Genetic aberrations in chronic myeloid neoplasms

Sophie R Laird

A thesis submitted to the University of Manchester for the degree of Doctor of Clinical Sciences (Life Sciences) in the Faculty of Biology, Medicine and Health

2021

The School of Medical Sciences

Contents

Final word count (main text including footnotes and endnotes): 52, 188

Abstract

Chronic myeloid neoplasms (CMNs) are a heterogeneous group of clonal blood disorders with diverse clinical outcomes. Genetic analysis plays a key role in the diagnosis of CMN, as well as identifying specific targets for therapy, indicators of prognosis and markers that can be used to gauge response to therapy. Until recently, such analysis mainly involved cytogenetics along with a limited number of targeted tests to detect specific somatic abnormalities. The development of genomic technologies, however, has greatly expanded our understanding of the pathogenesis of CMN, and presented new challenges to diagnostic laboratories which need to implement these complex tests in a way that fulfils the requirements of accreditation as well as producing quality controlled, reproducible and clinically useful results. Interpretation of genomic data is complicated by a number of factors, including uncertainty as to whether variants are drivers or passengers, somatic or germline and drivers of disease or age-related changes. Within this context I aimed to develop and apply myeloid panel next generation sequencing testing within the Wessex Regional Genetics Laboratory (WRGL), and utilise the panel along with other techniques to help understand the significance of *JAK2* V617F at low variant allele frequency (VAF) and to characterise a recurrent cytogenetic abnormality, the der(6)t(1;6).

Initially I describe the development of a standardised two-step process for the interpretation of variants detected by gene panels. A classification of pathogenicity is assigned using weighted evidence and then the clinical significance is assessed which informs whether the finding should be reported and how it is interpreted. This protocol has been successfully implemented into the WRGL and used to interpret >1,700 variants to date, with each classification recorded and auditable in accordance with strict quality management requirements for diagnostic laboratories. I undertook an audit to understand how the myeloid panel was being used clinically at an early stage of implementation which demonstrated clear clinical utility of the test, in particular when trying to formalise or exclude a diagnosis or when looking for prognostically significant markers in an already confirmed diagnosis. For cases of suspected myelodysplastic syndrome (MDS), the absence of mutations was a particularly useful piece of evidence for exclusion of neoplastic condition and enabled many patients to avoid an invasive bone marrow procedure.

In the second part of the study the clinical significance of low level *JAK2* V617F in suspected myeloproliferative neoplasm (MPN) referrals was investigated. Using digital droplet PCR I found no evidence that low level (<1% VAF) *JAK2* V617F was enriched in cases referred for MPN testing (prevalence 3% for controls [4/197] vs 2% for cases [20/662]; P=0.62, Fisher's exact test) suggesting that our standard diagnostic 1% VAF cut off is unlikely to be missing many true MPN cases. However, there was a significant enrichment of low level *JAK2* V617F in cases that had tested positive for *CALR* or *MPL* mutations by routine analysis (P=0.018; 7.4% [11/149] and P=0.006; 15% [8/52], respectively), and some cases with low level *JAK2* V617F tested positive for higher VAF driver mutations on myeloid panel analysis. A clinical audit suggested that routine reporting of *JAK2* V617F even at 1-5% VAF was often used as a major factor in the diagnosis of an MPN, but cases with lower level *JAK2* V617F <1% VAF (not reported) were usually not given a diagnosis. Unlike cases with MPN, low level *JAK2* V617F was not associated with the 46/1 haplotype suggesting a role for 46/1 in expansion of *JAK2* V617F mutated clones.

Finally, I used the myeloid panel (n=16), single nucleotide polymorphism (SNP; n=12) array and whole genome sequencing (WGS; n=6) to characterise patient samples (n=16) with CMN and a der(6)t(1;6). Although the breakpoints fell within a gene for 3 cases (*SNX27* and *PDEDIP* at 1q, and *ZSCAN9* at 6p), there was no evidence of a recurrent candidate fusion gene. Indeed, considerable diversity in breakpoint location was observed, with most breaks falling within 7.4 Mb and 5.9 Mb regions on 1q and 6p, respectively suggesting that the important consequence of the der(6)t(1;6) is likely to be gain of 1q, a recognised recurrent change in CMN. Interestingly, copy number variant analysis from the array data identified recurrent deletions at 13q/RB1 in 9/11 cases and 17q11.2/NF1 in 4/11 cases, and variant analysis performed by the myeloid panel and WGS identified *JAK2, CALR* or *MPL* driver mutations in 14/16 (88%) of cases with a range of other mutations typical of advanced MPN or MDS. These data support the hypothesis that the der(6)t(1;6) is a marker of disease progress in MPN, with recurrent genetic features that suggest it may be considered as a discrete subtype.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

Acknowledgements

This project was completed under the supervision of Prof NCP Cross (Laboratory Director, WRGL). The work detailed in this project was undertaken with the help of the following individuals.

Chapter 2: the implementation of a Myeloid Next Generation Sequencing panel into a diagnostic setting

The myeloid panel validation was performed and led by me with help in the experimental design and interpretation of results from Prof NCP Cross (Laboratory Director, WRGL), Dr Andy Chase (Research Scientist, Southampton University), Dr Feng Lin (Research Scientist, Southampton University), Chris Mattocks (Head of IT and Development Services, WRGL, Salisbury), Ahmed Dawoud (Research Scientist, Southampton University), Dr Laura Chiecchio (Lead Clinical Scientist, WRGL). Ahmed Dawoud was also involved in performing experiments used to calculate the limit of detection of this assay.

The somatic variant interpretation protocol was developed over time in a process led by me with significant input from Dr Laura Chiecchio (Lead Clinical Scientist, WRGL), Prof NCP Cross (Laboratory Director, WRGL), Alison Callaway (Principal Clinical Scientist, WRGL) and Kevin Baker (Senior Clinical Scientist, WRGL). Both Alison Callaway and Kevin Baker have extensive experience in utilising the ACMG guidelines for variant interpretation of constitutional variants in a diagnostic setting.

The clinical audit questionnaire was developed in liaison with Prof NCP Cross (Laboratory Director, WRGL) and Dr Sally Killick (Consultant Haematologist, Royal Bournemouth Hospital). Results returned were interpreted by myself, Prof NCP Cross and Dr Sally Killick.

Chapter 3: the significance of low level *JAK2* V617F mutations

A bioinformatic approach of preselection to identify candidate low level (<1%) *JAK2* V617F positive cases was designed in collaboration with Ben Sanders (Senior Clinical Scientist, WRGL). The ddPCR experiments were completed with the help of Katherine Waghorn (Research Scientist, Southampton University) and Sarah Yapp (Research Scientist, Southampton University).

Low level *JAK2* V617F positive samples were obtained from Guy's and St Thomas' Hospital following liaison with Dr Samah Alimam (Consultant Haematologist, Guy's and St Thomas's Hospital).

The ARMS PCR assay was designed and performed with the help of Katherine Waghorn (Research Scientist, University of Southampton).

Myeloid panel technical work was performed by Toby Mellows (Genetic Technologist, WRGL).

Chapter 4: characterisation of the der(6)t(1;6)

External samples were obtained from Prof Cristina Mecucci (University of Perugia), Prof Dr Claudia Haferlach (MLL Laboratory, Munich), Susan Rose (Genetic Technologist, WMRGL, Birmingham) and Joanne Mason (Head of Haemato-oncology, WMRGL, Birmingham).

SNP array analysis was performed at the WMRGL, Birmingham, with analysis completed by Emma Huxley (Principal Clinical Scientist, WMRGL).

Bioinformatic analysis of the whole genome sequencing data was performed by Dr Will Tapper (Research Scientist, University of Southampton).

Myeloid panel technical work was performed by Toby Mellows (Genetic Technologist, WRGL).

Breakpoint sequencing was optimised with the help of Prof NCP Cross and Katherine Waghorn (Research Scientist, University of Southampton).

Abbreviations

aCML	Atypical CML	**LoF**	Loss of function
ACGS	Association for Clinical Genomic Science	**MDS**	Myelodysplastic syndrome
ACMG	American College of Medical Genetics	**MDS/MPN RS-T**	MDS/MPN with ring sideroblasts and thrombocytosis
AML	Acute myeloid leukaemia	**MDT meeting**	Multi-disciplinary team meeting
AMP	Association for Molecular Pathology (AMP)	**MPN**	Myeloproliferative neoplasms
aUPD	Acquired uniparental disomy	**MPN-BP**	MPN- blast phase
ARCH	Age-related clonal haematopoiesis	**MPN-U**	Myeloproliferative neoplasm, unclassifiable
ARMS PCR	Amplification refractory mutation system PCR	**MRD**	Minimal residual disease
CEL	Chronic eosinophilic leukaemia	**MVL**	Managed variant list
CHIP	Clonal haematopoiesis of indeterminate potential	**NGS**	Next generation sequencing
CHOP	Clonal haematopoiesis of oncogenic potential	**OS**	Overall survival
CLL	Chronic lymphocytic leukaemia	**PMF**	Primary myelofibrosis
CML	Chronic myeloid leukaemia	**PV**	Polycythaemia vera
CMML	Chronic myelomonocytic leukaemia	**RC-PCR**	Reverse complement PCR
CMN	Chronic myeloid neoplasia	**RNAseq**	RNA sequencing
CN-LOH	Copy number neutral loss of heterozygosity	**RT-PCR**	Reverse transcriptase PCR
CNL	Chronic neutrophilic leukaemia	**SCT**	Stem cell transplant
CR	Complete remission	**SNP**	Single nucleotide polymorphism
ddPCR	Droplet digital PCR	**SM (-AHN)**	Systemic mastocytosis (with associated haematological neoplasia)
Epo	Erythropoietin	**Tpo**	Thrombopoietin
EpoR	Erythropoietin receptor	**TpoR**	Thrombopoietin receptor
ET	Essential thrombocythaemia	**TSMP**	Illumina® TruSight myeloid panel
FISH	Fluorescence *in situ* hybridisation	**qPCR**	Real-time quantitative PCR
FLT3-ITD	Internal tandem duplications of *FLT3*	**QA**	Quality assurance
GLH	Genomic Laboratory Hub	**VAF**	Variant allele frequency
GoF	Gain of function	**VUS**	Variant of uncertain significance
GWAS	Genome-wide association study	**WES**	Whole exome sequencing
Indel	Insertional deletion	**WGS**	Whole genome sequencing
IPSS (-R)	International Prognostic Scoring System (-Revised)	**WMRGL**	West Midlands Regional Genetics Laboratory
ISO	International Organisation for Standardization	**WRGL**	Wessex Regional Genetics Laboratory, Salisbury
LoB	Limit of blank		
LoD	Limit of detection		

1. Introduction to chronic myeloid neoplasia

Chronic myeloid neoplasms (CMNs) are a heterogeneous group of clonal disorders that arise in lineage-restricted cells or in a multipotent stem cell capable of differentiating into cells of granulocytic, monocytic, megakaryocytic or erythroblastic lineages, collectively referred to as myeloid cells (Figure 1.1).

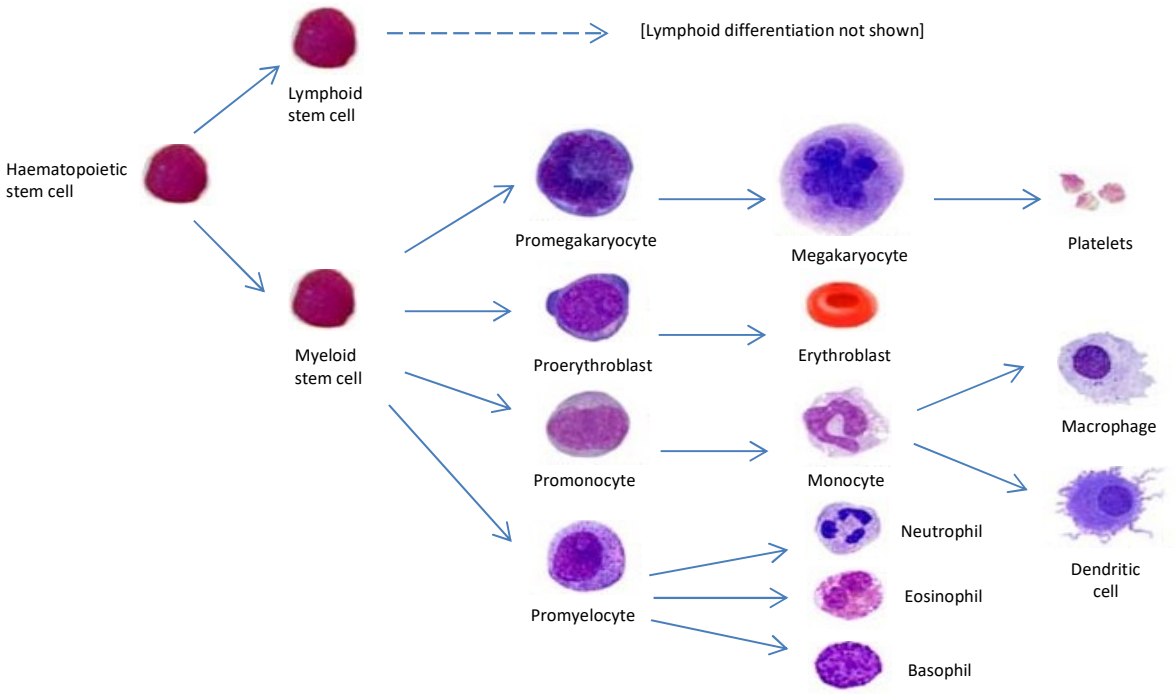


Figure 1.1 Haematopoiesis; showing myeloid differentiation.

Source: adapted from Sarvothaman*, et al*., (2015)

In CMN, cells retain their ability to differentiate into end cells but the pool of proliferating myeloid cells becomes expanded (Shepherd, et al., 2018; Van Egeren, et al., 2021). This group of diseases progress in an insidious fashion, leading in extreme cases to death within weeks or months if left untreated although in some cases the disease is relatively benign for several years. They are clinically and biologically distinct from acute leukaemia, including acute myeloid leukaemia (AML), which is characterised by a defect in cell differentiation, leading to expanding of pool of immature cells (blast cells) with reduced potential to die biologically and a clinically rapid disease onset.

Several subtypes of CMN are recognised, specifically myeloproliferative neoplasms (MPN), myelodysplastic syndromes (MDS), myelodysplastic/myeloproliferative neoplasms

(MDS/MPN) and mastocytosis, as defined by the classification of myeloid neoplasms within the World Health Organization (WHO) classification scheme (Swerdlow, et al., 2016). The WHO provides an internationally recognised system of disease classification which integrates clinical, morphological, immunophenotypic and genetic information into 9 main disease groups (Table 1.1). As will be described in the following pages, there are a number of overlapping haematological and molecular features between the different CMN subgroups- and even when comparing patients with CMN to apparently healthy individuals- which can complicate the diagnosis and prognostication in some patients with confirmed or suspected CMN.

WHO classification of myeloid neoplasms
Myeloproliferative neoplasms (MPN)
Mastocytosis
Myeloid/lymphoid neoplasms with eosinophilia and specific gene rearrangements (MLN-eo)
Myelodysplastic/myeloproliferative neoplasms (MDS/MPN)
Myelodysplastic syndromes (MDS)
Myeloid neoplasms with germline predisposition
Acute myeloid leukaemia and related precursor neoplasms
Blastic plasmacytoid dendritic cell neoplasm
Acute leukaemias of ambiguous lineage

Table 1.1 WHO classification for myeloid neoplasms**. A**dapted from Arber*, et al.,* (2016)

1.1 Myeloproliferative neoplasms

Myeloproliferative neoplasms (MPN) are a rare class of haematological disorders that occur primarily in adults, with the incidence peaking in the fifth and seventh decades of life, although some subtypes have been reported in children (Arber, et al., 2016). The combined annual incidence for all subtypes is 6 per 100,000 individuals (Titmarsh, et al., 2014). MPNs have a shared biology in that stem cell abnormalities transform myeloid progenitor cells resulting in an overproduction of both mature and immature cells in one or more cell types of the myeloid lineage (Figure 1.1). MPN subtypes include polycythaemia vera (PV), essential thrombocythaemia (ET), primary myelofibrosis (PMF), chronic neutrophilic leukaemia (CNL), chronic eosinophilic leukaemia (CEL), and myeloproliferative neoplasm, unclassifiable (MPN-U) (Table 1.2). PV, ET and PMF are often referred to as "classical" MPNs because they are relatively common and were included in the original description of myeloproliferative disorders in the 1950s (Dameshek, 1951). Prior to 2001, MPNs were categorized along with *BCR-ABL1* positive($^{+}$) chronic myeloid leukaemia (CML) under the broad category of chronic myeloproliferative disorders but *BCR-ABL1*$^{+}$ CML has subsequently been defined as a distinct entity and will not be discussed further (Wadleigh & Tefferi, 2010).

WHO classification for MPNs	CML, BCR-ABL1+
	Chronic neutrophilic leukaemia (CNL)
	Polycythaemia vera (PV)
	Primary myelofibrosis (PMF)
	Essential thrombocythaemia (ET)
	Chronic eosinophilic leukaemia (CEL), NOS
	Myeloproliferative neoplasm, unclassifiable (MPN-U)

Table 1.2 WHO classification for MPNs. Adapted from Arber*, et al.,* (2016).

Symptoms may differ between the subtypes but generally include bruising, excessive sweating, fatigue and headaches and thrombotic/ haemorrhagic events which contribute significantly to mortality (Titmarsh, et al., 2014); further detail regarding the clinical symptoms associated with the subtypes PV, ET and PMF are shown in Table 1.3. There is a risk of MPN transforming into acute myeloid leukaemia (where the proportion of immature

blasts in the peripheral blood or bone marrow is ≥20%) termed blast phase MPN (MPN-BP). MPN-BP is reported in 1.5% of patients with ET, 7.0% of patients with PV and 11% of patients with PMF; however, estimates may be confounded by the presentation of age-related *de novo* acute leukaemia and exposure to chemotherapy. Overall, MPN-BP is associated with a poor prognosis (Spivak, 2017).

MPN	Clinical features
PV	• Hypertension or vascular abnormalities caused by increased red blood cell (RBC) mass. • Headache, dizziness, visual disturbances, and paraesthesias • Pruritus • Erythromylalgia • Gout
ET	• >50% asymptomatic at diagnosis • Vascular occlusion or haemorrhage • Transient ischaemic attacks • Ocular migraine • Thrombosis of major arteries • Mild splenomegaly and hepatomegaly
PMF	• Bone marrow fibrosis with progressive clinical course leading to bone marrow and organ failure • Splenomegaly due to extramedullary haematopoiesis • Anaemia, leucocytosis and/or thrombocytosis • Fatigue, weight loss, dyspnoea, night sweats, low-grade fever • Up to 30% asymptomatic at the time of diagnosis

Table 1.3 Clinical features and WHO diagnostic criteria for the MPNs ET, PV and PMF. Adapted from Spivak, (2017) and Swerdlow, *et al*., (2016).

1.2 Myelodysplastic syndromes (MDS)

MDS is a heterogeneous set of stem cell disorders characterised by simultaneous apoptosis and proliferation of haematopoietic cells, resulting in normocellular or hypercellular bone marrow (Table 1.4). At diagnosis, there may be blast cells detectable in the peripheral blood but, as distinct from AML, the blast percentage must be ≤20%; these patients are at an increased risk of developing AML, although the actual risk varies between subtypes (Swerdlow, 2017). Cytopenia in at least one haematopoietic lineage must be present for a diagnosis to be made and this may be difficult to differentiate from other (reactive or non-neoplastic) causes of cytopenia. Clinical symptoms at presentation are usually dependent

upon the severity of the cytopenia and the number of cell lineages affected. Lineage specific dysplasia is often seen. At presentation, the subtypes of MDS are defined by the number of cytopenias, the number of myeloid lineages displaying dysplasia, the presence of ring sideroblasts, and the blast percentages in the bone marrow (Arber, et al., 2016).

WHO classification for MDS	MDS with single lineage dysplasia (MDS –SLD)
	MDS with ring sideroblasts (MDS-RS)
	MDS with multilineage dysplasia (MDS –MLD)
	MDS with excess blasts (MDS-EB)
	MDS with isolated del(5q)
	MDS, unclassifiable (MDS-U)
	Childhood MDS
	Refractory cytopenia of childhood

Table 1.4 WHO classification for MDS. Adapted from Arber*, et al.,* (2016).

1.3. Other chronic myeloid neoplasms

1.3.1 MDS/MPN

The MDS/MPN group of diseases have overlapping features of both MDS and MPN (Patnaik & Lasho, 2020). Subtypes include chronic myelomonocytic leukaemia (CMML); *BCR-ABL1* negative, atypical CML (aCML), juvenile myelomonocytic leukaemia (JMML), MDS/MPN with ring sideroblasts and thrombocytosis (MDS/MPN RS-T) and MDS/MPN unclassifiable. They usually present with hypercellularity in the bone marrow due to overproductions of one or myeloid lineage and patients can present simultaneously with cytopenia in at least one of the remaining lineages; some degree of morphological and/or functional dysplasia is also often seen.

1.3.2 Mastocytosis

Mastocytosis can vary in severity and presentation, ranging from indolent cutaneous presentation to overt, aggressive systemic disease and as such the symptoms can range from mild to life-threatening. Mastocytosis patients often have skin manifestations such as pruritus or urticaria and in advanced stages of disease there is usually organomegaly and organ impairment due to mast cell infiltration. Patients' haematological profile commonly includes eosinophilia and can also show anaemia, neutropenia and thrombocytopenia (Arber, et al., 2016).

1.3.3 MPN, unclassifiable (MPN-U)

MPN, unclassifiable (MPN-U) is designated to those patients with definite features of an MPN by clinical, morphological, genetic and/or immunophenotypic parameters but who do not fit the diagnostic criteria of a specific MPN category (Arber, et al., 2016). It is thought that this group accounts for 10-15% of all MPNs. MPN-U may also be applied where the disease exhibits features of more than one category of MPN. Of note, one of the main groups of patients that fall into this category are those that present with early stage disease e.g. pre-polycythaemic presentation of PV (Arber, et al., 2016) or early stage ET (Gisslinger, et al., 2016). Clinical phenotype and haematological features overlap with other MPNs.

1.4 Genetic aberrations in chronic myeloid neoplasia

The genomic landscape of CMN has been extensively investigated and a number of recurrent cytogenetic abnormalities (reciprocal translocations, deletions, copy number gains) and gene mutations (collectively referred to as genomic aberrations) have been associated with this group of diseases (Bejar, et al., 2011; Papaemmanuil, et al., 2013; Grinfeld, et al., 2018). As a result, specific genetic studies play a major role in the evaluation of patients with CMN in regard to confirming the presence of a clonal disorder (thus potentially excluding a reactive cause of the patient's phenotype), defining specific subtypes of disease, providing robust prognostic information and/or indicating the use of targeting therapies (National Comprehensive Cancer Network, 2019a; National Comprehensive Cancer Network, 2019b; Guglielmelli, et al., 2018; Papaemmanuil, et al., 2013; Welch, et al.,

2016). A wide range of technologies are currently employed in the genetic diagnostic work up of patients e.g. karyotyping by visual analysis of metaphase chromosomes or by genome-wide single nucleotide polymorphism (SNP) arrays, targeted sequencing for recurrent mutations and wider sequencing of a panel of genes by next generation sequencing (NGS). The technological landscape is changing rapidly and genomic approaches such as whole genome sequencing (WGS) and RNA sequencing (RNAseq) are expected to have a major impact in the near future.

Genes affected in CMN can generally be split into four main pathways with respect to cellular function: (i) transcription and cell cycle regulators, (ii) spliceosome components, (iii) cell signalling and (iv) epigenetic modifiers. A number of genetic aberrations are considered general markers of myeloid neoplasia in that they occur in a wide range of myeloid (but generally not lymphoid) disorders, such as mutations in *ASXL1* which have been reported in MDS, MDS/MPN, MPN and SM, and an extra copy of chromosome 8 (trisomy 8), gain of the long arm of chromosome 1 or deletions within the long arm of one chromosome 20. There are also a number of genetic aberrations that occur at different frequencies in specific sub-types of CMN; for example, *JAK2* V617F mutations can occur in PV, ET and MF (although most frequent in PV) whereas *JAK2* exon 12 mutations are detected in about 1% of PV but are not seen in ET or MF (Zoi & Cross, 2017). Furthermore, whilst no mutation or combination of mutations defines a CMN subtype (apart from *BCR-ABL1* and CML), the detection of some genetic aberrations can aid in the diagnosis of a disorder and/or direct further investigations, such as the combination of mutations in *SF3B1* and *JAK2,* which is strongly associated with the MDS/MPN refractory anaemia with ringed sideroblasts and thrombocytosis (RARS-T) (Jeromin, et al., 2013).

Although the identification of clonality by molecular genetic and cytogenetic testing can help in the diagnosis of CMN, it has been reported that up to 15% of MPN cases (Grinfeld, et al., 2018), 10-20% of established MDS cases (Haferlach, et al., 2014; Papaemmanuil, et al., 2013) and 5-7% of CMML cases (Itzykson, et al., 2013; Elena, et al., 2016) do not have detectable variants by sequencing analysis with panels of genes known to be mutated in myeloid neoplasia. In addition, the majority of these patients show a normal karyotype by cytogenetic analysis. Furthermore, improvements in the ability to perform large scale sequencing to detect variants at low allelic burden have revealed that clonally expanded somatic mutations in genes also known to be frequently mutated in myeloid neoplasia can

occur in the blood cells of apparently healthy individuals and these become more frequent with age (discussed further below), thus making the diagnosis of CMN in some patients particularly challenging.

1.4.1 Genetic aberrations in MPN

A number of genes are recurrently mutated in patients with MPN: over 90% of individuals carry at least one pathogenic mutation in *JAK2* (specifically *JAK2* V617F or *JAK2* exon 12 mutations), *CALR* exon 9 (+1 frameshift mutations) or *MPL* exon 10 (most commonly affecting p.Trp515) (Spivak, 2017) and as previously mentioned, the frequency of these mutations varies between subtypes (Figure 1.2). Mutations in these three genes are referred to collectively as 'MPN phenotypic driver mutations' whereas other mutations are referred to as 'initiation or landscaping mutations' or 'disease-modifying mutations' depending on whether they tend to be early events or late events associated with disease progression (Table 1.5). The reason why the same mutations can result in different clinical phenotypes is thought to be multifactorial, with constitutional genetic background, the phenotypic driver mutation burden, the presence of additional mutations and the order in which they are acquired considered significant influencers (Zoi & Cross, 2017). Given their frequency, the mutation status of these genes has been incorporated into the WHO diagnostic classification system for MPNs, summarised in Figure 1.2, and are therefore well established tests performed routinely in diagnostic genetics laboratories (Gong, et al., 2013; Arber, et al., 2016).

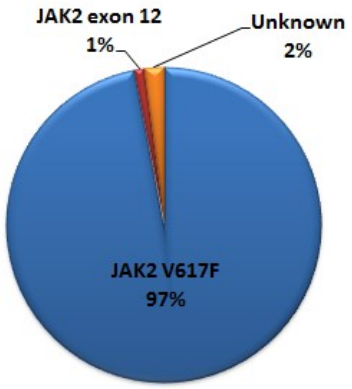
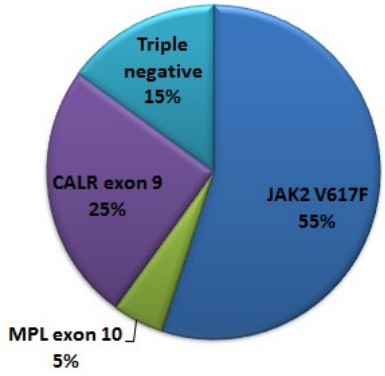
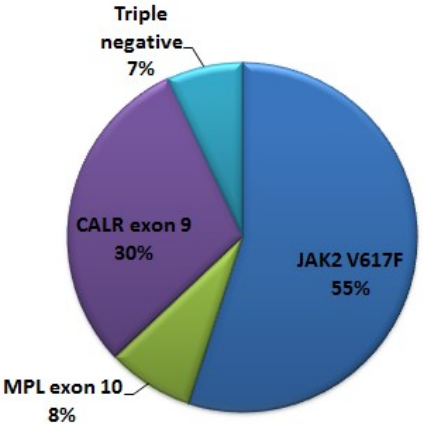
MPN	Summary of WHO diagnostic criteria	Phenotypic driver mutations
PV	A diagnosis requires all 3 major criteria or the first 2 major plus the minor **Major criteria:** (1) Elevated haemoglobin or increased red blood cell mass Bone marrow biopsy showing age-adjusted hypercellularity with trilineage growth. (2) Presence of *JAK2* V617F or *JAK2* exon 12 mutation **Minor criterion** Subnormal serum erythropoietin level	
ET	A diagnosis requires that either all or the first 3 major criteria plus the minor criteria are met **Major criteria** (1) Elevated platelet count. (2) Bone marrow biopsy showing proliferation mainly of the megakaryocytic lineage, with increased numbers of enlarged, mature megakaryocytes with hyperlobulated nuclei. (3) WHO criteria for ET, PV, *BCR-ABL1*-positive CML, MDS, or other myeloid neoplasm are not met. (4) *JAK2*, *CALR* or *MPL* mutation **Minor criterion** Presence of a clonal marker or absence of evidence of reactive thrombocytosis	
PMF	A diagnosis requires all 3 major criteria and at least 1 minor criterion are met **Major criteria** (1) Megakaryocytic proliferation and atypia, accompanied by reticulin and/collagen fibrosis grades 2 or 3 (2) WHO criteria for ET, PV, *BCR-ABL1*-positive CML, MDS, or other myeloid neoplasm are not met (3) *JAK2*, *CALR* or *MPL* mutation or presence of a clonal marker or absence of reactive myelofibrosis **Minor criteria** Presence of at least one of the following, confirmed in 2 consecutive determinations: unexplained anaemia, leukocytosis ≥11x10^{9}/L, palpable splenomegaly, increased lactate dehydrogenase, leukoerythroblastosis	

Figure 1.2 Summary of WHO diagnostic criteria and phenotypic driver mutations associated with the MPNs: PV, ET and PMF. Source: Adapted from Arber, et al., (2016) and Zoi & Cross (2017).

MPN patients that do not carry canonical mutations in *JAK2* V617F, *JAK2* exon 12, *CALR* and *MPL* are termed "triple-negative MPNs" and may have mutations in other genes which are widely mutated in MDS and MDS/MPN (Table 1.5). Interestingly, it has been reported that up to 19% of triple-negative MPN cases actually have a non-canonical mutation in *JAK2* (i.e. not the *JAK2* V617F or a *JAK2* exon 12 mutation) or *MPL* (i.e. not affecting p.Trp515 or p.Ser505) (Milosevic, et al., 2016).

Genetic aberration			**Mutation frequency in MPN (%)**			**Mutation frequency in MDS (%)**	**Mutation frequency in CMML (%)**
Gene	**Mutation location**	**Mutation effect**	**PV**	**ET**	**PMF**		
Phenotypic driver mutations							
JAK2	Exon 14	JAK/Stat signalling	95-97	60	60	3-4	1-10
JAK2	Exon 12	JAK/Stat signalling	1-2	0	Rare		
MPL	Exon 10	JAK/Stat signalling	0	3-5	5-10		6
CALR	Exon 9	JAK/Stat signalling	0	25	30	8	3
Initiation or landscaping mutations							
TET2	All coding regions	Epigenetic regulation	10-20	5	10-20	13-37	22-61
DNMT3A	Exons 7-23	Epigenetic regulation	5-10	1-5	8-12	3-13	2-13
Disease-modifying mutations							
IDH1/2	Exon 4/ exon 4	Epigenetic regulation	1-2	1-2	5-6	4-12	1-10
EZH2	All coding regions	Epigenetic regulation	1-2	1-2	7-10	3-11	6-13
ASXL1	Exon 12	Epigenetic regulation	2	2-5	10-35	5-46	22-60
SF3B1	Exons 12-16	mRNA splicing	2	2	5	10-33	4-10
SRSF2	Exon 1	mRNA splicing	Rare	Rare	5-17	12-33	28-52
U2AF1	Exons 2-7	mRNA splicing	Rare	Rare	2-8	5-17	5-15
ZRSR2	All coding regions	mRNA splicing	Rare	Rare	1	3-11	8-10
RUNX1	All coding regions	Transcription	Not known	Not known	Not known	8-20	7-37
IKZF1	All coding regions	Transcription	Rare	Rare	Rare		
TP53	All coding regions	DNA repair	Rare	Rare	Rare	5-18	<1-4
CUX1	All coding regions	Cell cycle	Not known	Not known	Not known		
CBL	Exons 8-9	JAK/Stat signalling	Rare	Rare	5-10	2-8	10-22
SH2B3	Exon 2	JAK/Stat signalling	1-2	3-6	3-6		

Table 1.5 Genes recurrently mutated in patients with MPN and MDS. Adapted from Zoi and Cross (2017) and McClure, et al., (2018).

1.4.2 Genetic aberrations in MDS

Diagnosing MDS can be challenging as the hallmark features of the disease, namely cytopenia and dysplasia, are known to occur in normal individuals due to reactive or non-neoplastic causes. This is further complicated by the observed positive association of

increased frequency of dysplasia in individuals with non-neoplastic causes of cytopenia (Parmentier, et al., 2012). Furthermore, whilst the classification of dysplasia has been well established, the identification and application of criteria has been reported to lack complete reproducibility even between experienced haematopathologists (Font, et al., 2013). Resultantly, genetic diagnostic testing has become an established part of the diagnostic work up of this group of disorders and is now incorporated into the WHO classification system for MDS (Arber, et al., 2016) and cytogenetic results have been integrated into the MDS International Prognostic Scoring System (IPSS) and revised-IPSS (Greenberg, et al., 2012).

Some of the most common cytogenetic findings are deletions within the long arm of chromosome 5 [del(5q)] and 7 [del(7q)], monosomy for chromosome 7, trisomy 8, loss of the Y chromosome and complex karyotypes defined by the presence of 3 or more independent abnormalities. The prognostic significance of these findings is well established (Table 1.6) and the need for cytogenetic testing has been formally incorporated into patient management pathways (National Comprehensive Cancer Network, 2019). Of note, the majority of cytogenetic abnormalities detected in MDS are copy number changes and as such can be tested for by SNP arrays as a replacement technology for G-banding analysis (Mikhail, et al., 2019); SNP arrays cannot detect the balanced translocations that characterise a small proportion of *de novo* and therapy related MDS. Gene mutations are yet to be included into the IPSS-R for MDS.

Chromosomal abnormality		Frequency (%)	Prognosis according to IPSS-R	
Unbalanced	Trisomy 8	10	Intermediate	
	Monosomy 7/deletion (del) 7q	10	del(7q): Intermediate	-7: Poor
	del(5q)	10	Good	
	del(20q)	5-8	Good	
	Loss of Y chromosome	5	Very good	
	Isochromosome 17q	3-5	Intermediate	
	Monosomy 13 or del(13q)	3	Intermediate	
	del(11q)	3	Very good	
	Abnormalities of 12p	3	Good	
	Deletions of 9q	1-2	Intermediate	
	Isodicentric Xp	1-2	Intermediate	
	Complex karyotype (≥3 abnormalities)	20-30	Poor/very poor	
Balanced	t(1;3)(p36.3;q21.2)	1	Intermediate	
	t(2;11)(p21;q23.3)	1	Intermediate	
	inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2)	1	Poor	
	t(6;9)(p23;q34.1)	1	Intermediate	
	Any other single or double independent clones		Intermediate	

Table 1.6 Commonly detected abnormalities in MDS and the cytogenetic prognostic subgroup associated with them according to the MDS IPSS-R. Adapted from Greenberg, *et al*., (2012) and Rossi, *et al.,* (2000).

In more recent years, there has been an increase in the number of publications demonstrating the clinical utility of mutational analysis in the diagnosis and stratification of patients with MDS. In 2011, Bejar *et al.* investigated the clinical effect of point mutations in MDS through a combination of genomic analysis techniques including next generation sequencing (7 genes), sanger sequencing (6 genes) and mass spectrometry-based genotyping (111 genes) on the bone marrow from 439 patients with MDS. They reported that at least 51% of all patients had at least 1 mutation, including 52% of patients with normal cytogenetics, and a genotype-phenotype correlation associated with some mutations, namely severe neutropenia and increased blast cell proportion in patients with *RUNX1*, *TP53* and *NRAS* mutations and poor overall survival in patients with *TP53*, *EZH2*, *ETV6*, *RUNX1* and *ASXL1* mutations (Bejar, et al., 2011). Subsequently, two major studies

were published describing the mutational landscape in MDS following more wide scale genomic analyses of larger cohorts of MDS patients (Papaemmanuil, et al., 2013; Haferlach, et al., 2014). Papaemmanuil et al. (2013) sequenced 111 genes in 738 patients with MDS (or closely related neoplasms including MDS/MPN). They reported that 78% of the patients tested has 1 or more mutation with >10% patients harbouring a mutation in *SF3B1*, *TET2*, *SRSF2* and *ASXL1,* with a further 3 genes carrying driver mutations. This group also demonstrated that the number of mutations is associated with clinical outcomes, with leukaemia-free survival (LFS) deteriorating as the number of driver mutations increased, with patients with $\geq$6 mutations having a particularly poor outcome. Furthermore, of those patients with follow up data available (n=595), 24 genes were mutated in >5 patients and of these, 8 were associated with significantly worse LFS and 1 (*SF3B1*) was associated with a better LFS. A recent large multicentre study has also shown that multi-hit *TP53* mutations are associated with a complex karyotype and predict death and leukaemic transformation independent of the revised-IPSS score (Bernard, et al., 2020). Of note, an *SF3B1* variant in the context of >5% ring sideroblasts now defines a sub-category of MDS, MDS with ring sideroblasts (Arber, et al., 2016).

Haferlach, et al (2014) performed targeted deep sequencing on 104 genes and array-based genomic hybridisation in 944 patients with MDS. This analysis demonstrated that 89.5% of patients had at least one mutation (range 0-12 per person) and 47 genes were demonstrated to be significantly mutated; most commonly *ASXL1, SRSF2, DNMT3A* and *RUNX1*, which were mutated in >10% of patients. This group also performed retrospective analysis on the clinical outcomes in 875 of these patients, and proposed a novel prognostic model ("Model-1") including clinical and genetic variables to stratify patients into 4 risk groups based upon 3 –year survival (shown in brackets): low (95.2%), intermediate (69.3%), high (32.8%) and very high (5.3%); later adapted into "Model-2" which includes a gene only model of predicting outcome based upon the mutation status of 14 genes.

Studies also show that the types of genes mutated in MDS fit into discrete functional pathways, including RNA splicing, DNA methylation, transcription, chromatin modification, signal transduction and DNA repair (Papaemmanuil, et al., 2013; Haferlach, et al., 2014). Papaemmanuil, et al., (2013) showed that progression of MDS resulted in emergence of new clones with additional mutations, with some genes being significantly enriched in secondary AML when compared with high-risk MDS, such as *FLT3, NPM1* and *NRAS.* A separate group have also shown that the presence of a mutation in *SRSF2, SF3B1, U2AF1,*

ZRSR2, ASXL1, EZH2, BCOR and/or *STAG2* in a patient with AML is suggestive of secondary AML, i.e. AML that has evolved from a previously undetected MDS or MDS/MPN (Lindsley, et al., 2015) rather than AML occuring *de novo* or following leukemogenic therapy (therapy-related AML, t-AML).

1.4.3 Genetic aberrations in other CMN

1.4.3.1 MDS/MPN

The majority of established cases of MDS/MPN have a normal karyotype but mutations are frequent, especially within non-kinase genes, such as *TET2* and *ASXL1* which encode epigenetic regulators (Mughal, et al., 2015). Although these mutations are not specific to this group of disease and cannot therefore be used to define a specific diagnosis, studies have shown clear differences in the mutation frequencies of individual genes within the different sub-types of MDS/MPN and therefore sequencing results can be used to support a clinical diagnosis, such as the presence of *SF3B1* and *JAK2* mutations in the context of MDS/MPN RS-T (Meggendorfer, et al., 2018; Patnaik & Tefferi, 2019). Furthermore, mutation status can be useful in determining prognosis for patients with CMML, as demonstrated in the prognostic scoring systems developed by Itzykson, et al., (2013) and Elena, et al., (2016).

1.4.3.2 Mastocytosis

The diagnosis of mastocytosis is usually dependent upon the morphological identification of clusters/aggregates of mast cells within biopsies; however, diffuse patterns of infiltration and the exclusion of reactive mast cell hyperplasia can be difficult. As such, the identification of the activating *KIT* D816V mutation, which is found in almost all cases of adult systemic mastocytosis (SM) along with other rarer *KIT* mutations, has become incorporated into the patient pathway to aid in the diagnosis of this disease. More in depth sequencing studies have also identified a number of mutations in genes associated with CMN in patients with SM and some of these have been reported as independent adverse prognostic factors for overall survival (OS) (Munoz-Gonzalez, et al., 2019; Schwaab, et al., 2013; Munoz-Gonzalez, et al., 2018). In the largest study to date, Munoz-Gonzalez, et al., (2019) showed that patients with indolent SM carrying *ASXL1*, *RUNX1*, and/or *DNMT3A*

mutations with a variant allele frequency (VAF) ≥ 30%, exhibited significantly shortened (P < .001) progression-free survival (PFS) and overall survival (OS). Many cases with additional mutations have SM in conjunction with an additional haematological neoplasm (SM-AHN), most commonly CMML or other CMN and in advanced SM, mutations in *SRSF2*, *ASXL1* and *RUNX1* were associated with an adverse prognosis (Jawhar, et al., 2016).

1.4.4 Hereditary predisposition to CMN

The majority of cases of CMN occur sporadically and are late onset, but familial clustering is known to occur. In MPN, it is reported that 7% involve a familial predisposition with a demonstrable relative risk of 5-7 in first degree-relatives of affected individuals (Landgren, et al., 2008; Rumi, et al., 2014) and there is evidence that common genetic variation increases the probability that MPNs will develop. Specifically, the constitutional *JAK2* haplotype, designated 46/1 or GGCC, is strongly associated with the development of both homozygous and heterozygous *JAK2* V617F-positive PV, ET or PMF (and to a lesser degree *MPL* W515-positive MPN) (Jones, et al., 2009). Moreover, MPN development has been associated with the presence of other single-nucleotide variants (SNVs) identified in *CALR*, *MPL* and other genes and it is thought that these predisposition factors may explain the co-occurrence of *JAK2, CALR* or *MPL* mutations within independent stem-cell clones within the same individual (Spivak, 2017).

In the 2016 revised WHO classification of tumours of haematopoietic and lymphoid tissues a new entity was described: 'myeloid neoplasms with germline predisposition'. This group contains a number of syndromic and non-syndromic causes of germline predisposition to MDS, MPN, MDS/MPN, AML and less frequently lymphoid neoplasia, involving genes such as *DDX41*, *CEBPA*, *RUNX1*, *AKRD26*, *ETV6* and *GATA2* (Arber, et al., 2016). Recognition of variants found in cancer-associated genes of germline in origin is imperative for informed clinical management of the patient and their relatives. Pertinent cancer susceptibility gene panels were constructed for common tumour types within the 100,000 Genomes Project; at the time of writing, the haematological malignancies cancer susceptibility panel (v2.5) contained 97 genes, of which 85 have a high level of evidence associated (Genomics England, 2020).

1.5 Somatic mutations in normal individuals

Evidence of clonal expansion of haematopoietic cells with aging was first demonstrated by X-inactivation studies in the 1970s (Fialkow, 1972). Since that time, the advent of NGS technologies has allowed for large numbers of patients to be screened for the presence of somatic mutations and, consequently, a wealth of information is being gathered about the presence of cancer-associated mutations in the peripheral blood of "normal" individuals that occur more frequently with age. It is now apparent that the acquisition of mutations within genes known to be associated with CMN is a general process of ageing. In two of the early studies published that investigated this topic, Jaiswal, *et al*. (2014) and Genovese, *et al*. (2014) performed whole exome sequencing (WES) or targeted sequencing on DNA from the peripheral blood of healthy individuals (i.e. individuals unselected for their haematological phenotypes). Jaiswal, *et al*. (2014) studied 17,182 individuals (median age 58 years; range from 19 to 108) and performed targeted analysis on 160 cancer-related genes; Genovese, *et al*. (2014) studied 12,380 Swedish individuals (median age 55 years; range 19 to 93) by WES. Both studies identified a large number of mutations dispersed across the exome consistent with the presence of clonal haematopoiesis of indeterminate potential, CHIP (also referred to as age-related clonal haematopoiesis [ARCH]) in these patients, including low-level, putative somatic mutations, the most common of which occurred within the genes *DNMT3A, TET2, ASXL1, TP53, SF3B1* and *JAK2* which have all been implicated in CMN.

Jaiswal, *et al*. (2014) and Genovese, *et al*. (2014) reported that CHIP was infrequent in persons less than 50 years old but the prevalence increased with age, to at least 10% of individuals greater than 65 years and 18% of individuals greater than 90 years. There is a clear association with an increased risk of haematological cancer and death in these individuals (Nielsen, et al., 2011; Genovese, et al., 2014; Jaiswal, et al., 2014), with approximately 1-2% per annum of individuals with CHIP going on to develop a haematological malignancy (Steensma, et al., 2015) and the excess of deaths at least in part attributable to the development of cardiovascular disease in these patients (Jaiswal, et al., 2017; Young, et al., 2019). The presence of CHIP-like clones have also been identified in the bone marrow of patients originally diagnosed with AML but who have achieved complete remission (Shlush, et al., 2014; Wong, et al., 2016; Jongen-Lavrencic, et al., 2018) indicating that the mutated clone is biologically distinct from the abnormal AML clone, possibly persisting due to slow proliferation rate and/or resistance to debulking agents used to treat

patients with AML and contributing to the relapse rate in these patients (Valent, et al., 2019).

Due to the error rate of NGS technologies utilised in these studies, only mutations down to a burden of approximately 1-2% can be detected. More recently, the use of error-corrected NGS which uses single molecule tagging with unique molecular identifiers to enable the detection of mutations down to an allelic burden of 0.03% has indicated that CHIP can be detected in 95% (n=19/20) of individuals 50-60 years old (Young, et al., 2016) and most commonly occur in the genes *DNMT3A* and *TET2.* In this study, the mutations were stable over time and presented in multiple haematopoietic compartments, consistent with the hypothesis that they represent mutations in long-lived stem/progenitor cells. These data raise questions regarding the order of mutation acquisition and cooperating events needed for neoplastic transformation (Genovese, et al., 2014).

One of the most well studied mutations in apparently normal individuals is the *JAK2* V617F mutation. In addition to presenting their own data on the prevalence of the *JAK2* V617F mutation in apparently normal individuals with incidentally detected erythrocytosis, Wouters, et al., (2020) performed an extensive literature search and summarised the published literature that reports the prevalence of *JAK2* V617F in individuals from the general population firstly in individuals with no abnormal haematological parameters (Levine, et al., 2005; Xu, et al., 2007; Rapado, et al., 2008; Martinaud, et al., 2010; Nielsen, et al., 2011; Weinberg, et al., 2012; Nielsen, et al., 2013; Genovese, et al., 2014; Jaiswal, et al., 2014; McKerrell, et al., 2015; Acuna-Hidalgo, et al., 2017; Buscarlet, et al., 2017; Zink, et al., 2017; Cordua, et al., 2019; Cook, et al., 2019; and the control samples from: Abelson, et al., 2018; Desai, et al., 2018) and secondly in individuals with abnormal measures of haemoglobin and/or haematocrit (Kralovics, et al., 2005; James, et al., 2005; Tefferi, et al., 2005; McClure, et al., 2006; Percy, et al., 2006; Bianchi, et al., 2007; Tagariello, et al., 2009; Magnussen, et al., 2013; Kamaruzzaman, et al., 2018). These studies used a number of different techniques for *JAK2* V617F detection [specifically, quantitative PCR (qPCR), targeted sequencing, WES, WGS, digital droplet PCR (ddPCR), amplification refractory mutation system (ARMS) PCR and MALDI-TOF mass spectrometry] and therefore the reported limit of detection (LoD; variable definition of LoD were used in different studies) ranged from 0.009% VAF by ddPCR to 10% VAF by WES/WGS. Nevertheless, these data indicate that the frequency of *JAK2* V617F in individuals with no abnormal haematological

parameters is between 0% and 3.1% (mean 0.73%; cohort sizes ranging from 142 to 49,488 individuals). Of note, in one study the prevalence was as high as 35.8% (n=29) but these individuals were cigarette smokers (n=81) admitted to a clinical ward in Israel (Weinberg, et al., 2012) suggesting that environmental factors can have a significant impact on the acquisition of this mutation. In those publications in which the individuals studied showed high haematocrit and/or haemoglobin, the prevalence of the *JAK2* V617F mutation was reported to be between 0% and 4.8% VAF (mean 1.1%; cohort sizes of 11 to 103).

Overall, the three largest cohorts studied for the prevalence of *JAK2* V617F in individuals with no abnormal haematological parameters were reported by Nielsen, et al., (2013) [n=49,488], Cordua, et al., (2019) [N=19,958] and Jaiswal, et al., (2014) [n=17,182]; in these studies, the prevalence was reported to be 0.1% (LoD 0.8% VAF), 3.1% (LoD 0.009% VAF) and 0.18% (LoD 3.5% VAF), respectively. This equates to at least a ten- fold higher prevalence of *JAK2* V617F mutation in the general population versus the estimated prevalence of MPNs (Titmarsh, et al., 2014). In "normal" individuals that carry this mutation, a 44-fold risk of haematological cancer and 221-fold increased risk of MPN development was reported following a study of approximately 10,000 individuals aged 20-95 years over a period of 17 years (Nielson, *et al*., 2011). When discussing the development of CMN, Valent *et al.* (2019) propose a differentiation between CHIP-type mutations and mutations that occur in healthy individuals but are more likely to be associated with the development of an overt neoplasm; they term this phenomenon clonal haematopoiesis of oncogenic potential (CHOP). They cite the observation that low-level and long-standing *JAK2* V617F clones have been detected in patients with cardiovascular events and the demonstrable increased risk of these patients developing an MPN versus both "normal" individuals and individuals with an alternate CHIP clone (Passamonti, et al., 2007; De Stefano, et al., 2005).

The occurrence of CHIP has also been investigated in individuals with a history of cancer that was not haematological. Xie, *et al*. (2014) performed WES on the peripheral blood of 2,728 individuals (median age 60 years; range 10 to 90) years known to have had one of 11 different cancers and no radiation or chemotherapy treatment. A total of 556 cancer-associated genes were interrogated for their mutation status and a number of low level variants (including *JAK2* V617F) were validated using deep sequencing. Consistent with the results from the studies discussed above, Xie, *et al*. (2014) identified CHIP with or without the concurrent presence of leukaemia in approximately 2% of individuals, including 64

mutations in 19 genes with VAFs greater than 10% and an additional 14 mutations detected at 2-10% VAF. Of these, 4 were the *JAK2* V617F mutation and these occurred down to 2% VAF. As the study participants had on this occasion been de-identified and de-coded, it was not possible to associate the presence of clonal haematopoiesis with the development of malignant haematological disease in the relevant individuals.

Within the population study reported by Nielsen, *et al*. (2014) [50,000 individuals from the general population in Denmark], the majority of *JAK2* V617F mutation positive individuals identified had a mutation burden <10% and, of note, the mutation burden appeared stable at the time of re-examination up to 9 years later. Furthermore, of the 8 *JAK2* V617F-positive individuals that showed no MPN, 7 had a mutation burden of <5% with one individual having evidence of splenomegaly and a mutation burden of 5.7%. Within this study, there was no evidence of a *JAK2* V617F mutation with a burden below 2% in any individual with a MPN at the first examination or follow-up examination. In a study of 4,067 individuals (including blood donors aged 17 to 70 years and unselected individuals aged 60 to 98 years), McKerrell, *et al*. (2015) was able to reliably detect the presence of mutated clones down to 0.8% VAF using a combination of NGS and barcoded multiplex PCR of mutational hotspots. *JAK2* V617F was detected in 25 out of 4067 individuals (0.6%) aged 34 to 90 years with the VAFs ranging from 0.83% to 25.94%. Cordua, et al., (2019) studied 19,958 individuals from the Danish population using a ddPCR assay with a sensitivity of 0.009% VAF. They found that age was positively associated not only with the prevalence of a *JAK2* V617F but also the VAF. The mean VAF was 2.1% overall, but variants >5% VAF only occurred in individuals greater than 80 years old and the mean VAF decreased in an apparently linear fashion to <1% in individuals 20 to 29 years old. Within this group, 16 individuals (0.08%) were subsequently identified as having MPN, but in those without an MPN diagnosis, the haematological parameters of *JAK2* V617F-positive individuals were investigated and results showed that when the VAF was ≥1%, individuals had significantly higher haematocrit, leukocyte, neutrophil, and thrombocyte counts and lower cholesterol levels when compared to those with a VAF <1%. In addition, leukocyte, neutrophil and thrombocyte counts were significantly higher in *JAK2* V617F-positive individuals generally *versus* positive controls. These findings suggest that the *JAK2* V617F mutations can occur at low level (<1% VAF) and that mutation burden is associated with MPN development, going from a continuum of no disease to symptomatic MPN with a *possible* intermediate stage of disease progression, as proposed by Nielsen, et al., (2014). Of note, Cordua, et al., (2019) also investigated the

presence of *CALR* mutations in the general population; these data indicate that *CALR* mutations are ~5-times less frequent but have a higher mean VAF and carriers are ~3-times more likely to have MPN.

In a study that investigated inherited factors that might affect the potential of individuals to develop CHIP and/or MPN, Hinds*, et al*. (2016) performed a genome-wide association study (GWAS) on 726 individuals with MPN and 252,637 population controls. In addition to the *JAK2* 46/1 haplotype, which is known to predispose to *JAK2* V617F (Jones, et al., 2009), the results indicated that germline variants in *TERT, SH2B3, TET2, ATM, CHEK2, PINT* and *GFI1B* are associated with *JAK2* V617F positive clonal haematopoiesis and the development of MPN independent of *JAK2* status suggesting that the same variants predispose to CHIP and/or MPN development. In addition, McKerrell, *et al*. (2015) comment that whilst *JAK2* V617F (and *DNMT3A*) mutations become more common with age, they were also found in younger individuals and this is consistent with the model of cumulative stochastic acquisition associated with aging. This is contrasted with spliceosome gene mutations that are found exclusively in individuals over 70 years old concordant with the rise in incidence of patients with MDS who carry these mutations (Haferlach, et al., 2014; Papaemmanuil, et al., 2013). Overall, the excess prevalence of *JAK2* V617F clonal haematopoiesis among "normal" individuals is likely to represent not only individuals with undiagnosed MPN but also those at risk of developing MPN and even individuals who will never go on to develop a haematologic disorder (Hinds, et al., 2016).

1.6 Methods for genomic analysis for patients with CMN

1.6.1 General overview of laboratory methods

As discussed in the preceding sections, laboratory testing for genetic aberrations is an integral part of the clinical workup for the diagnosis of CMN and can provide prognostic and therapeutic information that is crucial for patient management. Testing strategies currently applied in diagnostic laboratories have developed over time as the knowledge of genetic abnormalities in these disorders increases and the technological capabilities advance. At the time of writing, there is currently no single, economically feasible test that can be applied to haemato-oncology samples in a diagnostic laboratory setting for patients with confirmed or suspected CMN that can reliably identify and characterise all of the genetic aberrations

within this genetically diverse category of diseases. However, new technologies are being developed capable of simultaneously detecting structural changes, copy number changes and mutations, which will become increasingly feasible to apply in a diagnostic setting in the future. Until that time, diagnostic laboratories must continue to apply cytogenetic and/or molecular tests as required and combine these results to provide a comprehensive assessment of the genomic aberrations or clinical utility for patients with CMN.

G-banding analysis remains the gold standard for whole genome analysis to identify gross copy number changes (including whole chromosome gains/losses) and balanced structural rearrangements with a resolution of ~5-10 Mb depending on the quality of the metaphase cells available. Fluorescence *in situ* hybridisation (FISH) can also be used in isolation to identify locus specific copy number changes or rearrangements for those CMN with class defining lesions or as an adjunct to G-banding analysis to characterise abnormalities detected, compensate when there are insufficient dividing cells, or to identify sub-microscopic abnormalities/rearrangements (Rack, et al., 2019). However, testing strategies are evolving and a number of laboratories have replaced or are in the process of replacing G-banding analysis for copy number changes with SNP arrays [reviewed extensively by LaFramboise (2009)], which has a higher resolution than G-banding and the capability to detect copy number neutral loss of heterozygosity (CN-LOH). SNP array also has the benefit to detect ploidy, copy number changes and CN-LOH in the absence of viable cell division, which can be a problem to CMN in patients with a repressed BM, such as patients with myelofibrosis or hypoplastic MDS. SNP arrays are not able to detect structural abnormalities and must therefore be supplemented for additional techniques if exclusion/detection of rearrangements is required.

Molecular genetic analysis can include a number of different techniques, ranging from sequencing specific genomic regions to detect driver mutations to pan-genomic NGS mutation screens using targeted panels (commercial or bespoke) of relevant genes, WES or WGS. Sanger sequencing is another long standing method for targeted sequencing of specific regions of interest of DNA, usually <1 Kb in size, using “chain-termination” sequencing (Sanger, et al., 1977). It is a relatively cheap technique but due to the technical sensitivity (15-20% variant allele frequency) its uses are limited in haematological neoplasms which are prone to low levels of clonality. In addition, Sanger sequencing is not amenable to high-throughput screening, limiting its utility as a robust screening method for high numbers

of patients with aberrant blood counts requiring exclusion of neoplastic causes. Fragment analysis can also be used to size specific regions of DNA that are prone to variants that affect the size of the allele, such as internal tandem duplications within *FLT3* (*FLT3*-ITD) which may be acquired in transformation of CMN. Sizing is achieved by performing PCR with fluorescently labelled primers then separation by capillary electrophoresis with a size standard. This method is semi-quantitative and does not allow for full characterisation of any variants detected (ThermoFisher Scientific, 2020).

Whilst WGS and WES has proven to be clinically useful in identifying driver mutations in a research setting (Ley, et al., 2008; Merker, et al., 2012; Grinfeld, et al., 2018), the clinical utility of WGS/WES tests have not been proven in an upfront diagnostic setting. From 2021, Genomics England are making WGS available on the Genomics Test Directory for patients with acute leukaemia alongside standard of care genomic testing in order to address this issue (NHS England, 2020/2021). In time, one would expect that this may also be applied to CMN, but currently the genomic test directory recommends multi-target panels only for CMN. Targeting the sequencing in this manner allows for millions of sequencing reads to be focussed in specific areas of known clinical importance, theoretically increasing sensitivity of the assay through increased read depth.

Multi-target gene panels can vary in a number of ways: the NGS technology utilised for both library preparation and sequencing, the target regions of interest (few genes *versus* many genes; whole genes *versus* partial genes) and can be both commercially developed or developed in-house using adaptable commercial technologies. As such, as a number of different laboratories across the NHS can be broadly described as currently utilising NGS panels in a diagnostic setting for CMN but the number of genes sequenced, the regions sequenced and the technical limitations of the tests offered (such as LoD and the capability to detect complex variants) varies depending upon the methodologies applied. Currently, a limitation of a large number of gene panels when compared to WGS/WES is the inability to simultaneously detect copy number changes, structural rearrangements and mutations. Panels such as the Illumina© TruSight Oncology 500 panel, however, have been developed for comprehensive RNA and DNA analysis of solid tumours, and a similar approach may hold promise for haematological testing in the future (Illumina, 2020).

RNAseq can also be used in isolation to analyse differential gene expression and splicing of all mRNAs expressed in the cells being tested (termed the transcriptome). RNAseq is

particularly useful to identify fusion genes, but gene expression profiles also have potential clinical utility to identify biological subgroups of disease with distinct outcome predictions. This has been demonstrated in both MPN (Schischlik, et al., 2019) and MDS (Iacobucci, et al., 2019; Gerstung, et al., 2015). A very high proportion of the RNAseq technology used in published studies utilises the Illumina short-read technology, which is relatively cheap and easier to implement compared to gene-expression arrays and can produce high quality data across the transcriptome. However, short-read RNAseq is prone to biases and as such, developments have been made for long read cDNA sequencing and direct RNA sequencing without prior modification of the RNA (or the use of reverse transcription) to sequence full length mRNAs offers improved applications for RNA level data analysis which are likely to be incorporated into the genomic/transcriptomic analyses in the future (Stark, et al., 2019).

Of note, tests being implemented into genetic diagnostic laboratories in 2021 should take into consideration the National Genomics Test Directory for cancer which has been developed under the guidance of Cancer experts working with NHS England (NHS England, 2020/2021). This test directory is delivered by 7 genomic laboratory hubs operating as a national network in England to provide a core set of cancer tests (defined by disease) in an equitable manner according to pricing tariffs set by NHS England. Any new test implemented within diagnostic laboratories in England must consider these NHS England directives.

1.6.2 Methods for genomic analysis for patients with CMN at the WRGL

Prior to the onset of this study, the Wessex Regional Genetics Laboratory (WRGL) performed a limited number of routine cytogenetic and molecular screens for patients with CMN, summarised in Table 1.7.

Test	CMN(s) this test is used for	Sample type
Karyotype	New diagnosis CMN and CMN when transformation suspected	BM (or PB if blasts present)
BCR/ABL1 FISH	• ?CML • MPN, for CML exclusion	BM or PB
FIP1L1/PDGFRA (FISH and gDNA testing)	MPN with eosinophilia	BM or PB
KIT D816V by ddPCR	Suspected SM	BM or PB
KIT expanded hotspot analysis by NGS genotyping	Suspected SM but negative for *KIT* D816V	BM or PB
TP53 sequencing by NGS genotyping	Confirmed CMN	BM or PB
AML panel • *FTL3*-ITD by fragment analysis; *FLT3*-TKD, *NPM1*, *IDH1* and *IDH2* variant analysis by NGS genotyping pipeline.	Transformation to secondary AML (from MDS or MDS/MPN) or leukemic transformation of MPN	BM or PB
MPN Panel using NGS genotyping • *JAK2* V617F, *JAK2* exon 12, *CALR*, *MPL*	MPN or suspected MPN	BM or PB

Table 1.7 Tests that were in use in a diagnostic setting within the WRGL at the time of myeloid panel implementation, including the clinical indication for their application.

The tests applied are directed by the sample type, the clinical indication and the genetic test requested and sample testing is triaged by registered Clinical Scientists. The laboratory generally applies a number of different technologies for the analysis of patients with CMN. This included G-banding analysis, fluorescence *in situ* hybridisation (FISH), fragment analysis, ddPCR, reverse transcriptase PCR (RT-PCR), and NGS genotyping (described in more detail below).

Karyotype is undertaken when requested by clinicians on bone marrow samples from patients with a new diagnosis of CMN or in patients with known CMN and a suspicion of disease transformation. Karyotype would only be attempted on a peripheral blood sample if there were significant numbers of blasts present detected by flow cytometry or morphological assessment. *BCR/ABL1* FISH can also supplement karyotype or be used in isolation on bone marrow or peripheral blood in patients with MPN to exclude CML (as well as suspected new cases of classical CML). *KIT* D816V is performed using sensitive ddPCR for patients with suspected SM or SM- AHN; this test is preferably undertaken on bone marrow samples but it can also be performed using peripheral blood when bone marrow is not available. If this test is negative but SM still suspected, an expanded *KIT* test can be performed using the genotyping pipeline. The genotyping pipeline is an ISO15189-accredited NGS pipeline using an Illumina® MiSeq instrument. DNA is extracted from peripheral blood or bone marrow material and the input DNA is subjected to reverse complement PCR (RC-PCR) (NimaGen, 2020) that appends all the functional sequences necessary for sequencing on a MiSeq (sequencing primer binding, sample ID indexes and flow cell hybridisation adaptors). The data is then analysed bioinformatically using a bespoke pipeline generated by collating publicly available online tools and in-house bioinformatics tools and will call any variant detected at a mutation burden of greater than 1%. The AML panel (*FLT3*, *NPM1*, *IDH1* and *IDH2* analysis) is also undertaken using this genotyping pipeline and fragment analysis (for *FLT3*-ITD analysis) for patients with confirmed or suspected transformation of MPN.

In spite of an increasing portfolio of tests provided at WRGL, it was becoming increasingly apparent that molecular analysis of a much wider range of genes would provide useful diagnostic and prognostic information and therefore a myeloid gene panel was implemented into the diagnostic setting between 2019 and 2020. Further details of this panel, a summary of the validation outcomes and the development of strategies to interpret the results are provided in the next Chapter.

1.7 Aims of my project

The introduction above outlines the diversity of genetic aberrations that need to be considered during the diagnostic work up of patients with known or suspected CMNs, and describes the technologies employed at the WRGL to analyse the genomic aberrations in these patients. As technologies developed and the WRGL implemented more in-depth molecular analysis for these patients, namely analysis of MPN driver mutations by the NGS genotyping pipeline and the implementation of a myeloid gene panel, it became apparent that marked changes were required to the analytical approaches employed as well as the clinical interpretation. Specifically, during the validation stages of the myeloid panel it became clear that the output from such a molecular screen is far from straightforward and requires consideration of pathogenicity, mutation burden, referral reason, cytogenetic and other test results as well as technical considerations. In addition, it was clear that there was no unified approach for how to manage the detection of low level variants (e.g. low level *JAK2* V617F mutations <5% VAF).

Therefore, within this context, my project focuses on three interlinked aspects of CMN, with the following aims:

1. To understand how the results of the myeloid panel can be used to inform patient management and establish a robust interpretation and reporting strategy for the results from this test.

2. To understand the significance of low level *JAK2* V617F mutations in the context of diagnostic referrals for MPN testing.

3. To perform molecular characterisation of a rare recurrent cytogenetic abnormality in CMN, the der(6)t(1;6).

The three components are all of diagnostic relevance and are directly linked by myeloid panel analysis, but they also explore the application of other diagnostically relevant techniques including ddPCR, genome wide SNP array and WGS.

2 Implementation of a Myeloid Next Generation Sequencing panel into a diagnostic setting

2.1 Introduction

Within the National Health Service (NHS), genetic laboratories have to consider a number of practical considerations during development, validation and implementation of a new test to ensure accurate results for patient safety; there must also be a standardized approach to interpreting and reporting the test results. Taken together, these factors ensure that patient safety is paramount and that treating Clinicians receive accurate results that can be incorporated into patient pathways to inform management decisions.

In this introduction I detail the considerations required when implementing a myeloid NGS gene panel into a diagnostic setting. Firstly, I give a summary of the utility of genetics, including NGS panels, to help with the diagnosis of CMN for patients in the NHS; secondly, I provide details of the requirements for test validation when implementing a new assay into an NHS diagnostic laboratory; and finally, I discuss the current UK perspective on a standardised approach to somatic variant interpretation.

2.1.1 Diagnosis of chronic myeloid neoplasia (CMN) including the utility NGS panels

CMN are diagnosed using a combination of morphological, immunophenotypic and genetic features in accordance with the WHO criteria for the classification of myeloid neoplasms (Arber, et al., 2016). Since the integration of genetic aberrations into the classification system in 2001, cytogenetic analysis has been an established method of analysis for patients with myeloid neoplasia. In the most recent addition of this classification system, the molecular genetic basis of CMN has been integrated into the diagnostic criteria as new insights into the pathobiology of the myeloid disorders are gained. Even in the absence of a definitive clinical diagnosis, the finding of clonality by molecular genetic testing can be used to differentiate reactive symptoms from haematological neoplasia.

As discussed in the preceding Chapter, testing strategies currently applied in diagnostic laboratories have developed over time as the knowledge of genetic abnormalities in these disorders increases and the technological capabilities advance. At the time of writing, there

is currently no single, economically feasible test that can be applied to haemato-oncology samples in a diagnostic laboratory setting for patients with confirmed or suspected CMN that can reliably identify and characterise all of the genetic aberrations within this genetically diverse category of diseases. However, NGS gene panels have been adopted by diagnostic laboratories as economically feasible methods of sequencing multiple genes simultaneously with the potential to produce meaningful and clinically actionable results. One could argue that the affordability is limited by the requirement for upfront capital investment required to routinely process NGS panels, such as a benchtop NGS sequencers, but the fact that relatively large batches of samples can be sequenced at once reduces the cost per sample when the clinical uptake if sufficient.

2.1.2 Test implementation in a diagnostic setting

Within the United Kingdom (UK), genetic diagnostic tests such as NGS panels for patients with CMN must be accredited by the nationally appointed accreditation body, UK Accreditation Service (UKAS) (UKAS, 2020), in accordance with the regulations and requirements for quality and competence defined within International Organization for Standardization (ISO) standard 15189:2012 (International Organization for Standardization, 2012). The ISO 15189:2012 standards provide general criteria that a validated test must adhere to, where "validation" is defined as "confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled". According to clause 5.5.1.3 within this document, *"*[….] *the validation shall be as extensive as is necessary and confirm, through the provision of objective evidence (in the form of performance characteristics), that the specific requirements for the intended use of the examination have been fulfilled".* However, there is little guidance within this document about how to define the minimum requirements for a test. In response to a lack of clarity in the application of the ISO 15189 standards to genetic tests, EurogenTest published a framework for validation which can be applied with flexibility to design, perform and report validations in diagnostic laboratories (EuroGentest, 2011). This framework should be used in conjunction with the ISO standards and can also be applied alongside the ACCE framework (Haddow, 2003) which summarizes the main factors that should be addressed during the pre-implementation phase: analytical validation, clinical validation, clinical utility and consideration of ethical, legal and social implications of the test to ensure that the test

being implemented is clinically appropriate and feasible in the proposed setting. As part of the assessment of clinical utility and consideration of ethical, legal and social implications, laboratories must adopt (or design) robust methods of result interpretation and reporting so that results can be safely and equitably incorporated into patient pathways.

Of note, tests being implemented into genetic diagnostic laboratories after 2020 should take into consideration the National Genomics Test Directory for cancer which has been developed under the guidance of Cancer experts working with NHS England (NHS England, 2020/2021). This test directory is delivered by 7 Genomic Laboratory Hubs (GLHs) operating as a national network in England to provide a core set of cancer tests (defined by disease) in an equitable manner according to pricing tariffs set by NHS England. The WRGL is part of the West Midlands, Oxford & Wessex Genomic Laboratory hub, with laboratories based in Birmingham, Oxford and Salisbury/Southampton. Any new test implemented within diagnostic laboratories in England must consider these NHS England directives.

2.1.3 Interpreting the results from NGS panels for patients with CMN

2.1.3.1 Somatic variant interpretation

As discussed in more detail above, the diagnosis and prognostic stratification of patients with CMN now incorporates the results from a number of molecular genetic tests recommended in the genomic test directory for GLHs, including NGS gene panels. The incorporation of these results into diagnostic/prognostic criteria relies upon an *a priori* assumption that the diagnostic testing laboratories are applying robust variant interpretation strategies to the results from NGS panels (or independent sequencing). Furthermore, in order for the same standard of care to be available for patients irrespective of which GLH their sample is sent to, these strategies should be applied consistently between laboratories such that any patient being tested would receive the same result and the same interpretation no matter where the testing was performed. However, findings from an audit undertaken by Li, et al., (2017) indicated that at that time there were significant differences in the strategies applied to somatic variant interpretation. Examples including significant differences in the numerical cut-off for minor allele frequency (MAF) for polymorphic variants and the clinical implications of variants reported, with ~80% of participants saying that they *only* report variants with therapeutic implications (rather than

an alternative option, such as reporting all somatically acquired variants known to impact gene function).

In some scenarios, variant annotation is supported by a tranche of knowledge acquired from published literature for specific, well-characterised genes, examples of which include *JAK2, CALR,* or *MPL.* However, not all genes with a known association with myeloid neoplasia have been so well characterised from the perspective of understanding the range of mutations known to significantly impact gene function; *CBLB, GNAS* or *SMC3* may be considered as examples within this group of genes. Consequently, a laboratory's standardised approach to variant annotation must be applicable not only to known mutations but also to specific variants that may not have been reported in the literature before.

At the time of writing, there are no UK best practice guidelines that can be directly applied to the classification and reporting of somatic variants identified by NGS panels in the context of haematological malignancies. However, a number of key papers have been published in the recent years that describe standardised approaches to variant interpretation within a clinical laboratory setting.

In March 2015, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) published a joint consensus recommendation for the interpretation of sequence variants (Richards, et al., 2015). This document provided categories of acceptable evidence, and an associated weighting leading to a 5-tier classification system for variant annotation ("pathogenic," "likely pathogenic," "uncertain significance," "likely benign," and "benign"). In November 2016, the Association for Clinical Genomic Science (ACGS) published a consensus statement that the UK clinical genomics community should adopt this approach to variant interpretation, with publication of UK best practice guidelines in 2018 (Ellard, et al., 2018). However, in both the ACMG/AMP and the ACGS publications, it was clearly stated that these guidelines are designed for rare disease and familial cancer variant interpretation and were not intended for the interpretation of somatic variants due to complexities unique to cancer cells, such as tumour heterogeneity or clonality existing in more than one cell line. In addition, the authors cite the importance of consideration of the appropriateness of the 5-tier classification terminology in the somatic setting. Whilst this could be used to indicate the known (or predicted) biological impact of variants identified, it may be more appropriate to use terms such as "responsive", "resistant" or "actionable" when specifically referring to variants in genes known to be

therapeutic targets. “Actionable” could also be used in more general terms to apply to diagnosis, prognosis, targeted treatments or indication of disease relapse; and/or using the terms “driver” or “passenger” mutations to indicate the apparent biological impact on disease development may be clinically useful. Nevertheless, it is important that appropriate terms are considered and standardly applied to ensure equitable interpretation of results on a national level.

With respect to applying standardised criteria to the interpretation of variants identified in the somatic context, four papers have been particularly influential (Table 2.1).

Publication	Overview
Van Allen, et al. **Clinical analysis and interpretation of cancer genome data**. *J Clin Oncol* 2013. 31(15):1825-1833	**Summary**
	Provides a description of qualitative assessment parameters to be considered to identify high priority variants with clinical significance.
	Limitations
	No practical model provided which can be applied systematically to variants identified in a clinical setting.
Sukhai, et al. **A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer.** *Genet Med* 2016. 18(2):128-136	**Summary**
	Provides a 5-tier classification system that focusses on actionability by assessing the impact of genomic findings on patient management ranging from “Class 1” (previously reported and known to be actionable) to “Class 5 ” (not previously reported and no known actionability).
	Limitations
	Not intended to be used to determine pathogenicity of variants; more appropriate for solid tumour testing where the diagnosis is guided by histopatholigcal and immunophenotype examination. Assumes that germline variation not thought to impact the development of the patient’s cancer has already been filtered out (common in solid tumour testing); germline controls more difficult to obtain for haemato-oncology samples.
Li, et al. **Standards and guidelines for the interpretation and reporting of sequence variants in cancer**. *J Mol DIagn* 2017. 19(1):4-23	**Summary**
	Provides a framework to classify the clinical significance of variants identified into 4 tiers based on the compiled evidence; the strength of evidence is also categorised qualitative weighting. Guidance is also provided about which variants should be reported according to the assigned tier. These guidelines therefore focus on the clinical interpretation of the variants identified rather than the biological classifications.
	Limitations
	This framework is more appropriate for solid tumour cancers rather than haematological neoplasia, where the diagnosis is often known and genomic test is applied in order to find biomarkers that are likely to impact therapeutic interventions and/or germline variants

	that confer an increased risk of susceptibility. The application of this framework for haematological neoplasms may result in markers of clonality being not reported.
Froyen et al. **Standardization of somatic variant classifications in solid and haematological tumours by a two-level approach of biological and clinical classes: An initiative of the Belgian ComPerMed expert panel.** *Cancers* 2019. 11:2030	**Summary**
	A two-level classification workflow. Firstly, variants are assessed according to their biological significance according to the predicted impact on the gene function (adapted from the ACMG/AMP constitutional guidelines): results in a 5-class classification (benign to pathogenic). A semi-quantitative scoring model is proposed for those variants not clearly LoF which has the potential to be automated. Secondly, variants are assigned a classification of clinical significance based on criteria adapted from Li *et al*. (2017).
	Limitations
	Semi-quantitative scoring table may result in an over-representation of scoring variants as VUS and thus not reported. Whilst VUS cannot be used as proof of haematological neoplasia, they may represent passenger mutations/passive markers of clonality which may inform patient management. Furthermore, variants conferring resistance may not be reported in the absence of a driver mutation: again, this may be clinically informative as a representation of the presence of clonality.

Table 2.1 Summary of key papers identified via literature search of publications detailing somatic variant interpretation and reporting.

Firstly, Van Allen, *et al*., (2013) published a clinical analysis of cancer genome data and proposed a method for filtering variants to select those that are most likely clinically relevant. They wrote that classification of variants should "take into account" the type of biological effect it has (i.e. does the variant act as a biomarker for diagnosis, prognosis or predicting disease response to treatment?), what is the strength of evidence (i.e. what type of studies have been published and what is the quality of the publications?) and what is the known size of the biological effect (i.e. does the variant make a drug target more or less sensitive?). By qualitatively assessing these parameters, they propose that "high priority" variants that should be considered clinically significant when making clinical decisions are those with a strong evidence of a large predicted impact. However, this paper lacked a practical model that could be applied systematically to variants identified by NGS panels.

The next publication of note was Sukhai, *et al*., (2016), who developed an interpretation and classification system for somatic variants found from molecular testing of cancer. They presented a 5-tier classification system that focussed on actionability by assessing the impact of genomic findings on patient management ranging from "Class 1" variants that have been previously reported and that are known to be actionable to "Class 5 " variants that have not previously been reported and with no known actionability. Unlike the

ACMG/AMP/ACGS guidelines, these criteria are not intended to be used to determine pathogenicity of variants and as such are more appropriate for solid tumour testing where the diagnosis is guided by histopathological and immunophenotype examination and genetics is used for guiding therapeutic management decisions. In haematological neoplasms, identification of therapeutic targets is only one reason for genetic testing; variant classification criteria must also be applicable for variants that can help in the diagnosis and prognostication of patients with a known or suspected neoplasm. Furthermore, the Sukhai et al. (2016) classification system assumes that germline variation not thought to impact the development of the patient's cancer has already been filtered out, reflecting the common practice of sequencing solid tumour samples alongside a germline control (usually peripheral blood or buccal swab) from the same patient to leave only known acquired variants for assessment. For haematological neoplasms accessing suitable control tissue is more difficult since blood is clearly not applicable and buccal swabs are often substantially contaminated with clonal cells. In practice a germline control is not used for routine panel analysis and variant classification systems for somatic variants detected from haematological neoplasms and thus it cannot be assumed that germline variants have been excluded from the analysis.

In 2017, Li, *et al*., (2017) published a joint consensus recommendation from the AMP, the American Society of Clinical Oncology (ASCO) and College of American Pathologists (CAP) for the reporting of sequence variants in cancer. Building upon the ideas discussed in Van Allen, *et al*., (2013) and Sukhai, *et al*., (2016), Li *et al* (2017) present a classification system to identify clinically significant biomarkers based on the level of evidence available for the clinical impact of the variants identified (denoted Level A-D) based on the following factors: those known to confer therapeutic sensitivity, resistance or toxicity; those known to alter the function of the gene; variants that indicate an inclusion criteria for a clinical trial; variants that provide information relevant for diagnosis or prognosis; and variants that indicate that further disease surveillance is required. The evidence types include variant frequencies, assessment of variant allele frequency to assess whether the variant could be germline, frequency of the variant in population and somatic databases, *in silico* software predictions and published literature assessment. The levels of evidence are then categorised based on their clinical impact into 4 tiers (tier I – IV), with tier I having strong clinical significance and tier IV representing benign or likely benign variants. Recommendations were also provided on the reporting of results, with tier I to III being reported in descending

order of clinical importance and recommendations not to report tier IV variants. This group also address the issue of a systematic approach to the reporting of potential germline findings, with recognition that some genomic tests may be undertaken without a paired germline control. Much like the publication by Sukhai et al. (2016), one could argue that this framework for categorisation is more appropriate for solid tumour cancers rather than haematological neoplasia, where the diagnosis is often known and genomic test is applied in order to find biomarkers that are likely to impact therapeutic interventions and/or germline variants that confer an increased risk of susceptibility. These guidelines therefore focus on the clinical interpretation of the variants identified rather than the biological classifications and as such, the application of this framework for haematological neoplasms may result in markers of clonality being not reported.

In 2019, the Belgian ComPerMed expert panel published a standardised framework for somatic variant interpretation in solid *and* haematological tumours using a two-level approach which harmonises a biological classification and the clinical interpretation of somatic variants identified (Froyen, et al., 2019). This work was undertaken as part of a collaborative effort between members of 27 hospitals across the country. One of the main differences between this publication and those discussed above is the decision to adopt an adapted version of the ACMG/AMP 5-tier classification system for the biological classification of pathogenicity for variants identified, summarised in Figure 2.1. The steps allow decision making to determine whether the variant detected may have an impact on the function of the gene in question as is therefore "pathogenic" or is found in normal individuals with no apparent impact on gene function and is therefore "benign". The user is also recommended to check a list of consensus pathogenic variants (CPV), which is a list of mutations in solid or haematological malignancies known to be pathogenic. This list was curated locally by 6 NGS experts. It was not stated within the publication what the minimum criteria was for inclusion/exclusion and how often this list would be updated. For those variants that are *not* clear loss of function (LoF) variants (such as missense variants and in-frame deletions) the authors provide additional decision making criteria to apply in a semi-quantitative approach based on the frequency on the somatic database COSMIC (rationalising that tumour-related variants will occur at a higher frequency than passenger mutations), concordance between *in silico* prediction tools, evidence of functional studies, and whether it is present on a curated database. Due to the weighting of the scoring parameters, the frequency of the variant on the somatic variant database COSMIC is a

critical parameter, with variants always being annotated as at least likely pathogenic if they are very frequent (≥10 entries in haematological tumours) and for the majority of cases, variant of uncertain significance (VUS) if the number of COSMIC entries is very low (≤5). The authors do not provide evidence about the derivation of this cut-off and one may suppose that there will be scenarios where variants are classified as VUS because the gene region has been inadequately sequenced in tumour studies.

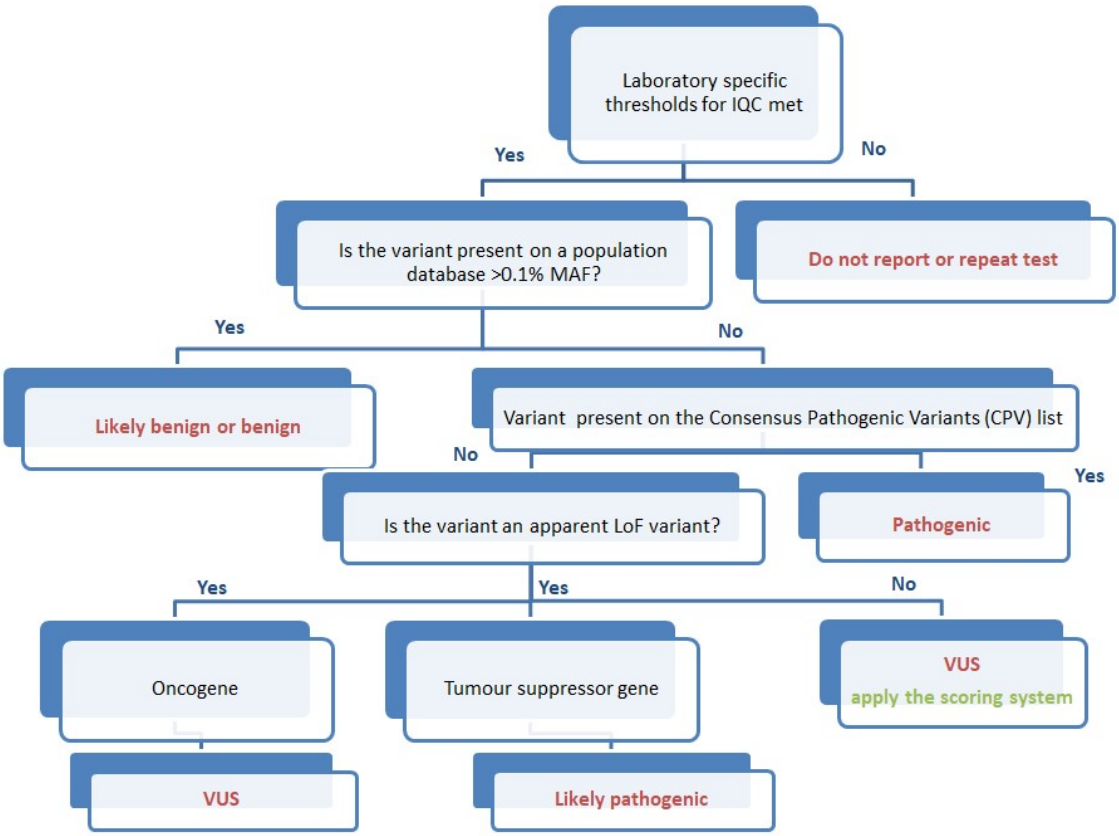


Figure 2.1 ComPerMed workflow adapted from Froyen *et al.* (2019).

The second element of the framework is a clinical interpretation of each pathogenic and likely pathogenic variant to decide whether it should be included on the clinical report or not; it is recommended by this group that the 4 tier classification system published by Li *et al*. (2017) is used. In accordance with Li et al. (2017), VUS and likely benign/benign variants would not be reported as they would always be classified as a Tier III or Tier IV, respectively. According to this framework, there may therefore be pathogenic or likely pathogenic variants that are not reported, such as variants known to confer resistance to therapy in patients without an activating mutation or a single mutation in *CEBPA* (where only double mutations are defined as a good prognostic indicator in AML).

2.1.3.2 Additional considerations for interpreting germline variants

It is reported that up to 10% of adults and children with a haematological neoplasms have an inherited risk factor and, due to the variation in clinical phenotype, it is predicted that they are underdiagnosed (Kohlmann & Schiffman, 2016). The identification of these variants not only allows for effective management of the primary patient presenting with a haematological aberration with respect to planning treatment and predicting future risks, but it also allows for identification of additional at risk family members. Identification of at risk family members also allows for improved treatment decisions in the primary patient in the context of planned allogeneic stem cell transplant (SCT) from a relative (DiNardo, et al., 2018).

When gene panel testing identifies a variant in a gene associated with a germline risk, laboratories are faced with not only determining the biological impact of the variant identified (LOF/gain of function [GoF]/no biological impact) but also determining whether the specific variant identified is germline or somatic in origin. For determining the biological significance, the Clinical Genome Resource (ClinGen) variant curation expert panels formed the ClinGen Myeloid Malignancy Variant Curation Expert Panel (MM-VCEP) to collaborate with the American Society of Hematology (ASH) to modify the ACMG/AMP guidelines for specific inherited myeloid malignancies. They have subsequently published two documents detailing recommendations for the gene specific recommendations for the interpretation of germline variants in *RUNX1* (Luo, et al., 2019) and *TP53* (ClinGen TP53 Variant Curation Expert Panel, 2019). In these documents, there are additions to the original ACMG guidelines which are specific to these genes such as adding granularity to the specific hotspot codons which should be considered in *TP53* and *RUNX1*, which is an amendment to the PM1 criterion ("*Located in a mutational hotspot and/or critical and well-established functional domain without benign variation*") and the required observed frequency of the variant in population controls in order to apply strong benign criteria (BS2; "*variant must be observed in >8 cancer free 60+ year old females"*) and supporting benign criteria (BS2_supporting; "*observed in 2-7 cancer free 60+ females").* In addition, the *RUNX1* criterion marks the move towards employment of REVEL, an Ensembl method for predicting the pathogenicity of missense variants (Ionnidis, et al., 2016), rather than the application of a consensus prediction from historically established tools such as SIFT (Vaser, et al., 2016), PolyPhen (Adzhubei, et al., 2010) and AGVGD (Tavtigian, et al., 2006) as there is evidence

that this approach increases the positive predictive power of *in silico* prediction (Tian, et al., 2019). They have also indicated those criteria that are not appropriate in the context of these genes, such as PM3 ("For recessive disorders, detected in *trans* with a pathogenic variant") and PM4 ("Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants"). For *RUNX1* variant assessment, these rules were tested using a pilot set of 52 germline variants which were previously classified as benign, likely benign, VUS, likely pathogenic, pathogenic and/or a conflicting interpretation (CONF) in the ClinVar database (Luo, et al., 2019). It was reported that the application of this standardised framework resulted in a reduction of VUS and CONF classifications by 33%, indicating the utility of robust models of classification.

For the majority of gene panels applied for patients CMN in diagnostic laboratories, testing does not include testing of a germline sample, mostly due to the difficulty in obtaining a true germline sample on a routine basis without contamination with leukocytes which are usually infiltrated with disease cells. One option for investigating the origin of these variants includes a blood sample when a patient is in complete remission (CR); however, it is sometimes difficult to obtain CR in patients at all, let alone in a timely manner. Alternatively, a buccal or fibroblast sample can be taken but saliva samples are typically contaminated with peripheral blood cells (Heinrichs, et al., 2010) and fibroblast sampling is not currently commonplace in the haematology clinic. In the absence of germline controls, DiNardo et al (2018) recommend using the VAF to identify those variants that may be of germline origin, with heterozygous variants primarily falling into the VAF range 40-60% and homozygous variants having a VAF >80%.

2.2 Aims and objectives

During the establishment of an NGS myeloid panel into the diagnostic portfolio of UKAS accredited tests at the WRGL it became clear that in the absence of national guidelines the laboratory required novel analytical approaches and clinical interpretation strategies to fully understand the results obtained and report them in a standardised manner. Furthermore, as the implementation of this assay represented the first gene panel offered for local patients with confirmed or suspected CMN, we aimed to assess how results were being utilised by referring clinicians.

The principal objectives of this part of my study were therefore as follows:

1. Develop a bespoke approach to classification/reporting variants identified by the assay in the absence of best practice guidelines that can be directly applied (i.e. applied without adaptation) in a diagnostic laboratory setting testing samples from patients with CMN.
2. Perform a clinical audit to assess how the myeloid panel results were being used to impact clinical management of patients with CMN in the NHS.

2.3 Materials and methods

2.3.1 Myeloid panel validation at the WRGL

2.3.1.1 My role in the myeloid panel validation

In 2017, the Research team based the WRGL (REC reference 10/H0102/61; IRAS project ID: 52340; Chief Investigator: Prof NCP Cross.) initiated a feasibility study looking into the utility of a myeloid panel for patients with CMN. The Illumina® TruSight Myeloid sequencing panel (TSMP) was chosen for implementation and in 2020, in collaboration with colleagues in the diagnostic and research teams based in the WRGL, I led the team effort to validate this assay according to UKAS standards. This assessment was considered as an "extension to scope" as the WRGL was active UKAS accredited laboratory performing panel tests for a range of non-malignant genetic disorders; extension to scope was granted on 02.01.2021. Further details are provided below.

In addition to defining and assuring the quality and accessibility of this test during the test validation process, prior to test implementation the WRGL had to ensure that the interpretation and reporting of results were appropriate for a diagnostic setting. I therefore also led the development of a somatic variant interpretation strategy for variants detected by the myeloid panel. This was developed from a strategy designed in 2018 (also led by me) for interpreting variants detected in *TP53* by the NGS genotyping pipeline in patients with CLL.[i] Whilst the test validation is referred to throughout and the key findings are summarised below, it is a discussion of the development of a somatic variant interpretation strategy that can be applied to any gene associated with CMN and a reporting strategy for those variants considered clinically relevant that are the main focus of this Chapter.

[i]The *TP53* NGS genotyping assay was initially developed for patients with chronic lymphocytic leukaemia (CLL) to supplement FISH testing. These patients have a variable disease course but disruptions of *TP53,* which can include whole gene and intragenic deletions and mutations, are known to define a "high risk" disease associated with short progression-free and overall survival in the context of conventional genotoxic therapy (Zainuddin, et al., 2011; Gonzalez, et al., 2011). The same poor prognosis has also been associated with low level variants (Rossi, et al., 2014) therefore mutation analysis requires an assay that is sensitive enough to detect variants <10% VAF, such as the diagnostic genotyping pipeline. In CLL, pathogenic mutations can be anywhere in the gene but occur most frequently in the DNA binding domain and are commonly missense changes therefore variant interpretation is required to ensure that only those variants thought to impact gene function are included in the diagnostic report.

2.3.1.2 Details of the myeloid panel implemented

When the implementation of a myeloid panel assay was being considered, the WRGL had already purchased two Illumina MiSeq™ machines for implementation of diagnostic rare disease panels as well as research studies. This effectively limited the choice for a diagnostic myeloid panel to (i) a bespoke, custom designed panel or (ii) the Illumina® TSMP; the only commercially available panel at the time that was compatible with Illumina sequencers. (The Illumina® MiSeq sequencing technology is summarised in Appendix 7.1). These options were considered based on the following criteria: panel design, clinical relevance of genes covered, technical performance from preliminary experiments conducted by the research group, ease of implementation in the diagnostic laboratory (including susceptibility of the workflow to error and contamination) and price. The situation was further complicated by the fact that National Genomics Test Directory had not been published and there was no clear consensus as to what genes needed to be covered and how to accommodate new research findings. Although a custom designed panel appeared to have the advantage of flexibility and adaptability, the stringent requirements for validation and accreditation suggested that a fixed panel designed to last for at least 1-2 years would be preferable. Consequently, and taking into account the factors above, the decision was made to adopt the TSMP.

The Illumina® TSMP (P/N FC-130-1010) is an amplicon based capture kit for preparation of targeted libraries for NGS on Illumina platforms. The capture amplifies 568 amplicons (equivalent of ~141 kb of genomic content) from 15 full genes plus exonic hotspots of an additional 39 genes known to be tumour suppressor genes or oncogenic hotspots associated with myeloid malignancies (Table 2.2).

Gene	Target region (exon)	Gene	Target region (exon)	Gene	Target region (exon)	Gene	Target region (exon)
ABL1	4-6	***DNMT3A***	Full	***KDM6A***	Full	***RAD21***	Full
ASXL1	12	***ETV6***	Full	***KIT***	2,8-11,13,17	***RUNX1***	Full
ATRX	8-10, 17-31	***EZH2***	Full	***KRAS***	2,3	***SETBP1***	4 (partial)
BCOR	Full	***FBXW7***	9-11	***KMT2A***	5-8	***SF3B1***	13-16
BCORL1	Full	***FLT3***	4,15,20	***MPL***	10	***SMC1A***	2,11,16,17
BRAF	15	***GATA1***	2	***MYD88***	3-5	***SMC3***	10,13,19,23,25,28
CALR	9	***GATA2***	2-6	***NOTCH1***	26-28,34	***SRSF2***	1
CBL	8,9	***GNAS***	8,9	***NPM1***	12	***STAG2***	Full
CBLB	9,10	***HRAS***	2,3	***NRAS***	2,3	***TET2***	3-11
CBLC	9,10	***IDH1***	4	***PDGFRA***	12,14,18	***TP53***	2-11
CDKN2A	Full	***IDH2***	4	***PHF6***	Full	***U2AF1***	2,6
CEBPA	Full	***IKZF1***	Full	***PTEN***	5,7	***WT1***	7,9
CSF3R	14-17	***JAK2***	12,14	***PTPN11***	3,13	***ZRSR2***	Full
CUX1	Full	***JAK3***	13				

Table 2.2 Gene regions assessed by the TruSight Myeloid Sequencing panel.

An overview of the TSMP workflow is shown in Figure 2.2. The process is described in detail in Appendix 7.2.

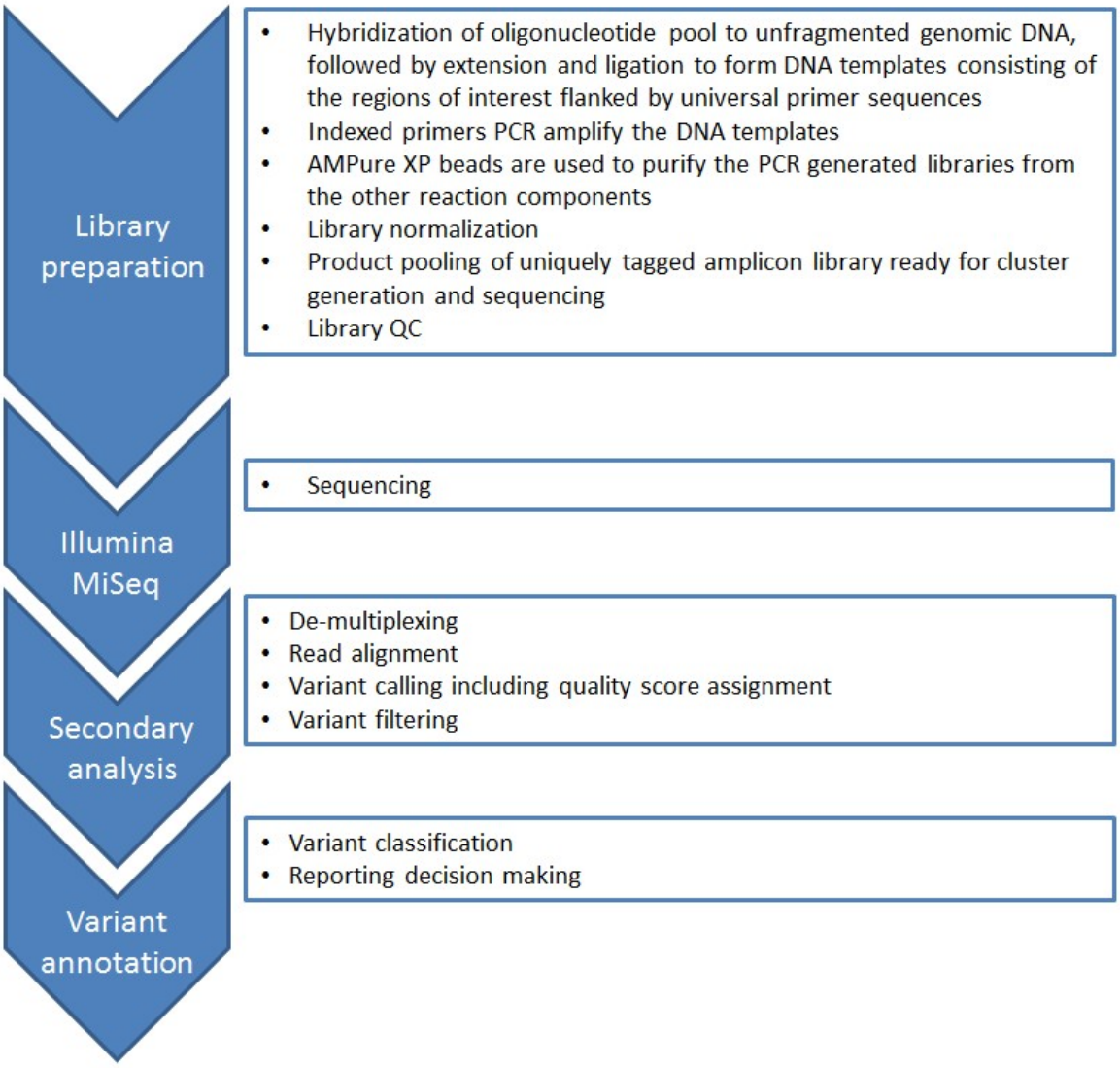


Figure 2.2 The Illumina TSMP sequencing panel workflow overview.

2.3.1.3 Summary of the myeloid panel validation outcomes

When considering the implementation of the myeloid panel for patients with confirmed or suspected CMN into the WRGL, in addition to the guidance provided by EurogenTest mentioned above, several recommendations published for the implementation of NGS techniques into diagnostic laboratories generally (Rehm, et al., 2013) and also specifically in the cancer setting were considered. Cancer specific guidelines included the 2017 joint consensus recommendation for the use of NGS-based sequencing panels for oncology panels published by the AMP and the College of American Pathologists (Jennings, et al., 2017) and Kanagal-Shamanna, *et al*., (2016) who describe recommendations for NGS for haematological neoplasms within the United States (US).

During the accreditation process, UKAS assesses the documentation that a diagnostic laboratory had collated and the procedures set in place within the laboratory to ensure that the test is safe, accurate to the reported limitations of the assay and accessible to referring clinicians via robust test request pathways, according to a number of standards defined in ISO 15189. The validation document is provided in Appendix 7.3 but the samples used and the key performance metrics are summarised below.

Samples used for the TSMP validation

'Gold standard' samples used included samples tested by the UKAS accredited NGS genotyping pipeline (n=32) and 2 commercially available standards: the Horizon Myeloid DNA Reference Standard (Horizon Discovery Ltd, Cambridge, UK) and the Coriell reference cell lines (NA19240 and NA12878).

Key performance criteria of the TSMP

Specificity (read depth ≥100): 99% to 100% (95% CI)

[Where analytical specificity was defined as: true negative/(true negative + false positive).]

Sensitivity (read depth ≥100): 95% to 100% (95% CI)

[Where analytical sensitivity was defined as: true positive/(true positive + false negative).]

Limit of detection: 5% VAF.

[Based on data from the commercial reference standards and Coriell cell line mixtures]

Minimum read depth: 100x

[Based on data from the commercial reference standards]

2.3.2 Developing a strategy for interpreting results from the TSMP for patients with CMN

2.3.2.1 Developing a strategy for variant filtering

Following initial bioinformatic processing of the TSMP using MiSeq reporter (integrated analysis software that produces information about alignment, structural variants, and contig assemblies for each genome and sample), approximately 300 variants are detected within the variant call file (.vcf) of each sample. The Alissa Interpret (Agilent) platform, used routinely at the WRGL for rare disease NGS panels, offers an option for variant filtering through user-defined Classification Trees, which can be version-controlled within the software and therefore fully auditable and reproducible.

A classification tree (described below) was adapted to filter out those variants which are unlikely to be relevant to the patient such as variants that have been reported at high frequency in unaffected individuals (polymorphisms) and low level variants with poor IQC metrics which are likely to represent technical artefacts; IQC cut-offs were defined within the validation of this assay, described above. Of note, the filtering process does not permanently remove the variants from the .vcf file or the analysis record within Alissa Interpret, thus allowing re-analysis in the future, if required.

2.3.2.2 Developing a somatic variant interpretation strategy

A literature review was performed to identify the current strategies employed in by using the following search terms +/-the keywords *cancer, haematological malignancy, haematology: somatic variant interpretation*; *somatic variant classification*; *interpretation of cancer data*; c*lassifying results from NGS panels; reporting results from NGS panels.* The UK Best Practice Guidelines for variant classification in rare disease (Ellard, et al., 2019), the ACMG/AMP US ClinGen Sequence Variant Interpretation (SVI) working group publications (Ghosh, et al., 2018; Tayoun, et al., 2018; Biesecker, 2018; Brinich, et al., 2020) and the ClinGen *RUNX1* (Luo, et al., 2019) and *TP53* (Fortuno, et al., 2021) germline variant curation recommendations were also reviewed in depth to assess how standardised variant interpretation in the germline setting could be applied in the context of somatic variants. As detailed above, a strategy was adapted from an existing variant interpretation strategy applied to interpreting variants identified within *TP53* for patients with CLL (presented at the International Workshop on CLL (iwCLL) 2017, New York; Appendix 7.4)

2.3.3 Clinical audit to assess the utility of the myeloid next generation sequencing panel

The Illumina® TSMP was first implemented at the WRGL on a purely research basis in September 2016. The audit aimed to review the initial use of the panel within the South Central (formerly Wessex) region for haemato-oncology patients with confirmed or suspected CMN. The overall aims were to assess (1) how TSMP results lead to changes in patient management and (2) how useful the clinician perceived the result to be to ensure this justifies the cost of the test.

The specific objectives of the audit were to:

(i) Capture a real world experience of the TSMP

(ii) Understand the clinical utility of the test

(iii) Collect data for further analysis

In order to collect this data, an audit questionnaire (Figure 2.3) was sent out to the referring clinicians who requested the first 200 myeloid gene panel tests (September 2016 - October 2017). Participants were requested to return completed forms within 6 weeks.

Salisbury NHS
NHS Foundation Trust

Wessex Regional Genetics Laboratory
Salisbury District Hospital
Salisbury
Wiltshire, UK
SP2 8BJ

Myeloid Gene Panel Audit

Patient Name:
WRGL accession number:
Centre:
Date of Sample:
Age:

1. Primary reason for submitting sample for myeloid gene panel (MGP) analysis (please circle):
   - *Diagnosis*
   - *Prognosis*
   - *Assess disease status, e.g. disease progression*
   - *?Relapse*
   - *?Response to treatment*

2. FBC results: Hb ..... g/dL; platelets ..... x 10^9/L; neutrophils ..... x 10^9/L (or attach report)

3. BM report: (please attach a copy)

4. WHO diagnosis .....

5. How was the diagnosis made? .....

6. How did the MGP results change the management of the patient (please circle)?
   - *Clarified diagnosis as bone marrow not diagnostic*
   - *Avoided a bone marrow procedure*
   - *Allowed patient to be discharged from clinic*
   - *Prognostic to guide active management decision e.g. SCT*
   - *del(5q), p53 status*
   - *Confirmed relapse or progression (or confirmed ongoing remission/absence of progression)*
   - *Interesting to know but did not change practice*
   - *Did not help*

7. How helpful did you find did you find the MGP result (please circle; 1 = useless; 10 = very useful)

1 2 3 4 5 6 7 8 9 10

Figure 2.3 Clinical audit questionnaire form

2.4 Results: somatic variant interpretation strategy

2.4.1 Overview

An adjustable variant filtering pipeline is first applied to variants detected in each sample to remove polymorphic variants and variants likely to represent technical artefacts. Sequence variants remaining are then classified by a two- step approach, adapted from the ACMG guidelines for constitutional variant detected in rare disease for biological classification and Li, et al., (2017) for clinical classification. Firstly, variants are classified as “pathogenic”, “likely pathogenic”, “VUS”, “likely benign” or “benign” according to a series of criteria with levels of evidence. Secondly, variants are categorised as either clinically significant or not clinically significant, and only the former are reported. This categorisation is based on the Li, et al., (2017) framework but with some adaptations, described more fully below. This two-tiered assessment approach was adopted as the pathogenicity of a given variant should not change depending on the patient; however, the interpretation of the finding(s) on the report may differ as it is applied to different clinical scenarios (Figure 2.4; discussed further below).

Of note, this protocol uses the interpretation software tool, Alamut® Visual (SOPHiA GENETICS, 2020) to perform *in silico* analysis and the collate information of the biological significance of variants and the clinical informatics software, Alissa® Interpret (Agilent, 2018) to manage variant curation and long term storage of variant classification. The use of these software tools are not fully described in this document but are summarised above and referenced throughout. The processes involved in variant filtering are summarised below.

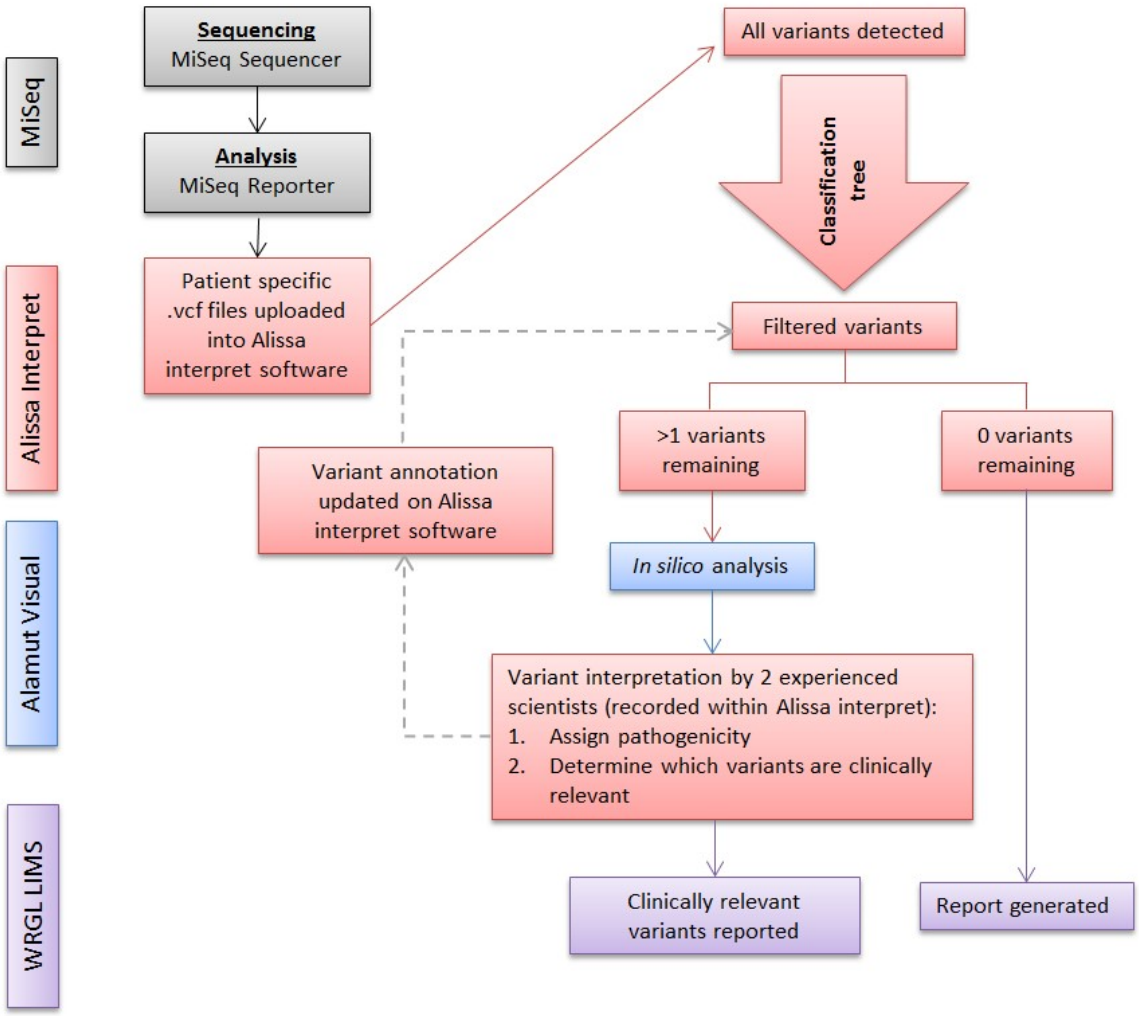


Figure 2.4 Overview of the diagnostic analysis workflow which is undertaken on the Illumina MiSeq (grey), the web-based software Alissa Interpret (red) and the laboratory information management system (LIMS) at WRGL (purple). Primary and secondary analysis of the sequencing data is done on the Illumina Miseq to generate patient specific .vcf files. These .vcf are uploaded into Alissa. Within Alissa, variants are filtered and interpreted; analysis notes and records of the interpretation and final conclusions are stored within Alissa in line with the requirement for diagnostic laboratory to store audit trails of analysis and decision making steps. Once the decisions have been made regarding which variants are to be reported and their clinical significance, a report is generated and authorised within the WRGL LIMS and distributed to clinicians.

2.4.2 Stage one: variant interpretation

2.4.2.1 Overview

This stage of variant interpretation aims to determine or predict the biological impact of a variant on the function of the gene for the variants that are presented for analysis for each sample following variant filtering within the Alissa interpret software. A template was produced to record the evidence collated for variant interpretation in order to ensure

standardised records; this is called the managed variant list (MVL) template (Appendix 7.5). This template is reviewed every 12 months to ensure that is up to date.

Evidence is compiled according to the following criteria, and each is discussed in more depth below: (i) databases; (ii) known mechanisms for pathogenicity in the affected gene; (iii) *in silico* prediction tools; (iv) literature evidence. The collated evidence is then reviewed to decide the pathogenicity of the variant.

2.4.2.2 Variant filtering

The classification tree designed from the quality assurance (QA) criteria indicated from the results of the validation experiments for the TSMP (specifically the minimum read depth required and the validated limit of detection) is an integral part of the somatic variant interpretation strategy at the WRGL and it acts as an initial filtering stage applied to variants detected in each sample (Figure 2.5). The classification tree is also dynamic, in that it incorporates previous variant assessments performed by analysts at the WRGL by inclusion of a filter step which interrogates the classification of variants inputted into the in-house MVL. Resultantly, the number of variants presented to analysts for variant interpretation is reduced from approximately 300 to less than 10 per sample. Filtering acts to remove those variants that are considered unreliable and most likely to represent technical artefacts (<100x read depth and/or a QC score[ii] <30) and variants known to be very frequent in the population (defined as ≥1% MAF according to the GnomAD dataset) and leave those variants that are known to impact gene function or have not yet been through variant interpretation but are likely to represent real changes (whether germline or somatic in origin).

Of note, analysts will review variants ≥100x read depth and/or with a QC score of ≥30 in case a clinically significant variant is identified that could be confirmed by another method, but a read depth of ≥500x *plus* a QC score of 100 was considered empirically to be the most reliable QA indicators based on the validation outcome and reported cut-off levels from other laboratories (Thomas, et al., 2017). In addition, there is a separate branch of the

[ii] The "QC score" is equivalent to the Q-score which is applied during secondary analysis by the MiSeq Reporter software, described in more detail in appendix 2. It is based upon the Phred scale to give the probability of error; Phred scale is defined as -10 * log(1-p), where a value of 10 indicates a 1 in 10 chance of error, while a 100 indicates a 1 in 10^10 chance.

classification tree which will, using different parameters, identify low level variants (i.e. those below the validated LoD of 5% VAF) at specific sites that are recurrently subjected to oncogenic changes (e.g. *JAK2* V617F) and for which we have an independent in-house assay available for confirmation of these low level findings.

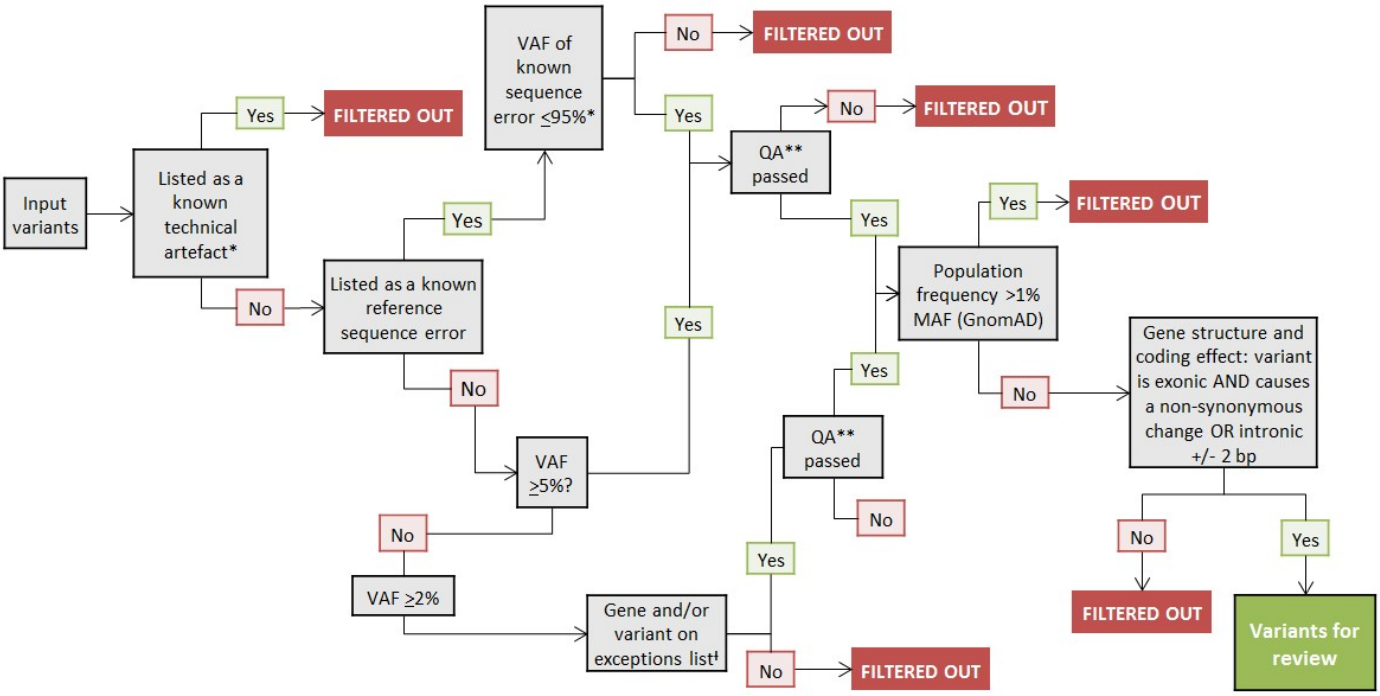


Figure 2.5 Variant filtering steps incorporated into the classification tree within Alissa Interpret. *This step was introduced to ensure that real somatic variants were not excluded in error; ** Read depth >100x AND call quality >30; ǂ Only pathogenic variants or genes with high association with actionable variants are on this list e.g. *FLT3*. Genes on this list have independent sequencing options at WRGL in order to confirm variants below the LoD for the TSMP (ie <5% VAF).

2.4.2.3 Collating evidence: databases

The variant identified is assessed using a number of different databases including cancer specific databases, population databases and an intra-laboratory database:

<u>Cancer-specific databases</u>

Cancer-specific databases will provide information about whether the specific variant has previously been reported in cancer. However, this is not considered as sufficient evidence in isolation to prove pathogenicity, as somatic variants may represent passenger mutations detected as they impact clinically relevant genes. Caution should particularly be applied when the variant has not been confirmed as a somatic change, as some databases (such as COSMIC) are not robustly curated and have loosely-controlled pre-requisites for submitting

data (Li, et al., 2017). As such, low frequency variants that have not been confirmed may represent technical artefacts or benign variants.

Catalogue of Somatic Mutations in Cancer *(COSMIC; http://cancer.sanger.ac.uk/cosmic).*

COSMIC contains millions of somatic variations across numerous tumour types. The frequency of the specific variant is recorded and also other variants affecting the same residue/residues within the same functional domain. This will indicate whether acquired variants in the affected nucleotides/amino acids have been reported in cancer previously, which cancer type and at what frequency.

The variant distribution within the whole gene is also assessed; this will indicate whether the gene has an apparent hotspot region(s) but must be considered alongside literature evidence as the variant distribution across a gene may have been biased by selective characterization of previously proven oncogenic regions within the gene.

IARC (WHO) TP53 mutation database *(http://p53.iarc.fr)*

IARC is a knowledgebase and includes statistical tools for the analysis of *TP53* gene mutations in human cancers. This *TP53* specific database includes records of both somatic and germline mutations in *TP53*, including validated polymorphisms.

Genomic databases

ClinVar *(https://www.ncbi.nlm.nih.gov/clinvar/):* aggregated information about genomic variation and its relationship to human health. This database incorporates somatic and germline variants and provides references to relevant publications. Review status indicates the level of confidence in any assertion (scored 0 [no assertion criteria provided] to 4 [accepted in best practice guidelines]) and reflects whether an interpretation is available, whether documentation of the assertion criteria provides transparency into the classification process, and whether there is consensus between submitters.
When a variant is present on this database as a somatic aberration (rather than germline) with a high review status (2 or more), it is considered as high strength evidence that the biological classification recorded on this database is accurate. In additional, germline records can also be considered as evidence if the gene affected is known to result in an inherited haematological disorder (discussed in more detail below).

Population databases

GnomAD Genome aggregation database *(http://gnomad.broadinstitute.org/).*

Many different ethnic populations are represented totalling 125,748 individuals (exome data) and a further 15,708 individuals (genome data), totalling 141,456 individuals (gnomAD v2.1, which uses genome build hg19). The data is aggregated from many individual projects, which include several disease cohorts such as myocardial infarction genetics consortium and the Swedish schizophrenia and bipolar studies.

GnomAD is normal used to establish if a variant of interest is present in apparently unaffected individuals, and if so, at what frequency. Three of the most commonly encountered scenarios are described below:

(a) *The frequency is >1%:* If either the overall allele frequency and the highest allele frequency listed in gnomAD is >1%, this can be considered as stand-alone evidence for a benign classification and no further criteria need to be satisfied (the identification of these variants has been incorporated into the variant classification tree within Alissa as an automated step to streamline the variant interpretation process).

(b) *The variant is absent from this database*: This can provide evidence that the variant is pathogenic; however, sequencing depth can vary between genes/gene regions and therefore it is important to ensure there is sufficient coverage over the affected region (Figure 2.6). Secondly, insertion/deletions (indels) may be poorly covered or may be miscalled in the gnomAD database, particularly in repetitive regions, so it is important to ascertain whether other indels have been detected within the region.

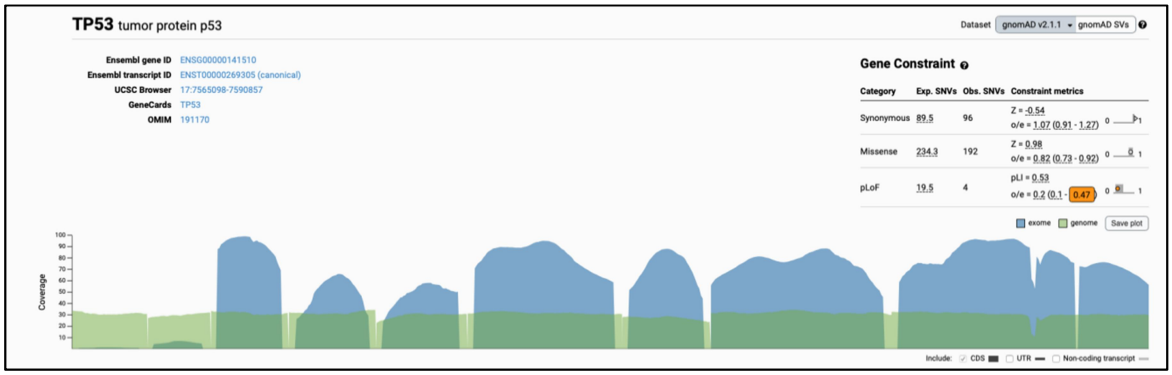


Figure 2.6 Example of coverage as displayed for each gene; coverage for the coding regions of the gene by WES is displayed by blue areas; mean coverage for WGS in this region is displayed in green.

(c) *The variant is present on GnomAD at low frequency (<<1%)*: it is important to remember that haematological malignancy-associated somatic driver mutations frequently appear in population databases due to CHIP. In most cases this occurs in older individuals (>40-50 years old) and the VAF is usually low (<10%). Variants that occur in younger individuals in genes not known to be associated with an inherited haematological disorder are likely to be benign.

Examples of differences in frequency and age distribution of variants detected in the GnomAD dataset are shown in Figure 2.7. In Figure 2.7-A, one can see an example of the frequency and age distribution of individuals carrying a common polymorphism (17-7579472-G-C; GRCh37): the variant has been detected in all age groups and the majority of the variant carriers identified are in the age groups most represented in this data set (i.e. most carriers are aged 40-75). In Figure 2.7-B, one can see an example of the frequency and age distribution of the known pathogenic variant, *JAK2* V617F, which is known to increase in prevalence with age due to both the occurrence of haematological neoplasia and through the acquisition of CHIP. In Figure 2.7-C, one can see an example of the frequency and age distribution of a variant of unknown clinical significance (5-141357912-C-A; GRCh37 in the gene *TET2*]: the carrier distribution does not appear to be the same as that seen for common polymorphisms which neither supports nor excludes that this variant is a rare polymorphism. Furthermore, *TET2* can be mutated in both haematological neoplasia and CHIP, but there is no definitive increased prevalence with age. The significance of this finding can therefore be difficult to interpret.

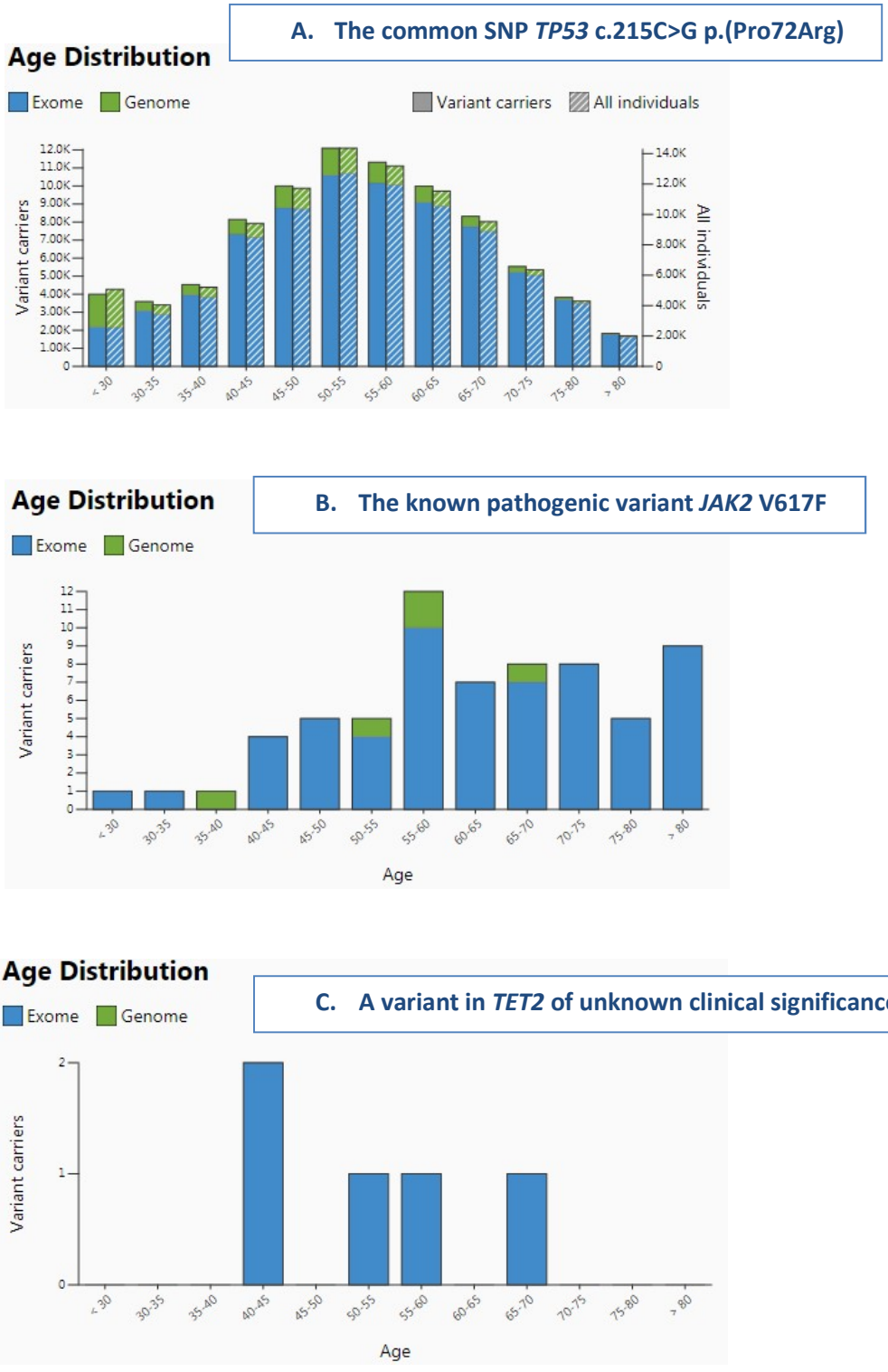


Figure 2.7 Example of age distribution of variants detected in GnomAD. **A.** The age distribution for individuals with the common polymorphism 17-7579472-G-C (GRCh37), which was detected in 186,832 out of 281,846 by WGS and WES (allele frequency 0.6629). In this image, the total number of individuals tested for this region for each age group is shown (bars with diagonal lines) and it can be seen the frequency is relatively evenly distributed across all age groups. **B.** The age distribution for individuals with the known pathogenic variant *JAK2* V617F [9-5073770-G-T (GRCh37)], which was detected in 97 out of 281,626 individuals (allele frequency 0.000344). This variant is more frequent in older individuals in this data set which is likely to be a combination

of individuals with MPN (known or unknown) or CHIP. **C**. The age distribution for individuals with a variant in *TET2* [5-141357912-C-A (GRCh37)], which was detected in 5 out of 250,606 individuals (allele frequency 0.00002). The variant has been seen in this cohort but it is rare and only in individuals >40 years old therefore this variant is of uncertain clinical significance.

NHLBI GO Exome Sequencing Project *(ESP; http://evs.gs.washington.edu/EVS/)*

ESP is an aggregated collection of exome and genome sequencing data from patients with heart, lung and blood disorders (>200,000 unrelated individuals). This database can be used in the same way as the GnomAD database, above; however, as there are individuals with blood disorders represented in this dataset, it is possible that low penetrance germline variants associated with a haematological phenotype may be present. As a result, the absence of a variant from this database may be considered as supporting evidence of pathogenicity but the presence of a variant in this dataset cannot be used as supporting evidence of being benign (unless >1% MAF).

Internal (laboratory database)

All laboratory findings are stored within Alissa, allowing for database interrogation in order to determine the frequency of a variant in the patients tested thus far. The internal quality control (IQC) cut-offs (described below) for the variant in an active sample and in previously tested samples should be considered. Variants are annotated within this database for both tracking variants identified within the laboratory and to provide consistency with respect to annotation (pathogenic, likely pathogenic, likely benign, benign, likely/confirmed artefact). If the specific variant under investigation has not been detected before, variants affecting the same amino acid or nearby regions in the same function domain can be reviewed. Some commonly encountered scenarios are described below:

(a) *The variant has been detected before (>10 times) and is always <10% VAF with poor IQC metrics (e.g. read depth <500; QC score <100):* these are considered as technical artefacts since true driver mutations would be expected to be present in some samples at VAFs >10%.

(b) *The variant has been seen multiple times (>5 times) with a VAF within a heterozygous/homozygous allele frequency range (i.e. close to 50% or 100%):* this would be considered supporting evidence that the variant represents a rare

polymorphism; this should be confirmed using population databases and literature and *in silico* evidence. However, as discussed above, the presence of a small number of reports of the variant on a population database (e.g. GnomAD) does not preclude that this variant could represent a somatic change (i.e. CHIP). Similarly, as there is increasing evidence that a significant proportion of variants in dbSNP represent false-positive calls (Mitchell, et al., 2004), the presence of a variant on this dataset cannot be taken as definitive proof that the origin is germline.

(c) *The variant has been reported >1x previously and passes IQC:* the analyst should check whether variant interpretation has been performed before and if so, when. If not previously classified variant interpretation should be performed. If the variant has been seen ≥1 year ago and classified as a VUS, a re-assessment should be performed. This involves reviewing the evidence compiled when an initial classification was made to ensure that it fulfils the most recent version of the variant interpretation protocol and performing an up-to-date literature review to identify any new publications which might up-/downgrade the classification.

(d) *The variant has not been detected before and the variant passes IQC:* variant interpretation should be performed to determine the biological significance of the variant.

2.4.2.4 Collating evidence: known mechanisms for pathogenicity in the affected gene

Accurate variant interpretation requires knowledge of the gene structure, protein structure/function, previously identified variants and mutational mechanism (e.g. whether a particular gene is pathogenic by haploinsufficiency, GoF or LoF). As all genes tested on the TSMP are disease-associated genes, they have a strong gene-disease relationship because they have been selected for the panel on this basis; however, there can be gene specific considerations. The mechanism of pathogenicity can vary depending on: (1) the gene function; for example, generally speaking tumour suppressors will require LoF mutations and oncogenes will require gain of function mutations to be considered pathogenic; (2) whether the variant is germline or somatically acquired; for example, *NRAS* gain of function mutations in the somatic context are known to be pathogenic when they are missense variants impacting the amino acids p.Gly12, p.Gly13 or p.Gln61 but in a germline context, the missense changes p.Ile24Asn, p.Pro24Leu, p.Thr50Ile have also been reported in affected Noonan-syndrome families (Atmuller, et al., 2017).

Consequently, resources such PubMed, NCCN guidelines (National Comprehensive Cancer Network, 2019a; National Comprehensive Cancer Network, 2019b), OMIM (McKusick-Nathans Institute of Genetic Medicine, 2020) and GeneReviews (NCB1 Resource Coordinators, 2018) were assessed for each gene to determine the mechanism of pathogenicity in the context of haematological neoplasia and in germline inherited haematological disorders, where appropriate (discussed further below). For genes with limited information available, the GnomAD pLI score was used; this score is the probability that the gene is loss-of-function intolerant based on the GnomAD (population) dataset after analysis of the number of protein truncating variants after adjusting to the size of the gene and sequence coverage (Lek *et al*., 2016). From this compiled information, a master list was generated for use by analysts in the diagnostic setting and incorporated into the MVL template (Appendix 7.5).

2.4.2.5 Collating evidence: *in silico* prediction tools

Prediction algorithms should not be used in isolation for variant classification and instead should be used as supporting lines of evidence, in line with Li et al (2017). This is because missense and splice site prediction tools have only a moderate specificity (approximately 60-80%) with a tendency to over-predict the deleterious impact on protein function (Thusber, et al., 2011; Houdayer, et al., 2012; Vreeswijk, et al., 2009).

At the WRGL, all *in silico* tool are accessed via Alamut Visual and these tools are summarised in Table 2.3. SIFT (Vaser, et al., 2016), PolyPhen-2 (Adzhubei, et al., 2010) and Align-GVGD (Tavtigian, et al., 2006) provide a prediction on the deleterious effect of missense changes based on algorithms that incorporate conservation data across multiple species. A consensus between SIFT, PolyPhen and AGVGD can provide strong supporting evidence of the impact of the variant on protein structure.

In silico tool	Summary
SIFT (sorting intolerant from tolerant)	This software provides prediction of the deleterious effect of missense changes. It does this by assuming that amino acids that are highly conserved between species are important for function, then predicting whether the inputted change would be tolerated depending on the degree of homology to the wild-type sequence.
PolyPhen-2	This software uses a combination of sequence- and structure-based predictive tools to compare the properties of the wild-type allele to the mutant allele. The probability (Naïve Bayes posterior probability) that this mutation is deleterious is given, and the estimate of error is also provided. Depending on these scores, the software annotates the variants as "benign", "possibly damaging", or "probably damaging".
Align- GVGD	This algorithm incorporates the biophysical characteristics of the wild-type and mutant amino acid/protein and applies the Grantham distance (Grantham, 1974) to multiple sequence alignments to predict whether the change is deleterious.
Splice Site Finder-like	Splice site detection is performed using algorithms that were developed using position weight matrices (PWM) for the relative frequencies of different amino acids in a specific sized window around known splice sites. Scores are then applied to windows of genomic sequence to determine the probability that it contains a splice site.
MaxEntScan	The 2 bp 5' (donor) and 3' (acceptor) splice sites single nucleotide variants are assessed against alternate sequence motifs using a maximum entropy distribution (MED) model to predict the likelihood of a retained splice site in the mutant sequence.
NNSPLICE (Neural network splice)	This splice site predictor is based on neural networks, inputted with a binary string that has been converted from a genomic sequence.
GeneSplicer	Potential splice sites are identified by combining Markov models that are trained to find dependencies between the coding and non-coding regions around known splice sites.
Human Splicing Finder	This model aims to predict 5' and 3' splice site and branch points using PWM [adapted from the work of Shapiro & Senapathy, (1987)] and MED [adapted from Yeo & Burge, (2004)].

Table 2.3 Description of the *in silico* tools applied at the WRGL

The following programs are used to predict the impact of variants detected at canonical splice sites: Splice Site Finder-like [adapted from (Shapiro & Senapathy, 1987)], MaxEntScan (Yeo & Burge, 2004), NNSPLICE (Reese, et al., 1997), GeneSplicer (Pertea, et al., 2001), Human Splicing Finder (Desmet, et al., 2009). Due to sub-optimal specificity of each program, a significant result is considered to be an alteration of ≥10% in ≥3 programs in either a splice donor or acceptor site, or more rarely the prediction of a new cryptic splice site, in order to use as strong supporting evidence of predicted biological impact (see

below). A 10% cut-off is recommended in multiple publications for individual splice site prediction programs (Houdayer, et al., 2012); however, a lower cut-off has been recommended for specific tools utilised at the WRGL such as MaxEntScan and NNSplice (Tang, et al., 2016), therefore a consensus is required from at least 3 programs to increase the reliability of results. At the WRGL, only variants affecting the canonical splice sites (±2 bp) are considered for variant interpretation as the most is known about mutations in these regions and *in silico* prediction tools have often been trained by sequence content in and around these regions. Whilst deep intronic variants have been reported to result in cryptic exons (or pseudoexons), the interpretation of such variants is challenging in a diagnostic setting without the use of functional assay and are therefore these variants are filtered out of the .vcf file output.

Following the publication of additional recommendations for applying the LoF (defined as PVS1) criterion of the ACMG/AMP guidelines (Abou Tayoun, et al., 2018) which presented a PVS1 decision making tree (not reproduced in this document) that was subsequently adopted into the ACGS Best Practice guidelines for constitutional variant analysis (Ellard, et al., 2019). Consequently the WRGL decided that this additional analysis should also be performed for apparent LOF variants in a somatic context to more accurately predict whether they result in a true null effect on the gene/protein. Specifically, this is applied for nonsense, frameshift, GT-AG (+/-1 or 2) splice site variants, deletions, intragenic duplications and initiation codon variants. Some of the key considerations for this decision tree are whether nonsense-mediated decay is predicted, whether the affected exon is present in biologically relevant transcripts, whether the truncated/altered region is critical to protein function and whether the variant is predicted to remove more or less that 10% of the protein. For splicing variants, the predicted impact of skipping an exon must also be considered, for example whether the deletion variant is in frame and whether the deleted sequence includes a functionally important domain.

The nucleotide and codon conservation are also considered when classifying a variant. In general, less conserved areas may tolerate DNA changes more than highly conserved regions. This piece of supporting evidence should be used in conjunction with an understanding of known protein domains. Hotspots corresponding to known gain of function mutations are recognized within oncogenes, but there is strong evidence that tumour suppressor genes also show accumulation of driver mutations in specific protein

domains (Yang, et al., 2015). To identifying protein domains, Uniprot is used (UniProt Consortium, 2020).

2.4.2.6 Collating evidence: literature

Up to date literature searches should be done with every variant interpretation with an aim to identify strong literature evidence of pathogenicity or neutrality of the specific variant identified or of variants affecting the same residues, such as variants within mutational hotspots. This should include where possible the identification of one or more publications from reputable sources providing evidence of pathogenicity. The functional studies utilised should also be assessed to determine whether they are well-established methods of *in vitro* or *in vivo* functional assessment. If two or more studies are identified of this nature with a consensus conclusion on the functional impact on the variant this can be used as stand-alone evidence of pathogenicity for a variant. Alternatively, when only one publication can be identified (even when recommended by professional guidelines) this should be used as strong supportive evidence but cannot be used as stand-alone evidence. An example of such a study is the p53 transactivation (TA) classification as determined by an *in vivo* functional assay in yeast employed by Kato, et al., (2003) to evaluate 2,314 p53 missense mutants. This paper is recommended by the European Research Initiative on CLL as a reliable reference for the functional impact of missense variants in *TP53*.

2.4.2.7 Additional considerations for suspected germline variants

Recognition that some of the variants found in cancer-associated genes routinely analysed for somatic mutations will in fact be germline in origin is imperative for informed clinical management, and the 2016 revised WHO classification of tumours of haematopoietic and lymphoid tissues now identifies 'myeloid neoplasms with germline predisposition' as a distinct entity. In the WRGL, germline samples are not analysed when samples are referred for genetic analysis from patients presenting with haematological malignancy due to technical and resource limitations. Furthermore, obtaining a truly representative germline sample not contaminated with tumour DNA from patients whose cancer affects the blood is challenging.

Despite the absence of a germline sample to accompany the tumour, hints to a potential germline, rather than somatic, origin can be obtained from the tumour sequence data. Germline variants will usually exist in a heterozygous state (unless mosaic) and will

therefore be present at an allelic frequency of approximately 50% in any tissue tested. As a general rule, any variants identified with a VAF 40-60% in a cancer susceptibility genes listed in Table 2.4 should be highlighted as having the potential to be germline [in line with the DiNardo, et al. (2018) recommendations] and, if the variant is considered (likely) pathogenic, future actions should be taken in order to confirm its origin particularly in patients <40 years of age. This list was adapted initially from the cancer susceptibility genes listed within the 100,000 Genomes project (Genomics England, 2020) and then from the NHSE national genomics test directory for cancer (NHS England, 2020).

Some germline variants are known to be associated with a syndromic phenotype or a spectrum of haematological conditions; however, for many of the samples received in the laboratory, the clinical information provided is very limited and, as the VAF alone cannot confirm the origin of the variant, efforts should be made to either sequence an alternative non-haematological sample (e.g. DNA extracted from fibroblasts, buccal cells or hair root), or a sample taken in morphological remission following treatment. The finding of a potential germline variant should prompt a discussion with the referring clinician, particularly if the patient is waiting for a stem cell transplant with a family member as a donor so that cascade testing can be undertaken, if required.

The origin of variants (somatic/germline) detected should be considered in the following genes:
CBL
ETV6
GATA1
GATA2
KRAS
NRAS
PTPN11
RUNX1
TP53

Table 2.4 List of genes found on the TSMP which are present on the list of genes included in the haematological malignancies pertinent cancer susceptibility panel

2.4.2.8 Weighting of evidence for variant classification

Once evidence has been collated in accordance with the protocol described above, the strength of the different lines of evidence is assessed and an overall conclusion about the biological classification is made. It is not possible in this document to provide an exhaustive list of all the possible combinations of outcomes. However, the strength of some of the most commonly encountered pieces of evidence are regarded as "stand-alone evidence", "strong supportive evidence" or "supportive evidence" and these are described below and summarised in Tables 2.5 to 2.7 and some examples are presented.

Stand-alone evidence (SAE)		
Description	Evidence that can be used in isolation to assign a classification of pathogenicity to a variant	
Classification	**Criteria**	**ACMG/AMP criterion from which this was adapted**
Benign	The variant detected has a minor allele frequency (MAF) >0.01 in GnomAD.	BA1
Pathogenic/Benign	Strong literature evidence of pathogenicity or neutrality of the variant or variants affecting particular residues (i.e. mutational hotspot). Strong evidence is considered as ≥2 publications from reputable sources providing evidence of pathogenicity which must include well-established *in vitro* or *in vivo* functional studies	PS3/BS3

Table 2.5 Evidence considered as strong enough to classify a variant when seen in isolation (i.e. with or without supporting evidence).

Strong supporting evidence (SSE)		
Description	Evidence that can be used in combination to provide a classification of pathogenicity to a variant[1]	
Supporting a classification of:	**Criteria**	**ACMG/AMP criterion from which this was adapted**
Likely pathogenic	Null variant in a gene where LoF is a known mechanism of disease in sporadic cancer or inherited cancer predisposition syndromes AND PVS1 decision tree consistent with predicted null effect.[2]	PVS1
(Likely) Pathogenic/ (Likely) Benign	*Some* literature evidence of pathogenicity or neutrality of the variant or variants affecting particular residues (i.e. mutational hotspot). This can include a single publication providing functional evidence that has been recommended by professional guidelines.	PS3/BS3
Likely pathogenic	Missense (novel or previously reported) variant affecting the same amino acid as a known pathogenic variant but resulting in a different amino acid change.	PS1, PM5
Likely pathogenic	≥2 publications from reputable sources demonstrating high prevalence of this variant in the disease population studied (e.g. patients with CMN) compared to population controls.	PS4, PP1
Likely pathogenic	Variant impacts a functional domain known to have high prevalence of other pathogenic mutations in a cancer setting. (NB this evidence is considered distinct from variants in well-established hotspots)	PM1
Likely benign	Variant that does not correlate with the known mechanism of pathogenicity in sporadic cancer or inherited cancer predisposition syndromes, such as: • A null variant in a gene where gain of function is a known mechanism of disease • Missense variants or in-frame deletions/insertion variants in a gene where LoF is the known mechanism of disease.	BP1, BP3
Likely pathogenic/ likely benign	The patient's clinical indication is consistent with a particular presentation that has been associated with known pathogenic variants in the affected gene	PP4
Likely benign	The MAF is greater than would be expected (but <0.01) in GnomAD across a wide range of age groups.[3]	PP4

Likely benign/ likely artefact	Low level deletion/insertion (<10% VAF) in a highly repetitive region PLUS no evidence of this variant in GnomAD PLUS no evidence or infrequently reported in COSMIC.[4]	BP3

Table 2.6 Evidence considered as strong supportive evidence requires additional supporting evidence (either SSE or SE) in order to determine a classification. If there is uncertain or multiple lines of conflicting evidence, a classification of VUS should be assigned. [1]Please note that conflicting strong supportive evidence cannot change a likely benign classification to a likely pathogenic classification or *vice versa* but there may be a scenario where the supportive evidence is so strong that the final classification is upgraded from a likely benign classification based on the stand alone criteria "MAF >0.01" to a VUS. There is unlikely to be a scenario where a likely pathogenic variant is downgraded to a VUS as the nature of stand-alone supportive evidence for pathogenicity means that independent verification of functional or clinical findings must exist from multiple groups or professional bodies in order to the evidence to be considered reliable as stand-alone.

[2]The additional criteria of predicted null effect according to the PVS1 decision tree (Abou Tayoun, et al., 2018) must be met in order for this piece of supportive evidence to be considered. [3]When variants are present at a higher frequency than expected but only in individuals <40 years old the possibility cannot be excluded that these represent individuals with an inherited predisposition to haematological disorders, especially in those genes with a known associated disorder. [4]COSMIC is not a curated database and may lack data from original study and/or variant may not have been a confirmed somatic variant. Low frequency of variants on this database may therefore represent artefactual calls submitted to this database.

Supporting evidence (SE)		
Description	Evidence that can be used in combination to provide a classification of pathogenicity to a variant; please note that: • the strength of this evidence is weaker than the SSE and therefore if any conflicting evidence present with SE alone, the variant must automatically be assigned VUS. • SE cannot be used to change a classification using SAE and/or SSE from P/LP to VUS/LB/B but other upgrades/downgrades to classification are possible.	
Supporting a classification of:	**Criteria**	**ACMG/AMP criterion from which this was adapted**
Likely pathogenic/ likely benign	Consensus in the predicted deleterious impact of the variant on the gene or gene product by *in silico* analysis*	PP3, BP4
Likely pathogenic	A single publication from a reputable source demonstrating high prevalence of this variant in the disease population studied (e.g. patients with CMN) compared to population controls	PP5, BP6

Table 2.7 Evidence considered as supportive requires additional supporting evidence (either SSE or multiple SE) in order to determine a classification of pathogenic (P), likely pathogenic (LP), variant of uncertain significance (VUS), likely benign (LB) or benign (B). If there is uncertain or multiple lines of conflicting evidence, a classification of VUS should be assigned. *Consensus defined as agreement between the output of PolyPhen, SIFT and AGVGD.

Generally, different lines of evidence are considered as weighted. Some evidence is considered strong enough to conclude the pathogenicity even without additional supportive evidence (although it is considered good policy to collect a full set of evidence for each variant even when a stand-alone piece of evidence is identified): this is termed stand-alone evidence (SAE). In the absence of SAE, supportive evidence (SE) should be considered in order to make an overall conclusion about the classification of a variant. Broadly, where there are conflicting pieces of evidence, a classification of VUS will be made. Where there are multiple lines of SE towards a pathogenic or benign classification but no definitive evidence of pathogenicity (such as the absence of functional *in vivo* analysis) a call of likely benign or likely pathogenic can be made. The assurance of this assertion is improved when there is more than one piece of concordant strong supportive evidence (SSE) or supportive evidence (SE).

2.4.2.9 Practical examples of variant classifications

In total, as of 31.01.2021, WRGL had generated 1,777 variant classifications according to the above protocol with the final classification of benign (n=9; 1%), likely benign (n=87; 5%), VUS (n=557; 31%), likely pathogenic (n=934; 53%) and pathogenic (n=187; 11%). The variants detected included non-synonymous missense (n=891; 50%), frameshift (n=458; 56%), nonsense (n=282; 16%), start-loss (n=2; <1%), stop-loss (n=4; <1%), in-frame deletion/duplication (n=56; 3%), splice site (i.e. ±2; n=78; 4%). The final classification applied to each type of variant is shown in Figure 2.8.

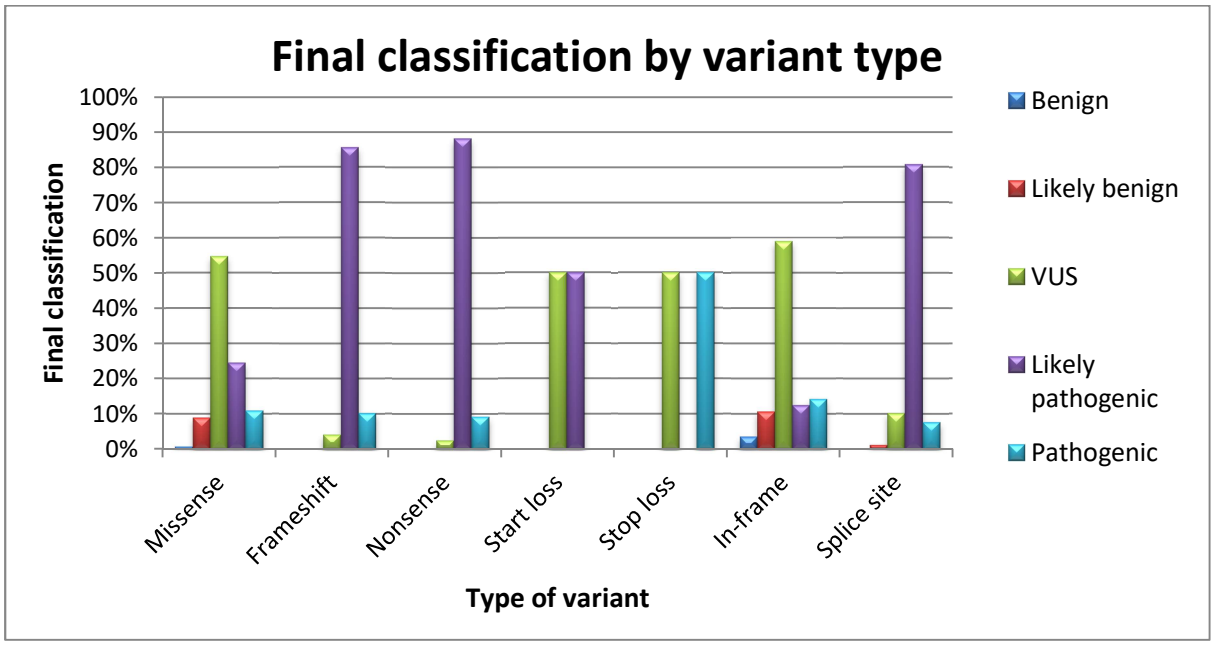


Figure 2.8 The final classification of variants interpreted at the WRGL up to 31.01.2021 (n=1,777) organised according to variant type.

Practical examples of compiled evidence, weighting of evidence and a summary of how a classification was derived are shown in Tables 2.8 to Table 2.11.

Table 2.8 shows the evidence collated for the interpretation of a missense variant [(c.1780G>A p.(Val594Ile)] detected in *BCOR* at 15% VAF. Known pathogenic variants in this gene are reported to be LoF variants i.e. nonsense, frameshift or splice site variants, therefore the clinical impact of this variant is unlikely to result in a pathogenic outcome in a somatic setting; this evidence meets the strong supporting evidence criteria for a classification of likely benign. However, the *in silico* evidence was conflicting. Furthermore, although the MAF was greater than would be expected in GnomAD across a range of age groups but less than 0.01 (MAF 0.0078% in age groups 30-65 with good overall coverage in

this region [total alleles 205,023]) leading to the application of a strong supportive evidence criteria towards a likely benign classification, there was no published evidence that this variant is particularly prevalent in CMN compared to population controls therefore no additional supporting evidence criteria was met. Overall, although there was some indication that this variant was likely benign, insufficient evidence was available to apply this classification and it was concluded that this variant was a VUS. (It is important to note that likely benign and benign variants will be filtered from all future .vcf uploaded to Alissa Interpret for review, therefore a likely benign classification is only made when there are multiple lines of concordant evidence and in the absence of this the classification of VUS will be made.)

***BCOR* (NM_001123385.1) c.1780G>A p.(Val594Ile)**		**VAF: 15%**
Databases		**Supporting evidence criteria**
COSMIC	4 entries Tissue types: haematopoietic and lymphoid, upper digestive tract FATHMM prediction: Pathogenic (score 0.95) Confirmed somatic: yes Total number of COSMIC records at this residue: 4	
GnomAD	Frequency: 0.0078% (16/205,023) **Ages**: 30-65	**Likely benign (SSE)**
ESP	No records	
ClinVar	Uncertain significance, 1 star review status (one submitter provided an interpretation with assertion criteria and evidence (or a public contact).	
Previous occurrences in WRGL	None	
Known mechanism for pathogenicity for this gene		
Truncating frameshift /nonsense/splice site (Grinfeld et al, NEJM 2018; 379:1416-1430; NCCN Guidelines MDS v 2.2020).		**Likely benign (SSE)**
In silico		
Consensus = conflicting (*In silico* from 12 species) Moderately conserved nucleotide (phyloP: 3.86 [-20.0;10.0]) Moderately conserved amino acid (considering 12 species) Small physicochemical difference between Val and Ile (Grantham dist.: 29 [0-215]) Align GVGD (v2007): C0 SIFT (v6.2.0): Tolerated (score: 0.32, median: 3.32) MutationTaster (v2013): disease causing (prob: 0.999) PolyPhen-2: Probably damaging (score 1.00)		
Literature		
None applicable		
SUMMARY		
• Two pieces of SSE for a classification of likely benign: the MAF is greater than would be expected in GnomAD across a range of age groups but less than 0.01. Known mechanism of pathogenicity for this gene is LoF but this variant is a missense change. • No experimental evidence about how this change impacts protein function • No SE (i.e. no consensus for *in silico* and no published evidence that this variant is particularly prevalent in CMN compared to population controls)		
CLASSIFICATION: VUS		

Table 2.8 The compiled evidence, weighting of evidence and a summary of how a classification of VUS was made for a variant in *BCOR*.

Table 2.9 shows the evidence collated for the interpretation of a missense variant [c.2141C>G p.(Ser714Cys)] detected in *DNMT3A* at 10% VAF. The most common pathogenic missense variant in this gene affects that residue p.(Arg882) but known pathogenic variants in this gene are reported to be LoF variants anywhere in the gene or missense variants in specific regions. The NCCN guidelines recommend that only variants affecting the following residues should be considered pathogenic p.(Gly543), p.(Arg635), p.(Ser741), p.(Arg736), p.(Arg739), p.(Ser770), p.(Met880), p.(Arg882), p.(Trp893), p.(Pro904), p.(Ala910); however, a number of alternate publications were found to indicate that missense variants within the methyltransferase domain were deleterious (Grinfeld, et al., 2018; Sandoval, et al., 2019). Furthermore, Sandoval et al. (2019) provided functional evidence that missense changes at p.(Ser714) impact protein function and there were 3 publications demonstrating a high prevalence of this variant in the disease population. These individual lines of evidence are all SSE criteria for a classification of likely pathogenic. In addition, the *in silico* evidence was concordant for a predicted deleterious impact on protein function, providing further SE for a likely pathogenic classification. Of note, although the MAF was greater than 0%, the overall frequency was much less than 0.01 MAF and all 4 individuals that were found to have this variant were >50 years old. Given that *DNMT3A* is one of the most commonly mutated genes in CHIP, it is not improbable that these detected mutant alleles representing CHIP clones in older individuals. Overall, a classification of likely pathogenic was applied. For a classification of pathogenic to have been applied in this instance, an additional piece of functional data would need to be identified; however, it is not considered a useful utilisation of resources to look for this additional functional evidence to upgrade the classification if analysts are assured with multiple lines of evidence to support a *likely* pathogenic classification.

DNMT3A (NM_175629.2) c.2141C>G p.(Ser714Cys) VAF: 10%		
Databases		**Supporting evidence criteria**
COSMIC	28 entries Tissue types: haematological and lymphoid, breast and liver FATHMM prediction: Pathogenic (score 0.95) Confirmed somatic: yes Total number of COSMIC records at this residue: 32 missense, 1 deletion	
GnomAD	Frequency: 0.0014% (4/282,850) **Ages**: 50 to 70	(Criteria for likely benign SSE not fully met)
ESP	No records	
ClinVar	Conflicting: 1 (germline) record assigned likely pathogenic according to the ACMG criteria; 1 record assigned uncertain clinical significance	
Previous occurrences in WRGL	None	
Known mechanism for pathogenicity for this gene		
Nonsense, frameshift, splice site, missense in codons G543, R635, S741, R736, R739, S770, M880, R882, W893, P904, A910 (NCCN Guidelines MDS v 2.2020). Methyltransferase domain mutations; 634-912 (Grinfeld et al, NEJM 2018; 379:1416-1430)		**Likely pathogenic (SSE)**
In silico		
Consensus = deleterious (*In silico* from 12 species) Highly conserved nucleotide (phyloP: 5.77 [-14.1;6.4]) Highly conserved amino acid, up to Zebrafish (considering 12 species) Moderate physicochemical difference between Ser and Cys (Grantham dist.: 112 [0-215]) This variant is in protein domain: C-5 cytosine methyltransferase Align GVGD (v2007): C65 (GV: 0 – GD: 111.67) SIFT (v6.2.0): Deleterious (score: 0, median: 4.32) MutationTaster (v2013): disease causing (p-value: 1) PolyPhen2: Probably damaging (score 0.996, sensitivity 0.36, specificity 0.97).		**Likely pathogenic (SE)**
Literature		
Sandoval et al. J Biol Chem. 2019; 294(13): 4898-4910 • Enzymatic activity of *DNMT3A* S714C was 2.5 fold lower than wild-type (methylation of poly(dI-dC) from at least 3 replicates, no statistical validation provided). • *DNMT3A* S714C appeared catalytically inactive with negligible DNA methylation detectible in dot blot assays (ESC purified from GFP conjugated mutant DNMT3A embryonic stem cells expressed in DKO mice). • Overall this data suggests S714C disrupts the catalytic function of *DNMT3A* No other primary studies of note to corroborate findings. There are papers identifying this variant in individual patients (see below); however no		**Likely pathogenic (SSE)**

statistical impact on disease progression or outcome was performed (likely due to small numbers). **Zhang Z-M et al., Nature, 2018, 554:387-391.** **Tatton-Brown K et al., Am J Hum Genet., 2017; 100:725-736** **Russler-Germain D et al., Cancer Cell, 2014, 25;442-454**	
SUMMARY	
• Like pathogenic SSE: a single publication providing functional evidence *plus* the location of this variant in an important functional domain as reported by multiple publications. There are also ≥2 publications from reputable sources demonstrating high prevalence of this variant in the disease population. • Likely pathogenic SE: consensus in the predicted deleterious impact of the variant on the gene or gene product by *in silico* analysis • Criteria for SAE not met	
CLASSIFICATION: LIKELY PATHOGENIC	

Table 2.9 The compiled evidence, weighting of evidence and a summary of how a classification of likely pathogenic was made for a variant in *DNMT3A*.

Table 2.10 shows the evidence collated for the interpretation of a frameshift variant [c.5618_5621del p.(Ile1873SerfsTer13)] detected in *TET2* at 42% VAF. As with *BCOR* variants, known pathogenic variants in *TET2* are reported to be LoF variants i.e. nonsense, frameshift or splice site variants, therefore one might immediately assume a classification of likely pathogenic for this variant. However, there was no record of this variant being reported previously in internal or external databases. Furthermore, assessment of this variant using the PVS1 decision tree in the ACMG/AMP recommendations for LoF variants (Abou Tayoun, et al., 2018) concludes that due to the proximity to the end of the gene, the frameshift affects <10% of the total protein and as such, nonsense mediated decay is likely to not occur. Consequently, one cannot exclude that mutant protein may be biologically available with (partial) retained function. There were no functional studies identified to provide further information about the impact of this variant on protein function. Overall, this variant was therefore classified as a VUS.

TET2 (NM_001127208.2) c.5618_5621del p.(Ile1873SerfsTer13) VAF: 42%		
Databases		**Supporting evidence criteria**
COSMIC	No records of this variant but 37 records of a variant affecting this amino acid [c.5618T>C p.(Ile1873Thr)]	
GnomAD	No records	
ESP	No records	
ClinVar	No records	
Previous occurrences in WRGL	None	
Known mechanism for pathogenicity for this gene		
Nonsense, frameshift, splice site. Plus missense: any codons 1134-1444 or 1842-1921 according to the NCCN Guidelines MDS v 2.2020; Grinfeld et al, NEJM 2018; 379:1416-1430: specify that missense mutations have to affect the following residues to be considered pathogenic T1184, I1873, R1261, C1271, R1359		
In silico		
Deletion (4 bps) in exon 11. This variation creates a frame shift starting at codon Ile1873. The new reading frame ends in a STOP codon 13 positions downstream • Not predicted to undergo NMD (final exon) • Variant removes <10% of protein • Exon present in biologically relevant transcripts **Conclusion = PVS1_moderate.**		(Criteria for SSE likely pathogenic not fully met; see below)
Literature		
None applicable		
SUMMARY		
• Criteria for SSE likely pathogenic not fully met: this variant is a null variant in a gene where LoF in a known mechanism of pathogenicity in cancer but the variant occurs close to the end of the protein, with the stop codon in final exon and removing <10% of protein. The protein is not predicted to undergo nonsense mediated decay, therefore there may be some retained function. • This variant has not been reported previously on cancer or population databases. • The variant allele frequency is 42%; cannot exclude that this is a rare germline variant with no phenotypic effect. • Criteria for SAE not met • Criteria for SE not met		
CLASSIFICATION: VUS		

Table 2.10 The compiled evidence, weighting of evidence and a summary of how a classification of variant of uncertain clinical significance (VUS) was made for a variant in *TET2*.

The example in Table 2.10 also provides a possible scenario of when other variants detected in the same sample might help classification. For example, if this *TET2* variant was the sole abnormality detected in a patient where the myeloid panel was done to exclude a neoplastic cause of idiopathic aberrant blood counts, the VAF of the *TET2* variant (42%)

would infer that the variant was either germline heterozygous or a somatic variant in almost all cells. In the absence of other supporting evidence of neoplasia in all cells of the marrow (such as clonality detected by immunophenotyping or morphology), this variant would be considered unlikely or uncertain to represent a somatic change. In the context of the affected gene being known to cause a severe constitutional disorder if germline and pathogenic, such as Noonan syndrome for *NRAS* variants, one can more confidently exclude that the variant is pathogenic, and downgrade it to a VUS or likely benign variant.

Finally, for variants in *TP53* and *RUNX1* that are considered a VUS by the WRGL variant interpretation policy and/or are within a VAF range that indicates that they *may* be germline in origin (40-60% VAF), it was decided that the ClinGen variant interpretation recommendations (Luo, et al., 2019; ClinGen *TP53* Variant Curation Expert Panel, 2019) should also be applied. It is expected that in the majority of cases, VUS will remain as this classification following re-assessment but that a minority will be able to be up- or down-graded to likely pathogenic or likely benign, respectively. An example of a VUS being upgraded to a likely pathogenic variant according to this framework (which was not published at the time of original variant assessment) is provided in Table 2.11. The NCCN guidelines are considered a reliable source of evidence for providing the mechanism of pathogenicity for genes; however, in the case of *RUNX1*, it does not provide information about which protein domains should be considered biologically important when a missense variant is identified. In the example provided below, the variant was on the edge of the RHD domain but there was insufficient literature evidence to ascertain whether this was sufficient to upgrade the variant from a VUS to a likely pathogenic variant. Re-assessment with the ClinGen criteria resulted in a classification of likely pathogenic, thus highlighting the utility of these guidelines even in a somatic context. Given that the WRGL classification criteria were framed around the consideration embedded within the ACMG criteria, it is not expected that there will complete discordancy (i.e. likely benign *versus* likely pathogenic or *vice versa* between the classifications from these two protocols.

RUNX1 (NM_001754.4) c.595G>T p.(Gly199Trp) VAF: 47%		
Databases		**Supporting evidence criteria**
COSMIC	No records	
GnomAD	No records	
ESP	No records	
ClinVar	This nucleotide change not listed, however c.596G>A (p.Gly199Glu) is listed as likely pathogenic (3 star review)	Likely pathogenic SSE
Previous occurrences in WRGL	None	
Known mechanism for pathogenicity for this gene		
Predominantly nonsense or frameshift (NCCN Guidelines MDS v 2.2020). Some pathogenic missense variants reported in the literature.		
In silico		
Consensus = Deleterious Highly conserved nucleotide and amino acid, up to C. elegans (considering 15 species) Large physicochemical difference between Gly and Trp (Grantham dist.: 184 [0-215]) This variant is in protein domains: RUNT domain Align GVGD (v2007): C15 (GV: 109.55 - GD: 86.50) SIFT (v6.2.0): Deleterious (score: 0, median: 3.40)		
Literature		
Variant **not present** in the following papers: Luo et al, Blood Advances 2019, 3(20):2962-2979; Sood et al, Blood 2017, 129:2070-2082		
SUMMARY		
• Likely pathogenic SSE: Missense (novel or previously reported) variant affecting the same amino acid as a known pathogenic variant but resulting in a different amino acid change. • This variant has not been reported previously on cancer or population databases. • The variant allele frequency is 47%; cannot exclude germline origin. • Variant in RUNT domain but insufficient literature evidence that this can be used as SAE for pathogenicity • Likely pathogenic SE: consensus for deleterious impact of the variant on the gene by *in silico* analysis • Criteria for SAE not met		
WRGL CLASSIFICATION: BORDERLINE VUS/LIKELY PATHOGENIC		
SUMMARY OF ClinGen EVIDENCE		
PM5_supp: Missense change at the same residue where a different missense change has been previously determined to be Likely PATH (c.596G>A (p.Gly199Glu)) **PS4_supp:** This patient meets the RUNX1-phenotypic criteria **PM2:** Completely absent from population databases **PP3 (supp):** REVEL score >0.7 **PM1_Supp:** Within hotspot residues p.His105-p.Arg204 within the RUNT domain. (Total= 1 moderate line of evidence, 4 supporting lines of evidence)		
CONCLUSION: LIKELY PATHOGENIC		

Table 2.11 The compiled evidence, weighting of evidence and a summary of how a classification of variant of likely pathogenic was made for a variant in *RUNX1*.

2.4.3 Stage two: Determining the clinical significance of variants

During stage one, variants identified during TSMP analysis have a biological classification assigned to them. During stage two, the clinical significance of these variants are determined and from this a decision is made about whether or not they should be reported. This stage has been adapted from Li, et al., (2017) but rather than 4 tiers, we have adopted a two tier system: clinically significant and likely to impact patient management or not clinically significant, with only the former being reported.

We consider all variants that are assigned as likely pathogenic or pathogenic as clinically significant; these variants are <u>reported</u>. The assignment of a likely pathogenic/ pathogenic classification indicates that it is thought that the variant identified is likely or is known to affect protein function. Taken alone, this classification is not an indication of whether the variant identified represents a malignant cell line or the presence of clonality (i.e. driver/passenger mutation or CHIP).

Variants assigned as benign (polymorphisms) or likely benign are <u>not reported</u>. These variants are likely to represent germline rather than somatic changes and would not affect patient management.

Variants assigned as VUS may be reported depending on the clinical context and the mutation pattern of the sample. Although the presence of a VUS in a specific gene should not be used in clinical decision making as it may represent a passenger (rather than driver) mutation, VUS will be reported as evidence of clonality in a sample with no likely pathogenic/pathogenic mutations identified, provided the VUS has a VAF of less than 40% and is therefore likely to be somatically acquired. This decision was made as result from the clinical audit which indicated that patients can be discharged from clinic without bone marrow assessment in the context of a normal TSMP result.

2.5 Results: clinical audit to assess the utility of the TSMP

2.5.1 Results returned

An audit was undertaken to understand how the TSMP was being used in clinical practice. Of 200 audit questionnaires sent out, 153 (77%) returns were received from 6 hospitals. From this cohort, the median age of patients was 66 years old (range 17-92).

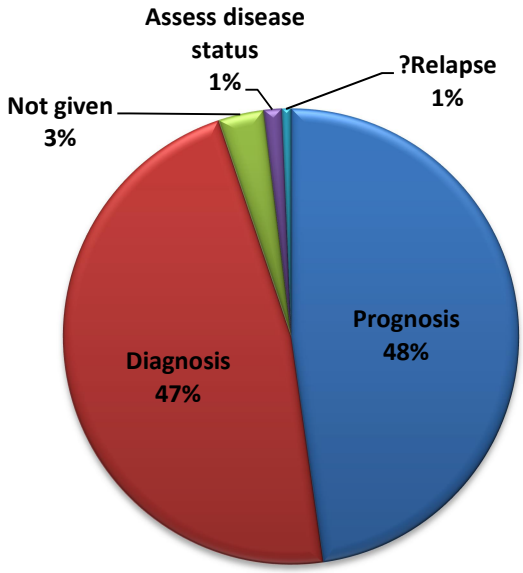
2.5.2 Previous genetic testing

Each patient had on average 2 “normal” genetic tests prior to the TSMP being undertaken, which may have included cytogenetic analysis or specific genotyping tests such as *JAK2* sequencing.

2.5.3 Primary reason for requesting a TSMP test

The majority of patients referred for TSMP testing were referred to aid diagnosis (47%) or to give further information on prognosis (48%). A small number of patients were referred to assess the disease status, such as suspected disease progression (1%), or for suspected relapse (1%) (Figure 2.9).

Figure 2.9 The primary reason for submitting a sample for the myeloid panel



2.5.4 Reason for referral: what disease?

The most common referral categories for TSMP were suspected (?) MDS (n=51), confirmed MDS (n=28), ?MPN (n=20) and known MPN (n=25); summarised in Table 2.12. Referrals were also received for ?triple negative MPN (n=2), MDS/MPN (n=2), ?MDS/MPN (n=5), AML (n=3), ?hypereosinophilic syndrome (n=6), systemic mastocytosis (n=5), ?BM involvement of sarcoma (n=1).

Referral category	?MDS	Confirmed MDS	?MPN	MPN	?triple negative MPN	MDS/MPN	?MDS/MPN	AML	?HES	SM	?BM involvement of sarcoma
Number of patients (%)	51 (34.5)	28 (18.9)	20 (13.5)	25 (16.9)	2 (1.4)	2 (1.4)	5 (3.4)	3 (2.0)	6 (4.1)	5 (3.4)	1 (0.7)

Table 2.12 The referral categories for samples referred for TSMP. HES= hypereosinophilic syndrome; SM= systemic mastocytosis; ICUS= idiopathic cytopenia of unknown significance; CCUS= clonal cytopenia of uncertain significance; CHIP = clonal haematopoiesis of indeterminate potential

2.5.5 How was the final diagnosis made in patients with suspected MDS

Respondents were asked how the final diagnosis was made in patients originally referred with ?MDS (n=51; Figure 2.9). Overall, there were 18 patients referred with ?MDS where respondents said that the final diagnosis was made using data from the TSMP. This was either in isolation (n=8) or in combination with the bone marrow aspirate (n=5), the BM aspirate and cytogenetic results (n=1), the full blood count (n=3) or the FBC plus clinical information (n=1). There were a number of patients for whom the final diagnosis was made without the TSMP result and instead was made using the following: BM aspirate (n=18), BM aspirate plus cytogenetics (n=2), clinically (n=3). No information was provided about how the diagnosis was made in 10 patients referred with ?MDS.

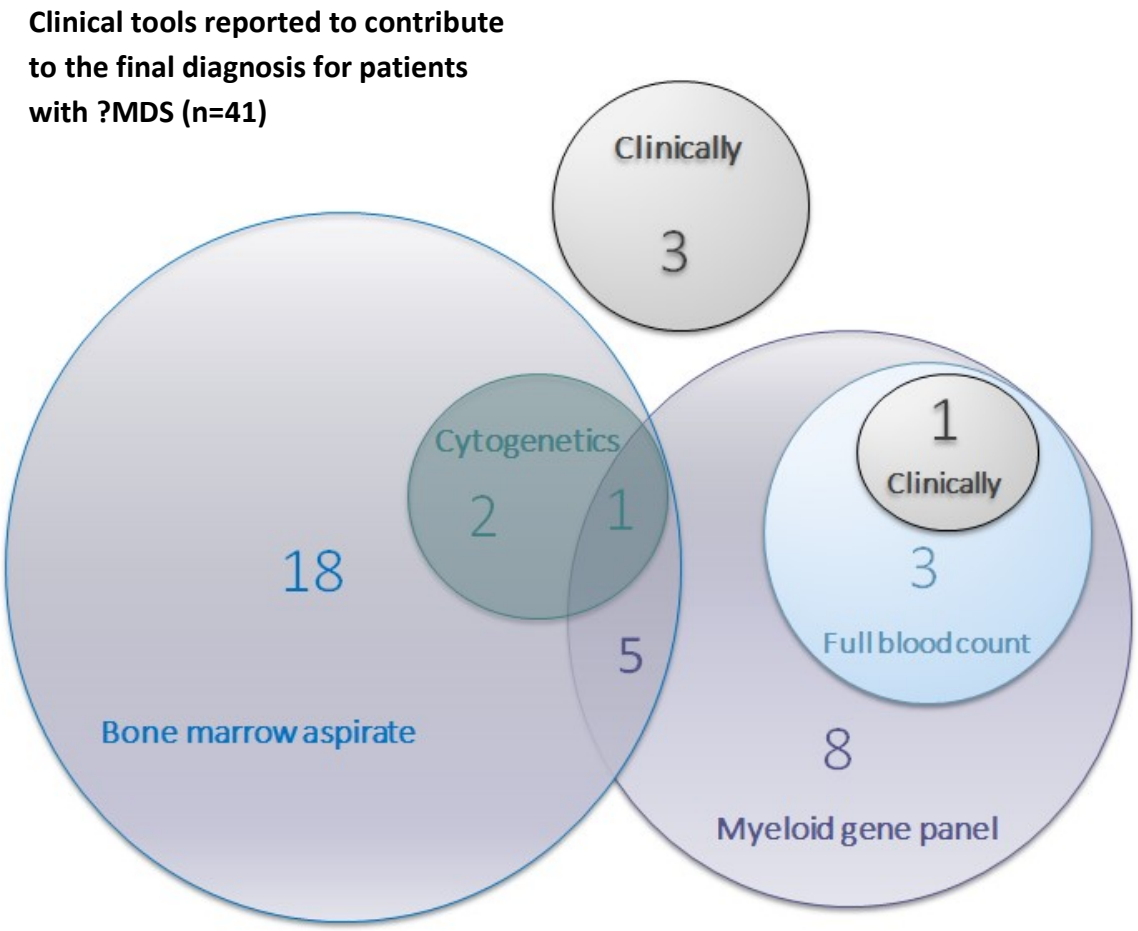
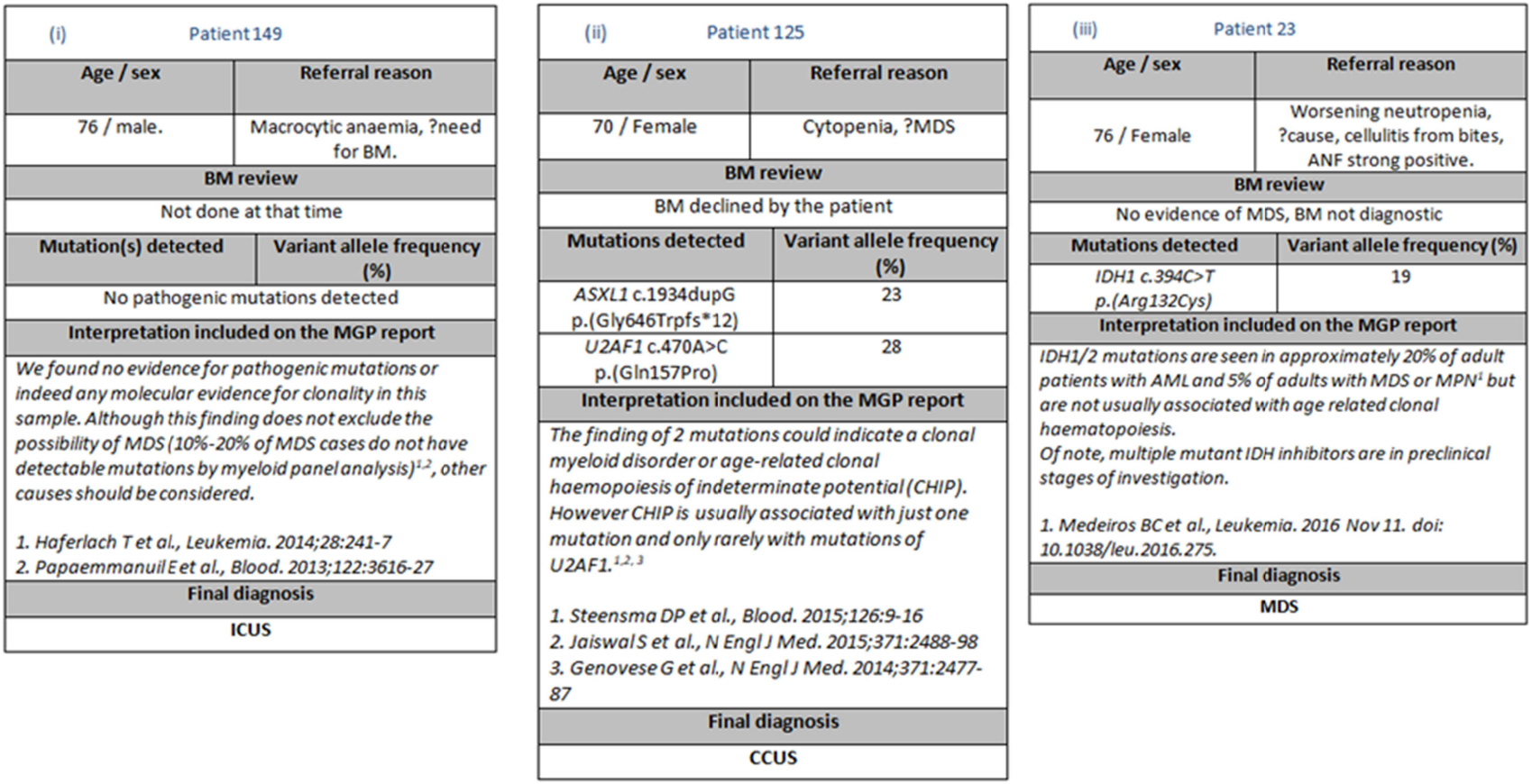


Figure 2.10 Clinical tools reported to contribute to the final diagnosis in 41 patients originally referred for TSMP testing with ?MDS. Some patients were reported to be diagnosed from one major source of information: BM aspirate (n=18), TSMP (n=8) or clinically (n=3). Some patients were diagnosed with the information from two investigations: BM and cytogenetics (n=20), BM and MGP (n=5), MGP and full blood count (n=11). One patient was diagnosed with data from the MGP, FBC and clinical information and 1 patient was diagnosed with information from the BM aspirate, cytogenetics and MGP. In total, 18 patients were diagnosed with data from the MGP. Please note that no information was returned for 10 patients originally referred for MGP testing with ?MDS.

Of the 8 patients who were given a final diagnosis based on results from the TSMP alone, the final diagnosis was: ICUS (n=1), CCUS (n=4), MDS (n=2), transient disorder (n=1); examples are shown in Figure 2.11. We had limited information returned from this audit questionnaire to ascertain how the data from the TSMP was used to make a final diagnosis.



(i) Patient 149

Age / sex	Referral reason
76 / male.	Macrocytic anaemia, ?need for BM.
BM review	
Not done at that time	
Mutation(s) detected	**Variant allele frequency (%)**
No pathogenic mutations detected	
Interpretation included on the MGP report	
We found no evidence for pathogenic mutations or indeed any molecular evidence for clonality in this sample. Although this finding does not exclude the possibility of MDS (10%-20% of MDS cases do not have detectable mutations by myeloid panel analysis)[1,2], other causes should be considered. *1. Haferlach T et al., Leukemia. 2014;28:241-7* *2. Papaemmanuil E et al., Blood. 2013;122:3616-27*	
Final diagnosis	
ICUS	

(ii) Patient 125

Age / sex	Referral reason
70 / Female	Cytopenia, ?MDS
BM review	
BM declined by the patient	
Mutations detected	**Variant allele frequency (%)**
ASXL1 c.1934dupG p.(Gly646Trpfs*12)	23
U2AF1 c.470A>C p.(Gln157Pro)	28
Interpretation included on the MGP report	
The finding of 2 mutations could indicate a clonal myeloid disorder or age-related clonal haemopoiesis of indeterminate potential (CHIP). However CHIP is usually associated with just one mutation and only rarely with mutations of U2AF1.[1,2,3] *1. Steensma DP et al., Blood. 2015;126:9-16* *2. Jaiswal S et al., N Engl J Med. 2015;371:2488-98* *3. Genovese G et al., N Engl J Med. 2014;371:2477-87*	
Final diagnosis	
CCUS	

(iii) Patient 23

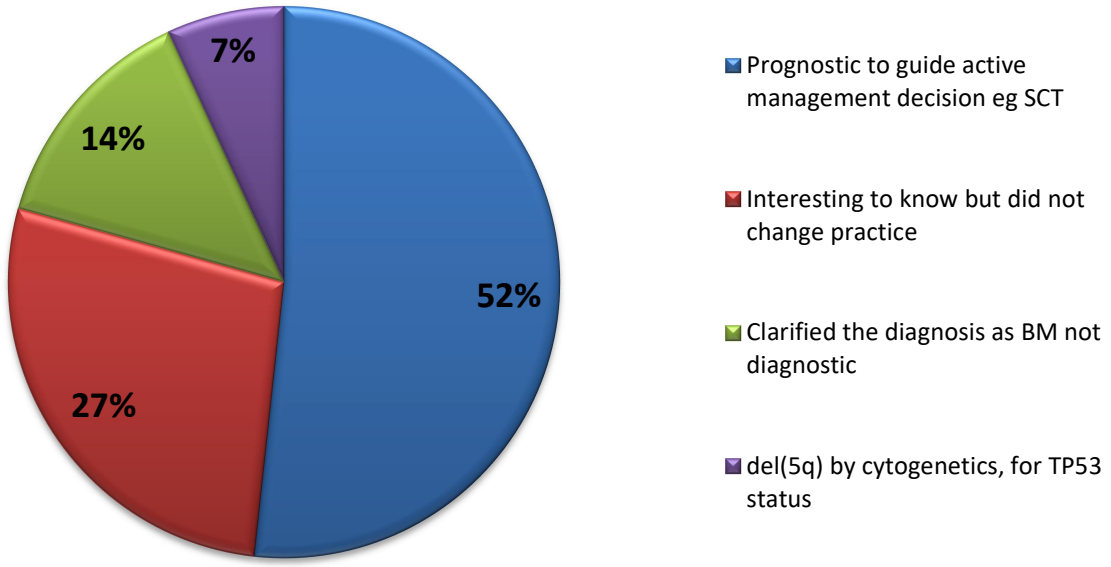
Age / sex	Referral reason
76 / Female	Worsening neutropenia, ?cause, cellulitis from bites, ANF strong positive.
BM review	
No evidence of MDS, BM not diagnostic	
Mutations detected	**Variant allele frequency (%)**
IDH1 c.394C>T p.(Arg132Cys)	19
Interpretation included on the MGP report	
IDH1/2 mutations are seen in approximately 20% of adult patients with AML and 5% of adults with MDS or MPN[1] but are not usually associated with age related clonal haematopoiesis. *Of note, multiple mutant IDH inhibitors are in preclinical stages of investigation.* *1. Medeiros BC et al., Leukemia. 2016 Nov 11. doi: 10.1038/leu.2016.275.*	
Final diagnosis	
MDS	

Figure 2.11 3 case examples patient 149, (i), patient 125 (ii) and patient 23 (iii)] of samples where respondents stated that the diagnosis was made based on the MGP result alone in patients referred originally with ?MDS.

2.5.6 How did the TSMP impact patient management in patients with confirmed MDS?

In patients with confirmed MDS (n=26), 73% (n=19) of respondents stated that the patient management was affected by the TSMP result: 52% (n=14) of analyses provided prognostic information which guided active management decisions such as SCT decisions, 14% (n=4) clarified the diagnosis as the BM was not diagnostic, 7% (n=2) further stratified risk by giving *TP53* status following the identification of del(5q) by cytogenetic analysis. The TSMP results were reported to be "interesting to know but did not change practice" in 27% (n=7) (Figure 2.12).

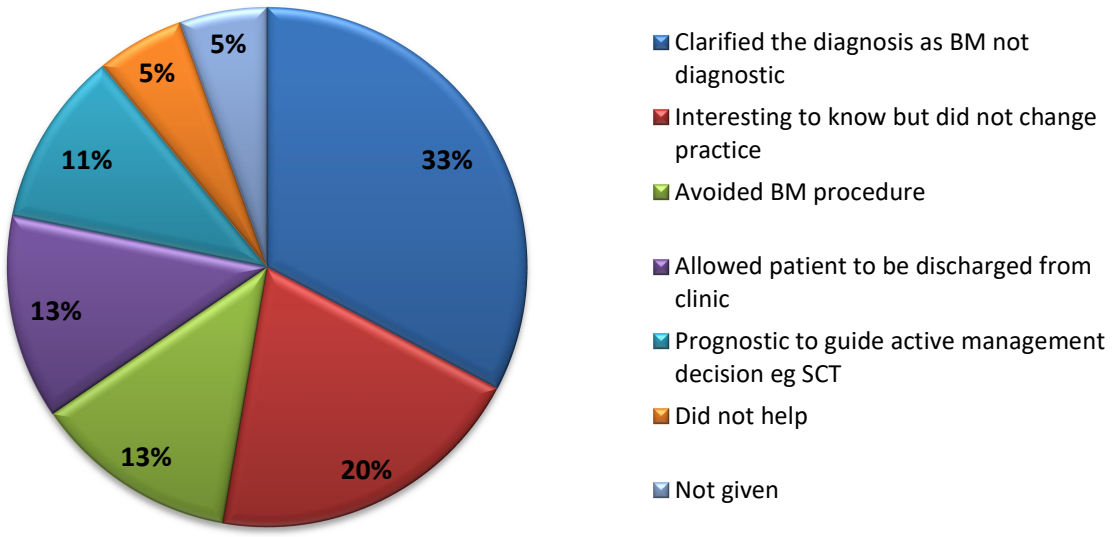
Figure 2.12 Reported changes to patient management following MGP result in patients with confirmed MDS (n=26)



2.5.7 How did the TSMP impact patient management in patients with suspected MDS?

In patients with ?MDS (n=51), 75% of respondents stated that the patient management was affected by the TSMP result (Figure 2.13): 33% clarified the diagnosis as the BM was not diagnostic, 13% allowed for avoidance of a BM procedure, 13% allowed the patient to be discharged from clinic, 11% gave prognostic information which guided active management decisions such as stem cell transplant decisions. Overall, the TSMP results were reported to be "interesting to know but did not change practice" or "not helpful" in 25%.

Figure 2.13 Reported changes to patient management following TSMP result in patients with ?MDS (n=51)



2.5.8 What was the final diagnosis in patients with suspected MDS?

Of the 51 patients referred for the TSMP with ?MDS, the final diagnosis was MDS in 29% (n=15) of cases ; Table 2.13. Other final diagnosis included reactive causes (n=4), idiopathic cytopenia of uncertain significance (ICUS; n=7), clonal cytopenia of uncertain significance (CCUS; n=4), CHIP; n=2, ?MDS (n=1), MDS/MPN (n=1), AML (n=3), aplastic anaemia (n=2), ITP (n=2), other (n=2), unknown (n=8).

Referral category	Number of patients	Final diagnosis (%)											
		MDS	Reactive causes	ICUS	CCUS	CHIP	?MDS	MDS/MPN	AML	Aplastic anaemia	ITP	Other	Unknown
?MDS	51	15 (29)	4 (8)	7 (14)	4 (8)	2 (4)	1 (2)	1 (2)	3 (6)	2 (4)	2 (4)	2 (4)	8 (16)

Table 2.13 The final diagnosis in those patients that were originally referred with ?MDS (n=51).

The clinical information provided within the returned questionnaire was reviewed by our team for the 3 patients where a final diagnosis of reactive marrow was given. All patients had no mutations identified by the TSMP and there was no evidence that this diagnosis was inaccurate.

2.5.9 Patients referred with suspected MPN

2.5.9.1 The final diagnosis in patients with suspected MPN

Of the 20 patients referred for the TSMP with suspected MPN, the final diagnosis was as follows: ET (n=2), PMF (n=2), MPN-U (n=1), SM-AHN [MPN-U] (n=1), MDS (n=1), idiopathic erythrocytosis (n=5), ?reactive (n=4), reactive (n=2) and unknown (n=2) (Table 2.14). In 2 of the 5 cases with a final diagnosis of idiopathic erythrocytosis, the bone marrow trephine and aspirate report was available and this clearly stated that the findings did not support a diagnosis of MPN.

Referral category	Number of patients	Final diagnosis (%)								
		ET	PMF	MPN-U	SM with ANH	MDS	Idiopathic erythrocytosis	?reactive	reactive	Unknown
?MPN	20	2 (10)	2 (10)	1 (5)	1 (5)	1 (5)	5 (25)	4 (20)	2 (10)	2 (10)

Table 2.14 The final diagnosis in those patients that were originally referred with ?MPN (n=20).

2.5.9.2 The mutations detected by TSMP in patients with suspected MPN

In these 20 cases, the number of likely pathogenic/pathogenic variants and VUS detected by the TSMP in was 0 (n=15; 75%), 1 (n=2; 10%) 2 (n=1; 5%), 3 (n=1; 5%), 4 (n=1; 5%). In those cases where no variants were detected, the final diagnosis was ET (n=1), MPN-U (n=1), MDS (n=1), idiopathic erythrocytosis (n=5), reactive/?reactive (n=5) and unknown (n=2). For a final diagnosis of ET, MPN-U and MDS, the diagnosis was made from the bone marrow aspirate/trephine review; when a final diagnosis was reported as idiopathic erythrocytosis, the diagnosis was made clinically (n=4) or by exclusion of other diagnoses (n=1). In those cases where one to 4 variants were detected by the TSMP, the final diagnosis was ET (n=2), PMF (n=2), SM with AHN (n=1) and ?reactive (n=1). The variants detected are given in table 2.15.

Variants detected by TSMP analysis in patients referred with ?MPN						
Sample name	**Number of variants detected**	**Pathogenic/likely pathogenic variants detected**		**VUS detected**		**Final diagnosis**
		Variant	**VAF**	**Variant**	**VAF**	
W1704722	1	None detected		*ASXL1* c.2957A>G p.(Asn986Ser)	44%	ET
W1703993	3	*DNMT3A* c.2644C>T p.(Arg882Cys)	46%	*GATA2* c.1168A>G p.(Lys390Glu)	43%	PMF
		SRSF2 c.284C>G p.(Pro95Arg)	20%			
W1702961	4	*SRSF2* c.284C>G p.(Pro95Arg)	29%	None detected		PMF
		SETBP1 c.2602G>A p.(Asp868Asn)	23%			
		ASXL1 c.1762C>T p.(Gln588*)	28%			
		RUNX1 c.485G>A p.(Arg162Lys)	14%			
W1705084	2	*KIT* c.2447A>T p.(Asp816Val)	38%	None detected		SM with AHN (MPN-U)
		ASXL1 c.2468T>G) p.(Leu823*)	37%			
W1701455	1	None detected		*BCORL1* c.1330A>C p.(Thr444Pro)	12%	?Reactive

Table 2.15 Variants detected in 5 cases that were originally referred for TSMP with ?MPN once benign/likely benign variants were excluded. The final diagnosis is provided for each case. The remaining 15 cases of ?MPN had no variants detected by TSMP. Transcripts: *ASXL1* NM_015338.5; *BCORL1* NM_021946.4; *DNMT3A* ; *KIT* NM000222.2; *RUNX1* NM_001754.4; *SETBP1* NM_015559.2; *SRSF2* NM003016.4.

2.5.10 Impact of the TSMP on patient management

Of the 153 respondents, 12% (n=18) stated that the results for the TSMP results allowed for avoidance of a bone marrow procedure; the median age of these patients was 55 years old (range 18-85 years). Whilst one of these patients was reported to have very mild anaemia, the haemoglobin levels for the remaining cases at the time that the TSMP referral was made

ranged from 100-180g/dL suggesting that this group of patients do not only represent patients with very mild anaemia.

The TSMP result from 16 patients showed no evidence of clonality (i.e. no mutations were detected); in 1 patient, 1 mutation was identified but as the patient had very mild anaemia we were informed that the clinician decided this would be monitored though regular clinical assessment (Table 2.14). Finally, 1 patient had 2 mutations and we were informed that the final diagnosis in this patient was CCUS; no further information was provided about why this result allowed for avoidance of a bone marrow procedure (Table 2.15).

Patient **127**	
Age / sex	**Referral reason**
70 / Male	Very mild anaemia
BM review	
BM not taken	
Mutations detected	**Variant allele frequency (%)**
EZH2 c.2051G>A p.(Arg684His)	53
Clinical outcome following MGP result	
The clinician reported that they decided not to do a BM procedure, but the patient would be routinely monitored in light of *EZH2* mutation identified.	

Table 2.14 Clinical details on the single patient where one mutation was identified, and the clinician reported that a BM was not undertaken following review of the MGP result on a peripheral blood sample.

Patient **9**	
Age / sex	**Referral reason**
78 / Male	Persistent cytopenia, progressive neutropenia, thrombocytopenia, no excess of blasts,?MDS
BM review	
BM not taken	
Mutations detected	**Variant allele frequency (%)**
SF3B1 c.2098A>G p.(Lys700Glu)	5
TET2 c.5103_5104delGCinsTT p.(Met1701_Gln1702delinsIleTer)	22
Clinical outcome following MGP result	
The MGP resulted in a diagnosis of CCUS; no further information provided.	

Table 2.15 Clinical details on the single patient where two mutations were identified, and the clinician reported that a BM was not undertaken following review of the MGP result on a peripheral blood sample.

Respondents stated that the TSMP result was interesting to know but did no change practice in 26% (n=40) patients. The final diagnosis in these patients was as follows: CHIP

(n=1), CCUS (n=2), MDS (n=14), MPN (n=11), MDS/MPN (n=3), AML (n=5), aplastic anaemia (n=1), hyper eosinophilia syndrome (n=1), SM (n=1), idiopathic erythrocytosis (n=1).

Respondents stated that the TSMP result allowed discharge from clinic in 8% (n=12) patients. Seven of these patients had had a BM taken for review, 3 patients had not had a BM taken and the information was not provided for 2 patients. All 12 patients had no mutations detected by the TSMP. The final diagnosis was provided for 6 of these patients: reactive cause (n=3), ?reactive cause (n=2), ITP (n=1).

2.5.11 Reported usefulness of the TSMP in patients with confirmed or suspected CMN

We asked respondents to score how useful the TSMP was from 1 (not useful) to 10 (very useful), 150 results were returned for this question with the majority of clinicians scoring the results as useful: 10, very useful (n=43), 9 (n=40), 8 (n=26), 7 (n=11), 6 (n=15), 5 (n=7), 4 (n=4), 3 (n=3), 2 (n=0), 1, not useful (n=1).

2.6 Discussion

In the absence of national guidelines for somatic variant interpretation for patients with haematological neoplasia, I describe in this Chapter the development of a robust protocol that can be applied systematically in a diagnostic setting for variant filtering, variant classification, and reporting of results from a myeloid panel assay (the Illumina® TSMP) for patients with confirmed or suspected CMN. Using practical examples, I have shown how this protocol enables objective classification of variants to inform patient management for patients with CMN. The protocol has been designed so that it can be applied to any gene panel and referral reason and is therefore not specific for the TSMP or patients with CMN. This standardisation allows for robust record keeping for each sample analysed, whilst maintaining the safe standards required for laboratory testing in an NHS laboratory.

This protocol was developed over time, starting from a strategy first conceived in 2018 for the interpretation of results from a *TP53* molecular assay for patients with CLL. It was then adapted to be suitable for the interpretation of results from multiple genes in a range of CMN, informed by a number of models presented in different publications.
Firstly, the process was informed by the ACMG/AMP (Richards, et al., 2015) and subsequently the ACGS (Ellard, et al., 2019) recommendations for variant interpretation in a constitutional setting, most notably in that the 5 tier classification system was adopted for assigning pathogenicity to variants (from benign to pathogenic, including VUS) rather than the assignment classifications relating to the "actionability" of mutations detected, as recommended by Li et al., (2017). It was felt that this was an important differentiation between the needs of variant interpretation in the context of haematological neoplasia *versus* solid tumours which was the focus of Li et al., (2017). Primarily this is because solid tumours are often diagnosed in a clinical setting using a combination of histopathological criteria and imaging, and molecular genetic testing is most commonly applied after diagnosis to identify aberrations that could be targeted therapeutically, such as testing for *EGFR* mutations in patients with non-small cell lung cancer to determine eligibility for tyrosine kinase inhibitor therapy such as gefitinib. Conversely, as described in depth in the preceding sections, CMN can be difficult to diagnose and genetic testing has become a key part of diagnosing and classifying these patients prior to the consideration of the most appropriate treatment strategies. In the context of identifying a mutational pattern that

would indicate a person has MDS, for example, the terms “actionable” or “targetable” may not necessarily apply unless the variants detected are themselves targetable by drugs. However, that is not to say that mutations identified by the TSMP for patients with CMN cannot be targetable, such as the identification of an *IDH1/IDH2* mutation in a patient with MDS or secondary AML. By incorporating a second stage of variant interpretation into the classification protocol, namely the categorising the variant as clinically significant or not, we are able to independently determine which variants should and should not be reported. This classification of clinical significance was adapted from Li, et al., (2017); however, a noticeable difference is that our protocol allows for scenarios where VUS may be reported to inform patient management. For example, by reporting a clearly somatic VUS as the sole abnormality detected by the TSMP in patient with an unconfirmed diagnosis of haematological neoplasia may prevent the referring clinician from discharging the patient on the basis of suspected reactive causes for their aberrant blood counts. A clinical audit undertaken to investigate the clinical utility of the TSMP in managing patients with CMN provided evidence that this has occurred in practice. In most other scenarios, the WRGL protocol for somatic variant assessment recommends that VUS are not reported because their significance might be misinterpreted.

The identification of VUS by gene panel analysis cannot be avoided and this can be particularly difficult when they impact a gene known to be prognostically significant in a number of different disorders, such as *RUNX1*. In an effort to reduce the number of VUS classifications at the WRGL, the ClinGen variant interpretation recommendations (Luo, et al., 2019; ClinGen *TP53* Variant Curation Expert Panel, 2019) were also adopted for the following specific scenarios: variants in *TP53* and *RUNX1* that are considered a VUS by the WRGL variant interpretation policy and/or are within a VAF range that indicates that they *may* be germline in origin (40-60% VAF). I have provided an example above of how this additional classification protocol can be helpful in these scenarios. We have had no instances of complete discordancy (i.e. likely benign *versus* likely pathogenic or *vice versa* between the classifications from these two protocols) which is not unexpected given that both protocols are adapted from the ACMG/AMP framework. However, in order to ensure that laboratory staff are using their time in the most effective way, we intend to review the policy of additional ClinGen protocol to assess what proportion of cases result in a change in classification using this method. Given that the aim of the myeloid panel is to identify *somatic* changes but it is also important to not miss any clinically relevant germline changes

in *RUNX1* or *TP53* that could impact patient management for the patient or their family members, an alternative option could be to apply this classification scheme to only those patients that are under 40 and more likely to have a germline disorder. It is worth noting that our laboratory was able to adopt these additional criteria due to close working relationships with our colleagues in the constitutional teams working at the WRGL; haemato-oncology laboratories with no connection to clinical scientist expertise in germline variant interpretation may find the adoption of these criteria challenging due to lack of experience.

Challenges also exist more generally for variants that may be germline in origin when applying a gene panel to samples from patients with a haematological malignancy. The TSMP is designed for testing the tumour sample only (i.e. blood or bone marrow sample) with no subtraction from a germline sample. The variant interpretation guidelines presented here take this into account and the finding of a potential germline variant in a defined list of genes included on the TSMP that are known to be strongly associated with germline haematological predisposition [adapted from the NHSE national genomics test directory for cancer (NHS England, 2020)] and recommends discussion with the referring clinician, particularly if the patient is waiting for a stem cell transplant with a family member as a potential donor so that cascade testing can be undertaken, if required.

Since the development of the WRGL's protocol for somatic variant interpretation, Belgium's ComPerMed group published a standardised framework for variant classification for solid and haematological tumours (Froyen, et al., 2019). This publication provides evidence that other centres are also applying a multistep approach to variant classification and, like the WRGL, they firstly apply a biological classification then to apply a clinical classification and only those variants that are considered clinically relevant will be reported. One of the main differences between our approach and the ComPerMed group is that the ComPerMed group recommend the inclusion of a semi-quantitative scoring system for interpreting missense variants. One of the scoring parameters is the frequency of variants in COSMIC. Further work to define when the frequency of variants in COSMIC becomes significant would be a particularly beneficial addition to the current WRGL protocol to act as a more robust method of determining whether a variant is more common in cancer *versus* the general population, given that some variants can be detected in GnomAD that may represent CHIP or an undiagnosed haematological neoplasia. Furthermore, variants in some genes may represent changes that have occurred in stem cells and therefore seen at slightly

higher frequency but do not themselves represent driver mutations and/or may be more commonly see in CHIP, therefore being able to apply a gene specific measures of significance to frequency would also be useful.

A limitation to the current WRGL protocol for somatic variant assessment include the staff resources required to collate evidence for assessment and review for a final classification in a meaningful timeframe. Improvements to this system would include automated collation of evidence and automation to the classification of pathogenicity. Examples of this have been published, such as the Variant Interpretation for Cancer (VIC) tool (He, et al., 2019) presented for the semi-automated application of the AMP/ASCO published guidelines for the interpretation of somatic variants in cancer (Li, et al., 2017). In addition, there is now a publically available automated classification tool to interpret the very strong pathogenic rating (PVS1) according to the updated recommendations by the ClinGen sequence variant interpretation working group (Xiang, et al., 2020). As it stands, there are no automated tools that could be applied for the WRGL protocol and it is likely that before automated tools can be used in a diagnostic laboratory setting, multi-centre beta testing and ratification would be required according to a nationally agreed framework.

Another limitation of this protocol as it is currently presented is that it has not been ratified by other diagnostic laboratory centres. Whilst the WRGL participate in EQA schemes for myeloid panel analysis, this pilot scheme requires that laboratories perform their panel on an external DNA samples and report only those variants that are considered to be clinically significant; this data is then collated and shared between participants for review. This scheme design reflects the differences in practices that exist between laboratories with respect to panel content, testing and analysis policies and acts to provide a broad overview of the capability of each laboratory's assay to detect variants and perform some form of variant interpretation. A nationally agreed framework, much like the ComPerMed workflow mentioned above, would be beneficial in the UK in that it would act to improve the equity of service provided by different diagnostic laboratories between the GLHs in England. In 2018, myself and colleagues at the WRGL joined a national somatic variant interpretation working group to try to define a shared recommendation document but the outcomes of this working group have been delayed due to the conflicting pressures of changes to clinical practice following the onset of the COVID-19 pandemic. In the interim period the WRGL has already commenced a process of sample swaps (both DNA and .vcf files) between

laboratories within the same GLH (Oxford and Birmingham) to assess the key differences between practice with an intention to create a unified approach.

It is my belief that an additional important outcome of this working group should be the concerted effort to initiate more active knowledge sharing between GLHs. For example, mandating that laboratories are required to routinely upload their data into open source (or curated) databases for professional access between GLH would act to provide a larger wealth of information for variant interpretation in the future. Such interaction has already begun between the individual genetic laboratories within GLHs as gene panels have become more standardised in order to comply with the NHS test directory to prevent duplication of work. Once professional guidelines are generated, these data will allow for prospective assessment of how easily applied they are between GLH, allowing updating versions to be developed in response to the identification of the causes of variation. National guidelines will be even more important as WGS enters clinical practice, with the potential for even greater variance in interpretation of results between centres.

In this Chapter, I also described the outcome of a clinical audit looking at how the TSMP can alter patient management with CMN. Overall, results indicated that clinicians found this panel very useful in the management of patients with CMN. In patients with confirmed MDS (n=26), the majority (73%) of respondent said that the patient management was affected by the TSMP result through additional prognostic information, guiding management decisions or clarifying diagnosis as the BM was not diagnostic. In patients referred for TSMP testing with suspected MPN (n=20), pathogenic/likely pathogenic variants were detected in all patients with a final diagnosis of a neoplastic condition (ET, PMF, MPN-U, MDS, or SM with AHN) whereas all patients with a final diagnosis of idiopathic erythrocytosis (n=5) and 5 out of 6 cases with a final diagnosis of reactive/suspected reactive marrow had no evidence of pathogenic/likely pathogenic variants or VUS. The remaining case was also given a final diagnosis of a suspected reactive marrow, and no variants were reported in the diagnostic report (due to reporting policy at the time), however a VUS was detected which indicated the presence of clonality and it would be interesting to know how this might have changed patient management and/or the final diagnosis in this patient. Overall, the TSMP results were reported to be “interesting to know but did not change practice” in 27%; as knowledge is gained about the prognostic impact of findings from this panel over time this group is likely to shrink.

Of note, in one case where the TSMP result informed the final diagnosis, the mutation detected was *JAK2* V617F in a patient with ET. There was also one patient diagnosed with SM and the mutation identified was *KIT* D816V which is almost invariably present in adults with SM. In general, where there is a differential diagnosis of a myeloid neoplasia strongly associated with a particular mutation, the appropriateness of gene panel testing should be questioned as a front line test as targeted testing is often cheaper and a result can be obtained more quickly.

The results also indicated that the TSMP is not being used primarily to determine patient's response to treatment in this cohort; however, we have noticed a change in referral patterns since the completion of the clinical audit. In particular, the WRGL now receives multiple samples over time from the same patients, with a request to monitor the presence of mutations detected previously or to look for the evidence of clonal evolution. Whilst this assay is very insensitive compared to MRD analysis by qPCR, for example, my experience of multi-disciplinary team (MDT) meetings to discuss these results indicates that a high level approach to monitoring mutations in this manner is proving to be clinically useful, especially when patients present with a mutational pattern that does not include a single primary abnormality (such as a *BCR-ABL1* fusion in patients with CML) but rather displays a complex pattern of sub-clonal mutations which may change over time as treatment is applied or as the disease progresses. It is expected that specific error-corrected panels will be introduced into routine practice in due course to enable more sensitive detection of MRD in patients with CMN, and particularly in patients with AML (Jongen-Lavrencic, et al., 2018).

In addition, from this audit, there was no evidence that the TSMP was being used inappropriately to manage patients with confirmed or suspected myeloid neoplasia. It had been a particular concern from some of our clinical colleagues that inexperience with panel results and complications such as CHIP might lead to inappropriate diagnosis of a CMN in some cases, but this concern was not borne out. However, there were examples of clinicians discharging patients on the basis of a negative TSMP result from a peripheral blood sample without evidence of repeat testing with a BM aspirate. Although this is not currently recommended practice, there is evidence that there is a high degree of concordance between the results from paired PB and BM samples from the same patient by clinical NGS testing. For example, a recent publication reported 98.9% concordance between variants detected by a 95-gene blood cancer panel in 164 paired PB and BM samples (Lucas, et al.,

2020). Some of the discordances reported between sample types were thought to largely be caused by subtle differences in the allelic burden of low level sub-clonal variants and regions of low coverage in one of the two samples. In order to reduce the chance of discordances, laboratories should thus play close attention to regions of poor coverage which may impact the limit of detection in specific regions, and consider supplemental testing when the genomic region is of particular clinical significance (e.g. a mutational hotspot).

3 The significance of low level *JAK2* V617F mutations

3.1 Introduction

3.1.1 *JAK2* structure and function

The gene *JAK2* (OMIM * 147796) is a member of the *janus* kinase (Jak) family of non-receptor tyrosine kinases (TKs) (OMIM, 2017). TKs are enzymes that phosphorylate tyrosine residues on signal transduction molecules thereby triggering signalling cascades (Vlahovic & Crawford, 2003). *JAK2* is involved in cytokine receptor signalling which regulates the proliferation and differentiation of haematopoietic cells and is essential for erythropoietin receptor (EpoR) and thrombopoietin receptor (TpoR; encoded by the *MPL* gene) signalling, which in turn are critical for red cell and platelet production (Stanley, 2009). Due to a lack of intrinsic catalytic activity within many cytokine receptors, such as the EpoR (Figure 3.1), receptor-associated Jak TKs are required for signal transduction following ligand binding (erythropoietin [Epo] in this example) and directly phosphorylate tyrosine residues in the cytokine receptors (Saharinen, et al., 2000). Once phosphorylated, these residues provide a docking site for downstream signalling molecules, such as Signal Transducers and Activators of Transcription (STATs) which are also phosphorylated by Jaks and subsequently translocate to the nucleus to elicit downstream transcription of target genes (Gnanasambandan & Sayeski, 2011). Jak-Stat signalling is highly regulated and is essential for myeloid cell development, proliferation and survival, and for the initial stages of immune response (Gnanasambandan & Sayeski, 2011; Bandaranayake, et al., 2012).

Jak proteins contain seven conserved Jak homology domains (named JH1-7) encoding four major functional domains (figure 3.1): (i) an N-terminal Band 4.1, ezrin, radixin, moesin (FERM) domain responsible for the association of Jak proteins with cytokine receptors; (ii) a Src homology-2 (SH2)-like domain, the precise role of which is unclear; (iii) a C-terminal tyrosine kinase domain (JH1), which is highly conserved and contains the activation loop, primary phosphorylation sites (Tyr1007 and Tyr1008) and the ATP binding site (Lys882); and (iv) a pseudo-kinase domain (JAK homology-2, JH2) which regulates the activity of JH1 (Bandaranayake, et al., 2012).

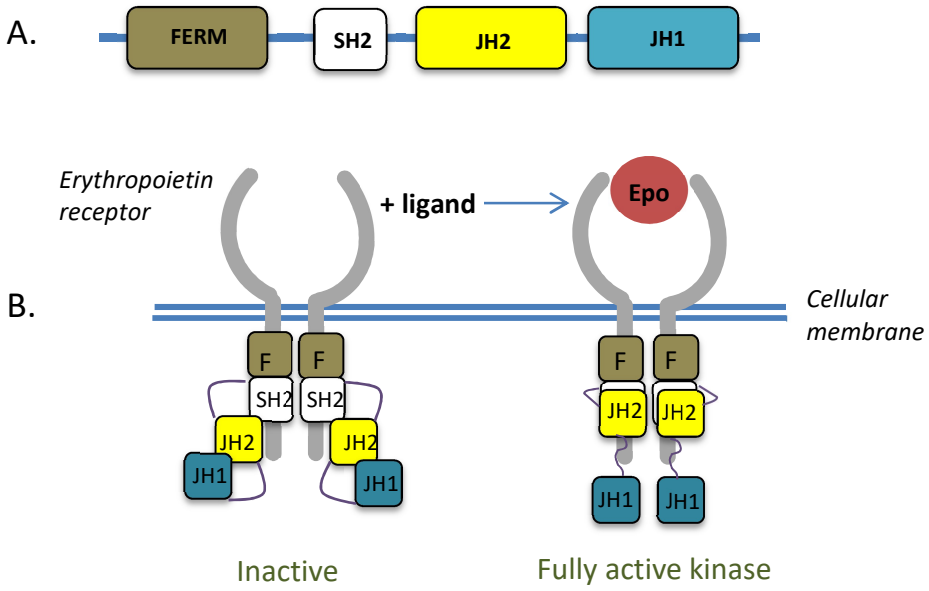


Figure 3.1 A. The Jak2 protein functional domains of the Jak2 protein: an N-terminal Band 4.1, ezrin, radixin, moesin (FERM) domain, an Src homology-2 (SH2)-like domain the precise role of which is unclear; a C-terminal tyrosine kinase domain (JH1) and a pseudo-kinase domain (JAK homology-2, JH2) which regulates the activity of JH1. **B.** Model of the interactions between the domains of the Jak2 protein bound to the erythropoietin (Epo) receptor. Image adapted from Skoda*, et al*. (2015).

3.1.2 Mutations of *JAK2* in MPN

JAK2 signalling requires strict regulation in order to stop the signalling process at the right time following activation and limit phosphotransferase activity in the absence of ligand activation (Gnanasambandan & Sayeski, 2011). Deregulation of *JAK2* via a mutation in this gene or other genes implicated in the Jak-STAT signalling pathway are reported to occur commonly in MPN and infrequently in number of other myeloid malignancies such as "atypical" MPNs, MDS and AML (Steensma, et al., 2005; James, et al., 2005; Patnaik, et al., 2010; Engle, et al., 2015; Spivak, 2017). This indicates that dysregulation of the Jak-Stat signalling pathway is a hallmark of MPN pathogenesis.

The *JAK2* p.(Val617) residue is within the JH2 pseudokinase domain which normally acts to inhibit JH1 kinase domain function; when *JAK2* contains the V617F mutation, the inhibitory influence is diminished, most likely by interfering with the mechanisms by which the pseudokinase domain negatively regulated the catalytically active kinase domain (Gnanasambandan & Sayeski, 2011). Of note, *JAK2* V617F homozygosity is seen in both PV and PMF (although infrequent in ET) and occurs through the mechanism of acquired uniparental disomy (aUPD) for chromosome 9p (Tefferi, et al., 2009). In the heterozygous

state, *JAK2* V617F-bound receptors can still respond to growth factors, but when in the homozygous state, these receptors can act autonomously from growth factors. Mutations (insertions or deletions) within *JAK2* exon 12 have also been reported in a further 5% of patients with PV (Scott, et al., 2007). *JAK2* exon 12 codes for the interface of the JAK2 SH2 and JH2 domains and when mutated, leads to constitutive activation. Activating mutations in *JAK2* (and also *MPL*, discussed further below) have been reported outside of these hotspot regions in a small proportion of ET and PMF patients (2% and 10%, respectively) (Milosevic Feenstra, et al., 2016). Recently, *JAK2* exon 13 indels have been described as drivers of occasional cases of primary eosinophilia (Patel, et al., 2019), and inherited weakly activating *JAK2* mutations are seen in rare cases of hereditary MPN-like disorders (Bellanné-Chantelot, et al., 2020).

3.1.3 Other mutations in MPN, including *CALR* and *MPL*

As mentioned above, somatic mutations have also been reported in *MPL,* which encodes for the TpoR, and in calreticulin (*CALR*), a multifunctional protein involved in glycoprotein folding, calcium homeostasis and cellular functions such as proliferation, phagocytosis and apoptosis (Spivak, 2017). In MPNs, somatic *MPL* mutations occur in exon 10, most commonly the residue p.(Tyr515) but also at p.(Ser505), and result in a conformational change of the receptor leading to activation of *JAK2* in the absence of Tpo stimulation. Mutations in *MPL* occur in 4% of patients with ET and 8% of patients with PMF and are uncommonly reported to co-occur with *JAK2* mutations (Spivak, 2017). Mutations have been rarely reported to occur outside of these hotspot regions in patients with ET or PMF (Milosevic Feenstra, et al., 2016), but do occur in occasional families with hereditary thrombocytosis (Bellanné-Chantelot, et al., 2020). Like *JAK2* mutations, mutations in *MPL* are not completely specific to MPNs and have been reported in other forms of myeloid malignancies (Patnaik, et al., 2010).

Calreticulin is a multifunctional protein that acts as a molecular chaperone and plays an important role in calcium homeostasis. *CALR* mutations reported to occur in MPNs include a wide range of mutations, most of which result in 52 bp deletions (type 1) or 5 bp insertions (type 2) within exon 9, generating a +1 frameshift. The resultant mutant protein has a novel highly-charged C-terminal which is obligatory for its transforming activity (Nangalia, et al.,

2013). CALR is reported to bind to a number of different proteins containing *N*-glycosylated residues; however, mutant CALR has been shown to result in specific (constitutive) activation of the TpoR and thereby activate downstream JAK2 signalling. *CALR* mutations are not seen in other cancers driven by mutated receptors that are *N*-glycosylated (EpoR, EGFR, KIT, for example) (Chachoua, et al., 2016). Thrombotic events in PV and ET are the leading cause of morbidity and were originally thought to occur in *CALR* mutation positive patients at a frequency of less than half of that seen in *JAK2* V617F positive patients. However, more recent data from a large cohort of ET patients (n=1053) indicates that this incidence of thrombosis is more significantly impacted by clinical characteristics such as age, sex and previous history of thrombosis than by *JAK2/CALR* mutation status (Finazzi, et al., 2014).

Mutations in a number of additional genes are thought to modify and enhance the effects of the "phenotypic driver mutations" in MPN (i.e. mutations in *JAK2*, *CALR* and *MPL*) and these account for the remaining mutations identified in MPNs. Such genes include *TET2*, *DNMT3A*, *ASXL1*, *EZH2*, *U2AF1, SH2B3* and *SF3B1* and are primarily associated with signal transduction or epigenetic regulation, tumour-suppressor or splicing proteins (Table 1.5; Chapter 1); mutations in these genes may be referred to as "disease-modifying". Whilst mutations in some of these genes (*RUNX1, TP53, IKZF1*, *CUX1*) tend to be acquired in individuals with MPN at the time of leukaemic transformation, others (such as mutations within *DNMT3A*) may be present prior to the acquisition of a phenotypic driver mutation thereby creating a "fertile ground" that is more conducive to the development of myeloid neoplasms and do not specifically promoting disease progression (Corces-Zimmerman, et al., 2014).

Cooperation between the "driver" mutations and the "disease-modifying" mutations in MPN has been demonstrated by Ortmann, *et al*. (2015) who genotyped stem cells and progenitor cells isolated from haematopoietic colonies and reported that "the order in which *JAK2* and *TET2* mutations were acquired influenced clinical features, the response to targeted therapy, the biology of stem and progenitor cells, and clonal evolution in patients with MPNs". For example, patients in whom the *JAK2* mutation was acquired first had a greater likelihood of presenting with PV than with ET and their *JAK2*-mutated progenitors showed an increased sensitivity to targeted therapy, ruxolitinib, *in vitro.*

3.1.4 The clinical utility of molecular genetic testing in MPNs

In order to undertake appropriate clinical management of patients presenting with haematological features such as those described above, it is important to distinguish between reactive (benign) and neoplastic (clonal) processes in the bone marrow and therefore accurate diagnosis is imperative. Classical MPNs can be diagnosed by the identification of driver mutations *JAK2, CALR*, and *MPL* and there is some evidence that the mutant allele burden can provide prognostic information (discussed below). Importantly, the mutation status of these genes cannot be used to define a particular subtype of MPN in the absence of clinical and/or morphological features and their absence does not exclude the presence of an MPN. However, as discussed above, the co-occurrence of driver mutations can further complicate decisions in the diagnostic setting.

The mutation burden of somatic mutations within MPNs can range from 100% to very low levels (<1%). The presence of low level mutations may represent one of two separate models of clonality: (i) a small mutated clone in the background of a larger population of non-mutated cells or (ii) a small subclone derived from a larger clone with overlapping genetic abnormalities (which may or may not have been fully characterised). Alternatively, apparently low level mutations may in fact represent technical artefact and are therefore false representations of clonality in an individual (Figure 3.2). For those low level mutations identified in patients with suspected or confirmed MPN that are proven to be real, there is still a lack of data regarding the clinical significance of these findings (Kouroupi, et al., 2012), and in particular the difference between scenarios (i) or (ii) above. Despite this, the presence of somatic mutations in MPN have proven clinical utility including distinguishing between clonal (neoplastic) and reactive (secondary) haematopoiesis and providing prognostic information and these scenarios are discussed below.

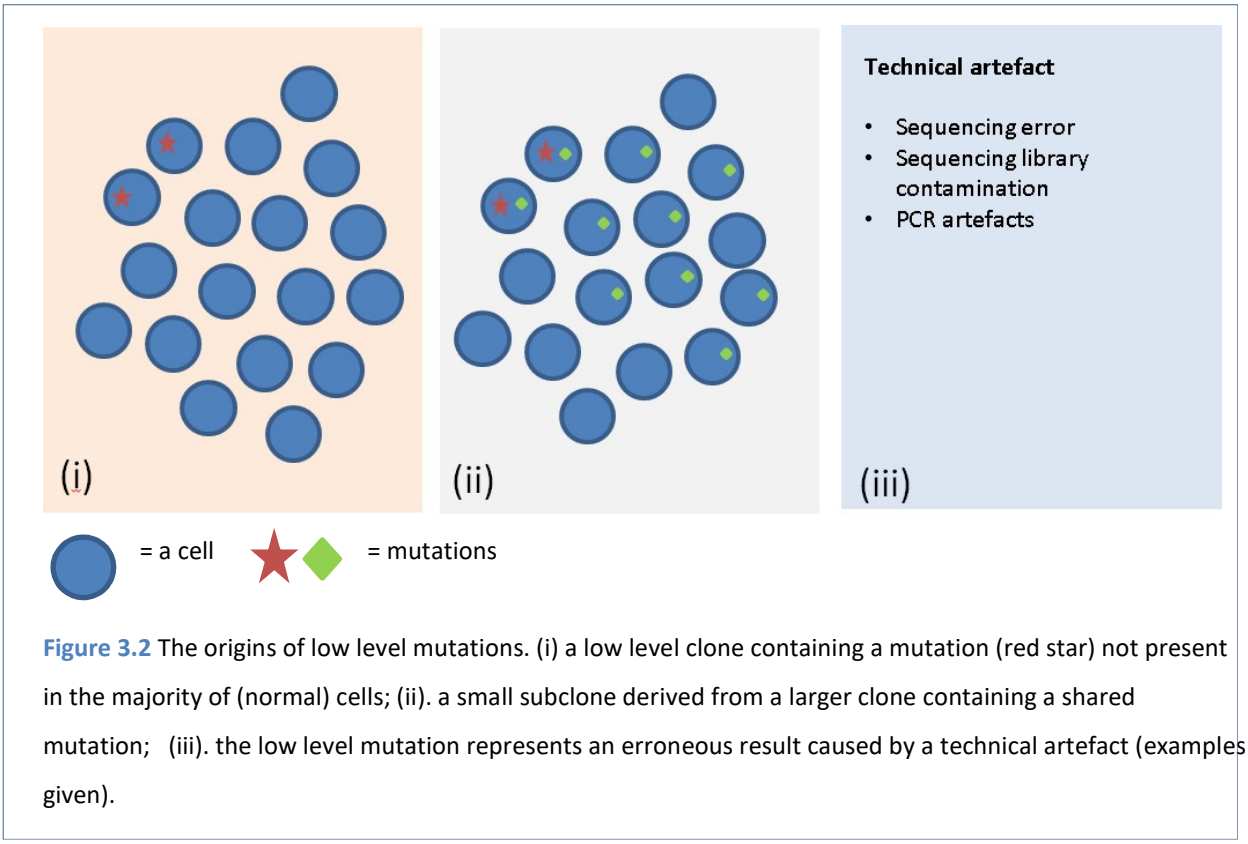


Figure 3.2 The origins of low level mutations. (i) a low level clone containing a mutation (red star) not present in the majority of (normal) cells; (ii). a small subclone derived from a larger clone containing a shared mutation; (iii). the low level mutation represents an erroneous result caused by a technical artefact (examples given).

3.1.4.1 Distinguishing between clonal (neoplastic) and reactive (secondary) haematopoiesis using molecular genetic tests

The identification of *JAK2, CALR* and *MPL* mutations can define the presence of a clonal disorder in an individual. PV and ET can be difficult to distinguish from secondary (non-clonal) erythrocytosis and thrombocytosis, respectively, and both neoplasms have a risk of thrombo-haemorrhagic complications, vasomotor disturbances, pruritis and a risk of disease progression into MPN-BP or MF. By identifying patients with PV and ET, they can be given therapy to prevent thrombotic complications (Tefferi, et al., 2011). The identification of mutations in *JAK2* exon 12 can also be helpful in excluding reactive erythrocytosis as they are found only in patients with PV, and particularly cases with features of erythrocytosis and supressed erythropoietin with non-specific bone marrow morphology (Scott, et al., 2007). However, whilst *JAK2* mutations are present in a large proportion of patients with PV, ET and PMF, there is a general lack of disease specificity based on the presence of this mutation alone (Gong, et al., 2013).

Furthermore, one could argue that the detection of a single driver mutation at low level cannot be used as definitive evidence that the patient has a neoplastic condition in the

context of borderline clinical features of MPN as the acquisition of somatic mutations as we age is now a well-established phenomenon in apparently normal individuals (CHIP, discussed in section 1.5) (Jaiswal, et al., 2014; Genovese, et al., 2014). Overall, in the largest cohorts studied, *JAK2* V617F has been reported in the general population at a prevalence of 0.1% to 3.1% depending on the LoD of the assays used (Nielsen, et al., 2011; Nielsen, et al., 2014; Hinds, et al., 2016, Cordua, et al. 2019), which equates to at least ten-times the prevalence of diagnosed MPNs. In one longitudinal study, the majority of *JAK2* V617F mutation positive individuals identified had a mutation burden <10% and the mutation burden appeared stable at the time of re-examination up to 9 years later (Nielsen, et al., 2014). Overall, there is a paucity of information in the literature about how to differentiate between CHIP and MPN when a low level driver mutation such as *JAK2* V617F is detected.

3.1.4.2 Deriving prognostic information from molecular genetic tests

Several phenotypic associations have been made between the driver mutation and disease phenotype. A number of groups have reported that the level of *JAK2* V617F mutation burden can have prognostic value in PV and PMF. Using a prospective study of 338 patients with PV, Passamonti, *et al.* (2010) identified a significant risk of developing myelofibrosis (P=0.029) in those patients with a *JAK2* V617F mutant allele burden of >50%. Conversely, in 129 patients with PMF and known *JAK2* status, Tefferi, *et al.* (2008) identified a significantly shortened overall (P=0.0008) and leukaemia-free (P=0.01) survival in those individuals with a *low JAK2* V617F mutant allele burden; these findings were replicated in an independent study published later (Guglielmelli, et al., 2009). Of note, both studies identified evidence of anaemia, leukopenia and/or thrombocytopenia rather than myeloproliferative phenotype in these patients suggesting the possibility of an undetected and presumably larger *JAK2* V617F negative clone which is more biologically aggressive and likely to be driven by driver mutations in other genes, i.e. model (ii) in Figure 3.2. Whilst these findings might be clinical useful for patient management, Tefferi, *et al*. (2011) suggested that a lack of standardisation regarding the quantitation of *JAK2* V617F mutation burden between diagnostic laboratories undermines the translation of this measurement into clinical practice at present.

Phenotypic and prognostic correlations have also been associated with individuals with *CALR* mutations. In patients with ET, *CALR* mutations are associated with higher platelet, lower haemoglobin levels, lower leukocyte counts and younger age at presentation; furthermore, the incidence of thrombotic events is lower in *CALR*-mutated ET when

compared to *JAK2* V617F-positive ET. Finally, a higher rate of progression to MF has been reported in patients with type 1 (52 bp deletion) versus type 2 (5 bp insertion) *CALR* mutations. In *CALR*-mutated PMF, patients present at a younger age, with higher platelet and leukocyte counts, reduced anaemia and are more commonly transfusion dependent when compared with *JAK2* V617F positive patients with PMF but they are also reported to have a better overall- and leukaemia-free survival (Nangalia, et al., 2013; Rumi, et al., 2014; Zoi & Cross, 2017).

Further clinical correlates have been reported relating to the somatic mutational events in "phenotypic driver" and "disease-modifying" genes in patients with MPN and their association with overall survival (OS) and risk of transformation to MPN-BP. For example, Lundberg, *et al.* (2014) used targeted NGS to assess the mutation status of 104 cancer-related genes in 197 patients with MPN and analysed the associations with clinical outcome; they reported that the presence of 2 or more somatic mutations significantly reduces the OS and increases the risk of transformation. These findings supports the hypothesis that alternative "driver mutations" drive poor prognosis neoplasia in the patients with a low level *JAK2* V617F mutation identified by Tefferi, *et al*. (2008) and Guglielmeli*, et al*. (2009). However, the mutations in "disease modifying" genes were individually at low frequencies and therefore their individual functional and prognostic significance was difficult to assess.

3.1.5 Co-occurring driver mutations in MPN

Immediately following the discovery of *CALR* mutations in *JAK2* V617F negative patients with PMF, it was thought that these two mutations were mutually exclusive. However, Tefferi, et al., (2014) was first to report the co-occurrence of these mutations (*JAK2* V617F+/CALR+) in single patient with PMF in a cohort of 254 PMF patients being studied to identify clinical and molecular sub-groups within this disease entity. Shortly after, the *JAK2* V617F+/CALR+ mutation pattern was reported to co-occur in both ET (Lundberg, et al., 2014) and PV, where *CALR* mutations are not commonly reported (Xu, et al., 2015; Xing, et al., 2016). Several studies have noted the co-occurrence of these mutations, and also the occasional co-occurrence of *JAK2* V617F with activating *MPL* mutation, *KIT* D816V, and *BCR-ABL1*.

Of those studies where co-mutation of MPN driver mutations are reported, Lim, et al., (2015) found the highest frequency of *JAK2* V617F+/CALR+ patients in their analysis of 92 Taiwanese patients with ET. They used high resolution melt (HRM) analysis to identify *CALR* mutations (reported sensitivity of 2.5% VAF) and allele specific PCR (AS-PCR) to identify *JAK2* V617F mutations (reported sensitivity of 5% VAF) in this cohort. Fifty-nine (64%) patients had a *JAK2* V617F mutation and of these, 16 (17%) patients were also shown to have a *CALR* alteration by HRM. Upon variant interpretation of the *CALR* alterations detected only 4 (4%) in fact result in the pathogenic "+1" frameshift (the remaining were a combination of truncating, missense and in-frame deletion variants); all four pathogenic variants were independently confirmed by either TA-cloning or Sanger sequencing. In these 4 patients, the *JAK2* V617F mutation was 7-83% VAF (median 22%); the VAF of the *CALR* mutations was not reported but 3 out of 4 samples were undetectable by Sanger sequencing indicating that they were low allelic burden.

Although the exact frequency of co-occurrence of driver mutation in Ph-negative MPN is unknown, estimates for *JAK2* V617F and *CALR* mutations is thought to be <1% of MPN cases (Ahmed, et al., 2016). Of note, Lim et al (2015) report the frequency of *JAK2*/*CALR* co-mutations to be 4% (when variants that are not "+1" pathogenic variants are excluded). It is likely that testing strategies employed by diagnostic laboratories (discussed in more detail below) do not usually pick up these cases since few laboratories test for *CALR*, *MPL* and *JAK2* V617F driver mutations in all cases Ph-negative MPN.

JAK2 V617F and *MPL* mutations have also been detected in MF and ET patients in a number of independent studies (Guglielmelli, et al., 2007; Beer, et al., 2008; Papaemmanuil, et al., 2013; Jang, et al., 2020). Guglielmelli, et al., (2007) identified a *MPL* mutation in 18 out of 217 (8.2%) MF patients; 4 (18%) of these patients also had a co-occurring *JAK2* V617F mutation identified by AS-PCR. Beer, et al., (2008) identified a single patient with co-mutation *JAK2* V617F and *MPL* in a cohort of 200 patients with MF or ET, and Pardanani, et al., (2006) identified this co-mutation in 6 out of 1182 (0.5%) patients with MF or ET, indicating that these co-mutations are relatively infrequent.

In the most extensive study on co-occurring driver mutations, the French intergroup of MPN (FIM) study characterised 47 patient identified across 4 French centres (Mansier, et al., 2018). Of these patient, 40 had ET, 5 had PMF, 1 had PV and 1 had MDS/MPN-RS-T; 68% were *JAK2* V617F+/*CALR*+, 23% had a *JAK2* V617F and a *MPL* mutation (*JAK2* V617F+/*MPL*+),

4% had a *CALR* and a *MPL* mutation (*CALR*+/*MPL*+), 2% (1 patient) had a *JAK2* V617 mutation and a co-occurring *JAK2* exon 12 mutation, and 2% (1 patient) with two independent *CALR* mutations. They compared the outcome of ET patients with double-mutations (DM-ET) to the outcome of "classical" ET with a single mutation (SM-ET) or triple-negative ET (TN-ET). The DM-ET patients were significantly older (median age 72 versus 61, P<0.001), especially when comared TN-ET patients (median age 56, P<0.001). For clinical comparisons, the DM-ET were compared to age-matched SM-ET: results showed that the haemoglobin and platelet counts did not significantly differ between SM(*CALR/MPL*)-ET and DM-ET but that lower haemoglobin and higher platelet counts were seen in DM-ET *versus* SM(*JAK2* V617F+)-ET. The frequency of thrombotic events and evolution to MF or acute leukaemia did not differ between DM-ET and SM-ET. Splenomegaly was seen more frequently in *JAK2* V617F+/*CALR*+ *versus JAK2* V617F+/*MPL*+ patients (17% versus 0%; n.b. that not all data was available for the *JAK2* V617F+/*MPL*+ cohort). This group previously reported that *JAK2* V617F+/*CALR*+ is more frequently encountered in patients with low *JAK2* V617F allelic burdens (Mansier, et al., 2016) and this was replicated in the analysis of allelic burden of *JAK2* V617F mutations in DM-ET versus SM-ET (median 1% vs. 26%, P<0.001). These findings led the group to suggest that low levels of *JAK2* V617F represents a CHIP clone with subsequent acquisition of an MPN driver mutation leading to the development of disease (Mansier, et al., 2018). This is supported by the increased age in DM-ET patients.

Given the scarcity of reports of co-mutated driver mutations, there is limited information about the clinical impact of this finding in Ph-negative MPN; however, the published literature is summarised below.

JAK2 V617F and CALR co-mutated

Kang, et al., (2016) examined 167 Korean patients with ET for the presence of *JAK2* V617F and *CALR* mutations: 4% (n=7) were *JAK2* V617F+/CALR+. The *JAK2* V617F mutational burden was very low for all *JAK2* V617F mutations in this group: mean 0.21% VAF; range 0.1-0.38% VAF, consistent with the findings of the FIM study (Mansier, et al., 2016). In the FIM study, *CALR* mutations were found in 2 out of 1895 (0.5%) patients with a *JAK2* V617F mutations between 5% and 99% VAF and in 19 out of 227 (14.3%) of patients with a *JAK2* V617F mutation at 0.01% to 4% VAF. Follow up was available from 7 of these patients between 1 and 101 months after the initial diagnostic testing was undertaken and the VAF of both the *JAK2* V617F and the *CALR* mutation stayed stable over time in all but 1 patient.

In this latter patient, the VAF of the *JAK2* V617F mutation dropped from 2.5% to 0.3% whilst the VAF for the *CALR* mutation rose from 43% to 50%, indicating that the mutations occurred in separate clones. For the remaining 6 patients, the sub-clonality was not investigated.

Kang, et al., (2016) also studied the clinical characteristics of *JAK2* V617F+/*CALR*+ patients when compared with other mutational patterns: *JAK2* V617F+/*CALR*-, *JAK2* V617F-/*CALR*+ and *JAK2* V617F-/*CALR*-). Overall, the *JAK2* V617F+/*CALR*- patients had a higher leukocyte count, haemoglobin level and higher frequency of thrombotic events when compare to the *JAK2* V617-/*CALR*- group. Infact, *JAK2* V617F positivity was an independent indicator of poor progression free survival (PFS) irrespective of *CALR* mutation status in this cohort. However, they found no significant differences between the *JAK2* V617F+/*CALR*+ group and other groups except that the white blood cell counts were lower when compare to the *JAK2* V617F+/*CALR*- group. Only 7 (4.2%) of the total 167 patients with ET progressed to MF or acute leukaemia during the monitoring period of 160 months and they were all in the *JAK2* V617F+/*CALR*- group. [Of note, Kang, et al. (2016) identify one PV patient with a *CALR* mutation but no *JAK2* V617F mutation down to the limit of sensitivity of the assay used (<1% VAF).]

A number of independent studies report the presence of single instances of co-occurrence of *JAK2* V617F and *CALR* mutations without significant analysis on the clinical impact of this finding (Tefferi, et al., 2014; Al Assaf, et al., 2015; McGaffin, et al., 2014; Xu, et al., 2015; Xing, et al., 2016; Ha & Kim, 2015). Of note, none of the major studies referenced here significantly addressed the issue of bi-clonality *versus* co-occurrence of mutations within the same clone, although Lundberg, et al., (2014) suggest that at least in a proportion of cases both mutations are acquired due to clonal evolution in a shared common progenitor with an epigenetic mutation, i.e. a mutation that affects the innate control mechanisms of DNA expression but does not alter the DNA sequence content.

JAK2 V617F and MPL co-mutated

Both *JAK2* V617F mutations and *MPL* mutations result in gain of function leading to promotion in the activation of STAT signalling, predominantly by aberrant and sustained phosphorylation of the JAK2 protein (Ahmed, et al., 2016). The observation that *MPL*

mutations occur in ET and PMF but not in PV suggests that *MPL* mutations favour megakaryocytic lineage differentiation as opposed to erythroid differentiation. Conversely, although *JAK2* V617F mutations can occur in PV, ET and PMF, the observations that JAK2 V617F over-expression in murine models results in a PV-like phenotype without thrombocytosis (James, et al., 2005) and that loss of heterozygosity in *JAK2* V7617F mutated patients occurs more frequently in PV (and PMF) than in ET (Akada, et al., 2014) , suggests that this mutation favours erythroid lineage differentiation at least when signalling levels are high.

In a study of 1182 patients with myeloid disorders (MPN and MF), Lasho, et al. (2006) performed *JAK2* V617F and *MPL* 515 genotyping and identified 6 patients with both a *JAK2* V617F and a *MPL* mutations. In all cases, the VAF was lower for the *JAK2* V617F variant and was only detectable due to the sensitive AS-PCR assay employed; the *MPL* variants were all >50% VAF. All patients presented with either ET or MF, suggesting the *MPL*-induced drive for megakaryocytic differentiation and thrombocytosis predominated in disease development over *JAK2* V617F-induced erythrocytosis; however, this group did not independently compare the haemoglobin levels in these patients compared to *MPL*-only mutated patients.

Lasho, et al. (2006) were unable to ascertain the cell of origin of the low level *JAK2* V617F mutations and as such were unable to comment on whether they represented secondary abnormalities in the *MPL*-mutated clone or independent clones; however, further studies were performed on multiple samples over a 4-8 year period from three patients to assess the VAF of these mutations over time. Results showed that the low level *JAK2* V617F and the high level *MPL* mutations were both present at diagnosis, indicating that they occurred as early events in the disease evolution. Furthermore, the allelic burden of these mutations stayed relatively stable over time, even in the presence of clinical and/or cytogenetic evolution in 2 patients. In one patient, the *JAK2* V617F allelic burden was shown to be higher in granulocyte-derived DNA; however it was not possible to determine whether the *JAK2* V617F –positive clone was independent from the *MPL*-positive clone or whether they shared a common progenitor clone (Lasho, et al., 2006). Nevertheless, these results do not support the notion that low level *JAK2 V617F* has a distinct role in the pathogenesis in all sub-types of MPN. More broadly, these data could be used as evidence that the VAF should

be taken into account when considering the predicted phenotype/disease course of MPN with co-occurring driver mutations.

In a separate study, Guglielmelli, et al., (2007) performed *JAK2* V617F and *MPL* testing in 217 MF patients; 18 (8.2%) had a *MPL* mutation, and of these, 4 (22%) also had a *JAK2* V617F mutation. Analysis of *MPL*-mutated versus *MPL*-unmutated patients showed that haemoglobin levels were significantly lower in *MPL*-mutated patients and this was not affected by the presence of *JAK2* V617F. This result suggests that the negative impact of the *MPL* mutation on erythropoiesis, which results in severe anaemia, is not "salvaged" by the presence of a *JAK2* V617F mutation in these co-mutated patients with MF. Again, the cell of origin of both mutations was not clarified in this cohort of patients.

A number of independent studies report the presence of single instances of co-occurrence of *JAK2* V617F and *MPL* mutations without significant analysis on the clinical impact of this finding (Beer, et al., 2008; Jang, et al., 2020). Of note, the case reported by Jang, et al. (2020) was *JAK2* V617F positive and *MPL* mutated in the bone marrow but the patient had previously tested negative for a *JAK2* V617F mutation in the blood, indicating that sub-clonal differences in cellular compartments within an individual may lead to some co-mutated individuals remaining undetected.

3.1.6 Timing of acquisition of *JAK* V617F mutations

The timing of acquisition of MPN driver mutations, the dynamics of their expansion and the biological and environmental factors required for clonal growth and the point at which this results in disease presentation is largely unknown. As discussed in Chapter 1, somatic mutations in driver genes have been detected in normal individuals (i.e. individuals without overt CMN presentation) and are reported to be more prevalent with age (Jaiswal, et al., 2014) with only a minority of these individuals going on to develop an overt MPN.

The timing of acquisition and the rate of clonal expansion of the *JAK2* V617F mutation has been studied using PCR based NGS (LoD $\geq$0.8% VAF) by McKerrell, et al., (2017) in 12 patients with MPN who had previously donated blood samples to the Cyprus Bone Marrow Donor Registry. The median time of blood donation prior to MPN diagnosis was 10.2 years. In 9 patients, the *JAK2* V617F mutation was detected in the historical samples and the rate of expansion was estimated to be 0.36% to 6.2% per annum. In this small cohort, the 46/1 T/C (rs12343867) haplotype was investigated and the T>C polymorphism was detected in

42% (10/24) alleles, with a homozygous/hemizygous change (likely due to aUPD) detected in the four MPN patients with the most rapidly expanding clones suggesting that this haplotype, particularly when hemizygous due to aUPD, may influence the rate of expansion of these clones.

3.1.7 Best practice guidelines for genetic testing in MPN

Recurrent mutations in *JAK2, CALR* and *MPL* are recognised hallmarks of *BCR-ABL1* negative MPN as indicated by their inclusion within the 2016 World Health Organisation (WHO) classification of haematopoietic neoplasms (Arber, et al., 2016). As *JAK2, CALR* and *MPL* mutations are not specific for MPN, these findings must not be used in isolation to diagnose MPN but incorporated into a diagnostic pathway. In spite of this, the identification of mutations within these genes (thus identifying clonality) can be clinically very useful in the scenario of possible reactive haematopoiesis. Laboratories can utilise a range of different techniques to investigate mutations within the recognised "phenotypic driver" and "disease modifying" genes, including commercially available kits and laboratory-developed assays, and results from different methods may vary (Verstovsek, et al., 2006; Cankovic, et al., 2009). Below I summarise a number of published guidelines available that make recommendations about the clinical utility of and laboratory approaches to genetic testing. Please note that whilst still relevant, these guidelines are widely considered to be out of date in places as they do not incorporate many of the more recent applications of mutation detection in MPN, such as the use of NGS gene panels to detect mutations in high molecular risk genes (*ASXL1, EZH2, SRSF2* and *IDH1/2*) in patients with PMF to add value to existed combined molecular and clinical prognostication systems (Guglielmelli, et al., 2014).

3.1.7.1 *JAK2*, *CALR* and *MPL* analysis at MPN diagnosis

In 2012-2013, the Association for Molecular Pathology published "laboratory practice guidelines for detecting and reporting *JAK2* and *MPL* mutations in myeloproliferative neoplasms" (Gong, et al., 2013); a UK MPN working group also published guidelines on this topic in 2013 (Bench, et al., 2013). Both groups recommend screening for *JAK2* mutations when clinical indicated by the presence of unexplained polycythaemia, neutrophilia, or thrombocytosis or splanchnic vein thrombosis and state that *MPL* mutation analysis is not required in PV as they do not occur in this disorder but is recommended in ET and PMF. A

diagnostic algorithm has been proposed by Tefferi and Pardanani (2014) in concordance with these recommendations and also incorporating the use of *CALR* mutation detection in ET and PMF, first reported in 2013 to occur in ET and PMF but not PV (Nangalia, et al., 2013)

3.1.7.2 *JAK2*, *CALR* and *MPL* analysis following treatment

Several papers have been published regarding the use of (i) high-resolution melt and Sanger sequencing or (ii) sensitive qPCR in order to monitor the mutant allele burden of *JAK2* V617F and *JAK2* exon 12 mutations with a detection limit of (i) 10-20% and (ii) ≥0.1% mutated alleles, respectively (Laughlin, et al., 2010; Carillo, et al., 2011; Kjaer, et al., 2012). Gong, *et al*., (2013) state that the clinical utility of follow-up testing is not currently established, not least because there is evidence that small molecule JAK2-inhibitors affect growth and viability of affected cells but do not generally target the specific *JAK2* mutations and therefore do not directly impact the mutation burden. Furthermore, although there is evidence that major and complete molecular response has been achieved in patients with *JAK2* exon 12 mutant ET or PV who were treated with interferon-α (Kjaer, et al., 2012), Gong, *et al*. (2013) argues that there is insufficient evidence of direct clinical correlations to the mutant allele burden and clinical response. Nevertheless, the clinical utility of sensitive quantitative *JAK2* V617F monitoring in patients with PMF following SCT has been established and provides information on the rate of disease eradication, overall survival and the risk of relapse (Lange, et al., 2013; Langabeer, et al., 2014). The monitoring of *CALR* and *MPL* mutation burdens in this context may also prove to be clinically useful and there is evidence to show that the eradication and persistence of these mutations mirrors results from the status of donor chimaerism testing (Langabeer, et al., 2014; Wolschke, et al., 2017).

Bench*, et al*. (2013) conclude that the use of *JAK2* testing for minimal residual disease (MRD) monitoring is still valuable despite marked differences in test performance between laboratories as reported by the European LeukemiaNet study group (Jovanovic, et al., 2011). Part of the reason for this conclusion is that the presence or absence of *JAK2* V617F after SCT is predictive of relapse rather than a specific threshold of positivity (Alchalby, et al., 2010). However, they also comment on the potential for false positivity caused by technical artefacts such as cross-reactivity of primers or probes and therefore recommended thorough validation to determine the sensitivity and specificity of the assay in use.

3.1.7.3 Laboratory methods for detecting *JAK2*, *CALR* and *MPL* mutations

DNA from peripheral blood or bone marrow is reported to be acceptable for *JAK2*, *CALR* and *MPL* mutation detection (Bench, et al., 2013; Gong, et al., 2013) but there is currently no recommended gold-standard method for the analysis of these genes. Traditional Sanger sequencing is thought to be an inappropriate technique due to its relatively poor level of detection (10-20% VAF) and does not allow for mutant load quantification. Alternatively, methods including high-resolution melt curve analysis (Rapado, et al., 2009), pyrosequencing (Ronaghi, 2001), and various allele-specific PCR systems can achieve a level of detection of around 5% and both qPCR (Poodt, et al., 2006) and ddPCR are able to detect mutants <0.1% VAF. It is recommended by Tefferi, *et al*. (2011) that a sensitive and quantitative method of analysis is utilised for the detection of *JAK2* V617F mutations in patients at diagnosis as the mutant allele burden can be low level (<10%) and in addition, quantitative methods can indicate homozygosity when ≥50%, which is atypical for ET (2-4% vs. 25-30% of PV) (Vannucchi, et al., 2007).

To ensure that >90% of *JAK2* mutated cases are detected at diagnosis, American (Gong, et al., 2013) and UK guidelines (Bench, et al., 2013) recommend that testing in a diagnostic setting has the analytical sensitivity to detect mutants to at least 1% VAF, referencing the findings of a number of independent studies that assessed the performance of *JAK2* mutation detection assays (Cankovic, et al., 2009; Wang, et al., 2008) and a study that assessed the correlations between *JAK2* V617F mutation burden and clinical phenotype in 260 patients with ET by qPCR which detected mutations down to 1% VAF (Antonioli, et al., 2008). Gong, *et al*., (2013) states that whilst there is evidence that *JAK2* mutants can be detected in affected individuals at <1% VAF (Kouroupi, et al., 2012), caution should be held due to the reported incidence of low level (<5% VAF) identified in the peripheral blood of apparently "normal" individuals (Nielsen, *et al*., 2014; Genovese, *et al*., 2014; Hinds, *et al*., 2016) and recommend follow-up testing in a further blood or bone marrow sample and clinical diagnosis only in the situation of supporting criteria according to the WHO classification system. Recommendations from the Canadian MPN Group (Busque, et al., 2016) agree that the threshold for detection should be low enough to pick up as many patients with MPN as possible but warn that this test should not be used as a basis for

population screening. The issue of low level variants detected in apparently normal individuals is discussed above.

ddPCR is a method of digital PCR that allows detection of *JAK2* V617F mutations <1% VAF (Link-Lenczowska, et al., 2018). This technique utilises template partitioning into ~20,000 droplets prior to amplification PCR to allow for absolute quantification. Each droplet theoretically contains either zero or one template and fluorescent probes are used to identify amplified DNA. Partitions are counted as positive or negative depending on the presence or absence of a fluorescent signals and proportion of droplets that are positive versus negative for a fluorescent signal analysis is then fitted to a Poisson distribution model to determine the absolute starting copy number in units of copies/µl in the input sample. Several studies have suggested that ddPCR is superior to qPCR for quantifying low allele burdens (La Rocca, et al., 2020).

3.2 Aims and objectives

The VAF of somatic mutations in cancer differ substantially depending on the clone size, subclonal structure and spatial heterogeneity. Defining a lower cut-off level for VAF in routine diagnostic practice is challenging, and depends in part on technical considerations (i.e. identifying technical artefacts) as well as understanding the clinical significance of different mutation burdens.

MPNs are a subset of myeloid malignancies driven in most cases by acquisition of the *JAK2* V617F mutation. For some MPN subtypes, a high *JAK2* V617F VAF is associated with more symptomatic disease but for myelofibrosis, paradoxically, a lower VAF is associated with more aggressive disease. In addition, broad population screens have determined that approximately 0.1-3.1% of "normal" individuals are *JAK2* V617F positive, usually at a low VAF, but only a proportion of these will go on to develop an MPN. Using MPN as a model, I aim to understand the significance of low level mutations and use this information to help validate and refine NGS based panel testing for myeloid malignancies.

The principal research questions I aim to address are as follows:

1. Do suspected MPN patients have an elevated frequency of very low VAF (<1%) *JAK2* V617F mutations compared to controls?
2. What is the clinical significance of low VAF (<5%) *JAK2* V617F in MPN?
3. Do mutations in additional genes explain the fact that some MPN patients have symptomatic disease despite low *JAK2* V617F VAF?
4. Is there sufficient evidence to define a quantitative cut-off for a positive result for *JAK2* V617F in relation to a diagnosis of an MPN?

The WRGL receives approximately 20,000 peripheral blood or bone marrow samples per annum for genetic, cytogenetic or genomic assessment of patients with malignant and non-malignant conditions. Ethical permission to utilise excess samples for research was granted by the NRES Committee South West (Central Bristol) in 2010: Study Title: 'The molecular pathogenesis of atypical chronic myeloproliferative neoplasms and related diseases'; REC reference 10/H0102/61; IRAS project ID: 52340; Chief Investigator: Prof NCP Cross. The

original application was successively amended and extended with new research funding with the most recent extension running to September 2021.

With more NHS laboratories replacing traditional sequencing methodologies with NGS technologies the interpretation of low level variants is becoming more of a challenge. It is anticipated that the findings of this study will help in the interpretation of low level *JAK2* variants and therefore in the clinical management of patients with MPN.

3.3 Materials and methods

3.3.1 Identification of cases with low level (<1%) *JAK2* V617F

At the WRGL, *JAK2* mutation status is analysed by an ISO15189-accredited NGS pipeline (referred to as the genotyping pipeline) using an Illumina MiSeq instrument. DNA from patients with suspected or possible MPN (referred to below as query MPN referrals) is extracted from peripheral blood or bone marrow material and the input DNA is subjected to reverse complement PCR (RC-PCR) (NimaGen, 2020) that appends all the functional sequences necessary for sequencing on a MiSeq (sequencing primer binding, sample ID indexes and flow cell hybridisation adaptors). The data is then analysed bioinformatically using a bespoke pipeline generated by collating publicly available online tools and in-house bioinformatics tools and will call any variant in *JAK2* exons 12 and 14, *CALR* exon 9 and *MPL* exon 10 detected at a mutation burden of greater than ≥1%. Two approaches were taken to identify *JAK2* V617F positive cases with a VAF <1%; (i) informatic identification of candidate cases (referred to below as preselection) followed by confirmation with ddPCR and (ii) direct ddPCR screening.

As a preselection step to identify candidate cases with *JAK2* V617F at a VAF of <1%, the diagnostic data obtained from suspected MPN referrals during the period March 2014 to Dec 2019 (approximately 1500 samples), were reanalysed to identify patients that had an apparent *JAK2* V617F at ≤1% VAF using modified bioinformatics tools. In the modified pipeline, the reads were aligned to the *JAK2* V617F amplicons from AmpliconAlignerInputV1.2.tx using AmpliconAlignerV2. The output SAM file was then converted to BAM file format, which was compressed, sorted, and indexed for improved access. Variants were called in secondary analysis using Lofreq (Wilm, 2012). This is more sensitive than the Illumina somatic variant caller used in the original bioinformatic pipeline. After secondary analysis, results were filtered to exclude positive calls with <1000x coverage. The specificity of this modified pipeline was not optimised as all positives indicated by these analyses were confirmed by ddPCR, as described below.

Samples were also tested by ddPCR that were not preselected via the modified bioinformatics pipeline but were instead selected randomly from a cohort of *JAK2* V617F negative MPN referrals. In addition, control samples from individuals unaffected by haematological neoplasia and therefore not previously tested for *JAK2* V617F status (n=303)

were also tested by ddPCR to determine the population background frequency for this technique.

Samples were selected for analysis from query MPN referrals to the WRGL as detailed in the Results. In addition, 32 anonymised low level *JAK2* V617F positive samples were received from Guy's and St Thomas's pathology laboratory.

3.3.2 *JAK2* V617F genotyping by droplet digital PCR (ddPCR)

The quality of genomic DNA was assessed using the Qubit® dsDNA BR Assay Kit and the Qubit Fluorometer according to the manufacturer's instructions; this assay is highly selective for double-stranded DNA (dsDNA) over RNA. DNA was diluted to 16.5 ng/µl in 50µl nuclease-free water. This sample was then sonicated in a Diagenode Bioruptor as follows: 20 cycles at 30 seconds on, 30 seconds off.

A PrimePCR™ ddPCR™ Mutation Detection Kit Assay (Bio-Rad Laboratories) was used to detect a *JAK2* V617F allele. The reaction volume was 20 µl using the following mix:

10 µl 2× ddPCR Supermix for Probes [no dUTP]

1 µl 20× target primers/probes (FAM or HEX) mix

1 µl 20× reference primers/probes (HEX or FAM) mix

8 µl of diluted DNA at 16.5 ng/µl

Droplet generation was performed using 20 µl reaction mix and 70 µl Droplet Generation Oil in a DG8™ Cartridge loaded onto a QX200™ Droplet Generator (Bio-Rad Laboratories) according to manufacturer's instructions. Droplets were then transferred into a 96-well plate and sealed with a PCR plate sealer and PCR was performed on a tetrad thermocycler according to the protocol described in Table 3.1.

Cycling step	Temperature, °C	Time	Ramp rate	Number of cycles
Enzyme activation	95	10 min	2 °C/sec	1
Denaturation	94	30 sec		40
Annealing/extension	55	1 min		
Enzyme deactivation	98	10 min		1
Hold	4	Infinite	1 °C/sec	1

Table 3.1 Thermocyler sequence used for the *JAK2* V617F ddPCR assay

After thermocycling, the quantification was performed using the QX200™ Droplet Reader (Bio-Rad Laboratories) and QuantaSoft™ software (Bio-Rad Laboratories) according to manufacturer's instructions. This software measures the total number of positive and negative droplets for each fluorophore in each sample, calculates the proportion of droplets with a positive signal, then applies a Poisson algorithm to calculate the concentration of DNA molecules in the starting sample (units: copies/µl input sample); of note, the DNA in the starting sample is diluted to 16.5 ng/µl to increase the probability that only one template molecule is present in each droplet. The VAF of the *JAK2* V617F mutation was calculated for each sample using the following calculation:

$$\text{Variant allele frequency (\%)} = \frac{\text{Concentration } \textit{JAK2} \text{ V617F}}{(\text{Concentration } \textit{JAK2} \text{ V617F} + \text{Concentration } \textit{JAK2} \text{ wild-type})}$$

3.3.2.1 Control samples

Each run included the 0% (negative control) and 0.03% (low level positive control samples from the WHO 1st International *JAK2* V617F Reference samples (supplied by NIBSC, South Mimms, UK) and a water blank as an additional negative control.

3.3.2.2 Determination of acceptable cut-off levels for ddPCR results analysis

The aim of this part of the analysis was to define the criteria required for a sample to be called positive or negative for the *JAK2* V617F mutation and estimate the specificity of this test.

(i) Limit of blank (LoB) was estimated using data from the WHO 1st International *JAK2* V617F Reference sample at allelic burden of 0% according to the following formula (Armbruster & Pry, 2008): **LoB = (mean $VAF_{0\%\ control}$ + 1.645)*$SD_{0\%\ control}$.**
A number of replicates using a blank (water) control were also performed to identify the rate of artefactual positive droplets in a system with no DNA template.

(ii) the LoD was estimated using replicate analyses of the WHO 1st International *JAK2* V617F Reference sample at 0.03% and according to the following formula (Armbruster & Pry, 2008): **LoD = LoB + 1.645($SD_{0.03\%\ control}$)**

(iii) data from replicate analyses performed on the WHO 1st International *JAK2* V617F Reference at 0% and 0.03% were used determine a positive result cut-off level.

(iv) results from the replicate analyses on the WHO 1st International *JAK2* V617F Reference at 0% and 0.03% were used to estimate the sensitivity and specificity of this test according to the following calculation:

Sensitivity = True positives / (true positives + false negatives) * 100

Specificity = True negatives / (true negatives + false positives) * 100

Please note that robust quantitation experiments were not undertaken to determine the measurement uncertainty as the aim of this study was to identify true positives at a level below 1% VAF, but not accurately quantify the mutation burden.

3.3.2.3 Statistical analysis of the frequency of low level *JAK2* V617F mutations

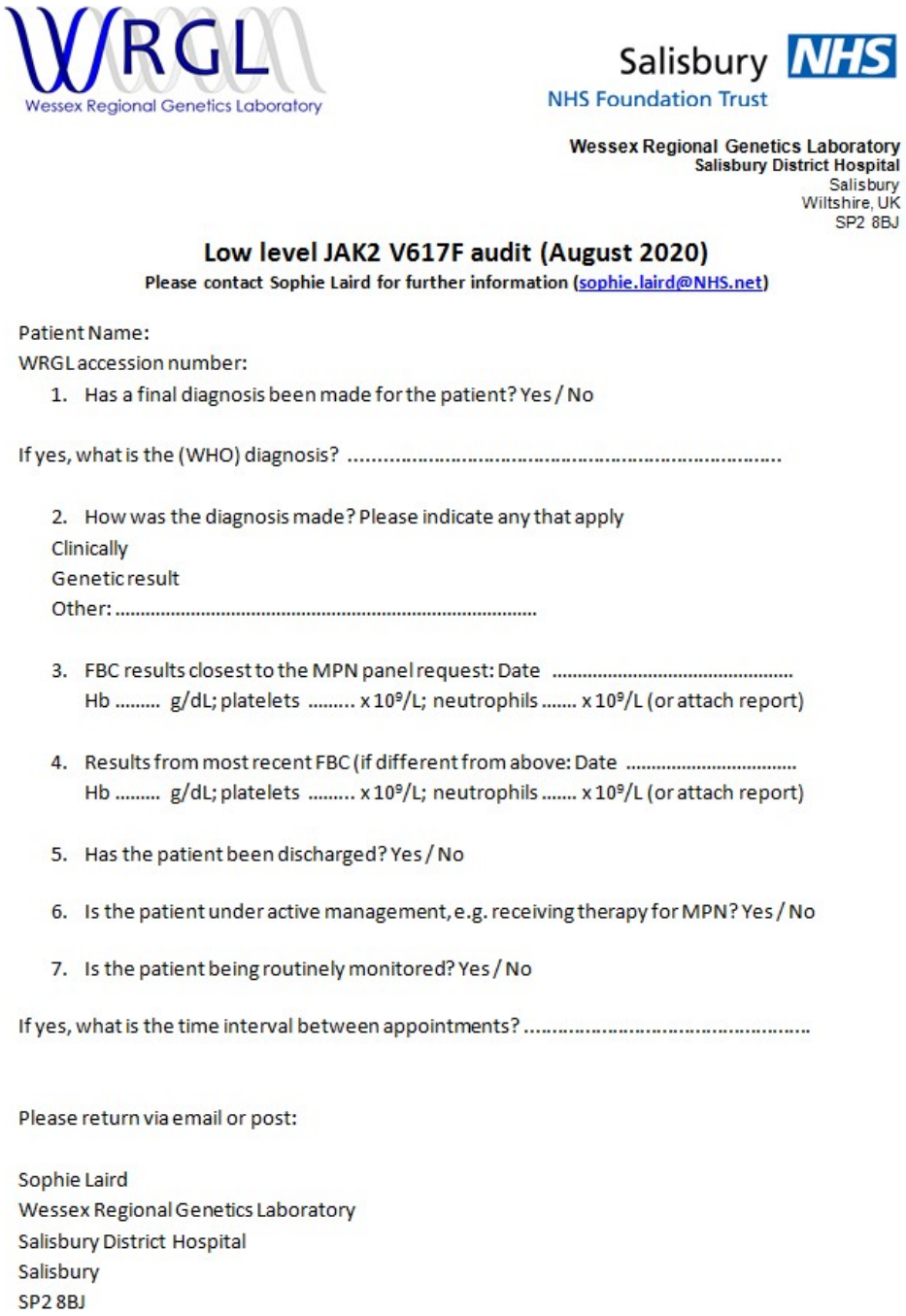
In order to determine whether low level *JAK2* V617F positive is more common in MPN or suspected MPN patients versus population controls, the two-tailed Fisher's exact test was performed using the online calculator supplied by GraphPad (GraphPad, 2020).

3.3.3 Assessing the clinical significance of low level *JAK2* V617F mutations

Clinical information for patients with low level *JAK2* V617F was assessed to identify genotype-phenotype trends in patients with a mutation in the range 0.03- 1% VAF identified by ddPCR (and therefore not reported in a diagnostic setting) and at 1-5% VAF identified by the MPN panel performed by the genotyping pipeline, which had been performed and

reported in a diagnostic setting. Clinicians were not informed if their patient subsequently had a low level (<1% VAF) *JAK* V617F mutation identified by this study.

The following questionnaire was sent out to the referring clinicians of these samples and the results collated (Figure 3.3).



WRGL
Wessex Regional Genetics Laboratory

Salisbury NHS
NHS Foundation Trust

Wessex Regional Genetics Laboratory
Salisbury District Hospital
Salisbury
Wiltshire, UK
SP2 8BJ

Low level JAK2 V617F audit (August 2020)

Please contact Sophie Laird for further information (sophie.laird@NHS.net)

Patient Name:
WRGL accession number:

1. Has a final diagnosis been made for the patient? Yes / No

If yes, what is the (WHO) diagnosis? .....

2. How was the diagnosis made? Please indicate any that apply
   Clinically
   Genetic result
   Other: .....

3. FBC results closest to the MPN panel request: Date .....
   Hb ..... g/dL; platelets ..... x 10^9/L; neutrophils ..... x 10^9/L (or attach report)

4. Results from most recent FBC (if different from above: Date .....
   Hb ..... g/dL; platelets ..... x 10^9/L; neutrophils ..... x 10^9/L (or attach report)

5. Has the patient been discharged? Yes / No

6. Is the patient under active management, e.g. receiving therapy for MPN? Yes / No

7. Is the patient being routinely monitored? Yes / No

If yes, what is the time interval between appointments? .....

Please return via email or post:

Sophie Laird
Wessex Regional Genetics Laboratory
Salisbury District Hospital
Salisbury
SP2 8BJ

Figure 3.3 Audit questionnaire sent to the referring clinicians of samples that had referred patients for MPN panel testing that had received a report detailing the detection of a *JAK2* V617F mutation between 1 and 5% VAF or who had received a *JAK2* V617F negative report for a sample that we had subsequently detected a low level (<1% VAF) *JAK2* V617F mutation.

3.3.4 Assessing the impact of additional mutations in patients with a low level *JAK2* V617F mutation

Selected cases with low *JAK2* V617F VAF were tested by the myeloid gene panel to determine if additional mutations may explain their clinical features. Clinicians were not informed that this test had been undertaken nor were they informed of the result unless the test was subsequently requested in a diagnostic setting. Details of the myeloid gene panel, namely the Illumina® TSMP are provided in Chapter 2.

3.3.5 46/1 haplotype analysis

ARMS PCR to identify the *JAK2* 46/1 haplotype (Jones, et al., 2009) was performed to see if there was an inherited contribution to the development of low level *JAK2* V617F acquisition. Primers were designed to detect rs12340895 (C>G at chr9:5076691; GRCh38.p1), a tag SNP that serves as a surrogate for 46/1 using an online design tool (http://primer1.soton.ac.uk/primer1.html).

Outer Forward Primer (OF):	TCAGATTATCATTAGCACCTTTTTTGG
Outer Reverse Primer (OR):	GGAGCCTCTCAGATACCTCCATATAAC
Inner Forward Primer (IF):	TCGAGGTATGCCTTTATTTTAGTCCC
Inner Reverse Primer (IR):	AACAATTTTCTTGAATGTAAACTTTGTCAC

The expected product sizes were as follows:

C allele 200 bp

G allele 252 bp

Product size for the two outer primers 396 bp

The PCR mastermix was made according to the volumes shown in Table 3.2.

Reagent	Volume per sample (µl)
OF 10mM	0.625
OR 10mM	0.625
IF 10mM	2.5
IR 10mM	2.5
Amplitaq Gold buffer	2.5
AmpliTaq $MgCl_2$	1.75
dNTP 10mM	0.5
Water	12.8
AmpliTaq Gold	0.2

Table 3.2 ARMS PCR mastermix

24 µl mastermix and 1 µl DNA was then combined and PCR was performed according to the following protocol:

95°C for 15 minutes
94°C for 20 seconds
60°C for 50 seconds
72°C for 50 seconds
} For 35 cycles
72°C for 10 minutes
15°C for infinity

Samples tested included those samples with a low level (<1% VAF) *JAK2* V617F mutation identified by ddPCR, age matched normal individuals unaffected by haematological neoplasia, and patients for whom a homozygous *JAK2* V617F mutation was identified by routine diagnostic screening.

3.4 Results

3.4.1 Determination of acceptable cut-off levels for ddPCR analysis

Replicate analyses were performed on the following samples; for each reference sample, the number of replicate analyses which met the minimum acceptable droplet cut-off level as recommended by BioRad (≥10,000) is also provided (full results from these analyses provided in Appendix 6):

(i) The WHO 1st International *JAK2* V617F Reference sample at allelic burden of 0% (n=82); Table 3.3. Of these, seventy four (90.2%) valid replicates showed no positive droplets whereas 8 (9.76%) valid replicates had a single positive droplet. In the replicates with a false positive droplet, the VAF was 0.003% to 0.005% (median 0.005%).

Replicate	Sample	Total Accepted Droplets	Positive droplets	Negative droplets	Concentration	Variant allele frequency
2	0%	12499	1	12498	0.09	0.005%
3	0%	13036	1	13035	0.09	0.005%
16	0%	10849	1	10848	0.11	0.005%
23	0%	11614	1	11613	0.1	0.004%
30	0%	15069	1	15068	0.08	0.003%
59	0%	14712	1	14711	0.08	0.005%
60	0%	13818	1	13817	0.09	0.005%
73	0%	15041	1	15040	0.08	0.005%

Table 3.3 ddPCR results from the replicates with one positive droplet (n=8) out of a total of 82 replicates tested with the WHO 1st International *JAK2* V617F Reference sample at an allelic burden of 0%. The remaining 76 replicates showed no positive droplets and therefore are not included in this table of results.

(ii) Blank (water) control (n=115); Table 3.4. Of these, 114 (99.1%) valid replicates showed no positive droplets whereas 1 (0.9%) valid replicate had two positive droplets leading to a (false positive) VAF of 0.012% in this sample.

Replicate	Sample	Total Accepted Droplets	Positive droplets	Negative droplets	Concentration	Variant allele frequency
90	BLANK	15768	2	15766	0.15	0.012%

Table 3.4 ddPCR results from the only replicate with positive droplets out of 115 blank control replicates tested.

(iii) The WHO 1st International *JAK2* V617F Reference sample at allelic burden of 0.03% (n=129). All 129 (100%) valid replicates were positive for a low level *JAK2* V617F variant at VAF 0.01% to 0.07% (median 0.04%) with 1 to 17 positive droplets in each valid replicate (median 7; Figure 3.4).

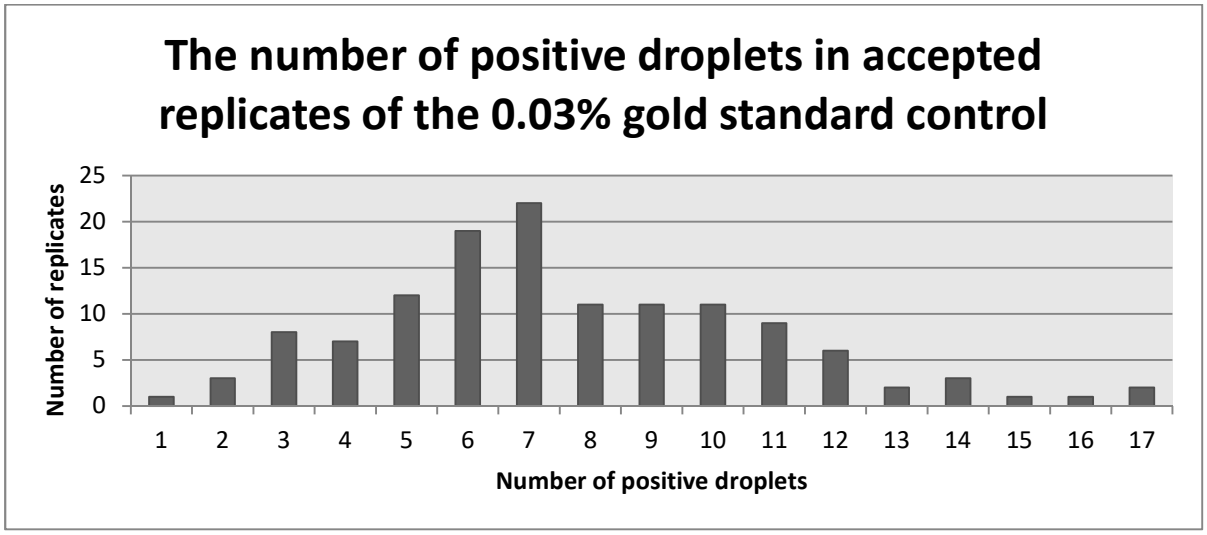


Figure 3.4 The number of positive droplets in accepted replicated (i.e. those with >10,000 droplets) of the 0.03% gold standard JAK2 V617F positive control

Estimation of LoB and LoD

The mean VAF for the *JAK2* V617F mutation in the 0% control (n=82) was 0.0004% and the SD was 0.0000135. The LoB was therefore estimated to be 0.00042% (=0.0004% + 1.645*0.0000135).
The SD for the 0.03% control (n=129) was 0.00013 and therefore the LoD was estimated to be 0.00063% (= 0.00042% + 1.645*0.00013).

Of note, >99% (n=114) of valid replicates of a blank water control showed no positive droplets indicating negligible interference of artefactual analytical signal that might result in a false positive result.

Determination of a cut-off level for a "positive" result

Please note that, as described more fully below, the modified bioinformatics pipeline, designed to identify a cohort of patients with a possible low level *JAK2* V617F variant for confirmation by ddPCR, identified a number of samples to be tested and, although no lower limit of positivity was defined, the lowest VAF indicated by this bioinformatic analysis was 0.04%. Consequently, I defined a cut-off level for a positive result for ddPCR down to 0.03% which allows for confirmation of the bioinformatic results and allows for a small margin of error in allelic quantification. This is well above the LoD for ddPCR defined above.

The cut-off level for a positive result includes a minimum accepted droplet count for each sample analysis, the minimum number of positive droplets within a sample analysis and the number of replicates required to consider this a reliable result.

The cut-off used was: **≥3 positive droplets in at least 2 replicates with a total ≥10,000 accepted droplets per replicate. The minimum VAF reported was 0.03%.**

The rationale for this cut-off is as follows:

In 82 replicates of the 0% control sample, 8 replicates with ≥10,000 total accepted droplets showed one positive droplet. No replicates showed 2 (or more) positive droplets. In 129 replicates of the 0.03% control sample, all 129 samples showed a positive result with VAF range 0.01% to 0.07% with a positive droplet count ranging from 1 to 17 (Figure 3.4); median 7. This data suggests that 2 positive droplets may be used as a positive cut-off but in order to reduce the chance of identifying false positive results (and in light of a single replicate in the blank control replicate dataset with 2 positive droplets), a cut-off level of 3 positive droplets was used. Results were also required to be consistent in at least two replicates to be considered reliable. Further, because the lowest positive reference control used was 0.03%, only those with an average VAF >0.03% in acceptable replicates would be considered positive as, even though the LoD has been estimated to be much lower than that.

Sensitivity and specificity

Positive and negative results were defined by the criteria defined above from the replicate analyses on the WHO 1st International *JAK2* V617F Reference at 0% and 0.03%.

As all negative controls tested negative by ddPCR, the specificity is estimated to be 100%.

As all positive controls tested positive by ddPCR, and there were no false positives, the specificity is estimated to be 100%.

3.4.2 Identifying patients for low level (<1% variant allele frequency) *JAK2* V617F ddPCR

As described above, all query MPN referrals to WRGL are simultaneously tested for *JAK2* V617F, *JAK2* exon 12 indels (c.1559_c.1641), *CALR* exon 9 indels (c.1097_1154) and *MPL* codons 505 and 515. The validated limit of detection for this assay is 1% VAF and a standard uncertainty of measurement of ±2%. Between March 2014 and October 2017, WRGL tested 4,026 query MPN referrals: 652 samples were *JAK2* V617F positive (16.2%) and 3372 (83.8%) had wild-type *JAK2* p.(Val617) status by this panel.[iii]

To identify low level *JAK2* V617F cases, 4 sample cohorts were tested:

Firstly, in a subgroup of WRGL *JAK2* V617F-negative query MPN referrals identified within this time-frame (n=615), a bioinformatic approach was designed to identify candidate low level (<1%) *JAK2* V617F positive cases; these samples were then tested by ddPCR for confirmation. These samples were labelled with the prefix "Low" for the purpose of this study. Of note, no sample selection was performed on the basis of the *JAK2* exon 12, *CALR* or *MPL* status in this cohort; of the 615 *JAK2* V617F negative samples, 90.2% (n=555) were apparently triple-negative, 0.2% (n=1) was *JAK2* exon 12 positive, 8% (n=49) were *CALR* mutated and 1.6% (n=10) were *MPL* mutated.

Secondly, 107 samples that were triple-negative by the MPN panel (*JAK2* V617F, *JAK2* exon 12, *CALR*, *MPL* negative) were tested by ddPCR (named samples 201 to 307). These samples were randomly selected from cases referred for testing between 1st November 2018 and 31st December 2018 in order to ensure that they did not over-lap with samples represented in other cohorts within this study. Samples within this cohort were not re-analysed by the modified bioinformatic pipeline so that the utility of this pipeline in identifying samples with low level (<1%) *JAK2* V617F could be assessed.

[iii] Please note that the referral criteria was not assessed for these samples as it was assumed that the vast majority would have been referred with suspected or clinically confirmed MPN for genotype analysis; furthermore, a significant proportion of samples referred for MPN panel testing are received by the laboratory with no specific clinical information relating to the suspected disorder. However, results indicate that a small proportion of these samples were referred with known or suspected MDS/MPN or other disorder.

Thirdly, it was noted within the first group of samples tested (the "Low" cohort), that some samples with a low level *JAK2* V617F mutations identified by ddPCR had co-occurring *CALR* and *MPL* variants. Therefore, a further 206 *CALR* or *MPL* positive samples identified by the diagnostic pipeline were tested by ddPCR to determine the frequency of co-mutated driver genes in this cohort; as there were insufficient *CALR/MPL* positive patients in the original 615 samples, these additional samples were randomly selected from cases identified by the diagnostic MPN panel between October 2017 and March 2020.

Finally, 32 anonymised samples were received from Guy's and St Thomas's pathology laboratory. This laboratory performs diagnostic *JAK2* V617F testing using ddPCR and as such they were able to provide a number of samples with low level *JAK2* V617F mutations. The only information provided with these external samples was that a proportion of them had a low level *JAK2* V617F mutation. These samples were labelled with the prefix "GST" (GST1 to GST 35) and were included to validate the ddPCR method utilised in this study and provide additional samples for myeloid gene panel analysis in those samples with a low level *JAK2* V617F mutation.

3.4.3 Prevalence of low level (<1%) *JAK2* V617F following preselection by re-analysis of samples negative for *JAK2* V617F genotyping test using modified bioinformatic pipeline

Samples that were wildtype *JAK2* p.(Val617) by the original genotyping pipeline (n=615; Low cohort) were re-analysed by modified bioinformatics pipeline. A low level *JAK2* V617F was indicated in 79 out of 615 samples (12.8%). There was no lower limit set within the modified pipeline for positivity but the VAF for mutations detected ranged from 0.03% to 1.14%.

Only one sample was detected by the modified pipeline with >1% VAF (the cut-off for the standard diagnostic MPN pipeline). Further analysis indicated that the sample was called as as wild-type by standard pipeline as there were <1% of *JAK2* V617F-positive reads that passed the diagnostics IQC parameters, which were not incorporated into the modified pipeline. Of note, 16 of the 79 samples with a candidate low level *JAK2* V617F mutation had a *CALR* mutation or *MPL* mutation identified by the diagnostic analysis; in these cases the VAF range was 10% to 53% for the *CALR* mutations, 3% to 52% for the *MPL* mutations and 0.04% to 1% for the *JAK2* V617 mutations.

JAK2 V617F ddPCR was undertaken to confirm the presence of the mutation and the results are shown in Table 3.5. Of the 79 samples, 2 failed and 24/77 (31%) were confirmed to have

a low level *JAK2* V617F mutation with VAFs ranging from 0.04% to 0.87%. Of the 24 true positives, 8 (33%) also had a *CALR* (n=7) or *MPL* (n=1) co-mutation.

Fifty three (69%) samples were considered negative for *JAK2* V617F; however, 6 of these samples showed some evidence of positivity but below the validated limit of sensitivity of the ddPCR assay (either >3 positive droplets in 2 or more than one replicates but <0.03% or 0.03% VAF with too few positive droplets). Of these 6 cases, 2 had a *CALR* of *MPL* mutation. Full results are shown in Appendix 7.

ddPCR results from samples for low level *JAK2* V617F confirmation following bioinformatic analysis of 615 samples *JAK2* V617F negative by the diagnostic MPN panel (n=77)				
Result	**Number of samples (% all samples, n=615)**	**Number of samples TN by the diagnostic MPN panel (%, n=555)**	**Number of *CALR*+ samples (%, n=49)**	**Number of *MPL*+ samples (%, n=10)**
Low level *JAK2* V617F	24 (3.9%)	16 (2.9%)	7 (14.3%)	1 (10%)
No *JAK2* V617F detected	591 (96.1%)	539 (97.1%)	42 (85.7%)	9 (90%)
Low level *JAK2* V617F positive samples				
Sample name: "Low…" number	**Number of replicates**	**Average VAF of the *JAK2* V617F detected by ddPCR (%)**	***CALR*+ or *MPL*+ by diagnostic MPN panel (VAF)**	
1	6	0.05%		
2	3	0.25%	*CALR*+ (35%)	
5	3	0.11%		
6	3	0.07%	*CALR*+ (34%)	
8	3	0.21%	*CALR*+ (12%)	
9	3	0.65%		
16	3	0.78%	*CALR*+ (10%)	
19	3	0.78%	*MPL*+ (3%)	
20	2	0.13%		
23	3	0.11%		
24	2	0.04%		
25	3	0.13%		
26	3	0.87%		
27	3	0.12%		
28	3	0.62%	*CALR*+ (43%)	
33	3	0.17%		
41	3	0.07%		
65	3	0.52%		
67	3	0.07%		
69	3	0.04%	*CALR*+ (27%)	
70	3	0.16%		
72	3	0.55%		
76	3	0.75%	*CALR*+ (32%)	
79	3	0.49%		

Table 3.5 Samples with a low level *JAK2* V617F detected by ddPCR on the "Low" sample cohort: these samples were originally *JAK2* V617F negative by the WRGL diagnostic pipeline (LoD=1%) but were identified as possible low level variant positive by bespoke bioinformatic analysis. Seventy nine samples were tested, but only 77

samples yielded a result. A low level *JAK2* V617F variant was detected in 24 (31%) samples. TN = triple-negative; *CALR*+ = *CALR* mutated; *MPL*+ = *MPL* mutated.

These results indicate that by this method of sample selection (preselection by a modified bioinformatic analysis followed by ddPCR confirmation), low level (<1% VAF) *JAK2* V617F mutations occur in 3.9% (=24/615) of all samples previously thought to be *JAK2* V617F negative. This equates to a low level *JAK2* V617F in 2.88% (=16/555) of samples triple negative by the diagnostic MPN panel, 14.29% (=7/49) of samples with a *CALR* mutation detected by the diagnostic MPN panel and 10% (=1/10) of the samples with a *MPL* mutation by the diagnostic MPN panel.

3.4.4 Prevalence of low level (<1% VAF) *JAK2* V617F in samples identified as triple-negative by the MPN panel

Samples (n=107) that were triple-negative by the diagnostic MPN panel (*JAK2* V617F, *JAK2* exon 12, *CALR*, *MPL* negative) were tested by ddPCR (named sample 201 to 307). Samples within this cohort were not re-analysed by the modified bioinformatic pipeline prior to selection. This cohort is referred to as the "Triple-negative (unselected)" cohort throughout the rest of this document.

Four of these samples (3.7%) were positive for a *JAK2* V617F mutation at 0.03% to 0.21% VAF (Table 3.6). Full results are shown in Appendix 8.

ddPCR results in samples TN by the diagnostic MPN panel (n=107)		
Result		**Number of samples (%)**
Low level *JAK2* V617F		4 (3.7%)
No *JAK2* V617F detected		103 (96.3%)
Low level *JAK2* V617F positive samples		
Sample name	**Number of replicates**	**Average VAF of the *JAK2* V617F by ddPCR (%)**
206	2	0.04%
268	3	0.04%
275	3	0.21%
298	3	0.03%

Table 3.6 Samples with a low level *JAK2* V617F detected by ddPCR on the randomly selected samples shown to be triple-negative (TN) by the MPN panel (n=107), i.e. *JAK2* V617F negative and no evidence of a *CALR*, *MPL* or *JAK2* exon 12 mutation by the validated WRGL genotyping assay (LoD=1%); termed the triple-negative (unselected) cohort.

3.4.5 Prevalence of low level (<1% VAF) *JAK2* V617F mutations in samples known to be *CALR/MPL* mutated

Samples that were positive for a +1 frameshift mutation in exon 9 of *CALR* but negative for a *JAK2* V617F mutation (*CALR*+/*JAK2* V617F-) by the standard genotyping pipeline with a LoD of 1% (n=152) or positive for a *MPL* W515/W505 mutation but negative for a *JAK2* V617F mutation (MPL+/JAK2 V617F-) by the MPN panel (n=52) were tested by ddPCR. Samples within this cohort were not re-analysed by the modified bioinformatic pipeline prior to selection. This cohort is referred to as the "*CALR*/*MPL* positive" cohort throughout the rest of this document. Full results are shown in Appendix 9.

Of the 152 *CALR*+/*JAK2* V617F- tested, three samples failed due to insufficient DNA (Samples 314, 489 and 476), 11 samples (7.24% [11/149]) were positive for a *JAK2* V617F mutation at 0.03% to 1.51% VAF (Table 3.7). [Please note that the 13 samples identified as low level *JAK2* V617F positive and *CALR* mutated in the "Low" cohort are not included in these data as those samples were pre-selected bioinformatically.]

ddPCR results in samples *CALR* positive by the diagnostic MPN panel (n=149)			
Result			**Number of samples (%)**
Low level *JAK2* V617F			11 (7.2%)
No *JAK2* V617F detected			138 (92.8%)
Low level *JAK2* V617F positive samples			
Sample name	**Number of replicates**	**Average VAF of the *JAK2* V617F by ddPCR (%)**	***CALR* mutation (VAF)**
318	3	0.04%	5 bp insertion (38%)
365	4	0.03%	5 bp insertion (34%)
384	3	0.34%	5 bp insertion (42%)
415	3	0.05%	5 bp insertion (37%)
421	3	0.90%	Other deletion (47%)
433	2	0.15%	Other deletion (36%)
453	3	0.25%	52 bp deletion (35%)
462	3	0.07%	Other insertion (40%)
463	3	0.47%	52 bp deletion (40%)
479	3	1.51%	52 bp deletion (12%)
484	3	1.32%	52 bp deletion (39%)

Table 3.7 Samples with low level *JAK2* V617F detected by ddPCR on the randomly selected samples shown to be *CALR* positive and *JAK2* V617F negative (LoD=1%) by the MPN panel (n=149); termed the CALR/MPL positive cohort.

Of the 52 *MPL*+/*JAK2* V617F- tested, 8 samples (15.7%) were positive for a *JAK2* V617F mutation at 0.04% to 0.21% VAF (Table 3.8). [Please note that the 3 samples identified as low level *JAK2* V617F positive and *MPL* positive in the "Low" cohort are not included in these data as those samples were pre-selected bioinformatically.

ddPCR results in samples *MPL* positive by the diagnostic MPN panel (n=52)			
Result			**Number of samples (%)**
Low level *JAK2* V617F			8 (15.7%)
No *JAK2* V617F detected			44 (84.3%)
Low level *JAK2* V617F positive samples			
Sample name	**Number of replicates**	**Average VAF of the *JAK2* V617F by ddPCR (%)**	***MPL* Mutation (VAF)**
313	3	0.09%	W515L (14%)
335	3	1.09%	W515K (18%)
336	3	0.21%	W515L (19%)
347	3	0.03%	W505N (17%)
367	2	0.15%	W515K (19%)
383	3	0.15%	W505N (30%)
387	3	0.04%	W515L (26%)
405	3	0.10%	W515L (5%)

Table 3.8 Samples with low level *JAK2* V617F detected by ddPCR on the randomly selected samples shown to be *MPL* positive and *JAK2* V617F negative (LoD=1%) by the MPN panel (n=52).

3.4.6 ddPCR results from the control samples

Three hundred and three DNA samples stored at WRGL for diagnostic or research testing for reasons other than haematological neoplasms were anonymised and tested for *JAK2* V617F by ddPCR testing to determine the background population frequency of this mutation.

This comprised of two groups of individuals: (i) 100 samples were from females of reproductive age recruited as part of the RAPID non-invasive prenatal testing (NIPT) evaluation study (R2000-R2100); (ii) 203 samples were received by WRGL for cystic fibrosis testing as part of cascade/familial studies or due to personal persistent unexplained chronic bronchitis (named CF1-CF203). In the latter group, all individuals were greater than 50 years old (median age 65) with representation from both sexes (male, n=75; female, n=125) and this group therefore represents an age and sex matched population for MPN cases. In both groups, DNA was extracted in-house from peripheral blood.

As for the analysis of the MPN patient datasets above, a positive result was considered when two or more replicates showed 3 or more positive droplets, with a minimum accepted

droplet count per replicate of 10,000; further, only samples with an average VAF of 0.03% were considered positive.

In total, *JAK2* V617F ddPCR analysis results were available for 282 control samples; ddPCR failed in 22 samples due to insufficient material.

A *JAK2* V617F mutation was detected in 5 samples (1.68%); Table 3.9. All 5 positives were from the older CF group; *JAK2* V617F was not detected in the younger RAPID cohort. The *JAK2* V617F mutation was low level (<1%) in three samples (CF61, CF159 and CF160). One sample had a mutation at 1.41% VAF (sample CF124) and the presence of this mutation was confirmed by the diagnostic genotyping pipeline. One sample unexpectedly had *JAK2* V617F at 24% VAF (sample CF45); further investigation elucidated that this individual is known to have a *JAK2*-mutated MPN.

Controls with *JAK2* V617F detected by ddPCR			
Control	**Age**	**Sex**	**Average VAF of the *JAK2* V617F by ddPCR (%)**
CF45	65	F	24.28%
CF124	78	M	1.41%
CF159	81	F	0.12%
CF61	78	M	0.06%
CF160	60	M	0.03%

Table 3.9 ddPCR was performed on 303 population control samples; ddPCR analysis was successful in 298 samples and analysis of results identified a *JAK2* V617F mutation in 5 samples. The age and sex (male, M; female, F) is shown here for the positive samples.

3.4.7 Comparison of the frequency of low level *JAK2* V617F mutations between groups

In order to determine whether low level *JAK2* V617F positive is more common in MPN or suspected MPN patients versus population controls, the Fisher's exact test was used. This was separated into a comparison of the prevalence of low level *JAK2* V617F mutations according to the following groups: (i) MPN referrals triple-negative by the WRGL MPN diagnostic assay that were pre-selected by a modified bioinformatic pipeline *versus* population controls; (ii) MPN referrals triple-negative by the WRGL MPN diagnostic assay that were randomly selected (i.e. no modified bioinformatic analysis) *versus* population controls; (iii) the prevalence of low level *JAK2* V617F in group (i) *versus* group (ii); (iv) the prevalence of low level *JAK2* V617F in group (i) plus group (ii) *versus* population controls; (v)

CALR mutated MPN (no bioinformatic preselection) *versus* population controls; (vi) *MPL* mutated MPN (no bioinformatic preselection) *versus* population controls.

To compare the prevalence of low level *JAK2* V617F mutations in the test samples, only the age matched controls were used (n=197). Within that cohort, 5 samples were shown to have a *JAK2* V617F mutation by ddPCR analysis; however, one had a VAF of 24% who was later identified as an MPN patient. Only those population controls with a low level *JAK2* V617F were considered in this analysis (i.e. the incidentally identified patient with MPN was excluded from the analysis) and thus the frequency of low level *JAK2* V617F in the population controls was considered as 4 out of 197 (2%).

(i) Prevalence of low level *JAK2* V617F in population controls (n=197) versus the Low cohort: query MPN referrals triple-negative by the WRGL diagnostic MPN panel that were preselected by a modified bioinformatics pipeline and then tested by ddPCR (n=555)

	Low level *JAK2* V617F positive	**Low level *JAK2* V617F negative**	**Total**
Control	4	193	197
MPN diagnostic panel negative (bioinformatically preselected)	16	539	555
Total	20	732	752

Table 3.10 Contingency table for Fisher's exact test for MPN referrals triple-negative by the WRGL MPN diagnostic panel that were pre-selected by a modified bioinformatic pipeline (n=61) from a cohort of 555 triple-negative samples.

The frequency of low level *JAK2* V617F is not significantly higher in these cases compared to controls (P=0.62; Fisher's exact test), i.e. no evidence for enrichment of low level *JAK2* V617F in the Low cohort.

(ii) MPN referrals triple-negative by the WRGL MPN diagnostic panel that were randomly selected (i.e. no bioinformatic preselection) (n=107).

	Low level *JAK2* V617F positive	Low level *JAK2* V617F negative	Total
Control	4	193	197
MPN diagnostic panel negative (no bioinformatic preselection)	4	103	107
Total	8	296	304

Table 3.11 Contingency table for Fisher's exact test for MPN referrals triple- negative by the WRGL MPN diagnostic panel that were randomly selected (i.e. no modified bioinformatic analysis).

The frequency of low level *JAK2* V617F is not significantly higher in these cases compared to controls (P=0.46; Fisher's exact test), i.e. no evidence for enrichment of low level *JAK2* V617F in the TN group.

(iii) Prevalence of low level *JAK2* V617F in MPN referrals triple-negative by the WRGL diagnostic MPN panel that were selected initially by a modified bioinformatics pipeline and then tested by ddPCR (n=555) versus MPN referrals triple-negative by the WRGL MPN diagnostic panel that were randomly selected (i.e. no bioinformatic preselection) (n=107).

	Low level *JAK2* V617F positive	Low level *JAK2* V617F negative	Total
MPN diagnostic panel negative (bioinformatically preselected)	16	539	555
MPN diagnostic panel negative (no bioinformatic preselection)	4	103	107
Total	20	642	662

Table 3.12 Contingency table for Fisher's exact test for MPN referrals triple-negative by the WRGL MPN diagnostic panel that were pre-selected by a modified bioinformatic pipeline (n=61) from a cohort of 555 triple-negative samples and MPN referrals triple- negative by the WRGL MPN diagnostic panel that were randomly selected (i.e. no modified bioinformatic analysis).

The frequency of low level *JAK2* V617F is not significantly different between these two case groups (P=0.55; Fisher's exact test) indicating that the bioinformatic filtering in the 'Low' group was unlikely to have missed any low level *JAK2* V617F positive cases and therefore the Low and TN groups could be combined to increase the power of the analysis.

(iv) Prevalence of low level *JAK2* V617F in (i) and (ii) combined compared to population controls.

	Low level *JAK2* V617F positive	Low level *JAK2* V617F negative	Total
Control	4	193	197
MPN diagnostic panel negative (bioinformatically preselected [n=555] plus no bioinformatic preselection [n=107])	20	642	662
Total	24	835	859

Table 3.13 Contingency table for Fisher's exact test for MPN referrals triple-negative by the WRGL MPN diagnostic panel (identified by either random selection or bioinformatics pre-selection).

The frequency of low level *JAK2* V617F is not significantly higher in these cases compared to controls (P=0.62; Fisher's exact test), i.e. no evidence for enrichment of low level *JAK2* V617F in the Low plus TN groups.

(v) *CALR* mutated MPN (no bioinformatic preselection)

	Low level *JAK2* V617F positive	Low level *JAK2* V617F negative	Total
Control	4	193	197
***CALR* positive (no bioinformatic preselection)**	11	137	149
Total	16	330	346

Table 3.14 Contingency table for Fisher's exact test for *CALR* mutated MPN (no modified bioinformatics analysis)

The frequency of low level *JAK2* V617F is significantly higher in these cases compared to controls (P=0.018; Fisher's exact test).

(vi) *MPL* mutated MPN (no bioinformatic preselection)

	Low level JAK2 V617F positive	Low level JAK2 V617F negative	Total
Control	4	193	197
MPL positive (no bioinformatic preselection)	8	44	52
Total	12	236	249

Table 3.15 Contingency table for Fisher's exact test for *MPL* mutated MPN (no modified bioinformatics analysis)

The frequency of low level *JAK2* V617F is significantly higher in these cases compared to controls (P=0.006; Fisher's exact test).

Overall, the prevalence of low level *JAK2* V617F mutations was not significantly different in query MPN cases that were triple-negative by the diagnostic MPN panel (with or without bioinformatic selection) compared to controls but there was evidence that low level *JAK2* V617F mutations are enriched in *CALR*- and *MPL*- positive MPN cases.

3.4.8 Assessing the clinical significance of low level *JAK2* V617F mutations

To try and understand the clinical significance, if any, of low level *JAK2* V617F in cases referred for investigation of a possible MPN, an audit questionnaire was sent to referring clinicians from 149 samples. This included three main groups of subjects:

(1) those who had sent samples for diagnostic testing by the diagnostic MPN panel and within which we had detected and reported a *JAK2* V617F mutation at 1-5% VAF (n=110)

(2) samples that were shown to be triple-negative by the MPN panel but had a low level (<1% VAF) *JAK2* V617F mutation detected by ddPCR (n=20)

(3) samples that had a *CALR/MPL* mutation but were negative for *JAK2* V617F mutation by the diagnostic MPN panel but were found to have a low level (<1% VAF) *JAK2* V617F mutation detected by ddPCR (n=19)

In total, I received 81 completed audits, consistent with a return rate of 54%, from 15 hospitals in the United Kingdom and abroad: The Royal Bournemouth Hospital; Poole NHS

Trust Hospital; Queen Alexandra Hospital, Portsmouth; Department of Pathology, Oxford; Southampton General Hospital; Salisbury NHS Foundation Trust, Upton Health Centre, Poole; Royal Hampshire County Hospital; Dorset County Hospital; Royal United Hospital NHS Trust, Bath; St Mary's Hospital IOW NHS Trust; Royal Berkshire Hospital; Antrim Area Hospital, Northern Ireland; Sir Anthony Mamo Oncology Centre; Malta.

The results from these returns are described below.

(1) Results from samples with a JAK2 V617F mutation at 1-5% VAF by the MPN panel

From the samples that showed a *JAK2* V617F mutation at 1-5% VAF by the MPN panel in a diagnostic setting at WRGL, I received 62 returns (56%). A summary of the key findings is provided below; the full results are displayed in Appendix 10.

The median age of the patients for whom I received information was 66 years (range 36-87) and the sex distribution was 1:1.6 male:female. Of the 62 cases, 55 were given a final diagnosis of ET (n=38), PV (n=10), PMF (n=2), MPN-U (n=3) or secondary MF (post-ET or –MDS/MPN, n=2). Respondents said that there was no final diagnosis in 7 cases. Of note, the genetic result (i.e. the finding of *JAK2* V617F) was considered to be an important factor in establishing a diagnosis of a CMN for 98% (n=54) cases, and the sole important factor in 42% (n=23) cases. The bone marrow was considered an important factor in the final diagnosis for only 4 (7%) cases and in one of these cases we were informed that the BM was the sole important factor in diagnosis (W1807472 diagnosed with PV).

A summary of the 7 cases without a final diagnosis are provided below:

(i) Sample W1604531 had a *JAK2* V617F mutation at 1% VAF and the patient was a 40 year old male referred in 2016 for testing with a family history of haematological neoplasm (mother and paternal grandfather). The clinician reported that it was unclear whether the patient had *JAK2*-positive ET or a familial inherited disorder. No further information was available. *CALR* and *MPL* variant analysis has not been undertaken on this sample (or another sample from this patient).

(ii) Sample W1819781 was a 48 year old male referred in 2018 with idiopathic erythrocytosis. He had a *JAK2* V617F mutation at 1% VAF but the clinician reported that there was insufficient clinical evidence for the diagnosis of a

haematological neoplasm given the longstanding nature of his erythrocytosis without significant progression and as such, the patient was reported to have been discharged from active monitoring and/or treatment. Of note, the WRGL received an additional sample from this patient in 2019 to investigate the persistence of this mutation in the absence of changes to the patient's clinical presentation and the *JAK2* V617F mutation was detected again at 1% VAF.

(iii) Sample W1806548 was a 54 year old male referred for MPN panel testing in 2018; a *JAK2* V617F mutation was detected at 2% VAF. The clinician reported that it is unclear whether this patient has an MPN based on clinical features and as such, the patient is not being treated but is being monitored on a yearly basis. Of note, the WRGL tested another sample from this patient in 2019 and the *JAK2* V617F variant was confirmed but showed no significant change to the VAF.

(iv) Sample W1809600 had a *JAK2* V617F mutation at 3% VAF but no diagnosis of haematological neoplasm was made before patient's demise due to bladder cancer.

(v) Sample W1417698 had a *JAK2* V617F mutation at 3% VAF but no other evidence of an MPN. The patient was reported to be being monitored yearly by their General Practitioner but not current in receipt of therapy.

(vi) Sample W1714687 had a *JAK2* V617F mutation at 3% VAF. The patient was female (aged 66 years old), referred for testing with portal vein thrombosis but no final diagnosis of MPN was made. The patient was reported to be being monitored every three months. We have never received a follow up sample for *JAK2* monitoring.

(vii) Sample W1909845 was a 36 year old male with a *JAK2* V617F at 2% VAF who was referred for testing with polycythaemia via the GP. The consultant haematologist reported that a final diagnosis had not been made but additional information was not available as the patient was being managed by their GP and hospital records had not been updated.

Please note that for all of these 7 cases listed above, the TSMP had not been undertaken on that patient previously therefore the presence of additional abnormal clones had not been excluded. However, four separate cases had additional mutations identified by either the diagnostic MPN panel or by an independent TSMP investigation; these are described below.

(i) Sample W1800499 was an 80 year old female who had a *JAK2* V617F mutation at 2% VAF and a *CALR* 52 bp deletion at 12% VAF. A final diagnosis of ET was reported and the patient is receiving aspirin and hydroxycarbamide therapy.

(ii) Sample W1710353 was a 71 year old male who had a *JAK2* V617F mutation at 2% VAF and a *CALR* 52 bp deletion at 39% VAF. A final diagnosis of PMF was made and the patient was reported to have progressed to AML in 2020. Interestingly, no additional mutations were detected by the TSMP in 2017 at the time of MPN panel testing and cytogenetics was undertaken in 2020 but no clonal abnormalities were detected.

(iii) Sample W1706237 was a 61 year old male who had a *JAK2* V617F mutation at 5% VAF and a +1 frameshift deletion in *CALR* at 66% VAF. A final diagnosis of MF progressed from MDS/MPN was made.

(iv) Sample W1808576 was a 64 year old female who had a *JAK2* V617F mutation at 5% VAF and final diagnosis of ET. The TSMP was also undertaken and the results showed a *DNMT3A* mutation.

In addition, two samples had also had the TSMP undertaken and no additional mutations were detected: (1) sample W1813661: 39 year of female, 2% *JAK2* V617F mutation, diagnosis of ET; (2) sample W1903488: 68 year old female, 5% *JAK2* V617F mutation, diagnosis of ET.

Of those patients reported to be under active patient management (currently [n=51] or up until the patient's demise [n=6]), the time interval between patient appointments for monitoring purposes ranged from 1-12 months.

Of those patients reported to be receiving therapy (n=48; which equates to 77% of all cases and 23% of those with a confirmed diagnosis) the specific therapies being administered were provided in 18 cases: 15 patients were reported to be receiving aspirin, of which 5 patients were also receiving hydroxycarbamide. Two patients were receiving hydroxycarbamide in isolation and one patient was reported to be on EPO to treat co-occurring MDS.

(2) Results from samples that were shown to be triple-negative by the diagnostic MPN panel but within which a low level (<1% VAF) JAK2 V617F mutation was detected by ddPCR

From the samples that were shown to be triple-negative by the diagnostic MPN panel but within which a low level (<1% VAF) *JAK2* V617F mutation was detected by ddPCR (n=20), I received 8 returns (40%). A summary of the key findings is provided below; the full results are displayed in Appendix 11.

The median age of the patients was 69 years (range 48-81) and the sex distribution equal between males and females. Of the 8 cases with audit forms returned, a final diagnosis was given in 5 samples. In one sample (Low 1) the final diagnosis of MPN (sub-type not provided) was made and this was diagnosed solely on clinical features. The only genetic test undertaken in a diagnostic setting at the WRGL for this sample was the diagnostic MPN panel, but the TSMP was undertaken on this sample as part of this study (see below) and identified no variants (pathogenic/likely pathogenic/VUS); a subsequent sample has not been received from this patient at the WRGL. Of note, we were told that this patient and only one further patient in this group (Low 72) were currently under active management. For Low 1, we were not informed of the treatment being administered; for sample Low 72, we were told that this patient had not been discharged in spite of the absence of a diagnosis of neoplasia as they are routinely in requirement of venesection to manage idiopathic erythrocytosis.

For the remaining 4 samples where a final diagnosis was given, the diagnoses were as follows: low protein S (n=1), secondary polycythaemia (n=1), idiopathic erythrocytosis (n=1) and giant cell arteritis and iron deficiency (n=1). Respondents said that there was no final diagnosis in 3 cases (Low 5, Low 20 and Sample 206); all three of these patients had been discharged from clinical follow up and were therefore not under active management. However, of note, in one of these patients (Low 5) we were informed that the aberrant counts were thought to be of reactive cause and that this patient is being monitored by their GP on a yearly basis. As described below, all samples in this group had the TSMP undertaken as part of this study (results not reported to clinicians). In 1 out of 3 samples without a final diagnosis (Low 5 but not Low 20 or sample 206), there was evidence of molecular clonality (as detailed below) by TSMP analysis; of note, a VUS was detected in sample 206 but the VAF was at a level that meant that it might have been of germline origin (~50%) and thus may represent a rare polymorphism rather than evidence of clonality.

For 6 out of 8 samples in this group, the diagnostic MPN panel was the only genetic test that was undertaken at the WRGL and reported clinically. For sample Low 20, a separate BM

sample was received 6 months after the original diagnostic MPN panel was performed for G-banding analysis and no clonal abnormalities were detected (46,XY[20]); this result was reported in a diagnostic setting. For sample 206, the diagnostic MPN panel was also undertaken on separate samples received 14 months prior to (Oct 2017) and 23 months following (Nov 2020) this sample and reported as no mutations in *JAK2* exon 12, *JAK2* exon 14, *CALR* or *MPL*; ddPCR was not undertaken on either of these samples.

Of the 8 sample results returned in this group, 3 patients were reported to be currently under routine monitoring (Low 1, Low 5 and Low 72), with a time interval ranging from 4 to 12 months. Of note, two of these samples (Low 5 and Low 72) had clonality indicated by the myeloid panel analysis (as detailed below).

(3) *Results from samples that had a CALR/MPL mutation but were negative for JAK2 V617F mutation by the diagnostic MPN panel but were found to have a low level (<1% VAF) JAK2 V617F mutation detected by ddPCR (n=19)*

From samples that had a *CALR/MPL* mutation and negative for *JAK2* V617F mutation by the diagnostic MPN panel but were found to have a low level (<1% VAF) *JAK2* V617F mutation detected by ddPCR (n=19), I received 11 (58%) returns: 4 returns from *CALR* positive samples and 7 returns from *MPL* positive samples. The median age of the patients for whom I received information was 83 years (range 43-90) and the sex distribution was 1:1.75 females:males. A summary of the key findings is provided below; the full results are displayed in Appendix 12.

Nine out of 11 (82%) patients were reported to have a final diagnosis of ET; a final diagnosis of MDS/MPN-RS-T was given in 1 patient (sample 405). One patient was reported to have not received a final diagnosis (sample 387); however, the respondent stated that a diagnosis of MPN was likely but unable to be confirmed as the patient refused a BM procedure. Of the 11 cases, a final diagnosis was supported by the genetic result from the diagnostic MPN panel in 10 (91%) and in 5 of these cases it was reported as the sole important factor in this decision. In 3 cases (27%), the genetic result was used in combination with clinical information to make a final diagnosis; in 1 case (sample 365), the genetic result, clinical information and the exclusion of other diagnoses were the important factors reported to impact the final diagnosis. In 1 case (sample 405), the genetic result was used in

combination with the BM results to make a final diagnosis (of MDS/MPN-RS-T) and in 1 case (sample 318) the BM results were used as the sole important factor in making a diagnosis.

The majority (n=10) of patients were under active management with monitoring occurring regularly (range every 1-6 months). One (sample 383) was reported to have been discharged; this patient was a 90 year old female who was reported to have declined treatment. Of note, the TSMP was undertaken on this sample as part of this study (results not reported to the referring clinician) and 5 mutations were detected, including the known *MPL* mutation, indicating that this patient may had advanced disease at the time of testing or an MDS/MPN overlap syndrome (further details provided below).

3.4.9 Additional mutations detected in patients with low level *JAK2* V617F

3.4.9.1 TSMP results from low level (<1% VAF) *JAK2* V617F positive samples

To determine if low level *JAK2* V617F may indicate the presence of a larger clone, the TSMP was performed on on all samples with a low level *JAK2* detected by ddPCR and sufficient DNA remaining for this test to be undertaken (n=43 samples). These samples included 27 internal samples and 16 external (GST) samples. The results are summarised in Table 3.16.

Sample name	Cohort	MPN panel result (VAF, if applicable)	*JAK2* V617F VAF by ddPCR	Pathogenic/ likely pathogenic variants detected	VUS detected
Low 1	Low	Triple negative (N/A)	0.05%	0	0
Low 5	Low	Triple negative (N/A)	0.11%	0	1*
Low 9	Low	Triple negative (N/A)	0.65%	0	1
Low 20	Low	Triple negative (N/A)	0.13%	0	0
Low 25	Low	Triple negative (N/A)	0.13%	0	1
Low 26	Low	Triple negative (N/A)	0.87%	1	0
Low 27	Low	Triple negative (N/A)	0.12%	0	0
Low 33	Low	Triple negative (N/A)	0.17%	2	0
Low 41	Low	Triple negative (N/A)	0.07%	0	0
Low 65	Low	Triple negative (N/A)	0.52%	2	0
Low 67	Low	Triple negative (N/A)	0.07%	0	0
Low 70	Low	Triple negative (N/A)	0.16%	0	0
Low 72	Low	Triple negative (N/A)	0.55%	0	2*
Low 79	Low	Triple negative (N/A)	0.49%	0	0
Sample 206	TN	Triple negative (N/A)	0.04%	0	1
Sample 268	TN	Triple negative (N/A)	0.04%	0	0
Sample 275	TN	Triple negative (N/A)	0.21%	0	0
Sample 298	TN	Triple negative (N/A)	0.03%	0	0
Low 6	Low	*CALR* mutated (34%)	0.07%	1	0
Low 16	Low	*CALR* mutated (10%)	0.78%	2	0
Low 28	Low	*CALR* mutated (43%)	0.62%	1	0
Low 69	Low	*CALR* mutated (27%)	0.04%	1	0
Sample 484	*CALR/MPL*	*CALR* mutated (39%)	1.32%	0	0
Sample 367	*CALR/MPL*	*MPL* mutated (19%)	0.15%	1	0
Sample 383	*CALR/MPL*	*MPL* mutated (30%)	0.15%	5	0
Sample 405	*CALR/MPL*	*MPL* mutated (5%)	0.10%	3	0
Low 19	Low	*MPL* mutated (3%)	0.78%	1	0
GST 1	External samples	Not done (N/A)	0.16%	0	1
GST 3	External samples	Not done (N/A)	0.23%	0	0
GST 4	External samples	Not done (N/A)	0.07%	2	0
GST 5	External samples	Not done (N/A)	0.11%	0	0
GST 6	External samples	Not done (N/A)	0.11%	0	0
GST 8	External samples	Not done (N/A)	1.67%	0	0
GST 10	External samples	Not done (N/A)	0.04%	0	0
GST 12	External samples	Not done (N/A)	0.03%	0	2
GST 13	External samples	Not done (N/A)	0.21%	0	0
GST 14	External samples	Not done (N/A)	1.17%	0	0
GST 15	External samples	Not done (N/A)	0.16%	0	0
GST 16	External samples	Not done (N/A)	0.13%	0	0
GST 24	External samples	Not done (N/A)	1.30%	1	0
GST 29	External samples	Not done (N/A)	0.08%	0	0
GST 30	External samples	Not done (N/A)	0.14%	0	0
GST 32	External samples	Not done (N/A)	0.07%	0	0

Table 3.16 Summary of the number of variants for each sample with a low level *JAK2* V617F variant tested by the TSMP, split by pathogenic or likely pathogenic variants and variants of uncertain significance (VUS). Where one or more of the VUS detected could be used as evidence of clonality (i.e. the VAF <40%), an asterisk (*) has been used. Cohorts: "Low": *JAK2* V617F negative by the MPN panel; modified bioinformatic analysis; "TN": Triple negative by MPN panel; no modified bioinformatic analysis; "*CALR*/*MPL*": *CALR*/*MPL* positive; no modified bioinformatic analysis; "External samples": anonymised samples sent from Guy's and St Thomas' Pathology laboratory for low level *JAK2* V617F testing.

In total, there were 18 out of 20 samples that were triple-negative by the diagnostic MPN panel and positive for a low level *JAK2* V617F by ddPCR that had TSMP analysis: 14 samples from the "Low" cohort and 4 samples from the triple-negative; 2 samples from the "Low" cohort had insufficient DNA remaining to do this test. Of these, 3 cases (17%) had one or more pathogenic or likely pathogenic variant (Low 26, Low 33, Low 65); 5 had one or more VUS (Low 5, Low 9, Low 25, Low 72, sample 206) but only in 2 of these cases could the VUS be used as evidence of clonality (Low 5 and Low 72; i.e. VAF <40% and therefore not apparently germline in origin). In 10 (56%) cases, no pathogenic/likely pathogenic variants or VUS were detected by the TSMP. Therefore, overall, only 5 out of 18 samples (28%) that were triple-negative by the diagnostic MPN panel and positive for a low level *JAK2* V617F by ddPCR showed the presence of abnormal clone(s) by TSMP analysis. Further details of the variants detected by TSMP in the samples that were triple-negative by MPN panel and positive for low level JAK2 V617F by ddPCR are provided in Table 3.17.

Variants detected by TSMP analysis of the samples that were triple-negative by the MPN panel but had a low level *JAK* V617F mutation detected by ddPCR					
Sample name	**VAF of the *JAK2* V617F by ddPCR**	**Pathogenic/likely pathogenic variants detected**		**VUS detected**	
		Variant	**VAF**	**Variant**	**VAF**
Low 26	0.87%	*CBL* c.1196T>G p.Leu399Arg	9%	None detected	
Low 33	0.17%	*DNMT3A* c.2644C>T p.(Arg882Cys)	33%	None detected	
		IDH1 c.394C>T p.Arg132Cys	33%		
Low 65	0.52%	*ASXL1* c.2419_2423del p.(Val807Cysfs*13)	33%	None detected	
		U2AF1 c.472_477dup p.(Tyr158_Glu159dup)	28%		
Low 5	0.11%	None detected		*TET2* c.5671_5676del p.(Arg1891_Asn1892del)	12%*
Low 9	0.65%	None detected		*CSF3R* c.2474G>A p.(Gly825Glu)	50%
Low 25	0.13%	None detected		*MYD88* c.538T>C p.(Tyr180His)	54%
Low 72	0.55%	None detected		*ASXL1* c.4562C>T p.(Ala1521Val)	39%
				DNMT3A c.2255_2257del p.(Phe752del)	17%*
Sample 206	0.04%	None detected		*CUX1* c.3118G>A p.(Val1040Met)	47%

Table 3.17 Variants detected by the TSMP in samples that were triple-negative by the MPN panel but showed a low level *JAK2* V617F mutation by ddPCR. Where a VUS detected indicates clonality at a higher level than the *JAK2* V617F clone, an asterisk (*) has been used. Those samples that showed no variants by the TSMP are not shown in this table (Low 1, Low 20, Low 27, Low 41, Low 67, Low 70, Low 79, Sample 268, Sample 275, Sample 298). Transcripts: *ASXL1* NM_015338.5, *CBL* NM_005188.3, *CSF3R* NM_156039.3, *CUX1* NM_001202543.1, *DNMT3A* NM_175629.2, *IDH1* NM_001282387.1, *MYD88* NM_001172567.1, *TET2* NM_001127208.2, *U2AF1* NM_006758.2

Of the 18 *CALR*+/low level *JAK2* V617F+ samples, 5 had TSMP analysis which included 4 samples from the Low cohort and 1 sample from the *CALR/MPL* positive cohort. In 3 of these samples, the sole abnormality by the TSMP was the known *CALR* variant (Low 6, Low 28, Low 69). In 1 sample (Low 16), the known *CALR* variant was detected at 8% was seen with a low level *CBL* variant, detected at 2% VAF. Of note, the *CBL* variant is below the validated level of detection for the TSMP and therefore if this variant was detected in a diagnostic setting, we would not have reported it. Further details of the variants detected in these samples are shown in Table 3.18. In 1 sample (Sample 484), no variants were detected including no evidence of the known *CALR* variant detected by the MPN panel at 39% VAF. For this sample, the TSMP was undertaken in the context of a clinically requested (rather

than research) TSMP investigation in 2017 and at that time, the secondary analysis of the sequencing results was limited in its ability to detect large imbalances such as this *CALR* variant.

Variants detected by TSMP analysis of the samples that were *CALR*+/*JAK2* V617F- by the diagnostic MPN panel but had a low level *JAK2* V617F mutation detected by ddPCR: *CALR*+/low level *JAK2* V617F+					
Sample name	**VAF of the *JAK2* V617F by ddPCR**	**Pathogenic/ likely pathogenic variants detected**		**VUS detected**	
		Variant	**VAF**	**Variant**	**VAF**
Low 6	0.07%	*CALR* c.1099_1150del p.(Leu367Thrfs*46)	44%	None detected	
Low 16	0.78%	*CALR* c.1103_1154del p.(Lys368Argfs*45)	8%	None detected	
		CBL c.1211G>A p.(Cys404Tyr)	2%		
Low 28	0.62%	*CALR* c.1099_1150del p.(Leu367Thrfs*46)	49%	None detected	
Low 69	0.04%	*CALR* c.1099_1150del p.(Leu367Thrfs*46)	13%	None detected	

Table 3.18 Variants detected by the TSMP in samples that were *CALR* positive/*JAK2* V617F negative by the MPN panel but showed a low level *JAK2* V617F mutation by ddPCR. One samples showed no variants by the TSMP (Sample 484) and is not shown in this table. Transcripts: *CBL* NM_005188.3, *CALR* NM_004343.3

Of the 9 *MPL*+/low level *JAK2* V617F+ samples (8 from the *CALR*/*MPL* cohort and 1 from the Low cohort), 4 had TSMP analysis. All four samples showed the known *MPL* variant. In two samples (Sample 383 and Sample 405), additional mutations were detected. In Sample 383, the *MPL* variant was seen alongside pathogenic/likely pathogenic variants in *SF3B1*, *SRSF2*, and *TET2*. In Sample 405, the *MPL* variant was seen alongside an *SF3B1* and a *TET2* pathogenic/likely pathogenic variant. Further detail is provided in Table 3.19.

Variants detected by TSMP analysis of the samples that were *MPL* +/*JAK2* V617F - by the MPN panel but had a low level *JAK* V617F mutation detected by ddPCR: *MPL*+/low level JAK2 V617F+					
Sample name	**VAF of the *JAK2* V617F by ddPCR**	**Pathogenic/likely pathogenic variants detected**		**VUS detected**	
		Variant	**VAF**	**Variant**	**VAF**
Sample 367	0.15%	*MPL* c.1543_1544delinsAA p.(Trp515Lys)	16%	None detected	
Sample 383	0.15%	*MPL* c.1544G>T p.(Trp515Leu)	30%	None detected	
		SF3B1 c.1997A>G p.(Lys666Arg)	5%		
		SRSF2 c.284C>G p.(Pro95Arg)	9%		
		TET2 c.1218_1221delTTCT p.(Ser407Profs*19)	8%		
		TET2 c.5665C>T p.(Pro1889Ser)	7%		
Sample 405	0.10%	*MPL* c.1543_1544delinsAA p.(Trp515Lys)	8%	None detected	
		SF3B1 c.2098A>G p.(Lys700Glu)	34%		
		TET2 c.4106C>A p.(Ser1369*)	5%		
Low 19	0.78%	*MPL* c.1544G>T p.(Trp515Leu)	3%	None detected	

Table 3.19 Variants detected by the TSMP in samples that were *MPL* positive/*JAK2* V617F negative by the diagnostic MPN panel but showed a low level *JAK2* V617F mutation by ddPCR. Transcript *MPL* NM_005373.2, *SF3B1* NM_012433.3, *SRSF2* NM_003016.4, *TET2* NM_001127208.2

Of the 18 external samples identified with a low level *JAK2* V617F mutation identified by ddPCR, TSMP was performed on 16 samples. No variants were detected in 12 samples (Table 3.16). In sample GST 4, a pathogenic variant in *MPL* and *DNMT3A* was detected, indicating that this sample was another *MPL*+/low level *JAK2* V617F+ co-mutated sample. In sample GST 24, the low level *JAK2* V617F variant which was detected at 1.3% VAF by ddPCR was the sole abnormality detected by the TSMP, seen at 3% VAF. The TSMP has a known margin of error for quantitation therefore the difference in allelic burden between the two techniques is likely to represent differences in the uncertainty of measurements by each method. The final two samples had VUS but no pathogenic or likely pathogenic variants (GST 1; *BCOR* and GST 12; *TET2* and *IKZF1*). Further details about the variants detected in these samples are shown in Table 3.20.

Variants detected by TSMP analysis of the external samples received that were low level *JAK* V617F positive by ddPCR					
Sample name	**VAF of the *JAK2* V617F by ddPCR**	**Pathogenic/likely pathogenic variants detected**		**VUS detected**	
		Variant	**VAF**	**Variant**	**VAF**
GST 1	0.16%	None detected		*BCOR* c.3692G>T p.(Arg1231Leu)	99%
GST 4	0.07%	*MPL* c.1544G>T p.(Trp515Leu)	6%	None detected	
		DNMT3A c.1792C>T p.(Arg598*)	17%		
GST 12	0.03%	None detected		*IKZF1* c.1085C>A p.(Pro353Gln)	55%
				TET2 c.4909C>G p.(Leu1637Val)	51%
GST 24	1.3%	*JAK2* c.1849G>T p.(Val617Phe)	3%	None detected	

Table 3.20 Variants detected by the TSMP in external samples sent by Guy's and St Thomas' Pathology Laboratory that showed a low level *JAK2* V617F mutation by ddPCR. Transcripts: *MPL* NM_005373.2, *SF3B1* NM_012433.3, *SRSF2* NM_003016.4, *TET2* NM_001127208.2. Transcripts: *BCOR* NM_001123385.1, *DNMT3A* NM_175629.2, *JAK2* NM_001322194.1, *MPL* NM_005373.2

3.4.9.2 Other molecular testing performed on samples with low level *JAK2* V617F

For each of the 27 internal samples that had a low level *JAK2* V617F mutation detected by ddPCR and TSMP analysis was performed, the patient records within the laboratory information management database at the WRGL were reviewed to try to gather information that could provide further understanding about the timing of origin of the variants detected. In 5 samples, additional molecular testing performed one a separate sample from the same patient provided additional information (Table 3.21).

Sample name	Date of sample receipt	MPN panel result (VAF, if applicable)	*JAK2* V617F VAF by ddPCR	Variants detected by TSMP (P=pathogenic, LP= likely pathogenic, VUS=variant of uncertain significance; VAF %)	Additional genetic testing performed at WRGL
Low 26	23/08/2017	Triple negative (N/A)	0.87%	*CBL* (LP; 9%)	**Test:** diagnostic MPN panel 20/02/2018 **Result**: *JAK2* V617F mutation at 1% VAF (W1800590)
Low 33	03/04/2017	Triple negative (N/A)	0.17%	*DNMT3A* (P; 33%) *IDH1* (P; 33%)	**Test:** TSMP 14/02/2017 **Result**: *DNMT3A* (9%) and *JAK2* V617F (2%) in peripheral blood
Sample 206	30/11/2018	Triple negative (N/A)	0.04%	*CUX* (VUS; 47%)	**Test:** diagnostic MPN panel 12/10/2017 and 12/11/2020 **Result**: Triple negative
Low 69	21/06/2017	*CALR* mutated (27%)	0.04%	*CALR* (P; 13%)	**Test (1):** *JAK2* ARMS testing 27/06/2013 **Result (1)** Negative **Test (2):** *BCR/ABL1* FISH testing 27/06/2013 **Result (2):** Negative
Low 19	28/07/2017	*MPL* mutated (3%)	0.78%	*MPL* (P; 3%)	**Test:** TSMP 07/01/2020 **Result:** *MPL* 2%, *DNMT3A* 6% and *JAK2* 7%

Table 3.21 5 samples from this study had previously had genetic investigations (other than the diagnostic MPN panel) at the WRGL; details are provided.

The results were as follows:

Sample Low 26 was received 23/08/2017. The diagnostic MPN panel was undertaken on this sample but no *JAK* V617F variant was detected at that time; subsequent ddPCR and the TSMP on this sample undertaken as part of this study identified the presence of a low level *JAK2* V617F variant at 0.87% and a likely pathogenic *CBL* variant at 9% indicating clonality. A separate sample was received from this patient almost 6 months later (20/02/2018) for repeat testing as the patient had ongoing polycythaemia; the *JAK2* V617F clone had risen to 1% and was thus detected by the diagnostic MPN panel. Of note, no follow up information about the final diagnosis was provided from the referring clinician about this patient in the clinical audit undertaken as part of this study.

Sample Low 33 was a bone marrow sample received 03/04/2017 and MPN testing was alongside the TSMP on the current sample. The MPN panel and TSMP showed no evidence of a *JAK2* V617F mutation but a *DNMT3A* and *IDH1* variant were identified; the LIMS records for this patient indicated that the *DNMT3A* and a *JAK2* V617F variant were detected at low

level (9% and 2%, respectively) in a separate blood sample from this patient 2 months earlier 14/02/2017 and the patient had been diagnosed with MDS-EB.

Sample 206 was received 30/11/2018 and the MPN panel showed no evidence of a *JAK2*, *CALR* or *MPL* mutation. A separate sample was received from this patient both one year before (12/10/2017) and two years after this sample (12/11/2020) with a referral reason ?MPN, indicating that unexplained symptoms in the patient are still being investigated; both tests were also negative for a *JAK2* V617F, *JAK2* exon 12, *CALR* and *MPL* variant. Results from the clinical audit indicated that no final diagnosis has been made in this patient presenting with persistent mild polycythaemia but otherwise normal blood counts. Although a *CUX1* VUS was detected by the TSMP, the VAF (47%) suggests that this variant may germline in origin rather than representing a heterozygous variant in all cells and therefore overt clonality has not yet been proven apart from the detection of the low level *JAK2* V617F variant.

Sample Low 69 was received 21/06/2017 and the MPN panel identified a *CALR* mutation at 27% VAF; ddPCR identified a low level *JAK2* V617F mutation at 0.04% VAF. A separate sample was received 27/06/2013 for *JAK2* V617F analysis and *BCR/ABL1* FISH, to exclude the diagnosis of CML. Both test results were normal; however, the WRGL performed *JAK2* V617F testing by ARMS in 2013 (LoD approximately 5% VAF) and therefore one cannot exclude that the low level JAK2 V617F variant detected by ddPCR in the current sample was not already present in 2013. *CALR* mutation analysis was not undertaken on the 2013 sample. Follow up from the referring clinician returned as part of the clinical audit (see above) indicates that this patient has ET and is being monitored every 3 months but no further samples have been received for genetic analysis.

Sample Low 19 was a blood sample received 28/07/2017 for MPN panel testing due to stroke at an early age. The MPN panel identified a *MPL* variant at 3% VAF and ddPCR and TSMP analysis undertaken as part of this study identified a low level *JAK2* V617F variant at 0.78% VAF but no additional variants. A separate bone marrow sample was received 07/01/2020 for TSMP testing, with referral information that the patient had ET and suspected transformation to MF. This analysis identified the *MPL* variant at 2% VAF, the *JAK2* V617F variant at 7% and a *DNMT3A* variant at 6% VAF. It is unclear whether the increased size of the *JAK2* V617F positive clone and the presence of variants in the most recent bone marrow sample not previously detected in the blood sample from this patient 3

years previously represented an expansion of this clone to detectable levels or differences in the sub-clonal composition of the blood and bone marrow compartments.

3.4.10 46/1 haplotype analysis

3.4.10.1 ARMS PCR results

The acquisition of *JAK2* V617F in MPN is known to be associated with the constitutional *JAK2* 46/1 haplotype (Jones, et al., 2009). To determine if this association holds true for low level *JAK2* V617F, ARMS analysis was performed on the samples with a low level *JAK2* V617F detected by ddPCR with sufficient DNA remaining (n=37) to determine the prevalence of the 46/1 haplotype in this group of individuals; this testing was also undertaken on population controls (n= 35) and samples known to have a homozygous *JAK2* V617F (n=29) to determine whether the prevalence was significantly different in these groups. The genotype for each sample was scored to be C/C, C/G or G/G at rs12340895 as determined by ARMS PCR with the G-allele marking 46/1. As part of the optimisation, the products of one sample from each haplotype was confirmed by Sanger sequencing of the ARMS PCR products (Figure 3.5).

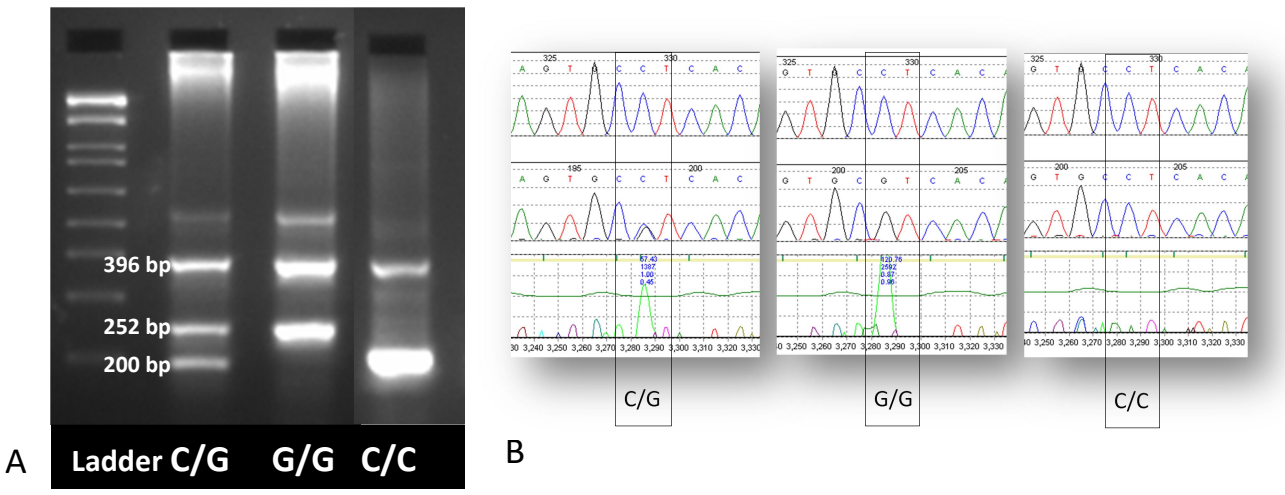


Figure 3.5 A. A gel image of ARMS products for C/G, G/G and C/C haplotype products. The product size for the C allele was 200 bp, the product size for the G allele was 252 bp and the product size of the two outer primers (control band) was 396 bp. B. Sanger sequencing confirmation of the ARMS products for each haplotype: C/G, G/G and C/C, as labelled.

Results from the samples with a low level *JAK2* V617F mutation by ddPCR showed that 15 samples (41%) showed a C/G genotype, 20 samples (54%) showed a C/C genotype and 2 samples (5%) showed a G/G genotype. One sample failed (Low 23).

46/1 haplotype analysis results for those samples with a low level *JAK2* V617F mutation detected by ddPCR (n=37)			
Sample name	**Cohort**	**Diagnostic MPN panel result**	**Result**
Low 1	Low	Triple negative	C/C
Low 5	Low	Triple negative	C/G
Low 9	Low	Triple negative	C/C
Low 20	Low	Triple negative	C/G
Low 24	Low	Triple negative	G/G
Low 25	Low	Triple negative	C/C
Low 26	Low	Triple negative	C/G
Low 27	Low	Triple negative	C/C
Low 33	Low	Triple negative	C/G
Low 41	Low	Triple negative	C/C
Low 65	Low	Triple negative	C/C
Low 67	Low	Triple negative	C/C
Low 70	Low	Triple negative	C/C
Low 72	Low	Triple negative	C/G
Low 79	Low	Triple negative	C/G
206	TN	Triple negative	C/C
268	TN	Triple negative	C/C
275	TN	Triple negative	C/C
298	TN	Triple negative	C/G
Low 6	Low	*CALR* mutated	C/G
Low 8	Low	*CALR* mutation	C/C
Low 16	Low	*CALR* mutation	C/C
Low 28	Low	*CALR* mutation	C/G
Low 76	Low	*CALR* mutation	C/C
Low 69	Low	*CALR* mutation	C/C
Sample 318	*CALR/MPL*	*CALR* mutation	G/G
Sample 365	*CALR/MPL*	*CALR* mutation	C/C
Sample 384	*CALR/MPL*	*CALR* mutation	C/G
Sample 313	*CALR/MPL*	*MPL* mutation	C/G
Low 19	Low	*MPL* mutation	C/G
Sample 335	*CALR/MPL*	*MPL* mutation	C/G
Sample 336	*CALR/MPL*	*MPL* mutation	C/C
Sample 347	*CALR/MPL*	*MPL* mutation	C/G
Sample 367	*CALR/MPL*	*MPL* mutation	C/C

Sample 383	*CALR/MPL*	*MPL* mutation	C/G
Sample 387	*CALR/MPL*	*MPL* mutation	C/C
Sample 405	*CALR/MPL*	*MPL* mutation	C/C

Summary

Genotype	Number samples (%)
C/G	15 (41%)
C/C	20 (54%)
G/G	2 (5%)

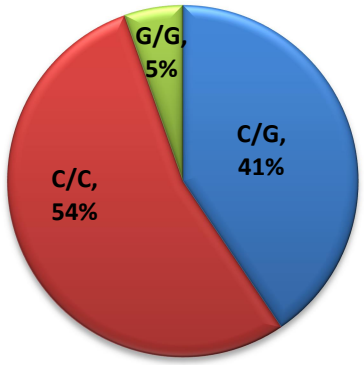


Figure 3.6 Summary of the 46/1 haplotype analysis results performed by ARMS for each sample with a low level *JAK2* V617F detected by ddPCR Cohorts: "Low": *JAK2* V617F negative by the MPN panel; modified bioinformatic analysis; "TN": Triple negative by MPN panel; no modified bioinformatic analysis; "CALR/MPL": CALR/MPL positive; no modified bioinformatic analysis.

Results from the population controls which comprised 35 samples referred to the WRGL for diagnostic testing for reasons other than a confirmed or suspected haematological neoplasia (namely suspected cystic fibrosis and infertility investigations) showed that 8 samples (23%) had a C/G genotype, 22 samples (63%) had a C/C genotype and 5 samples (14%) had a G/G genotype (Figure 3.7).

46/1 haplotype analysis results for control samples (n=35)		
Genotype	**Number samples (%)**	
C/G	8 (23%)	
C/C	22 (63%)	
G/G	5 (14%)	

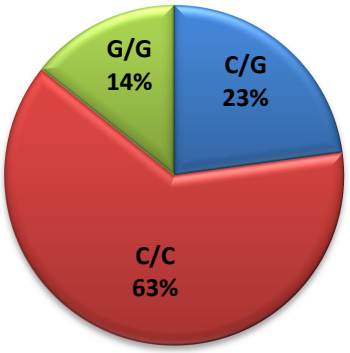


Figure 3.7 Summary of the 46/1 haplotype analysis results performed by ARMS for the population control samples (n=35). These samples were referred to the WRGL for diagnostic testing for reasons other than confirmed or suspected haematological neoplasia (cystic fibrosis investigations or infertility investigations).

Results from MPN cases known to be homozygous for a *JAK2* V617F mutation (VAF range 82-95%; median 87%) referred to the WRGL for MPN panel testing with confirmed or suspected MPN showed that 14 samples (48%) showed a C/G genotype, 9 samples (31%) showed a C/C genotype and 6 samples (21%) showed a G/G genotype (Figure 3.8).

46/1 haplotype analysis results for samples with a homozygous JAK2 V617F mutation (n=29)		
Genotype	**Number samples (%)**	
C/G	14 (48%)	
C/C	9 (31%)	
G/G	6 (21%)	

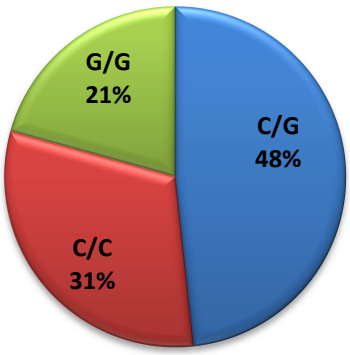


Figure 3.8 Summary of the 46/1 haplotype analysis results performed by ARMS for xxx

3.4.10.2 Comparison of allele frequencies between groups

For the 35 controls, the frequency of 46/1 (G allele for rs12340895) was 26% (=18/70). This is very similar to the frequency in the Wellcome Trust Case Control Consortium estimated by imputation (frequency = 26%; n=5195; data kindly provided by Dr William Tapper, University of Southampton) as well 1000 Genomes Project (frequency = 25%; n=5008; https://www.ncbi.nlm.nih.gov/snp/rs12340895).

For the MPN cases known to be homozygous for a *JAK2* V617F mutation, the frequency of 46/1 was 45% (=26/58). As expected from previous studies (Jones, et al., 2009), the frequency of 46/1 is significantly higher in these cases compared to controls (P=0.026; Fisher's exact test).

For the low level *JAK2* V617F mutation positive cases, the frequency of 46/1 was 26% (=19/74) which is indistinguishable from the controls but significantly less that the frequency in *JAK2* V617F homozygotes (P=0.027; Fisher's exact). Thus, based on these relatively small numbers, there is no evidence that low level *JAK2* is associated with the *JAK2* 46/1 haplotype.

3.5 Discussion

The acquisition of clonal mutations in apparently normal individuals is now a well accepted phenomenon (ARCH/CHIP) and is reported to occur in at least 10% of individuals greater than 65 years with increasing prevalence with age. *JAK2* V617F represents a known driver mutation in patients with MPN, but this mutation has also been reported in apparently normal individuals with a prevalence of 0.1 - 3.1% in population cohorts (Nielsen, et al., 2013; Cordua, et al., 2019; Jaiswal, et al., 2014), with a higher prevalence reported in studies that employed more sensitive assays. In the majority of cases identified, the VAF was <10%, with higher VAF being positively associated with increasing age, and reports indicate that low level *JAK2* V617F-positive clones can remain stable for several years (Cordua, et al., 2019; Nielsen, et al., 2014; Gale, et al., 2007). Only a proportion of *JAK2* V617F-positive individuals identified in populations studies are later identified as having an MPN but there is evidence that certain haematological characteristics are more common in *JAK2* V617F mutation individuals without a recognised MPN when compared to *JAK2* V617F-negative individuals, such as erythrocytosis, neutrophilia and thrombocytosis (Wouters, et al., 2020; Cordua, et al. 2019). A key question that arises from these observations from a genetic diagnostic perspective is what is the most appropriate cut off for *JAK2* V617F VAF to diagnose or help to diagnose MPN?

In this study, I add to what is currently understood about the clinical significance of low level *JAK2* V617F in individuals with confirmed or suspected MPN by investigating four main areas: firstly, the prevalence of low level *JAK2* V617F (<1% VAF) was investigated in individuals with confirmed or suspected MPN *versus* population controls. Secondly, I report on the outcomes collected from a clinical audit undertaken on individuals with confirmed or suspected MPN in whom a *JAK2* V617F was detected at 1-5% VAF in a diagnostic setting and individuals that were reported as *JAK2* V617F negative by diagnostic laboratory testing but were low level (<1% VAF) *JAK2* V617F positive by this study. Thirdly, I explore whether additional mutations identified in low level (<1% VAF) *JAK2* V617F positive are relevant to individuals with confirmed or suspected MPN. Fourthly, I explore whether the 46/1 haplotype predisposes to low level *JAK2* V617F mutations.

The prevalence of low level (<1%) JAK2 V617F in individuals with confirmed or suspected MPN

ddPCR (LoD 0.03% VAF) with or without a bioinformatic pre-screen was used to identify low level *JAK2* V617F in individuals that were negative for *JAK2* V617F by diagnostic laboratory testing (LoD 1% VAF). (Of note, the prevalence of low level (<1%) *JAK2* V617F in the control population studied was in line was the results from Cordua, et al. (2019), a population study that was able to detect *JAK2* V617F at allelic burden as low as 0.009% VAF.).

Overall, I found no evidence that low level *JAK2* V617F was enriched in cases referred for MPN testing that tested negative for all MPN driver mutations by the diagnostic pipeline compared to controls. This result suggests that our standard diagnostic 1% VAF cut off is unlikely to be missing many cases of true MPN with low level *JAK2* V617F. Although the difference was not significant, the prevalence of low level positives was actually higher in the query MPN group: 20/662 (3%) versus 4/197 (2%) in controls suggesting the possibility of a small difference that my study was not powered to detect. It is interesting to note that had the sample size been 10x larger for both the query MPN and control groups (i.e. control group, n=1970 and sample group, n=6620) and the prevalence had been the same then this would have been statistically different between these two groups (P=0.019). Since these numbers were prohibitively large, I did not attempt to expand my study cohort any further, although I validated that the bioinformatic pre-screen greatly reduced the number of query MPN cases that would need to be tested by ddPCR.

Conversely, there was a statistically significant enrichment of low level (<1%) *JAK2* V617F in both the *CALR*+ and *MPL*+ individuals identified by diagnostic laboratory testing *versus* population controls (P=0.018; 7.4% [11/149] and P=0.006; 15% [8/52], respectively). To my knowledge, extensive testing to determine the frequency of low level *JAK2* V617F has not previously been undertaken in *CALR*+ and *MPL*+ MPN patients; however, these frequencies are not dissimilar from Kang, et al., (2016) who reported low level *JAK2* V617F in 9% (n=7) of *CALR*+ ET patients (n=74) and Guglielmelli, et al., (2007) who reported a low level *JAK2* V617F in 22% (n=4) of *MPL*+ patients with MF (n=18). The reason why low level *JAK2* V617F is seen in association with a larger *CALR* or *MPL* positive clone is unclear. Potentially, *CALR* or *MPL* mutations might induce genomic instability, making the acquisition of a secondary *JAK2* mutation more likely. If this was the case it would be expected that *JAK2* V617F and *CALR/MPL* would always or generally be in the same clone. This would be challenging to

explore given the low level of the *JAK2* V617F mutation but could be addressed in principle by high throughput single cell sequencing. It would also be interesting to explore if cases of *JAK2* V617F mutated MPN had a higher prevalence of low level *CALR* or *MPL* mutations compared to controls, although a related study was performed by Cordua, et al., (2019) who demonstrated that *CALR* mutations are ~5-times less frequent in the Danish population but have a higher mean VAF and carriers are ~3-times more likely to have MPN. Alternatively, certain individuals might be predisposed to acquire multiple MPN driver mutations, in which case it would be expected that the mutations would generally occur in different clones. Several constitutional genetic variants have been reported to predispose to MPN (Tapper, et al., 2015) and it is possible that these variants also predispose to biclonal disease. Another possibility is that 'predisposition' might be somatic: individuals who develop a CHIP/ARCH clone have an increased risk of developing a haematological malignancy which is usually associated with the acquisition of additional mutations. In this scenario, independent MPN driver mutations may develop as subclones on a common clonal background. However my analysis does not support this: of the 9 cases in my study with low level *JAK2* V617F and either a *CALR* or *MPL* mutation who underwent myeloid gene panel analysis, only 3 cases had additional mutations. Furthermore, none of these 3 had a large clone with mutated *DNMT3A*, *TET2* or *ASXL1*, the 3 genes that account for the great majority of CHIP/ARCH cases (Jaiswal, et al., 2014).

The clinical significance of low level (<5%) JAK2 V617F in individuals with confirmed or suspected MPN

A clinical audit to establish the final diagnosis in individuals sent for diagnostic laboratory testing the *JAK2* V617F was undertaken. In individuals that were shown to have a *JAK2* V617F at 1-5% VAF by this assay (which was reported to the clinician), 89% (n=55) were reported to have been diagnosed with a sub-type of MPN. The finding of a *JAK2* V617F mutation was instrumental in arriving at a diagnosis of MPN in most cases; this is not surprising given that this would fulfil one of the 3 major criteria which must be met for the diagnosis of PV, ET or PMF according to the WHO classification and individuals are likely to be sent for *JAK2* V617F testing if clinical parameters are suggestive of an MPN. Respondents said that a final diagnosis had not been made in 7 individuals (11%). In four of these individuals, *JAK2* V617F was detected at 1-3% VAF and the respondents said that there was

insufficient evidence based on clinical features to classify the cause to be MPN (one patient was reported to have long standing idiopathic erythrocytosis and one patient was tested following a portal vein thrombosis); for two of these individuals a sample was received one year later and the VAF had not significantly changed over time. Of note, the TSMP had not been undertaken on these cases therefore the presence of additional clones that could be contributing to the clinical features in these individuals has not been excluded. In the remaining three patients, a diagnosis was not made for other reasons (such as insufficient records at the hospital centre contacted as patient was apparently being monitored by their GP or patient demise prior to diagnosis). This result is consistent with the findings of Wouters, et al., (2020) who studied the prevalence of erythrocytosis (with or without leukocytosis and/or thrombocytosis) and clonal haemopoiesis in individuals from the general population as part of a cross-sectional analysis of data collected as part of the LifeLines study in The Netherlands: they reported a relatively high prevalence (n=7; 5%) of *JAK2* V617F in this cohort but a diagnosis of MPN was made in only a proportion of these individuals (n=4). Taken together with the data from this study, it is clear that the detection of a *JAK2* V617F is not considered sufficient to diagnosis an MPN in a small proportion of individuals, even in those presenting with clinical features that are connected to MPN-like criteria (e.g. raised haematocrit).

In individuals that were reported to be triple-negative by the diagnostic MPN panel but shown to have a *JAK2* V617F at <1% VAF by ddPCR as part of this study (which were *not* reported to the clinician), I received information on the outcomes from 8 individuals. A final diagnosis of MPN was made in one patient on the basis of clinical features alone (Low 1). One might have expected this case to have a larger clone detected by TSMP but no variants were detected as part of this study; of course, mutations in genes not included within this panel cannot be excluded. These findings are consistent with the suggestion that low level *JAK2* V617F is an unreliable indicator of a diagnosis of MPN. Of note, one of the individuals without a final diagnosis in this cohort was reported to have such high levels of erythrocytosis that regular venesection was required for routine management; this finding is consistent with the previously reported association of low level *JAK2* V617F and erythrocytosis, even in the absence of a definitive diagnosis of MPN.

In individuals with a *CALR* or *MPL* mutation (which was reported to the clinician) but who were found to have a low level (<1% VAF) *JAK2* V617F mutation detected by ddPCR as part

of this study (which was *not* reported to the clinician), I received information of the outcomes from 11 individuals. With this small cohort, we were unable to determine meaningful information about whether the outcomes of these patients were significantly different from those commonly seen in *CALR*- or *MPL*-only mutated MPN patients. Further work could be undertaken to assess the full blood count with clinical colleagues in a larger cohort of samples to assess how the low level *JAK2* mutations might be modifying the clinical phenotype.

Additional mutations identified in low level (<1% VAF) JAK2 V617F- positive individuals with suspected MPN.

TSMP was undertaken on all 18 samples that were TN by the diagnostic MPN panel but low level (<1%) *JAK2* V617F positive by ddPCR and from 8 of these cases I received information about the final diagnosis as part of the clinical audit. Only 3 (17%) cases were shown to have pathogenic or likely pathogenic variants by the TSMP. In all of these samples there were mutations which indicated that the patient may have had a diagnosis of MDS/MPN or atypical CML rather than one of the classical MPN (specifically mutations detected in *CBL*, *IDH1*, and *U2AF1*); unfortunately, I did not received a response from the clinical audit for these cases. In these samples, the mutations were all detected above 5% VAF, indicating that the low level *JAK2* V617F was part of a sub-clone comprising only part of the bulk of the disease or co-occurred in a small, unrelated clone. There were 5 cases (28%) where only VUS were detected by the TSMP. In 2 of these (Low 5 and Low 72), the VUS could be used as evidence of clonality as they were detected at <40% VAF. For case Low 5, a *TET2* VUS at 12% VAF was detected and results from the clinical audit indicated that it was thought that there was a likely reactive cause to the aberrant blood counts detected clinically. For case 72, a final diagnosis of idiopathic erythrocytosis was made and the patient required management with regular venesection; however, an *ASXL1* VUS and a *DNMT3A* VUS were detected at 29% and 17% VAF, respectively. For both of these cases, the VUS detected by the TSMP occurred in genes commonly associated with CHIP therefore it is unclear whether these clones represent a neoplastic clone. For the remaining 3 cases where VUS were detected by the TSMP, the VAF was between 40% and 60% and therefore the germline origin of these variants could not be excluded. For one of these cases (sample 206), the clinical audit results detailed that a final diagnosis had not been made but laboratory records indicate that the

WRGL had received a total of 3 samples across 3 years for diagnostic MPN panel testing suggesting that there was a clinical suspicion of MPN which was persisting over time in spite of negative genetic results; for the remaining two cases, I did not receive audit data. Finally, there were 10 cases that were TN by the diagnostic MPN panel, low level (<1%) *JAK2* V617F positive by ddPCR and normal by the TSMP (i.e. no pathogenic/likely pathogenic variants or VUS detected). Of these samples, only 5 cases had information returned as part of the audit and the final diagnosis was given as follows: MPN (n=1), no final diagnosis (n=1), low protein S (n=1), secondary polycythaemia (n=1) and giant cell arteritis and iron deficiency (n=1). Therefore, to summarise, there was evidence of clonality (pathogenic, likely pathogenic variant or VUS) in only 28% of cases triple negative by the MPN panel and positive for a low level *JAK2* V617F by ddPCR. These findings add to the data collected during the clinical audit described in Chapter 2 (i.e. samples which were *not* tested for low level *JAK2* V617F by ddPCR), which showed that of 20 patients referred for TSMP with suspected MPN no variants were detected in 15 (75%) and the majority of these cases where given a final diagnosis of a non-neoplastic condition (idiopathic erythrocytosis, n=5; or reactive causes, n=5). The remaining 5 cases had a final diagnosis of MDS, MPN-U or ET made on the basis of the full blood count and bone marrow morphology review (n=3) or had an unknown final diagnosis (n=2). It would be interesting to re-visit the cases of idiopathic erythrocytosis from that cohort to look for the presence of a low level (<1%) *JAK2* V617F (which would not be detectable by TSMP) given the association of these two characteristics in population studies (Wouters, et al., 2020; Cordua, et al. 2019).

Overall, TSMP analysis on individuals with suspected MPN with a low level *JAK2* V617F identified by ddPCR indicated that additional mutations in genes other than *CALR* or *MPL* are only seen in a minority of individuals. In the small cohort of cases with follow information obtained as part of the clinical audit, there was no clear correlation between the diagnosis of MPN in spite of a negative diagnostic MPN panel and the presence of variants (pathogenic/likely pathogenic or VUS) by the TSMP. It is possible that additional low level clones were present but not detectable by the TSMP as they were below the LoD of this assay (5% VAF); however, if present, it would be unclear how these additional mutations would result in a clinical MPN-like phenotype when present in such a small proportion of cells.

46/1 haplotype analysis in individuals with a low level JAK2 V617F

Surprisingly, no significant difference in the frequency of 46/1 haplotype was detected in low level *JAK2* V617F positives individuals *versus* population controls. This is in striking contrast to *JAK2* V617F positive MPN patients which are known to be associated with 46/1 (Jones, et al., 2009), and confirmed in my analysis.

However, my finding is consistent with the suggestion that the 46/1 haplotype may in fact be a modifier that contributes to the likelihood of expansion of a low level *JAK2* V617F mutation (McKerrell, et al., 2017), particularly when seen in a homozygous or hemizygous state due to aUPD. In other words, low level *JAK2* V617F on the 46/1 haplotype is likely to expand relatively rapidly and give rise to an MPN phenotype whereas *JAK2* V617F on other haplotypes expands more slowly and is therefore more likely to be detected as a low level clone in the absence of an overt MPN. Additional work is required to further investigate this hypothesis and understand how non-genetic factors contribute to the likelihood of clones of expanding and the rate at which they do expand. This will act to further add to the understanding of the acquisition of driver mutations in apparently normal individuals which will be particularly useful in light of the recent suggestion that they can occur early in life, even *in utero* and later evolve into MPN (Williams, et al., 2020).

Is there sufficient evidence to define a quantitative cut-off for a positive result for JAK2 V617F in a diagnostic setting?

When the laboratory best practice guidelines were written for *JAK2* V617F testing a lower limit cut-off of 1-3% VAF was recommended (Bench, et al., 2013), and was largely informed by the level of sensitivity available for testing options at the time rather than clear clinical evidence that this cut off was clinically appropriate. Our diagnostic pipeline uses a hard LoD cut off of 1%, i.e. any variants below 1% are not seen or reported. Some other centres use tests with a lower LoD in patients at diagnosis but in my study there was no significant difference in the prevalence of low level (0.03-1% VAF) *JAK2* V617F in samples sent for MPN diagnostic panel testing (i.e. confirmed or suspected MPN) *versus* controls thus supporting the continued use of a 1% cut-off in a diagnostic setting.

Evidence from this study and published population studies indicate that low level variants can be detected in up to 3% of individuals from the general population, and whilst there is a reported increased frequency of these mutations in individuals with idiopathic erythrocytosis, it is not clear at this time whether it is possible to determine on an individual level whether the finding represents CHIP, whether the clone will be stable over time or whether the patient will develop MPN. Furthermore, it is not clear what factors could be used to give a reliable estimation of progression (such as the 46/1 haplotype or the presence of additional mutations). In the future, when further information is known about why the rate of expansion varies between individuals, detecting low level *JAK2* V617F might be clinically useful in a subset of cases to prevent thrombotic events and potentially limit the progression/expansion of the clone.

4. Characterisation of the der(6)t(1;6)

4.1 Introduction

4.1.1 Myelofibrosis

Myelofibrosis (MF) is characterised by the hallmark features of reactive deposition of fibrous connective tissue (reticulin) and extramedullary haematopoiesis (Arber, et al., 2016; Schieber, et al., 2019). It can occur as a primary disease (PMF) or secondary disease (SMF) following transformation from PV or ET (Tefferi, 2014); PMF is currently classified as a subtype of MPN along with PV and ET according to the WHO categories of MPN (Swerdlow, et al., 2016).

PMF usually impacts individuals with advanced age but can also occur in younger people. The median survival of PMF is 6 years when >60 but 15 years when <60 (Tefferi, et al., 2018). In all cases, there is a risk of leukaemic transformation to AML, and this occurs in up to 20% of PMF cases. Current drug regimens utilised for these patients tend to ameliorate the symptoms but do not impact the disease course or significantly extend survival, a prime example being the use of the *JAK2* inhibitor ruxolitinib to relieve symptoms and reduce spleen size (Tefferi, et al., 2011). The only known curative treatment for patients with PMF is allogeneic stem cell transplant (SCT); however, this is associated with adverse outcome in up to 50% of patients (Ballen, et al., 2010).

In 2009, an IPSS was developed for patients with PMF, to identify those at risk of inferior outcome (median survival: low risk,135 months; intermediate-1 risk, 95 month;intermediate-2 risk,48 months; high risk, 27 months) and further, to identify those who should be considered for SCT. This algorithm incorporated the following clinical characteristics: patient age, haemoglobin levels, leukocyte counts, circulating blasts and constitutional symptoms (Cervantes, et al., 2009). At this time, adverse cytogenetics was considered to be informative in further stratifying the intermediate risk group only. Later, additional risk models were proposed with more comprehensive consideration of genetic factors (discussed below).

4.1.2 Genetic characterisation of myelofibrosis

As described in more detail above (Chapter 1), MPN driver mutations can be found in ~93% of patients with PMF (*JAK2* V617F, 55%; *MPL* exon 10 mutations, 8%; *CALR* exon 9 mutations, 30%). Additional mutations in genes involved in DNA methylation regulators, splicing, epigenetic modification, RAS pathway, transcription factors, cell signalling and histone modification have also been reported in MF and occur at different frequencies within *JAK2* V617F-, *CALR*- and *MPL*-positive MF (Nangalia, et al., 2013). Overall, the number of mutations per patient has reported to be higher in MF when compared with PV and ET and increases with the age (Nangalia, et al., 2013; Grinfeld, et al., 2018; Vainchenker & Kralovics, 2017). In addition, mutations in spliceosome, epigenetic and RAS pathways are reported to be more prevalent in accelerated phase disease (Grinfeld, et al., 2018). Of note, there is some evidence that triple-negative PMF is more likely to represent MDS with secondary fibrosis rather than primary disease (Vainchenker & Kralovics, 2017); nevertheless, Grinfeld, et al. (2018) report that the outcome of patients with MF (n=309) did not significantly differ in patients with PMF versus MF post-ET or PV (secondary MF).

Cytogenetic and molecular genetic abnormalities are known to impact the prognosis of patients with MF and, over the years, a number of cytogenetic-risk categorisations have been proposed that were found to be independently significant from the stratification based on clinical parameters as part of the original IPSS model proposed by Cervantes, et al, (2009) (Tam, et al., 2009; Hussein, et al., 2010; Caramazza, et al., 2011). Caramazza, et al. (2011) presented a two-tiered cytogenetic risk stratification to identify those at high/low risk of reduced overall and leukaemia-free survival: “unfavourable” and “favourable” which had a 5-year leukemic transformation rate of 7% and 46%, respectively. “Unfavourable” cytogenetic abnormalities included a complex karyotype or sole or two abnormalities that include inv(3), -5/5q-, +8, -7/7q-, 12p-, 11q23 rearrangement, i(17q); all other cytogenetic abnormalities (and a normal karyotype) were categorised as “favourable”. At the time of publication, additional studies have been published with potential prognostic relevance of molecular genetic abnormalities in PMF but it was not until the publication of the following criteria that molecular genetic abnormalities were considered robust enough to incorporate into prognostic scoring systems for PMF: GIPSS (Tefferi, et al., 2017), MIPPS70 (Guglielmelli, et al., 2018) and MIPSS70+ version 2.0 (Tefferi, et al., 2018). In the MIPSS70+ version 2.0 (the most recent score), high risk mutations were defined as those affecting *ASXL1*, *SRSF2*,

EZH2, *IDH1*, *IDH2*, *U2AF1* Q157 and the absence of a *CALR* type 1/like mutation (Tefferi, et al., 2018). In addition, the cytogenetic risk levels were adjusted to 3-tiers (Tefferi, et al., 2018) with the incorporation of a very high risk group and refinement of the favourable risk group to normal karyotype or sole abnormalities of 13q-, +9, 20q-, chromosome 1 abnormalities and sex chromosome abnormalities only. It is advised that patients with a high or very high risk disease according to MIPSS70+ (v2) score are considered for allogeneic SCT and that treatment regimens to modulate disease symptoms are given to intermediate risk patients or those not eligible for SCT.

In addition to predicting overall survival, there is evidence that genetic abnormalities can be used to indicate disease progression. As mentioned above, gain of chromosome 1q is a recognised finding in myeloid neoplasia and is frequently reported in MPN, including PMF. As the sole abnormality, this finding is associated with a favourable prognosis according to MIPSS70+ (v2). However, this abnormality has been reported to be associated with disease progression when acquired during clonal evolution in MF (Najfeld, et al., 2010; Marcellino, et al., 2017); described more fully below. Clonal evolution at the molecular level has also been demonstrated in PMF patients that evolve to secondary AML (Engle, et al., 2015; Vallapureddy, et al., 2019). Engle, et al, (2015) performed WGS on a single patient with PMF transformed to secondary AML, including analysis of follow up samples and a skin sample for germline comparison to determine the order of acquisition of mutations, and showed that there were four main groups of mutations representing distinct stages of disease progression. For example, there were mutations present at the PMF stage but not in the transformation sample (e.g. *MYB*) and vice versa (e.g. *IDH1* and *RUNX1*) and mutations that represented an apparent founding clone (e.g. *JAK2* and *U2AF1*). Vallapureddy, et al. (2019) compiled cytogenetic, molecular genetic and clinical data on 1306 patients with PMF between 1976 and 2017. The median follow-up was 3.2 years (range 0-31) and transformation occurred in 11% of individuals (n=149) audited. In those patients that transformed to secondary AML, there was a higher incidence of a very high risk karyotype and of mutations in *ASXL1*, *SRSF2* and *IDH1*.

4.1.2.1 The der(6)t(1;6)

The der(6)t(1;6)(q21-23;p21.3) is a recurrent, unbalanced chromosome abnormality associated with primary and secondary myelofibrosis (Figure 4.1), resulting in partial trisomy of 1q21–23 to 1qter plus loss of 6p21.3 to 6pter. Although the der(6)t(1;6) has been reported as a recurrent rearrangement in the literature (Dingli, et al., 2005; Reilly, et al., 1997; Najfeld, et al., 2002; Hussein, et al., 2009; Andrieux, et al., 2003; Miller, et al., 1985; Reilly, et al., 1994; Tefferi, et al., 2001), it has not yet been characterised in any detail. What is known about this abnormality and the clinical phenotype associated with it is summarised below.

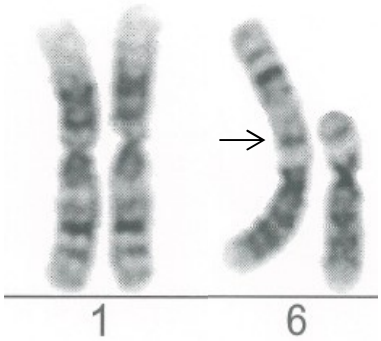


Figure 4.1 The der(6)t(1;6). From left to right: two normal chromosome 1 homologues, the derived (der) chromosome 6 and a normal chromosome 6 homologue. The breakpoint in the der(6) is arrowed, 6p21.3. In this karyogram, 1q material is gained from breakpoint 1q21 to 1qter but the proximal 1q breakpoint may range from q21-q23.

Generally speaking, abnormalities of chromosome 1 are a recognised finding in CMN and Caramazza, et al., (2010) found them the most frequently reported cytogenetic abnormality of *BCR-ABL1*-negative MPNs following extensive literature review. They also report that different subclasses of MPN harboured abnormalities of 1q clustered to specific breakpoints, with abnormalities in MF (post-PV or PMF) clustered around the following breakpoints: 1p13 to 1pter and 1q21-32 to 1q32-44 for duplications; 1p13-36 to 1pter and focally at 1q21 for deletions; and translocations with breakpoints at 1q21-q25, including the der(6)t(1;6)(q21-25;p21.3-23). They therefore conclude that there may be specific oncogenes or tumour suppressor genes within 1q, especially at 1q21-1q32 that explain why these abnormalities are seen in excess in this group of disorders, as well as other 1q abnormalities in other chronic and advanced MPN.

Dingli, et al., (2005) searched for the occurrences of this specific translocation within the samples received for cytogenetic testing by the Mayo Clinic, Rochester, and Royal Hallamshire Hospital, Sheffield, between January 1987 and December 2001 (n=17,791 and n=8,000, respectively). Fourteen (0.05%) individuals were shown to harbour the der(6)t(1;6) abnormality, with breakpoints at 6p21.3 and 1q21 to 1q23, as either the sole abnormality (n=6) or with additional abnormalities (n=8), including 2 patients with a complex abnormal karyotype. All patients were reported to have MF. The gene *FKBP51* on 6p (known to be overexpressed in megakaryocytes) was proposed to be a candidate gene involved in the pathogenesis of MF in these patients but FISH studies indicated that this gene was 8 Mb (centromerically) from the translocation breakpoint and therefore retained. FISH analysis to map the translocation also indicated that the 6p breakpoint was not identical between patients which would not support there being a specific chimeric fusion gene generated in all patients with this rearrangement, or a specific gene that is inactivated. However this analysis was very limited and did not exclude the possibility that a subset of cases might have a common breakpoint or that the rearrangement is more complex at the molecular level.

Reilly, et al. (1994) studied 69 patients with idiopathic MF; they identified one patient with the der(6)t(1;6) as the sole abnormality in 26% cells (9/35) in a 74 year old female with no evidence of leukemic transformation. In 1997, this group expanded their cohort to 106 cases but no additional cases of der(6)t(1;6) were detected (Reilly, et al., 1997).

Miller, et al. (1985) performed cytogenetic analysis on 8 patients with post-PV MF and 20 patients with idiopathic MF; one of the individuals with idiopathic MF had the der(6)t(1;6) and loss of the normal chromosome 6 homologue in 68% cells (20/29). This patient was reported to have deteriorating myelofibrosis and was previously heavily pretreated with cytotoxic therapy.

Andrieux, et al. (2003) reported the clinical and cytogenetic characteristics of 30 individuals with post-PV MF. One patient had the der(6)t(1;6) in 33% of cells (5/15) and an interstitial duplication of 1q (q21 to q42) in 67% of cells (10/15); this individual was a 47 year old male who had PV for 17 years prior to transformation to MF, had received pipobroman and hydroxyurea therapy and had a total survival of 18 years.

Najfeld, et al. (2002) collated the cytogenetic results for 220 patients with PV; abnormalities of 1q were detected in 11.5%. These individuals were further discussed in Najfeld, et al., (2010) where they report the presence of the der(6)t(1;6) in one patient in a cohort of MPN patients who had transformed to AML with gain of 1q in the form of jumping translocations resulting in 1q gain. The der(6)t(1;6) was detected in 2 out of 18 cells 6 years after diagnosis and over time this clone expanded such that this abnormality was later detected in all cells. The patient was treated with interferon and subsequent analyses detected persistence of the der(6)t(1;6) cell line with evidence of the 1q material jumping to other chromosomal locations (chromosome 7 and the Y chromosome) as well as an intrachromosomal duplication of chromosome material. Within this study, additional individuals were reported with unbalanced rearrangements resulting in gain of 1q leading to the authors proposing this as a mechanism of transformation of MPN.

It has been proposed that an important gene involved in MPN progression on chromosome 1q may be *MDM4* (1q32.1; OMIM *602704) (Marcellino, et al., 2017). It is known that the MDM4 protein acts as a negative regulator of the transcription factor p53 which has an important role in tumour suppression and that overexpression/amplification of *MDM4* (and its homologue, *MDM2* located on chromosome 12q) leads to inactivation of p53 (Shvarts, et al., 1996; Toledo & Wahl, 2006). Of note, mutations of *TP53*, particularly in the context of loss of heterozygyosity, have been reported in leukaemic progression of MPN (Lundberg, et al., 2014). Marcellino, et al. (2017) collected cytogenetic information on >600 individuals with MPN (ET, PV, MF or leukemic phase MPN) and cytogenetic abnormalities were detected in >400 patients. Approximately 25% of individuals with abnormal cytogenetics had gain of 1q or a rearrangements of 12q. Gain of 1q was seen as an unbalanced translocation (breakpoints at or around 1q21.2), an interstitial duplication or a jumping translocation. qPCR analysis found that there was overexpression of *MDM4* in 5 individuals with MF and gain of 1q compared with 5 patients with MF but no evidence of 1q gain, supporting the proposition of a role of *MDM4* in MPN transformation. Chromosome 1q and 12q abnormalities were found to be mutually exclusive in this cohort but were both seen more frequently in leukaemic phase MPN or MF rather than in PV or ET ($p<0.001$); of note, *MDM2* expression (by qPCR or another technique) was not investigated by this group in these samples. Abnormalities of 1q also tended to be gained over time, sometimes as secondary abnormalities (~5-7 years after diagnosis of PV or ET) whereas 12q abnormalities occurred only in MF patients as the sole abnormality suggesting different roles in the

development of this disease. The most frequent unbalanced translocation involving chromosome 1q was der(9)t(1;9)(q21;q12), resulting in gain of 1q *and* 9p material; gain of 9p is a known mechanism of transformation in *JAK2* V617F mutated patients.

4.2 Aims and Objectives

As described above, the der(6)t(1;6) is a recurrent abnormality in patients with MF but it has not yet been fully characterised. As such, it is unclear whether this rearrangement results in a specific chimeric gene fusion product and/or whether these patients frequently harbour additional cytogenetically-cryptic abnormalities or mutations which drive (or contribute to) disease pathogenesis.

The principal research questions I aimed to address were as follows:

(1) What is the diversity of der(6)t(1;6) rearrangements, based on cytogenetic, SNP array and WGS profiling?

(2) To what extent does the der(6)t(1;6) define a particular phenotype based on clinical information?

(3) Are there common mutations associated with this abnormality using the TSMP and WGS profiling?

(4) How well does WGS perform for identifying copy number changes and mutations, and does WGS provide any additional findings compared to conventional approaches?

4.3 Materials and Methods

4.3.1 Patient selection

Samples were selected from patients referred to the WRGL for cytogenetic analysis between 1987 and June 2019 with a confirmed or suspected CMN and the der(6)t(1;6). Additional samples with this abnormality were also requested from the Munich Leukemia Laboratory (MLL) (Germany), the University of Perugia (Italy) and the West Midlands Regional Genetics Laboratory (WMRGL, Birmingham, U.K). Clinical information from these samples was obtained from the original referral information received at the time of test request from each laboratory. DNA was extracted in the laboratory of origin, in some cases from cytogenetic cell suspensions.

4.3.2 WGS analysis

WGS was performed by Novagene Ltd. Data was returned and analysed in-house using a customised pipeline employing BWA-MEM, version 0.7.12 (Li, 2013) for alignment to hg38, Picard Tools, version 1.97 (Broad Institute, 2019) to remove duplicate reads from the .BAM file, GATK, version 3.7 (Poplin, et al., 2018) to recalibrate base quality scores in the .BAM file and Annovar, version 2015Dec14 (Wang, et al., 2010) to annotate variants. A virtual gene panel of 102 genes (shown in full in Appendix 13) was generated from publications relating to AML, MDS, MPN, AA, CHIP/ARCH and predisposition to MDS/AML to filter variants to leave only those in clinically relevant genes.

Secondary analysis of variants was performed as follows: variants were excluded if they were recorded on ESP, EXAC or 1000GP at ≥1% MAF or classified as likely benign/benign according to an in-house variant classification system (described in Chapter 2). Variants were also excluded if the alternate allele had <5 reads and/or the total depth was <20x to reduce the likelihood that of technical artefact since novel variants were not confirmed independently.

Candidate breakpoints for the der(6)t(1;6) rearrangements were identified using TopHat, version 2.1.0 (Kim & Salzberg, 2011), with fusions called using TopHat fusion-search and

fusions filtered using TopHat fusion-post. Breakpoints were also manually assessed using IGV and confirmed using specific PCR and Sanger sequencing, where possible.

Copy number changes were called from WGS data using QDNAseq (Scheinin, et al., 2014), DNACopy (Sheshan & Olshen, 2020) and CGHcall (van de Wiel & Vosse, 2020). The genome was divided into non-overlapping 15 kb bins and the number of reads in each bin were assessed (QDNAseq). Read counts were adjusted according to GC content and spurious reads were removed. Circular binary segmentation (CBS) was used (DNACopy) to merge regions with similar copy number (Sheshan & Olshen, 2020) and copy number changes were called using CGHcall.

4.3.3 Mutation analysis by myeloid gene panel analysis

Mutation analysis was performed on samples using the Illumina® TSMP. Further details of this panel are provided in Chapter 2, but in short, the panel amplifies 568 amplicons (equivalent of ~141 kb of genomic content) from 15 full genes plus exonic hotspots of an additional 39 genes known to be tumour suppressor genes or oncogenic hotspots associated with myeloid malignancies (Table 2.2; Chapter 2). Variant classification was undertaken using an in-house pipeline and Alissa Interpret support software (see Chapter 2 for further details).

4.3.4 SNP array analysis

SNP array analysis was performed by the WMRGL using the Affymetrix Cytoscan 750k array. Copy number calls were made using the internal algorithms of the Affymetrix Chromosome Analysis Suite (ChAS) v2.1 software using a virtual Affymetrix reference set.

4.3.5 MLPA

Further investigation of deletions of chromosome 17q detected by SNP array analysis was performed by Multiplex Ligation-dependent Probe Amplification (MLPA) using the MRC-Holland kits P081-D1 and P082-C2. For each sample, 1.7 µl sample DNA diluted to 20ng/µl,

covered in 10 µl of wax to prevent evaporation then denatured according to the following program: 98°C for 5 minutes and 25 °C hold. Once denatured, 0.5 µl MLPA buffer and 0.5 µl of the relevant probe mix were added to each reaction and run on a thermocycler according to the following program: 95 °C for 1 minute and 60 °C hold.

10 µl of the following master mix was then added to each reaction:

Ligase buffer A 0.95 µl/sample

Ligase buffer B 0.95 µl/sample

Water 7.78 µl/sample

Ligase-65 0.32 µl/sample

Ligation was then performed on a thermocycler according to the following protocol: 54 °C for 15 minutes, 98 °C for 5 minutes, 4 °C hold.

3 µl of the following master mix was then added to each reaction:

SALSA primers 0.64 µl/sample

Water 2.5 µl/sample

SALSA polymerase 0.16 µl/sample

The reaction was run on a thermocycler according to the following protocol: 95 °C for 30 seconds, 60 °C for 30 seconds and 72 °C for 1 minute for 33 cycles, then 72 °C for 20 minutes and a 4 °C hold.

Each sample was then run on an ABI3130 machine by adding 1µl of this PCR product to 10µl of Hi-Di formamide and 0.3µl of ROX500 size standards. Results were analysed using the visualisation software, Coffalyser (provided by MRC Holland).

4.4 Results

4.4.1 Patient details

Sixteen cases were selected for molecular analysis on the basis of availability of suitable material; of the 16 cases, 2 were from Italy, 7 from Germany and 7 from the UK (5 from WMRGL and 2 from the WRGL). The characteristics of the 16 cases are summarised in Table 4.1. Overall, there were 5 males and 11 females and the median age was 68 (range 47-79). In these cases, the der(6) was either in all cells (as the sole abnormality, n=2; or with one other abnormality, n=1), the sole abnormality in a proportion (45-85%) of cells (n=9), present with one other abnormality in a proportion (35-40%) of cells (n=2) or present as part of a complex karyotype in a proportion of cells (15-17%; (n=2). In the majority of cases (n=12), the karyotype was associated with a favourable prognosis according to the stratification defined by Tefferi, et al., (2018); in the remaining patients, the karyotype was associated with an unfavourable (n=3) or very high risk (n=1) prognosis. In the majority of cases (n=13), a diagnosis of MPN had been made, either generally defined as MPN (n=5) or more specifically MPN in acceleration/progressing to AML (n=3), MF (n=4) or PV progressing to MF (n=1). In two cases, the individual had been diagnosed with MDS RAEB and in one sample the individual had longstanding MDS (diagnosed in 2001) with suspected transformation to AML.

Clinical information			Cytogenetic results			Molecular analysis				
Sample name (lab)	Sex/ Age	Diagnosis	Karyotype	% of cells abnormal	Risk stratification	SNP array	TSMP	WGS	MLPA	Breakpoint sequencing
E14170 (G)	M/ 75	MPN	46,XY,der(6)t(1;6)(q21;p21)	100	Favourable	Y	Y	Y	N	Y
E14171 (G)	M/ 47	MDS RAEB	46,XY,der(6)t(1;6)(q21;p21)	57	Favourable	Y	Y	Y	N	N
E14172 (G)	F/ 77	PMF	46,XX,der(6)t(1;6)(q21;p21)	80	Favourable	Y	Y	N	Y	N
E14173 (G)	F/ 72	MPN in acceleration	46,XX,der(6)t(1;6)(q21;p22)	76	Favourable	Y	Y	Y	Y	N
E14174 (G)	F/ 75	MDS RAEB	46,XX,der(6)t(1;6)(q21;p22)	65	Favourable	Y	Y	N	Y	N
E14175 (G)	F/ 51	MPN	46,XX,der(6)t(1;6)(q22;p22)	73	Favourable	Y	Y	N	Y	N
E14176 (G)	F/ 57	MPN	46,XX,der(6)t(1;6)(q21;p22)	85	Favourable	Y	Y	N	N	N
E14210 (I)	F/ 60	PV/MF	46,XX,+1,der(6)t(1;6)(q12;p21)[7] /46,XX[13]	35	Favourable	N	Y	N	N	N
E14211 (I)	F/ 68	MF	46,XX,der(6)t(1;6)(q21;p21)[20]	100	Favourable	N	Y	Y	N	N
E14230 (B)	F/ 67	MF	47,XX,der(6)t(1;6)(q2;p21),+9[6]/ 46,XX[9]	40	Unfavourable	N	Y	N	N	N
E14231 (B)	F/ 73	MPN, splenomegaly	46,XX,der(6)t(1;6)(q2;p2)[13]/ 46,XX[7]	65	Favourable	N	Y	N	N	N
E14232 (B)	M/ 67	MF	46,XY,der(6)t(1;6)(q21;p21)[10]/4 6,XY[10]	50	Favourable	Y	Y	N	N	N
E14233 (B)	F/ 74	MPN transformed to AML	46,XX,der(6)t(1;6)(q2;p2)[2]/ 46,XX,der(5)t(3;5)(q2;q1), del(7)(q2q3)[2]/ 46,XX,-5,-7,+mar,+r[5]/ 45,XX,add(1)(q2),-5,-7,+mar[3]	17	Very high risk	N	Y	N	N	N
E14234 (B)	M/ 57	MF	46,XY,add(6)(p21)[3]/ 47,add(6)(p21),+9[5]/ 94,XXYY,add(6)(p21),add(6)(p21), +9,+9[2]/47,XY,+9[4]/ 46,XY,der(6)t(1;6)(q2;p2)[3]/ 46,XY,add(2)(q3)[1]/46,XY[2]	15	Unfavourable	Y	Y	N	N	N
M00002 67 (W)	M/ 67	MF diagnosed 1997. Increased WCC since, ?transformation to AML	46,XY,der(6)t(1;6)(q21;p22), del(13)(q13q31)[30]	100	Unfavourable	Y	Y	Y	N	N
M06066 34 (W)	F/ 79	MDS diagnosed 2001 ?AML blasts, low Hb and Plt	46,XX,der(6)t(1;6)(q21;p25)[9]/ 46,XX[11]	45	Favourable	Y	Y	Y	N	N

Table 4.1 Samples chosen for molecular characterisation in this study. The cytogenetic risk stratification provided was defined according to Tefferi, et al. (2018). The laboratory the sample was obtained from is indicated in the first column: MLL, Germany (G); University of Perugia, Italy (I); WMRGL, Birmingham (B); the WRGL (W). The molecular analyses performed on each sample are indicated by a Y (yes) or N (no).

There were also 7 individuals (2 males and 5 females; aged 56 to 79) identified in the WRGL samples with a der(6)t(1;6) but with insufficient material available for molecular characterisation as part of this study. Full details are provided in Appendix 14 but a summary is as follows: a suspected or confirmed diagnosis of MF or MPN was mentioned in the referral reason of 3 out of these 7 samples; this included MF (n=2), MPN (n=1). All of these patients were reported to be undergoing suspected transformation or had confirmed transformation over the time period that the laboratory received samples. In addition to the der(6)t(1;6), all of these 3 patients had additional abnormalities detected by cytogenetics. In two of these patients, the der(6)t(1;6) was detected in all cells at the time of known disease transformation but was only in a sub-clonal population of cells or not at all at the time of original referral for MF. In one patient, the der(6)t(1;6) was not detected at the time of original referral for MF, but was acquired in a proportion of cells when the patient was referred for suspected transformation. In one patient, we were informed that MPN had been diagnosed 8 years prior to the first and only sample received for cytogenetic analysis; at the time of sample receipt, the patient had suspected AML with a complex abnormal karyotype in two cell lines. This karyotype had several rearrangements, with the der(6)t(1;6) only in 85% of cells therefore, although the timing of acquisition of this finding is unknown, it is likely to represent a late event occurring on a cytogenetically abnormal background. One patient was referred with apparent MDS and was previously treated for non-Hodgkin's lymphoma. In two patients, the patient had MDS that over time progressed to AML ; of note, one of these patient was reported to have had a splenectomy due to an enlarged spleen, which is uncommon in both MDS and AML and is more commonly seen in patients with MPN. All of these patients had the der(6)t(1;6) as the sole abnormality detected by cytogenetics and the abnormality was also present with a normal cell line in all cases. One patient was initially referred with suspected MDS/AML but the immunophenotype was more consistent with B ALL; this patient had received chemotherapy for breast cancer 7 years prior. The complex abnormal karyotype contained 3 cytogenetically unrelated cells lines, including 1 cell with the der(6)t(1;6) as the sole abnormality.

4.4.2 Mutation analysis

4.4.2.1 Mutation analysis by TSMP

All sixteen samples in this study were tested by the TSMP (Table 4.2). Overall, after likely benign and known benign variants were excluded, a total of 43 variants were detected (pathogenic, n=23; likely pathogenic, n=14; VUS, n=6), with each sample having between 1 and 5 variants (median 3). An MPN driver variant (*JAK2, CALR* or *MPL*) was detected in 14 patients (88%). The *JAK2* V617F variant was detected in 9 patients (56%) as either the sole abnormality detected (n=5) or with additional variants (n=4). The *JAK2* V617F VAF was high (>50%) in all but one sample, in which the *JAK2* V617F variant was detected at 8% VAF; in this sample, a (pathogenic) 52 bp deletion in *CALR* was also detected at 63% VAF (E14230). Overall, a pathogenic *CALR* variant was detected in 5 samples as the sole abnormality (n=1, M0000267) or alongside additional variants (n=4), and a pathogenic *MPL* variant was detected in one sample with an additional variant (E14174). In those samples where an MPN driver variant was detected plus additional variants (n=8), the additional high risk variants were known to confer high risk in MF in 75% (n=6) samples [i.e. variants in *ASXL1*, *EZH2*, *SRSF2*, *IDH1*, *IDH2* or *U2AF1* Q157 (Tefferi, et al., 2018)]. In 3 samples, a *CALR* mutation was detected with a high risk variant (E14211, E14232, E14230). In 1 sample, the *JAK2* V617F variant was detected with an *SRSF2* variant but this was a VUS and therefore may not represent high risk in this patient. In two samples, the *JAK2* V617F variant was detected with genes known to be associated with leukemic transformation of MPN: *TP53* (E14233) and *SRSF2* and *IDH2* (E14234).

The 2 samples that showed no evidence of an MPN driver variant by TSMP analysis are described below:

E14171 A total of 5 variants were detected in this sample, of which 2 were the pathogenic variant *SRSF2* p.(Pro95_Arg102del) and a likely pathogenic variant, *TET2* p.(Leu1244dup)]. This patient was referred with MDS RAEB and this mutational pattern would be consistent with that finding; however, of note, this combination of variants has been reported to be commonly associated with CMML (Itzykson, et al., 2013).

M0606634 A total of 5 pathogenic or likely pathogenic variants were detected in this sample: *SRSF2* p.(Pro95Leu), *TP53* p.(Tyr220Cys), *RUNX1* p.(Leu102Cysfs*21), *TET2* p.(Trp954*) and *TET2* p.(Gln1191*). In MDS, variants in *TP53* and *RUNX1* have been

associated with decreased overall survival in multivariable models adjusted for IPSS or IPSS-R risk groups in several studies of distinct cohorts (National Comprehensive Cancer Network, 2019). As above, the combination of an *SRSF2* and *TET2* variant may indicate that this patient had CMML. In CMML, variants in *RUNX1* are associated with an adverse prognosis (Itzykson, et al., 2013; Elena, et al., 2016).

Sample	Total variants	Variants detected by the TSMP					
		Pathogenic		Likely pathogenic		VUS	
		Variant	VAF (%)	Variant	VAF (%)	Variant	VAF (%)
E14170	**1**	*JAK2* c.1849G>T p.(Val617Phe)	75				
E14171	**5**	*SRSF2* c.284_307del p.(Pro95_Arg102del)	57	*TET2* c.5456del p.(Leu1244dup)	49	*NRAS* c.190T>G p.(Tyr64Asp) *TET2* c.3729_3731dup p.(Leu1244dup) *CSF3R* c.2384T>G p.(Leu795Arg)	48 48 36
E14172	**4**	*CALR* c.1099_1150del p.(Leu367Thrfs*46)	54	*TET2* c.4546C>T p.(Arg1516*) *BCORL1* c.1429dup p.(Thr477Asnfs*58)	8 5	*HRAS* c.5C>G p.(Thr2Arg)	71
E14173	**1**	*JAK2* c.1849G>T p.(Val617Phe)	54				
E14174	**2**	*MPL* c.1544G>T p.(Trp515Leu) *DNMT3A* c.2644C>T p.(Arg882Cys)	89 44				
E14175	**1**	*JAK2* c.1849G>T p.(Val617Phe)	83				
E14176	**1**	*JAK2* c.1849G>T p.(Val617Phe)	87				
E14210	**4**	*JAK2* c.1849G>T p.(Val617Phe) *SF3B1* c.1998G>T p.(Lys666Asn)	94 35			*PTEN* c.895A>T p.(Thr299Ser) *SRSF2* c.260T>A p.(Val87Glu)	16 11
E14211	**4**	*CALR* c.1105_1138del p.(Flu369Argfs*50)	52	*ASXL1* c.1934dup p.(Gly646Trpfs*12) *CBL* c.1243G>A p.(Gly415Ser) *CBL* c.1259G>A p.(Arg420Gln)	33 53 3		
E14230	**3**	*JAK2* c.1849G>T p.(Val617Phe) *CALR* c.1099_1150del p.(Leu367Thrfs*46)	8 63	*ASXL1* c.2316_2317insTCTC p.(Glu773Serfs*15)	34		

E14231	**1**	*JAK2* c.1849G>T p.(Val617Phe)	96				
E14232	**3**	*CALR* c.1099_1150del p.(Leu367Thrfs*46)	93	*ASXL1* c.1772dup p.(Tyr591*) *ASXL1* c.1934dup p.(Gly646Trpfs*12)	7 27		
E14233	**2**	*JAK2* c.1849G>T p.(Val617Phe) *TP53* c.687T>A p.(Cys229*)	65 83				
E14234	**5**	*JAK2* c.1849G>T p.(Val617Phe) *IDH2* c.419G>A p.(Arg140Gln) *SRSF2* c.284C>G p.(Pro95Arg)	79 49 52	*CUX1* c.1071dup p.(Gly358Argfs*4) *PHF6* c.820C>T p.(Arg274*)	92 98		
M0000267	**1**	*CALR* c.1099_1150del p.(Leu367Thrfs*46)	57				
M0606634	**5**	*SRSF2* c.284C>T p.(Pro95Leu) *TP53* c.659A>G p.(Tyr220Cys)	48 10	*RUNX1* c.302_303dup p.(Leu102Cysfs*21) *TET2* c.2862G>A p.Trp954*) *TET2* c.6571C>T p.Gln1191*)	47 49 50		

Table 4.2 Mutation profile of samples tested by TSMP. Variant filtering was done according an in-house variant filtering pipeline and variants were also excluded if they were classified as likely benign/benign according to an in-house variant classification system. Transcripts: *BCORL1* NM_021946.4; *CALR* NM_004343.3; *CBL* NM_005188.3; *CSF3R* NM_156039.3; *DNMT3A* NM_175629.2; *HRAS* NM_176795.4; JAK2 NM_001322194.1; *MPL* NM_005373.2; *NRAS* NM_002524.3; *RUNX1* NM_001754.4; *SRSF2* NM_003016.4; *TET2* NM_001127208.2

4.4.2.2 Mutation analysis by WGS

WGS was performed on only 6 of the 16 samples due in part to limited availability of DNA and in part due to available funding. As mentioned above, variants were excluded if they were represented by <5 reads or were detected in a region with <20x total coverage. The average read depth was 46x for reported variants, with a max read depth of 214x. The theoretical (average) sensitivity was therefore 0.11 VAF (=5/46) but dependent upon coverage of that region; the maximum theoretical sensitivity was 0.02 VAF (=5/214).

In total, 41 variants were identified in these samples after initial filtering as described above (Appendix 15); 15 were classified as likely benign, 12 were classified as VUS, 8 were classified as likely pathogenic, and 6 were known pathogenic variants. Fifteen variants were also detected by the TSMP and 26 were not; of these 26 variants, 12 were classified as VUS or likely pathogenic (details provided in Table 4.3) and 14 were classified as likely benign (details provided in Appendix 15). All of the 26 variants that were detected by WGS but not

by the TSMP were in genes or gene regions not covered by the TSMP; all of these variants were unique within each sample, i.e. there were no frequently mutated genes identified in this cohort in addition to those seen by TSMP.

Of the variants detected by the TSMP in these 6 samples (n=17), 2 (12%) were not detected by WGS according to the variant caller used (described above). This included the likely pathogenic variant *CBL* c.1259G>A p.(Arg420Gln) at 3% VAF in sample E14211 and the pathogenic variant *TP53* c.659A>G p.(Tyr220Cys) at 10% VAF in sample M0606634. Analysis of the raw WGS data confirmed there was no evidence of the *CBL* or *TP53* variant by the variant caller used in the respective samples; however, analysis of the raw read pileup for each sample showed that 3/30 reads (~10%) had the expected mutant allele (relative to the positive strand) for *TP53* c.659A>G. There was no evidence in 37 reads for the *CBL* variant (sample E14211) by raw read pileup analysis.

Samples	Number of concordant variants (myeloid panel vs WGS)	Variants detected by the myeloid panel but not detected by WGS		(Possible) Reason for discordancy	Variants detected by WGS not seen by the myeloid panel			(Possible) Reason for discordancy
		Variant	VAF (%)		Variant	VAF (%)	Classification	
E14170	1	None		N/A	*MPL* c.1771T>C p.(Tyr591His)	43	VUS	Region not covered by the myeloid panel
E14171	5	None		N/A	*KMT2C* c.3499+1G>T	11	Likely pathogenic	Genes not included on the myeloid panel
					BRCC3 c.359G>A p.(Trp120*)	32	Likely pathogenic	
					SMC3 c.2535+1G>A	36	VUS	
					CTCF c.38G>A p.(Arg13His)	18	VUS	
					DDX41 c.1098+11delC	37	VUS	
					NF1 c.4514+19G>A	44	VUS	
					KDM5A c.1415A>G p.(Tyr472Cys)	44	VUS	
E14173	1	None		N/A	*STAT3* c.341G>A p.(Arg114His)	44	VUS	Gene not included on the myeloid panel
E14211	3	*CBL* c.1259G>A p.(Arg420Gln)	3	Low level	*PRPF40B* c.2456C>T p.(Ser819Phe)	73	Likely pathogenic	Genes not included on the myeloid panel
					KMT2D c.6742C>T p.(Arg2248Cys)	44	VUS	
M0000267	1	None		N/A	None	N/A	N/A	N/A
M0606634	4	*TP53* c.659A>G p.(Tyr220Cys)	10	Stringency of variant caller (variant detected in WGS data by analysis of raw read pileup)	*DIS3* c.231T>G p.(Ile77Met)	43	VUS	Gene not included on the myeloid panel

Table 4.3 Comparison of variants detected by WGS and the myeloid panel for 6 samples that were tested by both methodologies. Of note, the *TP53* p.(Y220C) variant detected in M0606634 was not detected by the variant caller used to analyse the WGS data but was detected in the analysis raw read pile up (3/30 reads).

4.2.2.3 Overview of mutation patterns detected

Overall, 55 pathogenic, likely pathogenic and VUS variants were detected in a total of 28 genes in the samples tested by TSMP and WGS (n=6) or TSMP only (n=10) (Figure 4.2).

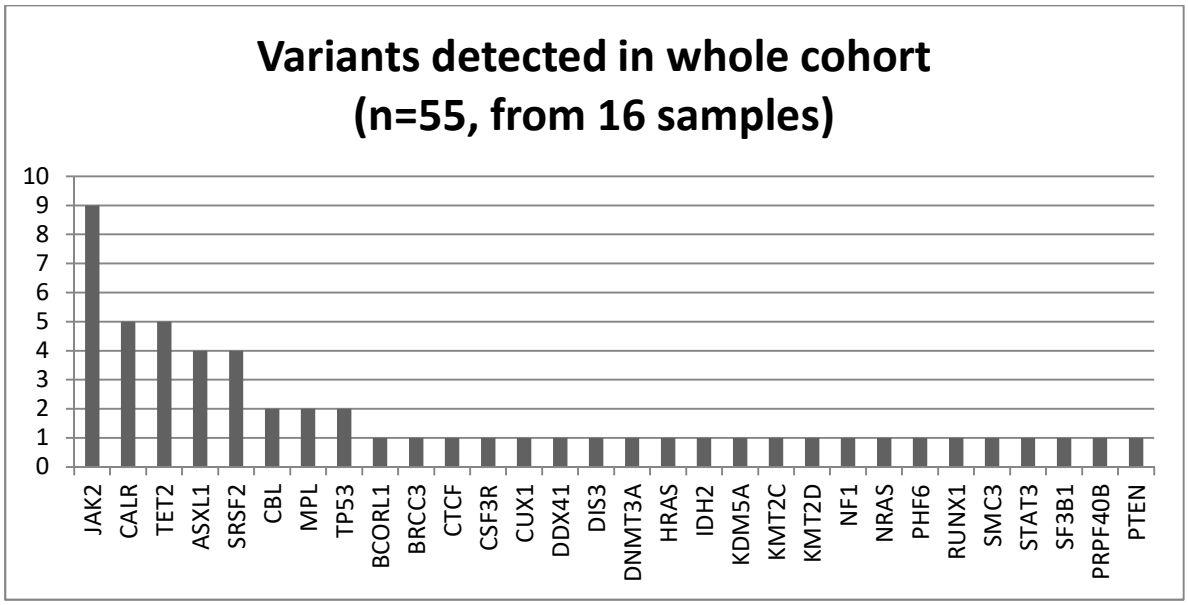


Figure 4.2 The total number of pathogenic (P), likely pathogenic (LP) variants or variants of known significance (VUS) detected in this study by TSMP and/or WGS.

The genes mutated in these samples included those involved in RAS/tyrosine kinase signaling (and other signaling pathways), DNA/chromatin modification, splicing pathways, and the cohesion complex as well as transcription factors (Figure 4.3). The most frequently mutated genes were *JAK2*, *CALR*, *TET2*, *ASXL1* and *SRSF2* with 9, 5, 5, 4 and 4 variants detected, respectively.

		Variants detected (pathogenic, likely pathogenic or VUS)															
		TSMP and/or WGS						TSMP only									
		E14171	M0606634	E14211	E14170	M0000267	E14173	E14234	E14172	E14210	E14230	E14232	E14174	E14233	E14175	E14176	E14231
RAS/TK signaling	*JAK2* V617F																
	CALR																
	MPL																
	CBL			2													
	CSF3R																
	HRAS																
	NRAS																
	STAT3																
DNA/ chromatin modification	*TET2*	2*	2														
	ASXL1											2					
	BCORL1																
	DNMT3A																
	IDH2																
	KMT2C																
	KMT2D																
	KDM5A																
	PHF6																
Splicing pathway	*SRSF2*																
	DIS3																
	DDX41																
	PRPF40B																
	SF3B1																
Transcription factor	*CUX1*																
	CTCF																
	RUNX1																
Other signalling	*PTEN*																
Cohesin complex	*SMC3*																
Other	*BRCC3*																
	NF1																
	TP53																

Total number of variants detected per sample	**12**	**6**	**6**	**2**	**1**	**2**	**5**	**4**	**4**	**3**	**3**	**2**	**2**	**1**	**1**	**1**

Figure 4.3 Summary of results from mutation analysis using WGS and TSMP or the TSMP only. Grey boxes indicate where one or more pathogenic or likely pathogenic variant was detected in each gene; purple boxes indicate where one or more VUS was detected in each gene. When more than one variant was detected, number of variants per gene is provided and in white. The genes underlined are those known to be associated with a poor prognosis in MF. (* for E14171: one *TET2* variant was likely pathogenic and the other was a VUS.)

When benign or likely benign variants were excluded, each sample had between 1 and 12 variants; the median number of variants detected in samples analysed using WGS was 4 versus a median of 3 for those analysed using the TSMP. There was no clear association between the number of mutated genes and the complexity of the karyotype (Figure 4.4).

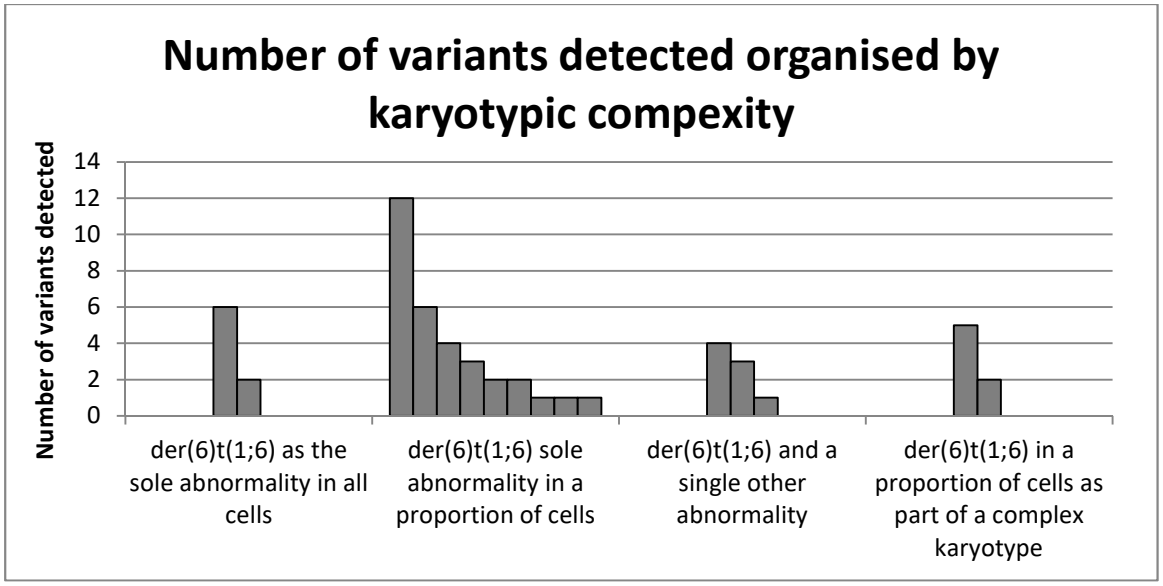


Figure 4.4 Number of variants detected by WGS and/or TSMP in 16 samples with der(6)t(1;6) (after exclusion of likely benign and benign variants) organised by karyotypic complexity, where each grey bar represent a specific sample.

4.4.3 Copy number analysis

4.4.3.1 SNP array results

SNP array results were obtained on 11 samples; aberrations of chromosome 1, 6, 7, 9, 12, 13, 14 and 17 were detected which included deletions, duplications, copy number neutral loss of heterozygosity (CN-LOH) and a whole chromosome trisomy (summarised in Figure 4.5). The specific abnormalities detected, including breakpoints, size of aberration and estimated proportion of cells affected is detailed in Table 4.5.

	Abnormalities detected by SNP array										
Chromosome	E14170	E14171	E14172	E14173	E14174	E14175	E14176	E14232	E14234	M0000267	M0606634
1	LOH										
6											
9	LOH						LOH				
12											
13											
14									LOH		
17											
Cells with der(6) by cytogenetic analysis	100	57	80	76	65	73	85	50	15	100	45

Figure 4.5 Chromosome aberrations detected by SNP array. Only those chromosomes with aberrations detected are shown. Green= segmental duplication; red= segmental deletion; LOH= copy number neutral loss of heterozygosity detected (segmental); blue= whole chromosome trisomy.

Sample name	Chromosome 1 duplication [breakpoints] (% cells)	Chromosome 6 deletion [breakpoints] (% cells)	Other deletions detected	Other duplications detected	Regions of CN-LOH detected
E14170	98 Mb [q21.3 to qter] (65%)	28 Mb [pter to p22.1] (70%)	**Chromosome 13:** 44 kb [within q14.2; including *RB1* ex3-13] (38% cells) **Chromosome 17**: 181 kb [q11.2q11.2; including NF1 ex8-60] (43%)	**Chromosome 12:** Trisomy 12 (24%)	**Chromosome 9:** 35 Mb [pter to p13.3] (~50-100%)
E14171	97 Mb [q21.3 to qter] (35%)	28 Mb (pter to p22.1) (37%).	**Chromosome 13:** 49 kb [within q14.2, including RB1 gene ex 3-7] (36%)	.	
E14172	104 Mb [q21.1 to qter] (53%)	27 Mb [pter to p22.1] (62%)	**Chromosome 13**: 30 kb [within q14.2; including *RB1* ex5 – 11] (54%) **Chromosome 17**: 173 kb [within q11.2; including *NF1* ex5 – 47] (66%)		
E14173	99 Mb [q21.2 to qter] (67%)	23 Mb [pter to p22.3] (68%)	**Chromosome 13**: 27 kb [within q14.2; including RB1 ex6-15] (24%) **Chromosome 17:** 1 Mb [q11.2q11.2; including the whole *NF1* gene] (77%)		
E14174	99 Mb [q21.2 to qter] (33%)	28 Mb [pter to p22.1] (40%)	**Chromosome 13**: 67 kb [within q14.2, including *RB1* exons 5 -15] (53%) **Chromosome 17**: 204 kb [q11.2q11.2; including *NF1* ex 5 – 55] (26%)		**Chromosome 1:** 120 Mb [pter to p11.2] (~100%)
E14175	100 Mb [q21.2 to qter] (65%)	28 Mb [pter to p22.1] (57%)	**Chromosome 13:** 68 kb [within q14.2; including RB1 gene ex3-16] (23%) **Chromosome 17:** 145 kb [q11.2q11.2; including *NF1* ex 13-53] (22%)		
E14176	99 Mb [q21.2 to qter] (67%)	26 Mb [pter to p22.2] (74%)	**Chromosome 13**: 59 kb [within q14.2; including *RB1* ex 3 – 11] (34%).		**Chromosome 9:** 33 Mb [pter to p13.3] (~50-100%)
E14232	101 Mb [1q21.2 to qter] (56%)	28 Mb [6pter to p22.1] (61%)	**Chromosome 13**: 4 Mb [q14.2 to q14.3; including *RB1* gene ex 10 – 28] (72%)		
E14234	**No evidence of a duplication but poor quality**	28 Mb [pter to p22.1] (16%)	**Chromosome 7:** 2 Mb [within q22.1; including whole *CUX1* gene] (93%)	**Chromosome 9:** Trisomy 9 (49%)	**Chromosome 14:** 29 Mb [q24.3 to qter] (~50 - 100%)
M0000267	103 Mb [q21.1 to qter] (92%)	26 Mb [pter to p22.2] (92%)	**Chromosome 13:** 42 Mb [q12.3 to q21.33; including RB1 gene] (89%).		
M0606634	104 Mb [q2.1 to qter] (48%)	**No evidence of a deletion**			

Table 4.5 The size and breakpoints of copy number changes and copy number neutral loss of heterozygosity (CN-LOH) detected by SNP array, as reported by the WMRGL.

The duplication of chromosome 1q material was detected in 91% (n=10) (Figure 4.6; A); the duplication was not detected in sample E14234 and the testing laboratory reported that a duplication of this region could not be excluded due to poor quality of the array. In this sample, the deletion of 6p was detected in 16% of cells, which is consistent with the cytogenetic result indicating that the der(6)t(1;6) was present in 15% of cells. Overall, the breakpoint on 1q was reported to be 1q21.1 to q21.3 by SNP array analysis; this is consistent with the results by cytogenetic analysis, where the breakpoints were reported to be 1q21 (n=9) or 1q22 (n=1). The size of the 1q duplication was reported to be between 97 and 104 Mb. The deletion of 6p was detected in 91% (n=10) with a size range of 23 to 28 Mb (Figure 4.6; B); the breakpoints were 6p22.1 to q22.3 (whereas the breakpoints detected by cytogenetics were 6p21 to p22 in these samples). A 6p deletion was not visible in sample M0606634. There was no apparent sample quality issue and the duplication of 1q was detected at an expected frequency (estimated 48% cells versus 45% seen by cytogenetics) therefore it is possible that the breakpoint, ascribed to 6p25 by cytogenetic analysis, was sub-telomeric and therefore not detectable by SNP array.

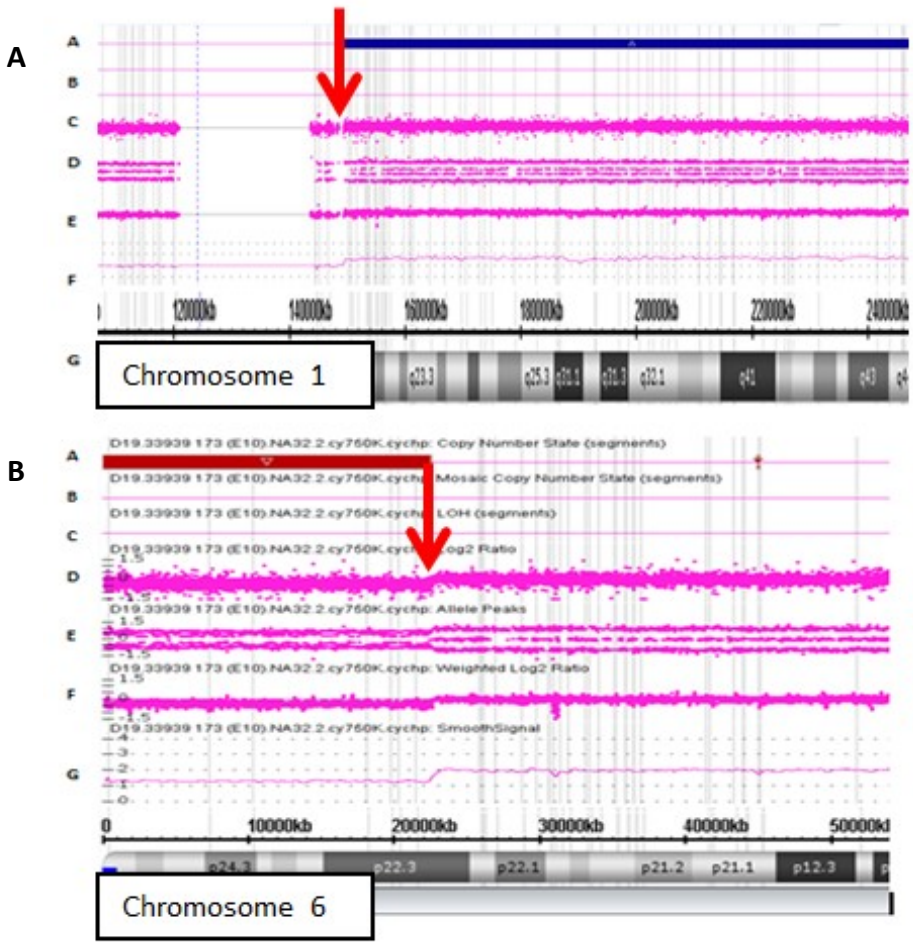


Figure 4.6 **A**. SNP array result for sample E14173 shows a duplication of 1q material, breakpoint q21.2 to qter in 67% of cells (red arrow). **B.** SNP array result for sample E14173 shows a deletion of 6p material, breakpoint pter to p22.3 in 68% of cells (red arrow). Tracks provide detail as follows: (A) copy number state; (B) mosaic copy number state; (C) loss of heterozygosity (LOH); (D) log2 ratio; (E) B allele peaks; (F) weighted log2 ratio; (G) smooth signal. Image provided by the WMRGL.

Strikingly, the array analysis detected recurrent (cytogenetically cryptic) deletions within 13q (example shown in Figure 4.7; A) and 17q (example shown in Figure 4.7; B) in 91% (n=10) and 45% (n=5) of samples tested, respectively (summarised in Figure 4.7; C). For 8 samples, the 13q deletions were less than 1 Mb and occurred within 13q14.2 (27 to 59 Kb) and deleted part of *RB1*; in one sample (E14232) the deletion was 4 Mb in size (13q14.2 to q14.3) also deleting part of *RB1*; in one sample (M0000267) the deletion was 42 Mb (13q12.3 to q21.33) and included deletion of the whole *RB1* gene in 89% of cells; this deletion was seen cytogenetically, alongside the der(6)t(1;6) in 100% of metaphases. The minimally deleted region of 13q14 encompassed exons 3 to 7 of *RB1* but the breakpoints were variable from exon 3 to exon 16 and a whole gene deletion in one instance. The 17q deletions detected in 5 samples were all within 17q11.2, were 145 Kb to 1 Mb in size

encompassing part of *NF1* in 4 samples (with no consistent minimal region of overlap) and the whole of *NF1* in one sample (E14173) in 77% of cells.

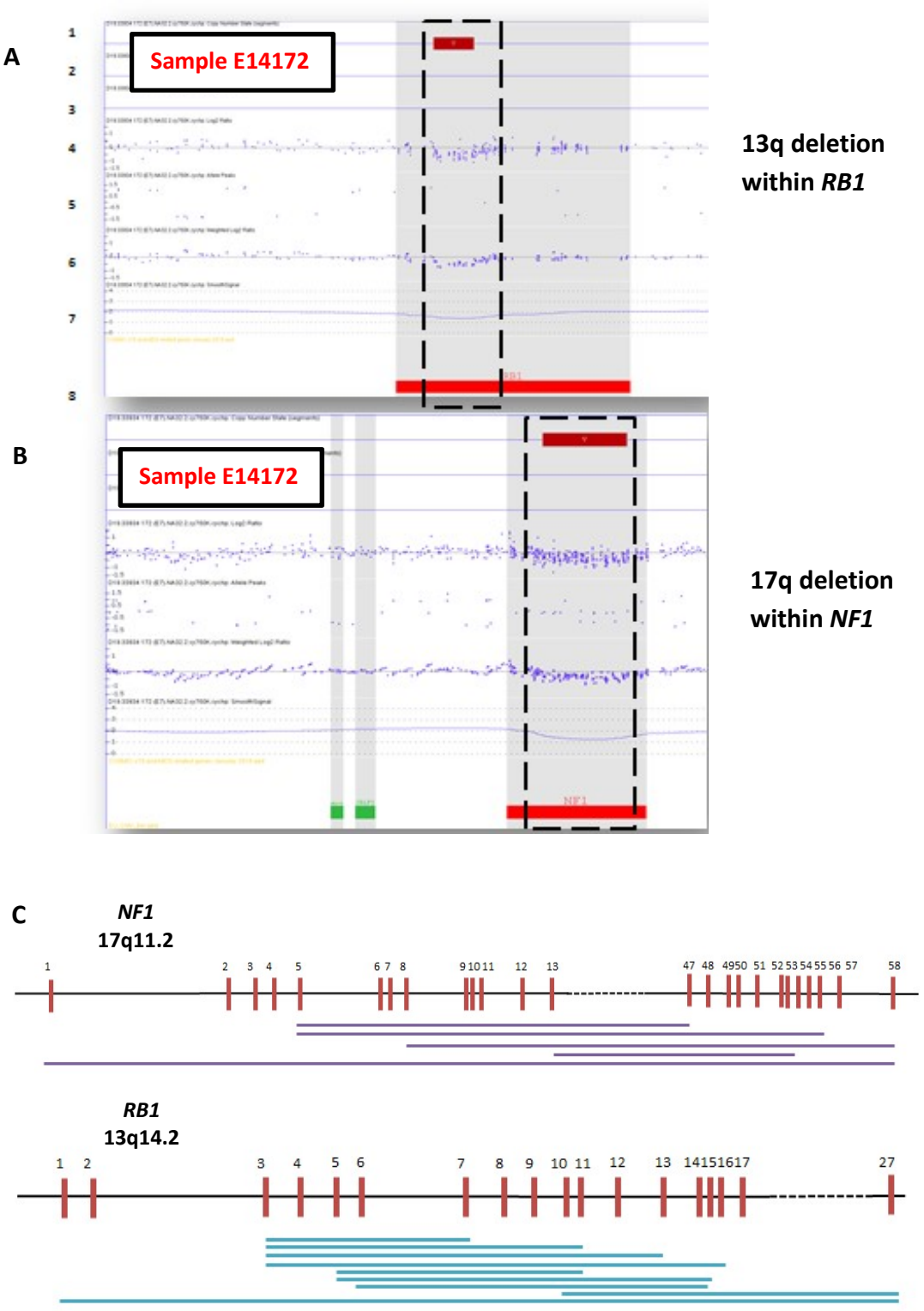


Figure 4.7 A. SNP array result for sample E14172 showed a deletion of 13q material within 13q14.2 in 54% of cells (black dashed line box). **B**. SNP array result for sample E14172 showed a deletion of 17q material within 17q11.2 in 66% of cells (black dashed line box). Tracks provide detail as follows: (1) copy number state; (2) mosaic copy number state; (3) loss of heterozygosity (LOH); (4) log2 ratio; (5) B allele peaks; (6) weighted log2 ratio; (7) smooth signal; (8) OMIM genes. Image provided by WMRGL**. C.** Deletions of *NF1* (purple) and *RB1*

(blue) detected by SNP array displayed as a schematic to demonstrate relative size and exonic involvement. Each line represents an individual sample and shows the size and exonic involvement of each copy number change.

Additional aberrations included a 120 Mb region of CN-LOH on chromosome 1 (pter to p11.2) detected in E14174. There was also a 2 Mb deletion on chromosome 7 (within q22.1) including *CUX1* and a 29 Mb region of CN-LOH on chromosome 14 (q24.3 to qter) in sample E14234 and a 49 Mb and an 81 Mb duplication of chromosome 12 material, from 12pter to 12p13.12 and 12q13.3 to qter, respectively, within sample E14170.

4.4.3.2 MLPA

Four samples were tested by MLPA to confirm the whole gene or partial gene deletions in *NF1* detected by SNP array (E14172, E41473, E14174 and E14175). This locus was chosen rather than *RB1* because of the availability of a validated MLPA kit for the genetic diagnosis of constitutional deletions that predispose to neurofibromatosis. A whole gene deletion was detected in one sample (E14173) by MLPA, consistent with the SNP array result for this sample which indicated a 1 Mb deletion encompassing this region in 77% of cells (Figure 4.8). Deletions were not confirmed in the 3 other cases, however it is important to point out that this technique has not been optimised or validated to detected deletions at sub-clonal levels and the estimated proportion of cells with an *NF1* deletion in the 3 samples that were normal by MLPA analysis was 22%, 26% and 66% by SNP array analysis.

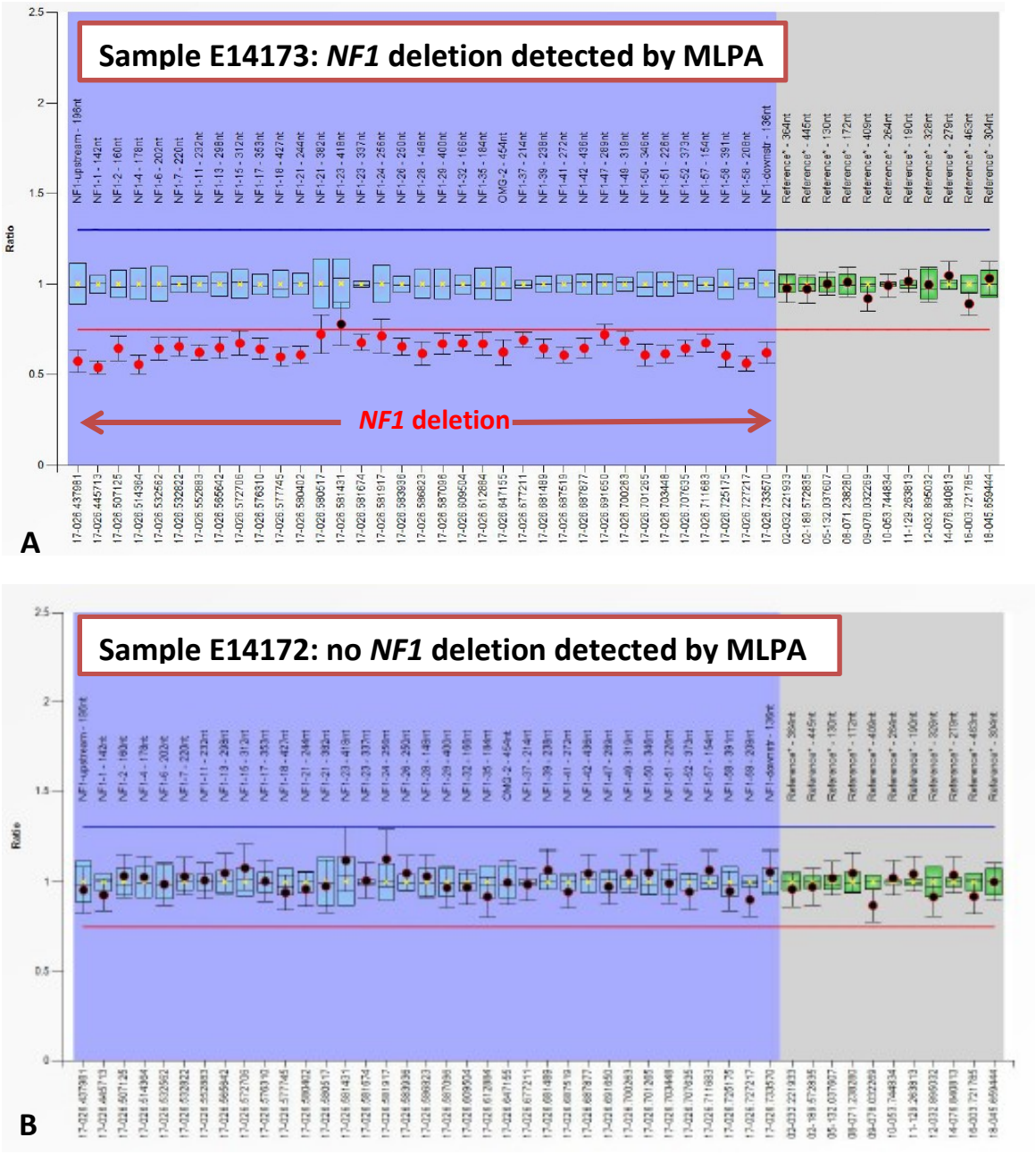


Figure 4.8 Example of *NF1* MPLA results: MRC-Holland kits P081-D1. The probes for *NF1* loci are highlighted by a purple box, the probes for control regions (expected to be in a normal diploid complement) are highlighted by a grey square. A normal diploid complement for each probe would result in a relative ratio of 1; loss of one copy and gain of one copy in all cells would result in a relative ratio of 0.5 and 1.5, respectively. The cut-off for reliable detection of gain and losses versus artefactual skewing in the context of detected *germline* deletions/duplications (calculated by the diagnostic team at WRGL) is shown by the blue and red line, respectively. **A.** Results from E14173 indicating a heterozygous deletion in a proportion of cells for all probes on *NF1* (red dots), confirming the SNP array result. **B.** Results from E14172 where no *NF1* deletion was detected by MLPA, which was discordant with the SNP array result.

4.4.3.3 Copy number analysis by WGS

Copy number analysis was performed on 6 samples (E14170, E14171, E14173, E14211, M0000267 and M0606634) with the principal aim to identify the t(1;6) breakpoints. The

genome wide copy number profiles after CBS (DNAcopy) is shown in Figure 4.9.

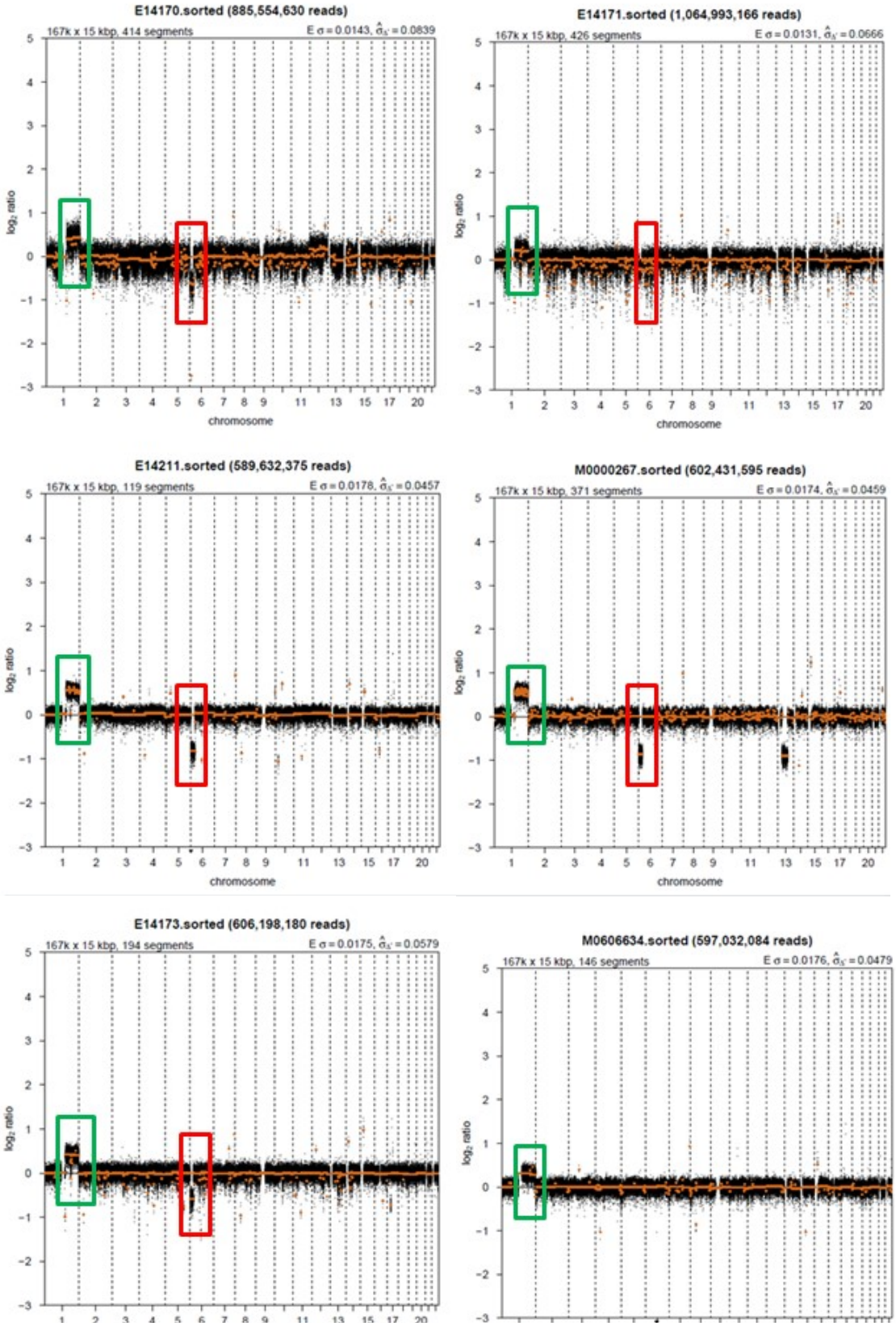
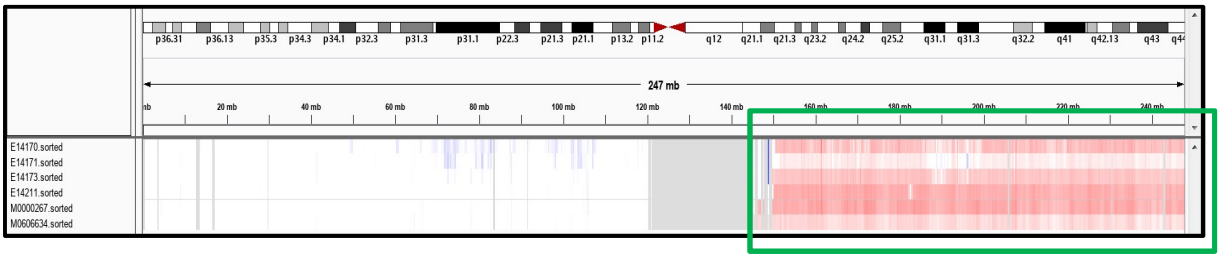


Figure 4.9 Genome wide view of copy number for changes called by QDNAseq following CBS using DNAcopy for 6 samples with a der(6)t(1;6). The gain of 1q material and loss of 6p material is indicated in each sample (where detected) by a green and red box, respectively.

Firstly, copy number changes of 1q and 6p due to the der(6)t(1;6) rearrangement were assessed. Gain of 1q was detected in all 6 samples (indicated by the green boxes in Figure 4.9 and Figure 4.10; A), and it was apparently at low level in two samples (E14171 and M0606634); loss of 6p (indicated by a red boxes in Figure 4.9 and Figure 4.10; B) was seen in 5 samples, and it appear to be at low level in sample E14171, but was not detected in M0606634, consistent with the SNP array findings (see above).

A. Chromosome 1



B. Chromosome 6

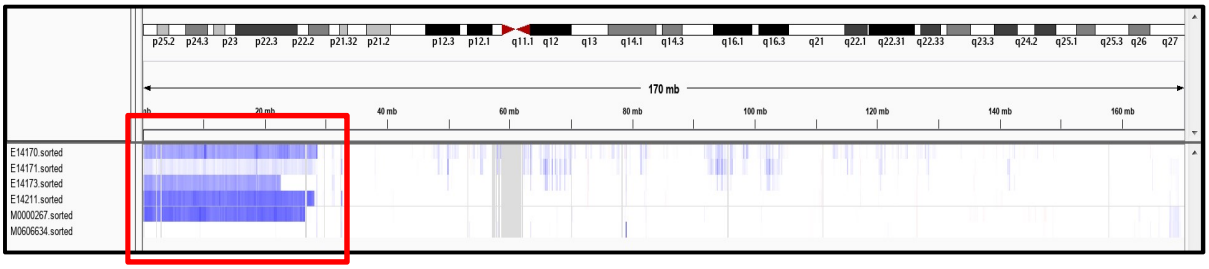


Figure 4.10 Copy number calls identified on chromosome 1 **(A)** and chromosome 6 **(B)** for each sample (identified and called using QDNAseq, DNACopy, CGHcall and visualized in IGV). Gains shown in red tiles and losses shown in blue tiles (each tile = 15 Kb); the confidence of the copy number change call (which is relative to the estimated proportion of cells with the copy number change) is indicated by the strength of the shading: dark shading = high proportion of cells with the aberration; light shading = low proportion of cells with the aberration. (A) gain of 1q was seen in all samples but at lower level in E14171 and M0606634 (green box); (B) loss of 6p was seen in 5 samples (but not M0606634), including at lower level in E14171.

Initially the breakpoints for the der(6)t(1;6) were estimated from the breakpoints of the copy number calls and are detailed in Table 4.6. Some breakpoints were considered less reliable as they were flanked by regions that could not undergo copy number assessment, due to low read depth and/or frequent occurrence of repetitive elements. (Of note, analysis of control samples without the der(6)t(1;6) also had similar stretches of uninterpretable copy number at this genomic location, indicating that this was unlikely to be artefactually caused by genomic complexity resulting from the 1;6 rearrangement.) In addition, one sample (E14171) had apparently low level copy number changes of 1q/6p which made the breakpoint approximation less reliable.

Sample	Breakpoint approximation from copy number analysis data (QDNAseq)	
	Chromosome 1q (hg38)	Chromosome 6p (hg38)
E14170	150,522,524	28,667,223
E14171	151,662,524	28,457,223
E14173	149,878,450	22,784,771
E14211	150,312,556	28,232,222
M0000267	147,063,456	26,669,772
M0606634	146,055,003	No copy number change detected

Table 4.6 Approximate breakpoints for the gain of 1q and loss of 6p detected by WGS. Those breakpoints in red/blue are considered less reliable as they were flanked by regions that could not be interrogated due to low depth or false positives (red) or were difficult to determine (blue) as the copy number changes were at low level.

Copy number changes were also assessed across the rest of the genome. Due to the difficulties of determining which changes were significant in this small cohort size, only those changes that were present in the majority of samples or were present in single samples but were large in size (>10 Mb) were considered for further investigation. Significant copy numbers changes were identified on chromosome 13 (loss), chromosome 15 (gain) and chromosome 21 (gain), shown in Figure 4.13.

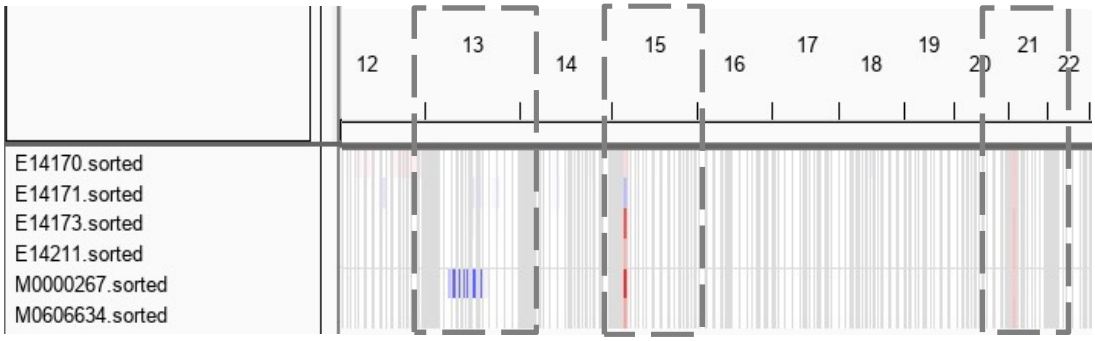


Figure 4.11 Significant copy number changes identified by QDNAseq whole genome analysis identified a deletion in 13q in one sample (M0000267) and duplications of chromosome 15 and chromosome 21 material in multiple samples, indicated by boxed regions. Other than the gain of 1q and loss of 6p seen as part of the der(6)t(1;6) rearrangement, no additional significant copy number changes were detected.

The chromosome 13 deletion detected in M0000267 was an apparently hemizygous deletion between 13q12.3 and 13q22.1, approximately 40 Mb in size. This deletion was also detected by SNP array analysis on this sample at 42 Mb (13q12.3 to q21.33). SNP array also detected smaller deletions within 13q at *RB1* less than 1 Mb in size in samples E14170, E14171 and E14173 but these were not detected by WGS analysis (Appendix 16). By WGS, there was no evidence of the 17q deletions detected by SNP array analysis in samples E14170 and E14173.

4.4.4 Gene fusion analysis

4.4.4.1 *In silico* gene fusion analysis

As expected from the diversity of breakpoints, no gene fusions were detected by TopHat fusion search within 100kb of the copy number changes called by QDNAseq. Therefore the breakpoints predicted by the copy number changes identified by QDNAseq were visualised in IGV for each sample to determine more precise breakpoint predictions. Rearrangement breakpoint candidates were determined in 5 out of 6 samples by this method, supported by split pair analysis (i.e. identifying mate reads mapping to the translocation partner chromosome), split reads (reads which map to chromosome 1 *and* chromosome 6 material) and the results from BLAT alignment. An example is shown in Figure 4.12.

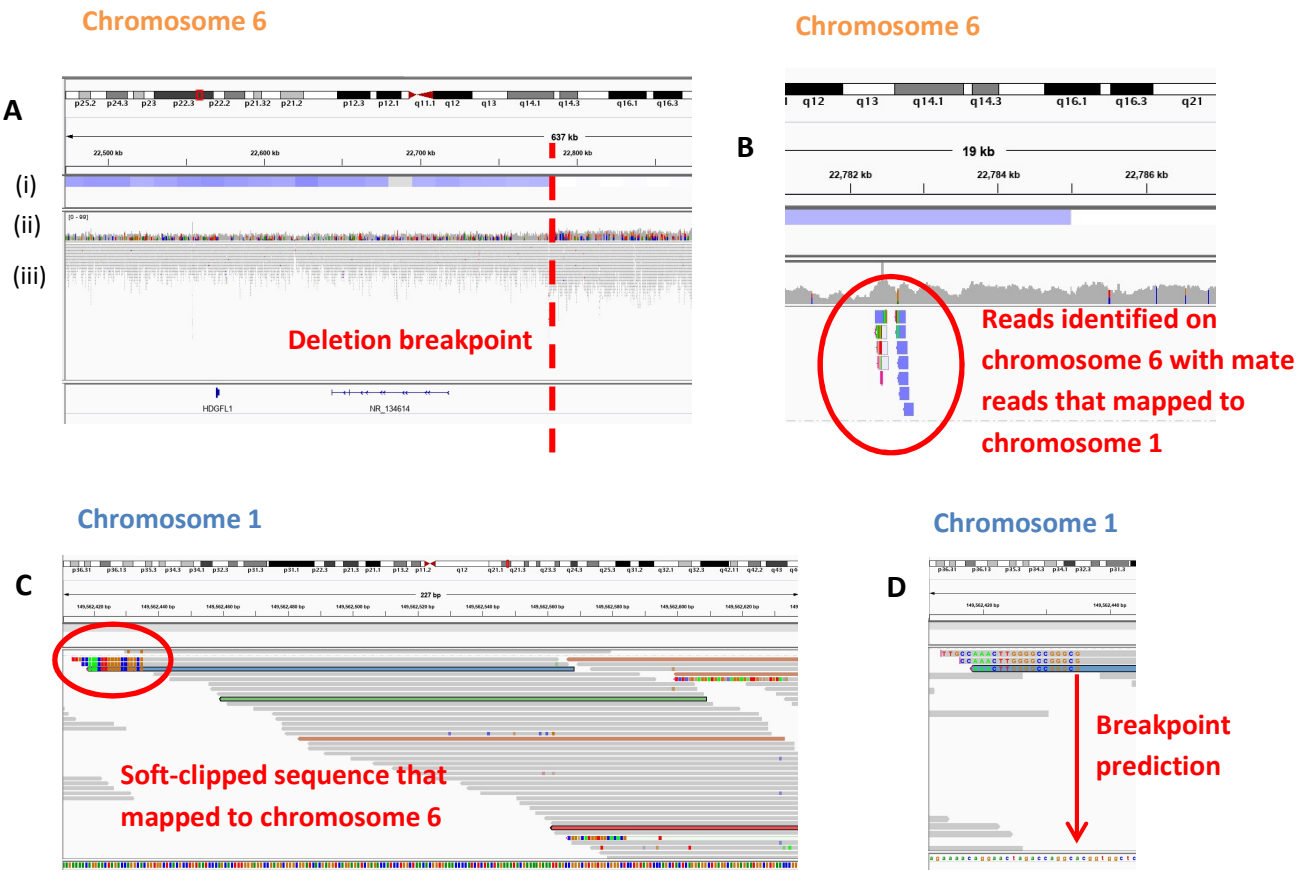


Figure 4.12 Using IGV to determine more accurate prediction of chromosome 6p breakpoints (sample E14173). **A.** Drop in read depth (red dashed line) corresponds with the deletion detected using QDNAseq (blue) [(i) QDNAseq copy number call; (ii) Compressed BAM coverage; (iii) Aligned reads]. **B.** 12 reads (circled) were detected with mate reads mapped to chromosome 1. Of these, 7 reads pointed towards the breakpoint (arrowed end of read) and mates mapped to predicted chromosome 1 deletion breakpoint**. C.** Reads are identified on chromosome 1 with mate reads mapped close to predicted 6p breakpoint. Reads are identified with soft-clipped sequence that map to sequence near to predicted 6p breakpoint (circled) and (**D)** this soft-clipped sequence is used to identify the more precise breakpoint prediction (arrowed).

Breakpoint locations identified this method in the 5 samples where copy number changes were detected by WGS (i.e. not sample M0606634) are summarised in Table 4.7. No candidate fusion genes were identified by this method but one breakpoint fell within a gene for 3 samples: *SNX27* at 1q (E14171); *PDE4DIP* at 1q (M0000267) and *ZSCAN9* at 6p (E14211).

	Breakpoint prediction by QDNAseq		Breakpoint prediction by IGV based on copy number					
Sample	**1q (hg38)**	**6p (hg38)**	**1q (hg38)**	**Band**	**Gene at breakpoint**	**6p (hg38)**	**Band**	**Gene at breakpoint**
E14170	150,522,524	28,667,223	150,523,527	1q21.2	none	28,666,224	6p22.1	none
E14171	151,662,524	28,457,223	151,668,582	1q21.3	***SNX27***	28,468,135	6p22.1	none
E14173	149,878,450	22,784,771	149,562,435	1q21.2	none	22,782,649	6p22.3	none
E14211	150,312,556	28,232,222	150,315,431	1q21.2	none	28,230,425	6p22.1	***ZSCAN9***
M0000267	147,063,456	26,669,772	148,980,499	1q21.2	***PDE4DIP***	26,852,068	6p22.2	none

Table 4.7 Breakpoint prediction by IGV analysis of WGS data. For comparison, the breakpoints predicted by copy number analysis using QDNAseq are also shown for each sample.

4.4.4.2 Localisation of breakpoints based on sequence data

Detailed inspection of the sequence data was performed to try and identify reads or mate pair that straddled the predicted breakpoints for the 5 cases shown in Table 4.7. For all cases, reads were identified at 1q and/or 6p close to the predicted breakpoint that had the correct orientation, i.e. reading towards the predicted breakpoint. For each sample, the 'soft clipped' sequences (i.e. sequence at the end of reads that had been masked by the alignment software due to mismatches with the expected sequence) near the predicted breakpoint were examined. 'Soft-clipped' sequence that was shared between more than one read were considered as candidate breakpoint-spanning reads; mismatches that were present in isolated reads only were not investigated as they could have represented sequencing errors. Mate pairs for reads mapping close to the predicted breakpoints were also examined for some cases. Some mapped to the same chromosome and therefore corresponded to the normal chromosome 1 or chromosome 6. For true breakpoint spanning mate pairs, one mate would be expected to map to chromosome 1 and the other to chromosome 6.

For E14170, 5 candidate reads were identified that mapped to within 1kb of the predicted chromosome 1 breakpoint and pointed towards the breakpoint. One of these reads had soft clipped sequence at the 5' end suggesting it might span the breakpoint; however the soft

clipped sequence corresponded to a repetitive Alu sequence and thus could not be mapped precisely. Two reads were identified that mapped close to the chromosome 6 breakpoint and had 5' soft clipped sequence but again this sequence was repetitive and could not be mapped precisely. In addition, 8 chromosome 1 reads were identified close to the breakpoint for which the mate pair mapped to chromosome 6, and for 7 of these reads the sequence was within 1kb of the predicted 6p breakpoint. Thus, the breakpoints for this case were defined to a small region but could not be identified precisely from the WGS data. To investigate this case further, PCR primers were designed to flank the region where the breakpoints were expected to be. These primers amplified a specific product for E14170 that was absent from ten controls. Sanger sequencing of the product and comparison to Genbank using blast revealed that both breaks had occurred by homologous recombination between Alu elements (Figure 4.12).

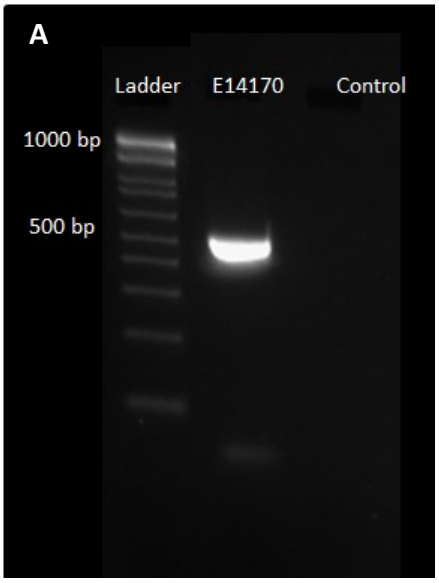


Figure 4.12 PCR primers designed to span predicted breakpoint for E14170. **A.** A product identified ~500 bp in size that was not dected in 10 control samples (only one control shown in this figure). The primer sequences were as follows: forward primer CCAGTCACCAATGTAAGCCAC and reverse primer ACCATACACTACTTTCCCACTCA. The initial part of the sequence of the E14170 amplicon showed a precise match to part of an Alu element (as identified using RepeatMasker) in sequence AL356356.17 at 1q21.3 (**B**) but then then the two sequences diverge and show multiple mismatches. The latter part of the sequence shows an exact match to part of an Alu element in sequence AL121932.19 at 6p22.1 (**C**). The highlighted 28bp sequence is identical in both Alu elements and therefore the breakpoints must presumably have occurred in this region by homologous recombination.

B

```
E14170 39      TTTTTTTTTTTTTTTTGACGGAGTCTCGCTCTGTTACCCAGGCTGGAGTGCAGTGGCATG  98
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1q21.3 50929   TTTTTTTTTTTTTTTTGACGGAGTCTCGCTCTGTTACCCAGGCTGGAGTGCAGTGGCATG  50870

E14170 99      ATCTTGGCTCACTACAACCTCTGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCC  158
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1q21.3 50869   ATCTTGGCTCACTACAACCTCTGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCC  50810

E14170 159     CGAGTAGCCGGGATTACAGGCTTGTGCCACCACGCCCTGCTAATTTTTTGTATTTTTAGT  218
               |||||||| |||| ||||||| ||  | |||| ||   ||||| ||||||||||||||||
1q21.3 50809   CGAGTAGCTGGGACTACAGGCATGCAC-ACCATGCATGGCTAA-TTTTTGTATTTTTAGT  50752

E14170 219     GGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACC-CA-GTGATCC  276
                ||||| |||||||||| |||| | |||| ||||||||| ||||||||| || ||||||
1q21.3 50751   AGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCT  50692

E14170 277     GCCCGCCTC-GGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCGCTCCAGG  331
                || ||||  |||||||||||||||| |||||||||  |||||||||| || ||
1q21.3 50691   TCCTGCCTTTGGCCTCCCAAAGTGCTAGGATTACAGGCATGAGCCACCGCGCCTGG  50636
```

C

```
E14170  3      TTTTTTTTTTTTTTGACGGAGTCTCGCTCTGTTACCCAGGCTGGAGTGCAGTGGCATGAT  62
               ||||| ||||||  ||||||||||||||||||  ||||| ||||||||  |  | |  ||
6p22.1 60653   TTTTTGTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGACTGGAGTG--G-AG-A-AAT  60707

E14170  63     CTTGGCTCACTACAACCTCTGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCG  122
               || |||||||| ||||||| ||||||||| |||| || |||||||||||||||||||||||
6p22.1 60708   CTCGGCTCACTGCAACCTCCGCCTCCCGGATTCAAGCGATTCTCCTGCCTCAGCCTCCCG  60767

E14170  123    AGTAGCCGGGATTACAGGCTTGTGCCACCACGCCCTGCTAATTTTTTGTATTTTTAGTGG  182
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6p22.1 60768   AGTAGCCGGGATTACAGGCTTGTGCCACCACGCCCTGCTAATTTTTTGTATTTTTAGTGG  60827

E14170  183    AGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACCCAGTGATCCGCCC  242
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6p22.1 60828   AGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACCCAGTGATCCGCCC  60887

E14170  243    GCCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCGCTCCAGG  293
               ||||||||||||||||||||||||||||||||||||||||||||||||||
6p22.1 60888   GCCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCGCTCCAGG  60938
```

For E14171, 4 reads were found that mapped close to the predicted chromosome 1 breakpoint that had matching 5' soft clipped sequence which mapped to chromosome 6. Two reads that mapped close to the chromosome 6 breakpoint had soft clipped bases that mapped *SNX27* exon 7 on chromosome 1q21.3. The breakpoints for this case were therefore precisely determined but, as shown in Figure 4.13, there was no homology between the two chromosomes.

```
Chr 1      TCGAAAGTGGCTTTTTACAACAGAAGAAGAAATTCTCTTAAATGAC
Chr 6      AAAGTGCTGGGATTACAAGTGTGAGCCACCCGTATAAAGCCTGCTA
Fusion     AAAGTGCTGGGATTACAAGTGTAATTAAGAAATTCTCTTAAATGAC
```

Figure 4.13 The t(1;6) genomic breakpoint for patient E14171. Chromosome 1 sequence is shown in black and is part of *SNX27* exon 7. Chromosome 6 sequence is shown in red. Four nucleotides of unknown origin shown in green were identified in the fusion between the chromosome 1 and 6 sequence.

For E14173 a total of 5 soft clipped reads mapping close to the predicted 1q or 6p breakpoints were identified but all the soft clipped sequence was repetitive and thus exact breakpoints could not be determined.

For E14211 a break was identified in *ZSCAN* exon 4 at 6p22.1 that read into sequence from 1q21.2 in multiple soft clipped reads in both directions. The fusion sequence is shown on Figure 4.14 and again there was no homology at the breakpoints.

```
Chr 1      TAGAGGCCATGAGGCGCTGGAGAGAGATGAATGGGGAGTTAACGCC
Chr 6      AGACACATTTTTATTTTCCAAACCTGTTGTGATCCCCCAGCTAAAA
Fusion     TAGAGGCCATCAGGCGCTGGAGCCTATTGTGATCCCCCAGCTAAAA
```

Figure 4.14 The t(1;6) genomic breakpoint for patient E14211. Chromosome 1 sequence is shown in black; chromosome 6 sequence is shown in red and is part of *ZSCAN* exon 4. Two mismatches, which may be polymorphisms or sequencing errors, are underlined.

Finally, for M0000267 multiple soft clipped reads identified a break within the 5' region of *PDE4DIP* at 1q21 that read into a repetitive sequence that could not be mapped. No mate pairs spanning the predicted breakpoints could be identified. Possible reasons for this are inadequate sequence coverage, long tracts of repetitive sequence in the region of the chromosome 6 breakpoint or a more complex genomic rearrangement than apparent by cytogenetics.

4.5 Discussion

The der(6)t(1;6) has been previously been reported as a rare but recurrent abnormality in patients with MF and related myeloid neoplasms but it has not yet been fully characterised. In this study, I report the molecular characterisation of 16 patients identified in 4 diagnostic laboratories within Europe to add to what is known about this rearrangement and determine to what extent this rearrangement is recurrent.

As previously reported, there was a clear association with an MPN phenotype in the cases identified with this aberration, in particular MF (primary or secondary). In addition, 3 out of 16 (19%) individuals were reported to have MDS with progression to AML indicated in one of these individuals. Four of the 16 cases (25%) were also reported to have had disease progression to AML, with evidence that the der(6) was a late acquisition in disease course in some patients. Six of the 7 additional cases identified at the WRGL (Appendix 13) that were not further investigated as part of this study due to insufficient DNA available also showed evidence of disease progression based on the clinical notes obtained by the laboratory at sample receipt, and in some, the progression was at the time point of der(6)t(1;6) acquisition. Taken together, these results are consistent with this aberration being common in, but not specific to, MF. These observations also support the der(6)t(1;6) being a marker of disease progression; first proposed more generally by the identification of gain 1q acquired during clonal evolution by Najfeld, et al., (2010) and Marcellino, et al., (2017). The karyotype obtained for each case demonstrated that this rearrangement was not always the sole abnormality but could also been seen in the context or simple abnormal (≤3 abnormalities) or complex abnormal karyotype (>3 abnormalities), including at a sub-clonal level. However, this does not prove the timing of acquisition of this rearrangement for all patients, and this was only known in the small proportion of cases from whom the WRGL had received more than one sample. To confirm that this abnormality represents a marker of disease progression following late acquisition in each patient's disease course, the cytogenetic results from more than one sample of each case would be required, ideally linked to the clinical information (including BM review) at different time points.

In this cohort, breakpoint characterisation was performed using standard cytogenetic analysis, SNP array, WGS and, in one patient, Sanger sequencing. By standard cytogenetic analysis, the breakpoints fell within the 1q21 and 6p21-p22 for the majority of cases, with one case reported to have a breakpoint more proximal on 1q at 1q12 and one case with a

more distal 6p breakpoint at 6p25. By SNP array and/or WGS, where detected, the start of the 1q duplication fell within a 7.4 Mb region from 1q21.1 (1:144,949,165) to 1q21.3 (1:152,371,259) and the end of the 6p loss fell within a 5.9 Mb region from 6p22.3 (6:22,741,451) to 6p22.1 (6:28,633,895), confirming a considerable degree of diversity with respect to breakpoints at a molecular level between individuals. In one case, the 1q duplication was not seen by SNP array (WGS not performed on this sample) but the karyotype was complex with der(6)t(1;6) seen in only 15% of cells therefore it is likely that this was not detected due to the limitations of SNP array in detecting sub-clonal duplications. In 1 case, the 6p deletion was not detected by SNP array or WGS. The rearrangement was seen in 45% of cells by cytogenetics but the breakpoint was assigned to the distal breakpoint 6p25; it is possible that the breakpoint was sub-telomeric and beyond the genomic region covered by SNP array. Where WGS was performed, breakpoint characterisation was also attempted via in manual interrogation of the data. From this analysis, there was no evidence of a recurrent candidate fusion gene which is not unsurprising given the observed molecular diversity at the breakpoints. The breakpoint fell within a gene for 3 cases (*SNX27* and *PDEDIP* at 1q, and *ZSCAN9* at 6p). In one sample, Sanger sequencing of the product and comparison to Genbank using blast revealed that both breaks had occurred by homologous recombination between Alu elements. Overall, therefore it appears that 1q gain is the important consequence of the der(6)t(1;6) and the reason this abnormality is recurrent is unclear.

In addition to the expected aberrations on 1q/6p, SNP array identified a deletion at 13q14.2 (27 kb – 42 Mb), resulting in whole or partial deletion of *RB1,* in 9 out of 11 samples (82%) and 17q11.2 (145 kb – 1 Mb; involving *NF1*) in 5 out of 11 samples (45%) tested. Of the 6 cases that underwent WGS, 5 also had SNP array analysis; of these, 4 cases had a 13q deletion and 2 cases had a 17q deletion detected by SNP array. WGS detected the 13q deletion in out of 4 cases (a 40 Mb deletion; M0000267); however, 17q deletions were not detected in the 2 cases which had this deletion by SNP array analysis. In addition, there were no mutations detected in *RB1* or *NF1*, as might be expected since both are considered to be classic tumour suppressor genes. Further work would be required to explore whether low level mutations might have been missed and to sequence addition cases with 13q/17q abnormalities. Of note, all 13q/17q deletions detected by SNP array but not confirmed by WGS were ≤1 Mb in size (and sub-clonal) therefore it is likely that CNV detection by WGS according to the strategy utilized in this study is an unreliable method for detecting small

deletions. A 1 Mb *NF1* deletion detected by SNP array in 77% of cells in one case was confirmed by MLPA; MLPA failed to confirm the deletion in the remaining samples but this method is not optimised for the detection of sub-clonal deletions. No orthogonal methods were available for confirmation of the remaining *NF1* or 13q deletions. Deletions within 13q are recurrently reported in myeloid neoplasia, particularly in MF (La Starza, et al., 1998; Tanaka, et al., 1999; Mehrotra, et al., 2015; Tefferi, et al., 2018). Deletions of 17q/*NF1* have also been reported in MPN (Rego de Paula, et al., 2018) and therefore may contribute to the disease pathophysiology in a proportion of these patients. These data suggest that overlapping 13q and 17q deletions may be a recurrent co-occurring abnormalities in patients with the der(6)t(1;6) but further independent confirmation is required. It would also be interesting to determine if 13q/17q deletions are seen more widely in the context of gain of 1q.

All samples showed one or more variant (pathogenic, likely pathogenic or VUS) by TSMP and WGS analysis: by TSMP, all 16 samples had between 1 and 5 variants and by WGS (performed on 6 samples), between 1 and 12 variants were detected. Overall the mutation pattern was consistent with MF or related myeloid neoplasms and apart from MPN driver mutations, there were no unusual mutation patterns associated with the der(6)t(1;6). In 14 (88%) samples, a *JAK2* V617F, *CALR* or *MPL* pathogenic variant was detected. The two samples that did not have one of these variants were E14171 and M0606634; these patients were referred with MDS RAEB and MDS progressing to AML, respectively. For both samples, the combination of *TET2* and *SRSF2* variants may indicate that the patient actually had MDS/MPN, as this pattern is commonly associated with CMML (Itzykson, et al., 2013). For M0606634 specifically, a *RUNX1* p.(Leu102Cysfs*21) variant was also detected. As *RUNX1* variants are associated with a poor prognosis in CMML, this finding would be consistent with the clinical diagnosis of progressive disease. Overall, the most frequently mutated genes were *JAK2*, *CALR*, *TET2*, *ASXL1* and *SRSF2* and when on one variant was detected in a sample (n=4; 25%), this was always a pathogenic variant in *JAK2, CALR* or *MPL*. Genes that are known to confer high risk in myelofibrosis where detected in 6 samples (38%): *ASXL1* (n=3), *SRSF2* (n=4), *IDH2* (n=1). Four of these 6 samples came from individuals reported to have MF (E14211, E14232 E14230, E14234); the other two samples with high risk variants were E14171 and M0606634, discussed above as possible MDS/MPN patients. The number of mutations did not appear to strictly correlate to the karyotypic complexity identified by G-banding (i.e. more mutations were not detected in those with a complex karyotype versus

those samples where the der(6)t(1;6) was the sole abnormality in all cells) or the clinical indication. For example, of the 3 samples with the most mutations detected by TSMP and WGS combined were E14171, M0606634 (MDS or MDS/MPN) and E14211 (MF). Finally, by WGS analysis there was no evidence of mutations in *MDM4* at 1q32 which was proposed as a critical gene in the region duplicated as part of this rearrangement (Marcellino, et al., 2017). The only other gene analysed on chromosome 1q as part of the virtual gene panel was *RIT1* (1q22), a gene encoding a GTP-binding protein involved in regulating p38 MAPK-dependent signaling reported to be associated with CMN according to Yoshizato, et al., (2015); no mutations were detected. With more time available, a more in-depth analysis of genes at 1q may be worthwhile.

5. Conclusion

As data is gathered about the genomic diversity of CMN and the technologies utilised in diagnostic laboratories are becoming more advanced, the approaches taken to interpret and report results must adapt. In particular, the knowledge that apparently normal individuals acquire clonal mutations as they age (ARCH/CHIP), reported to occur in at least 10% of individuals greater than 65 years (Xie, et al., 2014; Jaiswal, et al., 2014; Genovese, et al., 2014) has added to the complexity of interpreting results from small and large gene panels in the context of analysis of a sample from an individual with a suspected CMN.

At the WRGL, the implementation of a more in-depth molecular analysis for these patients, namely analysis of MPN driver mutations by the NGS genotyping pipeline (with a LoD 1%) and the implementation of a 54-gene myeloid panel, the TSMP, to analyse patients with confirmed or suspected CMN was a driver to initiate the process of developing a more robust interpretation and reporting strategy for the results from these tests. At the time of writing, there are no UK best practice guidelines that can be directly applied to the classification and reporting of somatic variants identified by NGS panels in the context of haematological malignancies, such as the TSMP. Furthermore, over time it became clear that there was no unified approach for how to interpret low level variants (e.g. low level *JAK2* V617F mutations <5% VAF). Therefore, my project aimed to address these issues through four main approaches. Firstly, I performed a clinical audit to gather information on how the results of the TSMP were being used in a research setting to inform patient management. Secondly, I led the development of a robust interpretation and reporting strategy for the results from the TSMP for use at the WRGL. Thirdly, I used ddPCR to understand the significance of low level *JAK2* V617F mutations in the context of diagnostic referrals for MPN testing. Fourthly, I performed molecular characterisation of a rare but recurrent cytogenetic abnormality in CMN, the der(6)t(1;6), including mutation analysis by the TSMP as well as additional diagnostically relevant techniques, namely ddPCR, SNP array and WGS.

The outcome of the clinical audit looking at the clinical utility of the TSMP in a research setting indicated that clinicians found this panel very useful in the management of patients with CMN, in particular when trying to formalise or exclude a diagnosis or when looking for prognostically significant markers in an already confirmed diagnosis. In patients with

confirmed MDS (n=26), the majority (73%) of respondent said that the patient management was affected by the TSMP result through additional prognostic information, guiding management decisions or clarifying diagnosis as the BM was not diagnostic. In cases of suspected MDS, the absence of mutations was a particularly useful piece of evidence for exclusion of a neoplastic condition. These findings are in line with what would be expected from a number of key studies which report on the mutation patterns in MDS and the NCCN guidelines (Papaemmanuil, et al., 2013; Malcovati, et al., 2015; National Comprehensive Cancer Network, 2021). However, it has been reported that up to 15% of MPN cases (Grinfeld, et al., 2018), 10-20% of established MDS cases (Haferlach, et al., 2014; Papaemmanuil, et al., 2013) and 5-7% of CMML cases (Itzykson, et al., 2013; Elena, et al., 2016) do not have detectable variants by panel analysis and thus it is imperative that genetic results are considered alongside supportive morphological criteria and immunophenotype results. In addition, from this audit, there was no evidence that the TSMP was being used inappropriately to manage patients with confirmed or suspected myeloid neoplasia. It had been a particular concern from some of our clinical colleagues that inexperience with panel results and complications such as CHIP might lead to inappropriate diagnosis of a CMN in some cases, but this concern was not borne out. This audit was undertaken in 2019 and since that time clinicians are becoming more used to incorporating (sometimes complex) molecular genetic results from gene panels in the diagnostic and prognostic scoring systems for patients with CMN, an example being the routine application of MIPSS70+ version 2.0 for patients with MF (Tefferi, et al., 2018)], and I have witnessed this in practice through participation in MDT meetings. An example of this is the more routine approach to testing multiple samples over time from the same patient with the TSMP, especially in patients with a complex pattern of abnormalities, to detect changes to the sub-clonal components and/or mutations burden as treatment is applied. It is expected that specific error-corrected panels will be introduced into routine practice in due course to enable more sensitive detection of MRD in patients with CMN, and particularly in patients with AML (Jongen-Lavrencic, et al., 2018).

Within this study, in the absence of national guidelines, I also present a standardised framework for the interpretation of variants detected by gene panels utilised in the context of testing for haemato-oncology samples. The protocol was developed over time at the WRGL and builds upon a number of publications including guidelines for the interpretation of variants in a germline context for non-cancer patients (Richards, et al., 2015; Ellard, et al.,

2019) and cancer patients (Luo, et al., 2019; ClinGen *TP53* Variant Curation Expert Panel, 2019) and somatically aquired variants for solid tumours (Li et al., 2017). The outcome is a two-step process, firstly to assign a pathogenicity of a variant based upon the application of weighted evidence (predominantly uncoupled from the reason for referral and applicable to any gene) and secondly to assign a clinical significance of the variant in the sample within which it was detected to allow decision making about whether or not the finding should be reported and how. Using this method, the WRGL has performed interpretation on >1,700 variants and each classification decision is recorded and auditable, in accordance with strict quality management requirements for diagnostic laboratories. Challenges still exist, such as the difficulty in determining the significance of evidence gathered from database searches. Specifically, further work to define when the frequency of variants in COSMIC becomes significant would be a particularly beneficial addition to the current WRGL protocol to act as a more robust method of determining whether a variant is more common in cancer *versus* the general population, given that some variants can be detected in GnomAD that may represent CHIP or an undiagnosed haematological neoplasia and some variants in COSMIC may be passenger mutations which were acquired in stem cells, or even technical artefacts which may occur more frequently in certain genes. It would therefore be particularly useful for there to be gene specific measures of significance. Challenges also exist in the interpretation of variants that may be germline in origin when gene panels are performed without subtraction of a germline sample; however, the WRGL protocol for variant interpretation includes a section on "additional considerations for germline variants" to address this which includes a list of defined genes to consider for further investigation and the approach that should be taken in determining the clinical significance.

A limitation to this framework is the absence of multicentre ratification. Whilst the WRGL mandate that internal staff undertaken competency assessment via independent review of variant interpretation of multiple variants types across a number of different genes by an experienced member of staff, and the laboratory participates in national EQA schemes for myeloid panel analysis, there has not been an independent critique of this protocol. Ratification would enable assessment of how easily this protocol can be applied across centres and identify any sources of inconsistency. A nationally agreed framework, much like the ComPerMed workflow in Belgium (Froyen, et al., 2019), would be beneficial in the UK in that it would act to improve the equity of service provided by different diagnostic laboratories between the GLHs in England. A national somatic variant interpretation

working group (initiated in 2018 but paused during the COVID-19 pandemic) has been convened and the main objective of this group is to define a shared recommendation document. The WRGL will participate in this process once activity is recommenced. Another limitation of this protocol is the amount of staff resources required to collate, weight and sort the evidence for each variant and, although this process does speed up as the analyst gains experience, automated collation and weighting of evidence would be required to improve this process. Examples of this exist for the application of the framework for solid tumour variant interpretation presented by Li et al., (2017) and, once a nationally ratified framework is established, efforts should be made to automate the formalised process, where possible. This in turn should help enable laboratories to more readily share the outcomes of their variant interpretation to keep databases up to date and ultimately reduce the amount of duplicated work undertaken between centres. More automated processes and wide data sharing will be particularly important as WGS becomes a mainstream diagnostic tool.

In this study, in addition to detailing the challenges associated with interpreting the clinical significance of newly identified or rare variants, I also discuss the challenges associated with interpreting the clinical significance of known pathogenic variants detected at low level (<5% VAF). In order to investigate this in more detail, I focussed on the *JAK2* V617F mutation. *JAK2* V617F is a key MPN driver mutation and as such can be used as a major criterion for the diagnosis of MPN according to the WHO categorisations (Arber, et al., 2016); however, it has also been reported to occur in ARCH/CHIP and has been identified in normal individuals with a prevalence of 0.1 - 3.1% in population cohorts (Nielsen, et al., 2013; Cordua, et al., 2019; Jaiswal, et al., 2014). In the majority of positive cases identified in these studies, the VAF was <10% and reports indicate that low level *JAK2* V617F-positive clones can remain stable for several years (Cordua, et al., 2019; Nielsen, et al., 2014; Gale, et al., 2007) with only a proportion of *JAK2* V617F-positive individuals identified in population studies later going onto develop MPN. When the laboratory best practice guidelines were written for *JAK2* V617F, testing a lower limit cut-off of 1-3% VAF was recommended (Bench, et al., 2013) and this was largely informed by the level of sensitivity available for testing options at the time rather than clear clinical evidence that this cut off was clinically appropriate. In order to add to what is known about the prevalence of low level (<1% VAF) *JAK2* V617F, this study assessed the frequency in suspected MPN panels referred to the WRGL for MPN panel testing. Overall, I found no evidence by ddPCR (LoD 0.03%) that low

level (<1% VAF) *JAK2* V617F was enriched in cases referred for MPN testing that tested negative for all MPN driver mutations by the diagnostic pipeline (LoD 1%) compared to controls (prevalence 3% for controls [4/197] *vs* 2% for TN samples [20/662]; P=0.62 Fisher's exact test). This result suggests that our standard diagnostic 1% VAF cut off is unlikely to be missing many cases of true MPN with low level *JAK2* V617F thus supporting the continued use of this allelic burden for a cut-off in a diagnostic setting. Of note, although the difference was not significant, the prevalence of low level positives was actually higher in the query MPN group: 20/662 (3%) versus 4/197 (2%) in controls suggesting the possibility of a small difference that my study was not powered to detect and expanding the study is likely to be prohibitively large at a single centre (estimated sample group size required: 6,600). Conversely, there was a statistically significant enrichment of low level (<1%) *JAK2* V617F in both the *CALR*+ and *MPL*+ inidividuals identified by diagnostic laboratory testing *versus* population controls (P=0.0175; 7.4% [11/149] and P=0.006; 15% [8/52], respectively) and further work to explore the biological mechanism behind this and the impact on MPN phenotype is required to fully understand the significance of this findings.

A clinical audit was also undertaken to establish the final diagnosis in individuals with low level *JAK2* V617F variants and given that the *JAK2* V617F <1% VAF would have been unknown to the referring clinician, the audit was also expanded to cases with a *JAK2* V617F up to 5% VAF which *were* externally reported by the WRGL. Overall, in individuals with a *JAK2* V617F reported at 1-5% VAF, 89% (n=55) were reported to have been diagnosed with a sub-type of MPN. Of the 7 individuals (11%) that were reported to have *not* been given a final diagnosis of MPN, 4 (57%) cases had a personal history of longstanding erythrocytosis or portal vein thrombosis but insufficient criteria to diagnose an MPN. These findings are consistent with the Danish population study by Cordua, et al., (2019) who found that *JAK2* V617F ≥1% VAF was associated with higher haematocrit, leukocyte, neutrophil, and thrombocyte counts *versus JAK2* V617F-negative individuals. In individuals that were reported to be triple-negative by the diagnostic MPN panel but shown to have a *JAK2* V617F at <1% VAF by ddPCR as part of this study (which were *not* reported to the clinician), I received information on the outcomes from 8 individuals, of which only 1 (13%) was given a final diagnosis of MPN. Again, these findings support keeping the exisiting cut-off of 1% VAF for *JAK2* V617F identified in a diagnostic setting. These data could also be considered as evidence to support the hypothesis by Nielsen, et al., (2014) that mutation burden is associated with MPN development, going from a continuum of no disease to symptomatic

MPN with a *possible* intermediate stage of disease progression. What is unclear by this study is how many cases would have been diagnosed with MPN if we had reported the *JAK2* V617F detected <1% VAF and how many of these cases will later go on to develop MPN. Continuining to gather long-term data on individuals identified as low level *JAK2* V617F positive from population level studies will help to gather further understanding of this.

Mutation assessment by the TSMP was also undertaken in the samples within which a low level (<1%) *JAK2* V617F was detected to ascertain whether there was an additional clone(s) that was driving an MPN-like phenotype in these cases. TSMP was undertaken on all 18 samples that were TN; only 5 samples (28%) showed the presence of abnormal clone(s) by TSMP analysis (i.e. no apparently somatic variants were detected in 72% of cases): in 3 cases (17%) pathogenic or likely pathogenic variants were detected; in 2 cases (11%), clonality was indicated by the detection of apparently somatic VUS. In the 3 cases with pathogenic/likely pathogenic variants, the mutation pattern suggested that the patient may have had a diagnosis of MDS/MPN or atypical CML rather than one of the classical MPN but a clinical audit response was not obtained for these cases to confirm this. These findings add to the data collected during the clinical audit described in Chapter 2, which showed that of 20 patients referred for TSMP with suspected MPN no variants were detected in 15 (75%) and the majority of these cases where given a final diagnosis of a non-neoplastic condition (idiopathic erythrocytosis or reactive causes). Overall, TSMP analysis on individuals with *suspected* MPN with a low level *JAK2* V617F identified by ddPCR indicated that additional mutations in genes other than *CALR* or *MPL* are only seen in a minority of individuals. Although additional low level mutations could not be excluded in these samples, it would be unclear how these additional mutations would result in a clinical MPN or MPN-like phenotype when present in such a small proportion of cells.

In order to investigate whether there may be a predisposing factor to acquiring a low level *JAK2* V617F, the 46/1 haplotype was investigated. Surprisingly, no significant difference in the frequency of 46/1 haplotype was detected in low level *JAK2* V617F positives individuals *versus* population controls. This is in striking contrast to *JAK2* V617F positive MPN patients which are known to be associated with 46/1 (Jones, et al., 2009), and confirmed in my analysis. However, McKerrell, et al. (2017) proposes that the 46/1 haplotype may be a modifier that contributes to the likelihood that a low level *JAK2* V617F clone expands. Additional work is required to further investigate this hypothesis and understand how non-genetic factors contribute to the likelihood of clones of expanding and the rate at which

they do expand; this will be particularly useful in light of recent evidence that *JAK2* V617F can occur early in life, even *in utero* and later evolve into MPN (Williams, et al., 2020). A more robust assessment of the association of low level *JAK2* V617F could involve *JAK2* V617F analysis on samples from individuals who had all had a diagnosis of MPN made via a bone marrow sample, with additional genetic and non-genetic factors which contribute to disease development also investigated. Given that results from clinical audit on cases with a *JAK2* V617F at 1-5% VAF indicated that only 7% of cases had a final diagnosis informed by a bone marrow sample, this would likely be challenging to coordinate in practice.

Finally, the TSMP was used as just one of several techniques to further characterise samples (n=16; from 4 diagnostic laboratory centres) with a recurrent cytogenetic abnormality: the der(6)t(1;6). This rearrangement is a recurrent abnormality in CMN but it was unclear whether it results in a specific chimeric gene fusion product and/or whether these patients frequently harbour additional cytogenetically-cryptic abnormalities or mutations which drive (or contribute to) disease pathogenesis. Altogether, the TSMP was used in combination with SNP array, WGS, MLPA and Sanger sequencing to investigate these samples. Overall, a considerable degree of diversity with respect to breakpoints at a molecular level between individuals was observed, with the majority of breakpoints falling within 7.4 Mb and 5.9 Mb regions on 1q and 6p, respectively. Furthermore, although the breakpoint fell within a gene for 3 cases (*SNX27* and *PDEDIP* at 1q, and *ZSCAN9* at 6p), and Sanger sequencing of one sample indicated that the breaks were likely to have occurred by homologous recombination between Alu elements, there was no evidence of a recurrent candidate fusion gene which is not unsurprising given the observed molecular diversity at the breakpoints. CNV analysis performed by SNP array identified a deletion at 13q14.2 (27 kb – 42 Mb), resulting in whole or partial deletion of *RB1,* in 9 out of 11 samples (82%) and 17q11.2 (145 kb – 1 Mb; involving *NF1*) in 5 out of 11 samples (45%) tested; although all samples were not tested by all methods, WGS and MLPA were undertaken on a proportion of these cases and confirmed a 13q and 17q deletion in one case each. Deletions within 13q are recurrently reported in myeloid neoplasia, particularly in MF (La Starza, et al., 1998; Tanaka, et al., 1999; Mehrotra, et al., 2015; Tefferi, et al., 2018). Deletions of 17q/*NF1* have also been reported in MPN (Rego de Paula, et al., 2018) and therefore may contribute to the disease pathophysiology in a proportion of these patients. Variant analysis was performed by TSMP and WGS. Samples had between 1 and 12 variants (when benign variants were excluded) and generally, WGS using a 102- gene virtual panel provided minimal additional

information to that which was available by TSMP analysis. Overall the mutation pattern was consistent with MF or related myeloid neoplasms and apart from MPN driver mutations, there were no unusual or unique mutation patterns associated with the der(6)t(1;6). A *JAK2* V617F, *CALR* or *MPL* pathogenic variant were detected in 88% (n=14) of cases, and where only one variant was identified it was always in one of these MPN driver genes. In the 2 cases where a *JAK2* V617F/*CALR*/*MPL* variant was not detected, the mutational pattern was indicative of a diagnosis of MDS/MPN (i.e. *TET2*, *SRSF2* and *RUNX1*). Assessment of the clinical indications provided by the clinicians and considered in the context of the known mutation patterns in these cases suggests that the der(6) has an association with an MPN phenotype, particularly MF but can also occur in MDS/MPN. There was also some evidence that this rearrangement may be a marker of disease progression but this would require further investigation using analysis from multiple samples from the same cases to pin down the timing of acquisition during the disease course. Taken together, these data add to the existing understanding of the der(6)t(1;6) in myeloid neoplasms and identify focal deletions of *RB1* (13q14.2) and *NF1* (17q11.2) as potential recurrent abnormalities.

Overall, my study adds to what is known about CMN and the diagnostic laboratory approaches that can be applied to the analysis and reporting of results obtained from genomic testing of these cases, in particular results from myeloid gene panel analysis, the significance of low level *JAK2* V617F in suspected MPN patients and molecular definition of the der(6)t(1;6).

6. References

Abelson, S. et al., 2018. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature,* 559:400-404.

Abou Tayoun, A. et al., 2018. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Human Mutation,* 39(11):1517-1524.

Acuna-Hidalgo, R. et al., 2017. Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *American Journal of Human Genetics,* 101(1):50-64.

Adzhubei, I. et al., 2010. A method and server for predicting damaging missense mutations. *Nature Methods,* 7(4):248-249.

Agilent, 2018. *Alissa Interpret v5.2: Instructions for Use.* [Online]
Available at:
https://wrgl.alissa.agilent.com/interpret/Alissa%20Interpret%20v5.2%20Instructions%20for%20Use%20-%20v1.pdf
[Accessed 26 January 2021].

Agilent, 2018. *Bioanalyzer DNA analysis.* [Online]
Available at: https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-dna-kits-reagents/bioanalyzer-dna-analysis-228248#productdetails
[Accessed 24 October 2020].

Ahmed, R. et al., 2016. Coexisting JAK2 V617F and CALR exon 9 mutations in myeloprolifertaive neoplasms- do they designate a new subtype? *Asian Pacific Journal of Cancer Prevention,*17:923-926.

Ahn, J. & Lee, J., 2008. X Chromosome: X Inactivation. *Nature Education,* 1(1):24.

Akada, H., Akada, S., Hutchinson, R. & Mohi, G., 2014. Loss of wild-type Jak2 allele enhances myeloid cell expansion and accelerates myelofibrosis in JakV617F knock-in mice. *Leukemia,* 28(8):1627-1635.

Al Assaf, C. et al., 2015. Analysis of phenotype and outcome in essential thrombocythemia with CALR or JAK2 mutations. *Haematologica,* 100(7):893-897.

Alchalby, H. et al., 2010. Impact of JAK2 V617F mutation status, allele burden, and clearance after allogeneic stem cell transplantation for myelofibrosis. *Blood,* 116(18):3572-3581.

Andrieux, J. et al., 2003. Karyotypic abnormalities in myelofibrosis following polycythemia vera. *Cancer Genetics and Cytogenetics,* 140(2):118-123.

Antonioli, E. et al., 2008. Influence of JAK2V617F allele burden on phenotype in essential thrombocythemia. *Haematologica,* 93(1):41-48.

Arber, D. et al., 2016. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood,*127(20):2391-2405.

Armbruster, D. & Pry, T., 2008. Limit of Blank, Limit of Detection and Limit of Quantitation. *The Clinical Biochemist Reviews,* 29(Suppl 1):S49-S52.

Atmüller, F. et al., 2017. Genotype and phenotype spectrum of NRAS germline variants. *European Journal of Human Genetics,* 25(7):823-831.

Ballen, K. et al., 2010. Outcome of transplantation for myelofibrosis. *Biology of Blood and Marrow Transplant,* 16(3):358-367.

Bandaranayake, R. et al., 2012. Crystal structures of the JAK2 pseudokinase domain and the pathogenic mutant V617F. *Nature structural & Molecular Biology,* 19(8):754-760.

Beer, P. et al., 2008. MPL mutations in myeloproliferative disorders: analysis of the PT-1 cohort. *Blood,* 112(1):141-149.

Bejar, R. et al., 2011. Clinical effect of point mutations in myelodysplastic syndromes. *New England Journal of Medicine,*364:2496-2506.

Bellanné-Chantelot, C. et al., 2020. Germline genetic factors in the pathogenesis of myeloproliferative neoplasms. *Blood Reviews,*42:100710.

Bench, A. et al., 2013. Molecular diagnosis of the myeloproliferative neoplasms: UK guidelines for the detection of JAK2 V617F and other relevant mutations. *British Journal of Haematology,*160:25-34.

Bernard, E. et al., 2020. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nature Medicine,* 26(10):1549-1556.

Bianchi, P. et al., 2007. Presence of JAK2 V617F mutation in peripheral blood of blood donors with upper-limit hematocrit and platelet values. *Blood,* 110(11):2533.

Biesecker, L. et al., 2018. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genetics in Medicine,* 20(12):1687-1688.

Brinich, S. et al., 2020. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome medicine,*12:3.

Broad Institute, 2019. *Picard toolkit, Github repository.* [Online]
Available at: http://broadinstitute.github.io/picard/
[Accessed 12 January 2021].

Buscarlet, M. et al., 2017. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood,* 130(6):753-762.

Busque, L. et al., 1996. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood,* 88(1):59-65.

Busque, L. et al., 2016. Laboratory Investigation of Myeloproliferative Neoplasms (MPNs): recommendations of the Canadian Mpn Group. *American Journal of Clinical Pathology,*146:408-422.

Cabagnois, X. et al., 2016. Presence of atypical thrombopoietin (MPL) mutations in triple-negative essential thrombocythemia patients. *Blood,* 127(3):333-342.

Campbell, P. et al., 2005. Definition of subtypes of essential thrombocythaemia and relation to polycythaemia vera based on JAK2 V617F mutation status: a prospective study. *Lancet,*366:1945-1953.

Campo, E. et al., 2011. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood,*117:5019-5032.

Cankovic, M. et al., 2009. Clinical performance of JAK2 V617F mutation detection assays in a molecular diagnostics laboratory: evaluation of screening and quantitation methods. *American Journal of Clinical Pathology,* 132(5):713-721.

Caramazza, D. et al., 2011. Refined cytogenetic-risk categorization for overall and leukemia-free survival in primary myelofibrosis: a single center study of 433 patients. *Leukemia,* 25(1):82-88.

Caramazza, D. et al., 2010. Chromosome 1 abnormalities in myeloid malignancies: a literature survey and karyotype–phenotype associations. *European Journal of Haematology,* 84(3):191-200.

Carillo, S. et al., 2011. Nested high-resolution melting curve analysis a highly sensitive, reliable, and simple method for detection of JAK2 exon 12 mutations- clinical relevance in the monitoring of polycythemia. *Journal of Molecular Diagnosis,* 13(3):263-270.

Cervantes, F. et al., 2009. New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. *Blood,* 113(13):2895-2901.

Chachoua, I. et al., 2016. Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. *Blood,*127(10):1325-1335.

ClinGen TP53 Variant Curation Expert Panel, 2019. *ClinGen TP53 expert panel specifications to the ACMG/AMP variant interpretation guidelines v1.* [Online]
Available at: https://clinicalgenome.org/affiliation/50013/#heading_documents
[Accessed 1 December 2020].

Cook, E. et al., 2019. Comorbid and inflammatory characteristics of genetic subtypes of clonal hematopoiesis. *Blood Advances,* 3(16):2482-2486.

Corces-Zimmerman, M. et al., 2014. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *PNAS,* 111(7):2548-2553.

Cordua, S. et al., 2019. Prevalence and phenotypes of JAK2 V617F and calreticulin mutations in a Danish general population. *Blood,* 134(5):469-479.

Courtier, F. et al., 2020. Targeted molecular characterization shows differences between primary and secondary myelofibrosis. *Genes Chromosomes Cancer,*59:30-39.

Dameshek, W., 1951. Some speculations on the myeloproliferative syndromes. *Blood,* 6(4):372-375.

De Stefano, V. et al., 2005. Incidence of the JAK2 V617F mutation among patients with splanchnic or cerebral venous thrombosis and without overt chronic myeloproliferative disorders. *Journal of Thrombosis and Haemostasis,*5(4):708-714.

Desai, P. et al., 2018. Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nature Medicine,*24:1015-1023.

Desmet, F.-O. et al., 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research,* 37(9):e67.

DiNardo, C. et al., 2018. Improving the detection of patients with inherited predispositions to hematologic malignancies using next-generation sequencing-based leukemia prognostication panels. *Cancer,* 124(13):2704-2713.

Dingli, D. et al., 2005. Der(6)t(1;6)(q21–23;p21.3): a specific cytogenetic abnormality in myelofibrosis with myeloid metaplasia. *British Journal of Haematology,* 130(2):229-232.

Elena, C. et al., 2016. Integrating clinical features and genetic lesions in the risk assessment of patients with chronic myelomonocytic leukemia. *Blood,* 128(10):1408-1417.

Ellard, S. et al., 2019. *ACGS Best Practice Guidelines for Variant Classification 2019,* London: ACGS.

Ellard, S. et al., 2018. *ACGS Best Practice Guidelines for Variant Classification 2018,* London: ACGS.

Engle, E. et al., 2015. Clonal evolution revealed by whole genome sequencing in a case of primary myelofibrosis transformed to secondary acute myeloid leukemia. *Leukemia,*29(4):869-876.

EuroGentest, 2011. *EuroGentest: EU Contract No.: FP6-512148.* [Online]
Available at: http://www.eurogentest.org/index.php?id=160
[Accessed 18 06 2020].

Ferrando, A. & López-Otín, C., 2017. Clonal evolution in leukemia. *Nature Medicine,* 23(10):1135-1145.

Fialkow, P., 1972. Use of genetic markers to study cellular origin and development of tumors in human females. *Anvances in Cancer Research,*15:191-226.

Finazzi, G. et al., 2014. Calreticulin mutation does not modify the IPSET score for predicting the risk of thrombosis amon 1150 patients with essential thrombocythemia. *Blood,*124(16):2611-2612.

Font, P. et al., 2013. Inter-observer variance with the diagnosis of myelodysplastic syndromes (MDS) following the 2008 WHO classification. *Annals of Hematology,* 92(1):19-24.

Fortuno, C. et al., 2021. Specifications of the ACMG/AMP variant interpretation guidelines for germline TP53 variants. *Human Mutation,* 42(3):223-236

Froyen, G. et al., 2019. Standardization of somatic variant classifications in solid and haematological tumours by a two-level approach of biological and clinical classes: an initiative of the Belgian ComPerMed expert panel. *Cancers (Basel),* 11(12):2030.

Gale, R., Allen, A., Nash, M. & Linch, D., 2007. Long-term serial analysis of X-chromosome inactivation patterns and JAK2 V617F mutant levels in patients with essential thrombocythemia show that minor mutant-positive clones can remain stable for many years. *Blood,* 109(3):1241-1243.

Genomics England, 2020. *Haematological malignancies cancer susceptibility (version 2.5).* [Online]
Available at: https://panelapp.genomicsengland.co.uk/panels/59/
[Accessed 29 November 2020].

Genovese, G. et al., 2014. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England Journal of Medicine,*371(26):2477-2487.

Gerstung, M. et al., 2015. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nature Communications,*6:5901.

Ghosh, R. et al., 2018. Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Human Mutation,* 39(11):1525-1530.

Gisslinger, H. et al., 2016. Clinical impact of bone marrow morphology for the diagnosis of essential thrombocythemia: comparison between the BCSH and the WHO criteria. *Leukemia ,* 30(5):1126-1132.

Gnanasambandan, K. & Sayeski, P., 2011. A structure-function perspective of Jak2 mutations and implications for alternate drug design strategies: the road not taken. *Current Medicinal Chemistry,* 18(30):4659-4673.

Gong, J. et al., 2013. Laboratory practice guidelines for detecting and reporting JAK2 and MPL mutations in myeloproliferative neoplasms. *The Journal of Molecular Diagnostics,* 15(6):733-744.

Gonzalez, D. et al., 2011. Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. *Journal of Clinical Oncology,* 29(16):2223-2229.

Grantham, R., 1974. Amino acid difference formula to help explain protein evolution. *Science,* 185(4154):862-864.

GraphPad, 2020. *Analyze a 2x2 contingency table.* [Online]
Available at: https://www.graphpad.com/quickcalcs/contingency1/
[Accessed 16 December 2020].

Greenberg, P. et al., 2012. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood,* 120(12):2454-2465.

Grinfeld, J. et al., 2018. Classification and personalized prognosis in myeloproliferative neoplasms. *The New England Journal of Medicine,*379:1416-1430.

Guglielmelli, P. et al., 2007. Anaemia characterises patients with myelofibrosis harbouringMplW515L/Kmutation. *British Journal of Hematology,*137:244-247.

Guglielmelli, P. et al., 2009. Identification of patients with poorer survival in primary myelofibrosis based on the burden of JAK2 V617F mutated allele. *Blood,*114(8):1477-1483.

Guglielmelli, P. et al., 2014. The number of prognostically detrimental mutations and prognosis in primary myelofibrosis: an international study of 797 patients. *Leukemia,*28(9):1804-1810.

Guglielmelli, P. et al., 2018. MIPSS70: Mutation-Enhanced International Prognostic Score System for transplantation-age patients with primary myelofibrosis. *Journal of Clinical Oncology,*36(4):310-318.

Haddow, J. & Palomaki, G. 2003. *ACCE: a model process for evaluating data on emerging genetic tests.* In: Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to improve Health and prevent disease. New York: Oxford University Press. pp. 217-233.

Haferlach, T. et al., 2014. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia,* 28(2):241-247.

Haferlach, T. et al., 2014. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia,* 28:241-247.

Ha, J.-S. & Kim, Y.-K., 2015. Calreticulin exon 9 mutations in myeloproliferative neoplasms. *Annals of laboratory medicine,*35(1):22-27.

Heinrichs, S., Li, C. & Look, A., 2010. SNP array analysis in hematologic malignancies: avoiding false discoveries. *Blood,*115(21):4157-4161.

Helliwell, T., Giles, T. & Fensome-Rimmer, S., 2015. *ISO 15189:2012 – An approach to the assessment of uncertainty of measurement for cellular pathology laboratories,* London: The Royal College of Pathologists.

He, M. et al., 2019. Variant Interpretation for Cancer (VIC): a computational tool for assessing clinical impacts of somatic variants. *Genome Medicine,*11:53.

Hinds, D. et al., 2016. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood,* 128(8):1121-1128.

Houdayer, C. et al., 2012. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Human Mutation,* 33(8):1228-1238.

Hussein, k. et al., 2010. International Prognostic Scoring System-independent cytogenetic risk categorization in primary myelofibrosis. *Blood,* 115(3):496-499.

Hussein, K., Van Dyke, D. & Tefferi, A., 2009. Conventional cytogenetics in myelofibrosis: literature review and discussion. *European Journal of Haematology,* 82(5):329-338.

Iacobucci, I. et al., 2019. Integrated transcriptomic and genomic sequencing identifies prognostic constellations of driver mutations in acute myeloid leukemia and myelodysplastic syndromes. *Blood,* 134(Supplement_2):LBA-4.

Illumina, 2010. *Illumina Sequencing Technology.* [Online]
Available at:
https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf
[Accessed 24 October 2020].

Illumina, 2016. *TruSight Myeloid Sequencing Panel Reference Guide.* [Online]
Available at: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/trusight/trusight-myeloid-sequencing-panel-reference-guide-15054779-02.pdf
[Accessed 18 November 2020].

Illumina, 2017. *MiSeq Reporter Software Guide.* [Online]
Available at: https://emea.support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/miseqreporter/miseq-reporter-software-guide-15042295-05.pdf
[Accessed 25 October 2020].

Illumina, 2020. *Illumina TruSight Oncology 500 Product Family.* [Online]
Available at: https://emea.illumina.com/products/by-brand/trusight-oncology-500.html?SCID=2020-271PPC2484
[Accessed 24 October 2020].

International Organization for Standardization, 2012. *ISO 15189:2012(E),* Switzerland: ISO.

International Organization for Standardization, 2019. *ISO 5725-2,* Switzerland: ISO.

Ionnidis, N. et al., 2016. REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics ,*99(4):877-885.

Itzykson, R. et al., 2013. Prognostic score including gene mutations in chronic myelomonocytic leukemia. *Journal of Clinical Oncology,* 31(19):2428-2436.

Jaiswal, S. et al., 2014. Age-related clonal hematopoiesis associated with adverse outcomes. *New England Journal of Medicine,*371(26):2488-2498.

Jaiswal, S. et al., 2017. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *The New England Journal of Medicine,* 377:111-121.

James, C. et al., 2005. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature,*434(7037):1144-1148.

Jang, M.-A. et al., 2020. A rare case of essential thrombocythemia with coexisting JAK2 and MPL driver mutations. *Journal of Korean Medical Science ,* 35(23):e168.

Jawhar, M. et al., 2016. Additional mutations in SRSF2, ASXL1 and/or RUNX1 identify a high-risk group of patients with KIT D816V(+) advanced systemic mastocytosis. *Leukemia,* 30(1):136-143.

Jennings, L. et al., 2017. Guidelines for validation of next-generation sequencing-based oncology panels: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. *The Journal of Molecular Diagnostics,* 19(3):341-365.

Jeromin, S. et al., 2013. High frequencies of SF3B1 and JAK2 mutations in refractory anemia with ring sideroblasts associated with marked thrombocytosis strengthen the assignment to the category of myelodysplastic/myeloproliferative neoplasms. *Haematologica,* 98(2):e15-e17.

Jones, A. et al., 2005. Widespread occurence of the JAK2 V617F mutation in chronic myeloproliferative disorders. *Blood,* 106(6):2162-2168.

Jones, A. et al., 2009. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nature Genetics,* 41(4):446-449.

Jongen-Lavrencic, M. et al., 2018. Molecular minimal residual disease in acute myeloid leukemia. *New England Journal of Medicine,* 378:1189-1199.

Jovanovic, J. et al., 2011. Systematic evalutation of DNA-based quantitative-polymerase chain reaction (Q-PCR) assays to track treatment response in patients with JAK2-V617F associated myeloproliferative neoplasms: a Joint European LeukemiaNet/MPN&MPNr-EuroNet study. *Blood,*118:2812.

Kamaruzzaman, S., Noor, N., Yusoff, S. & Abdullah, W., 2018. Detection of JAK2 V617F Mutation among Donors with Erythrocytosis. *Journal of Clinical and Health Sciences,* 3(2):19-25.

Kanagal-Shamanna, R. et al., 2016. Principles of analytical validation of next-generation sequencing based mutational analysis for hematologic neoplasms in a CLIA-certified labroatory. *Expert Review of Molecular Diagnostics,* 16(4):461-472.

Kang, M.-G.et al., 2016. Coexistence of JAK2 and CALR mutations and thei clinical implications in patients with essential thrombocythemia. *Oncotarget,* 7(35):57036-57049.

Kato, S. et al., 2003. Understanding the function-structure and function-mutation relationships of p53 tumor supressor protein by high-resoluytion missense mutation analysis. *Proc Natl Acad Sci USA,* 100(14):8424-9.

Kilpivaara, O. et al., 2009. A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms. *Nature Genetics,* 41(4):455-459.

Kim, D. & Salzberg, S., 2011. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biology,*12:R72.

Kjaer, L. et al., 2012. A highly sensitive quantitative real-time PCR assay for determination of mutant JAK2 exon 12 allele burder. *PLoS ONE,* 7(3):e33100.

Kohlmann, W. & Schiffman, J., 2016. Discussing and managing hematologic germ line variants. *The American Society of Hematology Education Program,* 2016(1):309-315.

Kouroupi, E. et al., 2012. Does increasing the JAK2 V617F assay sensitivity allow to identify more patients with MPN?. *Blood Cancer Journal,*2:e70.

Kouroupi, E. et al., 2012. Does increasing the JAK2 V617F assay sensitivity allow to identify more patients with MPN?. *Blood Cancer Journal,*2:e70.

Kralovics, R. et al., 2005. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *The New England Journal of Medicine,*352:1779-1790.

Kuo, F., Mar, B., Lindsley, R. & Lindeman, N., 2017. The relative utilities of genome-wide, gene panel, and individual gene sequencing in clinical practice. *Blood,* 130(4):433-439.

La Rocca, F. et al., 2020. Superiority of droplet digital PCR over real-time quantitative PCR for JAK2 V617F allele mutational burden assessment in myeloproliferative neoplasms: a retrospective study. *Diagnostics (Basel),* 10(3):143.

LaFramboise, T., 2009. Single nucleotide polymorphism arrays: a decade of biological computational and technological advances. *Nucleic Acids Research,* 37(13):4181-4193.

Landgren, o. et al., 2008. Increased risks of polycythemia vera, essential thrombocythemia, and myelofibrosis among 24,577 first-degree relatives of 11,039 patients with myeloproliferative neoplasms in Sweden. *Blood,*112:2199-2204.

Langabeer, S., Haslam, K. & Conneally, E., 2014. Monitoring residual disease in the Ph-negative myeloproliferative neoplasms post-allogeneic stem cell transplantation: more mutations and more methodologies. *Frontiers in Oncology,*4:212.

Lange, T. et al., 2013. JAK2 p.V617F allele burden in myeloproliferative neoplasms one moth ater allogenic stem cell transplantation significantly predicts outcome and risk of relapse. *Haematologica,* 98(5):722-728.

Lasho, T. et al., 2006. Concurrent MPL515 and JAK2V617F mutations in myelofibrosis: chronology of clonal emergence and changes in mutant allele burden over time. *British Journal of Haematology,*135:683-687.

Laughlin, T., Moliterno, A., Stein, B. & Rothberg, P., 2010. Detection of exon 12 mutations in the JAK2 gene: enhanced analytical senstivity using clamped PCR and nucleotide sequencing. *Journal of Molecular Diagnosis,* 12(3):278-282.

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):387-397.

Ley, T. et al., 2008. DNA sequencing of a cytogenetically normal acute myeloid leukemia genome. *Nature,* 456(7218):66-72.

Li, H., 2013. *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM,* s.l.: arXiv:1303.3997v1 [q-bio.GN].

Li, M. et al., 2017. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. *The Journal of Molecular Diagnostics,* 19(1):4-3.

Li, M. et al., 2017. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *The Journal of Molecular Diagnostics,* 19(1):4-23.

Lim, K.-H.et al., 2015. Frequent CALR exon 9 alterations in JAK2 V617F-mutated essential thrombocythemia detected by high-resolution melting analysis. *Blood Cancer Journal ,*5:e295.

Lindsley, R. et al., 2015. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood,* 125(9):1367-1376.

Link-Lenczowska, D. et al., 2018. A comparison of qPCR and ddPCR used for quanitification of the JAK2 V617F allele burden in Ph negative MPNs. *Annals of Hematology,*97:2299-2308.

Lucas, F., Michaels, P., Wang, D. & Kim, A., 2020. Mutational analysis of hematologic neoplasms in 164 paired peripheral blood and bone marrow samples by next-generation sequencing. *Blood advances,* 4(18):4362-4365.

Lundberg, P. et al., 2014. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood,*123:2220-2228.

Luo, X. et al., 2019. ClinGen Myeloid Malignancy Variant Curation Expert Panel recommendations for germline RUNX1 variants. *Blood advances,* 3(20):2962-2979.

Lyon, M., 1961. Gene action in the X chromosome of the mouse (Mus musculus L.). *Nature ,*190:372-373.

Magnussen, K., Hasselbalch, H., H., U. & Bjerrum, O., 2013. Characterization of blood donors with high haemoglobin concentration. *Vox Sanguinis,* 104(2):110-114.

Malcovati, L. et al., 2015. SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood,* 126(2):233-241.

Mansier, O. et al., 2016. JAK2 V617F and CALR double mutations are more frequently encountered in patients with low JAK2 V617F allelic burdens. *Leukemia & Lymphoma,* 57(8):1949-1951.

Mansier, O. et al., 2018. Clinical and biological characterization of MPN patients harboring two driver mutations, a French intergroup of myeloprolifertive neoplasms (FIM) study. *American Journal of Hematology,*93:e84-e109.

Marcellino, B. et al., 2017. Advanced Forms of the Myeloproliferative Neoplasms Are Associated with Chromosomal Abnormalities Involving 1q and 12q: Implicating MDM2 and MDM4 in Disease Progression. *Blood,* 130(Supplement 1):204.

Mardis, E., 2013. Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry,*6:287-303.

Martinaud, C., Brisou, P. & Mozziconacci, M.-J., 2010. Is the JAK2(V617F) mutation detectable in healthy volunteers?. *American Journal of Hematology,* 85(4):287-288.

Mattocks, C. et al., 2010. A standardized framework for the validation and verification of clinical molecular genetic tests. *European Journal of Human Genetics,* 18(12):1276-1288.

McClure, R., Hoyer, J. & Mai, M., 2006. The JAK2 V617F mutation is absent in patients with erythrocytosis due to high oxygen affinity hemoglobin variants. *Hemoglobin,* 30(4):487-489.

McClure, R. et al., 2018. Clinical Significance of DNA Variants in Chronic Myeloid Neoplasms: A report of the Association for Molecular Pathology. *Journal of Molecular Diagnostics,*20:717-737.

McGaffin, G., Harper, K., Stirling, D. & McLintock, L., 2014. JAK2 V61F and CALR mutations are not mutually exclusive; findings from a retrospective analysis of a small patient cohort. *British Journal of Haematology,*167:265-289.

McKerrell, T. et al., 2015. Leukemia-associated somatic mutations drive dinstinct patterns of age-related clonal hemopoiesis. *Cell Reports,*10:1239-1245.

McKerrell, T. et al., 2017. JAK2 V617F hematopoietic clones are present several years prior to MPN diagnosis and follow different expansion kinetics. *Blood Advances,* 1(14):968-971.

McKusick-Nathans Institute of Genetic Medicine, J. H. U., 2020. *Online Mendelian Inheritance in Man, OMIM®.*. [Online]
Available at: https://omim.org/
[Accessed 27 January 2021].

Meggendorfer, M. et al., 2018. The mutational landscape of 18 investigated genes clearly separates four subtypes of myelodysplastic/myeloproliferative neoplasms. *Haematologica,* 103(5):e192-e195.

Mehrotra, M. et al., 2015. Genomic and Clinicopathologic Features of Primary Myelofibrosis With Isolated 13q Deletion. *Clinical Lymphoma, Melanoma & Leukemia,* 15(8):496-505.

Merker, J. et al., 2012. Whole genome sequence analysis in primary myelofibrosis. *Blood,* 120(21):2863.

Mikhail, F. et al., 2019. Technical laboratory standards for interpretation and reporting of acquired copy-number abnormalities and copy-neutral loss of heterozygosity in neoplastic disorders: a joint consensus recommendation from the ACMG and the CGC. *Genetics in Medicine,* 21(9):1903-1916.

Miller, J., Testa, J., Lindgren, V. & Rowley, J., 1985. The pattern and clinical significance of karyotypic abnormalities in patients with idiopathic and postpolycythemic myelofibrosis. *Cancer,* 55(3):582-591.

Milosevic, J. et al., 2016. Whole-exome sequencing identifies novel MPL and JAK2 mutations in triple-negative myeloproliferative neoplasms. *Blood,* 127(3):325-332.

Mitchell, A., Zwick, M., Chakravarti, A. & Cutler, D., 2004. Discrepancies in dbSNP confirmation rates and allele frequency distributions from varying genotyping error rates and patterns. *Bioinformatics,* 20(7):1022-1032.

Mughal, T. et al., 2015. An International MDS/MPN Working Group's perspective and recommendations on molecular pathogenesis, diagnosis and clinical characterization of myelodysplastic/myeloproliferative neoplasms. *Haematologica,* 100(9):1117-1130.

Mullis, K. et al., 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbour Symposia on Quantitative Biology,* 51(1):263-273.

Munoz-Gonzalez, J. et al., 2018. Impact of somatic and germline mutations on the outcome of systemic mastocytosis. *Blood Advances,* 2(21):2814-2828.

Munoz-Gonzalez, J. et al., 2019. Frequency and prognostic impact of KIT and other genetic variants in indolent systemic mastocytosis. *Blood,*134:456-468.

Najfeld, V., Montella, L., Scalise, A. & Fruchtman, S., 2002. Exploring polycythaemia vera with fluorescence in situ hybridization: additional cryptic 9p is the most frequent abnormality detected. *British Journal of Haematology,* 119(2):558-566.

Najfeld, J. et al., 2010. Jumping translocations of the long arms of chromosome 1 in myeloid malignancies is associated with a high risk of transformation to acute myeloid leukaemia. *British Journal of Haematology,* 151(3):288-291.

Nangalia, J. et al., 2013. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *The New England Journal of Medicine,*369:2391-2405.

National Comprehensive Cancer Network, 2016. *NCCN Clinical Practice Guidelines in Oncology: Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma v1.2017.* [Online]
Available at: https://www.nccn.org/professionals/physician_gls/pdf/cll.pdf
[Accessed 14 February 2017].

National Comprehensive Cancer Network, 2019. *NCCN Guidelines: Myeloproliferative Neoplasms v3.2019.* [Online]
Available at: https://www.nccn.org/professionals/physician_gls/pdf/mpn.pdf
[Accessed 30 October 2019].

National Comprehensive Cancer Network, 2021. *NCCN Clinical Practcie Guidelines in Oncology: Myelodysplastci syndromes.* [Online]

Available at: https://www.nccn.org/professionals/physician_gls/pdf/mds.pdf
[Accessed 21 February 2021].

NCB1 Resource Coordinators, 2018. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research,* 46(D1):D8-D13.

NHS England, 2020. *National Genomic Test Directory for cancer.* [Online]
Available at: https://www.england.nhs.uk/publication/national-genomic-test-directories/
[Accessed 29 November 2020].

Nielsen, C. et al., 2011. The JAK2 V617F somatic mutation, mortality and cancer risk in the general population. *Haematologica,* 96(3):450-453.

Nielsen, C., Birgins, H., Nordestgaard, B. & Bojesen, S., 2013. Diagnostic value of JAK2 V617F somatic mutation for myeloproliferative cancer in 49 488 individuals from the general population. *British Journal of Haematology,* 160(1):70-79.

Nielsen, C. et al., 2014. JAK2V617F somatic mutation in the general population: myeloproliferative neoplasm development and progression rate. *Haematologica,* 99(9):1448-1455.

NimaGen, 2020. *EasySeq NGS: reverse complement PCR.* [Online]
Available at: https://www.nimagen.com/brands/easyseq-ngs-rc-pcr
[Accessed 4 January 2021].

Nowell, P., 1976. The clonal evolution of tumor cell populations. *Science,*194:23-28.

OGT, 2016. *Selecting the best NGS enrichment assay for your needs.* [Online]
Available at: https://www.ogt.com/assets/0000/6850/990215_SureSeq_Hyb_V_Aplicon_Low-res_261016.pdf
[Accessed 25 October 2020].

Olcardu, D. et al., 2009. A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nature Genetics,* 41(4):450-454.

OMIM, 2017. *JANUS KINASE 2; JAK2.* [Online]
Available at: http://omim.org/entry/147796
[Accessed 02 January 2018].

Ortmann, C. et al., 2015. Effect of mutation order on myeloprliferative neoplasms. *The New England Journal of Medicine,* 372(7):601-612.

Ortmann, C. et al., 2015. Effect of mutation order on myeloproliferative neoplasms. *The New England Journal of Medicine,* 372(7):601-612.

Papaemmanuil, E. et al., 2013. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood,* 122(22):3616-3627.

Pardanani, A. et al., 2006. MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood,* 108(10):3472-3476.

Parmentier, S. et al., 2012. Assessment of dysplastic hematopoiesis: lessons from healthy bone marrow donors. *Haematologica,* 97(5):723-730.

Passamonti, F. et al., 2010. A prospective study of 338 patients with polycythemia vera: the impact of JAK2 (V617F) allele burden and leukocytosis on fibrotic or leukemic disease transformation and vascular complications. *Leukemia,*24:1574-1579.

Passamonti, F. et al., 2007. JAK2 (V617F) mutation in healthy individuals. *British Journal of Haematology,*136:677-679.

Patel, A. et al., 2019. JAK2 ex13InDel drives oncogenic transformation and is associated with chronic eosinophilic leukemia and polycythemia vera. *Blood,* 134(26):2388-2398.

Patnaik, M. et al., 2010. WHO-defined 'myelodysplastic syndrome with isloated del(5q)' in 88 consecutive patients: survival data, leukemic transformation rates and prevalence of JAK2, MPL and IDH mutations. *Leukemia,*24:1283-1289.

Patnaik, M. & Lasho, R., 2020. Myelodysplastic syndrome/myeloproliferative neoplasm overlap syndromes: a focused review. *Hematology: The American Society of Hematology Education Program,* 2020(1):460-464

Patnaik, M. & Tefferi, A., 2019. Refractory anemia with ring sideroblasts (RARS) and RARS with thrombocytosis: "2019 Update on Diagnosis, Risk-stratification, and Management". *American Journal of Haematology,* 94(4):475-488.

Percy, M. et al., 2006. The incidence of the JAK2 V617F mutation in patients with idiopathic erythrocytosis. *Haematologica,* 91(3):413-414.

Pertea, M., Lin, X. & Salzberg, S., 2001. GeneSplicer: a new computational method for splice site prediction.. *Nucleic Acids Research,* 29(5):1185-1190.

Poodt, J., Fijnheer, R., Walsh, I. & Hermans, M., 2006. A sensitive and reliable semi-quantitative real-time PCR assay to detect JAK2 V617F in blood. *Hematological Oncology,* 24(4):227-233.

Poplin, R. et al., 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv:*201178.

Rack, K. et al., 2019. European recommendations and quality assurance for cytogenomic analysis of haematological neoplasms. *Leukemia,*33:1851-1867.

Rapado, I. et al., 2008. Validity test study of JAK2 V617F and allele burden quantification in the diagnosis of myeloproliferative diseases. *Annals of Hematology,* 87(9):741-749.

Rapado, I. et al., 2009. High resolution melting analysis for JAK2 exon 14 and exon 12 mutations: a diagnostic tool for myeloproliferative neoplasms. *The Journal of Molecular Diagnostics,* 11(2):155-161.

Rawstron, A. et al., 2008. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *The New England Journal of Medicine,* 359(6):575-583.

Reese, M., Eeckman, F., Kulp, D. & Haussler, D., 1997. Improved splice site detection in genie. *Journal of Computational Biology,*4:311-323.

Rego de Paula, M. et al., 2018. High Frequency of Copy-Neutral Loss of Heterozygosity in Patients with Myelofibrosis. *Cytogenetic and Genome Research,*154:62-70.

Rehm, H. et al., 2013. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med,* 15(9):733-747.

Reilly, J. et al., 1997. Cytogenetic abnormalities and their prognostic significance in idiopathic myelofibrosis: a study of 106 cases. *British Journal of Haematology,* 98(1):96-102.

Reilly, J. et al., 1994. Karyotypic and ras gene mutational analysis in idiopathic myelofibrosis. *British Journal of Haematology,*88:575-581.

Richards, S. et al., 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine,*17:405-423.

Risques, R. & Kennedy, S., 2018. Aging and the rise of somatic cancer-associated mutations in normal tissues. *PLOS Genetics,* 14(1):e1007108.

Ronaghi, M., 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Research,*11:3-11.

Rossi, D. et al., 2014. Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood,* 123(14):2139-2147.

Rossi, G. et al., 2000. Cytogenetic analogy between myelodysplastic syndrome and acute myeloid leukemia of elderly patients. *Leukemia,* 14(4):636-641.

Rumi, E. et al., 2014. CALR exon 9 mutations are somatically acquired events in familial cases of essential thrombocythemia or primary myelofibrosis. *Blood,*123:2416-2419.

Rumi, E. et al., 2014. Clinical effect of driver mutations of JAK2, CALR, or MPL in primary myelofibrosis. *Blood,*124:1062-1069.

Saharinen, P., Takaluoma, K. & Silvennoinen, O., 2000. Regulation of the Jak2 tyrosine kinase by its pseudokinase domain. *Molecular and Cellular Biology ,* 20(10):3387-3395.

Samorodnitsky, E. et al., 2015. Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Human Mutation,* 36(9):903-914.

Sandoval, J. et al., 2019. Mutations in the DNMT3A DNA methyltransferase in acute myeloid leukemia patients cause both loss and gain of function and differential regulation by protein partners. *Journal of Biological Chemistry,* 294(13):4898-4910.

Sanger, F., Nicklen, S. & Coulson, A., 1977. DNA sequencing with chain-terminating inhibitors. *PNAS,* 74(12):5463-5467.

Sarvothaman, S. et al., 2015. Apoptosis:role in myeloid cell development. *Blood Research,* 50(2):73-79.

Scheinin, I. et al., 2014. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Research,*24:2022-2032.

Schieber, M., Crispino, J. & Stein, B., 2019. Myelofibrosis in 2019: moving beyond JAK2 inhibition. *Blood Cancer Journal,*9:74.

Schischlik, F. et al., 2019. Mutational landscape of the transcriptome offers putative targets for immunotherapy of myeloproliferative neoplasms. *Blood,* 134(2):199-210.

Schwaab, J. et al., 2013. Comprehensive mutational profiling in advanced systemic mastocytosis. *Blood,* 122(14):2460-2466.

Scott, L. et al., 2007. JAK2 exon 12 mutations in polycythemia ver and idiopathic erythrocytosis. *The New England Journal of Medicine,* 356(5):459-468.

Scott, L. et al., 2007. JAK2 exon 12 mutations in polycythemia vera and idiopathic erthrocytosis. *The New England Journal of Medicine,*356:459-468.

Shapiro, M. & Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Research,* 15(17):7155-7174.

Shepherd, M. et al., 2018. Single-cell approaches identify the molecular network driving malignant haematopoietic stem cell self-renewal. *Blood*, 132(8):791-803.

Sheshan, V. & Olshen, A., 2020. *DNAcopy: DNA copy number data analysis. R Package version 1.64.0,* s.l.: Bioconductor.

Shlush, L. et al., 2014. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature,*506:328-333.

Shvarts, A. et al., 1996. MDMX: a novel p53-binding protein with some functional properties of MDM2. *European Molecular Biology Organization,* 15(19):5349-5357.

Sidon, P., El Housni, H., Dassars, B. & Heimann, P., 2006. The JAK2V617F mutation is detectable at very low level in peripheral blood of healthy donors. *Leukemia,*20:1622.

Skoda, R., Duek, A. & Grisouard, J., 2015. Pathogenesis of myeloproliferative neoplasms. *Experimental Hematology,*43:599-608.

SOPHiA GENETICS, 2020. *Alamut Visual.* [Online]
Available at: https://www.interactive-biosoftware.com/alamut-visual/
[Accessed 26 January 2021].

Spivak, J., 2017. Myeloproliferative neoplasms. *The New England Journal of Medicine,*376:2168-2181.

Stanley, E., 2009. Cytokines and cytokine receptors regulating cell survival, proliferation, and differentiation in heamtopoiesis. In: R. Bradshaw & E. Dennis, eds. *Handbook of cell signaling:3.* Oxford: Academic Press:2733-2742.

Stark, R., Grzelak, M. & Hadfield, J., 2019. RNA sequencing: the teenage years. *Nature Reviews: Genetics,*20:631-656.

La Starza, R. et al., 1998. Molecular delineation of 13q deletion boundaries in 20 patients with myeloid malignancies. *Blood,* 91(1):231-237.

Steensma, D. et al., 2005. The JAK2 V617F activating tyrosine kinase mutation is an infrequent event in both "atypical" myeloproliferative disorders and myelodysplastic syndromes. *Blood ,*106:1207-1209.

Steensma, D. et al., 2015. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood,* 126(1):9-16.

Strobbe, L., P., L., Hermans, M. & Fijnheer, R., 2007. Detection of the JAK2V617F mutation in patients with slightly elevated platelets or hemoglobin without a secondary cause. *Annals of Hematology,* 86(11):801-803.

Sukhai, M. et al., 2016. A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer. *Genetics in Medicine,* 18(2):128-136.

Swerdlow, S. et al., 2017. Myeloprliferative neoplasms. In: S. Swerdlow, et al. eds. *WHO classification of tumours of haematopoietic and lymphoid tissues.* Lyon: WHO Press:30-59.

Swerdlow, S. et al., 2016. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood,*127:2375.

Tagariello, G. et al., 2009. The JAK2V617F tyrosine kinase mutation in blood donors with upper-limit haematocrit levels. *Blood Transfusion,* 7(2):111-116.

Tam, C. et al., 2009. The role of cytogenetic abnormalities as a prognostic marker in primary myelofibrosis: applicability at the time of diagnosis and later during disease course. *Blood,* 113(18):4171-4178.

Tanaka, K. et al., 1999. Frequent allelic loss of the RB, D13S319 and D13S25 locus in myeloid malignancies with deletion/translocation at 13q14 of chromosome 13, but not in lymphoid malignancies. *Leukemia,*13:1367-1373.

Tang, R., Prosser, D. & Love, D., 2016. Evaluation of Bioinformatic Programmes for the Analysis of Variants within Splice Site Consensus Regions. *Advances in Bioinformatics,* 2016(5614058):1-10.

Tapper, W. et al., 2015. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. *Nature Communications,*6:6691.

Tavtigian, S. et al., 2006. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *Journal of Medical Genetics,* 43(4):295-305.

Tayoun, A. et al., 2018. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Human Mutation,* 39(11):1517-1524.

Tefferi, A., et al., 2001. Comparison of peripheral blood interphase cytogenetics with bone marrow karyotype analysis in myelofibrosis with myeloid metaplasia. *British Journal of Haematology,* 115(2):316-319.

Tefferi, A. et al., 2005. Concomitant neutrophil JAK2 mutation screening and PRV-1 expression analysis in myeloproliferative disorders and secondary polycythaemia. *British Journal of Haematology,* 131(2):166-171.

Tefferi, A., 2008. The history of myeloproliferative disorders: before and after Dameshek. *Leukemia,* 22(1):3-13.

Tefferi, A. et al., 2008. Low JAK2 V617F allele burden in primary myelofibrosis, compared to either a higher allele burden or unmutated status, is associated with inferior overall and leukemia-free survival. *Leukemia,*22:756-761.

Tefferi, A., Skoda, R. & Vardiman, J., 2009. Myeloproliferative neoplasms: contemporary diagnosis using histology and genetics. *Nature Reviews Clinical Oncology,*6:627-637.

Tefferi, A., 2014. Primary myelofibrosis: 2014 update on diagnosis, risk-stratification, and management. *American Journal of Hematology,*89:915-925.

Tefferi, A. et al., 2014. CALR vs JAK2 vs MPL-mutated or triple-negative myelofibrosis: clinical, cytogenetic and molecular comparisons. *Leukemia,*28:1472-1477.

Tefferi, A. et al., 2018. MIPSS70+ Version 2.0: Mutation and Karyotype-Enhanced International Prognostic Scoring System for Primary Myelofibrosis. *Journal of Clinical Oncology,*36:1769-1770.

Tefferi, A., et al., 2018. Myelofibrosis Treatment Algorithm 2018. *Blood Cancer Journal,*8:72.

Tefferi, A., et al., 2011. Long-Term Outcome of Treatment with Ruxolitinib in Myelofibrosis. *The New England Journal of Medicine,*365:1455-1457.

Tefferi, A., Noel, P. & Hanson, C., 2011. Uses and abuses of JAK2 and MPL mutation tests in myeloproliferative neoplasms: a paper from the 2010 William Beaumont Hospital Symposium on molecular pathology. *The Journal of Molecular Diagnostics,* 13(5):461-466.

Tefferi, A. & Pardanani, A., 2014. CALR mutations and a new diagnostic algorithm for MPN. *Nature Reviews Clinical Oncology,*11:125-126.

Tefferi, A. et al., 2017. Risk factors and a prognostic model for postsplenectomy survival in myelofibrosis. *American Journal of Hematology,* 92(11):1187-1192.

Tefferi, A. et al., 2018. Revised cytogenetic risk stratification in primary myelofibrosis: analysis based on 1002 informative patients. *Leukemia,* 32(5):1189-1199.

ThermoFisher Scientific, 2017. *Sanger Sequencing.* [Online]
Available at: https://www.thermofisher.com/uk/en/home/life-science/sequencing/sanger-sequencing/sanger_sequencing_method.html#
[Accessed 5 April 2017].

ThermoFisher Scientific, 2018. *Sanger Sequencing Workflow.* [Online]
Available at: https://www.thermofisher.com/us/en/home/life-science/sequencing/sanger-sequencing/sanger-sequencing-workflow.html
[Accessed 17 January 2018].

ThermoFisher Scientific, 2020. *What is fragment analysis.* [Online]
Available at: https://www.thermofisher.com/uk/en/home/life-science/sequencing/sequencing-learning-center/capillary-electrophoresis-information/what-is-fragment-analysis.html
[Accessed 20 January 2021].

Thomas, M. et al., 2017. Integration of Technical, Bioinformatic, and Variant Assessment Approaches in the Validation of a Targeted Next-Generation Sequencing Panel for Myeloid Malignancies. *Archive of Pathology Laboratory Medicine,*141:759-775.

Tian, Y. et al., 2019. REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. *Scientific Reports,*9:Scientific Reports.

Titmarsh, G. et al., 2014. How common are myeloprliferative neoplasms? A systematic review and meta-analysis. *American Journal of Haematology,* 89(6):581-587.

Toledo, F. & Wahl, G., 2006. Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nature Reviews Cancer,* 6(12):909-923.

UKAS, 2020. *United Kingdom Accreditation Service.* [Online]
Available at: https://www.ukas.com/
[Accessed 18 06 2020].

UniProt Consortium, 2020. *UniProt.* [Online]
Available at: https://www.uniprot.org/
[Accessed 29 November 2020].

Vainchenker, W. & Kralovics, R., 2017. Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. *Blood,* 129(6):667-679.

Valent, P. et al., 2019. Clonal Hematopoiesis with Oncogenic Potential (CHOP): Separation from CHIP and Roads to AML. *International Journal of Molecular Sciences,* 20(3):789.

Vallapureddy, R. et al., 2019. Leukemic transformation among 1306 patients with primary myelofibrosis: risk factors and development of a predictive model. *Blood Cancer Journal,*9:12.

Van Allen, E., Wagle, N. & Levy, M., 2013. Clinical analysis and interpretation of cancer genome data. *Journal of Clinical Oncology,* 31(15):1825-1833.

Van Egeren, D. et al., 2021. Reconstructing the lineage histories and differentiation trajectories of individual cancer cells in myeloproliferative neoplasms. *Cell Stem Cell*, 28(3):514-523

Van de Wiel, M. & Vosse, S., 2020. *CGHcall: Calling aberrations for array CGH tumor profiles. R package version 2.52.0,* s.l.: Bioconductor.

Vannucchi, A. et al., 2007. Clinical profile of homozygous JAK2 617V>F mutation in patients with polycythemia vera or essential thrombocythemia. *Blood,*110:840-846.

Vaser, R. et al., 2016. SIFT missense predictions for genomes. *Nature protocols,* 11(1):1-9.

Vassiliou, G., 2016. JAK2 V617F clonal disroders: fate or chance?. *Blood,* 128(8):1032-1033.

Verstovsek, S., Cross, N. & Tefferi, A., 2006. JAK2V617F mutational frequency in polycythemia vera: 100%, >90%, less?. *Leukemia,*20:2067.

Vlahovic, G. & Crawford, J., 2003. Activation of tryosine kinases in cancer. *The Oncologist,* 8(6):531-538.

Wadleigh, M. & Tefferi, A., 2010. Classification and diagnosis of myeloproliferative neoplasms according to the 2008 World Health Organization criteria. *Progress in Hematology,*91:174-179.

Wang, k., Li, M. & Hakonarson, H., 2010. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Research,*38:e164.

Wang, Y. et al., 2008. JAK2 mutations are present in all cases of polycythemia vera. *Leukemia,*22:1289.

Weinberg, I. et al., 2012. Janus Kinase V617F mutation in cigarette smokers. *American Journal of Hematology,* 87(1):5-8.

Welch, J. et al., 2016. TP53 and Decitabine in Acute Myeloid Leukemia and Myelodysplastic Syndromes. *New England Journal of Medicine,*375:2023-2036.

Williams, N. et al., 2020. Phylogenetic reconstruction of myeloproliferative neoplasm reveals very early origins and lifelong evolution. *BioRxiv,*11.09.374710.

Williams, N. et al., 2020. Driver Mutation Acquisition in Utero and Childhood Followed By Lifelong Clonal Evolution Underlie Myeloproliferative Neoplasms. *Blood,* 136(Supplement_2):LBA-1.

Wilm, A. et al., 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing dataset. *Nucleic Acids Research,* 40(22):11189-111201.

Wolschke, C. et al., 2017. Impact of molecular residual disease post allografting in myelofibrosis patients. *Bone Marrow Transplantation,*1526-1529:52.

Wong, T. et al., 2016. Rapid expansion of preexisting nonleukemic hematopoietic clones frequently follows induction therapy for de novo AML. *Blood,* 127(7):893-897.

Wouters, H. et al., 2020. Erythrocytosis in the general population: clinical characteristics and association with clonal hematopoiesis. *Blood advances,* 4(24):6353-6361.

Xiang, J., Peng, J., Baxter, S. & Peng, Z., 2020. AutoPVS1: An automatic classification tool for PVS1 interpretation of null variants. *Human Mutation,* 41(9):1488-1498.

Xie, M. et al., 2014. Age-related cancer mutations associated with clonal hematopoietic expansion. *Nature Medicine,* 20(12):1472-1478.

Xing, C.-Y., Li, H.-Y., Wu, J.-B. & Gao, S.-M., 2016. Co-occurence of JAK2 V617F and an uncommon CALR del (p.K368fs*51) mutation facilitates JAK2/STAT signaling in polycythemia vera. *Leukemia & Lymphoma,* 57(7):1743-1745.

Xu, N. et al., 2015. A report on the co-occurence of JAK2 V617F and CALR mutations in myeloproliferative neoplasm patients. *Annals of Hematology ,*94:865-867.

Xu, X. et al., 2007. JAK2V617F: prevalence in a large Chinese hospital population. *Blood,*109:339-342.

Yang, F. et al., 2015. Protein domain-level landscape of cancer-type-specific somatic mutations. *PLOS Computational Biology,* 11(3):e1004147.

Yeo, G. & Burge, C., 2004. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *Journal of Computational Biology,*11:377-394.

Yoshizato, T. et al., 2015. Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia. *The New England Journal of Medicine,*373:35-47.

Young, A., Challen, G., Birmann, B. & Druley, T., 2016. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in health adults. *Nature Communications,*7:12484.

Zainuddin, N. et al., 2011. TP53 Mutations are infrequent in newly diagnosed chronic lymphocytic leukemia. *Leukemia Research,* 35(2):272-274.

Zink, F. et al., 2017. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood ,* 130(6):742-752.

Zoi, K. & Cross, N., 2017. Genomics of myeloproliferative neoplasms. *Journal of Clinical Oncology,* 35(9):947-954.

7. Appendices

7.1 Appendix 1 – Illumina© sequencing by synthesis

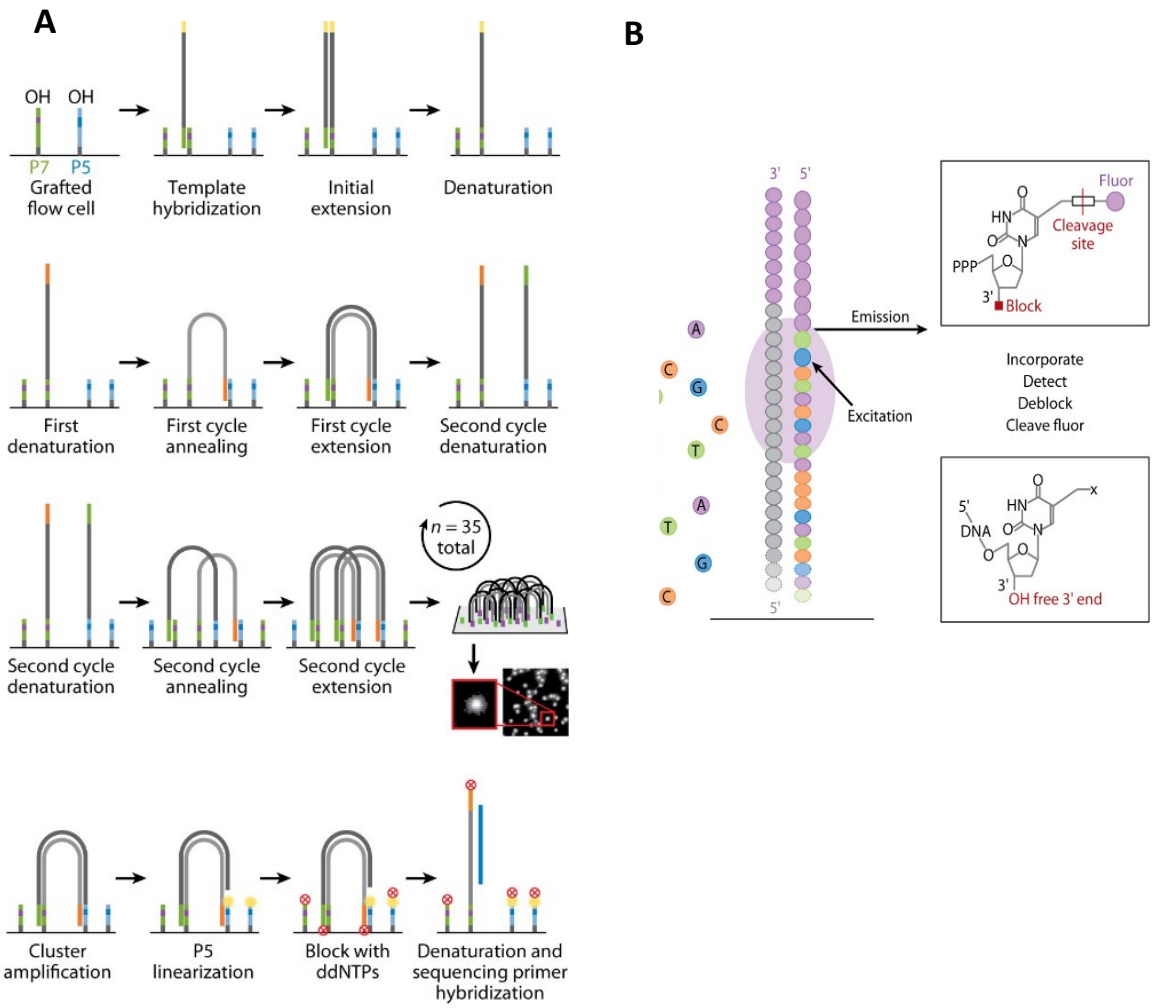


Figure 7.1 A sequencing library is prepared through either fragmentation of DNA or through PCR methodologies but both result in DNA sequences with specific adapters on the ends [image not shown]. At Wessex Regional Genetics Laboratory (WRGL), input DNA is subjected to a novel PCR methodology that appends all the functional sequences necessary for sequencing on a MiSeq instrument (sequencing primer binding, sample ID indexes and flow cell hybridisation adaptors). The Illumina microfluidic conduit is a flow cell composed of flat glass with eight microfluidic channels, each decorated by covalent attachment of adapter sequences complementary to the library adapters. The library is amplified in situ on the flow cell surfaces by use of a bridge amplification step to produce foci for sequencing (clusters) (**A**). The series of events in each step includes the following, in order of occurrence: (i) The nucleotide is added by polymerase, (ii) unincorporated nucleotides are washed away, (iii) the flow cell is imaged on both inner surfaces to identify each cluster that is reporting a fluorescent signal, (iv) the fluorescent groups are chemically cleaved, and (v) the 3'-OH is chemically deblocked (**B**). This series of steps is repeated.
Source: Mardis (2013)

7.2 Appendix 2 – The Illumina® TSMP protocol

Please note that all reagents are provided by Illumina at a dilution that is ready to use unless otherwise stated.

Firstly, libraries are prepared from DNA isolated from blood or bone marrow samples via extension and ligation between oligonucleotide probes that flank the regions of interest. Libraries are prepared in batches of 16 samples and 100 ng of genomic DNA (gDNA). The DNA is diluted (using TE buffer) to 10 ng/µl based on Qubit readings (alternatively DNA is used neat if <10ng/µl) and 10µl DNA is added to 40 µl of a mix of TruSight Oligos and Oligo Hybridisation for sequencing reagent as per the manufacturer's instructions (Illumina, 2016) and takes places on a thermocycler according to the following protocol: initial denaturation at 95°C for 1min, followed by hybridization using a decreasing temperature gradient from 95°C to 40°C (110 cycles: 30sec per cycle, 0.5°C decrement per cycle), and hold at 40°C. The samples are then prepared for extension ligation by removing unbound oligos by three wash processes through a filter plate unit provided by Illumina: twice with Stringent Wash 1 (SW1) and once with Universal Buffer 1 (UB1) centrifuged at 24000x g for 5 minutes. The gDNA remains bound to the filter plate membrane in each column.

The target DNA sequence between the upstream and downstream oligos is then in-filled using a DNA polymerase and the extended sequence is ligated to the 5' end of the downstream oligo using a DNA ligase. All components for this are in Extension Ligation Mix 4 (ELM4), provided by Illumina; 45 µl of ELM4 is added to each sample and incubated for 45 minutes at 37^{o}C. The filter plate is then sealed and centrifuged at 2400x g for 5 minutes and the pour-off discarded. The extension-ligation products are finally eluted into 30 µl of 50 mM sodium hydroxide (NaOH).

The extension-ligation products are then amplified and sample-specific index sequences are then tagged to the products of this reaction which enables pooling of samples into one library for higher-throughput sequencing. The indexes used are i7 and i5 adapters and sequences that are required for cluster formation on the MiSeq (described below); each sample in each library preparation requires a unique combination of index adapters to allow for sample identification of specific sequencing products after library pooling. Per sample, 22 µl of PCR Master Mix 2 (PMM2), 0.45 µl TruSeq DNA Polymerase 1 (TDP1) and a combined volume of 8 µl of i7/i5 indexes are loaded into a new PCR plate along with 20 µl of

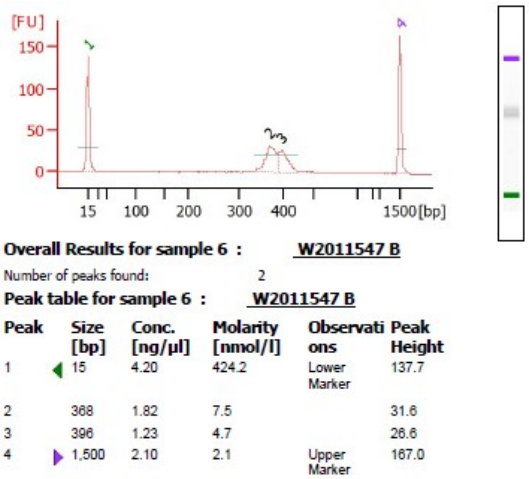
extension-ligation product which was eluted into NaOH in the previous step. The PCR amplification takes places on a thermocycler according to the following protocol:

95°C for 30 seconds
66°C for 30 seconds } for 27 cycles
72°C for 60 seconds

4°C hold

Library clean-up is then performed using AMPure XP beads in order to purify the PCR products from other reaction components. Beads with bound library amplicons are pulled to the side of wells using a magnetic stand and are washed using 80% ethanol. The purified library is then eluted from the beads ready for quality control checking and normalisation.

Quality control (QC) check is then performed for each sample. As recommended by Illumina, this laboratory uses Agilent 2100 Bioanalyzer, a microfluidics-based electrophoresis instrument which can be used for assessing the size, quantity and purity of the samples post-library preparation and prior to sequencing (samples at this stage have been PCR amplified, labelled with index primers and had a bead clean-up) according to manufacturer's instructions (Agilent, 2018). This method allows for assessment of the average amplicon length (peak size, bp), the amount of library product per sample (assessed by peak height) and that the adapters and other PCR components have been removed (shown to be successful by the absence of additional peaks). Preliminary experiments showed that samples with a peak height of <50 arbitrary fluorescence units (FU) consistently failed to sequence using this panel due to insufficient library product and as such this cut-off was used as a mandatory QC requirement prior to sequencing (Figure 7.2).



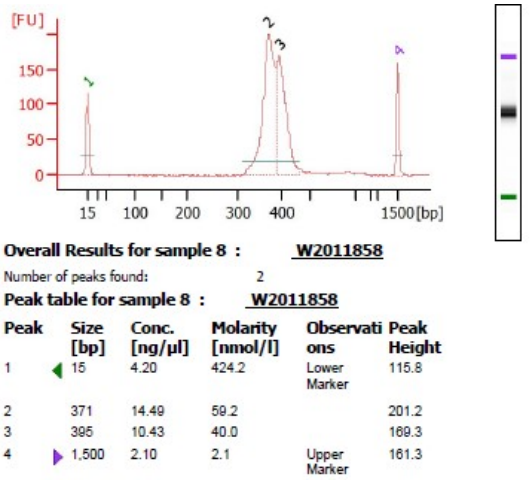
Overall Results for sample 6 : W2011547 B

Number of peaks found: 2

Peak table for sample 6 : W2011547 B

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Peak Height
1	15	4.20	424.2	Lower Marker	137.7
2	368	1.82	7.5		31.6
3	396	1.23	4.7		26.6
4	1,500	2.10	2.1	Upper Marker	167.0

A



Overall Results for sample 8 : W2011858

Number of peaks found: 2

Peak table for sample 8 : W2011858

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations	Peak Height
1	15	4.20	424.2	Lower Marker	115.8
2	371	14.49	59.2		201.2
3	395	10.43	40.0		169.3
4	1,500	2.10	2.1	Upper Marker	161.3

B

Figure 7.2 Examples of bioanalyzer readings for a sample that fails (A) to meet the required FU cut-off and a sample that meets the required cut-off (B). There are four peaks apparent in each trace: peaks 1 and 4 are the size markers; peaks 2 and 3 represent the sample PCR products of varying sizes, which should be ~350-400 bp in length. For each trace, the peak height (FU) and the estimated size (bp), concentration (ng/µl) and molarity (nmol/l) is provided in tabulated form as calculated by the Agilent Bioanalyzer software.

The next step is library normalisation to standardise the quantity of each sample's library prior to pooling. Normalisation is performed using 20 µl PCR product and 52 µl master mix of normalisation beads provided by Illumina. The beads are washed then the normalised library is chemically removed from the beads ready for library pooling.

An equal volume of each normalised library (9 µl) is pooled into a single tube and 9 µl of the pooled library is sequenced on an Illumina MiSeq™. The MiSeq utilises Illumina's sequencing by synthesis (SBS) on a flow cell which acts as a microfluidic conduit. The flow cell is composed of flat glass with eight microfluidic channels, each decorated by covalent attachment of adapter sequences complementary to the library adapters. The library is amplified *in situ* on the flow cell surfaces by use of a bridge amplification step to produce foci for sequencing (clusters). The series of events in each step includes the following, in order of occurrence: (i) The nucleotide is added by polymerase, (ii) unincorporated nucleotides are washed away, (iii) the flow cell is imaged on both inner surfaces to identify each cluster that is reporting a fluorescent signal, (iv) the fluorescent groups are chemically cleaved, and (v) the 3'-OH is chemically deblocked. This series of steps is repeated and base calls are made by interpretation of signal intensity measurements during each cycle (Illumina, 2010; Mardis, 2013). A PhiX Control v3 Library ("PhiX spike in") is also incorporated into each run to increase the diversity of the library and assist with overall increased run quality, as recommended by Illumina; it also acts as a control in the event of a run failure.

The whole process takes ~3 days including 3 hours of hands-on practical work to generate the library, perform QC checks and load the library onto the MiSeq.

The data produced is analysed (secondary analysis) using on-instrument software or equivalent to produce sample specific variant call files (.vcf) containing information about single nucleotide variants (SNV) or small insertion/deletions (indels) which can be used for result interpretation. At this stage, variants are usually classified and/or annotated and reported according to laboratory protocols

Using MiSeq Reporter software as the secondary analysis, quality scores are calculated for each base call, termed Q-scores; these are scored in base call (*.bcl) files after cycle 25, with scores for the base call quality per cycle. The minimum Q-score is an adjustable setting within the software to specify the minimum base call Q-score to use as input to variant calling, giving confidence that the variant is genuine: the higher the score, the more reliable the base call. The Q-score is based upon the Phred scale to give the probability of error according to the calculation shown in Figure 7.3.

Given a base call, x, the probability that x is not true, P(~x), results in a quality score, Q(x), according to:

$$Q(x) = -10 \log_{10}(P(\sim x))$$

Quality Score Q(x)	Error Probability P(~x)
Q40	0.0001 (1 in 10,000)
Q30	0.001 (1 in 1,000)
Q20	0.01 (1 in 100)
Q10	0.1 (1 in 10)

Figure 7.3 Q-score calculations within MiSeq Reporter. *Source: adapted from Illumina, (2017)*

Individual sample .vcf files are then uploaded into Agilent Technologies® Alissa Interpret, a web-based genomic data management software tool which provides functionality for variant storage and filtering, variant interpretation support and variant annotation. Variant filtering is managed via the editable classification tree (High-level screenshot shown in Figure 7.4). The software also uses role-based user access control to support audit trail record keeping within the software in real-time as samples are analysed (Figure 7.5).

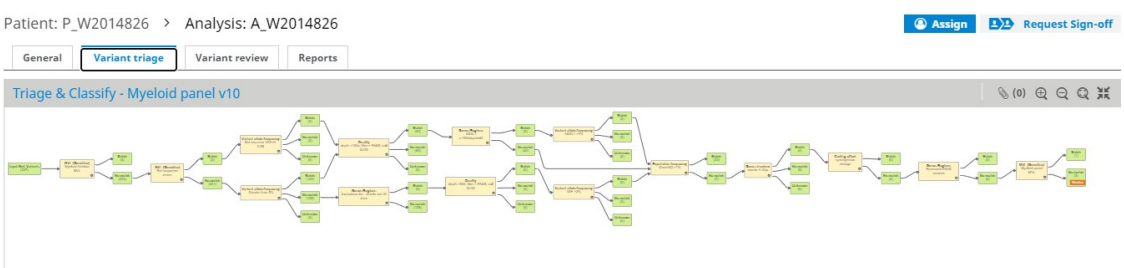


Figure 7.4 Example of variant filtering as displayed in Alissa interpret. Each step of the variant triage is recorded and accessible and Alissa Interpret allows visualisation of all variants present in the sample specific .vcf file prior to filtering.

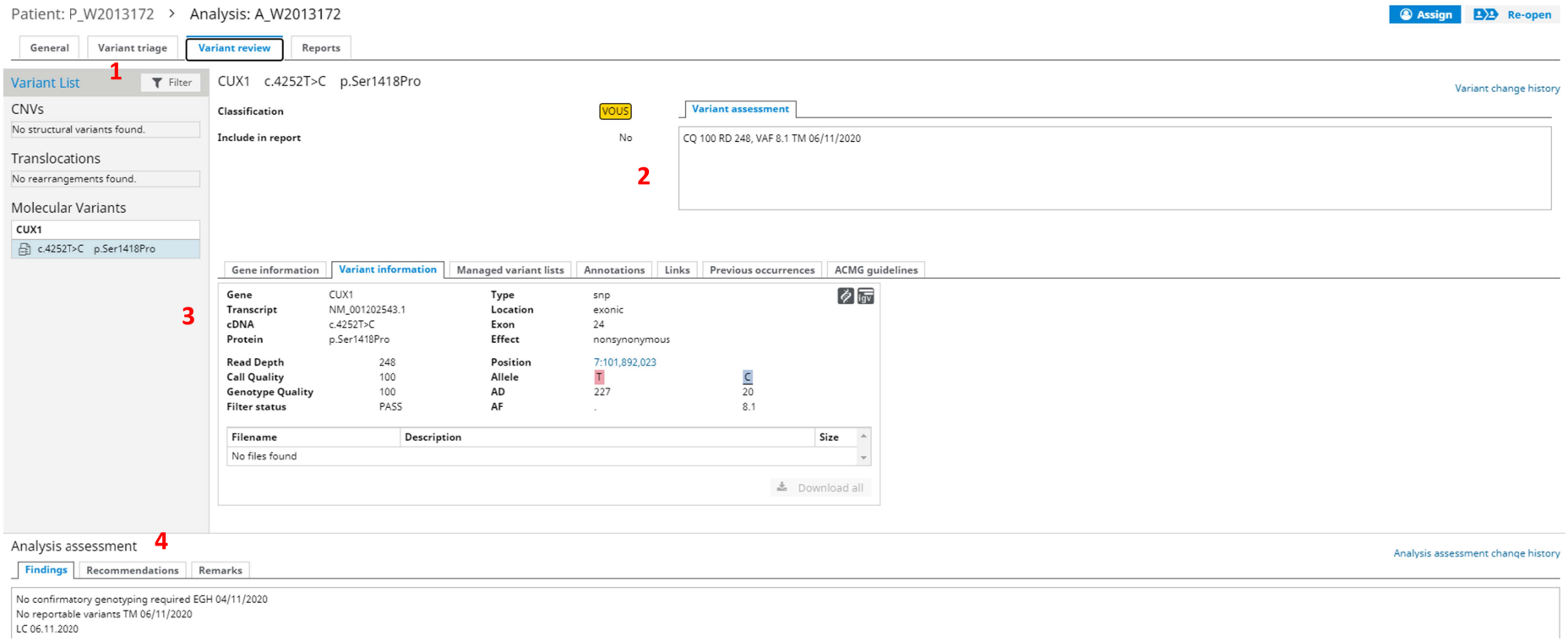


Figure 7.5 Example of the audit trail kept within Alissa Interpret. 1. Variants remaining after variant filtering has occurred are displayed here. 2. Variant assessment is displayed here in an editable box. 3. Variant information is displayed here. 4. Analysis assessment comments are recorded in this box by the analyst(s).

7.3 Appendix 3 – The TSMP validation

# Validation of diagnostic procedures	Wessex Regional Genetics Laboratory Salisbury NHS Foundation Trust Salisbury, SP2 8BJ

1. Validation details

Validation/verification number	**219**	**Date**	02/08/2019 (Updated 04/06/2020)
Procedure	Validation of Illumina Trusight myeloid sequencing panel		

1.1 Test details

Intended use or application	The Illumina TruSight Myeloid sequencing panel (P/N FC-130-1010) is an amplicon based capture kit for preparation of targeted libraries for next generation sequencing on Illumina platforms. This laboratory intends to use this kit for delivering mutational profiling of patients with clinical presentation which fall largely into (but are not exclusive to) the following categories in order to improve diagnostic accuracy: persistent cytopenia, suspected MDS, persistent unexplained eosinophilia, triple negative MPN and MDS/MPN. In addition, this panel can improve risk stratification and prognostication in cases with a confirmed diagnosis of MDS, myelofibrosis, systemic mastocytosis and aplastic anaemia. The method is intended to be used for DNA from peripheral blood or bone marrow aspirate samples. The laboratory has been using this test for the above application for research purposes since October 2016.
Locus / Gene / Marker	The capture covers 15 full genes plus exonic hotspots of an additional 39 genes known to be tumour suppressor genes or oncogenic hotspots associated with myeloid

	malignancy (further details appendix 1).
Reference Sequence	Reference sequences used at the time of writing this document are detailed in appendix 1.
Outline methodology	The TruSight panel leverages Illumina's TruSeq Custom Amplicon (TSCA) assay. The method involves a number of steps outlined below: 1. Hybridization of olionucleotide pool to unfragmented genomic DNA, followed by extension and ligation to form DNA templates consisting of the regions of interest flanked by universal primer sequences 2. Indexed primers PCR amplify the DNA templates 3. AMPure XP beads are used to purify the PCR generated libraries from the other reaction components 4. Library normalization 5. Product pooling of uniquely tagged amplicon library ready for cluster generation and sequencing 6. Library QC 7. Sequencing (Illumina MiSeq / NextSeq) 8. De-multiplexing (Miseq Reporter; v.3.5.1) 9. Read alignment (Miseq Reporter; v.3.5.1 utilising SmithWaterman aligner) 10. Variant calling (Miseq Reporter; Somatic Variant Caller v.3.5.2.1) including quality score assignment 11. Variant filtering (Cartagenia) 12. Variant annotation (Manual) **The Myeloid panel analysis pipeline is described in the WRGL DOC 033456 Oncology Myeloid panel- Analysis pipeline description.**
SOP	SOP 033309 Oncology myeloid panel – Trusight Myeloid Panel Sequencing (technical SOP) SOP 033458 Oncology myeloid panel – Post run QC analysis workflow SOP 033310 Oncology myeloid panel – Reporting of TruSight myeloid panel results (uploading vcf into Cartagenia, Starlims workflow and reporting writing) SOP 02506 Generating an NGS coverage report (panel coverage, myeloid panel coverage)
References	**Files associated with this validation/verification can be found at:** W:\Quality Management\Validation records\Records\Records #201 - #250\#219 Validation -ACTIVE

1.2 Validation details

Overall Aims	The aim of this validation is to estimate quality performance characteristics (defined below) to assess the appropriateness of application of this panel within a diagnostic setting.
Requirements	In the absence of specific national guidelines for performance characteristics: • the validation should meet or better UKGTN recommended overall sensitivity of 95% (95% CI) and specificity of 95% (95% CI) or better. • a lower limit of detection (LoD) of at least 5% is desirable.
Considerations relevant to COSHH	Reference SOP 033309 Oncology myeloid panel – Trusight Myeloid Panel Sequencing (technical SOP) for full details.
Scope / limitations	The test will be applicable to DNA extracted from peripheral blood and bone marrow. It expressly excludes detection in samples extracted from FFPE material. Samples are QC measured to ensure at least 100ng of DNA must be available for testing. Samples are also QC measured by Bioanalyzer prior to sequencing (samples at this stage have been PCR amplified, labelled with index primers and had a bead clean-up). Each sample must have a Bioanalyser reading of >50 FU to continue to sequencing. This methodology will be validated for the detection of single nucleotide variations and small indels - maximum theoretical detection size 60 – indels larger than this are out of scope. Internal tandem duplications are also out of scope.
Type	**Qualitative type D** – see Mattocks *et al* (2010) A standardized framework for the validation and verification of clinical molecular genetic tests. Eur J Hum Genet. 18, 1276-88
Turn around time	21 calendar days when in scope
Other considerations	

2. Validation of utility

Applicability of measurements	This methodology will be delivered as an adjunct to existing testing provided by WRGL for patients with suspect myeloid neoplasia in the absence of a confirmed diagnosis and will also be offered for patients with a confirmed diagnosis to provide additional prognostic information.
Selectivity	For this test, the secondary analysis on base calls and quality scores generated during the sequencing run is performed using MiSeq Reporter software v2.5.1 provided by Illumina. Version control of the software used for the analysis will be detailed in DOC 033457 Oncology Myeloid panel analysis software list Selectivity is largely a function of read mapping; the MinQScore is an adjustable setting within MiSeq Reporter that specifies the minimum base call Q-score to use as input to variant calling, giving confidence that the variant is genuine. The VariantMinimumQualCutoff setting has been adjusted to 20 but variants will only be considered reportable if they have a Q-score of 100 or they have been confirmed by another method.
Interferences	None noted
Cross-reactivity	Samples are identified via unique index primers, one appended to each end during library preparation. It is critical to ensure that correct ID tags are added to samples, and that no cross- contamination of sample material or ID primers occurs before this stage. These stages will be witness checked in order to reduce the risk of this happening. The assay was tested using a range of sample types (experimental detail provided in Sections 3.1 and 3.2, below). A number of reference samples were tested during this validation, including 32 DNA samples previously tested for specific genes(*JAK2* and *TP53*) by a validated in-house methodology, namely the genotyping pipeline. Analysis of results from these samples identified one false positive (FP) result to the estimated limit of detection of this assay (ref. Section 3.2). The sample (W1617187) had a known *TP53* variant detected by the genotyping pipeline and the myeloid panel detected an additional variant in *TP53* at 10% VAF; however, assessment of the results from the myeloid panel showed the presence of multiple low level variants after variant filtering within Cartagenia had occurred (n=17); there was also an additional 13 low level variants (<5% VAF) detected in *TP53* that were filtered out by the variant filtering pipeline but were present with variable read depth and call quality. Taken together, this result indicates that this sample either had DNA that was sub-optimal quality resulting in a high number of false-positive [artefact] variants and/or that cross-contamination had occurred. This frequency of low level variants did not occur in any of the other control samples during this validation; this result indicates that

	diagnostic samples should be considered as a fail when an excess of low level variants are detected and the potential for cross-contamination should be investigated. **<u>Update 04/06/2020:</u>** A QC tool has subsequently been implemented into the diagnostic pipeline that is designed to assess the number of low level variants in each sample and is detailed in Appendix 4. One of the main purposes of this tool is to highlight cross-reactivity occurring through cross-contamination introduced during the workflow.

Authorisation	
Authorised by	Chris Mattocks
Grade	Head of technical services
Date	11/06/2020
Comments	

3.1 Validation[§] for Sensitivity, specificity and accuracy

3.1.1 Work plan

<table>
<tr><td>Section aims</td><td>To estimate:

Analytical Sensitivity defined as: Se = TP/(TP+FN)

Analytical Specificity defined as: Sp = TN/(TN+FP)

Where: TP=True positive, FP=False positive, TN=True negative, FN=False negative

Required standard for all parameters = 95% (95%CI)

All these parameters are considered critical.</td></tr>
<tr><td>Samples</td><td>(a) A total of 33 samples from DNA from a range of sources (table 1) were used for this element of the validation; Positive controls: 14 variants in JAK2 [c.1849G>T p.(Val617Phe)] in 14 samples (VAF 2-96%) and 25 variants in TP53 in 11 samples (VAF 9-99%) defined by the validated WRGL genotyping pipeline; all variants were single nucleotide variants resulting in a missense amino acid change; total number of variants, n=39.
(b) Negative controls: defined as all nucleotides sequenced as wild-type/reference in both the positive control samples detailed above, plus an additional 7 samples with wild-type JAK2 exon 14 chosen to act as negative controls sequenced by the WRGL genotyping pipeline. Total number of samples, n=32.
(c) The Horizon diagnostics myeloid Reference Standard (Beta material): this reference material contains 22 validated variants across 19 genes, with a VAF of 5-70%; however, only 18 variants are detectable by this panel based on genomic regions covered by this assay (VAF of 5-70%).

<table>
<tr><th>Sample</th><th>Container</th><th>Number of samples</th></tr>
<tr><td>Bone marrow</td><td>Cytogenetic medium</td><td>12</td></tr>
<tr><td>Bone marrow</td><td>Lithium Heparin</td><td>1</td></tr>
<tr><td>External DNA</td><td>Eppendorf</td><td>3</td></tr>
<tr><td>Peripheral blood</td><td>EDTA</td><td>17</td></tr>
</table>
Table1 Sources of material used for testing</td></tr>
</table>

Methodology

Positive and negative controls from sample group (a) and (b) were sequenced using the WRGL genotyping pipeline and the Illumina Trusight myeloid sequencing panel. Both techniques involve sequencing on the Illumina Miseq but the library preparation for the genotyping involves subjecting input DNA to a novel PCR methodology that appends all the functional sequences necessary for sequencing on a MiSeq instrument (sequencing primer binding, sample ID indexes and flow cell hybridisation adaptors). The WRGL genotyping pipeline has been validated (W:\Quality Management\Validation records\Records\Records #101 - #150\#110 Validation – ACTIVE) and accredited by UKAS. These data were used for sensitivity and specificity calculations.

As mentioned above, some samples (group b) were analysed for more than one region of interest (ROI) and thus acted as both positive and negative controls. The target length in base pairs (bp) for the relevant amplicons sequenced by the WRGL genotyping pipeline are shown in table 2; total number of bp analysed by both WRGL genotyping and the Illumina Trusight myeloid sequencing panel, n=23616. These data were used for sensitivity and specificity calculations.

	Amplicons analyzed by the WRGL genotyping pipeline										
	JAK2 exon 14	**TP53 exon 2**	**TP53 exon 3**	**TP53 exon 4a**	**TP53 exon 4b**	**TP53 exon 5**	**TP53 exon 6**	**TP53 exon 7**	**TP53 exon 8**	**TP53 exon 9**	**TP53 exon 10**
Target length of amplicons (bp)	83	172	75	226	196	257	206	193	244	155	179
Number of samples	30	11	11	11	11	11	11	12	11	11	11
Number of bp analysed	2490	1892	825	2486	2156	2827	2266	2316	2684	1705	1969

Total number of nucleotides analyzed	23616

Table 2

All results from the Trusight myeloid panel for the regions of interest were compared to results from the WRGL genotyping pipeline as the 'gold standard' (GS).

Please note that variants detected by the myeloid panel at <5% VAF and/or <100x were excluded from the specificity calculations as the standard cut off for variant calling will be 0.05 VAF and 100x (ref. sections 3.2 and 3.4).

Horizon diagnostics myeloid Reference Sample: this sample was tested 3 times across 2 independent runs. The myeloid panel sequencing results from 2 independent analyses of this external DNA sample were assessed for concordance with the expected (externally validated) variants. As the analyses were performed in independent runs, we have considered the same variants detected in each replicate as individual data points. These data were used for sensitivity calculations only.

TP and TN were defined as concordance between the current analysis and either the GS or externally validated mutations.

FN were defined as any variants present in the GS or reported as externally validated mutations but not detected in current analysis

FP were defined as any variants present in the current analysis but not detected in GS or reported as externally validated mutations.

Sensitivity and specificity calculations were performed using Medcalc® online statistical tool (https://www.medcalc.org/calc/diagnostic_test.php)

Authorisation	
Performed by	Sophie Laird
Grade	Principal Clinical Scientist
Date	02/08/2020
Authorised by	Chris Mattocks
Grade	Head of technical services
Date	11/06/2020

3.1.2 Partial results and conclusions

Experimental results

Files associated with this validation/verification can be found at:

W:\Quality Management\Validation records\Records\Records #201 - #250\#219 Validation –ACTIVE

Full results for sample group (a) and (b) are provided in appendix 2. A summary of FP and TN used to calculate specificity of this test are provided below (Table 3).

Myeloid panel results for sample group a and b (read depth ≥100x)	
FP	1*
TN	23576

Table 3

*One false positive (FP) result was detected; please see the above section (section 2) for further details.

Results from the Horizon diagnostics myeloid Reference Sample analyses: all 18 validated mutations in the Horizon diagnostic myeloid Reference sample considered to be detectable by this assay based on genomic regions covered were detected in each analysis. No false negative results were detected.

A summary of the TP and TN used to determine sensitivity for this assay, which included results from sample group (a), (b) and (c), are shown in Table 4.

Myeloid panel results for sample group a, b and c (read depth ≥100x)	
TP	75
FN	0

Table 4 True positives were detected within the (a) and (b) sample groups (n=39) and the horizon sample analyses (n=36).

Based on these data, the specificity and sensitivity of this assay is estimated to be as

	follows: **Specificity (read depth ≥100): 99% to 100% (95% CI)** **Sensitivity (read depth ≥100): 95% to 100% (95% CI)**
Interpretation	Overall estimated sensitivity and specificity meet the required standard for blood and bone marrow samples.
Outcome / limitations	The sensitivity/specificity estimate applies only to missense variants.

Authorisation	
Performed by	Sophie Laird
Grade	Principal Clinical Scientist
Date	02/08/2019
Authorised by	Chris Mattocks
Grade	Head of technical services
Date	11/06/2020

3.2 Validation** for level of detection

3.2.1 Work plan

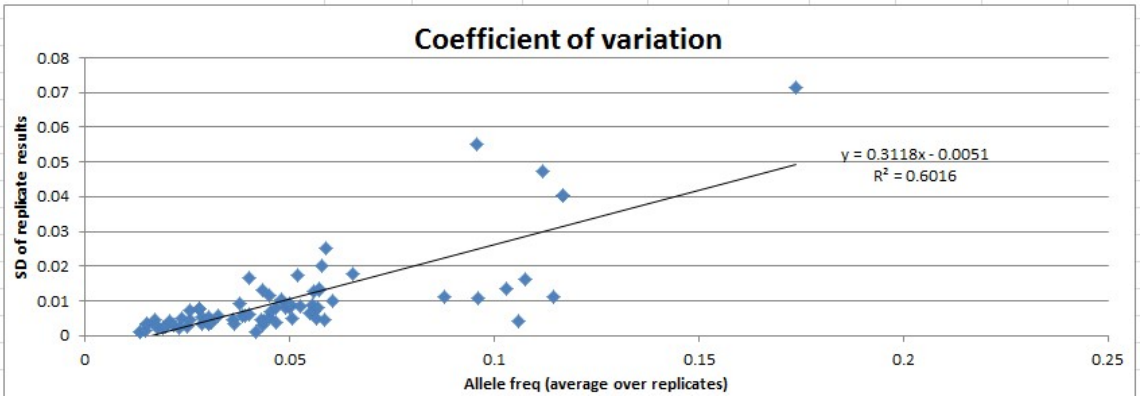
Section aims	To estimate the limit of detection (LoD) for low abundance alleles
Samples	(i)The Horizon diagnostics myeloid Reference Standard (Beta material): this reference material contains 22 validated variants across 19 genes, with a variant allele frequency of 5-70%; however, only 18 variants are detectable by this panel based on genomic regions covered by this assay. This sample was tested 3 times across 2 independent runs. (ii) One anonymised patient sequenced 10 times across 8 runs

Methodology	To estimate the LoD, defined as the lowest level of variant that can be distinguished above "background noise", filtered* sequencing results from (i) and (ii) were combined and the relationship between the allele frequency measured and the standard deviation (SD) of replicate measurements (n=278 measurements) was measured and modelled to a straight line regression and the formula was used to represent σ_s (standard deviation of replicate positive control measurements) to calculate the LoD in excel using the goal seek function. Limit of blank (LoB) was not calculated for this technology, therefore the LoB was assumed to be 1% which we predict is an overestimate of the actual value. *Variant filtering: • Only variants with a record on one of the following databases were considered to remove possible artefacts from analysis: dbSNP, 1000 genome project, HGMD, ClinVar, COSMIC, EXAC (European, Non-Finnish population). • Only exonic variants and variants +/-2bp were considered for analysis as they affect "clinically relevant" nucleotides. • (i): only variants detected in all three replicates were considered • (ii): only variants detected in at least 6 out of 10 replicates were considered based on an analysis of frequency (see W:\Quality Management\Validation records\Records\Records #201 - #250\#219 Validation –ACTIVE\LoD calcuations) Only variants <0.2 VAF were considered.

Authorisation	
Performed by	Sophie Laird
Grade	Principal Clinical Scientist
Date	02/08/2019
Authorised by	Chris Mattocks
Grade	Head of technical services
Date	11/06/2020

3.2.2 Partial results and conclusions

Experimental results	Files associated with this validation/verification can be found at: W:\Quality Management\Validation records\Records\Records #201 - #250\#219 Validation –

	ACTIVE\LoD calculations.xls **LoD was estimated to be 0.025-0.028 (95% CI) when the LoB is assumed to be 0.01.**  **σ_s = 0.3118[variant frequency] + 0.0051** **Please note:** The target value for the LoB at 1SD is very low due to apparent negative bias in the system i.e. test value is lower than it should really be. However, the value for 1SD is not used and the value of the target at 2SD is judged to be more realistic. The cause for this negative bias is unknown but please note that whilst extensive quantitation experiments have not been undertaken for this technique, a number of positive control measurements have shown high level concordance when compared to expected VAF from validated techniques or with commercially available reference material (ref. section 3.1).

Allele freq.	σ_s	Target - z*σ_s	LoB (Gausian)	95%CI -	95%CI +
0.010000	-0.001982	0.013885	0.010000	0.010869	0.009131
0.026700	0.003225	0.020379	0.020000	0.025287	0.028113

Interpretation	LoD was estimated to be 0.025-0.028 (95% CI) when the LoB is assumed to be 0.01. Variants detected at all allele frequency below this interval may represent technical artefacts and therefore these variants should not be reported unless confirmed by another method.
Outcome / limitations	The standard cut off for variant calling will be 0.05 VAF. As the LoD for the assay has been estimate to be 0.025-0.028 VAF, variants detected <0.05 VAF can be considered for analysis if they occur in clinically relevant genes but should be confirmed by an independent method where

	possible.

Authorisation	
Performed by	Sophie Laird
Grade	Principal Clinical Scientist
Date	02/08/2019
Authorised by	Chris Mattocks
Grade	Head of technical services
Date	11/06/2020

3.3 Validation for measurement of uncertainty

3.3.1 Work plan

Section aims	There are no guidelines specifying what accuracy is clinically acceptable with regard to quantitative results in this context. In general, detection/non-detection of clonality is the critical parameter, but quantitative data provides additional information, e.g. higher mutation burdens (i.e. >50% mutant allele burden) of *JAK2* V617F variants are associated with more symptomatic disease in large cohort studies. As the clinical requirement for accuracy is limited, we estimated the variation for allele frequency (VAF) detection so that this can be included in the clinical report.
Samples	We used four datasets to evaluate the quantitative performance: (i) The Horizon diagnostics myeloid Reference Standard (Beta material): this reference material contains 22 validated variants across 19 genes, with a variant allele frequency of 5-70%. Variants included SNV, deletions and duplications. This sample was tested 3 times across 2 independent runs. (ii) One anonymised patient sequenced 10 times across 8 runs. As the true genotype of this patient is unknown, only benign/likely benign polymorphisms (n=15; expected allele frequency 0.5 or 1.0) and one known pathogenic mutation in SF3B1 were included in this analysis (the mean allele frequency across all replicates [0.27] was used as the surrogate expected allele frequency for this variant).

	(iii) The Coriell reference cell lines (NA19240) was sequenced in the following dilution series (diluted with NA12878): 100%, 99%, 90%, 75%, 50%, 25%, 10% 1% and 0%. The results were compared to the reported variants by GetRM project that had been confirmed by 2 or more independent studies (https://www.ncbi.nlm.nih.gov/variation/tools/get-rm/). (iv) Positive controls: 14 variants in *JAK2* [c.1849G>T p.(Val617Phe)] in 14 samples (VAF 2-96%) and 25 variants in *TP53* in 11 samples (VAF 9-99%) defined by the validated WRGL genotyping pipeline; all variants were single nucleotide variants resulting in a missense amino acid change; total number of variants, n=39.
Methodology	1. The relative closeness of quantitative measurements of VAF by the myeloid panel to the actual values was estimated by determining the coefficient of variation (CV) for replicates of variants at a range of allele frequencies from 0.05-1.0. For the purpose of this estimate, variants that were expected to have the same allele frequency were considered replicates irrespective of whether they were true replicates of the same variant or measurements from unique variants with the same expected allele frequency*. The CV was calculated for variants with mutation burdens of 0.05-0.1 VAF and variants >0.11 VAF. *The number of unique variants in the replicate data, n=154 Total number of replicates, n= 358 2. R- squared regression was calculated for TP53 and JAK2 in-house positive controls (iv) and the Coriell reference cell line (iii).
Authorisation	
Performed by	Sophie Laird
Grade	Principal Clinical Scientist
Date	02/08/2019
Authorised by	Chris Mattocks

Grade	Head of technical services
Date	11/06/2020

3.3.2 Partial results and conclusions

Experimental results

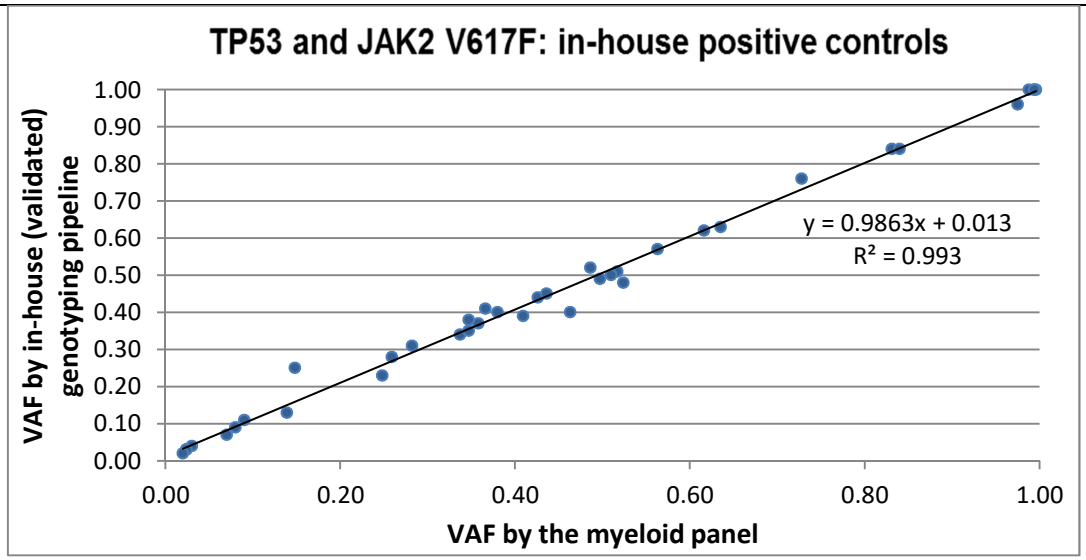
1. **CV for replicate analyses**

Intervals for analysis		Number of unique variants	Total number of replicates	Mean CV
Lower limit VAF	Upper Limit VAF			
0.05	0.1	34	63	12.36%
0.11	1	65	80	6.55%

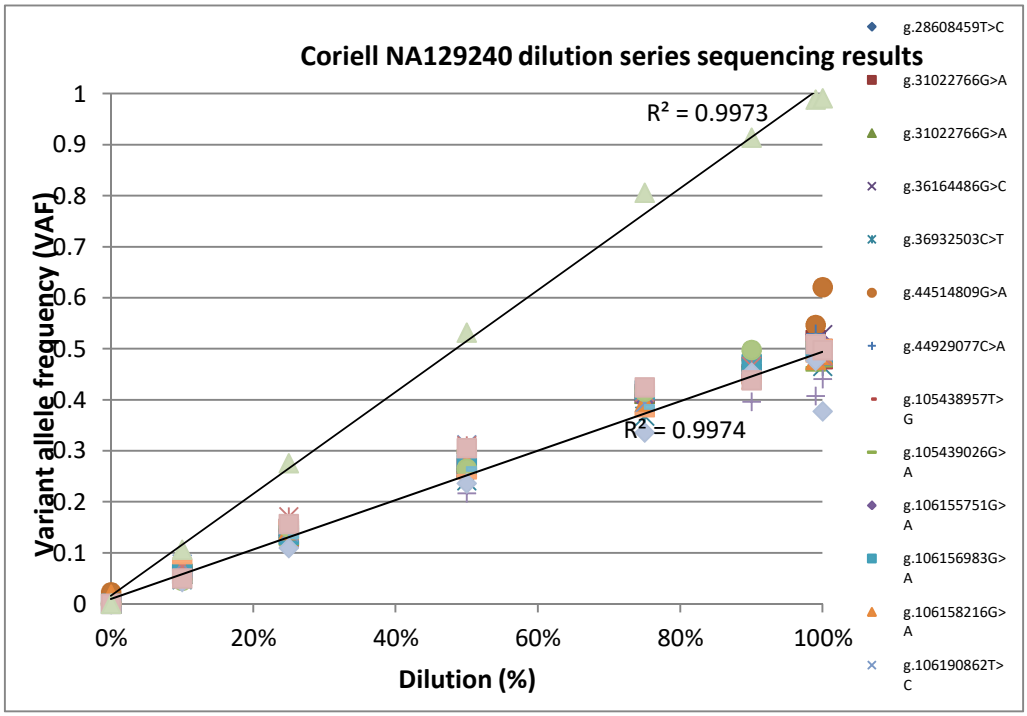
Files associated with this validation/verification can be found at:

W:\Quality Management\Validation records\Records\Records #201 - #250\#219 Validation –ACTIVE\MoU calculations.xls

2.



Files associated with this validation/verification can be found at: W:\Quality Management\Validation records\Records\Records #201 - #250\#219 Validation – ACTIVE\Final_genotyping comparison.xls



Interpretation	The estimated variation in quantitative assessment of VAF meets an acceptable standard. However, CV calculations indicate that variants of low allelic burn (0.05-0.1 VAF) show a greater dispersion of quantitative measurements than variants of higher allelic burden (0.11-1.0 VAF) by this assay.

Outcome / limitations	Estimated variation in allelic quantification will be stated on the report for low allelic burden (0.05-0.1 VAF) and higher allelic burden (0.11-1.0 VAF).
Authorisation	
Performed by	Sophie Laird
Grade	Principal Clinical Scientist
Date	02/08/2019
Authorised by	Chris Mattocks
Grade	Head of technical services
Date	11/06/2020

3.4 Validation for horizontal coverage

3.4.1 Work plan

Section aims	1. To determine the minimum coverage (read depth) required for an amplicon. 2. To determine the proportion of amplicons consistently covered to the minimum read depth. 3. Identify any amplicons that are consistently not covered to the minimum read depth. 4. To determine consistency of coverage for each ROI (comprised of ≥1 amplicon). This parameter is considered critical; any ROI that consistently fails to obtain >100x will be considered “not covered” by the myeloid panel as we do not intend to “gap fill” by alternative methodologies on a routine basis.
Samples	1. One anonymised patient sequenced 10 times across 8 runs and the Horizon diagnostics myeloid Reference Standard (Beta material): this reference material contains 22 validated variants across 19 genes, with a variant allele frequency of 5-70%. Variants included SNV, deletions and duplications. This sample was tested 3 times across 2 independent runs. Both of these samples were used to determine the LoD for this assay (0.025-0.028 VAF, ref. section 3.2.2).

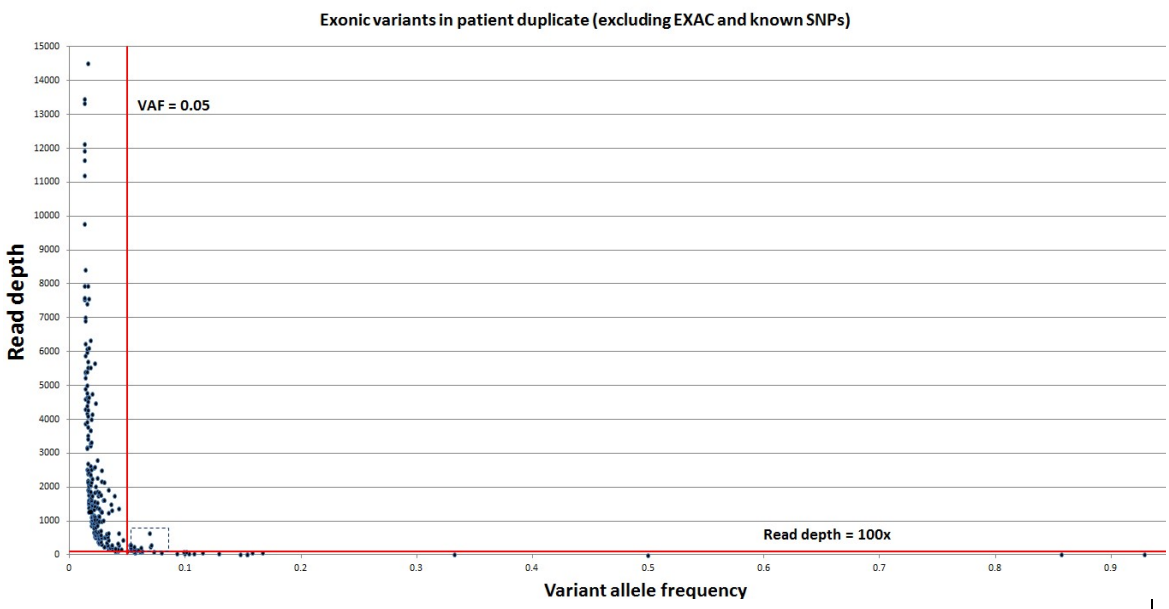
	2. Compiled coverage data from all patients that have been tested using this assay for research purposes across 62 consecutive (independent) runs (total number of samples, n=916). 3. Compiled coverage data from all patients that have been tested using this assay for research purposes across 48 consecutive (independent) runs (total number of samples, n=768). 4. ROI % coverage data was compiled from coverage data from patient samples (n=203) that have been tested using this assay for research purposes since October 2016.
Methodology	1. Data was compiled for each replicate from each dataset, respectively. Variants filtered to remove ExAC reported variants >0.1 frequency and known SNPs (heterozygous/homozygous variants classified as likely benign/benign on ClinVar); the read depth and allele frequency was plotted for each variant. 2-3. Compiled coverage data from all patients that have been tested using this assay for research purposes across consecutive (independent) runs was analysed to determine the proportion of amplicons that obtained coverage of at least 100x per sample tested. This data was also analysed to identify those amplicons that are consistently not covered to the minimum read depth and those amplicons that are more likely to not be covered to the minimum read depth. 4. ROI % coverage data was compiled from coverage data from 203 patient samples that have been tested using this assay for research purposes since October 2016.
Authorisation	
Performed by	Sophie Laird
Grade	Principal Clinical Scientist
Date	02/08/2019
Authorised by	Chris Mattocks
Grade	Head of technical services
Date	11/06/2020

3.4.2 Partial results and conclusions

Experimental results

1. Figure 1

(A)



(B).

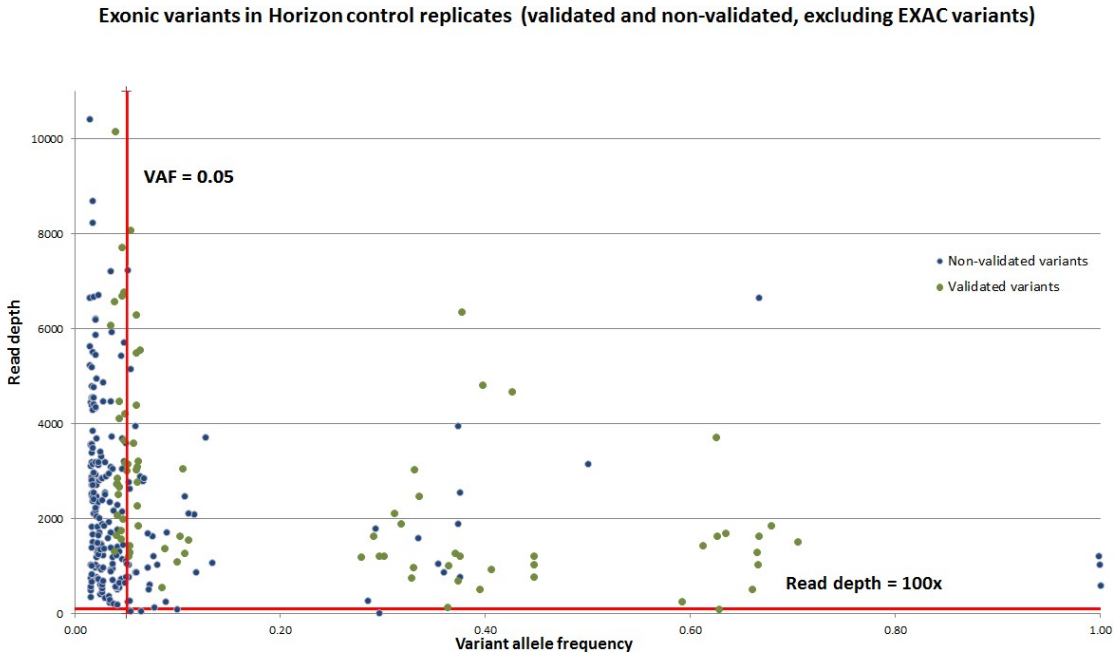


Figure 1 Variants <0.05 VAF have been indicated as this was the standard VAF cut-off decided for filtering variants detected based upon the LoD for this assay. (A) Exonic variants in one anonymised patient sequenced 10 times across 8 runs when EXAC variants with a frequency >0.1/known SNPs are filtered out (n=286). Seventeen variants (indicated by a dashed box) had a read depth >100 and a VAF of 0.05-0.07; however, the call quality for these variants was suboptimal (<100 CQ) therefore it was unclear whether they represented true variants or artefacts. (B) Exonic variants in the Horizon diagnostics myeloid Reference Standard (Beta material) sequenced 3 times when EXAC variants with a frequency >0.1/known SNPs are filtered out (n=299). Validated (green) and non-validated (blue) variants are shown.

The lowest read-depth for validated variants in sample (B) was 114x. For sample (A) a number of variants that may have been artefact calls (n=17) indicate that the minimum depth required could be increased above had a read depth >100x (n=17) but this was not replicated in the Horizon control dataset. **Based on these data, we consider a minimum read depth of 100x is required for amplicon coverage for variants >0.05 VAF.**

2.

Figure 2

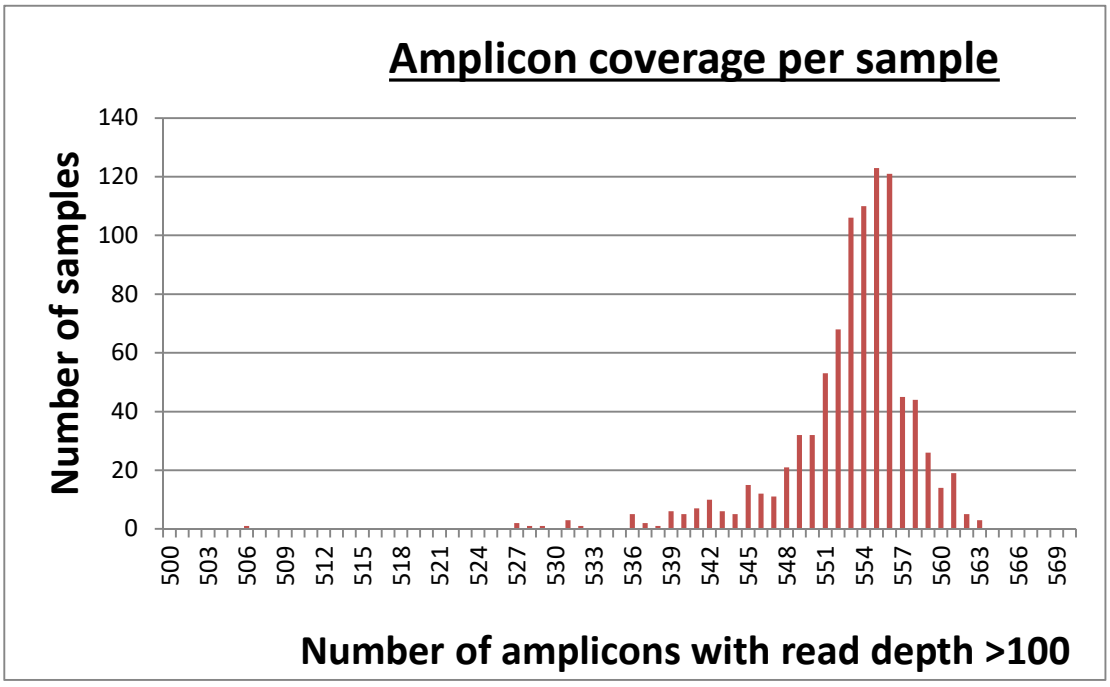


Figure 2 Number of amplicons with read depth >100. Total amplicons tested per sample, n=573. Total number of samples tested, n=916.

97.5% of samples tested (n=906) achieved a read depth of 100x in 540 amplicons.

3.

Number of amplicons that achieved ≥100x in all samples analysed: - 439 amplicons out of 573 within this assay (**77%**).Median number of amplicons ≥100x per sample, **99.6%.**

Further information is given in **figure 3**.

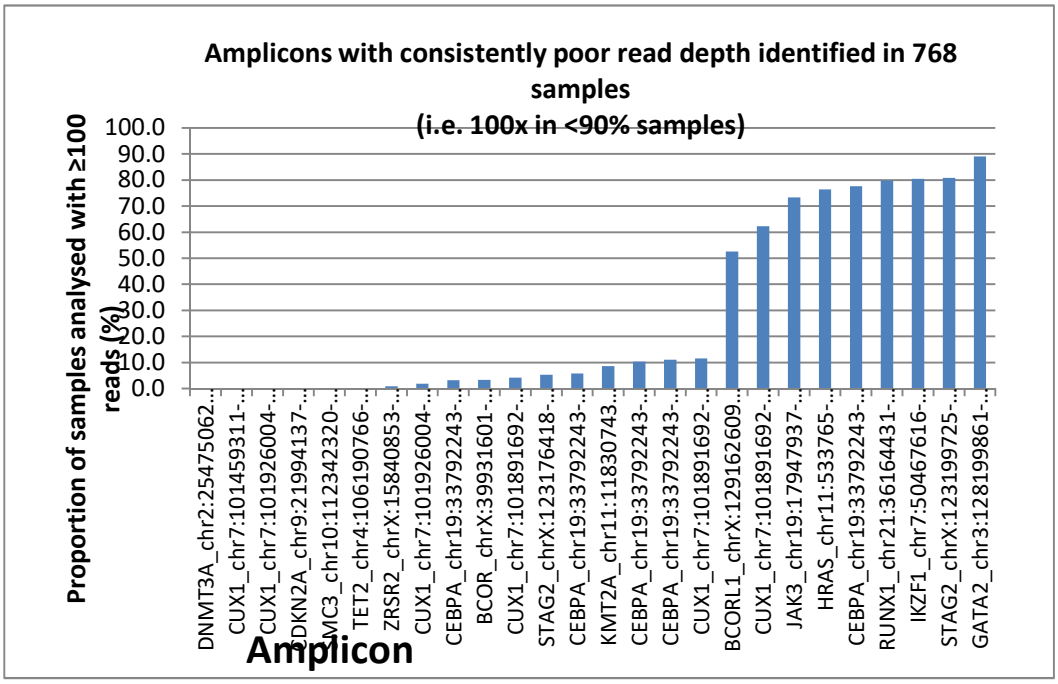


Figure 3

4. Determining the consistency of coverage for each ROI (comprised of ≥1 amplicon).

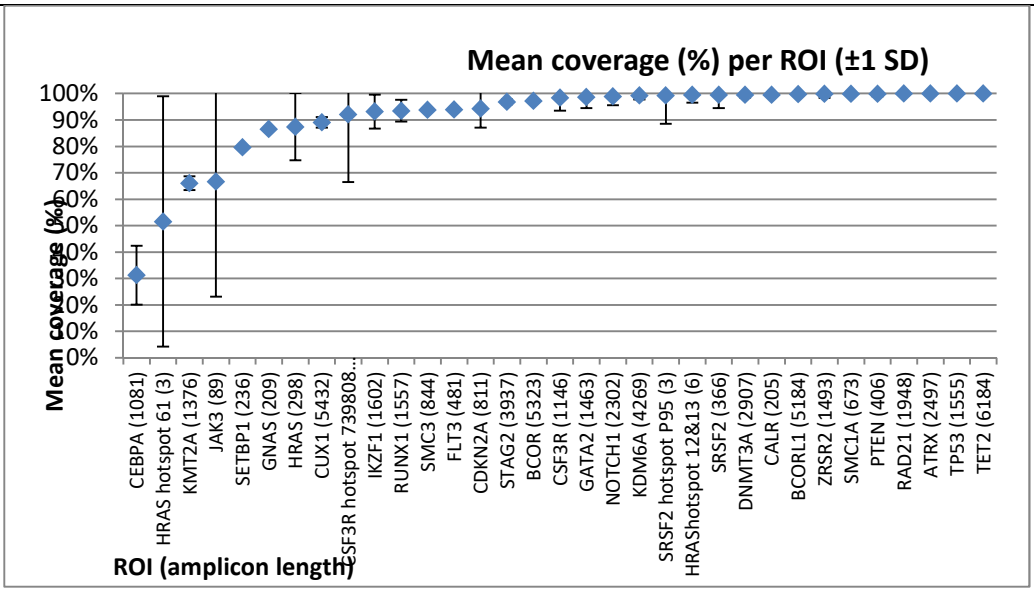


Figure 4

Mean coverage (%) per region of interest (ROI) (±1 standard deviation, SD). The remaing ROI not shown in this figure had full coverage in in all samples analysed (SD=0).

Interpretation

1. Based on these data, we consider a minimum read depth of 100x is required amplicon coverage.

2. Based on these data, we consider a minimum read depth of 100x is required in 540 out of 573 amplicons tested per sample (94.2%). For samples that do not meet this metric, the quality of the result will be considered sub-optimal, with the presence of variants within the region not excluded to the same performance criteria as defined in this document and a poor quality rider may be included in the report accordingly.

3. Based on these data, *CEBPA* will be considered as “not covered” by this panel.

4. The majority of ROI performed >90% coverage in the samples tested (n=203); a proportion of ROI (n=8 out of 79) had a mean coverage <90% with some variability seen. *CEBPA* performed particularly poorly across all cases and is therefore considered not covered by this panel; a caveat with go onto clinical reports stating that this gene is not covered by this assay.

Outcome / limitations

Consistency of coverage is given as a proportion of the region of interest but the consistency of coverage per nucleotide within each ROI was not assessed.
In lieu of gap filling, coverage per ROI will be given as a percentage on each patient report with clinically significant ROI (e.g. *ASXL1* and *RUNX1* which are associated with a poor prognosis in PMF) highlighted if their % coverage is sub-optimal. *CEBPA* gives consistently poor coverage across all regions of the gene sequenced by the myeloid panel and therefore is considered not covered by the panel.

Authorisation

Performed by	Sophie Laird
Grade	Principal Clinical Scientist
Date	02/08/2019
Authorised by	Chris Mattocks
Grade	Head of technical services
Date	11/06/2020

Supplementary work:

- Variants may be detected in genes within which germline variants have been reported. In order to indicate an acceptable range for allele frequency that may elicit further investigation into its origin (somatic/germline), the allele frequency of known heterozygous SNPs was investigated. **The allele frequency range was estimated to be 0.44-0.57 for heterozygous variants (further details are provided in appendix 3).**

4. Validation / Verification[††] final conclusions

Overall Conclusion	This methodology has been shown to perform to the required level of accuracy or above and is therefore validated for use in WRGL for use in mutation scanning for acquired variant detection subject to the limitations given below.
Estimates of accuracy and measures of uncertainty	**Specificity (read depth ≥100):** 99% to 100% (95% CI) **Sensitivity (read depth ≥100):** 95% to 100% (95% CI) **Limit of Detection:** estimate to be 0.025-0.028 VAF
Limitations and/or predictable interferences	The sensitivity/specificity estimate applies only to missense variants only. This validation applies to mutation scanning for acquired variants in samples extracted form peripheral blood or bone marrow material only. The standard cut off for variant calling will be 0.05 VAF. As the LoD for the assay has been estimated to be 0.025-0.028 VAF, variants detected with VAF <0.05 can be considered for analysis if they occur in clinically relevant genes but should be confirmed by an independent method where possible. In lieu of gap filling, coverage per region of interest (ROI) will be given as a percentage on each patient report with clinically significant ROI (e.g. *ASXL1* or *RUNX1* which are

	associated with a poor prognosis in PMF) highlighted if their coverage is sub-optimal. *CEBPA* gives consistently poor coverage across all regions of the gene sequenced by the myeloid panel and therefore is considered not covered by the panel.
Internal QC	Ongoing record of run metrics and coverage statistics.
External QA	This laboratory participates in UK NEQAS AML gene panel (pilot) scheme for this assay. Records can be found in W:\Share\EQA\NEQAS Oncology\Acute myeloid leukemia gene panel (Pilot)
Authorisation	
Performed by	Sophie Laird
Grade	Principal Clinical Scientist
Date	02/08/2019
Authorised by	Chris Mattocks
Grade	Head of technical services
Date	11/06/2020

Appendix 1: Regions covered by the myeloid panel including reference sequences used during analysis at the time of writing.

Gene	Target region (exon)	Reference sequence
ABL1	4-6	NM_007313.2
ASXL1	12	NM_015338.5
ATRX	8-10, 17-31	NM_000489.4
BCOR	full gene	NM_001123385.1
BRAF	full gene	NM_004333.4
CALR	9	NM_004343.3
CBL	8, 9	NM_005188.3
CBLB	9, 10	NM_001321807.1
CBLC	9, 10	NM_012116.3
CDKN2A	full gene	NM_001195132.1
CEBPA	full gene	NM_001285829.1
CSF3R	14-17	NM_156039.3

Gene	Target region (exon)	Reference sequence
JAK2	12, 14	NM_001322194.1
JAK3	13	NM_000215.3
KDM6A	full gene	NM_001291415.1
KIT	2, 8-11, 13, 17	NM_000222.2
KRAS	2, 3	NM_033360.3
MLL	5-8	NM_001197104.1
MPL	10	NM_005373.2
MYD88	3-5	NM_001172567.1
NOTCH1	26-28, 34	NM_017617.4
NPM1	12	NM_002520.6
NRAS	2,3	NM_002524.3
PDGFRA	12, 14, 18	NM_006206.4

CUX1	full gene	NM_001202543.1
DNMT3A	full gene	NM_175629.2
ETV6	full gene	NM_001987.4
EZH2	full gene	NM_004456.4
FBXW7	9-11	NM_033632.3
FLT3	14, 15, 20	NM_004119.2
GATA1	2	NM_002049.3
GATA2	2-6	NM_001145661.1
GNAS	8, 9	NM_080425.3
HRAS	2, 3	NM_176795.4
IDH1	4	NM_001282387.1
IDH2	4	NM_002168.3
IKZF1	full gene	NM_006060.5
U2AF1	2, 6	NM_006758.2
WT1	7, 9	NM_024426.3
ZRSR2	full gene	NM_005089.3

PHF6	full gene	NM_032458.2
PTEN	5, 7	NM_001304717.2
PTPN11	3, 13	NM_002834.3
SF3B1	13-16	NM_012433.3
SMC1A	2, 11, 16, 17	NM_006306.3
SMC3	10, 13, 19, 23, 25, 28	NM_005445.3
SRSF2	1	NM_003016.4
STAG2	full gene	NM_001042749.2
TET2	3-11	NM_001127208.2
TP53	2-11	NM_000546.5
RAD21	full gene	NM_006265.2
RUNX1	full gene	NM_001754.4
SETBP1	4 (partial)	NM_015559.2

Appendix 2: Samples tested for the sensitivity/specificity experiments detailed in section 3.

			Amplicons analysed by the WRGL genotyping pipeline										
Samples tested	**Sample type**	**Sample tube**	**JAK2 exon 14**	**TP53 exon 2**	**TP53 exon 3**	**TP53 exon 4a**	**TP53 exon 4b**	**TP53 exon 5**	**TP53 exon 6**	**TP53 exon 7**	**TP53 exon 8**	**TP53 exon 9**	**TP53 exon 10**
W1507308	External DNA	DNA tube	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
W1617187	BM	Medium	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
W1617244	PB	EDTA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
W1703166	BM	LiHep	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
W1705791	PB	EDTA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

W1707255	PB	EDTA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
W1708332	PB	EDTA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
W1708605	BM	Medium	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
W1712554	PB	EDTA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
W1713670	BM	Medium	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
W1705644	PB	EDTA	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
W1708769	BM	Medium	N	N	N	N	N	N	N	Y	N	N	N
W1615664	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1616361	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1700329	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1700955	BM	Medium	Y	N	N	N	N	N	N	N	N	N	N
W1701201	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1702309	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1703720	BM	Medium	Y	N	N	N	N	N	N	N	N	N	N
W1704537	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1704817	BM	Medium	Y	N	N	N	N	N	N	N	N	N	N
W1705351	BM	Medium	Y	N	N	N	N	N	N	N	N	N	N
W1707073	BM	Medium	Y	N	N	N	N	N	N	N	N	N	N
W1707205	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1707645	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1713849	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1713998	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1714877	BM	Medium	Y	N	N	N	N	N	N	N	N	N	N
W1717332	BM	Medium	Y	N	N	N	N	N	N	N	N	N	N
W1717965	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1800595	PB	EDTA	Y	N	N	N	N	N	N	N	N	N	N
W1800726	BM	Medium	Y	N	N	N	N	N	N	N	N	N	N
		Target length of amplicons (bp)	83	172	75	226	196	257	206	193	244	155	179

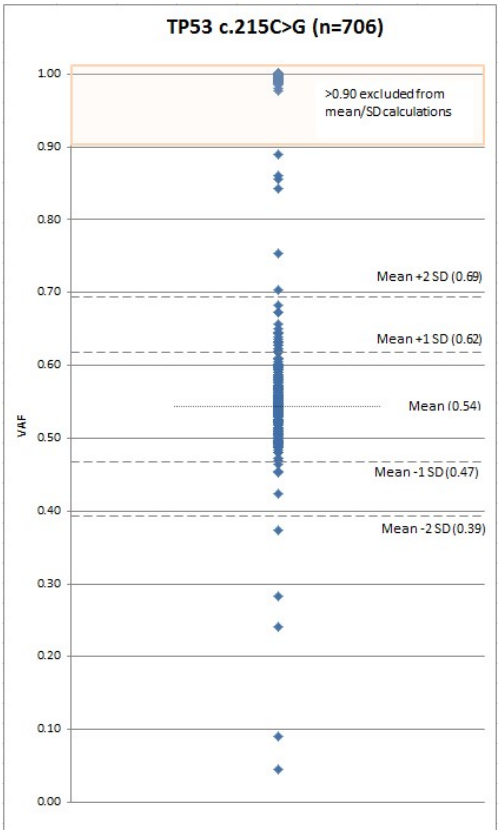
Number of samples	30	11	11	11	11	11	11	12	11	11	11
Number of bp analysed	2490	1892	825	2486	2156	2827	2266	2316	2684	1705	1969
Total analyses		23616									

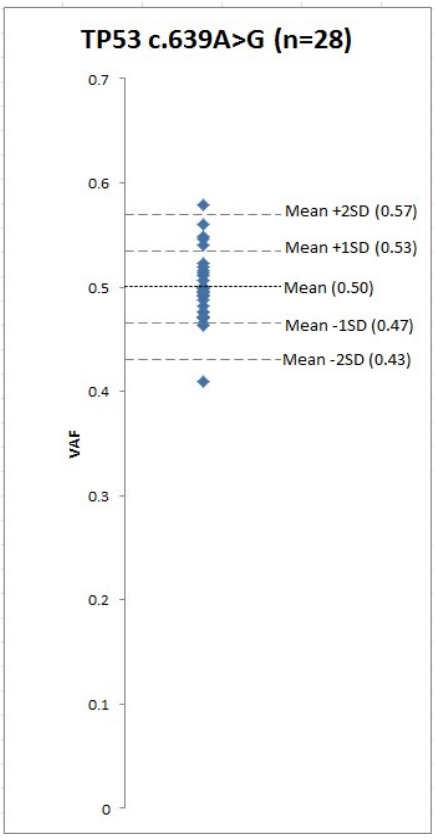
Appendix 3

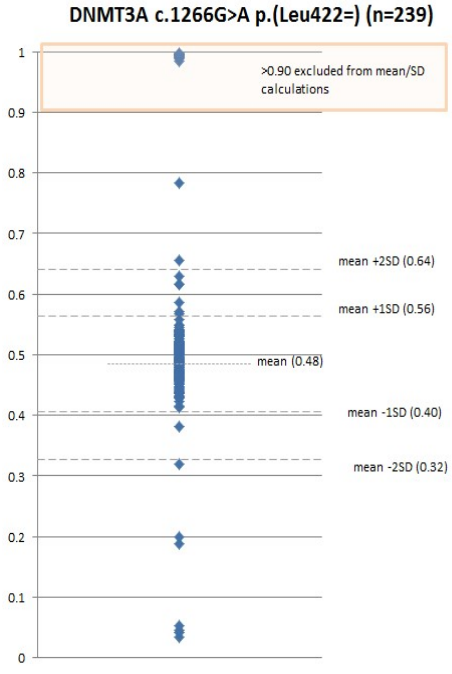
Variants may be detected in genes within which germline variants have been reported. In order to indicate an acceptable range for allele frequency that may elicit further investigation into its origin (somatic/germline), the allele frequency of known heterozygous SNPs was investigated.

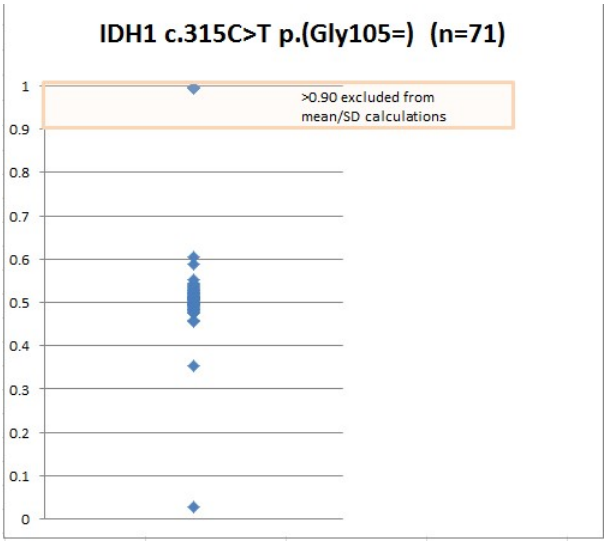
The allele frequency range was estimated to be 0.48-0.56 for heterozygous variants.

Variant	Number of samples analysed	Mean	Mean +1 SD	Mean -1SD
TP53 c.215C>G	706	0.54	0.62	0.47
TP53 c.639A>G	28	0.50	0.53	0.47
DNMT3A c.1266G>A p.(Leu422=)	239	0.48	0.56	0.40
IDH1 c.315C>T p.(Gly105=)	71	0.49	0.57	0.43
Total	1044			
Average (adjusted)		0.52	0.56	0.48


TP53 c.215C>G (n=706)
>0.90 excluded from mean/SD calculations
Mean +2 SD (0.69)
Mean +1 SD (0.62)
Mean (0.54)
Mean -1 SD (0.47)
Mean -2 SD (0.39)
VAF
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00


TP53 c.639A>G (n=28)
Mean +2SD (0.57)
Mean +1SD (0.53)
Mean (0.50)
Mean -1SD (0.47)
Mean -2SD (0.43)
VAF
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0


DNMT3A c.1266G>A p.(Leu422=) (n=239)
>0.90 excluded from mean/SD calculations
mean +2SD (0.64)
mean +1SD (0.56)
mean (0.48)
mean -1SD (0.40)
mean -2SD (0.32)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0


IDH1 c.315C>T p.(Gly105=) (n=71)
>0.90 excluded from mean/SD calculations
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Appendix 4 – added to this document 04/06/2020

Myeloid panel run low level variant QC tool: Description and use

Tool (file name): Myeloid panel RunQC template V20191125.xlsm

Located in: W:\Share\Oncology\Myeloid panel\TSMP Runs\QC-Low level variants script

Path for data: Path: W:\Share\Oncology\Myeloid panel\TSMP Runs\TSMP_[run number]_[run ID]

SOP: This process is also detailed in SOP 033458 Oncology myeloid panel- Post run QC analysis workflow.

Introduction and background

This tool was written in as one of the measures introduced following the reporting of 3 low level JAK2 variants in October 2019 that were subsequently shown to be false positives. Specific sources of contamination were not identified, but it was noted that the runs containing false positives contained unusually high numbers of low level variants compared to other runs. The tool was designed to bring together relevant global data from each run so that runs with unusual characteristics that may be indicative of poor quality data can be identified and more carefully scrutinised.

Basic functionality

The tool comprises an Excel spreadsheet with VBA modules that automate collection and visualisation of data from various source files associated with a selected run. The collected data are compiled and analysed and the results are presented in two 'output' tabs. The **Samples** tab is a tabulated format summarising read depth, the number variants in different frequency categories (<5%, 5-10% and <10%), the proportion variants passing/failing analysis filter and associated reasons for each sample. Significant statistical outliers are flagged in this sheet (red= high, purple = low). The **Run metrics** tab displays normalised median values for the run as a whole for the same metrics and compared to box and whisker plots representing a set of 9 runs that are known to have generated high quality results (data stored at W:\Share\Oncology\Myeloid panel\TSMP Runs\QC-Low level variants script\ Myeloid panel QC Baseline stats.xlsm). This tab is a graphical representation of the data calculated in the **Stats** tab.

RUN	SAMPLE	I5	I7	READS	<5%	5-10%	<10%	PASS	SB	LowDP	R8	SB;R8	ALL
TSMP Runs\TSMP_189_J3JFT	W1807875	A501	A705	889525	103	12	115	187	24	4	84	1	300
TSMP Runs\TSMP_189_J3JFT	W1601747	A502	A705	888956	98	11	109	188	16	6	84	0	294
TSMP Runs\TSMP_189_J3JFT	W2006796	A503	A705	1127274	101	9	110	186	23	5	85	0	299
TSMP Runs\TSMP_189_J3JFT	W2006795	A504	A705	1013560	92	16	108	190	19	7	87	2	305
TSMP Runs\TSMP_189_J3JFT	W2006784	A505	A705	1168815	109	12	121	200	23	4	88	0	315
TSMP Runs\TSMP_189_J3JFT	W2006797	A506	A705	1079989	115	18	133	219	30	6	77	2	334
TSMP Runs\TSMP_189_J3JFT	W2006955	A507	A705	796980	108	11	119	198	20	7	81	1	307
TSMP Runs\TSMP_189_J3JFT	W2006878	A508	A705	1518055	99	13	112	181	35	6	88	2	312
TSMP Runs\TSMP_189_J3JFT	W2006897	A501	A706	753645	89	20	109	179	20	11	78	2	290
TSMP Runs\TSMP_189_J3JFT	W2006906	A502	A706	759960	101	18	119	190	18	5	84	1	298
TSMP Runs\TSMP_189_J3JFT	W2006290	A503	A706	967632	99	16	115	182	21	6	82	0	291
TSMP Runs\TSMP_189_J3JFT	W2000256	A504	A706	1066656	104	7	111	194	27	5	85	1	312
TSMP Runs\TSMP_189_J3JFT	W2006058	A505	A706	791128	95	18	113	169	19	6	82	1	277
TSMP Runs\TSMP_189_J3JFT	W2006956	A506	A706	829957	101	31	132	220	31	0	76	2	329
TSMP Runs\TSMP_189_J3JFT	W2000321	A507	A706	717912	98	11	109	190	21	6	81	1	299
TSMP Runs\TSMP_189_J3JFT	W2003818	A508	A706	657532	81	20	101	171	17	20	79	1	288

Figure 1: Samples tab

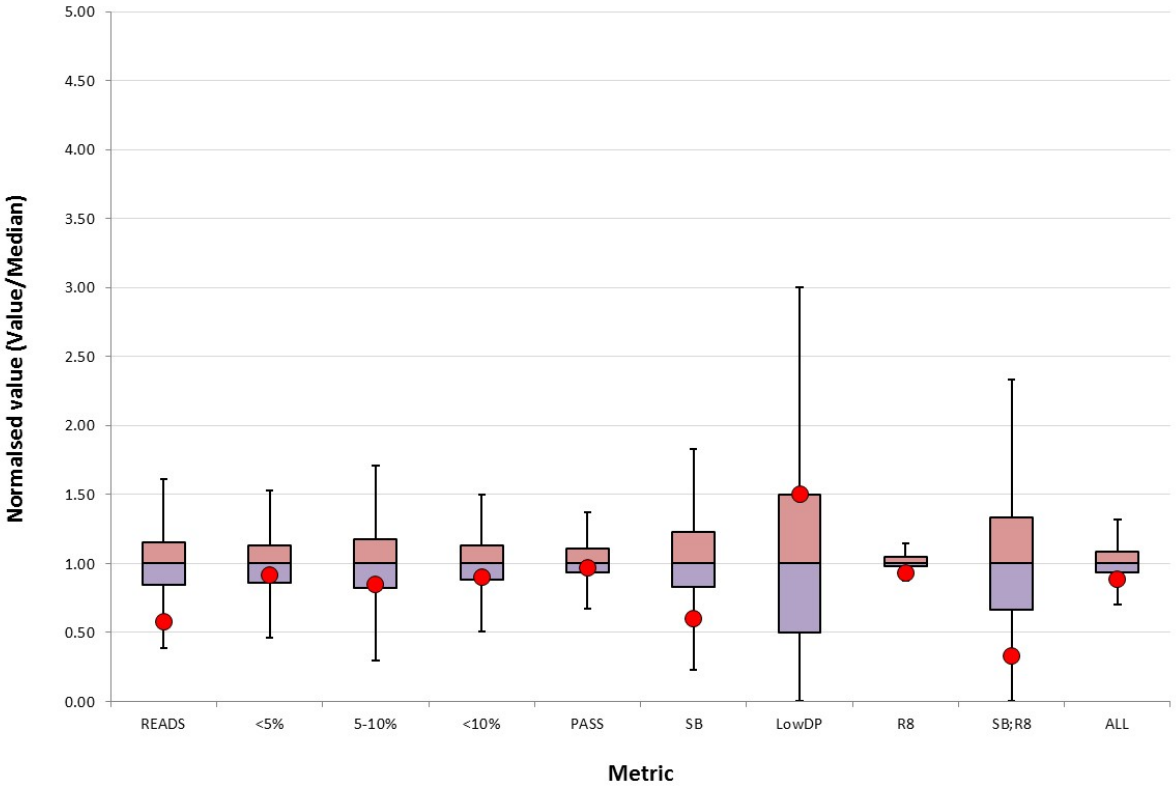


Figure 2: Run metrics tab. Red dots represent normalised median data for the current run. Box and whisker plots represent expected ranges for good quality runs.

Intended use

Data analysis is very focussed for this panel, therefore it will not necessarily be evident from the diagnostic analysis if the overall run was good quality or if a particular sample had an unusual data profile. Therefore, this tool is intended to give the user an overall picture of the quality of a given run and any samples within that run that may have performed unusually compared to what might be considered the ongoing normal. In conjunction with other lines of evidence, it may flag individual samples that merit additional scrutiny or a run that is particularly unusual in terms of performance. For

example, flagged outliers in most / all of the data categories in the **Samples** tab for one particular sample may indicate failure or poor quality data.

Other tabs (**Stats**, **Variants** and **Reads**) are primarily used to collect and analyse data for presentation, but can also be used for a more detailed analysis of the run data if required (ref. Figure 3 and Figure 4).

Further information is available in WRGL SOP 033309 Oncology myeloid panel – Trusight Myeloid Panel Sequencing (technical SOP).

RUN	SAMPLE	I5	I7	READS	<5%	5-10%	<10%	PASS	SB	LowDP	R8	SB;R8	ALL
TSMP_148_CP3F5	W1914926	A501	A705	1248890	230	25	255	334	21	1	92	1	449
TSMP_148_CP3F5	W1914925	A502	A705	823344	114	21	135	202	18	5	90	1	316
TSMP_148_CP3F5	W1915077	A503	A705	1040370	131	20	151	221	24	1	93	2	341
TSMP_148_CP3F5	W1915083	A504	A705	987432	95	15	110	173	19	6	90	3	291
TSMP_148_CP3F5	W1915080	A505	A705	1973392	106	15	121	190	24	2	92	4	312
TSMP_148_CP3F5	W1915079	A506	A705	2112972	108	15	123	193	32	4	101	3	333
TSMP_148_CP3F5	W1915338	A507	A705	1765533	101	13	114	185	38	1	100	2	326
TSMP_148_CP3F5	W1915218	A508	A705	1809657	232	26	258	341	29	2	96	1	469
TSMP_148_CP3F5	W1915097	A501	A706	1378703	128	18	146	205	27	3	91	1	327
TSMP_148_CP3F5	W1915204	A502	A706	1631470	123	11	134	192	28	2	93	1	316
TSMP_148_CP3F5	W1912671	A503	A706	1387833	137	17	154	231	29	2	94	1	357
TSMP_148_CP3F5	W1915249	A504	A706	1501789	153	19	172	234	28	1	94	3	360
TSMP_148_CP3F5	W1915007	A505	A706	1154131	252	20	272	349	30	3	89	2	473
TSMP_148_CP3F5	W1913131	A506	A706	816113	102	13	115	186	21	1	91	2	301
TSMP_148_CP3F5	W1914954	A507	A706	783072	109	18	127	216	24	13	91	1	345
TSMP_148_CP3F5	W1913607	A508	A706	1926602	541	39	580	658	36	5	97	2	798

Figure 3: Samples tab for Run TSMP_148. Note the extremely high numbers of low level variants seen in this analysis for sample W1913607 , flagged in red at the bottom.

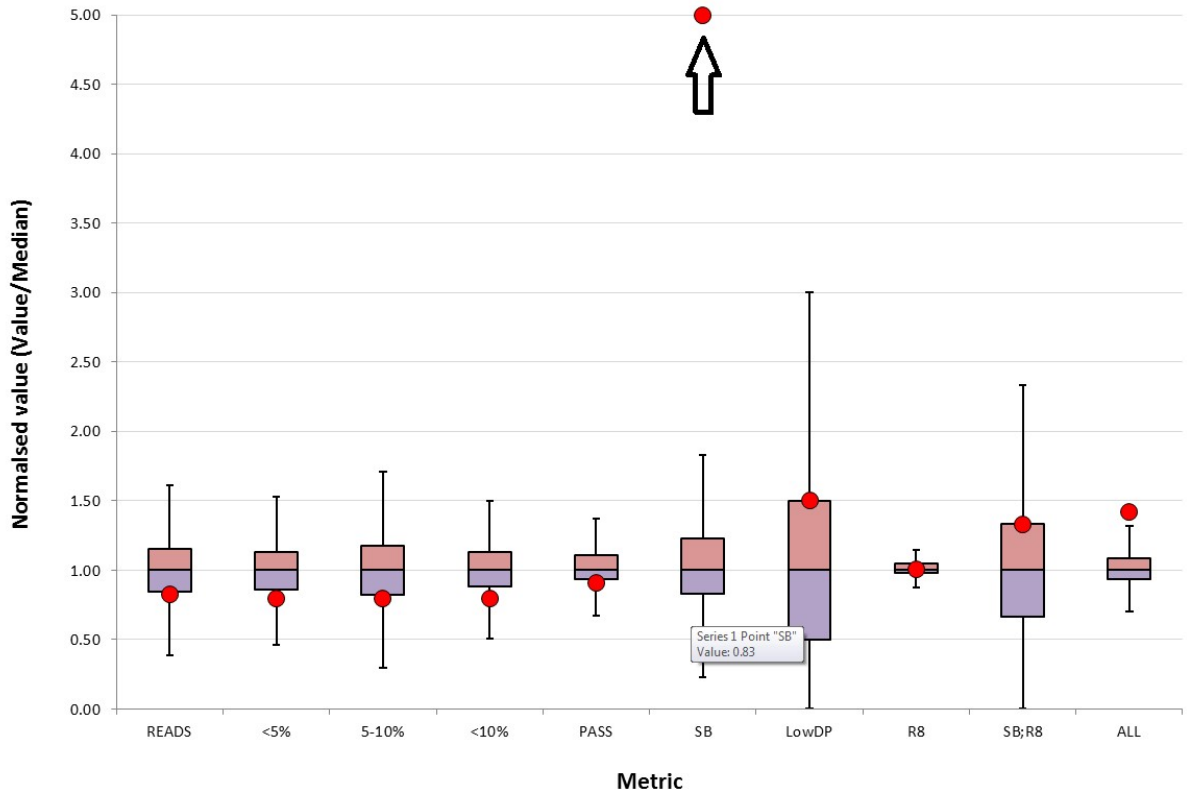
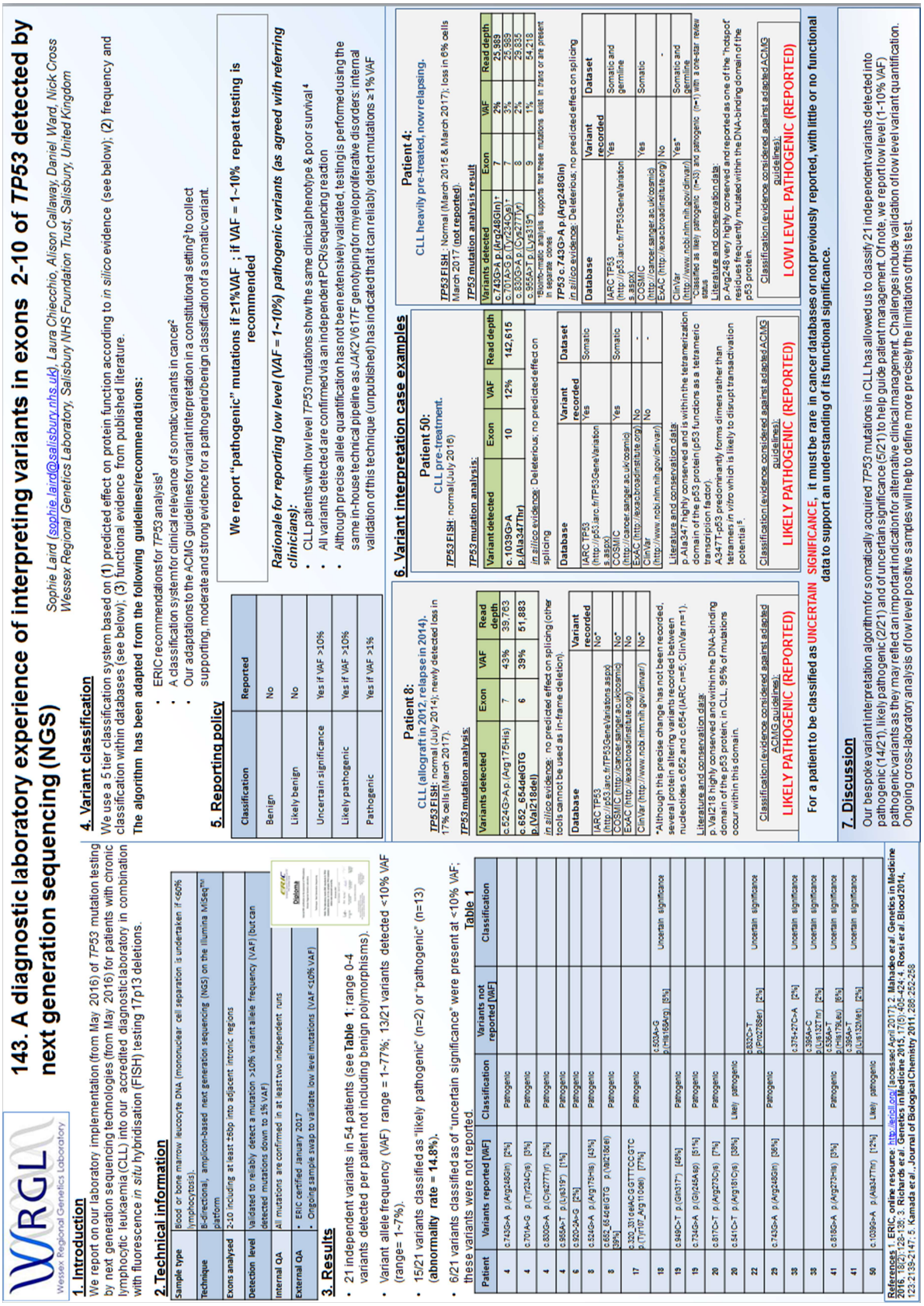


Figure 4: Run metrics tab for run TSMP_145. This run contained a reported false positive that was the subject of the original investigation. Note the extremely high level of variant that failed analysis due to strand bias [SB] (indicated with an arrow).

Operation

The tool will be used on every run for long to assess sample quality and run quality and monitor long term performance.

7.4 Appendix 4 – Poster presentation iwCLL 2016 meeting, New York. Title: A diagnostic laboratory experience of interpreting variants in exons 2-10 of TP53 detected by next generation sequencing (NGS)



7.5 Appendix 5- Somatic variant interpretation variant classification template

The first page provides a standardised framework for recording evidence for classification of pathogenicity. The second page provides some examples of the known mechanisms of pathogenicity for genes included on the panel; in the template utilised in a diagnostic setting, all genes on the panel are represented in this table.

Databases

COSMIC (exact variant): entries: tissue types: FATHMM prediction: ?Confirmed somatic:

Total number of COSMIC records at this residue:

GnomAD: Ages:

ESP:

ClinVar:

Previous occurrences WRGL (include frequency, VAF, read depth, quality):

Known mechanisms for pathogenicity in this gene:

In silico

Consensus = deleterious/neutral/conflicting

In *silico* from xx species (gaps= xx)

[paste in silico evidence here]

Or for splice site changes copy the summary information from Alamut and complete the following:

xx/5 splice prediction programs (SSF, MaxEnt, NNSPLICE, GeneSplicer, HSF) predict xxx

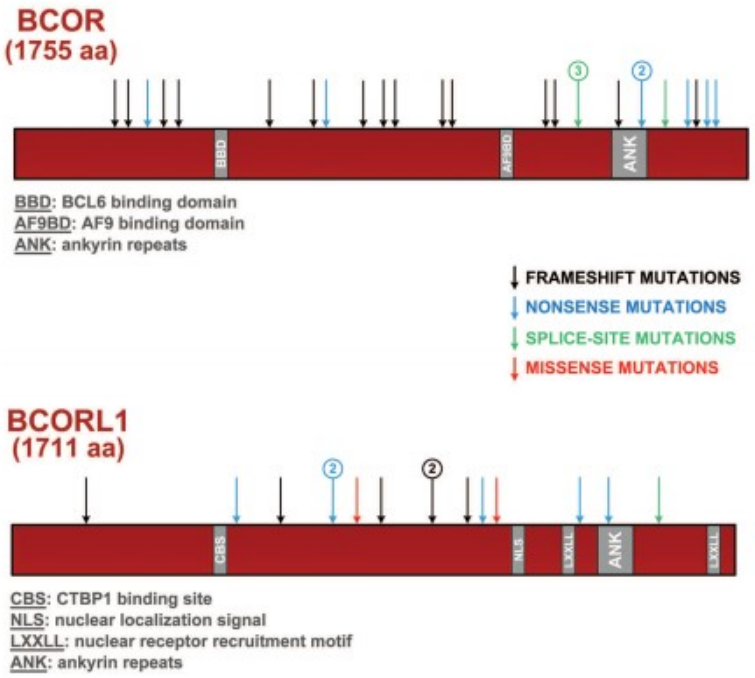
Exon xx is divisible by 3; skipping of exon xx may lead to an in-frame deletion of exon xx (however, please note that this is not always predictable).

or

Exon xx is NOT divisible by 3; skipping of exon xx may lead to a frameshift (however, please note that this is not always predictable).

Literature

Summary

Gene	Mechanism of pathogenicity
ASXL1	Truncating frameshift/nonsense/splice site mutations (Grinfeld et al, NEJM 2018; 379:1416-1430; NCCN Guidelines MDS v 2.2020)
BCOR	Truncating frameshift /nonsense/splice site (Grinfeld et al, NEJM 2018; 379:1416-1430; NCCN Guidelines MDS v 2.2020)  Figure 1. Schematic representation o BCOR (top) and BCORL1 (bottom) tran tional corepressors with their respe domains and the location and type of tions occurring in AML patients. The nu above the arrows indicate the presen more than one case, of the same type of tion occurring at the same exonic locati involving the splicing of the same i *BCORL1* mutations also include five muta that were found in four AML cell lines a the Jurkat T-lymphoblastic leukemia cel The specific type (i.e. frameshift, nons missense or splice-site) of *BCORL1* muta was inferred from Li *et al.*[9] Tiacci et al, Haematologica 2019, 97(1)
BCORL1	Similarly to mutations of *BCOR*, most of the alterations affecting *BCORL1* are nonsense mutations, out-of-frame insertions/deletions or splice site mutations that, although not resulting in clearly diminished mRNA levels of *BCORL1*, are predicted to encode truncated proteins lacking the last C-terminal LXXLL nuclear receptor recruitment motif [Tiacci et al, Haematologica 2012, 97(12)3-5]- *Cannot find exact position of this motif but Totoki (below) illustrates beyond residue 1618.*

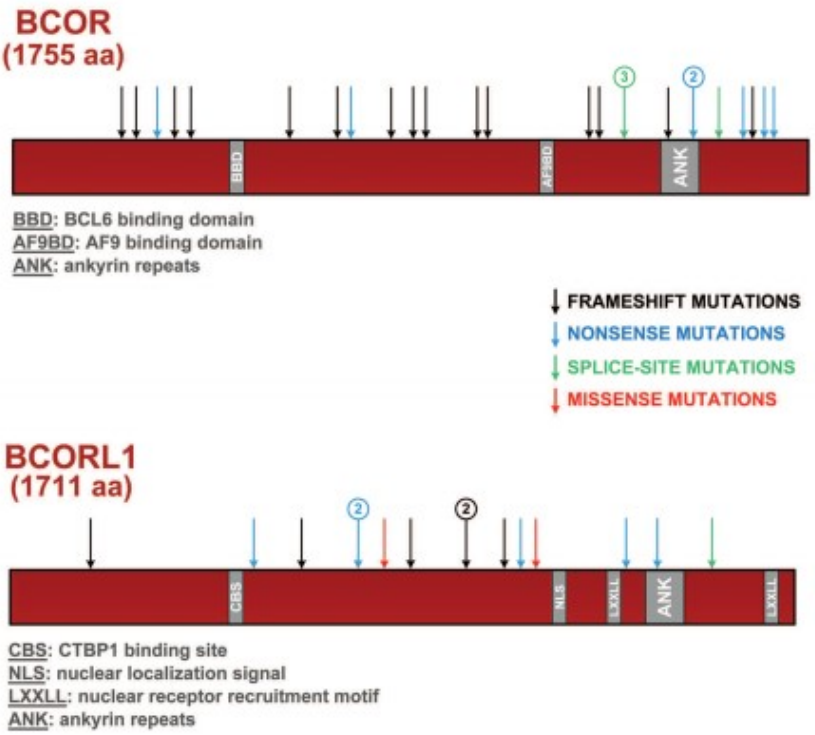


Figure 1. Schematic repres
BCOR (top) and BCORL1 (b
tional corepressors with
domains and the location a
tions occurring in AML patier
above the arrows indicate
more than one case, of the sa
tion occurring at the same e
involving the splicing of th
BCORL1 mutations also inclu
that were found in four AML
the Jurkat T-lymphoblastic le
The specific type (i.e. fram
missense or splice-site) of *BC*
was inferred from Li *et al.*[9]

Tiacci et al, Haematologica 2019, 97(1)

FIGURE 1

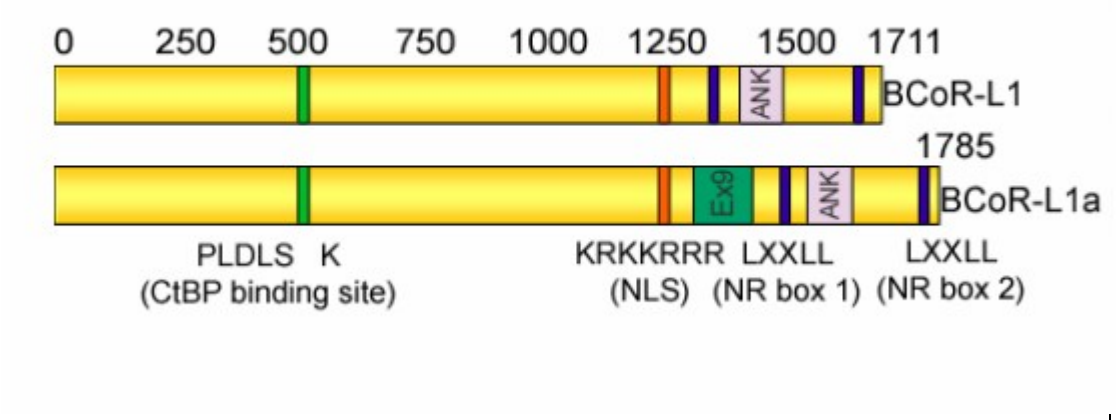


Figure taken from Pagan et al, J Biol Chem 2007 A NOVEL COREPRESSOR, BCOR-L1, REPRESSES TRANSCRIPTION THROUGH AN INTERACTION WITH CTBP

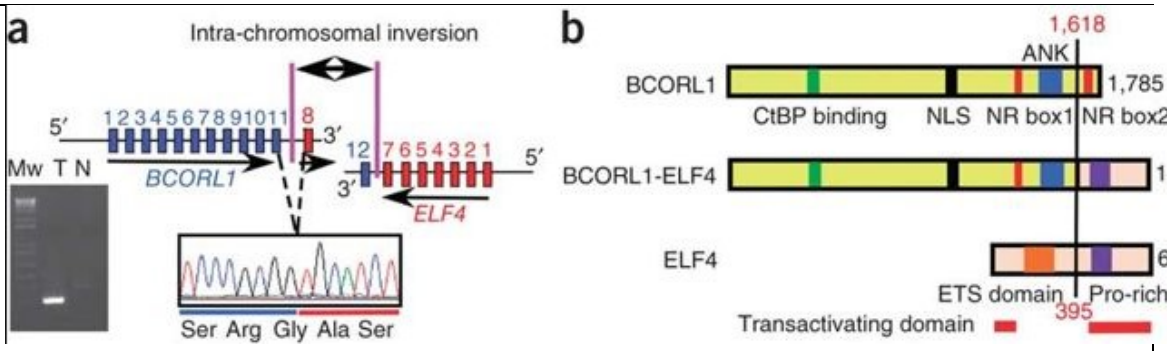


Image from Totoki et al, Nature Genetics 2011, 43: 464-469 included to show that the C-terminal LXXLL motif is beyond residue 1618.

7.6 Appendix 6- ddPCR results for the replicate analyses performed on control samples

For each reference sample, the number of replicate analyses which met the minimum acceptable droplet cut-off level as recommended by BioRad (≥10,000) is also provided:

(i) The WHO 1st International *JAK2* V617F Reference sample (supplied by NIBSC) at allelic burden of 0% (n=82); Table 7.4.1.

(ii) Blank (water) control (n=115); Table 7.4.2.

(iii) The WHO 1st International *JAK2* V617F Reference sample (supplied by NIBSC) at allelic burden of 0.03% (n=129); Table 7.4.3.

The results from these analyses are shown below:

Replicate	Sample	Total Accepted Droplets	Positive droplets	Negative droplets	Concentration	Variant allele frequency
1	0%	10716	0	10716	0	0.000%
2	0%	12499	1	12498	0.09	0.005%
3	0%	13036	1	13035	0.09	0.005%
4	0%	11555	0	11555	0	0.000%
5	0%	11916	0	11916	0	0.000%
6	0%	15024	0	15024	0	0.000%
7	0%	11147	0	11147	0	0.000%
8	0%	10907	0	10907	0	0.000%
9	0%	10415	0	10415	0	0.000%
10	0%	11067	0	11067	0	0.000%
11	0%	12733	0	12733	0	0.000%
12	0%	10896	0	10896	0	0.000%
13	0%	13830	0	13830	0	0.000%
14	0%	11915	0	11915	0	0.000%
15	0%	12340	0	12340	0	0.000%
16	0%	10849	1	10848	0.11	0.005%
17	0%	13132	0	13132	0	0.000%
18	0%	10650	0	10650	0	0.000%
19	0%	11351	0	11351	0	0.000%
20	0%	12416	0	12416	0	0.000%
21	0%	10766	0	10766	0	0.000%
22	0%	13198	0	13198	0	0.000%

23	0%	11614	1	11613	0.1	0.004%
24	0%	12872	0	12872	0	0.000%
25	0%	13339	0	13339	0	0.000%
26	0%	14666	0	14666	0	0.000%
27	0%	10212	0	10212	0	0.000%
28	0%	10003	0	10003	0	0.000%
29	0%	11277	0	11277	0	0.000%
30	0%	15069	1	15068	0.08	0.003%
31	0%	11711	0	11711	0	0.000%
32	0%	13155	0	13155	0	0.000%
33	0%	13213	0	13213	0	0.000%
34	0%	10622	0	10622	0	0.000%
35	0%	13236	0	13236	0	0.000%
36	0%	10684	0	10684	0	0.000%
37	0%	13422	0	13422	0	0.000%
38	0%	14437	0	14437	0	0.000%
39	0%	12775	0	12775	0	0.000%
40	0%	10255	0	10255	0	0.000%
41	0%	10086	0	10086	0	0.000%
42	0%	12897	0	12897	0	0.000%
43	0%	12253	0	12253	0	0.000%
44	0%	10205	0	10205	0	0.000%
45	0%	13639	0	13639	0	0.000%
46	0%	13724	0	13724	0	0.000%
47	0%	14488	0	14488	0	0.000%
48	0%	17715	0	17715	0	0.000%
49	0%	13422	0	13422	0	0.000%
50	0%	11961	0	11961	0	0.000%
51	0%	12496	0	12496	0	0.000%
52	0%	10818	0	10818	0	0.000%
53	0%	11931	0	11931	0	0.000%
54	0%	15144	0	15144	0	0.000%
55	0%	15511	0	15511	0	0.000%
56	0%	11075	0	11075	0	0.000%
57	0%	13332	0	13332	0	0.000%
58	0%	12670	0	12670	0	0.000%
59	0%	14712	1	14711	0.08	0.005%
60	0%	13818	1	13817	0.09	0.005%
61	0%	12249	0	12249	0	0.000%
62	0%	11757	0	11757	0	0.000%
63	0%	12171	0	12171	0	0.000%
64	0%	12026	0	12026	0	0.000%

65	0%	13031	0	13031	0	0.000%
66	0%	14811	0	14811	0	0.000%
67	0%	13286	0	13286	0	0.000%
68	0%	14845	0	14845	0	0.000%
69	0%	13204	0	13204	0	0.000%
70	0%	11742	0	11742	0	0.000%
71	0%	12394	0	12394	0	0.000%
72	0%	15059	0	15059	0	0.000%
73	0%	15041	1	15040	0.08	0.005%
74	0%	12224	0	12224	0	0.000%
75	0%	10075	0	10075	0	0.000%
76	0%	11244	0	11244	0	0.000%
77	0%	13469	0	13469	0	0.000%
78	0%	11197	0	11197	0	0.000%
79	0%	10798	0	10798	0	0.000%
80	0%	10977	0	10977	0	0.000%
81	0%	13949	0	13949	0	0.000%
82	0%	12002	0	12002	0	0.000%

Table 7.4.1 ddPCR results from 82 replicates of the WHO 1st International *JAK2* V617F Reference sample (supplied by NIBSC) at allelic burden of 0%

Replicate	Sample	Total Accepted Droplets	Positive droplets	Negative droplets	Concentration	Variant allele frequency
1	BLANK	10192	0	10192	0	0.000%
2	BLANK	11266	0	11266	0	0.000%
3	BLANK	12968	0	12968	0	0.000%
4	BLANK	11715	0	11715	0	0.000%
5	BLANK	13059	0	13059	0	0.000%
6	BLANK	11600	0	11600	0	0.000%
7	BLANK	12467	0	12467	0	0.000%
8	BLANK	10814	0	10814	0	0.000%
9	BLANK	11026	0	11026	0	0.000%
10	BLANK	12506	0	12506	0	0.000%
11	BLANK	10358	0	10358	0	0.000%
12	BLANK	14399	0	14399	0	0.000%
13	BLANK	14494	0	14494	0	0.000%
14	BLANK	15162	0	15162	0	0.000%
15	BLANK	13346	0	13346	0	0.000%
16	BLANK	13649	0	13649	0	0.000%
17	BLANK	10202	0	10202	0	0.000%
18	BLANK	11025	0	11025	0	0.000%
19	BLANK	11140	0	11140	0	0.000%
20	BLANK	12443	0	12443	0	0.000%

21	BLANK	12617	0	12617	0	0.000%
22	BLANK	12111	0	12111	0	0.000%
23	BLANK	12681	0	12681	0	0.000%
24	BLANK	14335	0	14335	0	0.000%
25	BLANK	14510	0	14510	0	0.000%
26	BLANK	14088	0	14088	0	0.000%
27	BLANK	10646	0	10646	0	0.000%
28	BLANK	13164	0	13164	0	0.000%
29	BLANK	13899	0	13899	0	0.000%
30	BLANK	11469	0	11469	0	0.000%
31	BLANK	13671	0	13671	0	0.000%
32	BLANK	13164	0	13164	0	0.000%
33	BLANK	13032	0	13032	0	0.000%
34	BLANK	14571	0	14571	0	0.000%
35	BLANK	14498	0	14498	0	0.000%
36	BLANK	15116	0	15116	0	0.000%
37	BLANK	13558	0	13558	0	0.000%
38	BLANK	10508	0	10508	0	0.000%
39	BLANK	10089	0	10089	0	0.000%
40	BLANK	12673	0	12673	0	0.000%
41	BLANK	12201	0	12201	0	0.000%
42	BLANK	13814	0	13814	0	0.000%
43	BLANK	11391	0	11391	0	0.000%
44	BLANK	10206	0	10206	0	0.000%
45	BLANK	12704	0	12704	0	0.000%
46	BLANK	13093	0	13093	0	0.000%
47	BLANK	13958	0	13958	0	0.000%
48	BLANK	13117	0	13117	0	0.000%
49	BLANK	12627	0	12627	0	0.000%
50	BLANK	11704	0	11704	0	0.000%
51	BLANK	10575	0	10575	0	0.000%
52	BLANK	14016	0	14016	0	0.000%
53	BLANK	14377	0	14377	0	0.000%
54	BLANK	14039	0	14039	0	0.000%
55	BLANK	12453	0	12453	0	0.000%
56	BLANK	12504	0	12504	0	0.000%
57	BLANK	13135	0	13135	0	0.000%
58	BLANK	11731	0	11731	0	0.000%
59	BLANK	12752	0	12752	0	0.000%
60	BLANK	13729	0	13729	0	0.000%
61	BLANK	10232	0	10232	0	0.000%
62	BLANK	14464	0	14464	0	0.000%
63	BLANK	14980	0	14980	0	0.000%
64	BLANK	13772	0	13772	0	0.000%
65	BLANK	12441	0	12441	0	0.000%

66	BLANK	15332	0	15332	0	0.000%
67	BLANK	14923	0	14923	0	0.000%
68	BLANK	10339	0	10339	0	0.000%
69	BLANK	14192	0	14192	0	0.000%
70	BLANK	10602	0	10602	0	0.000%
71	BLANK	16525	0	16525	0	0.000%
72	BLANK	16061	0	16061	0	0.000%
73	BLANK	15025	0	15025	0	0.000%
74	BLANK	15614	0	15614	0	0.000%
75	BLANK	15107	0	15107	0	0.000%
76	BLANK	13552	0	13552	0	0.000%
77	BLANK	13796	0	13796	0	0.000%
78	BLANK	14232	0	14232	0	0.000%
79	BLANK	13610	0	13610	0	0.000%
80	BLANK	14692	0	14692	0	0.000%
81	BLANK	13909	0	13909	0	0.000%
82	BLANK	13088	0	13088	0	0.000%
83	BLANK	14125	0	14125	0	0.000%
84	BLANK	12979	0	12979	0	0.000%
85	BLANK	11331	0	11331	0	0.000%
86	BLANK	12517	0	12517	0	0.000%
87	BLANK	10626	0	10626	0	0.000%
88	BLANK	11352	0	11352	0	0.000%
89	BLANK	12307	0	12307	0	0.000%
90	BLANK	15768	2	15766	0.15	0.012%
91	BLANK	15232	0	15232	0	0.000%
92	BLANK	15602	0	15602	0	0.000%
93	BLANK	15819	0	15819	0	0.000%
94	BLANK	11465	0	11465	0	0.000%
95	BLANK	11669	0	11669	0	0.000%
96	BLANK	10357	0	10357	0	0.000%
97	BLANK	13419	0	13419	0	0.000%
98	BLANK	14538	0	14538	0	0.000%
99	BLANK	12894	0	12894	0	0.000%
100	BLANK	10743	0	10743	0	0.000%
101	BLANK	11908	0	11908	0	0.000%
102	BLANK	10604	0	10604	0	0.000%
103	BLANK	11806	0	11806	0	0.000%
104	BLANK	10883	0	10883	0	0.000%
105	BLANK	14089	0	14089	0	0.000%
106	BLANK	12077	0	12077	0	0.000%
107	BLANK	10712	0	10712	0	0.000%
108	BLANK	11416	0	11416	0	0.000%
109	BLANK	11963	0	11963	0	0.000%
110	BLANK	11017	0	11017	0	0.000%

111	BLANK	11492	0	11492	0	0.000%
112	BLANK	12489	0	12489	0	0.000%
113	BLANK	12365	0	12365	0	0.000%
114	BLANK	12441	0	12441	0	0.000%
115	BLANK	12568	0	12568	0	0.000%

Table 7.4.2 ddPCR results from 115 replicates of a blank control.

Replicate	Sample	Total Accepted Droplets	Positive droplets	Negative droplets	Concentration	Variant allele frequency
1	0.03%	12388	1	12387	1.1	0.05%
2	0.03%	10131	2	10129	1	0.05%
3	0.03%	13323	2	13321	1	0.05%
4	0.03%	13174	2	13172	0.7	0.03%
5	0.03%	10018	3	10015	0.9	0.04%
6	0.03%	10945	3	10942	0.8	0.04%
7	0.03%	12463	3	12460	0.7	0.03%
8	0.03%	11370	3	11367	1.4	0.06%
9	0.03%	12161	3	12158	1.1	0.05%
10	0.03%	11238	3	11235	0.8	0.05%
11	0.03%	13691	3	13688	0.47	0.02%
12	0.03%	11222	3	11219	0.54	0.03%
13	0.03%	11093	4	11089	0.59	0.03%
14	0.03%	12396	4	12392	1.1	0.06%
15	0.03%	10490	4	10486	0.42	0.02%
16	0.03%	10589	4	10585	1	0.06%
17	0.03%	11750	4	11746	1	0.04%
18	0.03%	11901	4	11897	1	0.04%
19	0.03%	11873	4	11869	0.8	0.04%
20	0.03%	12414	5	12409	0.38	0.02%
21	0.03%	10036	5	10031	0.7	0.04%
22	0.03%	10127	5	10122	1.2	0.06%
23	0.03%	11135	5	11130	0.56	0.03%
24	0.03%	14102	5	14097	1	0.05%
25	0.03%	14025	5	14020	0.9	0.05%
26	0.03%	11616	5	11611	1	0.06%
27	0.03%	15031	5	15026	0.6	0.03%
28	0.03%	11613	5	11608	0.35	0.02%
29	0.03%	10459	5	10454	0.7	0.03%
30	0.03%	13094	5	13089	0.8	0.03%
31	0.03%	12073	5	12068	0.32	0.01%
32	0.03%	13077	6	13071	0.7	0.03%
33	0.03%	12043	6	12037	1.1	0.04%
34	0.03%	10077	6	10071	1.5	0.06%
35	0.03%	12615	6	12609	1.1	0.04%

36	0.03%	11575	6	11569	0.45	0.04%
37	0.03%	12507	6	12501	0.23	0.02%
38	0.03%	11304	6	11298	0.6	0.05%
39	0.03%	10202	6	10196	0.6	0.05%
40	0.03%	13344	6	13338	0.7	0.04%
41	0.03%	13988	6	13982	0.44	0.03%
42	0.03%	11368	6	11362	0.8	0.03%
43	0.03%	13089	6	13083	1	0.04%
44	0.03%	11590	6	11584	0.9	0.04%
45	0.03%	13252	6	13246	0.8	0.04%
46	0.03%	10659	6	10653	0.9	0.05%
47	0.03%	13118	6	13112	0.4	0.02%
48	0.03%	14488	6	14482	0.64	0.04%
49	0.03%	12152	6	12146	0.09	0.01%
50	0.03%	12652	6	12646	0.56	0.03%
51	0.03%	11409	7	11402	0.6	0.04%
52	0.03%	10847	7	10840	0.8	0.03%
53	0.03%	12225	7	12218	1.2	0.05%
54	0.03%	10347	7	10340	0.65	0.02%
55	0.03%	11399	7	11392	0.8	0.03%
56	0.03%	11394	7	11387	0.7	0.04%
57	0.03%	12409	7	12402	0.7	0.04%
58	0.03%	10803	7	10796	0.65	0.04%
59	0.03%	12923	7	12916	0.68	0.04%
60	0.03%	12660	7	12653	1.1	0.07%
61	0.03%	10614	7	10607	0.18	0.01%
62	0.03%	11172	7	11165	0.7	0.04%
63	0.03%	11784	7	11777	0.53	0.03%
64	0.03%	12692	7	12685	0.58	0.03%
65	0.03%	14176	7	14169	0.42	0.02%
66	0.03%	14901	7	14894	0.42	0.02%
67	0.03%	13626	7	13619	0.51	0.03%
68	0.03%	12855	7	12848	1	0.06%
69	0.03%	12825	7	12818	0.28	0.02%
70	0.03%	14405	7	14398	0.55	0.03%
71	0.03%	12278	7	12271	0.7	0.04%
72	0.03%	10140	7	10133	0.31	0.02%
73	0.03%	11216	8	11208	0.9	0.06%
74	0.03%	11326	8	11318	0.29	0.02%
75	0.03%	11713	8	11705	0.6	0.04%
76	0.03%	11082	8	11074	0.8	0.04%
77	0.03%	13933	8	13925	0.9	0.05%
78	0.03%	12821	8	12813	0.64	0.03%
79	0.03%	14082	8	14074	0.39	0.02%
80	0.03%	15796	8	15788	0.67	0.04%

81	0.03%	10589	8	10581	0.6	0.04%
82	0.03%	11799	8	11791	0.31	0.02%
83	0.03%	11617	8	11609	0.53	0.03%
84	0.03%	11177	9	11168	0.26	0.01%
85	0.03%	10325	9	10316	0.9	0.05%
86	0.03%	11203	9	11194	0.5	0.03%
87	0.03%	11498	9	11489	0.74	0.04%
88	0.03%	11127	9	11118	1.2	0.06%
89	0.03%	14314	9	14305	1.3	0.06%
90	0.03%	14217	9	14208	1.1	0.05%
91	0.03%	10611	9	10602	0.64	0.03%
92	0.03%	10965	9	10956	0.74	0.03%
93	0.03%	14836	9	14827	1	0.05%
94	0.03%	11099	9	11090	0.6	0.03%
95	0.03%	11796	10	11786	1.1	0.05%
96	0.03%	11555	10	11545	0.9	0.04%
97	0.03%	11162	10	11152	0.8	0.04%
98	0.03%	12124	10	12114	0.4	0.02%
99	0.03%	12302	10	12292	0.54	0.03%
100	0.03%	12848	10	12838	1	0.05%
101	0.03%	13998	10	13988	0.57	0.03%
102	0.03%	12555	10	12545	1	0.05%
103	0.03%	13456	10	13446	0.51	0.02%
104	0.03%	11601	10	11591	0.6	0.03%
105	0.03%	13610	10	13600	0.7	0.03%
106	0.03%	11638	11	11627	0.71	0.03%
107	0.03%	13309	11	13298	1.1	0.05%
108	0.03%	10808	11	10797	1	0.05%
109	0.03%	13400	11	13389	1.2	0.05%
110	0.03%	11591	11	11580	1.3	0.06%
111	0.03%	13844	11	13833	0.6	0.03%
112	0.03%	12218	11	12207	0.53	0.02%
113	0.03%	12007	11	11996	1.2	0.06%
114	0.03%	13566	11	13555	0.7	0.03%
115	0.03%	12360	12	12348	0.54	0.03%
116	0.03%	13028	12	13016	1	0.05%
117	0.03%	14050	12	14038	0.49	0.02%
118	0.03%	12546	12	12534	0.9	0.04%
119	0.03%	12074	12	12062	1.2	0.06%
120	0.03%	14533	12	14521	0.45	0.02%
121	0.03%	12734	13	12721	0.8	0.03%
122	0.03%	13042	13	13029	0.58	0.02%
123	0.03%	11572	14	11558	0.31	0.02%
124	0.03%	15137	14	15123	1	0.06%
125	0.03%	13910	14	13896	0.4	0.02%

126	0.03%	12108	15	12093	0.8	0.04%
127	0.03%	15663	16	15647	0.56	0.03%
128	0.03%	14853	17	14836	0.18	0.01%
129	0.03%	14908	17	14891	0.49	0.02%

Table 7.4.3 ddPCR results from 129 replicates of the WHO 1st International *JAK2* V617F Reference sample (supplied by NIBSC) at allelic burden of 0.03%

7.7 Appendix 7 – ddPCR results from samples following selection by re-analysis of samples negative for *JAK2* V617F genotyping test using modified bioinformatic pipeline

Results from ddPCR experiments on the "Low" sample cohort: these samples were originally *JAK2* V617F negative by the WRGL diagnostic pipeline (LoD=1%) but were identified as possible low level variant positive by bespoke bioinformatic analysis. Seventy nine samples were tested, but only 77 samples yielded a result. A low level JAK2 V617F variant was detected in 24 (31%) samples, and a further 6 cases had suggested positivity <0.03% VAF but were considered as negative.

Sample name: "Low…" number	Number of passed (>10,000) replicates	Average *JAK2* V617F VAF	*CALR* or *MPL* positive by diagnostic MPN panel (VAF)	Comments
1	6	0.05%		
2	3	0.25%	*CALR* mutated (35%)	
3	5	0.00%		
4	6	0.00%		
5	3	0.11%		
6	3	0.07%	*CALR* mutated (34%)	
7	6	0.00%		
8	3	0.21%	*CALR* mutated (12%)	
9	3	0.65%		
10	5	0.00%		
11	5	0.00%		
12	5	0.00%	*CALR* mutated (17%)	Below VAF cut-off (but >3 droplets in 2 replicates)
13	5	0.00%		
14	5	0.00%		
15	5	0.00%		
16	3	0.78%	*CALR* mutated (10%)	
17	5	0.00%		
18	6	0.00%		
19	3	0.78%	*MPL* mutated (3%)	
20	2	0.13%		
22	3	0.00%		
23	3	0.11%		
24	2	0.04%		
25	3	0.13%		
26	3	0.87%		
27	3	0.12%		
28	3	0.62%	*CALR* mutated (43%)	

29	3	0.00%		Below VAF cut-off (but >3 droplets in 2 replicates)
30	3	0.00%		
31	3	0.00%		
33	3	0.17%		
34	3	0.00%		
35	3	0.00%		
36	3	0.00%		
37	3	0.00%		
38	3	0.00%		
39	3	0.00%	*CALR* mutated (52%)	
40	3	0.00%		
41	3	0.07%		
42	3	0.00%	*CALR* mutated (53%)	
43	3	0.00%		
44	3	0.00%	*MPL* mutated (19%)	Below VAF cut-off (but >3 droplets in 2 replicates)
45	3	0.00%		Below VAF cut-off (but >3 droplets in 2 replicates)
46	3	0.00%		
47	2	0.00%		Below VAF cut-off (but >3 droplets in 2 replicates)
48	2	0.00%		
49	3	0.00%		
50	3	0.00%		
51	3	0.00%	*MPL* mutated (52%)	
52	2	0.00%		
53	3	0.00%	*CALR* mutated (51%)	
54	3	0.00%		
55	3	0.00%		
56	3	0.00%		
57	2	0.00%		
58	3	0.00%		
59	2	0.00%		
60	3	0.00%		
61	4	0.00%		
62	3	0.00%		
63	3	0.00%		
64	3	0.00%		
65	3	0.52%		
66	2	0.00%		
67	3	0.07%		
68	3	0.00%	*CALR* mutated (52%)	
69	3	0.04%	*CALR* mutated (27%)	
70	3	0.16%		
71	3	0.00%		Mean VAF 0.03% but below positive droplet cut-off
72	3	0.55%		

73	3	0.00%	*CALR* mutated (13%)	
74	3	0.00%		
75	3	0.00%		
76	3	0.75%	*CALR* mutated (32%)	
77	3	0.00%		
78	3	0.00%		
79	3	0.49%		

7.8 Appendix 8 – ddPCR results from samples identified as triple-negative by the diagnostic MPN panel

Results from ddPCR experiments on the randomly selected samples shown to be triple-negative by the MPN panel (n=107), i.e. *JAK2* V617F negative and no evidence of a *CALR*, *MPL* or *JAK2* exon 12 mutation by the validated WRGL genotyping assay (LoD=1%); termed the triple-negative (unselected) cohort. Where a possible low level variant below the validated cut-off was detected, a comment has been provided.

Sample name	Replicates	Average VAF	Comments
201	3	0.00%	
202	3	0.00%	
203	3	0.00%	
204	2	0.00%	
205	2	0.00%	
206	2	0.04%	
207	3	0.00%	
208	2	0.00%	
209	3	0.00%	
210	3	0.00%	
211	3	0.00%	
212	3	0.00%	
213	3	0.00%	
214	3	0.00%	
215	3	0.00%	
216	4	0.00%	
217	3	0.00%	
218	2	0.00%	
219	3	0.00%	
220	3	0.00%	
221	3	0.00%	
222	6	0.00%	
223	3	0.00%	
224	3	0.00%	
225	3	0.00%	
226	6	0.00%	
227	2	0.00%	
228	3	0.00%	
229	3	0.00%	
230	9	0.00%	
231	3	0.00%	
232	3	0.00%	
233	3	0.00%	
234	3	0.00%	
235	3	0.00%	
236	6	0.00%	

237	6	0.00%	
238	3	0.00%	
239	3	0.00%	
240	3	0.00%	
241	3	0.00%	
242	6	0.00%	
243	6	0.00%	
244	3	0.00%	
245	9	0.00%	Below VAF cut-off (but >3 droplets in 2 replicates)
246	6	0.00%	
247	3	0.00%	
248	6	0.00%	
249	9	0.00%	
250	3	0.00%	
251	6	0.00%	
252	9	0.00%	
253	9	0.00%	
254	6	0.00%	
255	3	0.00%	
256	3	0.00%	
257	3	0.00%	
258	3	0.00%	
259	6	0.00%	
260	6	0.00%	
261	3	0.00%	
262	3	0.00%	
263	3	0.00%	
264	6	0.00%	
265	3	0.00%	
266	3	0.00%	
267	6	0.00%	
268	3	0.04%	
269	3	0.00%	
270	3	0.00%	
271	3	0.00%	
272	6	0.00%	
273	6	0.00%	
274	3	0.00%	
275	3	0.21%	
276	3	0.00%	
277	3	0.00%	
278	3	0.00%	
279	3	0.00%	
280	3	0.00%	
281	3	0.00%	
282	3	0.00%	
283	3	0.00%	
284	3	0.00%	
285	3	0.00%	

286	3	0.00%	
287	3	0.00%	
288	3	0.00%	
289	3	0.00%	
290	3	0.00%	
291	3	0.00%	
292	3	0.00%	Below VAF cut-off (but >3 droplets in 2 replicates)
293	3	0.00%	
294	3	0.00%	
295	3	0.00%	
296	3	0.00%	
297	3	0.00%	
298	3	0.03%	
299	3	0.00%	
300	3	0.00%	
301	3	0.00%	
302	3	0.00%	
303	3	0.00%	
304	3	0.00%	
305	3	0.00%	
306	3	0.00%	
307	3	0.00%	

7.9 Appendix 9 – ddPCR results from samples known to be *CALR/MPL* mutated

7.9.1 ddPCR results from samples known to be *CALR* mutated

Results from ddPCR experiments on the randomly selected samples shown to be *CALR* positive and *JAK2* V617F negative (LoD=1%) by the MPN panel (n=149); termed the CALR/MPL positive cohort. Where a possible low level variant below the validated cut-off was detected, a comment has been provided.

Sample name	Replicates	Average VAF (%)	CALR mutation	Comments
308	3	0.00%	CALR 5 bp ins	
309	2	0.00%	CALR 52 bp del	
311	3	0.00%	CALR 5 bp ins	
315	3	0.00%	CALR other del	
316	3	0.00%	CALR 52 bp del	
318	3	0.04%	CALR 5 bp ins	
319	3	0.00%	CALR other del	
321	3	0.00%	CALR other del	
322	3	0.00%	CALR 52 bp del	
323	3	0.00%	CALR 52 bp del	
324	3	0.00%	CALR 5 bp ins	
327	3	0.00%	CALR other del	
328	3	0.00%	CALR 5 bp ins	
329	3	0.00%	CALR other del	
330	3	0.00%	CALR other del	
333	3	0.00%	CALR 52 bp del	
334	3	0.00%	CALR 52 bp del	
337	3	0.00%	CALR 52 bp del	
339	3	0.00%	CALR 5 bp ins	
345	3	0.00%	CALR 5 bp ins	
346	3	0.00%	CALR 5 bp ins	
348	3	0.00%	CALR 5 bp ins	
352	3	0.00%	CALR 5 bp ins	
353	3	0.00%	CALR 5 bp ins	
355	3	0.00%	CALR 52 bp del	
357	3	0.00%	CALR other del	
358	3	0.00%	CALR other del	
359	3	0.00%	CALR 5 bp ins	
360	3	0.00%	CALR 5 bp ins	
363	2	0.00%	CALR 52 bp del	
365	4	0.03%	CALR 5 bp ins	
366	3	0.00%	CALR 5 bp ins	

368	3	0.00%	CALR 5 bp ins	
370	3	0.00%	CALR 52 bp del	
372	3	0.00%	CALR 5 bp ins	
373	2	0.00%	CALR 5 bp ins	
378	2	0.00%	CALR other del	
381	3	0.00%	CALR 5 bp ins	
382	3	0.00%	CALR 5 bp ins	
384	3	0.34%	CALR 5 bp ins	
386	3	0.00%	CALR other del	
388	3	0.00%	CALR 52 bp del	
389	3	0.00%	CALR 5 bp ins	
391	3	0.00%	CALR other del	
397	3	0.00%	CALR other del	
399	3	0.00%	CALR other del	
404	2	0.00%	CALR 52 bp del	
407	3	0.00%	CALR 5 bp ins	
408	3	0.00%	CALR 5 bp ins	
409	3	0.00%	CALR 5 bp ins	
413	3	0.00%	CALR 5 bp ins	
415	3	0.05%	CALR 5 bp ins	
416	3	0.00%	CALR 52 bp deletion	
417	3	0.00%	CALR other del	
418	3	0.00%	CALR 52 bp deletion	
419	3	0.00%	CALR 5 bp ins	
420	3	0.00%	CALR other del	
421	3	0.90%	CALR other del	
422	2	0.00%	CALR 5 bp ins	
423	3	0.00%	CALR other del	
424	3	0.00%	CALR 5 bp ins	
425	2	0.00%	CALR 52 bp deletion	
426	3	0.00%	CALR 52 bp deletion	
427	3	0.00%	CALR 5 bp ins	
428	3	0.00%	CALR 52 bp deletion	
429	3	0.00%	CALR 52 bp deletion	
430	3	0.00%	CALR 52 bp deletion	
431	3	0.00%	CALR 52 bp deletion	Below VAF cut-off (but >3 droplets in 2 replicates)
432	3	0.00%	CALR 52 bp deletion	
433	2	0.15%	CALR other del	
434	3	0.00%	CALR 52 bp deletion	
435	3	0.00%	CALR 5 bp ins	
436	2	0.00%	CALR 5 bp ins	
437	3	0.00%	CALR 5 bp ins	
438	3	0.00%	CALR other del	
439	3	0.00%	CALR 52 bp deletion	
440	3	0.00%	CALR 5 bp ins	

441	3	0.00%	CALR 52 bp deletion	
442	3	0.00%	CALR 52 bp deletion	
443	3	0.00%	CALR other del	
444	3	0.00%	CALR other del	
445	3	0.00%	CALR 52 bp deletion	
446	3	0.00%	CALR 52 bp deletion	
447	3	0.00%	CALR other del	
448	3	0.00%	CALR 5 bp ins	
449	3	0.00%	CALR 52 bp deletion	
450	3	0.00%	CALR 5 bp ins	
451	3	0.00%	CALR 52 bp deletion	
452	3	0.00%	CALR other ins	
453	3	0.25%	CALR 52 bp deletion	
454	3	0.00%	CALR 52 bp deletion	
455	3	0.00%	CALR 52 bp deletion	
456	3	0.00%	CALR 52 bp deletion	
457	4	0.00%	CALR 5 bp ins	
458	3	0.00%	CALR 52 bp deletion	
459	3	0.00%	CALR 52 bp deletion	
460	2	0.00%	CALR 52 bp deletion	
461	3	0.00%	CALR 52 bp deletion	
462	3	0.07%	CALR other ins	
463	3	0.47%	CALR 52 bp deletion	
464	3	0.00%	CALR other del	
465	3	0.00%	CALR 5 bp ins	
466	3	0.00%	CALR 52 bp deletion	
467	3	0.00%	CALR 52 bp deletion	
468	2	0.00%	CALR 52 bp deletion	Below VAF cut-off (but >3 droplets in 2 replicates)
469	3	0.00%	CALR 5 bp ins	
470	3	0.00%	CALR 52 bp deletion	
471	3	0.00%	CALR 52 bp deletion	
472	3	0.00%	CALR 52 bp deletion	
473	4	0.00%	CALR 52 bp deletion	
474	3	0.00%	CALR 52 bp deletion	
475	3	0.00%	CALR 52 bp deletion	
477	3	0.00%	CALR 52 bp deletion	
478	3	0.00%	CALR 52 bp deletion	
479	3	1.51%	CALR 52 bp deletion	
480	3	0.00%	CALR 5 bp ins	
481	3	0.00%	CALR 52 bp deletion	
482	3	0.00%	CALR 52 bp deletion	
483	3	0.00%	CALR 5 bp ins	
484	3	1.32%	CALR 52 bp deletion	
485	3	0.00%	CALR 52 bp deletion	
486	3	0.00%	CALR 5 bp ins	Below VAF cut-off (but >3 droplets in 2

				replicates)
487	3	0.00%	CALR 5 bp ins	
488	3	0.00%	CALR 5 bp ins	
490	3	0.00%	CALR 52 bp deletion	
491	3	0.00%	CALR 52 bp deletion	
492	2	0.00%	CALR 52 bp deletion	
493	3	0.00%	CALR 5 bp ins	
494	3	0.00%	CALR 5 bp ins	
495	3	0.00%	CALR 5 bp ins	
496	3	0.00%	CALR 52 bp deletion	
497	3	0.00%	CALR 52 bp deletion	
498	3	0.00%	CALR 52 bp deletion	
499	3	0.00%	CALR 52 bp deletion	
500	3	0.00%	CALR 52 bp deletion	
501	3	0.00%	CALR 5 bp ins	
502	3	0.00%	CALR 5 bp ins	
503	3	0.00%	CALR 52 bp deletion	
504	3	0.00%	CALR 52 bp deletion	
505	3	0.00%	CALR 52 bp deletion	
506	3	0.00%	CALR 52 bp deletion	
507	3	0.00%	CALR 5 bp ins	
508	3	0.00%	CALR 52 bp deletion	
509	3	0.00%	CALR 52 bp deletion	
510	3	0.00%	CALR 52 bp deletion	
511	3	0.00%	CALR 52 bp deletion	
512	3	0.00%	CALR 5 bp ins	
513	3	0.00%	CALR 52 bp deletion	
514	3	0.00%	CALR 52 bp deletion	

7.9.2 ddPCR results from samples known to be *MPL* mutated

Results from ddPCR experiments on the randomly selected samples shown to be *MPL* positive and *JAK2* V617F negative (LoD=1%) by the MPN panel (n=52). Where a possible low level variant below the validated cut-off was detected, a comment has been provided.

Sample name	Replicates	Average VAF (%)	Mutation	Comments
310	3	0.00%	MPL W515L	
312	3	0.00%	MPL W515K	
313	3	0.09%	MPL W515L	
317	3	0.00%	MPL W505N	
320	3	0.00%	MPL W515K	
326	3	0.00%	MPL W515K	Below VAF cut-off (but >3 droplets in 2 replicates)
331	3	0.00%	MPL W505N	

332	3	0.00%	MPL W515L	Below VAF cut-off (but >3 droplets in 2 replicates)
335	3	1.09%	MPL W515K	
336	3	0.21%	MPL W515L	
340	2	0.00%	MPL W515L	
341	3	0.00%	MPL W515K	
342	2	0.00%	MPL W515K	
343	2	0.00%	MPL W515L	
344	2	0.00%	MPL W515L	
347	3	0.03%	MPL W505N	
349	3	0.00%	MPL W515L	
350	3	0.00%	MPL W515K	
351	3	0.00%	MPL W515K	
354	2	0.00%	MPL W515L	
361	3	0.00%	MPL W515L	
362	3	0.00%	MPL W515L	
364	3	0.00%	MPL W505N	
367	2	0.15%	MPL W515K	
369	3	0.00%	MPL W515K	
371	3	0.00%	MPL W515K	
374	3	0.00%	MPL W505N	
375	3	0.00%	MPL W515K	
376	3	0.00%	MPL W515L	
377	6	0.00%	MPL W505N	
379	3	0.00%	MPL W515L	
380	3	0.00%	MPL W515L	
383	3	0.15%	MPL W505N	
385	3	0.00%	MPL W515L	
387	3	0.04%	MPL W515L	
390	3	0.00%	MPL W515L	
392	3	0.00%	MPL W515L	
393	3	0.00%	MPL W515K	
394	3	0.00%	MPL W505N	Below VAF cut-off (but >3 droplets in 2 replicates)
395	3	0.00%	MPL W515L	
396	3	0.00%	MPL W515L	
398	3	0.00%	MPL W505N	
400	3	0.00%	MPL W505N	
402	3	0.00%	MPL W515L	
403	3	0.00%	MPL W515L	
405	3	0.10%	MPL W515L	
406	3	0.00%	MPL W515K	
410	3	0.00%	MPL W515K	
411	3	0.00%	MPL W515A	
412	3	0.00%	MPL W515L	
414	3	0.00%	MPL W515L	

7.10 Appendix 10 – Clinical audit results from samples with a *JAK2* V617F mutation at 1-5% VAF detected by the diagnostic MPN panel (n=62)

Sample ID	*JAK2* V617 VAF	Patient details		Date MPN panel	Final diagnosis made	WHO diagnosis	How was the diagnosis made?					Patient management					Comments
		Age	Sex				Genetics	Clinically	BM	Other	N/A	Discharged	Patient under active management	Receiving therapy	Patient being routinely monitored	Time interval (months)	
W1604531	1%	39	M	08/04/2016	**No**	ET or familial (mother, PAT GF)	Y	N	N	N	N	No	No	N/A	Yes	12	
W1504290	1%	87	F	26/03/2015	Yes	ET	Y	N	N	N	N	NA	Yes (patient now deceased)	Aspirin	Yes (patient now deceased)	N/A	Reason for demise: dementia
W1606484	1%	71	F	20/05/2016	Yes	ET	Y	N	N	N	N	No	Yes (patient now deceased)	N/A	Yes (patient now deceased)	N/A	
W1819781	1%	48	M	27/12/2018	**No**	N/A	Y	Y	N	N	N	Yes	No	N/A	No	N/A	
W1810984	1%	53	F	24/07/2018	Yes	ET	Y	N	Y	N	N	No	Yes	Aspirin	Yes	12	
W1807472	1%	43	F	18/05/2018	Yes	PV	N	N	Y	N	N	No	Yes	Info not provided	Yes	1.5-2	
W1711362	1%	57	M	11/08/2017	Yes	ET	Y	N	N	N	N	No	No	N/A	Yes	3	
W1501269	1%	52	F	28/01/2015	Yes	ET	Y	Y	N	N	N	No	Yes	Info not provided	Yes	2	
W1806548	2%	54	M	02/05/2018	**No**	?MPN	Y	Y	N	N	N	No	No	N/A	Yes	12	Diagnosis unclear
W1817196	2%	69	M	13/11/2018	Yes	ET	Y	N	N	N	N	No	Yes	Info not provided	Yes	1	
W1804013	2%	71	M	15/03/2018	Yes	PV	Y	Y	N	N	N	No	Yes	Info not provided	Yes	3	
W1800499	2%	80	F	11/01/2018	Yes	ET	Y	N	N	N	N	No	Yes	Aspirin and hydroxycar-bimide	Yes	3	Patient also has a *CALR* 52 bp mutation
W1606588	2%	46	F	24/05/2016	Yes	ET	Y	Y	N	N	N	No	Yes (patient now deceased)	N/A	Yes (patient now deceased)	N/A	Reason for demise: metastatic lung cancer

W1717160	2%	79	F	12/12/2017	Yes	ET	Y	N	N	N	N	No	Yes (patient now deceased)	N/A	Yes (patient now deceased)	3	Reason for demise: not provided
W1500514	2%	43	F	13/01/2015	Yes	ET	Y	Y	N	N	N	No	Yes	Aspirin	Yes	12	
W1601795	2%	77	F	08/02/2016	Yes	ET	Y	N	N	N	N	No	Yes	Aspirin and hydroxycar-bamide	Yes	3	
W1417772	2%	47	F	24/12/2014	Yes	ET	Y	N	N	N	N	No	Yes	Info not provided	Yes	6	
W1607522	2%	69	M	09/06/2016	Yes	ET	Y	N	N	N	N	No	Yes	Aspirin and hydroxycar-bamide	Yes	6	
W1710353	2%	71	M	25/07/2017	Yes	PMF	Y	Y	N	N	N	No	No	N/A	Yes	3-4 months originally; now more frequenctly as ?transformation to AML	Patient also has a *CALR* 52 bp mutation; no addition mutations identified by myeloid panel
W1813661	2%	39	F	12/09/2018	Yes	ET	Y	N	N	N	N	No	No	N/A	Yes	4	Normal myeloid panel
W1502131	2%	69	F	12/02/2015	Yes	ET	Y	Y	N	N	N	No	Yes	Info not provided	Yes	1.5	
W1501560	2%	43	F	03/02/2015	Yes	ET	Y	N	N	N	N	No	No	N/A	Yes	6	
W1807909	2%	46	M	30/05/2018	Yes	PV	Y	N	N	N	N	No	Yes	Info not provided	Yes	6	
W1611816	2%	63	F	08/09/2016	Yes	PV	Y	Y	N	N	N	Yes	No	N/A	Yes	UNK	Patient being monitored by GP; interval unknown.
W1809600	3%	75	F	27/06/2018	**No**	N/A	Y	Y	N	N	N	No	Yes (patient now deceased)	N/A	Yes (patient now deceased)	N/A	Reason for demise: bladder cancer
W1906645	3%	52	M	16/04/2019	Yes	ET - probable (Crohn's, HH, IDA)	Y	N	N	N	N	No	Yes	Aspirin	Yes	12	
W1814823	3%	71	M	03/10/2018	Yes	ET	Y	Y	N	N	N	No	Yes	Aspirin and hydroxycar-bamide	Yes	6	

W1417698	3%	51	F	23/12/2014	**No**	N/A	N	N	N	N	**Y**	Yes	No	N/A	Yes	12	*JAK2* positive but no evidence of MPN; patient being monitored by GP
W1714687	3%	66	F	17/10/2017	**No**	N/A	N	N	N	N	**Y**	No	Yes	Info not provided	Yes	3	
W1610764	3%	70	M	16/08/2016	Yes	PV	**Y**	**Y**	N	N	N	No	Yes (patient now deceased)	N/A	Yes until death	3	Reason for demise: not provided
W1713577	3%	75	F	26/09/2017	Yes	MPN-U	**Y**	**Y**	N	N	N	No	Yes	Info not provided	Yes	2	
W1515736	3%	62	F	20/11/2015	Yes	ET	**Y**	**Y**	N	N	N	No	Yes	Info not provided	Yes	3	
W1710543	3%	51	M	28/07/2017	Yes	ET	**Y**	**Y**	N	N	N	No	Yes	Aspirin	Yes	12	
W1800905	3%	79	M	18/01/2018	Yes	PMF	**Y**	**Y**	**Y**	N	N	No	Yes	Info not provided	Yes	1.5	Patient reported to be in an early fibrotic stage of PMF
W1814128	3%	44	F	20/09/2018	Yes	ET	**Y**	**Y**	N	N	N	No	Yes	Aspirin	Yes	3	
W1912016	3%	65	F	10/07/2019	Yes	ET	**Y**	N	N	N	N	No	No	N/A	Yes	6	
W1615957	4%	70	F	29/11/2016	Yes	ET	**Y**	N	N	N	N	No	Yes	Aspirin and hydroxycar-bamide	Yes	6	
W1701961	4%	77	F	08/02/2017	Yes	PV	**Y**	N	N	N	N	No	Yes	Info not provided	Yes	3	
W1909845	2%	36	M	07/06/2019	**No**	N/A	N	N	N	N	**Y**	N/A	No	N/A	No	UNK	Referred by GP with polycythemia
W1511042	4%	74	M	17/08/2015	Yes	MF post-ET	**Y**	N	N	N	N	No	Yes (patient now deceased)	N/A	N/A	N/A	Reason for demise: not provided but patient transformed to MF in 2016
W1507053	4%	66	M	26/05/2015	Yes	ET	**Y**	N	N	N	N	No	Yes	Hydroxycar-bamide	Yes	6	Patient also has CLL

W1604391	4%	72	F	05/04/2016	Yes	PV	Y	Y	N	N	N	No	Yes	Info not provided	Yes	2	
W1605844	1%	85	F	06/05/2016	Yes	Probable ET	Y	N	Y	N	N	N/A	UNK	UNK	UNK	UNK	Patient declined radiotherapy under oncologist for solid tumour, moved hospital, no intervention for *JAK2* clone.
W1703297	4%	86	F	07/03/2017	Yes	ET	Y	Y	N	N	N	No	Yes	Info not provided	Yes	2-3	
W1815558	4%	40	F	16/10/2018	Yes	ET	Y	N	N	N	N	No	Yes	Aspirin	Yes	6	
W1609968	4%	63	F	29/07/2016	Yes	ET	Y	Y	N	N	N	No	Yes	Info not provided	Yes	3	
W1710488	4%	77	F	27/07/2017	Yes	ET	Y	N	N	N	N	No	Yes	Info not provided	Yes	3	
W1515830	4%	69	F	24/11/2015	Yes	PV	Y	Y	N	N	N	No	No	N/A	Yes	2	
W1706213	4%	64	M	05/05/2017	Yes	ET	Y	N	N	N	N	No	Yes	N/A	Yes	3	
W1816198	4%	46	M	25/10/2018	Yes	Probable ET	Y	Y	N	N	N	No	Yes	N/A	Yes	6	
W1602433	4%	78	F	22/02/2016	Yes	PV	Y	Y	N	N	N	No	Yes	N/A	Yes	3	
W1606378	5%	79	F	18/05/2016	Yes	PV	Y	N	N	N	N	No	Yes	Hydroxyca-rbamide	Yes	4	
W1606505	5%	65	M	20/05/2016	Yes	MPN	Y	Y	N	N	N	No	Yes	N/A	Yes	2	Patient had dysplastic features on the BM
W1716256	5%	49	F	15/11/2017	Yes	ET	Y	Y	N	N	N	Not	Yes	N/A	Yes	4	
W1712658	5%	79	F	11/09/2017	Yes	ET	Y	Y	N	N	N	No	Yes	N/A	Yes	2-3	
W1706237	5%	61	M	05/05/2017	Yes	MF from MDS/MPN	Y	Y	N	N	N	No	No	On EPO for MDS	Yes	1.5	Patient also has a *CALR* mutation (66%)
W1903488	5%	68	F	26/02/2019	Yes	ET	Y	Y	N	N	N	Not	Yes	Info not provided	Yes	3	Myeloid panel also undertaken: results showed *JAK2*

																V617F only
W1803324	5%	73	M	05/03/2018	Yes	ET	Y	Y	N	N	N	No	Yes	Aspirin	Yes	3-4	
W1808576	5%	64	F	08/06/2018	Yes	ET	Y	Y	Y	N	N	No	Yes	Aspirin	Yes	4	Myeloid panel also undertaken: results showed *JAK2* V617F and *DNMT3A* variant
W1818538	5%	46	M	05/12/2018	Yes	MPN-U	Y	N	N	N	N	Yes	No	N/A	No	N/A	
W1714813	5%	71	F	19/10/2017	Yes	ET	Y	Y	N	N	N	No	Yes	Info not provided	Yes	4	
W1816392	5%	47	M	30/10/2018	Yes	ET	Y	Y	N	N	N	No	Yes	Aspirin	Yes	6	

7.11 Appendix 11 – Clinical audit results from samples that were shown to be triple-negative by the diagnostic MPN panel but within which a low level (<1% VAF) *JAK2* V617F mutation was detected by ddPCR (n=8)

Sample ID	*JAK2* V617 VAF	Patient details		Date MPN panel	Final diagnosis made	WHO diagnosis	How was the diagnosis made?						Patient management						Comments
		Age	Sex				Genetics	Clinically	BM	Other	N/A	Discharged	Patient under active management	Receiving therapy	Patient being routinely monitored	Time interval (months)	
Low 1	0.05%	74	F	11/07/20117	Yes	MPN	N	Y	N	N	N	No	Yes	Yes	Yes	4	No variants detected by myeloid panel (not reported)
Low 5	0.11%	79	F	12/07/2017	No (likely reactive)	N/A	N	N	N	N	Y	Yes	No	No	Yes (GP)	12	VUS detected in TET2 (12% VAF) by myeloid panel (not reported)
Low 20	0.13%	71	M	01/08/2017	No	N/A	N	N	N	N	Y	Yes	No	No	No	N/A	No variants detected by myeloid panel (not reported); G-banding analysis undertaken Feb 2018: 46,XY[20] (reported)
Low 27	0.12%	48	M	25/08/2017	Yes	Low protein S	N	Y	N	Y	N	Yes	No	No	No	N/A	No variants detected by myeloid panel (not reported)
Low 67	0.07%	51	M	20/06/2017	Yes	Secondary polycythaemia	N	Y	N	N	N	Yes	No	No	No	N/A	No variants detected by myeloid panel (not reported)
Low 72	0.55%	54	M	29/06/2017	Yes	Idiopathic erythrocytosis	N	Y	N	N	N	No	Yes	Yes (venesect-ions)	Yes	12	VUS detected in *ASXL1* (29% VAF) and *DNMT3A* (17% VAF) by myeloid panel (not reported)
Low 79	0.49%	81	F	13/12/2017	Yes	Giant cell arteritis and iron deficiency	N	Y	N	N	N	Yes	No	No	No	N/A	No variants detected by myeloid panel (not reported)
Sample 206	0.04%	67	F	20/12/2018	No	N/A	N	N	N	N	Y	Yes	No	No	No	N/A	VUS detected in *CUX1* (47% VAF) detected by myeloid panel (not reported) Diagnostic MPN panel also undertaken Oct 2017 and Nov 2020 and no mutations detected.

7.12 Appendix 12 – Clinical audit results from samples that had a *CALR/MPL* mutation but were negative for *JAK2* V617F mutation by the diagnostic MPN panel but were found to have a low level (<1% VAF) *JAK2* V617F mutation detected by ddPCR (n=11)

Sample ID	*JAK2* V617 VAF	Patient details		Date MPN panel	MPN panel result	Final diagnosis made	WHO diagnosis	How was the diagnosis made?					Patient management						Comments
		Age	Sex					Genetics	Clinically	BM	Other	N/A	Discharged	Patient under active management	Receiving therapy	Patient being routinely monitored	Time interval (months)		
Sample 318	0.04%	87	M	19/12/2014	*CALR* positive	Yes	ET	N	N	Y	N	N	No	Yes	Yes	Yes	3-4		
Sample 365	0.03%	69	M	06/04/2018	*CALR* positive	Yes	ET	Y	Y	N	Y*	N	No	Yes	Yes	Yes	6		
Low 8	0.21%	71	F	18/07/2017	*CALR* positive	Yes	ET	Y	N	N	N	N	No	Yes	Yes	Yes	3		
Low 69	0.04%	43	M	21/06/2017	*CALR* positive	Yes	ET	Y	N	N	N	N	No	yes	Yes	Yes	3		*CALR* mutation sole abnormality by myeloid panel on this sample (not reported) *JAK2* ARMS and BCR/ABL1 FISH testing requested in 2013; both negative (reported in a diagnostic setting).
Sample 313	0.09%	84	M	23/03/2020	*MPL* positive	Yes	ET	Y	Y	N	N	N	No	Yes	Yes	Yes	1-2		
Sample 336	0.21%	83	M	10/09/2020	*MPL* positive	Yes	ET	Y	N	N	N	N	No	Yes	Yes	Yes	3		
Sample 347	0.03%	84	F	16/07/2020	*MPL* positive	Yes	ET	Y	Y	N	N	N	No	Yes	Yes	Yes	2		
Sample 367	0.15%	64	F	29/05/2018	*MPL* positive	Yes	ET	Y	N	N	N	N	No	Yes	Yes (Aspirin)	Yes	4		*MPL* mutation sole abnormality by myeloid panel (not reported)
Sample 383	0.15%	90	F	30/10/2018	*MPL* positive	Yes	ET	Y	N	N	N	N	Yes	No	No (patient declined)	No	N/A		5 mutations detected by myeloid panel (not reported)
Sample 387	0.04%	85	M	18/12/2019	*MPL* positive	No	Probable MPN but patient refused BM so diagnosis not confirmed	Y	Y	N	N	N	No	No	Yes	Yes	4		
Sample 405	0.10%	63	M	17/07/2020	*MPL* positive	Yes	MDS/MPN-RS-T	Y	N	Y	N	N	No	Yes	Yes	Yes	1.5		3 mutations detected by myeloid panel (reported in a diagnostic setting; result c/w MDS/MPN-RS-T)

*Exclusion of other diagnoses

7.13 Appendix 13 – Virtual gene panel applied to WGS results

Gene (synonym)	Chromosome	Source (AML, MDS, MPN, AA, ARCH, predisposition to MDS/AML)
ANKRD26	10p	Obrochta E et al., Best Pract Res Clin Haematol. 2018 Dec;31(4):373-379
ASXL1	20q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-30
ASXL2	2p	Faber Z et al., Nat Genet. 2016 Dec;48(12):1551-1556
ATM	11q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
ATRX	Xq	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-59
BCOR	Xp	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-41
BCORL1	Xq	Li M et al. Blood. 2011 Nov 24;118(22):5914-7
BRAF	7q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-68
BRCC3	Xq	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
CALR	19p	Grinfeld J et al., N Engl J Med. 2018 Oct 11;379(15):1416-1435
CBL	11q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-39
CBLB	3q	Papaemmanuil E et al., N Engl J Med. 2016 Jun 9;374(23):2209-222
CCND2	12p	Faber Z et al., Nat Genet. 2016 Dec;48(12):1551-1557
CDKN2A	9p	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-60
CEBPA	19q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-63
CREBBP	16p	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-51
CSMD1	8p	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
CTCF	16p	Haferlach T et al., Leukemia. 2014 Feb;28(2):241-9
CUX1	7q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-43
DCLRE1C	10p	Haferlach T et al., Leukemia. 2014 Feb;28(2):241-10
DDX41	5q	Obrochta E et al., Best Pract Res Clin Haematol. 2018 Dec;31(4):373-378
DHX15	4p	Faber Z et al., Nat Genet. 2016 Dec;48(12):1551-1558
DIS3	13q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
DNMT3A	2p	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-31
EED	11q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
EP300	22q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-46
ETNK1	12p	Gambacorti-Passerini CB et al., Blood. 2015 Jan 15;125(3):499-503
ETV6	12p	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-61
EZH2	7q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-35
FANCL	2p	Haferlach T et al., Leukemia. 2014 Feb;28(2):241-11
FBXW7	4q	Papaemmanuil E et al., N Engl J Med. 2016 Jun 9;374(23):2209-222
FLT3	13q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-64
GATA2	3q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-49
GNAS	20q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-65
GNB1	1p	Grinfeld J et al., N Engl J Med. 2018 Oct 11;379(15):1416-1430
GPRC5A	12p	Haferlach T et al., Leukemia. 2014 Feb;28(2):241-12
IDH1	2q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-44
IDH2	15q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-36
IKZF1	7p	Papaemmanuil E et al., N Engl J Med. 2016 Jun 9;374(23):2209-222
IRF1	5q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-57
JAK1	1p	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
JAK2	9p	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-42

JAK3	19p	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
KDM5A	12p	Papaemmanuil E et al., N Engl J Med. 2016 Jun 9;374(23):2209-222
KDM6A	Xp	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-62
KIT	4q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-52
KMT2D (MLL2)	12q	Zink F et al., Blood. 2017 Aug 10;130(6):742-752
KRAS	12p	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-45
LAMB4	7q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
LUC7L2	7q	Haferlach T et al., Leukemia. 2014 Feb;28(2):241-8
MBD4	3q	Obrochta E et al., Best Pract Res Clin Haematol. 2018 Dec;31(4):373-383
MDM4	1q	Marcellino, et al., (2017)., Blood, 2017; 130(supplement 1):204
MECOM	3q	Obrochta E et al., Best Pract Res Clin Haematol. 2018 Dec;31(4):373-384
KMT2C (MLL3)	12q	Grinfeld J et al., N Engl J Med. 2018 Oct 11;379(15):1416-1431
KMT2E (MLL5)	7q	Papaemmanuil E et al., N Engl J Med. 2016 Jun 9;374(23):2209-222
MPL	1p	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-54
MYC	8q	Papaemmanuil E et al., N Engl J Med. 2016 Jun 9;374(23):2209-222
NCOR2	12q	Haferlach T et al., Leukemia. 2014 Feb;28(2):241-7
NF1	17q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-55
NFE2	12q	Grinfeld J et al., N Engl J Med. 2018 Oct 11;379(15):1416-1432
NPM1	5q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-47
NRAS	1p	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-40
PEG3	19q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
PHF6	Xq	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-48
PIGA	Xp	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
POT1	7q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
PPM1D	17q	Grinfeld J et al., N Engl J Med. 2018 Oct 11;379(15):1416-1434
PRPF40B	12q	Papaemmanuil E et al., N Engl J Med. 2016 Jun 9;374(23):2209-222
PRPF8	17p	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
PRR14L	22q	Chase A et al., Leukemia. 2019 May;33(5):1184-1194
PTEN	10q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-66
PTPN11	12q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-50
RAD21	8q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-58
RAP1A	1p	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
RB1	13q	Grinfeld J et al., N Engl J Med. 2018 Oct 11;379(15):1416-1433
RBBP4	1p	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
RIT1	1q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
RTEL1	20q	Obrochta E et al., Best Pract Res Clin Haematol. 2018 Dec;31(4):373-385
RUNX1	21q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-32
SAMD9	7q	Obrochta E et al., Best Pract Res Clin Haematol. 2018 Dec;31(4):373-380
SAMD9L	7q	Obrochta E et al., Best Pract Res Clin Haematol. 2018 Dec;31(4):373-381
SETBP1	18q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
SF1	11q	Papaemmanuil E et al., N Engl J Med. 2016 Jun 9;374(23):2209-222
SF3A1	22q	Papaemmanuil E et al., N Engl J Med. 2016 Jun 9;374(23):2209-222
SF3B1	2q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-27
SH2B3	12q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-67

SMC1A	Xp	Kon A et al., Nat Genet. 2013 Oct;45(10):1232-7
SMC3	10q	Kon A et al., Nat Genet. 2013 Oct;45(10):1232-8
SRP72	4q	Obrochta E et al., Best Pract Res Clin Haematol. 2018 Dec;31(4):373-382
SRSF2	17q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-28
STAG2	Xq	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-37
STAT3	17q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
STAT5B	17q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
SUZ12	17q	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
TERT	5p	Yoshizato T et al., N Engl J Med. 2015 Jul 2;373(1):35-47
TET2	4q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-29
TP53	17p	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-34
U2AF1	21q	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-33
U2AF2	19q	Haferlach T et al., Leukemia. 2014 Feb;28(2):241-13
WT1	11p	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-56
ZBTB7A	19p	Faber Z et al., Nat Genet. 2016 Dec;48(12):1551-1559
ZRSR2	Xp	Papaemmanuil E et al Blood. 2013 Nov 21;122(22):3616-38

7.14 Appendix 14 – Clinical details and cytogenetic results of the samples with der(6)t(1;6) identified in the WRGL

Two samples marked with * were chosen for further molecular investigations as part of this study.

Patient number	Sex	Age at time of first referral	N	Sample name	Date of sample received	Clinical details at time of sample receipt	Karyotype
WRGL-1	M	67	4	M9709000	06/12/1997	Myelofibrosis	46,XY,del(13)(q13q31)[7]/46,idem,dup(12)(q13q14) or ins(12;?)(q13;?)[8]/46,idem,der(6)t(1;6)(q21;p22)[4]/46,XY[1]
				M0000267*	12/01/2000	MDF, ?MDS, ?transformation	46,XY,der(6)t(1;6)(q21;p22),del(13)(q13q31)[30]
				M0002875	11/04/2000	Myelofibrosis diagnosed 1997. Increased WCC since, ?transformation to AML/RAEB	46,XY,der(6)t(1;6)(q21;p22),del(13)(q13q31)[30]
				M0203804	25/04/2002	Myelofibrotic	46,XY,der(6)t(1;6)(q21;p22),del(13)(q13q31)[30]
WRGL-2	M	56	17	M0303940	01/05/2003	Pancytopenia of unknown cause	
				M0306568	21/07/2003	Pancytopenia ?MDS	
				M0402451	11/03/2004	AML (secondary to MDS) x 3 courses of chemo, drop in neutrophils, ?relapse.	
				M0407384	10/08/2004	MDS/AML	
				M0412039	29/12/2004	AML	46,XY,der(6)t(1;6)(q21;p23)[19]/46,XY[1]
				M0502962	30/03/2005	AML relapsed, post first re-induction chemo. Clonal evolution at relapse ?remission	46,XY[20] No evidence of der(6)t(1;6) in 30 cells
				M0511012	15/11/2005	Relapsed AML	46,XY[60]
				M0512048	14/12/2005	AML post therapy, ?remission	46,XY,der(6)t(1;6)(q21;p23)[1]/46,XY[59]
				M0601674	20/02/2006	D+69 post MUD allograft for relapsed AML, not in CR at transplant ?any evidence of disease now.	46,XY[20] No evidence of der(6)t(1;6) in 30 cells
				W1804633	03/04/2018	VUD allo (male donor) Jan 2006 for secondary AML (previous MDS). Recent fall in counts ?cause.	46,XY,der(6)t(1;6)(q21;p23)[4]/46,XY[16]
				W1804729	29/03/2018	VUD allo (male donor) in Jan 2006 for secondary AML (previous MDS). Recent fall in counts ?cause/?secondary MDS.	Not done
				W1808390	07/06/2018	VUD allo 2006 for ref? AML, recent falling counts	46,XY,der(6)t(1;6)(q21;p23)[10]/46,XY[20]

						with recurrence of CGN clone. ?Progression.	
				W1814833	04/10/2018	VUD allo 2006 for AML; dysplastic relapse 06/18- post Aza/DL1 #3 -> response assessment.	46,XY,der(6)t(1;6)[3]/46,XY [17]
				W1900030	02/01/2019	VUD allo in 2006 for AML; dysplastic relapse in 2018, now post Aza x 6 + DLI x 7; response assessment.	46,XY,der(6)t(1;6)(q21;p23)[3]/46,XY[17]
				W1905921	04/04/2019	VUD allo 2006 for 20 AML. Dysplastic relapse June 18- post Aza/DCl x9 ?response assessment.	Failed
				W1913390	31/07/2019	VUD allo 2006 for AML, dysplastic relapse June 2018, R/Aza/ DCLI x12, for response assessment.	Failed
				W1920043	13/11/2019	VUD allo for refractory AML 2006, MDS relapse 2018, Rx Aza/DLI x 12; follow up marrow – to assess response.	46,XY[20] No evidence of der(6)t(1;6) in 30 cells (male BMT donor so origin of cells unknown)
WRGL-3	M	75	1	M0000317	04/01/2000	Myeloproliferative disorder diagnosed 1982. Rapid decline in counts in last 3/12. Blasts in PB. ?Transforming to AML.	46,XY,t(13;17)(q32;q21),del(20)(q?11q?13)[4]/46,idem,der(6)t(1;6)(q2?3;p2?2),-7,t(8;12)(q21;q21),der(14)t(7;14)(q11;q32),t(16;19)(q12;q13),der(18)t(18;?)(q11;?),+der(?)t(?;18)(?;q11)[17]
WRGL-4	F	65	2	M9502734	12/06/1995	MF FOR 2YRS. LOW PLTS, BRUISING & MOUTH ULCERS	48,XX,+8,+9 (cell numbers unknown)
				M0000641	25/01/2000	Myelofibrosis. Previous +8,+9. ?Progression.	48,XX,+8,+9[1]/48,XX,der(6)t(1;6)(q21;p22),+8,+9[18]/46,XX[1]
WRGL-5	F	70	3	M9603183	25/04/1996	Splenomegaly, raised WBC, myelocytes. ?Myelofibrosis ?CML.	46,XX[15]
				M9809001	18/11/1998	Myelofibrosis now transforming to ?AML.	46,XX[30]
				M0204638	21/05/2002	MF transformed to AML	46,XX,der(6)t(1;6)(q21;p21),del(20)(q?11q?13),add(21)(q22)[20]
WRGL-6	F	79	1	M0606634*	30/06/2006	MDS diagnosed 2001 ?AML transformation, blasts, decreased Hb and Plt	46,XX,der(6)t(1;6)(q21;p25)[9]/46,XX[11]
WRGL-7	F	77	1	M0612344	28/11/2006	MDS ?evolving to AML. Patient had splenectomy.	46,XX,der(6)t(1;6)(q25;p23)[17]/46,XX[3]
WRGL-8	F	53	1	M0810021	12/08/2008	Blast cells, ?AML/MDS. Subsequent information: markers indicate pre-B ALL. Previous	46,XX,der(7)t(1;7)(q21;p22)[1]/46,XX,der(6)t(1;6)(q21;p2?5)[1]/46,XX,der(16)t(1;16)(q21;q2?4)[1]/46,XX[17

						chemotherapy for breast cancer in 2001.	]
WRGL-9	F	79	1	M9909657	23/12/1999	?MDS, previous treatment for NHL	46,XX,der(6)t(1;6)(q11;p23)[11]/46,XX[19]

7.15 Appendix 15 Mutation analysis results from WGS data from der(6)t(1;6) cohort

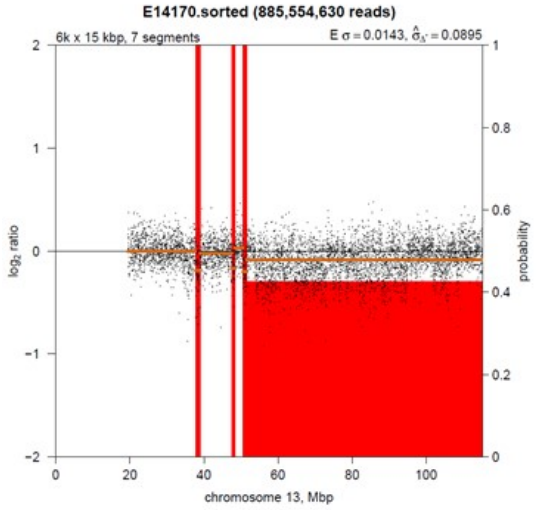
Rare variants detected by WGS. Table does not include those variants excluded due to low read depth (total or alt read depth).

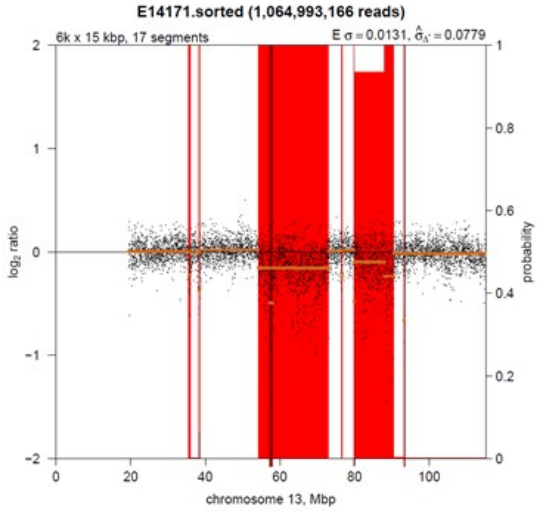
Sample	Gene	Transcript	c.	p.	Classification	Seen by TSMP
E14170	*ASXL2*	NM_018263	c.553A>G	p.S185G	Likely benign	No
E14170	*CREBBP*	NM_001079846	c.383C>G	p.S128C	Likely benign	No
E14170	*MPL*	NM_005373	c.1771T>C	p.Y591H	VUS	No
E14170	*JAK2*	NM_004972	c.1849G>T	p.V617F	Pathogenic	Yes
E14171	*KMT2C*	NM_170606	c.3955G>C	p.D1319H	Likely benign	No
E14171	*SAMD9*	NM_001193307	c.4666G>A	p.A1556T	Likely benign	No
E14171	*CSMD1*	NM_033225	c.4867+10C>T		Likely benign	No
E14171	*CTCF*	NM_001191022	c.38G>A	p.R13H	VUS	No
E14171	*CSF3R*	NM_156039.3	c.2384T>G	p.L795R	VUS	Yes
E14171	*NRAS*	NM_002524	c.190T>G	p.Y64D	VUS	Yes
E14171	*TET2*	NM_001127208	c.3728_3729insACT	p.K1243delinsKL	VUS	Yes
E14171	*DDX41*	NM_016222	c.1098+11delC		VUS	No
E14171	*NF1*	NM_000267	c.4514+19G>A		VUS	No
E14171	*KDM5A*	NM_001042603	c.1415A>G	p.Y472C	VUS	No
E14171	*KMT2C*	NM_170606	c.3499+1G>T		Likely pathogenic	No
E14171	*BRCC3*	NM_001018055	c.359G>A	p.W120X	Likely pathogenic	No
E14171	*SMC3*	NM_005445	c.2535+1G>A		Likely pathogenic	No
E14171	*TET2*	NM_001127208	c.5455delT	p.L1819X	Likely pathogenic	Yes
E14171	*SRSF2*	NM_001195427	c.284_307del	p.95_103del	Pathogenic	Yes
E14173	*CUX1*	NM_001202544	c.1633-14del		Likely benign	No
E14173	*STAT3*	NM_003150	c.341G>A	p.R114H	VUS	No
E14173	*JAK2*	NM_004972	c.G1849G>T	p.V617F	Pathogenic	Yes
E14211	*SRP72*	NM_006947	c.1640+5insA		Likely benign	No
E14211	*KMT2D*	NM_003482	c.10993C>G	p.P3665A	Likely benign	No
E14211	*CUX1*	NM_001202544	c.1633-14del		Likely benign	No
E14211	*ASXL1*	NM_015338	c.4183C>G	p.L1395V	Likely benign	Yes
E14211	*KMT2D*	NM_003482	c.6742C>T	p.R2248C	VUS	No
E14211	*CBL*	NM_005188	c.1243G>A	p.G415S	VUS	Yes
E14211	*PRPF40B*	NM_012272	c.2456C>T	p.S819F	Likely pathogenic	No
E14211	*CALR*	NM_004343	c.1104_1137del	p.K368fs	Pathogenic	Yes
M0000267	*CUX1*	NM_001202544	c.1633-14del		Likely benign	No
M0000267	*CALR*	NM_004343	c.1092_1143del	p.E364fs	Pathogenic	Yes
M0606634	*ATM*	NM_000051	c.1066-6T>G		Likely benign	No
M0606634	*GPRC5A*	NM_003979	c.544A>G	p.T182A	Likely benign	No
M0606634	*JAK3*	NM_000215	c.452C>G	p.P151R	Likely benign	No

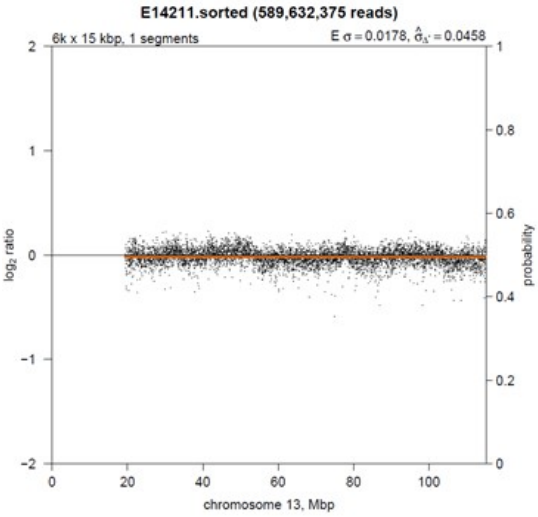
M0606634	*KMT2E*	NM_018682	c.5350C>T	p.P1784S	Likely benign	**No**
M0606634	*DIS3*	NM_001128226	c.231T>G	p.I77M	VUS	**No**
M0606634	*RUNX1*	NM_001001890	c.222_223insTG	p.L75fs	Likely pathogenic	**Yes**
M0606634	*TET2*	NM_001127208	c.3571C>T	p.Q1191X	Likely pathogenic	**Yes**
M0606634	*TET2*	NM_001127208	c.2862G>A	p.W954X	Likely pathogenic	**Yes**
M0606634	*SRSF2*	NM_001195427	c.284C>T	p.P95L	Pathogenic	**Yes**

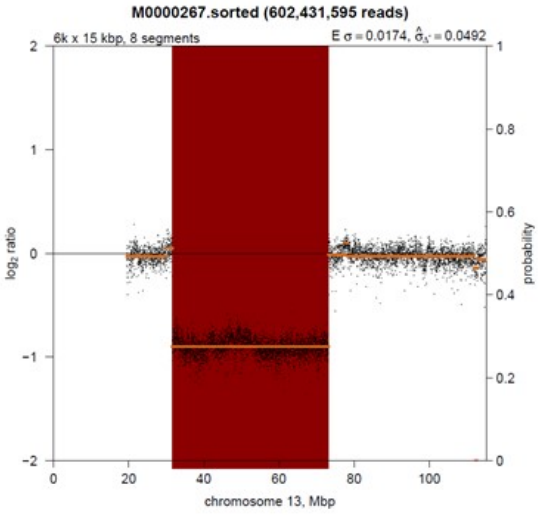
7.16 Appendix 16– WGS results for copy number changes on chromosome 13

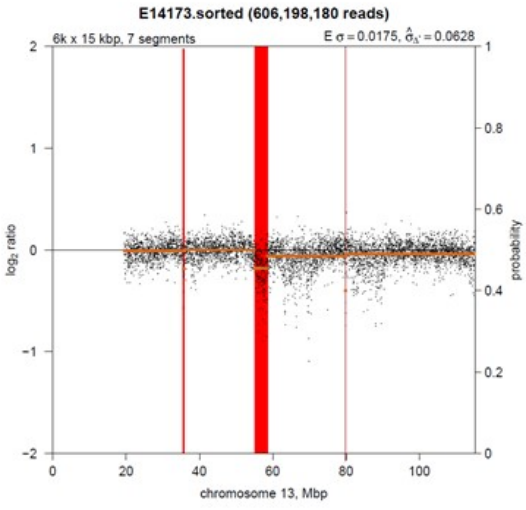
Segmental regions of possible loss are highlight in bright red in 3 samples. Only one sample (M0000267) showed a definite loss of 13q (highlighted in dark red).

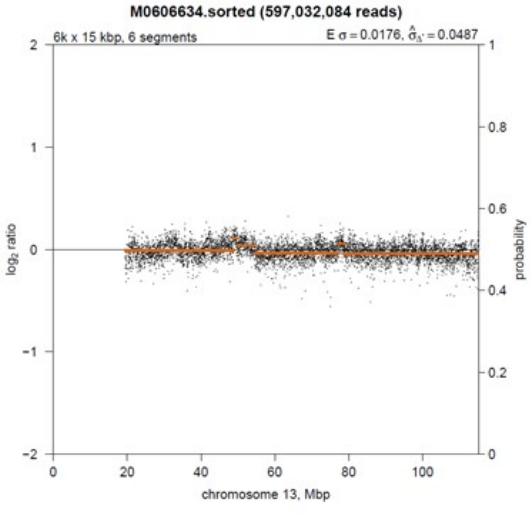












7.17 Appendix 17 - Details of additional work undertaken for fulfilment of DClinSci requirements

Additional work undertaken to meet the fulfilments of the DClinSci program includes:

1. Successful completion of the C1: Innovation project (Title: *TP53* acquired mutation analysis by next generation sequencing).
2. Successful completion of A units delivered by Alliance Manchester Business School (AMBS), as detailed below
3. Successful completion of the Part 1 (written and practical) and Part 2 (oral) exams under the specialism of Clinical Cytogenetics held by the Royal College of Pathologists. The date of completion of Part 2 oral examination was 30th September 2020.

A Units and C1 Credits for Appendix to DClinSci Thesis

Alliance Manchester Business School (AMBS) **A Units**		
Unit Title	**Credits**	**Assignment Word Count**
A1: Professionalism and Professional Development in the Healthcare Environment	30	Practice Paper – 2000 words A1 – Assignment 1 – 1500 words A1 – Assignment 2 – 4000 words
A2: Theoretical Foundations of Leadership	20	A2 – Assignment 1 – 3000 words A2 – Assignment 2 – 3000 words
A3: Personal and Professional Development to Enhance Performance	30	A3 – Assignment 1 – 1500 words A3 – Assignment 2 – 4000 words
A4: Leadership and Quality Improvement in the Clinical and Scientific Environment	20	A4 – Assignment 1 – 3000 words A4 – Assignment 2 – 3000 words
A5: Research and Innovation in Health and Social care	20	A5 – Assignment 1 – 3000 words A5 – Assignment 2 – 3000 words