THE UTILITY OF NOVEL TECHNOLOGIES IN THE DIAGNOSIS OF CHRONIC MYELOID MALIGNANCIES

Catherine Anne Cargo

Doctor of Philosophy

University of York

Health Sciences

September 2019

ABSTRACT

The approach to diagnosing the myelodysplastic syndromes (MDS) and myelodysplastic/myeloproliferative neoplasms (MDS/MPN) is primarily based on a subjective morphological assessment that is neither sensitive nor specific. Objective measures such as a cytogenetic assessment yield results in only a proportion of patients. Over the past 2 decades a range of new technologies have been developed which have the potential to revolutionise the diagnosis of these conditions by providing an objective measure of disease.

This research aimed to investigate the utility of novel technologies in the diagnosis of MDS and MDS/MPN overlap syndromes, in particular chronic myelomonocytic leukaemia (CMML). The technologies utilized throughout the research included flow cytometry, single nucleotide polymorphism (SNP) arrays and high throughput sequencing (HTS). With the latter, a novel HTS panel was designed to target genes commonly mutated in myeloid malignancies. This was initially used to investigate a cohort of patients in whom a diagnosis could not be reached on initial analysis but subsequently developed a myeloid malignancy. Somatic mutations were detected at a very high frequency in the pre-diagnostic sample suggesting that targeted sequencing, in particular, could confirm clonality. This technology was further investigated on large cohorts of patients presenting with a monocytosis or cytopenia in whom mutations correlated strongly with survival and blood count trajectories as well as being predictive of a subsequent diagnosis.

The detection of somatic mutations in those with persistent monocytosis or cytopenia, particularly with a clone size >20% and co-occuring mutations, is clinically important and patients should be managed as per those with confirmed disease. HTS is therefore essential in the diagnosis of these patients. This does however raise concerns regarding funding and change management and will likely necessitate rationalization of the diagnostic service. Overall, there is significant potential to transform the diagnostic approach to this group of disorders.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I wish to thank all those who assisted me in completing this research including my work colleagues, friends and family. This thesis is the result of many years of hard work and would not have been possible without the significant contributions made by many people within HMDS. In particular, I wish to thank Paul Evans for his contribution and support and imparting some of his vast knowledge and experience of molecular genetics. I also wish to thank Sharon Barrans, Jan Taylor and Matt Cullen who were all instrumental in the delivery of this research. I also wish to acknowledge the support of the entire staff within HMDS who not only performed much of the routine work but also kept me sane through this process. More widely I thank Nicola Rowbotham and Tumas Beinortas, academic trainees who provided support both within the lab and by collecting clinical information. I thank all the clinicians from the region who readily and willingly provided additional information if required and everyone at ECSG (Alex Smith, Dan Painter and John Blase) for help with data analysis and preparation of manuscripts.

A huge thankyou to my husband Ian who both assisted and supported me by sharing his knowledge of IT and taking on the domestic responsibilities. I promise this will be my last degree. I would also like to thank my son Max, who arrived towards the end of my research and always makes me smile.

Finally, I am extremely grateful to my supervisors, Andrew Jack, Eve Roman and Simon Crouch, for their advice, support, constructive feedback and encouragement. A special thanks to Simon who also provided much needed statistical support throughout.

AUTHOR'S DECLARATION

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

This research was undertaken within the Haematological Malignancy Diagnostic Service (HMDS), a large integrated diagnostic laboratory, and was performed with the assistance of many scientific staff within the department. My contributions have been the overall conception and design of the research, the development and testing of the targeted myeloid sequencing panel (in collaboration with Dr Paul Evans, Consultant Clinical Scientist), annotation and interpretation of all sequencing data, analysis of SNP-array data (in collaboration with Dr Sharon Barrans, Consultant Clinical Scientist), collection of clinical and follow-up data (in collaboration with HMRN and Dr Tumas Beinortas) and ultimate integration of data and analysis (with statistical analysis support from Dr Simon Crouch). As this research was performed in a clinical laboratory, many aspects of the sample processing and analysis were performed as part of the routine clinical work and many scientific staff contributed to this. This ultimately included the targeted sequencing panel developed as part of this research which was incorporated into routine practice within HMDS. Additional aspects of this research were performed by others including: flow cytometry analysis which was performed by Dr Matt Cullen, Clinical Scientist; and development of the in-house sequencing analysis pipeline which was created by Dr Jan Taylor, Clinical Scientist in Bioinformatics. In addition, the integration of new testing processes into the current laboratory informatics system was performed by Dr Richard Jones, Clinical Scientist.

To date there have been 2 publications arising from this research with associated commentaries which are detailed below. These papers are included in the Appendix.

Targeted sequencing identifies patients with preclinical MDS at high risk of disease progression.

Cargo CA, Rowbotham N, Evans PA, Barrans SL, Bowen DT, Crouch S, Jack AS.

Blood. 2015 Nov 19;126(21):2362-5.

The use of targeted sequencing and flow cytometry to identify patients with a clinically significant monocytosis.

Cargo C, Cullen M, Taylor J, Short M, Glover P, Van Hoppe S, Smith A, Evans P, Crouch S.

Blood. 2019 Mar 21;133(12):1325-1334.

CHAPTER 1: INTRODUCTION

The chronic myeloid malignancies are stem cell derived clonal disorders and include three broad categories – myelodysplastic syndromes (MDS), myeloproliferative neoplasms (MPN) and the "overlap" category of myelodysplastic/myeloproliferative neoplasms (MDS/MPN). The gold standard diagnostic criteria for these cancers are those recommended in the World Health Organisation's (WHO) most recent classification and are primarily based on peripheral blood counts and subjective morphological assessment of both peripheral blood and bone marrow smears (Swerdlow *et al.*, 2017). While the diagnosis of MPNs now centre on the detection of key molecular abnormalities, the diagnosis of MDS and MDS/MPNs are somewhat dated in the fast-moving area of diagnostic haematology. More recently a number of new technologies have been reported in the literature including array-based approaches and a new generation of high throughput sequencers. To date these have been most commonly utilized in the research setting however as experience and knowledge grows and costs fall, they are now reaching the clinical interface.

1.1 Disease context and research aims

The myelodysplastic syndromes are malignant haemopoietic disorders characterised by cytopenias, ineffective haematopoiesis and a propensity to evolve to acute myeloid leukaemia (AML). Apart from a small subset with low risk disease the majority of patients have a poor prognosis with a 5-year overall survival (OS) of only 21.2% (95% CI 18.7-23.8%) (Roman *et al.*, 2016). Confirming a diagnosis of MDS in a cytopenic patient has been notoriously difficult particularly in early disease, not only due to the extensive differential diagnoses in these patients but also the morphological challenges when assessing a bone marrow for dysplasia. There is reported poor inter-observer concordance when recognising dysplasia and numerous non-neoplastic conditions which can mimic MDS (Parmentier *et al.*, 2012; Steensma, 2012; Bejar, 2015). In addition, dysplasia is also reported in a significant proportion of healthy marrows (Parmentier *et al.*, 2012). As a result,

patients can be erroneously diagnosed with MDS or a diagnosis of MDS could be missed.

Similarly, in a patient presenting with a monocytosis, a diagnosis of CMML can be equally as challenging. The MDS/MPNs are a group of disorders with both dysplastic and proliferative features, with CMML being by far the most common of these disorders. In CMML, the presence of a monocytosis is the hallmark of disease however patients can present at any point on the dysplastic/proliferative spectrum with presenting features ranging from an MDS-like phenotype with a monocytosis to a marked proliferative phenotype with raised white cell count, constitutional symptoms and splenomegaly (Itzykson *et al.*, 2018). Similarly to MDS, the overall prognosis is poor with a 5 year OS of 13.3% (95% CI 9.1-18.4%) (Roman *et al.*, 2016). Again diagnosis remains largely centred on morphology though worryingly a diagnosis of CMML can be made even in the absence of definitive morphological findings providing the monocytosis is persistent and alternative causes have been excluded (Swerdlow *et al.*, 2017).

To date, cytogenetics has provided the only clonal markers of disease for both of these disease sub-groups though abnormalities are identified in only 30% and 50% of confirmed cases of CMML and MDS respectively (Schanz *et al.*, 2012; Palomo, Garcia, *et al.*, 2016). An objective measure of disease is therefore a much-needed core-criterion for diagnosis, particularly when the disease in question confers such a poor prognosis.

This research aims to address the issues surrounding the current approach to diagnosing chronic myeloid malignancies in adults, as well as investigating how new technologies can refine the diagnostic approach to those patients presenting with cytopenias and/or a monocytosis. This introductory chapter provides an overview of the current diagnostic and classification systems, discusses the history and evidence base behind them and describes the latest technologies and their impact to date.

1.2 Classification of Myeloid Malignancies

1.2.1 The FAB Classification – A Morphology Based Approach

In 1976, the French-American-British (FAB) co-operative group, while proposing a classification for acute leukaemias, described a distinct group of less acute disorders typically occurring in the over 50s with a propensity to evolve to AML. (Bennett *et al.*, 1976). These were collectively termed dysmyelopoietic syndromes and were divided into two broad subgroups, refractory anaemia with excess of blasts (RAEB) and chronic myelomonocytic leukaemia (CMML), the former defined as having a blast plus promyelocyte count of 10-30% (Bennett *et al.*, 1976). The diagnostic criteria for CMML included the presence of a monocytosis (>1x10^9/l at some stage during the disease), myeloblasts and promyelocytes up to 30% and a raised serum lysozyme (Bennett *et al.*, 1976).

Following reports of the heterogeneous nature of this group of disorders, the FAB group reconvened in 1982 to determine if specific morphological features could further subdivide MDS (Bennett *et al.*, 1982). Based on the review of 80 cases, five new subtypes of MDS were defined based on morphological features in both the peripheral blood and bone marrow (Table 1.1). In contrast to the 1976 FAB classification normal promyelocytes were excluded from the blast count and a detailed description of morphological blast and dysplastic features was provided (Bennett *et al.*, 1982). CMML remained within the MDS classification despite a number of cases demonstrating proliferative features with minimal dysplastic changes (Bennett *et al.*, 1982).

The FAB classification was internationally accepted as the standard method for classifying MDS and became widely used as a diagnostic guideline. Concerns were raised, however, regarding the specificity of the morphological features proposed by the FAB group. Abnormal megakaryocytes were shown to be present in both normal and pathologic control marrows (Kuriyama *et al.*, 1986; Wong and Chan, 1991). Studies of normal bone marrows (21-56yrs) showed mild dyserythropoiesis (dysplastic features ≤10% of erythroblasts) in 78% (n=50) while 38% of subjects had

Table 1.1. FAB Classification of the myelodysplastic syndromes (adapted from Bennett et al, 1982)

Category	Peripheral Blood		Bone Marrow
Refractory anaemia (RA) or refractory cytopenia	Anaemia* Blasts ≤ 1% Monocytes ≤ 1 x 10^9/l	and	Blasts < 5%, ringed sideroblasts ≤15% of erythroblasts
Refractory anaemia with ringed sideroblasts (RARS)	Anaemia Blasts ≤ 1% Monocytes ≤ 1 x 10^9/l	and	Blasts < 5%, ringed sideroblasts >15% of erythroblasts
Refractory anaemia with excess blasts (RAEB)	Cytopenias Blasts < 5% Monocytes ≤ 1 x 10^9/l	and	Blasts ≥ 5% but ≤ 20%
Chronic myelomonocytic leukaemia (CMML)	Blasts < 5 % Monocytes > 1 x 10^9/l Granulocytes often increased	and	Blasts up to 20% promonocytes often increase
Refractory anaemia with excess of blasts in transformation (RAEB-T)	Blasts ≥ 5% or Auer rods in blasts	or	Blast > 20% but < 30% or Auer rods in blasts

* Or in the case of refractory cytopenia, either neutropenia or thrombocytopenia

Dyserythropoiesis	Dysmegakaryocytopoiesis	Dysgranulopoiesis
Ringed sideroblasts ≥ 15% Multinuclearity Nuclear fragments Other nuclear abnormalities Cytoplasmic abnormalities Erythroblasts < 5%, > 60%	Micromegakaryocytes Large mononuclear forms Multiple small nuclei Reduced numbers	Nuclear abnormalities Hypogranular cells

occasional dysplastic megakaryocytes (Bain, 1996). In contrast those older than 50 demonstrated dyserythropoiesis and dysgranulopoiesis in up to 15% and 27% of

respective cell elements (Ramos *et al.*, 1999). This is particularly relevant as the median age of the MDS group is 75.7yrs (Roman *et al.*, 2016).

The most specific morphological markers of MDS were reported as pseudo-Pelger-Huet anomalies, micromegakaryocytes, hypogranular megakaryocytes and internuclear bridging in erythroid precursors (Kuriyama *et al.*, 1986; Head *et al.*, 1989; Wong and Chan, 1991). However these abnormalities were not disease specific and pseudo-Pelger-Huet forms have since been reported as an iatrogenic phenomenon in a proportion of patients (Wang *et al.*, 2011).

This initial attempt to define and classify this group of disorders was imperfect and limited to an approach which is neither sensitive nor specific. This was, however, restricted by the tools available at that time.

1.2.2 Advances in Technology Provide New Tools for Diagnosis

1.2.2.1 Cytogenetic Analysis

With the advent of G-banding (Drets and Shaw, 1971; Seabright, 1971) the ability to detect small structural abnormalities in chromosomes enabled researchers to identify underlying clonal karyotypes in a variety of malignancies. Recurrent abnormalities were identified in the MDS patient group providing both an additional diagnostic marker and prognostic information. A report from the Second International Workshop on Chromosomes in Leukaemia identified chromosomal aberrations in approximately 50% of patients with the most common abnormalities being +8, -7, 7q-, -5 and 5q- ('Second International Workshop on Chromosomes in Leukemia', 1980). The mortality rate was noted to be substantially higher in the abnormal karyotype group ('Second International Workshop on Chromosomes in Leukemia', 1980). Specific abnormalities were noted to have characteristic morphological features including hypolobated nuclei in the megakaryocytes of those patients with del(5q), and pseudo-Pelger-Huet anomaly and small vacuolated neutrophils in those with loss of 17p (Mahmood *et al.*, 1979; Lai *et al.*, 1995). An international prognostic scoring system (IPSS) was subsequently developed

incorporating cytogenetics results along with percentage of BM myeloblasts and number of cytopenias (Table 1.2). With the combination of these features, this system separated patients into 4 distinct subgroups based on median survival and risk of transformation to AML (Greenberg *et al.*, 1997).

Table 1.2: International Prognostic Scoring System (adapted from Greenberg et al, 1997)

Score	0	0.5	1	1.5	2
Prognostic variables					
% bone marrow blasts	<5%	5-10%		11-19%	20-30%
Karyotype*	Good	Intermediate	Poor		
Cytopenias**	0-1	2-3			

*Karyotype: Good = normal, -Y, del(5q), del (20q);
Poor = complex (≥3 abnormalities) or chromosome 7 anomalies;
Intermediate = other abnormalities

**Cytopenias: Hgb <10g/dL
Neutrophils <1.8x10^9/L
Platelets <100x10^9/L

1.2.2.2 Immunohistochemistry

The additional information provided by the bone marrow trephine biopsy was also highlighted particularly in the assessment of patients with hypocellular MDS and those with MDS associated with fibrosis (Lambertenghi-Deliliers *et al.*, 1991; Tuzuner *et al.*, 1995). The addition of immunohistochemical markers such as CD61 allowed a reproducible method to study megakaryocyte size, shape and position (Fox *et al.*, 1990) and increased numbers of CD34 positive blasts by immunohistochemistry was also shown to have an impact on prognosis (Verburgh *et al.*, 2003).

1.2.3 The WHO Classification - a Combined Approach to Diagnosis

Following the publication of the Revised European-American Lymphoma (REAL) classification in 1994, the European Association of Haematopathologists and the Society for Hematopathology developed a new WHO classification based on the principle that disease entities were defined by a combination of morphology, immunophenotype, genetic and clinical features (Harris *et al.*, 1999). This classification divided myeloid malignancies into four major groups – MPN, MDS/MPN, MDS and AML.

1.2.3.1 Myelodysplastic Syndromes

The AML blast threshold was lowered based on a number of studies which demonstrated that the prognosis in patients with 20-30% was similar blasts to those with >30% blasts (Bernstein *et al.*, 1996; Chan *et al.*, 1997; Estey *et al.*, 1997). Recommendations were also made to divide RAEB into 2 subcategories (Table 1.3) based on evidence that those patients with >10% blasts had a worse prognosis (Greenberg *et al.*, 1997). The presence of Auer rods alone to place patients into the RAEB-T subgroup was also disputed following a study which illustrated a higher probability of survival in those patients defined as RAEB-T solely on this basis (Seymour and Estey, 1993).

The need for minimal criteria to define MDS was highlighted particularly in the advent of the new category of Refractory Cytopenia with Multilineage Dysplasia (RCMD) (Greenberg *et al.*, 2000). The introduction of this category followed reports that dysplasia in 2 or more cell lines was associated with similar cytopenias to RAEB and a prognosis intermediate between RARS and RAEB (Rosati *et al.*, 1996; Balduini *et al.*, 1998). In particular, one study showed the presence of pseudo-Pelger-Huet neutrophils and micromegakaryocytes in the FAB-RA patients correlated with a poorer overall survival (OS) and leukaemia free survival (LFS) (Matsuda *et al.*, 1998, 1999). These studies used strict morphological criteria for trilineage dysplasia requiring dysplastic features in ≥3% of granulocytic and erythroid lineages and ≥10% in the megakaryocyte lineage (Rosati *et al.*, 1996).

Table 1.3. WHO Classification of Myelodysplastic Syndromes 2002 (adapted from Vardiman et al, 2002)

Disease	Blood findings	Bone marrow findings
Refractory anemia (RA)	Anemia No or rare blasts	Erythroid dysplasia only <5% blasts <15% ringed sideroblasts
Refractory anemia with ringed sideroblasts (RARS)	Anemia No blasts	Erythroid dysplasia only ≥15% ringed sideroblasts <5% blasts
Refractory cytopenia with multilineage dysplasia (RCMD)	Cytopenias (bicytopenia or pancytopenia) No or rare blasts No Auer rods <$1x10^9$/L monocytes	Dysplasia in ≥10% of cells in 2 or more myeloid cell lines <5% blasts in marrow No Auer rods <15% ringed sideroblasts
Refractory cytopenia with multilineage dysplasia and ringed sideroblasts (RCMD-RS)	Cytopenias (bicytopenia or pancytopenia) No or rare blasts No Auer rods <$1x10^9$/L monocytes	Dysplasia in ≥10% of cells in 2 or more myeloid cell lines ≥15% ringed sideroblasts <5% blasts No Auer rods
Refractory anemia with excess blasts-1 (RAEB-1)	Cytopenias <5% blasts No Auer rods <$1x10^9$/L monocytes	Unilineage or multilineage dysplasia 5%-19% blasts No Auer rods
Refractory anemia with excess blasts-2 (RAEB-2)	Cytopenias 5%-19% blasts Auer rods ± <$1x10^9$/L monocytes	Unilineage or multilineage dysplasia 10%-19% blasts Auer rods ±
Myelodysplastic syndrome, unclassified (MDS-U)	Cytopenias No or rare blasts No Auer rods	Unilineage dysplasia in granulocytes or megakaryocytes <5% blasts No Auer rods
MDS associated with isolated del(5q)	Anemia <5% blasts Platelets normal or increased	Normal to increased megakaryocytes with hypolobated nuclei <5% blasts No Auer rods Isolated del(5q)

With these findings and recommendations from Kouides and Bennett, the WHO therefore opted for a minimal quantitative threshold of 10% to define dysplasia

with the aim of providing consistency in diagnosis (Kouides and Bennett, 1996; Vardiman, Harris and Brunning, 2002).

Validation studies in large numbers of patients (n=1600) confirmed that the new subgroups were more homogeneous with respect to prognosis (Germing *et al.*, 2000). Others however failed to demonstrate a significant difference in survival or progression to AML in those patients re-categorized to RCMD (Nösslinger *et al.*, 2001). The latter study did not however use the WHO criteria of 10% for dysplasia.

1.2.3.2 The WHO Recognise Mixed Myelodysplastic/Myeloproliferative Neoplasms

The WHO classification introduced a new subgroup primarily to accommodate CMML which had been difficult to classify due to both its dysplastic and proliferative phenotypes. This group also included atypical chronic myeloid leukaemia (aCML), juvenile myelomonocytic leukaemia (JMML) and myelodysplastic/myeloproliferative neoplasms-unclassifiable (MDS/MPD-u) (Table 1.4) (Vardiman, Harris and Brunning, 2002). This thesis focuses on disorders in adulthood and therefore JMML will not be included in this review.

Following publication of the FAB classification, reports surfaced of distinct prognostic factors in the CMML patient group. Blast percentage was shown to have a prognostic impact similar to that of RAEB, while the initial leucocyte count was also shown to impact on survival (Fenaux *et al.*, 1988). The spectrum of patients presenting with a monocytosis was vast ranging from those with severe cytopenias to those with largely proliferative features or in some cases arising secondary to other MPD or MDS (Michaux and Martiat, 1993). Similarities were noted between these patients and the group of patients with BCR-ABL1 negative chronic myeloid

Table 1.4. WHO Classification of CMML and atypical CML 2002 (adapted from Vardiman et al, 2002)

CMML
Persistent peripheral blood monocytosis greater than $1x10^9$L
No Philadelphia chromosome or BCR/ABL fusion gene
Fewer than 20% blasts* in the blood or bone marrow
Dysplasia in one or more myeloid lineages. If myelodysplasia is absent or minimal, the diagnosis of CMML may still be made if the other requirements are present and: an acquired, clonal cytogenetic abnormality is present in the marrow cells, or the monocytosis has been persistent for at least 3 months and all other causes of monocytosis have been excluded
Diagnose CMML-1 when blasts fewer than 5% in blood and fewer than 10% in bone marrow
Diagnose CMML-2 when blasts are 5% to 19% in blood, or 10% to 19% in marrow, or if Auer rods are present and blasts are fewer than 20% in blood or marrow
Diagnose CMML-1 or CMML-2 with eosinophilia when the criteria above are present and when the eosinophil count in the peripheral blood is greater than $1.5x10^9$/L

*In this classification of CMML, blasts include myeloblasts, monoblasts, and promonocytes

Atypical CML
Peripheral blood leucocytosis (WBC ≥ $13x10^9$/L) due to increased numbers of neutrophils and their precursors with prominent dysgranulopoiesis
No Ph chromosome of BCR-ABL1 fusion gene
Neutrophil precursors (promyelocytes, myelocytes, metamyelocytes) ≥ 10% of leukocytes
Minimal absolute basophilia; basophils usually <2% of leukocytes
No or minimal absolute monocytosis: monocytes <10% of leukocytes
Hypercellular bone marrow with granulocytic proliferation and granulocytic dysplasia, with or without dysplasia in the erythroid and megakaryocytic lineages
Less than 20% blasts in the blood and in the bone marrow

leukaemia (CML), termed atypical CML (Galton, 1992). Initial reports suggested that these cases represented part of the spectrum of CMML (Martiat, Michaux and Rodhain, 1991) and some authors recommended a combined subgroup for these entities (Michaux and Martiat, 1993).

As a result of these reports, the FAB group published recommendations on diagnosing the chronic myeloid malignancies with particular reference to CML, aCML and CMML (Bennett *et al.*, 1994). They described distinct features which could be used to diagnose these conditions (Table 1.5) (Bennett *et al.*, 1994). With

respect to aCML, it was stated that these patients had significantly greater dysplasia in the granulocytic lineage when compared with both CML and CMML. They also felt it was appropriate to distinguish between the myelodysplastic and myeloproliferative forms of CMML using a white cell count of $13x10^9$/l as a cut-off. This was considered to represent a significant degree of leucocytosis (Bennett *et al.*, 1994).

Table 1.5: FAB Classification of chronic myeloid malignancies (adapted from Bennett et al, 1994)

	CGL	aCML	CMML
Basophils	≥2%	<2%	<2%
Monocytes	<3%	≥3-10%	≥3-10%
Granulocytic dysplasia	-	++	+
Immature granulocytes	>20%	10-20%	≤10%
Blasts	≤2%	>2%	<2%

The clinical impact of such an arbitrary cut-off was assessed in a number of studies and while distinct clinical features were identified between the dysplastic and proliferative variants there was minimal impact on either OS or acute leukaemic transformation (Germing *et al.*, 1998; Voglová *et al.*, 2001).

The arbitrary cut-off proposed by the FAB group was removed with the introduction of the new WHO MDS/MPN subcategory. This enabled clinicians to determine if dysplastic features or proliferative features predominated and treat accordingly (Vardiman, Harris and Brunning, 2002). In addition, 2 prognostic categories (CMML-1 and -2) were defined based on the blast count following evidence that the percentage correlated with prognosis (Tefferi *et al.*, 1989). In this classification the blast count included myeloblasts, monoblasts and promonocytes (Vardiman, Harris and Brunning, 2002).

The prognostic relevance of the CMML subgroups was confirmed in an analysis of 339 patients, the CMML-2 group having a significantly reduced median survival and

an increased risk of leukaemic transformation (Germing *et al.*, 2007). Regarding aCML, one report showed ongoing heterogeneity within this patient group with respect to both clinical and haematological findings despite the revised diagnostic criteria (Breccia *et al.*, 2006)

The category of MDS/MPN, unclassifiable was first introduced in the 2001 WHO classification. This category includes patients with clinical, morphological and laboratory features that overlap with both MDS and MPN that do not meet the criteria for other WHO classifications. Within this category a new subgroup of 'refractory anaemia with ringed sideroblasts associated with marked thrombocytosis (RARS-T)' was introduced as a provisional entity. This entity was to include patients who fulfilled the criteria for RARS but also had a platelet count over $600x10^9$/L (WHO, 2001). This phenomenon was first recognized as a favourable prognostic factor in 1977 and subsequent studies confirmed the presence of thrombocytosis in up to 20% of patients with this MDS subtype (Streeter, Presant and Reinhard, 1977; Juneja *et al.*, 1983; Gupta, Abdalla and Bain, 1999). The evidence that this was a separate entity, however, was based on case reports and small series leading some researchers to question its' validity (Schmitt-Graeff *et al.*, 2002). Patients in this subgroup have, however, been shown to have a better prognosis compared to others with MDS/MPN, with survival comparable to that of RARS but less favourable than patients with essential thrombocythaemia (Shaw, 2005; Atallah *et al.*, 2008). The presence of the *JAK2* V617F mutation has also been found in a high percentage of these patients and appears to confer a better prognosis (Atallah *et al.*, 2008). These studies supported the classification of RARS-T as a separate entity.

Despite the overall improvement on the FAB classification, by incorporating objective measures such as cytogenetics, subjective morphological assessment remained central to the diagnosis of the chronic myeloid malignancies.

1.2.4 The WHO Classification (2008) - Refinement of a Morphological Classification

In 2008 a 4th edition of the WHO classification was published (Swerdlow *et al.*, 2008). Minimal changes were noted in the MDS/MPN group; however, a number of changes were made in the MDS group (Table 1.6). Unfortunately, these primarily involved refining the morphological diagnosis.

The minimal amount of dysplasia required for a diagnosis of MDS remained uniform at 10% for each lineage, though proposals had been made to raise the threshold in the megakaryocyte lineage to 40% following a report that this was associated with an adverse prognosis on a multivariate analysis (Matsuda *et al.*, 2007). Guidance was provided in the literature on identifying blast cells, ring sideroblasts and the stages of monocytic maturation in an attempt to standardize the morphological assessment (Mufti *et al.*, 2008; Goasguen *et al.*, 2009). However, the reproducibility of these descriptions, as tested by a panel of experts, suggested a less than optimal agreement further highlighting the limitations of this technique (Mufti *et al.*, 2008; Goasguen *et al.*, 2009).

For those patients who failed to meet these criteria a presumptive diagnosis of MDS was permitted if one of the 'specific clonal chromosomal abnormalities' (Table 1.7) was detected (Vardiman *et al.*, 2009). To address the issue of patients who lacked either morphological or cytogenetic evidence of MDS, the newly coined term "idiopathic cytopenia of undetermined significance" (ICUS) was suggested as a descriptive phrase (Wimazal *et al.*, 2007).

Table 1.6: WHO classification of Myelodysplastic Syndromes 2008 (adapted from Swerdlow et al. 2008)

Disease	Blood Findings	Bone Marrow Findings
Refractory cytopenias with unilineage dysplasia (RCUD) Refractory anaemia; Refractory neutropenia; Refractory thrombocytopenia	Unicytopenia or bicytopenia No rare of blasts (<1%)	Unilineage dysplasia: ≥10% of the cells in one myeloid lineage <5% blasts <15% of erythroid precursors are ring sideroblasts
Refractory anaemia with ring sideroblasts (RARS)	Anaemia No blasts	≥15% of erythroid precursors are ring sideroblasts Erythroid dysplasia only <5% blasts
Refractory cytopenia with multilineage dysplasia (RCMD)	Cytopenia(s) No or rare blasts (<1%) No Auer rods <$1x10^9$/L monocytes	Dysplasia in ≥10% of the cells in ≥ two myeloid lineages (neutrophil and/or erythroid precursors and/or megakaryocytes) <5% blasts in marrow No Auer rods ±15% ring sideroblasts
Refractory anaemia with excess blasts-1 (RAEB-1)	Cytopenia(s) <5% blasts No Auer rods <$1x10^9$/L monocytes	Unilineage or multilineage dysplasia 5-9% blasts No Auer rods
Refractory anaemia with excess blasts-2 (RAEB-2)	Cytopenia(s) <5-19% blasts Auer rods ± <$1x10^9$/L monocytes	Unilineage or multilineage dysplasia 10-19% blasts Auer rods ±
Myelodysplastic syndrome - unclassified (MDS-U)	Cytopenias ≤1% blasts	Unequivocal dysplasia in <10% of cells in one or more myeloid cell lines when accompanied by a cytogenetic abnormality considered as presumptive evidence for a diagnosis of MDS <5% blasts
MDS associated with isolated del(5q)	Anaemia Usually normal or increased platelet count No or rare blasts (<1%)	Normal to increased megakaryocytes with hypolobated nuclei <5% blasts Isolated del(5q) cytogenetic abnormality No Auer rods

Table 1.7: Recurring chromosomal abnormalities considered as presumptive evidence of MDS (adapted from Vardiman et al, 2009)

Unbalanced	Balanced
-7 or del(7q)	t(11;16)(q23;13.3)
-5 or del(5q)	t(3;21)(q26.2;q22.1)
i(17q) or t(17p)	t(1;3)(p36.3;21.2)
-13 or del (13q)	t(2:11)(p21;q23)
del(11q)	inv(3)(q21q26.2)
del(12p) or t(12p)	t(6;9)(p23;q34)
del(9q)	
idic(X)(q13)	

In addition, to improve the classification of those in the MDS-unclassified subgroup, a new category was recommended for patients with unilineage dysplasia in the presence of either uni- or bi-cytopenia (RCUD) (Vardiman *et al.*, 2009).

1.2.5 The Integrated Approach to Diagnosing Myeloproliferative Neoplasms

The MPNs in the most recent WHO classification are highlighted in Table 1.8.

Table 1.8: WHO Classification of Myeloproliferative Neoplasms 2017 (adapted from Swerdlow et al, 2017)

Myeloproliferative neoplasms
Chronic myelogenous leukaemia, BCR-ABL1 positive
Chronic neutrophilic leukaemia
Polycythaemia vera
Primary myelofibrosis
Essential thrombocythaemia
Chronic eosinophilic leukaemia, not otherwise specified
Myeloproliferative neoplasm, unclassifiable

This group is described separately as it has seen the greatest advancements over the past 4 decades. This has led to a greater understanding of pathogenesis, specific diagnostic markers and disease specific treatments with dramatic success.

The term myeloproliferative disorder was first used in 1951 by William Dameshek, who described a collection of disorders with variable proliferative activity. This included chronic myeloid leukaemia, polycythaemia vera (PV), primary myelofibrosis (PMF), essential thrombocythaemia (ET) and Di Guglielmo syndrome, now recognized as erythroleukaemia (Dameshek, 1951). The latter has since been placed in the subgroup of acute leukaemia, though the others have remained as myeloproliferative disorders.

The discovery of the Philadelphia chromosome has revolutionized the diagnosis and treatment of chronic myeloid leukaemia (Nowell and Hungerford, 1960; Rowley, 1973). The mechanism of disease became firmly established and the concept of targeted therapy was realized with the development of a tyrosine kinase inhibitor in the form of STI571, now known as imatinib (O'Dwyer and Druker, 2000). The presence of the Philadelphia chromosome or BCR/ABL fusion gene is a pre-requisite for the diagnosis of CML according to the most recent WHO classification (Swerdlow *et al.*, 2017).

The progress in molecular and genetic diagnostics led to the discovery of a number of frequently occurring and disease specific abnormalities in the myeloproliferative disease group.

The discovery of the *JAK2* V617F mutation (G to T at nucleotide 1849 resulting in the substitution of valine to phenylalanine at position 617) in myeloproliferative disorders marked the beginning of a series of developments within this area (Baxter *et al.*, 2005; James *et al.*, 2005; Levine *et al.*, 2005; R Kralovics *et al.*, 2005). This gain of function mutation occurs in almost all patients with PV (approximately 95%) though with 50% of patients with either ET or PMF also harbouring the mutation as

well as a number of other myeloid neoplasms, it is not specific (Tefferi *et al.*, 2005, 2006; Wolanskyj *et al.*, 2005). In patients without the V617F mutation, an exon 12 mutation has been identified which is functionally similar and therefore *JAK2* aberrations are found in virtually all PV patients (Scott *et al.*, 2007). In addition, a small percentage of patients with PMF (5%) or ET (1%) were shown to have a functionally similar gain of function mutation of *MPL*, either *MPL*W515L or *MPL*W515K (A. Pardanani *et al.*, 2006; Pikman *et al.*, 2006). The presence of either of these mutations confirms the clonality of these disorders and is now one of the major criteria in the WHO classification. .

With respect to other recurrent abnormalities, somatic point mutations of the KIT protooncogene are found to be recurring abnormalities in mastocytosis the most common being D816V (Nagata *et al.*, 1995). This mutation has been identified in more than 95% of patients with systemic mastocytosis with other activating point mutations including D816V found in cutaneous mastocytosis (Swerdlow *et al.*, 2008). Due to it's unique clinical and pathologic features, mastocytosis is now considered in a separate disease category in the most recent WHO classification (Swerdlow *et al.*, 2017).

A separate category was also introduced in the 2008 WHO classification termed 'Myeloid and Lymphoid neoplasms with Eosinophilia and Abnormalities of PDGFRA, PDGFRB and FGFR1' (Swerdlow *et al.*, 2008). These are 3 rare specific disease groups all resulting from formation of a fusion gene encoding an aberrant tyrosine kinase (Swerdlow *et al.*, 2008). The PDGFRA related disorders most commonly present as chronic eosinophilic leukaemia with a prominent mast cell population and have been shown to be responsive to tyrosine kinase inhibitors, in particular imatinib (Cools *et al.*, 2003). Those with PDGFRB related disorders most commonly present as CMML with eosinophilia and again show responses to imatinib (Golub *et al.*, 1994; David *et al.*, 2007). The final group of patients with FGFR1 related disease commonly present with lymphomatous disease though specific therapy has yet to be developed (Xiao *et al.*, 1998).

While morphological assessment remains important in distinguishing these disorders the progress made in identifying the underlying molecular abnormalities has provided an objective diagnostic tool to aid in diagnosis and classification.

The remainder of this thesis will therefore focus on the diagnosis of those chronic myeloid diseases which remain dependent on subjective morphological assessment namely MDS and MDS/MPN. Reference to the MPDs will be minimal.

1.2.6 The WHO Classification (2017) – an opportunity to integrate objective parameters

The WHO classification underwent a further revision in 2016/2017. The major changes within MDS however were to the terminology rather than the core criteria for diagnosis, with the expanding literature on somatic mutations also being addressed (see section 1.4). To move the emphasis away from cytopenia or the specific type of cytopenia, all subtypes have been renamed with MDS rather than "refractory anaemia/cytopenia" (see Table 1.9). This was on the basis that diagnosis and classification are focused on the degree of dysplasia and blast percentages while cytopenias have little impact (Arber *et al.*, 2016).
The minimal threshold of 10% to define dysplasia was again retained within this revision, and the same 2008 cytogenetic abnormalities remained disease defining (Swerdlow *et al.*, 2017).

Despite the explosion of data gathered from mutation analysis (see section 1.4), del(5q) remains the only cytogenetic or molecular genetic abnormality that defines a specific MDS subtype. Following data published from cohort 1 (Chapter 4) of this research along with other studies (Cargo *et al.*, 2015; Kwok *et al.*, 2015; Steensma *et al.*, 2015), the WHO do acknowledge, in the most recent classification, that mutations are frequent in confirmed disease and also in patients who fail to meet minimal diagnostic criteria though at present these cannot be considered diagnostic (Arber *et al.*, 2016). This is due to reports of the same mutations detected in aging healthy individuals – see Chapters 4 & 6 (Genovese *et al.*, 2014; Jaiswal *et al.*,

2014). The only gene mutation that is incorporated into the most recent classification involves *SF3B1*, a spliceosome gene in which mutations correlate strongly with the presence of ring sideroblasts (Papaemmanuil *et al.*, 2011). In the presence of this mutation, MDS with ring sideroblasts can be diagnosed with as few as 5% ring sideroblasts, in contrast to 15% in previous iterations (Swerdlow *et al.*, 2017). This follows previous studies which confirm that the proportion of ring sideroblasts was irrelevant with regard to prognosis in those without excess blasts (Mrinal M Patnaik, Hanson, *et al.*, 2012).

With respect to MDS/MPN overlaps, RARS-T has been accepted as a full entity and is now termed MDS/MPN with ring sideroblasts and thrombocytosis to align with the updated MDS terminology (Swerdlow *et al.*, 2017). In this subtype the 15% cut-off for ring sideroblasts has been retained. The high frequency of recurrent somatic mutations across the MDS/MPN group was also acknowledged though for the same reasons as in MDS it has been stated that the presence of these should not be used alone as proof of neoplasia (Arber *et al.*, 2016). The strong correlations between mutations in *SRSF2, TET2, ASXL1* and CMML as well as mutations in *SETBP1, ETNK1* and atypical CML have however been highlighted (Swerdlow *et al.*, 2017).

In CMML, the morphological criteria have been refined to reflect the importance of blast percentage on prognosis with 3 blast-based groups now included (see Table 1.10) These are reported to show significant differences in outcome in both dysplastic and proliferative disease (Schuler *et al.*, 2014).

Table 1.9: WHO classification of Myelodysplastic Syndromes 2017 (adapted from Swerdlow et al. 2017)

Disease Entity	Dysplastic Lineages	No. of cytopenias	Ring sideroblasts	BM and PB blasts	Cytogenetics
MDS-SLD	1	1-2	<15%/<5%	BM<5% PB<1%	
MDS-MLD	2-3	1-3	<15%/<5%	BM<5% PB<1%	
MDS-RS					
MDS-RS-SLD	1	1-2	≥15%/≥5%	BM<5% PB<1%	
MDS-RS-MLD	2-3	1-3	≥15%/≥5%	BM<5% PB<1%	
MDS with isolated del(5q)	1-3	1-2	None or any	BM<5% PB<1%	del (5q) alone or with 1 additional abnormality
MDS-EB					
MDS-EB-1	1-3	1-3	None or any	BM 5-9% PB 2-4%	
MDS-EB-2	1-3	1-3	None or any	BM 10-19% PB 5-19% or Auer rods	
MDS-U					
with 1% PB blasts	1-3	1-3	None or any	BM<5% PB=1%	
with SLD and pancytopenia	1	3	None or any	BM<5% PB<1%	
defining cytogenetics	0	1-3	<15%	BM<5% PB<1%	MDS-defining abnormality

Abbreviations: MDS-SLD, myelodysplastic syndrome with single lineage dysplasia; MDS-MLD, myelodysplastic syndrome with multi-lineage dysplasia; RS, ring sideroblasts; EB, excess blasts; MDS-U, myelodysplastic syndrome unclassifiable

Table 1.10: Morphological criteria for CMML 2017 (adapted from Swerdlow et al. 2017)

Classification	PB Blasts	BM Blasts
CMML-0	<2%	<5%, no Auer rods
CMML-1	2-4%	5-9%, no Auer rods
CMML-2	5-19%	10-19%, or Auer rods present

While this revision of the WHO classification had the potential to incorporate objective criteria for diagnosis, this has been hindered by the lack of data and ongoing uncertainty, particularly in those without definitive morphological disease.

1.3 The Promise of New Diagnostic Technologies

It is apparent that a subjective morphological assessment is not an ideal approach to diagnosing these diseases. Extensive research is ongoing into objective measures using both routine techniques and a spectrum of new diagnostic technologies. The most extensive work has involved the role of multiparameter flow cytometry (MFC) as a potential diagnostic tool for MDS (Loken and Wells, 2008). This can be used to support a diagnosis of MDS in the most recent WHO classification, but is yet to form a definitive component of the diagnostic criteria (Swerdlow *et al.*, 2017).

More recently however advances in molecular technology have also seen the introduction of array based techniques to assess gene expression, microRNA expression and to perform whole genome scanning for cytogenetic abnormalities (Steensma and List, 2005; Maciejewski, Tiu and O'Keefe, 2009). The advent of next generation sequencers has also enabled researchers to unearth multiple mutations with both diagnostic and prognostic potential (Mardis, 2008).

The following sections will provide a review of the published literature available surrounding the technologies described and will discuss their potential role in the diagnosis of chronic myeloid malignancies.

1.4 The New Era of DNA Sequencing

The advent of high throughput sequencing has revolutionized the process of DNA sequencing. This was traditionally performed using the method described by Sanger et al. in 1977. More recently however a new generation of sequencers has provided a high throughput method with a significantly greater resolution to this traditional technique.

1.4.1 Sanger Sequencing

The seminal paper by Sanger et al. described a method of sequencing based on the synthesis of complementary DNA using DNA polymerase in the presence of natural deoxynucleotides (dNTPs) and dideoxynucleotides (ddNTPs) which act as chain terminating inhibitors (Sanger, Nicklen and Coulson, 1977). The randomly terminated oligonucleotide chains were then separated on a polyacrylamide gel electrophoresis with the ddNTPs determining the DNA sequence.

This method remains relatively unchanged today, though chain terminators are now labelled with fluorescent dyes and separation occurs using capillary gel electrophoresis (Morozova and Marra, 2008). Despite these technical improvements this method is limited particularly in large genome projects as this requires in vivo amplification of DNA fragments which is both labour intensive and expensive (Morozova and Marra, 2008).

1.4.2 High throughput Sequencing

The development of high throughput sequencers was driven by the high demand for low cost sequencing. Initially there were 3 platforms in widespread use – the Roche 454, Illumina Genome Analyser and the Applied Biosystems SOLiD system. More recently Illumina have monopolized the market with the powerful HiSeq platforms and the benchtop MiSeq option. The research described in this thesis has been performed using both the Roche and Illumina platforms and these are described below.

The Roche 454 was the first next generation sequencer to become commercially available in 2004 (Mardis, 2008). It uses bead amplification by emulsion PCR followed by sequencing using an alternative technology called pyrosequencing (Fig. 1.1). During pyrosequencing, each incorporation of a nucleotide by DNA polymerase naturally releases pyrophosphate. This initiates a series of downstream reactions which result in the production of light by the firefly enzyme luciferase (Mardis, 2008).

In contrast, the Illumina sequencers use bridge amplification of DNA fragments followed by sequencing by synthesis using labelled reversible terminators (Fig. 1.2). In this sequencing method all 4 nucleotides are labelled with a unique fluorescent dye and are then added simultaneously along with DNA polymerase to the DNA cluster fragments. Once incorporated further DNA synthesis is blocked and an imaging step allows for identification of the nucleotides (Mardis, 2008).

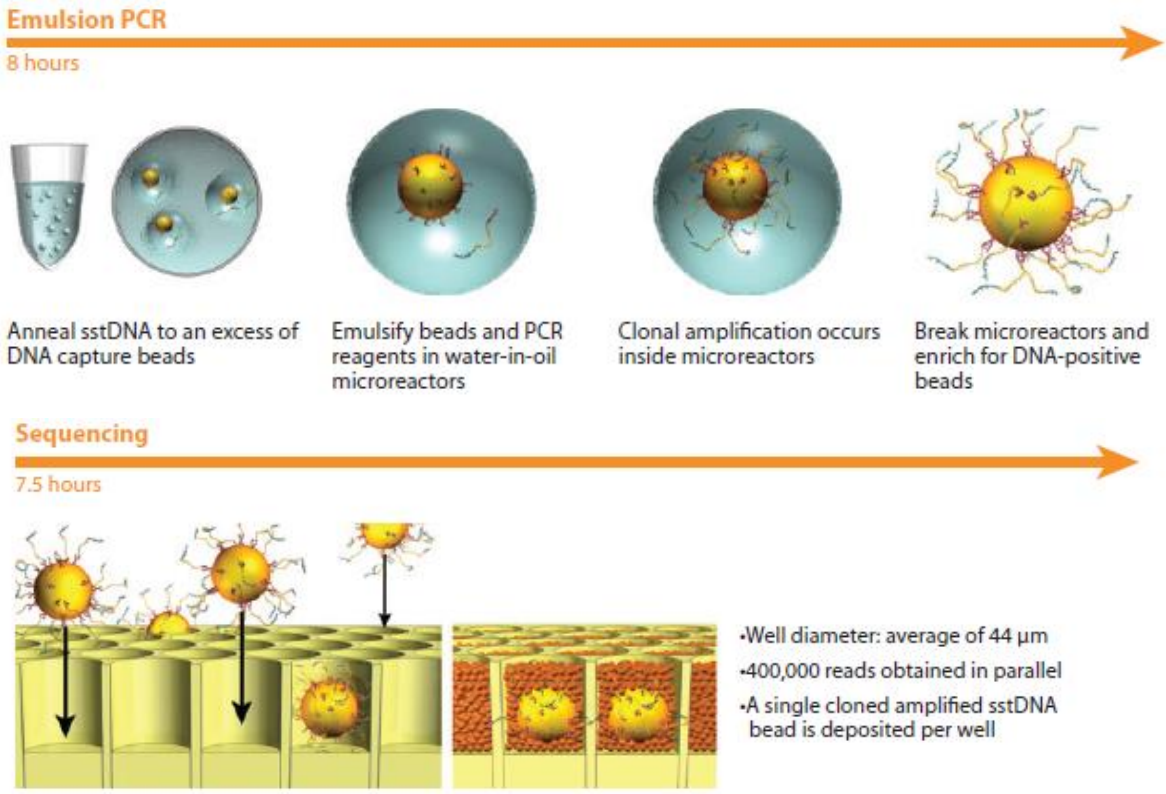


Figure 1.1. The Roche 454 sequencing approach (Mardis, 2008)

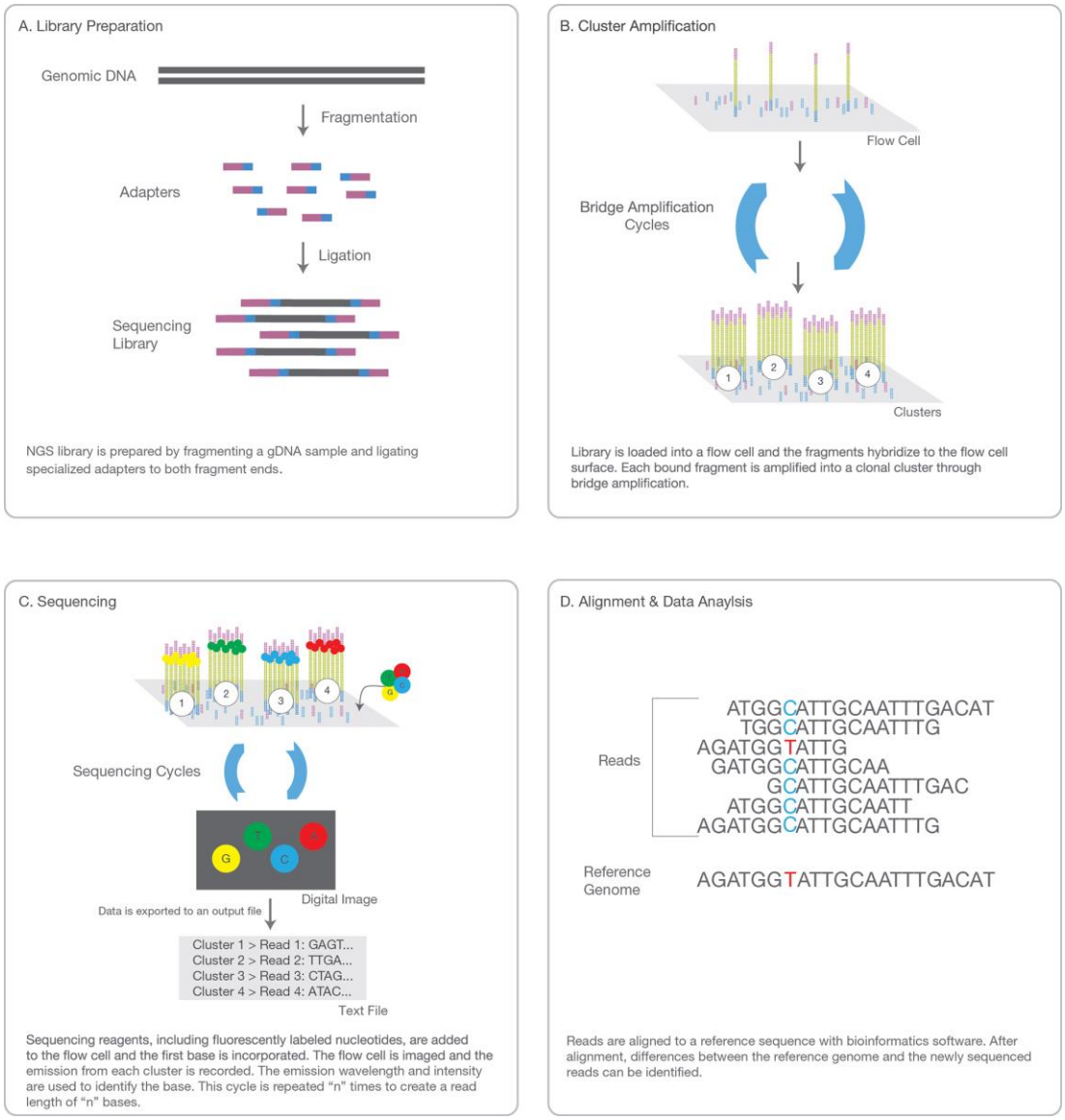


Figure 1.2. The Illumina sequencing by synthesis approach (Courtesy of Illumina, Inc.)

These technologies enable the acquisition of large amounts of sequenced data in much shorter periods of time. This has allowed researchers to sequence both whole genomes and also target large panels of recurrently mutated genes across the spectrum of cancer including myeloid malignancies. This technology has also been utilized for analysing gene expression, sequence variation (e.g. SNPs) and identifying small noncoding RNAs (Morozova and Marra, 2008).

1.4.3 Somatic Mutations are common in Myeloid Malignancies

Using both techniques described above, somatic mutations have been identified in a large number of patients with chronic myeloid malignancies.

Initial studies, using Sanger sequencing, either targeted known tumour suppressor genes or those involved in other haematological malignancies. The first mutations identified were in in *TP53*, a critical cell-cycle checkpoint regulator (Jonveaux *et al.*, 1991; Sugimoto *et al.*, 1993), which were shown to be present in up to 15% of MDS patients and reported to be associated with a worse prognosis (Kita-Sasai *et al.*, 2001a). Subsequent studies identified mutations in the RAS proto-oncogene family which encodes guanosine triphosphate hydrolases (GTPase), which are regulators of cellular growth related signals (Steensma and List, 2005). In myeloid malignancies *NRAS* mutations predominate, though overall these mutations are more common in CMML (40-60%) (Steensma and List, 2005). *RUNX1* (previously AML1), the transcription factor essential for normal haematopoiesis is the target of a number of translocations in acute leukaemia. Mutations were identified initially in a small number of patients with MDS (Imai *et al.*, 2000), however subsequent studies reported mutations in up to 25% of patients with RAEB and RAEBt and also more common in those with therapy related disease (Christiansen, Andersen and Pedersen-Bjergaard, 2004; Harada *et al.*, 2004). The presence of recurrent mutations of *FLT3* and nucleophosmin (*NPM1*) in AML led researchers to assess the presence of these mutations in MDS. Mutations were identified in only a small number of patients (Horiike *et al.*, 1997; Zhang *et al.*, 2007) though *FLT3* mutations were associated with a poorer prognosis and progression to AML (L. Shih *et al.*, 2004).

The introduction of high throughput sequencing allowed an unbiased approach to identifying mutations by sequencing the whole genome. Two novel mutations were identified using this method on 2 separate patients with AML (Mardis *et al.*, 2009; Ley *et al.*, 2010). The first mutation was isocitrate dehydrogenase 1 (*IDH1*), a tumour suppressor gene which had been previously reported in malignant gliomas and the other was DNA methyltransferase gene (*DNMT3A*) which encodes DNA

methyltranferases involved in methylation (Mardis *et al.*, 2009; Ley *et al.*, 2010). Both of these mutations were shown to have a negative prognostic impact in AML, and a small number of reports have identified these mutations in MDS. While they are present in only a small number of patients (*IDH1* 3.6%, *DNMT3A* 2.6-8%), both have been shown to be associated with a poorer prognosis and higher rate of transformation to AML (F Thol *et al.*, 2010; F Thol, Winschel, *et al.*, 2011; Walter *et al.*, 2011).

There has since been an explosion of research in this area and mutations have now been identified in a number of key pathways implicated in myeloid disease pathogenesis.

1.4.4 High throughput sequencing identifies key pathways in disease pathogenesis

High throughput sequencing has provided researchers with the ability to explore the genomes of large numbers of cancer patients. This has led to the discovery of multiple recurrently mutated genes and the functional pathways involved in disease pathogenesis. With respect to myeloid malignancies, and in particular MDS and MDS/MPN, the key pathways and genes involved are described below and include RNA splicing, epigenetic regulators, cell signalling, transcription factors, the cohesin complex, and tumour suppressor genes.

1.4.4.1 RNA Splicing

Genes involved in RNA splicing, most commonly *SF3B1, SRSF2, U2AF1* and *ZRSR2*, are the most frequently mutated genes in MDS. Splicing is critical for successful transcription and is the process by which introns are excised from pre-mRNA, enabling exon ligation and the formation of mature messenger RNA (Shukla and Singh, 2017). This process is essential for protein diversity as a number of alternate isoforms can be generated from a single pre-RNA transcript (Shukla and Singh, 2017). The splicing factors involved in myeloid malignancies form part of the E/A complex which coordinates 3' splice site recognition. Mutations have been shown to cause aberrant 3' splice site recognition and the generation of aberrantly spliced

mRNA transcripts (Armstrong *et al.*, 2018). Initial sequencing studies of MDS patients identified these mutations at high frequency and showed mutations to be both mutually exclusive and associated with disease phenotype (Kenichi Yoshida *et al.*, 2011; Papaemmanuil *et al.*, 2011). *SF3B1* mutations, in particular, correlated strongly with the presence of ring sideroblasts (Kenichi Yoshida *et al.*, 2011; Papaemmanuil *et al.*, 2011) while *SRSF2* or *ZRSR2* in combination with *TET2* is highly specific for a myelomonocytic phenotype (Malcovati *et al.*, 2014). Importantly, mutations also correlated with the presence of dysplasia, being infrequent in both de novo AML and myeloproliferative neoplasms (Kenichi Yoshida *et al.*, 2011). Research is ongoing to understand the link between mutations and disease pathogenesis though aberrant splicing of ABCB7, a mitochondrial iron exporter, by mutant *SF3B1* is thought to lead to the accumulation of mitochondrial iron seen in MDS with ring sideroblasts (Dolatshad *et al.*, 2016). Furthermore, mutations in *SRSF2* alter this genes preference for specific exonic splicing enhancer motifs leading to mis-splicing of key haemopoietic regulators (Kim *et al.*, 2015).

SF3B1 is one of the few mutations in MDS associated with a good prognosis. A study using unsupervised hierarchical clustering incorporating somatic mutations, identified MDS with *SF3B1* mutation as a distinct entity in patients with <5% blasts irrespective of WHO morphological classification (Malcovati *et al.*, 2014). These patients had a favourable prognosis which held true whether the mutation was clonal or subclonal (Malcovati *et al.*, 2014). A subsequent study of MDS patients with >1% ring sideroblasts, confirmed these findings with the presence of an *SF3B1* mutation associated with a significantly improved overall survival and lower risk of disease progression (Malcovati *et al.*, 2015). This was again restricted to those patients without an excess of blasts. *SF3B1* has a limited pattern of co-occurring mutations, mainly involving epigenetic regulators and *RUNX1*, however only the latter has a negative impact on prognosis within this group (Malcovati *et al.*, 2014).
In contrast, mutations in other spliceosome mutations have all been reported to be associated with a shorter overall survival and increased risk of progression to AML (Damm, Kosmider, Gelsi-Boyer, Renneville, Carbuccia, Hidalgo-Curtis, Della Valle, *et*

al., 2012; Graubert *et al.*, 2012; Makishima *et al.*, 2012a; Thol *et al.*, 2012; Wu *et al.*, 2012).

1.4.4.2 Epigenetic Regulators

Mutations in genes involved in DNA methylation were the first genes implicated in myeloid pathogenesis in the era of novel sequencing techniques. *TET2* mutations were first reported in 2009 by studies focused on identifying the underlying mutation associated with uniparental disomy (UPD) of chromosome 4q (Delhommeau *et al.*, 2009; Mohamedali *et al.*, 2009). *TET2* catalyzes the conversion of 5-methylcytosine to 5-hydroxymethylcytosine and depletion leads to skewing towards monocyte/granulocytic lineages and myeloid tumourigenesis (Ko *et al.*, 2010; Z. Li *et al.*, 2011). Soon after, mutations in 2 further genes in this pathway were identified by whole genome sequencing of de-novo AML cases (Mardis *et al.*, 2009; Ley *et al.*, 2010). This included *DNMT3A*, a DNA methyltransferase involved in de novo methylation (Ley *et al.*, 2010), and *IDH1*, which along with *IDH2* catalyses the conversion of isocitrate to α-ketoglutarate which itself regulates *TET2* (Figueroa *et al.*, 2010). Subsequent studies confirmed the presence of these mutations in MDS patients, albeit at a lower frequency (F Thol *et al.*, 2010; Kosmider *et al.*, 2010; Walter *et al.*, 2011).

Abnormal histone modification is also thought to contribute to epigenetic deregulation in MDS. Mutations in the histone modifiers *ASXL1*, a histone-binding protein and *EZH2,* a histone methyltransferase, are both reported with varying frequency (Gelsi-Boyer *et al.*, 2009; Nikoloski *et al.*, 2010a).

While *TET2* is one of the most commonly mutated genes in MDS, it appears to have no specific prognostic relevance. Initial studies suggested it may confer a favourable prognosis however this has not been confirmed in subsequent studies and a large meta-analysis showed no significant prognostic value (Kosmider *et al.*, 2009; Smith *et al.*, 2010; Lin *et al.*, 2017). Mutations in other epigenetic regulators do however appear to have prognostic relevance. *DNMT3A* has been shown to be

associated with worse overall survival and rapid progression to AML (F Thol, Winschel, *et al.*, 2011; Walter *et al.*, 2011). However, this gene fails to hold prognostic significance in MDS with ring sideroblasts and an *SF3B1* mutation, meaning the clinical context in which it occurs is important (Malcovati *et al.*, 2015).

By far, the genes that have the most clinical relevance in this group are the chromatin modifiers. *ASXL1* has been shown to be a poor prognostic feature across all myeloid malignancies including MDS (Gelsi-Boyer *et al.*, 2012). *ASXL1* mutations, in particular frameshift mutations, were independent prognostic markers with respect to survival and AML transformation (Felicitas Thol, Friesen, *et al.*, 2011). This was confirmed in patients with low risk MDS and was found to be one of the 5 poor prognostic mutations in initial large sequencing studies (Bejar *et al.*, 2011, 2012). Similarly, *EZH2* was also been shown to be associated with poor prognosis in these studies. Importantly mutations in this gene were the only one to retain prognostic significance in a multivariate model in low risk MDS (Bejar *et al.*, 2011, 2012).

1.4.4.3 Cell signalling and transcription factors

Mutations in the RAS proto-oncogene family have long been reported in myeloid malignancies. RAS genes encode guanosine triphosphate hydrolases (GTPase) which are regulators of cellular growth related signals (Steensma and List, 2005). In myeloid malignancies *NRAS* mutations predominate, though overall these mutations are more common in CMML rather than MDS (Al-Kali *et al.*, 2013).
CBL gives rise to cbl protein which targets a variety of tyrosine kinases for ubiquitination. Again, firstly reported in AML, mutations in this gene are most commonly reported in CMML but also in MDS (Bacher *et al.*, 2010).

Mutations in tyrosine kinase genes are more common in myeloproliferative mutations and AML, however they do occur at low frequencies in MDS. This includes *JAK2* mutations which, while are more common in MDS/MPN overlaps, also rarely occur in MDS (Ceesay *et al.*, 2006).

RUNX1 (previously AML1), the transcription factor essential for normal haematopoiesis is the target of a number of translocations in acute leukaemia (Harada *et al.*, 2004). Point mutations are also somatically acquired in MDS (Harada *et al.*, 2004) as well as being associated with familial myeloid neoplasms when inherited in the germline (Galera, Dulau-Florea and Calvo, 2019). *BCOR* (BCL6 corepressor) is a key transcriptional regulator of haemopoiesis and was first reported to be mutated in cytogenetically normal AML (Vera Grossmann *et al.*, 2011). Subsequent analysis of MDS showed mutations occur at a similar frequency in this patient group (Damm *et al.*, 2013).

Mutations in *TP53*, a critical cell-cycle checkpoint regulator are reported across the spectrum of haematological and non-haematological cancers (Wang and Sun, 2017). With respect to MDS they are associated with complex karyotype and worse prognosis (Kita-Sasai *et al.*, 2001a).

Transcription factors and *TP53* have been associated with poor outcome in MDS. *RUNX1* mutations have long been associated with a poor prognosis, particularly in high risk MDS and AML (Harada *et al.*, 2004). When looking at larger numbers of genes, this mutation was again shown to be a predictor of poor survival even in low risk MDS (Bejar *et al.*, 2011, 2012). Similarly *BCOR* has also been associated with a significantly inferior overall survival (Damm *et al.*, 2013).

1.4.4.4 COHESIN COMPLEX

One of the most recent pathways implicated in myeloid pathogenesis is the cohesin complex. This is involved in the cohesion of sister chromatids, regulating transcription and is involved in post-transcriptional DNA repair (Kon *et al.*, 2013). *STAG2* is by far the most frequently mutated gene within this complex, other genes include *RAD21, SMC3, SMC1A*. Mutations within this pathway are found most frequently in high-risk MDS and secondary AML (Thota *et al.*, 2014) and recent

studies have shown that mutant cohesin proteins block differentiation of haematopoietic stem and progenitor cells (Mazumdar *et al.*, 2015).

Mutations in this pathway, are most prevalent in high-risk MDS and secondary AML and are associated with poor overall survival (Thota *et al.*, 2014).

1.5 Array Based Whole Genome Scanning

Karyotyping is an important tool for both the diagnosis and identification of prognostic markers across the spectrum of myeloid malignancies. With relatively low sensitivity and variable resolution, the detection of abnormalities using time consuming labour intensive conventional cytogenetics (CC) is limited (Maciejewski, Tiu and O'Keefe, 2009). Array based technologies have been developed which have many advantages over this traditional technique. One particular benefit is the ability to screen for new lesions without the prerequisite of cells in metaphase (Maciejewski, Tiu and O'Keefe, 2009). Initially this was performed using comparative genomic hybridization arrays (CGH-A), however the development of SNP-A heralded a turning point in the use of array based techniques; this method having the advantage of being able to detect copy number neutral loss of heterozygosity or somatic uniparental disomy (UPD) (Maciejewski, Tiu and O'Keefe, 2009). Further studies in myeloid malignancy focused on this technique with only a handful using CGH-A for analysis.

1.5.1 Single Nucleotide Polymorphism Arrays

SNP-A are based on oligonucleotide probes which correspond to allelic variants of SNPs (Maciejewski, Tiu and O'Keefe, 2009). Hybridisation of DNA to the probes results in either signals for both alleles consistent with heterozygosity or only one signal in the case of hemi/homozygosity. The strength of the fluorescent signal also allows for determination of copy number (Maciejewski, Tiu and O'Keefe, 2009) (Fig. 1.3).

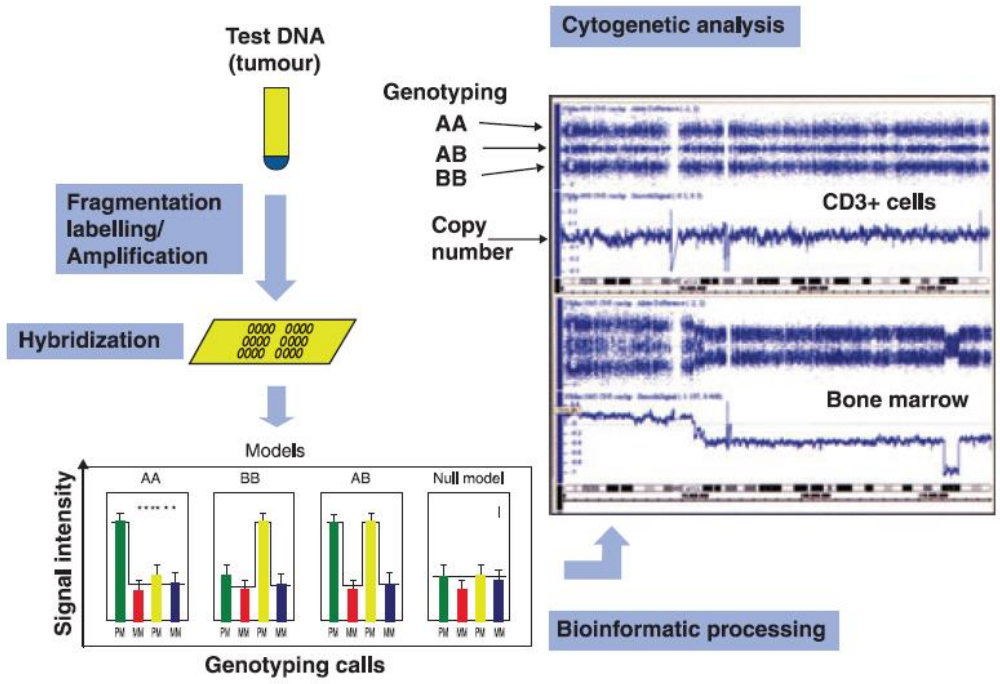


Figure 1.3. The Principle of SNP-A (Maciejewski et al, 2009)

A number of platforms are commercially available allowing the analysis of over 900,000 loci simultaneously enabling increased precision (Heinrichs, Li and Look, 2010).

The major advantage of this approach is the ability to detect diploid stretches of homozygosity. These can be due to acquired somatic UPD, autozygosity or early embryonic UPD (Maciejewski, Tiu and O'Keefe, 2009). With respect to malignancy, UPD results from errors during mitosis resulting in both copies of a chromosome or part of a chromosome being derived from one parent (Maciejewski, Tiu and O'Keefe, 2009). The loss of heterozygosity in UPD can be distinguished from deletions as the former occurs in the presence of a diploid chromosome while the latter results in loss of DNA copy number (Gondek, Dunbar, *et al.*, 2007). This important mechanism leading to point mutations or microlesions can be identified by this method (Heinrichs, Li and Look, 2010). One such example of this is the *JAK2* V617F gain of function mutation in myeloproliferative disorders. Initial genome wide analysis identified UPD at chromosome 9p as a recurring abnormality in this patient group (Kralovics, Guan and Prchal, 2002). This group went on to confirm

that UPD9p leads to homozygosity of the V617F mutation in one of the first papers to report on this mutation (R Kralovics *et al.*, 2005). Subsequent studies in AML identified homozygous mutations associated with UPD in genes known to be mutational targets – *WT1, FLT3, CEBPA* and *RUNX1* (Fitzgibbon *et al.*, 2005).

1.5.2.1 SNP-A has Increased Resolution with Clinical Impact

SNP-A are much more precise than CC, though have a comparable sensitivity with the minimal detectable clone size being 25-50% of total cells (Gondek *et al.*, 2008).

An early report by Gondek et al. using 50K SNP-A analysis confirmed the presence of the majority of abnormalities identified by CC in MDS patients, and discovered new defects including UPD (Gondek, Tiu, *et al.*, 2007). Abnormalities by SNP-A were present in 69% of patients with normal cytogenetics and only 12/66 (18%) patients had a normal karyotype by this technique (vs. 39% by cytogenetics, $p<0.001$). In addition, a higher proportion of patients had multiple defects which could place a significantly greater percentage of patients in the complex cytogenetic category (44 vs 6%, $p<0.001$). UPD was detected in 33% of patients and frequently involved areas affected by traditional cytogenetic deletions including 7q and 11q (Gondek, Tiu, *et al.*, 2007). Subsequent clinical correlation showed that patients with chromosome 7 lesions by SNP-A had worse survival than those with normal karyotype ($p=0.047$), and the survival of patients with SNP detected lesions was similar to 7/7q lesions by CC. In addition those patients with multiple lesions had more advanced disease, though comment on survival in this group was not made (Gondek, Haddad, *et al.*, 2007).

Mohamedali et al. showed that a higher resolution array, 250K, added additional information to that identified on the 50K array and used this to analyse a group of low-risk MDS patients (Mohamedali *et al.*, 2007). UPD was identified in 46% of MDS patients overall and copy number changes not detected by CC were identified in 18% of patients. UPD on chromosome 4q was identified in 9 patients, and GEP showed up-regulation of erythroid specific genes in these patients, though no

overtly downregulated genes in this region. Clinical follow-up of this patient group showed no correlation between UPD or copy number change and disease progression, though copy number change was correlated with a higher IPSS score and overall survival (Mohamedali *et al.*, 2007).

The Cleveland Clinic group proceeded to use the 250K array to analyse a larger group of patients with a spectrum of myeloid malignancies (Gondek *et al.*, 2008). Again, a higher percentage of chromosomal abnormalities were detected by SNP-A than by CC (78% vs. 59% MDS, 75% vs. 37% MDS/MPN, 77% vs. 53% AML). UPD was detected in 20%, 35% and 23% of MDS, MDS/MPN and AML patients respectively. When analyzing the clinical impact of these findings, those with normal karyotype by both CC and SNP-A had an improved OS compared with those with normal CC in whom additional lesions were identified by SNP-A (39 vs. 16 months, p=0.02). This survival impact was maintained within both the MDS/MPN and AML subgroups, but was not significantly affected in the MDS group; which may reflect a shorter follow-up. In addition, specific lesions also conferred a worse survival; patients with a new cryptic lesion on chromosome 7 had poorer prognosis than those with known deletions of 7/7q, and significantly worse outcomes than those with normal SNP-A (Gondek *et al.*, 2008).

The MD Anderson group prospectively analysed 51 newly diagnosed MDS patients, with the added benefit of matched normal DNA in all patients (Heinrichs *et al.*, 2009). In those patients with a normal karyotype 12% (4/33) had clonal regions of acquired UPD involving 3q, 4q, 7q and 17p. Unfortunately, follow-up was short, and clinical validation of these findings was not possible.

In a larger study by Tiu et al the combination of CC and SNP-A increased the detection of abnormalities from 44% to 74% (Tiu *et al.*, 2011). Of those with normal cytogenetics 130 out of 241 patients (54%) had cryptic abnormalities, and additional aberrations were identified in 117 out of 189 (62%) with an abnormal karyotype by CC. Overall survival was shown to be worse in those with either CC or SNP-A defects compared to those without (16 vs. 43 months, p<0.001). Regardless

of the karyotype by CC, the addition of abnormalities by SNP-A showed a significantly reduced OS, PFS and EFS, with both new and increased lesions by SNP-A being independent predictors of OS and EFS in a multivariate analysis. Deletions and UPD of chromosomes 7, 11 and 17 were associated with poor OS, similar to that if detected by CC; and the authors suggested that these abnormalities should be included in higher risk cytogenetic groups (Tiu *et al.*, 2011).

Similarly, in patients with CMML, a high frequency of abnormalities have been reported in patients with low risk cytogenetic features (normal karyotype or isolated -Y) or no metaphases at diagnosis (Palomo, Xicoy, *et al.*, 2016). Out of 128 patients, 86 (67%) had an abnormality reported, with a size >11Mb being associated with a shorter overall survival on univariate analysis. This was not, however, maintained on a multivariate analysis; SNP-A having no overall impact on outcome (Palomo, Xicoy, *et al.*, 2016).

1.5.2.2 SNP-A can Reveal Mechanisms of Disease Pathogenesis

SNP-A have also identified important mechanisms in disease pathogenesis, including the origin of MDS clones. By performing SNP-A analysis on fractionated cells one group determined that MDS clones were selected during differentiation and suggested some clones may arise from lineage committed progenitors (Huang *et al.*, 2009).

Investigation of regions of UPD and cryptic deletions has revealed potential candidates for disease pathogenesis including adenosine deaminase (ADA) and calreticulin (*CALR*) in MDS which showed significant differential gene expression by GEP (Nowak *et al.*, 2009). A further group discovered a mutation in *TET2* located at 4q24 following investigation of recurrent UPD and microdeletion of this region (Langemeijer *et al.*, 2009). Subsequent analysis confirmed this mutation was present in 26% (27/102) of MDS patients in their cohort. *TET2* has a proposed role in myeloid differentiation (Langemeijer *et al.*, 2009).

In a study by Dunbar et al, UPD was shown to occur more frequently in those with CMML (48%, 21/44) and MDS/MPN-u (38%, 14/37) (Dunbar *et al.*, 2008). One of the most common chromosome arms affected was 11q, which was involved in 12/301 patients. Further analysis of this region identified c-*CBL*, a gene which encodes a ubiquitin ligase involved in ubiquitylation and degradation of active protein tyrosine kinase receptors, mutations of which have been found in AML (Dunbar *et al.*, 2008). Direct sequencing identified 3 unique missense mutations in 7 out of 12 patients within this gene, including 2 new mutations not previously reported (Dunbar *et al.*, 2008).

Ernst et al. analysed a cohort of patients with MDS/MPN (n=148) and identified 13 who had UPD7q (T Ernst *et al.*, 2010). A homozygous mutation of one of these genes, *EZH2*, was shown to be present in 9 of those patients with UPD7q. Analysis of a larger cohort identified 49 variants of inactivating mutations including missense and frameshift mutations. *EZH2* forms part of the polycomb repressive complex 2, which is a highly conserved histone H3 lysine 27 (H3K27) methyltransferase which influences stem cell renewal (V Grossmann, Kohlmann, Eder, Haferlach, Kern, N. Cross, *et al.*, 2011). These findings suggest this gene is a tumour suppressor for myelopoiesis. Correlation with clinical outcome showed that mutations were associated with a poorer outcome in the MDS/MPN subgroup (T Ernst *et al.*, 2010). These findings were corroborated by subsequent studies (Makishima, Jankowska, *et al.*, 2010; Nikoloski *et al.*, 2010b)

Analysis of a small group of patients with RARS-T also revealed that UPD1p was associated with MPL W515L in a proportion of cases analogous with UPD9p and the *JAK2* V617F mutation (Szpurka *et al.*, 2009). This cohort was further studied and cryptic abnormalities were identified in 13/22 patients including deletions of 2p and 5q and UPD of 1p, 2p, 3q, 6p, 8p and 10p (Szpurka *et al.*, 2010).

1.5.2.3 Regions of UPD are Acquired During Transformation to AML

Paired analysis of a small number of patients who transformed from MDS to AML identified acquisition UPD during disease progression (Flach *et al.*, 2011). Thirty two percent of patients (11/34) had UPD during the MDS phase, however, one acquired a UPD(21q) and another acquired both UPD(17q) and UPD(19q) at the time of transformation. The acquisition of UPD(21q) was accompanied by a homozygous *RUNX1* mutation. Unfortunately this group did not determine if any of the original UPD were germline (Flach *et al.*, 2011).

1.5.2.4 Pitfalls of SNP-A

The major disadvantage of SNP-A is the inability to detect balanced translocations and inversions. In addition this technique has a lower sensitivity for smaller clones (Gondek, Tiu, *et al.*, 2007). As a result, not all abnormalities identified by CC are also detected with SNP-A (Gondek, Tiu, *et al.*, 2007; Gondek *et al.*, 2008). Makishima et al analysed the combination of CC, fluorescence in situ hybridization (FISH) and SNP-A to detect specific cytogenetic lesions, namely del(5q), monosomy 7, del(7q), trisomy 8, and del(20q) (Makishima, Rataul, *et al.*, 2010). Interestingly no single method identified all defects, with a combination of all 3 methods detecting the highest rates (5% increase in diagnostic yield). Disparity was noted between samples used for each technique, and it was reiterated that the sensitivity of SNP-A remains low. The authors stated however that this may not be a clinical issue, as smaller clones may not be clinically relevant (Makishima, Rataul, *et al.*, 2010).

A further disadvantage is the need to consider germline abnormalities. Mohamedali et al demonstrated a high incidence of constitutional UPD in their group of patients (12 out of the 13 patients tested) while most of the copy number changes were present at low frequency in the general population (Mohamedali *et al.*, 2007). Most early studies, including this one, failed to perform germline confirmation of all their findings. Heinrichs et al performed germline conformation in all samples, and found that only 3/31 copy number changes were true

microdeletions, and only 4/110 areas of LOH were true UPD (Heinrichs *et al.*, 2009). The absence of matched DNA can therefore lead to a vast overestimation of abnormalities, and it is essential to correlate findings with germline databases.

1.6 Flow Cytometry

MFC is an essential tool in the diagnosis of most haematological malignancies, particularly mature lymphoid malignancies and acute leukaemias. Extensive research has been performed to assess its utility in the diagnosis of chronic myeloid malignancies, most specifically MDS. While a number of laboratories currently employ this method in the diagnosis of MDS, it is yet to be a core criterion in the internationally recognized diagnostic criteria. This chapter will outline the history of this technique in the chronic myeloid malignancies and the evidence to date for its use in diagnosis.

1.6.1 What is MFC?

MFC is the process by which multiple objective measurements of single cells occur as they pass through the measuring apparatus in a fluid stream (Shapiro, 1988). The origins of this technique date back to the 1930s when primitive versions of a flow cytometer were used to count aerosol particles for analysis of mine dust, and used for the detection of bacteria and spores during World War II (Shapiro, 1988).

The principle of the technique relies on the measurement of scattered light and fluorescence from individual cells in the population (Watson, 1991). Fluorescently labelled antibodies are applied to specific markers on the cell population of interest. The stained cells are then passed, in single file in fluid suspension, through a high intensity light source (Watson, 1991). Light detectors are present in line with, and perpendicular to, the light beam allowing detection of both forward and side scatter. These determine both the size and complexity of the cell. Each fluorochrome has a characteristic peak excitation and emission wavelength allowing distinction of these labels by fluorescence detectors. The light flash is

converted to an electronic signal which is digitalized by an analogue-to-digital converter. The signal is then stored electronically (Watson, 1991).

1.6.2 Flow cytometry in MDS

1.6.2.1 Early Studies Identify Aberrant Phenotypes in MDS

Initial studies demonstrated the increased and reduced expression of antigens normally identified on myeloid cells. This included increased expression of CD34, HLA-DR, CD13 and CD33 on myeloid precursors (Baumann *et al.*, 1986; Hokland *et al.*, 1986; Kristensen and Hokland, 1990; Mittelman *et al.*, 1993; Fuchigami *et al.*, 2000; Karmon, Manaster and Chezar, 2002) and reduced CD11b, CD43 and CD10 expression on neutrophils (Mittelman *et al.*, 1993; Chang and Cleveland, 2000; Kyriakou *et al.*, 2001). The increased CD34 expression was later shown to directly correlate with the morphological blast percentage, and the blast phenotype in MDS was confirmed to be a committed myeloid precursor in the majority of cases (Maynadié *et al.*, 2002; Ogata *et al.*, 2002; Del Cañizo *et al.*, 2003).

Aberrant phenotypes have been widely reported in the literature, including asynchronous expression of immature and maturing markers and inappropriate expression of lymphoid markers such as CD3, CD7 and CD56 (Schlesinger *et al.*, 1996; Hansen, Meyer and Hokland, 1998; Ogata *et al.*, 2002; Del Cañizo *et al.*, 2003). Certain abnormal phenotypes were also shown to have prognostic significance, with high HLA-DR/low CD11b and CD7 expression being associated with a poor prognosis (Mittelman *et al.*, 1993; Ogata *et al.*, 2002).

1.6.2.2 Multiple Methods Are Proposed For Diagnosing MDS

In light of the above findings, the use of MFC as a diagnostic tool was considered. A number of different methods have been reported in the literature though are all based on the concept of analysing the expression of several antigens simultaneously. These include pattern recognition methods (Stetler-Stevenson *et al.*, 2001; Kussick *et al.*, 2005; Stachurski *et al.*, 2008; Truong *et al.*, 2009; Kern *et al.*,

2010), flow cytometric scoring systems (Wells *et al.*, 2003; Ogata *et al.*, 2006, 2009; Matarraz *et al.*, 2008; Satoh *et al.*, 2008; van de Loosdrecht *et al.*, 2008; Goardon *et al.*, 2009; Chu *et al.*, 2011; Cutler *et al.*, 2011) and classification functions (Malcovati *et al.*, 2005; Della Porta *et al.*, 2006), all of which show varying degrees of sensitivity and specificity for detecting MDS.

1.6.2.3 Standardization and Working Group Guidelines

With the expanding evidence of the diagnostic utility of MFC this technology was introduced as a co-criterion for the diagnosis of MDS following a working conference on MDS in 2006 (Valent *et al.*, 2007). The 2008 WHO classification of MDS also included MFC as an adjunct to diagnosis, with 3 or more aberrant features being highly suggestive of MDS (Swerdlow *et al.*, 2008). The 2017 WHO classification makes a similar statement, highlighting that flow cytometry findings alone are not sufficient to establish a diagnosis, pointing the reader to the consensus guidelines produced by the European LeukaemiaNet (ELN) MDS working group (Swerdlow *et al.*, 2017).

In 2008 a European LeukemiaNet workshop was held, with representatives from 18 European institutes, in an attempt to develop a much needed standardized approach (van de Loosdrecht *et al.*, 2009). Guidance was provided on sample type, red cell lysis techniques and antibody staining, and a list of recommended markers was provided. The group also listed those aberrancies regarded as most relevant to diagnosis (Table 1.11) (van de Loosdrecht *et al.*, 2009). This group however failed to determine the optimum approach to utilising this technique to diagnose MDS.

Further reports from this working group proposed a minimum consensus panel to analyse dysplasia, with the recommended core markers being sufficient to categorise cytopenic patients into normal, suggestive, or diagnostic of MDS (Westers *et al.*, 2012). This group went on to propose guidelines for the integration of MFC into the WHO classification (Porwit *et al.*, 2014) suggesting a mini-panel which can be used for screening purposes based on the 'Ogata score'. This had

previously been validated in a large multicentre study with moderate sensitivity (70%) and high specificity (93%) for diagnosing MDS (Della Porta *et al.*, 2012). A comprehensive panel was then suggested using the markers stated in Table 1.12.

Table 1.11. Relevant aberrancies in the blast, maturing myeloid, monocytic and erythroid lineages as recommended by the European LeukaemiaNet working conference *(adapted from van de Loosdrecht, 2009)*

Blasts	Maturing Myeloid	Monocytic	Erythroid
Increased percentage	Abnormal SSC	Decreased or increased proportion	Abnormal CD71/CD235a pattern
Abnormal granularity	Increased or decreased expression of CD45, CD13, CD33, CD11b, CD16, CD64	Abnormal expression of CD13 and CD33	
Abnormal intensity of CD45, CD34 and CD117	Lack of CD10	Abnormal CD116/ HLA-DR pattern	
Expression of CD11b or CD15	Expression of CD34 and CD14	Abnormal intensity of CD14, CD36, CD64	
Abnormal intensity of HLA-DR	Expression of lineage infidelity markers	Overexpression of CD56 (>1 log)	
Expression of lineage infidelity markers		Expression of lineage infidelity markers	

Table 1.12: Recommended panel of markers for the diagnostic work-up in MDS

CD45	**CD117**	**CD34**	**HLA-DR**	**CD11b**	**CD16**
CD13	**CD33**	**CD14**	**CD15**	**CD10**	**CD19**
CD36	**CD5**	**CD56**	**CD7**	**CD71**	**CD235**

Aberrant findings in at least 3 tested features comprising at least 2 cell compartments was the recommended definition to determine aberrant results (Porwit *et al.*, 2014). This statement has since been adopted in the most recent WHO classification, though flow cytometry does not form part of the core criteria for diagnosis (Swerdlow *et al.*, 2017).

1.6.3 Flow cytometry in MDS/MPN

The use of MFC within this group of disorders has been limited to CMML. Occasional studies have included MDS/MPN within the MDS subgroup, though extrapolation of data specific for subtypes other than CMML is difficult.

1.6.3.1 MFC Can Facilitate the Diagnosis of CMML

In the early studies of its use in MDS, CMML was categorized by the FAB classification as a myelodysplastic disorder and encompassed within this group. More recently, a small number of groups have identified aberrant phenotypes within the monocyte component which can aid in the distinction between reactive monocytosis, CMML, and acute monocytic leukaemia.

Studies have identified recurrent aberrancies in the monocytic component of CMML, including partial loss of CD13, CD14, CD15, HLA-DR and CD36 along with variable expression of CD56 and dim expression of CD2 (Dunphy, Orton and Mantell, 2004; Xu *et al.*, 2005; Lacronique-Gazaille *et al.*, 2007; Subirá *et al.*, 2008).

Comparison with reactive monocytosis and normal bone marrows showed that the combination of CD56 expression and reduced expression of myeloid antigens was specific for CMML, however it only had a sensitivity of 40%. Sensitivity was improved by including the presence of 20% or more immature monocytes, which in the presence of 2 additional aberrancies had 100% specificity and 67% sensitivity for CMML (Xu *et al.*, 2005). The use of CD56 alone has been suggested as a simple

method to detect CMML, though overexpression of this marker in reactive conditions limits this method (Lacronique-Gazaille *et al.*, 2007).

Most recently screening of peripheral blood monocyte subsets has been proposed as a screening tool to distinguish CMML from reactive monocytosis (Selimoglu-Buet *et al.*, 2015). By analyzing peripheral blood on patients with CMML, reactive monocytosis, other haematological malignancies and healthy donors it was confirmed that CMML patients have a characteristic increase in classical monocytes (CD14+/CD16-) with a cut-off value of 94%. This was both highly sensitive and specific (90.6% and 95.1%) for CMML (Selimoglu-Buet *et al.*, 2015) Subsequent studies have validated this method, and also confirmed the ability to distinguish CMML from MDS, as well as MPN presenting with a peripheral blood monocytosis (Patnaik *et al.*, 2017; Talati *et al.*, 2017).

Traditionally, it has also been difficult to distinguish CMML from acute myelomonocytic leukaemia due to the challenging morphology in these cases (Goasguen *et al.*, 2009). The use of 2 different anti-CD14 antibodies, which recognize the epitopes MO2 and MY4, has been reported to distinguish monoblasts, promonocytes and mature monocytes by their expression at different stages of maturation, providing a possible adjunct to the morphological assessment of this lineage (Yang *et al.*, 2005). A study from a German group compared 27 cases of acute monocytic/monoblastic leukaemia with 138 cases of CMML (Kern *et al.*, 2011). They identified CD56 aberrant expression in both diseases, however additional aberrancies were noted in CMML patients including CD2 co-expression and lack of both CD13 and HLA-DR. There was also a significantly greater percentage of granulocytic cells in the CMML cohort (Kern *et al.*, 2011).

The addition of MFC to the diagnostic armoury for CMML has great potential, particularly in a disease which is notoriously difficult to diagnose morphologically. In addition, if this method can be confidently applied to peripheral blood samples the impact would be favourable. Large prospective studies are required with clinical follow-up to fully assess this method.

1.7 Summary

The diagnosis of chronic myeloid malignancies remains centred on morphological assessment which is limited by subjectivity and poor interobserver concordance. There is, therefore, a real need for objective techniques in the diagnostic work-up of patients with suspected disease and the techniques described in this chapter offer great potential.

By using high throughput sequencing, somatic mutations can be identified in a significant majority of patients with both MDS and MDS/MPN meaning this technique could provide a potential objective marker of disease. This would improve the diagnostic capabilities dramatically for these diseases. Furthermore, targeted sequencing panels can provide a high-throughput and cost-effective option for analyzing multiple genes in parallel in the routine clinical setting.

The use of SNP-A also has the potential to dramatically improve the resolution of current cytogenetic techniques, allowing the identification of clonal abnormalities in a greater number of patients. It is possible that this approach could replace the current cytogenetic method, though the inability to detect balanced translocations and inversions raises concerns regarding its utility. Hence, combination would be required with other techniques that could perform this task.

MFC also has a place in the diagnosis of chronic myeloid malignancies though this remains difficult to implement in large busy laboratories, particularly in the setting of comprehensive panels for MDS. Concern also remains regarding the subjectivity of this method, particularly when an objective approach is required. A screening panel on peripheral blood samples, as described in CMML, is particularly appealing and has the potential to identify patients who require additional investigations.

Ultimately, further studies incorporating these techniques are required to refine the diagnostic process and identify the method of choice. This PhD aimed to

investigate the use of these technologies in the diagnosis of chronic myeloid malignancies, assessing whether they can refine current diagnostic criteria.

CHAPTER 2: METHODS

2.1 Haematological Malignancy Diagnostic Service

The research was undertaken within the HMDS (www.hmds.info), a large integrated diagnostic laboratory within the north of England. The laboratory is the sole provider of haematopathology for a population of ~6 million and utilizes morphology, flow cytometry, histology, and molecular techniques to analyse all peripheral blood, bone marrow and tissue biopsies from patients with a suspected haematological malignancy. The laboratory houses relevant expertise in all aspects of haematopathology including 7 principal and consultant grade clinical scientists across the laboratory sections and 6 consultant haematopathologists, of which I am one. The sample tracking, testing and reporting are all managed through the HMDS Integrated Laboratory Information System (HILIS). This system underpins all the laboratory processes, providing a searchable database of samples and test results.

The laboratory also provides the diagnostic information for patients included in the Haematological Malignancy Research Network (HMRN)(Smith *et al.*, 2018). HMRN is a population-based patient cohort encompassing just over half of the patient population served by the laboratory. HMRN collects detailed clinical data on patients diagnosed with a haematological malignancy including treatments, responses and outcomes.

2.1.2 Patient Samples

The samples included in the study comprised surplus material from patient samples referred to HMDS for investigation of a suspected myeloid malignancy. Ethical approval for use of surplus material in research and development of new techniques has been in place in the department since 2004. This ethical approval was however restricted to samples originating in the HMRN study location and so a further ethical application was completed and submitted by myself to expand the cohort to include samples referred from outside the Yorkshire and Humber network. This ethics application was granted in 2016 and further information, including the approval letter is available in Appendix 8.1.

2.2 Patient cohorts

To investigate the potential of new technologies in the diagnosis of myeloid malignancies, 3 patient cohorts (Figure 2.1) were investigated as described below. Cohort 1 was used as a proof of principle, to determine the frequency of clonal abnormalities in those cytopenic patients with the most clinically significant disease who failed to reach current minimum diagnostic criteria. These patients were identified by capturing all patients with a diagnosis of AML or MDS who had a previous non-diagnostic bone marrow over an 8-year period. For this cohort, both high throughput sequencing and SNP arrays were performed on matched pre-diagnostic and diagnostic samples. The 2 subsequent cohorts comprised all unselected patient samples sent for investigation of cytopenia or monocytosis over a 2-year period; these were used to determine the frequency of abnormalities and their clinical significance. For these cohorts, high throughput sequencing was utilized for both while flow cytometry was applied in patients with a monocytosis. Full details of criteria for sample selection can be found in each results chapter, and a general overview of this process is detailed below.

	Cohort 1		Cohort 2		Cohort 3
No. of Samples	69 paired samples		283 samples		2088 samples
Sample Description	Paired non-diagnostic and diagnostic samples		All patients referred for investigation of monocytosis		All patients referred for investigation of cytopenia
Time period	Diagnosed 2004 – 2012		July 2014 – July 2016		July 2014 – July 2016
Technologies Utilised	High Throughput Sequencing SNP-array		High Throughput Sequencing Flow Cytometry		High Throughput Sequencing

Figure 2.1. Summary of patient cohorts investigated as part of this research

2.2.1 Cohort 1

Samples were identified by performing an SQL search on HILIS. Searches were constructed to identify all patients with a confirmed diagnosis of MDS or AML during the time period (2004-2012) who had undergone a previous bone marrow biopsy without a confirmed diagnosis. These samples were then manually curated by me to ensure the study criteria were met.

2.2.2 Cohorts 2 and 3

Samples were identified in real time to apply additional novel tests to run in parallel with traditional testing. Samples were identified by the screening term given on receipt of the sample in the lab. 'Screening' is performed by a Consultant Clinical Scientist following review of the clinical details and morphology. The screening category is selected on HILIS and several linked investigations are triggered as a result. A weekly search was set-up through HILIS using the criteria below which I reviewed for the purposes of sample selection.

The selection criteria for these cohorts were –

Inclusion Criteria

- Age ≥18yrs
- Referred with a suspected chronic myeloid malignancy and screened within the following categories
    - 'Cytopenia', 'Suspected MDS', 'Suspected CMML'

Exclusion Criteria

- Prior diagnosis of a myeloid malignancy
- Sample of insufficient quality for molecular analysis.

2.3 Peripheral Blood Counts

Blood count parameters were obtained on all peripheral blood samples received as per routine laboratory protocol. This analysis was performed on a Sysmex K-1000

haematology analyser which produced a haemoglobin, white cell count (with 3-part differential) and platelet count.

2.4 Morphological Assessment

2.4.1 Staining

All peripheral bloods and bone marrow samples were prepared for morphological assessment by light microscopy. This was performed as part of the routine diagnostic work-up. Air dried slides of peripheral blood or bone marrow aspirate were stained using the May-Grunwald-Giemsa stain on the MIRA II staining machine. This involves fixing the sample using absolute methanol and then staining with the May-Grunwald's stain followed by the Giemsa stain. The slide is then rinsed in buffered water and dried. Buffered water is made by adding 3.955g Na_2HPO_4 and 3.79g KH_2PO_4 to 5L distilled water, ensuring the resultant pH is 6.8.

Trephine samples were fixed in formalin and embedded in resin. Samples were stained with both Haemotoxylin & Eosin and Giemsa. Immunohistochemistry analysis was performed at the discretion of the reporting haematopathologist.

2.4.2 Morphological Review

All samples were independently reviewed by 2 experienced haematopathologists as per laboratory protocol and the cases were reported as part of the routine case load in real time. As a consultant haematopathologist within HMDS, I was part of this reporting team. Each case was reported and then authorized only if there was agreement between haematopathologists. For cases in which there was disagreement, a 3rd haematopathologist was consulted for the casting vote.

2.5 Flow Cytometry

Flow cytometry allows for the assessment of single cells using fluorescently labelled antibodies to target specific antigens on the cell surface. This technique was specifically utilized in this research for the assessment of patients referred with a monocytosis, more detailed methodology is provided in Chapter 5.

In brief, all samples for immunophenotypic analysis were processed within 24 hours of receipt in the laboratory. Numerical studies and basic assessment of aberrant antigen expression were performed on BM or PB samples following a stain-lyse-wash procedure. This lysis technique retains nucleated red cells therefore providing a more accurate total cell percentage.

In addition, more extensive immunophenotyping was performed on PB samples following NH4Cl lysis of erythrocytes using a lyse-stain-wash procedure. A minimum of 10^5 leucocytes were acquired on a single cytometer for all cases.

2.6 High throughput Sequencing

2.6.1 DNA extraction

DNA was extracted from fresh peripheral blood and bone marrow samples on receipt at the laboratory. This was performed on all cases screened with the above screening terms as part of routine laboratory processes. The samples were firstly treated with a lysis solution (12mls 0.86% ammonium chloride), to lyse the red blood cells. Following 2 washes in phosphate buffered saline (PBS, Gibco, product no. 70011-051)) the remaining white cell pellet was incubated with 200µl buffer AL (lysis buffer Qiagen, product no. 51306) and 20µl proteinase K (along with 200µl of PBS) and incubated in a waterbath at 37°C.

For cohort 1, a proportion of samples did not have DNA extracted from fresh liquid samples and so DNA was extracted from air dried bed side smears. For these samples, cells were scraped from the slide into a micro tube and incubated at 37˚C with 200µl buffer AL and 20µl proteinase K.

DNA extraction was then performed using the Qiagen QIAamp DNA mini kit on the automated QIAcube platform as per manufacturer's instructions.

2.6.2 DNA quantification

It is essential that DNA concentrations are normalized before proceeding to library preparation and all DNA samples were therefore quantified using the GloMax

Detection System, which is a fluorescence-based method for determining DNA concentration. To quantify the sample, 1µl of each DNA sample was diluted in 99µl of 1xTE (Tris EDTA) buffer (Promega, product no. 2021-03-20). To this 100µl of QuantiFluor dsDNA dye was added and mixed thoroughly. DNA was quantified on Glomax system after being standardized using a negative control and a control sample prepared to a known DNA concentration (50ng/ml).

Samples were then normalized to an ideal concentration of 50ng/µl. For those with a higher concentration, samples were diluted with an appropriate amount of laboratory grade water. For those with a concentration below 50ng/µl, samples were added neat. Those with a concentration below 5ng/µl were excluded from further analysis.

2.6.3 Preparation of Sample Plate

In a 96-well plate, 30µl of 48 individual patient DNA samples were pipetted into columns 1-6, rows A-H with a concentration normalized to 50ng/µl as described above. Position and specific dilution of each sample was documented on a DNA quantity worksheet, an excel spreadsheet summarizing individual sample information, and the plate was prepared and checked by 2 individuals (Table 2.1).

Table 2.1. Example of DNA Quantity Worksheet

Sample	Last name	ng/ul	dilution	ul of sample	ul water	Screening	sample well
XXXX	Jones	42.15	0.843	29.7	-4.7	Suspected CMML	F5
XXXX	Jones	60.44	1.2087	20.7	4.3	Cytopenia	G5
XXXX	Jones	137.71	2.7543	9.1	15.9	Cytopenia	H5

2.6.4 Fluidigm library preparation

Library preparation was performed using the Fluidigm access array, an amplicon-based system for target enrichment. The amplicon tagging strategy combines tagged target specific (TS) primer pairs with sample-specific barcodes and the adapter sequences used by the specified high throughput sequencer (Fig. 1.2; using

Illumina adapters as an example). This is all combined using integrated fluidics circuits (IFC) on the Fluidigm 48.48 Access Array.

The primer pairs incorporate universal forward (common sequence 1 (CS1)) and reverse (common sequence 2 (CS2)) tags which act as the sequencing primer binding sites. The adapter sequences that are also incorporated allow library fragments to attach to the flow cell.

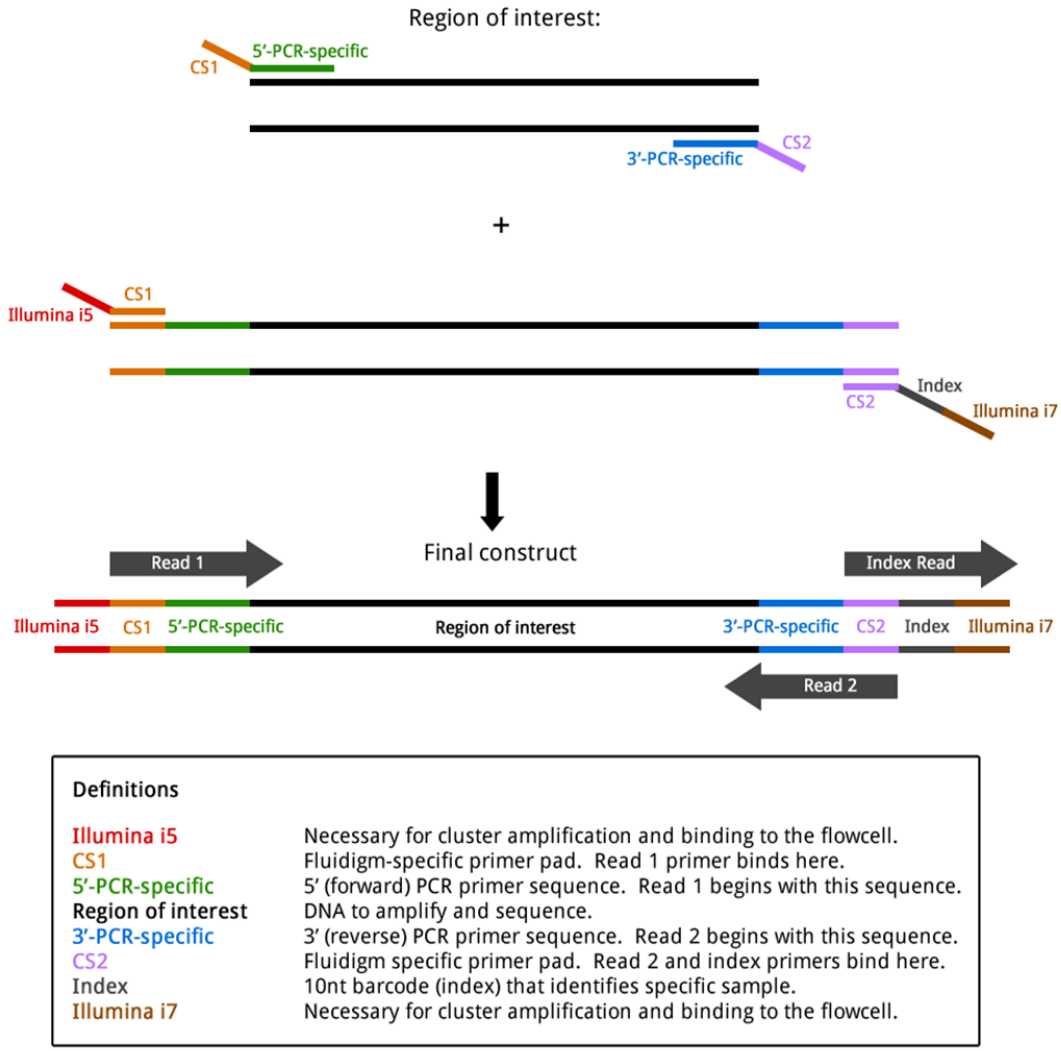


Figure 2.2. Fluidigm Amplicon Structure.

Targeted sequencing primers were designed for both the Roche 454 and Illumina MiSeq using the Fluidigm D3 design system; panel development and validation is discussed in Chapter 3.

The process for library preparation for both the Roche 454 (using 48 individual primer pairs) and the Illumina MiSeq (Multiplexed primer pairs) was performed according to manufacturer's instructions and is described in Appendix 8.2. This document also describes the process for sequencing on the Illumina MiSeq and together these form the basis of the HMDS standard operating procedure (SOP) for both techniques.

In brief, Fluidigm library preparation is a 2-step approach. In the first step, target regions are amplified with target specific primer pairs which have been tagged with common sequence tags. The amplification is carried out on the FC1 cycler. After harvesting the products from the 48.48 Access Array IFC, the second step is performed on a 96-well plate. During this PCR step a sample specific barcode and sequencer specific adapter are introduced.

Once the libraries were prepared, these were sequenced using either the Roche GS Junior or Illumina MiSeq. Both techniques are described below.

2.6.5 Sequencing on the Roche 454

The Roche GS Junior system uses an emulsion PCR based approach for target amplification followed by a pyrosequencing reaction. DNA fragments are annealed to DNA capture beads and then emulsified within a water-in-oil mixture. Each bead is captured within its own water droplet and amplification occurs within this microreactor. Following amplification, the beads are loaded onto a PicoTitrePlate along with sequencing enzymes. Sequencing is performed by a sequencing-by-synthesis method during which individual nucleotides flow across the PicoTitrePlate in a fixed order. If the nucleotide is complementary to the template DNA the polymerase extends the DNA by 1 or more nucleotides. This results in a reaction that generates a chemiluminescent signal which is recorded by the camera.

This process was performed according to manufacturer's instructions and these are included in Appendix 8.3.

2.6.5.1 Sequencing Analysis on the GS Junior

Analysis was performed using the GS Amplicon Variant Analyzer with paired end reads aligned to the reference amplicon sequence. Variants identified by this analysis software were then analysed using the Ensembl Variant Effect Predictor (VEP) software (McLaren *et al.*, 2010). Variant annotation was performed by myself. All synonymous, non-coding variants and germline polymorphisms were excluded. The latter polymorphisms were retained if previously confirmed somatic and recurrently reported in COSMIC database (Forbes *et al.*, 2011).

2.6.6 Sequencing on the Illumina MiSeq

The Illumina MiSeq using Illumina's sequencing by synthesis (SBS) technology. The prepared library is loaded onto a flow cell which contains a lawn of surface bound oligonucleotides which are complementary to the library adapters. The captured fragments then undergo bridge amplification to generate clonal clusters large enough to be visualized during sequencing. During sequencing all 4 reversible terminator bound dNTPs flow over the clusters and complementary bases are incorporated. This is followed by an imaging step before the cycle is repeated.

In preparation for sequencing on the MiSeq the harvested products, from library preparation, are pooled, purified, using Ampure beads, and then quantified using the QuantiFluor Fluorometer as described in section 2.6.2. Once the concentration is known, the library is diluted with hybridization buffer (Illumina, product no. 20015892) to a concentration of 7pM and added to the MiSeq reagent cartridge to be loaded onto the sequencer.

This process is performed according to manufacturer's instructions which is outlined in Appendix 8.2.

2.6.6.5 Sequencing Analysis for the Illumina MiSeq

Initially sequencing analysis was performed using Illumina commercial software (MiSeq reporter). The VCF (variant call file) was then annotated using same process as the output from the GS Amplicon Variant Analyzer as described in section

2.6.5.1. More recently this was performed using a bespoke analysis pipeline which was developed in-house for the targeted panel on the Illumina MiSeq. This pipeline runs automatically at the end of the sequencing run, producing VCF files for further variant annotation and interpretation. The pipeline is discussed further in Chapter 3.

2.6.7 Variant selection

Due to the nature of this study, matched germline samples were not received for analysis. It was therefore essential that extensive filtering was performed to exclude artefacts and previously documented single nucleotide polymorphisms (SNPs). All variant annotation was performed by myself using the following resources and criteria.

The following resources were used for variant filtering and annotation to identify drivers and exclude technical artefacts.

- Alamut
    - Data sources include –
    - COSMIC, Clinvar, DbSNP, gNomad, 1000 genomes, CentoMD, HGMD, DMuDB, SwissVar
- Correlation with an in-house database of >3000 myeloid malignancy samples referred to HMDS
- Online search tools – PubMed, Google

The following criteria were applied to ensure only high confidence variants were included-

Inclusion Criteria

- Variants previously reported in the literature, confirmed in haematopoietic tissue (as reported in COSMIC)

    - Well documented hotspot variants were retained irrespective of coverage however in those with low coverage these were only included in the analysis if reproducible across 2 separate runs.
- Truncating variants (nonsense, frameshift indels and essential splice site variants) in genes implicated in myeloid pathogenesis by loss-of-function – *ASXL1, BCOR, DNMT3A, EZH2, RUNX1, STAG2, TET2, TP53, WT1.*
- Previously unreported variants that cluster (±3aa) with a hotspot variant or at a recurrently mutated amino acid position, providing the variant is not reported in population databases.

Exclusion Criteria

- All non-coding variants
- All synonymous variants
- All known polymorphisms, as reported on germline databases
    - These were retained if previously confirmed somatic and recurrently reported in COSMIC and reported in population databases at <0.0014.
- Highly recurrent variants consistent with artefact
    - A database of recurrent PCR/sequencing artefacts was developed following analysis of multiple sequencing runs.
    - Each individual run was also analysed for run specific, highly recurrent variants likely to be artefact. These were also excluded.

2.6.8 Validation of detected variants

2.6.8.1 Sanger Sequencing

To ensure the variants detected were genuine, particularly on the initial sequencing runs, extensive validation was performed by Sanger sequencing. This technique is still considered the gold standard for detecting sequence changes, though it is limited by sensitivity when looking for variants present at low allele fraction. The same CS tagged Fluidigm primers that are used in high throughput sequencing were also used to Sanger sequence the PCR products.

Initially PCR of the region of interest was performed using CS1 and CS2 tagged primers, Amplitaq Gold in standard 20µl PCR mix and Biomed2 PCR program. This produces a CS1/2 tagged fragment of interest which is then cleaned up with Ampure beads allowing removal of excess primers and dNTPs. The DNA is eluted in 26µL of H2O and 5.7µL of this is added to 4.3µL of BIG DYE reaction mix containing CS1 or CS2 primers. This then undergoes PCR using the BIG-DYE sequencing program. This produces labelled fragments of interest which can be analysed. The product is cleaned up using Agencourt clean-seq beads and eluted in 20µL of water. For sequencing 10µL of each eluted DNA is resuspended in 20µL of HiDi Formamide reaction mix and loaded onto the ABI 3500 Genetic Analyser.

Electropherograms were reviewed manually on proprietary software and new or previously identified variants were confirmed.

2.6.8.2 Fragment Analysis for detection of *FLT3*-ITDs and validation of low level *JAK2* V617F mutations

Following extensive assessment of the sequencing panel (see Chapter 3 – Panel Development), it was clear that *FLT3*-ITDs were not detectable using this technique. These abnormalities were therefore detected using current laboratory procedures. Internal Tandem Duplications of the *FLT3* receptor gene can be detected in ~30% of AML cases and typically involve exon 11 and sometimes adjacent intron sequences (Kottaridis *et al.*, 2001). The duplicated DNA length varies between samples though is always in-frame and occurs in the juxtamembrane region. The current technique in HMDS detects both *FLT3*-ITDs and *NPM1* mutations, which are to date the most clinically relevant mutations detected in AML (Noguera *et al.*, 2005). The technique uses fragment analysis i.e. analyzing the length of an amplified targeted region of DNA. For *FLT3*, amplification is carried out across the boundary of exons 11 and 12 on genomic DNA. The reverse primer is labelled with FAM (6-carboxyfluorescein), a fluorescent dye, and combined with HiDi formamide and ROX500 size standard prior to electrophoresis on the Life Technologies 3130 platform. Amplification results in a wildtype peak of 329bp while a *FLT3*-ITD peak is larger, usually between

332-437bp in length. For *NPM1*, amplification occurs across the last exon of *NPM1*, with the reverse primer labelled with an alternative fluorescent dye, HEX. The *NPM1* wildtype product is 195bp while the mutant fragment is 199bp in length. Both these methods have a limit of detection of 5%.

A fragment analysis approach is also applied for the detection of the common *JAK2* variant V617F which is frequently detected in cases of MPN. This technique will detect variants down to 1% clone size and was therefore used to validate low level *JAK2* mutations detected in the study cohort. The technique has been developed to detect both *JAK2* V617F and *CALR* frameshift mutations. For *JAK2*, primers targeting exon 14 of *JAK2* are used to amplify this region in genomic DNA. A fluorescent reverse primer (labelled with FAM) is used in combination with an unlabelled mutant specific ASO (allele specific oligonucleotide) and a consensus *JAK2* normal forward primer. Amplification of the *JAK2* wildtype produces a product of 362bp, while *JAK2* V617F produces a band of 202bp. For *CALR*, amplification of exon 9 is performed using a fluorescently labelled (HEX fluorescent dye) forward primer and unlabelled reverse primer. The wildtype fragment measures 292bp while insertion/deletion mutations produce a variety of fragment sizes.

2.7. SNP-array analysis

Array based whole genome scanning was performed using the Illumina Infinium HD Assay Ultra technique and hybridized to the CytoSNP12 beadarray. This assay genotypes a locus using 2 colour read outs – one for each allele. This beadchip is coded with multiple copies of oligonucleotide probes with target specific loci across the genome. As DNA fragments pass over the BeadChip, each probe binds to its complementary sequence in the sample DNA. Single base extension then occurs which extends the probe using the sample DNA as a template and incorporates detectable labels. The BeadChip is then scanned using the Illumina BeadArray reader which uses a laser to excite the fluorophore of the nucleotide label causing it to emit a signal. The intensity values for each colour convey information about the alleleic ratio. This process was performed according to manufacturer's instructions

(http://emea.support.illumina.com/downloads/infinium_hd_ultra_assay_manual_euc_(11328095_b).html) and an outline of the protocol is included in the appendix (see Appendix 8.4).

2.7.1 Analysis of SNP array data

BeadChips were scanned using the BeadArray™ Reader. Data was visualised using Karyostudio™ software which applies the CNV Partition 2.4 algorithm to identify abnormalities. The LogR ratio and B-allele frequency plots were further scrutinised 'by eye', due to the mis-calling of some mosaic CNVs/CN-LOH.

Only regions of CN-LOH>10Mb were included in the analysis as per reported guidelines unless these regions included a gene sequenced on the targeted panel (Simons *et al.*, 2012). All copy number variants were compared with online databases of constitutional variants - Database Genomic Variants (http://dgv.tcag.ca/dgv/app/home) and db-Var (http://www.ncbi.nlm.nih.gov/dbvar/) with reported variants excluded from further analysis.

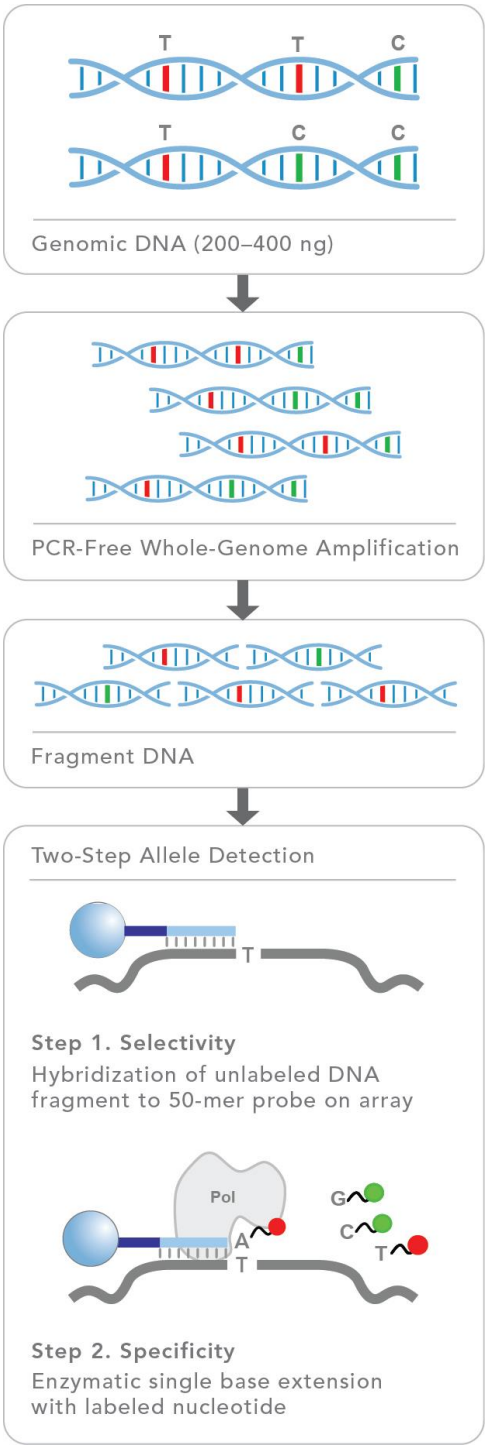


Figure 2.3. Overview of Illumina SNP array methodology (courtesy of Illumina, Inc)

2.8 Statistical Analysis

All statistical analyses were performed using standard tests and R (R Core Team, 2019); the specific methods are provided in each chapter of this thesis.

In general, comparison of features between groups was performed using parallel χ^2, univariate logistic regression or Mann-Whitney U tests. Sensitivity, specificity, positive and negative predictive values (PPV and NPV) were calculated using 2x2 contingency tables. Any pairwise associations were tested with the Fisher exact test corrected for multiple testing using the Benjamini–Hochberg procedure.

The impact of abnormalities on overall survival and risk of progression were estimated using Cox regression. Simple differences in survival were assessed using the log rank test.

CHAPTER 3: TARGETED MYELOID SEQUENCING PANEL DEVELOPMENT

3.1 Introduction

With the advent of high throughput sequencing, it became increasingly feasible to perform whole genome and targeted sequencing for gene discovery across all malignancies. Due to the relative ease of access to fresh DNA in haematological malignancies, in particular myeloid malignancies, these diseases have led the way in research in this area, and a number of novel genes have been implicated across the disease spectrum. Initial use of this technology focused on basic laboratory research, however the development of benchtop sequencers meant that it could be incorporated into routine clinical laboratories for translational research and ultimately for implementation into routine clinical practice. This technology has both diagnostic and prognostic potential, and could be cost effective due to the ability to sequence multiple genes across multiple samples in one sequencing run.

HMDS is well positioned to fully investigate the potential of new technologies in routine diagnostics. The laboratory is a fully integrated service, providing the benchmark for haematopathology in the UK. It provides diagnostic services for a well-defined geographical area with a population of ~6 million. All diagnostic and follow-up samples for patients within this region are referred to HMDS. The laboratory also has strong clinical links through both working relationships with clinical colleagues across the region as well as a collaborative relationship with HMRN. This provides the necessary framework to definitively assess the diagnostic potential of new technologies in the 'real-world' setting.

A benchtop sequencer (Roche GS Junior) was acquired by HMDS in 2010; this technology provided the basis for this research and high throughput sequencing was used across all 3 cohorts. This chapter outlines the development and validation of the myeloid panel within HMDS.

3.2 Rationale for the development of a myeloid gene panel

3.2.1 High throughput sequencing can identify multiple mutations in parallel

As discussed in the introductory chapter, high throughput sequencing has provided a platform to investigate the genes and functional pathways implicated in myeloid disease pathogenesis and these are now well established. Subsequent research has tended to focus on identifying mutations in parallel, not only looking at relationships between mutations but also assessing the prognostic relevance using a multivariate analysis.

3.2.1.1 MDS Patients

An initial study by Bejar et al, investigated 111 cancer associated genes by mass spectrometry as well as high throughput sequencing, and demonstrated mutations across 18 genes (Bejar *et al.*, 2011). Mutations were detected in a total of 51% of patients. Of these 18, mutations in 5 genes were associated with a poor prognosis including *TP53, EZH2, ASXL1, RUNX1* and *ETV6* by multivariate analysis (Bejar *et al.*, 2011) though importantly key genes such as the spliceosome genes were not targeted in this analysis. The same group analysed 22 genes in patients with low risk MDS to both validate the MD Anderson Lower risk Prognostic Scoring system (LRPSS) and to determine if any genes added additional prognostic information (Bejar *et al.*, 2012). Interestingly *TP53, RUNX1, EZH2* and *ASXL1* were all associated with shorter overall survival but only *EZH2* retained significance when the LRPSS was incorporated (Bejar *et al.*, 2012).

The following year, 2 seminal papers from England and Germany used more relevant panels to determine the genetic landscape in larger MDS cohorts (Haferlach *et al.*, 2013; Papaemmanuil *et al.*, 2013). By performing targeted sequencing of over 100 genes both groups reported a high frequency of driver mutations (up to 90% of cases), however only 4-6 genes were found in >10% of patients, with a long tail of low frequency mutations. Importantly however the 16 most frequently mutated genes were identical across both studies, albeit found in a

slightly different order (Fig. 3.1) Not only did these studies confirm the high frequency of mutations, but

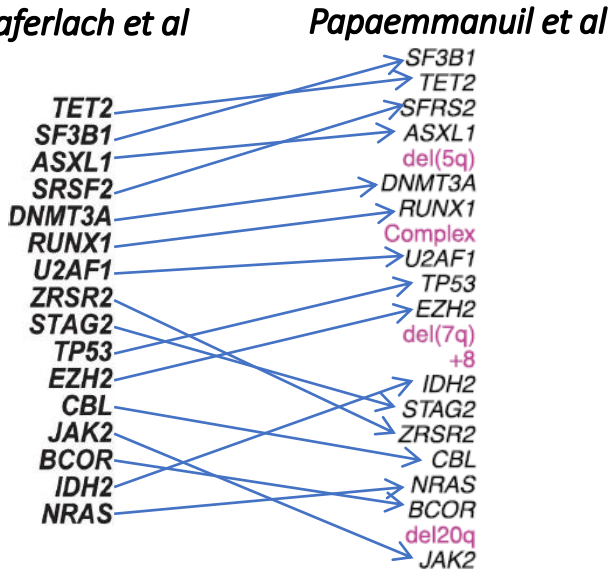


Figure 3.1 Genes most frequently mutated in MDS. *Adapted from Papaemmanuil et al, 2013 and Haferlach et al, 2013.*

also demonstrated important relationships between the mutation profile and disease phenotype, as well as identifying genes commonly co-mutated and those which are mutually exclusive (Papaemmanuil *et al.*, 2013).

With regards to prognosis, Haferlach et al proposed a prognostic model integrating both clinical and genetic information. This model incorporated 14 genes, as well as age, gender, IPSS-R, blood count parameters, bone marrow blasts and cytogenetics which could classify patients into 4 distinct groups which had significantly different ($p<0.001$) 3-year survivals of 95.2 vs 69.3 vs 32.8 vs 5.3% (Haferlach *et al.*, 2013). Importantly the 5 genes reported by Bejar et al were all included in this prognostic model, as well as other commonly mutated genes including *STAG2, NRAS, KRAS* and *CBL* (Haferlach *et al.*, 2013). This prognostic model has yet to be validated.

Interestingly in the other sequencing study published at this time, the addition of point mutation data to the IPSS and standard clinical variables had only a marginal non-significant increase in prognostic potential (Papaemmanuil *et al.*, 2013).

Mutations were however able to predict prognostically significant variables including haemoglobin, ring sideroblasts and bone marrow blast percentage, as well as confirming the strong association between number of driver mutations and outcome; the median leukaemia free survival being 49 months for patients with 1 mutation, dropping to only 4 months in those with ≥6 mutations (Papaemmanuil *et al.*, 2013).

A large international collaborative study by the IWG-PM (International Working Group for Prognosis in MDS) is currently underway to develop an 'IPSS-molecular' which will incorporate clinical, haematological and molecular parameters (Cazzola, Della Porta and Malcovati, 2013). This study will combine data from the largest MDS cohort to date and should provide a definitive prognostic model using this novel information.

3.2.1.2 CMML PATIENTS

The majority of early studies using HTS to analyse MDS mutations in parallel, have been performed in the CMML subtype.

An initial study using the Roche 454 sequenced 7 candidate genes (*CBL, JAK2, MPL, NRAS, KRAS, RUNX1* and *TET2*) in 81 CMML patients (Kohlmann *et al.*, 2010). Mutations were found in a surprisingly high number of samples despite the small panel, with 59 out of 81 patients (72.8%) habouring at least 1 mutation. The mean number of mutations was 1.6 (range 1-6) and *TET2* was the most frequently mutated gene (Fig. 9) (Kohlmann *et al.*, 2010). Interestingly, a *TET2* mutation was shown to have an impact on survival, with a better outcome in those with a mutation (median OS 130.4 vs. 53.6 months P=0.013) (Kohlmann *et al.*, 2010).

This same group extended their findings by studying an additional 5 genes – *IDH1, IDH2, NPM1, ASXL1* and *EZH2* (V Grossmann, Kohlmann, Eder, Haferlach, Kern, N. Cross, *et al.*, 2011). When combined with the previous study this brought the number of patients with at least one mutation to 81.5%. No specific pattern was

noted between these mutations, however a poor outcome was noted in patients that harboured an *EZH2* mutation compared to *EZH2* wildtype (3-year survival 33.3 vs 69.9%, P=0.001) (V Grossmann, Kohlmann, Eder, Haferlach, Kern, N. Cross, *et al.*, 2011).

A further study of both CMML and CMML-derived AML, while using the less sensitive method of Sanger sequencing, identified at least one mutation in 86% of patients in their cohort by analysing a range of mutations including *CBL, KRAS, NRAS, IDH1, IDH2, DNMT3A, TET2, EZH2* and a previously unrecognized gene UTX (Jankowska *et al.*, 2011). UTX encodes for a demethylase specific for H3K27, a histone mark thought to contribute to the pathogenesis of malignant evolution, and this study was the first report of mutations in CMML patients (4/52) and secondary AML (2/20). *EZH2* was mutated in 5.5% of patients while *DNMT3A* mutations were noted in 10%. While there was little impact on patient outcome, the findings highlight the molecular heterogeneity of this disease and the importance of molecular analysis (Jankowska *et al.*, 2011).

More recently a number of studies have utilized larger targeted panels to investigate CMML and fully elucidate the prognostic significance of somatic mutations. Mutations are now consistently identified in >90% of CMML patients (Itzykson *et al.*, 2013; Elena *et al.*, 2016; Mason *et al.*, 2016; Palomo, Garcia, *et al.*, 2016). While the genes mutated occur across the spectrum of haematological malignancies, *SRSF2, TET2* and *ASXL1* are by far the most commonly mutated, and the combination of *TET2* with either *SRSF2* or *ZRSR2* has been shown to be highly specific for a myelomonocytic phenotype (Malcovati et al., 2014).

A prognostic score was proposed following a study of 312 CMML patients in whom *ASXL1* was sequenced and up to 18 other genes (Itzykson *et al.*, 2013). Of the 173 in whom all genes were sequenced ≥1 mutation was detected in 95%. *ASXL1* mutations were confirmed to be a poor prognostic feature, which retained significance on multivariate analysis, and this was subsequently incorporated into a prognostic score along with age and blood count parameters (Itzykson *et al.*, 2013).

An alternative scoring system was subsequently proposed which was based on the established CMML prognostic scoring system (CPSS) developed in 2013 (Such *et al.*, 2013). Out of the 214 patients included in this study 93% harboured ≥1 mutation. The updated scoring system again incorporated *ASXL1* but also *RUNX1, NRAS* and *SETBP1*, which were all shown to be independently associated with a worse overall survival (Elena *et al.*, 2016). The scoring system combined these genetic factors with red cell transfusion dependency, white cell count and bone marrow blasts producing a CPSS-Molecular model which significantly improved risk stratification (Elena *et al.*, 2016). Mutations were also shown to be significant in patients with low risk or uninformative cytogenetics with *ASXL1, NRAS, EZH2* and *SRSF2* mutations being identified as adverse risk on multivariate analysis along with the CPSS score (Palomo, Garcia, *et al.*, 2016).

From the literature it is therefore clear that high throughput sequencing can facilitate the sequencing of large numbers of genes in parallel on multiple patients. This could not only provide important prognostic information but due to the high frequency of mutations had the potential to become a diagnostic tool particularly in difficult cases of suspected MDS and CMML. Hence, the design and implementation of a myeloid panel is a central feature of this PhD research, allowing for the investigation of its diagnostic potential.

3.2.2 Rationale for gene selection

At the outset, it was apparent that there is significant overlap clinically and morphologically between the different sub-classes of myeloid malignancy, and this is also apparent at the genomic level with certain genes mutated across multiple disease groups. A myeloid panel was designed with the ultimate goal of implementing this test into routine clinical practice. It was therefore essential that this panel was clinically relevant, evidence based, and applicable to a large number of patients to ensure cost effectiveness. A targeted sequencing panel was therefore designed to include the most commonly mutated genes across all myeloid malignancies.

This would –

- Allow patients across the spectrum of myeloid malignancies to be sequenced together
- Provide an all-encompassing panel for the investigation of patients with a suspected myeloid malignancy
- Ensure a high throughput of cases providing a more cost-effective approach
- Allow for a more meaningful analysis of results by standardizing the genes sequenced

3.2.3 Selection of genes for the panel

Prior to commencing this research, a small myeloid panel had been designed within HMDS for potential use in myeloid malignancies. However, this did not include many of the newly reported genes, and so I undertook a review of the literature to identify which genes should ideally be added. This review was to identify and record the frequency of the most commonly mutated genes across the spectrum of myeloid malignancies. A PubMed search was performed using the search term 'mutation' AND each of the following list of disease specific search terms –

- AML; acute myeloid leukaemia
- MDS; myelodysplas*
- CMML; chronic myelomonocytic leukaemia
- PV; polycythaemia vera
- ET; essential thrombocythaemia
- PMF; primary myelofibrosis

The gene implicated, the frequency reported and the number of samples analysed for each disease sub-group were recorded for each published study. The results are presented in Appendix 8.5 and summarized in the graphs below (Fig. 3a&b).

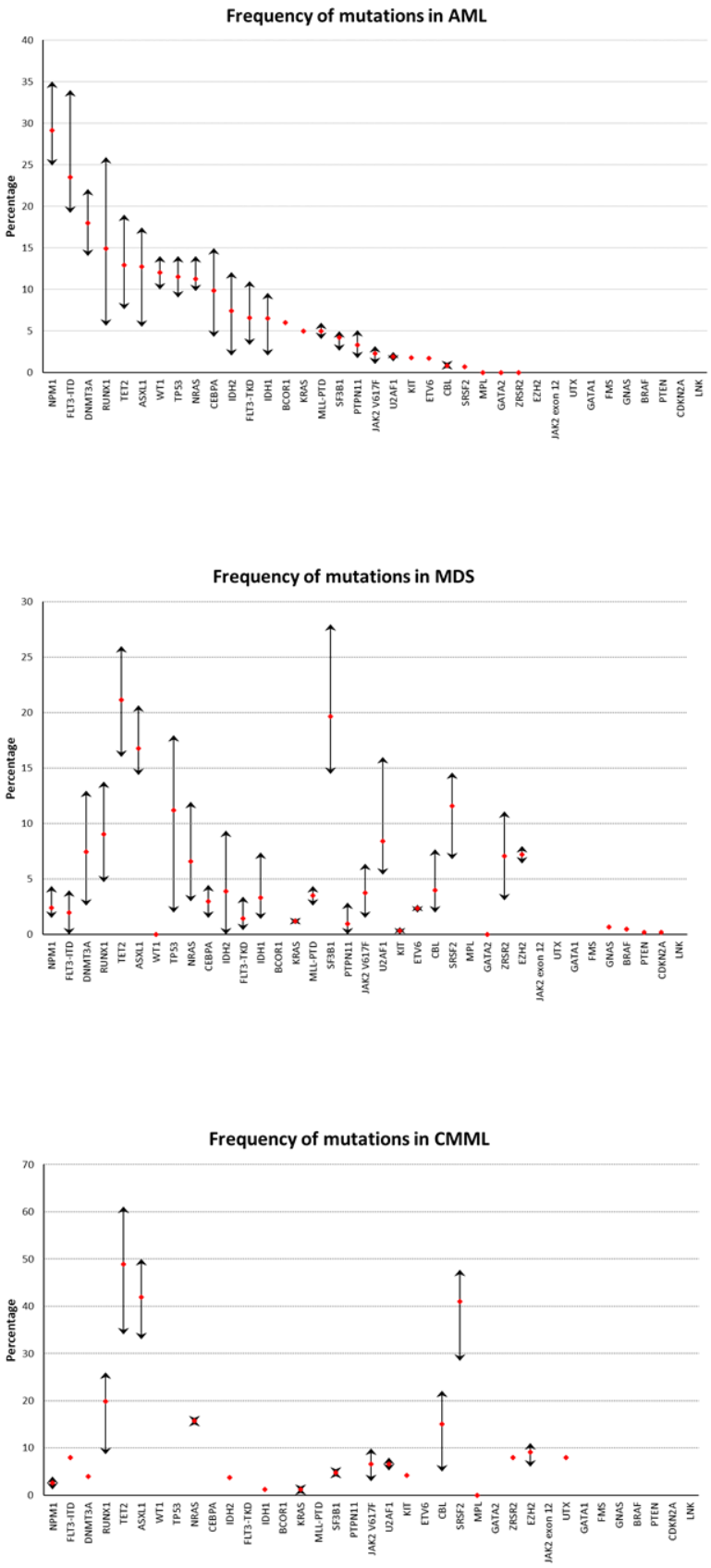


Figure 3.2a. Frequency of gene mutations across the spectrum of myeloid malignancies

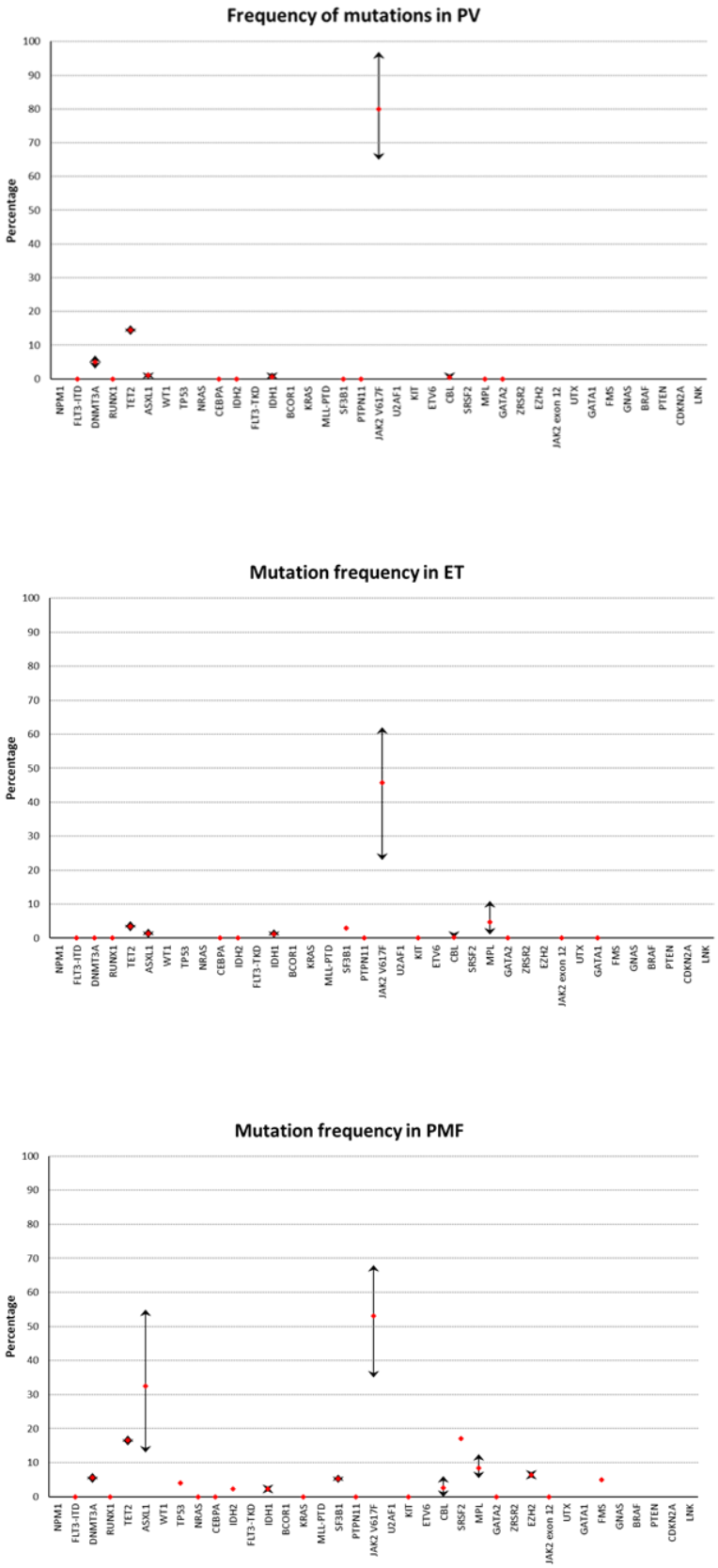


Figure 3.2b. Frequency of gene mutations across the spectrum of myeloid malignancies

The genes reported to be mutated in >5% of patients across the disease groups were captured from the literature review and prioritized for inclusion in the panel. These were *NPM1, FLT3, DNMT3A, RUNX1, TET2, ASXL1, WT1, TP53, NRAS, CEBPA, IDH1, IDH2, BCOR, KRAS, SF3B1, U2AF1, SRSF2, ZRSR2, EZH2, JAK2, MPL*. The rarer myeloid malignancies were then considered, and KIT was identified as an additional target for the panel due to the high frequency of mutations in systemic mastocytosis and its inclusion as a diagnostic criterion in this disease (Swerdlow *et al.*, 2017). This gene also has prognostic relevance for core binding factor leukaemias in which it infers a poor prognosis (Kim *et al.*, 2013).

It was also clear from the literature review that some genes contained hotspot regions in which mutations invariably occurred, while others could be mutated across the whole coding region. The targeted regions for each gene are listed in Table 3.1.

Gene	Targeted Region
TET2	exon 3-11
DNMT3A	exon 11-23
IDH1	exon 4
IDH2	exon 4 & 5
ASXL1	exon 12
*EZH2**	exon 2-20
SRSF2	exon 1 & 2
U2AF1	exon 2 & 6
ZRSR2	exon 2-11
SF3B1	exon 12-16
NPM1	exon 12
RUNX1	exon 4-8
*BCOR**	exon 2-15
TP53	exon 5-9
FLT3	exon 20
NRAS	exon 2 & 3
KRAS	exon 2 & 3
CBL	exon 8 & 9
JAK2	exon 12 & 14
WT1	exon 7 & 9
KIT	exon 8 & 17
MPL	exon 10

*Due to capacity issues, these genes were not included on the Roche 454 panel

Table 3.1. Genes and targeted regions for inclusion in myeloid panel

3.3 Roche GS Junior panel development

3.3.1 Amplicon design

Due to sequencing capacity (Table 3.2), and to ensure adequate coverage depth on the GS Junior, a maximum of 48 amplicons could be sequenced on each sequencing run.

	GS Junior
Total output per run	35Mb
Run time	10 hours
Read length	300b
Reads per run	100 000

Table 3.2. GS junior sequencing run parameters

To encompass as many of the desired genes as possible, 2 complementary panels were designed for use on the GS Junior. These panels included 20 genes in total. *EZH2* and *BCOR* were not included due to capacity issues.

Primers for the panels were designed using the D3™ Assay Design service (Fluidigm®, San Francisco, CA, USA). The list of genes and targeted regions were uploaded to the online system using the genomic coordinates and the desired genome build GRCh37/hg19.

Access array target specific primers were then designed by this service using proprietary design software which ensures optimal primer specificity and avoids placing primers over SNPs, which may impact on specificity, and avoids placing primers in GC-rich regions or repeats. The final list of primer pairs provided by the D3™ Assay Design service is provided in Appendix 8.6. The final layout of the primer plates is presented in Table 3.3 (a)&(b) below.

(a)

Column	1	2	3	4	5	6
A	ASXL1 exon 12.1	ASXL1 exon 12.9	KRAS exon 2.1	RUNX exon 5.1	JAK2 exon 12.1	p53 exon 9.5
B	ASXL1 exon 12.2	ASXL1 exon 12.10	KRAS exon 3.1	RUNX exon 6.1	JAK2 exon 14.1	DNMT3A exon 16
C	ASXL1 exon 12.3	ASXL1 exon 12.11	DNMT3A exon 11	RUNX exon 7b.1	WT exon7.2	DNMT3A exon 17
D	ASXL1 exon 12.4	ASXL1 exon 12.12	DNMT3A exon 12/13	RUNX exon 8a.1	WT exon 9.1	DNMT3A exon 18/19
E	ASXL1 exon 12.5	CBL exon 8.1	NRAS exon 2.1	RUNX exon 8b.1	p53 exon 5.1	DNMT3A exon 20
F	ASXL1 exon 12.6	CBL exon 9.1	NRAS exon 3.1	DNMT3A exon 14	p53 exon 6.2	DNMT3A exon 21
G	ASXL1 exon 12.7	cKit exon 17.1	RUNX exon 3.1	DNMT3A exon 15	P53 exon 7.3	DNMT3A exon 22
H	ASXL1 exon 12.8	cKit exon 8.1	RUNX exon 4.1	FLT3TKD.1	p53 exon 8.4	DNMT3A exon 23

(b)

Column	1	2	3	4	5	6
A	SRSF2 exon 1	ZRSR2 exon 2	ZRSR2 exon 10	TET2 exon 4.7	TET2 exon 5	TET2 exon 11.2
B	SRSF2 exon 2.1	ZRSR2 exon 3	ZRSR2 exon 11	TET2 exon 4.8	TET2 exon 6	TET2 exon 11.3
C	SRSF2 exon 2.2	ZRSR2 exon 4	TET2 exon 4.1	TET2 exon 4.9	TET2 exon 7	TET2 exon 11.4
D	SF3B1 exon 14	ZRSR2 exon 5	TET2 exon 4.2	TET2 exon 4.10	TET2 exon 8	TET2 exon 11.5
E	SF3B1 exon 15	ZRSR2 exon 6	TET2 exon 4.3	TET2 exon 4.11	TET2 exon 9	TET2 exon 11.6
F	SF3B1 exon 16	ZRSR2 exon 7	TET2 exon 4.4	TET2 exon 4.12	TET2 exon 10.1	IDH1 exon 1
G	U2AF1 exon 2	ZRSR2 exon 8	TET2 exon 4.5	TET2 exon 4.13	TET2 exon 10.2	IDH1 exon 2
H	U2AF1 exon 6	ZRSR2 exon 9	TET2 exon 4.6	TET2 exon 4.14	TET2 exon 11.1	IDH2 exon 1

Table 3.3 (a) & (b). Layout of 2 primer plates for sequencing on the GS Junior

3.3.2 Initial sample selection

A cohort of 69 paired samples was identified for initial sequencing using the 2 myeloid panels on the GS Junior. These samples formed cohort 1, the analysis of which is described in greater detail in Chapter 4. By using paired samples, increased confidence could be placed on those variants identified across 2 samples at 2 distinct timepoints.

A proportion of samples had DNA stored at the time of receipt in the laboratory and this sample was used for further analysis. On the remaining samples, DNA was extracted from unstained slides using the method described in Chapter 2. These samples were quantified and then checked for amplification quality using control gene PCR and run on 2% agarose gel. All samples amplified despite varying concentrations of DNA. These samples were then used for initial analysis of the targeted panels.

3.3.3 Pipeline development

Initial read trimming, alignment and variant calling were performed using commercial software – GS Amplicon Variant Analyzer (AVA; Roche). Reads were

aligned to the reference amplicon sequence which was uploaded to the software for each amplicon. The consequence of each variant was then analysed using Ensembl Variant Effect Predictor (VEP) software (McLaren *et al.*, 2010). This is an open access software which uses a number of annotation sources to predict the effect of individual variants. These sources include -

- COSMIC
- dbSNP, HGMD-PUBLIC, ClinVar, 1000 Genomes, NHLBI-ESP, gnomAD
- SIFT and PolyPhen

An application was developed within HMDS for the automation of this process. This enabled data to be retrieved from AVA, automatically converted into the correct format and uploaded to a local VEP database for effect prediction. This application was linked to the HMDS web-based laboratory system HILIS.

3.3.4 Variant Validation

All variants detected on the GS Junior were validated using Sanger sequencing. This was performed using the same primers as those designed for the sequencing panel (see Chapter 2, section 2.6.7.1). Using the sequencing analysis software, the electropherograms were visualized manually and the presence or absence of the variant recorded. The limitations of this validation technique were recognized, with variants only detected down to a variant allele fraction (VAF) of ~15%.

3.4 Illumina MiSeq myeloid panel development

3.4.1 Rationale for Development of Illumina MiSeq panel

It was clear from initial sequencing runs that the capacity of the GS Junior was not compatible with a high throughput laboratory and the process was both labour intensive and costly. Hence, alternative sequencing options were investigated, and an Illumina MiSeq was subsequently acquired for the department. While this sequencer offered shorter read lengths, the total output and reads per run were far superior to the GS Junior (Table 3.4); enabling a larger panel to be performed on an

individual sequencing run by multiplexing multiple targets during Fluidigm library preparation.

	Roche GS Junior	Illumina MiSeq
Total output per run	35Mb	4.5-5.1Gb
Run time	10 hours	24hrs
Read length	300b	150b
Reads per run	100 000	24-30 million

Table 3.4. Comparison of sequencing run parameters between Roche GS Junior and Illumina MiSeq

The 2 complementary myeloid panels were therefore reviewed and updated and consolidated into one myeloid panel.

3.4.2 New gene selection

Due to the increased capacity on the MiSeq, a larger panel was designed to incorporate all the desired genes from the previous literature review. Accordingly, an updated literature search was performed to identify any newly reported genes implicated across the myeloid malignancies. A number of important targets were identified for incorporation into the new panel, and these are discussed below.

3.4.2.1 *CALR*

In 2013, two seminal papers reported on the presence of recurrent somatic mutations in the Calreticulin gene in myeloproliferative neoplasms (Klampfl *et al.*, 2013; Nangalia *et al.*, 2013). Calreticulin is a highly conserved protein which is involved in calcium haemostasis, proliferation, apoptosis and immunogenic cell death. All mutations are frameshift and result in a novel C-terminal with loss of the KDEL motif. These mutations are frequent in both ET and PMF being identified in 25-32% and 14-35% of patients respectively (Klampfl *et al.*, 2013; Nangalia *et al.*, 2013).

3.4.2.2 *SETBP1*

Mutations in *SETBP1* were first reported in 2012 and were found to be enriched in atypical CML, and also reported in CMML and MDS/MPN-unclassifiable (Piazza *et al.*, 2012). *SETBP1* has recognized interactions with SET, through the SET-binding domain and mutations in this gene lead to SET stabilization resulting in PP2A inhibition, a phosphatase also inhibited in CML (Piazza *et al.*, 2012). Cells transduced with these mutations also demonstrate a higher proliferative capacity. Mutations in this gene were reported in 24% of atypical CML patients, and also in 17% of secondary AMLs and 15% of CMMLs (Makishima *et al.*, 2013).

3.4.2.3 *CSF3R*

CSF3R is the receptor for colony stimulating factor 3 and is implicated in granulocytic differentiation. Mutations in this gene had previously been reported in congenital neutropenia; and reports emerged in 2013 of a high frequency of recurrent somatic *CSF3R* mutations in chronic neutrophilic leukaemia and atypical CML (59%) (Maxson *et al.*, 2013). Two distinct classes of mutations were identified; truncation and membrane proximal mutations which respectively result in dysregulation of SRC family-TNK kinases and JAK family kinases and show contrasting responses to dasatinib and JAK kinase inhibitors (Maxson *et al.*, 2013). Subsequent studies confirmed that these mutations were highly specific for CNL when strict WHO criteria were applied (Pardanani *et al.*, 2013). This gene has since been incorporated into the WHO classification for chronic neutrophilic leukaemia (Swerdlow *et al.*, 2017).

3.4.2.4 *STAG2*

STAG2 is a key component of the cohesin complex and was reported to be mutated across the spectrum of myeloid malignancies, most commonly AML and MDS (Kon *et al.*, 2013) (see introduction). As the most commonly mutated component of this protein complex, this gene was identified for inclusion in the final panel.

Due to the significant overlap between mutations detected in myeloid malignancies and T-cell lymphomas(Palomero *et al.*, 2014), it was concluded that this panel could also be utilized in those with suspected T-cell lymphoma. As a result 2 genes were added to the sequencing panel – *STAT3* (Jerez *et al.*, 2012) and *RHOA* (Palomero *et al.*, 2014; Sakata-Yanagimoto *et al.*, 2014, p.) which are both frequently mutated in sub-types of T-lymphoproliferative disorders.

The final panel and list of targeted regions is reported below

Gene	Targeted Region
TET2	exon 3-11
DNMT3A	exon 2-23
IDH1	exon 4
IDH2	exon 4 & 5
ASXL1	exon 12
EZH2	exon 2-20
SRSF2	exon 1
U2AF1	exon 2 & 6
ZRSR2	exon 2-11
SF3B1	exon 12-16
NPM1	exon 12
RUNX1	exon 4-8
BCOR	exon 2-15
TP53	exon 5-9
FLT3	exon 20
NRAS	exon 2 & 3
KRAS	exon 2 & 3
CBL	exon 8 & 9
JAK2	exon 12 & 14
CSF3R	exon 14 & 17
STAG2	exon 3-35
WT1	exon 7 & 9
c-KIT	exon 8 & 17
MPL	exon 10
SETBP1	exon 4
CALR	exon 9
STAT3	exon 21 & 22
RHOA	exon 2

Table 3.5. Final Panel of genes and targeted regions for Illumina MiSeq

Primers were again designed for the panel using the D3™ Assay Design service (Fluidigm®, San Francisco, CA, USA). A large number of amplicons were required to cover the updated list of genes and primers were multiplexed with up to 10 primer pairs per well. The design layout for multiplexing was provided by Fluidigm along with the individual primer stock plates. The primers were then multiplexed in-house.

3.4.3 Validation cohort

A validation cohort of 48 samples was identified from those samples previously sequenced on the GS Junior. These samples were of good DNA quality and concentration and contained mutations across the spectrum of genes in the panel (see Appendix 8.7). The mutations identified on the GS Junior had been confirmed and validated by Sanger sequencing.

This validation group was used both to assess the performance of the panel and look for run to run variation. Initial assessment of the primer performance on initial runs identified poorly performing primers (coverage <10x). These were predominantly in GC rich regions and due to persistent poor performance *CEBPA* was removed from the panel. Other primers were re-designed in an attempt to improve coverage. These included *SRSF2, IDH1, IDH2* and *SF3B1*.

To ensure there was no run to run variation, the validation cohort was re-run using the following strategy

- Run 1 - Validation cohort with barcodes 1-48
- Run 2 - Library from Run 1 re-sequenced
- Run 3 - Validation cohort with barcodes 49-96
- Run 4 - Validation cohort with barcodes 1-48 – total process repeated
- Run 5 - Library from Run 4 re-sequenced

Comparison of the output from these sequencing runs confirmed that results were consistent irrespective of barcodes applied. The validation group also enabled comparison between the 2 sequencing platforms along with comparison of the commercial MiSeq analysis software with the new in-house pipeline (see section 3.4.6 and Appendix 8.7).

Once the final panel design was confirmed and initial validation of the processes was complete, proof of principle analysis was performed to confirm reproducibility and the limit of detection.

3.4.4 Dilution studies

Dilution studies were performed on 8 samples with 15 well characterised variants to determine the limit of detection of the panel. These included substitutions, small insertions and deletions ensuring all common variant types were analysed. The mean coverage for this study was 2615x (range 626x-5386x). Libraries were prepared on neat samples as well as 1:2, 1:4, 1:8, 1:16, 1:32 and 1:64 dilutions. The libraries were sequenced and analysed as detailed above. The minimum detectable VAF, detected by sequencing, for each variant was consistent across samples – mean 3% (range 2-7%). All variants were detected below 5%, except for 1 substitution with a lowest VAF of 7%. Furthermore, the predicted and actual VAF for the dilutions were comparable (Fig 3.3).

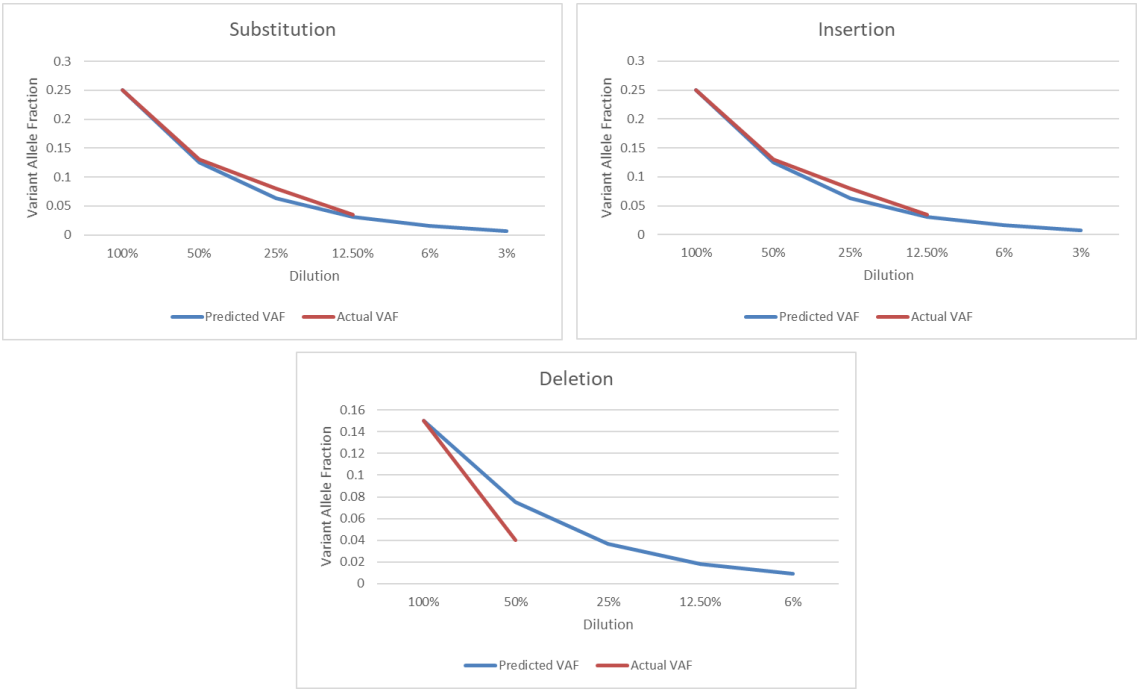


Figure 3.3. Examples of dilution study results. *Each line graph shows the predicted VAF for each variant dilution (blue line) and the actual VAF result (red line) from the MiSeq analysis.*

3.4.5 Repeat analysis

To assess the reproducibility of the sequencing panel, 144 samples were run in duplicate. All VCFs were processed and analysed using the filtering protocol below. Samples that had no variants after filtering were excluded from the analysis.

		VAF >5% 100-500x	VAF >5% >500x	VAF >20% >500x
Run 1	*Positive predictive value*	0.22	0.82	1.0
Run 2		0.47	0.83	1.0
Run 3		0.21	0.70	0.98

Table 3.6. Duplicate analysis of 144 samples

Variants detected with a VAF >20% and coverage >500x were invariably true, however the PPV of the test reduces in areas of low coverage and with VAF 5-20%. Using both the dilution studies and duplicate run analysis, minimum coverage for variant detection was set at 500x and minimum VAF at 5%. Variants with a VAF

between 5-20% or variants in areas of poor coverage are validated by repeat sequencing.

3.4.6 New Analysis pipeline

Initial analysis was performed using the Illumina commercial software (MiSeq reporter) however there was concern that variants were being missed with this pipeline and there were limited options to improve this process.

Accordingly, a bespoke in-house pipeline was developed within HMDS. This pipeline included initial trimming, read alignment and variant calling and consists of the following steps -

- pre-alignment QC using fastQC (v0.11.5, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
- alignment to GRCh37 (hg19) using BWA-mem (v0.7.12, http://bio-bwa.sourceforge.net/)
- aligned QC metrics using picard tools (v2.9.4, https://broadinstitute.github.io/picard/index.html)
- coverage analysis using samtools mpileup (v1.5, http://www.htslib.org/) and bespoke python scripts
- variant calling using VarScan2 (v2.4.3 http://dkoboldt.github.io/varscan/)
- annotation using Alamut Batch (v1.8), COSMIC, VEP

The various components are linked via a python script. The pipeline is hosted on an 8 x 2GHz Intel Xeon E5-2683 v3, 16GB RAM virtual server running Ubuntu 16.01 (LTS), hosted on a Windows enterprise server farm.

The pipeline was validated as part of the validation process described below. Initially the output from the pipeline was also compared to the Illumina commercial analysis tool (MiSeq reporter) on a run by run basis to ensure variants were not being missed.

3.4.7 Variant selection

Due to the nature of the samples received throughout this study, matched germline samples were not available for analysis. It was therefore essential that extensive filtering was performed to exclude artefacts and previously documented single nucleotide polymorphisms (SNPs).

3.4.7.1 Initial variant filtering

It is well recognized that sequencing errors commonly occur in high throughput sequencing particularly when using PCR based library preparations (Schirmer *et al.*, 2015). A significant number of these however are not random, occurring consistently at the same position across each run and it is essential to exclude these variants from further analysis. To assist with this process a list of highly recurrent variants, consistent with artefact, was developed which would be excluded as part of the analysis pipeline. These variants were identified initially using the on-instrument Illumina software 'Sequencing Analysis Viewer' through which each gene could be manually reviewed across each sequencing run. Subsequent run specific variants were identified on a run by run basis by reporting the frequency of each variant across the run in the analysis output. Those variants occurring in >5 samples were reviewed and added to the exclusion list if appropriate (providing these were not hotspot variants).

A complementary list of hot spot variants was compiled which included highly recurrent variants both from the literature and online databases (see Appendix 8.8). Due to the well documented somatic nature of these variants, these were retained by the analysis pipeline irrespective of coverage and VAF.

In addition to the coverage and VAF filters described above the following variants were also excluded –

- All non-coding variants, except essential splice site variants.
- All synonymous variants
- All known polymorphisms, as reported on germline databases

    - These were retained if previously confirmed somatic and recurrently reported in COSMIC and reported in population databases at <0.0014.
- Highly recurrent variants consistent with artefact
    - As described above
- Each individual run was also analysed for run specific, highly recurrent variants likely to be artefact. These were also excluded.

3.4.7.2 CONFIRMATION OF PATHOGENIC VARIANTS

To determine whether the remaining variants were likely pathogenic or variants of unknown significance, a number of resources were utilized and strict criteria were applied. The latter were based on previously applied criteria in one of the large sequencing studies in MDS (Papaemmanuil *et al.*, 2013).

The following resources were used for variant filtering and annotation

- Alamut
    - Data sources include – COSMIC1, Clinvar, DbSNP, gNomad, 1000 genomes, CentoMD, HGMD, DMuDB, SwissVar
- Correlation with an ever-expanding in-house database of >3000 myeloid malignancy samples referred to HMDS
- Online search tools – PubMed, Google

Alamut is a comprehensive variant annotation tool which provides information on the effects of variants on human genes, known variants and mutations and missense and splicing variant predictions.

The following criteria were applied to ensure only high confidence variants were included (Papaemmanuil *et al.*, 2013).

Inclusion Criteria

- Variants previously reported in the literature, confirmed in haematopoietic tissue (as reported in COSMIC)
    - Well documented hotspot variants were retained irrespective of coverage however in those with low coverage these were only included in the analysis if reproducible across 2 separate runs.
- Truncating variants (nonsense, frameshift indels and essential splice site variants) in genes implicated in myeloid pathogenesis by loss-of-function – *ASXL1, BCOR, DNMT3A, EZH2, RUNX1, STAG2, TET2, TP53, WT1*.
- Previously unreported variants that cluster (±3aa) with a hotspot variant or occur at a recurrently mutated amino acid position, providing the variant is not reported in population databases.

3.4.8 Validation of Myeloid panel

Once all these processes were in place, internal and external validation was performed specifically looking at variant detection.

3.4.8.1 Internal validation – comparison to Roche 454 Junior and Sanger sequencing

The sequencing panel was validated internally on a cohort of 48 samples with 139 well characterised variants across all genes included in the panel. These variants had been validated on 2 different sequencing platforms including the Illumina MiSeq, Roche 454 Junior and Sanger sequencing. The samples were used both to establish the wet laboratory procedures and to compare the in-house Bioinformatics Pipeline to Illumina commercial software.

Of the total 139 variants, 128 were detected using the Illumina MiSeq and in-house pipeline. Of the 11 variants not identified, 7 were *SRSF2* hotspot variants with very low coverage on the MiSeq. *SRSF2* is a GC rich gene and is notoriously difficult to amplify using an amplicon-based library preparation. In light of this, all samples

with *SRSF2* coverage <20x undergo Sanger sequencing for this hotspot location. The remaining 4 variants included 3 in *TET2* at areas of low coverage and a large 52bp deletion in *CALR*. The panel was therefore judged to perform well on internal validation with a sensitivity of 97% (when *SRSF2* is excluded). Variants can however, be undetectable in areas of low coverage, and the panel fails to detect large deletions.

3.4.8.2 External validation – comparison to a commercial sequencing platform and pipeline

To evaluate the robustness of the myeloid sequencing and analysis pipeline, concordance with an external sequencing facility was assessed. 37 samples from patients with a confirmed myeloid malignancy were sent to a commercial sequencing facility as part of a collaborative project. Although the commercial company used a larger targeted panel, the HMDS panel formed a subset of the genes included. The commercial partner also used a different library capture protocol (Agilent SureSelect), sequencing platform and bioinformatics pipeline. Both BAM and VCF files were made available to compare the independent variant calls.

A total of 96 variants were detected across the 33 samples in the regions covered by both panels. There was 80% concordance (77/96 variants) between the 2 panels with 11 variants detected only by the in-house HMDS panel and 4 variants only by the commercial panel. Data was unavailable for 4 variants. Of the discordant results, 12/15 were due either to low coverage at the respective position or a VAF below the threshold for reporting. The remaining 3 were filtered by the respective bioinformatic pipeline. It was therefore concluded that the in-house sequencing panel performed well when compared with an external partner. A small number of variants may be missed at the limit of detection and caution is required in the setting of low coverage.

3.4.9 Run Quality Control

To determine the quality of each run and identify poor quality samples, a run level and sample level quality control (QC) strategy was developed.

To gather QC metrics for sequencing runs, Picard an open source toolkit was used (https://broadinstitute.github.io/picard/index.html). Picard is a set of Java command line tools for manipulating high-throughput sequencing data (HTS) data and formats.

Picard version 1.129 tools CollectAlignmentSummaryMetrics and CalculateHsMetrics were used to analyse the sorted bam files, along with the panel manifest target locations.

The metrics for 64 runs (3072 samples) were analysed to find consistent parameters which distinguish between known failed runs (n=1) and known failed samples (n=9).

3.4.9.1 Run level metrics

Initially, metrics were identified at a run level which could distinguish the known failed run from the successful runs. The failed run had clustered poorly when analysed on the MiSeq.

- PF_READS (The number of reads that pass the vendor's filter)

  Threshold: 500,000. Runs with a mean PF_READS of below 500,000 should be flagged as a failed run.
- PCT_TARGET_BASES_100X (The percentage of ALL target bases achieving 100X or greater coverage)

  Threshold: Runs with a mean PCT_TARGET_BASES_100X below 0.9 will be flagged as a failed run.

3.4.9.2. SAMPLE LEVEL METRICS.

Metrics were then identified that could distinguish known failed samples from good quality samples within a machine run. The failed samples were identified based on DNA quantity, number of low-level non-recurring variants and coverage.

- ZERO_CVG_TARGETS_PCT (The number of targets that did not reach coverage=2 over any base)

  Threshold: mean + 2sd - if the sample value is greater than 2 x standard deviation above the mean for the run, the sample is flagged as an outlier.
- FOLD_80_BASE_PENALTY (The fold over-coverage necessary to raise 80% of bases in "non-zero-cvg" targets to the mean coverage level in those targets)

  Threshold: mean + 2sd - if the sample value is greater than 2 x standard deviation above the mean for the run, the sample is flagged as an outlier.
- PCT_TARGET_BASES_100X (The percentage of ALL target bases achieving 100X or greater coverage)

  Threshold: mean - 2sd - if the sample value is less than 2 x standard deviation below the mean for the run, the sample is flagged as an outlier.
- GC_DROPOUT (A measure of how undercovered >= 50% GC regions are relative to the mean. For each GC bin [50..100] we calculate a = % of target territory, and b = % of aligned reads aligned to these targets. GC DROPOUT is then abs(sum(a-b when a-b < 0)). E.g. if the value is 5% this implies that 5% of total reads that should have mapped to GC>=50% regions mapped elsewhere)

  Threshold: mean + 2sd - if the sample value is greater than 2 x standard deviation above the mean for the run, the sample is flagged as an outlier.

On review, no single metric was sufficient to flag a sample as a failure for a sequencing run. However, >2 metrics captured all failed samples.

In addition to the sequencing metrics, input DNA concentration is recorded, along with the number of annotated variants, in the VCF files. DNA concentration

correlates strongly with sequencing quality and samples with <5ng/µl consistently fail. The input DNA concentration was therefore included in the sample QC metrics. This was recorded, along with a flag, to show three levels. GREEN is above 15 ng/µl, AMBER is between 5 and 15 ng/µl, and RED is below 5 ng/µl.

For QC purposes it is also important to have an overview of coverage for each sample across the panel. This is particularly relevant with Fluidigm library preparation when single amplicons can fail due to issues within the microfluidics at the point of sample and primer loading. It is therefore necessary to know which amplicons have failed for each sample, and the number of failed amplicons and their labels are added to the QC record. Due to the importance of identifying well-characterized common 'hotspot' mutations, it is essential to know if these areas were well covered during the sequencing run. For each sample, the % of the hotspot locations which have over 100 reads are reported. The list of the genes where hotspot coverage fails to meet reporting criteria are included in the QC report.

3.4.9.3 QC REPORT

For each run, the QC parameters described are recorded in a tab-separated text file which can be imported easily into Microsoft Excel. In the report if a sample falls outside of the threshold for an individual sequencing metric, a '1' is recorded, otherwise '0'. The sequencing QC parameters are collated with the other metrics described to for the QC report.

Sample	ng.ul	dna.flag	num.var	zeroCVG	fold80	targets100x	gcDropout	QC_pass	hotspots .percent	hotspots. below.100	targets Failed
XX	32.3	GREEN	6	0	0	0	0	PASS	1		4
XX	104.08	GREEN	6	0	0	0	0	PASS	1		4
XX	11.442	AMBER	9	1	0	1	1	FAIL	0.93	MPL	7
XX	90.52	GREEN	4	0	0	0	0	PASS	1		1
XX	108.84	GREEN	1	0	0	0	0	PASS	1		2

Fig. 3.3. Example of QC report

3.5 Discussion

It is clear from the literature that many of the genes recurrently mutated in myeloid malignancies have prognostic relevance in those with confirmed disease. Targeted sequencing could also provide confirmatory evidence of disease in the diagnostic setting due to the high frequency of mutations across the chronic myeloid malignancies. The panel was designed to include the most commonly mutated genes across all myeloid malignancies, providing a platform to investigate the diagnostic potential. This would also capture important prognostic information, and importantly be applicable in a large number of patients. The latter is particularly relevant in the setting of a routine clinical laboratory, where high throughput and low costs are essential to ensure results are available in real time and the test is cost efficient.

The panel does not however include all genes reported across these malignancies. As demonstrated in the large MDS sequencing studies, and similarly across other disease types, only a small number of genes are mutated in >10% of cases and there is a long tail of genes mutated at low frequency (Haferlach *et al.*, 2013; Papaemmanuil *et al.*, 2013). At the time when this panel was in development it was not logistically or economically viable to include all the relevant genes in a single panel, particularly for the application into routine practice. The core genes which were included are, however, those which provide the most clinically relevant information both throughout this research and still to this day.

The panel performed well on both internal and external validation, though the limitations of a PCR based library preparation are recognized. The need for validation of low-level variants does add to the workload and turn-around times, but importantly ensures that the variants detected can be reported with confidence. The additional validation steps performed as part of this research provided the extensive information needed to move this platform into routine practice.

The most challenging aspect of the sequencing process remains that of variant annotation. As described above, there are many tools available which provide evidence for each variant, but there is still a subjective element to this process, particularly in the absence of a matched germline sample. Furthermore, online somatic databases including COSMIC are contaminated with germline variants while confirmed somatic variants are also reported in germline databases. It is also likely that variants are being under-reported in the literature due to the reliance on somatic databases for pathogenic confirmation. Historically, guidance on this area has been focused on constitutional variants, but more recently guidelines have been produced for the interpretation and reporting of variants in cancer samples (Li *et al.*, 2017). For this research, variant annotation was aligned with that performed in one of the largest MDS sequencing studies ensuring the data can be compared with the reported literature.

The following chapters will detail the studies performed to assess the clinical utility of this sequencing panel, particularly with respect to diagnosis in cytopenic patients and those presenting with a monocytosis.

CHAPTER 4: COHORT 1 – DETERMINING THE FREQUENCY OF CLONAL ABNORMALITIES IN PATIENTS WITH PRE-CLINICAL MDS

These data have been published as a plenary paper in Blood with an associated commentary (see Appendix 8.9)

4.1 Introduction

As previously described, the diagnosis of MDS, particularly in those without increased blasts, remains problematic due to the reliance on morphological assessment for diagnosis (Swerdlow *et al.*, 2017). There is reported poor inter-observer concordance when recognising dysplasia (Parmentier *et al.*, 2012), and numerous non-neoplastic conditions which can mimic MDS (Steensma, 2012) therefore impairing the sensitivity and specificity of this technique. To date the only objective marker of neoplastic haematopoiesis has been provided by conventional cytogenetics and/or fluorescence in situ hybridisation (FISH) (Schanz *et al.*, 2011, 2012) though these techniques are uninformative in a large proportion of cases.

With the advent of high throughput sequencing and array-based cytogenetics, pathogenic abnormalities can now be detected in a significant majority of MDS patients diagnosed by the currently established diagnostic criteria (Malcovati *et al.*, 2013). Large sequencing panels targeting recurrently mutated genes have identified driver mutations in up to 90% of patients (Haferlach *et al.*, 2013) and the addition of single nucleotide polymorphism arrays (SNP-A) to conventional cytogenetics has increased the diagnostic yield by identifying small copy number variants (CNV) and areas of copy neutral loss of heterozygosity (CN-LOH)(Tiu *et al.*, 2011). To date these techniques have been used to explore disease pathogenesis and identify prognostic markers. However, with the high frequency of abnormalities reported it may be feasible to use the presence of a pathogenic genetic abnormality as a core criterion for the diagnosis of MDS. This could substantially increase the

reproducibility of diagnosis and the sensitivity of detecting patients at an early stage of disease.

The diagnostic utility of mutational analysis has however been questioned due to reports of somatic mutations in healthy individuals (Genovese *et al.*, 2014; Jaiswal *et al.*, 2014; Xie *et al.*, 2014; McKerrell *et al.*, 2015). These mutations involve genes commonly mutated in myeloid malignancies, and it is estimated that they occur in ~4% of the population, though increase significantly with age, being found in over 10% of people over 70 (Jaiswal *et al.*, 2014). These individuals do have an increased risk of progression to a haematological malignancy, though only at a rate of ~1% per year, analogous to the transformation levels seen for MGUS (monoclonal gammopathy of undetermined significance) to myeloma (Jaiswal *et al.*, 2014). Distinction between those with age related clonal haematopoiesis and those with a disease defining mutation in the absence of definitive morphological disease would, therefore, be difficult.

The aim of this initial analysis was to molecularly characterise patients with the most clinically significant disease who fail to meet diagnostic criteria using conventional techniques. To achieve this, a cohort of cytopenic patients was identified who, despite having an initial bone marrow with non-diagnostic features, went on to develop progressive dysplasia or AML. In this setting it was postulated that the initial sample represented early or pre-clinical MDS. The paired samples were then studied using targeted sequencing and SNP-A analysis to determine if pathogenic abnormalities could be detected even in the absence of diagnostic features using current methods. It was hypothesised that characterising these patients would provide potential criteria to distinguish pre-clinical MDS from healthy individuals and importantly detect those patients at high risk of disease progression.

4.2 Methods

4.2.1 Patient Samples

A search was performed for all new diagnoses of MDS (with multilineage dysplasia) or AML made at HMDS, between 2004 and 2012 who had undergone previous non-diagnostic bone marrow analysis for investigation of cytopenia. The paired samples, termed pre-diagnostic and diagnostic, were dual reported by the same team of experienced haematopathologists. All diagnostic samples were classified according to the relevant WHO classification system at the time of diagnosis.

A total of 82 patients were identified. This represented 1.7% of patients referred for investigation of cytopenia with a non-diagnostic marrow during this time period (n=4835). Sixty-nine patients had adequate DNA from both samples for molecular analysis. The demographics of these 69 individuals are presented in Table 4.1. Of note, the initial presentation samples include both those with insufficient morphological features for diagnosis and those in whom morphological assessment was limited by poor sample quality. Blood count parameters contemporaneous with both bone marrow biopsies were available on 47 patients.

4.2.2 DNA extraction and targeted amplicon sequencing

DNA was extracted from fresh peripheral blood or bone marrow total nucleated cells, or retrieved from unstained smears as described in Chapter 2.

Targeted gene sequencing was performed initially on the GS Junior using the panel targeting 20 genes and was subsequently repeated using the more comprehensive 26 gene panel designed for the MiSeq. Twelve and forty-eight barcoded patient samples were pooled for the GS Junior and MiSeq and subjected to 300bp and 150bp paired-end sequencing respectively. Both library preparation and sequencing were performed as per the manufacturer's instructions. The mean coverage on GS Junior was 97x (range 30-354x) and 1297x (range 578-1816x) on MiSeq.

Table 4.1. Characteristics of patients included in cohort 1.

Patient Demographics				
No. Of patients		69		
Male:Female		46:23		
Median age (range)				
	Initial Presentation	74 (18-87)		
	Diagnosis	75 (18-88)		
Initial Presentation				
	Anaemia of chronic disease	5		
	Iron deficiency anaemia	2		
	ITP	3		
	No evidence of disease	12		
	Reactive changes only	27		
	Suspicious of malignancy†	17		
	Inadequate sample	3		
Final Diagnosis				
	MDS with 5q-	1		
	RCMD	29		
	RAEB-1	9		
	RAEB-2	9		
	AML	21		
		RCMD*(n=30)	**RAEB (n=18)**	**AML (n=21)**
No. of days between samples; median (range)		606 (19-2305)	439 (27-2037)	403 (26-2484)
Blood Count Parameters; median (range)				
Initial presentation				
	Hb (g/dL)	10.0 (6.1-15.1)	9.9 (7.2-13.2)	11.3 (7.5-13.5)
	WCC (10^9/L)	4.5 (1.9-18.2)	6.1 (1.7-15.2)	3.9 (2.6-10.1)
	Plt (10^9/L)	113 (15-409)	133 (10-226)	97 (46-364)
At diagnosis				
	Hb (g/dL)	9.1 (5.9-12.9)	7.9 (6.6-11.4)	9.1 (6.5-12.0)
	WCC (10^9/L)	3.6 (1.5-23.5)	3.8 (1.1-72.3)	8.1 (1.9-144.2)
	Plt (10^9/L)	91 (11-427)	94.5 (13-188)	72.5 (16-402)

ITP, immune thrombocytopenia; RCMD, refractory cytopenia with multilineage dysplasia; RAEB, refractory anaemia with excess blasts

† Includes patients with insufficient morphological features (n=10) and those with suboptimal sample quality (n=7)

* Also includes MDS with 5q- patient

4.2.2.1 VARIANT CALLING AND ANALYSIS

Initial trimming, read alignment and variant calling were performed using respective commercial software – GS Amplicon Variant Analyzer (Roche) and MiSeq reporter (Illumina) with paired end reads aligned to the reference amplicon sequence. Variants were analysed using the Ensembl Variant Effect Predictor (VEP) software (McLaren *et al.*, 2010). All synonymous, non-coding variants and germline polymorphisms were excluded. The latter were retained if previously confirmed somatic and recurrently reported in COSMIC database (Forbes *et al.*, 2011).

4.2.2.2 VALIDATION

Variants were accepted if identified by at least 2 methods – MiSeq, GS Junior - or by conventional Sanger sequencing. Sanger sequencing was performed for *SRSF2* across areas of poor coverage as previously described. *JAK2* Val617Phe mutations were confirmed using an ASO based approach as discussed in Chapter 2.

All patients were screened for *NPM1* mutations and *FLT3* internal tandem duplications (ITD) using conventional fragment analysis.

4.2.3 Single Nucleotide Polymorphism array analysis

Array based whole genome scanning was performed on all diagnostic samples using the Human CytoSNP-12 BeadChip according to manufacturer instructions. BeadChips were scanned using the BeadArray™ Reader. Data was visualised using Karyostudio™ software and reported as per published guidelines (See Chapter 2).

In those patients with a confirmed abnormality in the diagnostic sample, SNP-A analysis was performed on the corresponding previous non-diagnostic sample. Poor quality samples precluded detailed analysis on 7 diagnostic samples.

4.2.4 Statistical Analysis

Comparison of mutations between disease subgroups was performed on genes sorted into functional pathways and analysed using a parallel χ^2 test and univariate logistic regression. Survival curves were produced using the Kaplan Meier method (censored 01/06/2015) and the impact of abnormalities on overall survival and risk

of progression were estimated using Cox regression. Pairwise associations were tested with the Fisher exact test corrected for multiple testing using the Benjamini–Hochberg procedure, with false discovery rate (FDR) set at 5%. Associations between genetic mutations and progression were investigated with Fisher's exact test and by Cox regression within diagnostic groups.

4.3 Results

4.3.1 Mutational analysis demonstrates abnormalities in the majority of pre-diagnostic samples with a paucity of structural variants

A driver mutation and/or structural abnormality was identified in 91% (n=63/69) of pre-diagnostic samples, with all but one (62/63) harbouring a somatic mutation. The remaining patient had an isolated cryptic MLL rearrangement, which was retrospectively identified by FISH following progression to MLL-rearranged AML.

A total of 133 mutations (Fig. 4.1) were identified across 62 patients with a median of 2 mutations per sample (range 1-5; Fig. 4.4A). Mutations in epigenetic regulators and spliceosome genes were most commonly identified in the pre-diagnostic samples, with all but 3 patients (59/62) harbouring a mutation in one of these pathways. *TET2, SRSF2* and *ASXL1* were most frequently mutated being identified in 39%, 26% and 20% of patients respectively. While mutations in these pathways were common, they rarely occurred in isolation, with 75% patients harbouring >1 mutation. The spectrum of mutations at this time-point mirrored that reported in large MDS populations (Haferlach *et al.*, 2013; Papaemmanuil *et al.*, 2013) with the exception of *SF3B1* (n=3), though these mutations are strongly associated with ring sideroblasts which are easily identified morphologically (Papaemmanuil *et al.*, 2011). To determine if certain gene mutations commonly co-occurred or were mutually exclusive we performed pairwise association on both pre-diagnostic and diagnostic samples, though failed to identify any significant relationships, possibly due to small sample size.

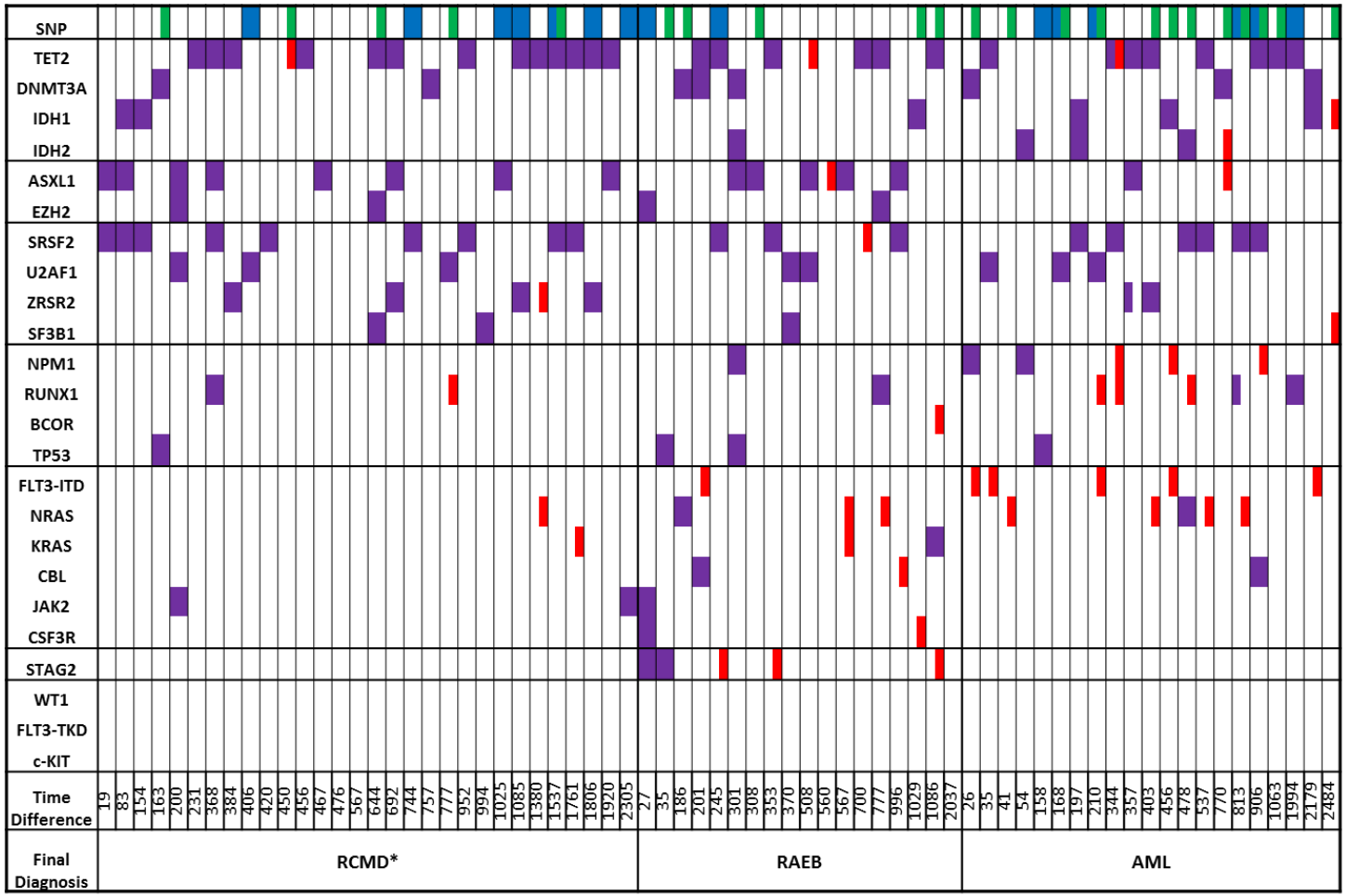


Figure 4.1. Mutations and structural variants identified per patient, grouped by final disease category and ordered by time between samples from shortest to longest (days)*. Purple depicts mutations identified in the non-diagnostic sample with red depicting those acquired at diagnosis. Similarly blue represents structural variants identified by SNP array in the non-diagnostic sample and green those acquired at diagnosis. *Includes the single patient with MDS with 5q- (time between samples 2305 days)*

Structural abnormalities were much less frequent being identified in only 23% (16/69) of pre-diagnostic samples. Most (14/16) possessed a single abnormality (range 1-7) with 9 regions of LOH and 13 copy number variants (CNV), most commonly deletions, identified. Recurrent abnormalities included trisomy 8, del20q, del5q, LOH4q and LOH11q. The majority of LOH regions were co-existent with a mutation in the alternate allele including *TET2* with LOH4q, *CBL* with LOH11q, *TP53* with LOH17p and *EZH2* with LOH7q.

Conventional cytogenetics had been performed in only 10 patients at this timepoint, all of which demonstrated a normal karyotype.

4.3.2 Mutational profile in pre-diagnostic samples differs from healthy individuals

When comparing the mutational profile in this patient cohort with that in healthy individuals distinct differences were noted. *DNMT3A* was by far the most frequently mutated gene in the healthy population (Genovese *et al.*, 2014; Jaiswal *et al.*, 2014; Xie *et al.*, 2014; McKerrell *et al.*, 2015), however this was only seen in 10% (n=7) of our patient cohort. There was a median of 2 mutations detected per patient (range 1-5) and the frequency of multiple mutations was significantly greater here than reported in healthy individuals (≥2 mutations; 64%vs.8% (Jaiswal *et al.*, 2014)). The median variant allele fraction (VAF), and inferred clone size in our series was also notably greater at 40% (range 2.31-100%) than the reported median of 9-10% in healthy individuals(Jaiswal *et al.*, 2014). Importantly, only 1 patient harboured an isolated mutation with a VAF less than 20%. These differences suggest that the clone must expand to an appropriate level and/or acquire cooperating mutations to cause cytopenias and subsequent disease. This is supported by the greater mean VAF (25.2%) observed in healthy individuals who subsequently developed a haematological malignancy(Jaiswal *et al.*, 2014). A driver gene mutation with a high allele fraction (>20%) and/or co-occurring mutations may therefore be disease defining but in the very least identifies clinically significant clonality requiring close follow-up as a minimum.

4.3.3 Progression to RAEB/AML is associated with acquisition of mutations/structural abnormalities

In 30 patients, new mutations were identified in the diagnostic samples that had not been detected in the pre-diagnostic samples. This included 2 of the patients who were mutation negative at initial investigation, increasing the final mutation frequency to 94% (65/69). This correlated strongly with progression to RAEB (n=11) and AML (n=15) with only 4 patients with a final diagnosis of RCMD having acquired a mutation. Mutations of transcription factors and cell signalling genes were most commonly acquired with *NRAS*, *FLT3* and *RUNX1* accounting for 18%, 15% and 10% of total acquired mutations respectively (Fig. 4.2). Importantly *FLT3* ITD was only identified at the point of diagnosis and only in those with RAEB or AML.

Structural abnormalities were acquired in 21 patients resulting in an overall frequency of 46% (32/69) at the point of diagnosis. These were again acquired most frequently in those patients progressing to AML (n=11; RAEB n=5; RCMD n=5). A median of 2 abnormalities were acquired (range 1-5) between samples though in contrast to the pre-diagnostic sample there was a striking predominance of CNVs (31 vs 7; CNV vs LOH).

4.3.4 Mutations in the pre-diagnostic sample predict disease progression and OS

To determine whether parameters in the pre-diagnostic sample could predict disease progression both blood count data and mutational spectrum were examined at this time point. Overall, the rate of progression to AML/RAEB was faster than to a diagnosis of RCMD (median 403 vs. 606 days respectively; HR, 3.7; 95% CI 2.1-6.6; $p < 0.001$). In those with blood count data available there was no significant difference in parameters between those that progressed to the 3 final disease subgroups; RCMD, RAEB and AML.

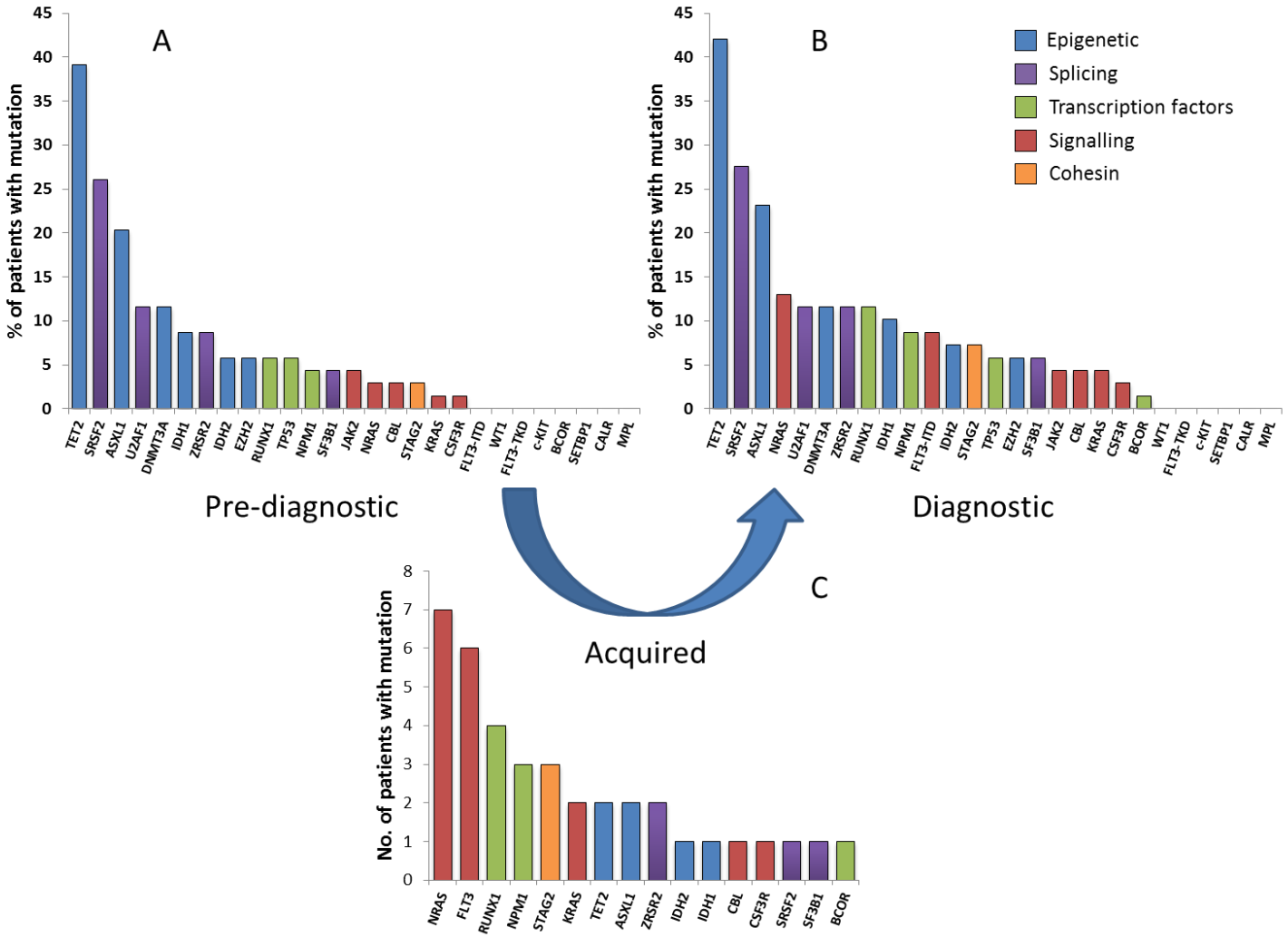


Figure 4.2. Frequency of driver mutations across all 69 patients in the (A) pre-diagnostic and (B) diagnostic samples coloured according to functional pathway. (C) Spectrum of mutations acquired between samples again coloured by functional pathway.

To assess any difference in the spectrum of mutations between these groups we performed parallel χ^2 tests and univariate logistic regression on genes sorted into functional pathways. Mutations in chromatin modifiers (*ASXL1, EZH2*) were found to be significantly less frequent in those patients who progressed to AML (p=0.032). There was no statistically significant difference in the frequency of mutations in other pathways. The presence of a structural abnormality at this time point did not impact on overall survival or the rate of disease progression.

The most frequently mutated genes (present in >5% of individuals; *TET2, SRSF2, ASXL1, U2AF1, DNMT3A, IDH1, ZRSR2, IDH2, TP53, RUNX1, EZH2*) were analysed to determine whether certain mutations were associated with progression to high risk MDS (RAEB) or AML as opposed to RCMD and also whether the rate of progression to either diagnostic group was affected by such mutations.

All four patients with *IDH2* mutations progressed to AML/RAEB providing weak evidence of a possible association (p=0.052). While *NPM1, CBL* and *NRAS* mutations were not included in this analysis due to low frequency, all patients carrying these mutations progressed to AML or RAEB. Specific mutations were found to be associated with a more rapid time from the pre-diagnostic analysis to diagnosis/progression. In those that progressed to AML/RAEB, this was true for both *IDH2* mutations (HR, 4.2; 95% CI 1.3-13.8; p = 0.017) and *TP53* mutations (HR, 5.5; 95% CI 1.1-27.7; p = 0.038). Similarly, there was evidence that both *TP53 and IDH1* mutations were associated with a more rapid time to diagnosis of RCMD (*TP53*; HR, 28.5; 95% CI 1.8-455.6; p = 0.018. *IDH1*; HR, 11.6; 95% CI 2.1-64.2; p = 0.0049). These associations with time to progression were adjusted linearly and non-linearly for age; all associations remained except for that between *TP53* and time to diagnosis of RCMD. Caution should be observed in interpreting these results as the number of patients with such mutations is small.

Survival data were available on 59 patients, of which only 10 were alive at the cut-off time point for data analysis. Median survival from the non-diagnostic sample was 43.6 months (95% CI 33.8-55.8) and from diagnosis, 13 months (95% CI 9.9-

24.6). Patients diagnosed with AML had a median survival of only 1.28 months (95% CI - 0.789-12.625) which was significantly shorter than those diagnosed with either RCMD (29.2 months, p<0.001; Fig 4.3) or RAEB (12.48 months, p=0.042). Interestingly the survival of the AML group was comparable to the RAEB group if determined from the time of the pre-diagnostic sample (29.1mths vs 35.3mths, p=0.442) (see Fig. 4.3). While this can clearly not be viewed as a survival advantage of early diagnosis it does highlight a potential period for earlier intervention in a group of patients with a poor prognosis.

Cox regression univariate analysis from the pre-diagnostic sample showed that, *TP53* (HR, 21.68; 95% CI 4.72-99.64, p<0.001), *U2AF1* (HR, 2.63; 95% CI 1.0-6.4, p=0.049) mutations, and the number of mutations identified (HR, 1.447; 95% CI 1.12-1.88; p=0.006) were all associated with a significantly poorer overall survival; with the latter most significant in those with >3 mutations (Fig. 4.4C).

4.3.5 Patterns of progression can be revealed by assessing changes in the variant allele fraction between samples

By assessing changes in the variant allele fraction (VAF) between samples, differing patterns of progression could be identified in patients with a demonstrable mutation pre-diagnosis. Examples of these are presented in Fig. 4.5. The majority of patients progressing to AML/RAEB showed acquisition of new mutations which in some cases drove expansion of the dominant clone (Fig. 4.5A), in others the pre-diagnostic mutation remained stable (Fig. 4.5B) while in further patients there was a decrease in the VAF of co-existing mutations (Fig. 4.5C). The latter likely represents competition between clonal populations and in the example in Fig. 4.5C the *SRSF2* mutation remains stable however the *RUNX1* mutation identified in the pre-diagnostic sample disappears with the acquisition of an *NRAS* mutation. While a proportion of RAEB/AML patients did not acquire mutations, there was evidence of progression demonstrated by growth of the dominant clone or acquisition of a structural

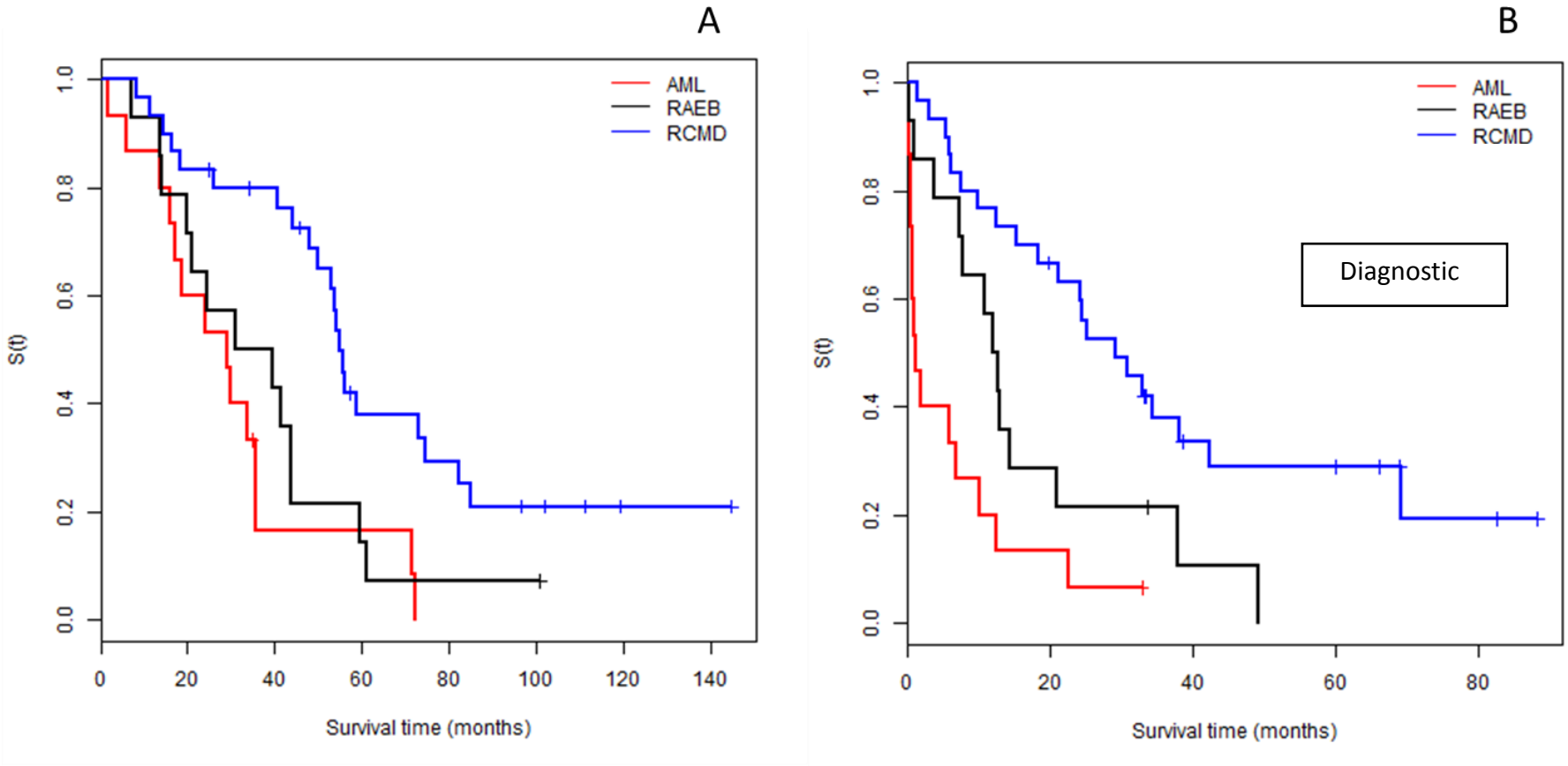


Figure 4.3. Overall survival in patients grouped by final disease category. *This was determined from time of (A) pre-diagnostic sample and (B) diagnostic sample. The survival of the AML group is comparable to the RAEB group if determined from the pre-diagnostic sample (p=0.442) though significantly worse if determined from the diagnostic sample (p<0.001). This highlights a potential period for earlier intervention.*

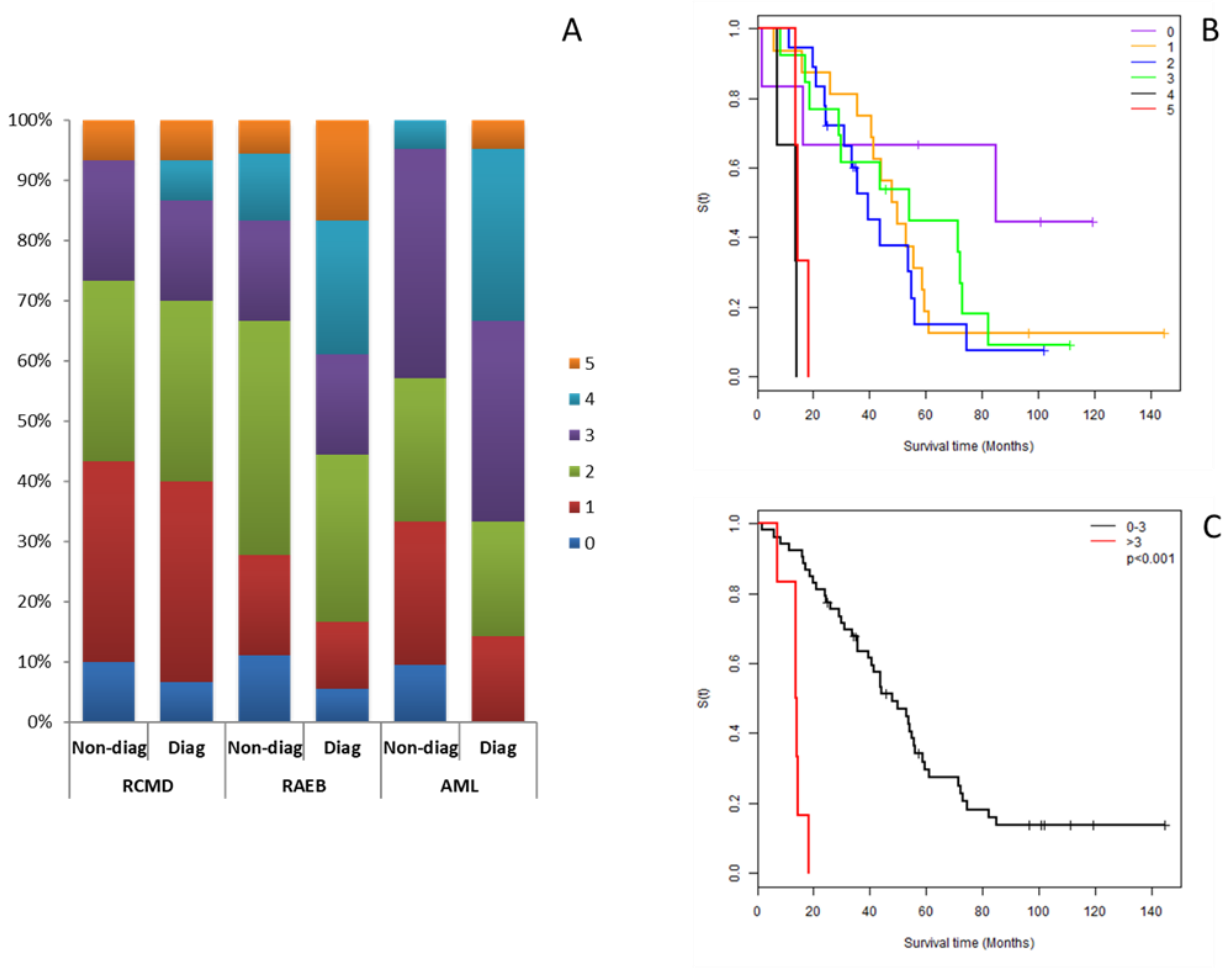


Figure 4.4. Distribution of mutations and correlation with overall survival. *(A) Distribution of no. of mutations according to final disease subgroups and across both non-diagnostic and diagnostic samples. (B) Overall survival in all patients according to no. of mutations identified in non-diagnostic sample and (C) when grouped into those with <3 and >3 mutations.*

abnormality. One patient showed only a decrease in the VAF of both mutations identified in the pre-diagnostic sample suggesting the emergence of a subclone containing a mutation not targeted in our panel.

In contrast, those with a final diagnosis of RCMD rarely acquired mutations and while a significant number showed no change in mutation VAF between samples a proportion showed clonal expansion at the point of diagnosis (Fig. 4.5D).
Copy number analysis provides additional information when interpreting changes in VAFs. Patient HMDS55 showed an increase in the VAF of a *TP53* mutation from 42.3% to 93.9% (Fig. 4.5E) though this occurred in the context of a del17p which was acquired as part of a complex karyotype between samples.

4.4 Discussion

Using a combination of sequencing and array-based cytogenetics this retrospective study has confirmed that a high frequency of abnormalities, predominantly driver mutations, can be detected in samples with indeterminate morphology from patients who were subsequently diagnosed with a myeloid malignancy. Importantly these techniques appear somewhat less affected by sample quality with results being obtained even in inadequate/suboptimal bone marrow samples.

This analysis used strict criteria along with multi-platform validation to ensure the abnormalities reported were genuine. However, it is likely that some abnormalities were excluded during this process, particularly those from small clones. It is also likely that a proportion of patients harbour mutations in genes not targeted with the panel.

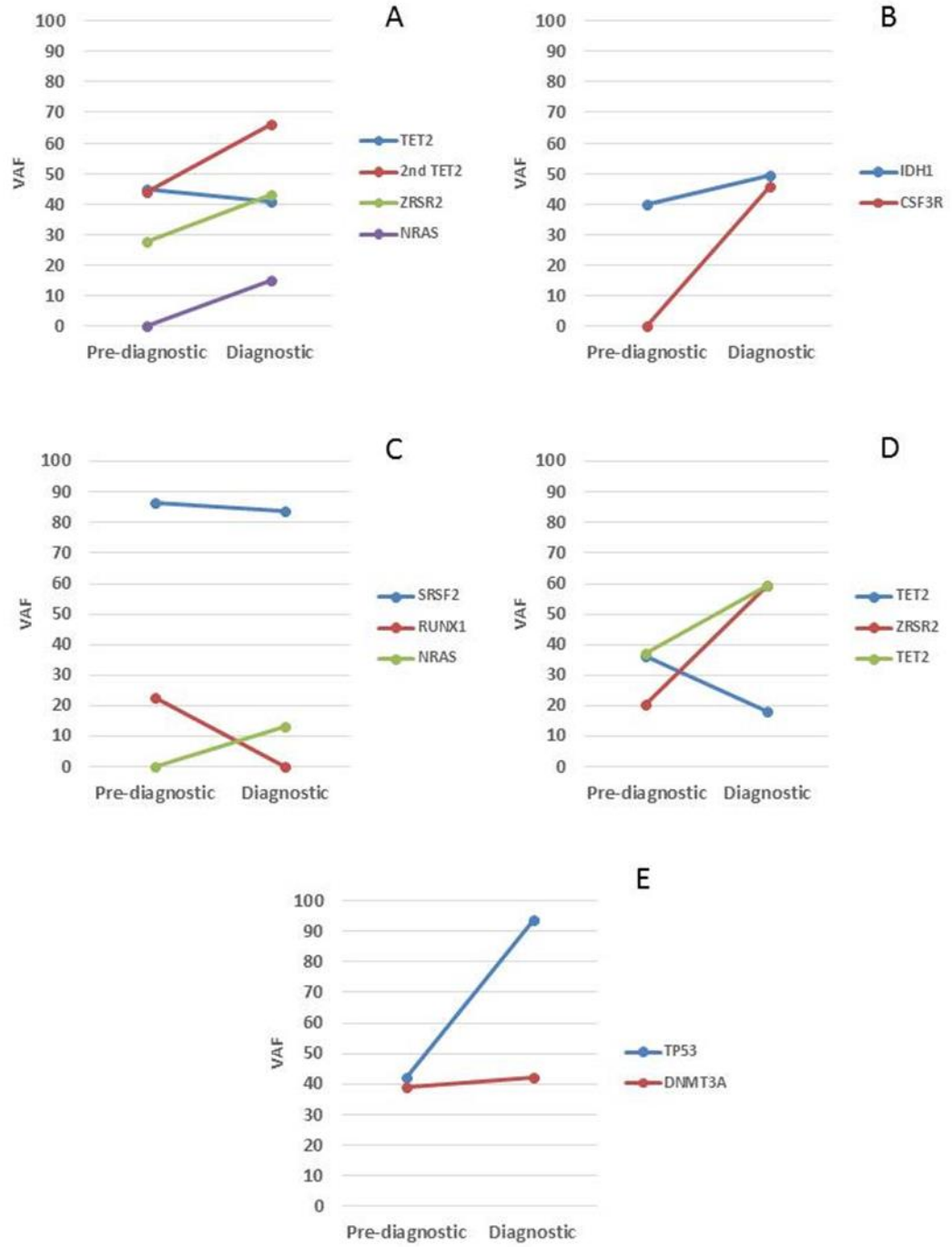


Figure 4.5. Changes in variant allele fraction between samples. *(A) Patient HMDS63 (AML) who showed expansion of the dominant TET2 clone following acquisition of an NRAS mutation. The ZRSR2 appears to be subclonal within this TET2 clone. The additional TET2 mutation remains stable suggesting this is non-ancestral. (B) Patient HMDS43 (RAEB) who acquired a CSF3R mutation at diagnosis though the pre-diagnostic IDH1 mutation remained constant. (C) Patient HMDS68 (AML) who showed likely competing subclones with loss of a RUNX1 mutation and acquisition of an NRAS mutation at diagnosis. (D) Patient HMDS22 (RCMD) showed expansion of both TET2 and ZRSR2 mutations between samples with a decrease in the VAF of the 2nd TET2 mutation again suggesting this is non-ancestral. (E) Patient HMDS55 (RCMD) showed an increase in the VAF of a TP53 mutation in the setting of del17p.*

The high mutation frequency in this cohort contrasts to previous reports in early MDS in which driver mutations were detected in ~70% of refractory anaemia (RA) and RCMD cases(Bejar *et al.*, 2012; Haferlach *et al.*, 2013). In light of the subjective nature of current diagnostic approaches it is possible that these cohorts are contaminated with non-MDS cases, which could explain a proportion of these negative results. Alternatively, mutation frequency may be significantly higher in those more likely to show disease progression. The results from this study suggest however that driver mutations are frequently identified even in early disease, and particularly in those with the most clinically significant disease.

The spectrum of mutations identified in the pre-diagnostic sample mirrors that reported in MDS populations with a predominance of DNA methylation and RNA splicing gene mutations. With the benefit of sampling early in the natural history of the disease the findings are consistent with these being early events in disease pathogenesis as previously speculated (Haferlach *et al.*, 2013; Papaemmanuil *et al.*, 2013). The high frequency of additional abnormalities also supports the hypothesis that a 2nd hit is required for the onset of disease(Busque *et al.*, 2012). Notably *SF3B1* mutations were rarely detected in this patient cohort (n=3) in contrast to much larger studies (Haferlach *et al.*, 2013; Papaemmanuil *et al.*, 2013). These mutations are however strongly associated with ring sideroblasts (Papaemmanuil *et al.*, 2011) which are easily identified morphologically.

In contrast there was a paucity of structural variants in the pre-diagnostic sample and this, in conjunction with increasing abnormalities between samples, is consistent with these being secondary events in MDS disease pathogenesis(Cazzola, Della Porta and Malcovati, 2013) possibly as a result of an underlying pathogenic mutation. Mutations in functional pathways are thought to lead to chromosomal instability arising either as a single cataclysmic event, such as in the context of *TP53* mutations(Rausch *et al.*, 2012) or as a more gradual degeneration(Lindsley and Ebert, 2013). Transcription factor and signalling pathway gene mutations were a more commonly acquired event in this cohort, and correlated strongly with progression to RAEB and AML. This is again consistent with previous reports of an increased frequency of mutations in *NPM1, FLT3, RUNX1* and *NRAS* genes in

secondary AML compared to MDS (L. Shih *et al.*, 2004; Bacher *et al.*, 2007a; C.-Y. Chen *et al.*, 2007; Dicker *et al.*, 2010; Schnittger *et al.*, 2011). Importantly specific mutations in the pre-diagnostic sample were associated with rapid disease progression and worse overall survival; identifying these patients early provides a window of opportunity for intervention. In the setting of a non-diagnostic marrow, mutational analysis can therefore identify high risk patients who require close clinical follow up as a minimum.

The high frequency of driver mutations in this cohort suggests that mutational analysis in the setting of appropriate peripheral blood count parameters could be confirmatory evidence of MDS. This has been successfully implemented in MPNs for which a *JAK2* mutation in conjunction with proliferative blood parameters is sufficient for diagnosis (McMullin *et al.*, 2007; Harrison *et al.*, 2010). Recent reports, however, of frequent somatic mutations(Genovese *et al.*, 2014; Jaiswal *et al.*, 2014; Xie *et al.*, 2014; McKerrell *et al.*, 2015) as well as large chromosomal abnormalities(Jacobs *et al.*, 2012; Laurie *et al.*, 2012) in the aging healthy population question the validity of this approach, particularly in the absence of definitive morphological disease. Importantly when compared with the pre-diagnostic patients in this cohort the reported median allele fraction of mutations in the healthy population was much smaller (9-10%(Jaiswal *et al.*, 2014) vs. 40%, see supplementary Fig.1), and the number of individuals with multiple mutations was also lower (≥2 mutations, 8%(Jaiswal *et al.*, 2014) vs. 64%). This would suggest that while somatic mutations are frequent in the aging population the clone size must expand to an appropriate level to be disease defining, a model somewhat analogous to monoclonal gammopathy of uncertain significance (MGUS) and myeloma. This is supported by the fact that in the healthy population, those who developed a haematological malignancy had a significantly greater mean allele fraction (25.2%)(Jaiswal *et al.*, 2014).

The implementation of mutational analysis in the diagnosis of MDS will however be more complicated than in diseases such as MPNs. There are more genes implicated and the majority are not specific for this disease subtype. Integration with clinical

factors is, therefore, critical. Defining the somatic nature of certain mutations is also paramount as it is unlikely that a germline sample would be available for comparison in the routine clinical setting. Extensive analysis of cytopenic individuals should help determine both the frequency and likely pathogenic potential of certain mutations/structural variants and the allele fraction required to confirm disease.

In conclusion this initial analysis has confirmed that targeted sequencing is more sensitive than morphological assessment in identifying early MDS, including those with high risk disease, and is more tolerant of sample quality than current methods. With the ever-reducing costs of these techniques, application into the routine setting is becoming increasingly feasible. However, analysis of a large unselected cytopenic population will be required to refine patient selection and clarify those mutations with pathogenic potential. In a group of patients in whom survival can be very short, early detection can provide an opportunity for intervention and improved outcomes.

CHAPTER 5: COHORT 2 – INVESTIGATING THE UTILITY OF TARGETED SEQUENCING AND FLOW CYTOMETRY IN PATIENTS REFERRED WITH A MONOCYTOSIS

These data have been published as a regular article in Blood with an associated commentary (See appendix 8.10)

5.1 Introduction

Distinguishing a reactive monocytosis from CMML is challenging for the haematopathologist. Using current WHO diagnostic criteria, a persistent monocytosis is the hallmark of disease and demonstrating clonality is not a definitive requirement (Swerdlow *et al.*, 2017). This leads to a greater risk of mis-diagnoses or mis-classification, particularly in patients with prolonged reactive changes.

As discussed in Chapter 1, alternative techniques, such as flow cytometry and HTS, have provided a potential objective tool to identify patients with disease. Recent studies have demonstrated a skewing of monocyte subsets in the peripheral blood of patients with CMML which appears to be both sensitive and specific for this disease (Selimoglu-Buet *et al.*, 2015). In addition, interest in molecular genetics has increased since the development of high throughput sequencing techniques. Large studies using extensive targeted sequencing panels have identified recurrent somatic mutations in >90% of patients with CMML (Elena *et al.*, 2016), providing a further potential tool for diagnosis. The presence of a *TET2* mutation along with a mutation in *SRSF2* (or *ZRSR2*) has been shown to be highly specific for a myelomonocytic phenotype (Malcovati *et al.*, 2014) and these, along with *ASXL1*, are the most frequently mutated genes within this disease group (Elena *et al.*, 2016). While the WHO, in the most recent update, have stated the presence of these mutations can support a diagnosis of CMML, there have been no studies directly assessing the use of this technology in diagnosis (Arber *et al.*, 2016). The aim of this study was to determine whether mutational analysis can provide

confirmatory evidence of disease and predict disease outcome in patients presenting with a monocytosis.

5.2 Methods

5.2.1 Patients and Samples

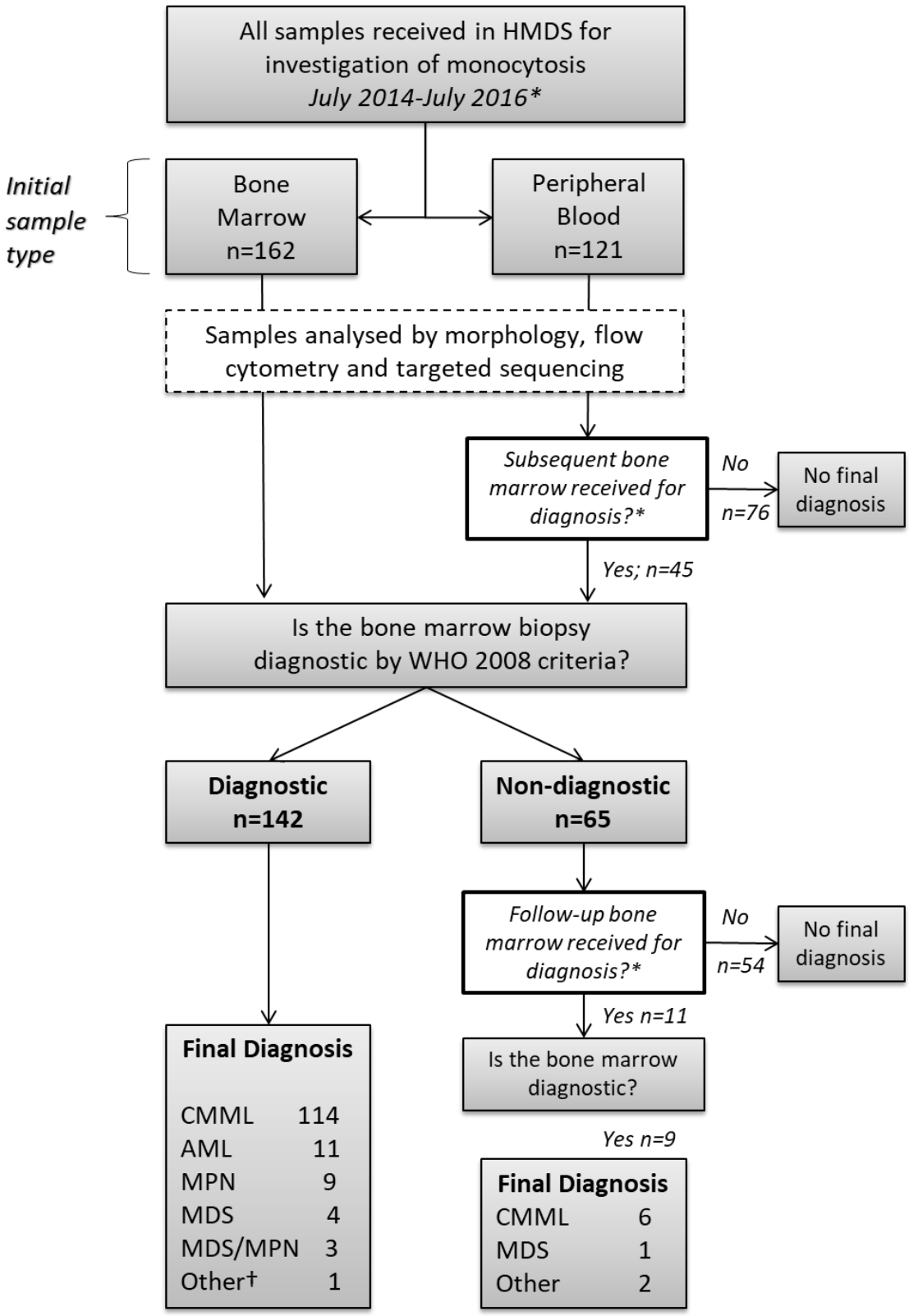
All consecutive samples (peripheral blood (PB) or bone marrow (BM)) received between July 2014 and July 2016 from patients ≥18yrs for the investigation of monocytosis were included; local ethical approval (REC ref-16/NE/0105). Patients with a confirmed myeloid diagnosis prior to July 2014 were excluded. The decision to investigate was at the discretion of the referring clinician, and the study cohort therefore reflects the variety of samples received in a routine laboratory for the investigation of a monocytosis. An absolute monocyte count was determined on all PB samples when received in HMDS (see Table 5.1) using flow cytometry (see Flow cytometry methods). Interestingly this was calculated to be $<1x10^9$/L on a proportion of samples (11%), however the vast majority of these were very close to this threshold and review of local blood count parameters and clinical details confirmed the presence of a PB monocytosis at some point and clinical suspicion of CMML. This highlights the recognised variation in monocyte counts between laboratories, and the difficulty when applying arbitrary cut-offs.

A total of 283 patients were referred during this time period (Table 5.1) of which 121 and 162 had an initial PB and BM sample respectively (Fig. 5.1). A confirmed diagnosis was only made on those patients who ultimately underwent a BM biopsy (n=207). All samples were processed according to gold standard techniques and dual reported by a team of experienced haematopathologists. Those with a confirmed diagnosis were classified in accordance with the WHO 2008 classification. Those failing to fulfil the morphological and genetic WHO 2008 criteria, as agreed by 2 haematopathologists were classified as 'non-diagnostic (ND)'. This term encompassed samples in which a diagnosis could not be reached due either to poor sample quality or insufficient/no abnormal morphological features. All samples were taken with full-informed patient consent for investigation of a suspected haematological disorder.

Table 5.1. Characteristics of patients included in the cohort 2

Patient Characteristics				
No. of patients				283
Male:Female				174:109
Median Age (range)				76 (24-96)
Final Diagnosis				
	PB only			76
	CMML			114
	AML			11
	MPN			9
	MDS			4
	Other			4
	Non-diagnostic			65
		CMML	Other haem malignancy	Non-diagnostic
Blood count parameters: Median (Range)				
Haemoglobin (g/L)		105.5 (38-161)	108 (53-174)	122 (84-163)
White cell count (x10^9/L)		13.6 (4-104.9)	10.6 (3.9-83.4)	7.9 (4.2-38.2)
Platelets (x10^9/L)		90 (1-442)	154 (39-1085)	150 (8-499)
Monocytes (x10^9/L)†		2.69 (0.47-23.59)	1.71 (0.23-9.57)	1.29 (0.72-4.08)
Mutation Frequency: no. of patients (%)				
TET2		72 (63%)	9 (32%)	31 (48%)
SRSF2		48 (42%)	9 (32%)	14 (22%)
ASXL1		39 (34%)	13 (46%)	10 (15%)
NRAS		17 (15%)	7 (25%)	5 (8%)
RUNX1		16 (14%)	6 (21%)	4 (6%)
DNMT3A		9 (8%)	5 (18%)	5 (8%)
CBL		18 (16%)	1 (4%)	4 (6%)
KRAS		9 (8%)	3 (11%)	2 (3%)
SETBP1		7 (6%)	2 (7%)	1 (2%)
JAK2		3 (3%)	7 (25%)	1 (2%)
EZH2		8 (7%)	2 (7%)	1 (2%)
SF3B1		6 (5%)	2 (7%)	0 (0%)

Abbreviations: PB, peripheral blood; CMML, chronic myelomonocytic leukaemia; AML, acute myeloid leukaemia; MPN, myeloproliferative neoplasm; MDS, myelodysplastic syndrome
† Monocyte count determined by flow cytometry



*Decision to investigate was at the discretion of the referring clinician
†Focal area of Diffuse Large B-cell Lymphoma noted in bone marrow, likely co-occurring with CMML

Figure 5.1. Summary of samples included in study. *Flowchart of cases referred for investigation of a monocytosis to HMDS*.

5.2.2 Flow Cytometry

All samples for immunophenotypic analysis were processed within 24 hours. Numerical studies and assessment of monocytic CD56 expression were performed on BM or PB samples following a stain-lyse-wash procedure (FACSLyse, Becton Dickinson) using an antibody cocktail containing CD14, CD64, CD34, CD56, CD45 and CD2. There was strong correlation between monocyte CD56 expression in the PB and BM (Fig 5.2) enabling analysis to include samples from either source.

PB CD14/CD16 "classical" monocytic subset studies were performed on samples following NH4Cl lysis of erythrocytes using a lyse-stain-wash procedure. The antibody cocktail for this tube contained CD16, CD45, CD64, CD14, CD56, IREM2 (CD300e) and HLA-DR.A minimum of 10^5 leucocytes were acquired on a single cytometer (FACSCanto II, Becton Dickinson) for all cases. Monocytes were identified using a combination of CD64, CD45, and scatter characteristics and analysed by a single operator (Dr Matt Cullen) for all analyses (see Appendix 8.11).

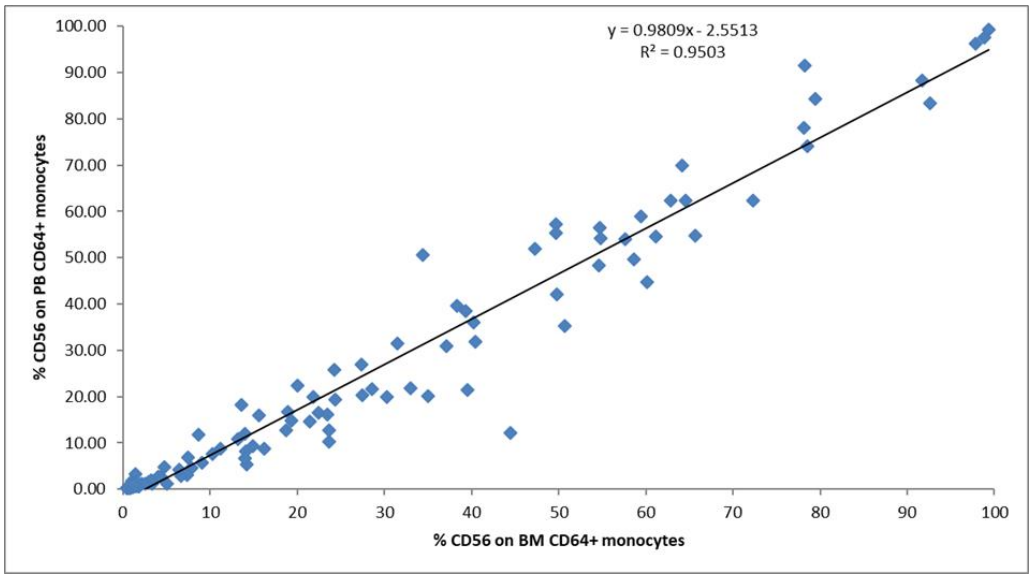


Figure 5.2. Correlation between CD56 expression on monocytes in peripheral blood and matched bone marrow samples.

5.2.3 DNA Extraction and Targeted Amplicon Sequencing

In parallel to the above analyses, samples were subjected to targeted high throughput sequencing using the previously described myeloid panel. Referring clinicians and haematopathologists were blinded to the results of this analysis to exclude reporting or treatment bias.

DNA was extracted from fresh blood or BM mononuclear cells and validation of low-level variants was performed as described in Chapters 2 and 3.

The mean coverage of identified variants was 1514x (range 52-5605x).

5.2.4 Clinical Follow-up

All follow-up BM assessments were performed as clinically indicated by the referring clinician. These samples were also processed according to gold standard techniques and underwent targeted sequencing in parallel, as described above. Any subsequent new diagnoses were recorded.

Survival data was available for all patients and censored on the date of extraction (08/08/2017). Additional clinical information, including serial full blood count data, was collected on a sub-cohort of patients (n=182) either directly from the referring hospital or through the HMRN (n=85).

5.2.5 Statistical Analysis

Survival curves were produced using the Kaplan Meier method and differences were assessed with the log rank test. The impact of abnormalities on overall survival (OS) and risk of progression were estimated using Cox regression; where variable selection was required to arrive at a multivariable regression, the lasso was used for variable selection and results were reported for the corresponding relaxed lasso model.

Sensitivity, specificity, positive and negative predictive values were calculated from 2x2 contingency tables.

Comparison between flow cytometric parameters in the main cohort was performed using Mann-Whitney U tests. Correlation between CD56 expression, M1

monocytes and mutational analysis was performed using both logistic and Poisson regression.

The effect of mutations on longitudinal blood counts was assessed using random effects models. Four different models were fitted using a full-factorial interaction between time and mutation status: (i) a random intercept model; (ii) a random intercept and slope model with uncorrelated random effects; (iii) a random intercept and slope model with correlated random effects; and (iv) a random intercept and slope model with correlated random effects, additionally adjusted for age and sex. For each mutation/blood count relationship the best-fitting model was chosen according to a likelihood ratio test. To limit any potential effect from periods of acute illness or intensive treatment, blood count trajectory analysis was restricted to those with <40 measurements over >100 days.

5.3 Results

5.3.1 Somatic mutations are detected at high frequency in patients with a monocytosis irrespective of diagnosis

To define the mutation spectrum in patients referred with a monocytosis, targeted sequencing results were analysed for the total cohort and correlated with the final diagnosis in those who underwent BM sampling. Of the total 283 patients, ≥1 mutation was detected in 78% of samples, the spectrum of which is presented in Fig. 5.3A. Of these patients, 207 underwent BM assessment for a definitive diagnosis.

A total of 142 patients had a confirmed diagnosis of a myeloid malignancy with ≥1 mutation detected in all but 2 cases (99%). Importantly 1 of the mutation negative cases had a complex karyotype detected by conventional cytogenetics including inversion 3 (with MECOM gene rearrangement), leaving only 1 case without a demonstrable clonal abnormality. The vast majority of cases were confirmed as CMML (n=114), however small numbers of patients had an alternative diagnosis including acute myeloid leukaemia (AML) (n=11) and myeloproliferative neoplasm (MPN) (n=9), highlighting the importance of a bone marrow assessment for disease classification in this patient group.

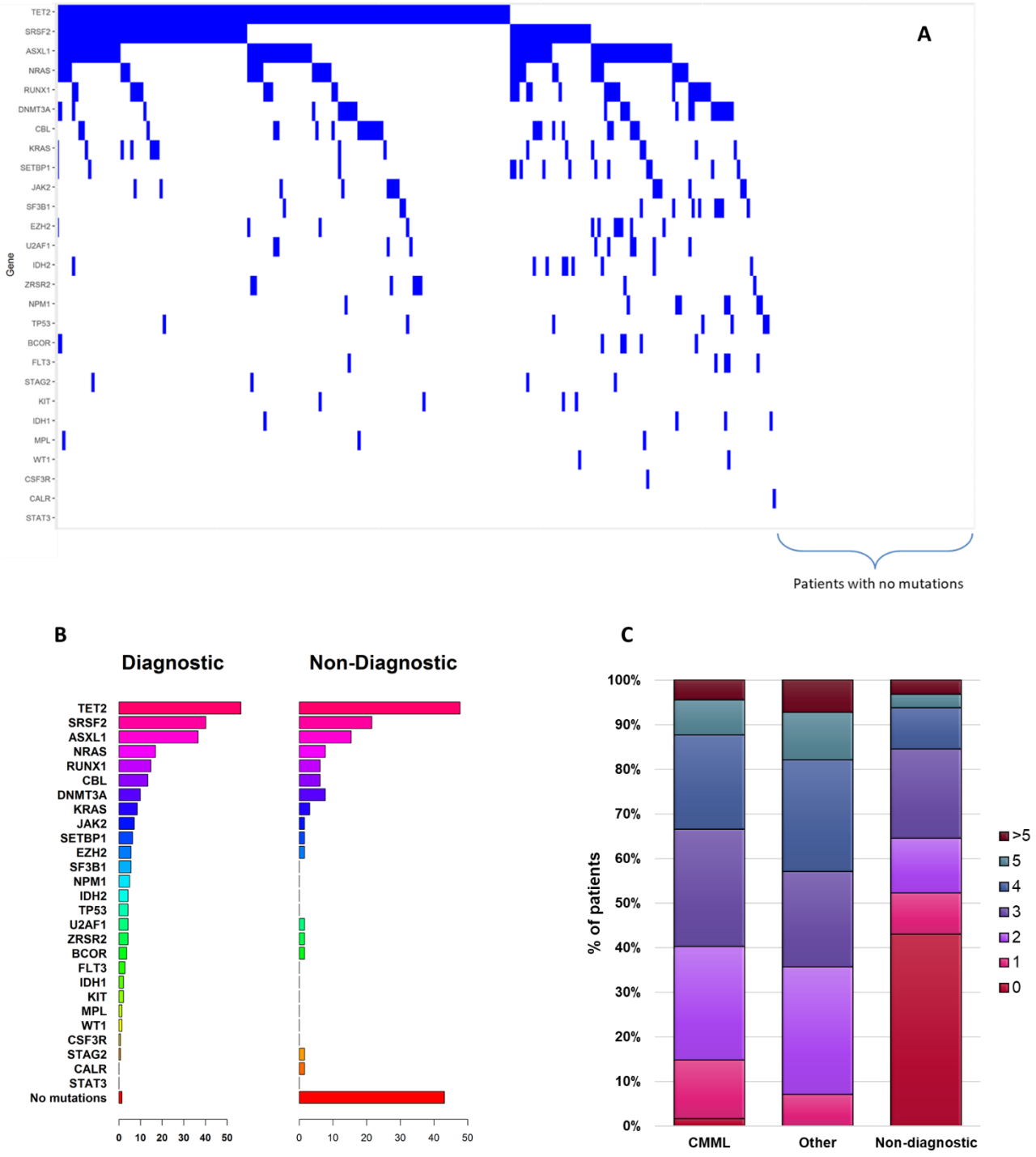


Figure 5.3. Characteristics of mutations detected in patient samples. *(A) Spectrum of mutations detected across all patients in study (n=283). (B) Comparison of mutations detected in those with a diagnostic bone marrow sample vs a non-diagnostic bone marrow sample (diagnostic (n=142) vs non-diagnostic (n=65)). (C) Distribution of no. of mutations according to final diagnostic category. 'Other' denotes those patients with an alternative haematological malignancy.*

Of note, somatic mutations were also detected at a high frequency in those patients without a confirmed diagnosis. At least 1 mutation was detected in 37/65 patients (57%) with indeterminate or reactive features. The spectrum of mutations in this group mirrored that detected in the diagnostic group (see Figure 5.3) with *TET2, SRSF2* and *ASXL1* being the most frequently mutated. The most notable differences in the non-diagnostic group were the absence of mutations associated with high risk disease, including *TP53* and *NPM1*, as well as those associated with specific morphological abnormalities i.e. *SF3B1* (associated with ring sideroblasts) and *JAK2* (associated with myeloproliferative features).

The median and mean number of mutations was higher in those with a confirmed diagnosis (median-3, range 0-8; mean-3) versus those without (median-1, range 0-6; mean-2) (Fig.5.3C). However, in patients with a confirmed mutation, the number of mutations did not differ significantly between diagnostic and non-diagnostic groups (p=0.62).

The median variant allele fraction (VAF) for all variants was 39% (range 5.2-100%) and there was no difference between the VAF in diagnostic and non-diagnostic cases (p=0.33). In those with an isolated mutation the median VAF was also noted to be high at 38.2% (range 6.3-97.1%) with only 2 variants having a VAF of <10%.

Mutations are therefore found at a very high frequency with a high clonal burden in patients with a monocytosis, and involve a similar spectrum of genes, irrespective of diagnosis.

5.3.2 Immunophenotypic features correlate strongly with the presence of a mutation and a subsequent diagnosis

To determine whether immunophenotyping can predict for the presence of a mutation or a BM diagnosis, flow cytometric analysis was performed alongside sequencing.

Firstly, comparison was made between the immunophenotypic features in the BM of patients with a confirmed diagnosis of CMML versus those with non-diagnostic samples. Key phenotypic abnormalities were demonstrated in samples with a confirmed diagnosis. These included increased CD64+ monocytes, reduced CD14 expression as well as aberrant CD56 expression on monocytes. These abnormalities were also noted in cases without a confirmed diagnosis, but importantly correlated strongly with the presence of a mutation. There was a significant difference in the proportion of CD64+ monocytes, CD14 expression and CD56 expression between those patients with a demonstrable mutation and those without (Fig. 5.4). This was most pronounced with CD56 expression which was significantly associated with the presence of a mutation (OR 12.2; 95% CI 4.8-41.3; p<0.0001) and the number of mutations (OR 1.4; 95% CI 1.2-1.6; p<0.0001). With respect to individual mutations, aberrant expression of CD56 was strongly associated with *TET2* mutations (OR 4.0; 95%CI 2.4-6.8; p<0.0001). The relationship between CD56 expression and a somatic mutation has not previously been described however bears similarities to the phenotypic aberrancies described by Seglimolu et al which were predictive of a CMML diagnosis.

5.3.2.1 Peripheral blood monocyte subsets and CD56 expression are predictive of a somatic mutation

To further analyse the relationship between peripheral blood monocyte subsets, CD56 expression and mutation profile a separate cohort was investigated for all 3 parameters (see methods). A total of 135 patients were analysed with 95 undergoing a bone marrow for definitive diagnosis (CMML=28, MDS=23, MPN=9, non-diagnostic=27, other=8). The presence of aberrant CD56 was again strongly associated with the presence of a mutation (OR 43.9; 95%CI 8.9-793.9; p=0.0003). This was also noted, to a lesser extent, with having greater than 94% M1 monocytes (OR 3.9; 95% CI 1.8-8.7; p=0.0007). There was some correlation between the presence of CD56 expression and >94% M1 monocytes (r=0.17, p=0.039) and combining both produced a stronger effect than >94% M1 monocytes alone (OR 8.5; 95%CI 3.9-19.5; p<0.00001). Importantly, combining these phenotypic

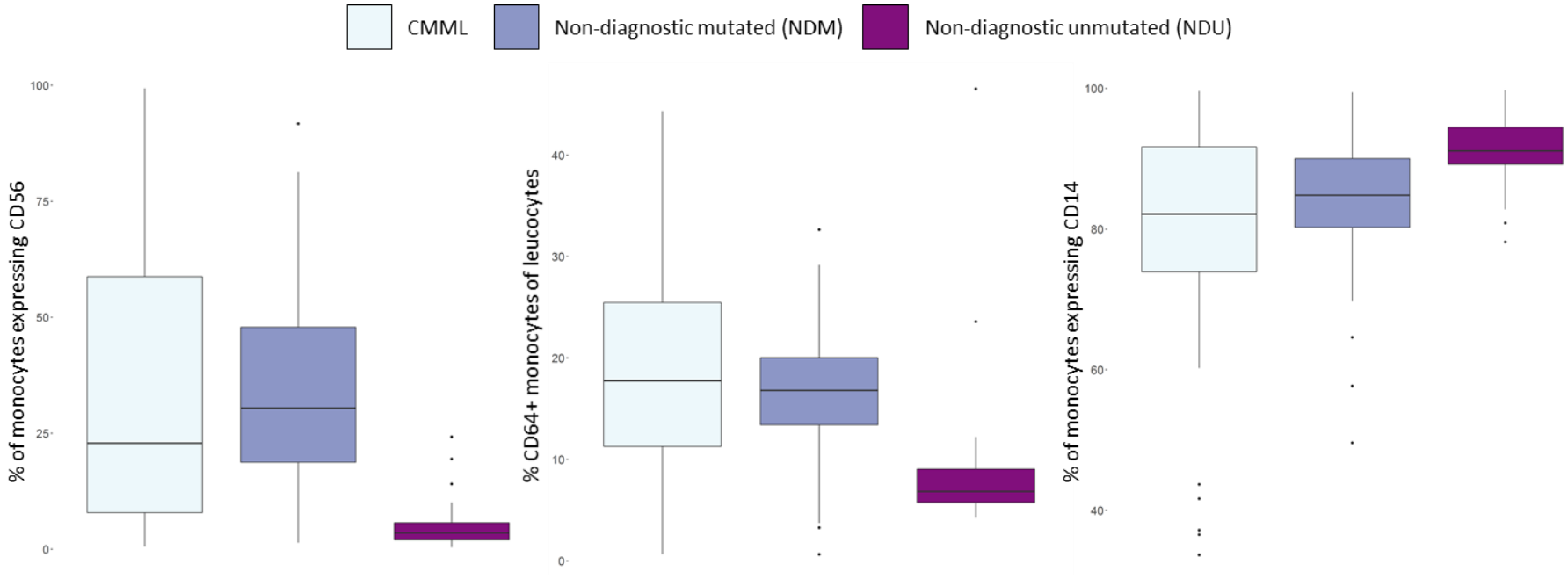
aberrancies did not capture all patients with a mutation. Whilst CD56 was highly specific for the presence of a mutation (98%), sensitivity was only 48%. Similarly, the presence of >94% M1 monocytes had a sensitivity of 56% and specificity of 75% for detecting a mutation.

The presence of either CD56 expression or >94% M1 monocytes was strongly associated with the presence of *ASXL1, TET2, and SRSF2* mutations. Interestingly all patients with *SRSF2* mutations had either CD56 expression or >94% monocytes or both. There was a weaker association with *NRAS* mutations for both aberrant features and *EZH2* mutations were only statistically associated with CD56 expression.

With respect to a confirmed diagnosis, both CD56 expression (OR 4.9; 95%CI 1.9-13; p=0.001) and >94% M1 monocytes (OR 4.2; 95%CI 1.7-11.5; p=0.003) were associated with a final diagnosis of CMML with a stronger effect again noted when combining both (OR 5.2; 95%CI 1.8-19; p=0.0056). It is interesting to note that four patients with CMML (with confirmed mutations) had neither of these phenotypic aberrancies.

5.3.3 Peripheral Blood mutation profiling is predictive of a bone marrow diagnosis

As well as providing key immunophenotypic information, mutation profiling on peripheral blood was also highly predictive of a bone marrow diagnosis. A total of 121 PB samples were received as the initial sample and somatic mutations were detected in 66% (80/121). Forty-five patients (37%) had a subsequent bone marrow performed for diagnosis, further samples have yet to be received on the remaining patients. There was high concordance between mutations detected in PB and subsequent BM with only 5 discordant results (119/124 variants concordant; 96%). Importantly these were low level variants at the limit of detection for the test or variants detected at areas of poor coverage (*SRSF2/ASXL1*). All 9 mutation negative cases were fully concordant.



	% CD56 expression on monocytes
CMML vs NDM	p=0.45; NS
CMML vs NDU	p<0.0001
NDM vs NDU	p<0.0001

	% CD64+ monocytes of leucocytes
CMML vs NDM	p=0.36; NS
CMML vs NDU	p<0.0001
NDM vs NDU	p<0.0001

	% CD14 expressing monocytes
CMML vs NDM	p=0.22; NS
CMML vs NDU	p=0.0004
NDM vs NDU	p=0.006

Figure 5.4. Relationship between immunophenotype and mutations. *Box and whisker plots comparing immunophenotypic features of CMML, non-diagnostic mutated (ND^{mut}) and non-diagnostic unmutated (ND^{unmut}) cases. The p-values refer to Mann-Whitney U tests comparing CMML with either non-diagnostic category.*
(A) % CD56 expression on monocytes. (B) % CD64+ monocytes of leucocytes. (C) % CD14 expressing monocytes.

The presence of a mutation in the peripheral blood was highly predictive of diagnosing a myeloid malignancy in bone marrow (PPV 0.96, NPV 1.0). Of note, none of the mutation negative cases (n=11) had a subsequent confirmed diagnosis.

5.3.4 Overall survival correlates strongly with mutation profile irrespective of diagnosis

Survival data were available on all patients. The median survival from the time of the first sample was 35.2 months and correlated strongly with the number of mutations; those with no mutations had a significantly better overall survival while those with >5 mutations had by far the worst (see Fig. 5.5). The difference between the presence or absence of a mutation was highly significant ($p<0.001$). On univariate analysis, age was strongly associated with survival. For mutations occurring in >5% of subjects, *ASXL1, CBL, DNMT3A, NRAS & RUNX1* were all strongly associated with survival, as were *EZH2 & STAG2* amongst the less frequently mutated genes (Table 5.2). In order to investigate multivariate significance, all genes mutated in >5% subjects were entered into a lasso survival regression. Taking the 1SE shrinkage parameter, age, *ASXL1, CBL, DNMT3A, NRAS & RUNX1* were selected by the lasso and retained significance in a relaxed lasso regression.

5.3.5 Patients without a confirmed diagnosis have similar outcome to CMML patients and have significantly worse blood counts

In patients who proceeded to a bone marrow biopsy, survival correlated with the final morphological diagnosis. Those without a confirmed diagnosis had a significantly better overall survival than those with CMML or another myeloid malignancy. However, on further investigation only patients without a demonstrable mutation retained a significantly improved survival ($p=0.0002$). Survival in non-diagnostic patients with a demonstrable mutation was similar to those with CMML ($p=0.118$) (Fig. 5.6).

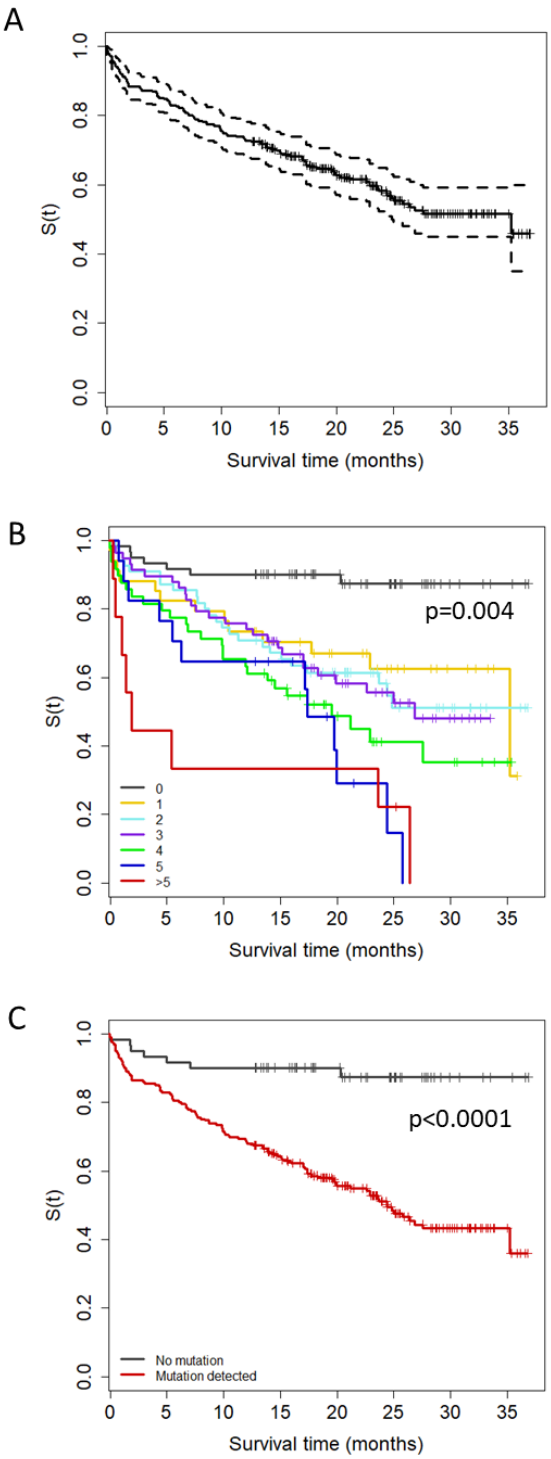


Figure 5.5. Overall survival according to mutation number. *(A) Overall survival in total cohort from time of initial sample. (B) Overall survival in total cohort by no. of mutations detected at time of initial sample. The p-value represents a log rank test comparing those without a mutation to those with a single mutation. (C) Overall survival in total cohort by the presence or absence of a mutation*

***Table 5.2. Univariate and multivariate overall survival analyses for total cohort**. For multivariable regression, the lasso was used for variable selection and results reported for the corresponding relaxed lasso model.*

Variable	Univariate Analysis	Multivariate Analysis
	HR, 95% CI	***HR, 95% CI***
Age	**1.04(1.02-1.06)**	1.04(1.02-1.06)
ASXL1	**2.10(1.45-3.05)**	1.59(1.07-2.38)
CBL	**2.12(1.30-3.47)**	2.24(1.35-3.73)
DNMT3A	**2.87(1.73-4.76)**	2.82(1.68-4.73)
KRAS	1.69(0.85-3.33)	
NRAS	**2.11(1.32-3.36)**	1.85(1.13-3.01)
RUNX1	**2.75(1.74-4.37)**	2.20(1.37-3.54)
SRSF2	1.18(0.80-1.72)	
TET2	0.82(0.57-1.19)	
EZH2	**4.88(2.52-9.44)**	
STAG2	**22.39(7.48-67.01)**	

HR – Hazard ratio; CI – Confidence Intervals

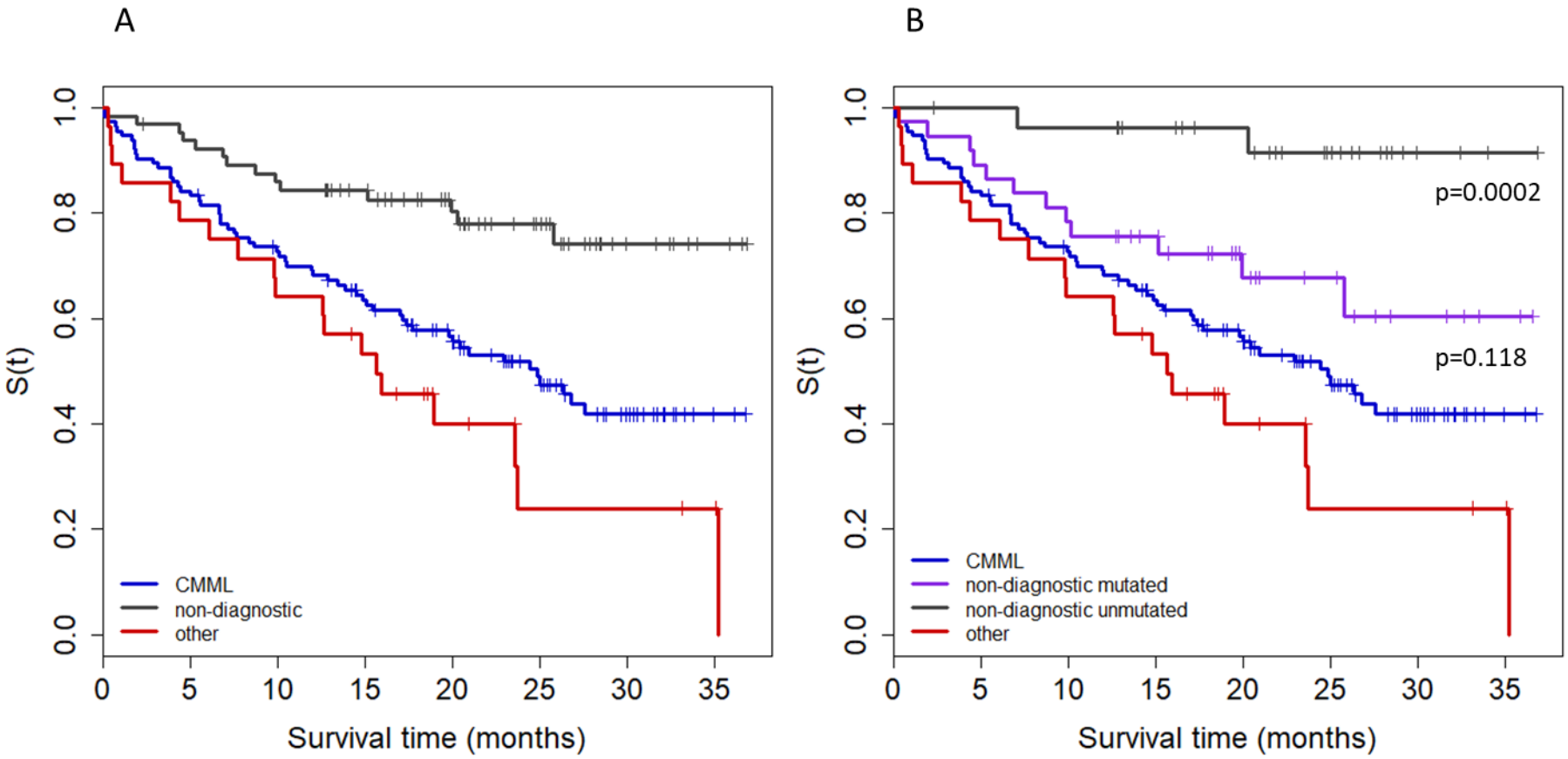


Figure 5.6. Overall survival according to final diagnosis. *(A) Overall survival by diagnosis on bone marrow sample (n=207)*
(B) Overall survival by diagnosis with non-diagnostic samples separated by the presence or absence of a mutation. The p-values refer to log-rank tests comparing CMML and non-diagnostic unmutated patients (p=0.0002) and comparing CMML with non-diagnostic mutated patients (p=0.118)

To further assess the impact of mutations on outcome, longitudinal blood count data were analysed in conjunction with mutation profile. Blood count data were available for 182 patients. Subsequent analysis was restricted to those with <40 measurements over >100 days (n=133), to exclude periods of acute hospital admissions (due to periods of acute illness/infection) or intensive chemotherapy (median follow-up 465 days; range 119-996 days). Patients with a demonstrable mutation had a significantly lower haemoglobin and platelet count as well as a higher monocyte count than those without a mutation (Fig. 5.7). This effect persisted over the time period measured and followed a divergent trajectory. With respect to individual mutations, certain mutations were associated with increasing or declining blood count parameters over time (Table 5.3). Monocyte counts were found to increase over time in *TET2, SRSF2, ASXL1, NRAS or RUNX1* mutated subjects relative to non-mutated; similarly, white blood counts increased in *ASXL1, NRAS & DNMT3A* mutated subjects; and platelet levels decreased in *ASXL1, CBL* and *RUNX1* mutated subjects relative to non-mutated.

In those without a confirmed diagnosis, follow-up BM biopsies were received on 11 patients. Importantly, of those with a subsequent diagnosis of CMML, all had a confirmed mutation on the original sample. In total 7/37 (19%) non-diagnostic mutated patients had a confirmed diagnosis (6 CMML, 1 MDS). Furthermore, none of the mutation negative cases went on to develop CMML, however 2 patients had confirmed alternative haematological diagnoses - DLBCL and Rosai Dorfman disease.

These findings confirm that the presence of a mutation has a significant impact on outcome with respect to both survival and blood count parameters.

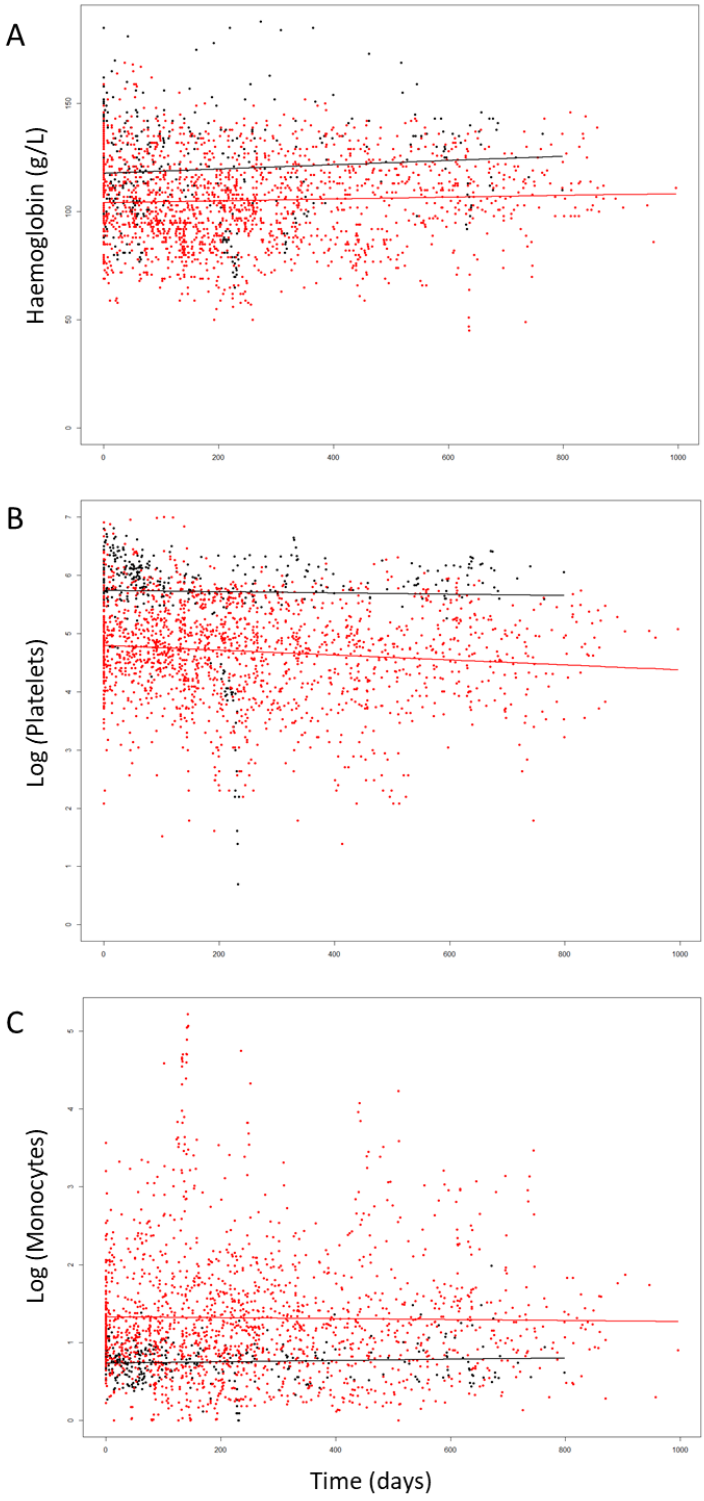


Figure 5.7. Longitudinal blood count trajectories in relation to mutation status. *Plots of all blood count trajectories averaged between mutated (red) and unmutated (black) groups with overlaid linear regression line.*
(A) Haemoglobin (g/L) in patients with or without a detectable mutation. (B) Platelet count (log transformed) in patients with or without a detectable mutation. (C) Monocyte count (log transformed) in patients with or without a detectable mutation.

Table 5.3. Effect of mutations on longitudinal blood counts, assessed using random effects models

	Haemoglobin	WCC	Platelets	Monocytes
Any mutation	– – Y	No effect	– – Y	+ + Y
TET2	No effect	No effect	– + Y	+ + Y
SRSF2	No effect	No effect	– + Y	+ + Y
ASXL1	– + Y	+ + Y	– – Y	+ + Y
NRAS	0 – Y	+ + Y	No effect	+ + Y
CBL	– – Y	No effect	– – Y	+ – Y
DNMT3A	No effect	0 + Y	No effect	No effect
JAK2	No effect	No effect	+ – Y	No effect
RUNX1	No effect	No effect	– – Y	0 + Y

+ + Y statistically significant positive difference with diverging trajectories that persist over the time of observation.

– – Y statistically significant negative difference with diverging trajectories that persist over the time of observation

– + Y statistically significant negative difference with converging trajectories that persist over the time of observation.

+ – Y statistically significant positive difference with converging trajectories that persist over the time of observation.

0 – Y no overall difference but a statistically significant decrease in the mutated relative to the non-mutated

0 + Y no overall difference but a statistically significant increase in the mutated relative to the non-mutated

5.4 Discussion

Using the myeloid panel on this cohort of patients, mutations are identified in virtually all patients with a morphological diagnosis of CMML, and in a significant proportion of patients with a monocytosis and non-diagnostic features. Importantly the latter group had a mutation spectrum, immunophenotype and outcome indistinguishable from CMML. Mutational analysis can therefore provide key diagnostic and prognostic information in the investigation of patients with a persistent monocytosis.

By analysing sequential samples referred to HMDS the cohort has included the typical patient population encountered in routine haematology practice. It is possible that the proportion of non-diagnostic samples with detectable mutations was inflated due to referral bias and a high pre-test probability of disease in those undergoing testing, however the use of objective outcome measures (longitudinal blood counts and OS) and an unselected patient population have otherwise minimised bias and ensured the results are applicable in the 'real-world' setting.

As described in Chapter 1, many studies have investigated the genetic profile in established CMML and its clinical significance, confirming that the profile is relatively homogenous involving a restricted number of genes. Looking at only 7 genes, initial studies identified mutations in >70% of patients, with *TET2* mutations being most frequent (Kohlmann *et al.*, 2010). Subsequent studies, with varying panel sizes (19-276 genes), have consistently reported mutations in >90% of patients (Itzykson *et al.*, 2013; Elena *et al.*, 2016; Mason *et al.*, 2016). While these mutations are identified in genes implicated across the spectrum of haematological malignancies, *SRSF2, TET2* and *ASXL1* are by far the most commonly mutated, with a combination of *TET2* and either *SRSF2* or *ZRSR2* being shown to be highly specific for a myelomonocytic phenotype (Malcovati *et al.*, 2014). The mutation profile in this cohort reflects these findings even in the absence of a confirmed diagnosis.

ASXL1 has also been shown to be consistently associated with a poorer prognosis(Itzykson *et al.*, 2013; Patnaik *et al.*, 2013, 2014, 2015, 2016; Elena *et al.*,

2016), a feature also replicated across this dataset. Despite a restricted panel, the mutation frequency was high and the presence of even an isolated mutation impacted significantly on outcome. Therefore, in the investigation of a monocytosis, even modest sized panels can provide key clinical information and be cost effective in a clinical laboratory setting. It is however likely that a proportion of these patients will have additional mutations in genes not sequenced in this study, though, to investigate this further would require a much larger patient population.

The findings of this study will be key to refining future diagnostic algorithms in the investigation of patients referred with a monocytosis. Mutational analysis has been incorporated into the recent amendment of the WHO diagnostic criteria, which now state that the presence of a mutation can support a diagnosis of CMML. Similar to the diagnosis of MDS, concerns have also been raised regarding the use of mutational analysis in this setting, due to reports of frequent somatic mutations in aging healthy individuals (Genovese *et al.*, 2014; Jaiswal *et al.*, 2014; Xie *et al.*, 2014; McKerrell *et al.*, 2015). As a result the WHO have stated that the presence of a mutation in CMML should not be used alone as proof of disease (Swerdlow *et al.*, 2017). Results from this cohort have however shown that even in the absence of morphological features, those patients with a mutation had a clinical phenotype and genotype indistinguishable from CMML, and comparably poor outcomes. Distinguishing features were also noted between the variants reported in healthy individuals and the mutations detected this cohort. The VAF or clone size of the mutations in our study were significantly higher than in healthy individuals (median 39.2% vs 9-10%), and this was demonstrated across both diagnostic and non-diagnostic samples. This finding was also described in those patients investigated in cohort 1 (see Chapter 4), in which a VAF >20% and the presence of co-occurring mutations could distinguish patients with clinically significant cytopenias from healthy individuals. While the higher VAF is replicated in this cohort, importantly our study has shown that even isolated mutations have a significant impact on survival in patients with a monocytosis. These findings provide strong evidence that in those without diagnostic morphological features the presence of a mutation,

irrespective of mutation number, could be disease defining. At the very minimum, it is imperative that these patients are identified and monitored closely.

Flow cytometry has also been proposed as a screening tool on PB samples to identify patients with CMML. Most recently, >94% M1 monocytes in the PB has been reported to be highly sensitive and specific for CMML (Selimoglu-Buet *et al.*, 2015). Subsequent studies have validated this method and also confirmed the ability to distinguish CMML from MDS as well as MPN presenting with a PB monocytosis (Patnaik *et al.*, 2017; Talati *et al.*, 2017). These studies however are centred on morphological diagnosis and they have not consistently performed mutation analysis on the patient cohorts. While investigation of cohort 2 has shown a strong correlation between skewed monocyte subsets and a diagnosis of CMML this did not capture all patients and was neither sensitive nor specific for the presence of a mutation. In contrast aberrant CD56 expression was highly specific for the presence of a mutation (98%), particularly involving *TET2*. Similar sensitivity and specificity for CMML diagnosis has been reported (100% and 67% respectively) previously when combining immunophenotypic features including CD56 expression, reduced expression of myeloid antigens and ≥20% immature monocytes(Xu *et al.*, 2005). Though concern was raised regarding the overexpression of CD56 in reactive conditions(Lacronique-Gazaille *et al.*, 2007). Analysis on this cohort has shown that CD56 expression at diagnosis is invariably associated with the presence of a somatic mutation, though again will not capture all patients (sensitivity 48%). Flow cytometry could, therefore, provide a screening tool for the investigation of PB monocytes but, ultimately, mutation analysis will be required to identify patients who require clinical follow-up.

Importantly there was high concordance between PB and BM mutational analyses, and the presence of a PB mutation was highly predictive of a subsequent BM diagnosis. This suggests that PB screening may be suitable for identifying or excluding significant mutations, however this could lead to a rise in inappropriate referrals and a significant burden on laboratory workload. Furthermore, the small proportion of mutated patients in this cohort with other haematological

malignancies in the BM, including AML, highlights the importance of a baseline BM assessment to definitively classify the disease. In contrast, the negative predictive value of PB screening was 100%, suggesting that those without a mutation should not undergo BM assessment. In the first instance, PB screening would be a practical option in those patients unfit for BM assessment or potentially to monitor for treatment response or disease evolution. The latter would require further investigation in a prospective study.

In conclusion, these analyses have confirmed that mutations are commonly detected in patients referred with a persistent monocytosis. The presence of a mutation impacts significantly on outcome irrespective of diagnosis, and patients with a mutation who fail to meet WHO criteria have CMML disease characteristics. These findings validate the inclusion of somatic mutations in the diagnostic criteria for CMML and, at the very minimum, suggest that those without a confirmed diagnosis require close clinical follow-up. While PB can be confidently used to detect mutations, a baseline BM biopsy is required for definitive disease classification in patients fit for treatment. Immunophenotypic assessment of monocytes may provide a potential screening tool to detect those with a mutation however it will miss a proportion of mutated patients. Ultimately, early identification of patients could provide an opportunity for intervention in this patient group, and this requires further investigation.

CHAPTER 6: COHORT 3: INVESTIGATING THE DIAGNOSTIC POTENTIAL OF HIGH THROUGHPUT SEQUENCING IN AN UNSELECTED CYTOPENIC PATIENT GROUP

6.1 Introduction

The high frequency of driver mutations in MDS cases has generated considerable interest in the utility of mutational analysis as a diagnostic tool. As previously discussed in Chapters 1 and 4, the current diagnostic criteria for MDS rely heavily on the morphological assessment of the bone marrow (Swerdlow *et al.*, 2017) and this is hampered by poor interobserver concordance (Parmentier *et al.*, 2012). To date cytogenetics has provided the only objective marker of disease though an abnormal karyotype is only found in around half of patients with a confirmed diagnosis (Schanz *et al.*, 2011). Nonetheless, concern has been raised regarding the diagnostic utility of mutational analysis due to reports of somatic mutations in healthy individuals (Genovese *et al.*, 2014; Jaiswal *et al.*, 2014; Xie *et al.*, 2014; McKerrell *et al.*, 2015). Hence, distinction between those with age related clonal haematopoiesis and those with a disease defining mutation in the absence of definitive morphological disease remains challenging.

The study performed on cohort 1 and discussed in Chapter 4 provided proof of principle that mutations were present in a very high percentage of patients who initially failed to meet diagnostic criteria but subsequently progressed to MDS or AML. Importantly these patients had a higher frequency of co-occurring mutations and a greater allele fraction or clone size than that reported in healthy individuals (39.9% vs 9-10%) (Jaiswal *et al.*, 2014; Cargo *et al.*, 2015) suggesting that mutation number and allele fraction could, perhaps, form part of the diagnostic requirements. Subsequent studies have confirmed a high frequency of mutations in cytopenic patients, even in the absence of a confirmed haematological malignancy with mutations reported in ~30% of ICUS patients (~30%) (Kwok *et al.*, 2015). Such an approach was further investigated by looking at the predictive value

of mutation analysis when performed on peripheral blood taken simultaneously with the bone marrow at the time of diagnosis (Malcovati *et al.*, 2017). This showed that certain mutations, particularly those involving spliceosome genes, were highly predictive of MDS, even in isolation. In contrast, mutations in epigenetic regulators had a much lower positive predictive value if identified in isolation, though this increased significantly when additional mutations were demonstrable. Importantly the latter paper also highlighted the negative predictive value of mutation analysis; reported to be 0.84 increasing to 0.92 when a negative cytogenetic analysis was also included (Malcovati *et al.*, 2017).

Interestingly all MDS studies to date, even those with large sequencing panels, have identified a proportion of patients (10-20%) who lacked a demonstrable somatic mutation (Haferlach *et al.*, 2013; Papaemmanuil *et al.*, 2013). In the absence of increased blasts, these patients have subsequently been shown to have a much better overall survival than other MDS patients, with none transforming to AML over the long follow-up period (Malcovati *et al.*, 2014, 2017). Furthermore, X-chromosome inactivation pattern analysis in a subset of these patients has confirmed polyclonal haematopoiesis (Malcovati *et al.*, 2017). A similar group was also identified by gene expression analysis during the MILE (Microarray Innovations in Leukaemia) study. In this study, a proportion of MDS cases had an expression profile which was classified as a non-leukemic condition, and these patients again showed no evidence of AML transformation over the 5 year follow-up (Mills *et al.*, 2009). The evidence suggests that these patients may not actually have MDS, despite the WHO criteria being fulfilled, though clearly further investigation of this group is required. Hence, while the diagnosis is unclear in these patients, the absence of a mutation or cytogenetic abnormality should certainly be viewed as a good prognostic feature.

It is therefore clear that mutations can be detected at a high frequency but the clinical significance of these in the absence of morphological disease remains uncertain. To further investigate the significance of somatic mutations in cytopenic patients, analysis was performed on this final cohort of unselected cytopenic

patients. The aim of the analysis was to both determine the frequency of mutations in cytopenic individuals, and investigate the clinical significance of detecting a somatic mutation. As with cohort 2, this was achieved by assessing objective outcome measures including longitudinal blood counts and overall survival as well as correlating with subsequent diagnoses.

6.2 Methods

6.2.1 Patients and Samples

The study cohort included all sequential bone marrow samples referred for investigation of cytopenia between July 2014 and July 2016. Only adult patients were included (≥18yrs) and patients with a confirmed diagnosis of a myeloid malignancy before July 2014 were excluded. Similar to cohort 2, the decision to investigate was at the discretion of the referring clinician, and the study cohort therefore reflects the variety of samples received in a routine laboratory for the investigation of a cytopenia.

A total of 2130 samples were received during this time period (see Fig.6.1). All samples were processed according to gold standard techniques and HMDS laboratory processes, with morphology being dual reported by a team of experienced haematopathologists. Those with a confirmed diagnosis were classified in accordance with the WHO 2008 classification. Due to recognized challenges when confirming a diagnosis of MDS, particularly in cases with unilineage dysplasia, these cases were captured as unilineage dysplasia but not given a formal diagnosis of MDS on the initial sample. This is in accordance with recommendations from the European LeukaemiaNet (Malcovati *et al.*, 2013). Those otherwise failing to fulfil the morphological and genetic WHO 2008 criteria, as agreed by 2 haematopathologists were classified as 'non-diagnostic' (ND) as described in Chapter 5 (section 5.2.1)

All samples were taken with full-informed patient consent for investigation of a suspected haematological disorder. This study had local ethical committee approval (REC ref-16/NE/0105).

6.2.2 DNA Extraction and Targeted Amplicon Sequencing

In parallel to the standard laboratory processes, samples were subjected to targeted high throughput sequencing using the previously described methods (see Chapters 2 and 3). Referring clinicians and haematopathologists were blinded to the results of this analysis to exclude reporting or treatment bias.

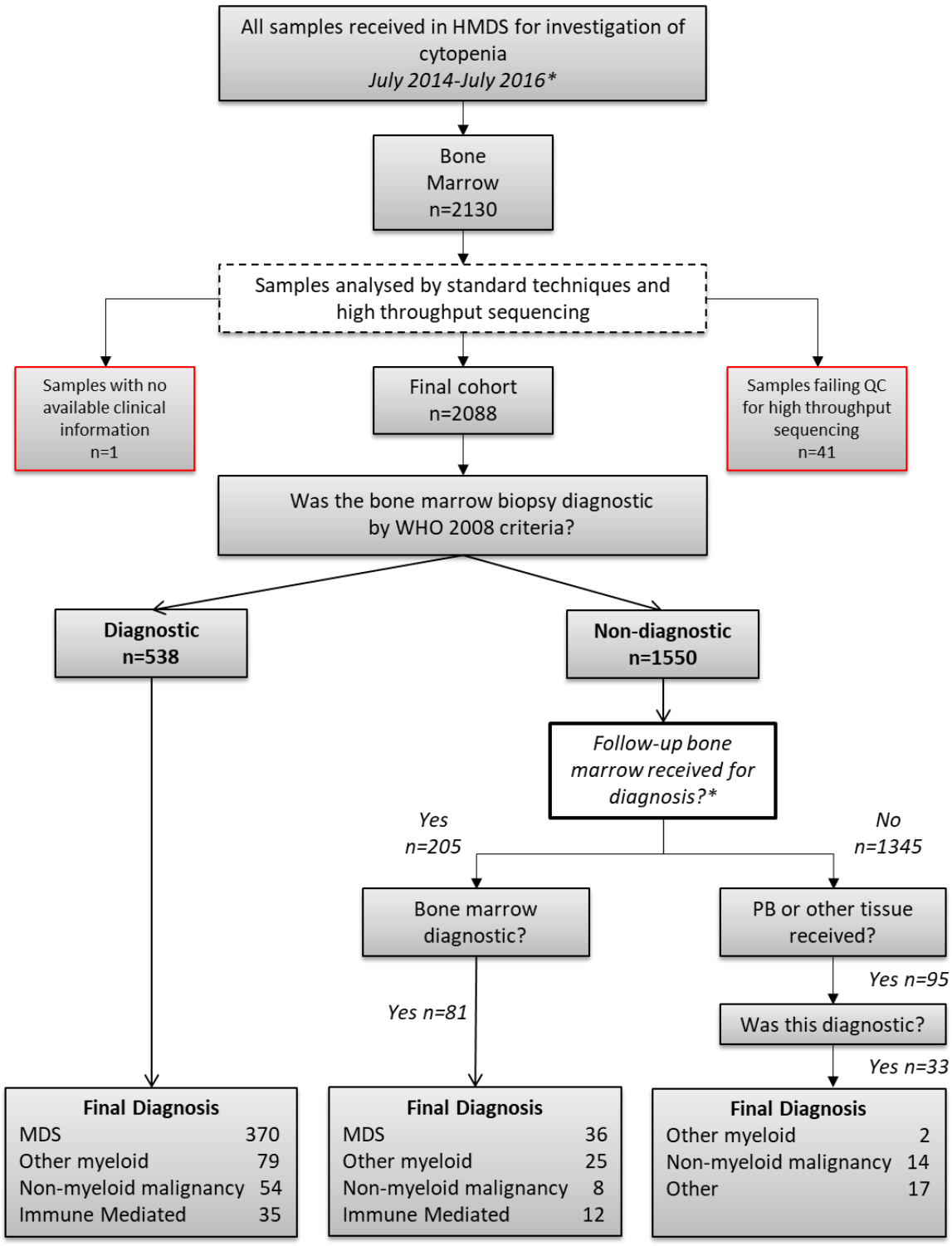
DNA was extracted from fresh blood or BM mononuclear cells using the methods described in Chapter 2. A total of 41 samples (1.9%) failed to meet DNA QC thresholds either to proceed to sequencing or for interpretation of results (see Chapter 3); most commonly due to low DNA quantity. These patients were excluded from further analysis, leaving a cohort of 2089 patients. Clinical information was not available for 1 patient, leaving a final cohort of 2088 patients. Characteristics of this patient group can be found in Table 6.1.

The mean coverage of identified variants was 1594x (range 51-5666x).

6.2.3 Clinical Follow-up

Follow-up BM assessments were performed as clinically indicated by the referring clinician and these samples were also processed according to gold standard techniques with subsequent new diagnoses recorded. Targeted sequencing was performed in parallel.

Survival data were available for all patients and censored on the date of retrieval (04/06/2018). Serial full blood count data was collected on 1272 patients directly from the referring hospital. This included all full blood counts performed from the beginning of 2014 until the time of data collection.



* Sample taken at the discretion of the referring clinician

Figure 6.1: Flowchart of cases referred for investigation of cytopenia

6.2.4 Statistical Analysis

As previously described in Chapter 5, the impact of abnormalities on OS and risk of progression were estimated using Cox regression, and the lasso method was used for variable selection when performing multivariate analysis with results reported for the corresponding relaxed lasso model. For the progression analysis, death was analysed as a competing risk using a simple 3-state markov multistate model (R library "msm").

The effect of mutations on longitudinal blood counts was again assessed using random effects models as described in Chapter 5 (section 5.2.5). In contrast to the previous analysis, no restrictions were placed on the frequency of blood count measurements.

6.3 Results

6.3.1 Primary bone marrow diagnosis is made in only a minority of cases

A total of 2088 patients were included in the final analysis and of these only 538 had a confirmed diagnosis (26%; see Table 6.1 and Fig. 6.1). The vast majority were diagnosed with a myeloid malignancy (449/538; 83%) with MDS being by far the most common diagnosis (370/449; 82%). Of the remaining cases, an alternative malignant diagnosis was made on the bone marrow sample in 55 patients, which included mostly lymphoproliferative disorders or metastatic carcinoma. Immune mediated causes for cytopenia were identified in 36 patients, including both aplastic anaemia and red cell aplasia.

Table 6.1: Characteristics of patients included in cohort 3

Patient Characteristics					
No. of patients	2088				
Male:Female	1232:856				
Median Age (range)	72 (18-99)				
Final Diagnosis					
MDS	370				
Other Myeloid	79				
AML	*31*				
MDS/MPN	*31*				
MPN	*10*				
Other†	*7*				
Non-myeloid malignancy	54				
Lymphoma/LPD	*36*				
Non-haemopoietic	*18*				
Immune mediated	35				
Non-diagnostic	1550				
	MDS n=370	**Other Myeloid Malignancy n=79**	**Non-Myeloid Malignancy n=54**	**Immune Mediated n=35**	**Non-diagnostic n=1550**
Blood count parameters: Median (Range)					
Haemoglobin (g/L)	96 (49-152)	93 (70-128)	102 (48-161)	96.5(55-108)	113 (33.9-174)
White cell count ($x10^9$/L)	3.8 (0.9-13.9)	6.00 (0.7-64.6)	4.55 (1-19.2)	3.1 (0.6-7.6)	4.9 (0-24.2)
Platelets ($x10^9$/L)	109 (0-484)	88 (12-894)	106.5 (8-401)	42.5 (0-425)	114 (0-764)
Neutrophils ($x10^9$/L)	1.89 (0.30-9.19)	4.56 (0.33-19.58)	2.53 (0.67-10.02)	1.65 (0.50-5.87)	2.89 (0.18-19)
Mutation Frequency: no. of patients (%)					
SF3B1	93 (25.1%)	11 (13.9%)	1 (1.9%)	-	24 (1.5%)
TET2	90 (24.3%)	22 (27.8%)	7 (13%)	-	209 (13.5%)
ASXL1	88 (23.8%)	24 (30.4%)	3 (5.6%)	-	58 (3.7%)
SRSF2	52 (14.1%)	12 (15.2%)	1 (1.9%)	-	101 (6.5%)
RUNX1	46 (12.4%)	6 (7.6%)	-	1 (2.9%)	19 (1.2%)
TP53	43 (11.6%)	8 (10.1%)	1 (1.9%)	1 (2.9%)	20 (1.3%)
U2AF1	41 (11.1%)	8 (10.1%)	2 (3.7%)	1 (2.9%)	36 (2.3%)
STAG2	39 (10.5%	9 (11.4%)	-	-	9 (0.6%)
DNMT3A	36 (9.7%)	12 (15.2%)	-	4 (11.4%)	83 (5.4%)
IDH2	24 (6.5%)	6 (7.6%)	-	-	17 (1.1%)
EZH2	22 (5.9%)	9 (11.4%)	1 (1.9%)	-	11 (0.7%)
ZRSR2	22 (5.9%)	2 (2.5%)	-	-	25 (1.6%)
IDH1	12 (3.2%)	2 (2.5%)	-	-	12 (0.8%)
CBL	9 (2.43%)	5 (6.33%)	-	-	10 (0.6%)
SETBP1	8 (2.16%)	4 (5.1%)	-	-	4 (0.3%)
BCOR	8 (2.2%)	5 (6.33%)	1 (1.9%)		14 (0.9%)
NRAS	8 (2.2%)	6 (7.6%)	-	-	7 (0.4%)
JAK2	4 (1.1%)	12 (15.2%)	-	-	6 (0.4%)
STAT3	-	-	7 (13%)	-	12 (0.8%)

Abbreviations: AML, acute myeloid leukaemia; MPN, myeloproliferative neoplasm; MDS, myelodysplastic syndrome; LPD, lymphoproliferative disorder

† Myeloid malignancy, not further classifiable

6.3.2 The frequency of mutations correlates with the morphological diagnosis

A somatic mutation was detected in 41% (865/2088) of samples referred for investigation of a cytopenia. The presence of a mutation correlated strongly with the morphological diagnosis with ≥1 mutation being detected in 87% of those with a confirmed myeloid malignancy. This increased to 91% when karyotypic abnormalities were included. In those with a confirmed diagnosis of MDS, the spectrum of mutations mirrored that reported in the literature with *SF3B1, TET2* and *ASXL1* being most frequently mutated (25%, 24%, 24% respectively; see Fig 6.2A). At 69% (38/55), those with unilineage dysplasia had a lower but still significant mutation frequency.

Mutations were detected at a much lower frequency in patients with an alternative diagnosis. In the 55 cases with a non-myeloid malignancy (see Table 6.1), mutations were detected in 35% of samples. The majority of these were STAT3 mutations associated with T-large granular lymphocytic leukaemia. In contrast, a mutation was detected in only 7/35 patients with an immune mediated cause for the cytopenia and these were all isolated mutations, most commonly involving DNMT3A.

Importantly, ≥1 mutation was detected in 28% of non-diagnostic cases (412/1496). Of those that harboured a mutation, the median number detected was 1 (range 1-6), significantly lower than that detected in those diagnosed with MDS (median 2; range 1-6; $p<0.0001$). Of note, the mutation spectrum differed from that seen in the MDS cohort, with *TET2* being by far the most frequently mutated gene (47% of mutated cases), followed by *SRSF2* (22%) and *DNMT3A (*19%). Similar to cohort 2 (see Chapter 5), *SF3B1* was infrequently mutated in these cases, likely due to the strong correlation with ring sideroblasts (Papaemmanuil *et al.*, 2011). There were also significantly fewer high-risk mutations including *RUNX1, STAG2* and *TP53*. The median VAF for the mutations detected in the non-diagnostic group was 17.7% (range 5.1-94.7%) and this was also significantly lower than that identified in the MDS group (35.1%; range 5.1-100%; $p<0.0001$) (see Fig. 6.2C).

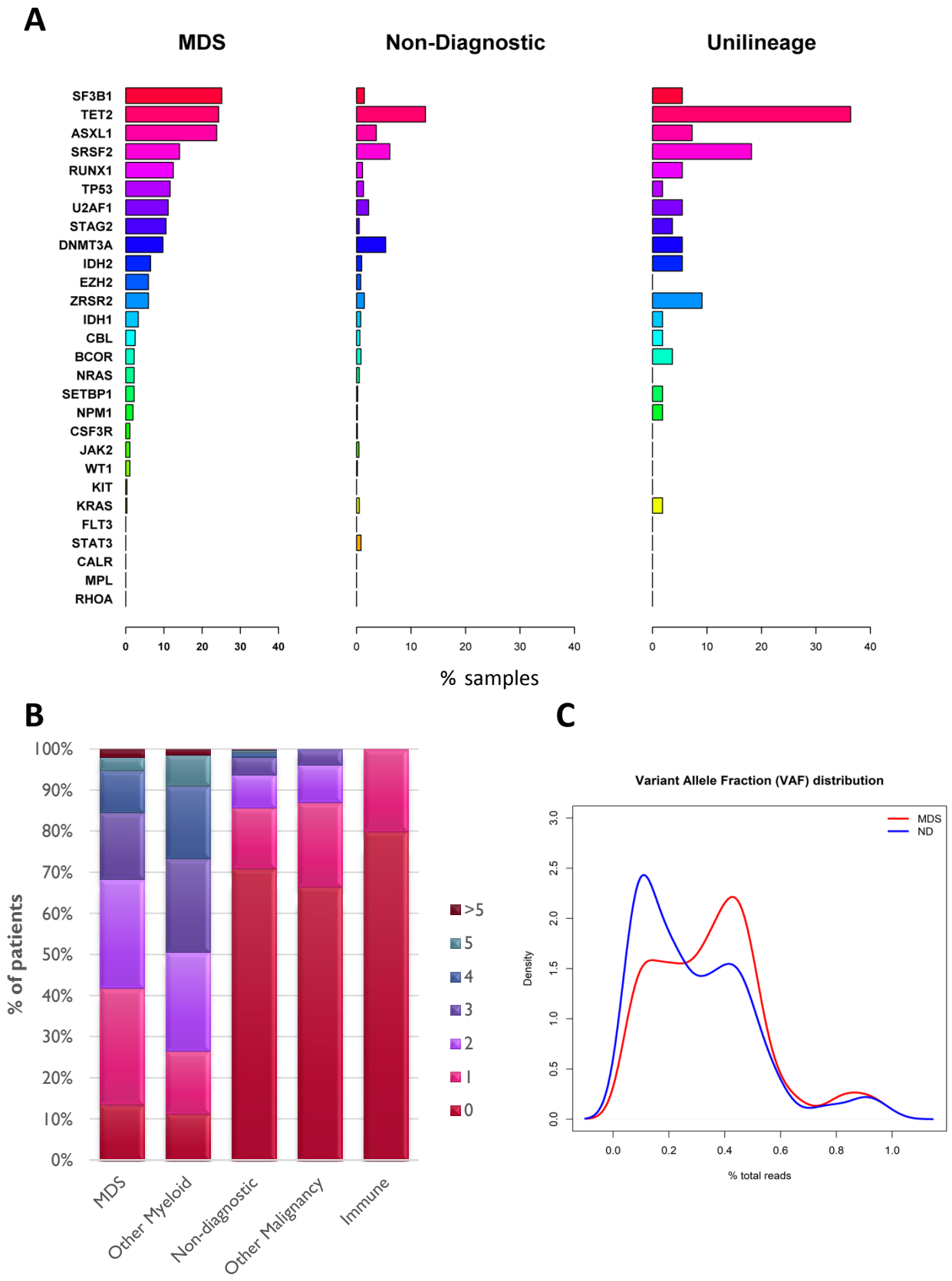


Figure 6.2. Characteristics of mutations detected in patient samples. *(A) Comparison of mutations detected in those with MDS vs a non-diagnostic bone marrow sample vs unilineage dysplasia. (B) Distribution of no. of mutations according to final diagnostic category. (C) VAF distribution in MDS samples (red) vs those with a non-diagnostic bone marrow sample (blue).*

6.3.3 The presence of somatic mutations correlates with overall survival

To determine the clinical impact of somatic mutations, the presence of these was correlated with objective outcome measures including overall survival and blood count trajectories.

The median overall survival for the total cohort was not reached with a median follow-up of 35 months (range 1-46 months). There was, however, a distinctive decline in survival for these patients which correlated strongly with the final morphological diagnosis (see Fig 6.3). This was by far the most pronounced in those with a confirmed myeloid malignancy; those without a malignant diagnosis having the best OS.

There was also a strong correlation between OS and the mutation profile for the total cohort. Survival was significantly reduced in those with a confirmed mutation ($p<0.0001$; Fig.6.4A). This was shown to be progressively worse with each additional mutation, with 5 or more mutations impacting most on survival (Fig.6.4B). The genes that impacted most on survival were assessed on both univariate and multivariate analysis, again using a relaxed lasso regression. As expected, age had a significant impact on survival, and of those genes mutated in >1% of cases, *ASXL1, BCOR, EZH2, IDH1/2, NRAS, RUNX1, SRSF2, STAG2, TP53, U2AF1* and *ZRSR2* were all significantly associated with a worse overall survival on univariate analysis.

Multivariate analysis was performed on all genes as well as restricted to those mutated in >1% of cases. With each analysis the selected genes were entered into a lasso survival regression. Taking the 1SE shrinkage parameter, age, *ASXL1, BCOR, IDH2, RUNX1 and TP53* were consistently selected by the lasso and retained significance in a relaxed lasso regression (Table 6.2).

Table 6.2. Univariate and multivariate overall survival analyses for total cohort. *For multivariable regression, the lasso was used for variable selection and results reported for the corresponding relaxed lasso model.*

Variable	Univariate Analysis	Multivariate Analysis
	HR, 95% CI	***HR, 95% CI***
Age	**1.04(1.036-1.050)**	1.04(1.03-1.05)
ASXL1	**2.82(2.32-3.43)**	1.97(1.58-2.45)
BCOR	**3.48(2.28-5.33)**	2.04(1.31-3.18)
CBL	1.65(0.95-2.85)	-
DNMT3A	1.30(1.00-1.68)	-
EZH2	**3.17(2.25-4.48)**	-
IDH1	**2.27(1.40-3.67)**	-
IDH2	**2.58(1.83-3.64)**	1.87(1.32-2.66)
JAK2	1.40(0.77-2.53)	-
KRAS	1.19(0.49-2.86)	-
NPM1	**3.65(2.01-6.63)**	-
NRAS	**3.50(2.13-5.75)**	-
RUNX1	**3.28(2.49-4.32)**	1.86(1.35-2.55)
SETBP1	**4.17(2.40-7.22)**	-
SF3B1	1.05(0.79-1.40)	-
SRSF2	**1.77(1.42-2.20)**	-
STAG2	**2.66(1.93-3.68)**	-
STAT3	0.36(0.12-1.14)	-
TET2	1.20(1.00-1.44)	-
TP53	**3.97(3.03-5.19)**	3.92(2.99-5.14)
U2AF1	**1.83(1.36-2.46)**	-
ZRSR2	**1.52(1.02-2.26)**	-

HR – Hazard ratio; CI – Confidence Intervals

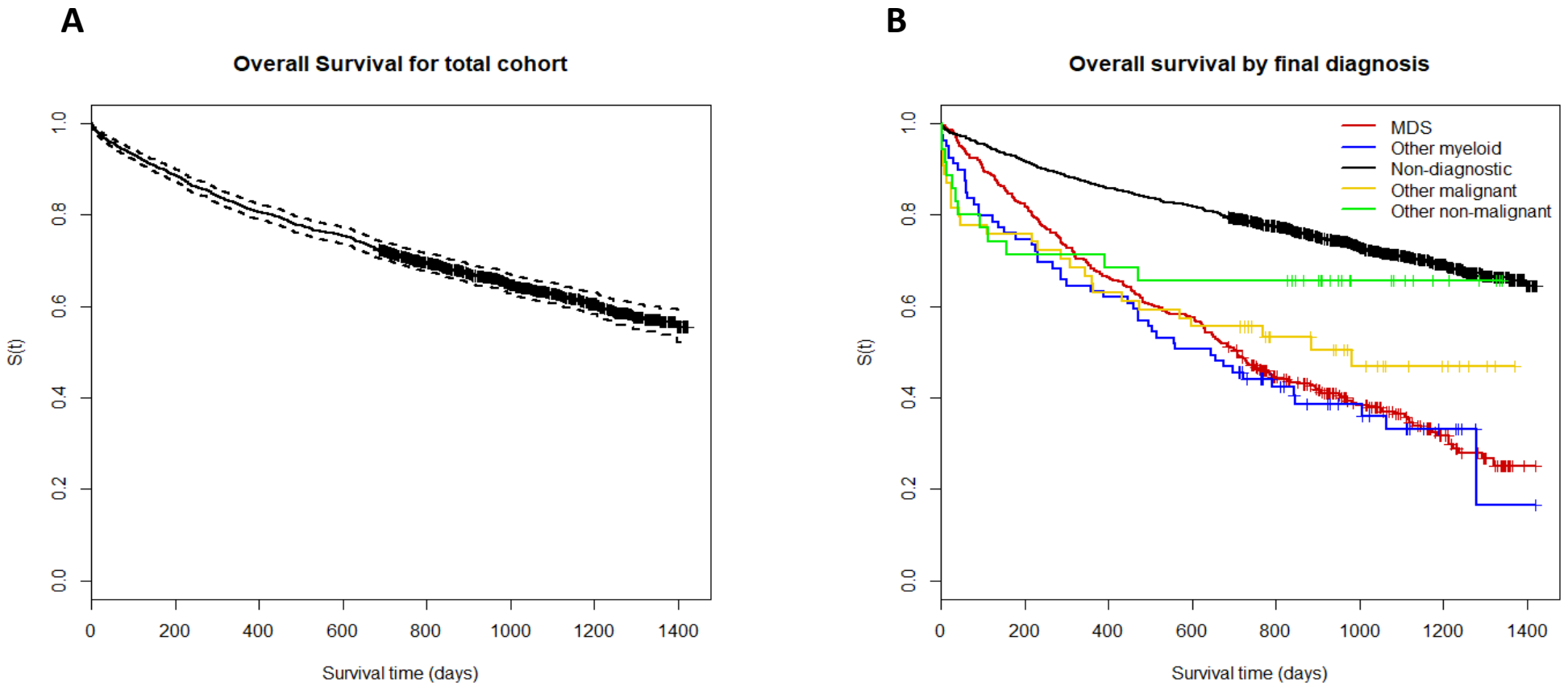


Figure 6.3. Overall survival according to final diagnosis. *(A) Overall survival for total cohort from time of initial sample. (B) Overall survival by final diagnosis on initial bone marrow sample.*

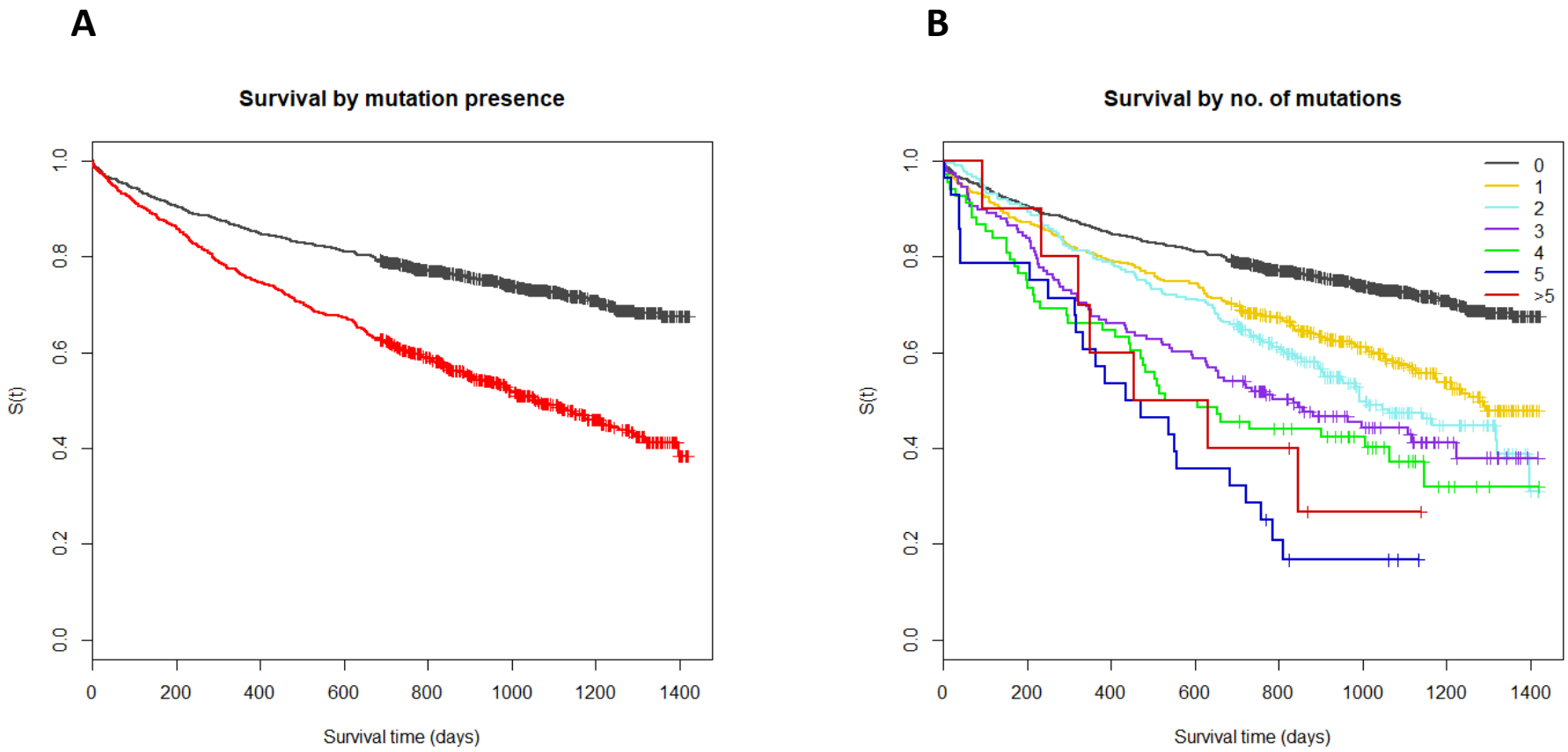


Figure 6.4. Overall survival according to mutation number. (A) Overall survival in total cohort by the presence or absence of a mutation. (B) Overall survival in total cohort by no. of mutations detected at time of initial sample.

6.3.4 Somatic mutations impact on outcome even in the absence of a diagnosis

To determine if somatic mutations also impact on survival in those without a confirmed diagnosis, this group was analysed separately. As highlighted in section 6.2.1, patients whose samples had unilineage dysplasia were not given a formal MDS diagnosis but are included in the analysis presented here. In the first instance, overall survival was assessed with respect to morphological diagnosis. Low risk MDS cases (<5% blasts) were used as a comparator arm as it is this group where diagnosis is particularly difficult and where the potential for overlap with the non-diagnostic and unilineage dysplasia group is highest.

Interestingly, cases with unilineage dysplasia had an outcome indistinguishable from non-diagnostic cases (p=0.914), both groups faring significantly better than those with MDS (p=0.0004) (Fig. 6.5). For this reason, these cases were combined with the ND group for further analysis.

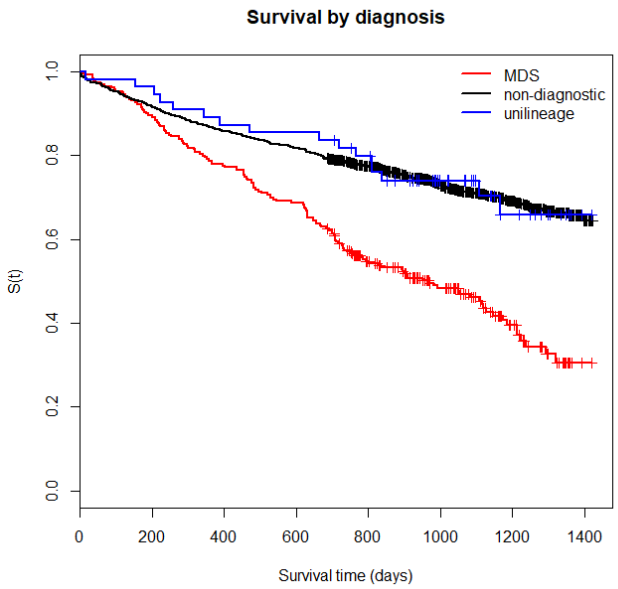


Figure 6.5. Overall survival by diagnosis comparing MDS vs a ND marrow vs unilineage dysplasia

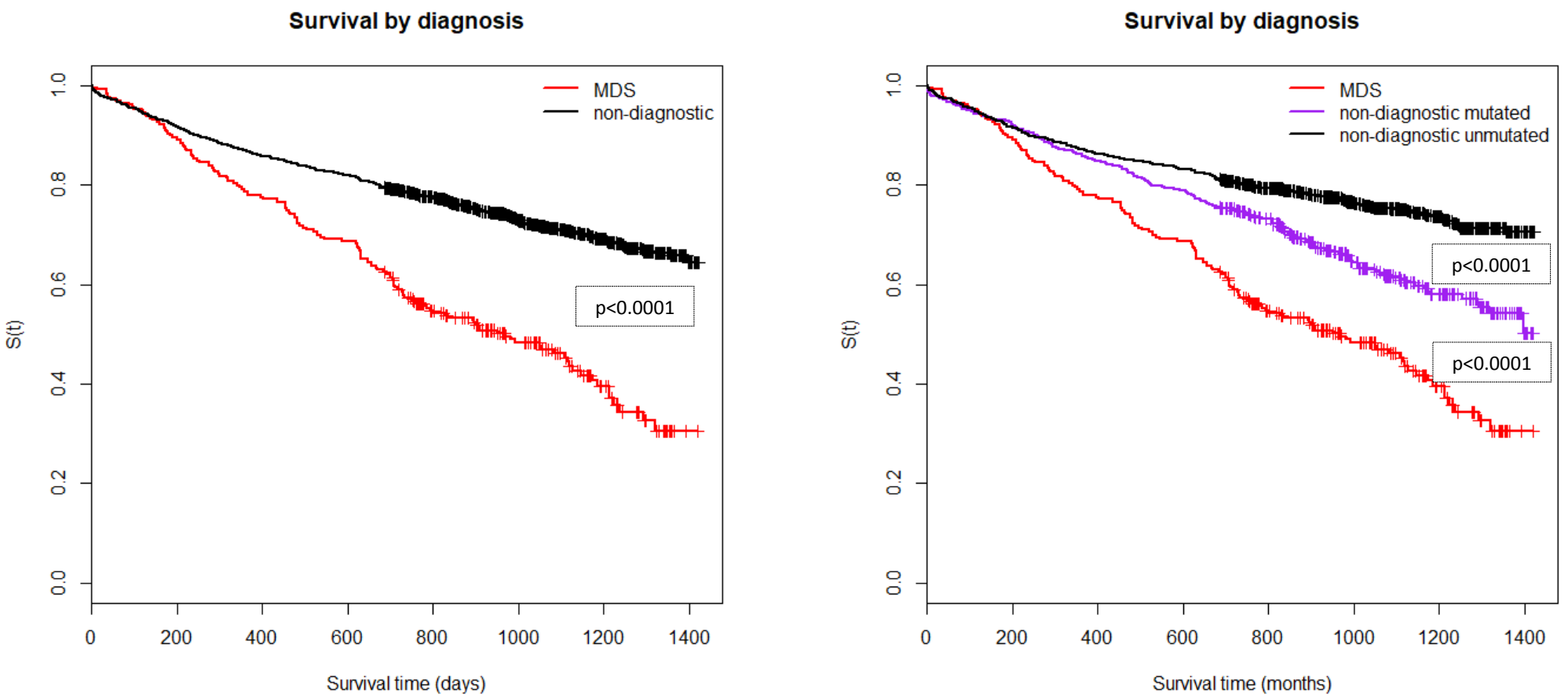


Figure 6.6. Overall survival according to final diagnosis and mutation status. *(A) Overall survival by diagnosis comparing low risk MDS vs. a non-diagnostic bone marrow. (B) Overall survival by diagnosis with non-diagnostic samples separated by the presence or absence of a mutation.*

When combined, the ND patients had a significantly better OS than patients with MDS ($p<0.0001$) with the median survival not reached for ND cases and 968 days for those with MDS (95% CI; 783-1147). However, this survival advantage changed significantly when mutational analysis was incorporated. Those with a mutation (non-diagnostic mutated; ND^{mut}) still had a significantly better OS than MDS patients ($p<0.0001$), though the survival was now significantly worse compared to those without a mutation (non-diagnostic unmutated; ND^{unmut}) ($p<0.0001$) (Fig. 6.6).

With respect to individual mutations in ND samples, on a univariate analysis, *ASXL1, BCOR, EZH2, IDH2, RUNX1, SRSF2* and *TP53* mutations significantly impacted on survival amongst genes mutated in >10 cases. Age had an overwhelming effect on the multivariate analysis, largely due to the lower mutation frequency in this group. To overcome this issue, multivariate analysis was performed initially without age as a variable. By again using a lasso survival regression on both the complete spectrum of genes and those mutated in >1% of cases, similar results were obtained. Mutations in *ASXL1, BCOR, IDH2* and *TP53* significantly impacted on survival with *EZH2* having a borderline effect. Once age was re-introduced, only *ASXL1, BCOR* and *TP53* retained significance.

Importantly, with the exception of *EZH2*, those mutations impacting on survival within the non-diagnostic group mirror those with prognostic relevance in the total cohort.

This analysis confirms that even in the absence of a morphological diagnosis, somatic mutations confer an adverse outcome.

Table 6.3. Univariate and multivariate overall survival analyses for non-diagnostic samples. *For multivariable regression, the lasso was used for variable selection and results reported for the corresponding relaxed lasso model.*

Variable	Univariate Analysis	Multivariate Analysis
	HR, 95% CI	***HR, 95% CI***
Age	**1.04(1.03-1.05)**	-
Mutation detected	**1.59(1.32-1.93)**	-
ASXL1	**2.46(1.71-3.54)**	2.15(1.46-3.16)*
BCOR	**3.78(2.02-7.08)**	3.05(1.59-5.82)*
CBL	0.65(0.16-2.60)	-
DNMT3A	1.24(0.86-1.80)	-
EZH2	**2.77(1.31-5.85)**	2.17(1.00-4.67)
IDH1	1.80(0.81-4.04)	-
IDH2	**1.95(1.01-3.76)**	2.10(1.08-4.06)
RUNX1	**2.53(1.39-4.60)**	-
SF3B1	1.29(0.67-2.50)	-
SRSF2	**1.47(1.06-2.03)**	-
STAT3	0.80(0.26-2.50)	-
TET2	1.15(0.88-1.49)	-
TP53	**2.32(1.27-4.22)**	2.49(1.37-4.53)*
U2AF1	1.28(0.72-2.28)	-
ZRSR2	1.63(0.90-2.97)	-

* Retained significance when age was re-introduced; HR – Hazard ratio; CI – Confidence Intervals

6.3.4 The presence of a mutation predicts for a subsequent diagnosis

To determine whether mutational analysis on ND bone marrows can predict a subsequent myeloid diagnosis, follow-up samples were captured for all patients and any confirmed diagnoses recorded. Out of the 1550 non-diagnostic patients, 205 (13%) had a subsequent follow-up bone marrow biopsy, on the basis of which 81 had a confirmed diagnosis. The follow-up diagnosis was most commonly a myeloid malignancy (61/81; 75%), however a small number had a confirmed diagnosis of aplastic anaemia/red cell aplasia (12/81; 15%), lymphoma (5/81), metastatic carcinoma (2/81) and plasma cell myeloma (1/81). Two patients had a subsequent diagnosis of AML made on a PB sample and 2 an MPN; all 4 were included in the myeloid malignancy group. In addition, 8 patients had a diagnosis of lymphoma and 2 a non-haemopoietic malignancy made on a subsequent lymph node biopsy while 12 patients had a PNH clone detected in PB.

Of those with a subsequent myeloid malignancy (in either PB or BM; n=65), 91% harboured a mutation at the time of the initial bone marrow biopsy. The presence of a mutation was strongly predictive of a myeloid diagnosis with only 0.5% (6/1101) of ND^{unmut} patients having a subsequent diagnosis versus 13% of ND^{mut} patients (59/449; $p<0.0001$). To determine whether certain mutations could predict a subsequent diagnosis, analysis was performed on those cases with a confirmed mutation (n=449). The presence of a mutation at baseline was found to be strongly associated with a subsequent diagnosis, though this was rarely seen in those with an isolated mutation (1 vs >1 mutation; $p<0.0001$). The clone size of the mutations detected in those that progressed also had a VAF that was comparable to MDS (median 31.2% vs 35.1%; $p=0.14$), again suggesting that increased clone size and acquisition of additional mutations is a requirement for overt dysplasia over time. Using Cox regression, multivariate analysis confirmed that *BCOR, EZH2, RUNX1* and *SRSF2* mutations were most predictive for a subsequent confirmed diagnosis (see Table 6.4). This appears to contrast with those mutations most strongly associated with overall survival.

Table 6.4. Univariate and multivariate progression analyses for non-diagnostic samples. *For multivariable regression, the lasso was used for variable selection and results reported for the corresponding relaxed lasso model.*

Variable	Univariate Analysis	Multivariate Analysis	
		Non-competing risk analysis	***With death as a competing risk***
	HR, 95% CI	***HR, 95% CI***	***HR, 95% CI***
Age	0.99(0.97-1.02)	-	-
Gender (M)	**2.57(1.33-4.95)**	3.83(0.99-3.80)	3.76(1.47-9.62)
ASXL1	**2.24(1.19-4.24)**	-	-
BCOR	**5.62(3.43-22.03)**	5.12(2.12-12.38)	6.17(2.34-16.3)
CBL	0.67(0.09-4.84)	-	-
DNMT3A	0.45(0.19-1.04)	-	-
EZH2	**4.63(1.85-11.58)**	6.13(2.30-16.34)	5.84(1.94-17.6)
IDH1	**2.72(0.99-7.52)**	-	-
IDH2	0.42(0.06-3.06)	-	-
RUNX1	**5.26(2.57-10.7)**	2.69(1.27-5.69)	3.19(1.48-6.85)
SF3B1	0.94(0.30-3.02)	-	-
SRSF2	**3.26(1.96-5.45)**	3.32(1.94-5.69)	3.69(2.07-6.59)
STAT3	0.56(0.08-4.02)	-	-
TET2	1.24(0.74-2.07)	-	-
TP53	1.31(0.41-4.17)	-	-
U2AF1	1.64(0.75-3.62)	-	-
ZRSR2	0.95(0.30-3.02)	-	-

HR – Hazard ratio; CI – Confidence Intervals; M - Male

6.3.5 Somatic mutations correlate with blood count parameters and trajectories

To further assess the clinical impact of somatic mutations in those with ND bone marrows, longitudinal blood counts were analysed on a subgroup of patients (n=878). Similar to cohort 2 (see Chapter 5), this analysis was performed using random effects models and the results are summarised in Table 6.5.

Patients with a confirmed mutation had a significantly lower WCC and platelet count, with platelets showing a divergent trajectory over time when compared to those without a mutation. While there was no significant difference in Hb levels at baseline between those with and without a mutation, Hb levels in ND^{mut} patients decreased significantly over time. Of note, the red cell distribution width was also significantly higher in those with a confirmed mutation, a phenomenon also reported in age related clonal haematopoiesis. This effect persisted over the time period measured and followed a divergent trajectory.

With respect to individual mutations, analysis was restricted to those genes mutated in greater than 10 individuals. *ASXL1* and *DNMT3A* were associated with an increasing WCC over time while a falling platelet count was noted in those with a *U2AF1* mutation relative to wild type patients. Haemoglobin was also noted to decrease significantly in those with either *ASXL1 or TET2* mutations versus non-mutated cases. Patients with *TET2, TP53 or U2AF1* had a significantly lower WCC versus non-mutated cases. Regarding RDW, only *ASXL1* mutations were associated with an increasing level over time.

This analysis provides further objective evidence of the clinical impact of somatic mutations in those without a confirmed haematological malignancy.

Table 6.5. Effect of mutations on longitudinal blood counts, assessed using random effects models

	Hb	WCC	Plts	Neut	Mon	Eos	Bas	MCV	MCH	RBC	RDW
Any mutation	0 – Y	– 0 Y	– – Y	– 0 Y	+ 0 Y	– – Y	– 0 Y	+ – Y	+ – Y	– 0 Y	+ + Y
ASXL1	0 – Y	+ + Y	No effect	0 + Y	0 + Y	– – Y	No effect	No effect	No effect	0 – Y	0 + Y
DNMT3A	+ 0 Y	+ + Y	+ – Y	+ 0 Y	+ 0 Y	No effect	0 + Y	No effect	0 + Y	+ – Y	– 0 Y
IDH2	+ 0 Y	No effect	+ – Y	No effect	No effect	+ + Y	– 0 Y	No effect	No effect	+ – Y	0 – Y
SF3B1	0 + Y	No effect	+ 0 Y	No effect	No effect	No effect	No effect	+ + Y	+ + Y	No effect	– 0 Y
SRSF2	– + Y	No effect	– + Y	+ 0 Y	+ 0 Y	No effect	No effect	No effect	0 – Y	– + Y	– 0 Y
TET2	0 – Y	– 0 Y	+ – Y	No effect	No effect	No effect	No effect	0 – Y	0 – Y	No effect	No effect
TP53	0 + Y	– 0 Y	No effect	– + Y	– 0 Y	– + Y	No effect	– + Y	– + Y	0 + Y	No effect
U2AF1	No effect	– 0 Y	– – Y	No effect	+ 0 Y	No effect	No effect	No effect	No effect	+ 0 Y	No effect
ZRSR2	No effect	No effect	+ – Y	No effect	No effect	No effect	No effect	– 0 Y	– 0 Y	No effect	– 0 Y

Abbreviations: Hb, haemoglobin; WCC, white cell count; Plts, platelets, Neut, neutrophils; Mon, monocytes; Eos, eosinophils; Bas, basophils; MCV, mean cell volume; MCH, mean cell haemoglobin; RBC, red blood cell count; RDW, red cell distribution width

+ + Y statistically significant positive difference with diverging trajectories that persist over the time of observation.

– – Y statistically significant negative difference with diverging trajectories that persist over the time of observation

– + Y statistically significant negative difference with converging trajectories that persist over the time of observation.

+ – Y statistically significant positive difference with converging trajectories that persist over the time of observation.

0 – Y no overall difference but a statistically significant decrease in the mutated relative to the non-mutated

0 + Y no overall difference but a statistically significant increase in the mutated relative to the non-mutated

+ 0 Y statistically significant positive difference with no convergence/divergence of trajectories.

– 0 Y statistically significant negative difference with no convergence/divergence of trajectories.

6.4 Discussion

Analysis of cohort 6 has confirmed that somatic mutations are identified at high frequency in patients referred for investigation of cytopenia. This includes almost a third of cases without a confirmed morphological diagnosis. Importantly, mutations in the latter group of patients show some overlap with diagnostic samples, however distinct differences were also noted. The median number of mutations was lower in the non-diagnostic group and the clone size of these mutations was also significantly lower, particularly in those without a recorded diagnosis in HMDS on follow-up. In the non-diagnostic group, it was very rare for a patient with an isolated mutation at baseline to have a subsequent diagnosis. However, the likelihood of progression increased with each additional mutation and those that did progress had a median VAF comparable to MDS patients. This is consistent with the findings from cohort 1, when it was postulated that >1 mutation and >20% VAF were a requirement for a confirmed diagnosis of MDS. This is also similar to that reported in aging healthy individuals in whom a mean VAF of 25.2% was noted in those that developed a haematological malignancy (Jaiswal *et al.*, 2014). Similarly, a study by Malcovati et al showed that >1 mutation had a very high positive predictive value for a myeloid diagnosis, while a VAF of 8.7% had the highest discriminatory ability (Malcovati *et al.*, 2017). It should be noted however, that the latter study, which predominantly analysed the predictive value of peripheral blood taken alongside the bone marrow specimen, was not longitudinal, and only included a small number of non-diagnostic bone marrow samples (n=154).

On further analysis, certain mutations were shown to be associated with a subsequent confirmed diagnosis including *BCOR, EZH2, RUNX1* and *SRSF2*. It is noted however, that follow-up bone marrow biopsies in the present study were at the discretion of the referring clinician, and hence the impact of selection bias is difficult to judge. It is possible that other cases would have been diagnosed had they undergone further sampling and it is possible these mutations are associated with more profound cytopenias leading clinicians to re-investigate at an earlier timepoint. Unfortunately, there were insufficient numbers to investigate this on

the longitudinal blood count analysis. It is also possible that these mutations are more strongly associated with morphological dysplasia over time, hence leading to a confirmed diagnosis, though this would be challenging to investigate. Of note, a similar spectrum of genes was found to be predictive of a myeloid diagnosis by Malcovati and colleagues, who identified a highly specific mutation pattern which included spliceosome mutations, and co-mutation patterns involving *TET2, ASXL1* and *DNMT3A* with any of *RUNX1, EZH2, CBL, BCOR, CUX1, TP53 or IDH1/IDH2* (Malcovati *et al.*, 2017). ICUS patients with these mutations had a significantly higher risk of developing a myeloid malignancy. When applied to cohort 3, this reported mutation pattern was also highly predictive of a subsequent diagnosis ($p<0.0001$), although there is significant overlap between the predictive mutations reported by Malcovati et al and those identified in cohort 6.

While these mutations predicted for a subsequent diagnosis, a relatively distinct set of mutations were associated with poor overall survival. Of note, the presence of a mutation had a significant impact on survival in the ND group, though these patients still had a significantly better survival than those with a confirmed MDS diagnosis. On multivariate analysis mutations in *ASXL1, BCOR, IDH2* and *TP53* had the greatest impact on survival; but it is unclear how these mutations may have contributed to patient mortality, particularly in the absence of cause of death information. Irrespective of the cause of death, analysis from this cohort suggests there is a detrimental effect of somatic mutations even in the absence of a diagnosis.

On analysis of longitudinal blood counts, the presence of a mutation impacted significantly on individual parameters, in the absence of a confirmed haematological malignancy, with progressive cytopenias over time. It is therefore essential that these patients are identified and monitored for the potential risks associated with progressive changes in blood counts.

Of interest, in aging healthy individuals, patients with mutations were found to have a higher red cell distribution width (RDW) and those with an RDW above the normal range along with a mutation had a markedly increased risk of death (Jaiswal

et al., 2014). The authors postulated that the increased RDW could be a marker of disturbed haematopoiesis, which could be related to the increased cumulative incidence of coronary heart disease and ischaemic stroke. Furthermore, increased RDW has been shown to correlate with AML risk in healthy individuals and increased levels can be present several years before an AML diagnosis (Abelson *et al.*, 2018). Indeed, this parameter has been incorporated into an AML prediction model (Abelson *et al.*, 2018).

Analysis of cohort 6 also showed an association between mutations and an increasing RDW which persisted over time with a divergent trajectory. On further analysis of individual genes, only *ASXL1* was associated with an increasing level. Data from this analysis would therefore support the theory that RDW may be a marker of disturbed haematopoiesis resulting from the underlying clone. This may also have contributed to the increased mortality in this patient group, similar to that reported in healthy individuals.

In summary, this analysis has provided key information regarding the mutational landscape in patients with unexplained cytopenia. Mutations have been shown to impact significantly on both blood count parameters and overall survival even in the absence of a confirmed diagnosis and these patients require close clinical follow-up at a minimum. While a proportion will be ultimately diagnosed with a myeloid malignancy, many succumb without a diagnosis and early detection and monitoring may provide an opportunity for intervention. To focus genetic testing, it may be possible to triage samples using parameters such as RDW plus other baseline blood counts. It may also be possible to use alternative techniques, including flow cytometry as used in those with a monocytosis; further analysis of this cohort using such techniques is underway.

CHAPTER 7: GENERAL DISCUSSION AND CONCLUSIONS

The diagnosis of MDS and CMML remains heavily reliant on the morphological assessment of the bone marrow and this subjective approach has well recognized limitations. The overall aim of this research was to investigate whether novel technologies could improve the diagnosis of these conditions, particularly in those with early disease who can be missed using current techniques. Over the past 15-20 years there have been significant advances in the field of molecular diagnostics with the introduction of array-based technologies and high throughput sequencing. This has led to an explosion of research into the underlying pathological mechanisms of all malignancies and the identification of numerous genetic abnormalities which are responsible for disease. Due to the high frequency of somatic mutations in chronic myeloid malignancies, this research focused on the use of high throughput sequencing with the addition of array-based cytogenetics and flow cytometry for certain aspects of the work. At the time this research commenced there was no published literature in this area, the findings from this thesis producing some of the earliest publications investigating these diagnostic approaches.

To achieve the main aim of this research, a targeted high throughput sequencing panel was developed for use in the investigation of patients with suspected or confirmed myeloid malignancies. This was ultimately developed for the Illumina MiSeq with the goal of implementing the procedures into routine clinical practice. As such, it was essential that the processes were not labour intensive, could be delivered in real time and were cost effective. The panel, which was designed by myself in collaboration with Dr Paul Evans, targeted the 26 most commonly mutated genes across the myeloid malignancies. This included MDS, MPN, MDS/MPN and AML. By encompassing genes implicated across the spectrum of diseases, the panel could be used on a much larger patient cohort allowing for a high throughput of cases. Since the initial design of this panel, further genes have been implicated in disease pathogenesis and there are now 40-50 genes reported to be recurrently mutated in MDS. However, only a small number are mutated in

greater than 10% of patients, and the myeloid panel presented here focuses on the most commonly mutated; importantly including the 16 most frequently mutated seen genes in MDS. By limiting the panel to these 26 genes, it is not only economically viable for routine use but has also identified mutations at a frequency comparable to much larger studies.

This research also provided the framework for implementation of this panel into routine clinical practice in 2015, and HMDS was one of the first laboratories in the UK to offer this as a UKAS (UK accreditation scheme) accredited test. To date, over 12,000 samples have been sequenced for both routine use and research using this panel.

7.1 Myeloid panel as a diagnostic tool

The main aim of this research was to determine if new technologies could improve the diagnosis of chronic myeloid malignancies. The initial analysis of cohort 1 confirmed that mutations could be detected at a very high frequency in cytopenic patients who failed to meet diagnostic criteria but subsequently progressed to confirmed MDS/AML. The mutation profile and clone size were similar to those of MDS and differed from that reported in healthy individuals. In contrast, structural abnormalities, identified by SNP-array, were comparatively rare and more commonly acquired at the point of diagnosis. This suggested that mutation analysis was the optimal tool to identify patients with clinically significant disease in the absence of morphological findings.

This was further supported by the analyses of patients being investigated for either persistent monocytosis or cytopenia. These studies confirmed a high frequency of mutations in patients without a confirmed diagnosis (57% in patients with a monocytosis, 28% in cytopenic patients); the presence of a mutation having clinical significance in these patients. In those presenting with a monocytosis, patients had a similar mutation profile, immunophenotype, and outcome irrespective of whether a morphological diagnosis was made. Importantly the presence of a

mutation had a significant impact on overall survival, with outcome being indistinguishable from those with CMML. This suggests that patients who do not fulfil current diagnostic criteria but harbour a somatic mutation, should be managed similarly to those with a confirmed diagnosis.

The findings were not however as clear cut in the cytopenic patient group. While mutations impacted significantly on both blood count parameters and overall survival the mutation spectrum, number of mutations and VAF of those mutations differed between the ND patients and those with confirmed MDS. The negative survival impact in the ND^{mut} patients was also not as profound as that seen for MDS. This may simply reflect earlier stage disease, however further follow-up of this patient group is needed to determine the ultimate impact of mutation detection in the absence of morphological disease.

Concern also remains regarding the detection of similar mutations in healthy aging individuals. The data from this research would suggest that there are distinct differences between healthy individuals and those with clinically significant disease. The latter are far more likely to harbour more than 1 mutation, and the VAF or clone size of these mutations is invariably greater than 20%. Caution must therefore be applied in patients with isolated low-level mutations in the absence of definitive morphological disease.

To ensure that mutation analysis is applicable in the routine diagnostic service, it will be essential to triage samples and restrict testing to those with a high pre-test probability of detecting an abnormality. With regards to patients presenting with a monocytosis, flow cytometry has shown potential as a screening tool while blood count parameters could be utilised in those with cytopenias. Further work is needed to develop predictive tools and optimize patient testing.

7.2 Myeloid panel as a prognostic tool

The panel utilized in this research can also provide important prognostic information even in those without a confirmed WHO diagnosis. In the analysis of

cohort 1, *IDH2* and *TP53* mutations were associated with a more rapid time to progression to AML/RAEB from the pre-diagnostic sample, while *TP53* and the number of mutations were associated with a worse OS.

Importantly mutations in both *TP53* and *IDH2* also had the greatest impact on survival in the ND patients in cohort 3, along with *ASXL1, BCOR* and *EZH2*. In patients with unexplained cytopenia, the presence of a mutation resulted in a significantly worse survival compared to those without a demonstrable mutation. In patients investigated for a monocytosis, survival was also significantly worse even in the presence of an isolated mutation and those mutated patients without a confirmed diagnosis had an outcome comparable to CMML patients.

This further validates the use of mutation analysis in the investigation of patients with either a monocytosis or cytopenias and has the potential to identify high risk patients much earlier in the natural history of their disease. This may provide an increased window of opportunity which, in the future, could be used for early intervention to possibly change the disease course.

7.3 Implications for routine practice

This work has several potentially significant implications for routine practice. With regards to the laboratory, this could significantly increase the number of samples referred for investigation of both cytopenias and monocytosis. To date, only those with a high pre-test probability of disease are referred for investigation; however, this is likely to increase as knowledge of the clinical impact of mutation detection grows.

As such testing becomes increasingly feasible in routine laboratories, it is essential that procedures are standardized, not only with respect to panel design but also in respect to variant filtering and annotation. The latter is particularly relevant in the absence of morphological disease, when the presence of a mutation is the only measure of disease.

To date researchers have applied their own criteria for variant annotation and as a result both somatic and germline databases have become extensively contaminated

with misplaced variants. Hence, in recognition of the fact that a high level of experience is needed to both analyse and interpret such data a concerted effort both nationally and internationally to optimize databases and standardize reporting processes is now underway.

In the UK, NHS England have recently reconfigured genomics services across England with the goal of making standardized genomic testing available to everyone. This strategy will hopefully provide a national platform to standardize all elements of laboratory practice and provide guidance on analysis and interpretation of high throughput sequencing data.

With respect to clinical work, the ability to detect increasing numbers of patients with mutations will significantly increase the numbers of patients requiring follow-up in haematology clinics. While it is essential that these patients are monitored, there is still uncertainty as to what these mutations mean in certain clinical situations. CHIP/CCUS clinics have already been established in large centres in the US and this is also being considered in the UK as patient numbers increase. These clinics will require expertise in the area of genomics and will need to be suitably resourced. This will become increasingly important as early intervention is considered in this patient group.

It will also be essential that all practicing clinicians and doctors in training are educated regarding genomic testing and the interpretation of the results. This information will be increasingly fed back to local haematologists and it is essential that this information is consistently and reliably fed back to patients.

7.4 Future Work

While this research has provided extensive insight into the genetic landscape of patients with both monocytosis and cytopenia, further work is needed. In the first instance, further analysis will be performed on cohort 3 in preparation for

publication. This will include the incorporation of flow cytometry data, and the development of predictive tools for triaging samples and predicting outcomes.
The follow-up of patients in both cohort 2 and 3 is relatively short, and future reanalysis should provide additional information on the long-term clinical impact of somatic mutations. This follow-up analysis, which is planned for both 5 years and 10 years from the date of the last sample, will include updated survival and any subsequent diagnoses.

The current myeloid panel is small and this has its limitations. A larger panel encompassing less frequently mutated genes would expand on the research to date and potentially provide additional key diagnostic and prognostic information. To investigate these less frequently mutated genes would however require much larger patient cohorts with long clinical follow-up. In HMDS, a large pan-haemonc panel has been developed encompassing 238 genes implicated across the spectrum of haematological malignancies. This panel also includes a copy number back bone which will allow copy number changes to be detected alongside somatic mutations. This panel will provide a more extensive platform to further investigate cytopenic patients. Providing funding can be acquired, samples included in this research will be considered for resequencing alongside the analysis of new samples referred for investigation. It is likely that our definitions of myeloid malignancy will change as knowledge of underlying disease pathogenesis increases. Further research using large unselected patient cohorts such as those utilized in this research should aid in the re-classification of these disorders.

It is also likely, as part of NHS England's genomic strategy, that whole genome sequencing will become increasingly available. This will provide an even greater wealth of knowledge for this patient group. It will, however, also lead to increasing numbers of patients in whom CHIP mutations are detected, not only in those with suspected haematological malignancies but also in the solid tumour group and this needs to be considered when targeting future populations.

7.5 Final Conclusions

In conclusion, this research has provided valuable information regarding the diagnostic potential of high throughput sequencing in patients with suspected MDS and CMML. This will help refine the diagnostic criteria for these diseases, potentially providing an opportunity for early intervention in high risk patients. This will, however, impact significantly on laboratory and clinical work requiring standardization and extensive resourcing. Further large studies will be needed to allow molecular classification of these diseases.

Appendix 8.I

Ethics Approval

Health Research Authority
North East - York Research Ethics Committee
Jarrow Business Centre
Viking Business Park
Rolling Mill Road
Jarrow, Tyne & Wear
NE32 3DT

Telephone: 0207 104 8085

24 March 2016

Dr Catherine Cargo
Consultant Haematologist
Leeds Teaching Hospitals NHS Trust
Haematological Malignancy Diagnostic Service
St James's Institute of Oncology
Beckett Street, Leeds
LS9 7TF

Dear Dr Cargo

Study title:	**Development of laboratory techniques to improve the diagnosis of patients with myeloid malignancies**
REC reference:	**16/NE/0105**
IRAS project ID:	**190526**

The Proportionate Review Sub-committee of the North East - York Research Ethics Committee reviewed the above application in correspondence.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager Miss Kathryn Murray, nrescommittee.northeast-york@nhs.net. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Ethical opinion

On behalf of the Committee, the sub-committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA Approval (England)/ NHS permission for research is available in the Integrated Research Application System, www.hra.nhs.uk or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publicly accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion").

Approved documents

The documents reviewed and approved were:

Document	*Version*	*Date*
REC Application Form [REC_Form_11032016]		11 March 2016
Research protocol or project proposal [Study Protocol]	v1.3	09 March 2016
Research protocol or project proposal [Study Protocol]	v1.3	09 March 2016
Summary CV for Chief Investigator (CI) [CV for Chief Investigator]		07 March 2016
Summary CV for supervisor (student research) [Supervisor CV]		07 March 2016

Membership of the Proportionate Review Sub-Committee

The members of the Sub-Committee who took part in the review are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document “After ethical review – guidance for researchers” gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:
http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

16/NE/0105	**Please quote this number on all correspondence**

Yours sincerely

pp.
Mr Chris Turnock
Chair

Email: nrescommittee.northeast-york@nhs.net

Enclosures: *List of names and professions of members who took part in the review*

"After ethical review – guidance for researchers" [SL-AR2]

Copy to: *Ms Anne Gowing, Leeds Teaching Hospitals NHS Trust*

North East - York Research Ethics Committee

Attendance at PRS Sub-Committee of the REC meeting in Correspondence

Committee Members:

Name	*Profession*	*Present*	*Notes*
Dr Mary Connor	Coaching & Mentoring Consultant	Yes	
Mr Chris Turnock (Chair)	Head of Technology Enhanced Learning	Yes	
Ms Lorraine Wright	Senior Research Nurse	Yes	

Also in attendance:

Name	*Position (or reason for attending)*
Miss Kathryn Murray	REC Manager

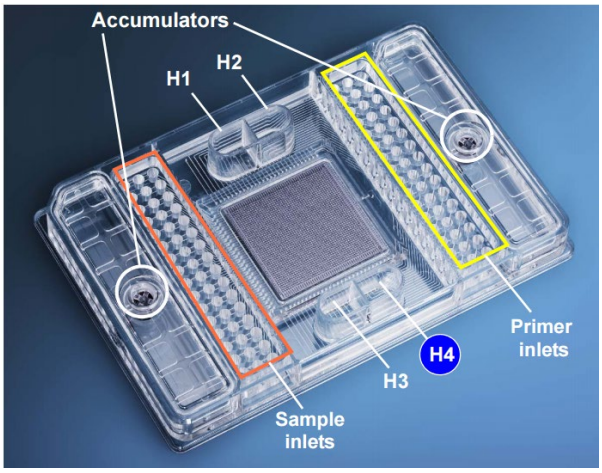
Appendix 8.2

Fluidigm and MiSeq Laboratory Protocol

Appendix 8.2 Fluidigm and MiSeq Laboratory Protocol

IN THE PCR LABORATORY

Priming the Access array

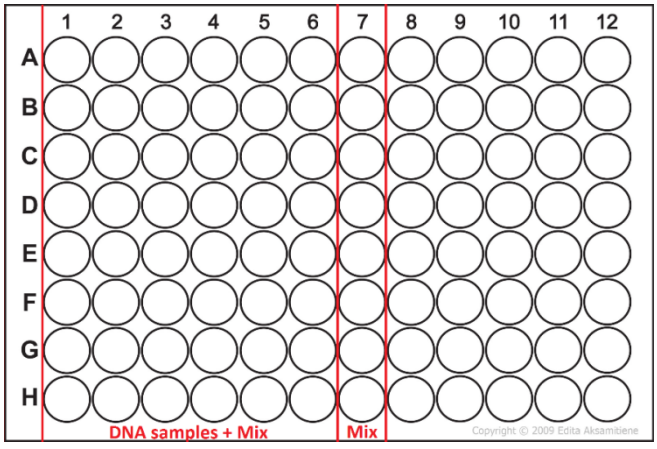


a. Inject control line fluid in both accumulators of the IFC.
b. Use one syringe for each accumulator. Add 500 µl of 1X Access Array Harvest Solution to the H1, H2, and H3 wells on the IFC.
c. Add 500 µl of 1X Access Array Hydration Reagent v2 to the H4 well on the IFC.
d. Remove the blue protective film from the bottom of the IFC.
e. Turn on the IFC Controller AX in the Pre-PCR using the on/off switch on the back of the machine.
f. Click on **Eject** to move the tray out of the IFC Controller AX.
g. Place the IFC onto the tray by aligning the notched corner of the IFC to the A1 mark.
h. Press **Load Chip** to register the barcode of the IFC and activate the script selection.
i. Select **Prime (151x)** and login as a **User**.
j. Click on **Run Script** to prime the IFC Access Array (Run time = 6 minutes).
k. Once the script is complete, press **Eject** to remove the IFC.

Sample preparation

a. **Reaction mix required for 48 samples.**

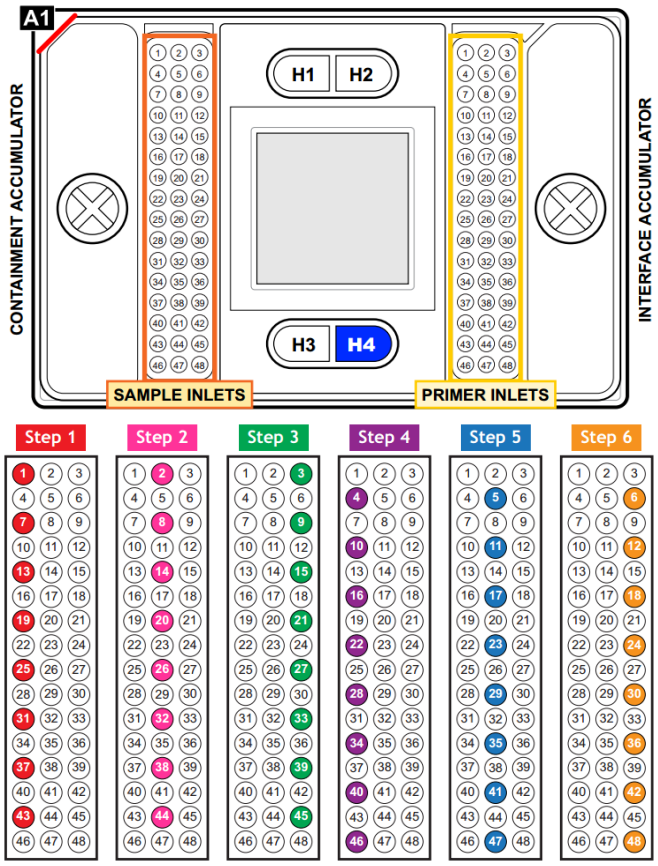
	Volume
10X FastStart High Fidelity Reaction Buffer without MgCl2	45µl
MgCl2	81 µl
DMSO	22.5µl
10 mM PCR Grade Nucleotide Mix	9µl
5 U/µL FastStart High Fidelity Enzyme Blend	4.5µl
20X Access Array Loading Reagent	22.5µl
PCR Certified Water	176 µl



b. Vortex sample Pre-Mix for 20 seconds.
c. Add 40 µl of sample Pre-Mix into column 7 (A7 – H7) of a 96-well PCR plate.
d. Transfer 5 µl of sample Pre-Mix from column 7 into columns 1-6 using a multi-channel pipette.
e. Save remaining Pre-Mix.
f. Add 1.25 µl of DNA sample from the created DNA plate into columns 1-6 of the same 96-well PCR plate.
g. Cover the plate with an adhesive seal.
h. Briefly vortex and centrifuge the sample mix plate.

Sample loading onto the 48:48 access array

a. All the 48 primer solutions are in a 96-well PCR plate (columns 1-6). Different primers plates are used for lymphoid and myeloid runs. Briefly vortex and centrifuge the primer plate. Add 4.3 µl primer solution into each of the primer inlets. The primers from column 1 of the primer plates should be added to the primer inlets (right side of the IFC) shown at step 1. Primers from column 2 into the inlets of step 2, etc. While pipetting, do not go past the first stop on the pipette. Doing so may introduce air bubbles into the inlets.



b. Add 4.3 µl of each sample mix into the Sample Inlets (left side of the IFC) using the same pipetting scheme as shown in the previous step. While pipetting, do not go past the first stop on the pipette. Doing so may introduce air bubbles into the inlets.
c. Use a magnifying glass to make sure there are no air bubbles near the middle dot in both the Sample and Primer Inlets.
d. Press **Eject** to move the tray out of the IFC Controller AX.
e. Place the IFC onto the tray by aligning the notched corner of the IFC to the A1 mark.
f. Press **Load Chip** to register the barcode of the IFC and activate the script selection.
g. Select **Load Mix (151x)** and login as a **User**. Click on **Run Script** (run time = 1 hour).
h. Once the script is complete, press **Eject** to remove the IFC.

Click on the **Logout** symbol at the bottom right of the screen followed by **Log off** and turn

off the IFC Controller AX.

Thermal cycling on the FC1™ Cycler

a. Turn on the FC1™ Cycler in the Pre-PCR using the on/off switch on the back of the machine.
b. Click on **Start** > **Login** > **User** > **Start**.
c. Open the lid and transfer the IFC access array onto the FC1™ Cycler. Close the lid.
d. Click on **Continue** and choose protocol **AA 48x48 Standard v1** (run time = 2 hours) and click on **Run**.
e. Once the script is complete, Press **OK** remove the IFC out of the FC1™ Cycler.

Click on the **Logout** symbol at the bottom right of the screen followed by **Log off** and turn off the FC1™ Cycler.

48:48 standard Thermal Cycling Programme

STEP 1	TEMP	TIME	STEP2	TEMP	TIME	STEP3	TEMP	TIME	STEP4	TEMP	TIME
FIRST CYCLE	50	120sec	2nd CYCLE	70	20 min	3rd cycle	95°	10min	1st PCR CYCLES	95 60 72	15 secs 30 secs 60 secs
1x			1x			1x			10x		
STEP 5	**TEMP**	**TIME**	**STEP 6**	**TEMP**	**TIME**	**STEP 7**	**TEMP**	**TIME**	**STEP 8**	**TEMP**	**TIME**
1st C0t cycle	50 80 60	15sec 30sec 30sec	2nd PCR CYCLE	95 60 72	15secs 30 secs 60secs	2nd C0t cycle	50 80 60	15sec 30sec 30sec	3rd PCR CYCLE	95 60 72	15secs 30 secs 60secs
2x	72	60sec	8x			2x	72	60sec	8x		
STEP 9	**TEMP**	**TIME**	**STEP 10**	**TEMP**	**TIME**	**STEP 11**	**TEMP**	**TIME**			
3rd C0t cycle	50 80 60	15sec 30sec 30sec	Ext	72	3 min	Hold	4°	Hold			
5x	72	60sec									

IN THE POST PCR LABORATORY

Harvesting the sample libraries

a. Move Access Array IFC into the Post-PCR laboratory.
b. Remove remaining fluids from the H1-H4 wells.
c. Add 600 µl of 1X Access Array Harvest Solution into wells H1-H4.
d. Add 3 µl of 1X Access Array Harvest Solution into each of the Sample Inlets on the IFC.
e. Turn on the IFC Controller AX in the Post-PCR using the on/off switch on the back of the machine.
f. Click on **Eject** to move the tray out of the IFC Controller AX.
g. Place the IFC onto the tray by aligning the notched corner of the IFC to the A1 mark.
h. Press **Load Chip** to register the barcode of the IFC and activate the script selection.
i. Select **Harvest v5 (151x)** > **User** > **Run Script** (Run time = 1 hour).

Sample barcoding

a. In the Pre-PCR laboratory, prepare a new Sample Pre-Mix using the table below. Vortex the Pre-Mix.

b. Barcoding reaction mix required for 48 samples.

	Volume
10X FastStart High Fidelity Reaction Buffer without MgCl2	120µl
25mM $MgCl_2$	216µl
DMSO	60µl
10 mM PCR Grade Nucleotide Mix	24µl
5 U/µL FastStart High Fidelity Enzyme Blend	12µl
PCR Certified Water	468µl

c. Add 102 µl of sample Pre-Mix into column 8 (A8 – H8) of a clean labelled 96-well PCR plate.

d. Transfer 15 µl of sample Pre-Mix from column 8 into columns 2-7 using a multi-channel pipette. Leave column 1 empty to try to prevent evaporation.

e. There are 8 different sets of barcodes available in four 96-well PCR plates (as shown below). Each set of barcodes should be used before starting with the first set of barcodes again, i.e. if the previous run used barcodes 97-144 then barcodes 145-192 should be used next.

Plate	Columns (1-6)	Columns (7-12)
A1	1-48	49-96
A2	97-144	145-192
A3	193-240	241-288
A4	289-336	337-384

f. Add 4 µl of all 48 barcode libraries to the wells with 15 µl of sample Pre-Mix. When the barcodes in columns 1-6 are used, the seal should only be partial removed and then re-used. The seal should be completely removed when using barcodes in columns 7-12 and a new clear plate seal should be applied.

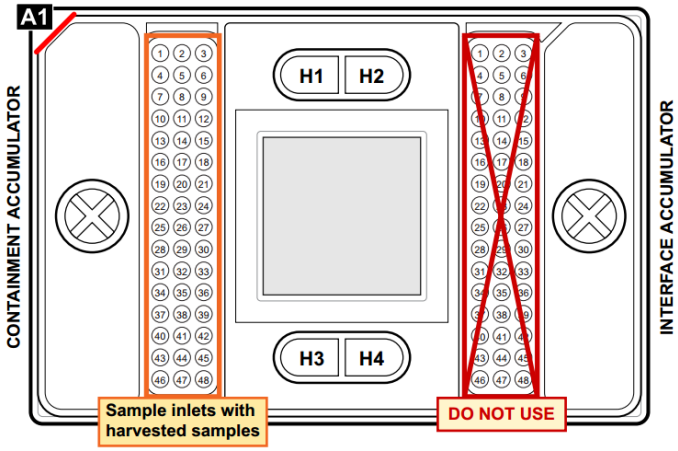
This is a sample transfer step and the tube/plate alignment must be checked and signed by a second person if it is a complex transfer. If it is tube to tube, tube to plate, or plate to plate, in the same position, then no 2nd check is needed providing the following safe transfer practices are adhered to.

Safe transfer

1. Split samples into small groups when transferring tube to tube.
2. Split tube samples into blocks of 8 to transfer into plates.
3. Put marker pen on plates where sample is added
4. Use a multichannel pipette where possible
5. Cap off rows once sample is added.

g. Seal the 96-well PCR plate with Pre-Mix and barcodes with a sticky clear plate seal and transfer to the Post-PCR laboratory.

h. In the Post-PCR, once the Harvest v5 script is complete, press **Eject** on the IFC Controller AX to remove the IFC.
i. Click on the **Logout** symbol at the bottom right of the screen followed by **Log off** and turn off the IFC Controller AX.
j. Use a clean labelled 96-well PCR plate and carefully transfer the harvested samples (10 µl) from the sample inlets into columns 1-6 using the same pipetting scheme as on page 3. Briefly centrifuge the plate.



k. In the same PCR plate, add 99 µl of distilled water into all the wells of columns 7-12. Transfer 1 µl of the harvested PCR products to the wells with water, i.e. column 1 to 7, column 2 to 8 etc.
l. Carefully mix the diluted PCR products wells by pipetting.
m. Add 1 µl of diluted harvested PCR products to the 96-well plate containing Sample Pre-Mix and the barcodes.

This is a sample transfer step and the tube/plate alignment must be checked and signed by a second person if it is a complex transfer. If it is tube to tube, tube to plate, or plate to plate, in the same position, then no 2nd check is needed providing the following safe transfer practices are adhered to.

Safe transfer

1. Split samples into small groups when transferring tube to tube.
2. Split tube samples into blocks of 8 to transfer into plates.
3. Put marker pen on plates where sample is added
4. Use a multichannel pipette where possible
5. Cap off rows once sample is added.

n. Seal the 96-well plate ready for PCR with a new clear plate seal. Briefly vortex and centrifuge.
o. Seal the plate with harvested PCR products and diluted PCR products and store in the freezer.
p. Transfer the 96-well PCR plate to the FISH Post-PCR laboratory.

q. Turn on the G-Storm thermal cycler by the on/off switch on the back in the top right corner. Open the thermal cycler by using the pressure dial on the lid.
r. Place the 96-well PCR plate into the thermal cycler and close the lid.
s. Click on **Run Program** > **Fluidigm barcodes** > **Run selected**.
t. Enter the volume as **20 µl** and click on **OK** (run time = 50 minutes).
u. Whilst the thermal cycling. in the Post-PCR remove the Quantifluor dye and Ampure beads from the fridge and leave at room temperature for the duration of the PCR run. Ensure that the reagents are brought to room temperature in the dark.
v. Remove a single reagent cartridge (part 1 of 2 of the 300v2 sequencing kit) from the freezer in the store room and place in a water bath in the post-PCR laboratory for 1 hour with the volume of water just below the line on the reagent cartridge.

Pooling of sample libraries

a. Once the thermal cycler run has finished, remove the PCR plate and transfer back to the Post-PCR laboratory.
b. Add 1 µl of each library (columns 1-6) into column 12 of the same PCR plate using a multi-channel pipette.
c. Add 96 µl of 1x TE into well A12 and carefully mix by pipetting.
d. Transfer solution from well A12 to B12 and carefully mix by pipetting. Transfer the solution from B12 to C12 and mix by pipetting. Repeat this step until all the libraries are mixed together in well H12.
e. Transfer the mixed libraries from well H12 into a labelled 1.5ml Eppendorf tube called "Pool" (144 µl).

Library purification.

a. Add an equal volume of Ampure beads (144 µl) to the pooled library.
b. Briefly vortex and incubate at room temperature for 10 minutes.
While waiting, make up some of the following solutions:

- 70% ethanol: 700 µl of 100% ethanol and 300 µl of distilled water
- Sodium hydroxide: 490 µl of distilled water and 10 µl of 10M sodium hydroxide
- Primers: 695 µl of hybridization buffer and 3.5 µl of each primer
  FL1 (2x): primers CS1 + CS2
  FL2 (1x): primers CS1 RC + CS2 RC

c. Place the 1.5ml Eppendorf tube with Ampure beads and pooled library on a magnet for a few minutes.
d. Discard the supernatant and add 180 µl of freshly made 70% ethanol. Vortex thoroughly and place the Eppendorf tube back on the magnet. Move the tube in 10-20 degrees in either direction while on the magnet to create a nicely shaped pellet. Discard the supernatant and again add 180 µl of freshly made 70% ethanol. Vortex thoroughly and place the Eppendorf tube back on the magnet. Move the tube in 10-20 degrees in either direction while on the magnet to create a nicely shaped pellet.
e. Discard supernatant and centrifuge the Eppendorf tube for 15 seconds at 2990 RPM. Remove the remaining supernatant.

f. Open the Eppendorf tube while placed in the magnet and incubate at room temperature for 10-15 minutes.
g. Meanwhile, make up Quantifluor dye by 995 µl of 1x TE and 5 µl of Quantifluor dye. Keep in the dark.

Add 80 µl of 1x TE to the air dried Ampure beads and vortex thoroughly. Place the tube back on the magnet and transfer the supernatant to a clean 1.5ml Eppendorf tube labelled: "Myeloid or Lymphoid + #run + neat".

Library quantification

a. Label six 0.5ml tubes with blank, 50 (2x), 1x, 2x and 5x.
b. Prepare the following tubes:

- (1x) Blank: 100 µl of 1x TE
- (2x) 50 standard: 95 µl of 1x TE and 5 µl of 2 ug/ml standard
- (1x) Library 1x: 99 µl of 1x TE and 1 µl of neat library
- (1x) Library 2x: 98 µl of 1x TE and 2 µl of neat library
- (1x) Library 5x: 95 µl of 1x TE and 5 µl of neat library

c. Add 100 µl of Quantifluor dye to all tubes and vortex thoroughly.
d. Use HMDS SOP MH61 how to use the QuantiFluor Fluorometer.
e. Calculate the concentrations of each of the libraries. The values from each of the three library samples will slightly differ from each other. Use the middle value as the actual concentration of the library.

Library	ng/µl	ng/ml
1x	divide by 5	divide by 5 * 1000
2x	divide by 10	divide by 10 * 1000
5x	divide by 25	divide by 25 * 1000

f. Open the EXCEL spreadsheet "MiSeq library - Multiplex protocol - conc calculator" in location i:\molecular documents\papers & presentations\Fluidigm access array information.
g. In the "Calculation" tab add the concentration of the library in ng/µl.
Note: The "2nM library" tab shows the required volumes to create a 2nM library. The spreadsheet will always show a fixed volume of 5 µl for the library but for the actual dilution use only 4.5 µl of library.
h. Label a 1.5ml Eppendorf tube with "Myeloid or Lymphoid + #run + 2mM". Make the 2nM library using EB/Tween buffer and the EXCEL spreadsheet. After making the library vortex thoroughly.

Note: To make the EB/Tween buffer, 1mL of EB buffer and 10µl 10% Tween

i. Add 10 µl of the 2nM library and 10 µl of sodium hydroxide into a clean 1.5ml Eppendorf tube labelled "20pM library". Briefly vortex and incubate at room temperature for 5 minutes.
j. Add 980 µl of hybridization buffer to the 20pM library and vortex thoroughly. The library will now have a concentration of 20pM.
k. Add 350 µl of the 20pM library and 650 µl of hybridization buffer into a clean 1.5ml Eppendorf tube labelled "7pM library". Vortex thoroughly.

Running the MiSeq sequence platform

Pre and post washes

a. Prior to every sequence run a maintenance wash should be performed as per the manufactures instructions
b. Following all sequence runs a post sequence wash should always be performed as per the manufactures instructions.

Creating a sample plate and sheet on the MiSeq.

a. Open the fluidigm spreadsheet in location i:\MiSeq\Myeloid\Fluidigm.
a. Open the DNA quantity worksheet that was made at the beginning of the procedure and paste the HMDS numbers into the fluidigm spreadsheet. Save the file on a portable USB hard drive.
b. Open **IEM** (Illumina Experiment Manager) on the MiSeq.
c. Click on **Create New Sample Plate**.
d. In the new screen add the following information followed by **Next**:

Sample Prep Kit Selection:	Select the correct Fluidigm barcodes
Unique plate name:	MiSeq_Myeloid or Lymphoid_year_#run
Index read:	1

e. Click on the **Plate** tab and paste the sample sheet from the Fluidigm EXCEL sheet (stored on the portable USB harddrive) into the plate view of the IEM software.
f. Click on **Table** (tab) > **Apply default index layou**t > **Finish** > **Save**.
g. In the Illumina Experiment Manager, click on **Create New Sample Sheet**.
h. Click on **MiSeq** > **Targeted resequencing** > **PCR amplicon** > **Next**.
i. In the Workflow Parameters Screen fill in the following information followed by **Next**:

Reagent Cartridge Barcode:	On the reagent cartridge starting with "MS".
Sample Prep-Kit:	Select correct Fluidigm barcodes
Index reads:	1
Experiment Name:	MiSeq_Myeloid or Lymphoid_year_#run
Read type:	Paired End
Cycle Read 1+2:	151

Box 1: Tick

Box 2: Tick
Box 3: Tick
Box 4: Tick

j. Unselect the **Maximize** button in the top right corner.
k. Click **Select Plate** > Plate corresponding to your run (that was just created) > **Open** > **Select All** > **Add Selected samples**.
l. On the right side of the screen, go to the "Nextera Manifest" column and click on the first empty row to select the manifest: *Myeloid Panel Manifest NEW* or *29.03.2016 Lymphoid Panel*.
m. Click on the "Nextera Manifest" column and all rows will be selected. Right click on any of the rows and select **Fill Down**. Sample sheet status should now be valid.
n. Click on **Finish** > **Save** > **No** and shut down the Illummina Experiment Manager.

Perform a restart of the system by using **Restart** in the Windows Start menu. This will re-establish the computer links between the software and the platform.

Starting a MiSeq sequence run

a. Open **MCS (MiSeq Control Software)** on the MiSeq instrument and click on **Sequence** > **Next**. The program will take you through the required steps to start the run.
b. Take the MiSeq v2 reagent kit from the cold room and take out the flow-cell.
c. Clean the flow-cell with distilled water and lint-free tissue. The imaging glass on the flow-cell should be completely clear.
d. Place flow-cell in the MiSeq and a green tick will be visible on the screen. Click **Next**.
e. Open the MiSeq door and push the white handle up (to remove the sipper) and take out the wash bottle. Place the Incorporation buffer bottle (PR2), provided in the MiSeq v2 kit, in the correct position and push the white handle down. Another green tick should be visible on the screen and click **Next**.
f. Mix the reagent cartridge by hand by turning it upside down 10 times. Pearce the seals of wells 17-20 with a 1000ml pipette tip.
g. Transfer 600 µl of 7pM library into well 17. Transfer 680 µl of FL1 primers into well 18 and 20. Transfer 600 µl of FL2 primers into well 19.
h. Open the small white door on the MiSeq next to the PR2 bottle and take out the washing cartridge. Place the loaded reagent cartridge in the MiSeq and close both the white and black door. After the green tick click **Next**.
i. The machine will now load the sample sheet and this may take a few minutes. Once the information is loaded on the screen click **Next**.
j. On the new screen, green ticks will start to appear in front of the different parameters. The last green tick will take a few minutes to appear. Meanwhile, make notes of the reagent ID, PR2 bottle ID and Flow cell ID on your worksheet. Click **Start run** when all parameters are green.

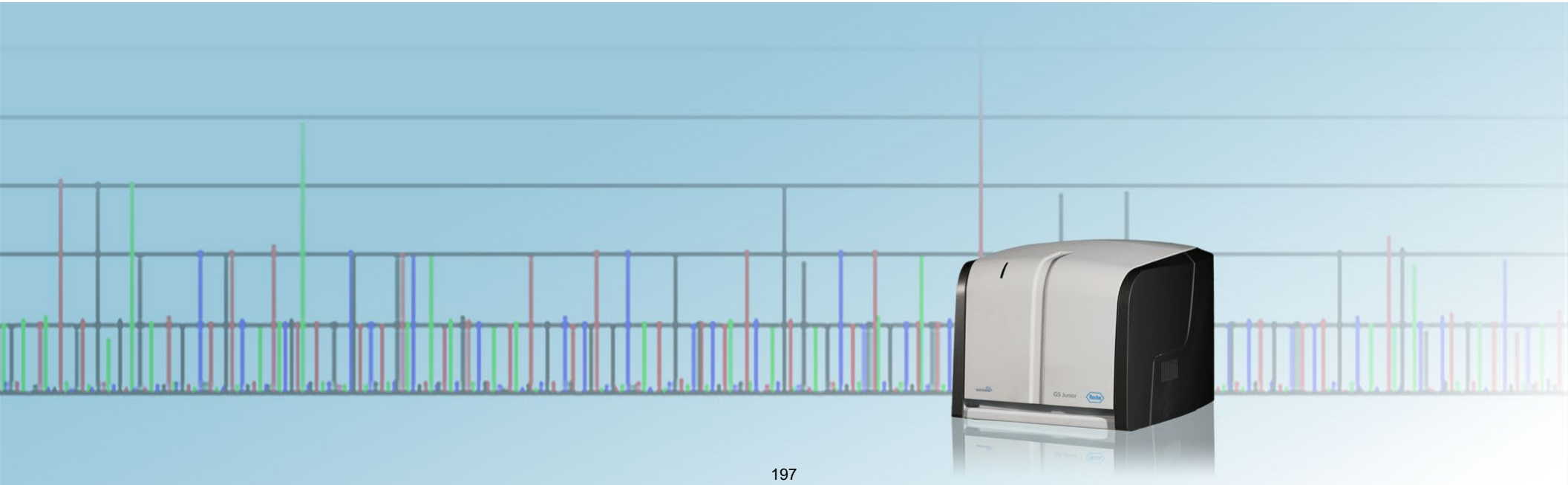
Appendix 8.3

GS Junior Emulsion PCR and Sequencing protocol

emPCR Amplification Method Manual – Lib-A

GS Junior Titanium Series

March 2012



1 WorkFlow

The emulsion-based clonal amplification (emPCR amplification) of a DNA library sample involves 7 major steps, described in Figure 1. The method described in this manual is for the GS Junior Titanium Series of PCR products with bi-directional fusion primers (Lib-A).

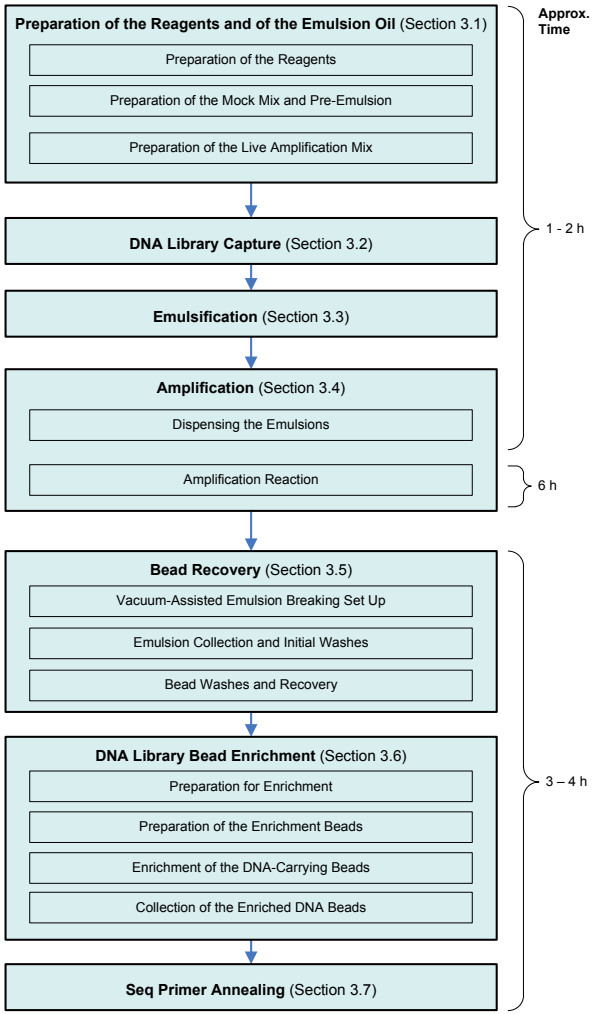


Figure 1: Workflow of the emPCR Amplification method for the GS Junior System

2 Before You Begin

- Room temperature is +15 to +25°C.
- When processing multiple emulsions, it is cost effective to use the GS FLX Titanium MV emPCR Lib-A Kit, as it contains eight emulsions. Refer to the *emPCR Method Manual Lib-A, MV - GS Junior Titanium Series* and to the TCB 010-002.01 to use this kit. In this case, the option to perform an enrichment titration is available and the emulsification takes place in a TissueLyser, which requires the GS FLX Titanium emPCR Shaker Adapter MV.
- To access the US power adaptor for the IKA Turrax, lift the part insert from the Turrax box and look at the reverse side. The adaptor is located in the reverse side of the part insert.

The GS Junior Titanium emPCR Kit (Lib-A) is intended for a single emulsion, either of a single sample processed with the Amplicon library protocol or of multiple samples processed with the Amplicon library protocol using MIDs. The present manual describes the method to process one sample.

2.1 Determining the Amount of Library to Use in emPCR Amplification

For most Amplicon library samples, an input of 2 molecules of library DNA per Capture Bead will yield a bead enrichment between 5% and 20%, and will generate satisfactory sequencing results.

When the % bead enrichment is above 20%, reduce the library input molecules 3-fold.

When the % bead enrichment is below 5%, increase the library input molecules 3-fold.

If the library yield can be predicted from past experience, one can use a pre-determined number of DNA molecules per bead.

3 PROCEDURE

3.1 Preparation of the Reagents and of the Emulsion Oil

3.1.1 Preparation of the Reagents

1. Open the emPCR Reagents box and thaw the kit components at room temperature, except the Enzyme Mix and the PPiase tubes, which should be kept at -15°C to -25°C. Once thawed, vortex for 5 seconds.
2. Vortex and heat the tube of Additive at 55ºC for 5 minutes to aid dissolving. If a precipitate persists, centrifuge the tube and use the supernatant.
3. Spin all the kit components (including enzymes) in a bench top mini centrifuge for 10 seconds.
4. Return the enzymes to -15 to -25°C. Leave the other reagents at room temperature.

3.1.2 Preparation of the Mock Mix and Pre-Emulsion

1. Vortex vigorously the tube of emulsion oil for 10 seconds at maximum vortex speed, and pour the entire content (4 ml) into the Turrax stirring tube.
2. Prepare 1x Mock Mix by adding **430 µl** of Mock Mix to **1.72 ml** of Molecular Biology Grade Water. Vortex to mix.
3. Add **2.0 ml** of 1x Mock Mix to the Turrax stirring tube containing the emulsion oil.
4. Set the Ultra Turrax Tube Drive (UTTD) to **4000 rpm for 5 minutes**.

Lock the UTTD settings: Because the setting knob is sensitive, it is critical that the UTTD settings be locked before use. Refer to the manufacturer's instructions to set and lock the settings of the UTTD.

5. Place the stirring tube in the UTTD and start the UTTD to mix the emulsion.
6. When finished, remove the stirring tube from the UTTD.

3.1.3 Preparation of the Live Amp Mixes A and B

1. Prepare the Live Amp Mix A and the Live Amp Mix B in two separate tubes, according to Table 1. Add the reagents in the order they are listed in the table.
2. Vortex the Live Amp Mixes A and B for 5 seconds, and store them on ice.

A: Live Amp Mix A

Reagent	Volume (µl)
Mol. Bio. Grade Water	205
Additive	260
Amp Mix	135
Amp Primer A	40
Enzyme Mix	35
PPiase	1
Total	**676**

B: Live Amp Mix B

Reagent	Volume (µl)
Mol. Bio. Grade Water	205
Additive	260
Amp Mix	135
Amp Primer B	40
Enzyme Mix	35
PPiase	1
Total	**676**

Table 1: Preparation of the Live Amp Mixes A and B

3.2 DNA Library Capture

Capture Beads:

- Do NOT mix Capture Beads A and B.
- Only mix Capture Beads A with Live Amp Mix A and Capture Beads B with Live Amp Mix B.

1. Prepare 1x Wash Buffer by mixing **0.5 ml** of Wash Buffer with **4.5 ml** of Molecular Biology Grade Water.
2. Add **1 ml** of 1x Wash Buffer to both tubes of Capture Beads and vortex.
3. Pellet the Capture Beads in a bench top minifuge, by spinning for 10 seconds, rotating the tube 180°, and spinning again for 10 seconds (spin-rotate-spin).
4. Carefully remove and discard the supernatants without disturbing the bead pellets.
5. Wash the Capture Beads once more with **1 ml** of 1x Wash Buffer. Vortex to resuspend the beads, spin-rotate-spin, and discard the supernatant.
6. Thaw an aliquot of the Amplicon DNA library to be amplified.
7. Calculate the volume of DNA library needed by using the following equation:

$$\mu l \text{ of DNA library per tube} = \frac{\text{Desired molecules per bead x 5 million beads}}{\text{Library concentration (in molecules/}\mu l)}$$

For example:

$$5\ \mu l \text{ of library} = \frac{\text{2 molecules per bead x 5 million beads}}{\text{2 million molecules/}\mu l}$$

If necessary, prepare a dilution of the library such that the volume to be added is between 5 µl and 30 µl.

8. Add the calculated volume of the Amplicon DNA library to the tube of washed Capture Beads A.
9. Add the same calculated volume of Amplicon DNA library to the tube of Capture Beads B.
10. Vortex the tubes for 5 seconds to mix.

3.3 Emulsification

1. Add **600 µl** of Live Amp Mix B to the tube of captured library B. Vortex, and transfer the entire content into the Turrax stirring tube (from Section 3.1.2).
2. Set the UTTD **to 2000 rpm for 5 minutes**.

Lock the UTTD settings: Because the setting knob is sensitive, it is critical that the UTTD settings be locked before use. Refer to the manufacturer's instructions to set and lock the settings of the UTTD.

3. Place the stirring tube in the UTTD and start the UTTD to mix the emulsion.

Time constraints for the emulsification: Immediately start the UTTD once the Live Amp Mix and the Capture Beads are added to the Turrax stirring tube. Do not allow the Live Amp Mix and the Capture Beads to sit in the Turrax tube for more than five minutes before stirring the tube.

4. Add **600 µl** of Live Amp Mix A to the tube of captured library A. Vortex, and transfer the entire content into the same Turrax stirring tube as in Step 1.
5. Place the stirring tube in the UTTD and start the UTTD to make the final emulsion.
6. When finished, remove the stirring tube from the UTTD.

3.4 Amplification

3.4.1 Dispensing the Emulsions

1. Using a Combitip, aliquot 100 µl of emulsion into nine 8-strip cap tubes or one 96-well plate (~70 wells), by slowly aspirating, taking care not to draw air.
2. Cap the wells and make sure that all the wells are properly sealed.
3. Clean up the area of any spilled reaction mix.

3.4.2 Amplification Reaction

1. Place the tube strip/plate in a thermocycler and start the amplification program with the heated lid turned on. The program takes ~6 h to complete.

1x	4 minutes	at 94°C
50x	30 seconds	at 94°C,
	4.5 minutes	at 58°C,
	30 seconds	at 68°C
End		at 10°C on hold

- **Do not freeze the DNA beads:** You can leave the amplification reactions at 10°C for up to 16 hours before further processing the samples.
- **Emulsion breakage:** Check all wells for emulsion breakage (*i.e* a clear middle layer). If the emulsion in any well appears broken, discard the entire well and do not recover the beads from it. See Section 4.1 for pictures of intact and broken emulsions.

3.5 Bead Recovery

3.5.1 Vacuum-Assisted Emulsion Breaking Set Up

1. Bring the GS Junior Titanium emPCR Oil and Breaking Kit to the externally ventilated hood.

Externally ventilated hood: The procedures from this point until Section 3.5.2, Step 4 are performed in the externally ventilated hood.

2. Attach a 50 ml conical tube to the lid from the GS Junior Titanium Oil and Breaking Kit. Save the cap of the 50 ml tube.
3. Insert the blue connector into the top opening of the transpette.
4. Connect the other end of the tubing to a vacuum source (with liquid trap to capture the isopropanol waste). See Figure 2 for a representation of the set-up.

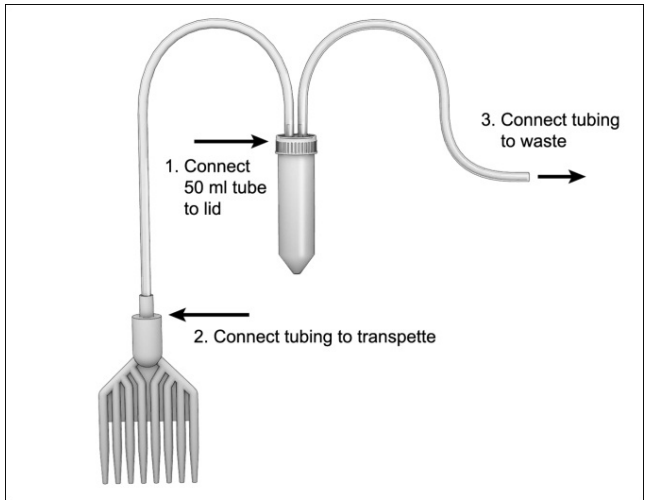


Figure 2: Schematic view of an assembled set up for vacuum-assisted emulsion breaking and bead recovery

3.5.2 Emulsion Collection and Initial Washes

1. Turn on the vacuum and aspirate the emulsions (A and B) from all the wells and collect them in the 50 ml tube, using a slow circular motion of the transpette tips at the bottom of the wells.
    - After aspirating all the emulsions, turn the transpette upside-down to help drain as much material as possible into the collection tube.
2. Rinse the wells **twice** with **100 µl** of isopropanol per well (using a reservoir and a multichannel pipet, if available). Aspirate the rinse and turn the transpette upside-down to retrieve as much material as possible.
3. **SLOWLY** aspirate an additional (approximate) **5 ml** of isopropanol to collect any beads that may remain in the tubing.
4. Turn off the vacuum, and remove and cap the 50 ml tube containing the amplified DNA beads. Take the 50 ml tube out of the hood.

3.5.3 Bead Washes and Recovery

Tube requirement: Use the 1.7 ml siliconized tubes provided in the GS Junior Titanium emPCR Oil and Breaking Kit.

1. Vortex the 50 ml tube of collected emulsions.
2. Add isopropanol to a final volume of **35 ml** and vortex to resuspend the pellet.
3. Pellet the beads in a centrifuge at **930 x *g* for 5 min** (2813 RPM for the Eppendorf 5430 centrifuge, rotor F-35-6-30) and carefully pour out the supernatant.
4. Add **10 ml** of Enhancing Buffer and thoroughly vortex to resuspend the pellet (it is important to properly rinse the beads. Use glass rod or a spatula to break the aggregates, if necessary).
5. Add isopropanol to a final volume of 40 ml and vortex well.
6. Pellet the beads in a centrifuge at **930 x *g* for 5 min** and carefully remove the supernatant.
7. Add isopropanol to a final volume of **35 ml** of and vortex well.
8. Pellet the beads in a centrifuge at **930 x *g* for 5 min** and carefully remove the supernatant.
9. Add ethanol to a final volume of **35 ml** of and vortex well.
10. Pellet the beads in a centrifuge at **930 x *g* for 5 min** and carefully remove the supernatant.
11. Add Enhancing Buffer to a final volume of **35 ml** of and vortex well.
12. Pellet the beads in a centrifuge at **930 x *g* for 5 min** and carefully remove the supernatant, leaving approximately 2 ml of Enhancing Buffer.
13. Transfer the DNA bead suspension using a 1000 µl pipette into a provided 1.7 ml micro-centrifuge tube.
14. Spin-rotate-spin and discard the supernatant.
15. Rinse the 50 ml tube with **1 ml** of Enhancing Buffer, and add this rinse to the 1.7 ml tube. Spin-rotate-spin and discard the supernatant.
16. Thoroughly rinse the bead pellet **twice** with **1 ml** of Enhancing Buffer. Spin-rotate-spin and discard the supernatant.

3.6 DNA Library Bead Enrichment

3.6.1 Preparation for Enrichment

Hazardous Chemical – Sodium Hydroxide Solution: Sodium hydroxide (present in the Melt Solution) is a highly corrosive chemical that may cause burns if it contacts eyes or skin. Read the Material Safety Data Sheet for handling precautions.

1. Turn on the heating dry-block and set it to 65°C.
2. Prepare the Melt Solution by mixing **125 µl** of NaOH (10 N) in **9.875 ml** of Molecular Biology Grade Water.
3. Add **1 ml** of Melt Solution to the 1.7 ml tube of beads and vortex. Incubate for **2 minutes at room temperature**. Spin-rotate-spin and discard the supernatant.
4. Repeat Step 3 once.
5. Add **1 ml** of Annealing Buffer to the 1.7 ml tube of beads and vortex. Spin-rotate-spin and discard the supernatant.
6. Repeat Step 5 **twice**.
7. Add **45 µl** of Annealing Buffer, **15 µl** of Enrich Primer A and **15 µl** of Enrich Primer B to the 1.7 ml tube of beads, and vortex.
8. Place the tube in a heat block at **65°C for 5 minutes**, and then promptly cool on **ice for 2 minutes**.
9. Add **1 ml** of Enhancing Buffer to the 1.7 ml tube of beads and vortex. Spin-rotate-spin and discard the supernatant.
10. Repeat Step 9 **two more times**.
11. Add **1 ml** of Enhancing Buffer to the 1.7 ml tube of beads and vortex.
12. Set the tube aside at room temperature until Section 3.6.3.

3.6.2 Preparation of the Enrichment Beads

1. Vortex the tube of brown Enrichment Beads for **1 minute** to resuspend its contents completely.
2. Place the tube in a Magnetic Particle Concentrator (MPC) and wait **~3 minutes** to pellet the Enrichment Beads.
3. Discard the supernatant, taking care not to draw off any Enrichment Beads.
4. Add **500 µl** of Enhancing Buffer and vortex.
5. Pellet the Enrichment Beads using an MPC.
6. Discard the supernatant, taking care not to draw off any Enrichment Beads.
7. Repeat Steps 4 to 6 once.
8. After discarding the supernatant, remove the tube from the MPC.
9. Add **80 µl** of Enhancing Buffer and vortex.

3.6.3 Enrichment of the DNA-Carrying Beads

1. Add **80 µl** of washed Enrichment Beads to the 1.7 ml tube of beads (from Section 3.6.1) and vortex to mix completely.
2. Rotate the tube on the LabQuake, at room temperature for **5 minutes**.
3. Place the tube in the MPC, and wait **3-5 minutes** to pellet the Enrichment Beads.
4. Invert the MPC several times and wait for the beads to pellet.
5. Carefully discard the supernatant using a 1000 µl pipette, taking care not to draw off any brown Enrichment Beads.
6. Wash the beads with Enhancing Buffer until there are no visible white beads remaining in the supernatant, as follows:
    a. Add **1 ml** of Enhancing Buffer to the tube.
    b. Remove the tube from the MPC and vortex well.
    c. Place the tube back into the MPC to pellet the beads on the wall of the tube with the magnet. Invert the MPC and wait for the beads to pellet.
    d. Carefully discard the supernatant with a 1000 µl pipette, taking care not to draw off any Enrichment Beads.
    e. Repeat **6 to 10 times** until white DNA beads are no longer being aspirated.
    - Optionally, collect the supernatant and spin to monitor when washes are complete.

3.6.4 Collection of the Enriched DNA Beads

1. Remove the tube of enriched beads (enrichment tube) from the MPC and resuspend the bead pellet in **700 µl** of Melt Solution.
2. Vortex for 5 seconds, and place the enrichment tube in the MPC until the Enrichment Beads have pelleted.
3. Transfer the **supernatant** containing the enriched DNA beads to a new 1.7 ml microcentrifuge tube.
4. Add once again **700 µl** of Melt Solution to the enrichment tube.
5. Vortex for 5 seconds, and place the enrichment tube in the MPC until the Enrichment Beads have pelleted.
6. Transfer the **supernatant** containing enriched DNA beads into the same 1.7 ml tube, from Step 3.
7. Discard the enrichment tube.
8. Spin-rotate-spin the 1.7 ml tube and discard the supernatant.
9. Add **1 ml** of Annealing Buffer and vortex for 5 seconds.
10. Spin-rotate-spin and discard the supernatant.
11. Repeat Steps 9 and 10 **two times**.
12. Add **100 µl** of Annealing Buffer and vortex to resuspend the beads.

3.7 Seq Primer Annealing

1. Add **15 µl** of Seq Primer A and **15 µl** of Seq Primer B, and vortex.
2. Place the 1.7 ml tube in a heat block at **65°C for 5 min**, and then promptly cool on **ice for 2 min**.
3. Add **1 ml** of Annealing Buffer and vortex for 5 seconds. Spin-rotate-spin, and discard the supernatant.
4. Repeat Step3 **two times**.
5. Add **1 ml** of Annealing Buffer to the bead pellet and vortex.
6. Spin-rotate-spin to pellet the beads.
7. A GS Junior sequencing Run requires an input of 500 000 enriched beads. Evaluate the amount of enriched beads using the GS Junior Bead Counter, as described in Section 4.1.
8. Store the beads at +2 to +8°C and sequence them within two weeks.

4 APPENDIX

4.1 The GS Junior Bead Counter

The recommended input bead number for a GS Junior sequencing Run is 500,000 enriched beads. To evaluate the amount of enriched beads, use the GS Junior Bead Counter v2, which is an accessory to the GS Junior Instrument.

1. Insert the 1.7 ml tube from Section 3.7, Step 7 in the GS Junior Bead Counter v2, such that the bottom of the tube touches the floor of the hole.
2. Hold the Bead Counter at eye level with the side reading Single Prep facing you. Look in the window to evaluate the amount of enriched beads. The top of the bead pellet should be within the window.
    a. If the top of the bead pellet is within the window and above the bottom edge of the window (500,000 beads) (Figure 3**A**), remove the excess beads and store them into a clean 1.7 ml microcentrifuge tube for later use, as they are valuable. Load the remaining 500,000 beads on the GS Junior Instrument.
    b. If the top of the bead pellet is below the bottom edge of the Bead Counter window (*i.e.* not visible), the preparation contains less than 500,000 beads. However, the beads are valuable and the amount could be sufficient for the experimental design.
    c. If the top of the bead pellet is above the top edge of the Bead Counter window, the preparation has failed. Discard the beads and repeat the emulsion process with 3 times less library sample.

- The lower edge of the window defines 500,000 beads while the upper edge defines 2 million beads.
- A front view of the bead counter may be helpful to visualize the bead pellet. Use the indentations on the side reading Single Prep to estimate the bead number (Figure 3**B**).

A B

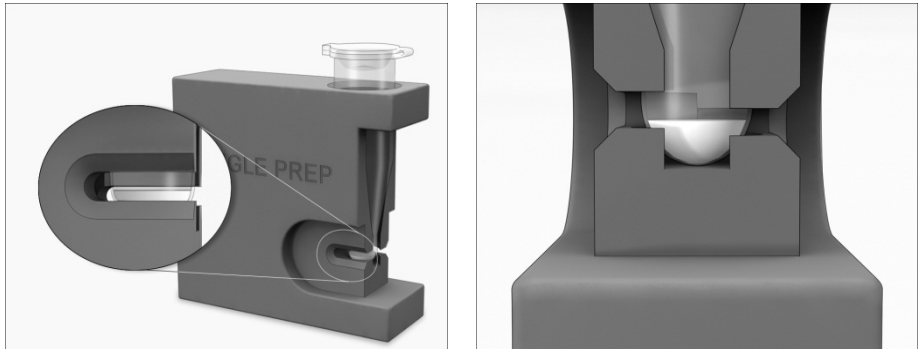
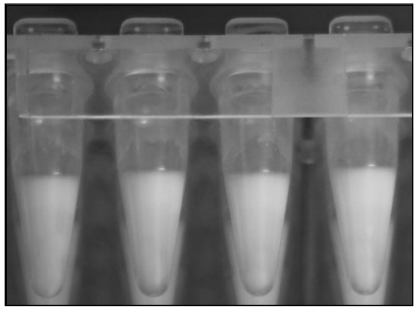


Figure 3: (A) The Single Prep side view of the GS Junior Bead Counter v2, (B) a front view of the GS Junior Bead Counter v2, with the Single Prep side on the left and the Multiple Prep side on the right

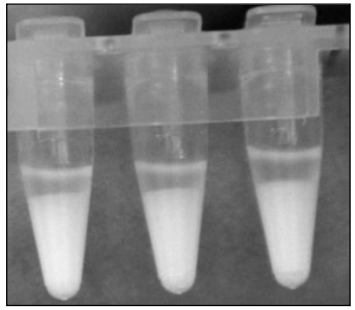
Customers can also use the original, unmarked GS Junior Bead Counter. It is equivalent to using the GS Junior Bead Counter v2 SINGLE PREP side.

4.2 Pictures of Emulsions

A



B



C

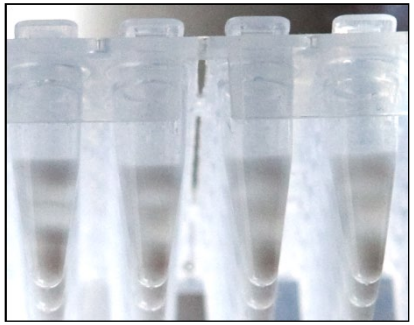


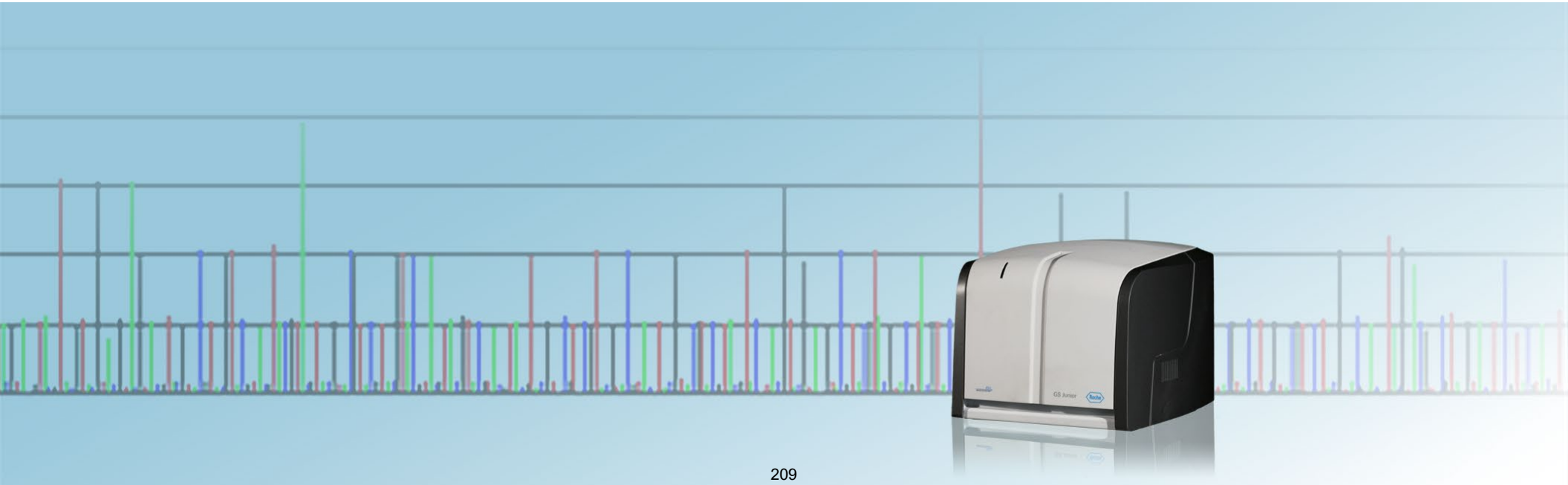
Figure 4: A and B show intact emulsions appearing as an homogenous suspension (A), at times layered by a clearer phase (B). The multiple bands in C are a sign of a broken emulsion.

Table of Contents

Sequencing Method Manual

GS Junior Titanium Series

March 2012



1 WorkFlow

The overall workflow of a GS Junior sequencing Run is shown in Figure 1. The tasks performed on the GS Junior Instrument are highlighted in green.

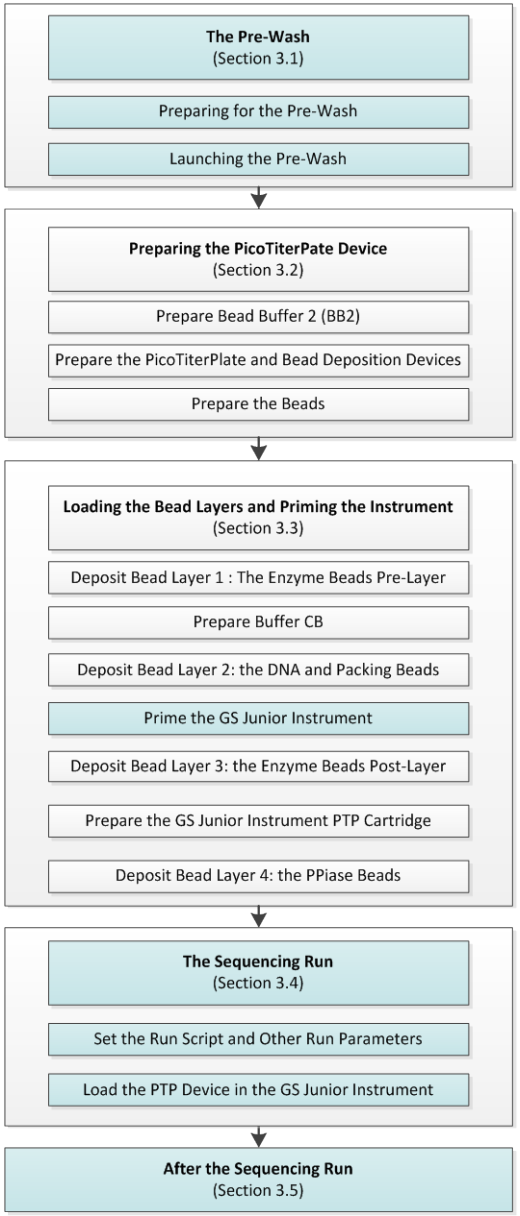


Figure 1: Workflow of a sequencing Run performed on the GS Junior Instrument

2 BEFORE YOU BEGIN

- Room temperature is +15 to +25°C.
- All 1.7 ml tubes must be siliconized.
- The BDD centrifuge speed noted in RPM is for the Eppendorf 5430 centrifuge.

2.1 What You Should Have Before Starting

2.1.1 Sample

The sample library being sequenced must have been prepared using GS FLX Titanium or GS Junior Titanium series methods and kits. The library will then have been amplified using one of the GS FLX Titanium or GS Junior Titanium series emPCR Kits.

2.1.2 Required GS Junior System Equipment and Reagents

Visit our customer-restricted web site at www.454.com/my454 for the lists of material required but not provided.

2.1.3 GS Junior Titanium Sequencing Kit and PicoTiterPlate Kit

The GS Junior Titanium Sequencing Kit is used in combination with the matching GS Junior Titanium PicoTiterPlate Kit. These kits provide reagents and components necessary for a single sequencing Run. Visit www.454.com/my454 for a complete list of kits and reagents.

The GS Junior Instrument holds a one-region PTP device.

Region per PTP Device	Bases per Region (Mbp)	Reads per Region
1	35	100,000

2.1.4 Reagents and Buffer

The GS Junior Titanium Sequencing Kit comes in three parts:

- The Sequencing Kit Reagents and Enzymes (stored at -15 to -25°C)
- The Sequencing Kit Buffers (stored at +15 to +25°C)
- The Sequencing Kit Packing Beads and Supplement CB (stored at +2 to +8°C)

Before beginning a sequencing Run, do the following:

1. Thaw the components of the Reagents and Enzymes box:
2. Peel open the seal of the Reagents and Enzymes container.
3. Take out the 2.0 ml tubes and the 50 ml conical tube. Place the 2.0 ml tubes on ice to thaw.
4. Lift the tray and fill the container that holds the 10-tubes Reagents cassette with room temperature tap water. Do not submerge the cassette, keep it upright and protected from direct sunlight.
5. Put the tray back in the container and place the 50 ml tube in the tray for thawing.
6. Periodically, invert the cassette and the 50 ml tube three times to mix their contents.
7. When the contents have thawed, transfer the Reagents cassette and the 50 ml tube to +2 to +8°C until use.
8. Retrieve the Packing Beads and Supplement CB box from the refrigerator and keep on ice until needed.
9. From the Buffers box, retrieve the bottles of Buffer CB and Pre-wash Buffer, as well as the Buffer Sipper Tube, the Reagent Sipper Tubes, and the Pre-wash cassette. Keep at room temperature.

3 PROCEDURE

Prior to beginning a sequencing Run, the instrument should be empty of used Sipper Tubes, Reagent cassette and Reagent bottle. If not, please see Section 3.5. Only an empty waste bottle should be in place on the right.

The GS Junior Sequencing procedure is a four step process that involves:

- Washing the instrument's fluidics with Pre-wash Buffer
- Preparing and depositing the beads in the Bead Deposition Device (BDD)
- Priming the instrument with reagents and buffers
- Performing a sequencing Run

3.1 The Pre-Wash

Under normal conditions of continuous operation, the instrument is kept on standby after a Run has completed. Therefore, the PTP device from the previous Run should still be in place in the instrument's PTP cartridge. If there is no PTP device in the cartridge, you must install a used but intact one (and a cartridge seal) before proceeding with the Pre-wash.

3.1.1 Preparing for the Pre-Wash

1. Close the previous run by clicking OK on the sequencing Run complete window.
2. If required, log in as follows:
   - Click on Operator in the Status area of the GS Junior Sequencer application window (Figure 2).

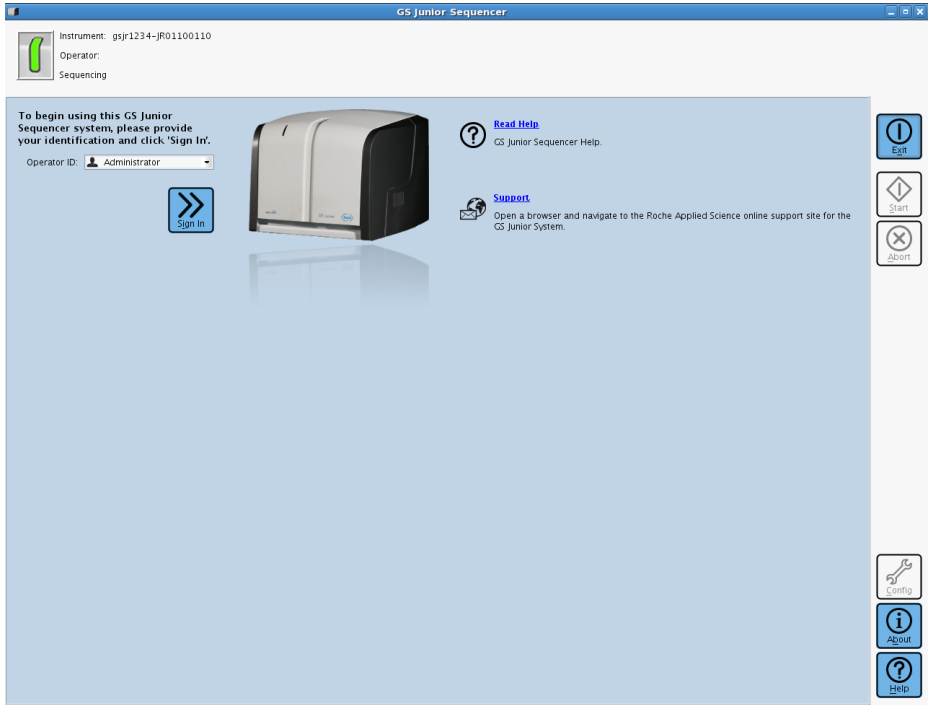


Figure 2: The sign in window of the GS Junior Sequencer application

- Sign in by selecting your user name and click Sign In. The main window of the GS Junior Sequencer application will open (Figure 3).

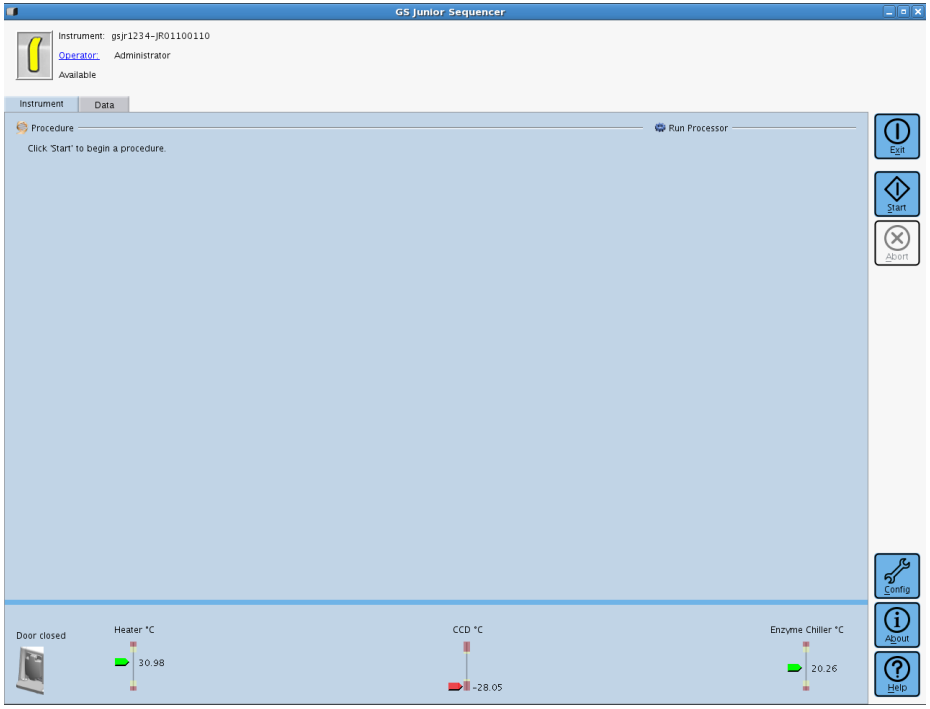


Figure 3: The GS Junior Sequencer application window after an Operator signs in

3. Begin by clicking the Start button on the right of the GS Junior Sequencer application window (Figure 3). This will open the Instrument Procedure wizard first window: Choose a procedure (Figure 4).

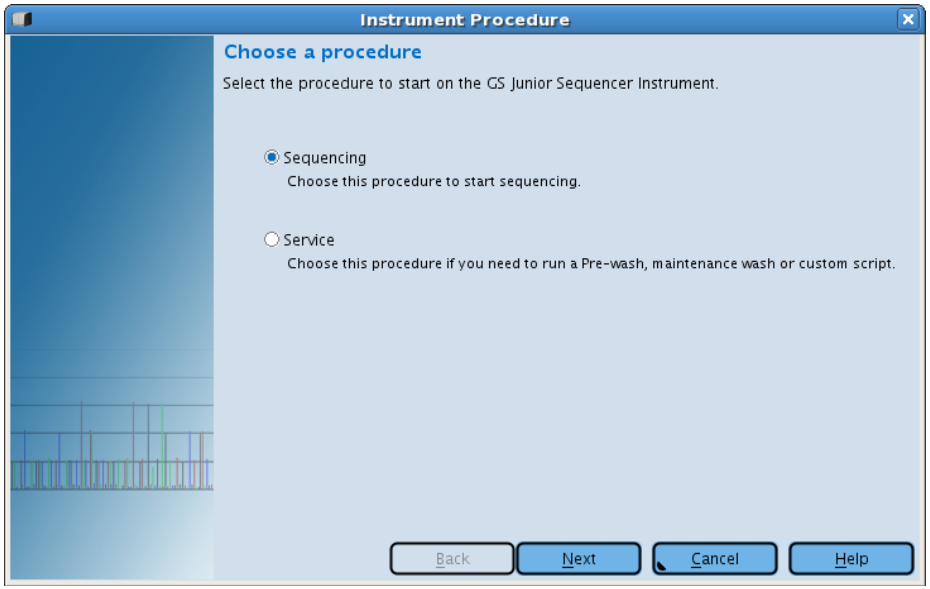


Figure 4: The Instrument Procedure wizard first window: Choose a procedure, Sequencing option selected

4. If the Sequencing option is not selected, select it, and click Next. This will open the Instrument Procedure second window Start Sequencing (Figure 5).

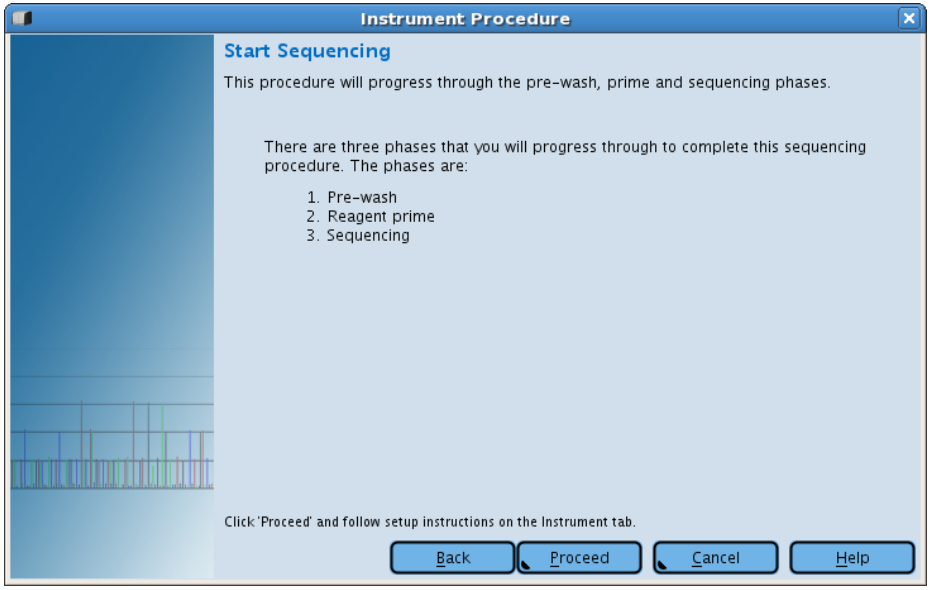


Figure 5: The Instrument Procedure second window: Start Sequencing

5. Click Proceed. The GS Junior Sequencer application window will open, in the Instrument tab. An abbreviated version of the instructions that are detailed below will appear on the screen (Figure 6).

Figure 6: The GS Junior Sequencer application window before launching the Pre-wash

6. Lift the instrument cover and raise the sipper manifold completely.

7. Affix the large filter on the Buffer Sipper Tube (Figure 7**A**), and the small filters on the ten Reagents Sipper Tubes (Figure 7**B**).

A B

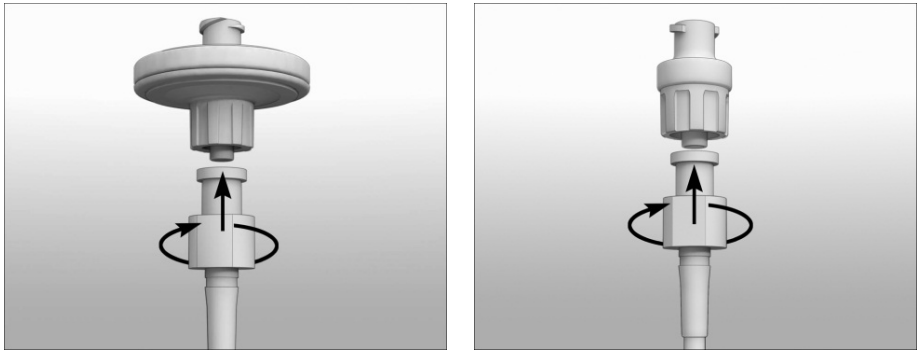


Figure 7: Assembly of (A) the Buffer Sipper Tube onto its filter and (B) a Reagent Sipper Tube onto its filter

8. Holding the assembled tubes by the top the sipper tubes like in Figure 8**A**, screw them in the manifold, finger tight.

A B

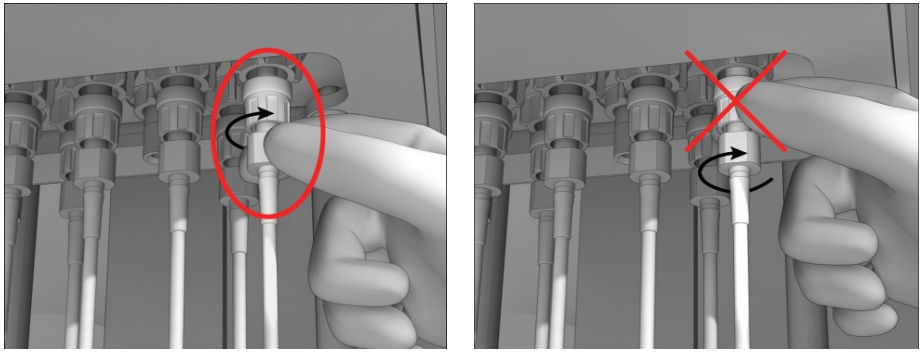


Figure 8: (A) Correct and (B) incorrect screwing in of the Reagent Sipper Tube

9. Change gloves to avoid contaminating other components.
10. From the GS Junior Titanium Sequencing Kit Buffers box, open the bag containing the Pre-wash cassette (Figure 9).

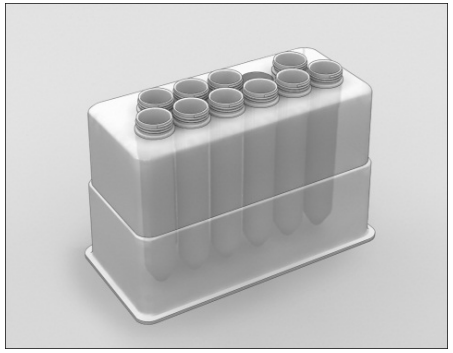


Figure 9: The Pre-wash cassette

11. Rinse the ten tubes in the Pre-wash cassette **twice** with nanopure water.
12. Fill the ten tubes with Pre-wash Buffer till to ~ 1 cm from the top of the tubes (Figure 10).

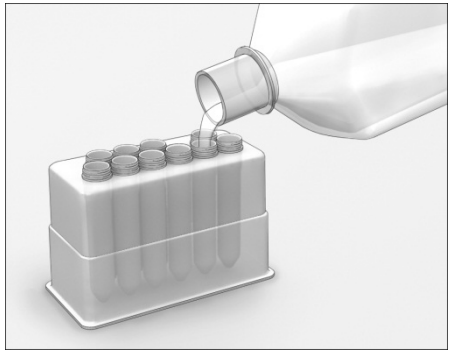


Figure 10: Filling the Pre-wash cassette

13. Mount the Pre-wash cassette in the instrument, with the vacant position in the back, as shown in Figure 11**A**. There should be no sipper tube in the vacant position.
14. Place the bottle of Pre-wash Buffer on the left.
15. Lower the sipper manifold and close the instrument cover (Figure 11**B**).

A B

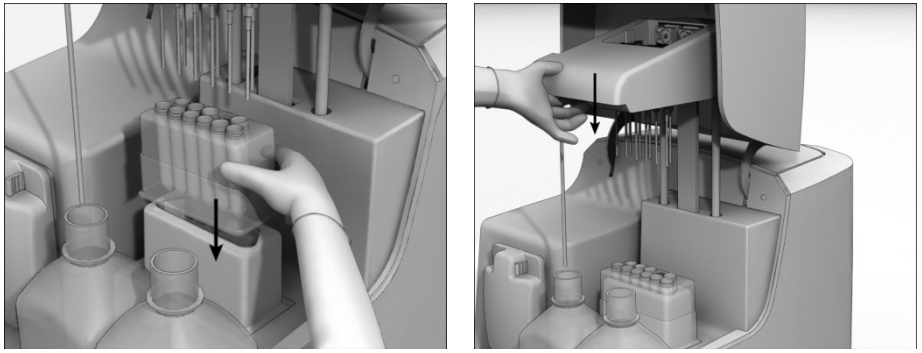


Figure 11: The Pre-wash cassette: (A) mounting the Pre-wash cassette and (B) lowering the sipper manifold

If necessary, replace any bent Sipper Tubes, as described in the *GS Junior Instrument Owner's Manual*.

16. If there isn't a PTP device in the instrument, install a used but intact PTP device (with a used but intact cartridge seal) following the instructions in Section 3.4.2.

3.1.2 Launching the Pre-Wash

1. Return to the Attendant PC and click Proceed in the GS Junior Sequencer application (Figure 6).
2. As the Pre-wash is progressing, the steps being performed will appear on the screen. A countdown will also show. The Pre-wash lasts approximately 30 minutes (Figure 12).

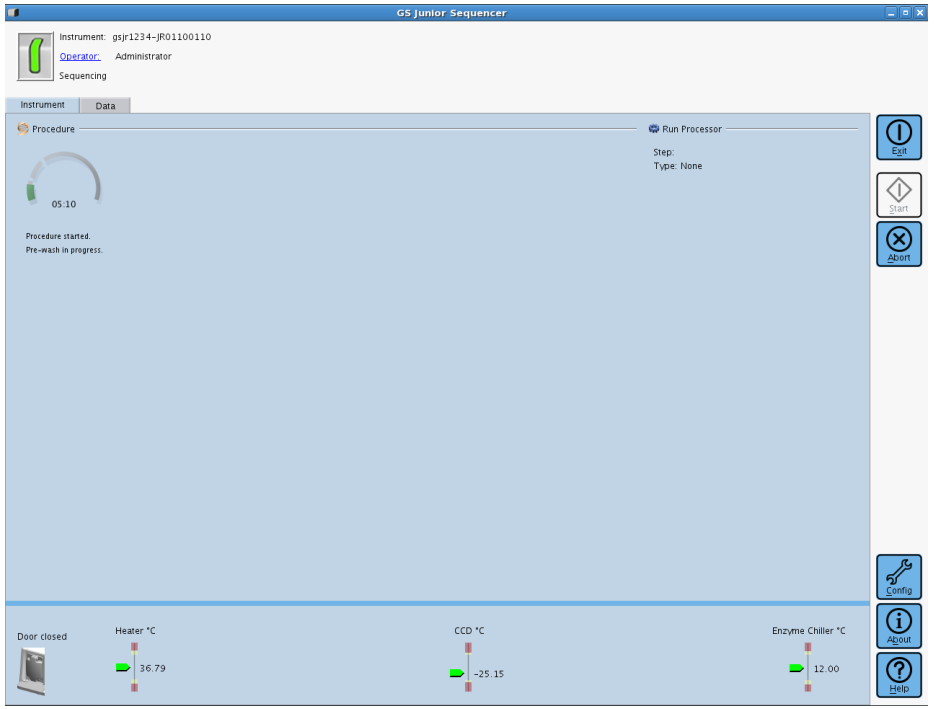


Figure 12: The GS Junior Sequencer application window while the Pre-wash is in progress

3. Once the Pre-wash is complete, the GS Junior Sequencer application window will read: Prepare instrument for Reagent prime (Figure 13). Before priming the instrument, a number of steps detailed in Section 3.2 to Section 3.3.3 need to be performed.

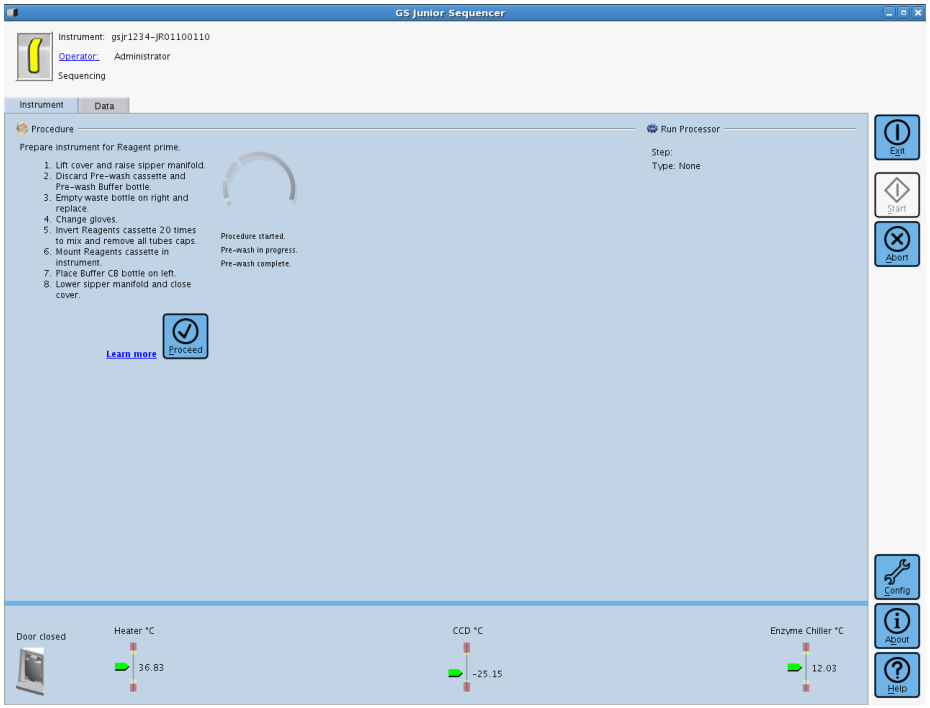


Figure 13: The GS Junior Sequencer application window before launching the instrument prime

Monitor the instrument until the Status LED on the instrument is blinking green. If the instrument encounters any problems during the initiation of the Pre-wash, a message describing the issue will appear over the Status area of the GS Junior Sequencer window.

PicoTiterPlate device preparation: Start the preparation of the components of a sequencing Run (Section 3.2) as soon as the Pre-wash is safely ongoing. The Pre-wash will proceed to completion without any further user intervention (approximately 30 minutes).

There is an *Abort* button available in the GS Junior Sequencer main window, which can be clicked if a problem occurs. Aborting a Pre-wash, Prime or sequencing Run will void the steps performed prior to aborting.

3.2 Preparing the PicoTiterPlate Device

3.2.1 Prepare Bead Buffer 2 (BB2)

1. Add **6.6 ml** of Supplement CB to the Buffer CB bottle. Thoroughly mix the contents by inverting the bottle 10 times.
2. With a 50 ml serological pipette, transfer **40 ml** of the Buffer CB into a clean 50 ml conical tube and place on ice.
3. Spin the Apyrase tube in a microcentrifuge for 5 seconds at **9,300 RCF** (10,000 RPM). Add **6.5 µl** of Apyrase to the 50 ml tube. Label the tube BB2, for Bead Buffer 2. Gently invert the tube 10 times to mix the contents and place on ice.

3.2.2 Prepare the PicoTiterPlate and Bead Deposition Devices

1. Retrieve the PTP device tray from the GS Junior Titanium PicoTiterPlate Kit. Peel open the seal on top of the tray.
2. Write down the PTP device ID to enter the 6-digit number in the Instrument Procedure third window, Enter IDs and barcodes (Figure 21).
3. Remove the bead loading gasket and cartridge seal from the tray. Wash them by a gentle shaking for 30 seconds in a Sparkleen solution. Rinse thoroughly with nanopure water and let air dry on a paper towel.
4. Wash the Bead Deposition Device (BDD) using a soft bristle brush and a Sparkleen solution. Rinse thoroughly with nanopure water and let the device air dry on a paper towel.
5. Assemble the BDD with the washed gasket and a PTP device, as follows:
6. Wearing gloves, orient the tray in front of you. Hold the tray with one hand and with the index of the other pry up the PTP device without touching either flat surface. Once the PTP device out, always hold it by its edges.
7. Place the PTP device onto the BDD base (Figure 14A), aligning the notched corner of the PTP device and the BDD base.
8. Secure the washed and dried bead loading gasket to the BDD base by laying it on top of the PTP device (Figure 14B). Align the notched corner of the bead loading gasket and the BDD base, as shown in the Figure.
9. Carefully place the BDD top over the assembled BDD base/PTP device/gasket (Figure 14C). Align the dowels on the BDD base so they slide into the holes in the BDD top and the BDD top is sitting flat across the PTP device.
10. Press down on the top of the BDD, and rotate the two latches from the BDD base into the grooves in the BDD top to firmly secure the assembly (Figure 14D). When you hear a 'click', the latches should be firmly seated in the grooves, providing the correct amount of pressure to maintain a liquid-tight seal.
11. There are two holes on the top of the BDD; the larger is a loading port and the smaller, an air vent.

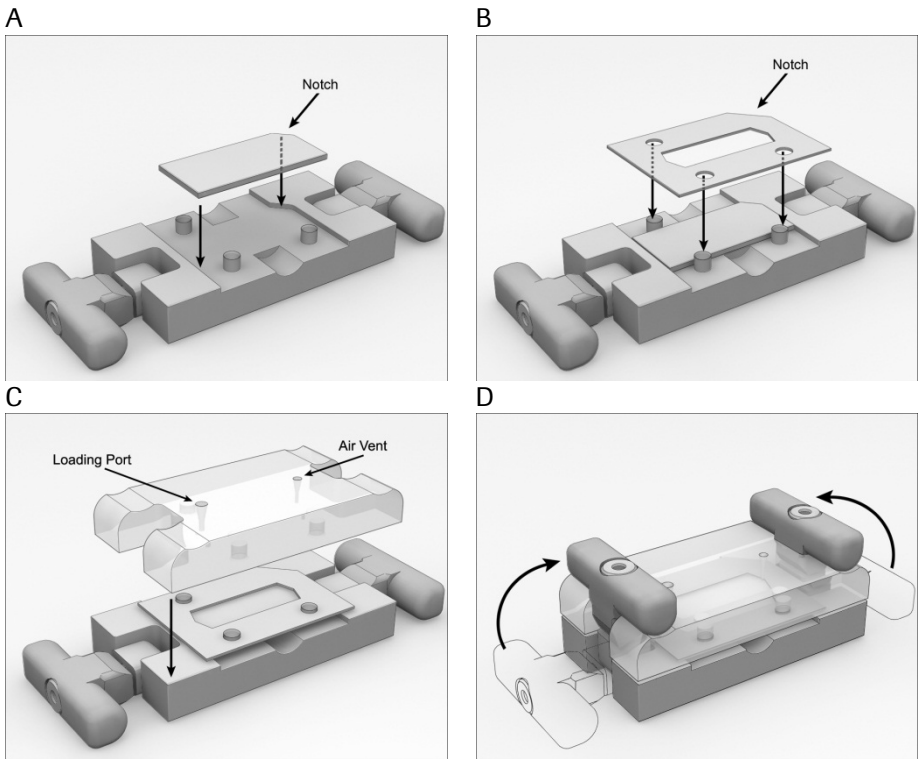


Figure 14: Assembly of the Bead Deposition Device

12. Slowly pipet **350 µl** of BB2 onto the PTP device through the loading port (the larger of the two holes, opposite of the notch, Figure 15) and spin the BDD in a centrifuge for **5 minutes** at **1,620 RCF** (4,013 RPM for the Eppendorf 5430).

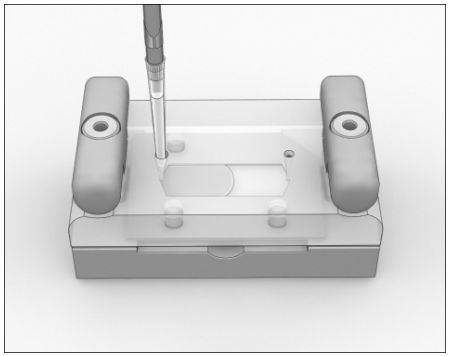


Figure 15: Loading the PTP device

- There are two GS Junior Bead Deposition Device (BDD) adapters. Adapter A is used with the Beckman centrifuges, and Adapter B is used with the Eppendorf 5430 centrifuge. Both adapters are supplied with the GS Junior Installation Kit, the GS Junior PM Kit, and the GS Junior BDD Counterweight. Use the one matching your microplate centrifuge brand.
- Balance the rotor with the BDD counterweight. These are supplied with the GS Junior Installation Kit and the GS Junior BDD Counterweight.
- Place adapters A or B in the swinging buckets such that the arrow on the adapter top points to the rotor axle. See Section 4.2 for pictures of the adapters in the centrifuges.

13. Remove the BDD from the centrifuge and leave it on the bench.

3.2.3 Prepare the Beads

The GS Junior System contains 4 kinds of microparticles (beads), as listed in the *Research Applications Guide – GS Junior Titanium Series*. Each type of bead must undergo a specific preparation procedure.

These beads will be loaded onto the PTP device in layers, in the order specified in Table 1, with layer 1 loaded first and layer 4 last.

Bead Layer	Bead Type
Layer 1	Enzyme Beads Pre-layer
Layer 2	DNA and Packing Beads
Layer 3	Enzyme Beads Post-layer
Layer 4	PPiase Beads

Table 1: Bead layers

3.2.3.1 Prepare the DNA Beads – Adding the Control Beads XLTF

1. Obtain an aliquot of enriched DNA beads in a proper amount for a sequencing Run on the GS Junior Instrument.
2. Spin the Control Beads XLTF in a microcentrifuge for 5 seconds at **9,300 RCF** (10,000 RPM) to collect the beads. Pipet up and down 5 times to resuspend the beads.
3. Add **6 µl** of Control Beads XLTF to the enriched DNA beads.
4. If the volume of sample DNA beads is greater than 100 µl, spin the DNA beads in a microcentrifuge for 10 seconds at **9,300 RCF** (10,000 RPM) to pellet the beads. Remove enough supernatant to leave 100 µl in the tube.
5. Add **500 µl** of BB2 to the DNA Beads, vortex gently, and incubate for **20 minutes** at **room temperature** on a lab rotator.

During this 20 minute incubation, prepare the Packing Beads, Enzyme Beads, and PPiase Beads.

3.2.3.2 Prepare the Packing Beads

1. Add **1 ml** of BB2 to the tube of Packing Beads and vortex at high speed.
2. Wash the Packing Beads **three times** with **1 ml** of BB2 and centrifuge at **9,300 RCF** (10,000 RPM) for **5 minutes** for each wash. Wash the beads thoroughly, by vortexing to break up aggregates, until a uniform suspension is achieved. After spinning, carefully remove the supernatant without disturbing the bead pellet.
3. After the third wash, add **200 µl** of BB2, resuspend the beads by vortexing at high speed, and keep the tube on ice.

3.2.3.3 Prepare the Enzyme and PPiase Beads (Bead Layers 1, 3 & 4)

These two types of beads can be washed in parallel. Make sure to change pipette tip to avoid contaminating the Enzyme and PPiase Beads with one another.

1. Add **1 ml** of BB2 to each tube and mix by vortexing at medium speed. Pellet the Enzyme Beads and the PPiase Beads using a Magnetic Particle Concentrator (MPC), waiting **30 seconds** for the beads to pellet. Invert the MPC several times and wait **30 seconds** again. Carefully remove the supernatants and remove the tubes from the MPC.
2. Wash both bead types **three times** with **1 ml** of BB2. Vortex and collect the beads using the MPC, as above.
3. After the third wash, add **400 µl** of BB2 to the Enzyme Beads and **410 µl** of BB2 to the PPiase Beads, and vortex at medium speed to resuspend the beads. Keep the tubes on ice.
4. Prepare two new 1.7 ml tubes, labeled Enzyme Pre-layer and Enzyme Post-layer, following Table 2. Vortex at medium speed and place the tubes on ice.

Reagents	BB2	Enzyme Beads	Total volume
Enzyme Pre-layer	300 µl	110 µl	410 µl
Enzyme Post-layer	180 µl	230 µl	410 µl

Table 2: Preparation of the Enzyme Beads for the pre and post-layers

3.2.3.4 Prepare the DNA and Packing Beads Mix

1. Retrieve the tube of DNA Beads from the lab rotator.
2. Spin the DNA Beads in a microcentrifuge for **10 seconds** at **9,300 RCF** (10,000 RPM) to pellet the beads.
3. Remove enough BB2 to leave 50 µl in the tube (calculate the volume of supernatant to remove from the known volume of enriched DNA beads, plus 6 µl of Control Beads XLTF, plus 500 µl of BB2).
4. Spin the Polymerase and Polymerase Cofactor tubes in a microcentrifuge for 5 seconds at **9,300 RCF** (10,000 RPM).
5. Add the following reagents to the DNA Beads:
   - 40 µl Polymerase
   - 20 µl Polymerase Cofactor
   - 65 µl BB2
6. Vortex the mixture at low speed for 5 seconds and incubate on the lab rotator **at room temperature** for **10 minutes**.
7. Once the incubation is complete, vortex the tube of Packing Beads at high speed.
8. Add **175 µl** of Packing Beads to the DNA mixture, vortex at low speed, and incubate on the lab rotator **at room temperature** for **5 minutes.**

3.3 Loading the Bead Layers and Priming the Instrument

The beads are deposited onto the PTP device by injecting the bead suspension through the loading port (the larger of the two holes, see Figure 15) of the assembled Bead Deposition Device (BDD), and then by using centrifugal sedimentation to settle the beads at the bottom of the PicoTiterPlate wells. This process is repeated for each of four layers:

- Bead layer 1: Enzyme Beads Pre-layer
- Bead layer 2: DNA and Packing Beads
- Bead layer 3: Enzyme Beads Post-layer
- Bead layer 4: PPiase Beads

During the centrifugation of the bead layers, prepare the Buffer CB (see Section 3.3.2), prime the GS Junior Instrument with the reagents (see Section 3.3.4), and clean the instrument (see Section 3.3.6).

- **Pipette tips**: Use 1000 µl pipette tips to load beads.
- **Time between centrifugations**: Minimize the time interval between loading the beads and starting the centrifugation.

- **Air bubbles:** Avoid injecting air into the BDD.
- **Bead delivery:** Use a single, even injection to fill the BDD.
- **Loading fill:** Fill the BDD completely but do not overflow. Discard any excess bead mix.

3.3.1 Deposit Bead Layer 1: the Enzyme Beads Pre-Layer

1. Retrieve the BDD from Section 3.2.2 and carefully pipet out and discard as much BB2 as possible through the port hole on the BDD.
2. Vortex the tube of Enzyme Beads Pre-layer (layer 1) at low speed for 5 seconds to obtain a homogeneous suspension.
3. Promptly load 350 µl of the bead suspension onto the PTP device, through the port hole on the BDD top (see Figure 15). Make sure to use a single, smooth dispensing action to ensure even distribution of the beads over the entire PTP device.
4. Centrifuge the BDD for **5 minutes at 1,620 RCF** (4,013 RPM).

3.3.2 Prepare Buffer CB

1. Retrieve the bottle of Buffer CB.
2. Vortex the tube of DTT tube at high speed for 5 seconds. Add 1 ml of DTT to the bottle of Buffer CB.
3. Invert the tube of Substrate TW 10 times to mix.
4. Using a serological pipette, transfer 44 ml of Substrate TW to the Buffer CB.
5. Invert the bottle 10 times to mix thoroughly.

3.3.3 Deposit Bead Layer 2: the DNA and Packing Beads

1. Remove the BDD from the centrifuge.
2. With a pipettor, gently remove as much of the supernatant as possible through the port hole on the BDD top.
3. Remove the tube of the DNA and Packing Beads mix suspension from the rotator.
4. Spin the tube in a microcentrifuge for 5 seconds at 9,300 RCF (10,000 RPM).
5. Pipet up and down 5 times to mix.
6. Promptly load 350 µl of this suspension onto the PTP device.
7. Centrifuge the BDD for **10 minutes at 1,620 RCF** (4,013 RPM).

3.3.4 Prime the GS Junior Instrument

- An abbreviated version of the following instructions appear on the screen of the Attendant PC (Figure 13).
- If the caps of tubes in the Reagent Cassette are tight and do not unscrew easily, use the GS Junior Reagent Decapping Tool to unscrew them.

1. Lift the instrument cover and raise the sipper manifold.
2. Remove and discard the Pre-wash cassette and Pre-wash bottle.
3. Empty the waste bottle on the right.
4. Change gloves.
5. Invert the Reagents cassette 20 times to mix.
6. Carefully remove all the tube caps, making sure that the caps do not pass over the tubes to reduce the risk of contamination.
7. Mount the Reagents cassette in the instrument.
8. Place the bottle of Buffer CB on the left.
9. Slowly lower the sipper manifold, checking that sippers plunge in the reagent tubes and Buffer CB bottle.
10. Close the instrument cover.
11. Click Proceed to begin priming the instrument. The Prime will last approximately 5 minutes (Figure 16).
12. Once the Prime is complete, the Instrument Procedure third window, Enter IDs and barcodes, will open automatically (Figure 21, Section 3.4).

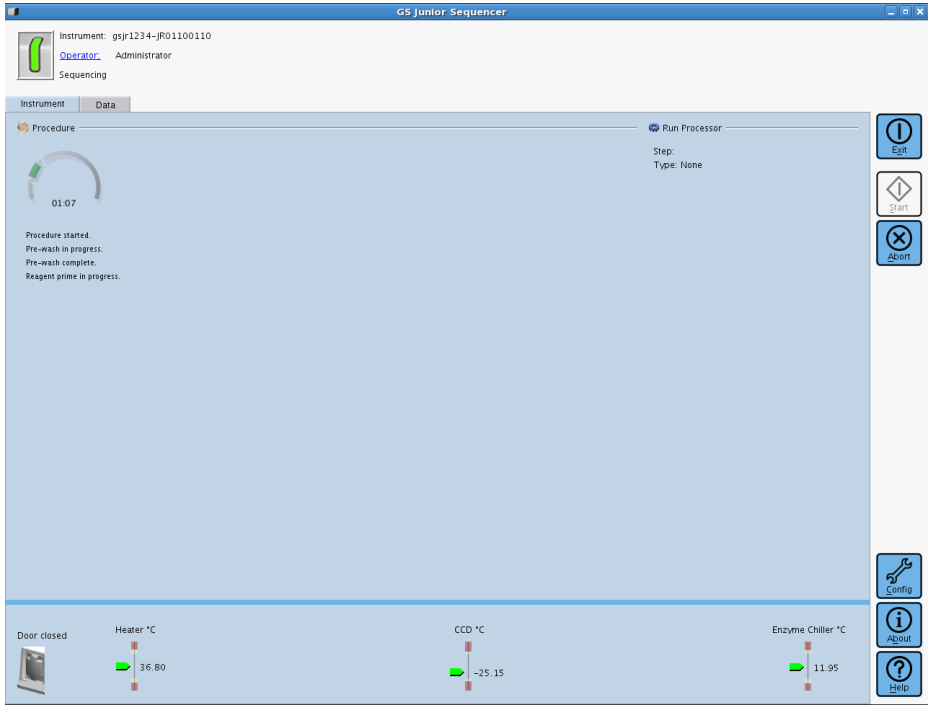


Figure 16: The GS Junior Sequencer application window when priming is in progress

3.3.5 Deposit Bead Layer 3: the Enzyme Beads Post-Layer

1. Change gloves after loading the Reagents cassette into the instrument.
2. Remove the BDD from the centrifuge.
3. With a pipettor, gently draw out all the supernatant from bead layer 2 through the port hole on the BDD top.
4. Vortex the Enzyme Beads Post-layer for 5 seconds at medium speed to obtain a uniform suspension.
5. Promptly load 350 µl of the bead suspension onto the PTP device, through the loading port hole on the BDD top.
6. Centrifuge the BDD for **10 minutes at 1,620 RCF** (4,013 RPM).

3.3.6 Prepare the GS Junior Instrument PTP Cartridge

1. Once the instrument priming is complete, the GS Junior application Sequencer window will read Prime Complete.
2. Lift the instrument cover and open the camera door by pressing on the two side latches and pulling the door in a downward motion (Figure 17).

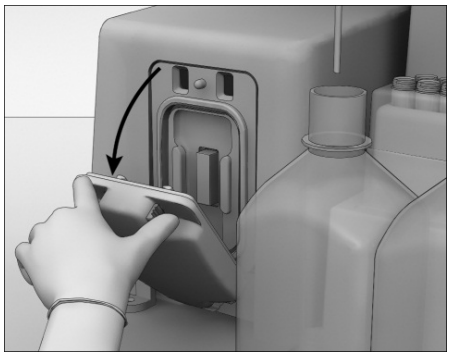


Figure 17: Opening the camera door

3. Remove the PTP device of the previous sequencing Run from the cartridge by first pressing the PTP frame spring latch to lift the frame from the cartridge (Figure 18**A**) and then lifting out the used PTP device (Figure 18**B**).

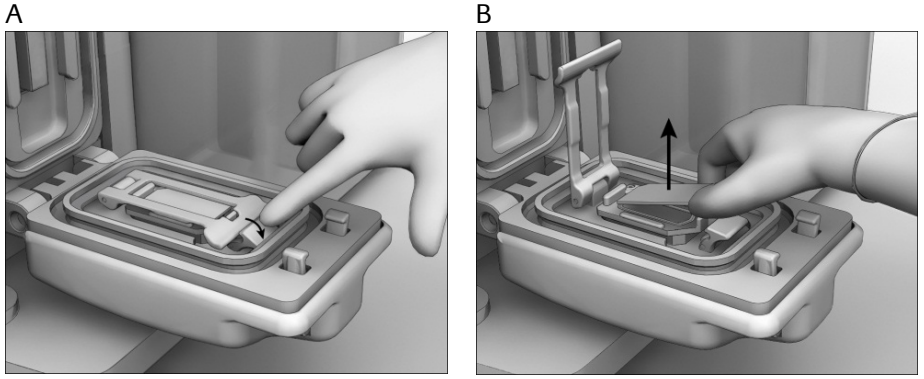


Figure 18: Removing the used PTP device

4. Carefully remove the cartridge seal with a clean pipet tip and discard it (Figure 19**A**).
5. Change gloves.
6. Wet a Kimwipe with 50% ethanol and wipe the surface of the cartridge to remove any bead and reagent residue (Figure 19**B**). Allow the cartridge to air dry completely.

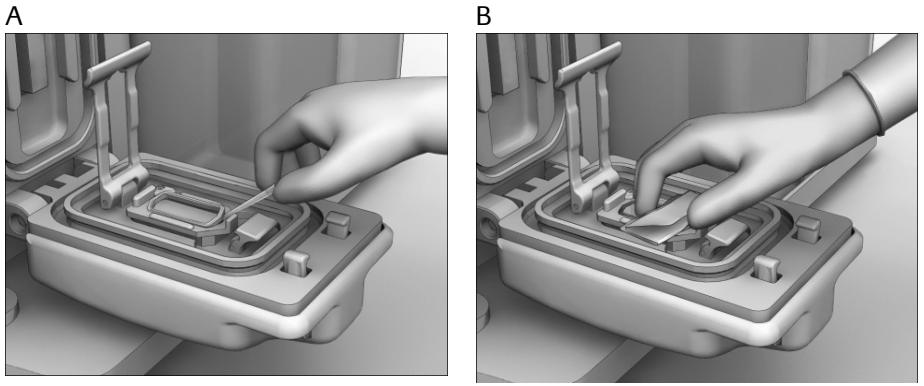


Figure 19 (A) Removing the PTP cartridge seal and (B) wiping the PTP cartridge

7. Insert the washed cartridge seal from Section 3.2.2 in its groove in the cartridge (Figure 20, **A** and **B**).

A

B

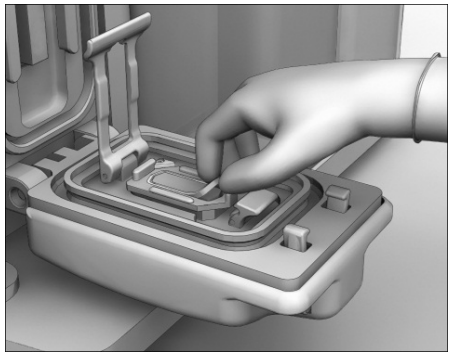


Figure 20: (A) The cartridge seal and (B) its placement in the groove in the cartridge

8. Moisten a Kimwipe with a user-prepared 10% solution of Tween-20 and wipe the surface of the PTP cartridge.

Sensitive camera faceplate: Always be extremely careful when handling or working near the camera faceplate. Never touch the camera faceplate with anything other than Zeiss moistened cleaning tissue or Lens paper from Thorlabs.

9. Use a new Zeiss pre-moistened cleaning tissue to gently wipe the camera faceplate.
10. Allow the camera faceplate to air dry completely.

3.3.7 Deposit Bead Layer 4: the PPiase Beads

1. Remove the BDD from the centrifuge.
2. With a pipettor, gently draw out and discard all the supernatant from the centrifuged bead layer 3, through the port hole on the BDD top.
3. Vortex the PPiase Beads tube for 5 seconds at medium speed to obtain a uniform suspension.
4. Promptly load 350 µl of the bead suspension onto the PTP device, through the port hole on the BDD top (see Figure 15).
5. Centrifuge the loaded PTP device in the BDD for **5 minutes at 1,620 RCF** (4,013 RPM).

3.4 The Sequencing Run

This Section describes how to manually set up and launch a sequencing Run. However, if the GS Junior Instrument is connected to a Laboratory Information Management System (LIMS), the PTP Lookup feature of the Instrument Procedure wizard will seek the information that describes the Run from your LIMS after you enter the PTP device ID.

3.4.1 Set the Run Script and Other Run Parameters (without LIMS)

1. The GS Junior Instrument Procedure third window, Enter IDs and barcodes, opens automatically at the end of instrument priming (Figure 21).

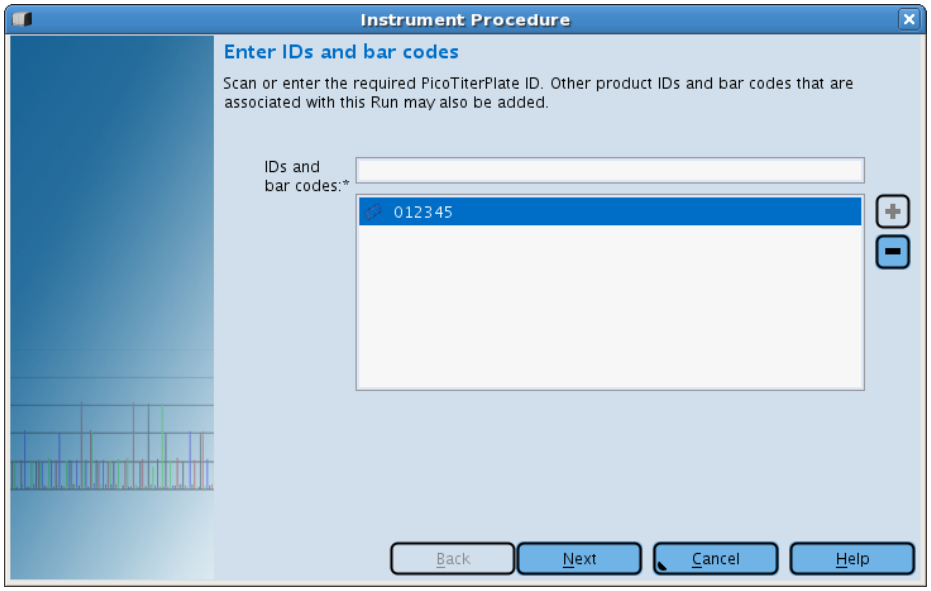


Figure 21: The Instrument Procedure third window: Enter IDs and barcodes

2. Enter the 6-digit ID of the PicoTiterPlate device to be used in this Run. You may also enter the product IDs or barcodes of other materials associated with this sequencing Run (*e.g.* Library Prep, emPCR Amplification, and Sequencing Kits); you will be asked for this information if you call Roche Customer Support for help if you encounter any difficulties with your Run. When you have entered all the information, click Next.
    - This opens the Instrument Procedure fourth window: Enter Run name and Run Group (Figure 22).

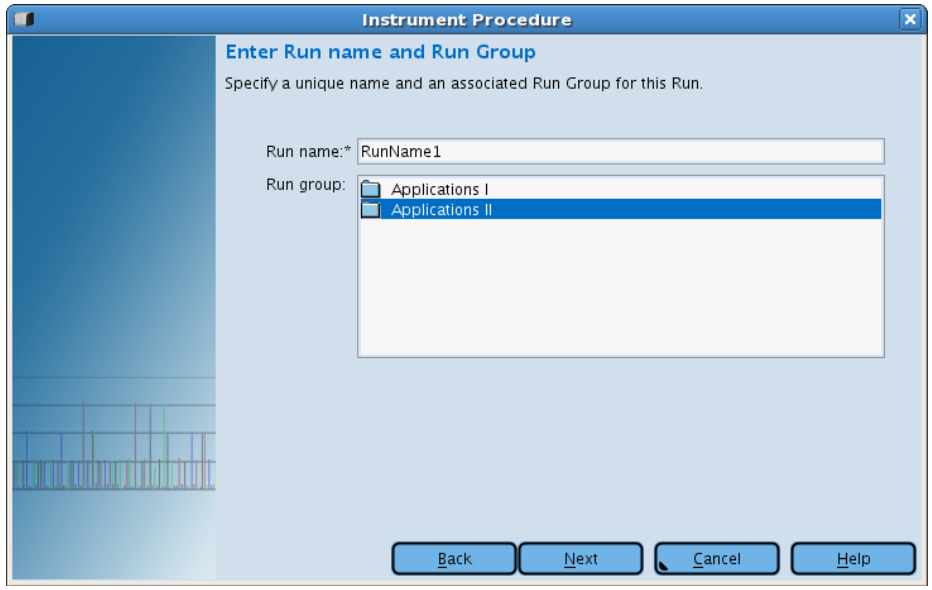


Figure 22: The Instrument Procedure fourth window: Enter Run name and Run Group

3. Enter a specific, unique name for this Run. Then find and select your Run Group in the Run group list. Click Next.
    - This opens the Instrument Procedure fifth window: Choose number of cycles (Figure 23).

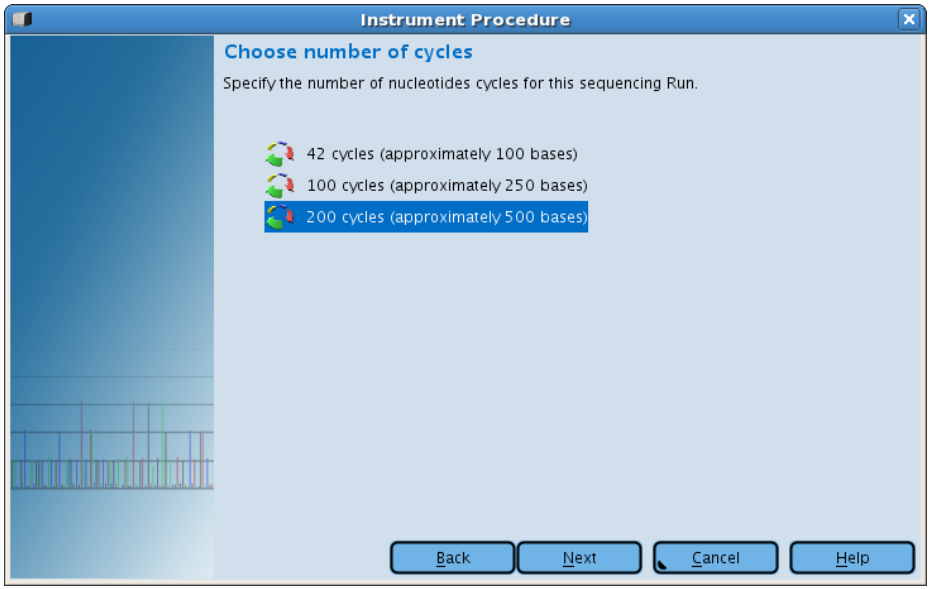


Figure 23: The Instrument Procedure fifth window: Choose number of cycles

4. Select the number of nucleotide cycles appropriate for this Run, and click Next.
5. With the GS Junior Titanium chemistry,
    - 42 cycles will produce reads of approximately 100 bases (~ 3 hours)
    - 100 cycles, approximately 250 bases (~ 5 hours and 30 minutes)
    - 200 cycles, approximately 500 bases (~ 9 hours and 20 minutes)
    - This opens the Instrument Procedure sixth window: Choose Run Processing type (Figure 24).

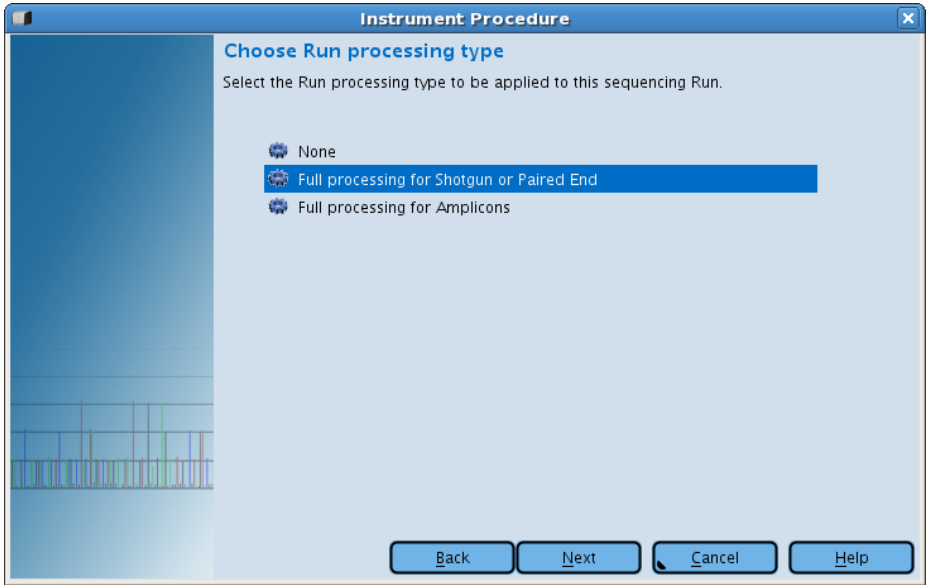


Figure 24: The Instrument Procedure sixth window: Choose Run Processing type

6. Select the data processing scheme appropriate for this Run, and click Next.
7. Three data processing schemes are available on the GS Junior System (see the *454 Sequencing System Software Manual* for a complete description of the GS Junior System's data processing and analysis processes, including a description of when to use each of these options):
    - None
    - Full processing for Rapid, Paired End or cDNA Rapid libraries
    - Full processing for Amplicon libraries
8. This opens the Instrument Procedure seventh window: Request data backup (Figure 25).

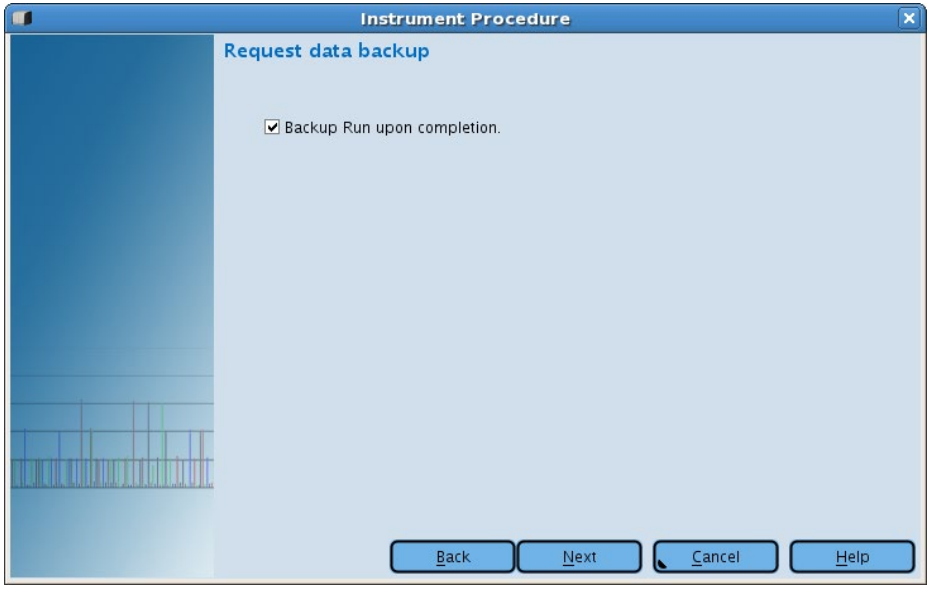


Figure 25: The Instrument Procedure seventh window: Request data backup

9. The Backup Run and Processor data upon completion checkbox is selected by default. Click Next.
    - With this checkbox selected, the data from the Run will be automatically backed up at the end of the Run (to a storage location pre-determined by an Administrator, see Appendix 4.1).
    - This opens the Instrument Procedure eighth window: Run comments (Figure 26).

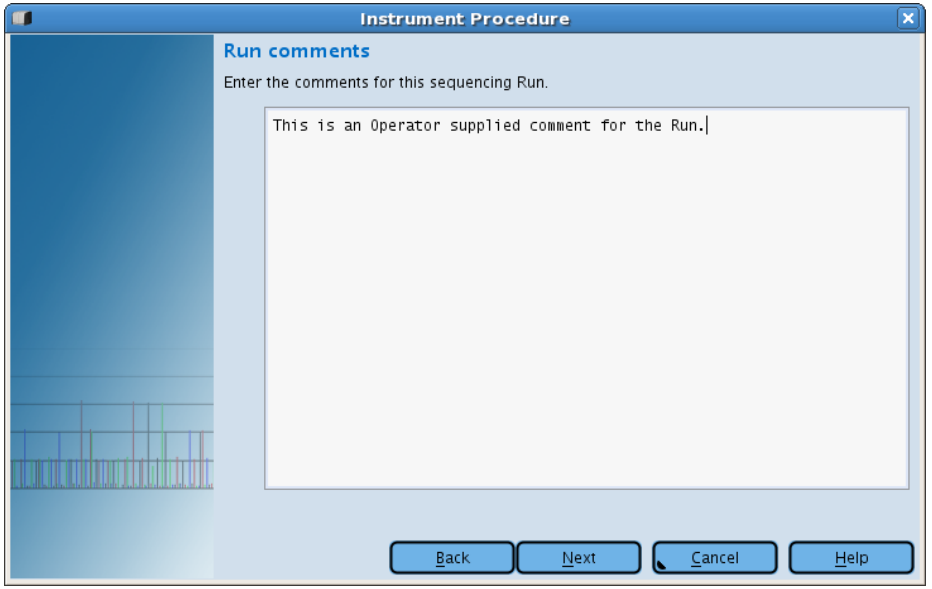


Figure 26: The Instrument Procedure eighth window: Run comments

10. Enter any comments about the Run, and click Next.
    - This opens the ninth and last window: Run Parameters Review (Figure 27).

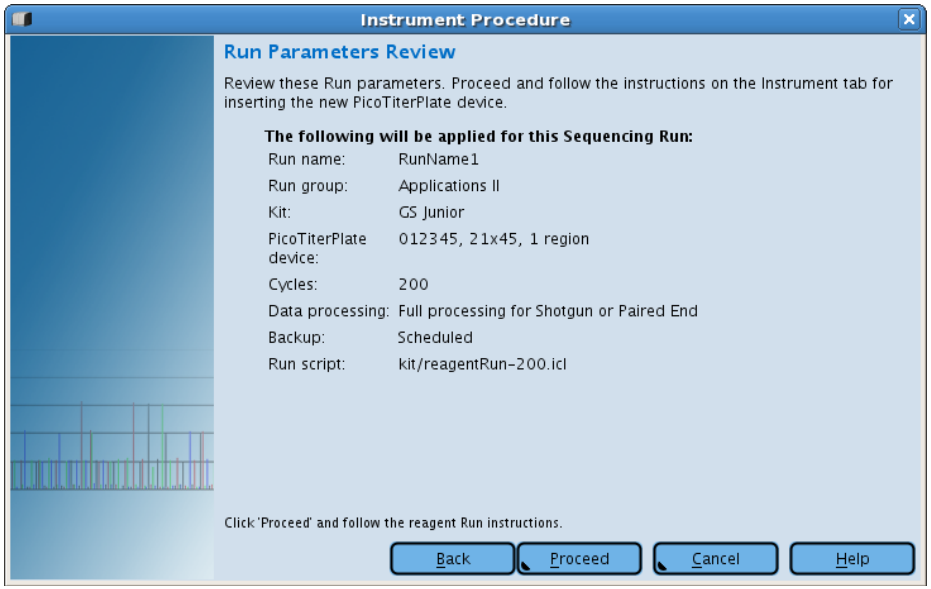


Figure 27: The Instrument Procedure ninth window: Run Parameters Review

- If the Proceed button is still grayed out, check for "x" in red circle icons in the Instrument Procedure window, and address any problem that may be listed.

11. Click Proceed in the Instrument Procedure Run Parameters Review window. This will navigate back to the Instrument tab on the GS Junior Sequencer application window (Figure 28).

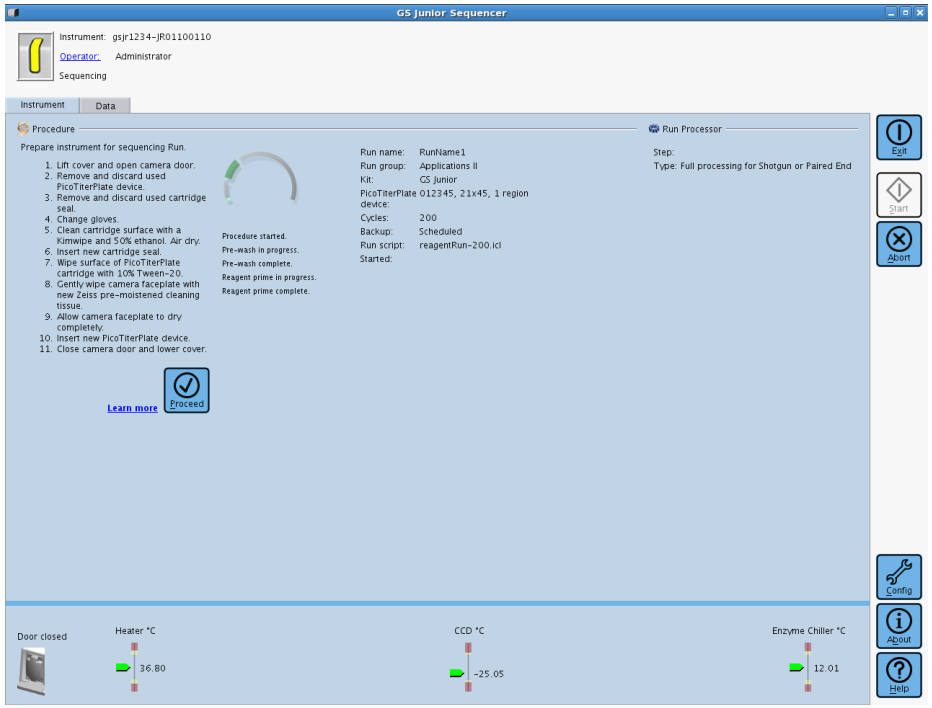


Figure 28: The GS Junior Sequencer application window before a sequencing Run is started

3.4.2 Load the PTP Device in the GS Junior Instrument

1. Remove the BDD from the centrifuge.
2. With a pipettor, gently draw out and discard all the supernatant from the centrifuged bead layer 4, through the port hole on the BDD top.
3. Remove the PTP device from the BDD, as follows:
4. Rotate down the latches of the BDD to unfasten them.
5. Carefully remove the BDD top.
6. Gently lift off and discard the bead loading gasket.
7. Remove the PTP device, being careful to handle it only by the edges.
8. Gently place the PTP device into the cartridge frame on top of the cartridge seal (Figure 29**A**), flipping it upside down so that the wells face downwards. Make sure that the PTP device notch is on the far right hand corner, matching the notch in the cartridge.
9. Close the PTP frame, making sure it is properly secured by the latch (Figure 29**B**).
10. Carefully wipe the back of the PTP device with a Kimwipe.
11. Close the camera door and lower the instrument cover (Figure 29**C**).

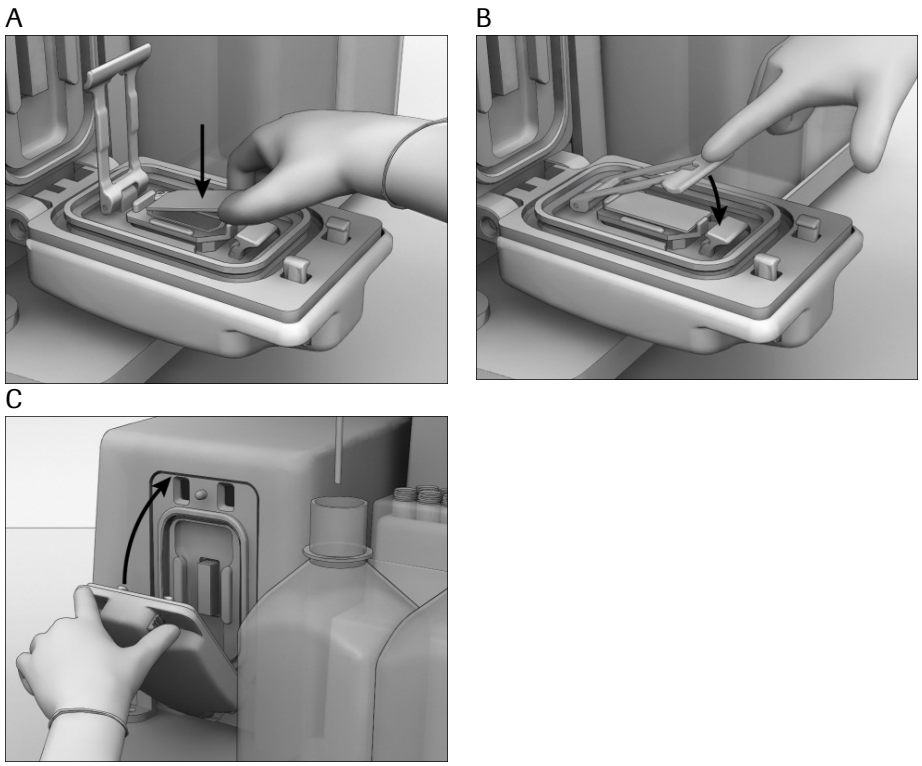


Figure 29: Loading the PTP device into the cartridge

12. In the GS Junior Sequencer application window, click Proceed to begin the sequencing Run. The Operator can monitor the progress of a sequencing Run by viewing the instrument status and the data images as they are being captured by the camera: Thumbnail images will appear on the Instrument tab during the Run (Figure 30).

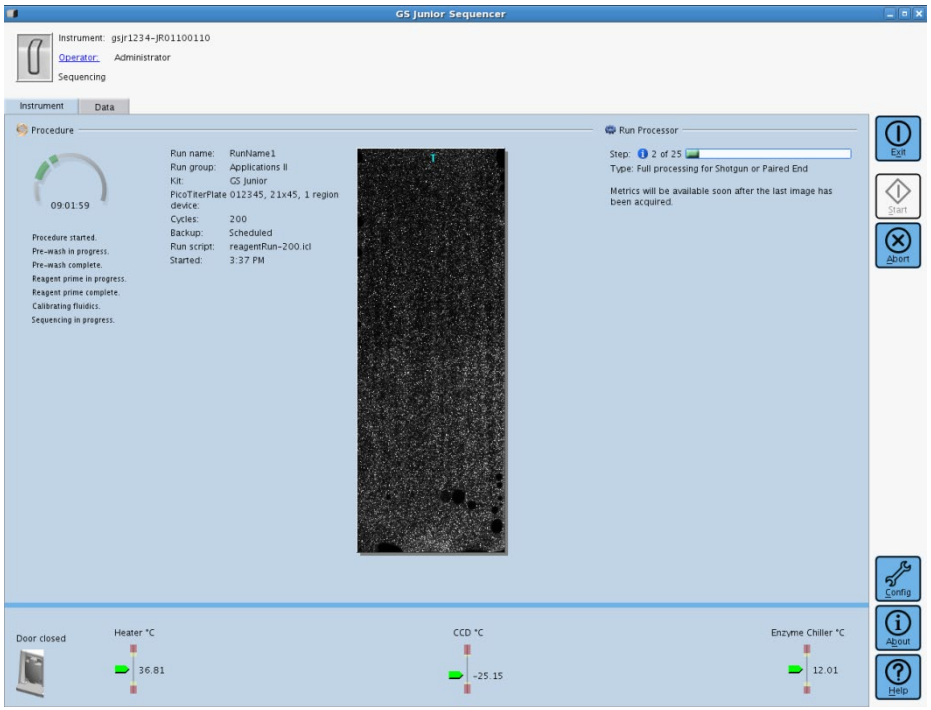


Figure 30: The GS Junior Sequencer application window with a sequencing Run in progress

- Monitor the instrument until the sequencing Run is under way and the Status LED is blinking green (also shown at the upper-left corner of the GS Junior Sequencer window). If the instrument encounters any problems during the initiation of the Run, a message describing the issue will appear in the Status area of the GS Junior Sequencer window, and user intervention will be required.
- In most cases, the software will offer to Abort or Proceed with the Run. If you are certain that the warning is unwarranted then you can decide to proceed.
- In a few special cases, however, you will not be offered the possibility to proceed. If this happens, you must restart the software, as follows:
  - Close the GS Junior Sequencer application.
  - Re-launch the GS Junior Sequencer application by double-clicking the GS Junior Sequencer icon on the desktop, and set up your Run again.
- There is an *Abort* button available in the Global Action area of the GS Junior Sequencer application main window (Figure 30) which can be clicked if a problem occurs during the Run. The Abort Run dialog box will ask you to verify that you want to stop the Run; if you confirm, the Run will be immediately terminated. There is no procedure for pausing and resuming a Run. If a Run is aborted, follow the abort with a Pre-wash before proceeding with another sequencing Run.

13. When the sequencing Run is complete, the GS Junior application window will show the message Run complete and an OK button (Figure 31). Clicking the OK button will navigate to the GS Junior application window at the beginning of a new sequencing Run (Figure 3).

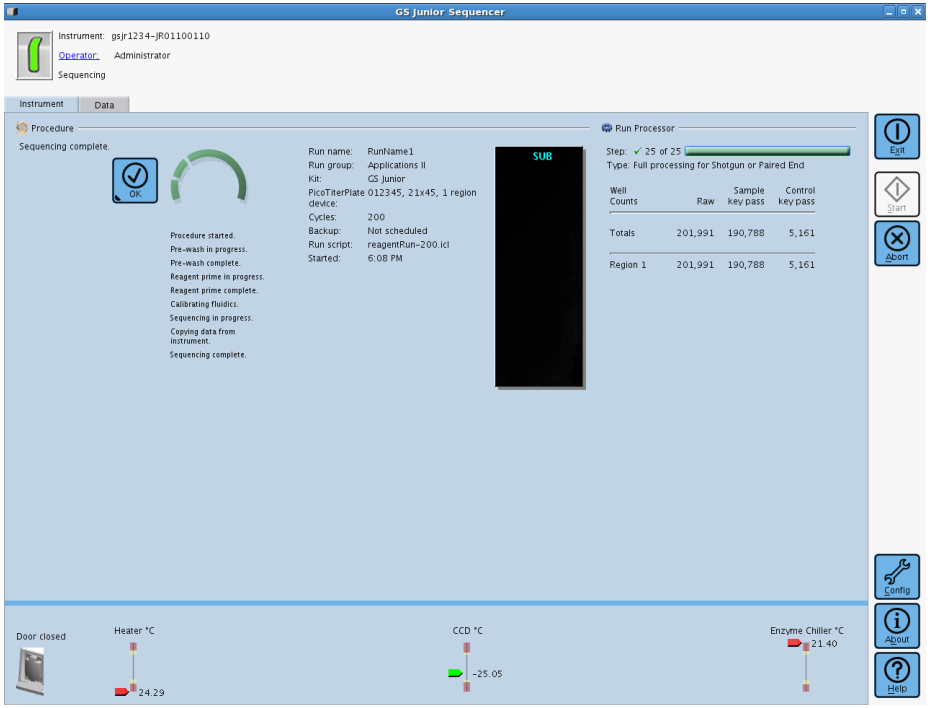


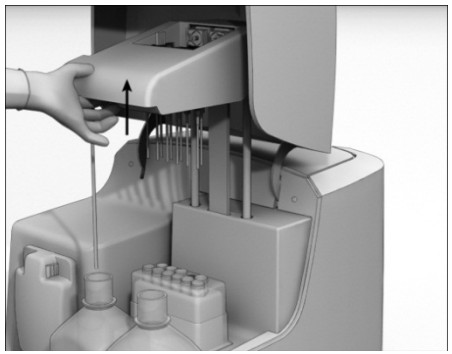
Figure 31: The Sequencing Run window once the sequencing Run is complete

3.5 After the Sequencing Run

Once the sequencing Run has ended, at your earliest convenience, remove Sipper Tubes and filters, as well as the Reagent bottle and Reagent cassette. Performed within three days, this will minimize the risk of biofilm buildup.

1. Lift the instrument cover and raise the sipper manifold completely (Figure 32**A**).
2. Remove and discard the used Reagents cassette (Figure 32**B**).

A



B

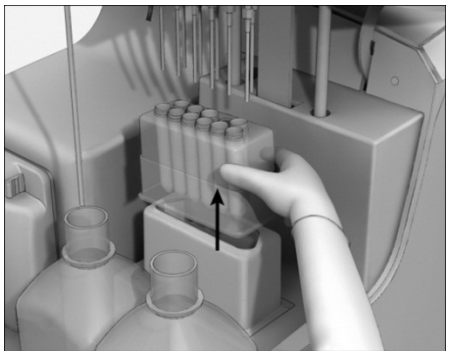


Figure 32: (A) Raising the sipper manifold and (B) Removing the Reagents cassette from the previous Run

3. Discard the waste bottle on the right.
4. Empty the bottle of Buffer CB on the left and place it on the right, to be used as the waste bottle for the next sequencing Run.
5. Unscrew all the used sippers for the reagent and the buffer and discard (Figure 33**A** and **B**).

For an instrument not in continual use, follow the instructions below:

- As soon as possible within three days of a sequencing Run, the instrument should be emptied of consumables, as described in Step 4 above. The instrument can stay in this configuration for up to 30 days after a sequencing Run.
- If the instrument is to stay idle between 30 and 60 days, perform a maintenance wash immediately prior the next sequencing Run.
- If the instrument is to stay idle for more than 60 days, prepare it for storage.

Performing a maintenance wash and preparing the instrument for storage are described in the *GS Junior Instrument Owner's Manual*.

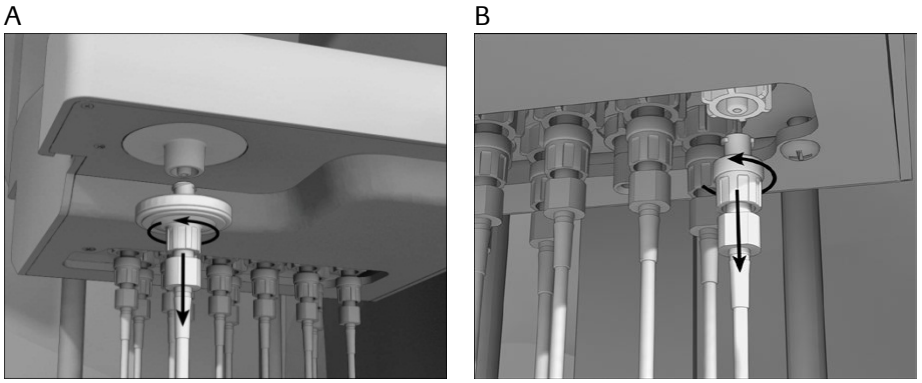


Figure 33: (A) Removing the Buffer Sipper Tube and (B) the Reagents Sipper Tubes

4 APPENDIX

4.1 Network Setup for Sequencing Data Backup

A network data backup mechanism to archive sequencing Run data is important because the Attendant PC and the instrument are not meant for the long-term storage of datasets. This is usually setup by, or in coordination with, your IT personnel. Refer to the *GS Junior System SysAdmin Guide* for a detailed description of recommended data backup setup.

Table 3 shows the approximate size of the data sets created during a sequencing Run for each data processing option (assuming the largest bead loading regions available on the PTP device).

Sequencing Kit	Number of Cycles	Raw Images Only	Raw Images Plus Image Processing	Raw Images Plus Full Processing
XLR70	200	10 GB	10-11 GB	11-12 GB

Table 3: Approximate amount of data generated by the XLR70 sequencing kit for the 200 cycle Run script and for each data processing option

4.2 BDD Centrifuge Adapters

In your swinging bucket centrifuge, place the BDD in the appropriate adaptor, with a counterweight when need. Make sure to use adapters A with the Beckman centrifuges (Figure 34**A**) and adapters B with the Eppendorf 5430 centrifuge (Figure 34**B**). In all cases, the arrows on the adapters should point toward the rotor axle. For information on centrifuges usable in this setting, refer to the *GS Junior System Laboratory Setup Guide*.

A B

Figure 34: (A) Adapters A in a Beckman centrifuge and (B) Adapters B in the Eppendorf 5430, holding a BDD and a BDD counterweight

Table of Contents

Appendix 8.4

SNP Array Protocol

Infinium® HD Assay Ultra, Manual

Experienced User Card

Day 1

Quantitate DNA

Hands-on: 30 min/plate
Fluorometer: 5 min/plate

Reagents
Lambda DNA
PicoGreen dsDNA
1X TE

Output
Sample QDNA Plate with Quantitated DNA

Make MSA3

Hands-on: ~60 min/ 96 samples

Reagents
0.1N NaOH
MA1
MA2
MSM

Output
MSA3 Plate

Incubate MSA3

Incubation: 20–24 hours

Output
MSA3 Plate

Day 2

Fragment MSA3

Hands-on: ~30 min/ 96 samples
Incubation: 1 hour

Reagents
FMS

Output
MSA3 Plate

Precip MSA3

Hands-on: ~30 min/ 96 samples
Incubation/Dry Time: 2 hours

Reagents
2-propanol
PM1

Output
MSA3 Plate

Resuspend MSA3

Hands-on: ~30 min/ 96 samples
Incubation: 1 hour

Reagents
RA1

Output
MSA3 Plate

Hyb Multi BeadChip

Hands-on: ~1 hour/ Up to 8 BeadChips
Incubation: 16–24 hours

Reagents
PB2

Output
BeadChip

Day 3

Wash BeadChip

Hands-on: ~30 min/ Up to 8 BeadChips

Reagents
PB1

Output
BeadChip

XStain HD BeadChip

Hands-on: ~3 hours/ 8 BeadChips
Dry Time: 1 hour

Reagents
RA1
95% Formamide / 1 mM EDTA
PB1
XC1
XC2
XC3
XC4
TEM
STM
ATM

Output
BeadChip

Image BeadChip

Scan times vary depending on Bead Pool complexity. See the Image BeadChip section for specific scan times.

Output
Image and Data Files

Optional
Pre-Amp
Post-Amp
Cold Storage Option
Overnight Incubation

Fill in the lab tracking form and the sample sheet as you perform the assay

Appendix 8.5

Mutation frequencies across myeloid malignancies

Appendix 8.5 – Mutation frequencies across myeloid malignancies

Mutation	AML			MDS			MDS/MPN			MPN		
NRAS	AML	9.8%	(n=2128)	MDS	6%	(n=252)	CMML	15.4%	(n=266)	PMF	0%	(n=40)
		14%	(n=201)		12%	(n=50)		16%	(n=273)			
		11%	(n=1106)		9%	(n=70)	aCML	0-55%				
		10.3%	(n=2502)		7%	(n=70)						
	NK-AML	13%	(n=872)		3.6%	(n=439)						
	s/tAML	5.9%	(n=591)	low/int	3%	(n=288)						
				+CMML	5.7%	(n=176)						
KRAS	AML	5%	(n=739)	MDS	1.5%	(n=65)	CMML	10.8%	(260)			
					0.9%	(n=439)		11%	(266)			
TET2	AML	12%	(n=91)	MDS	19%	(n=81)	CMML	61.1%	(n=157)	MPN	12%	(n=198)
		19%	(n=32)		23%	(n=96)		44.4%	(n=81)		13%	(n=239)
		13.2%	(n=486)		26%	(n=102)		44%	(n=52)		8%	(n=354)
	NK-AML	23%	(n=427)		18.5%	(n=65)		34%	(n=38)	MF	15%	(n=46)
	Young	7.6%	(n=783)		20.5%	(n=439)		61%	(n=160)		14%	(n=21)
				low/int	23%	(n=288)	RARS-T	9%	(n=22)		17%	(n=60)
											18%	(n=55)
										PV	16%	(n=89)
											13%	(n=71)
										ET	5%	(n=57)
											2%	(n=43)
SRSF2	AML	0.7%	(n=151)	MDS	12.4%	(n=193)	CMML	47.8%	(n=268)	PMF	17%	(n=187)
					14.6%	(n=233)		47%	(n=275)			
					6.8%	(n=88)		28.4%	(n=88)			
					11.1%	(n=221)						
				low/int	13%	(n=288)						
				MDS-RS	5.5%	(n=73)						
				w/o RS	11.6%	(n=155)						

ASXL1	AML	17.5%	(n=63)	MDS	18.5%	(n=65)	CMML	44.7%	(n=255)	PV	0	(n=10)
		10.8%	(n=501)		15.3%	(n=182)		34.2%	(n=79)		2%	(n=42)
	<60yrs	3.2%	(n=189)		20.7%	(n=193)		46%	(n=52)	ET	2.8%	(n=35)
	>60yrs	16.2%	(n=234)		14.4%	(n=439)		50%	(n=38)		0	(n=41)
		5.3%	(n=882)	low/int	15%	(n=288)		44%	(n=261)	MF	13%	(n=46)
		17.2%	(n=740)					33%	(n=51)		30%	(n=10)
							RARS-T	10%	(n=20)		32%	(n=47)
											55%	(n=42)
											18%	(n=22)
											43%	(n=30)
											22%	(n=23)
RUNX1	AML	25.9%	(n=814)	MDS	4.7%	(n=85)	CMML	22.8%	(n=267)	MPN	0	(n=14)
		13.2%	(n=470)		6.2%	(n=65)		8.7%	(n=81)			
		5.6%	(n=945)		8.7%	(n=439)		26%	(n=38)			
	NK-AML	32.7%	(n=449)	low/int	9%	(n=288)		22%	(n=274)			
				MDS+CMML	12%	(n=132)	MDS/MPN	14%	(n=187)			
				HR+CMML	13.8%	(n=188)						
				HR+AML	23.6%	(n=110)						
CBL	AML	0.6%	(n=319)	MDS	2.3%	(n=439)	CMML	19.1%	(n=267)	PV	0	(n=74)
		0.7%	(n=150)	low/int	2%	(n=288)		22.2%	(n=81)		0	(n=32)
		1.1%	(n=279)	HR	7.7%	(n=65)		13%	(n=52)		1.1%	(n=89)
	t(8;21)	5%	(n=40)					5%	(n=38)	ET	0	(n=24)
	Inv 16	16%	(n=37)					10%	(n=78)		0	(n=48)
								17.3%	(n=278)		0.8%	(n=245)
								19%	(n=274)	PMF	6%	(n=53)
							aCML	8%	(n=152)		2%	(n=49)
							MDS/MPN	2.3%	(n=222)		0	(n=19)
								8.1%	(n=301)			

EZH2	AML -7/7q	0	(n=54)	MDS	6.4%	(n=439)	CMML	9.3%	(n=205)	MF	7%	(n=46)
					6%	(n=148)		11.1%	(n=81)		5.9%	(n=370)
				low/int	8%	(n=288)		6%	(n=52)		1.2%	(n=84)
								10%	(n=208)		9.4%	(n=64)
								13%	(n=118)			
							aCML	13%	(n=70)			
							MDS/MPN	12%	(n=219)			
JAK2 V617F	AML	2.7%	(n=113)	MDS	5%	(n=101)	CMML	6.7%	(n=268)	PV	97%	(n=73)
		1%	(n=959)		1.5%	(n=65)		9.9%	(n=81)		74%	(n=164)
		3.2%	(n=339)		3%	(n=439)		3%	(n=38)		89%	(n=45)
				low/int	3%	(n=288)		7%	(n=275)		65%	(n=128)
				5q-	6.4%	(n=78)	RARS-T	49.5%	(n=111)		81%	(n=72)
											80%	(n=84)
											74%	(n=57)
										ET	57%	(n=51)
											32%	(n=115)
											43%	(n=21)
											23%	(n=93)
											41%	(n=59)
											48.7%	(n=73)
											62%	(n=243)
											58.8%	(n=68)
										PMF	50%	(n=16)
											35%	(n=46)
											43%	(n=7)
											57%	(n=23)
											43%	(n=35)
											45.3%	(n=117)
											68%	(n=22)
											54.6%	(n=152)

									66.7%	(n=12)
									63.5%	(n=304)
									58%	(n=603)
JAK2 exon 12 (V617F neg)								PV	20%	(n=20)
									10.5%	(n=19)
									58%	(n=26)
									14%	(n=58)
								ET	0	(n=75)
								PMF	0	(n=35)
U2AF1	AML	2.5%	(n=275)	MDS	6.3%	(n=96)	CMML	5.2%	(n=268)	MPN 1.9% (n=53)
		1.3%	(n=151)		5.4%	(n=221)		8%	(n=88)	
					6.8%	(n=88)				
					8.7%	(n=150)				
					7.3%	(n=193)				
				low/int	16%	(n=288)				
				w/o RS	11.6%	(n=155)				
				MDS-RS	0	(n=73)				
SF3B1	AML	5%	(n=57)	MDS	20%	(n=354)	CMML	5%	(n=240)	PMF 4% (n=136)
		5%	(n=38)		17%	(n=88)		5%	(n=106)	6.5% (n=155)
		2.6%	(n=151)		28%	(n=533)		4.5%	(n=88)	ET 3% (n=189)
					14.5%	(n=193)	MDS/MPN	19.3%	(n=83)	MPN 0 (n=53)
					16.4%	(n=221)	RARS-T	86.5%	(n=111)	
				low/int	22%	(n=288)				
				MDS-RS	50%	(n=107)				
				MDS-RS	53%	(n=104)				
				MDS-RS	75.3%	(n=73)				
				w/o RS	6.5%	(n=155)				
KIT	AML	1.8%	(n=2136)	MDS	0.7%	(n=269)	CMML	4.2%	(n=263)	MPN 0 (n=115)
		5%	(n=500)		0	(n=28)				
	s/tAML	1.4%	(n=513)							

TP53	AML	11.5%	(n=1000)	MDS	12%	(n=57)				PMF	4%	(n=107)
		14%	(n=235)		14%	(n=70)						
	elderly	9%	(n=140)		11%	(n=47)						
	complex	60%	(n=234)		14%	(n=118)						
					7.5%	(n=439)						
				low/int	2%	(n=288)						
				5q-	18%	(n=55)						
PTPN11	AML	1.7%	(n=173)	MDS	0.7%	(n=439)	MDS/MPN	10%	(n=30)	MPN	0	(n=14)
		3.5%	(n=167)	HR	7%	(n=28)		1%	(n=87)			
		5.1%	(n=272)	MDS+CMML	0	(n=107)						
		3%	(n=64)	MDS+AML	0	(n=96)						
				t-MDS/AML	2.9%	(n=140)						
MPL	AML	0	(n=126)	MDS 5q-	3.8%	(n=78)	CMML	0	(n=81)	PV	0	(n=32)
	M7	25%	(n=12)					0	(n=118)		0	(n=57)
							RARS-T	13%	(n=23)		0	(n=242)
										ET	1%	(n=100)
											11%	(n=143)
											3.5%	(n=199)
											1.3%	(n=318)
											7%	(n=417)
											4.1%	(n=776)
										PMF	9%	(n=11)
											12.5%	(n=24)
											9%	(n=32)
										JAK2 WT	9%	(n=45)
											6%	(n=96)
											5.5%	(n=290)
											8%	(n=100)
											8.1%	(n=603)

IDH1	AML	8%	(n=187)	MDS	3%	(n=65)	CMML	1.2%	(n=81)	PMF	1.7%	(n=301)
		2%	(n=198)		4%	(n=71)	IDH1/2	4%	(n=52)		2.9%	(n=35)
		6%	(n=199)		2.4%	(n=82)				IDH1/2	6%	(n=46)
		5.5%	(n=493)		3.6%	(n=193)				IDH1/2	3.9%	(n=77)
		7.6%	(n=805)		1.4%	(n=439)				IDH1/2	4.2%	(n=312)
		9.6%	(n=520)		2.9%	(n=277)				MPN	0	(n=263)
		6%	(n=893)		7.4%	(n=108)				PV	0	(n=38)
		8%	(n=1333)	low/int	2%	(n=288)					0	(n=33_
		6.6%	(n=1414)							IDH1/2	1.9%	(n=421)
		6.1%	(n=446)							ET	0	(n=47)
	NK	14%	(n=358)								2.7%	(n=73)
										IDH1/2	0.8%	(n=594)
IDH2	AML	8.7%	(n=805)	MDS	4.6%	(n=65)	CMML	3.7%	(n=81)	MPN	0	(n=263)
		2%	(n=196)		3.6%	(n=82)				PV	0	(n=38)
		5%	(n=198)		2.1%	(n=439)				ET	0	(n=47)
		3%	(n=520)		9.4%	(n=277)				PMF	2.3%	(n=301)
		11%	(n=893)		3.7%	(n=108)						
		12.1%	(n=446)		0	(n=193)						
		10%	(n=1473)									
	NK	19%	(n=358)									
	NK	12.1%	(n=272)									
NPM1	AML	35%	(n=591)	MDS	4.4% (n=160)		CMML	1.2%	(n=81)			
		24.9%	(n=257)	low/int	2% (n=288)			4%	(n=50)			
		27.5%	(n=1485)	MDS+MDS/MPN	1.8% (n=493)		MDS/MPN	3%	(n=187)			
	NK	48%	(n=300)	MDS+CMML	1.5% (n=66)							
	NK	52.9%	(n=401)									
	NK	53%	(n=872)									
	s/tAML	7%	(n=140)									
	sAML	12.5%	(n=350)									

DNMT3A	AML	22.1%	(n=281)	MDS	7.8%	(n=51)	CMML	4%	(n=52)	MPN	0	(n=57)
		17.8%	(n=489)		8%	(n=150)				PV	3%	(n=33)
		14%	(n=500)		2.6%	(n=193)					7%	(n=30)
				low/int	13%	(n=288)				ET	0	(n=56)
				HR	6%	(n=100)					0	(n=30)
										PMF	4%	(n=25)
											7%	(n=46)
											6%	(n=16)
UTX							CMML	8%	(n=52)			
FLT3-ITD	AML	22.4%	(n=2813)	MDS	0	(n=32)	CMML	8%	(n=38)	MPN	0	(n=115)
		23%	(n=201)		4%	(n=70)				PMF	0	(n=40)
		13.2%	(n=106)		3%	(n=97)						
				MDS+MDS/MPN								
		27%	(n=854)		1.5%	(n=1316)						
		20.4%	(n=979)	MDS+CMML	0.6%	(n=182)						
		23.5%	(n=1003)	HR+CMML	2.5%	(n=198)						
		19.2%	(n=956)	HR	2.3%	(n=338)						
		24%	(n=250)	MDS+AML	8%	(n=92)						
		28.3%	(n=60)									
	elderly	34%	(n=140)									
	NK	31%	(n=872)									
	s/tAML	11.7%	(n=605)									
FLT3-TKD	AML	5.5%	(n=2357)	MDS	0.4%	(n=237)						
		7.7%	(n=979)		3.4%	(n=29)						
		11%	(n=1107)	MDS+MDS/MPN								
		4.8%	(n=3082)		0.5%	(n=1316)						
		7%	(n=429)									
		3.3%	(n=60)									
	NK	11%	(n=872)									
	s/tAML	1.8%	(n=489)									

CEBPA	AML	11%	(n=135)	MDS	1.5%	(n=68)	MDS/MPN	4%	(n=187)	MPN 0 (n=14)
		4.3%	(n=277)		4.5%	(n=382)				
		15%	(n=236)							
		15%	(n=104)							
		6.9%	(n=598)							
		7%	(n=1427)							
	NK	12.8%	(n=1182)							
	NK	13%	(n=872)							
WT1	AML	14%	(n=34)	MDS	0	(n=27)	MDS/MPN	1%	(n=187)	
		10%	(n=70)							
	NK	10.7%	(n=196)							
	NK	10%	(n=470)							
	NK	12.6%	(n-617)							
MLL-PTD	AML	6%	(n=2735)	MDS	2.6%	(n=338)				
		5%	(n=956)		4.4%	(n=180)				
		4%	(n=250)							
	NK	7%	(n=872)							
	s/tAML	5.9%	(n=591)							
ETV6	AML	1.7%	(n=300)	MDS	2.7%	(n=439)				
				low/int	2%	(n=288)				
BCOR1	AML	6%	(n=173)							
	NK	3.8%	(n=262)							
GATA1										ET 0 (n=46)
GATA2	AML	0	(n=270)	MDS	0	(n=30)				MPN 0 (n=286)
	CEBPAmut	39%	(n=33)							
	CEBPA WT	0	(n=89)							
FMS				MDS+AML+CMML						PMF 5% (n=40)
					12.7%	(n=110)				
GNAS				MDS	0.7%	(n=439)				

BRAF		MDS 0.5% (n=439)		
PTEN		MDS 0.2% (n=439)		
CDKN2A		MDS 0.2% (n=439)		
ZRSR2	AML 0 (n=151)	MDS 3.1% (n=193) 11.1% (n=221) MDS-RS 1.4% (n=73) w/o RS 7.7% (n=155)	MDS/MPN 8% (n=88)	MPN 1.9% (n=53)

Abbreviations; NK, normal karyotype; int, intermediate; HR, high risk; RS, ring sideroblasts; sAML, secondary acute myeloid leukaemia; tAML, therapy related acute myeloid leukaemia

(Abdel-Wahab et al., 2011a; Bacher et al., 2010, 2007; Braun et al., 2011; Ernst et al., 2010a, 2010b; Gelsi-Boyer et al., 2008; Grossmann et al., 2011a; Jankowska et al., 2011; Kohlmann et al., 2010; Makishima et al., 2010; Malcovati et al., 2009; Muramatsu et al., 2012; Szpurka et al., 2010)(King-Underwood et al., 1996)(Kiyoi et al., 1999)(Abu-Duhier et al., 2000)(Stirewalt et al., 2001)(Thiede et al., 2002)(Kottaridis et al., 2001)(Schnittger et al., 2002)(Preudhomme et al., 2002)(Barjesteh van Waalwijk van Doorn-Khosrovani et al., 2003)(Fröhling et al., 2004)(Falini et al., 2005)(Lin et al., 2005)(Nomdedéu et al., 2005)(Hugues et al., 2005)(Barjesteh van Waalwijk van Doorn-Khosrovani et al., 2005)(Olesen et al., 2005)(Bowen et al., 2005)(Suzuki et al., 2005) (Döhner et al., 2005) (Schnittger et al., 2005)(Lee et al., 2006) (Bacher et al., 2006)(Thiede et al., 2006)(Illmer et al., 2007)(Summers et al., 2007)(Mead et al., 2007)(Vicente et al., 2007)(Bacher et al., 2008)(Andersen et al., 2008)(Hou et al., 2008)(Schlenk et al., 2008)(Haferlach et al., 2008)(Abbas et al., 2008)(Paschka et al., 2008)(Virappane et al., 2008)(Wouters et al., 2009)(Hussein et al., 2009)(Gaidzik et al., 2009)(Mardis et al., 2009)(Tang et al., 2009)(Carbuccia et al., 2010)(Chou et al., 2010a)(Marcucci et al., 2010)(Green et al., 2010b)(Thol et al., 2010a)(Paschka et al., 2010)(Boissel et al., 2010)(Abbas et al., 2010)(Green et al., 2010a)(Chou et al., 2010b)(Ley et al., 2010)(Schnittger et al., 2010)(Patel et al., 2011)(Chou et al., 2011b)(Taskesen et al., 2011)(Gaidzik et al., 2011)(Metzeler et al., 2011b)(Green et al., 2011)(Thol et al., 2011a)(Chou et al., 2011a)(Li et al., 2011)(Grossmann et al., 2011b)(Metzeler et al., 2011a)(Qian et al., 2012)(Hou et al., 2012)(Pratcorona et al., 2012)(Rücker et al., 2012)(Gaidzik et al., 2012)(Greif et al., 2012)(Schnittger et al., 2012b)(Ridge et al., 1990)(Kaneko et al., 1995; Misawa and Horiike, 1996)(Misawa et al., 1997)(Horiike et al., 1997)(Hosoya et al., 1998)(Tang et al., 1998)(Xu et al., 1999)(Yamamoto et al., 2001)(Fritsche-Polanz et al., 2001)(Kita-Sasai et al., 2001)(Kaeferstein et al., 2003)(Johan et al., 2004; Shih et al., 2004)(Nakao et al., 2004)(Watkins et al., 2004)(Christiansen et al., 2007)(Steensma et al., 2005)(Chen et al., 2007a)(Fuchs et al., 2009)(Patnaik et al., 2010; Rocquain et al., 2010)(Andrulis et al., 2010a)(Thol et al., 2010b)(Hussein et al., 2010b)(Lin et al., 2011)(Bains et al., 2011)(Ewalt et al., 2011)(Schnittger et al., 2011)(Jädersten et al., 2011)(Thol et al., 2011b)(Walter et al., 2011)(Bejar et al., 2011)(Papaemmanuil et al.,

2011)(Thol et al., 2011c)(Malcovati et al., 2011)(Graubert et al., 2012)(Patnaik et al., 2012a)(Patnaik et al., 2012b)(Lin et al., 2012)(Wu et al., 2012)(Thol et al., 2012)(Cui et al., 2012)(Abu-Duhier et al., 2002)(Abu-Duhier et al., 2003)(Pardanani et al., 2003)(Gandini et al., 2004)(Baxter et al., 2005)(Levine et al., 2005)(James et al., 2005)(Kralovics et al., 2005)(Jones et al., 2005)(Wolanskyj et al., 2005)(Tefferi et al., 2005)(Vizmanos et al., 2006)(Campbell et al., 2006)(Pikman et al., 2006)(Oki et al., 2006)(Fontalba et al., 2006)(Pardanani et al., 2006)(Chen et al., 2007b)(Martínez-Avilés et al., 2007)(Percy et al., 2007)(Barosi et al., 2007)(Pietra et al., 2008)(Beer et al., 2008)(Chaligné et al., 2008)(Ormazábal et al., 2008)(Tefferi et al., 2009)(Delhommeau et al., 2009)(Grand et al., 2009)(Zhang et al., 2009)(Carbuccia et al., 2009)(Ding et al., 2009)(Hussein et al., 2010a)(Ruan et al., 2010)(Boyd et al., 2010)(Glembotsky et al., 2010)(Pardanani et al., 2010)(Tefferi et al., 2010)(Andrulis et al., 2010b)(Ma et al., 2011)(Pietra et al., 2011)(Brecqueville et al., 2011)(Abdel-Wahab et al., 2011b)(Stegelmann et al., 2011)(Stein et al., 2011)(Raza et al., 2011)(Guglielmelli et al., 2011)(Pardanani et al., 2011)(Ricci et al., 2012)(Tefferi et al., 2012)(Lasho et al., 2012a)(Schnittger et al., 2012a)(Aranaz et al., 2012)(Meggendorfer et al., 2012)(Lasho et al., 2012b)(Wang and Chen, 1998)(Au et al., 2004)(Yoshida et al., 2011)(Boultwood et al., 2010; Dicker et al., 2010)(Bejar et al., 2012)(Makishima et al., 2012)(Damm et al., 2012)(Ernst et al., 2010a)

Appendix 8.6

Primer Pairs for GS Junior and MiSeq

Appendix 8.6 - Primer pairs for GS Junior Panel

Gene	Chr	Forward Primer	Reverse Primer
ASXL1	20	ACTCACACAGTCCCACCAGAA	GAGCACGGGCTTTAATGTCTGC
ASXL1	20	ACCAGATATGCCCCCGGAT	TCTCTCCTAGCTCTGGACATGG
ASXL1	20	AACTACTGCCGCCTTATCCTCTA	TCCTTTCTCTAATGTATCATCTCCCA
ASXL1	20	AATGTGAGTCTGGCACCACTTC	ATTCTGGTTTGGGCTGTTTCACT
ASXL1	20	AGGCTCTCGTTTCTAACAGTTCT	GCTGTCCTCCGTGAGGTG
ASXL1	20	TCAACAGGTGGACATTGAAAAGC	GACCCAGGCATGGACAGG
ASXL1	20	TACTGAGTACCAGCCAAGAGCC	GAGTCAAAACTTGGGACTGCCTT
ASXL1	20	CACTGGTCTTGCCAGGATTGA	CACATTCCCAGAGCCAAAAAGC
ASXL1	20	GGTGATCAGAGCAATGTTACAGG	TCTCGGGGTAATTTCCAGAAGG
ASXL1	20	AAATGCCGAGAACAGGAAAGCTA	GTGCTGCTGTCAGTGAACATTTG
ASXL1	20	CTTCTGGAAATTACCCCGAGAGC	GCTTTGAGGGTCCAATACAGTCA
ASXL1	20	AGTGCATCACTTTCCTTGCAAAT	GGCAGCCCAATGGGTTATAC
CBL	11	CCCTGGAGCTTAAAATAGGACCC	CACCCCTTGTATCAGTAAAGGCTA
CBL	11	CCTGGCTTTTGGGGTTAGGTTTA	TGTTTTACGGCTTTAGAAGACAAC
DNMT3A	2	CAGAGTTCCCAGGCAACAA	AGGTGGGAACAAGTTGGAGA
DNMT3A	2	CTCTCCTGGGTGGGTGTG	CCCTCCTGAGTAACCTGGAA
DNMT3A	2	CCCAGCTAAGGAGACCACTG	GCTGCTCTTTGGTTCTGTCC
DNMT3A	2	GCACAACAGGTCAGATGCAG	TCTGCTCACTGGGTCTCCTT
DNMT3A	2	TCTGCTCACTGGGTCTCCTT	TTCGCTAATAACCACGACCA
DNMT3A	2	GTGCAGGGAGGGGAAGAC	CTCTCCGGTCAGTTTTCTGC
DNMT3A	2	ATTAGTGAGCTGGCCAAACC	GCCTTTATCCTCCCAGATCC
DNMT3A	2	TCATCCCACCTGCAGTCC	CAGCTTGTGGAATGTGGCTA
DNMT3A	2	CCCAGCAGAGGTTCTAGACG	TAACGCTTGGTGGATTTGTG
DNMT3A	2	AGCAAGCACAGCAATCAGAA	GCTCACTCAGAGCCATACCC
DNMT3A	2	TTTGTGTCGCTACCTCAGTTTG	CTGAGTGCCGGGTTGTTTAT
FLT3TKD	13	CTGCTGTGAGGGTTTTTTGATGT	GGCACAGCCCAGTAAAGATAAGA
IDH1	2	AAATCACATTATTGCCAACATGAC	AGAATCGTGATGCCACCAAC
IDH2	15	GTGCCCAGGTCAGTGGAT	GACTCCAGAGCCCACACATT
JAK2	9	TTACTCCTCTTTGGAGCAATTCA	AACATCTAACACAAGGTTGGCAT
JAK2	9	GGACCAAAGCACATTGTATCCTCA	TCACAAGATATAACTGAATAGTCCTACAG
KIT	4	AGTGAAGTGAATGTTGCTGAGGT	TCAGCAAACAAAATTAATGTCTACCA
KIT	4	ATGTGAACATCATTCAAGGCGTA	TGAAACTAAAAATCCTTTGCAGGAC
KRAS	12	AGGAAAGTAAAGTTCCCATATTAATGGTT	TTGTATTAAAAGGTACTGGTGGAGT
KRAS	12	ACAGGGATATTACCTACCTCATAAACA	TTTTGAAGTAAAAGGTGCACTGTAAT
MPL	1	GGGGCCCTGACCTTGC	GAAGTGGCGAAGCCGTAGG
NPM1	5	ACTCTCTGGTGGTAGAATGAAAAA	AAACAGGCATTTTGGACAACACA
NRAS	1	TGGGTAAAGATGATCCGACAAGTG	GAGGCCGATATTAATCCGGTGTT
NRAS	1	ACAACCTAAAACCAACTCTTCCCA	AAAATTGAACTTCCCTCCCTCCC
TP53	17	AACCAGCCCTGTCGTCTCT	GTTTCTTTGCTGCCGTCTTC
TP53	17	GCCACTGACAACCACCCTTA	CATGAGCGCTGCTCAGATAG
TP53	17	GGGTCAGAGGCAAGCAGA	GAGCTTGCAGTGAGCTGAGA
TP53	17	TGCTAGGAAAGAGGCAAGGA	CAAGGGTGGTTGGGAGTAGA
TP53	17	CCACTTGATAAGAGGTCCCAAG	CTTTGAGGTGCGTGTTTGTG
RUNX1	21	GGGCCCCTTTCCAGAATCC	GTGATGCGTATCCCCGTAGATG
RUNX1	21	GGGATTCCATCACAGAAATCACT	TCATTGCTATTCCTCTGCAACCTAA
RUNX1	21	CAATGCAACTTTTTGGCTTTACGG	AACTGGTAACTTGTGCTGAAGGG

RUNX1	21	CAGTTGGTCTGGGAAGGTGTG	AAGCCCCAGTTTTAGGAAATCCA
RUNX1	21	TCAATAATGTTCTGCCAACTCCTTC	TGAACAAGGGCCACTCATTTCTT
RUNX1	21	CGGTGGAGGCGTTGGTG	CTCACTTCCGCTCCGTTCTCTT
RUNX1	21	GCGCGGGCCTGACCTA	CGGCTCCTACCAGTTCTCCAT
SF3B1	2	ACAAAGTTACATTACAACTTACCATGTTC	CCAACTCATGACTGTCCTTTCTT
SF3B1	2	GATAAATCAAAAGGTAATTGGTGGA	TGCTGACAGGCTATGGTTCA
SF3B1	2	TGTTAGAACCATGAAACATATCCAG	TATCCGCCAACACAGAGGA
SRSF2	17	TTAGGGTTATGTGTCTCGGATTC	CGCTGCCTGGAATTAACC
SRSF2	17	CCTCAGCCCCGTTTACCT	CAAGGTGGACAACCTGACCT
SRSF2	17	ATGGCATCCATAGCGTCCT	CAAGGCCTTTCCCAGTGTC
U2AF1	21	GAACACACTTATGAACACAAATGGA	GAGGTGCTTAATACCACGGAAA
U2AF1	21	CTCGTGTGCATTCTCTGTGG	ATTAAAGCGTGGATGGCAAG
WT1	11	AGCAGTGCTTACTTTCCATCCTG	CCCTCAAGACCTACGTGAATGTT
WT1	11	TTCCACCCTCCCCTTCTTTA	TGCAGGCATGGCAGGAAA
ZRSR2	X	TCCTTTCATTGGGCACAGA	TTTGACTTCCTTCTGACACTGG
ZRSR2	X	AAGGTTGATCAGAGACCTTTTTG	GACTGGTACTGGTTAGTAAAGGTTGA
ZRSR2	X	TTTGCTCTCGTGTGTGTGTG	CCTCCCAAGATAGGCAACTCT
ZRSR2	X	TGTGCGCTGTATGTGAAATG	GACCCGAAGAAGAGCATCAG
ZRSR2	X	AACTTGTGTGCGTGTGTGTG	TGAGTCCAGATATCCAAACATGA
ZRSR2	X	TGAAACATTTCGTCTTTCATGG	GCAGCAGATGTCTGATGACG
ZRSR2	X	TTTCAACTATTGGCCTAGTGAATTT	CAGGGTGAAGCAAACCAGTC
ZRSR2	X	GGGAATGTTAGCCTGGACAA	TCCCAGACATCAACAGAAACA
ZRSR2	X	CGGGGTTAATTAATAGTAGAGCTAATC	AAGCTATGCCTCACCACTTGA
ZRSR2	X	AAGTGCTGTTTCATCACTGTGC	CCTCTCCCGACTCTTTGATG
ZRSR2	X	CATCAAAGAGTCGGGAGAGG	AACCCATCTGCGTTCATAGC

Appendix 8.6 - Primer pairs for MiSeq panel

Gene	Chr	Forward Primer	Reverse Primer
ASXL1	20	CCCTAGGTCAGATCACCCAGTC	GGCTTTAATGTCTGCGAGGGTC
ASXL1	20	GTGGTTAAAGGTCAGCCCACTT	CCCCCTCCGATGGCAGT
ASXL1	20	CGAGGGGCGAGAGGTCA	GATCTGACGTACACTTTCCAGGG
ASXL1	20	CCACCGATGAGGGAGGTGG	CTCTCTCCTAGCTCTGGACATGG
ASXL1	20	TACGTCAGATCTACAGCGAACAC	TTGGGAGGCATCTCCTAGCC
ASXL1	20	AGGAGGAAAGCTGCCTACTACA	GCCAGACTCACATTCAGTTCTAA
ASXL1	20	CCACTGGGGACCAGCCAT	TCCCACTAGAGACAGAATGGGAC
ASXL1	20	TTCCTGGGAAAGTGATGATGAGG	CTCTGTTCTGCAGGCAATCAGTC
ASXL1	20	GAGAAAGGAACTGGCCAAGC	AGAGCCTTGGTTTTCAAGTTTTC
ASXL1	20	CCTGAATCCTCACCGACTGATTG	ATTCTGGTTTGGGCTGTTTCACT
ASXL1	20	AGGCTCTCGTTTCTAACAGTTCT	TAGACCCTCCTCAGCTGTCAAAT
ASXL1	20	GATGAGGTAGTGAAACAGCCCAA	CTCCTCGAGATGGCACAGTC
ASXL1	20	ATTTGACAGCTGAGGAGGGTCTA	TTCAAAGTCAGAGGCTGTATCCG
ASXL1	20	AATCAACGGAGACTCTGAAGCAC	ATGTCACCATTCACCTTGGACAG
ASXL1	20	CACTAGAGAAGCTGCAGTGACAA	AGCAGTAGGGAATCTGGGATCTT
ASXL1	20	GTCTGGTTACAAGGACAGATGGG	CAGCAACTGCATCACAAGTGG
ASXL1	20	AAGAGCCGTGTGCCTGTC	ACCATGTAAAGATCCCATGCGTA
ASXL1	20	TAGCTTGCCCCTAGAGAAGGTT	AAGAGGCTCCTTCAAAGCCCTTA
ASXL1	20	GGAAAAAACAGTGGCATGGTTGA	ATGGGATTTGTCACTGGATGGAG
ASXL1	20	CTGTGAAACAGGCACTGGTCTTG	ACGGACTTCCTTCTGATCTTCAC
ASXL1	20	CCTCTAGGAAACTGGAAGAAATGGA	GCTCTGGACCAAAGGAGATCACA
ASXL1	20	AGATCAGAAGGAAGTCCGTGCTA	TTCTTCCCTTGGCCTGTAACATT
ASXL1	20	CTTTGGTCCAGAGCAGACAGG	CACCAGACATGGAGTTTGTGCTT
ASXL1	20	CCCTTCAGCGCCCCAG	CTTAAGAGGACCCCCCACAAAAG
ASXL1	20	GGGGTACAGACTCCAAGGGAAG	AATTTCCAGAAGGGCAAGTCCAT
ASXL1	20	ATAGTCCCCTGGAACTGGTGG	TGGTGGAACTCAGTTGGAGTTTA
ASXL1	20	TTCTTCTCTCCCCTCCCAACTCA	GAACATTTGCAAGGAAAGTGATGC
ASXL1	20	CTTGCTGGAAGTGTGGTGCAG	GCTTTGAGGGTCCAATACAGTCA
ASXL1	20	GCTGCGGTGCGTTCTGTC	AGAGTGCTCCTGCCTAAAGAGTA
BCOR	X	TGACACATATGCACAAGGATTAACA	GGAAGCCTTCAACCCTGAAAGTA
BCOR	X	CCACTCTACAGAGGAGCCCAG	ATTTCGCTGCAATTTTCCAAACG
BCOR	X	GAACAAGAGACTTGCAGAAACCT	GGGCGCACTTTTCATTTTACTGG
BCOR	X	CTAACTCCTGTCACCTCAAGAGC	GATGTTTTAGCCAACCCCCCAG
BCOR	X	CTGAGCCACAGATACTTGGATGT	CCTTCCTCTCTTATTTGCCCTCC
BCOR	X	AGACAGAAATGGACTTGAACTTGT	CCATCATTTTGTTCTCAGCAGTAG
BCOR	X	AGCCTTTGTGATTTCTGCCTAGT	CTTCTCTCTTATGGTGCTGACCC
BCOR	X	TGTCATACCTGTTAAGAACTTTTCCA	AAGTGTGGCTTGCATGAGCATAA
BCOR	X	ACAGATAGGGAAGCTTGGTCTCA	TAAATCATCGGGACAACGCAGGT
BCOR	X	AGTTGACATCAGCGCCATATTCA	AGTGATCTGATGGCATTCTTCCC
BCOR	X	CCTCGCCCACCACAGTC	CCATGCTCAAGTTCCCCTCA
BCOR	X	GGTCTCCTGAGGGGAACTTGAG	GTGTGCGCTGCTTGGTTTTATG
BCOR	X	GGGGATGTGTTCGCACAGG	CTGCAGAAATACACCGACAACAG
BCOR	X	TCTCGCTGTTGTCGGTGTATTTC	AAGAGTCTTTCATCCACCAGTGC
BCOR	X	GCAATCCTCTTCTTCGTCTGCAC	CCGTCACCCCCTCCGTA
BCOR	X	GATCTCCTCAAAAGCCCTTTCCT	TTCCTCAGGGGACTAACATCACT
BCOR	X	GATGTTAGTCCCCTGAGGAATGG	TATAGTGAGCTGACCAACCTGAA
BCOR	X	CAGCAAGTGGCGTTGTTTTTT	ATGCGTAAGACAGTTTGTTCCC

BCOR	X	CCGCACATCCACATCTCCTG	GTGGACAGGAAACGCAAAGTC
BCOR	X	TTGCGTTTCCTGTCCACCC	TAATGCCCTCTTGTCTCCCCTC
BCOR	X	CCATTTCTCCAAGCAGATGCCAA	CAGCAACCAAAGACTCCGAGATG
BCOR	X	GACCGACTTTGGCTTTTTGTCC	AGCTTTTATGGTAAGGTGTTGTTTT
BCOR	X	ATGGCCCACAAACTTCCCTTT	ATGATCCATTGAGTGCCCATTTT
BCOR	X	TACCATACTCCCCCAATCCTGTT	GCTGGCAAAGAGAATCGCCAA
BCOR	X	GCCATCATTTGATTCAGCCTCAT	AAGAGAACCTAGGGTTGCCAGTC
BCOR	X	CTCTTTGTTGGTACCTGCCAGAA	CTTTGCAGCAGAGAGTGTTGG
BCOR	X	CTCCTCTCTCAGGGCGATGAAAT	GACTGTTGTCAAAAGCGACAAGC
BCOR	X	GTCTTCCCTTGATTCCAGTTGGG	CACACACGCCCATAGAGATTACT
BCOR	X	CTCGTAACGGGCTCTCTCATGG	GTCTGTTTCCTGGGCACCTT
BCOR	X	TGGTATCAACATGGGATGCAC	TTGCTGTAAGTCCCCTCTCCTTA
BCOR	X	AAACAGACTGCCATTGGGTAACA	GGCCTTCCACCAAGCTCTATATT
BCOR	X	CCGTTCTCGTTTGCTTTGAAACT	CATCACCCGCCCCCAAT
BCOR	X	GATGGTGTGGTTTCTACAGAGC	GAAAAACAAGGCATTGGACTGGG
BCOR	X	AGTCCAATGCCTTGTTTTTCAGC	CTGGAAGTGGCTTAGTGCTCTC
BCOR	X	GAGAGCACTAAGCCACTTCCAG	ACAGCGGTTCAAGACAGAAAAGA
BCOR	X	TCTTTGGTAACGGTCTGCTTCTC	GTTGCCCTGTCAAAGCCATACAT
BCOR	X	GAGCCTTGGGATACTTGCCATT	TCTCCTGTTGCCCCCCTC
BCOR	X	CAACAGGAGAGCTGTGTCCC	CGGCCTCCCCAGCCAT
BCOR	X	GATAGGCGTGGGAATCAACAGGA	CAATGGGGAGCGCTTTCTCTA
BCOR	X	ATTGGTGCAGACTGGAGAATACA	ACAAACAGAGCCCTCTCAACATC
BCOR	X	GCGTGGCACCCTCCAT	GAAAACCCCCAAATGGCTTCAG
BCOR	X	GCAGATATGGCATCAACAGAAGC	TCTGTTCTGGGGAGATCTGTGAA
BCOR	X	CGGTGGCTGTGAGAAGTTGAG	CTTGAAGCAAAGCTGCCATCC
CALR	19	GTAACAAAGGTGAGGCCTGGT	CCTCCTCCTCTTTGCGTTTCTTG
CALR	19	AATGAAGGACAAACAGGACGAGG	AGGCAGGCCTCTCTACAGC
CALR	19	ATGAGGAGGACAAGGAGGAAGAT	CAAAATCCACCCCAAATCCGAAC
CBL	11	CCCAGACTAGATGCTTTCTGGTT	ATGTGGAGCCCATCTCACAGTAT
CBL	11	TGAATTATACTGTGAGATGGGCTCC	AGGTTATTACATAGCTGAAAAAAGTCG
CBL	11	TGCATCTGTTACTATCTTTTGCTTCT	CTCTGCTCCTTGCCTCAACAG
CBL	11	CCCATCGTGGTAGATCCGTTTG	ACAATGGATTTTGCCAGTCTCCT
CSF3R	1	CCTCCATCCCATGGACCC	CTATCTCCGCTGTGACTCCACT
CSF3R	1	CTGGAACCAGAGGTTCTCATAGG	GGACCCAAGAGCAGTTTCCAC
CSF3R	1	GACCAGGGGATTCAAAGTCAGTC	CCCGCCAGTCTGTATCACATC
CSF3R	1	CTTACCTGGGGTCAAGGTCATCA	GGAGGCAGCTTTACCATCCAG
CSF3R	1	GCAGTGCAAGGAAATTCCCAATA	GGTCTTAAACAACCCTTCTGCCTA
DNMT3A	2	ATGTTTTGTGTTTTTTGTTTGTTTGTTT	CAGTCCACTATACTGACGTCTCC
DNMT3A	2	CCCCATGTCCCTTACACACAC	CTGCCCTCTCTGCCTTTTCTC
DNMT3A	2	TGGAAAACAAGTCAGGTGGGAAA	AACTCCATAAAGCAGGGCAAAGA
DNMT3A	2	TACCTTTCCATTTCAGTGCACCA	GCATATTTGGTAGACGCATGACC
DNMT3A	2	CTCATCCTGCCCTTCCTTCTC	CTTCCCGCTGTTATCCAGGTTTC
DNMT3A	2	CTCAGGGGCTTCCCCACTAT	TCATCTTCAAACCGTCTCCTGTT
DNMT3A	2	GATGAAGCAGCAGTCCAAGGTA	TCTTTGAGTTCTACCGCCTCCT
DNMT3A	2	AGATGTCCCTCTTGTCACTAACG	CACACCACTGTCCTATGCAGAC
DNMT3A	2	CAGGGCAGAAATATCCAAGGAGG	GCCCATCACGTTGCCTTTATC
DNMT3A	2	AACAAAATGAAAGGAGGCAAGGG	CTCGGAGGTGTGTGAGGACT
DNMT3A	2	TGGACATACATGCTTCTGTGTGA	GGAGATGGCTCCAAGTAACGG
DNMT3A	2	CATTTCGTTTTGCCAGAGTTGCC	CATCTGACCTGTTGTGCTCACT
DNMT3A	2	CTCAGGCCCCACAACCAA	ATTAAGGAAGACCCCTGGAACTG
DNMT3A	2	ATTAGCGAAGAACATCTGGAGCC	CTCCTCTGCTCACTGGGTCT
DNMT3A	2	CCAGCTAAGGAGACCACTGGAG	CCTGGTGGTTTCTGACCCTTC

DNMT3A	2	GTGGACACAGTCAGCCAGAAG	GGTACTCACCCCATCCCCTC
DNMT3A	2	TGGTCCCATGTCATTCAAACCTT	AGGAGGCCTGCATCCGAG
DNMT3A	2	CTCTGCAAGGGGAGGAGAGC	CAGCTGCCTACGCACCAC
DNMT3A	2	TGTGCGCTCATCAATAATCTCCT	CTGTCAGCCTGTAACTGACCTTG
DNMT3A	2	CCCTAAGCATGGCTTTCCCAG	CCACGACAGCGATGAGAGTG
DNMT3A	2	GTGCCCTCATTTACCTTCTGGTG	CCCAACCCTGGCGTGTC
DNMT3A	2	ACTTCCAGGCCTCCTAGTG	GCTTGTCCCCCCAGGT
DNMT3A	2	GGTGGAACGCACTGCAAAA	TATCACTGTATCTGGTCCCCTCC
DNMT3A	2	GGATCAAGAACCTTCCCCCAC	GCTTTGGCATTGGGGAGCTG
DNMT3A	2	TACCACTGAGAATTTGCCGTCTC	GTGTAATGATTTCTGCTCCTTGGG
DNMT3A	2	ATGGAGAGAGGAGAGCAGGAC	GCATCCCCCACTGTGGCTA
DNMT3A	2	ACCCCACGGGCTCAGG	GGAGAGGTCAAGGTGACTTTTTG
DNMT3A	2	TCTGCCATCCCACCAACAAATTA	TGGTTCTCAGAGCTAAGTATCAAGG
DNMT3A	2	GGGGAGGCATACTTCACTCTTTT	AGCTACTTCCAGAGCTTCAGGG
DNMT3A	2	CCCATCACTTCTGGTTTTCCAGT	GTGGCTTGGGCTGGGAG
DNMT3A	2	GCCTCACCTCCCTTTTCCAG	ACCCTAATGCCCTAATGTCTGTC
DNMT3A	2	CCACAGAGGGATGTGTAAAGAAGG	GAACTGGTCCCTTTGTTCTTCCC
DNMT3A	2	CTTTAGCCACGACCCAGACCAT	AAGCCTCAAGAGCAGTGGAAAAT
DNMT3A	2	TGAGGACTCACCCGCTTCT	TATTACCCAATGGGGACTTGGAG
DNMT3A	2	CAGGGCTCCCCTCCTCTG	CCCACTGATCTAACCCTCCTCTT
DNMT3A	2	CGTGTGTGTTGTGTGTGTGC	GCTGGGATCCACCTCTGG
DNMT3A	2	ACAGGGCTCTCCCTCTCC	ATAATTCCTTCCCCAAAGCCCAG
EZH2	7	CGAGGTTCCTGAAGCTAAGGC	AGCAGGCTTTGTTGTGTTAAGTC
EZH2	7	TCACAATCCAGTAGAAAAAGCCCTTA	GCCCGTCTTCATGCTCACT
EZH2	7	TCCCAGAAGTATTCAAGTCCATCA	TGAAATTATTCACTGGGCTGTGC
EZH2	7	AGAAGGCTGCCACATGCAA	TTCTGTCAGGCTTGATCACCTTT
EZH2	7	CCAATCAAACCCACAGACTTACC	ATCTATTGCTGGCACCATCTGAC
EZH2	7	GGTTATCAGTGCCTTACCTCTCC	AGTCTACTTTGTCCCCAGTCCAT
EZH2	7	AAGGCAATCCTGACATTTGCATC	CTACCTGGCTGTCCGAGAGTG
EZH2	7	GCAGTTCTTGCAGGACACATTTT	GGTGATGAGTGAAGAACCTCCAA
EZH2	7	CGTCAAGTAAACAAGGGAGTGCT	CCACGGCAGCCTTGTGA
EZH2	7	TGCATCAAAGCAACAAATACTTACACT	GATTCATTGGCCTGCATGATGTT
EZH2	7	ACCCAAGCTCTAATCCAGTTACTATTC	GTAGCTTCCCGCAGAAATTTGGT
EZH2	7	CCTTGCCTGCAGTGTCTATCTAT	ATAAGACTGTCCTCATGGCTCTG
EZH2	7	TGTTTGGACAACGAGTACAGTTT	GAACCTCCTGAGAATGTGGAGTG
EZH2	7	ACCTGTCTACATGTTTTGGTCCC	TGACCAGTGCTTACATTTGGTTT
EZH2	7	TGCATTATACATCCTTAATCCTCACAA	GCTGGAATCAAAGGATACAGACAG
EZH2	7	TAGACGTGTCTTACCAGAGGAGC	AGAAGAGGACGGCTTCCCAATAA
EZH2	7	TCCCTATCACTGTCTGTATCCTT	ACTTTGCCCTGATGTTGACATTT
EZH2	7	ACCCTCTGCAATAATTAGGCACT	CCAGTGGAACTGGAAGAGTGAAA
EZH2	7	AAATGATAGCACTCTCCAAGCTG	CCCAACATAGATGGACCAAATGC
EZH2	7	AACATCGCCTACAGAAAAGCGTA	TTTCAAAAAAGTGATTTTTGTTTCATGTTT
EZH2	7	TCCGCTACATTGATTCCATTTGT	TTAGATAAAGAAAGCCGCCCACC
EZH2	7	AGTTCTTCTGCTGTGCCCTTATC	TCAGCTTTGTTATAGAGACATAATTGGG
EZH2	7	AAACTAAGCCCTATTTCTACTCTTTCT	TTTTTGTGGAGTTGGTGAATGCC
EZH2	7	TCTCGGTGATCCTCCAGATCTTT	ATTGCTTCTCCTGTGTGTTTCTG
EZH2	7	TGCTTCATAAACAAAAGTGTCTCTCA	TTTCTTTTAGGTGGAAGATGAAACTG
EZH2	7	AGTTCTTCAATGAAAGTACCATCCTG	ATCTGGAGAACTGGGTAAAGACA
EZH2	7	ACTGTCTTGATTCACCTTGACAATA	GGATTTTCCAACACAAGTCATCCC
EZH2	7	ACCATAAAATTCTGCTGTAGGGGAG	ATGGCTACAGCTTAAGGTTGTCCT
EZH2	7	TGATTTCCTCCCAATAACCAAACA	TGGAAAGAACGGAAATCTTAAACCA
EZH2	7	ATGTTAACCAACCTCCCTAGTCC	ACAATTTCTCCTTTCCTCTCCTTCA

EZH2	7	AGGAGGGGAAAAAACCTATCCTT	ATAATCATGGGCCAGACTGGGAA
EZH2	7	TGAACCTCTTGAGCTGTCTCAGT	CTGATTGTTAGTTTGCTGCGGAT
FLT3	13	ACCACAGTGAGTGCAGTTGTTTA	AGGAACGTGCTTGTCACCC
FLT3	13	CTCACATTGCCCCTGACAACATA	TGCACTCCAGGATAATACACATCA
IDH1	2	CCTTGCTTAATGGGTGTAGATACCA	AGCCATTATCTGCAAAAATATCCC
IDH1	2	GCCAACATGACTTACTTGATCCC	CCTGATGAGAAGAGGGTTGAGGA
IDH1	2	TGCATTTCTCAATTTCATACCTTGCT	GGCTTGTGAGTGGATGGGTAAAA
IDH1	2	CCTTGCTTAATGGGTGTAGATACCA	AAAAATATCCCCCGGCTTGTGAG
IDH1	2	GCAAAATCACATTATTGCCAACATGAC	TGAAACAAATGTGGAAATCACCAA
IDH2	15	TGAAGAGACAAGCTGGGAGAGG	GCAGTACAAGGCCACAGACTT
IDH2	15	CCTCACCTCGTCGGTGTTGTA	CCTCGCCTAGCCATCCTCTT
JAK2	9	AAAAGAACAATTAGGAGTTATTAAGCATTTC	TAGGCCTCTGTAATGTTGGTGAG
JAK2	9	TCTAGTCTTCAGAACGAATGGTGT	TCACATGAATGTAAATCAAGAAAACAGA
JAK2	9	TCTGAACTATTTATGGACAACAGTCA	AAAACAGATGCTCTGAGAAAGGC
KIT	4	AGTGAAGTGAATGTTGCTGAGGT	CTCTGCTCAGTTCCTGGACAAAA
KIT	4	GGCCATTTCTGTTTTCCTGTAGC	TCAAGTGAATTGCAGTCCTTCCC
KIT	4	AATGGTTTTCTTTTCTCCTCCAACC	CTGTCAAGCAGAGAATGGGTACT
KRAS	12	AAACTATAATTACTCCTTAATGTCAGCTTATT	AAACCTGTCTCTTGGATATTCTCG
KRAS	12	ATGGTCCTGCACCAGTAATATGC	ATTATTTTTATTATAAGGCCTGCTGAAAATG
KRAS	12	TTAGCTGTATCGTCAAGGCACTC	AGGTGAGTTTGTATTAAAAGGTACTGGT
MPL	1	GGCTGGCTGGATGAGGG	GTCCACCGCCAGTCTCCT
NPM1	5	TGTCTATGAAGTGTTGTGGTTCC	AGGGAAAGTTCTCACTCTGCATT
NRAS	1	AGCTCTATCTTCCCTAGTGTGGT	GTACAGTGCCATGAGAGACCAAT
NRAS	1	GATGGCAAATACACAGAGGAAGC	AAAATTGAACTTCCCTCCCTCCC
NRAS	1	GACAAGTGAGAGACAGGATCAGG	AACAGGTTCTTGCTGGTGTGAAA
NRAS	1	TGGTTCTGGATTAGCTGGATTGT	GGAAGGTCACACTAGGGTTTTCA
RUNX1	21	CTCAGCTGCAAAGAATGTGTTTT	CAGATACAAGGCAGATCCAACCA
RUNX1	21	ATTGGTAGGACTGATCGTAGGAC	AACCCTGGTACATAGGCCACATA
RUNX1	21	GCCTTCCTCATAACGTGCATTCT	TTTTCGAACTTTCTCCCTGGTCA
RUNX1	21	GGCCAGTTGTGGGTGGT	CGGCGCACAGCCATGA
RUNX1	21	GGTTAAAGGCAGTGGAGTGGTTC	CCTCCCTGCTCCCCAC
RUNX1	21	TGTACCAGCCCCAAGTGGAT	ACAGATATGTTCAGGCCACCAAC
RUNX1	21	TTGAAATGTGGGTTTGTTGCCAT	CAGATGGCACTCTGGTCACT
RUNX1	21	TAACGTACCTCTTCCACTTCGAC	CTGCATTTGTCCTTTGACTGGTG
RUNX1	21	CCCTCGCGGATCTCCC	CGCAGCATGGTGGAGG
RUNX1	21	CAGAGGAAGTTGGGGCTGTC	CTTCACGCCGCCTTCCA
RUNX1	21	CCAGCACCTCCACCATGC	TTTGCAGGGTCCTAACTCAATCG
SETBP1	18	GGGTGGGGTCTGGATATTGGTA	GTTCCGGGGGCATGAGAAG
SETBP1	18	CAAGGCATCCCATTCAAAAAGCA	GGAAAGAACAGGGAATGACCACA
SETBP1	18	GGAATAAGAAGGATCCCCGTGTC	GATTTCTGGCAGCTTTCTGGATG
SETBP1	18	TATCCCAGGAGGTGTGTCTAAGC	TCAGTTTATCAGAGGTGGCCATA
SETBP1	18	CCATCCAGAAAGCTGCCAGAAAT	ACTGTCTGTGTTGGTGTAGAGAG
SETBP1	18	GTTTCTAGCCAGCCGGATGTTC	GATGGGCTGGAGATTTGGCATAG
SETBP1	18	AATGGCAACCTGAGCCCTG	AGCGTGATTTCCTTTAGGGAGC
SETBP1	18	AATACACAGTGGAACCTGGAAGC	CTTGGTCAGAAGTGCTGTTGTTG
SETBP1	18	CTAAAGGAAATCACGCTGTCCC	CGGTTCTTTGTGCTGGTGTC
SF3B1	2	AATTCTGTTAGAACCATGAAACATATCC	TGCTTTCTTGAAGGCTATTGGGT
SF3B1	2	AGTAGTTGGCATATTCTGCATCC	CGCCAACACAGAGGAAAGGTAAA
SF3B1	2	CTTCAAGAAAGCAGCCAAACCC	AGCAGCAACTCCTTATGGTATCG
SF3B1	2	TTGGCGGATACCCTTCCATAAAG	ACTTAGGTAATGTTGGGGCATAGT
SF3B1	2	TGTGTGTGTACCTCTAGTCCCAA	CTGTGCCATCTTGCCACATCTTA
SF3B1	2	AAAGACAAAGTTACATTACAACTTACCA	AAAGCTGTGTGCAAAAGCAAGAA

SF3B1	2	AAGATGGCACAGCCCATAAGAAT	TGGATGAGTATGTCCGTAACACA
SF3B1	2	TCTTGCTTTTGCACACAGCTTTT	ACCAACTCATGACTGTCCTTTCT
SF3B1	2	AACCATTTCTTTCCATAATCAATTCCA	TGATGTGAAAGTGTAGCTTCTTCT
SF3B1	2	TGTGCAAAGGAAAAGGTCTAGGA	ACTTGAGGATCAAGAGCGTCATT
SF3B1	2	TGGACGAACTAAGTCATCAAGTTT	ACCACACCTATTACTCTGCTCTTTT
SRSF2	17	CGACTCAAAAAGACCTACCCCAA	CAGGTCCCGGTCTCGGT
SRSF2	17	GATCGCGACCTGGATTTGGATT	GCCGCTCGAAGTCTCGG
SRSF2	17	GAGGACTTGGACTTGGACCTTC	GAATTAACCCCGCTGTGCTTG
SRSF2	17	GCCGCGGACCTTTGTGA	GAGCTGCGGGTGCAAATG
SRSF2	17	CACCGCCCCCGTACCT	AAGGAGTCCCGCGGCTT
SRSF2	17	CATGGCATCCATAGCGTCCTC	CTCAAGGTGGACAACCTGACCTA
SRSF2	17	CGCGGGACTCCTTGGTGTA	GGCCGCCACTCAGAGCTA
SRSF2	17	GTGCGGTAGGTCAGGTTGTC	CCTTTCCCAGTGTCCCCAC
SRSF2	17	CTTCGCCGCGGACCTT	GAGCTGCGGGTGCAAATG
SRSF2	17	GTCCCCTCAGCCCCGTTTA	CTGAGGACGCTATGGATGCC
SRSF2	17	CCCGTTTACCTGCGGCTC	CCTTCGTTCGCTTTCACGACAA
SRSF2	17	CACCGCCCCCGTACCT	AAGGAGTCCCGCGGCTT
STAG2	X	GAAAGAAGGCAAGCCACCATTTT	TCCCCATTTTGTGGAATAAACAAGT
STAG2	X	TTAGAAAATATGGCAAAAATTAGAAACTCAATA	TGAGATCTTAGGAAACAAAATAAAGTCAA
STAG2	X	AGCTGTGTTTTGAACTCTCAAGGA	TTTCCACTCCATTCTGTTGGTGA
STAG2	X	TAAAAAAGGCAAAAAGGGCCCAG	ACAAATGAACTGGGAAGAACAACA
STAG2	X	TTGATTTTGAGAAAATTAGAAGAAGCTAATG	AGTGACTATTTGAGAGCTGCTGATT
STAG2	X	TCACACCATATATTAACTTCTGACATTTGC	ACTTTCAAAAATAAGTCAAGATAAACATGC
STAG2	X	ACCTTTTCATGCTTTTGTCTAGGG	GCCGTACTAACACGCCAATGAAT
STAG2	X	AATTCATTGGCGTGTTAGTACGG	TCATCTCAAATCTAAGACAATATGCAGAA
STAG2	X	GCTCATTTCTGCTTAATATTTCTATTTGATCT	TCCAATCATTTTATTCCGTTCTGC
STAG2	X	AGTTGATGACAGCTTTGGTGAATG	CCAAGTTGGTCACACAATAGCC
STAG2	X	ACTTGGCATCTCTTGAATAAACCA	TCATGCATAGTCCAACCAACATA
STAG2	X	GCTGAAATTCGAGCTATTTGCATT	TCATGAATGAGAGGTGCAGACA
STAG2	X	AGCAAGGTGAAGTAAGACTCAAATG	AGGAAAGCAAGAGAAAAGTGTTTG
STAG2	X	TTGCAAACACTTTTCTCTTGCTT	AATATGCAACAAATACATACTTACTGTAAAAC
STAG2	X	TGTGTCTATGACCCTTGACAAAGAA	TCTAGCTGTAAACCTCCATGACG
STAG2	X	AGGACGTTACTAAAAGCACCTGT	ACTGTTATGAAAGTATGCCTTAGAAAATG
STAG2	X	AAATTTTTTGTTGTTGTTGTCGTAAAATAAA	CGAACCTAAGTATAAAATAAATACTATCAACTAACC
STAG2	X	AGTTTCCATTCTTTTGAGTTAAGGC	ACATCCTTACCTTCCTCTCCACT
STAG2	X	ACAAGTGGCATATAGGGAGAAGAA	AAAGACATTGATTGGCATACCCT
STAG2	X	GCAAGAGAGTGCTCTGATTGAAA	ACAAAATTGTTTCATGGCTAAAATGGAAT
STAG2	X	ACACTTAACAGTGCTAATGGGCTT	ACTTTTGCTAATAACTGAGGAAGGG
STAG2	X	AGTCTTTCTGTTCCTTTAGGTGCT	ATTCTGTGAGGCATTTAGGGAAA
STAG2	X	TTTTTTCCCTAAATGCCTCACAGAA	AATAATCTTACCTTTTCTAATCGTCCAGT
STAG2	X	GACTAACTTGTTGCAGTTGCCTC	AGGTTTCCTTTCTTAAAATACGTTCCA
STAG2	X	TGGTCATGTAGTCACTTTATACCTATC	TCTACTCTGTTGAAGATTGTGAACT
STAG2	X	TCAACAGAGTAGATATTTCAAGAAGTCA	ACTGCTAGGGACTATCACCAAGA
STAG2	X	AGTGACATTTTGATAGCCTAGGAG	ATTTCTACAGATTTAAGAAAACAGGAACT
STAG2	X	AGTTTACTGCATTTACTGCATTCTG	TCAGGCATGTCTCCATTTTCGAT
STAG2	X	CCATGACCTTTCAAAGTGGGATTT	CTGCAGTGCGTGAATAACAATCT
STAG2	X	TTTATCTTTGAAAAGTGAAGTCTTGTGA	TCTAATAAAAACACTACTGAACCACCA
STAG2	X	CAGAAGTAGAGTTAATAAAGCTATGAGATAATG	AACAGTAGTATTCACGTTGGTCA
STAG2	X	AGGACTTGCTGCGTTTAAAGAAA	AAGATATTTCTGCTTTGCTCAATCTT
STAG2	X	GGCAGTTAGTGAGAAACCTTGGA	CTGAGCAACTCAGACTGCAATGA
STAG2	X	TCTTCATTGCAGTCTGAGTTGCT	ACCTGTGTTTCACGTTTTTAAGGA
STAG2	X	ACAAGATGCTTAATGTTTGGGACT	TGCTATCCATTTTTCAACTTTACCTTCA

STAG2	X	TGTGACGTGTTTACATGACTAACCT	TGTTGCAGACTGAGAATAAGGGTC
STAG2	X	TGGAGATATCATCAAAGAAACAATGAGT	ATTGCACACTTTGCCCAATTTCA
STAG2	X	TGCCTCATTTATTGAACACCTGTA	AGCAAAACGTCGAGCAAGT
STAG2	X	GCTTTTTAATGAAATGATACAAGAAAATGGCTA	TTCATGACTAAAAAATATGAGAGCTGAAT
STAG2	X	TGGCAAAGGAAGTAGTGAGTGAA	TCACTCAGAATATCAAGAAATGCCAA
STAG2	X	TCCTTTGCAGAGATGGCATAGAA	TGCCCTTAAGAATCCCAAAATGAAA
STAG2	X	AATATGCCTATGCTCGCACAACT	CACCAGCTAGCAAAGAATTTCGG
STAG2	X	AATTCTTTGCTAGCTGGTGGTGA	ACAATTCAGTTGGTAAACATTAGGGA
STAG2	X	AGTGACTAAACCTCGTCGTTAATTTT	TGCATTGGATAAACACTCATGAAGC
STAG2	X	AGCTTCATGAGTGTTTATCCAATGC	GCTTCCTCTTGTTGTCTTTGAGC
STAG2	X	ACATAGAAAAAGTTAAGTTAAAAGAGGAATAAAAT	ACTCAGACAATAAGGCACTCTCAC
STAG2	X	AGCCACATACTGCTGCCTAGATA	CCATGGTGTCAAAATCCATTCCC
STAG2	X	AGCCTAATGGAAGATGATGAAGAGC	TCAAAAGATCTAAACTGAAAATCTTACAGAA
STAG2	X	TGGATATTTTCAGTTACTATCTGGTTTGT	TGATTTTGCAATCCTAGGGTACAT
STAG2	X	AGAGGGGAAGTTTTCAAAGTGGT	CTTAGAAAATGACTTCACCACAGATT
STAT3	17	GGCTTCCAACCTTTGGCAGATTA	GCTTACTGAATGCGAAGTCACAG
STAT3	17	GGCATTTGCCTATCTATCCTCCA	TGGTGTCTCCACTGGTCTATCTC
STAT3	17	TCCTCCTTGGGAATGTCAGGAT	CCTAGCTGTAGGTTCCATGATCT
TET2	4	CACATTTTAATTTTTGTTTCCATGCTCT	GGCAGTGGGCTTCCATTCT
TET2	4	CCATTCCTGATACCATCACCTCC	ACTCACACGACTATTCTGGCTTC
TET2	4	GCTCATCCAGAAGTAAATGGAGACA	TTGATCTGAAGGAGCCCAGAGAG
TET2	4	GTAGAGGGTATTCCAAGTGTTTGC	GAGACATTTGGTTGACTGCTTTC
TET2	4	CCAAAAGGCTAATGGAGAAAGACG	CTCTGGATTTTCAGGCCCACT
TET2	4	AGCCAAGAAAGAAATCCAGGTGA	CACTTTTCCCCTCCTGCTCATT
TET2	4	CAGTGGGCCTGAAAATCCAGAG	ACAATCTGGATAATATTGAGACAGTGTTTTT
TET2	4	TGCTAATGCCTAATGGTGCTACA	GGTGAGTGATCTCACAGGACAAC
TET2	4	AAATGCCATTAACAGTCAGGCTA	AGGTATTTAGCATTGCAGCTAGTTT
TET2	4	ATGCTGATGATGCTGATAATGCC	GCTTGCAAATTGCTGCTGGA
TET2	4	AAAGCTAGCGTCTGGTGAAGAAT	AAGAAGCAATTGTGATGGTGGTG
TET2	4	GCTTACTTCAAGCAAAGCTCAGT	AGTGTTGTGTTACTTTGGTTGGG
TET2	4	CAGGTTCCTCAGCTTCCTTCAG	TTTCAGAAAGCATCGGAGAAGGG
TET2	4	GGAAGTGAAAATAGAGGGTAAACCTGA	TGAGGTGTTCTGACATTGGTCTT
TET2	4	CTACACATGTATGCAGCCCTTCT	AGTTGTCCTGTAGCTCTCCACT
TET2	4	CACCTCAAGCATAACCCACCAAT	GATTCCGCTTGGTGAAAACGAG
TET2	4	GAGACAAGGAGCAAACACGAGAT	TCCAGTGTATTGTTTGGAGGTCA
TET2	4	TGAGGCATCACTGCCATCAATTC	TTGGGACTGCCCTTGATTCATTT
TET2	4	CAAGGCAAGCTTACACCCAGAA	ACAGTGACTGCACATGAGCTTTT
TET2	4	GTACAGTGGACCAACATCTCCAG	GCCTGTTGATTCAAGTGCTGTTT
TET2	4	AGTCACTGTGTGGCACTAGATTT	GAGGGAGATGTGAACTCTGGGAT
TET2	4	TGAGCCATTTTCAAACTCACACC	TGATCCTTCTCTTTGCTGATCATT
TET2	4	AAACCAGCAACAGCAGCAAAAAT	CCAGTCCCATTTGGACATTATGAG
TET2	4	GGCCAGACTAAAGTGGAAGAATGT	TCTGAAACTAGGTGTGTATTGTTTGA
TET2	4	ATCGTAGAAATTCCCCTTATAGTCAG	AGATCTTGCTTTGGGATCACATT
TET2	4	AGACTACACATCCTGAACTTTTTGC	CGCAGCTTGTTGACCAGACATA
TET2	4	GCTTTCAAGAACAGGAGCAGAAG	CTTTTGAGTGTCCTTCTGGGGAG
TET2	4	CCATGCAAATGTTTTTCCTGTGC	CACCTTAATTGGCCTGTGCATCT
TET2	4	CCCAAACTGAGTCTTGCCATAGT	TGGTCTCAATGATGCTCTTTTGC
TET2	4	AATCCACCTGCAAGCTGTGATAA	GTTCTGCAGCAGTGGTTTGTCTA
TET2	4	AAATCACAGAAGCAAGTAAAAGTTGA	GGAGTATCTAGTAATTTGGAAGGTGA
TET2	4	AGAGCAGCAAACAACTTCTTCAG	TGTCTCAGTACATTTCTGGCACT
TET2	4	CACCTTCCAAATTACTAGATACTCCT	AATTGCTGCCAGACTCAAGATTT
TET2	4	CTTGCAGATGTGTAGGTAAGTGC	CTTTCACAAGACACAAGCATCGG

TET2	4	GGTTAAGCTTTGTGGATGTAGCC	GTTAATCTGCCCTGTGCCTTTG
TET2	4	GTGTGTGTGTTTCTGTGGGTTTC	ACCCAATTCTCAGGGTCAGATTTAC
TET2	4	TTGTTTTGTTTTGGTTGGGGTGG	CGAGTAGAGTTTGTCAGCCAGAG
TET2	4	TGATTGTGATTCTCATCCTGGTGT	CCTTTCAACCAAAGATTGGGCTTT
TET2	4	TCAGCTGCACAGCCTATATAATG	ACTTCCTTGGGATCTTGCTTCTG
TET2	4	AGAAACCTGTGGTGCCTCCT	TGTCATATTGTTCACTTCATCTAAGCTA
TET2	4	TGGGATTCAAAATGTAAGGGGAA	ATATGCATCAGGTGCAAGTTTCT
TET2	4	GTCCACTCTTATGGCACCAACAT	TGCAGTGGTTTCAACAATTAAGAGG
TET2	4	TGTGTCATTCCATTTTGTTTCTGGA	CCTTCCTTCAGACCCAGACG
TET2	4	CATTCACACACACTTTTATTTTTCAGATT	GCTGCCATTCTGCATGTTGTG
TET2	4	AGAGTGCCGTCTGGGTCT	TACCCAGTCTTGCATATGTCTTT
TET2	4	ACACACACACACGTTTTCTTTGG	CTTCCACACTCCCAAACTCATCC
TET2	4	TCGAGAATTTGGAGGAAAACCTG	GTCTTGACTGGCTCTGCTAACAT
TET2	4	TCAGGAGGAGAAAAAACGGAGTG	TGGGGCTGACTTTTCCTTTTCAT
TET2	4	CGACAAAGGAAACTAGAAGCCAA	TTTAATATACCACACAACACATTTATCTACA
TET2	4	CTCACTAGCCTTCATAAAATAATCATCAA	CTCTGTCTGAGGGTGATGTG
TET2	4	AGTCATGCAGCAGTCCCAG	CTGAAGTGTGTGAAGAGTTTGGA
TET2	4	GGATCCACCAATCCATACATGA	TCTGATTCAAAAGCCCAGGGTAA
TET2	4	TATCTATGGAAGCACCAGCCCTA	CTGGGGAGAATAGGAACCCAGAT
TET2	4	CCCTTACCCTGGGCTTTTGAAT	ACCTTGGCTGGTAAAGTGTATGG
TET2	4	ATCTGGGTTCCTATTCTCCCCAG	GCTGAAACCATCTCCCTGCATATT
TET2	4	TCCATACACTTTACCAGCCAAGG	TAATCTAGAGGTGGCTCCCATGA
TET2	4	TCAGCAGTTGTACCATTAGACCAAA	GGAGCTGCACTGTAGTTATGGAT
TET2	4	CACCCAATCTGAGCAATCCAAAC	ATCATGGTTAAGAGCTGGAAGCA
TET2	4	CATGCTTTCCCACACAGCTAATG	CTGACCAGACCTCATCGTTGTC
TET2	4	ATGCTAATGGTCAGGAAAAGCAG	TTTAAAGGGGTTGTGGCATGCAG
TET2	4	GGTGCAGAGGACAACGATGAG	TGGTAAAAGACGAGGGAGATCCT
TET2	4	CTGCATGCCACAACCCCTTTA	TGGGATTTCTGAGGCACATAGTC
TET2	4	AAAATGGCTGAAAAAGCCCGTGA	GTCACGGACATGGTCCTTTCG
TET2	4	TGAGCCACATGAAACTTCAGAGC	ACAGGTTGGTTGTGGTCTTTTCA
TP53	17	AATGCCCCAATTGCAGGTAAAAC	TTATCACCTTTCCTTGCCTCTTT
TP53	17	AAGGAAAGGTGATAAAAGTGAATCTG	TGTTTGTGCCTGTCCTGGG
TP53	17	TTCTTGTCCTGCTTGCTTACCTC	CCTTACTGCCTCTTGCTTCTCTT
TP53	17	GTGTGCAGGGTGGCAAGT	CCACAGGTCTCCCCAAGG
TP53	17	CACTGACAACCACCCTTAACCC	CCTCTGATTCCTCACTGATTGCT
TP53	17	TGGGGACCCTGGGCAA	CTGTGCAGCTGTGGGTTGATT
TP53	17	CCTCACAACCTCCGTCATGTG	CTTGTGCCCTGACTTTCAACTCT
U2AF1	21	GGTGGGTTGGAAGGAGACATTTA	GGAAAAGGCTGTGATTGACTTGA
U2AF1	21	CTCCTCACTCACCCCATCTCAT	TGGATGGCAAGCACTTCTGTTTT
U2AF1UTR	21	ACGAGAGAAAAAATGACTTGCTTAAT	CATTTTCCCTTACAGAGTCAACTG
U2AF1UTR	21	TCGGTTTATTGTGCAACCGAGAG	TCTGTGTCATGTTTCTGTGAGGT
WT1	11	TCCCTCTCATCACAATTTCATTCCA	GAGGCTAGACCTTCTCTGTCCA
WT1	11	TGGAAATAACCTGGGTCCTTAGC	GGCATCTGAGACCAGTGAGAAA
WT1	11	ACACTTACCAGTGTGCTTCCTG	AAAGCCTCCCTTCCTCTTACTCT
ZRSR2	X	ACAGAGGAGCGGGGAGC	ACAGAGCAATTTCACAACACAGA
ZRSR2	X	CCTGAATTTTTGACCAAGGATTTGC	TAGCACTGACCTCTCTCTTTCCA
ZRSR2	X	TTCCCTAGGACTCTCACAGAAGG	TTATACAGAAGACTGGTACTGGTTAGT
ZRSR2	X	TCGTGTGTGTGTGTATTTGTGGA	ACGTGATTCCTACCATACCTCTT
ZRSR2	X	GTGCGCTGTATGTGAAATGTTTTT	CATCACACAGCACATCCCATGA
ZRSR2	X	GATGTGACTTACCTGACTTCTGGG	ACCACGAAACTAACATTACTGGA
ZRSR2	X	CTCCACCAGTAAAGTCATGGTTA	ATCACTTCATTCTGACATGATAAAACAA
ZRSR2	X	ATGCCTGGTCTAAAGCAGTTGTT	CTTGCGTCAGGGTCATAGTCATC

ZRSR2	X	AGCATGTTTACGACGTTTGGAAT	CAGTCCCCTCCTCCACA
ZRSR2	X	CTTTGGGGAATGTTAGCCTGGAC	CAGGAAGACATCCACAAGCAGAA
ZRSR2	X	ATCATTTGATTTTTGGTTTAAAGGGAAG	ATAGCAGTGGAACAATTGAGGAA
ZRSR2	X	ACCTTCGGAAAAGGATAAAGTAGCA	AGTCCGATCTGGAGACAAGTAGA
ZRSR2	X	GCACTGCAACTTTCTTCATGTGT	TTGTAGGAGTGGTCTGGACTAGG
ZRSR2	X	AGGAAAAGTAGTCGTCACAGGG	CTTCGGGACCTAGAGGAACTTT
ZRSR2	X	CAAAGTTCCTCTAGGTCCCGAAG	CCAGACTTGCGAACCCATCTG

Appendix 8.7

Validation Cohort

Sample no.	Gene	Variant	Chr	Start	End	MiSeq Reporter	GS Junior	In-house pipeline
						Mutation detected?		
1	DNMT3A	G/A	2	25464484	25464484	Y	Y	Y
	IDH1	G/C	2	209113113	209113113	Y	Y	Y
	DNMT3A	G/-	2	25470563	25470563	Y	N	Y
2	IDH1	G/C	2	209113113	209113113	Y	Y	Y
	NPM1	_/GTCG	5	170837547	170837548	Y	N	Y
3	IDH1	G/A	2	209113113	209113113	Y	Y	Y
	ASXL1	AC/--	20	31022430	31022431	Y	Y	N
	SRSF2	G/C	17	74732959	74732959	Y	N	N
4	RUNX1	G/A	21	36259208	36259208	Y	N	Y
	TET2	C/T	4	106156951	106156951	Y	Y	Y
	EZH2	T/-	7	148526850	148526850	Y	Not in panel	Y
5	TET2	A/T	4	106196294	106196294	Y	Y	Y
	NRAS	C/T	1	115258747	115258747	Y	Y	Y
	TET2	A/-	4	106156706	106156706	Y	Y	Y
	ZRSR2	_/A	X	15809096	15809097	Y	N	Y
6	DNMT3A	C/T	2	25457242	25457242	Y	Y	Y
	NRAS	C/T	1	115258747	115258747	Y	Y	Y
	NRAS	C/A	2	115258747	115258747	Y	N	N
	NRAS	C/T	2	115258748	115258748	Y	N	Y
	KIT	C/T	9	55599268	55599268	Y	N	Y
7	TET2	G/T	4	106190804	106190804	Y	Y	Y
	CBL	A/C	11	119148892	119148892	N	Y	N
	TET2	_/C	4	106157385	106157386	Y	N	Y
	STAG2	C/T	X	123179197	123179197	Y	Not in panel	Y
8	ASXL1	T/G	20	31024179	31024179	Y	Y	Y
	TET2	C/T	4	106196537	106196537	Y	Y	Y
	TET2	G/T	4	106197143	106197143	Y	Y	Y
	KRAS	C/T	12	25398281	25398281	Y	N	Y
	ZRSR2	G/-	X	15841165	15841165	Y	N	N
9	TET2	C/T	4	106164787	106164787	Y	Y	N
	ZRSR2	C/T	X	15838370	15838370	Y	Y	Y
	TET2	_/A	4	106164896	106164897	Y	N	Y
10	TET2	G/T	4	106193739	106193739	Y	Y	Y
	ASXL1	G/-	20	31024081	31024081	Y	N	Y
	ZRSR2	C/T	X	15822291	15822291	Y	Y	Y
11	TET2	C/T	4	106196726	106196726	Y	Y	N

	TET2	GATTA/-----	4	106196883	106196887	not the same del	Y	N
	TET2	C/A	4	106155620	106155620	Y	Y	Y
12	DNMT3A	C/T	2	25457242	25457242	Y	Y	Y
	TP53	C/G	17	7576853	7576853	Y	N	Y
13	DNMT3A	C/T	2	25457242	25457242	Y	Y	Y
	KRAS	C/T	12	25398281	25398281	Y	Y	Y
	TP53	C/G	17	7576853	7576853	Y	N	Y
14	TET2	C/T	4	106196267	106196267	Y	Y	Y
	SRSF2	24bp del	17	74732942	74732965	N	Y	Y
15	TET2	C/T	4	106196267	106196267	Y	Y	Y
	TET2	C/T	4	106196213	106196213	Y	Y	Y
	SRSF2	24bp del	17	74732942	74732965	N	Y	Y
16	IDH1	G/A	2	209113113	209113113	Y	Y	Y
	SF3B1	T/C	2	198266480	198266480	Y	Y	Y
17	NRAS	T/A	1	115256528	115256528	Y	Y	Y
	RUNX	C/T	21	36252865	36252865	Y	Y	Y
18	TET2	T/G	4	106180871	106180871	Y	Y	Y
	TET2	C/-	4	106156479	106156479	Y	N	Y
	RUNX1	T/G	21	36259160	36259160	Y	N	Y
19	U2AF1	T/C	21	44514777	44514777	Y	Y	Y
	RUNX	C/T	21	36231782	36231782	Y	Y	Y
20	SF3B1	C/A	2	198267359	198267359	Y	Y	Y
	TET2	T/G	4	106196829	106196829	Y	Y	Y
	KIT	C/T	4	55599268	55599268	Y	N	Y
21	SRSF2	G/T	17	74732959	74732959	Y	N	N
22	IDH1	G/A	2	209113113	209113113	Y	Y	Y
23	IDH1	G/A	2	209113113	209113113	Y	Y	Y
	CSF3R	G/A	1	36932209	36932209	Y	Not in panel	Y
24	ASXL1	G/T	20	31023821	31023821	Y	Y	Y
	P53	C/T	17	7578406	7578406	Y	Y	Y
	IDH2	C/T	15	90631934	90631934	Y	Y	Y
	DNMT3A	G/A	2	25469984	25469984	Y	N	Y
	NPM1	_/TGCA	5	170837546	170837547	Y	N	Y
25	ASXL1	G/T	20	31023821	31023821	Y	Y	Y
	KIT	C/T	4	55599268	55599268	Y	N	Y
26	NRAS	G/T	1	115256530	115256530	Y	Y	Y
27	TET2	C/T	4	106157527	106157527	Y	Y	Y
	U2AF1	G/A	21	44524456	44524456	Y	Y	Y
	TET2	A/-	4	106157346	106157346	Y	N	Y

28	CBL	G/A	11	119148991	119148991	Y	Y	Y
	TET2	C/-	4	106180812	106180812	Y	N	Y
	NPM1	_/TCTG	5	170837544	17083755	Y	N	Y
	SRSF2	G/T	17	74732959	74732959	N	N	N
29	ASXL1	_/T	20	31023329	31023329	Y	N	Y
30	ASXL1	G/T	20	31023821	31023821	Y	Y	Y
	TET2	G/A	4	106158509	106158509	Y	Y	Y
	TET2	G/A	4	106157961	106157961	Y	N	Y
31	TP53	C/T	17	7578406	7578406	Y	Y	Y
	DNMT3A	C/T	2	25457242	25457242	Y	Y	Y
32	DNMT3A	C/T	2	25463562	25463562	Y	Y	Y
	IDH2	C/T	15	90631934	90631934	Y	Y	Y
	ASXL	AGTC/----	20	31022884	31022887	Y	Y	Y
33	TET2	G/T	4	106196819	106196819	Y	Y	Y
	KRAS	C/G	12	25398285	25398285	Y	Y	Y
	SRSF2	G/T	17	74732959	74732959	Y	N	N
	TET2	_/A	4	106157291	106157292	Y	N	Y
34	TET2	T/A	4	106164865	106164865	Y	Y	Y
	STAG2	G/A	X	123179055	123179055	Y	Not in panel	Y
	SRSF2	G/T	17	74732959	74732959	N	N	N
35	ASXL1	A/G	20	31023663	31023663	Y	Y	Y
	SRSF2	G/T	17	74732959	74732959	Y	N	Y
36	NRAS	C/T	1	115258747	115258747	Y	N	Y
	ASXL1	A/-	20	31022418	31022418	Y	N	N
	KRAS	C/T	12	25398284	25398284	Y	N	Y
	ASXL1	C/T	20	31022425	31022425	Y	N	N
37	None					N	N	N
38	TET2	C/G	4	106196220	106196220	Y	Y	Y
	TET2	10bp del	4	106196361	106196370	N	Y	Y
	ZRSR2	CA/--	X	15840904	15840905	Y	Y	Y
	TET2	T/C	4	106157698	106157698	Y	N	Y
	TET2	C/T	4	106196834	106196834	Y	N	Y
	NRAS	A/C	1	115256521	115256521	Y	N	Y
39	SRSF2	G/T	17	74732959	74732959	Y	N	N
40	NRAS	C/T	1	115258744	115258744	Y	Y	Y
	KRAS	C/T	12	25398284	25398284	Y	N	Y
	SRSF2	G/T	17	74732959	74732959	Y	N	Y
41	NRAS	C/A	1	115258747	115258747	Y	Y	Y
	IDH2	C/T	15	90631934	90631934	Y	Y	Y

	RUNX	T/A	21	36206887	36206887	Y	Y	Y
	SRSF2	G/C	17	74732960	74732960	Y	N	Y
42	SRSF2	G/T	17	74732959	74732959	Y	N	Y
	CSF3R	_/G	1	36932123	36932124	Y	Not in panel	Y
	ASXL1	_/A	20	31022441	31022442	Y	N	Y
43	MPL	G/T	1	43815009	43815009	Y	N	Y
	ASXL1	_/T	20	31022450	31022451	Y	N	N
	ASXL1	C/T	20	31023159	31023159	Y	N	Y
	TET2	TATTC/-	4	106196346	106196350	Y	N	Y
	EZH2	A/G	7	148514438	148514438	Y	Not in panel	Y
44	KIT	A/T	4	55599321	55599321	Y	N	Y
	ASXL1	C/T	20	31024450	31024450	Y	N	Y
	RUNX1	G/T	12	36259260	36259260	Y	N	N
	RUNX1	C/-	12	36259265	36259265	Y	N	N
	RUNX1	G/A	12	36259265	36259265	Y	N	N
	TET2	G/C	4	106156021	106156021	Y	N	N
	TET2	T/G	4	106196829	106196829	Y	N	N
45	TET2	G/T	4	106190776	106190776	Y	N	Y
	JAK2	C/A	9	5070025	5070025	Y	N	Y
	JAK2	A/T	9	5070026	5070026	Y	N	Y
	JAK2	A/T	9	5070027	5070027	Y	N	Y
46	CALR	_/TTGTC	19	13054628	13054629	Y	Not in panel	Y
	U2AF1	T/G	21	44514777	44514777	Y	N	Y
47	CALR	52bp del	19			N	Not in panel	N
48	SRSF2	G/A	17	74732959	74732959	Y	N	N
	ASXL1	C/T	20	31023717	31023717	Y	N	Y
	TET2	A/-	4	106197145	106197145	Y	N	Y
	JAK2	G/T	9	5073770	5073770	Y	N	Y

Appendix 8.8

Hotspot variants in myeloid malignancies

chr	pos	wt	alt	gene	prot
1	43815009	G	T	MPL	p.W515L
1	115256528	T	A	NRAS	p.Q61H
1	115256529	T	A	NRAS	p.Q61L
1	115256530	G	T	NRAS	p.Q61K
1	115256532	C	A	NRAS	p.G60V
1	115258744	C	T	NRAS	p.G13D
1	115258745	C	A	NRAS	p.G13C
1	115258747	C	G	NRAS	p.G12A
1	115258748	C	A	NRAS	p.G12C
1	36933434	C	T	CSF3R	p.T618I
1	36932248	C	T	CSF3R	p.Q768*
1	36932224	C	T	CSF3R	p.Q776*
2	25457176	G	A	DNMT3A	p.P904L
2	25457209	C	G	DNMT3A	p.W893S
2	25457242	C	T	DNMT3A	p.R882H
2	25457243	G	A	DNMT3A	p.R882C
2	25458627	G	A	DNMT3A	p.P849L
2	25458649	G	A	DNMT3A	p.Q842*
2	25458669	G	A	DNMT3A	p.T835M
2	25463184	G	C	DNMT3A	p.S770W
2	25463248	G	T	DNMT3A	p.R749S
2	25463286	C	A	DNMT3A	p.R736L
2	25463287	G	A	DNMT3A	p.R736C
2	25463541	G	C	DNMT3A	p.S714C
2	25463562	C	A	DNMT3A	p.G707V
2	25464457	C	A	DNMT3A	p.D686Y
2	25467083	G	A	DNMT3A	p.R598*
2	25467134	A	T	DNMT3A	p.W581R
2	25467436	A	C	DNMT3A	p.L547R
2	25467449	C	A	DNMT3A	p.G543C
2	25470498	G	A	DNMT3A	p.R326C
2	25470516	G	A	DNMT3A	p.R320*
2	25470535	C	T	DNMT3A	p.W313*
2	198266834	T	C	SF3B1	p.K700E
2	198267359	C	G	SF3B1	p.K666N
2	198267360	T	A	SF3B1	p.K666M
2	198267361	T	G	SF3B1	p.K666Q
2	198267371	G	C	SF3B1	p.H662Q
2	198267373	G	C	SF3B1	p.H662D
2	198267483	C	A	SF3B1	p.R625L
2	198267484	G	A	SF3B1	p.R625C
2	198267491	C	G	SF3B1	p.E622D
2	209113112	C	T	IDH1	p.R132H

2	209113113	A	A	IDH1	p.R132C
4	55599320	G	C	KIT	p.D816H
4	55599321	A	T	KIT	p.D816V
4	55599340	T	G	KIT	p.N822K
4	106156246	C	T	TET2	p.Q383*
4	106156348	C	T	TET2	p.Q417*
4	106156478	C	T	TET2	p.S460F
4	106156687	C	T	TET2	p.Q530*
4	106156729	C	T	TET2	p.R544*
4	106156747	C	T	TET2	p.R550*
4	106156963	C	T	TET2	p.Q622*
4	106157002	C	T	TET2	p.Q635*
4	106157167	C	T	TET2	p.Q690*
4	106157212	C	T	TET2	p.Q705*
4	106157240	C	A	TET2	p.S714*
4	106157404	C	T	TET2	p.Q769*
4	106157527	C	T	TET2	p.Q810*
4	106157845	C	T	TET2	p.Q916*
4	106157971	C	T	TET2	p.Q958*
4	106157995	C	T	TET2	p.Q966*
4	106164061	C	T	TET2	p.Q1191*
4	106164778	C	T	TET2	p.R1216*
4	106182940	C	T	TET2	p.Q1327*
4	106193748	C	T	TET2	p.R1404*
4	106193778	C	G	TET2	p.Q1414E
4	106193853	A	T	TET2	p.K1439*
4	106193892	C	T	TET2	p.R1452*
4	106193931	C	T	TET2	p.R1465*
4	106196213	C	T	TET2	p.R1516*
4	106196556	C	A	TET2	p.S1630*
7	148506443	C	T	EZH2	p.R690H
9	5073770	G	T	JAK2	p.V617F
11	32413565	C	A	WT1	p.R462L
11	32413566	G	A	WT1	p.R462W
11	119149251	G	A	CBL	p.R420Q
11	119149242	C	T	CBL	p.P417L
11	119148990	T	G	CBL	p.C404G
11	119148991	G	A	CBL	p.C404Y
11	119148966	T	C	CBL	p.C396R
11	119148930	T	C	CBL	p.C384R
11	119148919	T	C	CBL	p.L380P
11	119148891	T	C	CBL	p.Y371H
12	25380275	T	G	KRAS	p.Q61H
12	25398281	C	T	KRAS	p.G13D

12	25398284	C	G	KRAS	p.G12A
12	25398285	C	A	KRAS	p.G12C
13	28592622	G	T	FLT3	p.N841K
13	28592623	T	A	FLT3	p.N841I
13	28592640	A	C	FLT3	p.D835E
13	28592641	T	G	FLT3	p.D835A
13	28592642	C	G	FLT3	p.D835H
13	28602329	G	A	FLT3	p.A680V
13	28602340	G	T	FLT3	p.N676K
13	28609758	C	A	FLT3	p.V491L
15	90631838	C	T	IDH2	p.R172K
15	90631934	C	T	IDH2	p.R140Q
15	90631935	G	A	IDH2	p.R140W
17	7578190	T	C	TP53	p.Y220C
17	74732959	G	T	SFRS2	p.P95H
17	74732960	G	C	SFRS2	p.P95A
17	74732961	G	C	SFRS2	p.P95H
18	42531907	G	A	SETBP1	p.D868N
18	42531913	G	A	SETBP1	p.G870S
18	42531917	T	C	SETBP1	p.I871T
20	31021250	C	T	ASXL1	p.R417*
20	31021472	C	T	ASXL1	p.Q491*
20	31022277	C	T	ASXL1	p.Q588*
20	31022288	C	A	ASXL1	p.Y591*
20	31022592	C	T	ASXL1	p.R693*
20	31022839	T	A	ASXL1	p.L775*
20	31022847	C	T	ASXL1	p.Q778*
20	31022902	G	A	ASXL1	p.W796*
20	31022988	A	T	ASXL1	p.K825*
20	31023408	C	T	ASXL1	p.R965*
21	36231773	C	A	RUNX1	p.R204L
21	36231782	C	T	RUNX1	p.R201Q
21	36231783	G	A	RUNX1	p.R201*
21	36252865	C	A	RUNX1	p.R166L
21	36252866	G	A	RUNX1	p.R166*
21	36252877	C	T	RUNX1	p.R162K
21	36259171	C	T	RUNX1	p.R107H
21	36259172	G	A	RUNX1	p.R107C
21	44514777	T	G	U2AF1	p.Q157P
21	44514780	C	T	U2AF1	p.R156H
21	44524456	G	A	U2AF1	p.S34F

Appendix 8.9

Plenary paper and associated commentary from Cohort 1 data

MYELOID NEOPLASIA

Targeted sequencing identifies patients with preclinical MDS at high risk of disease progression

Catherine A. Cargo,[1] Nicola Rowbotham,[1] Paul A. Evans,[1] Sharon L. Barrans,[1] David T. Bowen,[2] Simon Crouch,[3] and Andrew S. Jack[1]

[1]Haematological Malignancy Diagnostic Service, St James's University Hospital, Leeds, United Kingdom; [2]St James's Institute of Oncology, Leeds Teaching Hospitals, Leeds, United Kingdom; and [3]Epidemiology and Cancer Statistics Group, University of York, York, United Kingdom

Key Points

- The mutational profile of patients with preclinical MDS is distinct from that reported in healthy individuals.
- In the absence of morphologic disease, mutational analysis can predict those patients at high risk of disease progression.

The diagnosis of myelodysplastic syndromes (MDS) remains problematic due to the subjective nature of morphologic assessment. The reported high frequency of somatic mutations and increased structural variants by array-based cytogenetics have provided potential objective markers of disease; however, this has been complicated by reports of similar abnormalities in the healthy population. We aimed to identify distinguishing features between those with early MDS and reported healthy individuals by characterizing 69 patients who, following a nondiagnostic marrow, developed progressive dysplasia or acute myeloid leukemia. Targeted sequencing and array-based cytogenetics identified a driver mutation and/or structural variant in 91% (63/69) of prediagnostic samples with the mutational spectrum mirroring that in the MDS population. When compared with the reported healthy population, the mutations detected had significantly greater median variant allele fraction (40% vs 9% to 10%), and occurred more commonly with additional mutations (≥2 mutations, 64% vs 8%). Furthermore, mutational analysis identified a high-risk group of patients with a shorter time to disease progression and poorer overall survival. The mutational features in our cohort are distinct from those seen in the healthy population and, even in the absence of definitive disease, can predict outcome. Early detection may allow consideration of intervention in poor-risk patients. (*Blood*. 2015;126(21):2362-2365)

Introduction

The morphologic diagnosis of myelodysplastic syndromes (MDS) is problematic due to poor inter-observer concordance[1] and the difficulty in distinguishing MDS from non-neoplastic conditions.[2] Cytogenetics can provide objective evidence of disease, although reports of frequent driver mutations[3,4] and/or structural variants detected by single nucleotide polymorphism (SNP) arrays[5] have provided potential core criteria for the diagnosis of MDS. However, this approach has been complicated by reports of frequent somatic mutations[6-8] and large chromosomal abnormalities[9,10] in the aging healthy population. The term "clonal hematopoiesis of indeterminate potential" has been proposed for patients with somatic mutations but without evidence of hematologic malignancy, and "clonal cytopenias of undetermined significance" to encompass the subset with cytopenias.[11] Although these individuals show an increased risk of developing hematologic malignancies,[6-8] those with clinically significant mutations are currently indistinguishable from those who will not progress.

The aim of this study was to molecularly characterize those patients with the most clinically significant disease, who fail to meet diagnostic criteria using conventional techniques. We retrospectively identified patients who, despite having an initial bone marrow with nondiagnostic features, developed progressive dysplasia or acute myeloid leukemia (AML). We hypothesized that characterizing these patients would provide potential criteria to distinguish preclinical MDS from healthy individuals, and importantly, detect those patients at high risk of disease progression.

Study design

Patients and samples

A retrospective search was performed for new patients diagnosed with AML/MDS at the Haematological Malignancy Diagnostic Service between 2004 and 2012, with a previous nondiagnostic bone marrow performed for investigation of cytopenia (more recently termed "idiopathic cytopenia of undetermined significance"). The Haematological Malignancy Diagnostic Service provides a centralized hematopathology service for ~6 million population with all cases dual-reported by experienced hematopathologists. Eighty-two patients were identified with both prediagnostic and diagnostic samples, representing 1.7% of patients with idiopathic cytopenia of undetermined significance during this time period (n = 4835). Sixty-nine patients had adequate molecular material at both time points for analysis (see supplemental Table 1 on the *Blood* Web site) and survival data were available for 59 patients. Samples were

Submitted August 7, 2015; accepted September 8, 2015. Prepublished online as *Blood* First Edition paper, September 21, 2015; DOI 10.1182/blood-2015-08-663237.

Presented as an oral presentation at the 56th annual meeting of the American Society of Hematology, San Francisco, CA, December 4-9, 2014.

The array data reported in this article have been deposited in the Gene Expression Omnibus database (accession number GSE73074).

The online version of this article contains a data supplement.

There is an Inside *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

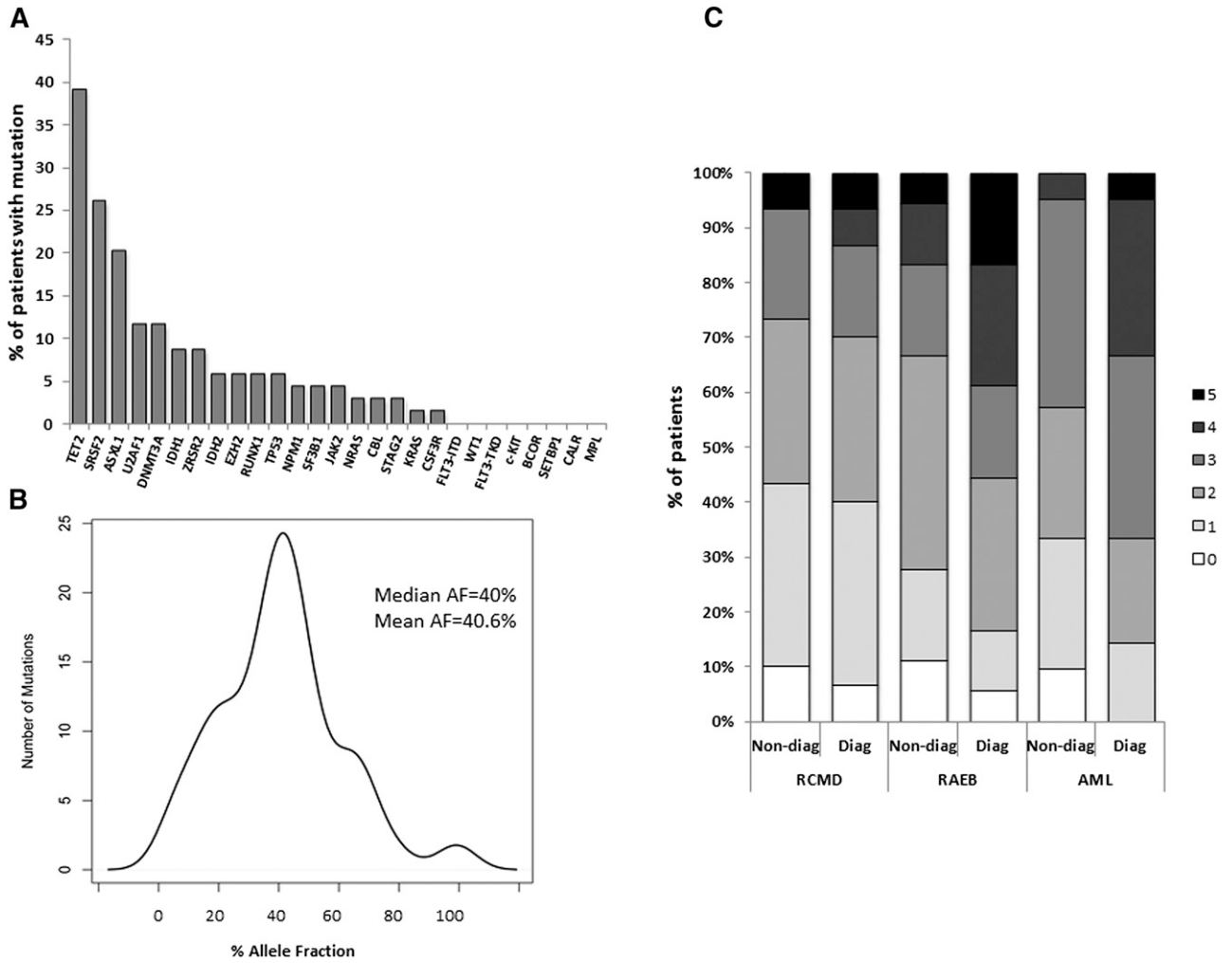


Figure 1. Characteristics of mutations detected in a prediagnostic sample. (A) Frequency of driver mutations across all 69 patients in the prediagnostic sample. (B) VAF (%) of somatic variants in the prediagnostic sample, including median and mean. (C) Distribution of the number of mutations according to final disease subgroups, and across both prediagnostic and diagnostic samples. AF, allele fraction.

taken with informed consent, in accordance with the Declaration of Helsinki, for investigation of a suspected hematologic disorder. The study had local Institutional Review Board approval.

DNA extraction, targeted sequencing, and SNP-array analysis

DNA was either extracted at the point of referral or from stored, unstained smears. Targeted sequencing of 26 commonly mutated genes in myeloid malignancies was performed on MiSeq (Illumina, Chesterford, United Kingdom) using custom Fluidigm panels (Fluidigm, San Francisco, CA) to construct DNA libraries. Read alignment and variant calling were performed using MiSeq reporter (Illumina) and variants annotated using Ensembl Variant Effect Predictor software.[12] Following exclusion of synonymous, noncoding variants and germline polymorphisms (unless recurrently reported in COSMIC database[13]), variants were validated by GS Junior (Roche, Burgess Hill, United Kingdom), conventional Sanger sequencing, or allele-specific oligonucleotide analysis (Janus kinase 2). All patients were screened for NPM1 and FLT3 mutations using conventional fragment analysis.

SNP-A analysis was performed on diagnostic samples using a HumanCytoSNP-12 BeadChip Kit (Illumina) with data visualized using KaryoStudio software (Illumina). Abnormalities were called using published guidelines.[14] In those with a documented abnormality, SNP-A was performed on the corresponding prediagnostic sample.

See supplemental Methods for details.

Statistical analysis

Overall survival (OS) was estimated using the Kaplan–Meier method (censored on 01/06/2015). The impact of abnormalities on OS, and associations between genetic mutations and progression were investigated with Fisher's exact test and Cox regression.

Results and discussion

Mutational profile in prediagnostic samples differs from healthy individuals

A somatic mutation and/or structural abnormality were identified in 91% (n = 63/69) of prediagnostic samples. This included 133 mutations across 62 patients, most commonly involving epigenetic regulators or spliceosome genes with TET2, SRSF2, and ASXL1 mutated in 39%, 26%, and 20% of patients, respectively (Figure 1A). The spectrum of mutations at this time point mirrored that reported in large MDS populations,[3,4] with the exception of SF3B1 (n = 3), though these mutations are strongly associated with ring sideroblasts, which are easily identified morphologically.[15] Although DNMT3A was the most frequently mutated gene in the healthy population,[6-8] this was seen in only

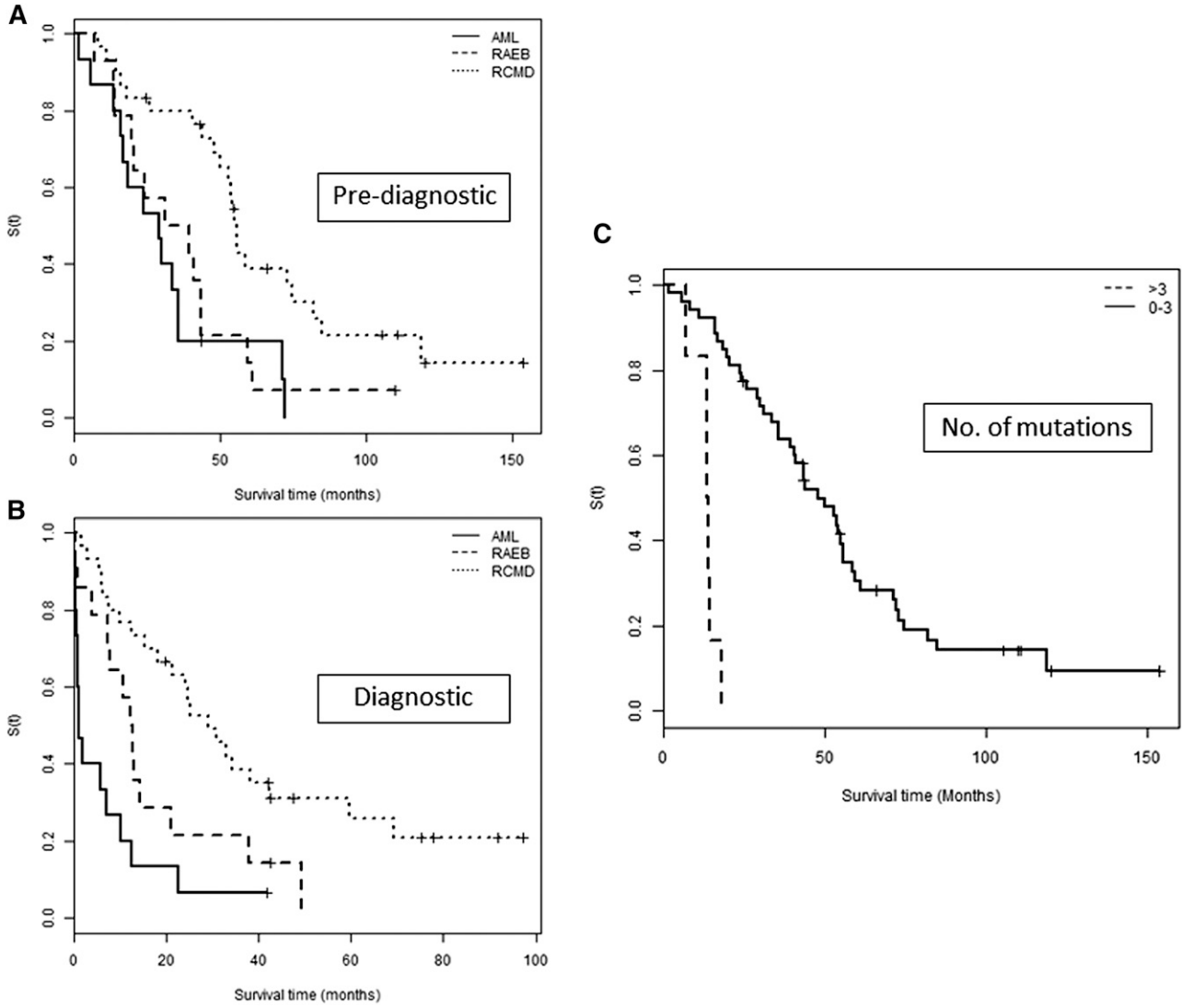


Figure 2. OS according to sample time and mutation number. OS in patients grouped by final disease category from time of (A) prediagnostic sample and (B) diagnostic sample. The survival of the AML group is comparable to the RAEB group if determined from the prediagnostic sample ($P = .442$) and highlights a potential period for earlier intervention. (C) OS according to the number of mutations detected in a prediagnostic sample grouped into those with ≤3 and >3 mutations.

10% (n = 7) of our patient cohort. A median of 2 mutations was detected per patient (range, 1-5; Figure 1C) and the frequency of multiple mutations was significantly greater here than reported in healthy individuals (≥2 mutations; 64% vs 8%[6]). The median variant allele fraction (VAF) and inferred clone size in our series was also notably greater at 40% (range, 2.31% to 100%; Figure 1B) than the reported median of 9% to 10% in healthy individuals.[6] Importantly, only 1 patient harbored an isolated mutation with a VAF <20%. These differences suggest that the clone must expand to an appropriate level and/or acquire cooperating mutations to cause cytopenias and subsequent disease. This is supported by the greater mean VAF (25.2%) observed in healthy individuals who subsequently developed a hematologic malignancy.[6] Therefore, although driver gene mutations with a high AF (>20%) and/or co-occurring mutations may not be disease-defining, it at least identifies clinically significant clonality requiring close follow-up.

In contrast, structural variants were identified in only 23% (16/69) of prediagnostic samples with all but one co-occurring with a somatic mutation (supplemental Table 5).

Mutations predict progression to high-risk disease and OS

Thirty-nine patients progressed to refractory anemia with excess blasts (RAEB) or AML, in a significantly shorter time than those with refractory cytopenia with multilineage dysplasia (median, 403 vs 606 days; hazard ratio [HR], 3.7; 95% confidence interval [CI], 2.1-6.6; $P < .001$). By analyzing the most frequently mutated genes (supplemental Methods), IDH2 was weakly associated with disease progression ($P = .052$). NPM1, CBL, and NRAS were mutated at low frequency, however, all progressed to AML or RAEB. Furthermore, IDH2 (HR, 4.2; 95% CI, 1.3-13.8; $P = .017$) and TP53 mutations (HR, 5.5; 95% CI, 1.1-27.7; $P = .038$) were associated with a more rapid time to progression. These observations require confirmation in larger cohorts.

Thirty patients (43%) acquired a mutation between samples, most commonly involving transcription factors and cell-signaling genes, and correlated strongly with progression to RAEB and AML (supplemental Results and supplemental Figure 2). Mutations across individual samples (supplemental Figure 1) and changes in VAF between samples (supplemental Figure 3) are presented in supplemental Results.

Of those with survival data available (n = 59), only 10 were alive at the point of analysis. Median OS from the prediagnostic sample was 43.6 months (95% CI, 33.8-55.8) and, as expected, much shorter from diagnosis (13 months; 95% CI, 9.9-24.6). This was most significant in those diagnosed with AML (1.28 months; 95% CI, 0.789-12.625; Figure 2) and importantly, all but 1 of these patients harbored a mutation in the prediagnostic sample.

Prediagnostic mutations were also associated with significantly worse OS, namely TP53 (HR, 21.68; 95% CI, 4.72-99.64; $P < .001$), U2AF1 (HR, 2.63; 95% CI, 1.0-6.4; $P = .049$), and the number of mutations (HR, 1.447; 95% CI, 1.12-1.88; $P = .006$). The latter was most significant in those with >3 mutations (Figure 2C).

The mutational profile in our cohort differs significantly from that of the healthy population, and has the potential to identify patients with clonal hematopoiesis of indeterminate potential/clonal cytopenias of undetermined significance who are at greater risk of progression, even in the absence of morphologic disease. Early detection would provide an increased window for therapeutic intervention in those with very poor prognosis. The diagnostic utility of these findings is however limited by the lack of a control group, including patients who did not progress to AML/MDS. Our study design included only those patients re-investigated for cytopenia, and both the mutation and disease frequency is likely to be greater. The optimal way to explore this and refine molecular criteria for diagnosis is to prospectively study an unselected cytopenic patient cohort, and this is currently in progress.

Acknowledgments

The authors thank Dr Jan Taylor for providing additional bioinformatic support.

Authorship

Contribution: C.A.C. and A.S.J. designed the study; C.A.C., N.R., P.A.E., and S.L.B. performed the research; C.A.C., S.L.B., and S.C. analyzed the data; and C.A.C., D.T.B., and A.S.J. wrote the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Catherine A. Cargo, Haematological Malignancy Diagnostic Service, St James's University Hospital, Beckett St, Leeds LS9 7TF, United Kingdom; e-mail: catherine.cargo@nhs.net.

References

1. Parmentier S, Schetelig J, Lorenz K, et al. Assessment of dysplastic hematopoiesis: lessons from healthy bone marrow donors. *Haematologica.* 2012;97(5):723-730.
2. Steensma DP. Dysplasia has A differential diagnosis: distinguishing genuine myelodysplastic syndromes (MDS) from mimics, imitators, copycats and impostors. *Curr Hematol Malig Rep.* 2012;7(4):310-320.
3. Papaemmanuil E, Gerstung M, Malcovati L, et al; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood.* 2013;122(22):3616-3627, quiz 3699.
4. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia.* 2014;28(2):241-247.
5. Tiu RV, Gondek LP, O'Keefe CL, et al. Prognostic impact of SNP array karyotyping in myelodysplastic syndromes and related myeloid malignancies. *Blood.* 2011;117(17):4552-4560.
6. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med.* 2014;371(26):2488-2498.
7. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med.* 2014;371(26):2477-2487.
8. Xie M, Lu C, Wang J, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med.* 2014;20(12):1472-1478.
9. Laurie CC, Laurie CA, Rice K, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet.* 2012;44(6):642-650.
10. Jacobs KB, Yeager M, Zhou W, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet.* 2012;44(6):651-658.
11. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood.* 2015;126(1):9-16.
12. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics.* 2010;26(16):2069-2070.
13. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 2011;39(Database issue):D945-D950.
14. Simons A, Sikkema-Raddatz B, de Leeuw N, Konrad NC, Hastings RJ, Schoumans J. Genome-wide arrays in routine diagnostics of hematological malignancies. *Hum Mutat.* 2012;33(6):941-948.
15. Papaemmanuil E, Cazzola M, Boultwood J, et al; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med.* 2011;365(15):1384-1395.

2015 126: 2362-2365
doi:10.1182/blood-2015-08-663237 originally published online September 21, 2015

Targeted sequencing identifies patients with preclinical MDS at high risk of disease progression

Catherine A. Cargo, Nicola Rowbotham, Paul A. Evans, Sharon L. Barrans, David T. Bowen, Simon Crouch and Andrew S. Jack

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.

inside blood commentary

19 NOVEMBER 2015 | VOLUME 126, NUMBER 21

MYELOID NEOPLASIA

Comment on Kwok et al, page 2355, and Cargo et al, page 2362

Cytopenias + mutations − dysplasia = what?

David P. Steensma HARVARD MEDICAL SCHOOL

Increasingly, clinicians are obtaining molecular genetic tests when evaluating patients with unexplained cytopenias. In this issue of *Blood*, Kwok et al and Cargo et al describe results of mutation testing in 2 series of patients with nondiagnostic marrows. These reports raise an important question: if a clonal mutation is discovered, yet diagnostic criteria for a hematologic neoplasm are not met, what is the diagnosis?[1,2]

Hematologists learn early in training how to evaluate patients with cytopenias, and this remains a core task of hematology clinical practice. But when common cytopenia-inducing causes (eg, vitamin or essential mineral deficiency, infection, immune dysregulation, drug effect, bleeding, organ dysfunction, neoplasia) have been systematically excluded, yet cytopenias persist, clinicians may be puzzled and patients left feeling unsettled. The term "idiopathic cytopenias of undetermined significance" (ICUS), a frank admission of pathophysiological ignorance, has been proposed to describe such patients.[3]

ICUS includes a hodgepodge of conditions. By definition, patients with ICUS are not known to have a clonal hematopoietic disorder. Some people with ICUS will spontaneously recover normal hematopoiesis, others will eventually be diagnosed with a nonhematologic disease, and a few will turn out to have myelodysplastic syndromes (MDS), especially older patients.

Because morphologically dysplastic cells are common in the marrow of healthy people older than 50 years of age,[4] detection of mild dysplasia is not enough to diagnose MDS, even in a cytopenic patient. Instead, in order for MDS to be diagnosed according to World Health Organization (WHO) criteria, patients must exhibit either extensive cellular dysplasia, an increase in marrow blast proportion, or an MDS-associated karyotypic abnormality.[5] Long-term natural history studies of large ICUS cohorts are lacking, so we cannot yet counsel patients with ICUS about their risk of developing MDS, acute myeloid leukemia (AML), or another hematologic neoplasm.

Enter mutation testing. More and more, academic institutions and commercial pathology laboratories are offering clinicians the opportunity to test their cytopenic patients for gene mutations commonly associated with hematologic neoplasms. Given the high frequency of such mutations in patients with bona fide MDS,[6] it would be easy to label cytopenic patients with mutations but without other findings as having MDS, or at least a disease state similar MDS.

But recent data indicate that at least 10% of patients older than 70 years of age have clonal mutations detectable at a variant allele frequency (VAF) of ≥2%.[7,8] We have termed this state, "clonal hematopoiesis of indeterminate potential" (CHIP).[9] Like the parallel precursor conditions monoclonal gammopathy of undetermined significance (MGUS) and monoclonal B-cell lymphocytosis, CHIP confers a risk of subsequent diagnosis of overt hematologic malignancy of 0.5% to 1% per year, and CHIP is associated with increased all-cause mortality. The presence of mutations in apparently healthy people with normal blood counts should give the astute clinician pause before assuming cytopenia plus mutation is equivalent to MDS.

The articles in this issue help define the relationship between ICUS, CHIP, and MDS. Using conventional hematopathology techniques supplemented by a 22-gene panel, Kwok and colleagues prospectively analyzed 144 patients with unexplained cytopenias whose samples were sent to a large commercial pathology laboratory for diagnostic testing.[1] Although 17% of cases met criteria for MDS, 15% had ICUS with mild dysplasia and 69% had ICUS without dysplasia; thus, ICUS was fivefold more common than MDS. Mutations were detected in 71% of MDS, 62% of ICUS with dysplasia, and 20% of ICUS without dysplasia; the frequency of mutations in the latter group is still higher than that for the general population.

The authors then confirmed these findings in a comparison of 91 lower-risk MDS (79% mutation rate) and 249 ICUS (45% mutation rate in cases with dysplasia, 17% without dysplasia), and proposed the term "clonal cytopenias of undetermined significance" (CCUS) to describe ICUS accompanied by a clonal mutation (see table). The most common mutations observed in the CCUS population included *DNMT3A*, *TET2*, *ASXL1*, and *TP53*, similar to CHIP. *SF3B1*, in contrast, was overrepresented in MDS compared with ICUS/CCUS/CHIP, probably because of the strong association of *SF3B1* mutations with the morphologic hallmark of ring sideroblasts.[10] Unfortunately,

Table: Cytopenic and clonal states and their relationship to MDS

State	Key features	Comment
Normal hematopoiesis	No cytopenias, no clonal mutation. Mild cellular dysplasia may be present, especially in older persons.	Healthy state.
ICUS	Cytopenias are present. Dysplasia may be present, but is minimal (<10% of cells per lineage). By definition, a clonal mutation is not known to be present, either because testing was not performed or because testing was unrevealing.	Heterogeneous cluster of pathophysiologically unrelated disorders. May resolve with time, or a diagnosis may become clearer. Some patients with ICUS have MDS but do not meet current diagnostic criteria.
CHIP	Cytopenias are not necessarily present. Dysplasia may be present, but is minimal (<10% of cells per lineage). By definition, patients with CHIP do not meet WHO criteria for a hematological neoplasm. A clonal mutation is present.	Common in the healthy aging population. Confers a 0.5-1.0% per year risk of progression to MDS, AML or another neoplasm. Most patients have just 1 mutation detectable. Certain mutations may confer a higher risk of progression, but this is not yet clear.
CCUS	Required: both cytopenia(s) and clonal mutation(s) in a gene or genes associated with myeloid neoplasia. By definition, patients with CCUS do not meet WHO criteria for MDS or another hematological neoplasm.	Could be considered a subset of CHIP, probably with a higher risk of progression to hematological neoplasia. Some patients with CCUS may have clonal mutations that are not actually responsible for the cytopenia, therefore having CHIP plus a reactive, non-clonal cause of cytopenias. In other patents, the clone contributes to ineffective hematopoiesis. Patients often have 2 or more mutations detectable.
MDS without blast increase	Cytopenias are required for diagnosis. Clonal disorder. Usually, extensive dysplasia is seen; using WHO criteria, diagnosis currently can be made in the absence of dysplasia if certain karyotypic markers are present (e.g., monosomy 7 or del(5q)).	In the future, specific mutations or combinations of mutations may define MDS, even in the absence of dysplasia.
MDS with blast increase	Clonal hematopoietic neoplasm in which myeloid cell differentiation is impaired and blast cells accumulate in the marrow or blood. The WHO calls this "refractory anemia with excess blasts" (RAEB), but anemia is not necessarily present (though cytopenias are typical), nor is dysplasia always present.	Biologically similar to AML. AML is currently defined by the WHO as requiring ≥20% marrow or blood blasts (or else one of a short list of AML-defining karyotypes such as t(8;21), regardless of blast proportion), but MDS with excess blasts can be considered an oligoblastic form of AML.

Colors signify relative risk to the patient. Red is the highest risk state, followed by orange; green is the healthy state; and yellow states are lower risk.

follow-up data on these ICUS/CCUS patients is not available.

Cargo and colleagues, from a large regional hematology service in the United Kingdom, looked at this diagnostic challenge from another direction. Using a 26-gene MiSeq assay, the investigators assessed 69 patients who developed MDS or AML who had previously undergone a nondiagnostic marrow.[2] This group represented 1.7% of 4835 patients referred to the service over an 8-year period for evaluation of unexplained cytopenias. The investigators found that 91% of these patients had mutations detectable in the nondiagnostic sample: a higher rate than in the series by Kwok et al. Unfortunately, the number of nondiagnostic samples with mutations among patients who did not progress to MDS or AML is unknown. Although most patients with CHIP have only 1 detectable mutation and VAF is often <10%,[7] 64% of patients in the Cargo series had 2 or more mutations and the median VAF was 40%, suggesting they were at higher risk of subsequent hematologic neoplasia evolution.

It is clear that some individuals with clonal mutations, cytopenias, or both develop MDS, whereas most others do not. A critical next step will be to conduct longitudinal studies of cohorts of ICUS and CCUS. Eventually, we may come to understand that patients with certain mutations or combinations of mutations have a natural history equivalent to MDS, such that they should be considered to have MDS despite the absence of morphologic dysplasia or a karyotypic abnormality. Other mutations may turn out to be less consequential, just as most elderly men with acquired loss of the Y chromosome do not have MDS or another disorder, even though clonal cells from MDS patients may have a −Y karyotype.[11]

In addition, mutation testing at the DNA level does not tell the whole story, and clinicians will need to learn how to use results from RNA-Seq and assays of epigenetic marks or other biomarkers. Eventually, it will also be possible to abort an emerging clonal process before it evolves to MDS or AML, thereby practicing a form of preventive hematology. Until then, the declaration of doubt inherent in words such as "uncertain" and "indeterminate" will remain a fundamental part of the definitions of ICUS, CHIP, and CCUS.

Conflict-of-interest disclosure: D.P.S. consulted for Genoptix in 2014 and has unrelated technology (an erythropoiesis stimulating agent hemodialysis dosing algorithm) licensed to Mayo Clinic Ventures, which is affiliated with Mayo Medical Laboratories. ■

REFERENCES

1. Kwok B, Hall JM, Witte JS, et al. MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance. *Blood.* 2015;126(21):2355-2361.

2. Cargo CA, Rowbotham N, Evans PA, et al. Targeted sequencing identifies patients with preclinical MDS at high risk of disease progression. *Blood.* 2015;126(21):2362-2365.

3. Wimazal F, Fonatsch C, Thalhammer R, et al. Idiopathic cytopenia of undetermined significance (ICUS) versus low risk MDS: the diagnostic interface. *Leuk Res.* 2007;31(11):1461-1468.

4. Bain BJ. The bone marrow aspirate of healthy subjects. *Br J Haematol.* 1996;94(1):206-209.

5. Brunning RD, Orazi A, Germing U, et al. Myelodysplastic syndromes/neoplasms, overview. In: Swerdlow SH, Campo E, Harris NL, et al, eds. WHO Classification of Tumours of the Haematopoietic and Lymphoid Tissues. 4th ed. Lyon, France: IARC Press; 2008:89-93.

6. Papaemmanuil E, Gerstung M, Malcovati L, et al; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood.* 2013;122(22):3616-3627.

7. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med.* 2014;371(26):2488-2498.

8. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med.* 2014;371(26):2477-2487.

9. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126(1):9-16.

10. Malcovati L, Karimi M, Papaemmanuil E, et al. SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood*. 2015;126(2):233-241.

11. Loss of the Y chromosome from normal and neoplastic bone marrows. United Kingdom Cancer Cytogenetics Group (UKCCG) [published correction appears in *Genes Chromosomes Cancer*. 1992;5(4):411]. *Genes Chromosomes Cancer*. 1992;5(1):83-88.

DOI 10.1182/blood-2015-10-672659

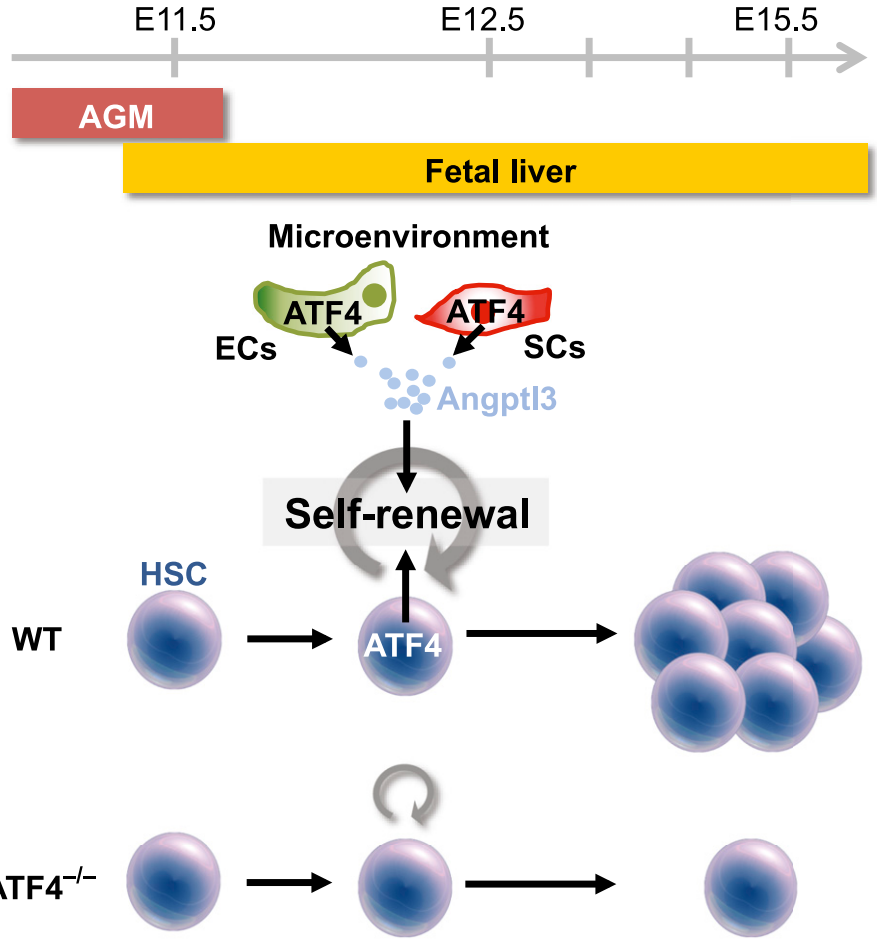
HEMATOPOIESIS AND STEM CELLS

Comment on Zhao et al, page 2383

ATF4, a new player in fetal HSC expansion

Michael A. Rieger GOETHE UNIVERSITY FRANKFURT; GERMAN CANCER CONSORTIUM (DKTK); GERMAN CANCER RESEARCH CENTER (DKFZ)

In this issue of *Blood*, Zhao et al have identified the basic region-leucine zipper transcription factor activating transcription factor 4 (ATF4) as a key molecule for the intrinsic and extrinsic regulation of the extensive expansion of fetal liver (FL) hematopoietic stem cells (HSCs) (see figure).[1]



The transcription factor ATF4 is essential for HSC self-renewal in the FL. Although the emergence and migration of definitive HSCs are largely independent of the presence of ATF4, the transcription factor is required for the intrinsic and extrinsic regulation of HSC self-renewal during their massive expansion phase between embryonic day 11.5 (E11.5) and E15.5 in the FL. ATF4 activates the expression and secretion of Angptl3 in the FL microenvironment, particularly in endothelial cells (ECs) and stroma cells (SCs), which further promotes the self-renewal of HSCs. AGM, aorta-gonad-mesonephros; WT, wild-type.

HSCs undertake an exciting journey during embryonic development from their first emergence to their final destination. The first definitive HSCs arise from hemogenic endothelial cells in the AGM region and in the placenta at E10.5 of mouse development.[2,3] They migrate to the FL and also to the spleen at E11.5 and start to massively renew themselves and thereby expand their numbers >100-fold within 4 to 5 days. This wave of HSC expansion provides the lifelong pool of stem cells for adulthood. From E15.5 until shortly after birth, HSCs settle in the bone marrow, where they largely reside in quiescence during adult hematopoiesis. In mice, there is a marked switch in their molecular program and in their functional behavior 3 to 4 weeks after birth. FL HSCs rapidly divide (1 division every 12-14 hours), while adult HSCs are deeply dormant and rarely divide (1 division every 145 days).[4]

The molecular profiling of FL long-term repopulating HSCs (LT-HSCs) during this massive expansion phase may provide the clues as to how we can manipulate adult HSCs to restart their self-renewal expansion for regenerative medicine. In adult HSCs, once they are "awoken" from their dormancy and enter cell cycle, they must rapidly return to their quiescent state to prevent exhaustion and long-term organ failure. Furthermore, extended replicative stress in adult HSCs holds the danger of DNA damage, which may occur once these cells are forced into cell cycle.[5] What makes fetal HSCs special enough to facilitate extensive self-renewal divisions without suffering from exhaustion, differentiation induction, or genomic instability? These are important questions that, once answered, may show us how to enforce HSC expansion for medical needs.

The study by Zhao et al introduces an important new player in the cell-intrinsic and -extrinsic regulation of FL HSC self-renewal and expansion. ATF4 is a basic region-leucine zipper transcription factor belonging to the ATF family which consists of 7 members in mice and humans. More than 10 years ago, the phenotype of the ATF4 homozygous knockout was reported in mice to result in perinatal lethality with a severe anemia and a low hematocrit.[6] The authors here revisited the consequences of ATF4 deletion on the biology of HSCs in great detail by investigating the intrinsically and extrinsically controlled cell fate decisions in the absence of ATF4 (see figure). They demonstrated that the FLs of

2015 126: 2349-2351
doi:10.1182/blood-2015-10-672659

Cytopenias + mutations – dysplasia = what?

David P. Steensma

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.

Appendix 8.10

Published paper and associated commentary from Cohort 2 data

blood®
Regular Article

MYELOID NEOPLASIA

The use of targeted sequencing and flow cytometry to identify patients with a clinically significant monocytosis

Catherine Cargo,[1] Matthew Cullen,[1] Jan Taylor,[1] Mike Short,[1] Paul Glover,[1] Suzan Van Hoppe,[1] Alex Smith,[2] Paul Evans,[1] and Simon Crouch[2]

[1]Haematological Malignancy Diagnostic Service, St James's University Hospital, Leeds, United Kingdom; and [2]Epidemiology and Cancer Statistics Group, University of York, York, United Kingdom

KEY POINTS

- Somatic mutations are detected at high frequency in patients with a monocytosis and are associated with significantly reduced survival.
- In those without a WHO-defined diagnosis, patients with a mutation have laboratory and clinical features indistinguishable from CMML.

The diagnosis of chronic myelomonocytic leukemia (CMML) remains centered on morphology, meaning that the distinction from a reactive monocytosis is challenging. Mutational analysis and immunophenotyping have been proposed as potential tools for diagnosis; however, they have not been formally assessed in combination. We aimed to investigate the clinical utility of these technologies by performing targeted sequencing, in parallel with current gold standard techniques, on consecutive samples referred for investigation of monocytosis over a 2-year period (N = 283). Results were correlated with the morphological diagnosis and objective outcome measures, including overall survival (OS) and longitudinal blood counts. Somatic mutations were detected in 79% of patients, being invariably identified in those with a confirmed diagnosis (99%) but also in 57% of patients with nondiagnostic bone marrow features. The OS in nondiagnostic mutated patients was indistinguishable from those with CMML (*P* = .118) and significantly worse than in unmutated patients (*P* = .0002). On multivariate analysis, age, ASXL1, CBL, DNMT3A, NRAS, and RUNX1 mutations retained significance. Furthermore, the presence of a mutation was associated with a progressive decrease in hemoglobin/platelet levels and increasing monocyte counts compared with mutation-negative patients. Of note, the immunophenotypic features of nondiagnostic mutated patients were comparable to CMML patients, and the presence of aberrant CD56 was highly specific for detecting a mutation. Overall, somatic mutations are detected at high frequency in patients referred with a monocytosis, irrespective of diagnosis. In those without a World Health Organization–defined diagnosis, the mutation spectrum, immunophenotypic features, and OS are indistinguishable from CMML patients, and these patients should be managed as such. (*Blood*. 2019;133(12):1325-1334)

Introduction

Distinguishing a reactive monocytosis from chronic myelomonocytic leukemia (CMML) is challenging for the hematopathologist. Using current World Health Organization (WHO) diagnostic criteria, a persistent monocytosis is the hallmark of disease, and demonstrating clonality is not a definitive requirement.[1] This leads to a greater risk for misdiagnoses or misclassification, particularly in patients with prolonged reactive changes.

More recently, alternative techniques, in particular flow cytometry, have provided a potential objective tool to identify patients with disease. Skewing of the distribution of monocyte subsets in the peripheral blood (PB; >94% M1 monocytes) has been reported to be sensitive and specific for detecting CMML.[2] In addition, large studies using targeted sequencing panels have identified recurrent somatic mutations in >90% of patients with CMML,[3] providing a further potential tool for diagnosis. The presence of a *TET2* mutation, in combination with a *SRSF2* (or *ZRSR2*) mutation, has been shown to be highly specific for a myelomonocytic phenotype[4]; these, along with *ASXL1*, are the most frequently mutated genes within this disease group.[3] Although the 2016 WHO diagnostic criteria have stated that these mutations can support a diagnosis of CMML, no study has directly assessed the use of this technology in a diagnostic setting. The aim of this study was to determine whether mutational analysis and flow cytometry can provide confirmatory evidence of disease and predict outcome in patients presenting with a monocytosis.

Methods

Patients and samples

The research was undertaken within the Haematological Malignancy Diagnostic Service (HMDS), a fully integrated laboratory that serves a population of ~6 million and is the benchmark for hematopathology services within the United Kingdom. All

consecutive samples (PB or bone marrow [BM]) received between July 2014 and July 2016 from patients ≥18 years old for the investigation of monocytosis were included. Patients with a confirmed myeloid diagnosis prior to July 2014 were excluded. The decision to investigate was at the discretion of the referring clinician; therefore, the study cohort reflects the variety of samples received in a routine laboratory for the investigation of a monocytosis. An absolute monocyte count was determined for all PB samples when received at HMDS (Table 1) using flow cytometry (see "Flow cytometry"). Interestingly, the absolute count was calculated to be $<1 \times 10^9$/L in a proportion of samples (11%); however, the vast majority were very close to this threshold, and review of local blood count parameters and clinical details confirmed the presence of a PB monocytosis and clinical suspicion of CMML. This highlights the recognized variation in monocyte counts between laboratories and the difficulty when applying arbitrary cutoffs as diagnostic criteria.

A total of 283 patients was referred during this time period (Table 1), of which 121 and 162 had an initial PB or BM sample, respectively (Figure 1). A confirmed diagnosis was only made on those cases with an ultimate BM sample (n = 207). All samples were processed according to gold standard techniques and were double reported, meaning the diagnosis was agreed upon by 2 experienced hematopathologists. Those with a confirmed diagnosis were classified in accordance with the WHO 2008 classification. Those failing to fulfill the morphological and genetic WHO 2008 criteria, as agreed by 2 hematopathologists, were classified as "nondiagnostic."

All samples were taken with fully informed patient consent for investigation of a suspected hematological disorder. This study had local Institutional Review Board approval (REC reference-16/NE/0105) and was performed in accordance with the Declaration of Helsinki.

Flow cytometry

All samples for immunophenotypic analysis were processed within 24 hours. Numerical studies and assessment of monocytic CD56 expression were performed on BM or PB samples following a stain–lyse–wash procedure (FACS Lyse; Becton Dickinson; supplemental Figures 1-2; supplemental Table 1, available on the *Blood* Web site). There was a strong correlation between monocyte CD56 expression in the PB and BM (supplemental Figure 5), enabling analysis using samples from either source.

PB CD14/CD16 "classical" monocytic subset studies were performed on samples following NH_4Cl lysis of erythrocytes using a lyse–stain–wash procedure. A minimum of 10^5 leukocytes was acquired on a single cytometer (FACSCanto II; Becton Dickinson) for all cases. Monocytes were identified using a combination of CD64, CD45, and scatter characteristics, and a single operator (M.C.) performed all analyses (supplemental Figures 3-4).

DNA extraction and targeted amplicon sequencing

In parallel with the above analyses, samples were subjected to targeted high-throughput sequencing. Referring clinicians and hematopathologists were blinded to the results of this analysis to exclude reporting or treatment bias.

DNA was extracted from fresh blood or BM mononuclear cells using a QIAamp DNA Mini Kit (QIAGEN, Manchester, UK).

Targeted gene sequencing of 27 genes recurrently mutated in myeloid malignancies was performed on a MiSeq System (Illumina, Chesterford, UK). Panel design, validation, and variant filtering criteria are included in supplemental Methods and supplemental Tables 2 and 3. The mean coverage of identified variants was 1514× (range, 52-5605×).

Clinical follow-up

All follow-up BM assessments were performed as clinically indicated by the referring clinician. These samples were also processed according to gold standard techniques and underwent targeted sequencing in parallel, as described above. Any new diagnoses were recorded.

Survival data were available for all patients and censored on the date of extraction (8 August 2017). Additional clinical information, including serial full blood count data, was collected on a subcohort of patients (n = 182), directly from the referring hospital or through the HMRN (n = 85).[5]

Statistical analysis

Survival curves were produced using the Kaplan-Meier method, and simple differences in survival were assessed with the log-rank test. The impact of abnormalities on overall survival (OS) and risk of progression were estimated using Cox regression; in cases in which variable selection was required to arrive at a multivariable regression, the lasso was used for variable selection, and results were reported for the corresponding relaxed lasso model.

Sensitivity, specificity, and positive and negative predictive values were calculated using 2×2 contingency tables.

Comparison between flow cytometric parameters in the main cohort was performed using the Mann-Whitney *U* test. Correlation among CD56 expression, M1 monocyte, and mutational analysis was performed using logistic and Poisson regression.

The effect of mutations on longitudinal blood counts was assessed using random effects models. Four models were fitted using a full-factorial interaction between time and mutation status: (1) a random intercept model, (2) a random intercept and slope model with uncorrelated random effects, (3) a random intercept and slope model with correlated random effects, and (4) a random intercept and slope model with correlated random effects, additionally adjusted for age and sex. For each mutation/blood count relationship, the best-fitting model was chosen according to a likelihood ratio test. To limit any potential effect from periods of acute illness or intensive treatment, blood count trajectory analysis was restricted to patients with <40 measurements over >100 days.

Results

Somatic mutations are detected at high frequency in patients with a monocytosis, irrespective of diagnosis

To define the mutation spectrum in patients referred with a monocytosis, targeted sequencing results were analyzed for the total cohort and correlated with the final diagnosis in those who underwent BM sampling. Of the total 283 patients, ≥1 mutation was detected in 78% of samples (the spectrum is presented in Figure 2A; see also supplemental Table 4). Of these patients,

Table 1. Patient characteristics

Characteristic	Distribution in cohort		
No. of patients	283		
Males/females, n	174/109		
Age, median (range), y	76 (24-96)		
Final diagnosis, n			
PB only	76		
CMML	114		
AML	11		
MPN	9		
MDS	4		
Other	4		
Nondiagnostic	65		
	CMML	**Other hematologic malignancy**	**Nondiagnostic**
Age, median (range), y	76 (24-91)	76 (42-93)	73 (34-93)
Blood count parameters: median (range)			
Hemoglobin (g/L)	105.5 (38-161)	108 (53-174)	122 (84-163)
White cell count ($\times 10^9$/L)	13.6 (4-104.9)	10.6 (3.9-83.4)	7.9 (4.2-38.2)
Platelets ($\times 10^9$/L)	90 (1-442)	154 (39-1085)	150 (8-499)
Monocytes ($\times 10^9$/L)*	2.69 (0.47-23.59)	1.71 (0.23-9.57)	1.29 (0.72-4.08)
Mutation frequency: no. of patients (%)			
TET2	72 (63)	9 (32)	31 (48)
SRSF2	48 (42)	9 (32)	14 (22)
ASXL1	39 (34)	13 (46)	10 (15)
NRAS	17 (15)	7 (25)	5 (8)
RUNX1	16 (14)	6 (21)	4 (6)
DNMT3A	9 (8)	5 (18)	5 (8)
CBL	18 (16)	1 (4)	4 (6)
KRAS	9 (8)	3 (11)	2 (3)
SETBP1	7 (6)	2 (7)	1 (2)
JAK2	3 (3)	7 (25)	1 (2)
EZH2	8 (7)	2 (7)	1 (2)
SF3B1	6 (5)	2 (7)	0 (0)

AML, acute myeloid leukemia; MDS, myelodysplastic syndrome; MPN, myeloproliferative neoplasm.

*Monocyte count was determined by flow cytometry (see "Methods").

207 underwent BM assessment for a definitive diagnosis. HMDS provides a centralized integrated hematopathology service, and all BMs were reviewed independently by 2 hematopathologists to ensure consistent and high-quality BM reporting for this purpose.

In those with a confirmed myeloid malignancy (142/207 cases; 69%), a mutation was almost invariably detected (140/142; 99% of cases). Of the 2 mutation-negative cases, 1 had a complex karyotype, including inv3 (involving *MECOM*), leaving only 1 case with no demonstrable clonal abnormality. The significant majority of diagnostic cases (80%; 114/142) were classified as CMML. The remaining samples were classified with a spectrum of myeloid malignancies, although, importantly, 11 patients were diagnosed with acute myeloid leukemia (AML) (n = 11), highlighting the importance of a BM assessment in patients referred with a monocytosis.

Somatic mutations were also detected at a high frequency in nondiagnostic samples. At least 1 mutation was detected in 37 of 65 patients (57%) with indeterminate features. The spectrum of mutations in this group mirrored those detected in the diagnostic group, with *TET2*, *SRSF2*, and *ASXL1* being the most frequently mutated (Figure 2B). The most notable differences in the nondiagnostic group were the absence of high-risk mutations, including *TP53*, *FLT3*, and *NPM1*, as well as those associated with specific morphological abnormalities, such as *SF3B1*, which correlates strongly with the presence of ring sideroblasts.[6] The median and mean number of mutations were higher in those with a confirmed diagnosis (median, 3; range, 0-8; mean, 3) vs those without (median, 1; range, 0-6; mean, 2) (Figure 2C). However, in patients with a confirmed mutation, the number of mutations did not differ significantly between diagnostic and nondiagnostic groups ($P = .62$).

The median variant allele fraction (VAF) for all variants was 39% (range, 5.2-100%; supplemental Figure 6), and there was no difference between the VAF in diagnostic and nondiagnostic cases ($P = .33$). In those patients with an isolated mutation, the

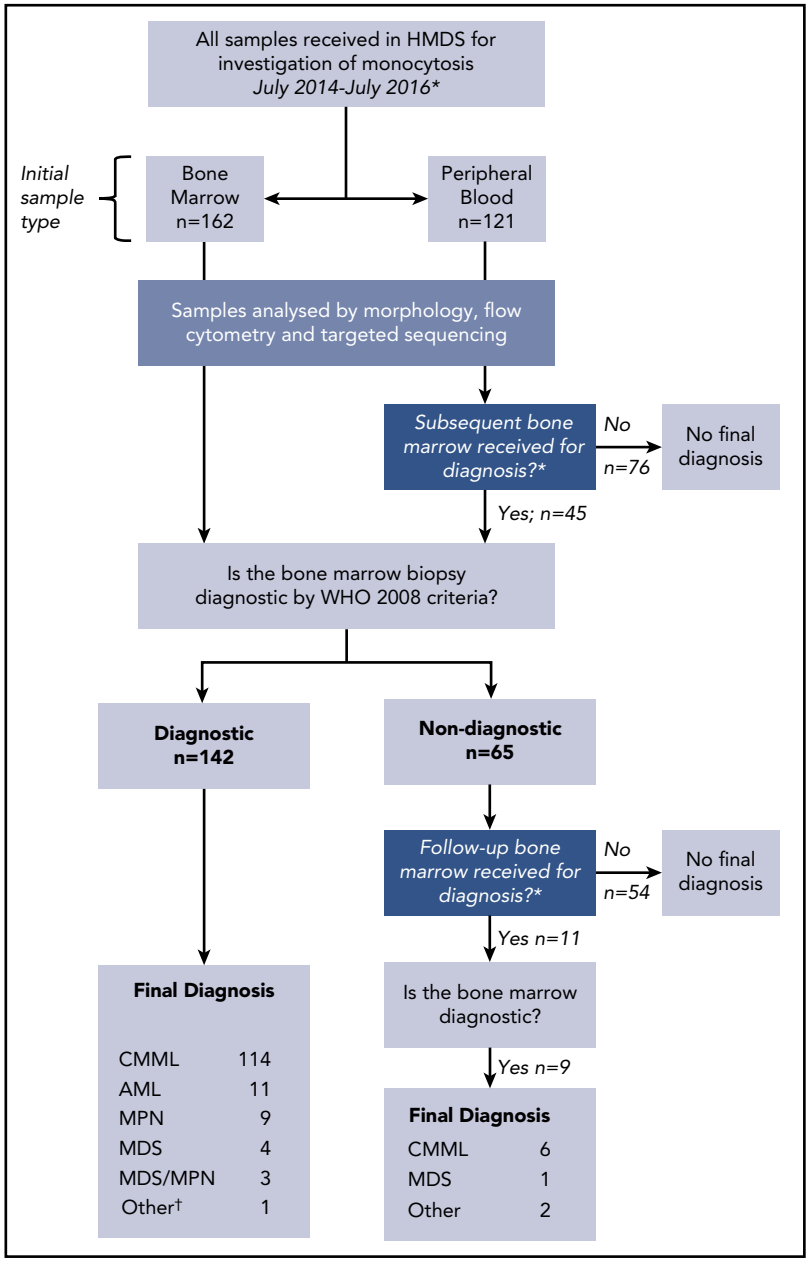


Figure 1. Summary of samples included in the study. Flowchart of cases referred to HMDS for investigation of a monocytosis. *Decision to investigate was at the discretion of the referring clinician. †Focal area of diffuse large B-cell lymphoma noted in BM, likely co-occurring with CMML.

median VAF was also noted to be high (38.2%; range, 6.3-97.1%), with only 2 variants having VAF $<$ 10%.

Therefore, mutations are found at a very high frequency with a high clonal burden in patients with a monocytosis, and they involve a similar spectrum of genes, irrespective of diagnosis.

OS and blood count trajectory correlate strongly with mutation profile

To understand the long-term clinical impact of detecting these mutations, objective outcome measures, including OS and longitudinal blood count analysis, were assessed in the total cohort and correlated with the final diagnosis.

The median survival of all patients from the time of first sampling was 35.2 months (95% confidence interval [CI] 25 months-not reached; Figure 3A). Survival correlated strongly with the number of mutations. Those without a mutation had a significantly better OS, and even the presence of a single mutation resulted in a significant reduction in survival (P = .004; Figure 3B). On univariate analysis, age was strongly associated with survival. For mutations occurring in >5% of subjects, *ASXL1*, *CBL*, *DNMT3A*, *NRAS*, and *RUNX1* were all strongly associated with survival, as were *EZH2* and *STAG2* among the less frequently mutated genes. To investigate multivariate significance, all genes mutated in >5% subjects were entered into a lasso survival regression. Taking the 1SE shrinkage parameter, age, *ASXL1*, *CBL*, *DNMT3A*, *NRAS*, and *RUNX1* were selected by the lasso and retained significance in a relaxed lasso regression (supplemental Table 5).

In those patients who proceeded to a BM biopsy, survival correlated with the final morphological diagnosis. Those without a confirmed diagnosis had a significantly better OS than did

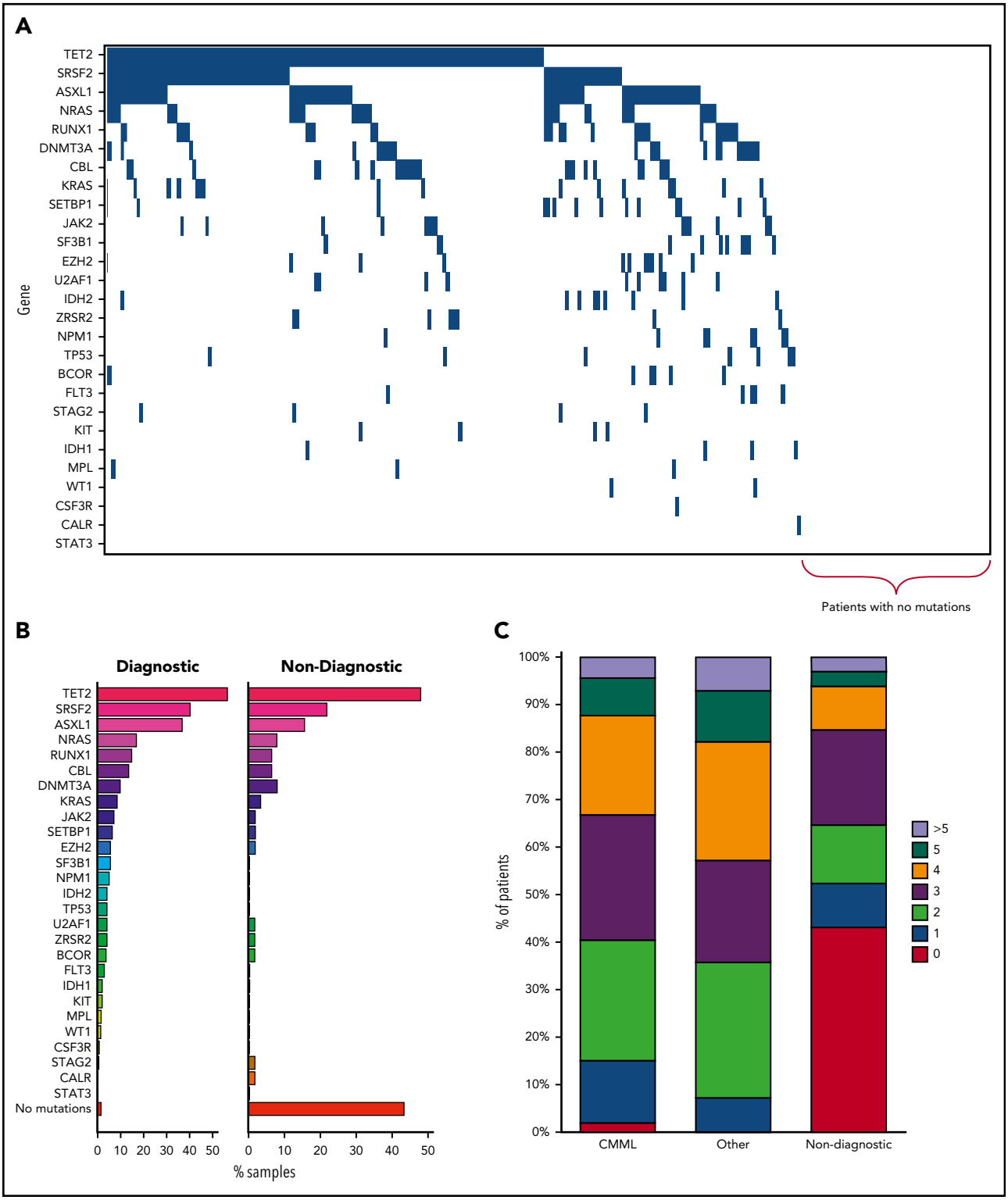


Figure 2. Characteristics of mutations detected in patient samples. (A) Spectrum of mutations detected across all patients in the study (N = 283). (B) Comparison of mutations detected in those with a diagnostic BM sample (n = 142) vs a nondiagnostic BM sample (n = 65). (C) Distribution of the number of mutations according to final diagnostic category. "Other" denotes those patients with an alternative hematological malignancy.

those with CMML or another myeloid malignancy. However, this survival benefit was retained only in those without a demonstrable mutation (P = .0002), with mutated patients having a similar survival to CMML patients (P = .118; not statistically significant) (Figure 4).

Longitudinal blood count data were available for 182 patients, although they were restricted to those with <40 measurements over >100 days (n = 133) to exclude periods of acute hospital admissions (due to periods of acute illness/infection) or intensive chemotherapy (median follow-up, 465 days; range, 119-996).

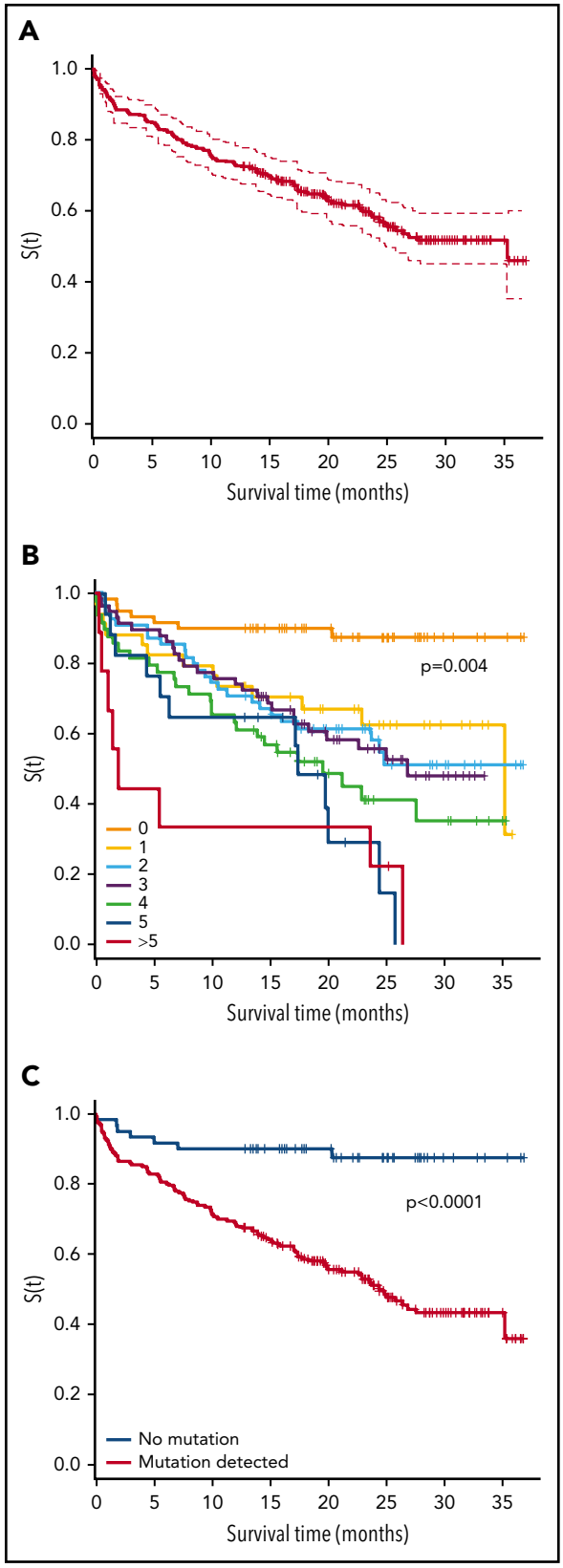


Figure 3. OS according to mutation number. (A) OS in total cohort from time of initial sample. (B) OS in total cohort by number of mutations detected at the time of initial sample. The *P* value was determined using the log-rank test to compare subjects without a mutation and subjects with a single mutation. (C) OS in total cohort by the presence or absence of a mutation. $P < .0001$, log-rank test.

The presence of a mutation was associated with a significantly lower hemoglobin and platelet count and a higher monocyte count relative to those without a mutation, which persisted over time and followed a divergent trajectory (Figure 5). With respect to individual mutations, certain mutations were associated with increasing or declining blood count parameters over time (supplemental Table 6). Monocyte counts were found to increase over time in subjects with *TET2*, *SRSF2*, *ASXL1*, *NRAS*, or *RUNX1* mutations relative to nonmutated subjects; similarly, white blood counts increased in subjects with *ASXL1*, *NRAS*, and *DNMT3A* mutations, and platelet levels decreased in subjects with *ASXL1*, *CBL*, and *RUNX1* mutations relative to nonmutated subjects.

In those without a confirmed diagnosis, follow-up BM biopsies were received from 11 patients. Importantly, of those with a subsequent diagnosis of CMML, all had a confirmed mutation on the original sample. In total, 7 of 37 (19%) nondiagnostic mutated patients had a confirmed diagnosis (6 CMML, 1 myelodysplastic syndrome [MDS]). Furthermore, none of the mutation-negative cases went on to develop CMML; however, 2 patients had confirmed alternative hematological diagnoses: diffuse large B-cell lymphoma and Rosai-Dorfman disease.

These findings confirm that the presence of a mutation has a significant impact on outcome with respect to survival and blood count parameters.

PB mutation profiling is predictive of a BM diagnosis

PB mutational analysis has been shown to correlate strongly with BM analysis in MDS, providing a potential alternative to BM sampling. To determine whether this is also true in CMML, matched PB and BM samples were analyzed. A total of 121 PB samples was received as the initial sample, and somatic mutations were detected in 66% (80/121). Forty-five of 121 patients (37%) had a subsequent BM biopsy performed for diagnosis. Sequencing failed on 2 of the matched BM samples. Of the 124 variants detected in the remaining 43 patients, there was high concordance between PB and BM (96%), with only 5 discordant results. Importantly, these were low-level variants at the limit of detection for the test or variants detected at areas of poor coverage (*SRSF2/ASXL1*). All 9 mutation-negative cases were fully concordant.

The presence of a mutation in the PB was highly predictive of diagnosing a myeloid malignancy in BM, with all but 1 case with a demonstrable mutation having a subsequent diagnosis (positive predictive value, 0.97; negative predictive value, 1.0; supplemental Figure 7). Of note, none of the mutation-negative ($n = 11$) cases had a subsequent confirmed diagnosis.

Immunophenotypic features correlate strongly with the presence of a mutation and a subsequent diagnosis

Flow cytometry has been proposed as a potential diagnostic tool in the investigation of patients with a monocytosis. To determine whether immunophenotyping can predict for the presence of a mutation or a BM diagnosis, flow cytometric analysis was performed alongside sequencing.

First, a comparison was made between the immunophenotypic features in the BM of those patients with a confirmed diagnosis

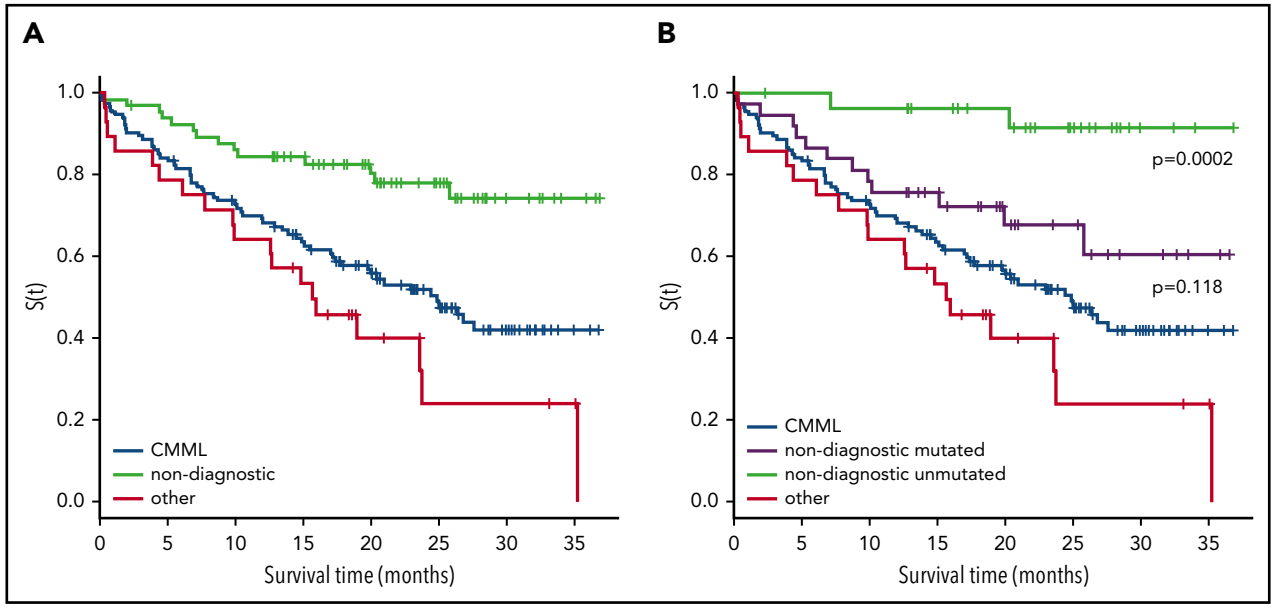


Figure 4. OS according to final diagnosis. (A) OS by diagnosis on BM sample (n = 207). (B) OS by diagnosis with nondiagnostic samples separated by the presence or absence of a mutation. $P = .0002$, CMML vs nondiagnostic unmutated patients, log-rank test; $P = .118$, CMML vs nondiagnostic mutated patients, log-rank test.

of CMML vs nondiagnostic samples. Importantly, nondiagnostic mutated patients had immunophenotypic features indistinguishable from CMML with respect to increased CD64$^+$ monocytes, reduced CD14 expression, and aberrant CD56 expression on monocytes (Figure 6). This was most pronounced with regard to CD56 expression (in PB or BM), which was found almost exclusively in those with a mutation. With respect to individual mutations, aberrant expression of CD56 was strongly associated with *TET2* mutations (odds ratio [OR], 4.0; 95% CI 2.4-6.8; $P < .0001$).

PB monocyte subsets and CD56 expression are predictive of a somatic mutation

The presence of >94% classical (M1) monocytes has been shown to be highly sensitive and specific for a diagnosis of CMML.[2] PB monocyte subset analysis was not available for every patient in the main cohort; therefore, to analyze the relationship among M1 monocytes, CD56 expression, and the mutation profile, a separate cohort of 135 patients was investigated. Of these 135 patients, 95 underwent a subsequent BM biopsy for definitive diagnosis (CMML = 28, MDS = 23, myeloproliferative neoplasm = 9, nondiagnostic = 27, other = 8). The presence of aberrant CD56 was again strongly associated with the presence of a mutation (OR, 43.9; 95% CI, 8.9-793.9; $P = .0003$). This was also noted, to a lesser extent, in association with having >94% M1 monocytes (OR, 3.9; 95% CI, 1.8-8.7; $P = .0007$) (supplemental Table 7). There was some correlation between the presence of CD56 expression and >94% M1 monocytes ($r = 0.17$; $P = .039$), and combining both produced a stronger effect (OR, 8.5; 95% CI, 3.9-19.5; $P < .00001$). Importantly, combining these phenotypic aberrancies did not capture all patients with a mutation. Although CD56 was highly specific for the presence of a mutation (98%), sensitivity was only 48%. Similarly, the presence of >94% M1 monocytes had a specificity of 75% for detecting a mutation, but the sensitivity was only 56%.

With respect to a confirmed diagnosis, CD56 expression (OR, 4.9; 95% CI, 1.9-13; $P = .001$) and >94% M1 monocytes (OR, 4.2; 95% CI, 1.7-11.5; $P = .003$) were associated with a final diagnosis of CMML; however, of note, 4 patients with CMML did not have either of these phenotypic aberrancies.

Discussion

This is the first study to formally examine the use of mutational analysis of patients presenting with a monocytosis. This was performed in combination with current gold standard techniques, including recently described flow cytometric analyses, in a large patient cohort. By analyzing sequential samples referred to a regional diagnostic laboratory, this study has investigated the typical patient population encountered in routine hematology practice. The use of objective outcome measures (longitudinal blood counts and OS) and an unselected patient population have minimized bias and ensured that the results are applicable in the "real-world" setting. Using a targeted sequencing panel of recurrently mutated genes, this study confirms that somatic mutations are identified in virtually all patients with a morphological diagnosis of CMML, as well as in a significant proportion of patients with a monocytosis and nondiagnostic features. It is possible that the proportion of nondiagnostic samples with detectable mutations was inflated as a result of referral bias and a high pretest probability of disease in those undergoing testing; however, these patients had a mutation spectrum, immunophenotype, and outcome indistinguishable from CMML. The presence of a mutation significantly impacted on survival, irrespective of the final diagnosis.

A number of technical limitations of this study should be highlighted. First, because these were routine samples referred for investigation, a corresponding germline sample was not available for analysis. The absence of reference material means that the distinction between germline variants or private single nucleotide polymorphisms and somatic variants is challenging; however, sequencing was limited to well-documented driver genes, and the landscape of mutations in these genes is well

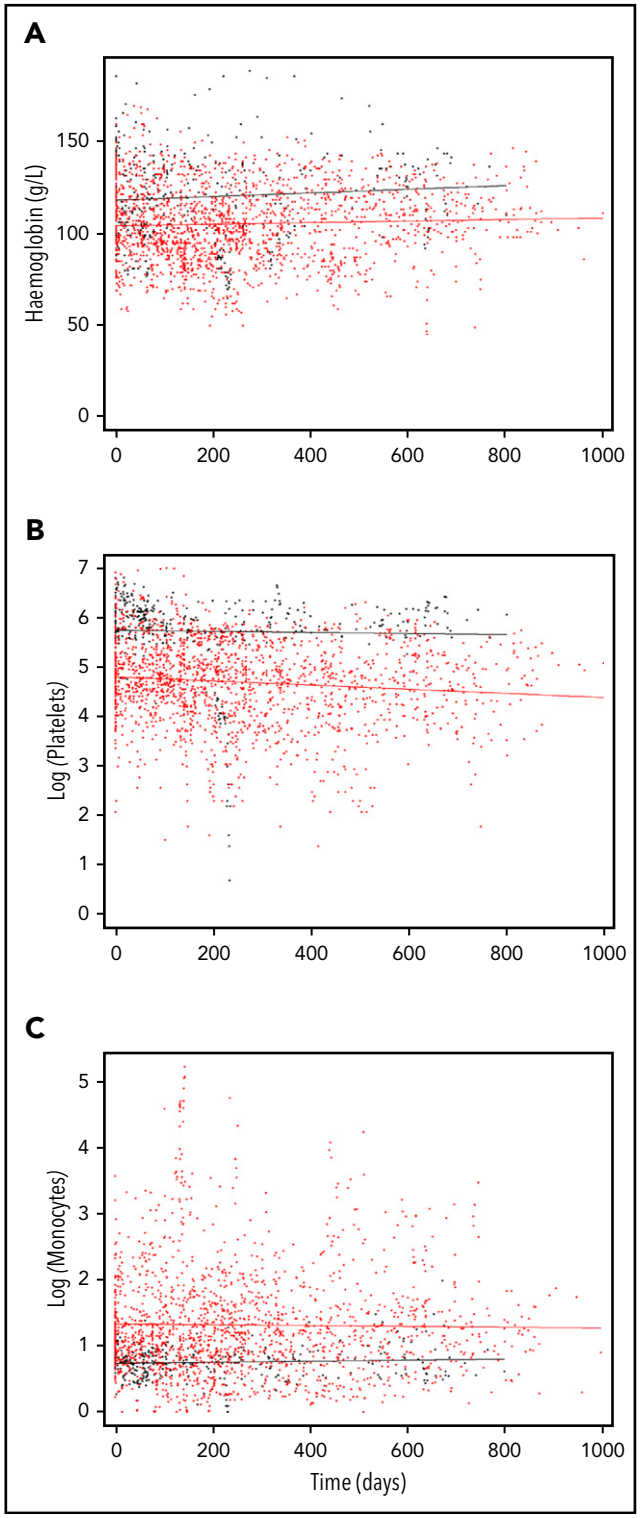
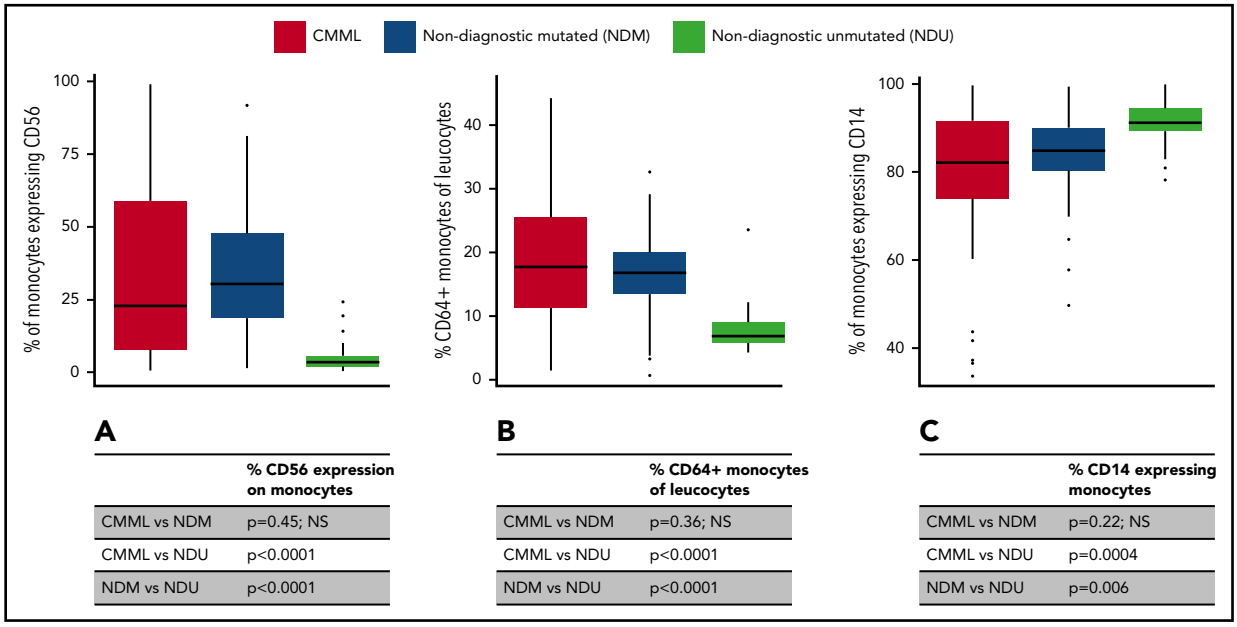


Figure 5. Longitudinal blood count trajectories in relation to mutation status. Plots of all blood count trajectories averaged between mutated (red) and unmutated (black) groups with overlaid linear regression line. Hemoglobin (g/L) (A), platelet count (log transformed) (B), and monocyte count (log transformed) (C) in patients with or without a detectable mutation.

established. Strict filtering criteria were applied (see supplemental Methods) to ensure that only high-confidence variants were included. Second, the sequencing analysis used amplicon-based library preparation, which has recognized limitations with respect to polymerase chain reaction errors and false-positive results, particularly at low VAF; however, the panel was validated internally and externally (see supplemental Methods), and only reproducible variants were included if detected at low VAF or in areas of low coverage. Therefore, the results are, to the best of our ability, accurate. In the future, deeper sequencing should enable more accurate variant calling at low VAF.

The findings of this study will be key to refining future diagnostic algorithms in the investigation of patients referred with a monocytosis. Mutational analysis has been incorporated into the recent amendment of the WHO diagnostic criteria, which now state that the presence of a mutation can support a diagnosis of CMML. However, concerns have been raised regarding the use of mutational analysis in this setting because of reports of frequent somatic mutations in aging healthy individuals.[7-10] As a result, the WHO has stated that the presence of a mutation in CMML or MDS should not be used alone as proof of disease.[1] However, our study has shown that, even in the absence of morphological features, those patients with a mutation had a clinical phenotype and genotype indistinguishable from CMML and a comparably poor outcome. Distinguishing features were also noted between the variants reported in healthy individuals and the mutations detected in our study group. The VAF or clone size of the mutations in our study was significantly higher than in healthy individuals (median 39.2% vs 9%-10%), and this was demonstrated across diagnostic and nondiagnostic samples. This finding has also been described in patients with unexplained cytopenias, and several studies have shown that VAF $>$ 10% and the presence of co-occurring mutations can distinguish clinically significant cytopenias from healthy individuals.[11-13] Although the higher VAF is replicated in our patient group, importantly, our study has shown that even isolated mutations have a significant impact on survival in patients with a monocytosis. These findings provide strong evidence that, in those subjects without diagnostic morphological features, the presence of a mutation, irrespective of mutation number, could be disease defining. At the very minimum, it is imperative that these patients are identified and monitored closely.

It has become increasingly feasible to perform mutational analysis in routine clinical practice, and this study has demonstrated how modest-sized gene panels can provide significant diagnostic and prognostic information. The panel used in the study targeted genes implicated in myeloid malignancies and was incorporated into the routine workload and performed in "real time" in a cost-effective manner. The genetic profile in CMML is now well established and is noted to be relatively homogeneous, involving only a restricted number of genes. Mutation frequencies in $>$90% of patients have been consistently reported using varying panel sizes, including as few as 19 genes.[3,14-16] The mutation profile in our cohort mirrored that reported in the literature; despite the restricted panel, the mutation frequency was high, and a significant impact on outcome was demonstrated. The recognized poor prognostic impact of *ASXL1* mutations[3,14,17-20] was also replicated across this data set. Therefore, mutational analysis is viable in a routine diagnostic laboratory. It is also likely that a proportion of these patients will have additional mutations in genes not sequenced in this study. To further investigate this would require more extensive sequencing on much larger patient populations.



	% CD56 expression on monocytes
CMML vs NDM	p=0.45; NS
CMML vs NDU	p<0.0001
NDM vs NDU	p<0.0001

	% CD64+ monocytes of leucocytes
CMML vs NDM	p=0.36; NS
CMML vs NDU	p<0.0001
NDM vs NDU	p<0.0001

	% CD14 expressing monocytes
CMML vs NDM	p=0.22; NS
CMML vs NDU	p=0.0004
NDM vs NDU	p=0.006

Figure 6. Relationship between immunophenotype and mutations. Box-and-whisker plots comparing immunophenotypic features of CMML, nondiagnostic mutated cases (NDM), and nondiagnostic unmutated cases (NDU). (A) Percentage of monocytes expressing CD56. (B) Percentage of $CD64^+$ monocytes among leukocytes. (C) Percentage of CD14-expressing monocytes. The *P* values were determined using the Mann-Whitney *U* test. NS, not statistically significant.

The potential for PB to be used as a screening tool for monocytosis has also been addressed in this study. This is an attractive option, particularly in a disease commonly presenting in the older patient population. Using flow cytometry, the presence of >94% M1 monocytes in the PB was reported to be highly sensitive and specific for CMML.[2] Subsequent studies have validated these findings and also confirmed the ability to distinguish CMML from MDS and myeloproliferative neoplasm cases presenting with a monocytosis.[21,22] However, these studies are centered on morphological diagnoses, and mutational analyses have not been performed consistently. Although our study has shown a strong correlation between skewed monocyte subsets and a diagnosis of CMML, this did not capture all patients and was neither sensitive nor specific for the presence of a mutation. In contrast, aberrant CD56 expression was highly specific for the presence of a mutation (98%), particularly involving *TET2*. CD56 expression has been reported to be highly sensitive and specific for a diagnosis of CMML (100% and 67%, respectively) when combined with other immunophenotypic features, including reduced expression of myeloid antigens and ≥20% immature monocytes[23]; however, subsequent studies raised concerns regarding the overexpression of CD56 in reactive conditions.[24] Our data show that CD56 expression at diagnosis is invariably associated with the presence of a somatic mutation, although sensitivity was low (48%). Therefore, flow cytometry could provide a screening tool for the investigation of PB monocytes; however, ultimately, mutational analysis will be required to identify patients who require clinical follow-up.

Importantly, there was high concordance between PB and BM mutational analysis, and the presence of a PB mutation was highly predictive of a subsequent BM diagnosis. This suggests that screening of the PB may be a suitable method for identifying or excluding significant mutations; however, this could lead to an increase in inappropriate referrals and a significant burden on laboratory personnel. Furthermore, the small proportion of mutated patients in our cohort with other hematological malignancies in the BM, including AML, highlights the importance of a baseline BM assessment to definitively classify the disease. In contrast, the negative predictive value of PB screening was 100%, suggesting that those without a mutation should not undergo BM assessment. In the first instance, PB screening would be a practical option in those patients unfit for BM assessment or potentially to monitor for treatment response or disease evolution. The latter would require further investigation in a prospective study.

In conclusion, this study has confirmed that mutations are commonly detected in patients referred with a persistent monocytosis. The presence of a mutation impacts significantly on outcome, irrespective of diagnosis, and patients with a mutation who fail to meet WHO criteria have CMML disease characteristics. These findings validate the inclusion of somatic mutations in the diagnostic criteria for CMML, and, at the very minimum, those without a confirmed diagnosis require close clinical follow-up. Although PB can be confidently used to detect mutations, a baseline BM biopsy is required for definitive disease classification in patients fit for treatment. Immunophenotypic assessment of monocytes may provide a potential screening tool to detect those with a mutation; however, it will miss a proportion of mutated patients. Ultimately, early identification of patients could provide an opportunity for intervention in this patient group, and this requires further investigation.

Acknowledgments

The authors thank all clinicians and laboratory staff who assisted in collecting longitudinal and outcome data.

A.S. and S.C. are supported by a Bloodwise program grant.

Authorship

Contribution: C.C. and S.C. designed the study and wrote the manuscript; C.C., M.C., M.S., S.V.H., and P.E. performed the research; and C.C., M.C., J.T., P.G., A.S., and S.C. analyzed the data.

Conflict-of-interest disclosure: C.C. and P.E. have served on advisory boards for Novartis. The remaining authors declare no competing financial interests.

ORCID profile: C.C., 0000-0002-5185-0190.

Correspondence: Catherine Cargo, Haematological Malignancy Diagnostic Service, St James's University Hospital, Beckett St, Leeds LS9 7TF, United Kingdom; e-mail: catherine.cargo@nhs.net.

Footnotes

Submitted 3 August 2018; accepted 18 December 2018. Prepublished online as *Blood* First Edition paper, 3 January 2019; DOI 10.1182/blood-2018-08-867333.

Presented in poster form at the 58th American Society of Hematology Annual Meeting and Exposition, San Diego, CA, 4 December 2016.

The online version of this article contains a data supplement.

There is a *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

REFERENCES

1. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405.
2. Selimoglu-Buet D, Wagner-Ballon O, Saada V, et al; Francophone Myelodysplasia Group. Characteristic repartition of monocyte subsets as a diagnostic signature of chronic myelomonocytic leukemia. *Blood*. 2015;125(23):3618-3626.
3. Elena C, Gallì A, Such E, et al. Integrating clinical features and genetic lesions in the risk assessment of patients with chronic myelomonocytic leukemia. *Blood*. 2016;128(10):1408-1417.
4. Malcovati L, Papaemmanuil E, Ambaglio I, et al. Driver somatic mutations identify distinct disease entities within myeloid neoplasms with myelodysplasia. *Blood*. 2014;124(9):1513-1521.
5. Roman E, Smith A, Appleton S, et al. Myeloid malignancies in the real-world: occurrence, progression and survival in the UK's population-based Haematological Malignancy Research Network 2004-15. *Cancer Epidemiol*. 2016;42:186-198.
6. Papaemmanuil E, Cazzola M, Boultwood J, et al; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med*. 2011;365(15):1384-1395.
7. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*. 2014;371(26):2477-2487.
8. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371(26):2488-2498.
9. McKerrell T, Park N, Moreno T, et al; Understanding Society Scientific Group. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Reports*. 2015;10(8):1239-1245.
10. Xie M, Lu C, Wang J, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med*. 2014;20(12):1472-1478.
11. Cargo CA, Rowbotham N, Evans PA, et al. Targeted sequencing identifies patients with preclinical MDS at high risk of disease progression. *Blood*. 2015;126(21):2362-2365.
12. Kwok B, Hall JM, Witte JS, et al. MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance. *Blood*. 2015;126(21):2355-2361.
13. Malcovati L, Gallì A, Travaglino E, et al. Clinical significance of somatic mutation in unexplained blood cytopenia. *Blood*. 2017;129(25):3371-3378.
14. Itzykson R, Kosmider O, Renneville A, et al. Prognostic score including gene mutations in chronic myelomonocytic leukemia. *J Clin Oncol*. 2013;31(19):2428-2436.
15. Mason CC, Khorashad JS, Tantravahi SK, et al. Age-related mutations and chronic myelomonocytic leukemia. *Leukemia*. 2016;30(4):906-913.
16. Kohlmann A, Grossmann V, Klein HU, et al. 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. *J Clin Oncol*. 2010;28(24):3858-3865.
17. Patnaik MM, Padron E, LaBorde RR, et al. Mayo prognostic model for WHO-defined chronic myelomonocytic leukemia: ASXL1 and spliceosome component mutations and outcomes [published correction appears in *Leukemia*. 2013;27(10):2112]. *Leukemia*. 2013;27(7):1504-1510.
18. Patnaik MM, Lasho TL, Vijayvargiya P, et al. Prognostic interaction between ASXL1 and TET2 mutations in chronic myelomonocytic leukemia. *Blood Cancer J*. 2016;6(1):e385.
19. Patnaik MM, Wassie EA, Padron E, et al. Chronic myelomonocytic leukemia in younger patients: molecular and cytogenetic predictors of survival and treatment outcome [published correction appears in *Blood Cancer J*. 2015;5:e280]. *Blood Cancer J*. 2015;5(1):e270.
20. Patnaik MM, Itzykson R, Lasho TL, et al. ASXL1 and SETBP1 mutations and their prognostic contribution in chronic myelomonocytic leukemia: a two-center study of 466 patients. *Leukemia*. 2014;28(11):2206-2212.
21. Patnaik MM, Timm MM, Vallapureddy R, et al. Flow cytometry based monocyte subset analysis accurately distinguishes chronic myelomonocytic leukemia from myeloproliferative neoplasms with associated monocytosis. *Blood Cancer J*. 2017;7(7):e584.
22. Talati C, Zhang L, Shaheen G, et al. Monocyte subset analysis accurately distinguishes CMML from MDS and is associated with a favorable MDS prognosis. *Blood*. 2017;129(13):1881-1883.
23. Xu Y, McKenna RW, Karandikar NJ, Pildain AJ, Kroft SH. Flow cytometric analysis of monocytes as a tool for distinguishing chronic myelomonocytic leukemia from reactive monocytosis. *Am J Clin Pathol*. 2005;124(5):799-806.
24. Lacronique-Gazaille C, Chaury MP, Le Guyader A, Faucher JL, Bordessoule D, Feuillard J. A simple method for detection of major phenotypic abnormalities in myelodysplastic syndromes: expression of CD56 in CMML. *Haematologica*. 2007;92(6):859-860.

2019 133: 1325-1334
doi:10.1182/blood-2018-08-867333 originally published online January 3, 2019

The use of targeted sequencing and flow cytometry to identify patients with a clinically significant monocytosis

Catherine Cargo, Matthew Cullen, Jan Taylor, Mike Short, Paul Glover, Suzan Van Hoppe, Alex Smith, Paul Evans and Simon Crouch

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.

Comment on Cargo et al, page 1325

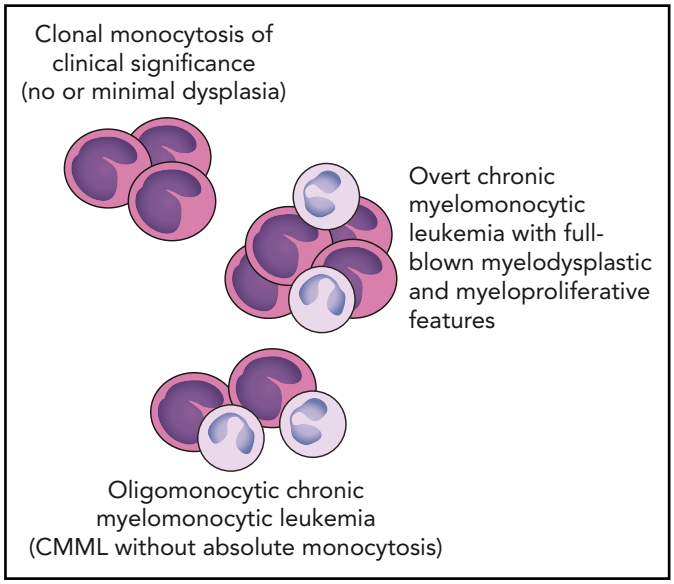
Clonal monocytosis of clinical significance

Mario Cazzola | University of Pavia

The World Health Organization (WHO) diagnostic criteria for chronic myelomonocytic leukemia (CMML) include clinical and morphological features; however, demonstrating clonality is not an absolute requirement for making the diagnosis.[1] In this issue of *Blood*, Cargo et al show that patients with clonal monocytosis identified by targeted gene sequencing have a clinical outcome similar to that of overt WHO-defined CMML.[2]

In the WHO criteria, CMML is classified as a myelodysplastic/myeloproliferative neoplasm (MDS/MPN), a category that also includes atypical chronic myeloid leukemia, juvenile myelomonocytic leukemia, and the MDS/MPN with ring sideroblasts and thrombocytosis.[1] These disorders have both myelodysplastic (dysplasia and cytopenia) and myeloproliferative features ("cytosis" of 1 or more myeloid lineages) at the time of diagnosis.

CMML is characterized by the accumulation of monocytes in the peripheral blood, and therefore, the initial diagnostic approach involves the differential diagnosis of monocytosis. Once reactive monocytosis has been excluded, the possibility of CMML should be considered, especially if the elevated monocyte count has persisted for ≥3 months. According to the WHO criteria, diagnosis of CMML requires an absolute monocyte count ≥1 × 10^9/L with monocytes accounting for ≥10% of circulating leukocytes. These cutoffs are arbitrary: the natural history of disease is that the monocyte count increases from normal to elevated in a continuous manner. Monocytosis can be present in other myeloid malignancies, such as MPNs, and therefore, diagnosis of CMML requires the exclusion of these conditions. To establish the myelodysplastic nature of the disease, the presence of dysplasia involving ≥1 myeloid lineages is required, whereas blasts must constitute <20% of the cells in the peripheral blood and bone marrow.



Relationship between clonal monocytosis of clinical significance, oligomonocytic CMML, and overt CMML. The number of monocytes reflects monocytic proliferation, whereas myelodysplasia is represented by neutrophils with hypogranulated cytoplasm and bilobed nucleus. Somatic mutations in genes like *TET2*, *SRSF2*, *ASXL1*, *NRAS*, *KRAS*, *CBL*, or *SETBP1* represent the common thread of these chronic myeloid neoplasms, whereas epigenetic factors may be responsible for the phenotypic variability. Professional illustration by Patrick Lane, ScEYEnce Studios.

As CMML lacks a unique disease-defining genetic lesion, genetic data have so far played a minor role in the diagnosis.[1] About three-quarters of patients have a normal karyotype, which means that cytogenetic abnormalities can be used as clonal markers only in a subset of patients.[3] Somatic gene mutations have been identified only in the last few years. A recent study using a panel of 38 recurrently mutated genes in myeloid malignancies has detected somatic mutations in 199 of 214 CMML patients (93%).[4] The most frequently mutated genes were *TET2*, *SRSF2*, *ASXL1*, *NRAS*, *KRAS*, and *SETBP1*. A significant association was found between mutations in *TET2* and spliceosome genes, and one-fifth of patients showed cooccurrence of *TET2* and *SRSF2* mutations, a comutation pattern that can be considered relatively typical of CMML. Quantification of monocyte subsets by flow cytometry has recently provided a new tool for the diagnosis of CMML.[5] An increase in the fraction of classical monocytes ($CD14^{++}/CD16^{-}$) to >94.0% of total monocytes has been found to be a biomarker that helps distinguish CMML from reactive monocytosis.

Cargo et al conducted a study that generated from routine hematology practice. They studied samples of patients referred to a hematology service for monocytosis. Through targeted sequencing of 27 genes recurrently mutated in myeloid malignancies, they detected ≥1 somatic mutation in 221 of 283 samples (78%). Overall, 207 subjects underwent additional tests, including bone marrow assessment, for a definitive diagnosis. Virtually all patients with a confirmed myeloid neoplasm carried a somatic mutation (140/142; 99% of cases), and most of them had CMML (114/142; 80% of cases). Of the 65 subjects who did not have a definitive diagnosis but just indeterminate features, 37 (57% of cases) carried at least 1 somatic mutation, with *TET2*, *SRSF2*, and *ASXL1* being the most frequently mutated genes. In terms of variant allele frequency (VAF), there was no significant difference between the diagnostic and nondiagnostic/indeterminate features groups, with average values ~40%. More importantly, the overall survival of mutated nondiagnostic patients

was indistinguishable from that of patients with WHO-defined CMML and worse than that of subjects with monocytosis without somatic mutations. Flow cytometry analysis of circulating monocytes showed overlapping features in mutated nondiagnostic subjects and CMML patients.

The study by Cargo et al validates the current WHO diagnostic criteria for CMML, showing that when myelodysplasia is absent or minimal, the diagnosis of CMML may still be made if a somatic mutation is present. In fact, many of the cases that were initially considered nondiagnostic based on morphological criteria were identified by demonstrating an acquired clonal genetic abnormality, as stipulated in the recently revised WHO criteria.[1] The conclusions of the study, however, go beyond this validation and suggest that the presence of a somatic mutation should become an absolute requirement for diagnosis of CMML, irrespective of the presence or absence of dysplasia. The fact that somatic mutations of myeloid genes can be found also in healthy individuals with age-related clonal hematopoiesis (ARCH)[6] does not represent a valid reason for not using them as markers of clonality in myeloid neoplasms. In both patients with clonal monocytosis and those with CMML, the VAF of somatic mutations was much higher (~40% on average) than that commonly observed in healthy subjects with ARCH (<10%), indicating a much more advanced clonal disease.

Through their investigations, Cargo et al have illuminated a condition that can be defined as "clonal monocytosis of clinical significance." The relationship between this condition and CMML resembles that between clonal cytopenia of undetermined significance (CCUS) and MDS.[7,8] The overall survival and the risk of disease progression of patients with CCUS and highly specific mutation patterns are indistinguishable from those of patients with a myeloid neoplasm with myelodysplasia.[9] While clonal monocytosis of clinical significance lacks overt myelodysplasia, an oligomonocytic CMML has also been described that displays a similar clinicopathologic and mutational profile to classical CMML.[10] Somatic mutations represent the common thread of all these conditions, which are schematically represented (see figure).

In conclusion, the available evidence suggests that demonstrating somatic mutations and defining their patterns may provide presumptive evidence of myeloid malignancies, specifically, of CMML, even in the absence of definitive morphological criteria. In addition, integrating clinical features, morphology, immunophenotyping, and gene mutations may also improve risk stratification of these patients, providing a robust basis for clinical decision making and a reliable tool for clinical trials.[4]

Conflict-of-interest disclosure: The author declares no competing financial interests. ■

REFERENCES

1. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405.
2. Cargo C, Cullen M, Taylor J, et al. The use of targeted sequencing and flow cytometry to identify patients with a clinically significant monocytosis. *Blood*. 2019;133(12):1325-1334.
3. Such E, Cervera J, Costa D, et al. Cytogenetic risk stratification in chronic myelomonocytic leukemia. *Haematologica*. 2011;96(3):375-383.
4. Elena C, Gallì A, Such E, et al. Integrating clinical features and genetic lesions in the risk assessment of patients with chronic myelomonocytic leukemia. *Blood*. 2016;128(10):1408-1417.
5. Selimoglu-Buet D, Wagner-Ballon O, Saada V, et al; Francophone Myelodysplasia Group. Characteristic repartition of monocyte subsets as a diagnostic signature of chronic myelomonocytic leukemia. *Blood*. 2015;125(23):3618-3626.
6. Shlush LI. Age-related clonal hematopoiesis. *Blood*. 2018;131(5):496-504.
7. Kwok B, Hall JM, Witte JS, et al. MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance. *Blood*. 2015;126(21):2355-2361.
8. Cargo CA, Rowbotham N, Evans PA, et al. Targeted sequencing identifies patients with preclinical MDS at high risk of disease progression. *Blood*. 2015;126(21):2362-2365.
9. Malcovati L, Gallì A, Travaglino E, et al. Clinical significance of somatic mutation in unexplained blood cytopenia. *Blood*. 2017;129(25):3371-3378.
10. Geyer JT, Tam W, Liu YC, et al. Oligomonocytic chronic myelomonocytic leukemia (chronic myelomonocytic leukemia without absolute monocytosis) displays a similar clinicopathologic and mutational profile to classical chronic myelomonocytic leukemia. *Mod Pathol*. 2017;30(9):1213-1222.

DOI 10.1182/blood-2019-01-896084

PLATELETS AND THROMBOPOIESIS

Comment on Marconi et al, page 1346

PTPRJ: a novel inherited thrombocytopenia gene

Renren Wen and Demin Wang | BloodCenter of Wisconsin

In this issue of *Blood*, Marconi et al use a high-throughput exome-sequencing approach to identify 2 biallelic loss-of-function mutations in PTPRJ that caused autosomal-recessive thrombocytopenia and a bleeding disorder in 2 siblings.[1]

Inherited thrombocytopenia (IT) is an extremely heterogeneous group of thrombocytopenic conditions. Classification of IT based on the inheritance pattern or clinical symptoms other than thrombocytopenia is not always reliable due to the high frequency of sporadic cases with de novo gene mutations, partial penetrance of the mutations, and variable modes of presentation in patients with the same gene mutations. Classification of IT based on platelet size can be helpful and relatively reliable.[2] However, characterization of clinical and laboratory findings that correlate with an identified genetic abnormality is essential to define a particular IT as a specific disease entity.

ITs have various phenotypes and are caused by mutations in many different genes. The genetic defects responsible for an IT were first defined in 2 conditions: Bernard-Soulier syndrome (BSS) and Wiskott-Aldrich syndrome (WAS). BSS typically is associated with a severe bleeding tendency and is caused by mutations of genes encoding the components of glycoprotein (GP) complex

2019 133: 1271-1272
doi:10.1182/blood-2019-01-896084

Clonal monocytosis of clinical significance

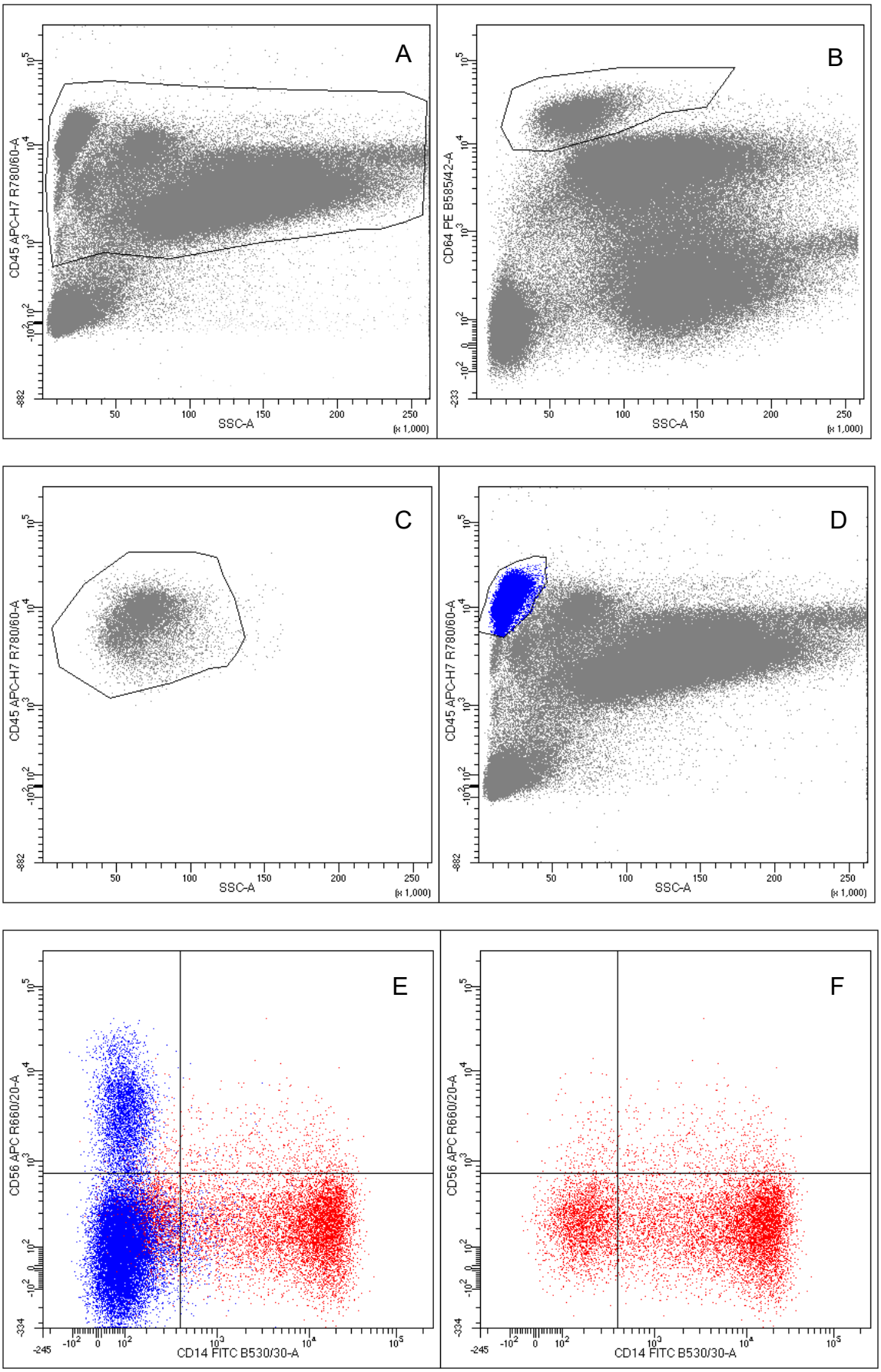
Mario Cazzola

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.

Appendix 8.11

Gating strategy for monocyte analysis of Cohort 2 data

Appendix 8.12 - Gating strategy for monocyte analysis of Cohort 2



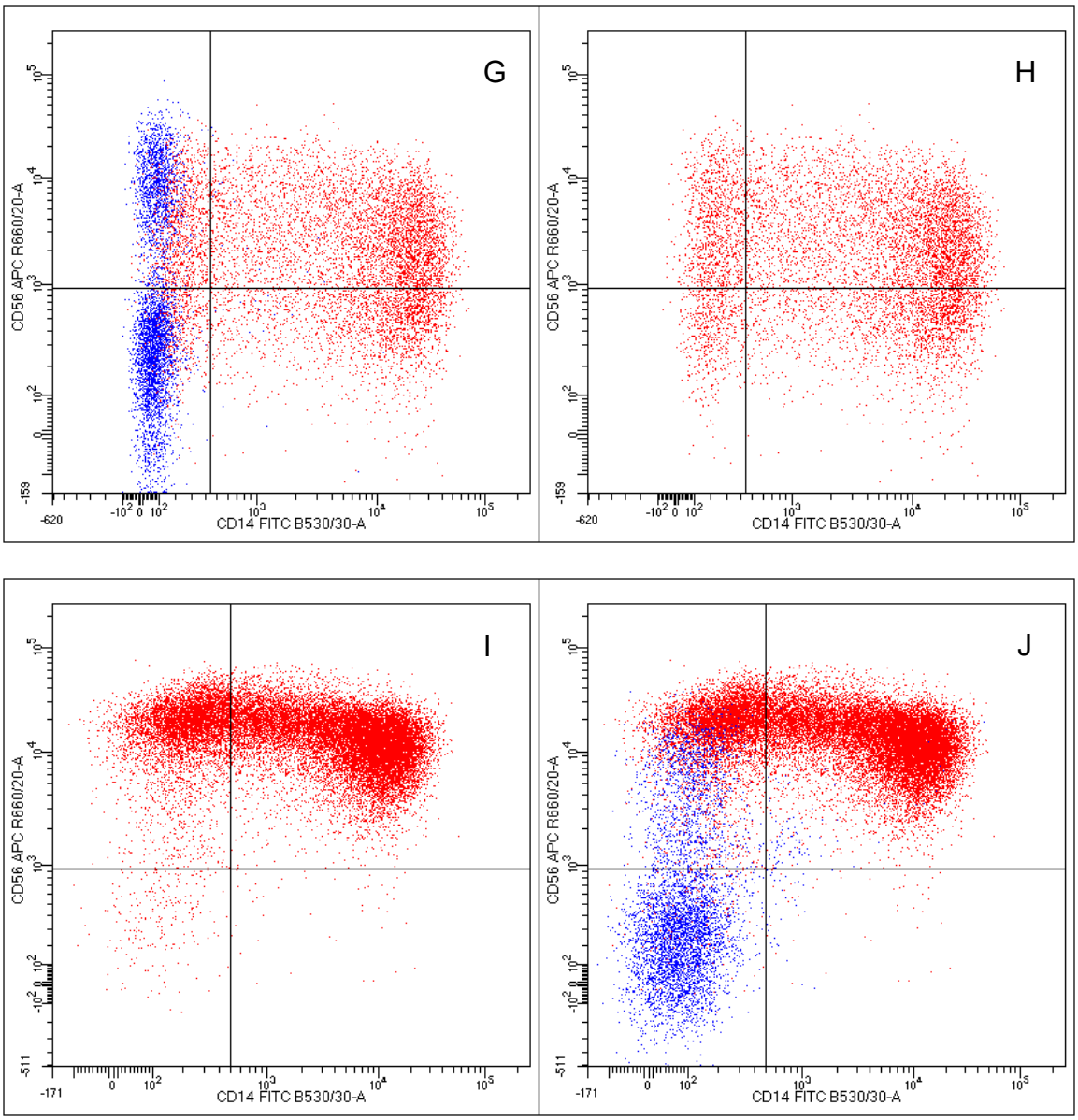


Figure S1. Flow cytometry gating strategy used to discriminate monocytes. Leucocytes were identified on the basis of CD45 and side scatter (SSC) expression (A). Within the leucocyte population, monocytic cells were defined on the basis of CD64 expression and side scatter characteristics (B). Confirmation was performed using CD45 expression (C). Monocytes were then enumerated as a percentage of $CD45^{+}$ leucocytes . Lymphocytes were used to control for the expression of CD56 and CD14 on the monocytic cells and these were identified on the basis of CD45 and side scatter characteristics (D)
Lymphocytes (blue) were used to define negative population quadrants for CD14 and CD56 expression on the monocytes (red) (E). In this example, 73% of monocytes express CD14 whilst <20% (4.8%) express CD56 (F).
In this example, 71% of monocytes express CD14 whilst >20% (78%) express CD56 (G&H).
In this example, 72% of monocytes express CD14 whilst <20% (99%) express CD56 (I&J).

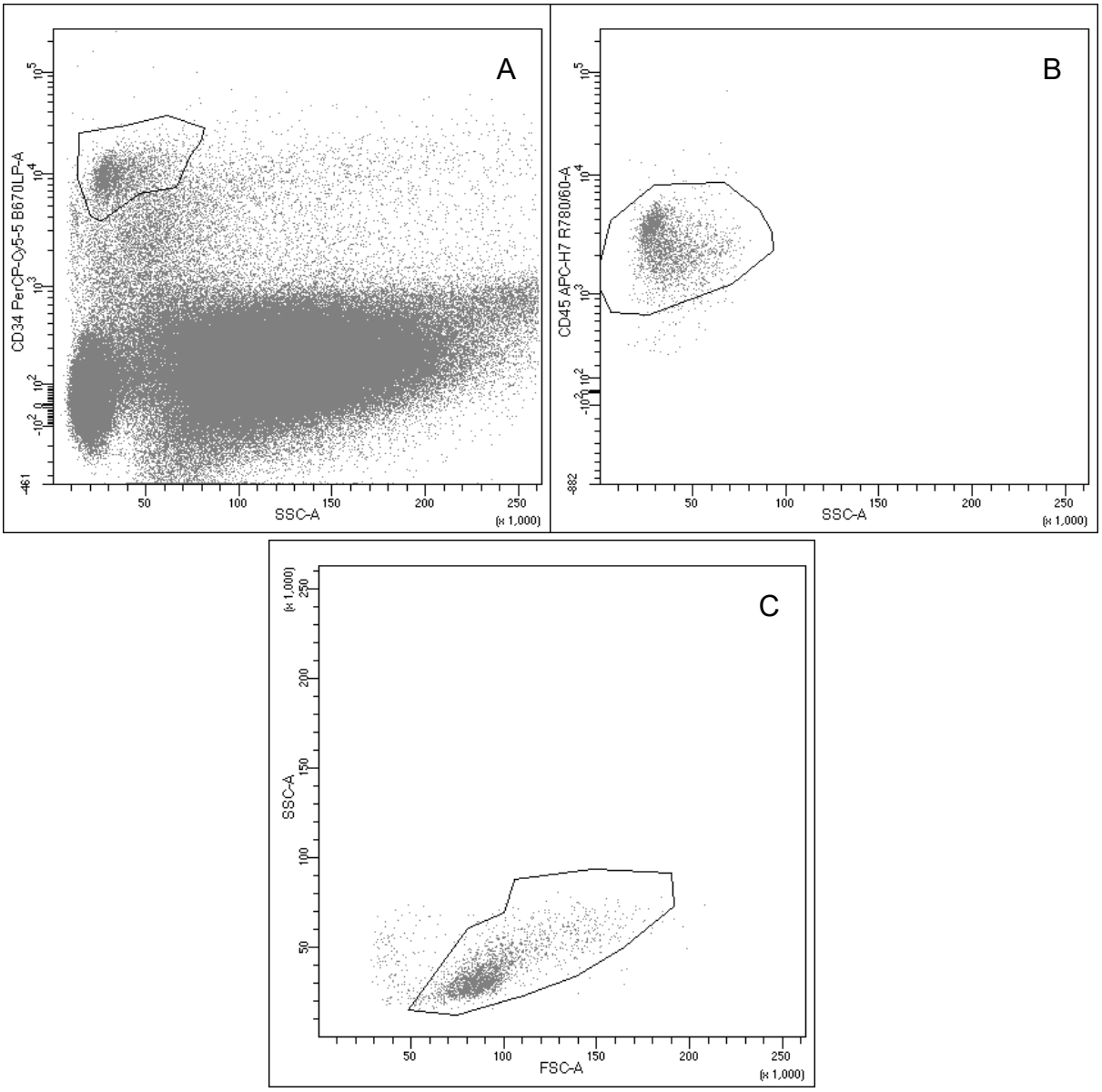


Figure S2. Flow cytometry gating strategy used to discriminate CD34⁺ myeloid progenitors. Leucocytes were identified on the basis of CD45 and side scatter (SSC) expression ***as shown in previous flow schema***. Within the leucocyte population, CD34⁺ myeloid progenitors were defined on the basis of CD34 expression and side scatter characteristics (A). Confirmation was performed using CD45 expression (B) and forward scatter and side scatter characteristics (C). CD34⁺ myeloid progenitors were then enumerated as a percentage of CD45⁺ leucocytes.

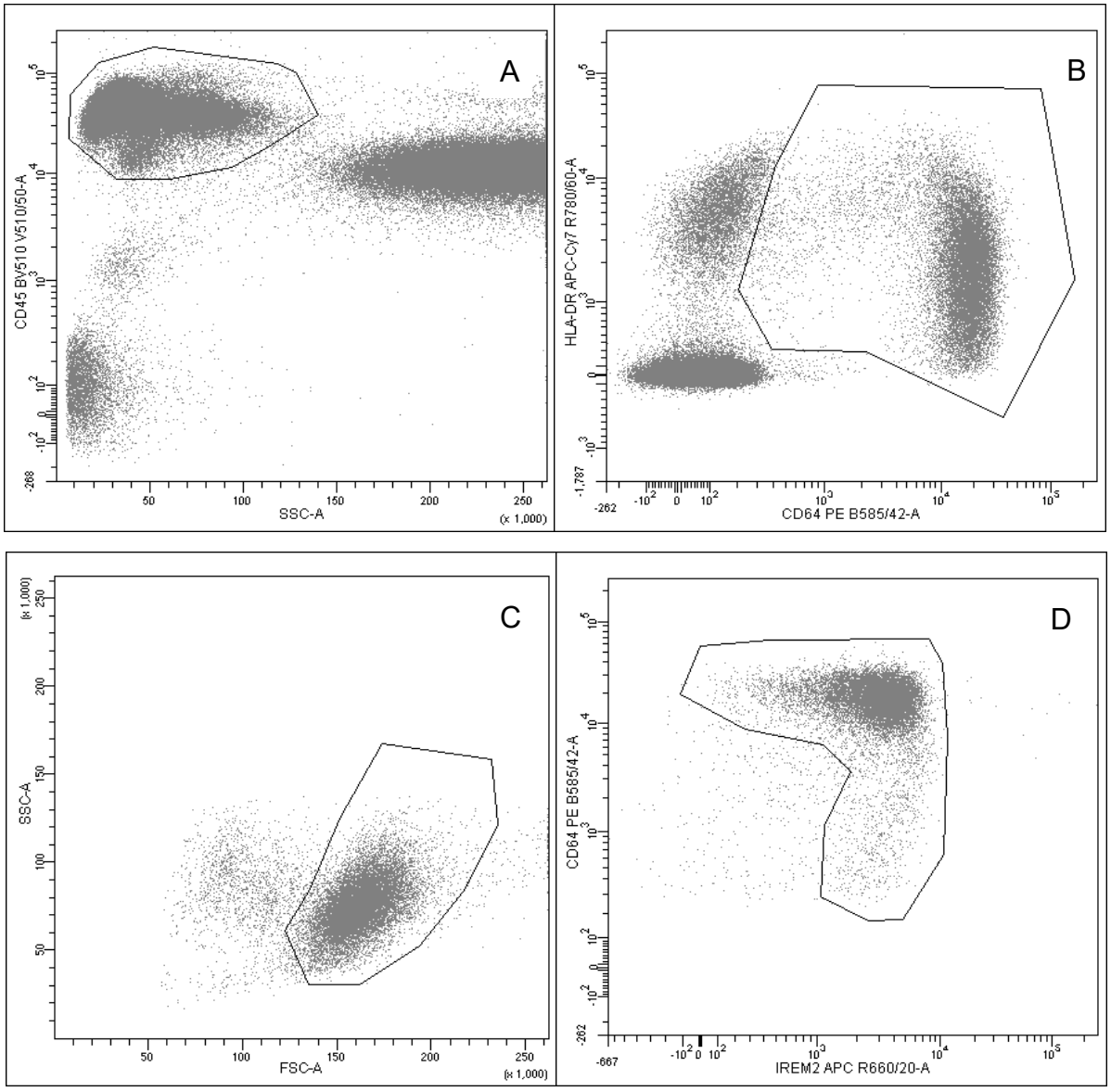


Figure S3. For the assessment of CD14 and CD16 expression, monocytes were positively identified on the basis of the following gating strategy: Mononuclear cells were identified on the basis of CD45 and side scatter (SSc) expression (A). A CD64 and HLA-DR inclusive gating strategy (B) included all monocytes before a combination of forward and side scatter expression (C) and IREM and CD64 expression (D) was employed to exclude all non-monocytic cells.

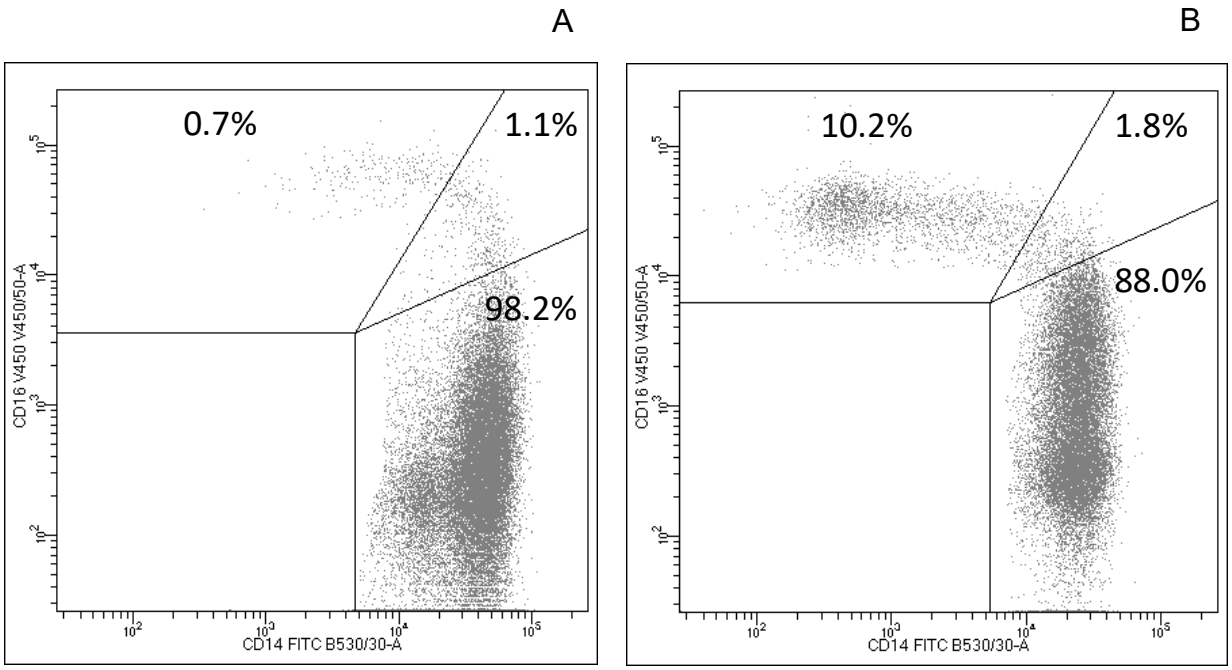


Figure S4. Examples of peripheral blood monocyte subset analysis. (A) Case of CMML with 98.2% M1 monocytes. (B) Reactive monocytosis with 88% M1 monocytes.

List of Abbreviations

aCML	Atypical chronic myeloid leukaemia
AML	Acute Myeloid Leukaemia
AVA	Amplicon Variant Analyzer
BM	Bone Marrow
CC	Conventional Cytogenetics
CCUS	Clonal Cytopenia of Undetermined Significance
CD	Cluster of Differentiation
CGH-a	Comparative Genomic Hybridization Arrays
CHIP	Clonal Haematopoiesis of Indeterminate Potential
CI	Confidence Intervals
CML	Chronic Myeloid Leukaemia
CMML	Chronic Myelomonocytic Leukaemia
CNL	Chronic Neutrophilic Leukaemia
CNV	Copy Number Variation
COSMIC	Catalogue of Somatic Mutations in Cancer
CS1	Common Sequence 1
CS2	Common Sequence 2
DNA	Deoxyribonucleic Acid
ddNTPs	Dideoxynucleotides
dNTPs	Deoxynucleotides
ET	Essential Thrombocythaemia
FAB	French American British
FDR	False Discovery Rate
MFC	Multiparameter Flow Cytometry
Hb	Haemoglobin
HMDS	Haematological Malignancy Diagnostic Service
HMRN	Haematological Malignancy Research Network
HR	Hazard Ratio

HTS	High Throughput Sequencing
ICUS	Idiopathic Cytopenia of Undetermined Significance
IFC	Integrated Fluidics Circuits
IPSS	International Prognostic Scoring System
LOH	Loss of Heterozygosity
MDS	Myelodysplastic Syndrome
MPN	Myeloproliferative Neoplasm
ND	Non-diagnostic
ND^{mut}	Non-diagnostic mutated
ND^{unmut}	Non-diagnostic unmutated
NS	Not Significant
OR	Odds Ratio
OS	Overall survival
PB	Peripheral Blood
PCR	Polymerase Chain Reaction
Plt	Platelet
PMF	Primary Myelofibrosis
PV	Polycythaemia Vera
QC	Quality Control
RA	Refractory anaemia
RARS	Refractory anaemia with ring sideroblasts
RARS-T	Refractory anaemia with ring sideroblasts & thrombocytosis
RCMD	Refractory cytopenia with multilineage dysplasia
RAEB	Refractory anaemia with excess blasts
RAEB-T	Refractory anaemia with excess blasts in transformation
SNP-a	Single Nucleotide Polymorphism – array
UKAS	United Kingdom Accreditation Scheme
UPD	Uniparental Disomy
VAF	Variant Allele Fraction
VCF	Variant Call Format
VEP	Variant Effect Predictor
WBC	White Blood Cell

WHO	World Health Organisation

REFERENCES

Abbas, S. *et al.* (2008) 'Exon 8 splice site mutations in the gene encoding the E3-ligase CBL are associated with core binding factor acute myeloid leukemias', *Haematologica*, 93(10), pp. 1595–1597. doi: 10.3324/haematol.13187.

Abbas, S. *et al.* (2010) 'Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value', *Blood*, 116(12), pp. 2122–2126. doi: 10.1182/blood-2009-11-250878.

Abdel-Wahab, O., Pardanani, A., Patel, J., *et al.* (2011) 'Concomitant analysis of EZH2 and ASXL1 mutations in myelofibrosis, chronic myelomonocytic leukemia and blast-phase myeloproliferative neoplasms', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 25(7), pp. 1200–1202. doi: 10.1038/leu.2011.58.

Abdel-Wahab, O., Pardanani, A., Rampal, R., *et al.* (2011) 'DNMT3A mutational analysis in primary myelofibrosis, chronic myelomonocytic leukemia and advanced phases of myeloproliferative neoplasms', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 25(7), pp. 1219–1220. doi: 10.1038/leu.2011.82.

Abelson, S. *et al.* (2018) 'Prediction of acute myeloid leukaemia risk in healthy individuals', *Nature*, 559(7714), pp. 400–404. doi: 10.1038/s41586-018-0317-6.

Abu-Duhier, F. M. *et al.* (2000) 'FLT3 internal tandem duplication mutations in adult acute myeloid leukaemia define a high-risk group', *British journal of haematology*, 111(1), pp. 190–195.

Abu-Duhier, F. M. *et al.* (2002) 'FLT3 internal tandem duplication mutations are rare in agnogenic myeloid metaplasia', *Blood*, 100(1), p. 364.

Abu-Duhier, F. M. *et al.* (2003) 'Mutational analysis of class III receptor tyrosine kinases (C-KIT, C-FMS, FLT3) in idiopathic myelofibrosis', *British journal of haematology*, 120(3), pp. 464–470.

Al-Kali, A. *et al.* (2013) 'Prognostic impact of RAS mutations in patients with myelodysplastic syndrome', *American Journal of Hematology*, 88(5), pp. 365–369. doi: 10.1002/ajh.23410.

Andersen, M. T. *et al.* (2008) 'NPM1 mutations in therapy-related acute myeloid leukemia with uncharacteristic features', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 22(5), pp. 951–955. doi: 10.1038/leu.2008.17.

Andrulis, M *et al.* (2010) 'Detection of isocitrate dehydrogenase 1 mutation R132H in myelodysplastic syndrome by mutation-specific antibody and direct sequencing', *Leukemia Research*, 34(8), pp. 1091–1093. doi: 10.1016/j.leukres.2010.02.014.

Andrulis, Mindaugas *et al.* (2010) 'IDH1 R132H mutation is a rare event in myeloproliferative neoplasms as determined by a mutation specific antibody', *Haematologica*, 95(10), pp. 1797–1798. doi: 10.3324/haematol.2010.024430.

Aranaz, P. *et al.* (2012) 'CBL mutations in myeloproliferative neoplasms are also found in the gene's proline-rich domain and in patients with the V617FJAK2', *Haematologica*, 97(8), pp. 1234–1241. doi: 10.3324/haematol.2011.052605.

Arber, D. A. *et al.* (2016) 'The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia', *Blood*, 127(20), pp. 2391–2405. doi: 10.1182/blood-2016-03-643544.

Armstrong, R. N. *et al.* (2018) 'Splicing factor mutations in the myelodysplastic syndromes: target genes and therapeutic approaches', *Advances in Biological Regulation*. (Symposium Issue 2017), 67, pp. 13–29. doi: 10.1016/j.jbior.2017.09.008.

Atallah, E. *et al.* (2008) 'Prognostic interaction between thrombocytosis and JAK2 V617F mutation in the WHO subcategories of myelodysplastic/myeloproliferative disease-unclassifiable and refractory anemia with ringed sideroblasts and marked thrombocytosis', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 22(6), pp. 1295–1298. doi: 10.1038/sj.leu.2405054.

Au, W. Y. *et al.* (2004) 'Low frequency of FLT3 gene internal tandem duplication and activating loop mutation in therapy-related acute myelocyticleukemia and myelodysplastic syndrome', *Cancer genetics and cytogenetics*, 149(2), pp. 169–172. doi: 10.1016/j.cancergencyto.2003.07.007.

Bacher, U. *et al.* (2006) 'Implications of NRAS mutations in AML: a study of 2502 patients', *Blood*, 107(10), pp. 3847–3853. doi: 10.1182/blood-2005-08-3522.

Bacher, U. *et al.* (2007a) 'A comparative study of molecular mutations in 381 patients with myelodysplastic syndrome and in 4130 patients with acute myeloid leukemia', *Haematologica*, 92(6), pp. 744–752.

Bacher, U. *et al.* (2007b) 'A comparative study of molecular mutations in 381 patients with myelodysplastic syndrome and in 4130 patients with acute myeloid leukemia', *Haematologica*, 92(6), pp. 744–752.

Bacher, U. *et al.* (2008) 'Prognostic relevance of FLT3-TKD mutations in AML: the combination matters--an analysis of 3082 patients', *Blood*, 111(5), pp. 2527–2537. doi: 10.1182/blood-2007-05-091215.

Bacher, U. *et al.* (2010) 'Mutations of the TET2 and CBL genes: novel molecular markers in myeloid malignancies', *Annals of Hematology*, 89(7), pp. 643–652. doi: 10.1007/s00277-010-0920-6.

Bain, B. (1996) 'The bone marrow aspirate of healthy subjects', *British Journal of Haematology*, 94(1), pp. 206–209.

Bains, A. *et al.* (2011) 'FLT3 and NPM1 mutations in myelodysplastic syndromes: Frequency and potential value for predicting progression to acute myeloid leukemia', *American Journal of Clinical Pathology*, 135(1), pp. 62–69. doi: 10.1309/AJCPEI9XU8PYBCIO.

Balduini, C. *et al.* (1998) 'Multilineage dysplasia without increased blasts identifies a poor prognosis subset of myelodysplastic syndromes', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 12(10), pp. 1655–1656.

Barjesteh van Waalwijk van Doorn-Khosrovani, S. *et al.* (2003) 'Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML', *The hematology journal: the official journal of the European Haematology Association / EHA*, 4(1), pp. 31–40. doi: 10.1038/sj.thj.6200216.

Barjesteh van Waalwijk van Doorn-Khosrovani, S. *et al.* (2005) 'Somatic heterozygous mutations in ETV6 (TEL) and frequent absence of ETV6 protein in acute myeloid leukemia', *Oncogene*, 24(25), pp. 4129–4137. doi: 10.1038/sj.onc.1208588.

Barosi, G. *et al.* (2007) 'JAK2 V617F mutational status predicts progression to large splenomegaly and leukemic transformation in primary myelofibrosis', *Blood*, 110(12), pp. 4030–4036. doi: 10.1182/blood-2007-07-099184.

Baumann, M. *et al.* (1986) 'Myeloid cell surface phenotype in myelodysplasia: evidence for abnormal persistence of an early myeloid differentiation antigen', *American Journal of Hematology*, 22(3), pp. 251–257.

Baxter, E. *et al.* (2005) 'Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders', *Lancet*, 365(9464), pp. 1054–1061. doi: 10.1016/S0140-6736(05)71142-9.

Beer, P. A. *et al.* (2008) 'MPL mutations in myeloproliferative disorders: analysis of the PT-1 cohort', *Blood*, 112(1), pp. 141–149. doi: 10.1182/blood-2008-01-131664.

Bejar, R. *et al.* (2011) 'Clinical effect of point mutations in myelodysplastic syndromes', *The New England Journal of Medicine*, 364(26), pp. 2496–2506. doi: 10.1056/NEJMoa1013343.

Bejar, R. *et al.* (2012) 'Validation of a Prognostic Model and the Impact of Mutations in Patients With Lower-Risk Myelodysplastic Syndromes', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. doi: 10.1200/JCO.2011.40.7379.

Bejar, R. (2015) 'Myelodysplastic Syndromes Diagnosis: What Is the Role of Molecular Testing?', *Current Hematologic Malignancy Reports*, 10(3), pp. 282–291. doi: 10.1007/s11899-015-0270-5.

Bennett, J. *et al.* (1976) 'Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group', *British Journal of Haematology*, 33(4), pp. 451–458.

Bennett, J. *et al.* (1982) 'Proposals for the classification of the myelodysplastic syndromes', *British Journal of Haematology*, 51(2), pp. 189–199.

Bennett, J. *et al.* (1994) 'The chronic myeloid leukaemias: guidelines for distinguishing chronic granulocytic, atypical chronic myeloid, and chronic myelomonocytic leukaemia. Proposals by the French-American-British Cooperative Leukaemia Group', *British Journal of Haematology*, 87(4), pp. 746–754.

Bernstein, S. *et al.* (1996) 'Acute myeloid leukemia-type chemotherapy for newly diagnosed patients without antecedent cytopenias having myelodysplastic syndrome as defined by French-American-British criteria: a Cancer and Leukemia Group B Study', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 14(9), pp. 2486–2494.

Boissel, N. *et al.* (2010) 'Prognostic impact of isocitrate dehydrogenase enzyme isoforms 1 and 2 mutations in acute myeloid leukemia: a study by the Acute Leukemia French Association group', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 28(23), pp. 3717–3723. doi: 10.1200/JCO.2010.28.2285.

Boultwood, J. *et al.* (2010) 'Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 24(5), pp. 1062–1065. doi: 10.1038/leu.2010.20.

Bowen, D. T. *et al.* (2005) 'RAS mutation in acute myeloid leukemia is associated with distinct cytogenetic subgroups but does not influence outcome in patients younger than 60 years', *Blood*, 106(6), pp. 2113–2119. doi: 10.1182/blood-2005-03-0867.

Boyd, E. M. *et al.* (2010) 'Clinical utility of routine MPL exon 10 analysis in the diagnosis of essential thrombocythaemia and primary myelofibrosis', *British journal of haematology*, 149(2), pp. 250–257. doi: 10.1111/j.1365-2141.2010.08083.x.

Braun, T. *et al.* (2011) 'Molecular predictors of response to decitabine in advanced chronic myelomonocytic leukemia: a phase 2 trial', *Blood*, 118(14), pp. 3824–3831. doi: 10.1182/blood-2011-05-352039.

Breccia, M. *et al.* (2006) 'Identification of risk factors in atypical chronic myeloid leukemia', *Haematologica*, 91(11), pp. 1566–1568.

Brecqueville, M. *et al.* (2011) 'Rare mutations in DNMT3A in myeloproliferative neoplasms and myelodysplastic syndromes', *Blood cancer journal*, 1(5), p. e18. doi: 10.1038/bcj.2011.15.

Busque, L. *et al.* (2012) 'Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis', *Nature genetics*, 44(11), pp. 1179–1181. doi: 10.1038/ng.2413.

Campbell, P. J. *et al.* (2006) 'V617F mutation in JAK2 is associated with poorer survival in idiopathic myelofibrosis', *Blood*, 107(5), pp. 2098–2100. doi: 10.1182/blood-2005-08-3395.

Carbuccia, N. *et al.* (2009) 'Mutations of ASXL1 gene in myeloproliferative neoplasms', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 23(11), pp. 2183–2186. doi: 10.1038/leu.2009.141.

Carbuccia, N. *et al.* (2010) 'Mutual exclusion of ASXL1 and NPM1 mutations in a series of acute myeloid leukemias', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 24(2), pp. 469–473. doi: 10.1038/leu.2009.218.

Cargo, C. A. *et al.* (2015) 'Targeted sequencing identifies patients with preclinical MDS at high risk of disease progression', *Blood*, 126(21), pp. 2362–2365. doi: 10.1182/blood-2015-08-663237.

Cazzola, M., Della Porta, M. G. and Malcovati, L. (2013) 'The genetic basis of myelodysplasia and its clinical relevance', *Blood*. doi: 10.1182/blood-2013-09-381665.

Ceesay, M. M. *et al.* (2006) 'The JAK2 V617F mutation is rare in RARS but common in RARS-T', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 20(11), pp. 2060–2061. doi: 10.1038/sj.leu.2404373.

Chaligné, R. *et al.* (2008) 'New mutations of MPL in primitive myelofibrosis: only the MPL W515 mutations promote a G1/S-phase transition', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 22(8), pp. 1557–1566. doi: 10.1038/leu.2008.137.

Chan, G. *et al.* (1997) 'Myelodysplastic syndrome in children: differentiation from acute myeloid leukemia with a low blast count', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 11(2), pp. 206–211.

Chang, C. and Cleveland, R. (2000) 'Decreased CD10-positive mature granulocytes in bone marrow from patients with myelodysplastic syndrome', *Archives of Pathology & Laboratory Medicine*, 124(8), pp. 1152–1156. doi: 10.1043/0003-9985(2000)124<1152:DCPMGI>2.0.CO;2.

Chen, C.-Y. *et al.* (2007) 'RUNX1 gene mutation in primary myelodysplastic syndrome--the mutation can be detected early at diagnosis or acquired during disease progression and is associated with poor outcome', *British journal of haematology*, 139(3), pp. 405–414. doi: 10.1111/j.1365-2141.2007.06811.x.

Chen, S. *et al.* (2007) 'Analysis of JAK2(V617F) mutation in Chinese patients with myeloproliferative disorders', *American journal of hematology*, 82(6), pp. 458–459. doi: 10.1002/ajh.20872.

Chou, W.-C., Hou, H.-A., *et al.* (2010) 'Distinct clinical and biologic characteristics in adult acute myeloid leukemia bearing the isocitrate dehydrogenase 1 mutation', *Blood*, 115(14), pp. 2749–2754. doi: 10.1182/blood-2009-11-253070.

Chou, W.-C., Huang, H.-H., *et al.* (2010) 'Distinct clinical and biological features of de novo acute myeloid leukemia with additional sex comb-like 1 (ASXL1) mutations', *Blood*, 116(20), pp. 4086–4094. doi: 10.1182/blood-2010-05-283291.

Chou, Wen-Chien *et al.* (2011) 'TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics', *Blood*, 118(14), pp. 3803–3810. doi: 10.1182/blood-2011-02-339747.

Chou, W-C *et al.* (2011) 'The prognostic impact and stability of Isocitrate dehydrogenase 2 mutation in adult patients with acute myeloid leukemia', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 25(2), pp. 246–253. doi: 10.1038/leu.2010.267.

Christiansen, D., Andersen, M. and Pedersen-Bjergaard, J. (2004) 'Mutations of AML1 are common in therapy-related myelodysplasia following therapy with alkylating agents and are significantly associated with deletion or loss of chromosome arm 7q and with subsequent leukemic transformation', *Blood*, 104(5), pp. 1474–1481. doi: 10.1182/blood-2004-02-0754.

Christiansen, D. H. *et al.* (2007) 'Mutations of the PTPN11 gene in therapy-related MDS and AML with rare balanced chromosome translocations', *Genes, chromosomes & cancer*, 46(6), pp. 517–521. doi: 10.1002/gcc.20426.

Chu, S. *et al.* (2011) 'Flow cytometric scoring system as a diagnostic and prognostic tool in myelodysplastic syndromes', *Leukemia Research*, 35(7), pp. 868–873. doi: 10.1016/j.leukres.2011.02.016.

Cools, J. *et al.* (2003) 'A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome', *The New England Journal of Medicine*, 348(13), pp. 1201–1214. doi: 10.1056/NEJMoa025217.

Cui, R. *et al.* (2012) 'Clinical importance of SF3B1 mutations in Chinese with myelodysplastic syndromes with ring sideroblasts', *Leukemia research*, 36(11), pp. 1428–1433. doi: 10.1016/j.leukres.2012.08.015.

Cutler, J. *et al.* (2011) 'Phenotypic abnormalities strongly reflect genotype in patients with unexplained cytopenias', *Cytometry. Part B, Clinical Cytometry*, 80(3), pp. 150–157. doi: 10.1002/cyto.b.20582.

Dameshek, W. (1951) 'Some speculations on the myeloproliferative syndromes', *Blood*, 6(4), pp. 372–375.

Damm, F., Kosmider, O., Gelsi-Boyer, V., Renneville, A., Carbuccia, N., Hidalgo-Curtis, C., Della Valle, V., *et al.* (2012) 'Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes', *Blood*, 119(14), pp. 3211–3218. doi: 10.1182/blood-2011-12-400994.

Damm, F., Kosmider, O., Gelsi-Boyer, V., Renneville, A., Carbuccia, N., Hidalgo-Curtis, C., Della-Valle, V., *et al.* (2012) 'Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes', *Blood*. doi: 10.1182/blood-2011-12-400994.

Damm, F. *et al.* (2013) 'BCOR and BCORL1 mutations in myelodysplastic syndromes and related disorders', *Blood*. doi: 10.1182/blood-2012-11-469619.

David, M. *et al.* (2007) 'Durable responses to imatinib in patients with PDGFRB fusion gene-positive and BCR-ABL-negative chronic myeloproliferative disorders', *Blood*, 109(1), pp. 61–64. doi: 10.1182/blood-2006-05-024828.

Del Cañizo, M. *et al.* (2003) 'Immunophenotypic analysis of myelodysplastic syndromes', *Haematologica*, 88(4), pp. 402–407.

Delhommeau, F. *et al.* (2009) 'Mutation in TET2 in myeloid cancers', *The New England journal of medicine*, 360(22), pp. 2289–2301. doi: 10.1056/NEJMoa0810069.

Della Porta, M. *et al.* (2006) 'Flow cytometry evaluation of erythroid dysplasia in patients with myelodysplastic syndrome', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 20(4), pp. 549–555. doi: 10.1038/sj.leu.2404142.

Della Porta, M. G. *et al.* (2012) 'Multicentre validation of a reproducible flow cytometric score for the diagnosis of low-risk myelodysplastic syndromes: results of a European LeukemiaNET study', *Haematologica*. doi: 10.3324/haematol.2011.048421.

Dicker, F. *et al.* (2010) 'Mutation analysis for RUNX1, MLL-PTD, FLT3-ITD, NPM1 and NRAS in 269 patients with MDS or secondary AML', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 24(8), pp. 1528–1532. doi: 10.1038/leu.2010.124.

Ding, Y. *et al.* (2009) 'AML1/RUNX1 point mutation possibly promotes leukemic transformation in myeloproliferative neoplasms', *Blood*, 114(25), pp. 5201–5205. doi: 10.1182/blood-2009-06-223982.

Döhner, K. *et al.* (2005) 'Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics:

interaction with other gene mutations', *Blood*, 106(12), pp. 3740–3746. doi: 10.1182/blood-2005-05-2164.

Dolatshad, H. *et al.* (2016) 'Cryptic splicing events in the iron transporter ABCB7 and other key target genes in SF3B1-mutant myelodysplastic syndromes', *Leukemia*, 30(12), pp. 2322–2331. doi: 10.1038/leu.2016.149.

Drets, M. and Shaw, M. (1971) 'Specific banding patterns of human chromosomes', *Proceedings of the National Academy of Sciences of the United States of America*, 68(9), pp. 2073–2077.

Dunbar, A. *et al.* (2008) '250K single nucleotide polymorphism array karyotyping identifies acquired uniparental disomy and homozygous mutations, including novel missense substitutions of c-Cbl, in myeloid malignancies', *Cancer Research*, 68(24), pp. 10349–10357. doi: 10.1158/0008-5472.CAN-08-2754.

Dunphy, C., Orton, S. and Mantell, J. (2004) 'Relative contributions of enzyme cytochemistry and flow cytometric immunophenotyping to the evaluation of acute myeloid leukemias with a monocytic component and of flow cytometric immunophenotyping to the evaluation of absolute monocytoses', *American Journal of Clinical Pathology*, 122(6), pp. 865–874. doi: 10.1309/BH58-8HVG-6UHN-2RF2.

Elena, C. *et al.* (2016) 'Integrating clinical features and genetic lesions in the risk assessment of patients with chronic myelomonocytic leukemia', *Blood*, 128(10), pp. 1408–1417. doi: 10.1182/blood-2016-05-714030.

Ernst, T *et al.* (2010) 'Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders', *Nature Genetics*, 42(8), pp. 722–726. doi: 10.1038/ng.621.

Ernst, Thomas *et al.* (2010) 'Transcription factor mutations in myelodysplastic/myeloproliferative neoplasms', *Haematologica*, 95(9), pp. 1473–1480. doi: 10.3324/haematol.2010.021808.

Estey, E. *et al.* (1997) 'Effect of diagnosis (refractory anemia with excess blasts, refractory anemia with excess blasts in transformation, or acute myeloid leukemia [AML]) on outcome of AML-type chemotherapy', *Blood*, 90(8), pp. 2969–2977.

Ewalt, M. *et al.* (2011) 'DNMT3a mutations in high-risk myelodysplastic syndrome parallel those found in acute myeloid leukemia', *Blood cancer journal*, 1(3), p. e9. doi: 10.1038/bcj.2011.7.

Falini, B. *et al.* (2005) 'Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype', *The New England journal of medicine*, 352(3), pp. 254–266. doi: 10.1056/NEJMoa041974.

Fenaux, P. *et al.* (1988) 'Prognostic factors in adult chronic myelomonocytic leukemia: an analysis of 107 cases', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 6(9), pp. 1417–1424.

Figueroa, M. E. *et al.* (2010) 'Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation', *Cancer cell*, 18(6), pp. 553–567. doi: 10.1016/j.ccr.2010.11.015.

Fitzgibbon, J. *et al.* (2005) 'Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias', *Cancer Research*, 65(20), pp. 9152–9154. doi: 10.1158/0008-5472.CAN-05-2017.

Flach, J. *et al.* (2011) 'An accumulation of cytogenetic and molecular genetic events characterizes the progression from MDS to secondary AML: an analysis of 38 paired samples analyzed by cytogenetics, molecular mutation analysis and SNP microarray profiling', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 25(4), pp. 713–718. doi: 10.1038/leu.2010.304.

Fontalba, A. *et al.* (2006) 'Identification of c-Kit gene mutations in patients with polycythemia vera', *Leukemia research*, 30(10), pp. 1325–1326. doi: 10.1016/j.leukres.2005.12.020.

Forbes, S. A. *et al.* (2011) 'COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer', *Nucleic acids research*, 39(Database issue), pp. D945-950. doi: 10.1093/nar/gkq929.

Fox, S. *et al.* (1990) 'Megakaryocytes in myelodysplasia: an immunohistochemical study on bone marrow trephines', *Histopathology*, 17(1), pp. 69–74.

Fritsche-Polanz, R. *et al.* (2001) 'Mutation analysis of C-KIT in patients with myelodysplastic syndromes without mastocytosis and cases of systemic mastocytosis', *British journal of haematology*, 113(2), pp. 357–364.

Fröhling, S. *et al.* (2004) 'CEBPA mutations in younger adults with acute myeloid leukemia and normal cytogenetics: prognostic relevance and analysis of cooperating mutations', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 22(4), pp. 624–633. doi: 10.1200/JCO.2004.06.060.

Fuchigami, K. *et al.* (2000) 'Absolute number of circulating CD34+ cells is abnormally low in refractory anemias and extremely high in RAEB and RAEB-t; novel pathologic features of myelodysplastic syndromes identified by highly sensitive flow cytometry', *Leukemia Research*, 24(2), pp. 163–174.

Fuchs, O. *et al.* (2009) 'Nature of frequent deletions in CEBPA', *Blood cells, molecules & diseases*, 43(3), pp. 260–263. doi: 10.1016/j.bcmd.2009.07.001.

Gaidzik, V. I. *et al.* (2009) 'Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group', *Blood*, 113(19), pp. 4505–4511. doi: 10.1182/blood-2008-10-183392.

Gaidzik, V. I. *et al.* (2011) 'RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group', *Journal of*

clinical oncology: official journal of the American Society of Clinical Oncology, 29(10), pp. 1364–1372. doi: 10.1200/JCO.2010.30.7926.

Gaidzik, V. I. *et al.* (2012) 'TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 30(12), pp. 1350–1357. doi: 10.1200/JCO.2011.39.2886.

Galera, P., Dulau-Florea, A. and Calvo, K. R. (2019) 'Inherited thrombocytopenia and platelet disorders with germline predisposition to myeloid neoplasia', *International Journal of Laboratory Hematology*, 41 Suppl 1, pp. 131–141. doi: 10.1111/ijlh.12999.

Galton, D. (1992) 'Haematological differences between chronic granulocytic leukaemia, atypical chronic myeloid leukaemia, and chronic myelomonocytic leukaemia', *Leukemia & Lymphoma*, 7(5–6), pp. 343–350. doi: 10.3109/10428199209049789.

Gandini, D. *et al.* (2004) 'No mutations in the GATA-1 gene detected in patients with acquired essential thrombocythemia', *Haematologica*, 89(5), pp. 613–615.

Gelsi-Boyer, V. *et al.* (2008) 'Genome profiling of chronic myelomonocytic leukemia: frequent alterations of RAS and RUNX1 genes', *BMC cancer*, 8, p. 299. doi: 10.1186/1471-2407-8-299.

Gelsi-Boyer, V. *et al.* (2009) 'Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia', *British journal of haematology*, 145(6), pp. 788–800. doi: 10.1111/j.1365-2141.2009.07697.x.

Gelsi-Boyer, V. *et al.* (2012) 'Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases', *Journal of Hematology & Oncology*, 5(1), p. 12. doi: 10.1186/1756-8722-5-12.

Genovese, G. *et al.* (2014) 'Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence', *The New England Journal of Medicine*, 371(26), pp. 2477–2487. doi: 10.1056/NEJMoa1409405.

Germing, U. *et al.* (1998) 'Problems in the classification of CMML--dysplastic versus proliferative type', *Leukemia Research*, 22(10), pp. 871–878.

Germing, U. *et al.* (2000) 'Validation of the WHO proposals for a new classification of primary myelodysplastic syndromes: a retrospective analysis of 1600 patients', *Leukemia Research*, 24(12), pp. 983–992.

Germing, U. *et al.* (2007) 'Chronic myelomonocytic leukemia in the light of the WHO proposals', *Haematologica*, 92(7), pp. 974–977.

Glembotsky, A. C. *et al.* (2010) 'Screening for MPL mutations in essential thrombocythemia and primary myelofibrosis: normal Mpl expression and absence

of constitutive STAT3 and STAT5 activation in MPLW515L-positive platelets', *European journal of haematology*, 84(5), pp. 398–405. doi: 10.1111/j.1600-0609.2010.01421.x.

Goardon, N. *et al.* (2009) 'Reduced CD38 expression on CD34+ cells as a diagnostic test in myelodysplastic syndromes', *Haematologica*, 94(8), pp. 1160–1163. doi: 10.3324/haematol.2008.004085.

Goasguen, J. *et al.* (2009) 'Morphological evaluation of monocytes and their precursors', *Haematologica*, 94(7), pp. 994–997. doi: 10.3324/haematol.2008.005421.

Golub, T. *et al.* (1994) 'Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation', *Cell*, 77(2), pp. 307–316.

Gondek, L., Haddad, A., *et al.* (2007) 'Detection of cryptic chromosomal lesions including acquired segmental uniparental disomy in advanced and low-risk myelodysplastic syndromes', *Experimental Hematology*, 35(11), pp. 1728–1738. doi: 10.1016/j.exphem.2007.08.009.

Gondek, L., Tiu, R., *et al.* (2007) 'Single nucleotide polymorphism arrays complement metaphase cytogenetics in detection of new chromosomal lesions in MDS', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 21(9), pp. 2058–2061. doi: 10.1038/sj.leu.2404745.

Gondek, L., Dunbar, A., *et al.* (2007) 'SNP array karyotyping allows for the detection of uniparental disomy and cryptic chromosomal abnormalities in MDS/MPD-U and MPD', *PloS One*, 2(11), p. e1225. doi: 10.1371/journal.pone.0001225.

Gondek, L. *et al.* (2008) 'Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML', *Blood*, 111(3), pp. 1534–1542. doi: 10.1182/blood-2007-05-092304.

Grand, F. H. *et al.* (2009) 'Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms', *Blood*, 113(24), pp. 6182–6192. doi: 10.1182/blood-2008-12-194548.

Graubert, T. A. *et al.* (2012) 'Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes', *Nature genetics*, 44(1), pp. 53–57. doi: 10.1038/ng.1031.

Green, C. L., Koo, K. K., *et al.* (2010) 'Prognostic significance of CEBPA mutations in a large cohort of younger adult patients with acute myeloid leukemia: impact of double CEBPA mutations and the interaction with FLT3 and NPM1 mutations', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 28(16), pp. 2739–2747. doi: 10.1200/JCO.2009.26.2501.

Green, C. L., Evans, C. M., *et al.* (2010) 'The prognostic significance of IDH1 mutations in younger adult patients with acute myeloid leukemia is dependent on FLT3/ITD status', *Blood*, 116(15), pp. 2779–2782. doi: 10.1182/blood-2010-02-270926.

Green, C. L. *et al.* (2011) 'The prognostic significance of IDH2 mutations in AML depends on the location of the mutation', *Blood*, 118(2), pp. 409–412. doi: 10.1182/blood-2010-12-322479.

Greenberg, P. *et al.* (1997) 'International scoring system for evaluating prognosis in myelodysplastic syndromes', *Blood*, 89(6), pp. 2079–2088.

Greenberg, P. *et al.* (2000) 'Problematic WHO reclassification of myelodysplastic syndromes. Members of the International MDS Study Group', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 18(19), pp. 3447–3452.

Greif, P. A. *et al.* (2012) 'GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia', *Blood*. doi: 10.1182/blood-2012-01-403220.

Grossmann, V, Kohlmann, A., Eder, C., Haferlach, C., Kern, W., Cross, N., *et al.* (2011) 'Molecular profiling of chronic myelomonocytic leukemia reveals diverse mutations in >80% of patients with TET2 and EZH2 being of high prognostic relevance', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 25(5), pp. 877–879. doi: 10.1038/leu.2011.10.

Grossmann, V, Kohlmann, A., Eder, C., Haferlach, C., Kern, W., Cross, N. C. P., *et al.* (2011) 'Molecular profiling of chronic myelomonocytic leukemia reveals diverse mutations in >80% of patients with TET2 and EZH2 being of high prognostic relevance', *Leukemia*, 25(5), pp. 877–879.

Grossmann, Vera *et al.* (2011) 'Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype', *Blood*, 118(23), pp. 6153–6163. doi: 10.1182/blood-2011-07-365320.

Guglielmelli, P. *et al.* (2011) 'EZH2 mutational status predicts poor survival in myelofibrosis', *Blood*, 118(19), pp. 5227–5234. doi: 10.1182/blood-2011-06-363424.

Gupta, R., Abdalla, S. and Bain, B. (1999) 'Thrombocytosis with sideroblastic erythropoiesis: a mixed myeloproliferative myelodysplastic syndrome', *Leukemia & Lymphoma*, 34(5–6), pp. 615–619. doi: 10.3109/10428199909058492.

Haferlach, C. *et al.* (2008) 'Mutations of the TP53 gene in acute myeloid leukemia are strongly associated with a complex aberrant karyotype', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 22(8), pp. 1539–1541. doi: 10.1038/leu.2008.143.

Haferlach, T. *et al.* (2013) 'Landscape of genetic lesions in 944 patients with myelodysplastic syndromes', *Leukemia*. doi: 10.1038/leu.2013.336.

Hansen, I., Meyer, K. and Hokland, P. (1998) 'Flow cytometric identification of myeloid disorders by asynchronous expression of the CD14 and CD66 antigens', *European Journal of Haematology*, 61(5), pp. 339–346.

Harada, H. *et al.* (2004) 'High incidence of somatic mutations in the AML1/RUNX1 gene in myelodysplastic syndrome and low blast percentage myeloid leukemia with myelodysplasia', *Blood*, 103(6), pp. 2316–2324. doi: 10.1182/blood-2003-09-3074.

Harris, N. *et al.* (1999) 'World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 17(12), pp. 3835–3849.

Harrison, C. N. *et al.* (2010) 'Guideline for investigation and management of adults and children presenting with a thrombocytosis', *British Journal of Haematology*, 149(3), pp. 352–375. doi: 10.1111/j.1365-2141.2010.08122.x.

Head, D. *et al.* (1989) 'Pathogenetic implications of internuclear bridging in myelodysplastic syndrome. An Eastern Cooperative Oncology Group/Southwest Oncology Group Cooperative Study', *Cancer*, 64(11), pp. 2199–2202.

Heinrichs, S. *et al.* (2009) 'Accurate detection of uniparental disomy and microdeletions by SNP array analysis in myelodysplastic syndromes with normal cytogenetics', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 23(9), pp. 1605–1613. doi: 10.1038/leu.2009.82.

Heinrichs, S., Li, C. and Look, A. T. (2010) 'SNP array analysis in hematologic malignancies: avoiding false discoveries', *Blood*, 115(21), pp. 4157–4161. doi: 10.1182/blood-2009-11-203182.

Hokland, P. *et al.* (1986) 'Analysis of leukocyte differentiation antigens in blood and bone marrow from preleukemia (refractory anemia) patients using monoclonal antibodies', *Blood*, 67(4), pp. 898–902.

Horiike, S. *et al.* (1997) 'Tandem duplications of the FLT3 receptor gene are associated with leukemic transformation of myelodysplasia', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 11(9), pp. 1442–1446.

Hosoya, N. *et al.* (1998) 'Mutation analysis of the WT1 gene in myelodysplastic syndromes', *Japanese journal of cancer research: Gann*, 89(8), pp. 821–824.

Hou, H.-A. *et al.* (2008) 'Characterization of acute myeloid leukemia with PTPN11 mutation: the mutation is closely associated with NPM1 mutation but inversely

related to FLT3/ITD', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 22(5), pp. 1075–1078. doi: 10.1038/sj.leu.2405005.

Hou, H.-A. *et al.* (2012) 'DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications', *Blood*, 119(2), pp. 559–568. doi: 10.1182/blood-2011-07-369934.

Huang, W. *et al.* (2009) 'Multiple distinct clones may co-exist in different lineages in myelodysplastic syndromes', *Leukemia Research*, 33(6), pp. 847–853. doi: 10.1016/j.leukres.2008.10.008.

Hugues, L. *et al.* (2005) 'Mutations of PTPN11 are rare in adult myeloid malignancies', *Haematologica*, 90(6), pp. 853–854.

Hussein, K. *et al.* (2009) 'MPLW515L mutation in acute megakaryoblastic leukaemia', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 23(5), pp. 852–855. doi: 10.1038/leu.2008.371.

Hussein, Kebede *et al.* (2010) 'Cytogenetic correlates of TET2 mutations in 199 patients with myeloproliferative neoplasms', *American journal of hematology*, 85(1), pp. 81–83. doi: 10.1002/ajh.21562.

Hussein, K *et al.* (2010) 'IDH mutation analysis is not suitable for the routine molecular diagnostic algorithm in myeloproliferative and myelodysplastic neoplasms', *Blood*, 116(23), pp. 5073–5074. doi: 10.1182/blood-2010-06-290791.

Illmer, T. *et al.* (2007) 'Tyrosine kinase mutations of JAK2 are rare events in AML but influence prognosis of patients with CBF-leukemias', *Haematologica*, 92(1), pp. 137–138.

Imai, Y. *et al.* (2000) 'Mutations of the AML1 gene in myelodysplastic syndrome and their functional implications in leukemogenesis', *Blood*, 96(9), pp. 3154–3160.

Itzykson, R. *et al.* (2013) 'Prognostic Score Including Gene Mutations in Chronic Myelomonocytic Leukemia', *Journal of Clinical Oncology*, 31(19), pp. 2428–2436. doi: 10.1200/JCO.2012.47.3314.

Itzykson, R. *et al.* (2018) 'Diagnosis and Treatment of Chronic Myelomonocytic Leukemias in Adults: Recommendations From the European Hematology Association and the European LeukemiaNet', *HemaSphere*, 2(6), p. e150. doi: 10.1097/HS9.0000000000000150.

Jacobs, K. B. *et al.* (2012) 'Detectable clonal mosaicism and its relationship to aging and cancer', *Nature Genetics*, 44(6), pp. 651–658. doi: 10.1038/ng.2270.

Jädersten, M. *et al.* (2011) 'TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 29(15), pp. 1971–1979. doi: 10.1200/JCO.2010.31.8576.

Jaiswal, S. *et al.* (2014) 'Age-related clonal hematopoiesis associated with adverse outcomes', *The New England Journal of Medicine*, 371(26), pp. 2488–2498. doi: 10.1056/NEJMoa1408617.

James, C. *et al.* (2005) 'A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera', *Nature*, 434(7037), pp. 1144–1148. doi: 10.1038/nature03546.

Jankowska, A. *et al.* (2011) 'Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2 and DNMT3A', *Blood*. doi: 10.1182/blood-2010-10-311019.

Jerez, A. *et al.* (2012) 'STAT3 mutations unify the pathogenesis of chronic lymphoproliferative disorders of NK cells and T cell large granular lymphocyte leukemia', *Blood*. doi: 10.1182/blood-2012-06-435297.

Johan, M. F. *et al.* (2004) 'Mutations in PTPN11 are uncommon in adult myelodysplastic syndromes and acute myeloid leukaemia', *British journal of haematology*, 124(6), pp. 843–844.

Jones, A. *et al.* (2005) 'Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders', *Blood*, 106(6), pp. 2162–2168. doi: 10.1182/blood-2005-03-1320.

Jonveaux, P. *et al.* (1991) 'Mutations in the p53 gene in myelodysplastic syndromes', *Oncogene*, 6(12), pp. 2243–2247.

Juneja, S. *et al.* (1983) 'Haematological features of primary myelodysplastic syndromes (PMDS) at initial presentation: a study of 118 cases', *Journal of Clinical Pathology*, 36(10), pp. 1129–1135.

Kaeferstein, A. *et al.* (2003) 'The emergence of a C/EBPalpha mutation in the clonal evolution of MDS towards secondary AML', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 17(2), pp. 343–349. doi: 10.1038/sj.leu.2402805.

Kaneko, H. *et al.* (1995) 'TP53 mutations emerge at early phase of myelodysplastic syndrome and are associated with complex chromosomal abnormalities', *Blood*, 85(8), pp. 2189–2193.

Karmon, Y., Manaster, J. and Chezar, J. (2002) 'Immunophenotypic characterization of myelopoiesis in early and late myelodysplastic syndromes: use of CD44 as an aid in early diagnosis', *Cytometry*, 50(4), pp. 225–230. doi: 10.1002/cyto.10110.

Kern, W. *et al.* (2010) 'Clinical utility of multiparameter flow cytometry in the diagnosis of 1013 patients with suspected myelodysplastic syndrome: correlation to cytomorphology, cytogenetics, and clinical data', *Cancer*, 116(19), pp. 4549–4563. doi: 10.1002/cncr.25353.

Kern, W. *et al.* (2011) 'Acute monoblastic/monocytic leukemia and chronic myelomonocytic leukemia share common immunophenotypic features but differ in the extent of aberrantly expressed antigens and amount of granulocytic cells', *Leukemia & Lymphoma*, 52(1), pp. 92–100. doi: 10.3109/10428194.2010.533801.

Kim, E. *et al.* (2015) 'SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition', *Cancer Cell*, 27(5), pp. 617–630. doi: 10.1016/j.ccell.2015.04.006.

Kim, Hee-Jin *et al.* (2013) 'KIT D816 mutation associates with adverse outcomes in core binding factor acute myeloid leukemia, especially in the subgroup with RUNX1/RUNX1T1 rearrangement', *Annals of Hematology*, 92(2), pp. 163–171. doi: 10.1007/s00277-012-1580-5.

King-Underwood, L., Renshaw, J. and Pritchard-Jones, K. (1996) 'Mutations in the Wilms' tumor gene WT1 in leukemias', *Blood*, 87(6), pp. 2171–2179.

Kita-Sasai, Y. *et al.* (2001a) 'International prognostic scoring system and TP53 mutations are independent prognostic indicators for patients with myelodysplastic syndrome', *British Journal of Haematology*, 115(2), pp. 309–312.

Kita-Sasai, Y. *et al.* (2001b) 'International prognostic scoring system and TP53 mutations are independent prognostic indicators for patients with myelodysplastic syndrome', *British journal of haematology*, 115(2), pp. 309–312.

Kiyoi, H. *et al.* (1999) 'Prognostic implication of FLT3 and N-RAS gene mutations in acute myeloid leukemia', *Blood*, 93(9), pp. 3074–3080.

Klampfl, T. *et al.* (2013) 'Somatic mutations of calreticulin in myeloproliferative neoplasms', *The New England Journal of Medicine*, 369(25), pp. 2379–2390. doi: 10.1056/NEJMoa1311347.

Ko, M. *et al.* (2010) 'Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2', *Nature*, 468(7325), pp. 839–843. doi: 10.1038/nature09586.

Kohlmann, A. *et al.* (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: Official Journal of the American Society of Clinical Oncology*, 28(24), pp. 3858–3865. doi: 10.1200/JCO.2009.27.1361.

Kon, A. *et al.* (2013) 'Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms', *Nature genetics*, 45(10), pp. 1232–1237. doi: 10.1038/ng.2731.

Kosmider, O. *et al.* (2009) 'TET2 mutation is an independent favorable prognostic factor in myelodysplastic syndromes (MDSs)', *Blood*, 114(15), pp. 3285–3291. doi: 10.1182/blood-2009-04-215814.

Kosmider, O. *et al.* (2010) 'Mutations of IDH1 and IDH2 genes in early and accelerated phases of myelodysplastic syndromes and MDS/myeloproliferative neoplasms', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 24(5), pp. 1094–1096. doi: 10.1038/leu.2010.52.

Kottaridis, P. D. *et al.* (2001) 'The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials', *Blood*, 98(6), pp. 1752–1759.

Kouides, P. and Bennett, J. (1996) 'Morphology and classification of the myelodysplastic syndromes and their pathologic variants', *Seminars in Hematology*, 33(2), pp. 95–110.

Kralovics, R *et al.* (2005) 'A gain-of-function mutation of JAK2 in myeloproliferative disorders', *The New England Journal of Medicine*, 352(17), pp. 1779–1790. doi: 10.1056/NEJMoa051113.

Kralovics, Robert *et al.* (2005) 'A gain-of-function mutation of JAK2 in myeloproliferative disorders', *The New England journal of medicine*, 352(17), pp. 1779–1790. doi: 10.1056/NEJMoa051113.

Kralovics, R., Guan, Y. and Prchal, J. (2002) 'Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera', *Experimental Hematology*, 30(3), pp. 229–236.

Kristensen, J. and Hokland, P. (1990) 'Monoclonal antibody ratios in malignant myeloid diseases: diagnostic and prognostic use in myelodysplastic syndromes', *British Journal of Haematology*, 74(3), pp. 270–276.

Kuriyama, K. *et al.* (1986) 'Diagnostic significance of detecting pseudo-Pelger-Huët anomalies and micro-megakaryocytes in myelodysplastic syndrome', *British Journal of Haematology*, 63(4), pp. 665–669.

Kussick, S. *et al.* (2005) 'Four-color flow cytometry shows strong concordance with bone marrow morphology and cytogenetics in the evaluation for myelodysplasia', *American Journal of Clinical Pathology*, 124(2), pp. 170–181. doi: 10.1309/6PBP-78G4-FBA1-FDG6.

Kwok, B. *et al.* (2015) 'MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance', *Blood*, 126(21), pp. 2355–2361. doi: 10.1182/blood-2015-08-667063.

Kyriakou, D. *et al.* (2001) 'Reduced CD43 expression on the neutrophils of MDS patients correlates with an activated phenotype of these cells', *International Journal of Hematology*, 73(4), pp. 483–491.

Lacronique-Gazaille, C. *et al.* (2007) 'A simple method for detection of major phenotypic abnormalities in myelodysplastic syndromes: expression of CD56 in CMML', *Haematologica*, 92(6), pp. 859–860.

Lai, J. *et al.* (1995) 'Myelodysplastic syndromes and acute myeloid leukemia with 17p deletion. An entity characterized by specific dysgranulopoïesis and a high incidence of P53 mutations', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 9(3), pp. 370–381.

Lambertenghi-Deliliers, G. *et al.* (1991) 'Myelodysplastic syndrome with increased marrow fibrosis: a distinct clinico-pathological entity', *British Journal of Haematology*, 78(2), pp. 161–166.

Langemeijer, S. *et al.* (2009) 'Acquired mutations in TET2 are common in myelodysplastic syndromes', *Nature Genetics*, 41(7), pp. 838–842. doi: 10.1038/ng.391.

Lasho, T L *et al.* (2012) 'SF3B1 mutations in primary myelofibrosis: clinical, histopathology and genetic correlates among 155 patients', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 26(5), pp. 1135–1137. doi: 10.1038/leu.2011.320.

Lasho, Terra L *et al.* (2012) 'SRSF2 mutations in primary myelofibrosis: significant clustering with IDH mutations and independent association with inferior overall and leukemia-free survival', *Blood*. doi: 10.1182/blood-2012-05-429696.

Laurie, C. C. *et al.* (2012) 'Detectable clonal mosaicism from birth to old age and its relationship to cancer', *Nature Genetics*, 44(6), pp. 642–650. doi: 10.1038/ng.2271.

Lee, J. W. *et al.* (2006) 'The JAK2 V617F mutation in de novo acute myelogenous leukemias', *Oncogene*, 25(9), pp. 1434–1436. doi: 10.1038/sj.onc.1209163.

Levine, R. *et al.* (2005) 'Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis', *Cancer Cell*, 7(4), pp. 387–397. doi: 10.1016/j.ccr.2005.03.023.

Ley, T. *et al.* (2010) 'DNMT3A mutations in acute myeloid leukemia', *The New England Journal of Medicine*, 363(25), pp. 2424–2433. doi: 10.1056/NEJMoa1005143.

Li, M. *et al.* (2011) 'Somatic mutations in the transcriptional corepressor gene BCORL1 in adult acute myelogenous leukemia', *Blood*, 118(22), pp. 5914–5917. doi: 10.1182/blood-2011-05-356204.

Li, M. M. *et al.* (2017) 'Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer', *The Journal of Molecular Diagnostics : JMD*, 19(1), pp. 4–23. doi: 10.1016/j.jmoldx.2016.10.002.

Li, Z. *et al.* (2011) 'Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies', *Blood*, 118(17), pp. 4509–4518. doi: 10.1182/blood-2010-12-325241.

Lin, J. *et al.* (2011) 'Recurrent DNMT3A R882 mutations in Chinese patients with acute myeloid leukemia and myelodysplastic syndrome', *PloS one*, 6(10), p. e26906. doi: 10.1371/journal.pone.0026906.

Lin, J. *et al.* (2012) 'IDH1 and IDH2 mutation analysis in Chinese patients with acute myeloid leukemia and myelodysplastic syndrome', *Annals of hematology*, 91(4), pp. 519–525. doi: 10.1007/s00277-011-1352-7.

Lin, L.-I. *et al.* (2005) 'Characterization of CEBPA mutations in acute myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells', *Clinical cancer research: an official journal of the American Association for Cancer Research*, 11(4), pp. 1372–1379. doi: 10.1158/1078-0432.CCR-04-1816.

Lin, Y. *et al.* (2017) 'Prognostic role of TET2 deficiency in myelodysplastic syndromes: A meta-analysis', *Oncotarget*. doi: 10.18632/oncotarget.17177.

Lindsley, R. C. and Ebert, B. L. (2013) 'The biology and clinical impact of genetic lesions in myeloid malignancies', *Blood*, 122(23), pp. 3741–3748. doi: 10.1182/blood-2013-06-460295.

Loken, M. and Wells, D. (2008) 'The role of flow cytometry in myelodysplastic syndromes', *Journal of the National Comprehensive Cancer Network: JNCCN*, 6(9), pp. 935–941.

van de Loosdrecht, A. *et al.* (2008) 'Identification of distinct prognostic subgroups in low- and intermediate-1-risk myelodysplastic syndromes by flow cytometry', *Blood*, 111(3), pp. 1067–1077. doi: 10.1182/blood-2007-07-098764.

van de Loosdrecht, A. *et al.* (2009) 'Standardization of flow cytometry in myelodysplastic syndromes: report from the first European LeukemiaNet working conference on flow cytometry in myelodysplastic syndromes', *Haematologica*, 94(8), pp. 1124–1134. doi: 10.3324/haematol.2009.005801.

Ma, W. *et al.* (2011) 'MPL mutation profile in JAK2 mutation-negative patients with myeloproliferative disorders', *Diagnostic molecular pathology: the American journal of surgical pathology, part B*, 20(1), pp. 34–39. doi: 10.1097/PDM.0b013e3181ecd261.

Maciejewski, J., Tiu, R. V. and O'Keefe, C. (2009) 'Application of array-based whole genome scanning technologies as a cytogenetic tool in haematological malignancies', *British Journal of Haematology*, 146(5), pp. 479–488. doi: 10.1111/j.1365-2141.2009.07757.x.

Mahmood, T. *et al.* (1979) 'Macrocytic anemia, thrombocytosis and nonlobulated megakaryocytes: the 5q-syndrome, a distinct entity', *The American Journal of Medicine*, 66(6), pp. 946–950.

Makishima, H., Rataul, M., *et al.* (2010) 'FISH and SNP-A karyotyping in myelodysplastic syndromes: improving cytogenetic detection of del(5q), monosomy 7, del(7q), trisomy 8 and del(20q)', *Leukemia Research*, 34(4), pp. 447–453. doi: 10.1016/j.leukres.2009.08.023.

Makishima, H., Jankowska, A., *et al.* (2010) 'Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 24(10), pp. 1799–1804. doi: 10.1038/leu.2010.167.

Makishima, H. *et al.* (2012a) 'Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis', *Blood*, 119(14), pp. 3203–3210. doi: 10.1182/blood-2011-12-399774.

Makishima, H. *et al.* (2012b) 'Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis', *Blood*. doi: 10.1182/blood-2011-12-399774.

Makishima, H. *et al.* (2013) 'Somatic SETBP1 mutations in myeloid malignancies', *Nature genetics*. doi: 10.1038/ng.2696.

Malcovati, L. *et al.* (2005) 'Flow cytometry evaluation of erythroid and myeloid dysplasia in patients with myelodysplastic syndrome', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 19(5), pp. 776–783. doi: 10.1038/sj.leu.2403680.

Malcovati, L. *et al.* (2009) 'Molecular and clinical features of refractory anemia with ringed sideroblasts associated with marked thrombocytosis', *Blood*, 114(17), pp. 3538–3545. doi: 10.1182/blood-2009-05-222331.

Malcovati, L. *et al.* (2011) 'Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms', *Blood*, 118(24), pp. 6239–6246. doi: 10.1182/blood-2011-09-377275.

Malcovati, L. *et al.* (2013) 'Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet', *Blood*, 122(17), pp. 2943–2964. doi: 10.1182/blood-2013-03-492884.

Malcovati, L. *et al.* (2014) 'Driver somatic mutations identify distinct disease entities within myeloid neoplasms with myelodysplasia', *Blood*, 124(9), pp. 1513–1521. doi: 10.1182/blood-2014-03-560227.

Malcovati, L. *et al.* (2015) 'SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts', *Blood*. doi: 10.1182/blood-2015-03-633537.

Malcovati, L. *et al.* (2017) 'Clinical significance of somatic mutation in unexplained blood cytopenia', *Blood*, p. blood-2017-01-763425. doi: 10.1182/blood-2017-01-763425.

Marcucci, G. *et al.* (2010) 'IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 28(14), pp. 2348–2355. doi: 10.1200/JCO.2009.27.3730.

Mardis, E. (2008) 'Next-generation DNA sequencing methods', *Annual Review of Genomics and Human Genetics*, 9, pp. 387–402. doi: 10.1146/annurev.genom.9.081307.164359.

Mardis, E. *et al.* (2009) 'Recurring mutations found by sequencing an acute myeloid leukemia genome', *The New England Journal of Medicine*, 361(11), pp. 1058–1066. doi: 10.1056/NEJMoa0903840.

Martiat, P., Michaux, J. and Rodhain, J. (1991) 'Philadelphia-negative (Ph-) chronic myeloid leukemia (CML): comparison with Ph+ CML and chronic myelomonocytic leukemia. The Groupe Français de Cytogénétique Hématologique', *Blood*, 78(1), pp. 205–211.

Martínez-Avilés, L. *et al.* (2007) 'JAK2 exon 12 mutations in polycythemia vera or idiopathic erythrocytosis', *Haematologica*, 92(12), pp. 1717–1718. doi: 10.3324/haematol.12011.

Mason, C. C. *et al.* (2016) 'Age-related mutations and chronic myelomonocytic leukemia', *Leukemia*, 30(4), pp. 906–913. doi: 10.1038/leu.2015.337.

Matarraz, S. *et al.* (2008) 'The immunophenotype of different immature, myeloid and B-cell lineage-committed CD34+ hematopoietic cells allows discrimination between normal/reactive and myelodysplastic syndrome precursors', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 22(6), pp. 1175–1183. doi: 10.1038/leu.2008.49.

Matsuda, A. *et al.* (1998) 'Refractory anemia with severe dysplasia: clinical significance of morphological features in refractory anemia', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 12(4), pp. 482–485.

Matsuda, A. *et al.* (1999) 'New system for assessing the prognosis of refractory anemia patients', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 13(11), pp. 1727–1734.

Matsuda, A. *et al.* (2007) 'Improvement of criteria for refractory cytopenia with multilineage dysplasia according to the WHO classification based on prognostic significance of morphological features in patients with refractory anemia according to the FAB classification', *Leukemia: Official Journal of the Leukemia Society of*

America, Leukemia Research Fund, U.K, 21(4), pp. 678–686. doi: 10.1038/sj.leu.2404571.

Maxson, J. E. *et al.* (2013) 'Oncogenic CSF3R mutations in chronic neutrophilic leukemia and atypical CML', *The New England journal of medicine*, 368(19), pp. 1781–1790. doi: 10.1056/NEJMoa1214514.

Maynadié, M. *et al.* (2002) 'Immunophenotypic clustering of myelodysplastic syndromes', *Blood*, 100(7), pp. 2349–2356. doi: 10.1182/blood-2002-01-0230.

Mazumdar, C. *et al.* (2015) 'Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation', *Cell Stem Cell*, 17(6), pp. 675–688. doi: 10.1016/j.stem.2015.09.017.

McKerrell, T. *et al.* (2015) 'Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis', *Cell Reports*, 10(8), pp. 1239–1245. doi: 10.1016/j.celrep.2015.02.005.

McLaren, W. *et al.* (2010) 'Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor', *Bioinformatics (Oxford, England)*, 26(16), pp. 2069–2070. doi: 10.1093/bioinformatics/btq330.

McMullin, M. F. *et al.* (2007) 'Amendment to the guideline for diagnosis and investigation of polycythaemia/erythrocytosis', *British Journal of Haematology*, 138(6), pp. 821–822. doi: 10.1111/j.1365-2141.2007.06741.x.

Mead, A. J. *et al.* (2007) 'FLT3 tyrosine kinase domain mutations are biologically distinct from and have a significantly more favorable prognosis than FLT3 internal tandem duplications in patients with acute myeloid leukemia', *Blood*, 110(4), pp. 1262–1270. doi: 10.1182/blood-2006-04-015826.

Meggendorfer, M. *et al.* (2012) 'SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML)', *Blood*. doi: 10.1182/blood-2012-01-404863.

Metzeler, K. H., Becker, H., *et al.* (2011) 'ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category', *Blood*, 118(26), pp. 6920–6929. doi: 10.1182/blood-2011-08-368225.

Metzeler, K. H., Maharry, K., *et al.* (2011) 'TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 29(10), pp. 1373–1381. doi: 10.1200/JCO.2010.32.7742.

Michaux, J. and Martiat, P. (1993) 'Chronic myelomonocytic leukaemia (CMML)--a myelodysplastic or myeloproliferative syndrome?', *Leukemia & Lymphoma*, 9(1–2), pp. 35–41. doi: 10.3109/10428199309148501.

Mills, K. *et al.* (2009) 'Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome', *Blood*, 114(5), pp. 1063–1072. doi: 10.1182/blood-2008-10-187203.

Misawa, S. *et al.* (1997) 'Genetic aberrations in the development and subsequent progression of myelodysplastic syndrome', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 11 Suppl 3, pp. 533–535.

Misawa, S. and Horiike, S. (1996) 'TP53 mutations in myelodysplastic syndrome', *Leukemia & lymphoma*, 23(5–6), pp. 417–422. doi: 10.3109/10428199609054848.

Mittelman, M. *et al.* (1993) 'High Ia (HLA-DR) and low CD11b (Mo1) expression may predict early conversion to leukemia in myelodysplastic syndromes', *American Journal of Hematology*, 43(3), pp. 165–171.

Mohamedali, A. *et al.* (2007) 'Prevalence and prognostic significance of allelic imbalance by single-nucleotide polymorphism analysis in low-risk myelodysplastic syndromes', *Blood*, 110(9), pp. 3365–3373. doi: 10.1182/blood-2007-03-079673.

Mohamedali, A. *et al.* (2009) 'Novel TET2 mutations associated with UPD4q24 in myelodysplastic syndrome', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 27(24), pp. 4002–4006. doi: 10.1200/JCO.2009.22.6985.

Morozova, O. and Marra, M. (2008) 'Applications of next-generation sequencing technologies in functional genomics', *Genomics*, 92(5), pp. 255–264. doi: 10.1016/j.ygeno.2008.07.001.

Mufti, G. *et al.* (2008) 'Diagnosis and classification of myelodysplastic syndrome: International Working Group on Morphology of myelodysplastic syndrome (IWGM-MDS) consensus proposals for the definition and enumeration of myeloblasts and ring sideroblasts', *Haematologica*, 93(11), pp. 1712–1717. doi: 10.3324/haematol.13405.

Muramatsu, H., Makishima, H. and Maciejewski, J. P. (2012) 'Chronic myelomonocytic leukemia and atypical chronic myeloid leukemia: novel pathogenetic lesions', *Seminars in oncology*, 39(1), pp. 67–73. doi: 10.1053/j.seminoncol.2011.11.004.

Nagata, H. *et al.* (1995) 'Identification of a point mutation in the catalytic domain of the protooncogene c-kit in peripheral blood mononuclear cells of patients who have mastocytosis with an associated hematologic disorder', *Proceedings of the National Academy of Sciences of the United States of America*, 92(23), pp. 10560–10564.

Nakao, M. *et al.* (2004) 'Novel loss-of-function mutations of the haematopoiesis-related transcription factor, acute myeloid leukaemia 1/runt-related transcription

factor 1, detected in acute myeloblastic leukaemia and myelodysplastic syndrome', *British journal of haematology*, 125(6), pp. 709–719. doi: 10.1111/j.1365-2141.2004.04966.x.

Nangalia, J. *et al.* (2013) 'Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2', *The New England Journal of Medicine*, 369(25), pp. 2391–2405. doi: 10.1056/NEJMoa1312542.

Nikoloski, G. *et al.* (2010a) 'Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes', *Nature Genetics*, 42(8), pp. 665–667. doi: 10.1038/ng.620.

Nikoloski, G. *et al.* (2010b) 'Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes', *Nature Genetics*, 42(8), pp. 665–667. doi: 10.1038/ng.620.

Noguera, N. I. *et al.* (2005) 'Simultaneous detection of NPM1 and FLT3-ITD mutations by capillary electrophoresis in acute myeloid leukemia', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 19(8), pp. 1479–1482. doi: 10.1038/sj.leu.2403846.

Nomdedéu, J. *et al.* (2005) 'Low frequency of exon 3 PTPN11 mutations in adult de novo acute myeloid leukemia. Analysis of a consecutive series of 173 patients', *Haematologica*, 90(3), pp. 412–413.

Nösslinger, T. *et al.* (2001) 'Myelodysplastic syndromes, from French-American-British to World Health Organization: comparison of classifications on 431 unselected patients from a single institution', *Blood*, 98(10), pp. 2935–2941.

Nowak, D. *et al.* (2009) 'Genome-wide DNA-mapping of CD34+ cells from patients with myelodysplastic syndrome using 500K SNP arrays identifies significant regions of deletion and uniparental disomy', *Experimental Hematology*, 37(2), pp. 215–224. doi: 10.1016/j.exphem.2008.10.012.

Nowell, P. and Hungerford, D. (1960) 'A minute chromosome in human chronic granulocytic leukaemia', *Science*, 132, p. 1497.

O'Dwyer, M. and Druker, B. (2000) 'STI571: an inhibitor of the BCR-ABL tyrosine kinase for the treatment of chronic myelogenous leukaemia', *The Lancet Oncology*, 1, pp. 207–211.

Ogata, K. *et al.* (2002) 'Clinical significance of phenotypic features of blasts in patients with myelodysplastic syndrome', *Blood*, 100(12), pp. 3887–3896. doi: 10.1182/blood-2002-01-0222.

Ogata, K. *et al.* (2006) 'Diagnostic application of flow cytometric characteristics of CD34+ cells in low-grade myelodysplastic syndromes', *Blood*, 108(3), pp. 1037–1044. doi: 10.1182/blood-2005-12-4916.

Ogata, K. *et al.* (2009) 'Diagnostic utility of flow cytometry in low-grade myelodysplastic syndromes: a prospective validation study', *Haematologica*, 94(8), pp. 1066–1074. doi: 10.3324/haematol.2009.008532.

Oki, Y. *et al.* (2006) 'Mutations and promoter methylation status of NPM1 in myeloproliferative disorders', *Haematologica*, 91(8), pp. 1147–1148.

Olesen, L. H. *et al.* (2005) 'Delineation and molecular characterization of acute myeloid leukemia patients with coduplication of FLT3 and MLL', *European journal of haematology*, 75(3), pp. 185–192. doi: 10.1111/j.1600-0609.2005.00502.x.

Ormazábal, C. *et al.* (2008) 'Low frequency of JAK2 exon 12 mutations in classic and atypical CMPDs', *Leukemia research*, 32(9), pp. 1485–1487. doi: 10.1016/j.leukres.2008.01.002.

Palomero, T. *et al.* (2014) 'Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas', *Nature Genetics*, 46(2), pp. 166–170. doi: 10.1038/ng.2873.

Palomo, L., Xicoy, B., *et al.* (2016) 'Impact of SNP array karyotyping on the diagnosis and the outcome of chronic myelomonocytic leukemia with low risk cytogenetic features or no metaphases', *American Journal of Hematology*, 91(2), pp. 185–192. doi: 10.1002/ajh.24227.

Palomo, L., Garcia, O., *et al.* (2016) 'Targeted deep sequencing improves outcome stratification in chronic myelomonocytic leukemia with low risk cytogenetic features', *Oncotarget*, 7(35), pp. 57021–57035. doi: 10.18632/oncotarget.10937.

Papaemmanuil, E. *et al.* (2011) 'Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts', *The New England journal of medicine*, 365(15), pp. 1384–1395. doi: 10.1056/NEJMoa1103283.

Papaemmanuil, E. *et al.* (2013) 'Clinical and biological implications of driver mutations in myelodysplastic syndromes', *Blood*, 122(22), pp. 3616–3627. doi: 10.1182/blood-2013-08-518886.

Pardanani, A. *et al.* (2003) 'Flt-3 and c-kit mutation studies in a spectrum of chronic myeloid disorders including systemic mast cell disease', *Leukemia research*, 27(8), pp. 739–742.

Pardanani, A. *et al.* (2006) 'MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients', *Blood*, 108(10), pp. 3472–3476. doi: 10.1182/blood-2006-04-018879.

Pardanani, A. *et al.* (2010) 'IDH1 and IDH2 mutation analysis in chronic- and blast-phase myeloproliferative neoplasms', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 24(6), pp. 1146–1151. doi: 10.1038/leu.2010.77.

Pardanani, A. *et al.* (2011) 'Primary myelofibrosis with or without mutant MPL: comparison of survival and clinical features involving 603 patients', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 25(12), pp. 1834–1839. doi: 10.1038/leu.2011.161.

Pardanani, A. *et al.* (2013) 'CSF3R T618I is a highly prevalent and specific mutation in chronic neutrophilic leukemia', *Leukemia*, 27(9), pp. 1870–1873. doi: 10.1038/leu.2013.122.

Pardanani, A. D. *et al.* (2006) 'MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients', *Blood*, 108(10), pp. 3472–3476. doi: 10.1182/blood-2006-04-018879.

Parmentier, S. *et al.* (2012) 'Assessment of dysplastic hematopoiesis: lessons from healthy bone marrow donors', *Haematologica*, 97(5), pp. 723–730. doi: 10.3324/haematol.2011.056879.

Paschka, P. *et al.* (2008) 'Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 26(28), pp. 4595–4602. doi: 10.1200/JCO.2007.15.2058.

Paschka, P. *et al.* (2010) 'IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 28(22), pp. 3636–3643. doi: 10.1200/JCO.2010.28.3762.

Patel, K. P. *et al.* (2011) 'Acute myeloid leukemia with IDH1 or IDH2 mutation: frequency and clinicopathologic features', *American journal of clinical pathology*, 135(1), pp. 35–45. doi: 10.1309/AJCPD7NR2RMNQDVF.

Patnaik, M. M. *et al.* (2010) 'WHO-defined "myelodysplastic syndrome with isolated del(5q)" in 88 consecutive patients: survival data, leukemic transformation rates and prevalence of JAK2, MPL and IDH mutations', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 24(7), pp. 1283–1289. doi: 10.1038/leu.2010.105.

Patnaik, M M *et al.* (2012) 'Differential prognostic effect of IDH1 versus IDH2 mutations in myelodysplastic syndromes: a Mayo Clinic study of 277 patients', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 26(1), pp. 101–105. doi: 10.1038/leu.2011.298.

Patnaik, Mrinal M, Hanson, C. A., *et al.* (2012) 'Prognostic irrelevance of ring sideroblast percentage in World Health Organization defined myelodysplastic syndromes without excess blasts', *Blood*. doi: 10.1182/blood-2012-03-415356.

Patnaik, Mrinal M, Lasho, T. L., *et al.* (2012) 'SF3B1 mutations are prevalent in myelodysplastic syndromes with ring sideroblasts but do not hold independent prognostic value', *Blood*, 119(2), pp. 569–572. doi: 10.1182/blood-2011-09-377994.

Patnaik, M. M. *et al.* (2013) 'Mayo Prognostic Model for WHO-Defined Chronic Myelomonocytic Leukemia- ASXL1 and Spliceosome Component Mutations and Outcomes', *Leukemia*. doi: 10.1038/leu.2013.88.

Patnaik, M. M. *et al.* (2014) 'ASXL1 and SETBP1 mutations and their prognostic contribution in chronic myelomonocytic leukemia: a two-center study of 466 patients', *Leukemia*, 28(11), pp. 2206–2212. doi: 10.1038/leu.2014.125.

Patnaik, M. M. *et al.* (2015) 'Chronic myelomonocytic leukemia in younger patients: molecular and cytogenetic predictors of survival and treatment outcome', *Blood Cancer Journal*, 5, p. e270. doi: 10.1038/bcj.2014.90.

Patnaik, M. M. *et al.* (2016) 'Prognostic interaction between ASXL1 and TET2 mutations in chronic myelomonocytic leukemia', *Blood Cancer Journal*, 6, p. e385. doi: 10.1038/bcj.2015.113.

Patnaik, M. M. *et al.* (2017) 'Flow cytometry based monocyte subset analysis accurately distinguishes chronic myelomonocytic leukemia from myeloproliferative neoplasms with associated monocytosis', *Blood Cancer Journal*, 7(7), p. e584; doi: 10.1038/bcj.2017.66.

Percy, M. J. *et al.* (2007) 'The frequency of JAK2 exon 12 mutations in idiopathic erythrocytosis patients with low serum erythropoietin levels', *Haematologica*, 92(12), pp. 1607–1614. doi: 10.3324/haematol.11643.

Piazza, R. *et al.* (2012) 'Recurrent SETBP1 mutations in atypical chronic myeloid leukemia', *Nature genetics*, 45(1), pp. 18–24. doi: 10.1038/ng.2495.

Pietra, D. *et al.* (2008) 'Somatic mutations of JAK2 exon 12 in patients with JAK2 (V617F)-negative myeloproliferative disorders', *Blood*, 111(3), pp. 1686–1689. doi: 10.1182/blood-2007-07-101576.

Pietra, D. *et al.* (2011) 'Deep sequencing reveals double mutations in cis of MPL exon 10 in myeloproliferative neoplasms', *Haematologica*, 96(4), pp. 607–611. doi: 10.3324/haematol.2010.034793.

Pikman, Y. *et al.* (2006) 'MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia', *PLoS Medicine*, 3(7), p. e270. doi: 10.1371/journal.pmed.0030270.

Porwit, A. *et al.* (2014) 'Revisiting guidelines for integration of flow cytometry results in the WHO classification of myelodysplastic syndromes-proposal from the International/European LeukemiaNet Working Group for Flow Cytometry in MDS', *Leukemia*, 28(9), pp. 1793–1798. doi: 10.1038/leu.2014.191.

Pratcorona, M. *et al.* (2012) 'Acquired mutations in ASXL1 in acute myeloid leukemia: prevalence and prognostic value', *Haematologica*, 97(3), pp. 388–392. doi: 10.3324/haematol.2011.051532.

Preudhomme, C. *et al.* (2002) 'Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: a study from the Acute Leukemia French Association (ALFA)', *Blood*, 100(8), pp. 2717–2723. doi: 10.1182/blood-2002-03-0990.

Qian, J. *et al.* (2012) 'U2AF1 Mutations in Chinese Patients with Acute Myeloid Leukemia and Myelodysplastic Syndrome', *PloS one*, 7(9), p. e45760. doi: 10.1371/journal.pone.0045760.

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramos, F. *et al.* (1999) 'Myelodysplastic syndrome: a search for minimal diagnostic criteria', *Leukemia Research*, 23(3), pp. 283–290.

Rausch, T. *et al.* (2012) 'Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations', *Cell*, 148(1–2), pp. 59–71. doi: 10.1016/j.cell.2011.12.013.

Raza, S. *et al.* (2011) 'TP53 mutations and polymorphisms in primary myelofibrosis', *American journal of hematology*. doi: 10.1002/ajh.22216.

Ricci, C. *et al.* (2012) 'ASXL1 mutations in primary and secondary myelofibrosis', *British journal of haematology*, 156(3), pp. 404–407. doi: 10.1111/j.1365-2141.2011.08865.x.

Ridge, S. A. *et al.* (1990) 'FMS mutations in myelodysplastic, leukemic, and normal subjects', *Proceedings of the National Academy of Sciences of the United States of America*, 87(4), pp. 1377–1380.

Rocquain, J. *et al.* (2010) 'Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNX1, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias', *BMC Cancer*, 10, p. 401. doi: 10.1186/1471-2407-10-401.

Roman, E. *et al.* (2016) 'Myeloid malignancies in the real-world: Occurrence, progression and survival in the UK's population-based Haematological Malignancy Research Network 2004-15', *Cancer Epidemiology*, 42, pp. 186–198. doi: 10.1016/j.canep.2016.03.011.

Rosati, S. *et al.* (1996) 'Refractory cytopenia with multilineage dysplasia: further characterization of an "unclassifiable" myelodysplastic syndrome', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 10(1), pp. 20–26.

Rowley, J. (1973) ‘Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining’, *Nature*, 243(5405), pp. 290–293.

Ruan, G.-R. *et al.* (2010) ‘MPL W515L/K mutations in 343 Chinese adults with JAK2V617F mutation-negative chronic myeloproliferative disorders detected by a newly developed RQ-PCR based on TaqMan MGB probes’, *Hematological oncology*, 28(1), pp. 33–39. doi: 10.1002/hon.899.

Rücker, F. G. *et al.* (2012) ‘TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome’, *Blood*, 119(9), pp. 2114–2121. doi: 10.1182/blood-2011-08-375758.

Sakata-Yanagimoto, M. *et al.* (2014) ‘Somatic RHOA mutation in angioimmunoblastic T cell lymphoma’, *Nature Genetics*, 46(2), pp. 171–175. doi: 10.1038/ng.2872.

Sanger, F., Nicklen, S. and Coulson, A. (1977) ‘DNA sequencing with chain-terminating inhibitors’, *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp. 5463–5467.

Satoh, C. *et al.* (2008) ‘Flow cytometric parameters with little interexaminer variability for diagnosing low-grade myelodysplastic syndromes’, *Leukemia Research*, 32(5), pp. 699–707. doi: 10.1016/j.leukres.2007.08.022.

Schanz, J. *et al.* (2011) ‘Coalesced Multicentric Analysis of 2,351 Patients With Myelodysplastic Syndromes Indicates an Underestimation of Poor-Risk Cytogenetics of Myelodysplastic Syndromes in the International Prognostic Scoring System’, *Journal of Clinical Oncology*, 29(15), pp. 1963–1970. doi: 10.1200/JCO.2010.28.3978.

Schanz, J. *et al.* (2012) ‘New Comprehensive Cytogenetic Scoring System for Primary Myelodysplastic Syndromes (MDS) and Oligoblastic Acute Myeloid Leukemia After MDS Derived From an International Database Merge’, *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*. doi: 10.1200/JCO.2011.35.6394.

Schirmer, M. *et al.* (2015) ‘Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform’, *Nucleic Acids Research*, 43(6), p. e37. doi: 10.1093/nar/gku1341.

Schlenk, R. F. *et al.* (2008) ‘Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia’, *The New England journal of medicine*, 358(18), pp. 1909–1918. doi: 10.1056/NEJMoa074306.

Schlesinger, M. *et al.* (1996) ‘Analysis of myeloid and lymphoid markers on the surface and in the cytoplasm of mononuclear bone marrow cells in patients with

myelodysplastic syndrome', *Journal of Clinical & Laboratory Immunology*, 48(4), pp. 149–166.

Schmitt-Graeff, A. *et al.* (2002) 'Essential thrombocythemia with ringed sideroblasts: a heterogeneous spectrum of diseases, but not a distinct entity', *Haematologica*, 87(4), pp. 392–399.

Schnittger, S. *et al.* (2002) 'Analysis of FLT3 length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease', *Blood*, 100(1), pp. 59–66.

Schnittger, S. *et al.* (2005) 'Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype', *Blood*, 106(12), pp. 3733–3739. doi: 10.1182/blood-2005-06-2248.

Schnittger, S. *et al.* (2010) 'IDH1 mutations are detected in 6.6% of 1414 AML patients and are associated with intermediate risk karyotype and unfavorable prognosis in adults younger than 60 years and unmutated NPM1 status', *Blood*, 116(25), pp. 5486–5496. doi: 10.1182/blood-2010-02-267955.

Schnittger, S. *et al.* (2011) 'Characterization of NPM1-mutated AML with a history of myelodysplastic syndromes or myeloproliferative neoplasms', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 25(4), pp. 615–621. doi: 10.1038/leu.2010.299.

Schnittger, S *et al.* (2012) 'ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*. doi: 10.1038/leu.2012.262.

Schnittger, Susanne *et al.* (2012) 'Use of CBL exon 8 and 9 mutations in diagnosis of myeloproliferative neoplasms and myeloproliferative/myelodysplastic disorders: an analysis of 636 cases', *Haematologica*. doi: 10.3324/haematol.2012.065375.

Schuler, E. *et al.* (2014) 'Refined medullary blast and white blood cell count based classification of chronic myelomonocytic leukemias', *Leukemia Research*, 38(12), pp. 1413–1419. doi: 10.1016/j.leukres.2014.09.003.

Scott, L. *et al.* (2007) 'JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis', *The New England Journal of Medicine*, 356(5), pp. 459–468. doi: 10.1056/NEJMoa065202.

Seabright, M. (1971) 'A rapid banding technique for human chromosomes', *Lancet*, 2(7731), pp. 971–972.

'Second International Workshop on Chromosomes in Leukemia' (1980) *Cancer Research*, 40(12), pp. 4826–4827.

Selimoglu-Buet, D. *et al.* (2015) 'Characteristic repartition of monocyte subsets as a diagnostic signature of chronic myelomonocytic leukemia', *Blood*, 125(23), pp. 3618–3626. doi: 10.1182/blood-2015-01-620781.

Seymour, J. and Estey, E. (1993) 'The prognostic significance of auer rods in myelodysplasia', *British Journal of Haematology*, 85(1), pp. 67–76.

Shapiro, H. (1988) *Practical Flow Cytometry*. 2nd Edition. New York: Alan R. Liss, Inc.

Shaw, G. (2005) 'Ringed sideroblasts with thrombocytosis: an uncommon mixed myelodysplastic/myeloproliferative disease of older adults', *British Journal of Haematology*, 131(2), pp. 180–184. doi: 10.1111/j.1365-2141.2005.05747.x.

Shih, L. *et al.* (2004) 'Acquisition of FLT3 or N-ras mutations is frequently associated with progression of myelodysplastic syndrome to acute myeloid leukemia', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 18(3), pp. 466–475. doi: 10.1038/sj.leu.2403274.

Shih, L.-Y. *et al.* (2004) 'Acquisition of FLT3 or N-ras mutations is frequently associated with progression of myelodysplastic syndrome to acute myeloid leukemia', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 18(3), pp. 466–475. doi: 10.1038/sj.leu.2403274.

Shukla, G. C. and Singh, J. (2017) 'Mutations of RNA splicing factors in hematological malignancies', *Cancer Letters*, 409, pp. 1–8. doi: 10.1016/j.canlet.2017.08.042.

Simons, A. *et al.* (2012) 'Genome-wide arrays in routine diagnostics of hematological malignancies', *Human mutation*, 33(6), pp. 941–948. doi: 10.1002/humu.22057.

Smith, A. *et al.* (2018) 'Cohort Profile: The Haematological Malignancy Research Network (HMRN); a UK population-based patient cohort', *International Journal of Epidemiology*. doi: 10.1093/ije/dyy044.

Smith, A. E. *et al.* (2010) 'Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with early origins, but indicates no definite prognostic value', *Blood*, 116(19), pp. 3923–3932. doi: 10.1182/blood-2010-03-274704.

Stachurski, D. *et al.* (2008) 'Flow cytometric analysis of myelomonocytic cells by a pattern recognition approach is sensitive and specific in diagnosing myelodysplastic syndrome and related marrow diseases: emphasis on a global evaluation and recognition of diagnostic pitfalls', *Leukemia Research*, 32(2), pp. 215–224. doi: 10.1016/j.leukres.2007.06.012.

Steensma, D. and List, A. (2005) 'Genetic testing in the myelodysplastic syndromes: molecular insights into hematologic diversity', *Mayo Clinic Proceedings. Mayo Clinic*, 80(5), pp. 681–698.

Steensma, D. P. *et al.* (2005) 'The JAK2 V617F activating tyrosine kinase mutation is an infrequent event in both "atypical" myeloproliferative disorders and myelodysplastic syndromes', *Blood*, 106(4), pp. 1207–1209. doi: 10.1182/blood-2005-03-1183.

Steensma, D. P. (2012) 'Dysplasia Has A Differential Diagnosis: Distinguishing Genuine Myelodysplastic Syndromes (MDS) From Mimics, Imitators, Copycats and Impostors', *Current hematologic malignancy reports*. doi: 10.1007/s11899-012-0140-3.

Steensma, D. P. *et al.* (2015) 'Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes', *Blood*, 126(1), pp. 9–16. doi: 10.1182/blood-2015-03-631747.

Stegelmann, F. *et al.* (2011) 'DNMT3A mutations in myeloproliferative neoplasms', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 25(7), pp. 1217–1219. doi: 10.1038/leu.2011.77.

Stein, B. L. *et al.* (2011) 'Disruption of the ASXL1 gene is frequent in primary, post-essential thrombocytosis and post-polycythemia vera myelofibrosis, but not essential thrombocytosis or polycythemia vera: analysis of molecular genetics and clinical phenotypes', *Haematologica*, 96(10), pp. 1462–1469. doi: 10.3324/haematol.2011.045591.

Stetler-Stevenson, M. *et al.* (2001) 'Diagnostic utility of flow cytometric immunophenotyping in myelodysplastic syndrome', *Blood*, 98(4), pp. 979–987.

Stirewalt, D. L. *et al.* (2001) 'FLT3, RAS, and TP53 mutations in elderly patients with acute myeloid leukemia', *Blood*, 97(11), pp. 3589–3595.

Streeter, R., Presant, C. and Reinhard, E. (1977) 'Prognostic significance of thrombocytosis in idiopathic sideroblastic anemia', *Blood*, 50(3), pp. 427–432.

Subirá, D. *et al.* (2008) 'Immunophenotype in chronic myelomonocytic leukemia: is it closer to myelodysplastic syndromes or to myeloproliferative disorders?', *Translational Research: The Journal of Laboratory and Clinical Medicine*, 151(5), pp. 240–245. doi: 10.1016/j.trsl.2008.03.001.

Such, E. *et al.* (2013) 'Development and validation of a prognostic scoring system for patients with chronic myelomonocytic leukemia', *Blood*, 121(15), pp. 3005–3015. doi: 10.1182/blood-2012-08-452938.

Sugimoto, K. *et al.* (1993) 'Mutations of the p53 gene in myelodysplastic syndrome (MDS) and MDS-derived leukemia', *Blood*, 81(11), pp. 3022–3026.

Summers, K. *et al.* (2007) 'Wilms' tumour 1 mutations are associated with FLT3-ITD and failure of standard induction chemotherapy in patients with normal karyotype AML', *Leukemia: official journal of the Leukemia Society of America, Leukemia*

Research Fund, U.K, 21(3), pp. 550–551; author reply 552. doi: 10.1038/sj.leu.2404514.

Suzuki, T. *et al.* (2005) 'Clinical characteristics and prognostic implications of NPM1 mutations in acute myeloid leukemia', *Blood*, 106(8), pp. 2854–2861. doi: 10.1182/blood-2005-04-1733.

Swerdlow, S. H. *et al.* (2008) *WHO Classification of Tumours of Haemopoietic and Lymphoid tissues*. 4th Edition. Lyon, France: International Agency for Research on Cancer.

Swerdlow, S. H. *et al.* (2017) *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Revised 4th Edition. Lyon, France: International Agency for Research on Cancer.

Szpurka, H. *et al.* (2009) 'UPD1p indicates the presence of MPL W515L mutation in RARS-T, a mechanism analogous to UPD9p and JAK2 V617F mutation', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 23(3), pp. 610–614. doi: 10.1038/leu.2008.249.

Szpurka, H. *et al.* (2010) 'Spectrum of mutations in RARS-T patients includes TET2 and ASXL1 mutations', *Leukemia Research*, 34(8), pp. 969–973. doi: 10.1016/j.leukres.2010.02.033.

Talati, C. *et al.* (2017) 'Monocyte subset analysis accurately distinguishes CMML from MDS and is associated with a favorable MDS prognosis', *Blood*, 129(13), pp. 1881–1883. doi: 10.1182/blood-2016-12-753210.

Tang, J. L. *et al.* (1998) 'P53 mutation in advanced stage of primary myelodysplastic syndrome', *Anticancer research*, 18(5B), pp. 3757–3761.

Tang, J.-L. *et al.* (2009) 'AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations', *Blood*, 114(26), pp. 5352–5361. doi: 10.1182/blood-2009-05-223784.

Taskesen, E. *et al.* (2011) 'Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity', *Blood*, 117(8), pp. 2469–2475. doi: 10.1182/blood-2010-09-307280.

Tefferi, A. *et al.* (1989) 'Chronic myelomonocytic leukemia: natural history and prognostic determinants', *Mayo Clinic Proceedings. Mayo Clinic*, 64(10), pp. 1246–1254.

Tefferi, A. *et al.* (2005) 'The JAK2(V617F) tyrosine kinase mutation in myelofibrosis with myeloid metaplasia: lineage specificity and clinical correlates', *British Journal of Haematology*, 131(3), pp. 320–328. doi: 10.1111/j.1365-2141.2005.05776.x.

Tefferi, A. *et al.* (2006) 'The clinical phenotype of wild-type, heterozygous, and homozygous JAK2V617F in polycythemia vera', *Cancer*, 106(3), pp. 631–635. doi: 10.1002/cncr.21645.

Tefferi, A. (2008) 'The history of myeloproliferative disorders: before and after Dameshek', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 22(1), pp. 3–13. doi: 10.1038/sj.leu.2404946.

Tefferi, A. *et al.* (2009) 'TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 23(5), pp. 905–911. doi: 10.1038/leu.2009.47.

Tefferi, A. *et al.* (2010) 'IDH1 and IDH2 mutation studies in 1473 patients with chronic-, fibrotic- or blast-phase essential thrombocythemia, polycythemia vera or myelofibrosis', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 24(7), pp. 1302–1309. doi: 10.1038/leu.2010.113.

Tefferi, A. *et al.* (2012) 'IDH mutations in primary myelofibrosis predict leukemic transformation and shortened survival: clinical evidence for leukemogenic collaboration with JAK2V617F', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 26(3), pp. 475–480. doi: 10.1038/leu.2011.253.

Thiede, C. *et al.* (2002) 'Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis', *Blood*, 99(12), pp. 4326–4335.

Thiede, C. *et al.* (2006) 'Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML)', *Blood*, 107(10), pp. 4011–4020. doi: 10.1182/blood-2005-08-3167.

Thol, F *et al.* (2010) 'IDH1 mutations in patients with myelodysplastic syndromes are associated with an unfavorable prognosis', *Haematologica*, 95(10), pp. 1668–1674. doi: 10.3324/haematol.2010.025494.

Thol, Felicitas *et al.* (2010) 'Prognostic impact of IDH2 mutations in cytogenetically normal acute myeloid leukemia', *Blood*, 116(4), pp. 614–616. doi: 10.1182/blood-2010-03-272146.

Thol, Felicitas, Damm, F., *et al.* (2011) 'Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 29(21), pp. 2889–2896. doi: 10.1200/JCO.2011.35.4894.

Thol, Felicitas, Friesen, I., *et al.* (2011) 'Prognostic significance of ASXL1 mutations in patients with myelodysplastic syndromes', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 29(18), pp. 2499–2506. doi: 10.1200/JCO.2010.33.4938.

Thol, F, Friesen, I., *et al.* (2011) 'Prognostic significance of ASXL1 mutations in patients with myelodysplastic syndromes', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 29(18), pp. 2499–2506. doi: 10.1200/JCO.2010.33.4938.

Thol, F, Winschel, C., *et al.* (2011) 'Rare occurence of DNMT3A mutations in myelodysplastic syndromes', *Haematologica*. doi: 10.3324/haematol.2011.045559.

Thol, F. *et al.* (2012) 'Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes', *Blood*, 119(15), pp. 3578–3584. doi: 10.1182/blood-2011-12-399337.

Thota, S. *et al.* (2014) 'Genetic alterations of the cohesin complex genes in myeloid malignancies', *Blood*, 124(11), pp. 1790–1798. doi: 10.1182/blood-2014-04-567057.

Tiu, R. *et al.* (2011) 'Prognostic impact of SNP array karyotyping in myelodysplastic syndromes and related myeloid malignancies', *Blood*, 117(17), pp. 4552–4560. doi: 10.1182/blood-2010-07-295857.

Truong, F. *et al.* (2009) 'The utility of flow cytometric immunophenotyping in cytopenic patients with a non-diagnostic bone marrow: a prospective study', *Leukemia Research*, 33(8), pp. 1039–1046. doi: 10.1016/j.leukres.2009.01.012.

Tuzuner, N. *et al.* (1995) 'Hypocellular myelodysplastic syndromes (MDS): new proposals', *British Journal of Haematology*, 91(3), pp. 612–617.

Valent, P. *et al.* (2007) 'Definitions and standards in the diagnosis and treatment of the myelodysplastic syndromes: Consensus statements and report from a working conference', *Leukemia Research*, 31(6), pp. 727–736. doi: 10.1016/j.leukres.2006.11.009.

Vardiman, J. *et al.* (2009) 'The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes', *Blood*, 114(5), pp. 937–951. doi: 10.1182/blood-2009-03-209262.

Vardiman, J., Harris, N. and Brunning, R. (2002) 'The World Health Organization (WHO) classification of the myeloid neoplasms', *Blood*, 100(7), pp. 2292–2302. doi: 10.1182/blood-2002-04-1199.

Verburgh, E. *et al.* (2003) 'Additional prognostic value of bone marrow histology in patients subclassified according to the International Prognostic Scoring System for myelodysplastic syndromes', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 21(2), pp. 273–282.

Vicente, C. *et al.* (2007) 'JAK2-V617F activating mutation in acute myeloid leukemia: prognostic impact and association with other molecular markers', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 21(11), pp. 2386–2390. doi: 10.1038/sj.leu.2404812.

Virappane, P. *et al.* (2008) 'Mutation of the Wilms' tumor 1 gene is a poor prognostic factor associated with chemotherapy resistance in normal karyotype acute myeloid leukemia: the United Kingdom Medical Research Council Adult Leukaemia Working Party', *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 26(33), pp. 5429–5435. doi: 10.1200/JCO.2008.16.0333.

Vizmanos, J. L. *et al.* (2006) 'JAK2 V617F mutation in classic chronic myeloproliferative diseases: a report on a series of 349 patients', *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 20(3), pp. 534–535. doi: 10.1038/sj.leu.2404086.

Voglová, J. *et al.* (2001) 'Myelodysplastic and myeloproliferative type of chronic myelomonocytic leukemia--distinct subgroups or two stages of the same disease?', *Leukemia Research*, 25(6), pp. 493–499.

Walter, M. *et al.* (2011) 'Recurrent DNMT3A mutations in patients with myelodysplastic syndromes', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 25(7), pp. 1153–1158. doi: 10.1038/leu.2011.44.

Wang, E. *et al.* (2011) 'Pseudo-Pelger-Huët anomaly induced by medications: a clinicopathologic study in comparison with myelodysplastic syndrome-related pseudo-Pelger-Huët anomaly', *American Journal of Clinical Pathology*, 135(2), pp. 291–303. doi: 10.1309/AJCPVFY95MAOBKRS.

Wang, J. C. and Chen, C. (1998) 'N-RAS oncogene mutations in patients with agnogenic myeloid metaplasia in leukemic transformation', *Leukemia research*, 22(7), pp. 639–643.

Wang, X. and Sun, Q. (2017) 'TP53 mutations, expression and interaction networks in human cancers', *Oncotarget*, 8(1), pp. 624–643. doi: 10.18632/oncotarget.13483.

Watkins, F. *et al.* (2004) 'Mutations in PTPN11 are rare in adult myelodysplastic syndromes and acute myeloid leukemia', *American journal of hematology*, 76(4), p. 417. doi: 10.1002/ajh.20134.

Watson, J. (1991) *Introduction to Flow Cytometry*. 1st Edition. Cambridge: Cambridge University Press.

Wells, D. *et al.* (2003) 'Myeloid and monocytic dyspoiesis as determined by flow cytometric scoring in myelodysplastic syndrome correlates with the IPSS and with outcome after hematopoietic stem cell transplantation', *Blood*, 102(1), pp. 394–403. doi: 10.1182/blood-2002-09-2768.

Westers, T. M. *et al.* (2012) 'Standardization of flow cytometry in myelodysplastic syndromes: a report from an international consortium and the European leukemianet working group', *Leukemia: Official Journal of the Leukemia Society of America, Leukemia Research Fund, U.K*. doi: 10.1038/leu.2012.30.

Wimazal, F. *et al.* (2007) 'Idiopathic cytopenia of undetermined significance (ICUS) versus low risk MDS: the diagnostic interface', *Leukemia Research*, 31(11), pp. 1461–1468. doi: 10.1016/j.leukres.2007.03.015.

Wolanskyj, A. *et al.* (2005) 'JAK2 mutation in essential thrombocythaemia: clinical associations and long-term prognostic relevance', *British Journal of Haematology*, 131(2), pp. 208–213. doi: 10.1111/j.1365-2141.2005.05764.x.

Wong, K. and Chan, J. (1991) 'Are "dysplastic" and hypogranular megakaryocytes specific markers for myelodysplastic syndrome?', *British Journal of Haematology*, 77(4), pp. 509–514.

Wouters, B. J. *et al.* (2009) 'Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome', *Blood*, 113(13), pp. 3088–3091. doi: 10.1182/blood-2008-09-179895.

Wu, S.-J. *et al.* (2012) 'The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution', *Blood*, 120(15), pp. 3106–3111. doi: 10.1182/blood-2012-02-412296.

Xiao, S. *et al.* (1998) 'FGFR1 is fused with a novel zinc-finger gene, ZNF198, in the t(8;13) leukaemia/lymphoma syndrome', *Nature Genetics*, 18(1), pp. 84–87. doi: 10.1038/ng0198-84.

Xie, M. *et al.* (2014) 'Age-related mutations associated with clonal hematopoietic expansion and malignancies', *Nature Medicine*, 20(12), pp. 1472–1478. doi: 10.1038/nm.3733.

Xu, F. *et al.* (1999) 'Tandem duplication of the FLT3 gene is found in acute lymphoblastic leukaemia as well as acute myeloid leukaemia but not in myelodysplastic syndrome or juvenile chronic myelogenous leukaemia in children', *British journal of haematology*, 105(1), pp. 155–162.

Xu, Y. *et al.* (2005) 'Flow cytometric analysis of monocytes as a tool for distinguishing chronic myelomonocytic leukemia from reactive monocytosis', *American Journal of Clinical Pathology*, 124(5), pp. 799–806. doi: 10.1309/HRJ1-XKTD-77J1-UTFM.

Yamamoto, Y. *et al.* (2001) 'Activating mutation of D835 within the activation loop of FLT3 in human hematologic malignancies', *Blood*, 97(8), pp. 2434–2439.

Yang, D. *et al.* (2005) 'Flow cytometric analysis of different CD14 epitopes can help identify immature monocytic populations', *American Journal of Clinical Pathology*, 124(6), pp. 930–936.

Yoshida, K *et al.* (2011) 'A nonsense mutation of IDH1 in myelodysplastic syndromes and related disorders', *Leukemia: official journal of the Leukemia Society of*

America, Leukemia Research Fund, U.K, 25(1), pp. 184–186. doi: 10.1038/leu.2010.241.

Yoshida, Kenichi *et al.* (2011) 'Frequent pathway mutations of splicing machinery in myelodysplasia', *Nature*, 478(7367), pp. 64–69. doi: 10.1038/nature10496.

Zhang, S.-J., Shi, J.-Y. and Li, J.-Y. (2009) 'GATA-2 L359 V mutation is exclusively associated with CML progression but not other hematological malignancies and GATA-2 P250A is a novel single nucleotide polymorphism', *Leukemia research*, 33(8), pp. 1141–1143. doi: 10.1016/j.leukres.2009.02.025.

Zhang, Y. *et al.* (2007) 'NPM1 mutations in myelodysplastic syndromes and acute myeloid leukemia with normal karyotype', *Leukemia Research*, 31(1), pp. 109–111. doi: 10.1016/j.leukres.2006.03.013.