60	*BCR-ABL1*	✓		✓	✓		
25894	63	Male	UKALL60	*IGH*@-r	✓					
25895	67	Male	UKALL60	B-other	✓	✓				
25907	70	Male	UKALL60	*IGH*@-r	✓	✓				
25925	61	Female	UKALL14	*BCR-ABL1*	✓					
25949	61	Female	UKALL14	B-other	✓					
25953	71	Female	UKALL60	*BCR-ABL1*	✓		✓	✓		
25967	60	Male	UKALL14	B-other	✓	✓	✓	✓	✓	
26062	61	Male	UKALL14	*BCR-ABL1*	✓		✓	✓		
26609	83	Male	UKALL60	*BCR-ABL1*	✓		✓	✓		
26610	65	Male	UKALL60	*BCR-ABL1*	✓		✓	✓		
26611	68	Male	UKALL60	B-other	✓	✓		✓		
26612	66	Male	UKALL60	B-other	✓	✓				
26613	66	Male	UKALL60	B-other	✓					
26614	75	Male	UKALL60	B-other	✓		✓	✓	✓	
26621	69	Female	UKALL60	T-cell	✓		✓	✓		
26659	60	Male	UKALL14	HoTr	✓		✓			
26660	62	Female	UKALL14	*BCR-ABL1*	✓		✓	✓	✓	
26682	63	Female	UKALL60	*BCR-ABL1*	✓		✓	✓		
26706	60	Male	UKALL14	HoTr	✓		✓			
26726	66	Male	UKALL60	B-other	✓		✓	✓		
26732	64	Female	UKALL14	*BCR-ABL1*	✓					
26768	70	Female	UKALL60	B-other	✓					
26971	67	Male	UKALL60	B-other	✓	✓	✓	✓		
26990	63	Female	UKALL60	B-other	✓	✓				
26995	70	Male	UKALL60	B-other	✓	✓	✓	✓		
27026	63	Female	UKALL14	*BCR-ABL1*	✓		✓	✓		

Patient	Age (years)	Sex	Trial	Genetic subgroup	Cytogenetics	Extended FISH	SNP arrays	MLPA	Targeted NGS	Exome
27033	72	Male	UKALL60	B-other	✓	✓				
27043	65	Female	UKALL14	*BCR-ABL1*	✓		✓	✓		
27071	60	Female	UKALL14	B-other	✓	✓	✓	✓		
27085	63	Female	UKALL60	*BCR-ABL1*	✓		✓			
27121	64	Female	UKALL60	HoTr	✓			✓		
27147	69	Female	UKALL60	*BCR-ABL1*	✓					
27181	65	Male	UKALL14	*IGH@*-r	✓	✓	✓	✓	✓	
27219	65	Female	UKALL60	B-other	✓	✓				
27298	66	Male	UKALL60	B-other	✓	✓	✓	✓		
27333	63	Female	UKALL14	*BCR-ABL1*	✓		✓	✓		
27389	73	Female	UKALL60	*KMT2A*-r	✓		✓		✓	
27391	65	Male	UKALL60	B-other	✓		✓	✓		
27392	73	Female	UKALL60	HoTr	✓		✓	✓		
27395	60	Male	UKALL14	T-cell	✓					
27407	69	Female	UKALL60	HoTr	✓					
27408	70	Male	UKALL60	T-cell	✓					
27409	74	Male	UKALL60	B-other	✓					
27441	61	Female	UKALL14	No data			✓	✓		
27452	62	Male	UKALL14	B-other	✓	✓				
27490	64	Male	UKALL60	B-other	✓	✓	✓			
27508	61	Male	UKALL14	HoTr	✓					
27509	69	Male	UKALL60	*BCR-ABL1*	✓		✓	✓		
27537	67	Female	UKALL60	HoTr	✓		✓	✓		
27554	78	Female	UKALL60	B-other	✓	✓	✓	✓		
27555	64	Male	UKALL60	HoTr	✓		✓	✓		
27556	75	Female	UKALL60	B-other	✓	✓				
27557	61	Male	UKALL14	T-cell	✓					
27579	60	Female	UKALL60	*BCR-ABL1*	✓					
27583	61	Female	UKALL60	*BCR-ABL1*	✓		✓	✓		
27584	76	Female	UKALL60	*BCR-ABL1*	✓		✓	✓		
27585	66	Female	UKALL60	*BCR-ABL1*	✓		✓	✓		
27596	62	Male	UKALL60	HoTr	✓		✓	✓		
27640	67	Male	UKALL60	B-other	✓	✓				
27642	72	Female	UKALL60	T-cell	✓		✓	✓	✓	
27668	68	Male	UKALL60	*BCR-ABL1*	✓					
27752	73	Female	UKALL60	B-other	✓	✓	✓	✓		
27754	63	Female	UKALL14	*BCR-ABL1*	✓		✓	✓		
27810	79	Female	UKALL60	No data						

Patient	Age (years)	Sex	Trial	Genetic subgroup	Cytogenetics	Extended FISH	SNP arrays	MLPA	Targeted NGS	Exome
27811	64	Male	UKALL14	HeH	✓					
27812	65	Male	UKALL14	*TCF3-PBX1*	✓		✓	✓		
27819	65	Female	UKALL60	B-other	✓	✓	✓	✓		
27833	73	Female	UKALL60	*IGH@*-r	✓	✓	✓	✓	✓	
27836	63	Male	UKALL14	*BCR-ABL1*	✓		✓	✓		
27887	68	Female	UKALL60	*BCR-ABL1*	✓					
27919	82	Female	UKALL60	B-other	✓					
27930	60	Male	UKALL60	B-other	✓					
27978	66	Male	UKALL60	No data	✓					
28011	61	Male	UKALL14	B-other	✓	✓	✓	✓		
28032	69	Female	UKALL60	T-cell	✓					
28033	63	Female	UKALL60	*KMT2A*-r	✓					
28034	74	Male	UKALL60	No data						
28039	76	Male	UKALL60	*CRLF2*-r	✓	✓		✓		
28050	66	Male	UKALL60	B-other	✓	✓				
28051	66	Male	UKALL60	*BCR-ABL1*	✓					
28057	64	Female	UKALL14	*BCR-ABL1*	✓		✓	✓	✓	
28076	63	Male	UKALL60	B-other	✓					
28091	64	Female	UKALL14	T-cell	✓					
28092	71	Female	UKALL60	HoTr	✓					
28093	72	Male	UKALL60	B-other	✓					
28104	64	Female	UKALL60	*TCF3-PBX1*	✓					
28105	78	Female	UKALL60	B-other	✓					
28135	79	Female	UKALL60	*BCR-ABL1*	✓					
28149	64	Female	UKALL14	B-other	✓		✓	✓		
28150	65	Male	UKALL60	*BCR-ABL1*	✓					
28168	65	Male	UKALL14	B-other	✓	✓	✓	✓		
28182	60	Female	UKALL14	*BCR-ABL1*	✓		✓	✓	✓	
28194	66	Male	UKALL60	B-other	✓	✓				
28196	61	Male	UKALL14	*BCR-ABL1*	✓		✓	✓		
28235	65	Female	UKALL60	*CRLF2*-r	✓	✓				
28310	66	Female	UKALL60	B-other	✓	✓				
28312	67	Female	UKALL60	*BCR-ABL1*	✓					
28317	80	Male	UKALL60	B-other	✓	✓				
28334	64	Male	UKALL14	HoTr	✓					
28335	63	Male	UKALL14	B-other	✓		✓	✓	✓	
28350	62	Female	UKALL60	*BCR-ABL1*	✓		✓	✓	✓	
28361	67	Male	UKALL60	HoTr	✓					

Patient	Age (years)	Sex	Trial	Genetic subgroup	Cytogenetics	Extended FISH	SNP arrays	MLPA	Targeted NGS	Exome
28370	81	Male	UKALL60	*KMT2A*-r	✓					
28403	71	Male	UKALL60	*KMT2A*-r	✓					
28404	63	Female	UKALL60	HoTr	✓					
28406	74	Male	UKALL60	HoTr	✓					
28456	62	Male	UKALL14	B-other	✓					
28581	65	Female	UKALL14	*KMT2A*-r	✓		✓	✓	✓	
28620	60	Male	UKALL14	*BCR-ABL1*	✓					
28644	64	Female	UKALL14	HoTr	✓		✓			
28670	61	Female	UKALL14	*BCR-ABL1*	✓		✓	✓	✓	
28905	60	Male	UKALL14	B-other	✓	✓				
28934	60	Female	UKALL14	*KMT2A*-r	✓		✓	✓		
28945	64	Female	UKALL14	B-other	✓	✓				
29089	60	Female	UKALL14	*BCR-ABL1*	✓		✓	✓		
29202	61	Male	UKALL14	B-other	✓					
29407	60	Female	UKALL14	HoTr	✓		✓		✓	
29453	63	Male	UKALL14	T-cell	✓					
29454	65	Female	UKALL14	B-other	✓					
29481	69	Female	UKALL60	B-other	✓					
29517	65	Male	UKALL14	*BCR-ABL1*	✓					
29519	61	Female	UKALL14	B-other	✓					
29589	61	Male	UKALL14	*BCR-ABL1*	✓		✓	✓		
29655	65	Female	UKALL60	*BCR-ABL1*	✓					
29710	67	Male	UKALL60	B-other	✓	✓				
29741	77	Male	UKALL60	*BCR-ABL1*	✓					
29779	87	Male	Non-trial	B-other	✓	✓				✓
29780	55	Female	Non-trial	B-other	✓	✓				✓
29808	66	Male	UKALL60	*IGH@*-r	✓					
29809	63	Male	UKALL60	*BCR-ABL1*	✓					
29848	68	Female	UKALL60	B-other	✓					
29849	70	Female	UKALL60	HoTr	✓					
29854	75	Female	Non-trial	*BCR-ABL1*	✓					✓
29881	63	Male	UKALL60	*BCR-ABL1*	✓					
29882	71	Female	UKALL60	HoTr	✓					
29908	62	Female	UKALL14	*KMT2A*-r	✓		✓		✓	
29958	79	Male	UKALL60	*BCR-ABL1*	✓					
30031	73	Female	UKALL60	*BCR-ABL1*	✓					
30032	-	Male	Non-trial	*BCR-ABL1*	✓		✓			
30033	-	Female	Non-trial	B-other	✓		✓			

Patient	Age (years)	Sex	Trial	Genetic subgroup	Cytogenetics	Extended FISH	SNP arrays	MLPA	Targeted NGS	Exome
30034	-	Female	Non-trial	B-other	✓		✓			
30035	-	Male	Non-trial	*BCR-ABL1*	✓		✓			
30036	70	Male	Non-trial	B-other	✓		✓			
30063	69	Male	UKALL60	No data						
30066	63	Male	UKALL14	HeH	✓					
30085	67	Male	UKALL60	*ZNF384*-r	✓	✓				
30086	69	Female	UKALL60	B-other	✓	✓				
30102	67	Female	UKALL60	*CRLF2*-r	✓	✓				
30103	64	Female	UKALL14	*BCR-ABL1*	✓					
30142	66	Female	Non-trial	B-other	✓	✓				✓
30175	63	Female	UKALL14	B-other	✓					
30236	73	Male	UKALL60	T-cell	✓					
30237	74	Female	UKALL60	B-other	✓					
30297	64	Female	UKALL60	*CRLF2*-r	✓	✓				
30298	69	Female	UKALL60	B-other	✓					
30299	74	Female	UKALL60	*CRLF2*-r	✓	✓				
30300	74	Female	UKALL60	*KMT2A*-r	✓					
30315	69	Female	UKALL60	HoTr	✓					
30331	77	Female	UKALL60	HoTr	✓					
30334	60	Male	UKALL14	B-other	✓					
30347	70	Male	UKALL60	B-other	✓					
30378	71	Male	UKALL60	B-other	✓	✓				
30389	83	Female	UKALL60	B-other	✓					
30390	73	Male	UKALL60	HoTr	✓					
30402	62	Female	UKALL60	*TCF3-PBX1*	✓					
30403	71	Male	UKALL60	B-other	✓					
30419	64	Female	UKALL60	B-other	✓					
30426	72	Male	UKALL60	B-other	✓					
30428	61	Female	UKALL14	HoTr	✓					
30438	81	Male	UKALL60	B-other	✓	✓				
30476	71	Male	UKALL60	B-other	✓					
30487	60	Male	UKALL14	*IGH@*-r	✓					
30521	66	Male	UKALL60	*BCR-ABL1*	✓					
30556	75	Male	UKALL60	B-other	✓					
30557	78	Female	UKALL60	*BCR-ABL1*	✓					
30623	64	Male	UKALL14	*KMT2A*-r	✓					
30641	62	Male	UKALL60	B-other	✓					
30643	52	Male	Non-trial	*CRLF2*-r	✓	✓				✓

Patient	Age (years)	Sex	Trial	Genetic subgroup	Cytogenetics	Extended FISH	SNP arrays	MLPA	Targeted NGS	Exome
30721	61	Female	UKALL14	No data						
31044	63	Female	Non-trial	HoTr			✓			✓
31085	62	Female	UKALL14	No data						
31095	64	Male	UKALL14	T-cell	✓					
31145	62	Male	UKALL14	*BCR-ABL1*	✓					

Supplementary table 2

Nexus segmentation settings used for SNP array patient cohort.

Patient	SNP array	Systematic correction	Significance threshold	Gain	Loss	Min probes
24309	Illumina	Sequential Loess	1.00E-14	0.06	-0.06	6
24813	Illumina	Sequential Loess	1.00E-18	0.1	-0.1	6
24890	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
24919	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25082	Illumina	Sequential Loess	1.00E-13	0.06	-0.06	6
25100	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25130	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25208	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25246	Illumina	Sequential Loess	1.00E-14	0.04	-0.04	6
25247	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25267	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25344	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25346	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25371	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25373	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25437	Illumina	Sequential Loess	1.00E-15	0.09	-0.09	6
25451	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
25548	Illumina	Sequential Loess	1.00E-12	0.1	-0.1	6
25552	Illumina	Sequential Loess	1.00E-15	0.05	-0.05	6
25793	Illumina	Sequential Loess	1.00E-25	0.05	-0.05	6
25893	Affymetrix	Quadratic	1.00E-20	0.15	-0.15	10
25953	Affymetrix	Quadratic	1.00E-20	0.15	-0.15	10
25967	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
26062	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
26609	Affymetrix	Quadratic	1.00E-20	0.15	-0.15	10
26610	Affymetrix	Quadratic	1.00E-20	0.15	-0.15	10
26614	Affymetrix	Quadratic	1.00E-15	0.15	-0.15	10
26621	Affymetrix	Quadratic	1.00E-18	0.15	-0.15	10
26659	Illumina	Sequential Loess	1.00E-25	0.06	-0.06	6
26660	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
26682	Affymetrix	Quadratic	1.00E-18	0.15	-0.15	10
26706	Illumina	Sequential Loess	1.00E-35	0.08	-0.08	6
26726	Affymetrix	Quadratic	1.00E-20	0.15	-0.15	10
26971	Affymetrix	Quadratic	1.00E-18	0.07	-0.07	10
26995	Affymetrix	Quadratic	1.00E-15	0.15	-0.15	10
27026	Illumina	Sequential Loess	1.00E-25	0.04	-0.05	6
27043	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6

Patient	SNP array	Systematic correction	Significance threshold	Gain	Loss	Min probes
27071	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
27085	Affymetrix	Quadratic	1.00E-22	0.15	-0.15	10
27181	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
27298	Affymetrix	Quadratic	1.00E-30	0.1	-0.1	10
27333	Illumina	Sequential Loess	1.00E-18	0.05	-0.05	6
27389	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
27391	Affymetrix	Quadratic	1.00E-30	0.15	-0.15	10
27392	Affymetrix	Quadratic	1.00E-25	0.15	-0.15	10
27441	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
27490	Affymetrix	Quadratic	1.00E-30	0.15	-0.15	10
27509	Affymetrix	Quadratic	1.00E-30	0.05	-0.05	10
27537	Affymetrix	Quadratic	1.00E-30	0.05	-0.05	10
27554	Affymetrix	Quadratic	1.00E-28	0.08	-0.08	10
27555	Illumina	Sequential Loess	1.00E-32	0.02	-0.02	6
27583	Affymetrix	Quadratic	1.00E-25	0.08	-0.08	10
27584	Affymetrix	Quadratic	1.00E-20	0.15	-0.15	10
27585	Affymetrix	Quadratic	1.00E-28	0.08	-0.08	10
27596	Affymetrix	Quadratic	1.00E-18	0.15	-0.15	10
27642	Affymetrix	Quadratic	1.00E-25	0.15	-0.15	10
27752	Affymetrix	Quadratic	1.00E-18	0.15	-0.15	10
27754	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
27812	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
27819	Affymetrix	Quadratic	1.00E-25	0.15	-0.15	10
27833	Affymetrix	Quadratic	1.00E-22	0.15	-0.15	10
27836	Illumina	Sequential Loess	1.00E-15	0.06	-0.06	6
28011	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
28057	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
28149	Illumina	Sequential Loess	1.00E-10	0.08	-0.08	6
28168	Illumina	Sequential Loess	1.00E-13	0.15	-0.15	6
28182	Illumina	Sequential Loess	1.00E-16	0.07	-0.07	6
28196	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
28335	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
28350	Illumina	Sequential Loess	1.00E-20	0.05	-0.05	6
28581	Illumina	Sequential Loess	1.00E-20	0.15	-0.15	6
28644	Illumina	Sequential Loess	1.00E-20	0.05	-0.05	6
28670	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
28934	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
29089	Illumina	Sequential Loess	1.00E-12	0.08	-0.08	6
29407	Illumina	Sequential Loess	1.00E-16	0.05	-0.05	6

Patient	SNP array	Systematic correction	Significance threshold	Gain	Loss	Min probes
29589	Illumina	Sequential Loess	1.00E-16	0.05	-0.05	6
29908	Illumina	Sequential Loess	1.00E-12	0.1	-0.1	6
30032	Affymetrix	Quadratic	1.00E-20	0.15	-0.15	10
30033	Affymetrix	Quadratic	1.00E-22	0.15	-0.15	10
30034	Affymetrix	Quadratic	1.00E-22	0.15	-0.15	10
30035	Affymetrix	Quadratic	1.00E-22	0.15	-0.15	10
30036	Affymetrix	Quadratic	1.00E-22	0.15	-0.15	10

Supplementary table 3

Genomic targets for full regions included in SureSelect XT2 capture library design.

Gene	Interval	Size	Coverage (%)	Total RNA probes
KDM6A	chrX:44732411-44971867	239,457	69.8	1,202
COL11A1	chr1:103342013-103574062	232,050	82.2	1,422
PTEN	chr10:89622860-89731697	108,838	77.8	636
LEMD3	chr12:65569801-65589942	20,142	90.0	134
MBNL1	chr3:151880890-151998320	117,431	82.1	720
NF1	chr17:29421935-29709144	287,210	73.9	1,562
CDC73	chr1:193071243-193132293	61,051	77.8	361
IKZF1	chr7:50343669-50472809	129,141	66.1	683
PAX5	chr9:36833262-37034486	201,225	88.5	1,424
FLT3	chr13:28577401-28674739	97,339	60.3	402
MEF2C	chr5:88013965-88199932	185,968	88.9	1,244
ETV6	chr12:11802778-12048346	245,569	90.1	1,688
JAK2	chr9:4985023-5128193	143,171	69.8	785
TCF4	chr18:52889552-53332028	442,477	90.5	3,063
ARID2	chr12:46110000-46301833	191,834	72.4	1,009
RAG1	chr11:36532249-36614716	82,468	63.1	394
NIPBL	chr5:36876851-37066525	189,675	79.8	1,143
VLDLR	chr9:2230000-2480000	250,001	71.0	1,337
ZEB2	chr2:145092000-145180000	88,001	91.3	612
MEF2D	chr1:156433503-156470644	37,142	91.3	286
ABL1	chr9:133589258-133763072	173,815	70.5	881
PDGFRB	chr5:149493390-149535445	42,056	95.0	337
TCF3	chr19:1609279-1652614	43,336	86.2	389
PAR1	chrx:1321616-1397693	76,078	34.7	174
DGKH	chr13:42614162-42830726	216,565	83.7	1,359
ABL2	chr1:179068452-179198829	130,378	66.9	619
CSF1R	chr5:149432844-149492945	60,102	75.5	343
CXCR4	chr2:136860000-136886000	26,001	90.4	179
ROCK1	chr18:18526785-18708000	181,216	71.5	956
6p22.1	chr6:28310000-28333000	23,001	84.7	151

Supplementary table 4

Genomic targets for coding regions only (exons) included in SureSelect XT2 capture library design.

Gene	Interval	Regions	Size	Coverage (%)
ASXL1	chr20:30946569-31025151	17	5,060	99.5
ATM	chr11:108098342-108236245	62	10,411	100
CREBBP	chr16:3777709-3929927	31	7,988	100
DNMT3A	chr2:25457138-25536863	25	3,388	100
FOXO1	chr13:41133650-41240359	2	2,008	100
IKZF2	chr2:213872074-214014917	12	2,065	100
IKZF3	chr17:37922033-38020389	9	1,793	100
IL7R	chr5:35857070-35876598	8	1,716	100
JAK1	chr1:65300235-65351957	24	3,945	100
JAK3	chr19:17937542-17955236	23	3,913	100
KMT2C	chr7:151833907-152132881	60	16,122	100
KRAS	chr12:25362719-25398328	6	828	100
NOTCH1	chr9:139390513-139440248	34	8,348	100
NR3C1	chr5:142658919-142780414	9	2,573	100
NRAS	chr1:115251146-115258791	4	650	100
NT5C2	chr10:104849419-104934725	22	2,373	100
PTPN11	chr12:112856906-112942578	16	2,142	97.4
RB1	chr13:48878039-49054217	27	3,327	100
RUNX1	chr21:36164422-36421206	11	1,804	100
SH2B3	chr12:111855940-111886116	9	2,144	100
TET2	chr4:106111617-106197686	10	6,365	100
TFDP3	chrX:132351060-132352297	1	1,238	100
TOX	chr8:59720296-60031556	9	1,761	100
TP53	chr17:7565247-7579922	14	1,658	94.2
ZFHX3	chr16:72821053-72994054	9	11,292	100

Supplementary table 5

Sequencing depth of coverage of samples included in SureSelect XT2 library prep calculated as follows: (average read length X number of reads) / genome length

Patient ID	Depth of coverage (x)
26614	220.99
25552	277.12
26660	294.66
26706	301.68
25967	305.18
28893	308.69
25451	312.20
27400	315.71
26910	319.21
25208	322.72
24890	329.74
28670	333.25
29407	333.25
25247	340.26
28581	343.77
24401	347.28
27478	350.79
28350	361.31
29491	364.82
27642	371.83
27833	371.83
28057	389.37
25267	396.39
25082	406.91
28182	406.91
25100	410.42
25437	410.42
25130	427.96
28335	427.96
24813	438.48

Supplementary table 6

Cytogenetic data for all cases included in the CART analysis of ploidy status in ALL. Genetic subgroups based on cytogenetics and SNP array are also shown as well as CART input and output classes.

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
4949	HoTr	LOH-LCN	HoTr	Non-ploidy	34~35,XX,der(2)t(7;13;2)(p15;?;p11),-3,-4,-5,der(6)t(6;7)(q2?5;q22),-7,-9,der(9)t(9;18)(p?1;p1),del(9)(p2?),dic(12;22)(p1;p1),-13,der(14)t(12;14)(p?10;p10),-15,-16,-17,-18,-20,-22[cp8]
4950	HoTr	Inc	HoTr	Non-ploidy	31-38,Y,-X,-2,-8,-9,-10,-13,-14,-15,-16,-17,-19,-22,inc[cp8]
24401	HoTr	Inc	HoTr	HoTr	37,XX,-2,-3,-7,-12,-13,-15,-16,-17,-20[6]
26076	HoTr	LOH-LCN	HoTr	HoTr	37,XY,-2,-3,-4,-7,-12,-13,-15,-16,-17[8]
26082	HoTr	LOH-LCN	HoTr	HoTr	39,X,-X,-3,der(7)t(7;11)(q11.2;q13),-8,-9,+10,der(15)t(15)t(15;17)(p11;q11.2),-16,-17,-17,der(18)t(8;18)(q?22;q2?3),-20[6]
26706	HoTr	LOH-LCN	HoTr	HoTr	39,XY,add(2)(p13),-3,-4,-7,-8,i(9)(q10),-12,-13,-16,-17,+2mar[cp3]
27069	HoTr	LOH-LCN	HoTr	HoTr	34,XX,-3,-4,-5,-7,-8,-9,?inv(12)(q13q24),-13,-14,-15,-16,-17,-18,-20,+mar[8]
27121	HoTr	Inc	HoTr	HoTr	36,X,-X,-2,-3,-4,-7,-12,-13,-15,-16,-17[7]
27596	HoTr	LOH-LCN	HoTr	HoTr	38~39,XY,-3,-4,-5,-7,-9,-15,-16,inc[cp4]
29588	HoTr	LOH-LCN	HoTr	HoTr	34,X,-Y,-2,-3,-4,-5,-7,-9,-13,-15,-16,-17,-20[8]
923	HoTr	Inc	HoTr	HoTr	38,XY,-3,-4,add(4)(q21),add(5)(q13),-7,-9,add(10)(p),-13,-15,-16,-17,-20[23]
11669	HoTr	LOH-LCN	HoTr	HoTr	32,Y,-X,-2,-3,-4,-6,-7,-9,-12,-13,-16,-17,-19,-20,-21[1]/ 34,X,-Y,-2,-3,-4,-7, -9,-12,-13,-15,-16,-17,-20[1]

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
21894	HoTr	LOH-LCN	HoTr	HoTr	32~36,XY,-2,-3,-4,-7,-12,-13,-14,-15,-16,-17,-18,-20[cp8]
25296	HoTr	LOH-LCN	HoTr	HoTr	30~34,X,-Y,-2,-3,-4,-5,-7,-9,-11,-12,-13,-15,-16,-17,-20,-21,-22[cp11]/51~60,X,+X,-Y,+1,+6,+8,+8,+10,+14,+18,+18,+19,+21,inc[cp5]
26624	HoTr	LOH-LCN	HoTr	HoTr	35,X-Y,-2,-3,-4,-7,-9,-12,-13,-15,-16,-17,-20,+mar[cp5]/62~69,idemx2,+4,-5,-6,+9,-11,+12,-14,-18,-19,+21[cp11]
26902	HoTr	LOH-LCN	HoTr	HoTr	36,XX,-3,-4,-5,-7,-9,-13,-15,-16,-17,add(19)(p13),-20[cp4]/70~73,XX,+X,+X,+1,+1,+2,+3,add(3)(q2),+5,+6,+6,+8,+10,+10,+11,+11,+12,+12,+13,+14,+14,+18,+18,+19,+19,+21,+21,+22,+22[cp7]
28336	HoTr	LOH-LCN	HoTr	HoTr	34,X,-X,-3,-4,-5,-7,-8,-9,-14,-15,-16,-17,-20[4]/68~70,idemx2[cp2]
28430	HoTr	LOH-LCN	HoTr	HoTr	33~34,X,-X,-2,-3,-4,-5,-6,-7,-9,-13,-15,-16,-17,-20[cp9]/66,idemx2,+3,+3,+20,+20[1]
28508	HoTr	LOH-LCN	HoTr	HoTr	~36,XX,inc[4]/~70,XX,inc[5]
28644	HoTr	LOH-LCN	HoTr	HoTr	37,XX,-3,-4,-5,-7,-9,add(14)(q32),-15,-16,-17,-20[5]/68,XX,+1,+2,+3,+8,+8,+10,+11,+12,+13,+14,+14,+16,+18,+18,+19,+19,+20,+21,+21,+22,+22,+mar[cp3]
29407	HoTr	LOH-LCN	HoTr	HoTr	31~35,X,-X,-2,dic(8;13)(q10;q10),+19,-20[cp3]/57~67,XX,+1,+1,+2,+4,+6,+6,dic(8;13)(q10;q10)x2,+10,+11,+11,+i(11)(q10),+12,+14,+14,+18,+18,+19,+19,+21,+21,+22,+22,+1~6mar[cp11]
22435	HoTr	LOH-LCN	HoTr	HoTr	37,XX,-2,-3,-4,-7,-12,-13,-15,-16,-17[6]/73,XX,+X,+X,+1,+1,+5,+5,+6,+6,+8,+8,+9,+9,+10,+11,+11,+

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
					14,+14,+18,+18,+19,+19,+20,+20,+21,+21,+22,+22[8]
1142	HoTr	Inc	HoTr	HoTr	72,XX,+X,+1,+1,+2,+3,+4,+5,+8,+8,+10,+11,+12,+14,+14,t(14;14)(q32;q32),+15,+16,+17,+17,+18,+18,+19,+20,+21,+21,+22,+mar[12]
5954	HoTr	LOH-LCN	HoTr	HoTr	57-69,XY,+X,+Y,+1,+2,+6,+6,+8,+8,+11,+11,+12,+12,+14,+14,+18,+19,+19,+21,+22,+mar,+mar[cp10]
25437	HoTr	LOH-LCN	HoTr	HoTr	66~69,XX,+X,+2,der(1;3)(p10;q10)x2,+4,+5,+6,+6,+10,+11,+12,+12,+18,+18,+19,+20,+21,+21,+22,+22,+mar1,+mar2[cp8]
25614	HoTr	LOH-LCN	HoTr	HoTr	63~67,XX,+add(1)(p36),+add(1)(p12),+2,+6,+6,add(9)(q34),+add(10)(q26)x2,+11,+12,+12,+14,+14,+18,+18,+19,+22,+22,+1~9mar[cp7]
25950	HoTr	LOH-LCN	HoTr	HoTr	64~65,XX,+X,+1,+4,+5,+6,+6,+8,+8,+10,+10,+10,+11,+11,+12,+14,+18,+19,+19,+20,+21,+21,+22[cp7]
27400	HoTr	Inc	HoTr	HoTr	76,XX,+X,+1,+1,+2,+3,+4,+5,+7,+8,+8,+10,+11,+12,+14,+14,+15,+16,+i(17)(q10),+?18,+19,+20,+20,+21,+21,+22,+22,+3~6mar[cp2]
27537	HoTr	Inc	HoTr	HoTr	69,XX,+1,+1,+2,+4,+4,+5,+6,+6,+8,+10,+10,+11,+11,+12,+14,+14,+18,+18,+19,+21,+21,+22,+mar[3]
27555	HoTr	Inc	HoTr	HoTr	67,XY,+X,+Y,+1,+1,+2,+4,+5,+6,+?del(6)(q?25),+8,?add(9)(p21),+10,+11,+11,+12,+14,+18,+19,+21,+21,+22,+22,inc[1]
27873	HoTr	LOH-LCN	HoTr	HoTr	63~65,XY,+Y,+del(X)(q22q2?6),+1,+1,+2,+6,+6,+8,+10,+11,+12,+12,+18,+19,+19,+21,+21,+22,+22[cp7]
28056	HoTr	Inc	HoTr	HoTr	71~73,XX,+X,+1,+1,+2,+4,+5,+6,+8,+8,+10,+11,+12,+12,+13,+13,+14,+14,+15,?add(15)(p10),+16,+18,add(19)(q13.3)x2,

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
					+20,+21,+22,+~5~7mar,inc[cp 5]
28885	HoTr	LOH-LCN	HoTr	HoTr	62~66,XY,+X,+Y,+1,+del(1)(q ?32),+2,+2,+4,+4,+5,+6,+8,+1 0,+13,+14,+15,- 16,+?i(17)(q10),+18,+19,add(19)(p13),?+20,+21,+21,+22,+ 1~6mar[cp9]
28890	HoTr	LOH-LCN	HoTr	HoTr	59~66,XY,+X,+1,+2,+6,+6,+8, +9,+10,+10,+11,+11,+12,+13, +14,+16,+17,+18,+19,+19,+21 ,+21,+22[cp6]
28893	HoTr	Inc	HoTr	HoTr	75~80,XY,+X,+Y,+Y,+Y,+1,+1 ,+2,+2,+3,+4,+5,+5,+6,+7,+8,+ 9,+10,+11,+12,+13,+14,+14,+ 15,+15,+16,+16,+17,+17,+18, +18,+19,+19,+20,+20,+21,+21 ,+22,+22[cp4]
29434	HoTr	Inc	HoTr	Non-ploidy	66~70,XX,-X,- 4,?del(6)(p22),+?del(6)(p22),- 7,+8,-9,+11,+12,-13,-15,-16,- 17,+18,-20,+21[cp3]
3071	HoTr	LOH-LCN	HoTr	HoTr	62-64,X,- Y,+1,+1,+2,+4,+5,+6,+6,+8,+8 ,der(9;13)(q10;q10),+der(9;13) ,+10,+10,+11,+12, add(14)(p11),+add(14),+15,+1 8,+21,+21,+22,+22[cp15]/ 64,idem,add(21)[2]/ 72,XY,+X,+1,+2, +2,+4,+5,+5,+6,+6,+9,der(9;1 3)(q10;q10),+10,+10,+11,+11, +12,+12,add(12)(p),+13,+14,+ 14, add(14)(p11),+15,+16,+18,+2 1,+21,+22,+22[1]
27058	HoTr	HET-CNG	HeH	HeH	64~66,XX,+X,add(1)(p?2),+1, +3,+4,+5,+6,+8,+10,+11,+12,+ 14,+14,+17,+18,+19,+20,+21, +21,+22,+mar[cp9]/46,XX[1]
25434	HoTr	LOH-LCN	HoTr	HoTr	38~39,XY,-3,- 7,add(9)(p21),del(12)(q2?1),- 13,-15,- 16,der(16)t(3;16)(p12;q13),- 17,- 19[cp4]/75~76,idemx2[cp2]
27392	HoTr	LOH-LCN	HoTr	HoTr	46,XX[20]

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
26659	HoTr	LOH-LCN	HoTr	HoTr	Failed
28486	HoTr	LOH-LCN	HoTr	HoTr	76,XX,+X,+1,+1,+2,+2,+4,+4,+5,+6,+6,+8,+9,+10,+11,+11,+12,+13,+14,+16,+16,+18,+19,+20,+21,+21,+22,+22,inc[1]
31044	HoTr	LOH-LCN	HoTr	HoTr	Not done
M18/1850	HoTr	Inc	HoTr	HoTr	Failed
M18/3283	HoTr	LOH-LCN	HoTr	HoTr	Failed
M18/3510	HoTr	LOH-LCN	HoTr	HoTr	46,XY[20]
M18/3469	HoTr	LOH-LCN	HoTr	HoTr	Failed
M18/3777	HoTr	LOH-LCN	HoTr	HoTr	46,XX[20]
24805	HeH	Inc	HeH	Non-ploidy	53,XX,+5,+6,+10,+11,+20,+21,i(21)(q10),+22[8]
25852	HeH	HET-CNG	HeH	HeH	51,XY,+X,+4,+14,+21,+mar[12]
26081	HeH	HET-CNG	HeH	HeH	51~53,XX,+X,+X,+6,+14,+17,+21,+21[cp10]
26910	HeH	LOH-LCN	HoTr	HoTr	54~56,XY,+1,add(2)(q3)x2,+3,add(3)(q2),+5,+6,?del(6)(q?2),+10,+11,+14,+?16,+18,+2mar,inc[cp8]
27478	HeH	LOH-LCN	HoTr	HoTr	59,XX,+X,+1,+2,+4,+6,+10,+12,+18,+19,+21,+21,+22,+22[10]
27599	HeH	HET-CNG	HeH	HeH	56,XY,+X,+Y,+4,+6,+17,+21,inc[cp4]
27921	HeH	HET-CNG	HeH	HeH	52,XY,+X,+4,+14,+?18,+21,+21[3]
28195	HeH	HET-CNG	HeH	HeH	59~61,XY,+X,+Y,?dup(1)(q21q25),+4,?dic(4;12)(p?15,p?13),?+del(5)(q?21q35),+6,?+8,+9,+10,+10,+11,+14,?dic(17;20)(p13;q13),+18,+18,+20,+20,+21,+21,+1~4mar[cp10]

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
28233	HeH	Inc	HeH	Non-ploidy	56,XY,+X,+Y,+?5,+6,-8,+10,+18,+21,+21,+22,+22,+mar[3]/46,XY[7]
28894	HeH	HET-CNG	HeH	HeH	52,XY,+X,del(1)(p32p36),+4,+8,+17,+18,+i(21)(q10)[2]
29491	HeH	LOH-LCN	HoTr	HoTr	58~59,XY,+?X,+1,+2,+6,add(8)(q2)x2,+10,+11,+12,+12,+14,idic(15)(p1),+18,add(18)(p1),+19,+21,+21,+22,+mar,inc[cp10]
27847	HeH	HET-CNG	HeH	HeH	59~60,inc[4]/46,inc[6]
M17/2484	HeH	HET-CNG	HeH	HeH	53,X,+X,-Y,+6,+10,+14,+17,+18,+18,+21[2]
M18/3960	HeH	Inc	HeH	HeH	50,XX,+6,+14,+21,+21[10]
M18/766	HeH	HET-CNG	HeH	HeH	55,X,+X,-Y,+6,+10,+14,+17,+18,+18,+21,+21,+22[1]
M18/968	HeH	HET-CNG	HeH	Non-ploidy	53-54,XY,+X,dup(1)(q21q32),+4,+6,+10,+18,+21,+21,+mar[cp4]
M19/525	HeH	HET-CNG	HeH	HeH	53~56,XY,add(3)(q2?),+5,+7,+16,+20,+21,+21,+2~5mar[cp5]
26726	HeH	HET-CNG	HeH	HeH	Not done
27044	HeH	HET-CNG	HeH	Non-ploidy	Failed
27212	HeH	Inc	HeH	Non-ploidy	Failed
25654	HeH	HET-CNG	HeH	Non-ploidy	Failed
26690	HeH	HET-CNG	HeH	HeH	Failed
27194	HeH	HET-CNG	HeH	HeH	Failed
27214	HeH	HET-CNG	HeH	HeH	Failed

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
27971	HeH	HET-CNG	HeH	HeH	Not done
28003	HeH	HET-CNG	HeH	HeH	46,XX[20]
28472	HeH	HET-CNG	HeH	HeH	Failed
28668	HeH	HET-CNG	HeH	HeH	Failed
28983	HeH	HET-CNG	HeH	HeH	Failed
M18/1246	HeH	HET-CNG	HeH	HeH	Failed
M18/1732	HeH	HET-CNG	HeH	HeH	46,XX[20]
M18/3111	HeH	HET-CNG	HeH	HeH	Failed
M19/1212	HeH	HET-CNG	HeH	HeH	46,XX[20]
M19/1483	HeH	HET-CNG	HeH	HeH	Failed
M19/1512	HeH	HET-CNG	HeH	HeH	Not done
M19/1519	HeH	HET-CNG	HeH	Non-ploidy	Failed
M19/1636	HeH	HET-CNG	HeH	HeH	Not done
M19/185	HeH	Inc	HeH	HeH	Not done
M19/526	HeH	HET-CNG	HeH	HeH	Failed
M19/548	HeH	HET-CNG	HeH	HeH	Not done
28011	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,-2,add(7)(q3),add(12)(p11),+mar[3]
30036	B-other	Non-ploidy	Non-ploidy	Non-ploidy	47,XY,+5[4]

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
27752	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XX[20]
28149	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,dic(9;12)(p13;p12)[5]
25373	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,del(9)(q13q22)[4]
27181	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,inv(14)(q11q32)[2]
25451	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XY[20]
25130	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XX[20]
25246	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,der(19)t(1;19)(q12;p13.3)[2]/46,idem,t(5;18)(q33;q23)[8]
24890	B-other	Non-ploidy	Non-ploidy	Non-ploidy	Failed
27833	B-other	Non-ploidy	Non-ploidy	Non-ploidy	Failed
28168	B-other	Non-ploidy	Non-ploidy	Non-ploidy	45,XY,inv(2)(p1?5q1?3),dic(9;12)(p13;p13)[10]
27071	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XX[20]
26614	B-other	Non-ploidy	Non-ploidy	HoTr	Failed
27819	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XX[20]
28335	B-other	Non-ploidy	Non-ploidy	Non-ploidy	Failed
27391	B-other	Non-ploidy	Non-ploidy	Non-ploidy	Failed
27554	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,add(12)(q13)[3]
25967	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,+5,-6,dic(7;9)(p13;p11),add(14)(q32),del(17)(p11),+21[18]
25552	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XY[20]

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
30033	B-other	Non-ploidy	Non-ploidy	Non-ploidy	45,X,-X,del(8)(q1?q2?),-11,-14,add(16)(q2?),-17,del(20)(q1?),+3mar[6]
26621	B-other	Non-ploidy	Non-ploidy	Non-ploidy	Failed
27490	B-other	Non-ploidy	Non-ploidy	Non-ploidy	~92,inc[11]
25344	B-other	Non-ploidy	Non-ploidy	Non-ploidy	Failed
25371	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XX[20]
25267	B-other	Non-ploidy	Non-ploidy	Non-ploidy	45~47,XX,+1,dic(1;17)(p32;q25),inc[cp3]
26971	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,add(1)(q1)[5]
24309	B-other	Non-ploidy	Non-ploidy	Non-ploidy	45~48,XY,t(2;18)(p11;q21),-4,-5,der(9)t(1;9)(q2;q2),del(13)(q12q14),del(14)(q1q2),add(16)(q1),i(17)(q10),+21,+1~3mar[cp11]
26995	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XY[20]
27298	B-other	Non-ploidy	Non-ploidy	Non-ploidy	46,XY[20]
26611	B-other	Non-ploidy	Non-ploidy	Non-ploidy	47,XY,+X,del(16)(?q1),i(17)(q10)[8]
28934	*KMT2A*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,der(4)t(4;11)(q21;q23)ins(4;13)(q21;q1?q34),der(11)t(4;11)
25100	*KMT2A*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(1;11)(p32;q23)[11]
29908	*KMT2A*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(4;11)(q21;q23),inc[2]
27389	*KMT2A*	Non-ploidy	Non-ploidy	Non-ploidy	47,XXX,t(11;19)(q23;p13)[7]
28581	*KMT2A*	Non-ploidy	Non-ploidy	Non-ploidy	47,XX,t(4;11)(q21;q23),+6[11]

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
27584	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(9;22)(q34;q11)[7]/45,idem,-7[3]
26682	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	44,XX,-7,der(9;12)(q10;q10),t(9;22)(q34;q11.2),-13,+mar[8]
25548	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(9;22)(q34;q11)[10]
27509	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	45,XY,der(7;9)(q10;q10)t(9;22)(q34;q11.2),der(22)t(9;22)(q34;q11.2)[11]
24919	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(9;22)(q34;q11)[8]
27043	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(2;9)(p21;p23),add(6)(q21),t(9;22)(q34;q11),add(21)(q21)[8]
26610	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,der(9)del(q)(p1)t(9;22)(q34;q11),der(22)t(9;22)(q34;q11)[13]
25953	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	Failed
25208	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,-9,t(9;22)(q34;q11),add(10)(q2?2),+der(22)t(9;22)(q34;q11)[2]
27333	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(9;22)(q34;q11.2)[13]/47,XX,der(9)t(9;22)(q34;q11.2)add(9)(p13),der(22)t(9;22)(q34;q11.2),+der(22)t(9;22)[3]
30035	*BCR-ABL1*	Non-ploidy	Non-ploidy	HeH	46,XY[20]
27836	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	45,XX,-7,t(9;22)(q34;q11.2),t(12;21)(p13;q22)[5]/46,XY,idem,+der(22)t(9;22)(q34;q11.2)[5]
29589	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	47,XY,+2,t(9;22)(q34;q11.2)[4]/48,XY,+2,t(9;22)(q34;q11.2),+der(22)t(9;22)(q34;q11.2)[9]
28196	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,add(9)(q34),der(17)t(17;22)(q21;q11),der(22)t(9;22)(q34;q11)[4]/46,XY,add(4)(q31),add(5)(q31),add(9)(q34),der(17)t(7;22)(q21;q11),der(22)t(9;22)(q34;q11)[6]
26660	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(9;22)(q34;q11.2)[8]

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
25247	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	47,XY,t(9;22)(q34;q11),+10[10]
28182	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(9;22)(q34.1;q11)[2]/46,idem,der(5)t(1;5)(q23;q21)[7]
27754	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	Failed
28670	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	45,XX,t(2;7)(p1;p1),der(3)t(3;5)(q13;q15),-5,der(9)add(9)(p1)t(9;22)(q34;q11),add(11)(q15),add(12)(p1),der(22)t(9;22)(q34;q11)[4]
27026	*BCR-ABL1*	HET-CNG	HeH	HoTr	67~73,XX,+X,t(9;22)(q34;q11),+ider(22)t(9;22)x2,inc[cp5]
28057	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(9;22)(q34;q11),inc[3]/46,XX,inc[2]
28350	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	45,XX,-7,ins(9;?)(q13;?),t(9;22)(q34;q11.2)[9]
25082	*BCR-ABL1*	HET-CNG	HeH	HeH	53,XX,+X,+2,+6,t(9;22)(q34;q11.2),+14,+18,+21,+der(22)t(9;22)[7]/54,idem,+der(22)t(9;22)[3]
26609	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	45,XY,-7,t(9;22;11)(q34;q11;q13)[19]
26062	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,t(9;22)(q34;q11)[5]/48~51,XY,+X,+6,+8,t(9;22)(q34;q11),+16,+der(22)t(9;22)[cp12]
25346	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,t(9;22)(q34;q11)[9]
25793	*BCR-ABL1*	Non-ploidy	Non-ploidy	HeH	46,XX,der(9)t(9;22)(q34;q11)t(9;21)(q34;q22.3),der(21)t(9;22)t(9;21),der(22)t(9;22)[8]
27085	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(9;22)(q34;q11.2),der(19)t(8;19)(q13;p13.3)[8]
27583	*BCR-ABL1*	HET-CNG	HeH	HeH	51,XX,+X,+4,t(9;22)(q34;q11.2),+14,+17,+der(22)t(9;22)[10]
30032	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,t(9;22)(q34;q11)[22]
25893	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	47,XX,+X,?add(3)(q21),i(9)(q10)t(9;22)(q34;q11)[5]

Patient ID	Cyto	SNP array	CART input	CART output	Cytogenetics analysis at diagnosis
27585	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(9;22)(q34;q11)[11]
29089	*BCR-ABL1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XX,t(9;22)(q34;q11.2)[9]
27812	*TCF3-PBX1*	Non-ploidy	Non-ploidy	Non-ploidy	46,XY,der(19)t(1;19)(q23;p13)[10]
30034	T-ALL	Non-ploidy	Non-ploidy	Non-ploidy	Failed
27642	T-ALL	Non-ploidy	Non-ploidy	Non-ploidy	Failed

Supplementary table 7

Demographic and clinical details of all cases included in the CART analysis of ploidy status in ALL.

Patient ID	Sex	Age	% BM Blasts
4949	Female	20	99
4950	Male	51	98
24401	Female	54	54
26076	Male	46	60
26082	Female	48	80
26706	Male	60	89
27069	Female	46	96
27121	Female	64	Not known
27596	Male	62	64
29588	Male	52	66
923	Male	15	74
11669	Male	16	100
21894	Male	12	50
25296	Male	28	94
26624	Male	50	39
26902	Female	43	60
28336	Female	58	Not known
28430	Female	42	37
28508	Female	59	57
28644	Female	64	95
29407	Female	60	94
22435	Female	12	100
1142	Female	20	96
5954	Male	50	95
25437	Female	64	88
25614	Female	56	75
25950	Female	40	73
27400	Female	46	89
27537	Female	67	87
27555	Male	64	Not known
27873	Male	39	85
28056	Female	55	90
28885	Male	59	27
28890	Female	49	Not known
28893	Male	27	95
29434	Female	49	97

Patient ID	Sex	Age	% BM Blasts
3071	Male	15	Not known
27058	Female	7	Not known
25434	Male	54	27
27392	Female	73	72
26659	Male	60	0
28486	Female	55	59
31044	Female	63	Not known
M18/1850	Male	87	Not known
M18/3283	Female	78	Not known
M18/3510	Male	77	Not known
M18/3469	Male	31	Not known
M18/3777	Female	8	Not known
24805	Female	46	39
25852	Male	40	95
26081	Female	28	85
26910	Male	43	68
27478	Female	58	88
27599	Male	29	82
27921	Male	48	18
28195	Male	30	89
28233	Male	58	56
28894	Male	41	100
29491	Male	51	90
27847	Female	3	Not known
M17/2484	Male	4	Not known
M18/3960	Female	2	Not known
M18/766	Male	3	Not known
M18/968	Male	6	Not known
M19/525	Male	8	Not known
26726	Male	66	88
27044	Female	54	95
27212	Male	33	97
25654	Male	8	Not known
26690	Female	3	Not known
27194	Female	5	Not known
27214	Female	2	Not known
27971	Female	4	Not known
28003	Female	2	Not known
28472	Male	12	Not known
28668	Male	7	Not known

Patient ID	Sex	Age	% BM Blasts
28983	Male	3	Not known
M18/1246	Female	7	Not known
M18/1732	Female	3	Not known
M18/3111	Male	3	Not known
M19/1212	Male	7	Not known
M19/1483	Female	3	Not known
M19/1512	Male	1	Not known
M19/1519	Female	4	Not known
M19/1636	Male	16	Not known
M19/185	Female	3	Not known
M19/526	Female	2	Not known
M19/548	Male	2	Not known
28011	Male	61	Not known
30036	Male	70	Not known
27752	Female	73	48
28149	Female	64	80
25373	Male	65	32
27181	Male	65	40
25451	Male	63	88
25130	Female	62	100
25246	Male	64	42
24890	Male	65	100
27833	Female	73	98
28168	Male	65	Not known
27071	Female	60	90
26614	Male	75	98
27819	Female	65	90
28335	Male	63	Not known
27391	Male	65	Not known
27554	Female	78	75
25967	Male	60	95
25552	Male	61	80
30033	Female		Not known
26621	Female	69	90
27490	Male	64	40
25344	Female	61	80
25371	Female	60	50
25267	Female	63	1
26971	Male	67	50
24309	Male	63	80

Patient ID	Sex	Age	% BM Blasts
26995	Male	70	Not known
27298	Male	66	90
26611	Male	68	Not know
28934	Female	60	9
25100	Female	63	79
29908	Female	62	Not known
27389	Female	73	Not known
28581	Female	65	88
27584	Female	76	90
26682	Female	63	35
25548	Female	60	16
27509	Male	69	60
24919	Female	64	95
27043	Female	65	90
26610	Male	65	63
25953	Female	71	90
25208	Male	62	66
27333	Female	63	73
30035	Male	Not known	Not known
27836	Male	63	61
29589	Male	61	82
28196	Male	61	99
26660	Female	62	84
25247	Male	64	95
28182	Female	60	90
27754	Female	63	Not known
28670	Female	61	94
27026	Female	63	Not known
28057	Female	64	Not known
28350	Female	62	72
25082	Female	62	95
26609	Male	83	50
26062	Male	61	84
25346	Male	64	40
25793	Female	67	100
27085	Female	63	92
27583	Female	61	95
30032	Male	Not known	Not known
25893	Female	78	92
27585	Female	66	62

Patient ID	Sex	Age	% BM Blasts
29089	Female	60	90
27812	Male	65	99
30034	Female	Not known	Not known
27642	Female	72	90

Supplementary table 8

Standardised whole chromosomal log2 ratios of all cases including in CART analysis of ploidy status (values in table rounded to 2 decimal places). HeH: High hyperdiploidy, HoTr: Low hypodiploidy/near triploidy, NP: non-ploidy

Case	27392	27121	25437	26659	27555	27596	29407	27537	28644
Status	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr
chr1	0.14	0.01	0.14	0.06	0.06	0.12	0.14	0.03	0.07
chr2	0.15	-0.02	0.05	-0.03	0.02	0.14	0.05	0.02	0.08
chr3	-0.23	-0.01	-0.18	-0.03	-0.01	-0.16	-0.11	-0.06	-0.14
chr4	-0.12	0.00	0.06	-0.02	0.04	-0.15	-0.13	0.05	-0.14
chr5	-0.24	0.03	0.05	0.02	0.03	-0.15	-0.12	0.03	-0.14
chr6	0.17	0.03	0.17	0.06	0.08	0.15	0.15	0.05	0.19
chr7	-0.23	-0.01	-0.18	-0.03	-0.01	-0.16	-0.12	-0.05	-0.14
chr8	0.10	0.03	-0.18	0.07	0.04	0.16	0.03	0.05	0.08
chr9	-0.24	0.02	-0.18	0.01	-0.02	-0.16	-0.12	0.01	-0.14
chr10	0.15	0.02	0.05	0.02	0.02	0.13	0.07	0.03	0.08
chr11	0.15	0.02	0.05	0.01	0.06	0.13	0.14	0.03	0.08
chr12	0.14	-0.02	0.17	-0.04	0.02	0.13	0.05	-0.02	0.07
chr13	-0.23	-0.01	-0.19	-0.02	-0.01	0.14	0.13	-0.03	0.15
chr14	0.12	0.02	-0.18	0.01	0.02	0.13	0.14	0.03	0.08
chr15	-0.24	-0.02	-0.18	-0.04	-0.02	-0.17	-0.12	-0.06	-0.14
chr16	-0.24	-0.02	-0.15	-0.03	-0.01	-0.17	-0.11	-0.06	-0.10
chr17	-0.24	-0.03	-0.16	-0.03	-0.02	-0.17	-0.10	-0.06	-0.11
chr18	0.09	0.03	0.18	0.07	0.03	0.15	0.13	0.04	0.19
chr19	0.14	0.01	0.06	0.05	0.02	0.11	0.08	-0.04	0.15
chr20	0.15	0.02	0.05	0.01	-0.01	0.14	-0.11	-0.05	-0.13
chr21	0.17	0.03	0.19	0.07	0.07	0.15	0.16	0.04	0.19
chr22	0.17	0.03	0.05	0.04	0.08	0.16	0.13	0.08	0.06

25296	29588	24401	28893	25950	25434	28486	27069	28430	26082	26706	**Case**
HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	**Status**
0.16	0.15	0.02	0.02	0.04	0.07	0.09	0.15	0.13	0.02	0.09	**chr1**
-0.06	-0.09	-0.03	-0.01	-0.11	0.07	0.01	0.16	-0.09	0.02	0.09	**chr2**
-0.06	-0.09	-0.03	-0.01	-0.10	-0.11	-0.11	-0.11	-0.09	-0.07	-0.18	**chr3**
-0.06	-0.09	0.02	0.01	0.05	0.06	0.02	-0.10	-0.09	0.03	-0.18	**chr4**
-0.06	-0.09	0.02	0.04	0.04	0.07	0.01	-0.10	-0.09	0.02	0.11	**chr5**
0.18	0.16	0.02	-0.01	0.08	0.01	0.15	0.17	0.13	0.02	0.10	**chr6**
-0.05	-0.08	-0.03	-0.01	-0.10	-0.16	-0.11	-0.10	-0.09	-0.07	-0.18	**chr7**
0.21	0.16	0.03	0.00	0.10	0.04	0.06	-0.10	0.14	-0.01	0.04	**chr8**
-0.06	-0.09	0.02	-0.01	-0.11	-0.01	-0.04	-0.10	-0.09	-0.07	0.10	**chr9**
0.17	0.15	0.02	0.03	0.20	0.07	0.09	0.16	0.13	0.10	0.09	**chr10**
-0.05	0.15	0.02	-0.01	0.13	0.07	0.07	0.16	0.12	0.06	0.17	**chr11**
-0.06	0.15	-0.04	-0.02	0.04	0.04	0.00	0.13	0.12	0.02	-0.19	**chr12**
-0.06	-0.09	-0.02	0.00	-0.11	-0.18	-0.11	-0.10	-0.09	0.03	-0.18	**chr13**
0.16	0.15	0.01	0.03	0.04	0.07	0.00	0.05	0.13	0.02	0.09	**chr14**
-0.06	-0.09	-0.03	-0.02	-0.11	-0.15	-0.11	-0.11	-0.09	0.01	0.08	**chr15**
-0.05	-0.09	-0.03	0.02	-0.09	-0.14	-0.05	-0.10	-0.09	-0.06	-0.17	**chr16**
-0.06	-0.08	-0.04	0.03	-0.08	-0.14	-0.10	-0.10	-0.09	-0.01	-0.17	**chr17**
0.21	0.16	0.02	0.04	0.04	0.06	0.01	-0.11	0.14	-0.05	0.11	**chr18**
0.14	0.12	0.01	0.04	0.07	-0.12	0.01	0.13	0.10	0.02	0.06	**chr19**
-0.05	-0.09	-0.03	0.02	-0.08	0.07	0.00	-0.10	-0.09	-0.06	0.07	**chr20**
0.19	0.16	0.02	0.04	0.14	0.07	0.16	0.18	0.14	0.02	0.11	**chr21**
-0.05	0.11	0.01	-0.01	-0.03	0.05	0.07	0.12	0.10	0.01	0.06	**chr22**

28885	27873	28508	25614	28056	27400	26076	28890	28336	29434	26624	**Case**
HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	**Status**
0.11	0.14	0.04	0.09	0.01	0.09	0.08	0.05	0.16	0.00	0.09	**chr1**
0.04	0.10	-0.05	0.06	0.01	-0.02	-0.09	0.02	0.13	0.00	-0.09	**chr2**
-0.13	-0.12	-0.05	-0.09	-0.03	-0.02	-0.09	-0.06	-0.09	0.00	-0.10	**chr3**
0.17	-0.11	-0.05	-0.03	0.02	-0.01	-0.08	-0.05	-0.05	0.00	-0.05	**chr4**
0.05	-0.12	0.05	-0.10	0.00	-0.02	0.10	-0.06	-0.19	0.00	0.05	**chr5**
0.05	0.16	0.04	0.15	0.00	0.00	0.09	0.12	0.13	0.00	0.13	**chr6**
-0.13	-0.13	-0.04	-0.10	-0.02	-0.03	-0.08	-0.06	-0.18	0.00	-0.07	**chr7**
0.08	0.02	0.05	-0.08	0.02	0.09	0.09	0.06	-0.05	0.01	0.10	**chr8**
-0.12	-0.13	0.03	-0.11	-0.01	-0.16	0.09	-0.06	-0.18	-0.01	0.02	**chr9**
0.03	0.15	0.05	0.15	0.02	-0.01	0.08	0.09	0.07	0.00	0.08	**chr10**
-0.04	0.10	0.05	0.09	0.01	-0.05	0.08	0.10	0.13	0.00	0.09	**chr11**
-0.13	0.15	-0.05	0.14	0.01	-0.03	-0.09	0.02	0.13	0.00	-0.05	**chr12**
-0.05	0.05	-0.04	-0.09	0.02	-0.01	-0.08	-0.04	0.03	0.00	-0.06	**chr13**
0.05	-0.04	0.02	0.14	0.01	-0.02	0.08	0.10	-0.06	0.00	0.08	**chr14**
0.04	-0.12	-0.06	-0.07	-0.03	0.04	-0.09	-0.07	-0.07	-0.01	-0.09	**chr15**
0.02	-0.12	-0.04	-0.09	-0.02	-0.03	-0.08	-0.06	-0.17	-0.01	-0.06	**chr16**
-0.07	-0.12	-0.05	-0.06	-0.02	0.00	-0.08	-0.06	-0.02	-0.01	-0.10	**chr17**
0.05	0.05	0.05	0.08	0.00	0.01	0.10	0.01	0.13	0.00	0.10	**chr18**
-0.02	0.13	0.04	0.04	-0.01	0.01	0.07	0.09	0.10	0.01	0.05	**chr19**
-0.10	-0.11	0.04	-0.09	-0.02	0.07	0.07	-0.06	-0.16	-0.01	-0.08	**chr20**
0.17	0.18	0.06	0.17	0.03	0.09	0.09	0.11	0.19	0.01	0.14	**chr21**
0.05	0.04	0.04	0.12	0.02	0.05	0.06	0.07	0.10	0.01	0.10	**chr22**

27478	27921	28894	26910	26081	25852	24805	28233	27599	28195	26902	**Case**
HoTr	HeH	HeH	HoTr	HeH	HeH	HeH	HeH	HeH	HeH	HoTr	**Status**
0.09	-0.01	-0.02	0.06	-0.02	-0.02	0.00	0.00	-0.04	-0.02	0.10	**chr1**
0.10	-0.01	-0.02	0.06	-0.02	-0.02	0.00	0.01	-0.04	-0.07	0.09	**chr2**
-0.11	0.00	-0.02	-0.07	-0.02	-0.02	0.00	0.01	-0.04	-0.07	-0.09	**chr3**
0.11	0.08	0.17	-0.06	-0.01	0.19	0.00	0.02	0.15	0.14	-0.17	**chr4**
-0.10	-0.01	-0.01	0.02	-0.01	-0.01	0.00	0.01	-0.04	0.07	-0.17	**chr5**
0.11	-0.01	-0.02	0.07	0.08	-0.02	0.00	0.00	0.13	0.12	0.11	**chr6**
-0.10	-0.01	-0.01	-0.07	-0.01	-0.01	0.00	0.01	-0.04	-0.05	-0.16	**chr7**
-0.10	-0.01	-0.01	0.08	-0.01	-0.01	0.01	0.01	-0.04	-0.06	0.11	**chr8**
-0.11	-0.01	-0.01	-0.09	-0.01	-0.01	0.00	0.01	-0.04	-0.06	-0.16	**chr9**
0.09	-0.01	-0.02	0.06	-0.02	0.03	0.00	0.01	0.13	0.24	0.11	**chr10**
-0.04	-0.01	-0.02	0.07	-0.02	-0.01	-0.01	0.00	-0.04	0.13	0.11	**chr11**
0.09	-0.01	-0.02	0.06	-0.02	-0.02	-0.01	0.00	-0.05	-0.10	0.10	**chr12**
-0.10	0.00	-0.01	-0.06	-0.01	-0.01	0.00	0.01	-0.05	-0.07	-0.12	**chr13**
-0.11	0.07	0.14	-0.01	0.07	0.16	0.00	0.00	0.13	0.12	0.11	**chr14**
-0.11	-0.01	-0.02	-0.07	-0.03	-0.02	0.00	0.00	-0.04	-0.07	-0.17	**chr15**
-0.10	-0.01	-0.01	-0.07	-0.02	-0.01	0.00	0.01	-0.03	-0.04	-0.15	**chr16**
-0.10	0.01	0.12	-0.08	0.05	0.02	-0.02	0.00	0.12	-0.04	-0.15	**chr17**
0.11	0.08	0.05	0.07	-0.01	-0.01	0.01	0.01	0.13	0.24	0.12	**chr18**
0.10	0.00	0.00	0.04	-0.01	0.00	-0.02	0.01	-0.02	-0.02	0.10	**chr19**
-0.11	-0.01	-0.02	-0.07	-0.02	-0.01	0.00	0.00	-0.03	-0.05	-0.15	**chr20**
0.22	0.13	0.28	0.16	0.15	0.18	0.01	0.01	0.25	0.33	0.12	**chr21**
0.14	-0.02	-0.02	0.09	-0.02	-0.01	-0.01	0.01	-0.02	-0.03	0.09	**chr22**

27212	27044	4949	923	22435	21894	5954	3071	1142	4950	29491	**Case**
HeH	HeH	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	HoTr	**Status**
0.01	-0.08	0.12	0.01	0.14	0.08	0.05	0.15	0.10	0.01	0.09	**chr1**
-0.01	-0.08	0.02	0.00	-0.18	-0.06	0.05	0.06	0.00	0.01	0.04	**chr2**
-0.01	-0.08	-0.21	0.01	-0.18	-0.06	-0.15	-0.24	0.01	0.02	-0.08	**chr3**
0.00	-0.08	-0.22	-0.01	-0.18	-0.06	-0.14	0.04	0.01	0.01	-0.11	**chr4**
0.02	0.16	-0.21	0.00	0.14	0.09	-0.16	0.02	0.01	0.01	-0.03	**chr5**
0.09	0.15	0.13	0.00	0.15	0.08	0.16	0.15	-0.16	0.01	0.09	**chr6**
-0.01	-0.07	0.01	-0.01	-0.18	-0.06	-0.13	-0.25	-0.16	0.01	-0.06	**chr7**
0.00	0.16	0.13	0.01	0.16	0.09	0.12	0.14	0.11	0.02	0.04	**chr8**
-0.01	-0.07	-0.21	-0.01	0.09	0.08	-0.17	-0.24	-0.15	0.01	-0.13	**chr9**
0.07	0.14	0.12	0.00	0.10	0.08	0.12	0.16	0.00	0.02	0.08	**chr10**
-0.01	-0.07	0.13	0.02	0.12	0.08	0.13	0.03	0.01	0.02	0.08	**chr11**
-0.02	-0.08	0.11	0.00	-0.19	-0.07	0.14	0.07	-0.01	0.01	0.17	**chr12**
-0.01	-0.08	-0.23	-0.02	-0.19	-0.06	-0.15	0.01	-0.17	0.01	-0.12	**chr13**
-0.01	-0.07	0.12	0.01	0.12	-0.06	0.15	0.04	0.11	0.01	0.09	**chr14**
-0.02	0.12	-0.20	0.02	-0.18	-0.07	-0.15	0.05	0.00	0.02	0.02	**chr15**
-0.02	-0.06	-0.18	0.01	-0.17	-0.07	-0.15	-0.21	-0.01	0.01	-0.12	**chr16**
-0.02	0.11	-0.20	-0.02	-0.20	-0.08	-0.17	-0.17	0.06	-0.02	-0.11	**chr17**
-0.01	0.16	0.14	0.01	0.14	0.09	0.15	0.00	0.11	0.03	0.09	**chr18**
-0.01	-0.08	0.10	0.00	0.06	0.05	0.12	-0.19	-0.04	-0.02	0.04	**chr19**
-0.01	-0.07	-0.18	0.02	0.10	-0.06	-0.14	-0.08	0.01	0.02	-0.12	**chr20**
0.12	0.29	0.13	-0.02	0.13	0.09	0.14	0.15	0.11	-0.01	0.20	**chr21**
0.03	0.00	-0.16	-0.02	0.06	0.03	0.03	0.11	-0.03	-0.03	0.02	**chr22**

27194	27058	25654	27847	28668	28003	27971	28472	26690	27214	28983	**Case**
HeH	HeH	HeH	HeH	HeH	HeH	HeH	HeH	HeH	HeH	HeH	**Status**
-0.07	-0.06	-0.01	-0.07	-0.03	-0.02	-0.07	0.02	-0.04	-0.03	-0.05	**chr1**
-0.13	-0.03	-0.03	-0.07	-0.04	-0.02	-0.06	-0.06	-0.09	-0.05	-0.05	**chr2**
-0.13	0.05	-0.01	-0.04	-0.04	-0.02	-0.06	-0.06	-0.09	-0.06	-0.05	**chr3**
0.10	0.04	0.13	0.07	-0.03	-0.02	0.16	0.16	0.14	0.16	0.17	**chr4**
0.08	0.05	0.12	0.06	0.16	-0.02	-0.06	0.04	-0.09	-0.05	-0.05	**chr5**
0.08	0.04	-0.01	0.06	0.15	0.14	0.10	0.13	0.12	0.14	0.15	**chr6**
0.08	-0.18	-0.01	-0.06	-0.04	-0.02	-0.06	-0.15	0.12	-0.05	-0.05	**chr7**
0.08	0.05	-0.05	0.06	-0.03	-0.02	0.15	0.15	0.13	-0.05	0.16	**chr8**
-0.12	-0.18	-0.01	-0.07	-0.04	-0.02	-0.06	-0.08	-0.09	-0.05	-0.05	**chr9**
0.03	0.05	-0.01	0.06	0.15	-0.02	0.15	0.13	0.12	0.14	0.15	**chr10**
0.08	0.05	-0.01	0.06	-0.04	-0.02	-0.06	-0.04	-0.09	-0.05	-0.04	**chr11**
0.07	0.04	-0.02	0.05	-0.05	-0.03	-0.07	-0.06	-0.09	-0.06	-0.05	**chr12**
-0.13	-0.19	-0.02	-0.08	-0.04	-0.03	-0.06	-0.06	-0.09	-0.05	-0.05	**chr13**
0.08	0.17	-0.01	0.08	0.15	0.13	0.14	0.13	0.24	0.14	0.15	**chr14**
-0.12	-0.17	-0.01	0.04	-0.04	-0.03	-0.06	-0.06	-0.08	-0.05	-0.05	**chr15**
-0.10	0.00	0.00	-0.06	-0.03	-0.02	-0.05	-0.05	-0.07	-0.04	-0.04	**chr16**
0.08	0.05	-0.01	0.05	0.13	-0.03	0.12	0.12	0.10	0.13	0.14	**chr17**
0.08	0.05	0.12	0.06	-0.03	0.15	0.15	-0.04	0.13	0.15	-0.05	**chr18**
-0.06	-0.13	0.00	-0.05	-0.02	-0.02	-0.05	-0.04	-0.06	-0.02	-0.03	**chr19**
-0.11	-0.15	-0.01	-0.06	-0.38	-0.02	-0.05	-0.05	-0.07	-0.04	-0.03	**chr20**
0.21	0.27	0.22	0.15	0.35	0.25	0.28	0.26	0.26	0.27	0.28	**chr21**
0.08	0.06	0.00	-0.05	0.13	-0.01	0.12	-0.04	0.10	-0.03	-0.02	**chr22**

M18/766	M18/968	M18/1246	M18/1732	M18/1850	M18/3283	M18/3510	M18/3469	M18/3777	M18/3960	M19/185	**Case**
HeH	HeH	HeH	HeH	HoTr	HoTr	HoTr	HoTr	HoTr	HeH	HeH	**Status**
-0.02	-0.01	-0.07	-0.06	0.03	0.06	0.15	0.10	0.06	-0.01	-0.01	**chr1**
-0.03	-0.04	-0.12	-0.06	0.02	0.00	-0.14	-0.12	-0.15	-0.01	-0.01	**chr2**
-0.02	-0.04	0.09	-0.07	0.00	-0.09	-0.14	-0.12	-0.15	-0.01	-0.01	**chr3**
-0.02	0.16	-0.11	0.15	0.00	-0.09	-0.16	-0.11	0.08	0.00	0.04	**chr4**
-0.03	-0.04	0.09	-0.06	0.03	0.06	0.17	0.08	0.07	0.00	-0.01	**chr5**
0.11	0.14	0.09	0.13	0.02	0.06	0.16	0.14	0.06	0.03	0.03	**chr6**
-0.03	-0.04	-0.11	-0.06	-0.03	-0.08	-0.14	-0.12	-0.14	0.00	0.00	**chr7**
-0.02	-0.04	0.10	0.13	0.00	0.06	0.17	0.15	0.07	0.00	-0.01	**chr8**
-0.02	0.06	0.03	-0.06	-0.03	0.06	0.14	0.02	0.07	0.00	-0.01	**chr9**
0.10	0.14	0.09	0.13	0.01	0.06	0.16	0.09	0.07	0.00	-0.01	**chr10**
-0.03	-0.04	0.08	-0.06	0.03	0.06	0.15	0.12	0.06	0.00	-0.01	**chr11**
-0.03	-0.05	-0.11	-0.07	0.02	-0.09	-0.14	-0.12	-0.15	-0.01	-0.01	**chr12**
-0.02	-0.04	-0.11	-0.07	-0.01	-0.09	-0.15	-0.12	0.07	0.00	0.00	**chr13**
0.10	-0.05	0.08	0.13	-0.02	0.06	0.16	0.03	0.06	0.03	0.02	**chr14**
-0.03	-0.04	-0.11	-0.06	-0.03	-0.09	-0.13	0.12	-0.14	-0.01	-0.02	**chr15**
-0.03	-0.03	0.08	-0.04	-0.03	-0.08	-0.13	-0.12	-0.12	0.00	-0.01	**chr16**
0.08	0.08	0.07	0.13	-0.03	-0.08	-0.14	-0.13	-0.12	-0.01	0.02	**chr17**
0.21	0.15	0.09	0.13	0.03	0.07	0.17	0.11	0.06	0.00	0.02	**chr18**
-0.02	-0.01	-0.09	-0.01	0.02	0.06	0.09	0.06	0.08	0.01	0.01	**chr19**
-0.03	-0.03	-0.10	-0.05	0.00	-0.01	-0.12	0.10	0.06	0.00	-0.01	**chr20**
0.21	0.26	0.22	0.26	0.03	0.07	0.16	0.14	0.08	0.04	0.06	**chr21**
0.08	-0.02	-0.09	-0.03	0.00	0.05	0.10	0.05	0.07	0.02	-0.01	**chr22**

26614	M18/3111	M19/526	M19/525	M19/548	M19/1212	M19/1483	M19/1512	M19/1519	M19/1636	M17/2484	**Case**
NP	HeH	HeH	HeH	HeH	HeH	HeH	HeH	HeH	HeH	HeH	**Status**
0.04	-0.02	0.02	-0.02	-0.05	-0.02	-0.03	-0.03	-0.02	-0.04	-0.03	**chr1**
0.04	-0.04	-0.05	-0.02	-0.05	-0.02	-0.03	-0.03	-0.03	-0.04	-0.03	**chr2**
0.04	-0.05	-0.05	-0.02	-0.05	-0.02	-0.03	-0.03	-0.04	-0.04	-0.03	**chr3**
0.05	0.12	0.12	0.14	0.15	0.19	0.08	0.15	0.05	0.17	-0.02	**chr4**
0.04	0.04	-0.05	-0.02	-0.05	-0.02	-0.03	-0.02	0.04	-0.04	-0.03	**chr5**
0.04	0.02	0.11	0.05	0.11	-0.01	0.06	0.13	0.03	0.14	0.14	**chr6**
-0.13	-0.05	-0.04	0.00	-0.04	-0.01	-0.02	-0.02	-0.03	-0.03	-0.03	**chr7**
0.05	0.04	-0.04	-0.08	0.13	-0.02	-0.03	-0.03	-0.03	-0.04	-0.03	**chr8**
0.00	-0.06	-0.04	-0.01	-0.05	-0.02	-0.02	-0.02	-0.03	-0.03	-0.03	**chr9**
0.05	0.04	0.09	-0.01	0.12	-0.01	0.06	0.12	0.03	0.15	0.14	**chr10**
0.04	0.03	-0.04	-0.01	-0.05	-0.02	-0.03	-0.02	0.03	-0.03	-0.03	**chr11**
-0.04	0.03	-0.05	-0.02	-0.05	-0.02	-0.03	-0.03	0.02	-0.04	-0.03	**chr12**
0.05	-0.05	-0.04	-0.01	-0.04	-0.01	-0.02	-0.02	-0.02	-0.04	-0.03	**chr13**
0.04	0.03	0.10	0.12	0.12	0.16	0.05	0.12	0.03	0.14	0.16	**chr14**
0.03	-0.06	-0.05	-0.02	-0.06	-0.03	-0.04	-0.04	-0.04	-0.05	-0.04	**chr15**
0.00	0.03	-0.03	-0.01	-0.04	-0.02	-0.03	-0.02	-0.03	-0.03	-0.02	**chr16**
-0.10	0.04	-0.03	-0.01	0.13	0.18	0.05	-0.01	0.03	0.13	0.12	**chr17**
0.06	0.11	0.11	0.12	0.12	-0.02	0.07	-0.03	0.04	0.15	0.25	**chr18**
0.00	-0.02	-0.01	0.01	0.00	0.03	-0.01	0.02	-0.01	0.00	-0.01	**chr19**
0.05	-0.04	-0.04	-0.02	-0.05	-0.02	-0.03	-0.03	-0.04	-0.03	-0.03	**chr20**
0.05	0.11	0.21	0.22	0.22	0.29	0.07	0.30	0.22	0.26	0.17	**chr21**
0.04	0.00	-0.02	-0.01	-0.03	0.00	-0.03	-0.01	-0.03	-0.03	-0.01	**chr22**

30033	25953	28149	26611	25208	25548	27490	28168	27584	27043	28934	**Case**
NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	**Status**
-0.01	0.01	0.00	0.02	0.08	0.00	0.03	0.00	0.00	0.00	0.00	**chr1**
0.01	0.02	0.00	0.02	0.01	0.00	0.02	0.00	0.00	0.00	0.00	**chr2**
0.01	0.02	0.00	0.02	0.01	0.00	0.05	0.01	0.01	0.01	0.00	**chr3**
0.02	0.03	0.01	0.03	0.02	0.01	0.04	0.02	0.01	0.01	0.01	**chr4**
0.01	0.02	0.01	0.02	0.01	0.00	0.01	0.01	0.01	0.01	0.00	**chr5**
0.01	0.02	0.00	0.02	0.00	0.00	0.05	0.00	0.02	-0.02	0.00	**chr6**
0.01	-0.39	0.01	0.02	0.01	0.00	0.04	0.01	0.00	0.01	0.00	**chr7**
0.01	0.03	0.01	0.02	0.01	0.01	0.05	0.04	0.02	0.01	0.00	**chr8**
0.00	0.02	-0.07	0.02	-0.32	0.00	0.04	-0.08	0.00	-0.06	0.00	**chr9**
0.00	0.02	0.00	0.03	0.01	0.00	0.04	0.01	0.00	0.01	0.00	**chr10**
0.00	0.02	0.00	0.02	0.00	0.00	-0.03	0.01	0.00	0.00	0.00	**chr11**
-0.01	0.01	-0.04	0.02	0.00	-0.01	0.03	-0.02	-0.01	0.00	-0.01	**chr12**
0.00	0.03	0.01	0.03	0.02	0.01	0.03	0.02	0.00	0.01	0.00	**chr13**
0.00	0.01	0.00	0.02	0.00	-0.01	0.03	0.00	0.00	0.00	0.00	**chr14**
0.01	0.02	0.00	0.02	0.00	-0.01	-0.08	0.00	0.00	0.00	-0.01	**chr15**
0.00	0.02	0.00	0.01	0.00	0.00	-0.07	0.01	0.00	0.00	0.00	**chr16**
-0.02	0.00	-0.01	0.00	0.00	0.00	0.03	0.00	0.00	-0.01	0.00	**chr17**
0.01	0.03	0.01	0.02	0.01	0.00	0.04	0.01	0.01	0.01	0.00	**chr18**
-0.01	0.00	0.00	0.00	0.00	0.01	0.02	0.01	-0.01	-0.01	0.01	**chr19**
0.01	0.03	0.00	0.02	-0.07	0.00	0.05	0.00	0.01	0.00	0.00	**chr20**
0.00	0.02	0.01	0.02	0.01	0.01	0.08	0.01	0.00	-0.09	0.01	**chr21**
0.02	0.02	-0.01	-0.01	0.02	0.00	0.07	0.00	0.03	-0.01	0.00	**chr22**

27836	30035	26995	27389	27583	25082	27391	25346	26660	25552	27333	**Case**
NP	NP	NP	NP	HeH	HeH	NP	NP	NP	NP	NP	**Status**
0.01	-0.04	0.01	-0.01	-0.06	-0.04	0.01	0.00	0.00	0.05	0.00	**chr1**
0.02	-0.03	0.02	-0.01	-0.05	0.16	0.02	0.01	0.00	0.00	0.01	**chr2**
0.02	-0.01	0.02	-0.01	0.00	-0.04	0.03	0.00	0.00	0.00	0.01	**chr3**
0.02	0.28	0.03	0.00	0.27	-0.04	0.04	0.02	0.01	0.01	0.02	**chr4**
0.02	-0.01	0.03	0.00	-0.05	-0.04	0.03	0.01	0.00	0.01	0.01	**chr5**
0.01	-0.02	0.03	-0.01	-0.05	0.16	0.03	0.00	0.00	0.00	0.00	**chr6**
-0.24	-0.02	0.02	0.00	-0.04	-0.04	0.02	0.01	0.00	0.00	-0.10	**chr7**
0.02	-0.01	0.03	0.00	-0.04	-0.04	0.03	0.01	0.00	0.01	0.01	**chr8**
0.02	-0.02	0.02	0.00	-0.04	-0.03	0.02	0.01	0.00	0.00	-0.08	**chr9**
0.02	-0.03	0.02	0.00	-0.06	-0.04	0.02	0.00	0.00	0.00	0.01	**chr10**
0.01	-0.03	0.02	0.00	-0.06	-0.03	0.02	0.00	0.00	0.00	0.00	**chr11**
0.01	-0.04	0.01	-0.01	-0.07	-0.04	0.01	0.00	-0.01	-0.01	0.00	**chr12**
0.02	-0.04	0.02	0.00	-0.07	-0.04	0.03	0.01	0.01	0.01	0.01	**chr13**
0.01	0.26	0.02	-0.01	0.24	0.15	0.02	0.00	-0.01	-0.01	0.00	**chr14**
0.01	-0.02	0.02	-0.01	-0.05	-0.04	0.02	0.00	-0.01	0.00	0.00	**chr15**
0.01	-0.02	0.01	0.00	-0.04	-0.02	0.03	0.00	0.00	0.00	0.00	**chr16**
0.00	-0.02	0.00	-0.01	0.22	-0.03	0.02	0.00	0.00	-0.01	0.00	**chr17**
0.02	-0.03	0.03	0.00	-0.05	0.16	0.03	0.01	0.00	0.00	0.01	**chr18**
-0.01	-0.03	0.00	0.00	-0.03	0.00	0.01	0.01	0.02	0.00	0.00	**chr19**
0.01	-0.02	0.02	0.00	-0.03	-0.03	0.04	0.00	0.00	0.00	0.00	**chr20**
0.01	0.26	0.02	0.00	-0.02	0.17	0.03	0.01	0.01	0.01	0.01	**chr21**
0.00	0.04	0.00	0.00	0.03	0.00	0.05	0.01	0.00	-0.01	0.01	**chr22**

25267	26610	28182	25893	25247	28011	27181	30034	28196	27298	29589	**Case**
NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	**Status**
0.00	0.02	0.05	-0.02	0.00	0.00	0.00	-0.01	0.00	0.03	-0.01	**chr1**
0.00	0.02	0.00	-0.02	0.00	0.00	0.00	0.00	0.00	0.04	0.13	**chr2**
0.00	0.03	0.00	-0.06	0.00	0.01	0.01	0.01	0.00	0.05	0.00	**chr3**
0.01	0.03	0.00	-0.01	0.01	0.01	0.01	0.02	0.01	0.05	0.00	**chr4**
0.00	0.03	-0.10	-0.01	0.00	0.13	0.01	-0.16	0.01	0.05	0.00	**chr5**
0.00	0.04	-0.01	-0.01	0.00	0.00	0.00	0.02	0.00	0.05	0.00	**chr6**
0.00	0.02	0.00	-0.01	0.00	-0.03	0.01	0.01	0.01	-0.14	0.00	**chr7**
0.00	0.02	0.00	0.00	0.00	0.01	0.01	0.09	0.01	0.05	0.00	**chr8**
0.00	-0.03	0.00	0.10	0.00	0.00	0.01	0.01	0.00	-0.01	0.00	**chr9**
0.00	0.01	0.00	-0.01	0.17	0.00	0.01	-0.01	0.01	0.03	0.00	**chr10**
0.00	0.01	0.00	-0.02	-0.01	-0.01	0.00	0.00	0.00	0.03	-0.01	**chr11**
-0.01	0.01	-0.01	-0.02	-0.01	0.00	0.00	-0.01	0.00	0.03	-0.01	**chr12**
0.00	0.03	0.01	-0.01	0.01	0.01	0.01	0.00	0.01	0.04	0.00	**chr13**
-0.01	0.02	0.00	-0.02	-0.01	0.00	0.00	0.00	0.00	0.03	-0.01	**chr14**
-0.01	0.01	-0.01	-0.02	-0.01	0.00	0.00	0.01	0.00	0.04	-0.01	**chr15**
0.00	0.00	0.00	-0.01	0.00	0.00	0.01	0.01	0.00	-0.07	-0.01	**chr16**
0.00	0.00	-0.01	-0.03	0.00	0.00	0.00	0.01	0.00	0.03	0.00	**chr17**
0.00	0.02	0.00	-0.01	0.00	0.01	0.01	0.01	0.01	0.04	0.00	**chr18**
0.01	0.00	0.01	-0.03	0.01	0.00	0.00	0.00	0.01	0.03	0.00	**chr19**
0.00	0.01	0.00	-0.01	-0.01	0.00	0.00	0.02	0.00	0.05	-0.01	**chr20**
0.01	0.03	0.00	-0.01	0.01	0.01	0.01	0.01	0.01	0.04	0.00	**chr21**
0.00	0.02	0.00	-0.01	0.00	0.00	0.00	0.05	0.00	0.06	0.01	**chr22**

25371	28335	27752	29908	28057	27026	28670	26971	25967	27754	27642	**Case**
NP	NP	NP	NP	NP	HeH	NP	NP	NP	NP	NP	**Status**
0.00	0.01	-0.01	0.00	0.00	0.08	-0.01	0.06	0.01	0.00	0.01	**chr1**
0.00	0.01	0.00	0.00	0.00	-0.03	0.02	0.01	0.01	0.00	-0.01	**chr2**
0.00	0.01	0.01	0.00	0.00	-0.03	-0.01	0.03	0.02	0.00	0.02	**chr3**
0.01	0.01	0.02	0.00	0.01	-0.02	0.03	0.03	0.03	0.00	0.03	**chr4**
0.00	0.01	0.01	0.00	0.00	0.10	0.01	0.05	0.20	0.00	0.02	**chr5**
0.00	0.00	0.00	-0.01	0.00	0.00	0.02	0.02	-0.26	-0.01	0.02	**chr6**
0.00	0.01	0.01	0.00	0.00	-0.03	0.02	0.02	-0.08	0.00	0.01	**chr7**
0.00	0.01	0.02	0.00	0.00	-0.04	0.02	0.04	0.02	0.00	0.02	**chr8**
0.00	-0.03	0.00	0.00	0.00	-0.05	-0.08	0.01	-0.07	0.00	0.00	**chr9**
0.00	0.01	0.00	0.00	0.00	-0.03	0.00	0.01	0.01	0.00	0.00	**chr10**
0.00	0.00	0.00	0.00	0.00	0.04	-0.10	0.00	0.01	0.00	0.00	**chr11**
-0.01	0.00	-0.01	-0.01	-0.01	0.08	-0.01	0.01	0.01	-0.01	-0.10	**chr12**
0.00	0.01	0.00	0.00	0.00	-0.02	0.00	0.02	0.02	0.00	0.01	**chr13**
0.00	0.00	-0.01	0.00	0.00	-0.03	-0.01	0.01	0.01	0.00	0.00	**chr14**
-0.01	0.00	0.00	0.00	-0.01	-0.03	-0.01	0.01	0.01	0.00	0.00	**chr15**
0.00	0.01	0.00	0.00	0.00	-0.03	0.01	0.01	0.01	0.00	0.00	**chr16**
0.00	0.00	-0.02	0.01	-0.01	-0.02	-0.01	0.00	-0.05	-0.01	0.01	**chr17**
0.00	0.01	0.01	0.00	0.00	-0.02	-0.04	0.02	0.02	0.00	0.01	**chr18**
0.01	0.00	-0.02	0.02	0.01	0.04	0.01	-0.01	0.00	0.00	-0.01	**chr19**
0.00	0.01	0.00	0.00	0.00	-0.03	0.01	0.02	0.01	0.00	0.02	**chr20**
0.01	0.01	0.00	0.00	0.00	0.09	0.04	0.02	0.19	0.00	0.00	**chr21**
0.00	0.00	0.03	0.01	0.00	0.06	0.01	0.04	0.00	0.00	0.03	**chr22**

25130	30036	27833	30032	27554	27509	27441	25100	25373	25451	28350	**Case**
NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	**Status**
0.00	-0.01	-0.02	0.01	-0.02	0.01	0.00	0.00	0.00	0.00	0.00	**chr1**
0.00	0.02	0.01	0.03	0.01	0.02	0.01	0.00	0.01	0.01	0.01	**chr2**
0.00	0.03	0.02	0.04	0.01	0.03	0.00	0.00	0.01	0.01	-0.04	**chr3**
0.01	0.04	0.02	0.04	0.01	0.04	0.00	0.01	0.02	0.01	0.02	**chr4**
0.00	0.29	0.02	0.03	-0.05	0.03	0.00	0.00	0.01	0.01	0.11	**chr5**
0.00	0.02	0.02	0.03	-0.01	0.03	0.00	0.00	0.01	0.01	0.00	**chr6**
0.00	0.03	0.01	0.03	0.01	0.00	-0.01	0.00	0.01	0.01	-0.29	**chr7**
0.00	0.03	0.02	0.04	0.01	0.06	0.00	0.00	0.01	0.01	0.01	**chr8**
-0.01	0.00	0.00	0.03	0.01	-0.01	-0.01	0.00	-0.04	0.01	-0.03	**chr9**
0.00	0.00	0.00	0.02	-0.03	0.02	0.01	0.00	0.01	0.01	0.01	**chr10**
0.00	0.01	0.00	0.01	0.11	0.02	0.00	0.00	0.00	0.01	0.01	**chr11**
-0.01	-0.01	-0.01	0.01	0.00	0.01	0.00	-0.01	0.00	0.00	0.00	**chr12**
0.00	0.01	-0.16	0.01	0.01	0.03	0.00	0.01	0.02	0.01	0.01	**chr13**
-0.01	-0.01	0.00	0.02	0.00	0.02	0.00	-0.01	0.00	0.00	0.00	**chr14**
0.00	-0.04	0.00	0.03	-0.01	0.02	-0.01	-0.01	0.00	0.00	0.00	**chr15**
0.00	-0.01	0.00	0.02	-0.01	0.03	-0.01	0.00	0.00	0.00	0.01	**chr16**
-0.01	-0.03	-0.01	0.01	-0.05	0.02	-0.01	0.00	0.00	-0.01	-0.01	**chr17**
0.00	0.02	0.01	0.03	0.01	0.03	0.01	0.00	0.01	0.01	0.01	**chr18**
0.00	-0.03	0.03	0.00	0.00	0.02	0.01	0.02	0.01	-0.01	0.01	**chr19**
0.00	-0.01	0.01	0.03	0.01	0.04	-0.01	0.00	0.00	0.00	0.01	**chr20**
0.01	-0.01	0.25	0.02	0.01	0.02	0.01	0.01	0.02	0.00	0.01	**chr21**
0.00	-0.02	0.07	0.01	0.00	0.05	0.00	0.00	0.01	-0.01	0.00	**chr22**

26621	27585	24813	24919	25246	28581	24309	26609	27085	25793	27071	**Case**
NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	**Status**
0.00	0.01	0.00	0.00	0.03	0.00	0.09	0.03	-0.01	-0.01	0.00	**chr1**
0.01	0.07	0.00	0.00	0.00	-0.02	0.01	0.05	0.00	-0.01	0.00	**chr2**
-0.03	-0.12	0.00	0.00	0.01	0.00	0.02	0.05	0.01	0.00	0.00	**chr3**
0.02	0.08	0.01	0.01	0.01	-0.05	0.00	0.06	0.01	0.00	0.01	**chr4**
0.01	0.07	0.01	0.00	0.01	-0.01	-0.28	0.06	0.01	0.00	0.00	**chr5**
-0.06	0.04	0.00	0.00	0.00	0.24	0.01	0.05	0.00	-0.01	0.00	**chr6**
0.03	-0.12	0.00	0.00	0.00	-0.03	0.01	-0.41	0.01	0.00	0.00	**chr7**
0.02	0.05	0.01	0.00	0.01	0.00	0.13	0.06	0.06	0.01	0.00	**chr8**
0.00	-0.29	-0.01	0.00	0.00	-0.02	0.00	0.04	0.00	-0.13	0.00	**chr9**
0.01	0.03	0.00	0.00	0.00	-0.01	0.01	0.04	-0.01	0.00	0.00	**chr10**
0.00	0.03	0.01	0.00	0.00	0.00	0.00	0.03	-0.01	0.08	0.00	**chr11**
0.00	0.02	-0.01	-0.01	0.00	-0.01	0.01	0.04	-0.01	0.00	-0.01	**chr12**
0.01	-0.28	0.01	0.01	0.01	-0.04	-0.03	0.05	-0.01	-0.01	0.00	**chr13**
0.00	0.03	0.00	-0.01	0.00	-0.02	0.00	0.04	-0.01	0.10	-0.01	**chr14**
0.01	0.00	-0.01	-0.01	0.00	0.01	0.01	0.04	0.00	-0.01	-0.01	**chr15**
0.00	0.04	0.00	0.00	0.00	0.00	-0.19	0.03	0.00	-0.01	0.00	**chr16**
0.00	0.03	-0.01	0.00	-0.01	-0.04	0.06	0.02	-0.01	-0.01	-0.01	**chr17**
0.01	0.04	0.01	0.00	0.01	-0.01	0.01	0.05	0.00	-0.01	0.00	**chr18**
-0.01	0.02	-0.02	0.02	-0.01	-0.08	-0.02	0.01	-0.02	-0.01	0.01	**chr19**
0.01	0.04	0.01	0.00	0.00	0.01	0.00	0.04	0.01	-0.01	0.00	**chr20**
0.01	0.04	0.01	0.01	0.00	-0.07	0.18	0.04	0.00	-0.02	0.01	**chr21**
0.01	0.06	0.00	0.00	-0.01	-0.04	-0.01	0.03	0.02	-0.01	0.00	**chr22**

31044	11669	26726	29089	26062	27812	27819	25344	26682	24890	**Case**
HoTr	HoTr	HeH	NP	NP	NP	NP	NP	NP	NP	**Status**
0.17	0.10	-0.07	0.00	0.00	0.10	0.00	0.00	0.02	0.00	**chr1**
-0.08	-0.08	-0.07	0.00	0.00	0.01	0.00	0.01	0.03	0.00	**chr2**
-0.08	-0.08	-0.07	0.00	0.00	0.01	0.02	0.01	0.03	0.01	**chr3**
-0.08	-0.07	0.23	0.01	0.01	-0.01	0.02	0.02	0.05	0.01	**chr4**
-0.08	0.11	-0.07	0.00	0.01	0.00	0.01	0.01	0.04	0.01	**chr5**
0.19	0.11	0.23	0.00	0.01	0.00	0.02	0.01	0.04	0.00	**chr6**
-0.08	-0.08	-0.06	0.00	0.00	0.00	0.01	-0.28	-0.25	0.01	**chr7**
0.19	0.11	-0.06	0.00	0.02	0.01	0.02	0.01	0.04	0.01	**chr8**
-0.08	-0.08	-0.06	0.00	0.00	0.00	-0.05	0.01	-0.05	0.01	**chr9**
0.17	0.10	0.22	0.00	0.01	0.01	-0.01	0.01	0.02	0.00	**chr10**
-0.08	0.10	-0.07	0.00	0.00	0.00	0.00	0.00	0.02	0.00	**chr11**
-0.08	-0.08	-0.07	-0.01	-0.01	0.00	0.00	0.00	-0.05	0.00	**chr12**
-0.07	-0.07	-0.07	0.00	0.02	0.00	0.00	0.01	-0.18	0.01	**chr13**
0.17	0.10	0.44	-0.01	0.00	0.00	0.00	0.01	0.02	0.00	**chr14**
-0.09	-0.08	-0.06	-0.01	-0.01	0.01	0.00	0.00	0.02	0.00	**chr15**
-0.08	-0.08	-0.06	0.00	0.00	0.00	0.00	0.00	0.02	0.00	**chr16**
-0.09	-0.08	0.22	0.00	0.01	-0.01	0.00	-0.01	0.00	0.00	**chr17**
0.20	0.12	0.23	0.00	0.00	0.01	0.01	0.01	0.04	0.01	**chr18**
0.13	0.08	-0.06	0.01	0.02	-0.03	0.00	0.00	0.00	0.00	**chr19**
-0.08	-0.08	-0.05	0.00	0.00	0.01	0.02	0.01	0.02	0.00	**chr20**
0.20	0.12	0.23	0.01	0.01	-0.01	0.00	0.01	0.03	0.01	**chr21**
0.12	0.06	-0.04	0.00	0.01	-0.01	0.02	-0.01	0.02	0.00	**chr22**

Supplementary table 9

Genetic details of Vienna validation cohort and output from decision tree classifier.

Patient	Classifier output	Genetic classification Vienna
1	Non-ploidy	diploid/t(1;19)/*TCF3-PBX1*
2	Non-ploidy	diploid/t(12;21)/*ETV6-RUNX1*
3	HoTr	diploid/t(12;21)/*ETV6-RUNX1*
4	Non-ploidy	diploid/B-other/*DUX4*
5	Non-ploidy	diploid//*KMT2A-AFF1*
6	Non-ploidy	diploid/B-other/*IKZF1*plus
7	Non-ploidy	diploid/B-other/*IKZF1*plus
8	Non-ploidy	high hyperdiploid
9	HeH	high hyperdiploid
10	HeH	high hyperdiploid
11	HeH	high hyperdiploid
12	HeH	high hyperdiploid
13	HeH	high hyperdiploid
14	HeH	high hyperdiploid
15	Non-ploidy	masked near-haploid with genome-wide LOH
16	HeH	masked near-haploid with genome-wide LOH
17	Non-ploidy	masked near-haploid with genome-wide LOH
18	Non-ploidy	masked near-haploid with genome-wide LOH
19	HeH	masked near-haploid with genome-wide LOH
20	HeH	near-haploid/low hyperdiploid mosaic
21	HeH	near-haploid/low hyperdiploid mosaic
22	HeH	masked near-haploid with genome-wide LOH
23	HoTr	low hypodiploid
24	HoTr	low hypodiploid
25	HoTr	low hypodiploid
26	HoTr	low hypodiploid
27	HoTr	low hypodiploid
28	HoTr	low hypodiploid
29	HoTr	low hypodiploid

Supplementary table 10

Coding regions included in SureSelect XT HS2 capture library design due to prevalence in clonal haematopoiesis (*DNMT3A*, *TET2*, *ASXL1*) or relapsed ALL (*TP53*, *KRAS*, *NRAS*, *PTPN11*).

Gene	Genomic co-ordinates	Exons
NRAS	chr1:115256299-115258874	2, 3
DNMT3A	chr2:25457138-25536863	all
TET2	chr4:106111617-106197686	all
PTPN11	chr12:112887835-112927579	3, 8, 13
KRAS	chr12:25380027-25398352	2, 3
TP53	chr17:7565247-7579922	all
ASXL1	chr20:31020911-31027928	11, 12

Supplementary table 11

Detailed mutational data of the SNVs (n=32) and indels (n=14) identified using the targeted NGS panel (library prep performed with SureSelect XT2 target enrichment kit) in 30 patients.

Gene	Location	Ref allele	Variant allele	Class	COSMIC ID	Patient
TP53	chr17:7577094	G	A	SNV	COSM1636702	27478
TP53	chr17:7577094	G	A	SNV	COSM1636702	29407
TP53	chr17:7577025	T	A	SNV	COSM99951	29491
TP53	chr17:7578479	G	A	SNV	COSM3378358	26910
NF1	chr17:29585375	G	A	SNV		28893
NF1	chr17:29685516-29685517	T	GGATAAG	Ins		25437
NF1	chr17:29676240-29676241	T	G	Ins		27478
NF1	chr17:29552243-29552244	G	TA	Ins		29407
NF1	chr17:29552244-29552245	G	GTCTCCGT	Ins		29407
JAK2	chr9:5078361	G	C	SNV	COSM29637	25130
JAK2	chr9:5080555	G	T	SNV		24890
JAK2	chr9:5089726	C	A	SNV	COSM23940	28893
PAX5	chr9:37006526	C	A	SNV		26614
PAX5	chr9:36882049	C	T	SNV		28670
PAX5	chr9:37020733	G	A	SNV		28893
ETV6	chr12:12038950-12038951	C	T	Ins		25130
ETV6	chr12:11905474-11905477	CAGGA	-	Del		26660
ETV6	chr12:11905492-11905493	GAT	-	Del		26660
CREBBP	chr16:3786037	C	A	SNV		27833
CREBBP	chr16:3786715	A	T	SNV	COSM220497	27833
CREBBP	chr16:3786036-3786037	C	TT	Ins		27833
ATM	chr11:108175544	C	T	SNV		25437

ATM	chr11:108098576	C	G	SNV		29491
CSF1R	chr5:149449827	C	T	SNV	COSM956	25100
CSF1R	chr5:149460542	A	C	SNV	COSM51393	25552
IKZF1	chr7:50450373	A	G	SNV		25451
IKZF1	chr7:50444255-50444256	A	TTCC	Ins		25130
FLT3	chr13:28626716	C	T	SNV	COSM28039	25437
FLT3	chr13:28626716	C	T	SNV	COSM28039	29407
KRAS	chr12:25380256	T	A	SNV		25451
KRAS	chr12:25398284	C	T	SNV	COSM1135366	28581
NRAS	chr1:115258748	C	T	SNV	COSM563	25967
NRAS	chr1:115258747	C	T	SNV	COSM564	24813
IL7R	chr5:35876251	A	C	SNV		24890
IL7R	chr5:35874574-35874575	A	TG	Ins		24890
KDM6A	chrX:44896923	T	C	SNV		25082
KDM6A	chrX:44938411	A	C	SNV		28670
ABL1	chr9:133760106	C	T	SNV		25130
ARID2	chr12:46246611	G	A	SNV		25967
ASXL1	chr20:31022441-31022442	A	G	Ins		27833
CXCR4	chr2:136872466-136872467	A	GGGGGCC	Ins		25437
DNMT3A	chr2:25464556	A	C	SNV		29407
IKZF3	chr17:37949083-37949084	A	T	Ins		26910
JAK3	chr19:17937610	T	C	SNV		27400
KMT2C	chr7:151859899	G	A	SNV	COSM1581234	25100
RAG2	chr11:36614521	C	G	SNV		27642
RUNX1	chr21:36206783-36206784	T	G	Ins		28182
TET2	chr4:106196381	C	T	SNV	COSM211627	25100