Genetic lineage tracing in clonal haematopoiesis and myeloproliferative neoplasms



Nikolaos Sousos

Keble College

DPhil in Medical Sciences

Supervisors: Prof. Adam J. Mead, Dr. Andrew J. Peniket

Thesis Committee: Assoc. Prof. Anne Goriely, Prof. Paresh Vyas

Acknowledgments

Firstly, I would like to thank my supervisor Professor Adam J. Mead, for welcoming me into his research group and providing guidance throughout this project. I would also like to thank my co-supervisor Dr Andrew J. Peniket, and my Thesis Committee, Professors Anne Goriely and Paresh Vyas, for their valuable feedback. Special thanks to Associate Professor Bethan Psaila, for the encouragement and advice throughout my clinical research fellowship.

I would like to thank all lab colleagues for their day-to-day support in conducting my DPhil project research and all external collaborators for their significant contributions. I would especially like to thank Ms Gemma Buck for introducing me to the basic lab techniques, which served as a foundation upon which I could later develop and apply the methods and techniques I used in my project. I would also like to express my gratitude to Dr Sally-Ann Clark (Flow Cytometry Specialist) for all her assistance with flow cytometry analysis and fluorescence-activated cell sorting. Special thanks to Dr Maire Ní Leathlobhair and Professor David C. Wedge as well as Dr Wei Xiong (Sean) Wen, who conducted all bioinformatic analyses for the twin study and targeted genetic analysis of platelet lineage cells, respectively. I am also grateful to Dr Angela Hamblin for her significant contribution in the analysis of the custom clonal haematopoiesis panel next-generation sequencing results. I would also like to thank all my colleagues in hospital and of course all patients included in the project. This work would not have been possible without the generous support from Cancer Research UK Oxford Centre. Lastly (but most importantly) I would like to thank my family for their continuous, unconditional support to all my endeavours.

It should be mentioned that Chapter 1—Introduction—/section 1.3.5 Recent advances in the understanding of developmental timing and evolution of myeloproliferative neoplasms,

was adapted from Sousos et al. The Hematologist: ASH News and Reports. 2023;20(1). doi: 10.1182/hem.V20.1.2023112.

Furthermore, parts of Chapter 2—Materials and Methods—, and Chapter 3—Genetic lineage tracing in monozygotic twins with adult-onset myeloproliferative neoplasms—were adapted from Sousos, Ní Leathlobhair et al. Nat Med. 2022;28(6). doi: 10.1038/s41591-022-01793-4.

Lastly, paragraphs 2.13 Next-generation sequencing/2.13.2 *Custom next-generation sequencing panel*, and 2.17 Bioinformatic analysis/2.17.1.2 *Custom next-generation sequencing panel*, in Chapter 2, were adapted from Khan et al. Cancer Discov. 2023;13(2). doi: 10.1158/2159-8290.cd-22-0199.

I declare that all experiments described in the thesis were designed, performed, and analysed by myself. Contributions of colleagues for certain experiments are acknowledged in the relevant sections.

Table of Contents

Acknowledgments ..... *iii*

Table of Contents ..... *v*

Abstract *1*

List of abbreviations *3*

Chapter 1 Introduction *5*

1.1 Haematopoiesis **5**

1.1.1 The reformed model of haematopoiesis 5

1.1.2 Haematopoiesis and ageing 7

1.1.3 Abnormal megakaryopoiesis in human myeloproliferative neoplasms 8

1.2 Clonal haematopoiesis **9**

1.2.1 Definition of clonal haematopoiesis 9

1.2.2 Distinction between myeloid clonal haematopoiesis and lymphoid clonal haematopoiesis 10

1.2.3 Mutational landscape of clonal haematopoiesis 11

1.2.4 Prevalence of clonal haematopoiesis 12

1.2.5 Role of clonal haematopoiesis in human aging and disease 14

1.2.5.1 Clonal haematopoiesis and venous thromboembolism 16

1.2.5.2 Clonal haematopoiesis and progression to myeloid malignancy 18

1.3 Myeloproliferative neoplasms **23**

1.3.1 World Health Organization (WHO) classification of myeloproliferative neoplasms 23

1.3.2 *BCR-ABL1*-negative myeloproliferative neoplasms and classical myeloproliferative neoplasms 24

1.3.3 Molecular landscape of classical myeloproliferative neoplasms 26

1.3.4 Developmental timing and evolution of myeloproliferative neoplasms – insights from twin studies 30

1.3.5 Recent advances in the understanding of developmental timing and evolution of myeloproliferative neoplasms 31

1.3.6 Overview of the treatment of classical myeloproliferative neoplasms 33

1.3.7 Evidence supporting early intervention treatment strategies in classical myeloproliferative neoplasms....34

1.4 Thesis aims and hypotheses....39

Chapter 2 Materials and Methods....41

2.1 Study participants....41

2.1.1 Cohort design....41

2.1.1.1 Twin study....41

2.1.1.2 Targeted genetic analysis of platelet lineage cells....42

2.1.2 Informed consent....44

2.1.2.1 Twin study....45

2.1.2.2 Targeted genetic analysis of platelet lineage cells....45

2.1.3 Sample handling and storage....45

2.2 Cell isolation....46

2.2.1 Platelet isolation protocol....46

2.2.2 Platelet isolation method optimisation....48

2.2.3 Platelet purity studies....49

2.2.3.1 Cytocentrifugation analysis....49

2.2.3.2 Gene expression analysis....49

2.2.3.3 Flow cytometry analysis....50

2.2.4 Platelet recovery/survival studies....50

2.3 Cell lines....51

2.4 Microscopy....51

2.5 Nucleic acid extraction....52

2.5.1 DNA extraction....52

2.5.2 RNA extraction....52

2.5.3 cDNA synthesis....52

2.6 Nucleic acid quantification....52

2.7 Whole genome amplification....53

2.8 Polymerase chain reaction (PCR) 53

2.9 Real-time quantitative PCR 54

2.10 Droplet digital PCR 54

2.10.1 Method overview 54

2.10.2 *Custom JAK2V617F* assays for parallel gDNA and cDNA analysis 55

2.10.2.1 Sample preparation 55

2.10.2.2 Droplet digital PCR instrumentation and configurations 57

2.11 Electrophoresis 59

2.11.1 Agarose gel electrophoresis 59

2.11.2 Automated electrophoresis 59

2.12 Sanger DNA Sequencing 60

2.13 Targeted next-generation sequencing 60

2.13.1 Extended next-generation sequencing myeloid-gene panel 60

2.13.1.1 Panel characteristics 60

2.13.1.2 Library preparation 60

2.13.1.3 Sequencing platform 61

2.13.2 Custom next-generation sequencing panel 61

2.13.2.1 Panel characteristics 61

2.13.2.2 Library preparation 63

2.13.2.3 Sequencing platform 64

2.14 Whole genome sequencing 64

2.14.1 Library preparation and sequencing platform 64

2.15 Flow cytometry and cell sorting 64

2.15.1 Sample preparation for flow cytometry 64

2.15.2 Flow cytometry antibody panel characteristics 65

2.15.3 Flow cytometry instrumentation and configurations 67

2.15.3.1 Twin study 67

2.15.3.2 Targeted genetic analysis of platelet lineage cells 67

2.15.4 Surface marker definitions of haematopoietic stem and progenitor cells and mature blood compartments 67

2.16 Single-cell cloning assay 68

2.17 Bioinformatic analysis 69

2.17.1 Extended next-generation sequencing myeloid-gene panel analysis 69

2.17.2 Custom next-generation sequencing panel analysis 69

2.17.2.1 Preprocessing of DNA- and RNA-sequencing data 69

2.17.2.2 Variant detection, annotation, and filtering 70

2.17.2.3 Next-generation sequencing data analysis and visualisation 71

2.17.3 Whole genome sequencing analysis 71

2.17.3.1 Whole genome data processing and alignment 71

2.17.3.2 Variant calling 72

2.17.3.3 Copy-number estimation 72

2.17.3.4 Mutation clustering 73

2.17.3.5 Structural variation analysis 74

2.17.3.6 Extraction of mutational signatures 75

2.17.3.7 Putative driver mutation analyses 75

2.17.3.8 Bayesian estimates of myeloproliferative neoplasm developmental timing 75

2.17.4 Single-cell index-sorting 79

2.18 Statistical analysis 79

Chapter 3 Genetic lineage tracing in monozygotic twins with adult-onset myeloproliferative neoplasms 81

3.1 Proof of concept: a study of twins with *CALR*-mutant myelofibrosis 81

3.1.1 Results 81

3.1.1.1 Clinical findings 81

3.1.1.2 Cell lineage specific mutation analysis 83

3.1.1.3 In utero origin of myeloproliferative neoplasm 85

3.1.1.4 Phenotypic and genetic analysis of haematopoietic stem and progenitor cells 94

3.2 Extension project: Investigation of the molecular status at birth of patients diagnosed with myeloproliferative neoplasms as children or young adults 97

3.3 Discussion 99

Chapter 4 JAK2 mutational analysis of platelet lineage cells in clonal haematopoiesis and myeloproliferative neoplasms 103

4.1 Characteristics of the cohorts studied 103

4.1.1 Haematologically normal elderly individuals 103

4.1.2 Unprovoked venous thromboembolism patients 104

4.1.3 Triple negative and low-variant-allele-frequency *JAK2V617F* myeloproliferative neoplasm patients 105

4.2 Platelet isolation studies 107

4.2.1 Platelet isolation methodology 107

4.2.2 Platelet purity studies 109

4.2.3 Platelet recovery survival studies 111

4.2.4 Platelet-derived nucleid acid quality control 112

4.3 Droplet digital PCR lineage tracing for *JAK2V617F* 113

4.3.1 Assay performance studies 113

4.3.2 Definition of *JAK2V617F* positivity and biased/restricted clonal haematopoiesis 116

4.3.3 Results of droplet digital PCR lineage tracing for *JAK2V617F* 116

4.3.4 Triple negative and low-variant-allele-frequency *JAK2V617F* myeloproliferative neoplasm patients 125

4.4 Discussion 127

Chapter 5 Targeted next-generation sequencing analysis of platelet lineage cells in clonal haematopoiesis and myeloproliferative neoplasms 133

5.1 Introduction 133

5.2 Assay performance studies 134

5.3 Detection and quantification of additional clonal haematopoiesis associated mutations through targeted next-generation sequencing 137

5.4 Discussion 151

Chapter 6 Conclusions and future directions *155*

6.1 Summary of findings **155**

6.1.1 Genetic lineage tracing in monozygotic twins with adult-onset myeloproliferative neoplasms 155

6.1.2 Mutational analysis of platelet lineage cells in clonal haematopoiesis and myeloproliferative neoplasms 156

6.2 Implications of this work **158**

6.3 Thesis limitations **159**

6.4 Future directions **160**

6.5 Conclusion **163**

Appendix A Supplementary Material and Methods *165*

A.1 Blood product isolation protocol **165**

A.1.1 Equipment and Reagents 165

A.1.1.1 Reagents 165

A.1.1.2 Equipment 166

A.1.2 Method 166

A.1.2.1 Plasma isolation 166

A.1.2.2 Platelet isolation 167

A.1.2.3 MNC, GRA, T cell and/or HSPC isolation 167

Appendix B Supplementary Tables *175*

B.1 Catalogue of the genes targeted in the extended next-generation sequencing myeloid-gene panel **175**

B.2 Summary of multidimensional mutation clustering **177**

B.3 Additional candidate cancer genes **197**

B.4 Catalogue of all somatic substitutions and indels in twin A and twin B **199**

B.5 Catalogue of reagents **233**

Bibliography *235*

Abstract

Clonal haematopoiesis (CH) is a common age-related condition in which somatic mutations accumulate in haematopoietic stem and progenitor cells (HSPCs), leading to the clonal expansion of mutated cells. This process has been causatively linked with atherosclerotic cardiovascular disease and related complications, while it has also been associated with an increased risk of development of myeloid neoplasms, including myeloproliferative neoplasms (MPNs). This DPhil thesis aimed to investigate the pathogenesis of CH and MPNs utilising genetic lineage tracing.

I initially focused on the study of granulocytes (as the most abundant type of myeloid cells) against germline DNA through a twin study in a pair of monozygotic twins who presented with adult-onset *CALR* mutant MPN. I aimed to investigate the origin of their premalignant clones and their subsequent molecular evolution. Lineage tracing of the MPN clone by whole genome sequencing revealed a common clonal origin of the MPN clone, which occurred *in utero* followed by twin-to-twin transplacental transmission and subsequent similar disease latency.

In the next part of the project, I focused on the study of granulocytes against platelets as a surrogate of the megakaryocyte lineage. The aim of this part of my work was to assess whether platelet-biased CH exists in humans and determine its prevalence in normal individuals and myeloid neoplasms. Targeted detection of *JAK2V617F* and other mutations commonly involved in CH and MPN revealed the presence of platelet-biased/restricted CH, accounting for a significant proportion of the overall prevalence of CH.

In conclusion, this DPhil thesis highlights the importance of genetic lineage tracing in CH and MPNs and provides insight into their molecular pathogenesis. It introduces a novel model of MPN pathogenesis, involving a prolonged disease latency between the acquisition of the premalignant MPN clone and the development of established disease, thus opening the potential for early diagnosis and preventive precision treatment strategies. It also describes for the first time presence of platelet-biased CH with direct clinical implications in the detection of CH and relevance within the wider context of diagnosis and monitoring of haematological disorders.

List of abbreviations

Abbreviation	Expansion
ACD	Acid citrate dextrose
Allo-HSCT	Allogeneic stem cell transplantation
AML	Acute myeloid leukaemia
ARCH	Age-related clonal haematopoiesis
ARMS	Amplification refractory mutation system
CH	Clonal haematopoiesis
CHIP	Clonal haematopoiesis of indeterminate potential
CVD	Cardiovascular disease
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DVT	Deep vein thrombosis
EDTA	Ethylenediaminetetraacetic acid
ET	Essential thrombocythaemia
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
GRA	Granulocyte
HSC	Haematopoietic stem cell
HSCT	Haematopoietic stem cell transplantation
HSPC	Haematopoietic stem and progenitor cell
IL	Interleukin
IMDM	Iscove's Modified Dulbecco's Medium
JAKi	JAK inhibitor
L-CH	Lymphoid clonal haematopoiesis
LN2	Liquid nitrogen
M-CH	Myeloid clonal haematopoiesis
MF	Myelofibrosis
MNC	Mononuclear cell

MPN	Myeloproliferative neoplasm
NGS	Next-generation sequencing
NHS	National Health Service
NIHR	National Institute of Health and Care Research
OHTC	Oxford Haemophilia and Thrombosis Centre
OUH	Oxford University Hospitals
ORB	Oxford Research Biobank
PB	Peripheral blood
PBS	Phosphate-buffered saline
PE	Pulmonary embolism
PLA	Plasma
PLT	Platelet
PET-MF	Post essential thrombocythaemia myelofibrosis
PMF	Primary myelofibrosis
PPV-MF	Post polycythaemia vera myelofibrosis
PV	Polycythaemia vera
RBC	Red blood cell
SNV	Single-nucleotide variant
sMF	Secondary myelofibrosis
TN	Triple negative
uVTE	Unprovoked venous thromboembolism
VAF	Variant allele frequency
VTE	Venous thromboembolism
WGA	Whole genome amplification
WGS	Whole genome sequencing
WHO	World Health Organization

Chapter 1 Introduction

1.1 Haematopoiesis

Haematopoiesis is the formation of blood cellular components; it is a stem cell hierarchy originating with haematopoietic stem cells (HSCs), which are rare and highly potent cells, that undergo a process of cell divisions allowing self-renewal to maintain the HSC pool, and gradual lineage restriction and further differentiation to result in functionally diverse blood cell populations.

1.1.1 The reformed model of haematopoiesis

The classical model of haematopoiesis suggests that multipotent HSCs reside at the apex of the haematopoietic differentiation tree hierarchy and maintain life-long blood production via gradual lineage restriction obtained through an ordered series of branching steps(1). However, recent advances in single-cell omics have revealed substantial cellular heterogeneity within previously distinct haematological stem and progenitor cell (HSPC) populations(2), highlighting that bulk analysis can only provide a coarse-grain image of the heterogeneity of lineage potential and bias within HSCs. Single-cell resolution has shown that priming to different cell fate choices is present from the earliest HSCs(3-5) and cell differentiation occurs over a continuum of gradually increasing priming towards the specific lineage alongside gradual loss of multipotency(1, 2, 4)(Figure 1-1). In this model, lineage-primed and committed unipotent cells exist within the HSC compartment.

Interestingly, mouse haematopoiesis studies have shown presence of HSCs that are megakaryocyte (Mk)-biased, i.e., committed to Mk differentiation, bypassing the megakaryocyte-erythroid bipotent progenitor(6-12). These HSCs are transcriptionally

primed towards megakaryocytic differentiation, though retaining multipotency. Moreover, a distinct pathway to Mk production directly from HSCs has also been described(3, 13) (Figure 1-1). The exact role of Mk-biased HSCs and their contribution to normal haematopoiesis is unclear. Given platelets' important role in haemostasis, inflammation, and immunity (alongside several other biological processes), their short lifespan, rapid recruitment and extensive consumption at the site of an acute event, a readily available source of platelet reserve could be biologically advantageous(14, 15). Indeed, Haas et al.(15) showed that acute inflammation in mice induced upregulation of post-transcriptional protein synthesis in those "stem-like Mk progenitors", driving cell cycle activation and maturation, which in turn resulted in effective replenishment of platelet numbers during the acute event. Notably, during homeostasis the Mk-biased HSCs remained quiescent though transcriptionally primed. Thus, platelet-biased HSCs may be an efficient means of rapid replenishment of the platelet pool in stress haematopoiesis.

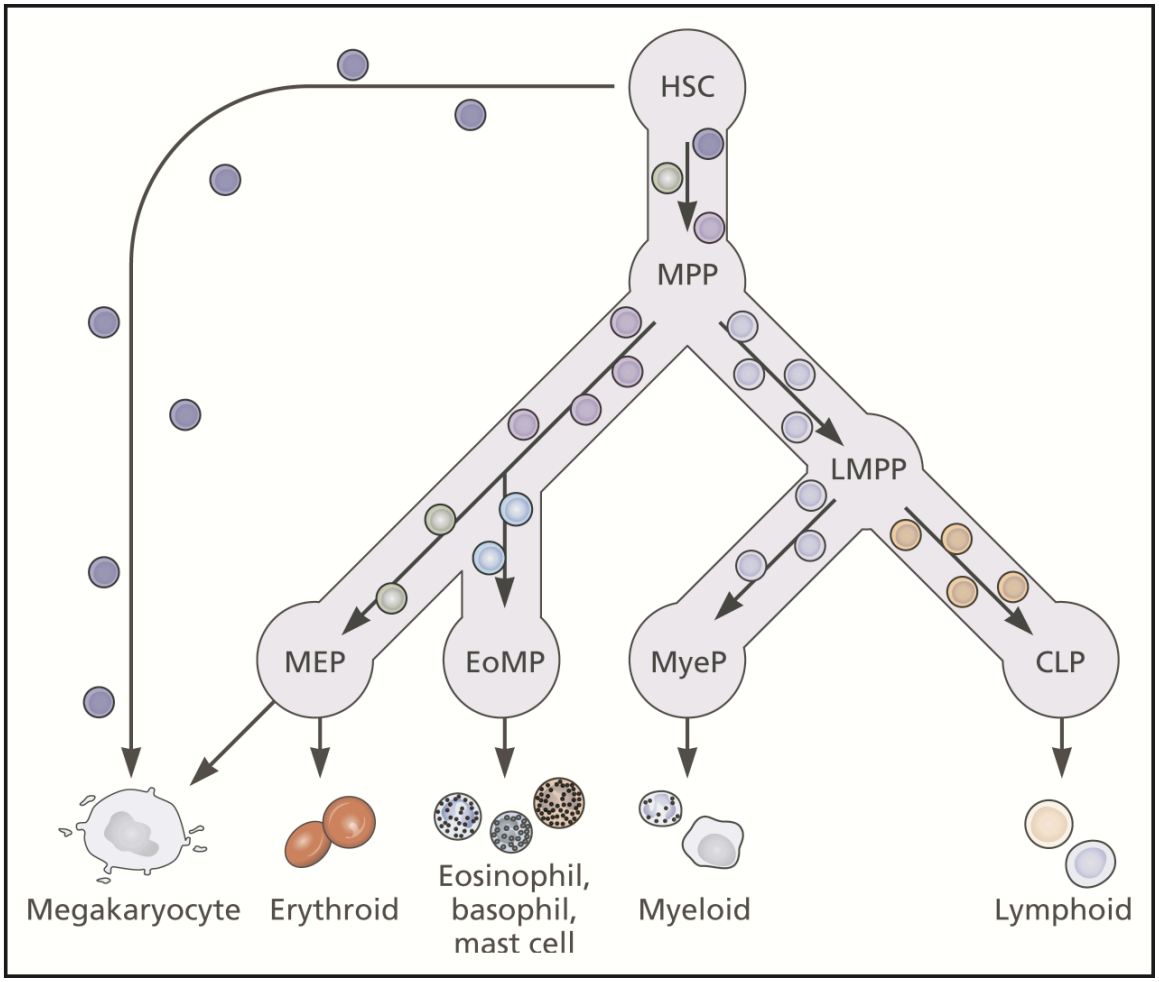


Figure 1-1. The current model of human haematopoiesis. *Transition from HSCs to unipotent cells occurs over a continuum of gradual increasing priming towards a specific lineage in parallel with the gradual loss of multipotency across cell differentiation. Importantly, accumulating evidence in both mice and humans support the presence of Mk-biased HSCs, committed to Mk differentiation, though retaining multipotency, as well as Mks arising directly from HSCs. CLP, common lymphoid progenitor; HSC, haematopoietic stem cell; EoMP, eosinophil-basophil-mast cell progenitor; LMPP, lymphoid-primed multipotent progenitor; MEP, megakaryocyte-erythroid progenitor; MPP, multipotent progenitor; MyeP, myeloid progenitor. Reprinted from Blood, 133(13), Psaila and Mead, Single-cell approaches reveal novel cellular pathways for Mk and erythroid differentiation, 1427-35, 2019; https://doi.org/10.1182/blood-2018-11-835371 (ref. (2)); with permission from The American Society of Hematology.*

1.1.2 Haematopoiesis and ageing

As humans age significant changes occur in HSCs as a result of cell-intrinsic changes, microenvironmental alterations and systemic aging that ultimately cause stem cell exhaustion(16, 17). These changes involve their overall cell number and composition as

well as senescence, lineage bias, and function(18). Aged human HSCs have increased cell division and proliferation activity, although functionality is overall impaired. In view of lineage bias and cell composition, bone marrow of aged individuals shows a skewed myeloid-to-lymphoid ratio, with predominance of myeloid- and Mk-biased HSCs, mostly due to a decrease in the number of lymphoid progenitors(19-21). This decrease contributes to the observed age-related lymphopenia and decreased immunity in older ages(18, 22, 23). Importantly, this skewing might also provide a rationale behind the increased incidence of myeloid malignancies in the elderly (compared to both young age and lymphoid malignancies in the age-matched population)(24). Interestingly, aged mouse HSCs show increased expression of genes involved in growth and proliferation as well as genes associated with the development of myeloid malignancies(25). Moreover, increased interleukin (IL)-1β-mediated inflammation and impaired phagocytic function in aged mouse marrow expands Mk/platelet-biased HSCs(26). Altogether, aged HSCs contribute to the increased propensity towards the acquisition of CH and initiation of myeloid neoplasms (including myeloproliferative neoplasms [MPNs] presenting with increased numbers of mature cells of the myeloid and megakaryocytic lineage), both associated with increasing age.

1.1.3 Abnormal megakaryopoiesis in human myeloproliferative neoplasms

Megakaryocytes play a key role in the development and progression of MPNs, and they are considered to carry the same mutations as the other cells of myeloid lineage. However, due to the multiple rounds of endoreplication during megakaryopoiesis, they are susceptible to the acquisition of additional mutations. Guo et al.(27) showed that Mks from MPN patients carry unique somatic mutations. Interestingly, the detected variants involved variants in cell signalling genes (such as *JAK1*, *JAK2*, *SH2B3*, *BRAF*, *RAF1*, *TLR4*, and

FGFR1), some of them in high variant allele frequencies (VAFs) reaching 100%, being suggestive of a Mk-biased HSC origin.

Furthermore, most recently, Psaila, Wang et al.(28) focused on the investigation of the abnormal Mk differentiation and function in the setting of human myelofibrosis (MF). The authors provided evidence of an abnormal pathway for the differentiation of Mks from uncommitted HSPCs in MF haematopoiesis, in addition to expansion of the normal shared trajectory between the erythroid and Mk lineages, and increased proliferation of Mk progenitors.

Overall, the presence of Mk-/platelet-biased HSCs in human haematopoiesis could have important implications for normal haematopoiesis, aging, and disease, and requires further research, including phenotypic studies. In particular, it is reasonable to assume that this haematopoietic pathway would have a significant role in age-related clonal haematopoiesis (ARCH) and MPNs.

1.2 Clonal haematopoiesis

1.2.1 Definition of clonal haematopoiesis

Haematopoietic stem cells gradually acquire somatic mutations during life with approximately 17 genome-wide base substitutions occurring per year of life(29, 30). A small, well-defined subset of these mutations can provide mutant HSCs a selective advantage and result in preferential proliferation and clonal expansion of the mutant HSC and its progeny – a process broadly termed clonal haematopoiesis (CH)(31-33).

Acquisition of CH-associated mutations is an inevitable consequence of aging(34). The term “age-related clonal haematopoiesis” (ARCH) reflects that strong correlation between

CH prevalence and aging, and it is used to describe CH caused by any mutation caused by any type of acquired clonal event, in individuals with no diagnosis of haematological malignancy. Notably, the term that was initially used to describe CH was clonal haematopoiesis of indeterminate potential (CHIP) and was defined as CH caused by a mutation in a candidate driver gene with a VAF greater than 2%, always in the absence of blood count abnormalities(35, 36). However, it should be mentioned that the VAF cut-off of 2% was related to the technical limitations of next-generation sequencing (NGS) platforms at the time of initial CH studies; it has no clinical relevance and should possibly be reconsidered in relation to the risk of morbidity and mortality of CH at lower VAFs(36-38). The umbrella term of CH has recently been introduced as a framework to better reflect the stepwise expansion of the mutant clone, irrespective of the clinical outcome of the process. Henceforth CH will be used for all cases unless use of an alternative term is most appropriate.

1.2.2 Distinction between myeloid clonal haematopoiesis and lymphoid clonal haematopoiesis

Typically, CH mutations confer lineage bias on the mutated cells towards either myeloid or lymphoid lineages with distinct sequelae for each lineage(39, 40). Unlike myeloid CH (M-CH), lymphoid CH (L-CH) can occur both within the HSPC compartment and downstream differentiated precursors and mature lymphocytes(39). In all cases, CH mutational events consist of either somatic point mutations or rarely mosaic chromosomal alterations.

Myeloid CH is approximately five-fold more frequent than L-CH in individuals between 40 and 70 years of age, and point mutations account for up to 95% of the mutational

events(40). Within point mutations, events occurring in three certain genes, namely *DNMT3A*, *TET2*, and *ASXL1*, account for the vast majority of the cases(36, 40, 41).

On the other hand, mosaic chromosomal alterations are more frequent in L-CH, accounting for more than half of the mutational events(39, 40), while recurring point mutations account for approximately only 20% of L-CH variants. These point mutations involve mostly *DUSP22*, *FAT1*, *KMT2D*, *SYNE1*, and *ATM* genes.

Henceforth I will be focusing on M-CH, and therefore, CH will be used for all cases unless distinction between M-CH and L-CH is required.

1.2.3 Mutational landscape of clonal haematopoiesis

Clonal haematopoiesis involves mutations that have been associated with haematological malignancies as well as non-driver candidates. It predominantly affects genes involved in epigenetic regulation (*DNMT3A*, *TET2* and *ASXL1*), cytokine signalling (*JAK2, GNAS*), splicing (*SF3B1* and *SRSF2*), and DNA repair and cell cycle control (*TP53* and *PPM1D*)(31-33, 42, 43). Among them, *DNMT3A* and *TET2* mutations account for approximately 95% of the identified candidate driver mutations(36, 41).

Although there is no single hotspot in either gene, one third of the *DNMT3A*-associated mutations occurs at the arginine 882 (R882) residue (particularly NM_022552.5[DNMT3A]:c.2645G>A p.R882H)(31, 32). *DNMT3A* and *TET2* mutations that are found in CH are mainly loss-of-function mutations, and they include indels, frameshift, nonsense, splice site, and missense mutations which abrogate catalytic function. It is, however, unclear whether those genes are inherently hypermutable or there are specific factors affecting the HSC or its microenvironment leading to a hypermutable state. Mutations in different genes can result in different levels of fit advantage for the

mutant HSPCs and lead to different trajectories(37, 44). All clonal mutations arise in long-lived HSPCs(45). However, *DNMT3A* and *TET2* mutations have different lineage involvement. *DNMT3A* lineage restriction patterns are compatible with a multipotent stem cell origin, whereas *TET2* mutations occur in a more committed HSC, mainly in myeloid cells and sometimes in B-cells(46). Notably, *DNMT3A* mutations have no impact on blood counts or indices, while mutations in *TET2* have been associated with a modest neutropenic effect(41, 46).

1.2.4 Prevalence of clonal haematopoiesis

Prevalence of CH varies in different studies, depending on the age of the cohort and its clinical characteristics as well as the screened regions and sensitivity of the method that is used to detect CH(44). Clonal haematopoiesis of indeterminate potential shows an increased incidence mostly in elderly people of more than 50 to 60 years old(47-49). It is rarely detected in individuals under 40 years old, whereas it is detected in 10-20% of individuals older than 70 years. However, Young et al. using targeted error-corrected sequencing methods, which allowed the accurate detection of clonal mutations at 0.03% VAF, showed that *DNMT3A*- and *TET2*-variant-driven CH was detected in 95% of the healthy 50- to 60-year-old individuals studied (median VAF 0.24%)(45). Furthermore, a WGS approach by Zink et al.(34) confirmed a high prevalence of mutations in driver candidates and an even higher prevalence of CH involving non-driver candidates. Further studies have confirmed that some low-VAF CH is inevitable for almost all adults, and moreover, the clones can emerge earlier in life than initially thought, although the dynamics and clinical significance of these clones is yet to be fully elucidated(37, 44, 45, 50).

Notably, there is an age-related pattern for the kinetics of haematopoietic clones driven by different gene mutations(51, 52). McKerrell et al.(51) analysed 4,067 blood DNA samples from individuals aged 17 to 98 years for 15 known leukaemogenic mutations in nine genes (namely *DNMT3A*, *JAK2*, *SRSF2*, *SF3B1*, *IDH1*, *IDH2*, *KIT*, *KRAS*, *NRAS*, *NPM1*, and *FLT3*), and using a VAF detection threshold of 0.8%, they showed that although the overall prevalence of CH increased with age, *DNMT3AR882X* and *JAK2V617F* mutations were detectable from young age, while spliceosome gene mutations (*SRSF2P95* and *SF3B1* K666 / K700) were restricted to individuals aged 70 years or older. Acuna-Hidalgo et al.(52) confirmed that CH occurs throughout adult life in a population-based sample of 2,000 individuals aged 20 to 69 years, using a more extended panel and a VAF detection threshold of 0.1%. Notably, *JAK2* (alongside *DNMT3A*, *TET2*, *CBL*, *KRAS* and *NRAS*) was also among the driver genes with early adulthood representation.

Besides aging, our understanding of the factors that predispose to acquisition/progression of these clones is currently incomplete. Genetic predisposition(34, 41, 53), longer leukocyte telomere length(53), genotoxic exposure (including radiotherapy and cytotoxic chemotherapy, such as carboplatin, and topoisomerase II inhibitors(54-56), involving mutations almost specifically in *TP53*, and *PPM1D*, which provide mutant HSPCs with resistance to genotoxic stress and subsequently, promote their clonal expansion)(36), smoking(53, 57), inflammaging (a chronic low-grade inflammation that occurs with aging and is associated with age-related diseases; mostly related to *TET2* CH, conferring tumour necrosis factor (TNF)-a resistance to HSPCs, associated with myeloid skewing and resistance to apoptosis)(58), immune dysregulation (mostly related to mutations in *PIGA*, *BCOR/BCORL1*, *HLA* deletions in younger patients, plus *DNMT3A* and *ASXL1* but rarely *TET2* mutations in elderly individuals)(59) have all been associated with an increased risk of acquisition and progression of CH.

1.2.5 Role of clonal haematopoiesis in human aging and disease

Clonal haematopoiesis has been associated with a wide range of pathological states (including but not limited) to cardiovascular disease (CVD), haematological malignancy, late-onset forms of autoimmunity, altered immune function, and inflammaging) and increased all-cause mortality (Figure 1-2)(32, 39, 60-64).

Most recently, CH and particularly M-CH, has also been associated with solid tumours, crossing the borders of distinct haematopoietic lineages and solid tumour oncology. Accumulating evidence reveals high rates of CHIP in patients with solid tumours(54, 55, 57, 65-70), possibly contributing towards adverse outcomes(57, 65), while in a single study of metastatic colorectal cancer presence of CHIP (especially *DNMT3A*-CH) was associated with favourable prognosis(70). Prevalence of CHIP in the four largest solid tumour studies ranged between 14% and 65%(57, 68, 71, 72). However, in most studies, CH was assessed post-treatment, which, as described above, may have affected CHIP prevalence. Interestingly, certain cancers are associated with increased prevalence of CHIP. In treatment-naïve cohorts, cancers associated with higher prevalence of CHIP included non-small cell lung cancer, as well as breast, pancreas, and prostate cancers, while renal cell carcinoma and ovarian cancer were associated with the lowest prevalence(54, 69).

Other age-related conditions that have been associated with CH include chronic obstructive pulmonary disease(73), osteoporosis(74), and advanced chronic kidney disease(75). In all conditions CH has been associated with increased risk of disease and/or more severe disease, typically driven by the CH-imposed hyperinflammation. Interestingly, in Alzheimer disease the hyperinflammatory milieu was found to have a protective role, significantly reducing the risk of dementia ($OR = 0.64$; $p < .001$)(76).

Except for CVD, where a genuine causal relationship between CH and disease has been established, most of the aforementioned associations have been based solely on epidemiological studies, and whether CH has a true causal relationship with those conditions or not warrants further investigation. Moreover, association between presence of CH and increased disease-related mortality in some of these conditions might be indirect and not always easily discernible. Further research is also required to elucidate the interplay between CH and these conditions, its role as a prognostic biomarker, its interaction with certain treatments, and in turn, how CH can be incorporated into precision medicine strategies.

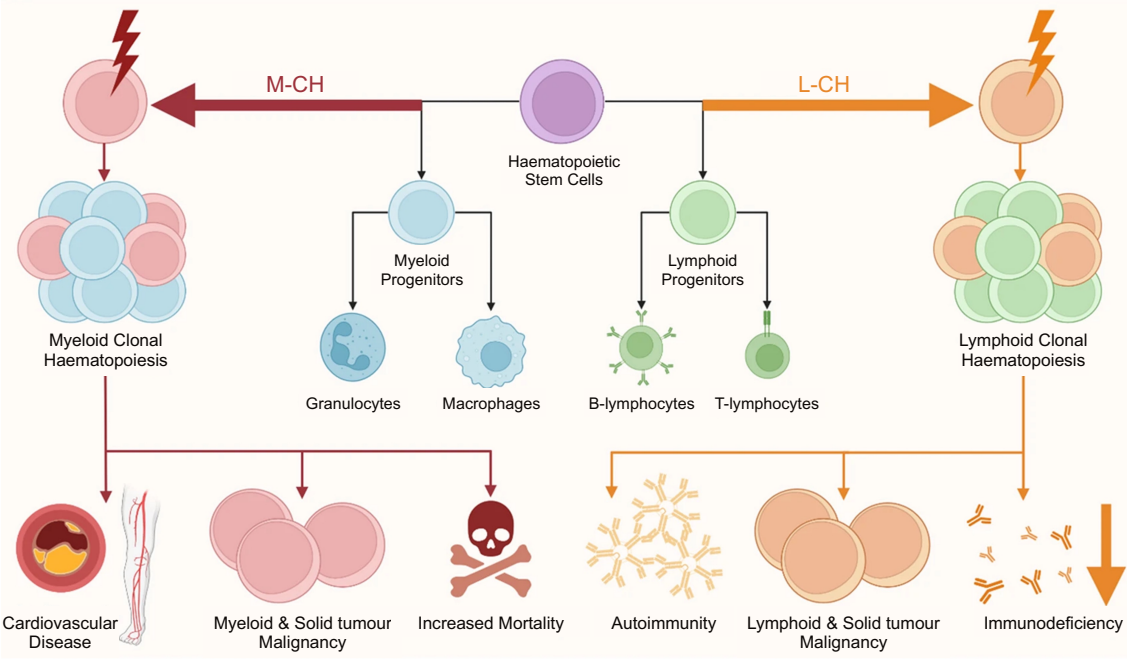


Figure 1-2. Schematic representation of CH clinical sequelae grouped by CH lineage bias. *M-CH has been associated with increased risk of cardiovascular disease, myeloid and solid tumour malignancy, and increased all-cause mortality, while L-CH has been associated with increased risk of late-onset autoimmune disease, lymphoid and solid tumour malignancy, and immunodeficiency. CH, clonal haematopoiesis; L-CH, lymphoid clonal haematopoiesis; M-CH, myeloid clonal haematopoiesis. Adapted from Blood Cancer J., 13(1), von Beck et al. Lymphoid clonal hematopoiesis: implications for malignancy, immunity, and treatment, 2023; https://doi.org/10.1038/s41408-022-00773-8 (ref. (39)); Creative Commons Attribution 4.0 International (CC BY 4.0).*

1.2.5.1 Clonal haematopoiesis and venous thromboembolism

The presence of CH confers an increased risk of all-cause mortality, beyond the level attributable to haematological cancers alone. Instead, the observed increased mortality has also been associated with atherosclerotic CVD and downstream complications, such as ischaemic heart failure, peripheral artery disease, and venous thromboembolism (VTE)(32, 60). Particularly, CHIP has been associated with an almost two-fold increased risk of coronary heart disease(60), while CH has also been associated with an adverse prognosis in chronic ischaemic heart failure(77-79). Clone size has been associated with a prognostic significance in this context too(77, 79). Assmus et al.(79) calculated that in such patients *DNMT3A*- and *TET2*-driven CH at a respective VAF of more than 1.15% and 0.73% was associated with increased 5-year mortality.

The underlying mechanisms in the pathogenesis of CH-associated CVD have not yet been fully elucidated, and they may vary depending on specific genes and/or mutations implicated. Multiple lines of evidence suggest that inflammation, and particularly a macrophage-driven proinflammatory cytokine and chemokine cascade, has a critical role in the pathophysiology of CVD in general. Accumulating evidence from both in vitro and in vivo studies have provided a mechanistic insight into CH-related CVD pathogenesis, showing that this proinflammatory state is further aggravated in presence of CH(60, 80, 81). Although *TET2* is the second most common gene affected in CH (only after *DNMT3A*), initial CVD studies have focused on *TET2*-related CH due to its strongest correlation with aging(82). These studies have collectively showed that *TET2*-related CH accelerates atherosclerosis through increased macrophage-induced NLP3 inflammasome activation associated with increased processing and release of proinflammatory cytokines such as IL-1β, IL-6, and IL-18 (ref. (60, 80)), alongside elevated concentrations of relevant biomarkers(83, 84). Importantly, Fuster et al.(80) showed that treatment with an NLRP3

inhibitor was able to reverse atherosclerosis in a *TET2*$^{-/-}$ CH mouse model. Subsequent studies have partially confirmed the above findings, while revealing distinct pathophysiological mechanisms depending on the driver gene. *DNMT3A*-CH was associated with higher expression of *NLP3* and higher levels of circulating IL-6 (ref. (83)), while *JAK2*-driven CH was shown to induce atherosclerosis mostly via the AIM2 inflammasome pathway, resulting in increased IL-1β and IL-18 release through gasdermin D pores, and increased pyroptotic cell death of *JAK2V617F* mutation-positive macrophages(85). Importantly, knockout of *AIM2* or *GSDMD* resulted in amelioration of atherosclerosis in *JAK2V671F* mouse models, while *NLRP3* deletion had only a minor effect. Moreover, massively parallel RNA sequencing of elderly individuals with CH (n = 6) versus controls revealed altered transcriptomic signatures in CH implicating neutrophil degranulation and dysregulation of the innate immunity(86). Notably, compared to other genes *JAK2V617*F-related CH has been associated with a 12-fold increase in the risk of coronary artery disease compared to the approximately 2-fold increase that has been estimated for the most common CH-genes, namely *TET2*, *DNMT3A*, and *ASXL1* (ref. (60, 87, 88)). Collectively, the above studies shed light on the pathophysiology of CH-associated atherosclerosis and demonstrate a proof of concept supporting the role of precision intervention to reduce the risk of atherosclerotic complications in selected high-risk patients.

In addition to coronary artery disease, *JAK2V617F* is also associated with a higher risk of atherothrombotic complications, particularly VTE, in both CH and MPNs(87, 89). Venous thromboembolism includes both deep vein thrombosis (DVT) and pulmonary embolism (PE) and refers to a blood clot that forms in a vein which partially or completely obstructs blood flow. In the absence of a known transient or persisting risk factor VTE is classified as unprovoked VTE (uVTE), accounting for approximately 50% of cases of VTE(90).

Notably, VTE has been associated with increased mortality risk, with a median survival as low as eight years, with patients with uVTE having an 1.4-fold higher mortality risk compared to other VTE subtypes(91). Unprovoked VTE has also been associated with occult cancer(92), while current cancer screening strategies following uVTE typically detect late-stage cancers and fail to detect 30% of cancers altogether(93, 94). Interestingly, in a large case-control study including 10,893 individuals, *JAK2V617F*-associated CHIP was present in 25% of patients with major thrombotic events, compared to only 4.4% of patients with CHIP carrying other mutations ($p = .02$), and 2.5% of patients where CHIP was not detected ($p < .001$)(87). The authors showed that *JAK2V617F*-induced venous thrombosis was driven by platelet-neutrophil crosstalk, resulting in increased NETosis, which in turn drove thrombosis in a PAD4-dependent manner. This study suggests a strong correlation between *JAK2V617F* CH and VTE and supports a potential role of *JAK2V617F*-driven CH as a risk factor for the development of uVTE.

Deeper understanding of the pathophysiological mechanisms in CH-associated CVD through mechanistic and omics studies, alongside genome-wide association and targeted mutation phenotypic studies, could open up new prospects for novel personalised medicine in the pandemic of chronic CVD and its complications, with early studies showing a potential role for anti-inflammatory therapy(85, 95-97).

1.2.5.2 Clonal haematopoiesis and progression to myeloid malignancy

Throughout the aging process, less and less HSCs contribute to haematopoietic output, although overall blood cell production remains largely stable(30, 32, 61). On the other hand, studies in mice have demonstrated that loss of function of either of the main genes involved in CH, *DNMT3A* or *TET2*, leads to increased HSC self-renewal capacity(98-101). Therefore, this increased self-renewal capacity possibly allows them to outcompete non-

mutated HSCs, leading to their clonal expansion(36). In the first place, CH may support an increased fitness of involved HSCs compared with wild ones, suggesting a potential utility for this phenomenon in maintaining normal blood counts.

However, there is robust evidence that these clones may progress to haematologic cancers. The risk of transformation has been calculated approximately 0.5% to 1% per year (initial studies on CHIP reported a 10-fold increased risk of development of haematological cancer, which was, however, overestimated, as they studied clones with VAF > 2%, while the control population was mismatched for age)(31, 32, 102-104). Notably, the most commonly mutated genes in CH (such as *DNMT3A*, *TET2*, *ASXL1*, *JAK2*, and *TP53)* are all also commonly involved in cancer pathogenesis, while the vast majority of CH clones hold only a single mutation, being in line with the hypothesis that CH represents the first step in the clonal evolution towards the development of leukaemia(61).

Interestingly, there is also evidence that haematological cancer may not always stem from the detected CH clone(s), but instead, CH may signify a biomarker of perturbed haematopoietic state, susceptible to malignant transformation(105). In particular, this applies to haematopoiesis post genotoxic exposure, including post-chemotherapy and/or radiotherapy haematopoiesis, as described in section 1.2.4. Prevalence of clonal haematopoiesis(106). It has also been shown that in the post-HSCT CH, the variants that are detected at the time of the HSCT are not always the ones that drive post-HSCT MPN. Moreover, the above hypothesis is supported by the observation that CH with no known driver mutations is still associated with increased risk of progression to haematological cancer(34), and it is further underscored by the fact that in all fundamental large CH studies, prevalence of lymphoid malignancy was higher compared to myeloid malignancy despite the higher frequency of M-CH (compared to L-CH)(32).

Overall, CH is considered a premalignant inflammatory state, holding an increased risk for progression to blood cancer (Figure 1-3). Importantly, it is only a small percentage of the individuals with CH that will finally develop a haematological malignancy, but the parameters of prognostic significance for the development of full-blown disease are not fully understood. The progression of CH to haematological malignancies has been associated with clone size(32, 103, 104), specific gene (stronger association with *JAK2*, spliceosome mutations and co-mutation patterns involving *TET2*, *DNMT3A*, or *ASXL1, and TP53*)(54, 103, 104, 107), specific mutation(44), the number of the mutations(54, 103, 104, 106), as well as age of the individual and presence of haematopoietic stress such as genotoxic exposure or immune attack(36, 108). Kar et al.(53) analysed 10,924 individuals with CH within a pool of 200,453 UK Biobank participants with whole-exome sequencing data available to show that genetic predisposition to CH increases the risk of haematological disorders, including MPN and blood epigenetic age acceleration, importantly revealing significant different risks between individuals harbouring *DNMT3A* or *TET2* mutations. Most recently, Weeks et al. developed a prognostic tool, the so-called "Clonal Hematopoiesis Risk Score", which enables classification of the risk of individuals with CHIP to develop a myeloid neoplasm(109). Moreover, CH holds a prognostic significance in the risk of acute myeloid leukaemia (AML) relapse post chemotherapy-induced remission(110-112) as well as the outcome of autologous(106, 113, 114) and allogeneic stem cell transplantation (allo-HSCT)(115, 116).

It should be stressed that chronic inflammation is a key component of all aging, CH, and MPN and is considered to have an important role in CH acquisition, clone expansion, and further evolutionary trajectory towards the development of MPN. The vicious cycle of inflammation and CH has been best characterised in the context of *DNMT3A*- and *TET2*-related CH (see also section 1.2.5.1 Clonal haematopoiesis and venous thromboembolism).

These epigenetic regulators are keys to restraining inflammation in the immune system. This is particularly evident in monocytes / macrophages, where mutations leading to loss of function in these regulators result in an overabundant release of proinflammatory cytokines, such as tumour necrosis factor (TNF)-a, IL-1β, IL-18, and IL-6 (ref. (60, 61, 80, 83, 117-122)). Notably, the CH-imposed inflammatory environment synergizes with inflammatory stressors, including additional components of inflammaging and disease, perpetuating the expansion and clonal evolution of CH, and further aggravation of inflammation(117). Exposure of HSCs to chronic inflammation in mouse models results in loss of quiescence and biases HSCs towards a myeloid-lineage output, which in turn leads to functional decline of HSCs(123, 124). On the contrary, CH mutations, such as *DNMT3A* and *TET2* loss-of-function mutations or the *JAK2V617F* mutation, result in an increased HSC self-renewal capacity and an overall competitive advantage under the hyperinflammatory environment(58, 98-101, 124, 125). The fitness advantage of these clones, in combination with their ability to evade immunosurveillance, drives their establishment and further expansion, gradually resulting to their clonal dominance and conversion to MPN disease(58, 117, 126, 127). Furthermore, the hyperinflammatory milieu of CH and MPN increases the risk of a second hit mutation, potentially resulting in clonal evolution and disease progression/transformation(124). Notably, in a recent longitudinal population-based cohort study of CH evolutionary landscape, peripheral blood analysis from 3,359 individuals aged 60 to 91 years showed that *JAK2*-mutant clones (alongside spliceosome-mutant clones) showed highest growth rates on top of relatively increased baseline VAFs(107). Besides, chronic inflammation is a typical feature of MPNs, arising from disruption of immune signalling pathways, leading to a proinflammatory state, which confers on the malignant clone a selective advantage(128).

Importantly, presence of CH can precede haematological malignancy years before diagnosis(103, 104). Therefore, deeper understanding of the pathophysiological mechanisms driving CH-initiated disease and precise identification of patients at increased risk of developing such conditions could provide a clinically actionable time frame for early detection, monitoring, and disease-modifying early intervention.

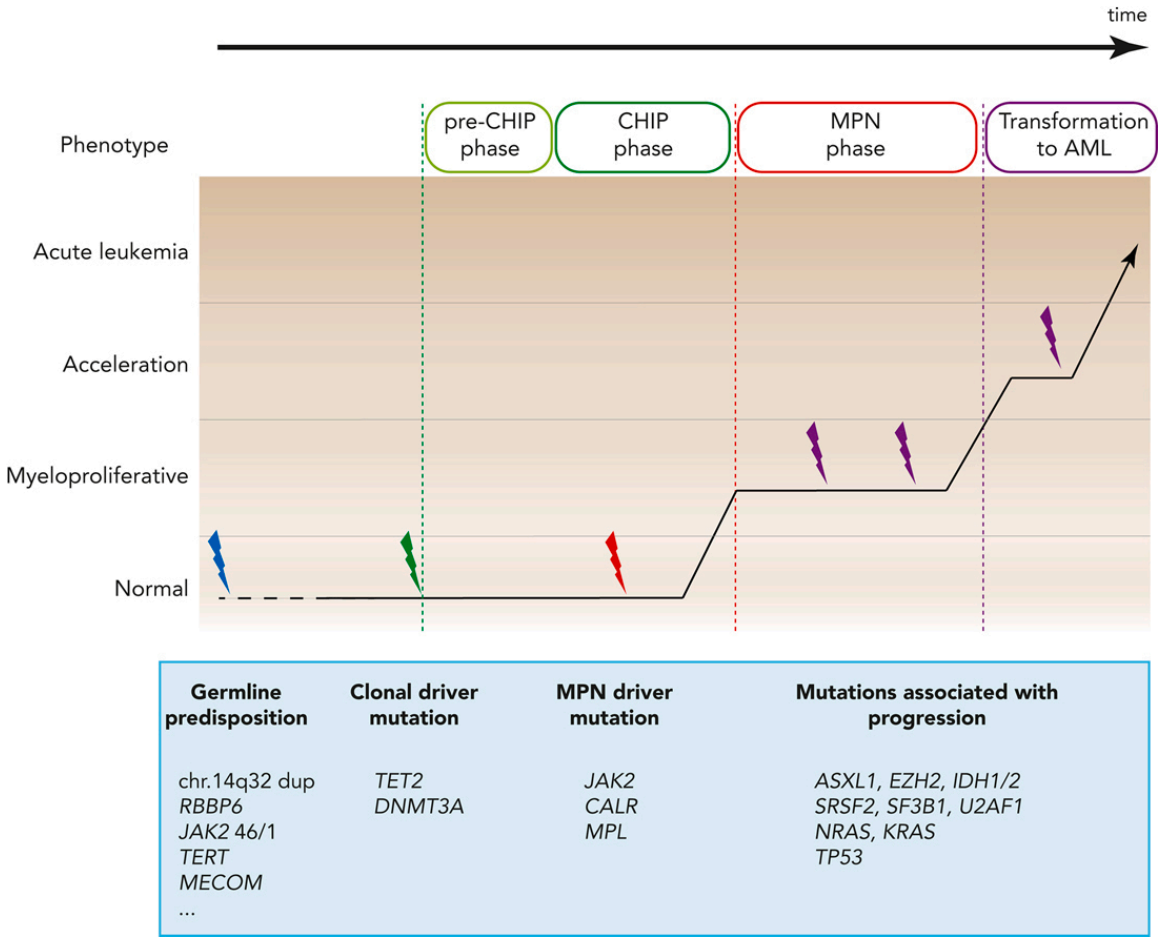


Figure 1-3. Schematic representation of the current model for the acquisition and evolutionary trajectory of premalignant CH towards the development of MPN. *The model involves early acquisition of a premalignant clone, which gradually progresses to established CH and subsequent development of overt disease over a prolonged period of time. Factors that control this process include inherited predisposition, acquisition of additional mutations, and age-related inflammation. AML, acute myeloid leukaemia; CH, clonal haematopoiesis; CHIP, clonal haematopoiesis of indeterminate potential; MPN, myeloproliferative neoplasm. Reprinted from Blood, 141(16), Luque Paz et al., Genetic basis and molecular profiling in myeloproliferative neoplasms, 1909-21, 2023; https://doi.org/10.1182/blood.2022017578 (ref. (126)); Creative Commons Attribution-NonCommercial-No Derivatives 4.0 International (CC BY-NC-ND 4.0).*

1.3 Myeloproliferative neoplasms

Myeloproliferative neoplasms are a heterogeneous group of blood cancers, classified within the myeloid neoplasms. They are clonal stem cell disorders characterised by aberrant chronic proliferation of terminal myeloid cell lines, resulting in blood count alterations in various combinations, typically associated with organomegaly and propensity for vascular complications and leukaemic transformation. They are rare, yet potentially life-threatening conditions, with a significant symptom burden, including splenomegaly-associated symptoms, and quality of life decrements(129, 130).

1.3.1 World Health Organization (WHO) classification of myeloproliferative neoplasms

"Myeloproliferative syndromes" were first recognised as a group of disorders by William Dameshek in 1951 (ref. (131)), while in 2008 the World Health Organization (WHO), in collaboration with the European Association for Haematopathology and the Society for Hematopathology, published a revised edition of the WHO classification of haematolymphoid tumours, where the diseases that made up the then called "chronic myeloproliferative diseases" were officially classified as neoplasms(132, 133).

The current revision of WHO classification of haematolymphoid tumours was released in 2016 (ref. (134)), while the upcoming 5th edition (being in beta version ahead of print as of 01 April 2023) is expected to have only minor changes regarding MPNs(135). Table 1-1 shows the classification of MPNs as provided in the 2016 revision to the WHO classification of myeloid neoplasms and acute leukaemia(134). Myeloproliferative neoplasms are broadly separated into *BCR-ABL1*-positive chronic myeloid leukaemia (defined by the *BCR*::*ABL1* fusion stemming from t(9;22)(q34;q11)), and the *BCR-ABL1*-negative clinical entities listed in the above table. The proposed MPN classification in the

5th edition of WHO Classification of Tumours retains the same clinical entity structure with the sole addition of juvenile myelomonocytic leukaemia (previously classified within the myelodysplastic/myeloproliferative neoplasms) as an MPN of early childhood(135).

Table 1-1. WHO classification of MPNs

WHO myeloid neoplasm classification
Myeloproliferative neoplasms (MPN)
Chronic myeloid leukaemia (CML), *BCR-ABL1*$^{+}$
Chronic neutrophilic leukaemia (CNL)
Polycythaemia vera (PV)
Primary myelofibrosis (PMF)
PMF, prefibrotic / early stage
PMF, overt fibrotic stage
Essential thrombocythaemia (ET)
Chronic eosinophilic leukaemia, not otherwise specified (NOS)
MPN, unclassifiable
Mastocytosis

Note: Adapted from Blood, 127(20), Arber DA et al., The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia, 2391-405, 2016; https://doi.org/10.1182/blood-2016-03-643544 (ref. (134)); with permission from The American Society of Hematology.

1.3.2 *BCR-ABL1*-negative myeloproliferative neoplasms and classical myeloproliferative neoplasms

BCR-ABL1-negative MPNs are a heterogeneous group of blood cancers, typically driven by somatic mutations affecting the MPL/JAK/STAT signalling pathway, which results in excessive production of differentiated myeloid cells(136). Distinction between different types of *BCR-ABL1*-negative MPNs is based on the integration of a patient's clinical features, peripheral blood counts, bone marrow histology, and mutation status(134, 135). As per WHO classification, *BCR-ABL1*-negative MPNs consist of essential thrombocythaemia (ET), polycythaemia vera (PV), primary myelofibrosis (PMF), chronic neutrophilic leukaemia, chronic eosinophilic leukaemia, and MPN unclassifiable.

Among them, the more frequently encountered *BCR-ABL1*-negative MPN types (also referred to as the classical MPNs) are ET, PV, and PMF(137). Essential thrombocythaemia is the most common type of MPN with an estimated annual incidence of 1.03 cases per 100,000 (with the reported annual incidence ranging from 0.2 to 2.5 cases per 100,000 in various studies) and a prevalence of 38 to 57 cases per 100,000 (ref. (138-143)). The incidence is higher in female patients (ratio 2:1), and it is increasing with age with a median age of diagnosis between 50 to 60 years, though 20% of patients are younger than 40 years old(143, 144). Polycythaemia vera is rarer with an estimated annual incidence of 0.84 cases per 100,000 (ranging between 0.01 to 2.8 per 100,000 in various studies)(141, 142) and an estimated prevalence of 22 cases per 100,000 (ref. (138)). There seems to be no significant difference in incidence between male and female patients(145). The median age of diagnosis is 60 years, and it is uncommon before the age of 40 (ref. (138)). Lastly, PMF has an estimated annual incidence of 0.47 cases per 100,000 (ref. (142)) (ranging between 0.1 to 1 case per 100,000 in various studies)(141-143, 146, 147), and a prevalence of 1 to 9 cases per 100,000 (ref. (143)). The median age of diagnosis is 65 years, and it is uncommon before the age of 50, while it affects men and women equally. All three classical MPNs are chronic in nature, presenting with variable degree of symptoms mostly due to splenomegaly and cytokine-driven constitutional symptoms, carrying a risk of complications such as thrombosis or bleeding, and progression to more advanced forms of myeloid neoplasms, including transformation to AML. The chronic phase is typically longer for ET and PV rather than PMF.

Despite having a typical benign clinical course and favourable prognosis, ET and PV hold a risk of progression to MF (post-ET MF [PET-MF], and post-PV MF [PPV-MF], respectively). Secondary MF (sMF) occurs in approximately 15% of patients(148)

(estimated 13% for ET, and 21% for PV)(149-152), and it has largely similar clinical characteristics and prognosis with PMF and is treated in the same manner.

Accordingly, the risk of transformation to AML is lower for the patients with ET or PV, with the vast majority of patients with well-controlled disease living a near-normal life(153), while the relevant risk is significantly increased for patients with MF. It ranges from 0.7% to 4% at 10 years for ET, and 2.3% to 14.4% for PV, increasing up to 20% to 30% for MF(148, 154-156). Notably, post-MF AML has a poor prognosis with overall survival of only three to six months(157).

1.3.3 Molecular landscape of classical myeloproliferative neoplasms

Myeloproliferative neoplasm-associated mutations are thought to originate in an individual HSPC, resulting in initiation and progression of clonal expansion, which eventually culminates in the development of overt MPN(136). The population of HSCs that initiate and drive the premalignant clone to overt disease are called MPN stem cells. In more than 95% of classical MPNs, the disease is driven by mutations in one of three genes: *JAK2*, *CALR*, or *MPL*.

A point mutation in the *JAK2* gene, NM_004972.4[JAK2]:c.1849G>T p.V617F, first described in 2005 (ref. (158-161)), accounts for more than 95% of cases of PV and approximately 50% of cases of ET or PMF. Most of the patients with *JAK2V617F*-negative PV will have a mutation in *JAK2* exon 12 (accounting for an additional 2% to 3% of all PV cases)(162, 163).

In 2006 it was found that part of the remaining 50% of patients with ET and PMF who are *JAK2V617F*-negative carry a mutation in the *MPL* gene (aka thrombopoietin receptor [*TPOR*]), mainly represented by mutations at codons 515 and 505 within exon 10 of *MPL*,

namely NM_005373.3(MPL):p.W515L/K/A and NM_005373.3(MPL):c.1514G>A p.S505N), which account for approximately 5% to 10% of all ET and PMF cases(164-166). However, in the majority of *JAK2V617F*-negative patients, the disease is driven by mutations in *CALR*(167, 168), initially described in 2013, and accounting for 20% to 25% of all ET and PMF. Two variants, namely *CALR* type 1: NM_004343.4(CALR):c.1092_1143del52 p.L367Tfs*46, and *CALR* type 2: NM_004343.4(CALR):c.1154_1155insTTGTC p.K385Nfs*47, account for more than 80% of *CALR*-driven MPNs.

Mutations in *JAK2*, *CALR*, and *MPL* occur in a largely mutually exclusive manner, but they all result in constitutive activation of the MPL/JAK/STAT signalling pathway in the affected cells. Mutations in *JAK2* and *MPL* are gain-of-function point mutations while the *CALR* mutations are +1 base-pair frameshifts in the coding region of the 3′ end exon of the gene, which result in the generation of a novel C-terminus(167, 168). However, irrespective of the specific MPN-driver mutation, they all result in increased and persistent activation of the JAK/STAT pathway via MPL(169).

Despite the discovery of *JAK2*, *CALR*, and *MPL* mutations, there is a remaining number of classical MPN cases, where no driver mutations in those three genes are detected(170). These “triple negative” (TN) cases account for approximately 2% of PV and 10% to 15% of ET and PMF cases(170). In the absence of an objective molecular marker diagnosis can be challenging, particularly in the cases where bone marrow testing is not feasible, declined by the patient, or inconclusive. Notably, TN ET patients are typically young female patients, while TN mutational status in PMF has been associated with poorer prognosis compared to *JAK2*-, *MPL*- or *CALR*-mutant disease(171). In a recent study by Alimam et al.(172), TN ET and *JAK2*/*CALR* mutation-positive patients, shared similar patterns of gene expression, DNA methylation, and transcriptional misregulation, associated with

JAK/STAT signalling pathway activation. Overall, TN MPN remains a diagnostically challenging subgroup of disease with adverse prognosis, leaving a large area of unmet need.

Patients with TN MPNs may carry mutations in other genes that are also involved in the pathogenesis of MPNs, though less commonly. A subgroup of these so-called myeloid genes holds a prognostic significance in different ways and variable levels of evidence and occasionally therapeutic significance, too. The genes with relatively frequent variants of clinical significance alongside current evidence on prognosis are provided in Table 1-2.

Table 1-2. Myeloid genes with common variants of clinical significance in MPNs

Gene	Polycythemia vera	Essential thrombocythemia	Primary myelofibrosis	"Secondary" myelofibrosis
***JAK2* V617F**		• Higher risk of thrombosis (included in revised IPSET)	• Intermediate prognosis and higher risk of thrombosis compared to patients with *CALR* mutation	
CALR		• Lower risk of thrombosis compared to *JAK2* mutated	• *CALR* type 1/like: improved survival compared to *JAK2* V617F and "triple negative" • Improved outcome after SCT • Absence of *CALR* type 1/like included in MIPSS70/plus v2.0 • Absence of *CALR* mutation included in MYSEC-PM and MTSS scores	• Improved OS compared to *JAK2* V617F and "triple negative" • Improved outcome after SCT • Absence of *CALR* mutation included in MYSEC-PM and MTSS scores
***MPL* W515**			• Intermediate prognosis and higher risk of thrombosis compared to patients with *CALR* mutation	
"Triple negative"			• Inferior OS and LFS compared to *JAK2/CALR* mutation (especially in pre-PMF)	
***JAK2* exon 12**	• Similar rates of thrombosis, evolution to post–PV-MF and BP, and survival to *JAK2* V617F			
ASXL1	• "Adverse variant" associated with inferior OS and MFS • Included in MIPSS-PV		• "HMR variant" associated with inferior OS/LFS, lower PFS following SCT • Included in MIPSS70/plus v2.0 and GIPSS	
EZH2		• "Adverse variant" associated with inferior OS/LFS	• "HMR variant" associated with inferior OS • Included in MIPSS70/plus v2.0	

IDH1 ***IDH2***	• *IDH1*: "adverse variant" associated with inferior OS and LFS	• *IDH2*: "adverse variant" associated with inferior OS	• "HMR variants" associated with inferior OS/LFS and as lower PFS following SCT • Included in MIPSS70/plus v2.0	• "Adverse variant" associated with lower PFS following SCT
SRSF2	• "Adverse variant" associated with inferior OS • Included in MIPSS-PV	• "Adverse variant" associated with inferior OS/LFS • Included in MIPSS-ET	• "HMR variant" associated with inferior OS/LFS • Included in MIPSS70/plus v2.0 and GIPSS	
***U2AF1* Q157**			• "HMR variant" associated inferior OS • Included in MIPSS70/plus v2.0 and GIPSS	
U2AF1* or *DNMT3A* or *CBL		• *U2AF1*: "adverse variant" associated with inferior OS/MFS • Included in MIPSS-ET	• "Adverse variant" associated with worse OS in SCT	• "Adverse variant" associated with worse OS in SCT
RUNX1	• "Adverse variant" associated with inferior OS and LFS	• "Adverse variant" associated with inferior OS/LFS		
TP53	• "Adverse variant" associated with inferior LFS	• "Adverse variant" associated with inferior OS/LFS • Included in MIPSS-ET	• "Adverse variant" associated with leukemic transformation	
LNK (SH2B3)		• "Adverse variant" associated with inferior OS		
SF3B1		• "Adverse variant" associated with inferior OS and MFS • Included in MIPSS-ET		
RAS			• "Adverse variant" associated with inferior OS • Associated with JAKi-refractoriness	

Note: Gene names are provided in italics in accordance with the current HUGO Gene Nomenclature Committee (HGNC) guidelines(173). GIPSS, genetically inspired prognostic scoring system; HMR, high molecular risk; IPSET, international prognostic scoring system for ET; MIPSS, mutation-enhanced international prognostic scoring system; MTSS, myelofibrosis transplant scoring system; MYSEC-PM, Myelofibrosis SECondary to PV and ET prognostic model; JAKi, JAK inhibitor; ET, essential thrombocythaemia; MPN, myeloproliferative neoplasm; OS, overall survival; PV, polycythaemia vera; LFS, leukaemia-free survival; MFS, myelofibrosis-free survival; PFS, progression-free survival; SCT, stem cell transplant. Table reproduced with permission from Vannucchi et al.: Hematology Am Soc Hematol Educ Program, Vannucchi et al. Molecular prognostication in Ph-negative MPNs in 2022, 2022; https://doi.org/10.1182/hematology.2022000339 (ref. (170)); © 2022 by The American Society of Hematology.

1.3.4 Developmental timing and evolution of myeloproliferative neoplasms – insights from twin studies

Despite the advances in the understanding of MPN molecular landscape over the last couple of decades, little is known about the developmental timing and evolutionary trajectories of the premalignant clones and their association with the clinical presentation of an established MPN, which typically occurs in the later decades of life.

Studies of blood cancers in monozygotic twins have provided significant insights into the pathogenesis of haematological cancers. The concept that leukaemia could potentially develop in one twin during fetal life and then get horizontally inherited to the other through intraplacental anastomoses was suggested based on initial clinical findings of twin pairs with the disease exhibiting a high level of concordance(174). In a proof-of-concept study involving genetic analysis of leukaemia clonal markers in twins Greaves et al.(175) showed that high levels of concordance are due to horizontally twin-to-twin transfer rather than vertical inheritance of susceptibility genes. So far, more than 100 cases of concordant acute leukaemia in monozygotic twins have been reported. Concordance for infants reach almost 100% while it has been calculated to approximately 15% in older children(175, 176). These comprise mostly studies of acute lymphoblastic leukaemia or AML and provide a unique window into the biology of haematological cancers. However, these cases consist of early-life leukaemia with a fast disease progression in months or few years from birth, that do not resemble MPN pathophysiology(177-179).

There have also been sporadic twin studies of paediatric MPN, while most recently, Valdes-Mas et al.(180) reported a case of transplacental transmission between monozygotic twins of a premalignant MPN clone, which resulted in the development of ET at late adolescence. A Danish cohort study focusing on MPN reported concordance for

MPN in 15% of monozygotic twins;(176) although the authors concluded that this might reflect genetic predisposition to MPN, it may be that these cases were due to an *in utero* origin and transplacental transmission, as concordant cases were not observed in dizygotic twins.

Furthermore, concordance of CH with identical somatic mutations in elderly twins has also been reported, suggesting that clones that arise *in utero* might even persist lifelong until old age without causing disease(181, 182).

Larger studies of blood cancers in monozygotic twins in sporadic MPNs are warranted, as this will provide a unique window into the evolutionary dynamics of MPNs and fitness advantage exerted by MPN-associated driver mutations.

1.3.5 Recent advances in the understanding of developmental timing and evolution of myeloproliferative neoplasms[1]

Over the last year, seminal studies shed new light into the tempo of MPN evolution, applying advanced genomics to unravel the clonal trajectories and disease latency, and revealing that the founding mutations were likely acquired much earlier in life than was previously suspected.

Prior to the availability of techniques such as those applied in these papers, our understanding of the temporal relationship between the clinical manifestation of myeloid neoplasms with the initiating genetic lesion was based on studies of individuals who developed secondary cancers after a specific genotoxic insult. Such observations, such as in survivors of Hiroshima and Nagasaki atomic bombings, suggested that the secondary cancers typically manifest after 2 to 8 years, after which the risk of malignancy

[1]Adapted from The Hematologist: ASH News and Reports, 20(1), Sousos et al., The clocks start ticking earlier than we thought, 2022; https://doi.org/10.1182/hem.V20.1.2023112.

decreases(183). The recent data in which the lineage histories of individual HSCs were reconstructed using somatic mutation patterns challenges this model, indicating that the latency between the acquisition of a driver mutation and the progression to overt myeloid malignancy can, in many cases, be many decades(184-186).

In the current studies, the timing of the original genetic lesion was estimated using the acquisition of somatic mutations across the genome as a "genetic clock", as these are presumed to occur at a constant rate over human ontogeny. The first paper was published at the end of 2021 by a group from Harvard(184), who performed whole genome sequencing (WGS) of single-cell-derived colonies grown from HSPCs isolated from two patients with ET and reconstructed the phylogenies of the cells. This estimated that the initiating *JAK2V617F* mutation was acquired two to four decades prior to clinical presentation – substantially earlier than previously thought.

In a comprehensive paper published in *Nature* by the Wellcome Sanger Institute group in February 2022, Williams *et al*(185) performed WGS of single-cell-derived colonies in a larger cohort of 12 patients with MPNs between 20 and 81 years of age. In total, they sequenced 1,013 colonies, identifying 560,978 single-nucleotide variants (SNVs) and 19,155 small indels. Tracing the detected SNVs across colonies for individual patients, they were able to decipher the phylogenetic trees, and arrange the colonies in "clades". Remarkably, for all cases studied, the acquisition of the driver mutation occurred early in the patients' life—in childhood, or in some cases, even *in utero*—. In the five patients where *JAK2V617F* was the first driver event within the MPN clone, the mean latency between *JAK2V617F* acquisition and development of disease was over three decades (ranging from 19 to 54 years). In all cases, within the premalignant clades, the earliest driver mutations were acquired within the first 200 mutations after the start of life. Equally important, the colonies harbouring no driver mutations shared only few variants with each

other, indicating that the remaining wild-type HSPCs remain highly polyclonal with a mutation rate similar to that in heathy individuals.

Collectively, the above studies suggest a new model of prolonged latency between the emergence of the premalignant clone and the development of overt disease, revealing a prolonged disease phase, in which no blood count changes or clinical phenotype have occurred, as the fitness advantage imposed by the MPN-driver mutation(s) has not yet reached such levels required for development of clinical symptomatology. Further studies in the evolutionary dynamics during this phase will improve early detection of MPNs.

1.3.6 Overview of the treatment of classical myeloproliferative neoplasms

Despite the recent advances in the treatment of *BCR-ABL1*-negative MPNs(187-190), the only curative treatment for these conditions remains allo-HSCT. However, due to the significant morbidity and mortality of transplant, allo-HSCT is only rarely an option in ET or PV, unless they progress to sMF or transform into AML(191). Even for MF (either primary or secondary), regimen-associated toxicity, high rates of non-relapse mortality due to graft failure, frequency of MPN relapse post allo-HSCT, and graft-versus-host disease remain significant concerns, preventing the widespread use of allo-HSCT, and limiting it to selected cases of advanced MF disease(192-198).

Therefore, treatment in ET and PV are mostly aimed at reducing the risk of thromboembolic complications associated with these conditions, with antithrombotic (primarily antiplatelet) treatment given to most of the patients. Apart from antithrombotic treatment, higher risk patients are treated with cytoreductive agents (with hydroxyurea or interferon in various forms and doses being the first-line agents), while lower risk patients are typically treated with active monitoring and venesections as needed(151).

Similarly, in MF, conventional chemotherapeutic treatments are primarily aimed at ameliorating MF-associated symptoms, including symptomatic splenomegaly, in patients with severe disease. Ruxolitinib, a JAK1/JAK2 inhibitor, was the first agent to be approved for the treatment of advanced MF in 2011 (ref. (199)), while two additional JAK inhibitors (JAKis), namely fedratinib (a selective JAK2 inhibitor, used as first-line therapy or second-line therapy in patients with advanced MF who have previously treated with ruxolitinib and maintain a platelet count ≥ 50 $\times 10^9$ per litre) and pacritinib (a dual JAK2 and FLT3 inhibitor, used in patients with advanced MF with a platelet count < 50 $\times 10^9$ per litre), were approved by the U.S. Food and Drug Administration (FDA) in 2019 and 2022, respectively(200, 201).

1.3.7 Evidence supporting early intervention treatment strategies in classical myeloproliferative neoplasms

Notably, real-world long-term analysis of interferon treatment in PV supports its disease-modifying effect and suggests that longer treatment with interferon agents reduces the risk of progression of the disease to MF and is also associated with reduced mortality(202-206). Abu-Zeinah et al.(204), in a single-centre retrospective study of 470 PV patients, showed that longer treatment with interferon was associated with higher MF-free survival (by 9% per year of treatment [hazard ratio (HR), 0.91; 95% confidence interval (CI), 0.87 to 0.95; $p < .001$]) and reduced mortality (by 6% per year of treatment [HR, 0.94; 95% CI, 0.9 to 0.99; $p = .012$]). Importantly, MF-free survival benefit remained significant within the low-risk group of patients when compared with phlebotomy only or hydroxyurea cytoreduction (log-rank $p = .001$). However, it should be mentioned that this was not a randomised trial, and the patients on interferon treatment were significant younger compared to the patients on hydroxycarbamide or phlebotomy alone (median age was 50 versus 58, and 52 years,

respectively; $p < .001$), with overall lower-risk disease (high-risk patients were 24% versus 56% and 45%, respectively; $p < .001$).

Interferon-induced disease modification has been attributed to its ability to induce molecular responses and/or remission. Regarding *JAK2V617F* mutation-positive MPNs, the predominant hypothesis for the mechanism of interferon-induced molecular response is that this is achieved by gradual clonal exhaustion, triggering *JAK2V617F*-mutant HSCs to exit from quiescence and differentiate into progenitors(205-208). A prospective study comparing interferon treatment (ropeginterferon alfa-2b) to hydroxyurea cytoreduction in high-risk PV patients (PROUD-PV and CONTINUATION-PV)(209), involving six-monthly quantification of the *JAK2V617F* mutation, showed that interferon treatment achieved a gradual decrease in median *JAK2V617F* allele burden throughout the study period, with the median VAF being 8.5% at 60 months, compared to 44.4% in the hydroxyurea arm ($p < 0.0001$; baseline VAFs were 37.3% and 38.1%, respectively)(206). Notably, at 60 months of treatment, 18/92 patients (19.6%) on interferon treatment achieved a VAF < 1% compared to only one patient (1.4%) on the hydroxyurea arm ($p = 0.0002$). Importantly, long-term treatment with interferon has not been associated with increased risk of leukaemia transformation or development of solid tumours (particularly cutaneous)(210). However, further studies on the clinical relevance of interferon-induced molecular response are required.

Furthermore, a randomised phase 2 trial of ruxolitinib versus best available therapy in high-risk PV patients with resistance or intolerance to hydroxycarbamide (MAJIC-PV)(211) showed that patients on ruxolitinib and those who achieved an early complete response (within one year) had an improved event-free survival (defined as follow-up time with no occurrence of major thrombosis, haemorrhage, disease transformation and/or death) (HR, 0.58; 95% CI, 0.35 to 0.94; $p = .03$ for ruxolitinib, and HR, 0.41; 95% CI, 0.21 to 0.78;

p = .01 for early complete remission). Importantly, in this population, a 50% reduction in *JAK2* mutant allele burden (which was more frequent in the ruxolitinib arm) was correlated with improved event-free survival (p = .001), progression-free survival (p = .001), and overall survival (p = .01), and clearance of MPN stem cells.

In addition to PV, there is also accumulating evidence that non-curative treatment with JAKis, particularly ruxolitinib, may be associated with an overall survival benefit in MF(156, 212-220). Data stem from case-control studies, including real-world data from 1677 MF patients of the Medicare database(216) and 1010 patients of the ERNEST project(218), favouring an overall survival advantage compared to other non-curative, drug-based treatment options(212-218). However, given the limitations of observational studies, additional evidence is required to draw a definitive conclusion on ruxolitinib's and other JAKis' disease-modifying effect(219, 220). Importantly, a recent data pooled analysis of the ruxolitinib registrational COMFORT-I/II trials(221, 222) showed a significant improvement in overall survival at week 240 for the patients who initiated ruxolitinib treatment within the first 12 months from diagnosis compared to the patient starting treatment more then 12 months after (63% [95% CI, 51% to 73%] vs 57% [95% CI, 49% to 64%]; p = .043)(223). This suggests that earlier initiation of ruxolitinib therapy in advanced MF may have a positive impact on survival. Moreover, analysis of the same dataset showed that treatment of early-stage though symptomatic MF was more effective, leading to better patient outcomes, alongside advanced improvement in allele burden (as evidence of disease modification), and potentially improved overall survival(224, 225). In addition to COMFORT-I/II data, a real-world retrospective study evaluating the possible predictors of response to ruxolitinib found that disease severity and delayed ruxolitinib initiation (by ≥ 2 years from MF diagnosis) resulted in inferior outcomes, highlighting the

importance of early initiation of JAK inhibition, although any survival advantage in patients with lower-risk disease remains to be seen in further studies.

Collectively, the above data on the disease-modifying effect of early therapeutic intervention in classical MPNs, in combination with the advances in our understanding of developmental timing and clonal evolution of these conditions, may result in preventive intervention strategies aiming to modify clonal evolution and drastically alter the clinical course of the disease (Figure 1-4).

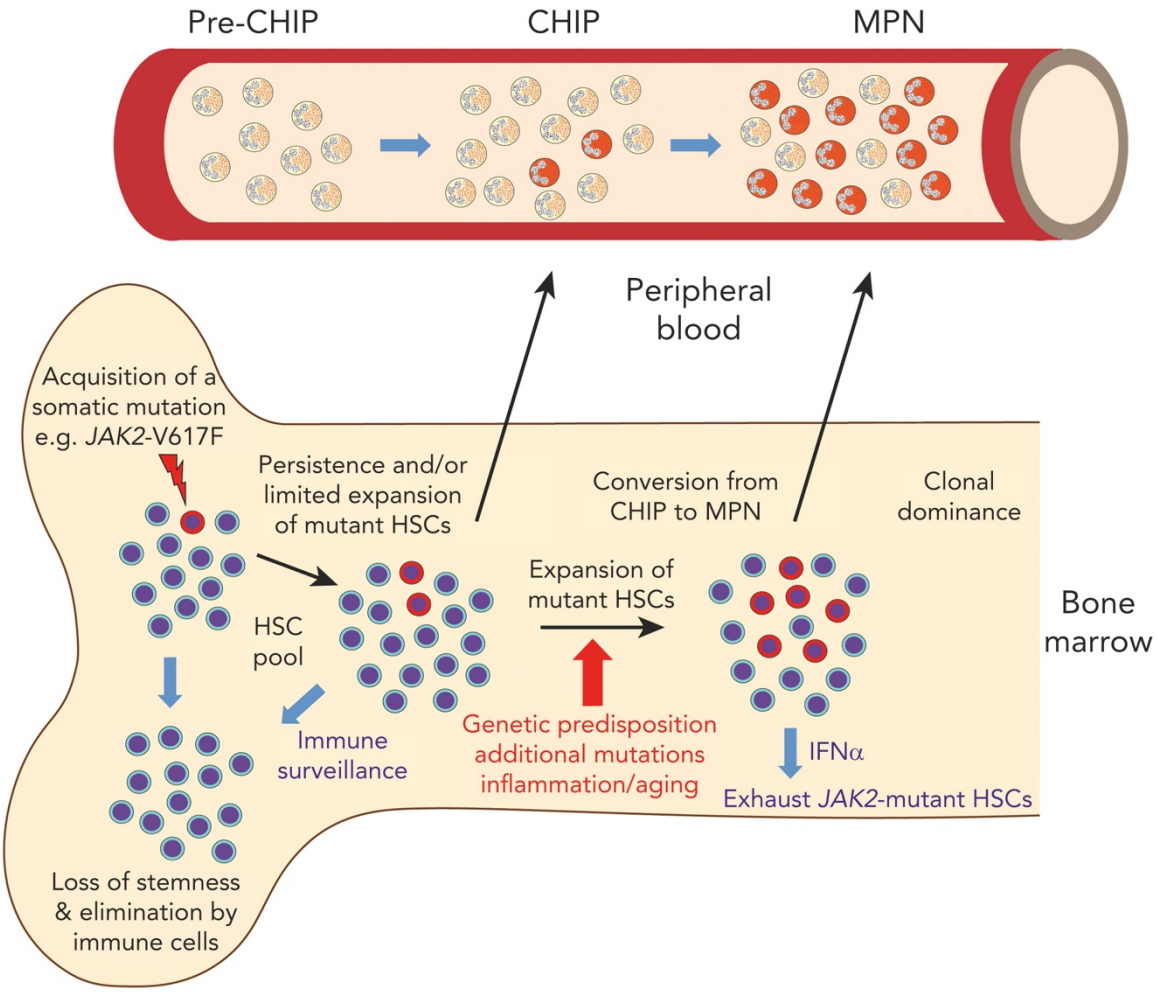


Figure 1-4. Evolutionary dynamics of MPN from the acquisition of a disease driver mutation to conversion to MPN. *Pre-CHIP: In the first place, a single HSC within the bone marrow acquires a disease driver mutation, conferring fitness advantage and an initial limited expansion. At this phase, the mutated HSCs can be eliminated due to loss of stemness and immunosurveillance. During this "pre-CHIP phase", the mutation is not detectable in the peripheral blood. CHIP: In the next step, in a process that can span decades, the clone gradually expands in various patterns of dynamics. At the phase of CHIP, the mutation is*

detectable in the peripheral blood at varying VAFs. MPN: In a minority of cases, the fitness advantage imposed by the MPN-driver mutation(s) can reach such levels that results in clonal dominance of the mutated HSPCs in the marrow, accumulation of mature cells in the peripheral blood, and ultimately, conversion to MPN with blood count abnormalities and clinical symptomatology. The factors that can contribute to the establishment of the MPN are listed in red font. Early treatment with disease-modifying agents, such as interferon, can trigger the JAK2V617F-mutant HSCs to exit from quiescence and differentiate into progenitors, resulting in gradual clonal exhaustion. CHIP, clonal haematopoiesis of indeterminate potential; HSC, haematopoietic stem cell; MPN, myeloproliferative neoplasm; VAF, variant allele frequency. Reprinted from Blood, 141(16), Luque Paz et al., Genetic basis and molecular profiling in myeloproliferative neoplasms, 1909-21, 2023; https://doi.org/10.1182/blood.2022017578 (ref. (126)) Creative Commons Attribution-NonCommercial-No Derivatives 4.0 International (CC BY-NC-ND 4.0).

1.4 Thesis aims and hypotheses

My DPhil project aimed to highlight the importance of genetic lineage tracing and articulate its benefits in the study of CH and MPNs.

The first part of this work focused on linage tracing in granulocytes (as the most abundant type of myeloid cells) versus T cells (as the germline DNA) and involved a twin study in a pair of monozygotic twins who presented with *CALR*-mutant PMF in the fourth decade of their life. I aimed to investigate the origin of their premalignant clones and their subsequent molecular evolution. To address these questions, I initially conducted lineage tracing of the *CALR* mutation by fragment analysis, followed by lineage tracing of the MPN clone through WGS of granulocytes (derived by the MPN clone) against T cells (as the germline control) for each of the siblings. Given the very low likelihood that the twins independently acquired the very same mutation during their life, I hypothesised that the twins shared the mutation through either vertical or horizontal gene transfer, and subsequently, the premalignant clone independently evolved to established disease in their adult life. The results of this study are presented in Chapter 3.

In the next part of the project, I shifted the focus from myeloid lineage against germline DNA to myeloid lineage against the study of platelets as a surrogate of the Mk lineage. The aim of this part of my work was to assess whether platelet-biased CH exists in humans and determine its prevalence in normal haematopoiesis and myeloid neoplasms. The hypothesis was that platelet-biased CH does exist, and therefore, the study of platelets can increase sensitivity in the detection of CH and MPNs compared to the currently applied methodology, which involves whole-blood- or granulocyte-derived DNA. This work involved two distinct substudies: As a first step, given the central role of the *JAK2V617F* mutation in both CH and MPNs, I aimed to track *JAK2V617F* in the granulocytes and

matched platelets from selected individuals. To conduct this proof-of-concept study I developed a method for high-purity blood product isolation, and a droplet digital PCR (ddPCR) assay for the detection of the *JAK2V617F* mutation in either gDNA or cDNA template. The results of this part of the work are presented in Chapter 4. Next, I aimed to investigate whether platelet-biased CH exists for genes—other than JAK2—that are commonly involved in CH and MPNs. In view of this, I developed a custom, targeted NGS panel for parallel sequencing of both gDNA and cDNA, covering the major genes or hotspot mutations involved in CH and MPNs. The results of this study are presented in Chapter 5.

Chapter 2 Materials and Methods[2]

2.1 Study participants

2.1.1 Cohort design

2.1.1.1 Twin study

Regarding the twin studies, one of the brothers was referred to the Myeloid Service, Department of Clinical Haematology, Cancer and Haematology Centre, Oxford University Hospitals (OUH) National Health Service (NHS) Foundation Trust, where I reviewed him and subsequently followed him up while living in the UK. His twin brother, who lives abroad, was contacted via phone and sample collection was arranged in collaboration with the centre where he is followed up.

For the extension project, that aimed to assess the mutational status at birth for patients diagnosed with MPNs as children or young adults, I set up a study to enable me to get access to their Guthrie cards and perform molecular studies on their neonatal blood spot DNA. The study (ANNB_NBS_027: "Investigating the molecular status at birth of patients diagnosed with myeloproliferative neoplasms as children or young adults") was approved under the Public Health England Antenatal and Newborn screening ethics and utilised the infrastructure of an already existing biobank study within the Mead Lab (The INForMeD

[2]Parts of Chapter 2 were adapted from Nat Med, 28(6), Sousos, Ní Leathlobhair et al., In utero origin of myelofibrosis presenting in adult monozygotic twins, 1207-11, 2022; https://doi.org/10.1038/s41591-022-01793-4. Additionally, parts of sections 2.13 Next-generation sequencing/2.13.2 Custom next-generation sequencing panel, and 2.17 Bioinformatic analysis/2.17.2 Custom next-generation sequencing analysis were adapted from Cancer Discov, 13(2), Khan et al., Human Bone Marrow Organoids for Disease Modeling, Discovery, and Validation of Therapeutic Targets in Hematologic Malignancies, 364-85, 2023; https://doi.org/10.1158/2159-8290.CD-22-0199.

Study: “Investigating the genetic and cellular basis of sporadic & Familial Myeloid Disorders”; REC reference 16/LO/1376). As Oxford Screening Laboratory, UK Newborn Screening Laboratories Network, have been storing Guthrie cards since 1976, I identified all MPN patients followed up in the Myeloid Service, Department of Clinical Haematology, Cancer and Haematology Centre, OUH NHS Foundation Trust, and the Children’s Haematology and Oncology, OUH NHS Foundation Trust, who were born from 1976 onwards. Based on this list, I looked for and retrieved the available Guthrie cards from the off-site archive they are stored in and analysed them for presence of the MPN-driver mutation in a cross-sectional study approach.

2.1.1.2 Targeted genetic analysis of platelet lineage cells

Targeted genetic analysis of platelet lineage cells in CH and MPNs consisted of three substudies:

2.1.1.2.1 Haematologically normal elderly individuals

The first one involved a cross-sectional study of haematologically normal elderly individuals, where I aimed to look for presence and the estimated prevalence of platelet-biased CH, as a proof-of-concept study. Given the presence of an already existing biobank study in the Mead Lab (The INForMeD Study; REC reference 16/LO/1376), I submitted a major revision to enable (among other changes) the collection of fully anonymised blood samples from the Clinical Diagnostic Laboratories, such as the National Institute for Health and Care Research (NIHR) Oxford Biomedical Research Centre, OUH NHS Foundation Trust, that are excess to clinical requirements. Eligibility included all individuals with a normal full blood count who have not explicitly declined the use of their samples in research, with the sole exception of children under 2 years of age. For the specific project additional inclusion criteria were i) individual’s age ≥ 70 years with ii) excess EDTA-

anticoagulated peripheral blood sample available within 48 hours after collection, while additional exclusion criterion was history of a haematological neoplasm, as assessed on the electronic patient record system at the time of sample selection. The acceptable ranges for blood counts were in line with the WHO diagnostic criteria for MPNs(134, 226): 4.0 $\times 10^9$ per litre ≤ white cell count ≤ 11.0 $\times 10^9$ per litre; 100 $\times 10^9$ per litre ≤ platelet count ≤ 450 $\times 10^9$ per litre; 110 g per litre ≤ haemoglobin < 160 g per litre for women, and 110 g per litre ≤ haemoglobin < 165 g per litre for men. In accordance with Human Tissue Authority requirements, no clinical information associated with these samples was collected.

2.1.1.2.2 Unprovoked venous thromboembolism patients

The second substudy was a case-control study of platelet-biased CH in patients with uVTE versus age- and sex-matched controls. Given the central role of platelets in thrombosis, we hypothesised that the presence of platelet-biased CH might increase the risk of VTE. Therefore, we aimed to determine the prevalence of platelet-biased/restricted CH in patients with uVTE ("thrombosis group") compared to its prevalence in an age- and sex-matched control group with no history of uVTE or known CVD ("control group"). In the thrombosis group we aimed to recruit 200 patients, while in the latter one we aimed to recruit 50 individuals. An interim analysis after recruitment of at least 100 subjects in the thrombosis group was planned as a proof-of-concept analysis and assessment of an emerging trend. The results of this interim analysis are herewith presented. All study subjects were identified by clinical staff at Oxford Haemophilia and Thrombosis Centre (OHTC) when attending either DVT diagnostic clinic or 3-month VTE review clinic. Clinical data were collected by the clinical teams and provided link-anonymised to the research team, using a patient-unique number to link samples and data. The data that were retrieved from the subjects' electronic patient record were date and site of uVTE

(DVT/PE), age, sex, body mass index, smoking history, history of hypertension, dyslipidaemia, and/or diabetes, significant past co-morbid history (previous history of venous/arterial thrombosis, cancer, and/or inflammatory condition), medications (including antiplatelet, anticoagulant, NSAIDs, antihypertensive, statins, and diabetes medications). Peripheral blood (10-ml EDTA-anticoagulated whole blood) was collected for each study participant and processed for full blood count, blood product isolation, and CH studies. Mutational analysis data were interpreted blinded to the clinical information. Clonal haematopoiesis data were then compared between samples from events (participants with uVTE) and non-events (those without VTE), after adjustment for the clinical data. Patient recruitment, and sample and data collection were coordinated by Associate Professor Susan Shapiro (Head of OHTC). Dr Akriti Rastogi (research assistant in Shapiro group), and Mr Nawshad Hayder (research assistant in Mead group) significantly contributed to sample processing.

2.1.1.2.3 Triple negative and low-variant-allele-frequency *JAK2V617F* myeloproliferative neoplasm patients

In the third substudy I aimed to assess the prevalence of platelet-biased CH in overt MPN disease. Therefore, I identified the patients with well-defined TN and low-VAF *JAK2V617F* mutation-positive MPNs (VAF < 10%) followed-up in the Myeloid Service, Department of Clinical Haematology, Cancer and Haematology Centre, OUH NHS Foundation Trust. Patients were consented to the INForMeD study (REC reference 16/LO/1376), and EDTA-anticoagulated samples were taken for CH studies.

2.1.2 Informed consent

All procedures followed in the present study were performed in accordance with the ethical standards of the current revision of the Declaration of Helsinki(227). Biological sex was

defined as binary (female, male) as per the source database (NHS), which did not include the intersex option. All patients provided written informed consent.

2.1.2.1 Twin study

Regarding twin A, samples were obtained from the center in which he is followed up (Christian Medical College, Vellore, India) for diagnostic purposes. Other patients were enrolled to The INForMeD Study (REC reference 16/LO/1376). All research analyses were conducted according to The INForMeD Study (REC reference 16/LO/1376) and ANNB_NBS_027 study (Public Health England Antenatal and Newborn screening research advisory committee, 13 March 2020).

2.1.2.2 Targeted genetic analysis of platelet lineage cells

For the "haematologically normal elderly individual" cohort as well as the MPN patient cohort, clinical samples were collected under The INForMeD Study (REC reference 16/LO/1376). All subjects in the "uVTE" cohort were consented to the "platelet-biased clonal haematopoiesis and unprovoked venous thromboembolism" study, approved under the Oxford Radcliffe Biobank (ORB) research tissue bank ethics (REC reference 19/SC/0173).

2.1.3 Sample handling and storage

As the first step, I developed a blood and/or bone marrow sample processing method to enable the simultaneous isolation of platelets, plasma, granulocytes, mononuclear cells, T cells, and/or HSPCs, as needed. The protocol is provided in Appendix A.1. All samples were processed within 48 hours after collection. Samples were stored until analysis in the Medical Research Council (MRC) Molecular Haematology Unit (MHU), MRC Weatherall

Institute of Molecular Medicine (WIMM), University of Oxford, under The INForMeD Study (REC reference 16/LO/1376) in accordance with the study protocol.

2.2 Cell isolation

Given the importance of effective blood product isolation for the lineage-specific genetic analysis project, I developed and optimised a method for the simultaneous isolation of platelet, plasma, mononuclear cells, granulocytes, T cells and/or HSPCs, as needed. The resulted protocol is provided in Appendix A.1.

Regarding the twin studies, in addition to the aforementioned protocol, granulocytes were also sorted in mini-bulk by fluorescence-activated cell sorting (FACS) and whole-genome amplified, as described in sections 2.7 Whole genome amplification and 2.15 Flow cytometry and cell sorting.

Moreover, dermal fibroblast cell cultures were set up following a skin biopsy from the proximal thigh in twin B. Cells were cultured in Gibco AmnioMAX C-100 Complete Medium (Gibco; Thermo Fisher Scientific Inc.) and Chang medium D (FUJIFILM Irvine Scientific) without additives. They were cultured at 37°C, 5% CO_2, and subcultured when confluent. At 22 days in culture, cells were removed from the subcultures in Chang medium using trypsin, and DNA was extracted. DNA from the cells in the AmnioMAX medium was extracted at day 27. Fibroblast cell culture was performed in Oxford Genetic Laboratories, OUH NHS Foundation Trust.

2.2.1 Platelet isolation protocol

As platelet isolation and downstream analysis was a key element for this project, I put special effort into platelet isolation optimisation and validation of the final method. The

final protocol was concluded after comparison of purity of the isolated platelet population among a number of different methodologies studied. The final two-step centrifugation protocol is provided in Appendix A.1 and schematically shown in Figure 2-1.

Thaw process involved the following steps: the previously cryopreserved platelet pellet was retrieved from storage at -80ºC and immediately incubated for 1 min at 37°C at the water bath. Next, it was resuspended in 1 ml of normal saline (previously equilibrated to room temperature), and gently mixed by pipetting. This step aimed to minimize the cytotoxic effect of dimethyl sulfoxide (DMSO) contained in the platelet freezing medium (Appendix A.1). The resultant platelet solution was then centrifuged at 1,000 g for 7 min at room temperature (20ºC-22ºC). After this wash step the supernatant was discarded and the remaining platelet pellet was used for downstream analyses.

Overall, the platelet isolation as well as the platelet freezing and thaw methods were based on previous published protocols for cryopreservation of platelet concentrates in the clinical setting(228-230), adjusted to blood collection tube volumes and research lab equipment.

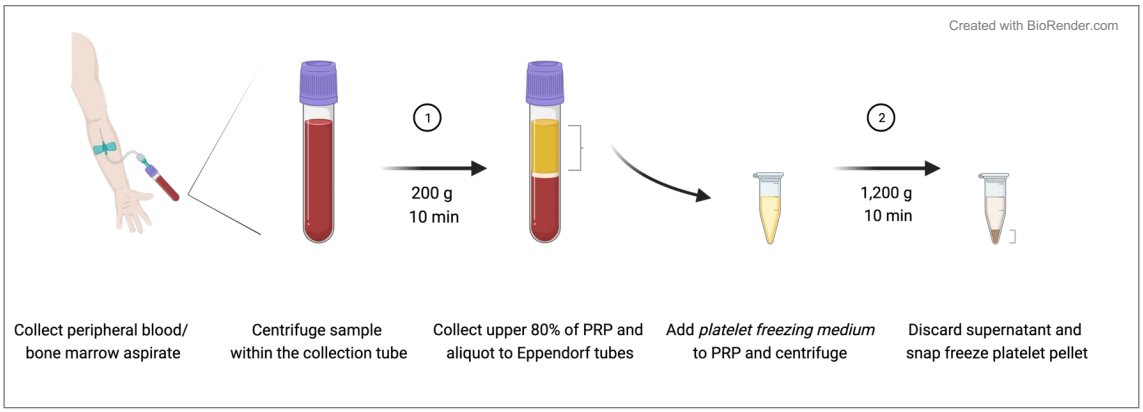


Figure 2-1. Schematic representation of the platelet cryopreservation method. *Platelet isolation involved a two-step centrifugation process: 1) In the first step (200 g, 10 min, room temperature, decelerator off) the platelet-rich plasma was separated from the blood cells and treated with 6% DMSO. 2) In the second step (1,200 g, 10 min, room temperature) the platelet-rich plasma with DMSO was centrifuged to give the platelet pellet and the supernatant DMSO-containing platelet-poor plasma. The supernatant was discarded, and the platelet pellet was snap frozen at -80ºC. Platelet freezing medium: 0.9% (w/v) NaCl with 27% DMSO, diluted*

in the platelet-rich plasma to a final DMSO concentration of 6%. DMSO, dimethyl sulfoxide; PRP, platelet-rich plasma.

2.2.2 Platelet isolation method optimisation

The aforementioned two-step centrifugation method was selected as the most effective one after application of various methodologies and comparison of the resulting product for platelet purity. In more detail, in addition to the two-step centrifugation of EDTA-anticoagulated samples, I assessed the use of (i) one or (ii) two additional centrifugation steps (600 g, 10 min, room temperature); (iii) heparin- or (iv) acid-citrate-dextrose (ACD)-anticoagulated samples; (v) addition of platelet stabilisers (inhibitors of platelet adhesion, activation and aggregation) in EDTA-anticoagulated samples, namely addition of ACD (MilliporeSigma; Merck) in the whole blood (16.66% [v/v]) prior to first centrifugation, and prostacyclin (Cayman Chemical Company) in the PRP (2×10^{-4} μg per litre) prior to the second centrifugation; leukodepletion filtering of either (vi) whole blood or (vii) the platelet-rich plasma (PRP), by using the Pall Sterile Acrodisc WBC syringe filter (Pall Corporation) before the first or the second centrifugation respectively; (viii) negative selection of $CD45^+$ cells (as a non-platelet, non-red-blood-cell, haematopoietic cell marker) or (ix) combined negative selection of $CD45^+$ cells and positive selection of $CD42b^+$ cells (as a platelet marker) with use of Dynabeads magnetic beads in the platelet product (Invitrogen; Thermo Fisher Scientific Inc.). Platelet isolation studies were done in collaboration with Ms Lauren Murphy (DPhil student in Psaila group).

Regarding the thaw process, ratio of platelet recovery was also assessed for HEPES-buffered Tyrode's solution instead of normal saline.

2.2.3 Platelet purity studies

Purity of the isolated platelet population was assessed by cytocentrifugation, gene expression, and flow cytometry analysis, as described below:

2.2.3.1 Cytocentrifugation analysis

Blood cell contamination in the isolated platelet population was assessed by cytocentrifugation. Cytospins were prepared in a two-step cytocentrifugation process, using polylysine-coated microscope slides (VWR International; Avantor Inc.) and the Thermo Scientific Shandon Cytospin®4 Cytocentrifuge (Thermo Fisher Scientific Inc.). Initially, cytospin filter cards were prepared with 50 µl phosphate-buffered saline (PBS) (Life Technologies; Thermo Fisher Scientific Inc.) centrifuged at 350 rpm for 2 min, and then 100 µl of the platelet resuspension, adjusted to an expected platelet concentration of 0.1 $\times 10^9$ per litre, were centrifuged at 1,000 rpm for 10 min. Cytospins were air dried and subsequently stained using the Richard-Allan Scientific Three-Step Stain Set (Thermo Fisher Scientific Inc.). Fixation and staining times were optimised to 60 s for methanol fixative (Fixative), 2.5 min for aqueous Eosin Y solution (Solution A), and 3 min for aqueous Methylene Blue/Azure A solution (Solution B). After completion of the staining, the slides were washed with deionised water, air dried, and stored for microscopy.

2.2.3.2 Gene expression analysis

Purity of the isolated platelet population was also assessed by real-time quantitative PCR (qPCR) using predesigned TaqMan Gene Expression assays (Thermo Fisher Scientific Inc.) for the following target-genes: *PF4*, *SELP*, and *TGFB1* as genes with markedly increased expression in platelets, *APOC1*, *KLF1*, and *TMOD1* as erythroid cell-associated genes, and *ELANE*, *CSF3R*, *CD79A*, and *CD3G* as white blood cell markers. Assay probes

were FAM(5′) labelled. The expression of the above target-genes was calculated relative to *GAPDH* (with VIC-MGB probe modification) as the reference gene control. All assays were performed using the TaqMan Fast Advanced Master Mix (Applied Biosystems; Thermo Fisher Scientific Inc.) with the following conditions: 10 µl TaqMan Fast Advanced Master Mix (2×), 1 µl TaqMan gene expression assay, 2 µl template cDNA, and PCR-grade water up to 20 µl; and the following cycling protocol: hold stage for 20 s at 95°C, followed by 40 cycles of denaturation for 3 s at 95°C, and annealing-extension for 30 s at 60°C on the Quantitation – Comparative C_T ($\Delta\Delta C_T$) platform and Fast ramp speed configuration. All assays were run simultaneously on the same qPCR plate in two technical replicates plus two PCR-grade water template controls per run.

2.2.3.3 Flow cytometry analysis

Purity of the isolated platelet population was further assessed via flow cytometry. Details on sample preparation, antibody panel characteristics, flow cytometry instrumentation, and data analysis are provided in section 2.15 Flow cytometry and cell sorting.

2.2.4 Platelet recovery/survival studies

Platelet preservation studies and RNA/cDNA yield studies were conducted to assess whether the number of platelets isolated from a typical 4-ml EDTA-anticoagulated sample was sufficient for downstream procedures. Nucleic acid extraction, cDNA synthesis, and DNA quantification details are provided in sections 2.5 Nucleic acid extraction and 2.6 Nucleic acid quantification.

2.3 Cell lines

A Jurkat cell line (CVCL_0065) was used as the *CALRdel52bp*-negative control. TF-1 (CVCL_0559) and HEL (CVCL_0001) cell lines were used as *JAK2V617F* negative and positive controls, respectively. All cell lines were purchased from American Type Culture Collection (catalog numbers TIB-152, CRL-2003, and TIB-180 for Jurkat, TF-1 and HEL, respectively). Cells were cultured according to American Type Culture Collection recommendations.

2.4 Microscopy

Peripheral blood and isolated blood product microscopy was done using the Olympus BX60 microscope with the Olympus UplanFl 40×/0.75 Infinity/0.17 and Olympus UplanSApo 100×/1.40 Oil Immersion Infinity/0.17/FN26.5 objectives (Olympus). Imaging was performed with the INFINITY3S-1UR camera and use of the Infinity ANALYZE software, release 6.5 (Lumenera). Tissue microscopy was performed using a Nikon Eclipse E400 microscope (40×/20× objectives) (Nikon) on routinely prepared 4-µm sections cut from formalin-fixed paraffin-embedded blocks. Microscope fields were selected from whole-slide scanned images using the NanoZoomer 2.0-HT scanner in 40× mode and NDP.view2 viewer software version 2.9.25 (Hamamatsu Photonics). Image analysis was done using Adobe Photoshop version 21.2 (Adobe) and was limited to white balance of complete image and addition of scale bars. Bone marrow microscopy was performed by Dr Daniel Royston in Cellular Pathology, OUH NHS Foundation Trust.

2.5 Nucleic acid extraction

2.5.1 DNA extraction

Blood product DNA was extracted using the DNeasy Blood & Tissue kit (QIAGEN), as per the manufacturer recommendations. Nail DNA from nail clippings of twin A and neonatal dried blood spot DNA were exceptionally extracted using the QIAamp DNA Micro kit (QIAGEN) according to the manufacturer's instructions.

2.5.2 RNA extraction

RNA was extracted using the RNeasy Mini Kit (QIAGEN), as per the manufacturer recommendations.

2.5.3 cDNA synthesis

cDNA synthesis was performed using the EvoScript Universal cDNA Master kit (F. Hoffmann-La Roche AG), as per the manufacturer recommendations. Total reaction volume was 30 µl.

2.6 Nucleic acid quantification

Extracted RNA quantification was performed using the NanoDrop One Spectrophotometer (Thermo Fisher Scientific Inc.), while all DNA products were quantified using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific Inc.). Depending on the expected DNA yield samples were prepared for fluorometry using either the Qubit 1X dsDNA HS Assay Kit (quantitation range 0.1-120 ng) or the Qubit dsDNA BR Assay Kit (quantitation range 4-2,000 ng).

2.7 Whole genome amplification

Whole genome amplification for each sorted population and single-cell-derived colony DNA was performed using the REPLI-g Mini/Midi Kits (QIAGEN) as per the manufacturer recommendations.

2.8 Polymerase chain reaction (PCR)

Polymerase chain reaction target regions and corresponding primer set sequences for the lineage-specific *CALR* genotyping, and validation of the targeted sequencing and WGS results are shown in Table 2-1. All PCRs were performed using the KAPA2G Robust HotStart ReadyMix PCR Kit (Merck) with the following conditions: 12.5 µl KAPA2G Robust HotStart ReadyMix (2×), primer set at 0.5 µM each, template DNA as required and PCR-grade water up to 25 µl; and the following cycling protocol: initial denaturation for 3 min at 95°C, 35 cycles of denaturation for 15 s at 95°C, annealing for 15 s at 60°C, extension for 15 s at 72°C and final extension for 3 min at 72°C. All PCR amplicons were purified using either Agencourt AMPure XP Magnetic Beads (Beckman Coulter) or the Monarch DNA Gel Extraction Kit (New England Biolabs).

Neonatal dried blood spot mutational analysis was performed with use of ddPCR in a two-step PCR. *JAK2* exon 14 (chr9[GRCh38]:5073730–5073813) was amplified in the first reaction, and the amplicon was then tested for presence of the NM_004972.4(JAK2):c.1849G>T p.V617F mutation using a ddPCR assay. The same primers were used for both reactions. Primer and probe details are provided in Table 2-1. The first round of PCR was performed using the KAPA HiFi HotStart ReadyMix PCR Kit (Merck) with the following conditions: 10 µl KAPA HiFi HotStart ReadyMix (2×), primers

at 0.3 µM, 3 µl template DNA and PCR-grade water up to 20 µl; and the following cycling protocol: initial denaturation for 3 min at 95°C, 20 cycles of denaturation for 20 s at 98°C, annealing for 15 s at 62.5°C, extension for 30 s at 72°C and final extension for 1 min at 72°C. Details pertained to the ddPCR assay, sample preparation, instrumentation and configurations are provided in section 2.10 Droplet digital PCR.

2.9 Real-time quantitative PCR

Real-time quantitative PCR (qPCR) was performed on the 7500 Fast Real-Time PCR System using predesigned TaqMan Gene Expression assays and the TaqMan Fast Advanced Master Mix (Applied Biosystems; Thermo Fisher Scientific Inc.). Details pertaining to the specific TaqMan Gene Expression assays used, and the reaction and cycling conditions are provided in section 2.2.3 Platelet purity studies. Real-time quantitative PCR data analysis was done using the 7500 Software, version 2.3 (Applied Biosystems, Thermo Fisher Scientific Inc.).

2.10 Droplet digital PCR

2.10.1 Method overview

Droplet digital PCR was developed to provide high-precision, absolute quantification of nucleic acid molecules encapsulated in discrete, volumetrically defined water-in-oil droplet partitions(231, 232). Droplet digital PCR reactions are prepared in a similar manner as real-time PCR reactions that use TaqMan hydrolysis probes labelled with reporter fluorophores or an intercalating dye such as EvaGreen. The droplet generator partitions samples into ~20,000 nanolitre-sized droplets, followed by PCR amplification that is

carried out within each droplet using a thermal cycler, and subsequent readout of individual droplets on the droplet reader. Fluorescent-positive droplets (containing at least one copy of the target DNA molecule) and negative droplets are counted to provide absolute quantification of target DNA in digital form (Figure 2-2).

2.10.2 *Custom JAK2V617F* assays for parallel gDNA and cDNA analysis

As the first step of the lineage specific genetic analysis project, I aimed for mutational analysis of the *JAK2V617F* mutation. In view of simultaneous analysis of gDNA and cDNA (from matched cell populations derived from the same donor), I developed and optimised a gDNA-specific *JAK2V617F* ddPCR assay alongside a cDNA-specific *JAK2V617F* ddPCR assay sharing similar technical properties. That way, the two assays can be loaded in parallel on a same run and the results of the samples included in the run are directly comparable.

Both gDNA and the cDNA assays shared the same pair of probes to target the wild-type and mutant sequences (being HEX-labelled and FAM-labelled, respectively), and the same forward primer. Distinct reverse primers provided specificity for gDNA or cDNA templates. A schematic representation of the ddPCR methodology is provided in Figure 2-2, while primer and probe details are provided in Table 2-1. Both Dr Christina Simoglou Karali and Dr Zemin Ren, postdoctoral researchers in Mead group, contributed to the optimisation of the assay.

2.10.2.1 Sample preparation

Sample preparation was done with Bio-Rad's Supermix for Probes (no dUTP) with the following conditions: 11 µl of the supermix (2×), primers at 900 nM each, probes at 250 nM each, DNA template (either gDNA or cDNA, aiming for 10-60 ng and 10-20 ng

input, respectively, to ensure that both target and reference copies were within the dynamic detection range of the system), and PCR-grade water to top up to a final reaction volume of 22 μl. Notably, given the typically low concentrations of platelet-derived cDNA, I modified the manufacturer recommended protocol, increasing primer working concentration from 10 μM to 100 μM, so as to allow more volume available in the reaction for the DNA template to reach the desired input of 10-20 ng. Paired blood product DNA (granulocyte gDNA and platelet cDNA template) for each patient were included within the same run to enable direct comparison of the results. For all runs TF-1 cell DNA was used as negative control, whereas HEL cell DNA was used as the positive control, with multiple replicates per run. Samples were loaded on semi-skirted 96-well plates and processed immediately at the Automated Droplet Generator (Bio-Rad Laboratories) for droplet generation and transfer of the emulsified samples to PCR plates. The amplification was performed using the Bio-Rad C1000 Touch Thermal Cycler with the following cycling protocol: initial enzyme activation for 10 min at 95°C, 43 cycles of denaturation for 30 s at 94°C, annealing/extension for 1 min at 57°C and enzyme deactivation for 10 min at 98°C (ramp rate 2°C/s at all steps).

Ms Jessica Whittle (MSc student in Mead group) as well as Ms Lauren Murphy (DPhil student in Psaila group) contributed to the analysis of the samples of the "uVTE" cohort.

As far as neonatal blood spot analysis is concerned, 5 ng of pre-amplified gDNA template in a 1:100,000 dilution was used as the DNA input in otherwise identical conditions to the aforementioned protocol. DNA extracted from the Guthrie card of a patient with *JAK2* wild-type systemic mastocytosis (DBS_1) and TF-1 cells were used as negative controls, whereas HEL cell DNA was used as the positive control, with multiple replicates per run. Presence of *JAK2V617F* in one of the samples was independently validated by nested PCR.

2.10.2.2 Droplet digital PCR instrumentation and configurations

All ddPCR reactions were performed using the Bio-Rad QX200 AutoDG Droplet Digital PCR System (Bio-Rad Laboratories), and ddPCR data analysis was done using the QuantaSoft Software, version 1.7.4, and the QX Manager Software, Standard Edition, version 1.2 (Bio-Rad Laboratories).

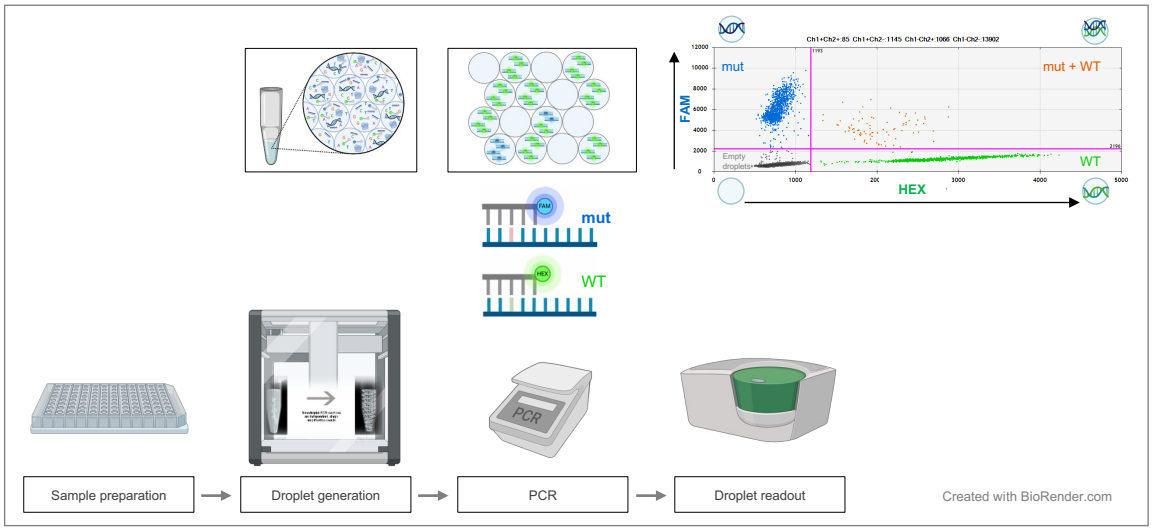


Figure 2-2. Droplet digital PCR analysis workflow. *Schematic representation of the instrumentation and process steps in ddPCR methodology: ddPCR reactions are prepared in 96-well plates with addition of primers, dual-labelled probes, DNA template, and the ddPCR supermix. Next, the plate is transferred to the droplet generator, which partitions samples into ~20,000 nanolitre-sized droplets per well. PCR amplification is carried out within each droplet using a thermal cycler. Lastly, the plate is transferred to the droplet reader where droplets are streamed in single file; fluorescent positive and negative events are counted to calculate target DNA concentration. Results can be visualised in the plot shown in the upper right corner of the figure, detailing positivity of each droplet for the fluorochromes used in the assay. ddPCR, droplet digital PCR; mut, mutated; WT, wild type.*

Table 2-1. Catalogue of oligonucleotides used in the PCR assays

Target_ nucleotide(s)	Nearest_ gene	Forward_primer_ sequence_5 → 3	Reverse_primer_ sequence_5 → 3	Probe_sequence	Product_ size (bp)	Company
chr19: 12943750-1294801	*CALR*	GCAGCAGAGAAACAAATG	GCCTCTCTACAGCTCGTCCTT		207	Invitrogen™
chr4: 105234449	*TET2*	ACGTAACTTCGGGGTAAGCC	TCCACGGAAGAGGCAGAAAC		291	Invitrogen™
chr4: 105259633	*TET2*	tgtgtggttatgccacagctt	acatcaggaagtaaacaaaacCTCT		312	Invitrogen™
chr1: 237523679	*RYR2*	ccagctgggagagaatatttgtga	ggagctgctaaaggttattctgc		1300	Invitrogen™
chr2: 236024598	*AGAP1*	ttgtccttttccccagaccc	catgaatccgggctgtcaaac		259	Invitrogen™
chr3: 187366417	–	gaatttaagccccaggaggaca	ccagtgtcagaggaaccgtag		268	Invitrogen™
chr16: 85313584	*AC092275.1*	tccagcgtgattccgtgttt	ttctgaccgcccagactttg		346	Invitrogen™
chr10: 18976904	–	aactcaacccttgtccccatc	cttttccctcccactgacaca		213	Invitrogen™
chr11: 78991748	*TENM4*	gcacttgttgaatgcttgtgc	ggtctgggcccacatttctt		390	Invitrogen™
chr7: 116155279	–	ttctagcccctgctatgcac	ggcccaatggcacagtaaga		452	Invitrogen™
chr8: 20863668	–	gccttcgattggggctgttta	aaggggttctcttccagcaaa		441	Invitrogen™
chr9: 5073770	*JAK2*	aggcttacacaggggtttcc	ttgtttgggcattgtaaccttc		505	Life Technologies™
chr9: 5073770 (gDNA)*	*JAK2*	GCTTTCTCACAAGCATTTG	gcattagaaagcctgtagtttta		84	Thermo Fisher Scientific
chr9: 5073770 (cDNA)	*JAK2*	GCTTTCTCACAAGCATTTG	ATTTTACAAACTCCTGAACCAGAA		83	Thermo Fisher Scientific
chr9: 5073770(G)	*JAK2*			[HEX]TCGTCTCCACAGACACATACTCCATGAGACGA[BHQ1]		Merck
chr9: 5073770(T)	*JAK2*			[6FAM]TCGTCTCCACAGAAACATACTCCATGAGACGA[BHQ1]		Merck

Note: *The *JAK2* ddPCR assay for gDNA was previously published by Cordua et al.(233) All target coordinates according to the Genome Reference Consortium Human Build 38 (GRCh38). All primers were checked against RefSeq and human genome assembly databases using NCBI Primer-BLAST® (National Library of Medicine, Bethesda) as well as SnapGene Viewer, version 5.3 (Dotmatics). The ddPCR assay probes were labelled with HEX and FAM for wild-type *JAK2* and the NM_004972.3[JAK2]:c.1849G>T p.V617F, respectively. Gene names are provided in italics in accordance with the current HUGO Gene Nomenclature Committee guidelines(173). Uppercase letters are used for exonic regions, while lowercase letters represent intronic regions. Invitrogen™, Life Technologies™, Thermo Fisher Scientific: Thermo Fisher Scientific Inc. ddPCR, droplet digital polymerase chain reaction. Table adapted from Nat Med, 28(6), Sousos, Ní Leathlobhair et al. In utero origin of myelofibrosis presenting in adult monozygotic twins, 1207-11, 2022; https://doi.org/10.1038/s41591-022-01793-4 (ref. (186)).

2.11 Electrophoresis

2.11.1 Agarose gel electrophoresis

Initial assessment of DNA fragment size was performed using gel electrophoresis on a 3% ethidium bromide-containing (10^{-4} g per litre) agarose gel in TAE buffer (Tris acetate-EDTA, Tris acetate 40 mM and 1 mM EDTA, pH 8.3) at 150 mV. Approximate DNA fragment size was determined using a 50-bp ladder (PCRBIO Ladder III, PCR Biosystems).

2.11.2 Automated electrophoresis

Separation, identification and quantitation of CALR type 1 / type 1-like mutations within the different cell populations examined was performed with use of the Agilent Fragment Analyzer automated fluorescence-based capillary electrophoresis and the DNF-905 dsDNA kit of 35–500 bp sizing range (Agilent Technologies), as well as the Agilent 2200 and Agilent 4200 TapeStation automated electrophoresis systems (with the A.0202 SR1 and the TapeStation Analysis Software 3.2, respectively), using the HS D1000 ScreenTape System (Agilent Technologies).

For nucleic acid quality control apart from NanoDrop One spectrophotometry of RNA (Thermo Fisher Scientific Inc.) and Qubit 2.0 fluorometry of DNA (Thermo Fisher Scientific Inc.), the Agilent 2100 Bioanalyzer system (Agilent Technologies) with either the RNA 6000 Pico kit (for blood product extracted RNA testing during platelet isolation optimisation studies) or the DNA 7500 kit (for NGS library quality control) was also used. In library preparation quality control studies TapeStation automated electrophoresis with the D5000 ScreenTape System (Agilent Technologies) was also used.

2.12 Sanger DNA Sequencing

Sanger DNA sequencing was performed in the MRC WIMM Sequencing facility using the Applied Biosystems 3730 DNA Analyzer (Thermo Fisher Scientific Inc.) with use of the BigDye Terminator version 3.1 chemistry. Results were visualised and annotated using the SnapGene Viewer software, version 5.3 (Dotmatics) and the Heatmapper visualiser(234).

2.13 Targeted next-generation sequencing

2.13.1 Extended next-generation sequencing myeloid-gene panel

2.13.1.1 Panel characteristics

Next-generation sequencing for the twins study, diagnostic accuracy studies of the custom-made NGS panel in CH studies, and molecular characterisation of the "TN and low-VAF *JAK2V617F* MPN patient" cohort was performed using an adapted method based on the previously described method by Silveira et al.(235). The method involves a custom designed SeqCap EZ HyperCap (Roche Sequencing Solutions, Inc.) probe library. The genes and transcript accessions covered by the panel are shown in Appendix B.1. The sensitivity (% mutant detectable in a background of wild type DNA) of SNV detection is between 5% to 10% and of insertion/deletion detection is 20%.

2.13.1.2 Library preparation

Library preparation performed on genomic DNA according to SeqCap EZ HyperCap Workflow, utilising the KAPA HyperPlus methodology with Dual-Indexed Adapters (Roche Sequencing Solutions, Inc.). Library preparation for this panel was performed by

the Molecular Diagnostics Centre, NIHR Biomedical Research Centre, OUH NHS Foundation Trust.

2.13.1.3 Sequencing platform

Sequencing carried out on the Illumina NextSeq 500 System and the NextSeq 500/550 Mid Output (300 cycles) kit version 2.5 (Illumina, Inc.)(236).

2.13.2 Custom next-generation sequencing panel

2.13.2.1 Panel characteristics

I developed and optimised a targeted, hybridisation capture NGS panel for parallel sequencing of both gDNA and cDNA, covering the major genes or hotspot mutations involved in CH and MPNs(44). I aimed for a sensitivity (% mutant detectable in a background of wild type DNA) of 1%, at least for gDNA, as VAFs of more than 1% in whole blood or the granulocyte compartment are most likely to carry a clinical significance in the setting of CH(37), MPNs(237), and other myeloid neoplasms(104, 238). I opted for a hybridisation capture rather than amplicon sequencing approach, as capture hybridisation allows for greater genomic coverage with higher depth and uniformity, providing superior sensitivity, while maintaining high specificity(239, 240). Oligos synthesis as well as scientific and technical support was provided by Twist Bioscience. Oligos were 120-nucleotide-long, the combined probe footprint size of the panel was 25,083 bp, and the full details of the areas covered in the panel are available in Table 2-2. Due to the hybrid nature of the panel, aiming to capture both gDNA and cDNA templates for the regions of interest, in the cases of exons less than 120 bp long, two separate probes (one gDNA-specific and another exon-exon junction cDNA-specific) were designed. In total, the panel consisted of 250 probes, including 27 cDNA-specific probes. Figure 2-3 provides an example of the

panel design methodology to enable parallel targeted sequencing of both gDNA and cDNA samples.

Table 2-2. Target regions for the hybridisation capture NGS panel enabling parallel sequencing of both gDNA and cDNA

Chr	Start	End	Strand	Gene_ name	Gene_ ID	Species	Assembly	Gene_ location	**Principal_ isoform(s)_ (Gencode33 /Ensembl99;_ APPRIS 2020_02.v30)**	Transcript_ accession_ID	CCDS	Target
chr 20	32358 330	32439 319	+	*ASXL1*	ENSG00000 171456	Homo Sapiens	GRCh38. p13	20: 32358330- 32439319	ENST00000375687	NM_015338.6	CCDS13201	**Selected Exons (12,13)**
chr 19	12938 578	12944 489	+	*CALR*	ENSG00000 179218	Homo Sapiens	GRCh38. p13	19: 12938578- 12944489	ENST00000316448	NM_004343.4	CCDS12288	**Selected Exon (9)**
chr 11	11920 6298	11931 3926	+	*CBL*	ENSG00000 110395	Homo Sapiens	GRCh38. p13	11: 119206298- 119313926	ENST00000264033	NM_005188.4	CCDS8418	**Selected Exons (8, 9)**
chr 2	25227 855	25342 590	-	*DNMT3A*	ENSG00000 119772	Homo Sapiens	GRCh38. p13	2: 25227855- 25342590	ENST00000264709; ENST00000321117	NM_022552.5; NM_175629.2	CCDS33157	**All Exons**
chr 1	18158 62	18157 56	-	*GNB1*	ENSG00000 078369	Homo Sapiens	GRCh38. p13	1: 1785285- 1891117	ENST00000378609	NM_002074.5	CCDS34	**Selected Exon (5)**
chr 15	90083 045	90102 504	-	*IDH2*	ENSG00000 182054	Homo Sapiens	GRCh38. p13	15: 90083045- 90102504	ENST00000330062	NM_002168.4	CCDS10359	**Selected Exon (4)**
chr 9	49843 90	51281 83	+	*JAK2*	ENSG00000 096968	Homo Sapiens	GRCh38. p13	9: 4984390- 5128183	ENST00000381652	NM_004972.4	CCDS6457	**Selected Exons (12, 14)**
chr 12	25205 246	25250 936	-	*KRAS*	ENSG00000 133703	Homo Sapiens	GRCh38. p13	12: 25205246- 25250936	ENST00000256078	NM_033360.4	CCDS8703	**Selected Exons (2, 3)**
chr 1	43337 818	43354 466	+	*MPL*	ENSG00000 117400	Homo Sapiens	GRCh38. p13	1: 43337849- 43352772	ENST00000372470	NM_005373.3	CCDS483	**Selected Exon (10)**
chr 17	60600 183	60666 280	+	*PPM1D*	ENSG00000 170836	Homo Sapiens	GRCh38. p13	17: 60600183- 60666280	ENST00000305921	NM_003620.4	CCDS11625	**Selected Exon (6)**
chr 2	19738 8515	19743 5091	-	*SF3B1*	ENSG00000 115524	Homo Sapiens	GRCh38. p13	2: 197388515- 197435079	ENST00000335508	NM_012433.4	CCDS33356	**Selected Exons (14, 15)**
chr 17	76734 115	76737 374	-	*SRSF2*	ENSG00000 161547	Homo Sapiens	GRCh38. p13	17: 76734115- 76737374	ENST00000359995; ENST00000392485	NM_001195427.2; NM_003016.4	CCDS11749	**Selected Exon (1)**
chr 4	10514 5875	10527 9816	+	*TET2*	ENSG00000 168769	Homo Sapiens	GRCh38. p13	4: 105145875- 105279816	ENST00000380013; ENST00000540549	NM_017628.4; NM_001127208.3	CCDS47120	**All Exons**
chr 17	76617 79	76875 50	-	*TP53*	ENSG00000 141510	Homo Sapiens	GRCh38. p13	17: 7661779- 7687550	ENST00000269305; ENST00000445888	NM_000546.5; NM_001126112.2	CCDS11118	**All Exons**
chr 21	43092 956	43107 587	-	*U2AF1*	ENSG00000 160201	Homo Sapiens	GRCh38. p13	21: 43092956- 43107587	ENST00000291552; ENST00000380276	NM_006758.3; NM_001025203.1	CCDS13694; CCDS33574	**Selected Exons (2, 6)**

Note: The list is alphabetised by gene name. Gene names are provided in italics in accordance with the current HUGO Gene Nomenclature Committee (HGNC) guidelines(173).

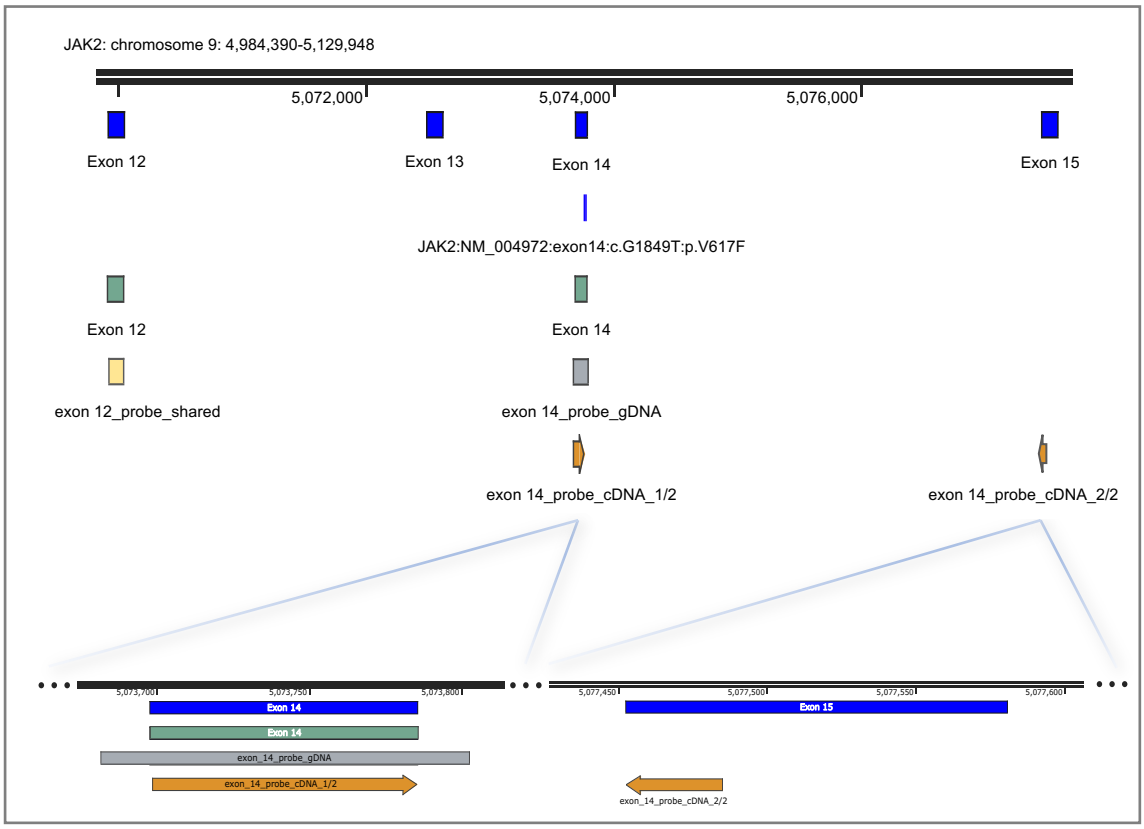


Figure 2-3. Custom targeted NGS panel design. *Example of the panel design methodology for parallel targeted sequencing of both gDNA and cDNA samples. Exonic regions are shown in blue. JAK2 target regions (shown in green) included exon 12 and exon 14. As exon 12 is 128 bp long, a single centrally-aligned 120-nucleotide-long probe (shown in yellow) could serve detection of variants within exon 12 for both gDNA and cDNA samples. However, exon 14 is only 88 bp long. Therefore, two separate 120-nucelotide probes, one gDNA-specific (shown in grey), extending to the flanking intronic regions, and one exon-exon junction cDNA-specific (shown in orange) were required for parallel analysis of gDNA and cDNA templates. NGS, next-generation sequencing.*

2.13.2.2 Library preparation

Library preparation was performed as per the Twist NGS Workflow manufacturer instructions, including enzymatic DNA fragmentation, end repair and dA-tailing; universal adapter ligation; PCR amplification using unique dual-indexed primers; bead-based, capture probe – pool hybridisation; and post-hybridisation target enrichment(241, 242).

2.13.2.3 Sequencing platform

Sequencing carried out in the MRC WIMM Sequencing facility using the Illumina NextSeq 500 System and the NextSeq 500/550 Mid Output (150 cycles) kit version 2.5 (Illumina, Inc.)(236).

2.14 Whole genome sequencing

2.14.1 Library preparation and sequencing platform

Paired-end sequencing of matched tumour and germline DNA samples for each twin was performed by Novogene (Novogene Co, Ltd.), and 350-bp insert libraries were prepared for all samples and run on the NovaSeq 6000 System (Illumina, Inc.) to generate 150-bp reads, aiming for 30× sequencing coverage.

2.15 Flow cytometry and cell sorting

2.15.1 Sample preparation for flow cytometry

Processed blood products were processed for flow cytometry analysis and fluorescence-activated cell sorting (FACS) as described below: previously cryopreserved peripheral blood mononuclear cell samples were retrieved from storage at the liquid nitrogen tank and kept at -80ºC for short-term storage. To initiate the thawing process, sample vials were incubated for up to 2 min at 37°C at the water bath, and immediately diluted in 30 ml of pre-chilled thawing medium (Iscove's Modified Dulbecco's Medium, IMDM [Gibco; Thermo Fisher Scientific Inc.] supplemented with 10% fetal bovine serum, FBS [MilliporeSigma; Merck)] and pre-heated/activated Dnase I [MilliporeSigma; Merck] at

an initial concentration of 1,000 mg per litre). This step aimed to minimize the cytotoxic effect of DMSO contained in the freezing medium (Appendix A.1). The volume of the medium was added in a dropwise manner, whilst ensuring that the sample preparations were loaded at a pre-chilled centrifuge (4°C) and spun at 200 g for 15 min with minimal centrifugal acceleration and deceleration settings for this step. Following the removal of the supernatants, cell pellets were resuspended in chilled "FACS-enrichment" medium (PBS supplemented with 5% FBS and 2 mM EDTA [Corning Inc]) and spun at 300 g for 10 min. After this wash step, cells were counted, and resuspended in FcR block solution (Miltenyi Biotec) for the preparation of controls that were used in the downstream flow cytometry analysis (Single Stain, Fluorescence Minus One, and unstained controls) and the fully stained samples. Volumes were adjusted to match half of the final staining volume of the above preparations. Samples were incubated for 10 min at 4°C and then subjected to their relevant antibody cocktails to be stained for 20 min at 4°C, spun at 300 g for 5 min, and resuspended in the "FACS-enrichment" medium supplemented with a viability (DNA-intercalating) dye for flow cytometry analysis and/or cell sorting. Depending on the appearance or re-emergence of clumps, samples were exposed to increasing amounts of dNase I (up to 10 mg per sample) and filtered using round bottom polystyrene FACS tubes equipped with 35 µm strainer caps. Samples were kept on ice until further processing.

2.15.2 Flow cytometry antibody panel characteristics

Staining panel details for granulocyte, T cell, HSPC, and platelet analysis and sorting are shown in Table 2-3. Antibody validation and overall quality performance of each panel was done with use of Single Stain, Fluorescence Minus One, and unstained controls.

Table 2-3. Flow cytometry and cell sorting panels used in granulocyte, T cell and HSPC analysis and sorting

(a) Granulocyte panel					
Antigen	**Clone**	**Conjugate**	**Dilution**	**Company**	**Cat. No.**
CD19	HIB19	V450	5:100	BD Biosciences§	560354
CD3	SK7	FITC	5:100	BD Biosciences§	345763
CD33	WM53	PE	5:100	BioLegend	303404
7AAD (7-Aminoactinomycin D)			1:100*	Cayman Chemical†	11397
CD11b	ICRF44	APC	5:100	eBioscience™#	17-0118-42
CD14	61D3	APC	5:100	eBioscience™#	17-0149-42
CD71	M-A712	Alexa Fluor® 700	5:100	BD Biosciences§	563769
CD34	4HI1	APC-eFluor® 780	2:100	eBioscience™#	47-0349-42
(b) T cell panel					
Antigen	**Clone**	**Conjugate**	**Dilution**	**Company**	**Cat. No.**
DAPI (4′,6-Diamidino-2-Phenylindole, Dilactate)			1:100*	Invitrogen™#	D3571
CD3	SK7	FITC	5:100	BD Biosciences§	345763
CD11b	ICRF44	PE-Cyanine5	5:100	BioLegend	301308
CD14	61D3	PE-Cyanine5	5:100	eBioscience™#	15-0149-42
CD4	SK3	APC	5:100	BD Pharmingen™§	565994
CD19	HIB19	APC-Cyanine7	5:100	BioLegend	302218
(c) HSPC panel					
Antigen	**Clone**	**Conjugate**	**Dilution**	**Company**	**Cat. No.**
CD90	5E10	Brilliant Violet 421™	3.33:100	BioLegend	328122
CD45RA	HI100	PE	0.66:100	BioLegend	304108
CD38	HIT2	PE-Texas Red®	4.66:100	Invitrogen™#	MHCD3817
7AAD (7-Aminoactinomycin D)			1:100*	Cayman Chemical†	11397
CD123	6H6	PE-Cyanine7	1.66:100	BioLegend	306010
CD34	4HI1	APC-eFluor® 780	0.66:100	eBioscience™#	47-0349-42
Lineage mix			21.66:100		
CD8a^	RPA-T8	FITC	1:100 of Lineage mix	BioLegend	301006
CD10^	HI10a	FITC	3.33:100 of Lineage mix	BioLegend	312208
CD20^	2H7	FITC	0.66:100 of Lineage mix	BioLegend	302304
CD66b^	G10F5	FITC	6.66:100 of Lineage mix	BioLegend	305104
(d) Platelet panel					
Antigen	**Clone**	**Conjugate**	**Dilution**	**Company**	**Cat. No.**
CD235a	HIR2 (GA-R2)	eFluor® 450	5:100	eBioscience™#	68-9987-42
CD3	SK7	FITC	5:100	BD Biosciences§	345763
CD42b	HIP1	PE	5:100	BioLegend	303905
7AAD (7-Aminoactinomycin D)			1:100*	Cayman Chemical†	11397
CD41a	HIP8	PE-Cy7	5:100	BioLegend	303718
CD11b	ICRF44	APC	5:100	BioLegend	301309
CD33	WM53	APC	5:100	BioLegend	303407
CD19	HIB19	APC-eFluor® 780	5:100	Invitrogen™#	47-0199-42

Note: *Viability stains were previously diluted 1:200 for 7-AAD and 1:100 for DAPI from their 5,000 mg per litre stock solutions. ^Lineage mix. §Becton, Dickinson and Company; #Thermo Fisher Scientific Inc.;

†Cayman Chemical Company. Cat. No., catalog number; HSPC, haematopoietic stem and progenitor cell. Table adapted from Nat Med, 28(6), Sousos, Ní Leathlobhair et al. In utero origin of myelofibrosis presenting in adult monozygotic twins, 1207-11, 2022; https://doi.org/10.1038/s41591-022-01793-4 (ref. (186)).

2.15.3 Flow cytometry instrumentation and configurations

2.15.3.1 Twin study

Flow cytometry and cell sorting was performed using a BD FACSAria Fusion Cell Sorter (Becton, Dickinson and Company) with FACSDIVA software version 8.0.1. Sorting of single cells into 96-well plates was performed using the automated cell deposition unit. Verification of the single-cell sorting mode was established by sorting single fluorescent beads (Alignflow Flow Cytometry Alignment Beads; Thermo Fisher Scientific Inc.) into a flat-bottomed 96-well tissue culture plate. A fluorescence microscope was used to visualise the beads to verify that no wells contained more than one fluorescent bead and that they were centrally positioned in the wells. Further flow cytometry analysis was performed using FlowJo software version 10.7 (Becton, Dickinson and Company).

2.15.3.2 Targeted genetic analysis of platelet lineage cells

Flow cytometry and cell sorting was performed using a BD FACSAria III Cell Sorter (Becton, Dickinson and Company) with FACSDIVA software version 8.0.1. BD FACS Accudrop technology was employed to determine drop delay and centre alignment for sorting in Eppendorf tubes at a mini-bulk scale. Further flow cytometry analysis was performed using FlowJo software version 10.7 (Becton, Dickinson and Company).

2.15.4 Surface marker definitions of haematopoietic stem and progenitor cells and mature blood compartments

Haematopoietic stem and progenitor cell subpopulations were phenotypically defined as follows: common myeloid progenitor (CMP): $Lin^{-}CD34^{+}CD38^{+}CD123^{+}CD45RA^{-}$;

granulocyte macrophage progenitor (GMP): $Lin^{-}CD34^{+}CD38^{+}CD123^{+}CD45RA^{+}$; HSC: $Lin^{-}CD34^{+}CD38^{-}CD90^{+}CD45RA^{-}$; lymphoid-primed multipotent progenitor (LMPP): $Lin^{-}CD34^{+}CD38^{-}CD90^{-}CD45RA^{+}$; megakaryocyte-erythroid progenitor (MEP): $Lin^{-}CD34^{+}CD38^{+}CD123^{-}CD45RA^{-}$; multipotent progenitor (MPP): $Lin^{-}CD34^{+}CD38^{-}CD90^{-}CD45RA^{-}$.

Regarding mature blood compartments, T cells were defined as $CD11bCD14^{-}CD19^{-}CD3^{+}$, granulocytes were defined as $CD3^{-}CD71^{-}CD19^{-}CD11bCD14^{+}CD33^{+}$, while platelets were defined as $CD3^{-}CD235^{-}CD19^{-}CD11bCD33^{-}CD41a^{+}CD42b^{+}$.

Flow cytometry was performed in the MRC WIMM Flow Cytometry facility with the significant contribution of Dr Sally-Ann Clark (Flow Cytometry Specialist).

2.16 Single-cell cloning assay

Single HSPCs ($7AAD^{-}$, Lin^{-}, $CD34^{+}$) were sorted into 96-well plates with 50 µl of MethoCult H4435 Enriched (STEMCELL Technologies). The mean fluorescence intensities of CD34, CD38, CD90, CD45RA and CD123 were also recorded for each individual HSPC sorted using the index sorting feature. Colony output was assessed on day 14 under direct light microscopy, and individual colonies were picked. All colonies were flash-frozen and stored at -80^{o}C for subsequent whole genome amplification.

2.17 Bioinformatic analysis

2.17.1 Extended next-generation sequencing myeloid-gene panel analysis

Bioinformatic analysis for the extended NGS myeloid-gene panel was carried out on an in-house pipeline developed within Molecular Diagnostics Centre, NIHR Biomedical Research Centre, OUH NHS Foundation Trust. It consists of alignment to build GRCh38 of the human genome using Burrows-Wheeler Aligner(243), variant calling with a combination of VarDict(244), Mutect2 (ref. (245)), and Pindel(246), and variant annotation with ANNOVAR(247).

Please note that SNVs with a minor allele frequency < 3% (< 5% for insertion/deletion variants) were not automatically called by the bioinformatic pipeline, but they were identified and reported following manual interrogation of alignment files when clinically indicated.

2.17.2 Custom next-generation sequencing panel analysis

Bioinformatic analysis for the custom myeloid-gene panel—namely preprocessing of DNA- and RNA-sequencing data, and subsequent variant detection, annotation, and filtering—was conducted by Dr Wei Xiong (Sean) Wen as described in paragraphs 2.17.2.1 and 2.17.2.2.

2.17.2.1 Preprocessing of DNA- and RNA-sequencing data

FASTQ files were preprocessed for downstream variant detection as previously described(248, 249). Nucleotide bases at the 3′ end of the sequencing reads with Phred score < 20 were trimmed using TrimGalore (version 0.6.5)(250). Trimmed reads were aligned to the human reference genome (GRCh38 build) using Burrows-Wheeler Aligner (version 0.7.17)(243) and STAR(251) for DNA- and RNA-sequencing data, respectively.

The resulting Sequence Alignment Map (SAM) were converted to Binary Alignment Map (BAM) and indexed using Samtools (version 1.9)(252). The sample IDs were added to the sequencing reads using the AddorReplaceReadGroups module from Picard (version 2.3.0). Next, duplicate reads were marked and removed using the MarkDuplicates module from Picard. For RNA-sequencing data, the splicing intervals represented by Ns in the CIGAR (Compact Idiosyncratic Gapped Alignment Report) string were additionally hard-clipped using the *SplitNCigarReads* from the Genome Analysis Toolkit (GATK) (version 4.2.0.0)(253). Finally, the base quality scores were subsequently recalibrated based on known single-nucleotide and indel polymorphism sites using the BaseRecalibrator and ApplyBQSR modules from GATK.

2.17.2.2 Variant detection, annotation, and filtering

Variant detection, annotation, and filtering were performed as previously described(249). Variant detection was performed using Mutect2 module from GATK with the following options to enable detection of variants with low VAF: --tumor-lod-to-emit 2.0– --min-base-quality-score 20– --disable-read-filter MateOnSameContigOrNoMappedMateReadFilter– --mitochondria-mode– --callable-depth 10– --max-reads-per-alignment-start 0– --annotation AS_FisherStrand– --annotation AS_QualByDepth– --annotation FisherStrand– --annotation QualByDepth. Variants detection was performed along the exons and the flanking 10 bp nucleotides (for splice site variants). Then, functional annotation of variants was performed using ANNOVAR(247), and the flanking 10 bp nucleotides of the variants were retrieved using BEDTools (version 2.29.2)(254). Finally, the following variants were excluded: (1) variants reported as germline polymorphisms, (2) variants located on the non-coding regions such as introns and 3′ and 5′ untranslated regions (UTRs), (3) variants located within homopolymer regions defined as regions with six or more consecutive identical nucleotides, (4) occurrence of at least three variants that are within a window of

20 bases, (5) variants that do not alter the open reading frame such as non-frameshift indels and synonymous variants, (6) variants with fisher strand bias > 20, (7) variants supported by less than five reads, (8) total coverage at variant site supported by less than 10 reads, (9) variants with VAF < 1%, and (10) variants located on soft-clipped bases (only for RNA-sequencing data).

2.17.2.3 Next-generation sequencing data analysis and visualisation

Automated variant calling was followed by variant curation, involving visual inspection of the detected variants on Integrated Genomics Viewer (IGV)(255, 256) and checking them against COSMIC (the "Catalogue of Somatic Mutations in Cancer") database(257), release v97 (November 2022) (https://cancer.sanger.ac.uk/cosmic). Dr Angela Hamblin contributed to the NGS data curation process.

Downstream analysis and visualisation of the NGS data were performed with use of Maftools(258) in RStudio version 2022.12.0+353 (ref. (259)) (R 4.2.2 [ref. (260)]), and Circos(261).

2.17.3 Whole genome sequencing analysis

Bioinformatic analysis of the WGS data was conducted by Dr Maire Ní Leathlobhair under the supervision of Professor David C. Wedge, as described in paragraphs 2.17.3.1 to 2.17.3.8.

2.17.3.1 Whole genome data processing and alignment

Whole genome sequence FASTQ files were processed and aligned as follows. The quality of short insert paired-end reads was assessed by FASTQC version 0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequencing reads were aligned to the human reference genome (GRCh38.p13) using the BWA-MEM (version

0.7.17)(243) aligner in default mode. Duplicates were marked using Picard tools version 2.3.0 (https://broadinstitute.github.io/picard/); indels were realigned and base quality scores were recalibrated according to GATK best practice to generate analysis-ready BAM files. The resulting sequencing coverage ranged from 36.5× to 42.4×.

2.17.3.2 Variant calling

We reported somatic SNVs and indels that overlapped across three different software packages: MuTect2 (GATK version 4.1.2.0)(245), strelka2 (version 2.8.4)(262), and Octopus (version 0.6.3-beta)(263). Random forest filtering was applied to both somatic and germline calls generated using the Octopus algorithm(263). Somatic variants notated as "somatic" and "PASS" in the VCF files generated by Octopus and strelka2 and "PASS" in the VCF files generated by MuTect2 were retained and constitute the high-confidence call set. Variants were classified as shared or unique based on overlap between the twin call sets.

2.17.3.3 Copy-number estimation

The Battenberg package (v.2.2.8; https://github.com/Wedge-lab/battenberg) was implemented in R (version 3.3.0)(260) to estimate tumour cellularity and detect clonal and subclonal copy-number aberrations (CNAs) for each sample(264). An approach similar to the R-only WGS pipeline described at https://github.com/Wedge-lab/battenberg/blob/master/inst/example/battenberg_wgs.R was used, including the preprocessing, allele-counting and phasing steps and using genome assembly hg38. Replication timing correction was not implemented, and structural variation (SV) breakpoint information was not included. Default parameters were used, but the PHASING_GAMMA parameter was increased to 2 to correct for overfitting. To run Battenberg in a way that was compatible with the hg38 genome build, Impute2 inputs were

converted to hg37 using liftOver(265). Heterozygous SNP positions were then phased using Beagle 5 (ref. (265)) with use of the 1000 Genomes genotypes as a reference panel and the output was converted to hg38. The "ntot" values for nearly all CNAs identified by Battenberg were approximately 2, suggesting that these were likely to be homozygous SNPs rather than somatic CNAs. To confirm whether these were heterozygous SNPs in the normal, distributions of allele frequencies of all SNPs within these CNAs in both the tumour and normal samples were inspected; copy-number segments were excluded where these distributions did not correspond to "real" somatic CNAs.

2.17.3.4 Mutation clustering

Copy-number changes called by Battenberg (as described in paragraph 2.17.3.3 Copy-number estimation) along with VAFs for each mutation were used to calculate CCF and prepared as input for DPClust(266), a Bayesian Dirichlet algorithm available at https://github.com/Wedge-lab/dpclust. Before running DPClust, we used the vafCorrect realignment tool (https://github.com/cancerit/vafCorrect) to calculate VAF values for all somatic variant calls (see also paragraph 2.17.3.2 Variant calling) from tumour and germline BAM files directly(267). The vafCorrect read mapping and base quality thresholds were set to 30. Variants with VAF > 0.05 in either germline sample or VAF < 0.1 in a tumour sample were flagged and removed prior to mutation clustering.

For each twin, we inferred the number of subclones and the fraction of tumour cells within each subclone using DPClust (version 2.2.2) implemented in R (version 3.4.0)(260) to cluster mutations according to their CCF. A pipeline similar to that described at https://github.com/Wedge-lab/dpclust/blob/master/inst/example/dpclust_pipeline.R was used with 1,000 iterations to run the Markov chain Monte Carlo simulation (no.iters) and

discarding 200 iterations as burn-in (no.iters.burn.in or burnin). Clusters that included less than 1% total mutations were removed.

DPClust was also used to cluster mutations according to their CCFs across samples from both twins simultaneously (Appendix B.2). DPClust input files for multidimensional DPClust contained all mutations across both samples. Variant sites with total coverage below 20 reads or with a mutation multiplicity estimate (no.chrs.bearing.mut) of zero in either sample were flagged and removed prior to multidimensional clustering. Depth at variant sites was determined using vafCorrect.

To perform statistical tests on individual mutations and establish CIs for CCF estimates, the binomial probability function pbinom was used in R. This returns the likelihood pbinom(x, y, z) of observing x or fewer mutant reads from a depth of y given an expected VAF of z.

2.17.3.5 Structural variation analysis

Structural variations were called using Manta (version 1.6.0; https://github.com/Illumina/manta)(268) and LUMPY (version 0.2.13; https://github.com/arq5x/lumpy-sv)(269). The LUMPY express wrapper was used to run LUMPY jointly on tumour-normal pairs. Structural variation breakpoints identified by the algorithm were then genotyped using SVTyper (version 0.7.1; https://github.com/hall-lab/svtyper)(270), run as a command line python script, and variants with any alt (AO) evidence in the normal were removed. In Manta, tumour-normal analysis workflows were configured using the configManta.py script with default settings; SVs were then filtered to include only calls marked as "PASS". For each twin, high-confidence somatic calls were derived by requiring either both callers to support an SV or one caller with additional support from a nearby CNV changepoint. In comparing alignments between Manta and

LUMPY, a bidirectional "slop" of 200 bp was used to account for possible imprecision in some reads. Structural variations were annotated using the online AnnotSV interface (version 3.0.2; https://lbgi.fr/AnnotSV/runjob)(271) using default settings.

2.17.3.6 Extraction of mutational signatures

Single-base substitution, doublet-base substitution and indel signatures were extracted by non-negative matrix factorisation, as implemented in SigProfilerMatrixGenerator version 1.1.23 (https://github.com/AlexandrovLab/SigProfilerMatrixGenerator)(272) using Python (version 3.8.3)(273). SigProfilerExtractor (version 1.1.0; https://github.com/AlexandrovLab/SigProfilerMatrixExtractor) was used to determine the proportion of mutations in each sample attributable to specific mutational signatures within the COSMIC database of signatures (version 3.1 in COSMIC release v91 (June 2020), available at https://cancer.sanger.ac.uk/cosmic/signatures)(257).

2.17.3.7 Putative driver mutation analyses

To identify non-silent coding mutations of interest, we considered 82 genes with relevant biological evidence(274-276) (Appendix B.3) as well as genes from the COSMIC Cancer Gene Census list (August 2020, http://cancer.sanger.ac.uk/census). Somatic variants were annotated using Ensembl Variant Effect Predictor release 103 (ref. (277)). Variants were considered pathogenic if their clinical significance was pathogenic, their impact was high or their CADD score was ≥ 30 (ref. (278)).

2.17.3.8 Bayesian estimates of myeloproliferative neoplasm developmental timing

An estimate for the time to the most recent common ancestor (MRCA) of MPN cells in both twins, tMRCA (that is the time at which *in utero* transmission took place), was derived using a Poisson Bayesian model. Known data used in the model included the ages of both

twins at sampling (39 years 11 months 26 days for twin A; 38 years 5 months 14 days for twin B) and mutation count data generated in this study. Using the pigeonhole principle, it was inferred that mutations from clusters 1 and 2 with CCF > 0.50 occurred on the same individual cell lineages in twin B and twin A, respectively (Appendix B.2). Mutations shared by the twins and occurring on the same cell lineage were assigned to cluster 3 (Appendix B.2). An initial model was built using these lineage-restricted counts (Table 2-4). The mutation rate and the number of SNVs (somatic variation data obtained from twin A and twin B) was modelled as mutations per genome per year. The prior on the mutation rate was exponential with mean of 10.6 clonal substitutions per genome per year; this estimate was based on reported clonal point mutation counts and donor ages in patients with MPN within the Pan-Cancer Analysis of Whole Genomes (PCAWG; see Gerstung et al., clonal counts from source data for Fig. 2 and donor ages Extended Data Fig. 8b)(279). Pre-MRCA mutation count data (SNVs shared by the twins and occurring on the same cell lineage) had observed values of 7. Post-MRCA mutation count data (lineage-restricted SNVs unique to each twin) had observed values of 91 and 90 for twin A and twin B, respectively. An upper bound on the rate parameter of the Poisson process (mutations per genome per year) was set based on the highest reported clonal substitution rate in a patient with MPN in PCAWG.

Results presented in the main text derive from a model using C>T somatic mutations in a NpCpG context (N[C>T]pG) (Table 2-4a), as these mutations are hypothesised to arise in a clock-like manner as a result of spontaneous deamination of 5-methylcytosine and can be used been to estimate MRCA timing in cancers(279, 280). N[C>T]pG mutation counts were extracted using the R/Bioconductor package MutationalPatterns(281) (version 1.12.0) implemented in R (version 3.6.0)(260). An important assumption of this model is that the N[C>T]pG mutation rate has remained constant over time in the MPN lineages in

both twins and that MPN lineages in twin A and B shared a common ancestor in the past. A further limitation of this model is that N[C>T]pG counts estimated from bulk data will not necessarily correspond to those arising on a single-cell lineage; 89 and 114 C>T somatic mutations in an NpCpG context were identified uniquely in twin A and twin B, respectively. Two shared N[C>T]pG somatic mutations were identified from the set of all shared somatic mutations (see also Chapter 3 Genetic lineage tracing in monozygotic twins with adult onset myeloproliferative neoplasms/3.1 Proof of study concept: a study of twins with *CALR* mutant myelofibrosis/3.1.1 Results/3.1.1.3 *In utero* origin of myeloproliferative neoplasm). Mutation counts were based on vafCorrect genotyping data after filtering (as described in paragraph 2.17.3.4 Mutation clustering). The prior on the mutation rate was exponential with a mean of 1.7 N[C>T]pG substitutions per genome per year; the mean estimated rate was based on reported N[C>T]pG counts and donor ages in patients with MPN within PCAWG (see Gerstung et al., Extended Data Fig. 8b)(279). Forty weeks of gestation time was also added to donor ages when generating mutation rate estimates from PCAWG data. An upper bound on the rate parameter of the Poisson process (mutations per genome per year) was set based on the highest reported N[C>T]pG rate in a patient with myeloid-MPN in PCAWG. The prior on tMRCA (that is, the span of time after zygote over which shared N[C>T]pG mutations arose) was defined as uniform between the time of fertilisation and the age of earliest presentation in the twins. t[1] and t[2] are time intervals during which subsequent (post-tMRCA) N[C>T]pG mutations arose in twin A and twin B, respectively. The prior on these intervals was defined as uniform between tMRCA and the time of sampling (approximately 40 years of age).

Mutation counts were modelled using a Poisson likelihood with rates given by the product of the mutation rate and the relevant time interval (that is, the time interval from fertilisation up to tMRCA for pre-MRCA count data and the time interval between tMRCA

up to the time of sampling for post-MRCA count data). The No-U-Turn sampler, implemented in Stan using the rstan package (https://CRAN.R-project.org/package=rstan; version 2.21.2, GitRev: 2e1f913d3ca3), was used to draw samples from the model's posterior distribution through Markov chain Monte Carlo sampling using five chains, each with 100,000 iterations, of which the first 20,000 are warmup(282). Log marginal likelihoods and error measures for marginal likelihood estimates were computed via bridge sampling using the bridgesampling package in R(283). Stan model outputs, likelihood estimates and error measures are reported in Table 2-4. A model using lineage-restricted or clonal N[C>T]pG mutations could not be implemented, as the mutation count data were too sparse.

Table 2-4. Data for time to the most recent common ancestor (MRCA) estimation

(a) Using N[C>T]pG mutation count data							
	mean	**se_mean**	**SD**	**95%_CI_lower_ confidence _limit**	**95%_CI_upper_ confidence _limit**	**n_eff**	**Rhat**
t_{MRCA}	0.95	0.00	0.54	0.20	2.25	287761.60	1.0000
t[1]	2.80	0.00	0.15	2.45	2.99	211004.80	1.0000
t[2]	31.87	0.01	3.37	25.33	38.16	188785.40	1.0000
rate	36.69	0.00	1.93	32.04	39.30	294264.00	1.0000
lp__	732.06	0.01	1.64	727.95	734.12	106813.70	1.0000
Bridge sampling estimate of the log marginal likelihood: -10.93478; Relative Mean-Squared Error: 1.756122e-06.							
(b) Using lineage-restricted SNV count data							
	mean	**se_mean**	**SD**	**95%_CI_lower_ confidence _limit**	**95%_CI_upper_ confidence _limit**	**n_eff**	**Rhat**
t_{MRCA}	2.16	0.00	0.98	0.61	4.37	78069.74	1.0000
t[1]	4.12	0.01	2.06	2.43	9.95	44277.66	1.0001
t[2]	25.22	0.03	7.53	9.06	36.49	51312.44	1.0000
rate	24.97	0.03	7.50	8.95	36.41	51542.02	1.0000
lp__	644.68	0.01	1.65	640.58	646.79	77872.47	1.0001
Bridge sampling estimate of the log marginal likelihood: -9.01966; Relative Mean-Squared Error: 9.335533e-06.							

Note: t_{MRCA} is an estimate of time (years) between fertilisation and the branch point of MPN lineages in twin A and B. t[1] and t[2] are estimates of the time (years) to the MRCA of cells in twin A and twin B. lp__ corresponds to the logarithm of the (unnormalised) posterior density as calculated by Stan. n_eff is a crude measure of effective sample size and "Rhat" is the potential scale reduction factor on split chains (Rhat = 1 at convergence). se_mean is the estimated standard error of the mean of the simulations. rate is the (a) N[C>T]pG (b) clonal mutation rate per genome per year. SNV, single-nucleotide variant.

The bioinformatic pipeline for WGS data analysis was developed and run by Dr Maire Ní Leathlobhair, under the supervision of Professor David C. Wedge.

2.17.4 Single-cell index-sorting

Visualisation of the cell surface HSPC markers expression for each cell and genotyping information was performed with in-house bioinformatic pipelines using ggplot package and performed in R (version 3.6.3)(260). Analysis was conducted by Dr Eleni Louka.

2.18 Statistical analysis

For all statistical analyses significance level was set to .05. All analyses were performed using GraphPad Prism versions 8 and 9 (Dotmatics).

Chapter 3 Genetic lineage tracing in monozygotic twins with adult-onset myeloproliferative neoplasms[3]

3.1 Proof of concept: a study of twins with *CALR*-mutant myelofibrosis

As a first step in my DPhil project, I aimed to highlight the importance of genetic lineage tracing in the study of CH and MPNs. In this proof-of-concept work, I conducted a twin study in a pair of monozygotic twins who presented with *CALR*-mutant PMF in their late 30s. Given the very low likelihood that the twins independently acquired the very same mutation during their life, I hypothesised that the twins shared the mutation through either vertical or horizontal gene transfer. To test this hypothesis, I initially conducted lineage tracing of the *CALR* mutation by fragment analysis followed by lineage tracing of the MPN clone by WGS of granulocytes (derived by the MPN clone) against T cells (as the germline control) from both twins.

3.1.1 Results

3.1.1.1 Clinical findings

Twin A presented at the age of 37 years with symptomatic splenomegaly, measuring 21 cm on computed tomography imaging. Blood parameters were haemoglobin 138 g per litre, white cell count 5.9 $\times 10^9$ per litre and platelets 183 $\times 10^9$ per litre. Blood film examination revealed leukoerythroblastic picture (Figure 3-1a,b). The bone marrow showed typical features of PMF with distorted intertrabecular spaces, prominent dilated sinuses and

[3]Chapter 3 was adapted from Nat Med, 28(6), Sousos, Ní Leathlobhair et al., In utero origin of myelofibrosis presenting in adult monozygotic twins, 1207-11, 2022; https://doi.org/10.1038/s41591-022-01793-4.

evidence of heavily disrupted haematopoiesis with moderate numbers of highly atypical, small- to medium-sized Mks forming small clusters, and WHO MF-3 fibrosis (Figure 3-1c,d). Genetic testing revealed presence of a type 1 *CALR* mutation (NM_004343.4[CALR]:c.1092_1143del52 p.L367fs*46), a highly recurrent mutation in MPN(167, 168). Targeted NGS revealed the presence of a frameshift *TET2* mutation (NM_001127208.2[TET2]:c.509del p.N170Tfs*13). The overall clinical presentation, genetic and morphological features were in keeping with a diagnosis of PMF. Due to disease-associated symptoms, twin A commenced treatment with the JAK1/2 inhibitor ruxolitinib with good clinical response. He remains on a stable dose of ruxolitinib (10 mg twice per day), with no disease-associated symptoms and reduction in spleen size to 17.6 cm on the most recent imaging. Blood parameters at the time of sampling for our studies were haemoglobin 138 g per litre, white cell count 7.3×10^9 per litre and platelets 62×10^9 per litre.

Twin A's identical twin brother, twin B, presented one year later at the age of 38 years with marked splenomegaly, measuring 20 cm below the costal margin, and associated early satiety. Blood parameters were normal apart from borderline leukopenia (white cell count, 3.88×10^9 per litre). Blood film examination revealed the presence of leukoerythroblastic picture (Figure 3-1e,f). Bone marrow biopsy showed a hypocellular marrow with markedly abnormal bony trabeculae with irregular outlines and areas of notable thickening, alongside areas of increased reticulin fibres (Figure 3-1g,h). Genetic testing revealed presence of the same 52-bp deletion type 1 *CALR* mutation as seen in twin A. Targeted NGS revealed no additional mutations in twin B. Taking clinical, genetic and morphological features into account, a diagnosis of PMF was made. He remains well, with no remarkable disease-associated symptoms, and to date, he has required no treatment.

There was no history of transfusion for either twin or other family history of MPN or other blood cancer. Information regarding whether the twins shared a monochorionic placenta was not available. The twins lived in the same household until the age of 25 years. Both are non-smokers who do not drink alcohol, and they both have a normal body mass index and no autoimmune or other chronic medical conditions apart from MPN. They report no known exposure to ionizing radiation or occupational and chemical exposure.

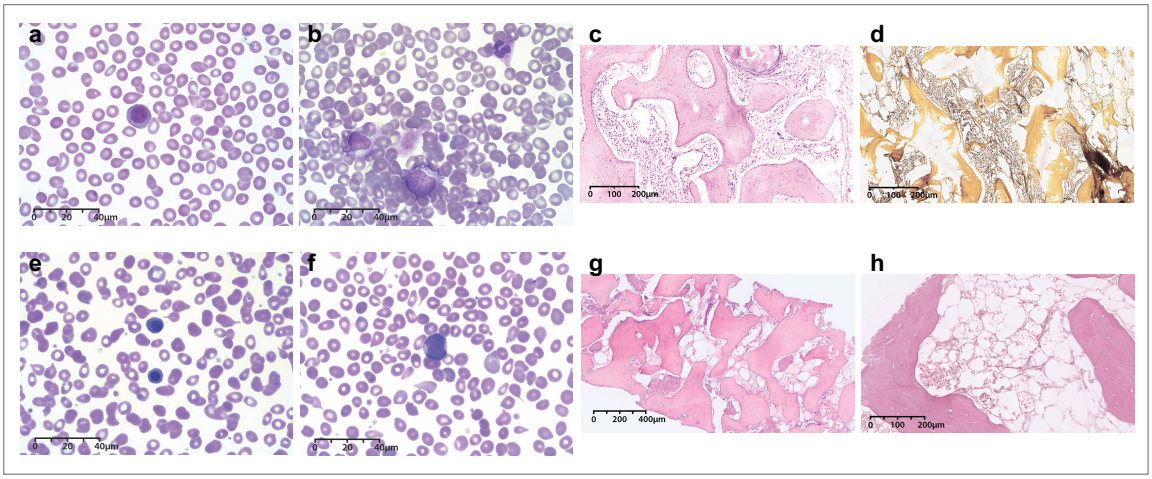


Figure 3-1. Blood film and bone marrow examination findings. ***a,b****, Leukoerythroblastic blood film of twin A (May-Grunwald-Giemsa stain; 100×).* ***C****, Bone marrow trephine examination from twin A at diagnosis (haematoxylin and eosin stain). Reticulin stain (****d****) showed diffuse fibre network with scattered coarse fibres (WHO MF-3) (reticulin stain).* ***e,f****, Leukoerythroblastic blood film of twin B (May-Grunwald-Giemsa stain; 100×).* ***G****, Bone marrow trephine examination from twin B (haematoxylin and eosin stain). Reticulin stain (****h****) revealed evidence of patchy, irregular fibrosis and a hypocellular marrow (reticulin stain).*

3.1.1.2 Cell lineage specific mutation analysis

As the concordant disease in both twins might have occurred due to germline *CALR* mutation, I carried out mutational analysis of *CALR* in purified T cells and myeloid cells for each twin and skin fibroblasts from twin B by fragment analysis. Immunomagnetic separation through double-filtration CD3-enrichment of cryostored MNCs was selected as the best source of germline DNA after comparison with other sources and techniques of T cell purification by fragment analysis (Figure 3-2a), while purity was further confirmed by

flow cytometry analysis (Figure 3-2b). The *CALR* mutation was present in myeloid cells, but only at a very low level in T cells from both twins, while it was not detected in skin fibroblasts from twin B (Figure 3-3). These findings confirm that the *CALR* mutation was not present in the germline and was somatically acquired.

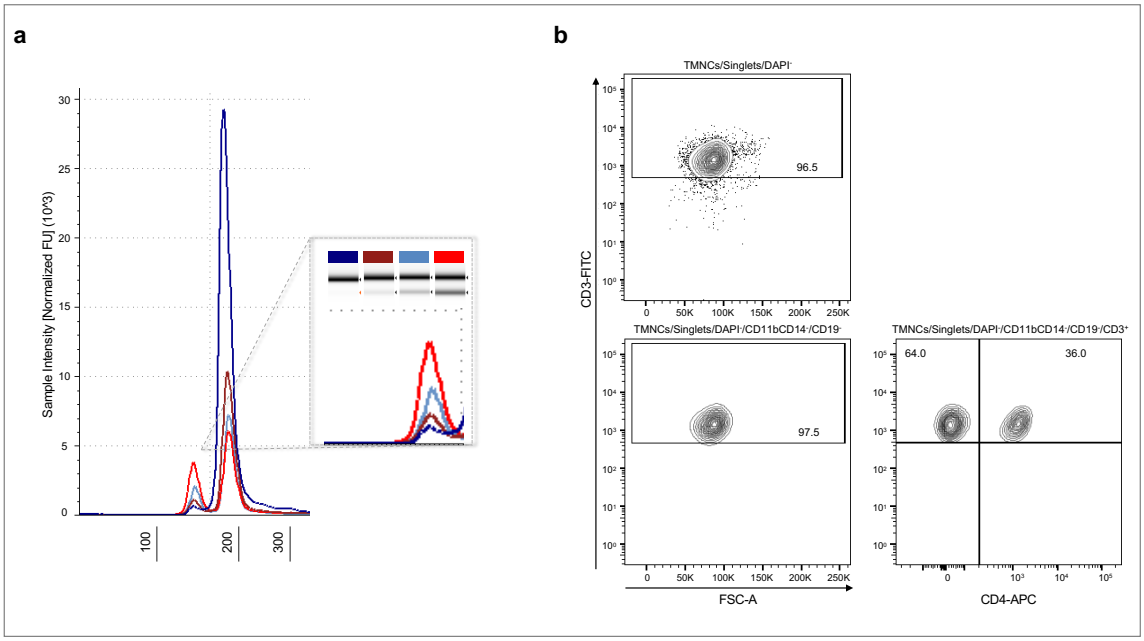


***Figure 3-2. Optimisation of germline DNA isolation**. **a**, Comparison of different sources of germline DNA for twin A through mutational analysis of CALR. Automated electrophoresis for double-filtration CD3-enriched cell DNA is shown in dark blue, nail DNA in brown, single-filtration CD3-enriched cell DNA in light blue, while CD34-enriched cell DNA is shown in red. CALRdel52bp VAF was calculated 1.8%, 9.0%, 10.8%, and 37.1%, respectively. **b**, T cell purity assessment of the double-filtration CD3-enriched population by flow cytometry. VAF, variant allele frequency.*

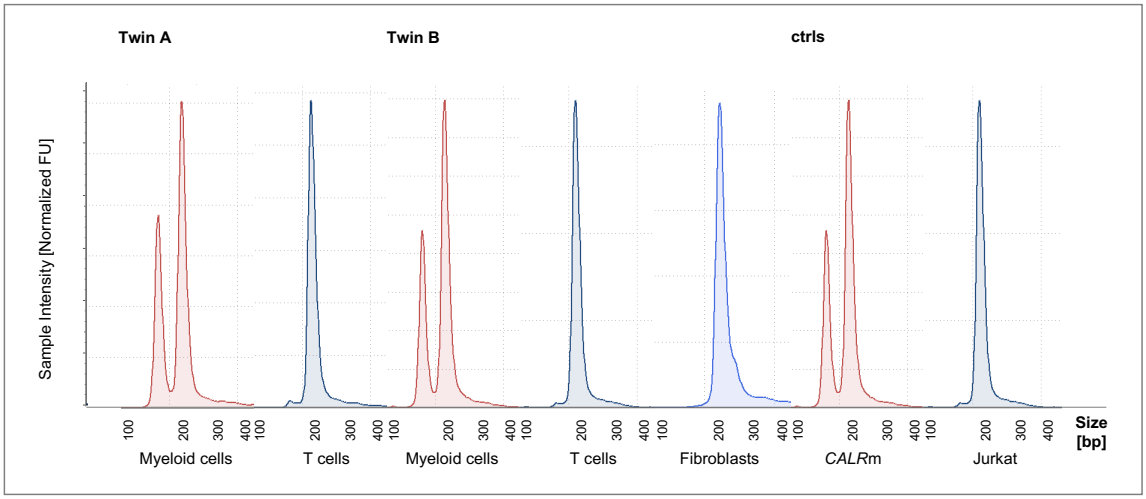


Figure 3-3. Genetic lineage tracing confirms that CALR was somatically acquired in both twins. a*, Cell-lineage-specific CALR mutational analysis. T cell DNA electropherograms (blue) have a single peak at 207 bp, whereas DNA extracted from myeloid cells (red) shows two peaks, one at 207 bp and another one corresponding to the 52-bp deletion fragment (CALRdel52bp VAF was calculated 28.4%, 32.6% and 43.7 for twin A, twin B and CALR type 1 myelofibrosis (CALRm) control, respectively). The absence of CALR 52-bp deletion in germline DNA was confirmed by analysis of dermal fibroblast DNA from twin B. DNA from a patient with known CALR type 1 myelofibrosis and DNA from Jurkat cells were used as positive and negative controls, respectively. Rearranged electropherograms from parallel experiments, all scaled to sample. ctrls, controls; CALRm, CALR mutation-positive; FU, fluorescence units; VAF, variant allele frequency.*

3.1.1.3 In utero origin of myeloproliferative neoplasm

I next reasoned that either the somatic *CALR* mutation could have been independently acquired by both twins or the MPN was transmitted by intraplacental twin-to-twin transfer. To examine these two possibilities, I carried out lineage tracing of the MPN clone by WGS of peripheral blood granulocytes (derived from the MPN clone) and T cells (germline control) from both twins. If the MPN cells in both twins had a shared clonal origin, then we would expect to find a number of shared somatic variants in the granulocytes of both twins. We identified 514 and 705 somatic SNVs, 240 and 44 somatic indels and 5 SVs in twin A and twin B, respectively (Figure 3-4,5,

Table 3-1,2, and Appendix B.4)(245, 262, 263, 268-270). The mean genome-wide substitution rate was 0.24 mutations per megabase, which represents a low mutational load, in line with the silent mutational landscape of MPNs(284). In addition to CALR p.L367fs*46, which was the only shared mutation affecting a coding region, nine high-confidence SNVs were also detected in both twins, strongly supporting the *in utero* origin of the MPN (Figure 3-4 and Table 3-2).

No significant difference in overall somatic mutation burden was observed between the twins, although SNV and indel burdens differed when considered separately. Non-silent somatic coding mutations exclusive to twin A included a frameshift indel not previously reported in *TET2*, an epigenetic regulator frequently mutated in haematological cancers, and a stop codon in *OTOG* associated with non-syndromic hearing loss(285), whereas twin B carried missense mutations in *MED27*, *ENKUR* and *CEP290* (Figure 3-4a and Table 3-3)(277). The *ENKUR* variant is reported as a driver in the COSMIC database(286), but it is not associated with haematological cancer. Notably, the targeted panel analysis for twin A showed presence of two *TET2* mutations at chr4(GRCh38):105234449 (NM_001127208.2[TET2]:c.509del p.N170Tfs*13) and chr4(GRCh38):105259633 (NM_001127208.2[TET2]:c.3818G>A p.C1273Y), with VAFs 9.21% and 8.63%, respectively. Both mutations were validated by Sanger sequencing both in bulk populations and in single-HSPC colonies (data not shown). The frameshift mutation was also detected in the WGS data, but the substitution was supported from a single caller only and therefore did not meet the variant calling criteria set in the study (as described below). This is a missense mutation that is not annotated in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) database, but it appears in COSMIC as pathogenic (https://cancer.sanger.ac.uk/cosmic/mutation/overview?id=145018270), with a CADD score of 27.2 (ref. (278)), while the SIFT(287) and PolyPhen(288) tools predict its effect

to be deleterious and probably damaging, respectively. However, because it was not included in the high-confidence set for WGS data, it was excluded from the downstream analysis. Analysis of the mutational signatures by non-negative matrix factorisation(272) revealed the presence of signatures 1, 5 and 19, all previously reported in MPNs, with near-identical contributions in twin A and twin B (Figure 3-5a)(279). However, indel signatures differed markedly between the twins; ID1 and ID12 accounted for respectively 57% and 43% of indels in twin A, whereas all indels in twin B were attributable to ID9 (Figure 3-5b)(286). This finding supports that the mutational processes were distinct, even though disease latency was almost identical in both twins. To assess clonal architecture, we clustered mutations based on cancer cell fraction (CCF), determined by adjusting the variant allele frequencies of SNVs for copy-number status and sample purity estimates (Figure 3-4a, Figure 3-6)(266).

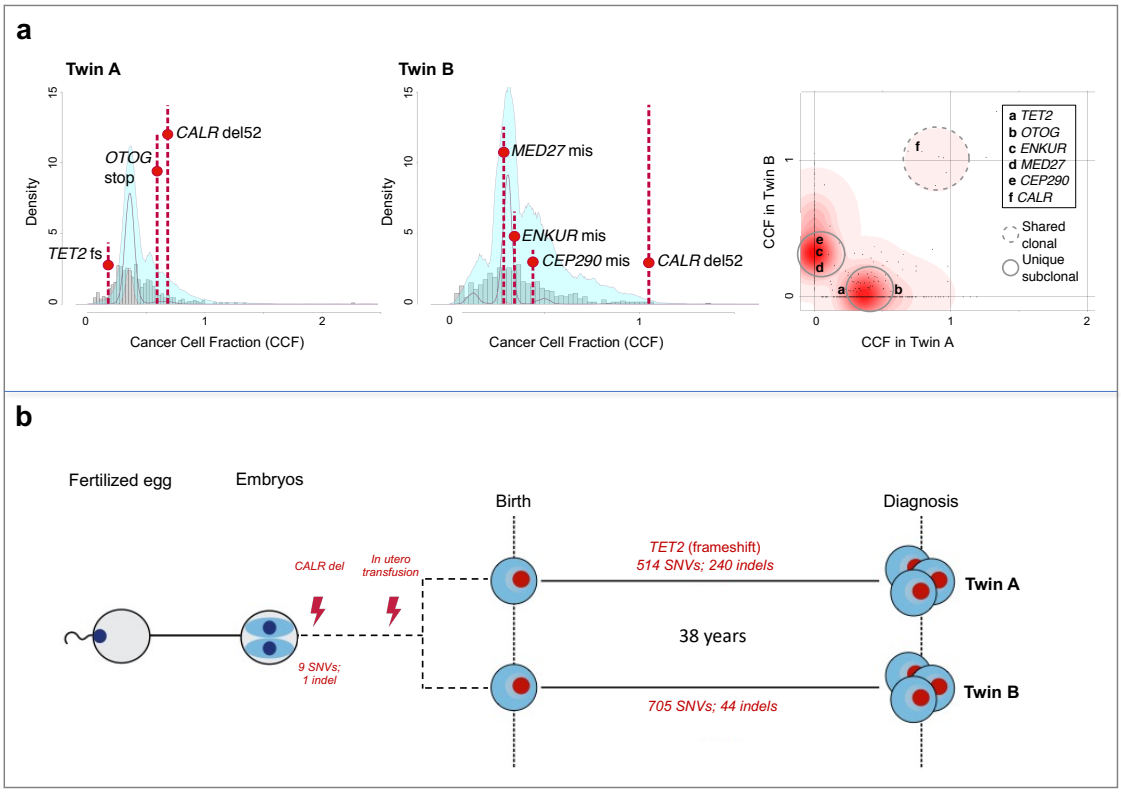


Figure 3-4. Genetic lineage tracing confirms a common in utero clonal origin of CALR mutation. ***a****, Statistical modelling of the distribution of subclonal and clonal mutations by Dirichlet-process clustering. The histogram of mutations is represented with grey bars, with the fitted distribution as a grey line; 95% posterior confidence intervals for the fitted distribution are also shown (pale blue area). In the rightmost panel, a two-dimensional density plot shows Dirichlet clustering of the fraction of cancer cells (CCF) within each twin for all somatic mutations detected (black dots). Higher posterior probability of a cluster is indicated by increasing intensity of red. The cluster indicated around (1,1) corresponds to mutations present in all cells in both twins; clusters along the axes correspond to twin-specific clones. A subset of non-silent coding mutations found in TET2, OTOG, ENKUR, MED27, CEP290 and CALR are highlighted.* ***b****, Schematic representation of the shared in utero clonal origin and number of high-confidence shared variants and separate postnatal clonal evolution with total number of SNV and indels for each twin shown. fs, frameshift; mis, missense; SNV, single-nucleotide variant.*

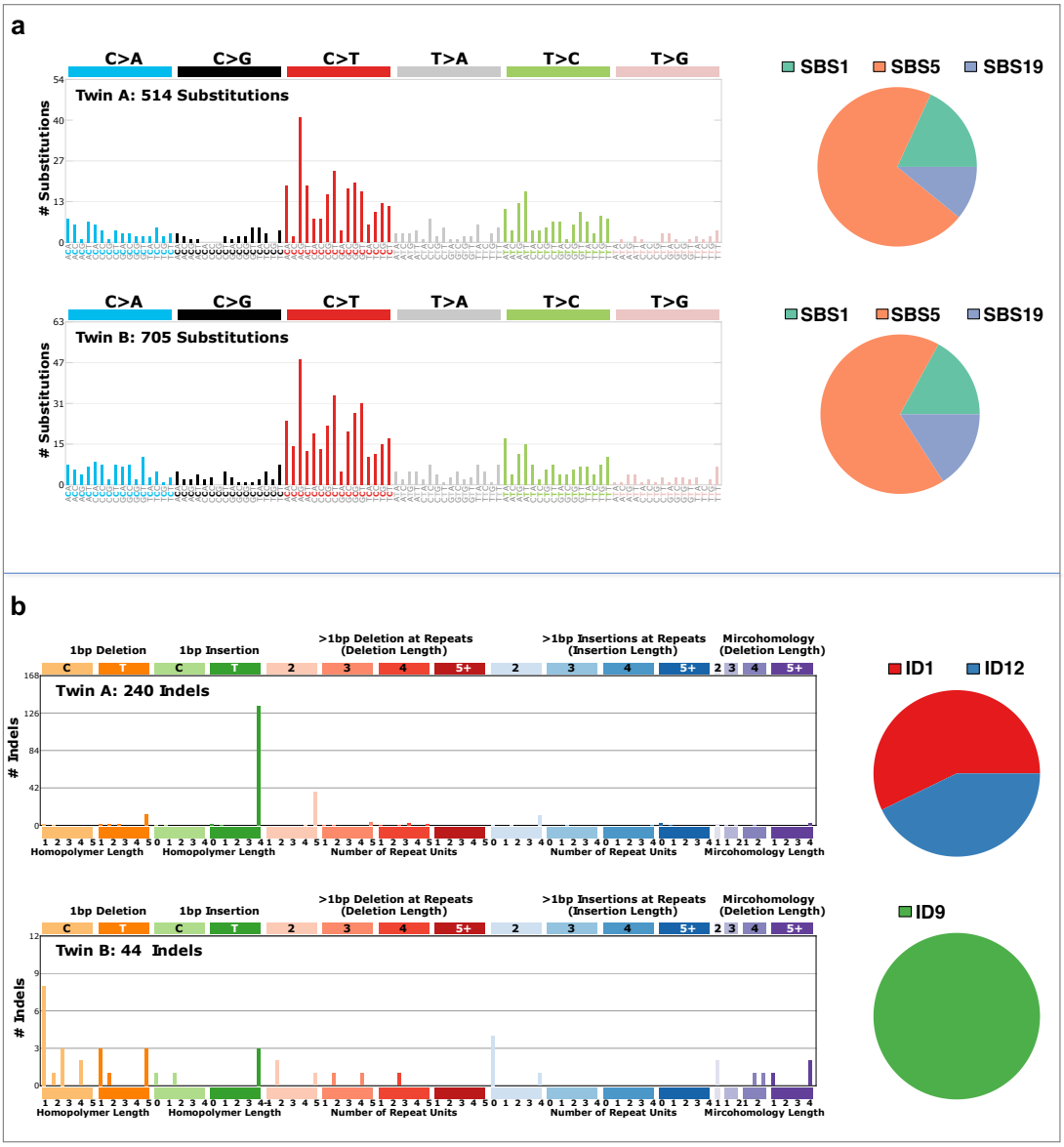


Figure 3-5. (a) Somatic point substitution and (b) indel spectra with corresponding mutational signatures for each of the twins. *We identified 514 and 705 somatic SNVs, 240 and 44 somatic indels, and 5 SVs unique to twin A and twin B, respectively. Analysis of the somatic point substitution signatures revealed shared signatures with near identical contributions in twin A and B. Indel signatures differed markedly between the twins; ID1 and ID12 accounted for respectively 57% and 43% of indels in twin A, while all indels in twin B were attributable to ID9. ID, indel signature; SBS, single-base substitution signature; SNV, single-nucleotide variant; SV, structural variation.*

Table 3-1. Catalogue of somatic SV in twin A and twin B

Individual	SV_class	SV_length	Chr1	Pos1	Chr2	Pos2
Twin_A	tandem_duplication	165127	11	54384289	11	54549416
Twin_A	tandem_duplication	71117	12	11054568	12	11125685
Twin_A	tandem_duplication	75902	20	23727447	20	23803349
Twin_A	translocation	.	6	[chr6:50867003[G	12	[chr12:93869517[C
Twin_A	translocation	.	3	]chr3:11573465]A	15	T[chr15:78040845[
Twin_B	tandem_duplication	272076	11	26184499	11	26456575
Twin_B	deletion	397816	11	48727000	11	55124816
Twin_B	translocation	.	6	[chr6:43581762[C	19	[chr19:41643986[C
Twin_B	deletion	604488	2	92998101	1	93602589
Twin_B	translocation	.	5	G]chr5:248313]	5	T]chr5:1579393]

SV, structural variation.

Table 3-2. Summary of somatic variants shared between twin A and twin B

#	Chr	Pos	Ref	Alt	Mutation_ type	Annotation	Shared_ annotation	Gene	Existing_ variation	CCF_in_ Twin_A	CCF_in_ Twin_B	Used_in_ Validation
1	1	237523679	G	A	Substitution	intron_ variant	High-confidence call set; vafCorrect genotyping	*RYR2*	rs1165384474	0.93	0.54	Yes
2	2	236024598	A	G	Substitution	intron_ variant	High-confidence call set; vafCorrect genotyping	*AGAP1*	-	1.05	0.99	Yes
3	3	9646697	T	A	Substitution	upstream_gene_ variant	High-confidence call set; vafCorrect genotyping	*MTMR14*	rs1001727401	0.93	1.06	No
4	3	187366417	G	A	Substitution	upstream_gene_ variant	High-confidence call set; vafCorrect genotyping	*RTP4*	-	1.01	1.16	Yes
5	5	18474168	C	T	Substitution	intergenic_ variant	High-confidence call set; vafCorrect genotyping	-	-	1.17	0.92	No
6	7	54587046	C	A	Substitution	intergenic_ variant	High-confidence call set; vafCorrect genotyping	-	-	0.89	0.37	No
7	10	103273122	G	A	Substitution	upstream_gene_ variant	High-confidence call set	*INA*	rs1286474689	0.39	0.57	No
8	16	67608941	A	G	Substitution	intron_ variant	High-confidence call set	*CTCF*	rs892105462	0.97	0.25	No
9	16	85313584	C	A	Substitution	downstream_gene_ variant	High-confidence call set; vafCorrect genotyping	*AC092275.1*	rs1483404740	1.03	1.01	Yes

10	19	12943750	AGCAGAGGCTTAAGGAGGAGGAAGAAGACAAGAAAACGCAAAGAGGAGGAGGA	A	Deletion	frameshift_variant	High-confidence call set	*CALR*	rs1555760738	0.69*	1.07	Yes
11	18	9601220	A	G	Substitution	intron_variant	vafCorrect genotyping	*PPP4R1*	rs982592192, COSV53280449	0.49	0.33	No
12	2	93865725	G	C	Substitution	intergenic_variant	vafCorrect genotyping	-	-	2	1.4	No
13	4	119161162	A	C	Substitution	intron_variant	vafCorrect genotyping	*MYOZ2*	-	0.75	0.2	No
14	4	16908350	G	T	Substitution	intergenic_variant; regulatory_region_variant	vafCorrect genotyping	-	rs750655685	0.79	1.06	No
15	5	130265666	C	T	Substitution	intergenic_variant	vafCorrect genotyping	-	rs62391477	0.61	0.68	No
16	7	71219697	C	T	Substitution	intron_variant	vafCorrect genotyping	*GALNT17*	-	0.77	0.82	No
17	1	93828702	C	CT	Insertion	intron_variant; regulatory_region_variant	vafCorrect genotyping	*BCAR3*	rs969811086	0.654	0.31	No
18	2	193562677	CGT	C	Deletion	intergenic_variant	vafCorrect genotyping	-	rs59053823	0.53	0.26	No
19	4	26343555	G	GT	Insertion	upstream_gene_variant	vafCorrect genotyping	*RBPJ*	rs969811086	0.43	0.26	No
20	4	61895466	AAAT	A	Deletion	upstream_gene_variant	vafCorrect genotyping	*ADGRL3*	rs969811086	0.62	0.26	No
21	6	7238595	C	CT	Insertion	intron_variant	vafCorrect genotyping	*RREB1*	rs796560101	0.41	0.28	No
22	7	43378810	C	CA	Insertion	intron_variant	vafCorrect genotyping	*HECW1*	rs202007736	0.5	0.23	No
23	8	17691704	G	GT	Insertion	regulatory_region_variant	vafCorrect genotyping	-	rs200014137	0.3	0.32	No
24	9	67205531	TG	T	Deletion	intron_variant	vafCorrect genotyping	*CNTNAP3P2*	rs368975415	0.3	0.5	No

Note: Gene annotation not provided for intergenic variants. CCF, cancer cell fraction, i.e., estimate of the fraction of tumour cells that carry the mutation. Somatic shared variants from the High-confidence call set are reported in both Twin A and Twin B in Appendix B.4. Somatic shared variants identified after vafCorrect genotyping are reported in either Twin A or Twin B in Appendix B.4 (see Chapter 2 Materials and Methods/2.17 Bioinformatic analysis/2.17.2 Whole genome sequencing analysis: 2.17.3.2 Variant calling and 2.17.3.4 Mutation clustering). Gene names are provided in italics in accordance with the current HUGO Gene Nomenclature Committee (HGNC) guidelines(173). *95% CI for *CALR* CCF: 0.42 to 1.00.

Table 3-3. Summary of non-silent coding variants in twin A and twin B

Individual	Chr	Pos	Ref	Alt	Mutation_ type	Annotation	Impact	Gene	Existing_ variation	Clinical_ significance	CADD_ PHRED
Twin_A	1	53327841	CA	C	Deletion	frameshift_ variant	HIGH	*LRP8*	rs761955852, COSV60099910	-	22
Twin_A	4	105234449	TA	T	Deletion	frameshift_ variant	HIGH	*TET2*	-	-	-
Twin_A	4	105259633*	G	A	Substitution	missense_ variant	MODERATE	*TET2*	COSV54404036	pathogenic	27.2
Twin_A	11	17634218	C	T	Substitution	stop_gained	HIGH	*OTOG*	rs866476223	pathogenic	40
Twin_A	19	12943750	AGCA GAGG CTTAA GGAG GAGG AAGA AGAC AAGA AACG CAAA GAGG AGGA GGAG	A	Deletion	frameshift_ variant	HIGH	*CALR*	rs1555760738	pathogenic	35
Twin_B	9	131860557	T	A	Substitution	missense_ variant	MODERATE	*MED27*			31
Twin_B	10	24984824	G	A	Substitution	missense_ variant	MODERATE	*ENKUR*	rs200773323, COSV58633510	-	25.9
Twin_B	12	88071791	G	A	Substitution	missense_ variant	MODERATE	*CEP290*	rs761572735	-	25.1
Twin_B	14	62950522	C	T	Substitution	missense_ variant	MODERATE	*KCNH5*	rs587777164, CM1310653, COSV59758318, COSV59758899	uncertain_ significance, pathogenic	28
Twin_B	19	12943750	AGCA GAGG CTTAA GGAG GAGG AAGA AGAC AAGA AACG CAAA GAGG AGGA GGAG	A	Deletion	frameshift_ variant	HIGH	*CALR*	rs1555760738	pathogenic	35

Note: Gene names are provided in italics in accordance with the current HUGO Gene Nomenclature Committee (HGNC) guidelines(173). *An additional variant from the myeloid-gene panel is also included for Twin A (marked with an asterisk). This is a variant detected in the targeted myeloid-gene panel sequencing but with no high-confidence support from whole genome sequencing data; therefore, it was excluded from the downstream analysis.

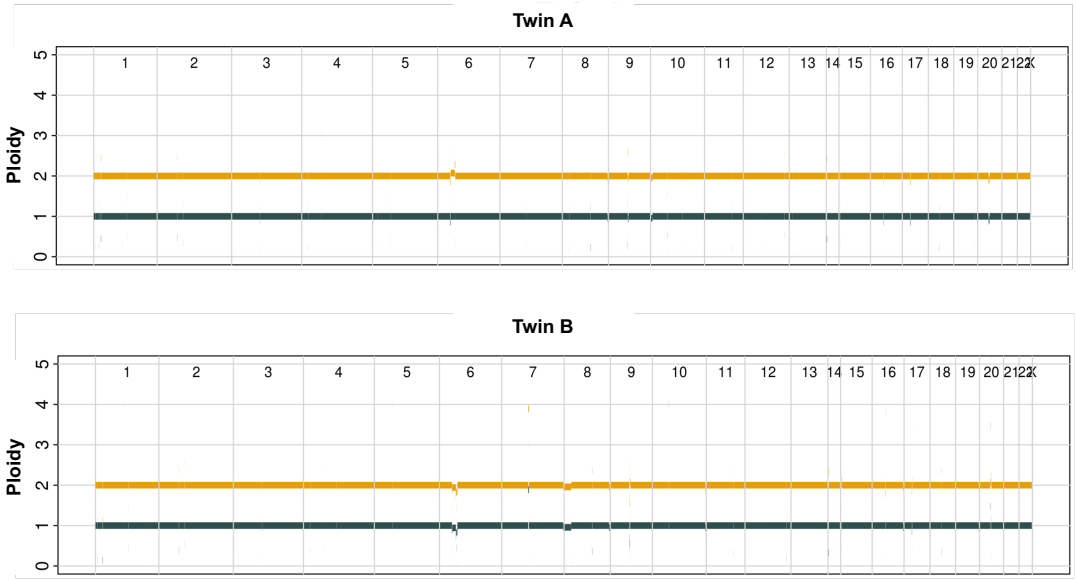


Figure 3-6. Genome-wide copy number profiles of twin A and twin B. *Copy number profiles of twin A and twin B estimated using the Battenberg algorithm. The gold line corresponds to total copy number, the dark blue line corresponds to copy number of the minor allele. Average ploidy was estimated to be 1.98 and 2.03 for twin A and twin B, respectively. Aberrant cell fraction (cellularity) was estimated as 100% and 92% for twin A and twin B, respectively.*

A number of studies have used mutational processes as a molecular clock to carry out chronological time estimates of disease latency in cancer(279, 280). The current study provided a unique opportunity to orthogonally validate such an approach for a blood cancer with a relatively silent mutational landscape such as MPN. We derived an estimate for the time to the most recent common ancestor of MPN cells in both twins (that is, the time of twin-to-twin transmission of the *CALR*-mutant clone) from WGS-identified SNV data arising from a clock-like mutational process. Time to the most recent common ancestor was consistent with an *in utero* origin of MPN (Table 2-4 and Figure 3-7).

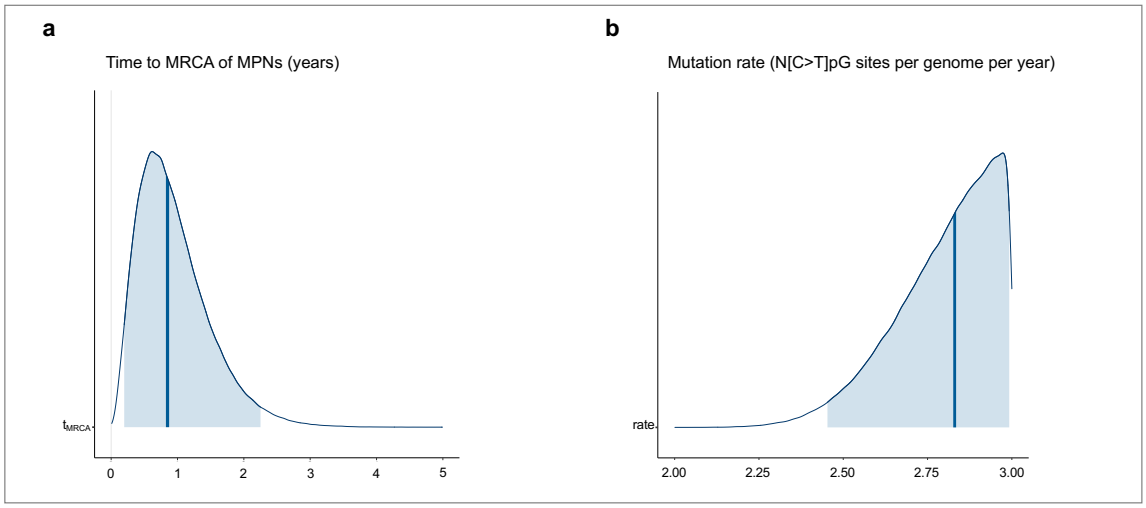


Figure 3-7. Posterior interval estimates from Markov chain Monte Carlo draws for (a) time to MRCA of MPNs in twin A and twin B and (b) mutation rate. *Posterior distributions with median (thick dark blue line) and 95% CI (shaded blue area) for (**a**) MRCA timing and (**b**) mutation rate estimated using the numbers of somatic cytosine-to-thymine single-nucleotide variants (SNVs) at CpG sites (N[C>T]pG) in each twin. **a**. Time to MPN origin is estimated as years post-zygote (horizontal axis). **b**. Mutation rate is estimated as number of N[C>T]pG sites per genome per year (horizontal axis). N[C>T]pG mutation rate is truncated at an upper bound of 3 N[C>T]pG per genome per year; this bound was based On the highest N[C>T]pG rate reported in any Myeloid-MPN case in Gerstung et al.(279). MPN, myeloproliferative neoplasm; MRCA, most recent common ancestor.*

3.1.1.4 Phenotypic and genetic analysis of haematopoietic stem and progenitor cells

Studies in monozygotic twins have provided unique insights into the cancer-propagating cells in childhood leukaemia(289). Some evidence suggests that the cell of origin in MPNs is the phenotypic HSC(136). Twin-to-twin transfusion of the MPN clone provided an opportunity to study the cell-of-origin in MPN. If the *CALR* mutation originally occurred in an HSC, then the *CALR* mutation should be present in HSCs from both twins. I performed index sorting of single $Lin^{-}CD34^{+}$ HSPCs into individual wells for colony-forming assay, recording the surface phenotype for CD90, CD45RA and CD123. Resulting colonies were picked, and DNA was whole-genome amplified. I first confirmed that a number of the shared somatic mutations present in both twins were also present in the colony-derived material for twin A and bulk material for both twins. As expected, for the

four shared mutations analysed, they were all present in both twins, whereas somatic mutations exclusive to each twin were only detected in twin A and twin B, respectively (Table 3-4 and Figure 3-8). The *CALR* mutation was present in all colonies from twin A with shared mutations. One wild-type *CALR* colony from twin A showed presence of one of the shared mutations, supporting that this mutation occurred before the *CALR* mutation.

Table 3-4. Summary of the shared somatic variants and unique somatic variants for each twin, used in whole genome sequencing validation at the single-cell level

Variant studied	Chr	Pos	Ref	Alt	Mutation_ type	Annotation	Nearest_ gene	Existing_ variation	VAF_ Twin_A	VAF_ Twin_B
somatic_ shared	19	12943750	AGCAGA GGCTTA AGGAGG AGGAAG AAGACA AGAAAC GCAAAG AGGAGG AGGAG	A	Deletion	frameshift_ variant	*CALR*	rs1555760738	_	_
somatic_ shared_1	1	237523679	G	A	Substitution	intron_ variant	*RYR2*	rs1165384474	0.32	
somatic_ shared_2	2	236024598	A	G	Substitution	intron_ variant	*AGAP1*	-	0.41	
somatic_ shared_3	3	187366417	G	A	Substitution	upstream_ gene_ variant		-	0.39	
somatic_ shared_4	16	85313584	C	A	Substitution	downstream_ gene_ variant	*AC092275.1*	rs1483404740	0.36	
somatic_ unique_ Twin_A_1	10	18976904	C	T	Substitution	Intergenic_ variant		rs181225799	0.31	NA
somatic_ unique_ Twin_A_2	11	78991748	C	T	Substitution	intron_ variant	*TENM4*	.	0.29	NA
somatic_ unique_ Twin_B_1	7	116155279	G	C	Substitution	intergenic_ variant		rs1461424202	NA	0.19
somatic_ unique_ Twin_B_2	8	20863668	T	A	Substitution	intergenic_ variant		-	NA	0.23

Note: Gene names are provided in italics in accordance with the current HUGO Gene Nomenclature Committee (HGNC) guidelines(173). VAF, variant allele frequency.

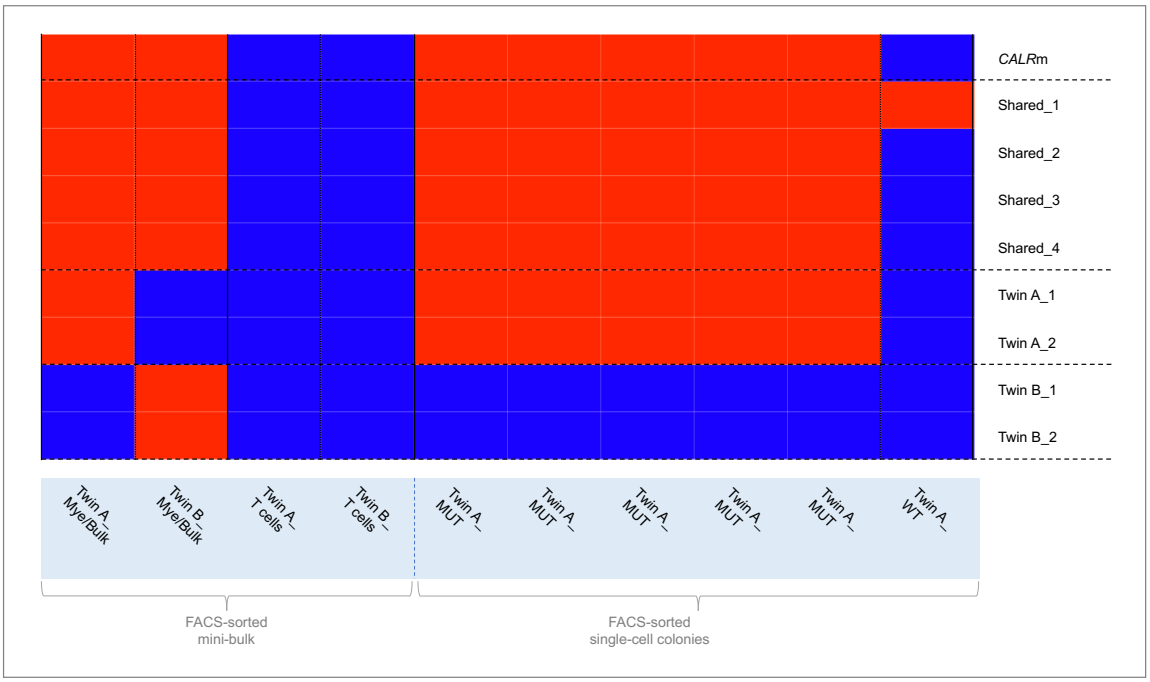


Figure 3-8. Orthogonal validation of the WGS results at the single-cell level*. Somatic variants shared between the twins and somatic variants unique for twin A or twin B were assessed in DNA from FACS-sorted single-HSPC colonies (from twin A) or mini-bulk populations using Sanger sequencing. The samples with presence of the studied variant are shown in red, whereas absence of the variant is shown in blue. Details of the tested variants are shown in Table 3-4. FACS, fluorescence-activated cell sorting; HSPC, haematopoietic stem and progenitor cell; WGS, whole genome sequencing.*

Analysis of 95 colonies demonstrated high clonal dominance of *CALR*-mutant cells, with 29 of 31 colonies in twin A and 64 of 64 in twin B showing presence of the *CALR* mutation. The two *CALR* mutation-negative colonies in twin A likely represent residual nonclonal haematopoiesis, as they were negative for additional shared variants between the twins. The index sorting data (Figure 3-9a) allowed colonies to be traced back to canonical HSPC hierarchies, revealing presence of *CALR* mutation-positive phenotypic HSCs ($Lin^{-}CD34^{+}CD90^{+}CD45RA^{-}$) in both twins (Figure 3-9b), supporting that the HSC is the propagating cell for the MPN clone.

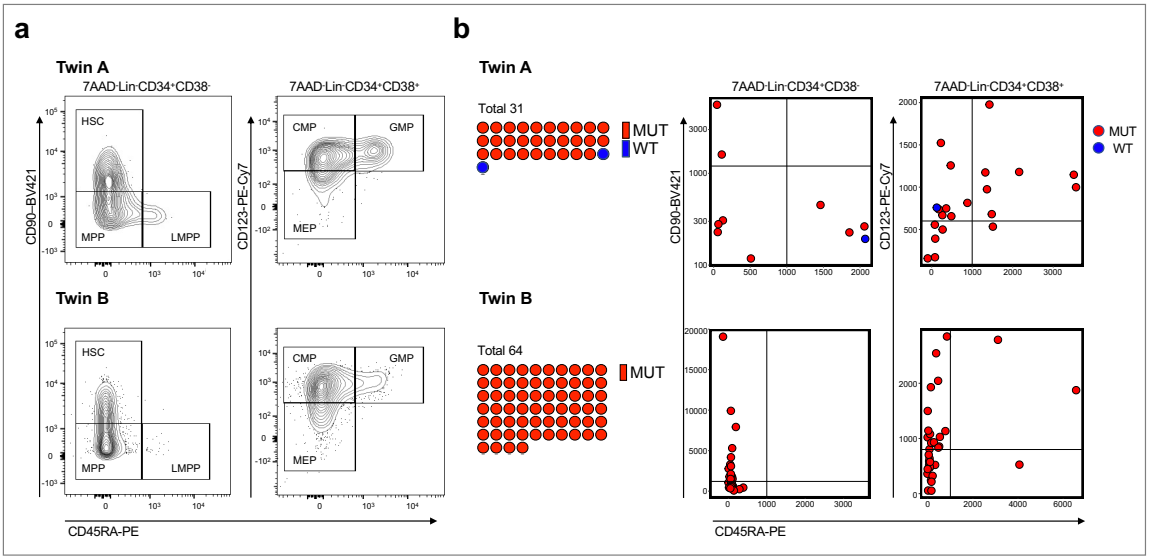


Figure 3-9. Single-cell-derived colony CALR genotyping. ***a****, Flow cytometry profiles of HSPCs for each of the twins.* ***b****, CALR genotyping of single-HSPC-derived colonies, integrated with index sorting data. CMP, common myeloid progenitor (Lin$^-$CD34$^+$CD38$^+$CD123$^+$CD45RA$^-$); GMP, granulocyte macrophage progenitor (Lin$^-$CD34$^+$CD38$^+$CD123$^+$CD45RA$^+$); HSC, haematopoietic stem cell (Lin$^-$CD34$^+$CD38$^-$CD90$^+$CD45RA$^-$); LMPP, lymphoid-primed multipotent progenitor (Lin$^-$CD34$^+$CD38$^-$CD90$^-$CD45RA$^+$); Lin, lineage; MEP, megakaryocyte-erythroid progenitor (Lin$^-$CD34$^+$CD38$^+$CD123$^-$CD45RA$^-$); MPP, multipotent progenitor (Lin$^-$CD34$^+$CD38$^-$CD90$^-$CD45RA$^-$); MUT, mutated; Mye, myeloid; PE-Cy7, PE-cyanine7; WT, wild-type.*

3.2 Extension project: Investigation of the molecular status at birth of patients diagnosed with myeloproliferative neoplasms as children or young adults

To determine whether *JAK2* mutation-positive MPN might also originate *in utero* yet present clinically decades later, I carried out *JAK2V617F* mutation analysis of neonatal blood spots stored from patients subsequently diagnosed with MPN as adults. We were able to identify three *JAK2V617F* mutation-positive MPN patients whose Guthrie cards were available (Table 3-5). *JAK2V617F* mutation detection and quantification was performed by ddPCR. The analysis revealed presence of the *JAK2V617F* mutation in the

neonatal blood spot of one of the three patients studied, with a VAF of 1.38% (Figure 3-10). This patient presented with *JAK2V617F* mutation-positive PV at the age of 34 years. These findings support that an *in utero* origin of the MPN clone, presenting after a prolonged disease latency, may be a widespread phenomenon in MPN, warranting further studies.

Table 3-5. Overview of patient clinical information for the MPN cases included in the JAK2V617F neonatal blood spot analysis

Sample	Sex	MPN Dx	Age at Dx (years)	Driver mutation	Driver mutation VAF at Dx (%)	BMAT	CGN	NGS	*JAK2V617F* detection on DBS (Yes / No)
DBS_1	Female	SM	31.9	*KITD816V*	NA	SM	NA	Nil else	No
DBS_2	Male	PV	28.9	*JAK2V617F*	16.0	NA	NA	NA	No
DBS_3	Female	PV	32.7	*JAK2V617F*	65.0	PV	46,XX[20]	Nil else	No
DBS_4	Male	PV	34.1	*JAK2V617F*	28.0	NA	NA	Nil else	Yes

Note: NSG analysis as described in Chapter 2/2.13 Next-generation sequencing/2.13.1 Extended next-generation sequencing myeloid-gene panel. Gene names are provided in italics in accordance with the current HUGO Gene Nomenclature Committee (HGNC) guidelines(173). BMAT, bone marrow aspirate and trephine; CGN, cytogenetics; DBS, neonatal dried blood spot; Dx, diagnosis; MPN, myeloproliferative neoplasm; NGS, next-generation sequencing; NA, not available; PV, polycythaemia vera; SM, systemic mastocytosis; VAF, variant allele frequency.

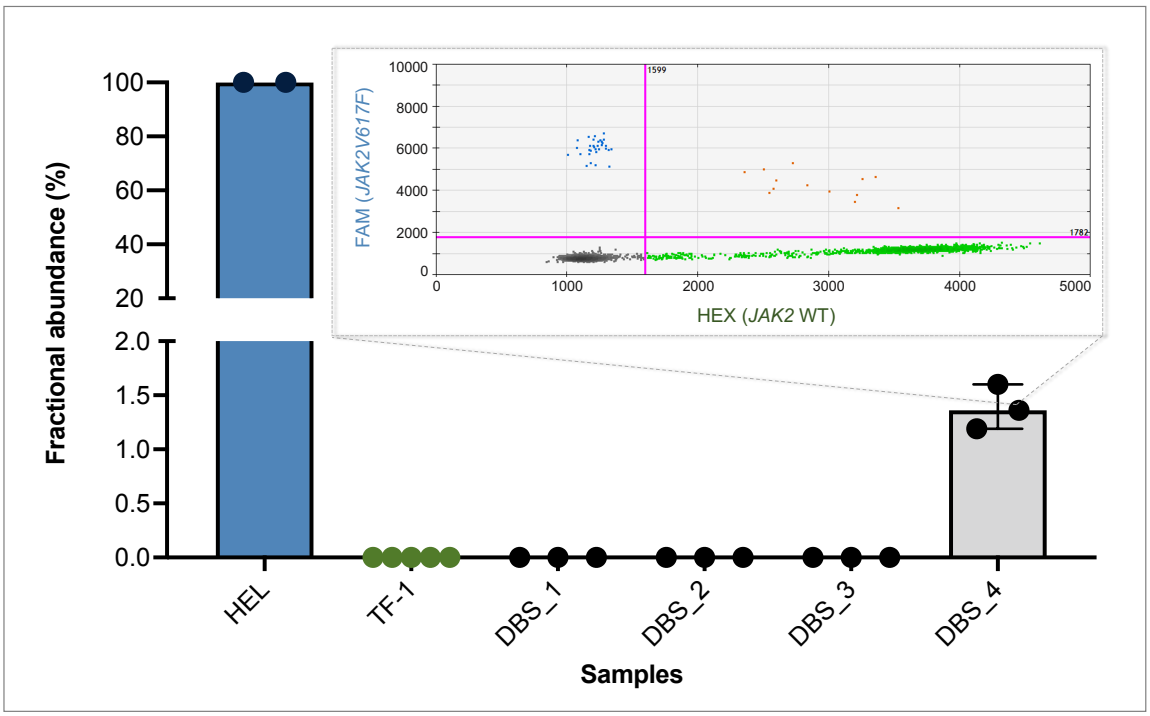


Figure 3-10. Neonatal blood spot analysis. *Results of ddPCR analysis for the detection and quantification of JAK2V617F in DNA extracted from neonatal dried blood spots from patients diagnosed with JAK2-mutant myeloproliferative neoplasms as adults. In one out of three patients studied (DBS_4) JAK2V617F was detected in neonatal blood with a fractional abundance of 1.38% (three technical replicates in one experiment, independently validated by nested PCR). FAM channel-positive events on the y axis correspond to JAK2V617F positivity, and HEX channel-positive events on the x axis correspond to JAK2 WT events. JAK2V617F-positive DNA from HEL cells, and JAK2 wild-type DNA from TF-1 cells and a dried blood spot from a patient with systemic mastocytosis (DBS_1), were used as positive and negative controls, respectively. All results were independently validated by a nested PCR method. Bars and error bars show median and 95% CI, respectively. Relevant clinical information for the patients studied is provided in Table 3-5. DBS, dried blood spot; WT, wild type.*

3.3 Discussion

Emerging evidence from WGS studies of solid tumours supports that the latency between acquisition of an initiating driver mutation and presentation with overt cancer can be prolonged, with important implications for early diagnosis and intervention(279, 290). However, these studies are based on multiple-site biopsies of tumours(290) or analysis of tumours with high genetic heterogeneity and require certain assumptions to be made about

linearity of mutational processes in tumours(279, 290). Transplacental transmission of haematological neoplasms between twins has provided a unique opportunity to study disease latency and genetic evolution of blood cancers but was hitherto considered to be a phenomenon occurring exclusively during childhood, including paediatric MPN(180, 291). In this work I describe transplacental transmission of a *CALR*-mutant clone between a pair of monozygotic twins, which resulted in the development of overt MF after a period of almost four decades. This remarkable case provides definitive evidence of a very early developmental origin of *CALR*-mutant blood cancer, presenting in adults as overt disease decades later. I also carried out neonatal blood spot analysis, which confirmed the *in utero* origin of *JAK2V617F* mutation in a patient presenting with PV at the age of 34 years. These findings definitively confirm long disease latency and *in utero* origin can occur for both *JAK2* and *CALR* mutation-driven MPN.

Single-colony-derived WGS data support that a prolonged disease latency might be a widespread phenomenon in *JAK2V617F*-mutated MPNs(185, 292) and that *JAK2* mutations might even occur *in utero* in some patients(185). Moreover, Hermange et al.(293) have recently proposed a mathematical model to infer disease initiation and evolutionary dynamics of MPNs. The model confirmed early acquisition of the MPN driver mutation, though it was calculated that relative to *JAK2V6176F CALR* mutations are acquired later in life (mean acquisition time for JAK2V617F was 15 years of age, expanding to potentially *in utero* origin, while mean age of acquisition for *CALR* mutations was calculated 25 years of age, excluding acquisition during fetal life). Furthermore, they calculated that *CALR* mutations confer a stronger HSC growth advantage compared with *JAK2V617F*. In addition to studies in MPN, recent evidence from single-colony-derived WGS phylogenomics and longitudinal data in CH revealed that acquisition of the majority of clones occurs before 40 years of age, following distinct gene-specific patterns of

evolutionary dynamics, and resulting in a dramatic decrease of clonal diversity in ages over 75 years(30, 294). Altogether, these findings suggest that the time period from the acquisition of an MPN driver mutation to the establishment and expansion of the clone, as well as the period from established *JAK2V617F*- or *CALR*-driven CH to the development of MPN can in many cases be in the range of decades.

In the present twin study, the unique ability to follow the same clone in the two siblings revealed a strikingly similar and prolonged disease latency for *CALR* mutation-positive MPN. This finding is compatible with the fitness advantage exerted by the clone being established early and not being contingent upon subsequent clonal evolution. However, it also remains possible that the *TET2* or other putative driver mutations identified in the twins collectively resulted in the same latency in the two twins. Moreover, the similar disease latency in both twins suggests that its primary determinants are cell-intrinsic rather than extrinsic factors, although environmental exposures in the twins would be largely shared in childhood. Notably, the similar disease latency might reflect the identical germline genetics, which may influence HSC response to MPN-associated mutations(295). This work also provides additional evidence that the phenotypic HSC is the key propagating population for MPN(136).

Of note, overexpression of fetal-associated genes (such as *HMGA2*) in HSPCs has been demonstrated to promote the development of MPN in model systems, suggesting that fetal haematopoiesis might be permissive for the acquisition of MPN-associated mutations(296). Besides, compared to adult haematopoiesis, fetal bone marrow shows higher number of HSPCs and significantly higher proliferative capacity(297), which contributes to increased replication stress and genetic instability. Most importantly, the nascent immune system during early gestation can provide mutation-specific immunotolerance to MPN-associated driver mutations (such as the *CALRdel52bp*) as a

result of negative selection of the tumour-reactive T cells, following presentation of the mutation-derived neoantigens to the fetal thymus(297-301).

A Danish cohort study previously reported concordance for MPN in 15% of monozygotic twins(176); although the authors concluded that this might reflect genetic predisposition to MPN, it may be that these cases were due to an *in utero* origin and transplacental transmission, as concordant cases were not observed in dizygotic twins. Furthermore, concordance of CH with identical somatic mutations in elderly twins has also been reported, suggesting that clones that arise *in utero* might even persist lifelong until old age without causing disease(181, 182). Whether an *in utero* versus postnatal acquisition of an MPN-driver mutation might influence the subsequent phenotype is unknown. Larger studies of blood cancers in monozygotic twins and of neonatal blood spots in sporadic MPNs are warranted, as this will provide a unique window into the evolution of MPNs and fitness advantage exerted by MPN-associated driver mutations. The long disease latency in combination with a better understanding of the evolutionary dynamics in MPN and availability of robust polygenic risk scores(295) might open up opportunities for early intervention. Although currently the only curative treatment for MPN remains allo-HSCT, there is accumulating evidence that certain treatments such as interferon or targeted inhibitors can induce molecular remissions, and it is also possible that the fetal origin in some cases of MPN might also present unanticipated therapeutic vulnerabilities, paving the way for disease-modifying early intervention strategies in MPN(204).

Chapter 4 *JAK2* mutational analysis of platelet lineage cells in clonal haematopoiesis and myeloproliferative neoplasms

As the next step in the study of genetic lineage tracing in CH and MPNs, I shifted the focus from myeloid lineage against germline DNA to myeloid lineage against the study of platelets as a surrogate of the Mk lineage. The aim of this part of my work was to assess whether platelet-biased CH exists in humans and determine its prevalence in normal haematopoiesis and myeloid neoplasms. The hypothesis was that platelet-biased CH does exist, and therefore, the study of platelets can increase sensitivity in the detection of CH and MPNs compared to the currently applied methodology, which involves whole-blood- or granulocyte-derived DNA. Given the central role of the *JAK2V617F* mutation in both CH and MPNs we aimed to track *JAK2V617F* in the granulocytes and matched platelets from selected individuals. To conduct this proof-of-concept study I developed a method for high-purity blood product isolation, and a ddPCR assay for the detection of the *JAK2V617F* mutation in either gDNA or cDNA template.

4.1 Characteristics of the cohorts studied

4.1.1 Haematologically normal elderly individuals

The "haematologically normal elderly individual" cohort consisted of 108 subjects recruited from 18 November 2019 to 07 October 2022. This cohort included individuals who i) had not explicitly declined the use of their samples in research, ii) aged ≥ 70 years with iii) excess EDTA-anticoagulated peripheral blood sample available within 48 hours

after collection, and iv) normal blood counts. Exclusion criterion was history of a haematological neoplasm, as assessed on the electronic patient record system at the time of sample selection. The acceptable ranges for blood counts were in line with the WHO diagnostic criteria for MPNs(134, 226): 4.0 $\times 10^9$ per litre ≤ white cell count ≤ 11.0 $\times 10^9$ per litre; 100 $\times 10^9$ per litre ≤ platelet count ≤ 450 $\times 10^9$ per litre; 110 g per litre ≤ haemoglobin < 160 g per litre for women, and 110 g per litre ≤ haemoglobin < 165 g per litre for men. In accordance with Human Tissue Authority requirements, no clinical information associated with these samples was collected. Descriptive statistics of demographic variables and blood counts are shown in Figure 4-1a. In summary, the cohort included 40 (37.0%) female and 68 (63.0 %) male patients, with a mean age of 80.5 (95% CI, 70.8 to 93.8) years. Blood counts were distributed as follows: haemoglobin, mean = 129.1 (95% CI, 112.5 to 154.0) g per litre; white cell count, mean = 8.0 (95% CI, 5.2 to 10.8) $\times 10^9$ per litre; platelet count, mean = 241.2 (95% CI, 140.4 to 381.5) $\times 10^9$ per litre.

4.1.2 Unprovoked venous thromboembolism patients

The "platelet-biased clonal haematopoiesis and unprovoked venous thromboembolism" study cohort included 88 subjects with uVTE and 48 age- and sex-matched controls, recruited between 06 July 2021 and 06 June 2022. For this cohort, study subjects were identified by clinical staff at Oxford Haemophilia and Thrombosis Centre (OHTC) when attending either DVT diagnostic clinic or 3-month VTE review clinic. Clinical data were collected by the clinical teams and provided link-anonymised to the research team. The mean age was calculated 58.7 (95% CI, 29.0 to 79.0) years for the uVTE group, and 60.1 (95% CI, 33.7 to 82.9) years for the control population, respectively. Additional descriptive statistics of demographic variables and blood counts are shown in Figure 4-1b, while distribution of demographic, thrombosis, and CVD risk factors between the two groups is provided in Table 4-1. Overall, no significant differences were found between the two

groups except for heart disease, which was more frequent in the uVTE group (10% compared to 0 patients in the control group; $p = .001$), and estimated glomerular filtration rate, which reached marginal statistical significance, though lacking clinical importance.

4.1.3 Triple negative and low-variant-allele-frequency *JAK2V617F* myeloproliferative neoplasm patients

The "TN and low-VAF *JAK2V617F* MPN patient" cohort included 15 well-defined MPN patients. For this cohort, TN and low-VAF *JAK2V617F* mutation-positive MPN patients followed-up in the Myeloid Service, Department of Clinical Haematology, Cancer and Haematology Centre, OUH NHS Foundation Trust, were identified and consented to the INForMeD Study (REC reference 16/LO/1376). Eligibility criteria included the clinical criteria of TN or low-VAF *JAK2V617F* MPN (VAF < 10%), but nil else. Relevant clinical data were collected from the patient electronic records system. The demographics, and clinical and molecular characteristics for this cohort are shown in Figure 4-1c.

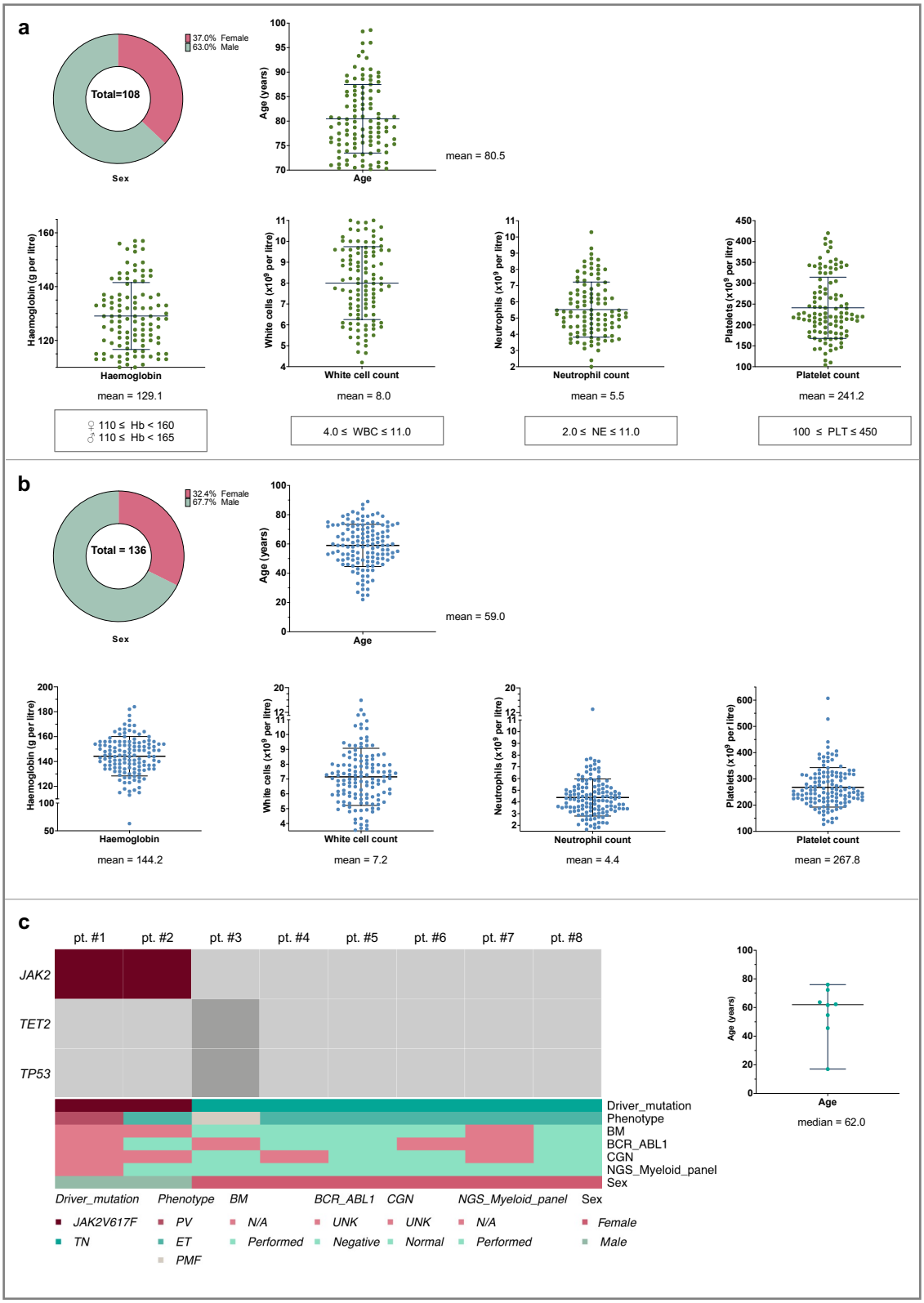


Figure 4-1. Descriptive statistics for each of the cohorts studied in the mutational analysis of platelet lineage cells: ***a***, *Demographics and blood count distribution of the "haematologically normal elderly individual" cohort (n = 108);* ***b***, *Demographics and blood count distribution of the "uVTE" cohort (n = 136); and* ***c***, *Demographics, and clinical and molecular characteristics of the "TN and low-VAF JAK2V617F*

MPN patient" cohort (n = 8). BM, bone marrow testing; CGN, cytogenetics; ET, essential thrombocythaemia; N/A, not available; NGS, next-generation sequencing; PMF, primary myelofibrosis; PV, polycythaemia vera; TN, triple negative; UNK, unknown.

Table 4-1. Distribution of demographic, thrombosis, and cardiovascular risk factors between uVTE and the control groups

		VTE		Ctrl		Sig. (2-tailed)
Age	(years)	58.7	(29.0 – 79.0)	60.1	(33.7 – 82.9)	.564
Sex	(Female = 0, Male = 1)	67.0%		72.0%		.580
Body mass index	(kg per m^2)	30.5	(23.1 – 41.2)	30.6	(19.8 – 42.6)	.960
Haemoglobin	(g per litre)	145.0	(117.8 – 170.6)	143.7	(119.9 – 165.2)	.617
White cell count	($\times 10^9$ per litre)	7.0	(3.9 – 10.4)	7.3	(4.9 – 11.0)	.228
Neutrophil count	($\times 10^9$ per litre)	4.3	(2.0 – 7..0)	4.5	(2.5 – 7.3)	.463
Platelet count	($\times 10^9$ per litre)	264.8	(149.0 – 399.6)	267.0	(168.0 – 377.4)	.865
CRP	(mg per litre)	7.9	(0.5 – 29.4)	10.4	(0.5 – 70.5)	.439
eGFR	(ml/min per 1.73m^2)	80.5	(59.6 – 90.0)	84.5	(55.0 – 90.0)	**.037***
Hypertension	(No = 0, Yes = 1)	25.0%		36.0%		.182
Hypercholesterolaemia	(No = 0, Yes = 1)	25.0%		20.0%		.545
DM2	(No = 0, Yes = 1)	9.0%		8.0%		> .999
Heart disease	(No = 0, Yes = 1)	10.0%		-		**.031**
Cancer	(No = 0, Yes = 1)	12.0%		6.0%		.387

Continuous variables are described with mean (95% CI); dichotomous variables are described with percentages. Comparison of continuous variables was performed with upaired t-test with Welch's correction, while comparison of the dichotomous variables was done with the Fisher's exact test. *Significance held true on nonparametric testing. CRP, C-reactive protein; DM2, diabetes mellitus type 2; eGFR, estimated glomerular filtration rate; VTE, venous thromboembolism.

4.2 Platelet isolation studies

4.2.1 Platelet isolation methodology

During optimisation of the product isolation protocol, special emphasis was put on the isolation of platelets due to their particular importance in the study of platelet-biased CH. As platelets lack nuclear DNA, genetic tracing in platelets was done in the platelet-contained RNA. Given the difference in the volume between platelets and leukocytes

(namely ~10 fl compared to ~400 fl, respectively)(302, 303), leukocyte contamination of the isolated platelet population would have an evident effect in our ability to investigate lineage-bias of detected variants. Therefore, it was crucial to confidently isolate pure platelet populations with minimal leukocyte contamination.

As described in Chapter 2 Materials and Methods/2.2 Cell isolation, a number of different methods were tested for purity of the isolated platelet population by flow cytometry. Figure 4-2 shows the results of comparison of these methods for purity of the isolated product in terms of myeloid, lymphoid, erythroid, and total cell contamination. Apart from FACS-sorted platelets, two-step centrifugation of EDTA-anticoagulated samples was the most effective method, outperforming the alternatives. Although the difference in contamination between the selected method and the alternative methods did not reach statistical significance in all cases, there was a repeating trend, consistent among all cell populations, favouring this method over the other methods studied. Importantly, this method was also the less labour-intensive and time-consuming one.

Regarding the thaw protocol, comparison of normal saline to HEPES-buffered Tyrode's solution revealed no difference in the platelet recovery ratio (data not shown).

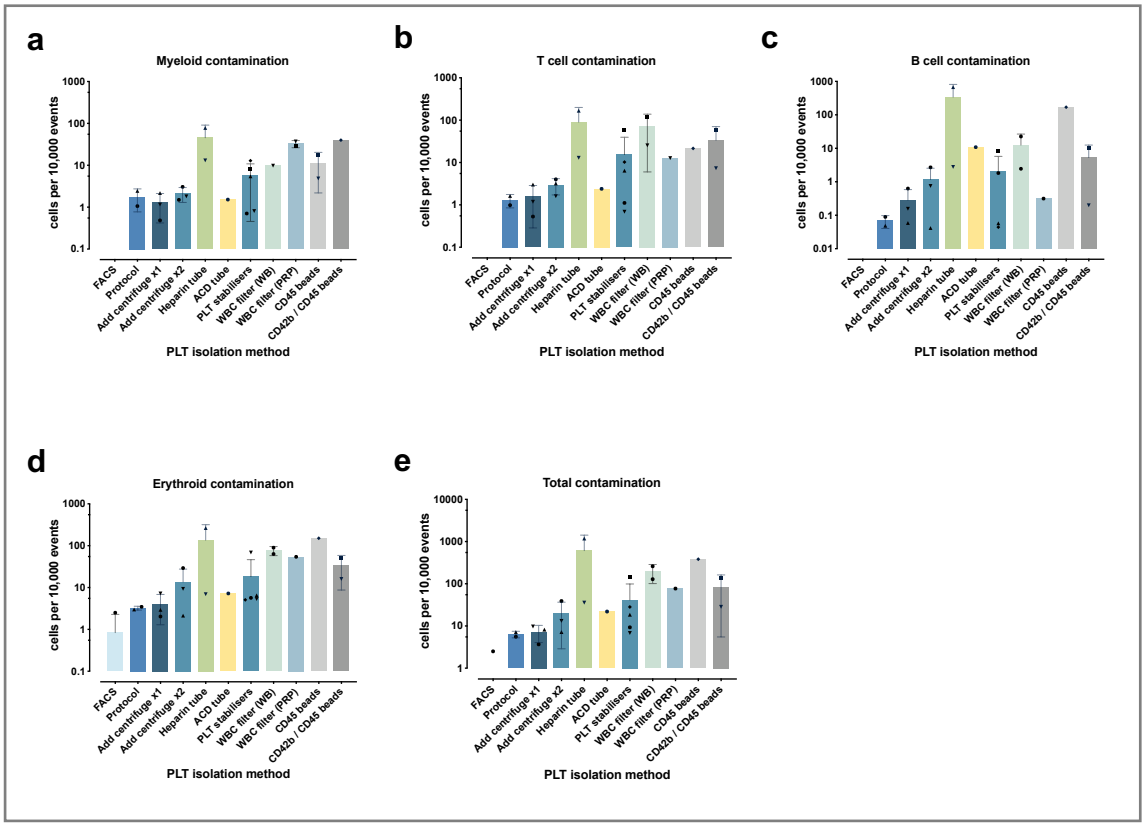


Figure 4-2. Platelet isolation studies. *Flow cytometry analysis of contamination of the isolated product with (a) myeloid cells, (b) B cells, (c) T cells, (d) erythroid cells and (e) combined figures for the various isolation methods studied. On the x-axis from left to right: FACS (n = 4); the two-step centrifugation of EDTA-anticoagulated samples (protocol method) (n = 2); one additional centrifugation step (n = 3); two additional centrifugation steps (n = 3); heparin-anticoagulated samples (n = 2); ACD-anticoagulated samples (n = 1); addition of platelet stabilisers (namely ACD and prostacyclin) in EDTA-anticoagulated samples (n = 5); leukodepletion filtering of whole blood (n = 1); leukodepletion filtering of PRP (n = 2); immunomagnetic negative selection of $CD45^+$ cells in the platelet product (n = 2); and combined immunomagnetic negative selection of $CD45^+$ cells and positive selection of $CD42b^+$ cells (n = 1). Flow cytometry antibody panel details and immunophenotypic definition of cell populations are provided in Chapter 2 Materials and Methods/2.15. Flow cytometry and cell sorting. Scatter dot plots display mean and standard deviation, while scatter plot symbols correlate to the methods that were parallelly assessed in each independent experiment. ACD, acid-citrate-dextrose; EDTA, ethylenediaminetetraacetic acid; FACS, fluorescence-activated cell sorting; PLT, platelet; PRP, platelet-rich plasma; WB, whole blood; WBC, white blood cell.*

4.2.2 Platelet purity studies

Purity of the platelet population isolated by two-step centrifugation of EDTA-anticoagulated samples was further assessed by cytocentrifugation, gene expression, and flow cytometry. A summary of the results is shown in Figure 4-3. Flow cytometry data

analysis showed that contamination of nuclear cells within the isolated platelet population was 0.031%, while myeloid cell contamination was calculated 0.018%. Notably, platelet volume was maintained across the steps of platelet isolation (Figure 4-4).

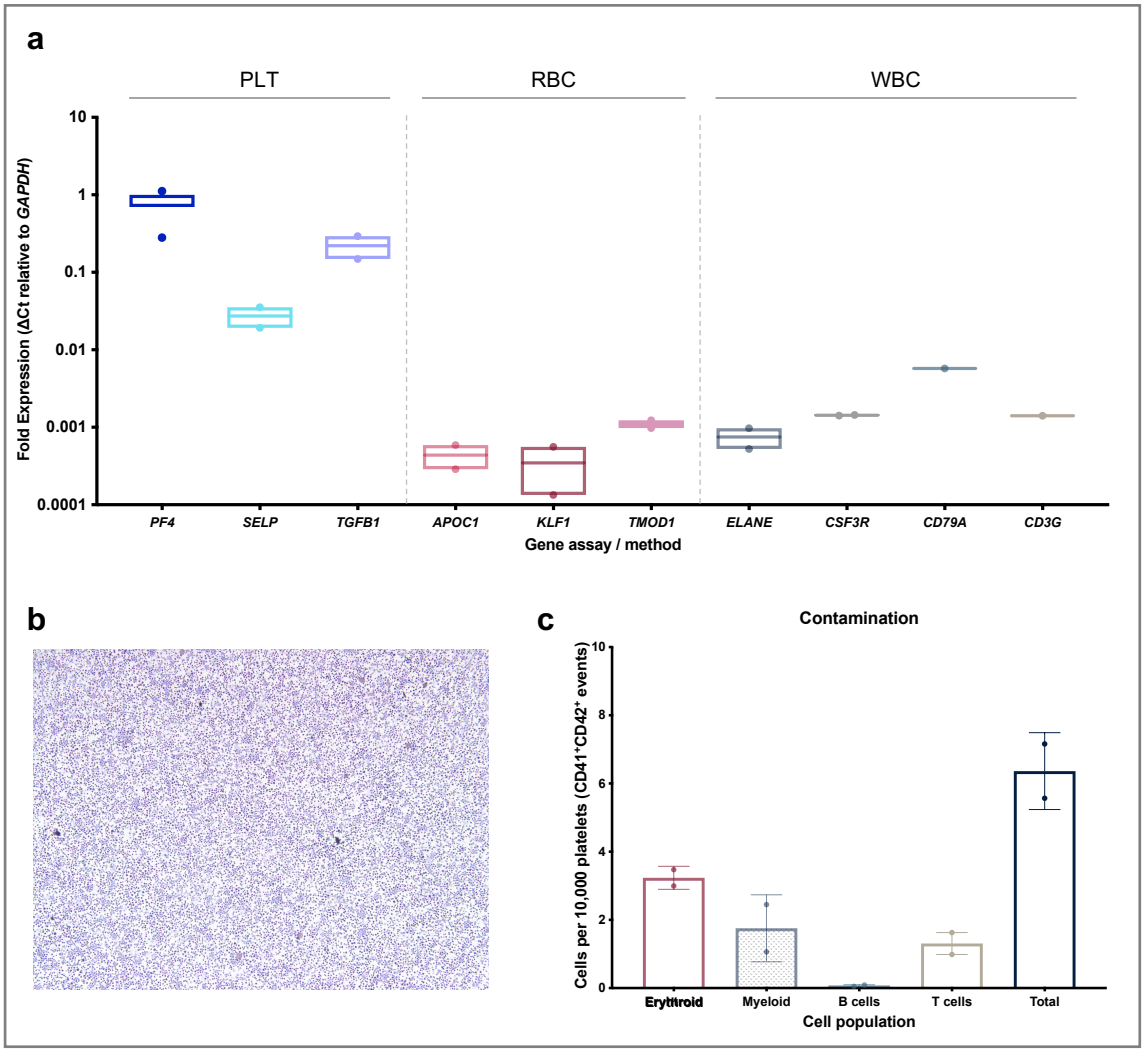


Figure 4-3. Platelet purity studies. ***a****, Gene expression quantification by qPCR, using TaqMan Gene Expression assays for differentially expressed genes within different blood populations (PF4, SELP, and TGFB1 as genes with markedly increased expression in platelets; APOC1, KLF1, and TMOD1 as erythroid cell-associated genes; and ELANE, CSF3R, CD79A, and CD3G as white blood cell markers). Expression of each target-gene was calculated relative to GAPDH as the reference gene control (two independent experiments, including two technical replicates per assay). Floating bars display mean and minimum to maximum range of estimated fold expression per target-gene.* ***b****, Cytocentrifugation confirmed minimal leukocyte contamination of the isolated platelet population.* ***c****, Flow cytometry analysis for contamination of the isolated platelet population with erythroid, myeloid, and lymphoid cells. Flow cytometry antibody panel details and immunophenotypic definition of cell populations are provided in Chapter 2 Materials and*

Methods/2.15. Flow cytometry and cell sorting. Scatter dot plot displays mean and standard deviation. PLT, platelet; RBC, red blood cell; WBC, white blood cell.

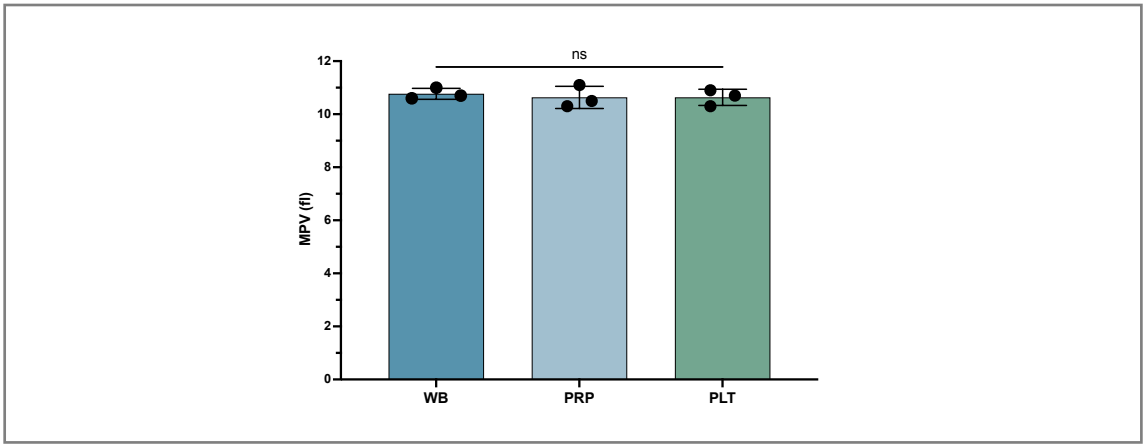


Figure 4-4. Platelet volume across platelet isolation. *Platelet volume measurement across platelet isolation showed no significant changes in the size of platelets between the different stages of the platelet isolation process (three samples processed and analysed in parallel in a single experiment) (ANOVA; p = .441). Scatter dot plot displays mean and standard deviation. ANOVA, analysis of variance; MPV; mean platelet volume; ns; not significant; PLT, platelet; PRP, platelet-rich plasma (post first centrifugation); WB, whole blood.*

Given that it would be very difficult to perform processing and downstream assays using fresh samples, freezing platelet pellets would be very advantageous, providing that this did not majorly affect their quality for downstream assays. Therefore, I assessed post-thawing platelet recovery, and platelet-extracted RNA and corresponding cDNA yield.

4.2.3 Platelet recovery survival studies

Platelet preservation studies were performed to assess whether the number of platelets isolated from a typical 4-ml EDTA-anticoagulated sample was sufficient for downstream procedures. The results are shown in Figure 4-5, showing good platelet recovery, being in line with the ratio described in the pertinent literature(229, 304).

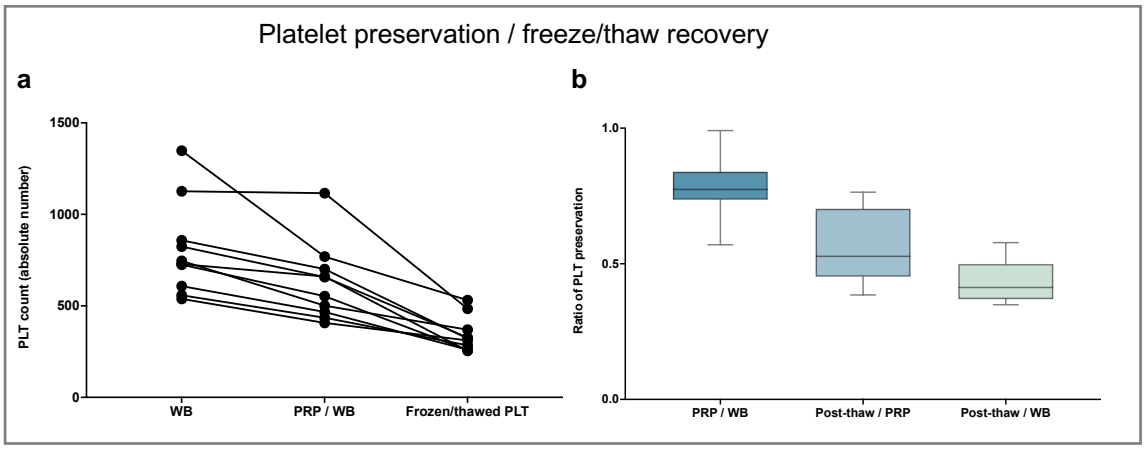


***Figure 4-5. Platelet preservation studies. a**, Platelet absolute count in different stages of platelet isolation was assessed using an automated cell counter; **b**, The box and whisker plot shows the ratio of platelet preservation between the different stages of the platelet isolation process (10 samples processed and analysed in parallel in a single experiment); PLT, platelet; PRP, platelet-rich plasma (post first centrifugation); WB, whole blood.*

4.2.4 Platelet-derived nucleid acid quality control

I then aimed to assess whether the RNA and corresponding cDNA yield resulting from a typical 4-ml EDTA-anticoagulated sample was sufficient for downstream procedures. In 10 samples with a mean platelet count of 230.3 (95% CI, 150.0 to 321.0) $\times 10^9$ per litre, mean RNA yield was calculated 111.2 ng (95% CI, 86.0 to 141.0) resulting in a cDNA yield of a mean concentration of 2.2 (95% CI, 1.2 to 3.7) ng per µl (30 µl elution). Comparison of RNA quality and cDNA yield derived from fresh platelets and frozen/thawed platelets showed no significant changes (Figure 4-6).

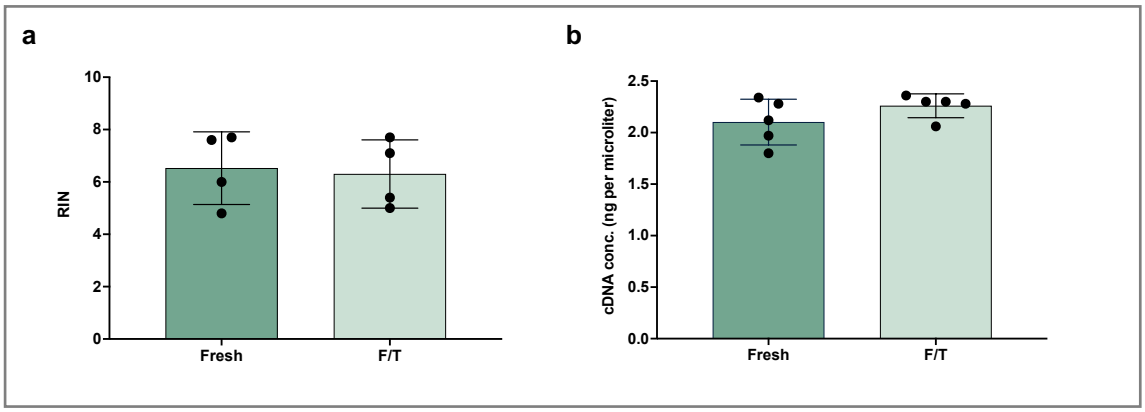


Figure 4-6. Comparison of RNA quality and cDNA yield between fresh platelets and frozen/thawed platelets. ***a****, Comparison of RNA integrity number for RNA extracted from fresh platelets and frozen/thawed platelets, assessed by automated electrophoresis (Agilent 2100 Bioanalyzer system).* ***b****, Comparison of cDNA concentration in 30 µl elution derived from fresh platelets and frozen/thawed platelets, as assessed by fluorometry (Qubit 2.0 fluorometer) (five samples [each consisting of two 4-ml EDTA-anticoagulated samples; one for each of the methods compared] run in one experiment). Automated electrophoresis of RNA provided no result for one sample from each method. Conc., concentration; F/T, frozen/thawed; RIN, RNA integrity number.*

Altogether, the above findings showed that the freezing/thawing procedure had no significant effect on the platelet yield, RNA isolation, and resulting cDNA yield, which was overall in line with the expected yield and considered sufficient for downstream analysis.

4.3 Droplet digital PCR lineage tracing for *JAK2V617F*

Given the central role of the *JAK2V617F* mutation in both CH and MPNs, I opted to initially focus my research for presence of platelet-biased CH on the *JAK2V617F*. To conduct this, I developed a ddPCR assay for high-sensitivity detection of the mutation in either gDNA or cDNA template.

4.3.1 Assay performance studies

Measurement error studies showed excellent performance of the assay for both gDNA and cDNA input. Reproducibility and accuracy assessment for cDNA is shown in Figure 4-7a,

revealing an analytical sensitivity of at least 0.1%, while sensitivity of the gDNA assay was previously calculated 0.009%(233). Given the relatively low yield of platelet-derived DNA, I assessed repeatability of the assays for low DNA input (Figure 4-7b,c). Both assays provided precise results for all studied conditions. To assess specificity of the method we retrospectively analysed TF-1 control samples (negative controls for *JAK2V617F*) of all successful runs and assessed them for positive events in the FAM channel. False FAM-positive events were rarely detected, and if so, there were no more than two FAM-positive events per sample, with more than 80% of them having a single event. In more detail, regarding gDNA false positive rate was calculated 0.0005% (specificity 99.9995%) for double-distilled water and 0.0357% (specificity 99.9995%) for TF-1 DNA, while for cDNA false positive rate was calculated 0.0009% (specificity 99.9991%) for double-distilled water and 0.0091% (specificity 99.9909%) for TF-1 DNA. When dividing the false FAM-positive events with the total number of the HEX positive events for the corresponding TF-1 control samples, mean fractional abundance was calculated 0.0357% for gDNA and 0.0091% for cDNA.

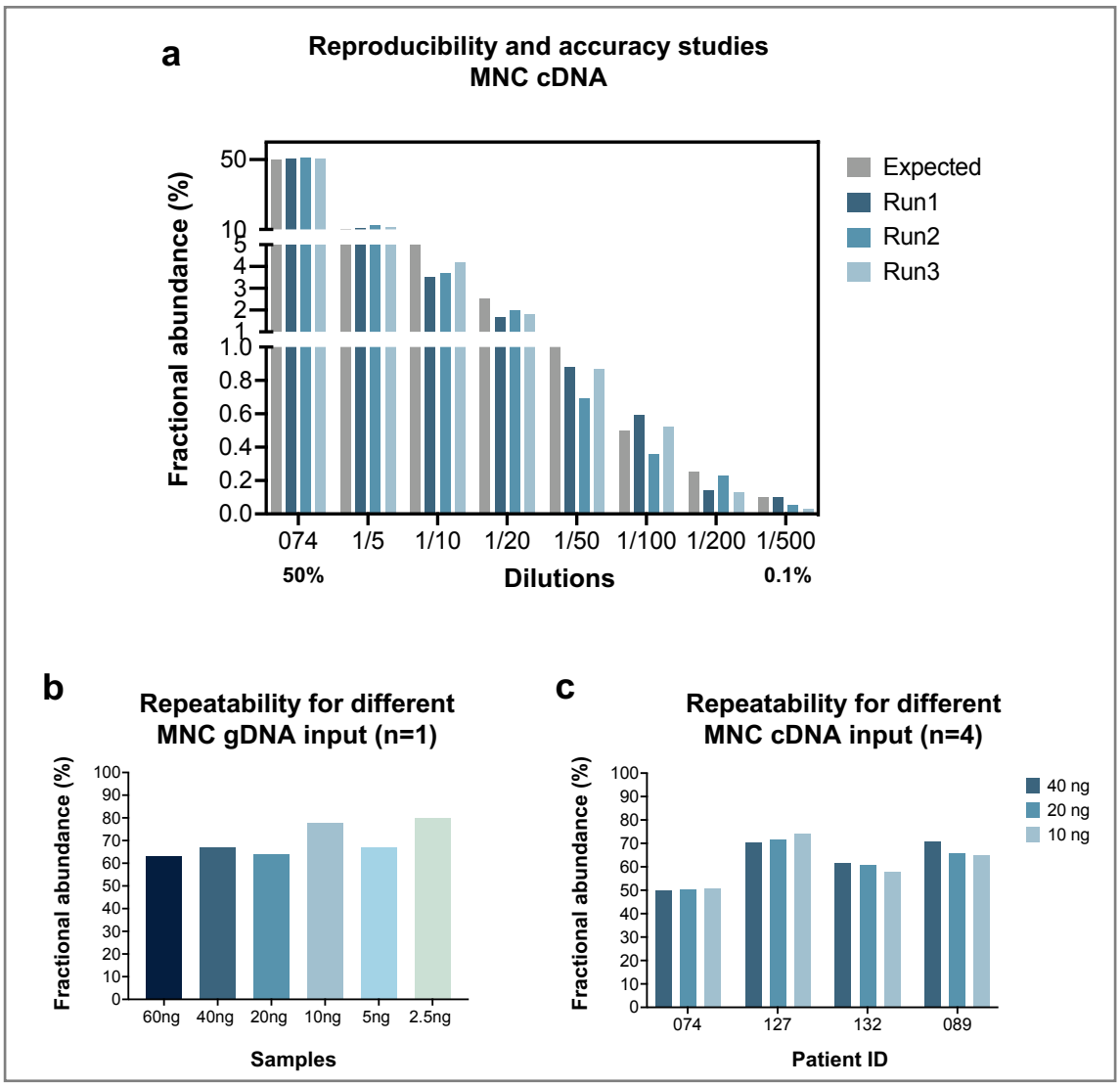


Figure 4-7. JAK2V617F ddPCR assay accuracy studies. *Measurement error studies showed that both (gDNA and cDNA) assays were both accurate and precise.* ***a****, Results of accuracy and reproducibility studies for the cDNA assay for serial dilutions down to a final expected fractional abundance of 0.1%, that was calculated as the assay's positivity threshold. Reproducibility was assessed in three independent ddPCR runs. Repeatability studies for different* ***b****, gDNA and* ***c****, cDNA input (down to 2.5 ng and 10 ng, respectively) were conducted to assess performance of the assays for low DNA input. ddPCR, droplet digital PCR; MNC, mononuclear cell.*

4.3.2 Definition of *JAK2V617F* positivity and biased/restricted clonal haematopoiesis

Based on the aforementioned performance studies, the criteria for *JAK2V617F* positivity for both gDNA and the cDNA assay were stringently set as: (i) presence of ≥ 3 FAM-positive events, (ii) accounting for a fractional abundance of ≥ 0.1%, and a 95% Poisson CI lower limit > 0, (iii) in a run including ≥ 10,000 events, a quality requirement for all samples analysed.

Furthermore, linage-restricted CH was defined as presence of CH in one cell lineage but not the other, while lineage-biased CH was defined as presence of CH in both cell lineages, when fractional abundance in one of the lineages is more than three times the fractional abundance in the other and at least 0.3%.

4.3.3 Results of droplet digital PCR lineage tracing for *JAK2V617F*

Haematologically normal elderly individuals. Given the strong correlation of CH with age, as a first step of this proof-of-concept study I aimed to assess presence of platelet-biased CH in the elderly population, defined as individuals of more that 70 years of age. In total, 108 matched granulocyte gDNA and platelet cDNA samples from elderly individuals were successfully analysed by ddPCR for presence of the *JAK2V617F* mutation. Within these 108 individuals, 14 (13.0%) were found to carry the *JAK2V617F* (Figure 4-8a). These cases consisted of one case where the mutation was present in both granulocyte-derived DNA and the platelet-derived DNA, while in the remaining of the cases *JAK2V617F* was lineage-restricted, being present either in the granulocytes (n = 5) or the platelets (n = 8) alone. Importantly, the median *JAK2V617F* allele burden was significantly higher in the platelet-derived DNA compared to the granulocyte DNA (1.00% versus 0.25%, respectively; p < .001). In the case of CH present in both cell lineages, fractional abundance

in platelet DNA was 0.90% versus 0.27% in the granulocyte DNA, i.e., more than three-fold higher, being suggestive of platelet-biased CH. Furthermore, in the cases of platelet-restricted CH median fractional abundance was 1.90% versus 0.25% for the granulocyte-restricted CH (p = .003) (Figure 4-8b). In four cases (3.7%) platelet analysis revealed presence of *JAK2V617F*-related CHIP (VAF > 2%), while none of the cases of granulocyte-restricted CH reached CHIP VAF levels as none had a VAF > 2%. Based on this cohort, there was no significant relationship between age and the probability of CH, considering either all cases (Figure 4-8c) of CH all platelet-biased CH alone (Figure 4-8d). Notably, age, sex, and blood counts for the individuals with CH had no significant differences compared with the total cohort (Figure 4-8e).

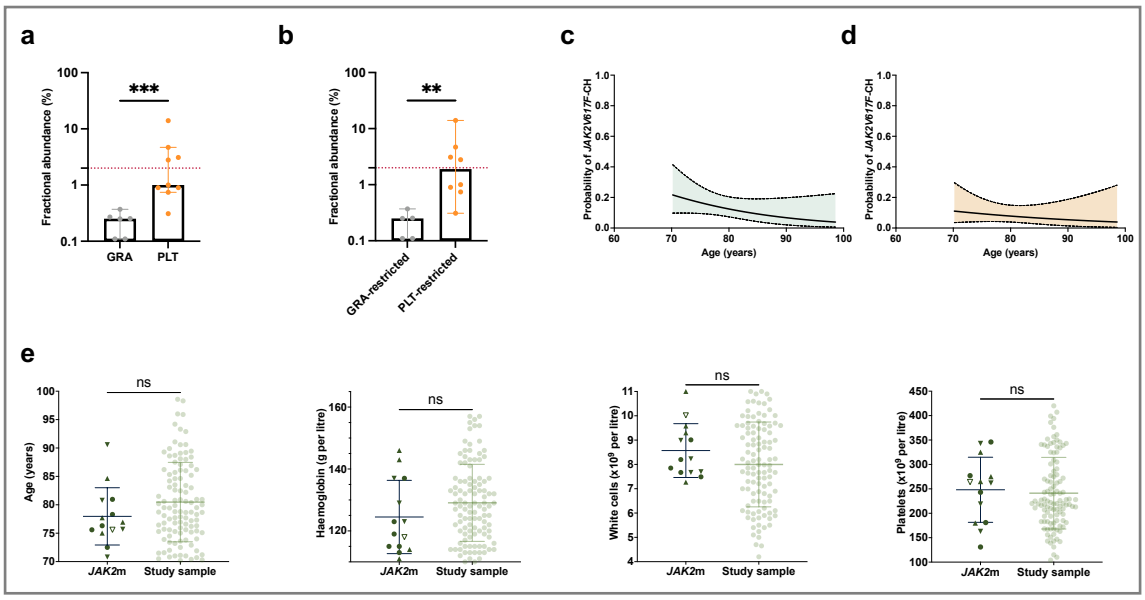


Figure 4-8. Results of the JAK2V617F lineage tracing in the "haematologically normal elderly individual" cohort (n = 108). ***a****, JAK2V617F was detected in 14 of the 108 individuals analysed, with a significant higher fractional abundance in the platelet-derived cDNA compared to the measured fractional abundance in the granulocyte-derived gDNA (Mann-Whitney test; p < .001).* ***b****, Similarly, in the cases of platelet-restricted CH (n = 8), JAK2V617F fractional abundance was significantly higher compared to the fractional abundance in the cases of granulocyte-restricted CH (n = 5) (Mann-Whitney test; p = .003). Bars and error bars show median and 95% CI, respectively.* ***c****,* ***d****, Simple logistic regression analysis of age (as the predictor) and the probability of presence of JAK2V617F-related CH (as the predicted outcome) for (c) all cases of CH (OR = 0.935, p = .154) and (d) platelet-restricted CH (OR = 0.962, p = .485). Estimated probability (curve fitted to data) is represented with the continuous line, while 95% CI for the fitted distribution is shown with the shaded area within error bands. No significant odds ratios were calculated for the present database.* ***e****, Comparison of age and blood counts between the individuals with CH and the total of screened population revealed no significant differences. Circle dots depict granulocyte-restricted CH, while triangles correspond to platelet-related CH: up-pointing triangles stand for cases of CHIP (fractional abundance ≥ 2%), while down-pointing triangles stand for cases of CH with JAK2V617F fractional abundance < 2%; blank (no fill) triangle corresponds to the cases of non-lineage-restricted CH. Scatter plots with light green dots show distribution of the variable among all study subjects. All scatter plots display mean and standard deviation. Comparisons were made using the unpaired t test with Welch's correction. **, p ≤ .01; ***, p ≤ .001; CH, clonal haematopoiesis; CHIP, clonal haematopoiesis of indeterminate potential; GRA, granulocyte; JAK2m, JAK2 mutation-positive; ns, not significant; PLT, platelet.*

Unprovoked venous thromboembolism patients. As the next step in the study of platelet-biased *JAK2*-related CH, and given the high correlation of the *JAK2V617F* mutation with VTE, I aimed to assess a group of patients with uVTE adjusted for age, sex, and common risk factors for CVD, for presence of *JAK2V617F*-associated CH. A total of 136 matched granulocyte gDNA and platelet cDNA samples, consisted of 88 patients with uVTE and 48 age- and sex-matched controls, were analysed by ddPCR for presence of the *JAK2V617F* mutation. In total, 14 (10.3%) participants, namely seven (8.0%) patients with uVTE and seven (14.6%) in the control group, were found to carry the mutation (Figure 4-9a). These cases consisted of five (3.7%) study participants where the mutation was present in both granulocyte-derived DNA and the platelet-derived DNA, while in the remaining of the cases *JAK2V617F* was lineage-restricted, being present in either the granulocytes (n = 4) or the platelets (n = 5) alone. Fractional abundance reached CHIP levels in two cases of platelet-restricted CH, accounting for 1.5% of the total cohort (fractional abundance being 14.8% and 5.2%, respectively), while one only case of granulocyte-restricted CH, accounting for 0.74% of the total cohort, had a fractional abundance of > 2% (being 2.8%). The comparison of *JAK2V617F* fractional abundance between the platelet cDNA and granulocyte gDNA for all CH, the non-lineage-restricted CH, and the lineage-restricted CH cases revealed no statistically significant difference (Figure 4-9b-d). However, in two of the cases of non-lineage-restricted CH *JAK2V617F* fractional abundance in platelet cDNA was more than three-fold higher compared to that in granulocyte gDNA, being suggestive of platelet-biased CH (Figure 4-9c). Notably, one of the remaining cases of non-lineage-restricted CH was a 77-year-old man presenting with pulmonary embolism alongside increased blood counts (haemoglobin 160 g per litre; platelet count 607 $\times 10^9$ per litre) and presence of the *JAK2V617F* mutation at a fractional abundance of approximately 35%, thus possibly representing bona fide MPN rather than

CH. Demographics and blood counts for the individuals with CH had no significant differences compared with the total of the study sample (Figure 4-9e).

Based on the available data at this phase of sample collection and analysis, no significant relationship between the *JAK2* mutational status (as the predictor parameter) and the probability for presence of uVTE (as the predicted outcome) was found ($OR = 0.392$, $p = .110$). However, it should be stated that this was an exploratory data analysis, and at this phase, the study was underpowered for that analysis.

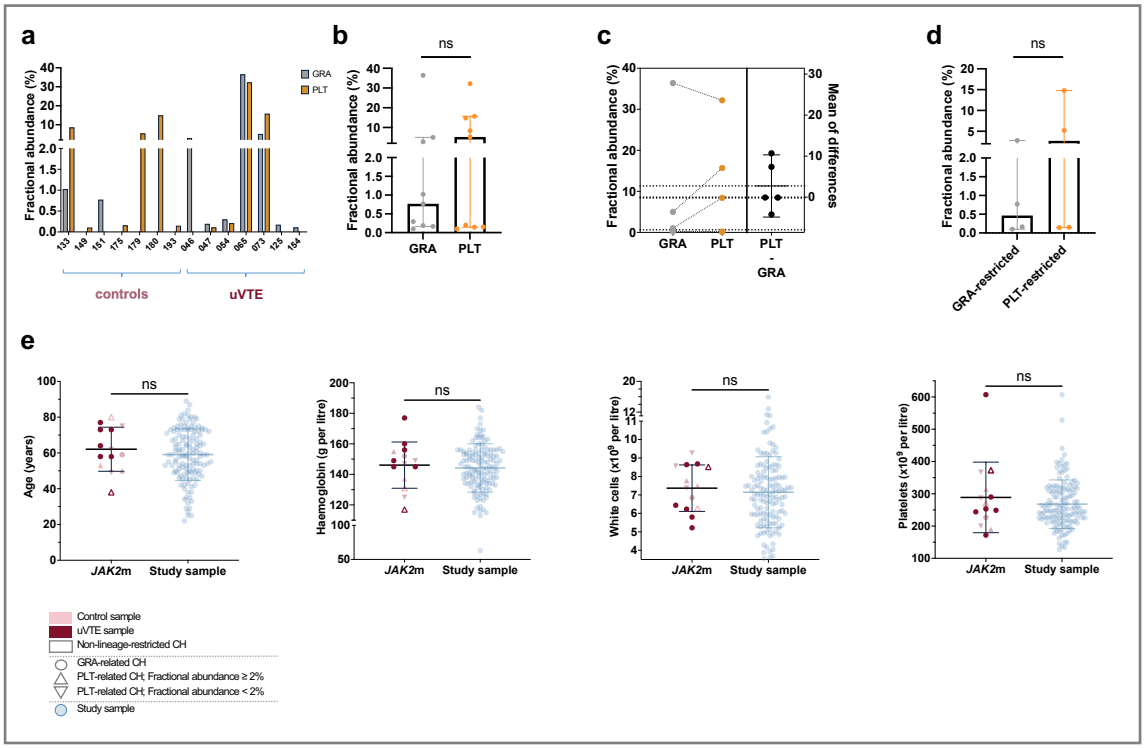


Figure 4-9. Results of the JAK2V617F lineage tracing in the "uVTE" cohort (n = 135). a, *JAK2V617F was detected in 14 of the 135 individuals analysed, consisting of seven cases of uVTE and seven cases in the control group.* ***b****, Comparison of the fractional abundance of the JAK2V617F mutation in granulocyte-derived gDNA and the platelet-derived cDNA for the total number of CH revealed no statistically significant difference. Bar and error bars show median and 95% CI, respectively (Mann-Whitney test; p = .685).* ***c****, Estimation plot, showing the mean of JAK2V617F fractional abundance differences between the platelet cDNA and the granulocyte gDNA in the study subjects who had non-lineage-restricted (JAK2V617F) CH (paired t-test; p = .372).* ***d****, Scatter plot showing the JAK2V617F fractional abundance in the study subjects with lineage-restricted JAK2-related CH (n = 4 in each group). Bar and error bars show median and 95% CI, respectively (Mann-Whitney test; p = .686).* ***e****, Age, haemoglobin, white cell and platelet counts were not significantly different between the individuals with CH and the total cohort. Circle dots, blank (no fill) triangles, and coloured triangles correspond to granulocyte-restricted, non-lineage-restricted CH, and platelet-restricted CH, respectively. Up-pointing triangles stand for cases of CHIP (fractional abundance ≥ 2%), while down-pointing triangles stand for cases CH with JAK2V617F fractional abundance < 2%. Dark red symbols stand for individuals with CH who were diagnosed with uVTE, while pink symbols stand for individuals with CH in the control group. Scatter plots with light blue dots show distribution of the variable among all study subjects. All scatter plots display mean and standard deviation. Comparisons were performed using unpaired t test with Welch's correction. ns, not significant; CH, clonal haematopoiesis; GRA, granulocyte; JAK2m, JAK2 mutation-positive; PLT, platelet; uVTE, unprovoked venous thromboembolism.*

Broad age general population group. Given the focus of the first cohort on elderly individuals (aged more than 70 years) and in view of studying CH in the general population with a broader age distribution, an ad hoc analysis of the merged "haematologically normal elderly individual" cohort and the control participants of the "uVTE" cohort with normal blood counts was conducted. In this group, I aimed to assess the effect of age on the probability of *JAK2V617F*-related CH. In total, the merged cohort included 151 individuals, aged 22.0 to 98.6 years (mean = 74.9; 95% CI, 50.0 to 91.8) (Figure 4-10). Clonal haematopoiesis was detected in 21 (13.9%) individuals at a mean age of 72.4 (95% CI, 49.1 to 90.0) years (Figure 4-11a,b), consisting of 2 cases of non-lineage-restricted CH, 6 cases of granulocyte-restricted CH, and 13 cases of platelet-restricted CH, including 6 cases of CHIP (i.e., fractional abundance > 2%). Comparison of average *JAK2V617F* fractional abundance in platelet-cDNA with the fractional abundance in the granulocyte-gDNA revealed significantly higher fractional abundances in the platelet-derived material (Figure 4-11c). No significant association between age and probability of *JAK2V617F*-related CH was detected in this sample ($OR = 0.995$, $p = .758$) (Figure 4-11d). This finding held true for platelet-biased/restricted CH, platelet-restricted CH, and platelet-restricted CHIP sub-analyses, too. Based on this sample, age between individuals with granulocyte-restricted CH (mean age 73.8 years) and those with platelet-restricted CH (mean age 70.9 years) was not significantly different either ($p = .558$). Lastly, comparison of age, sex, and blood counts between the cases of granulocyte-related CH, platelet-restricted CH, total cases of CH, and the total study sample showed no significant differences (Figure 4-12a,b).

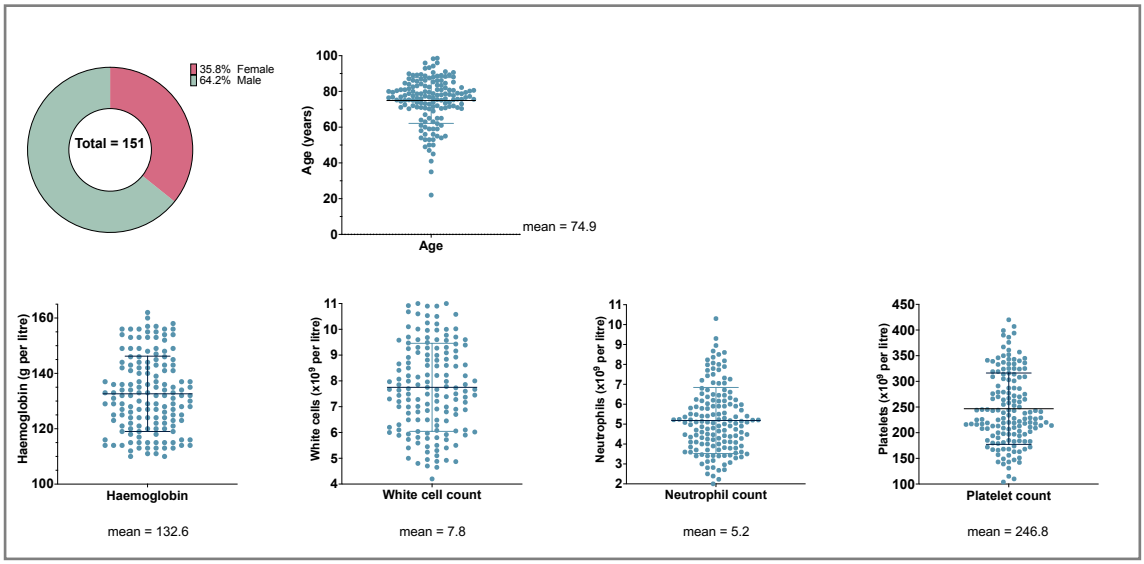


Figure 4-10. Descriptive statistics of age and blood counts of the broad age general population group (n = 151). *All scatter plots display mean and standard deviation.*

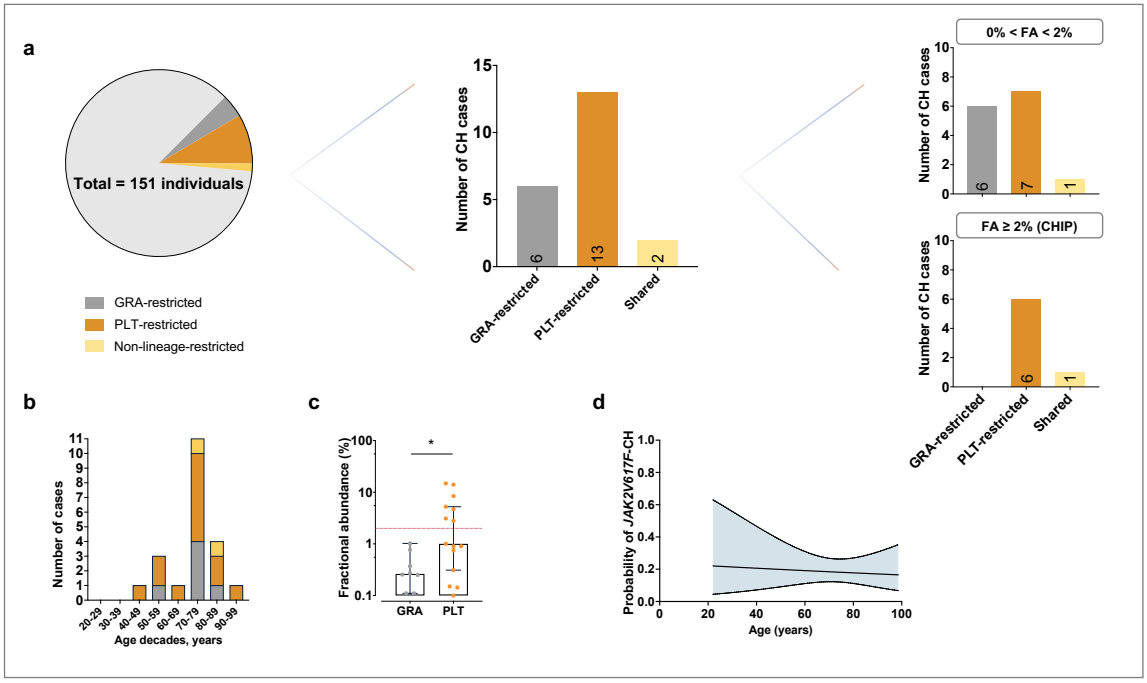


Figure 4-11. Results of the JAK2V617F lineage tracing in the broad age general population group (n = 151). a, *Pie chart showing overall prevalence of JAK2V617F-related CH detected in the study sample, and bar plots showing its lineage distribution. In total, CH was detected in 21 (13.9%) individuals, consisting of 2 cases of non-lineage-restricted CH, 6 (4.0%) cases of GRA-restricted CH, and 13 (8.6%) cases of PLT-restricted CH, including 6 cases of CHIP (fractional abundance ≥ 2%).* ***b****, Bar plot showing the lineage distribution of JAK2V617F-related CH cases against age of the screened individuals.* ***c****, Comparison of the fractional abundance of the JAK2V617F mutation in granulocyte-derived gDNA and the platelet-derived cDNA for the total number of CH revealed significantly higher fractional abundance for platelet-cDNA (Mann-Whitney test; p = .038). Bar and error bars show median and 95% CI, respectively, while the red grid line corresponds to 2% fractional abundance.* ***d****, Simple logistic regression showed no significant effect of age on the probability of JAK2V617F-related CH in this sample (OR = 0.995, p = .760). Estimated probability (curve fitted to data) is represented with the continuous line, while 95% CI for the fitted distribution is shown with the shaded pale blue area within error bands. *, p ≤ .05; CH, clonal haematopoiesis; CHIP, clonal haematopoiesis of indeterminate potential; FA, fractional abundance; GRA, granulocyte; PLT, platelet.*

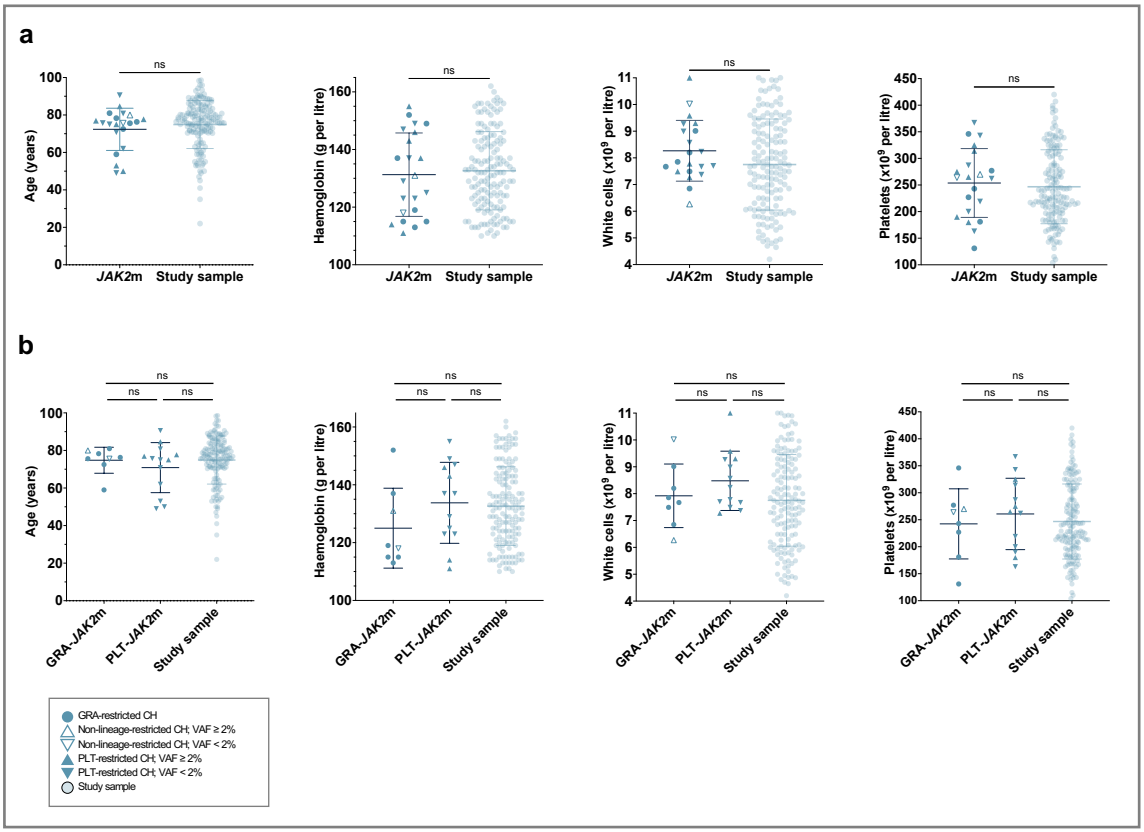


Figure 4-12. Association of CH with age and blood counts. a, *Age, haemoglobin, white cell and platelet counts were not significantly different between the individuals with CH and the total study sample, while* ***b****, separately focusing on granulocyte-related CH and platelet-restricted CH did not reveal significant differences either. Circle dots correspond to granulocyte-restricted CH, while blank (no fill) triangles show non-lineage-restricted CH. Teal triangles correspond to platelet-restricted CH. Up-pointing triangles stand for cases of CHIP (fractional abundance ≥ 2%), while down-pointing triangles stand for cases of JAK2V617F fractional abundance < 2%. Scatter plots with light teal dots show distribution of the variable among all study subjects. All scatter plots display mean and standard deviation. CH, clonal haematopoiesis; CHIP, clonal haematopoiesis of indeterminate potential; GRA, granulocyte; JAK2m, JAK2 mutation-positive; ns, not significant; PLT, platelet.*

4.3.4 Triple negative and low-variant-allele-frequency *JAK2V617F* myeloproliferative neoplasm patients

Lastly, we aimed to assess the presence of platelet-biased *JAK2V617F*-related CH in the setting of established MPN. We screened a group of patients with well-characterised TN or low-VAF *JAK2V617F* mutation-positive disease for presence of the mutation in their

granulocytes and matched platelets. In total, we screened eight patients, consisting of two cases of low-VAF disease and six patients with TN MPN.

Notably, in one of the cases of low-VAF *JAK2V617F* (Figure 4-1c; pt. #2), lineage tracing of the mutation in granulocytes and platelets, showed evidence of platelet-biased CH, with the VAF being measured 11.3% in the platelet-derived cDNA against 2.6% in granulocyte gDNA (Figure 4-13a). Notably, targeted NGS had previously detected the mutation at VAFs of 2% (Molecular Diagnostics Centre, NIHR Biomedical Research Centre, OUH NHS Foundation Trust) and 3% (Wessex Genomics Laboratory Service, University Hospitals Southampton), while amplification refractory mutation system PCR (ARMS-PCR) had detected the mutation at VAF < 5%.

Moreover, in one of the six cases of TN disease (Figure 4-1c; pt. #7), there was evidence of platelet-restricted CH, as *JAK2V617F* was detected in the platelets of the patient (VAF 1.3%) but not in her granulocytes (Figure 4-13b). Next-generation sequencing had previously revealed no variants of clinical significance on two occasions, while ARMS-PCR had detected the mutation (VAF < 5%) in one of two measurements. The patient has had persistent mild to moderate thrombocytosis, with normal morphological appearances on a bone marrow aspirate, but she was not able to tolerate a bone marrow trephine biopsy. Cytogenetic analysis showed normal karyotype (46,XX[20]).

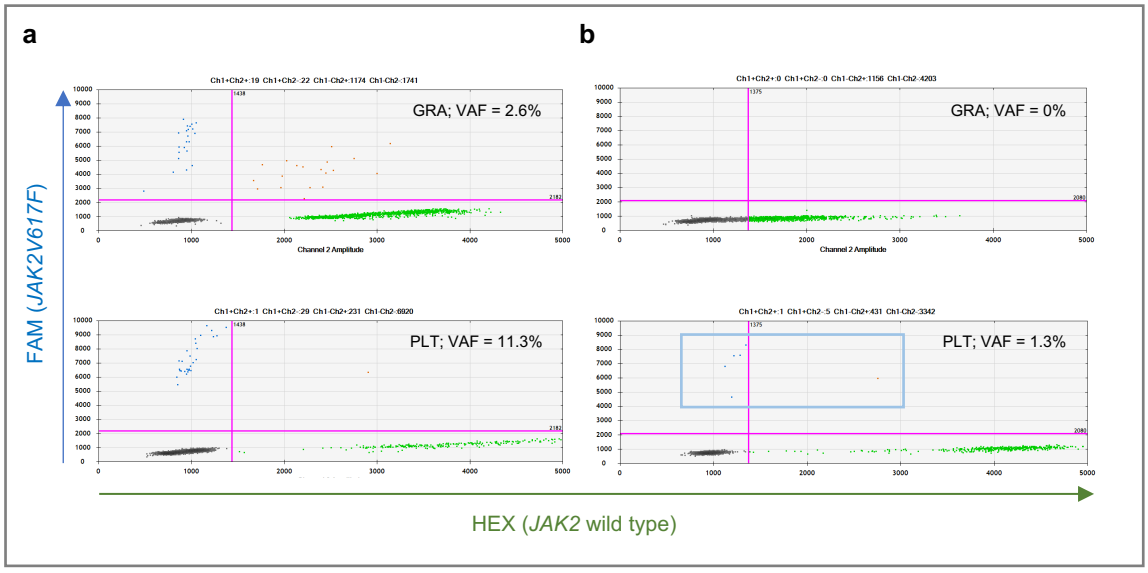


Figure 4-13. Results of JAK2 mutational analysis in patients with TN or low-VAF JAK2V617F mutation-positive MPN. a*, Low-VAF JAK2V617F ET with evidence of platelet-biased CH. JAK2V6176F VAF was calculated 2.6% in the granulocytes and 11.3% in the platelet cDNA.* ***b****, In one of the six cases of TN MPN, lineage tracing of JAK2V6167F revealed presence of platelet-restricted CH, as evidenced by presence of the mutation in the platelets cDNA at VAF 1.3%, but not in matched granulocyte gDNA. FAM channel corresponds to JAK2V6167F signal, while HEX channel corresponds to JAK2 wild type. GRA, granulocyte; PLT, platelet; TN, triple negative; VAF, variant allele frequency.*

4.4 Discussion

The present work shows for the first time the presence of platelet-restricted *JAK2V617F*-associated CH in normal individuals. Importantly, using granulocyte gDNA alone the prevalence of *JAK2V6167F*-related CH was 5.3% (8 cases), while study of platelet-derived cDNA enabled the detection of an additional 8.6% (13 cases), including 4.0% (6 cases) of CHIP (versus no cases of CHIP in the granulocyte-derived material).

Importantly, the detection of *JAK2V7617F* in granulocytes alone was in line with previous studies of *JAK2V617F* prevalence in general population. Cordua et al.(233) have recently reported a prevalence of 3.1% in the Danish general population using the same assay we applied for detection of granulocyte-related *JAK2V617F*-CH. Given the age distribution of

that registry compared to the age distribution in our study (mean = 74.9 years), the reported prevalence is highly comparable to the detected granulocyte-related CH in our study. The increased levels of overall prevalence of *JAK2*-related CH in our study is due to the presence of additional cases of platelet-restricted CH.

This is a novel finding and has a profound, direct implication for CH detection. Given that a key complication of CH is vascular disease as well as the central role of both platelets and *JAK2V617F* mutation in thrombosis, platelet-biased clones may have significant clinical sequelae. Based on the "uVTE" cohort data, there was no association between *JAK2V617F*-related CH and the probability of uVTE, but this was an early exploratory data analysis, knowing that at this stage of the project, the study was underpowered for such analysis.

Furthermore, as the *JAK2V617F* mutation accounts for the vast majority of MPNs, platelet-restricted CH may also have significant implications for the subsequent development of MPN. Notably, in a recent prospective population-based cohort study of the CH evolutionary landscape, analysis of whole-blood-derived DNA from 3,359 individuals aged 60 to 91 years showed that *JAK2*-related CH had the highest risk for incident myeloid neoplasms(107). Interestingly, *JAK2V617F* is among the most common CH-associated mutations, and in most cases it is detected as an isolated mutation with no presence of additional myeloid neoplasm-associated mutations, suggesting that *JAK2V617F* alone is sufficient to induce CH(136). However, the fact that the prevalence of *JAK2*-mutant MPNs is significantly lower than the estimated prevalence of *JAK2V617F*-positive CH is suggestive that *JAK2V617F* alone may be sufficient to induce CH but insufficient to drive MPN(136). This discrepancy might be explained (at least to a certain extent) by a retrospective cohort study among 49,488 individuals from the Copenhagen General Population Study(305). In this project Nielsen et al. assessed the above cohort for

JAK2V617F levels against the history of development of MPN disease. The authors reported that higher *JAK2V617F* allele burden was required for the development of overt MPN, estimating that a cut-off point of 2% was indicative of disease versus no disease. However, since no additional information about the mutational status of the participants was available, it is not possible to draw a definitive conclusion regarding the sufficiency of *JAK2V617F* alone to induce MPN.

Previous studies on CH have revealed an age-related pattern for the kinetics of haematopoietic clones driven by different gene mutations(51, 52) and *JAK2*-related CH has been shown to occur throughout life, including early adulthood representation. The results of the present work are in line with the above finding, as the age of *JAK2*-related CH ranged from 49.0 to 90.6 years, while no significant relationship was found between age and the probability of (*JAK2*-related) CH. The combination of early acquisition of the *JAK2V617F* mutation with the recent advances in the pathogenesis of MPNs, that support prolonged latency between the emergence of the premalignant clone and the development of overt disease, highlights the unique opportunity for early detection, monitoring and preventive, disease-modifying intervention.

Lineage tracing of the *JAK2V617F* in MPN supported the hypothesis that platelet studies in MPN can increase sensitivity in the diagnosis of challenging cases in which patients have no molecular markers on conventional testing or driver mutations are detected in only very low VAFs, and bone marrow testing is either not feasible or results in equivocal findings. In one of the cases of low-VAF *JAK2V617F*, lineage tracing of the mutation in granulocytes and platelets showed evidence of platelet-biased CH, with the VAF being measured 11.3% in the platelets against 2.6% in the granulocytes of the patient, and remarkably, in one out of six TN MPN patients, there was presence of a platelet-restricted clone at a VAF of 1.3%.

Importantly, the presence of platelet-restricted CH supports the hypothesis of presence of platelet-biased HSCs in human haematopoiesis, which would have considerable implications for genomic-based approaches for early detection and monitoring of disease, which, in haematology, are primarily based on isolation of nucleic acid from leukocytes and would fail to detect emerging platelet-biased clones. Importantly, the presence of clones with VAF ≥ 2% is suggestive of established CH rather than transient clones, as clones with VAF ≥ 2% have been associated with further clonal expansion and CHIP evolution(238), while specifically *JAK2V617F*-CH with VAF ≥ 2% has been associated with increased risk of progression from CHIP to MPN(233, 305, 306). However, bone marrow HSC analysis from individuals with presence of platelet-restricted clones is required to confirm the presence of Mk-biased HSCs, while longitudinal follow-up of such platelet-biased clones alongside phenotypic analyses are required to confirm their evolutionary dynamics and clinical relevance.

The above findings should be interpreted in light of some limitations. These primarily include the relatively small number of subjects, selection bias, and the above-mentioned absence of longitudinal analysis of the detected clones. In more detail, the limited sample size did not allow a precise estimation of the prevalence of platelet-biased *JAK2V617F*-CH, especially in younger individuals, who were under-represented in the studied cohorts, and the small series of MPN patients. Generalisability is also affected by the clinical characteristics of the "haematologically normal elderly individual" cohort (that included patients with cardiovascular disease and cancer), which might have also affected the study of association between age and blood count parameters of the study participants and the probability for presence of *JAK2V617F*-related CH. Limited sample size was also a limitation for the exploratory data analysis for association between *JAK2* mutational status and the probability for presence of uVTE in the "uVTE" cohort. Furthermore, the one-off,

non-longitudinal study design did not allow for assessment of longevity and longitudinal dynamics of the detected clones. However, it should be emphasised that this was a proof-of-concept study, aiming to investigate the presence of platelet-biased/restricted CH, which was evidently proven. The exact prevalence of the first-time described platelet-biased/restricted *JAK2*-related CH as well as its specific characteristics and clinical relevance should be further explored in larger cohorts of the general population or specific subgroups of normal individuals and patients.

Chapter 5 Targeted next-generation sequencing analysis of platelet lineage cells in clonal haematopoiesis and myeloproliferative neoplasms

5.1 Introduction

Following validation of presence of *JAK2V617F* platelet-biased CH, I aimed to assess whether platelet-biased CH also exists for other genes involved in CH and MPNs. In view of this, I developed and optimised a custom, targeted, hybridisation capture NGS panel for parallel sequencing of both gDNA and cDNA, covering the major genes or hotspot mutations involved in CH and MPNs, as described in Chapter 2 Materials and Methods/2.13 Targeted next-generation sequencing/2.13.2 Custom next-generation sequencing panel and summarised in Figure 5-1.

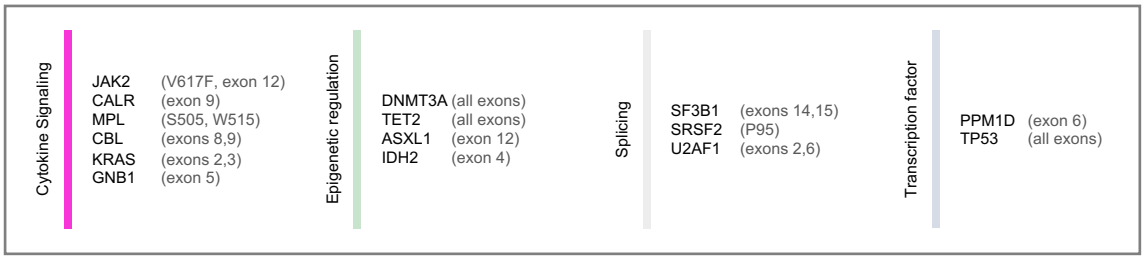


Figure 5-1. Custom targeted NGS panel design. *Summary of the genes and target regions included in the panel (detailed in Chapter 2 Materials and Methods/2.13 Targeted next-generation sequencing/2.13.2 Custom next-generation sequencing panel/2.13.2.1 Panel characteristics).*

5.2 Assay performance studies

As the first step, I assessed the performance of the panel using gDNA and cDNA templates from MPN patients with known mutations, previously identified and quantified in the Molecular Diagnostics Centre, NIHR Biomedical Research Centre, OUH NHS Foundation Trust. Eight patient samples involving a total of 15 mutations (in *JAK2*, *CALR*, *MPL*, *TET2*, *SF3B1*, *SRSF2*, *U2AF1*, and *TP53*) were used in the quality control studies, including variants of low allelic burden (VAF < 5%), which were used to infer specificity of the assay for either gDNA or cDNA (Table 5-1).

Performance studies of the initially designed panel revealed poor performance of certain probes due to either increased off-target effect or low coverage of the target region. Three of the probes accounted for approximately 80% of the off-target effect with one of them (namely a probe targeting the region of interest in *PPM1D* gene, though expanding to the 3′ untranslated region) being responsible for 63.6% of it. In the updated panel the probes of increased off-target effect were removed, while the remaining of low-performance probes were ranked by normalised mean coverage, and depending on that value, they were handled with 2×, 3×, or 4× tiling, adding additional staggered overlapping probes to boost coverage closer to the mean. The updated, final panel comprised a total of 250 probes with a combined footprint size 25,083 bp.

Quality control of the updated panel showed acceptable off-target effect for all probes, improved alignment rate, and satisfactorily results in terms of coverage, uniformity, and complexity. Sensitivity of the panel was 100%, while coverage was at least 500× for 95% of target bases for granulocyte gDNA, 1000× for 95% of target bases for granulocyte cDNA, but barely 100× for 95% of target bases for platelet cDNA. Given the uniform deep sequencing of granulocyte-derived cDNA and the gene-specific variation in platelet-

derived cDNA, the observed variability should be attributed to stochasticity in gene expression. Overall, the detection threshold for granulocyte gDNA was set at 1% VAF, while for platelet cDNA no fixed universal threshold could be applied. Notably, correlation and agreement for the NGS-measured VAFs of granulocyte gDNA and paired granulocyte cDNA was high (correlation: r = .962, p < .001; and Lin's concordance correlation coefficient: ρ_c = .952, alpha = .05) (Figure 5-2).

Table 5-1. Custom NGS panel performance studies

UPN	Gene	Mutation	VAF% (OUH)	gDNA_VAF%	gDNA_coverage	mRNA_VAF%	mRNA_coverage
pt_1	***JAK2***	c.1849G>T p.V617F		33.4	1397	Not sequenced	Not sequenced
pt_2	***JAK2***	c.1849G>T p.V617F	< 5 ^	Not sequenced	Not sequenced	0.57	1311
pt_3	***JAK2***	c.1849G>T p.V617F	< 5 ^	Not sequenced	Not sequenced	1.5	774
pt_4	***JAK2***	c.1849G>T p.V617F	< 5 ^	0.39 (5 reads)	1294	0.64 (5 reads)	785
pt_5	***JAK2***	c.1849G>T p.V617F		43.1	1250	46.1	1336
	SF3B1	c.2098A>G p.G700E		42.5	1157	37.3	1715
	TET2	c.3882C>G p.Y1294*		23.4	1353	18.7	637
	TET2	c.4317dup p.R1440Tfs*38		16.3	1014	18.0	572
	TP53	c.817T p.R273C		7.0	962	5.6	654
pt_6	***CALR***	c.1092_1143del52 p.L367fs*46		18.8	1116	18.6	1630
pt_7	***CALR***	c.1092_1143del52 p.L367fs*46		20.1	1868	11.1	2695
	U2AF1	c.470A>G p.Q157R		15.4	712	13.5	624
pt_8	***MPL***	c.1544G>T pW515L		35.8	1192	40.1	1168
	SRSF2	c.284C>G p.P95R		37.9	947	38.4	963
	TET2	c.4354C>T p.R1452*		35.7	1189	38.2	953

Note: Gene names are provided in italics in accordance with the current HUGO Gene Nomenclature Committee (HGNC) guidelines(173). ^Samples used in the calculation of VAF detection threshold. NGS, next-generation sequencing; VAF, variant allele frequency; OUH, Oxford University Hospitals (Molecular Diagnostics Centre, NIHR Biomedical Research Centre, OUH NHS Foundation Trust).

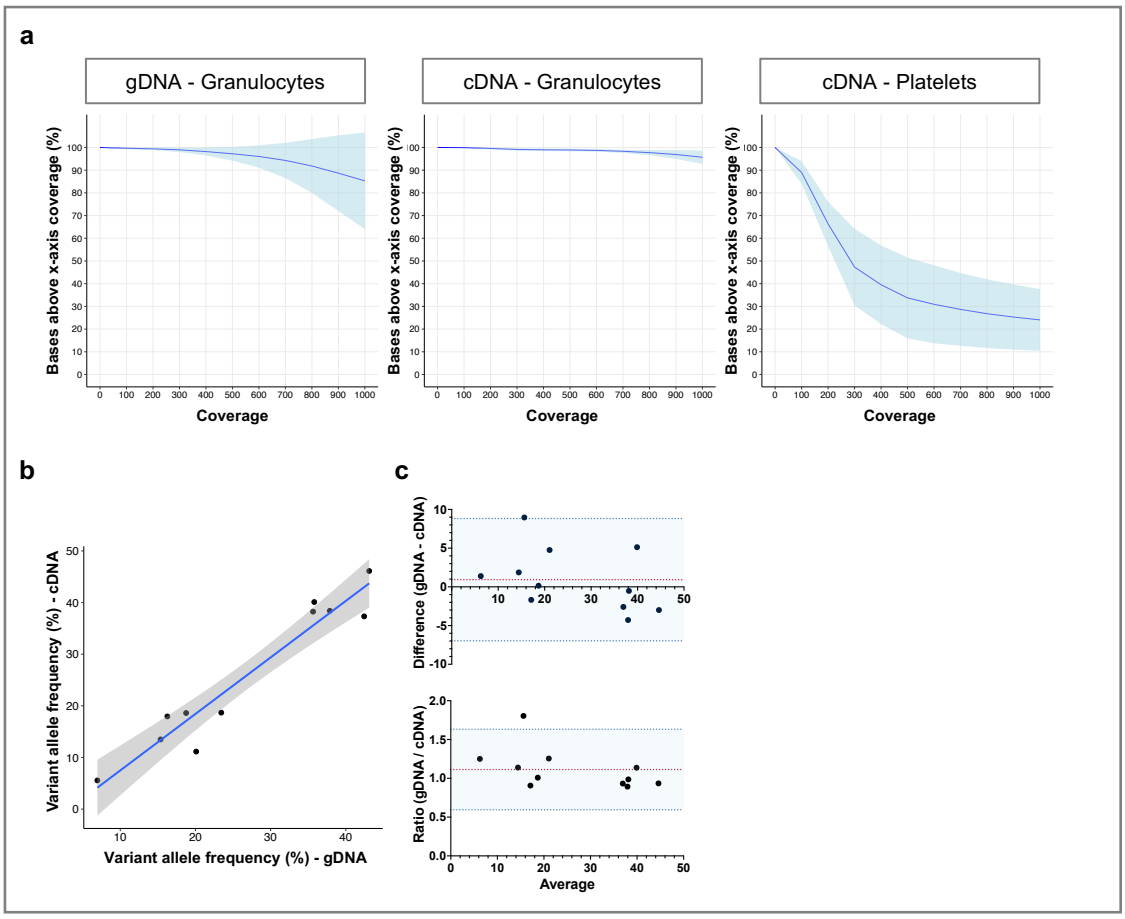


Figure 5-2. Performance studies for the custom targeted NGS panel. a, *Coverage per target base for (left panel) granulocyte gDNA, (middle panel) granulocyte cDNA, and (right panel) matched-sample platelet cDNA.* ***b****, Correlation between the NGS-measured VAFs for granulocyte gDNA and granulocyte cDNA showed strong relationship (Pearson correlation; r = .962, p < .001).* ***c****, Agreement between the NGS-measured VAFs for granulocyte gDNA and granulocyte cDNA (n = 11). (upper panel) Bland-Altman plot showing the difference between the paired readings versus their average. Bias was calculated 0.93% towards gDNA (SD = 4.03; 95% limits of agreement -6.97% to 8.82%). (lower panel) Bland-Altman plot showing the ratio of paired readings versus their average. Systematic bias was calculated 1.1× for gDNA (SD = 0.3; 95% limits of agreement 0.6 to 1.6). The red grid lines correspond to the calculated bias, while the blue grid lines correspond to 95% limits of agreement. NGS, next-generation sequencing; VAF, variant allele frequency.*

5.3 Detection and quantification of additional clonal haematopoiesis associated mutations through targeted next-generation sequencing

Forty-two paired samples of granulocyte-derived gDNA and platelet-derived cDNA were successfully sequenced. The samples were randomly selected from the "haematologically normal elderly individual" cohort and library preparation and sequencing was performed as described in Chapter 2 Materials and Methods/2.13 Targeted next-generation sequencing/2.13.2 Custom next-generation sequencing panel, while subsequent variant detection, annotation, and filtering was performed as described in Chapter 2 Materials and Methods/2.17 Bioinformatic analysis/2.17.2 Custom next-generation sequencing panel analysis.

Based on the results of the performance studies, and specifically the absence of consistent coverage of more than 500× for cDNA, a VAF cut-off of 2% was used for downstream analysis (Figure 5-3). Furthermore, outliers (n = 2) and extreme VAF measurements (n = 2) were checked against the Integrated Genomics Viewer (IGV)(255, 256). The extreme values were found to be artefacts due to low coverage, while the outliers were heterozygous SNPs. Next up, I checked all detected variants against COSMIC(257) and rejected the variants that were recurringly called in each run, but not previously reported in COSMIC. The resulting dataset included 53 variant calls. To minimise interpretation bias, I proceeded in a joint variant calling approach with the analysis of those variants without further filtering(307).

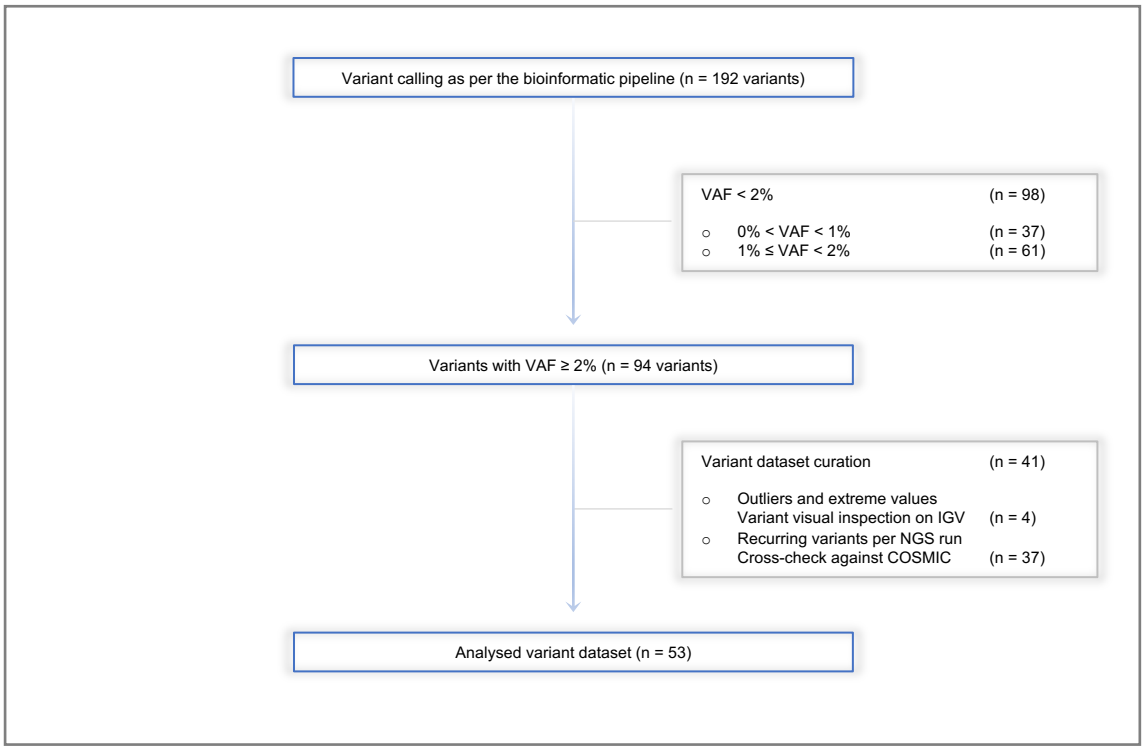


Figure 5-3. Flowchart of the variant curation in the NGS data analysis. *As the first step, variants with VAF < 2% were excluded from the analysis. In the next step, outliers and extreme values in the remaining dataset were detected and handled by visual inspection of the variants on IGV. Lastly, recurring variants which were not previously reported in COSMIC were also excluded. Altogether, the above algorithm resulted from the initial set of 192 variant calls to 53 variants that were included in the analysis. COSMIC, Catalogue Of Somatic Mutations In Cancer; IGV, Integrated Genomics Viewer; NGS, Next-generation sequencing; VAF, variant allele frequency.*

Overall, at the resolution of VAF 2%, granulocyte gDNA analysis revealed presence of CH in 12 (28.6%) samples, involving *DNMT3A*, *ASXL1*, *JAK2*, and *PPM1D* genes (Figure 5-4).

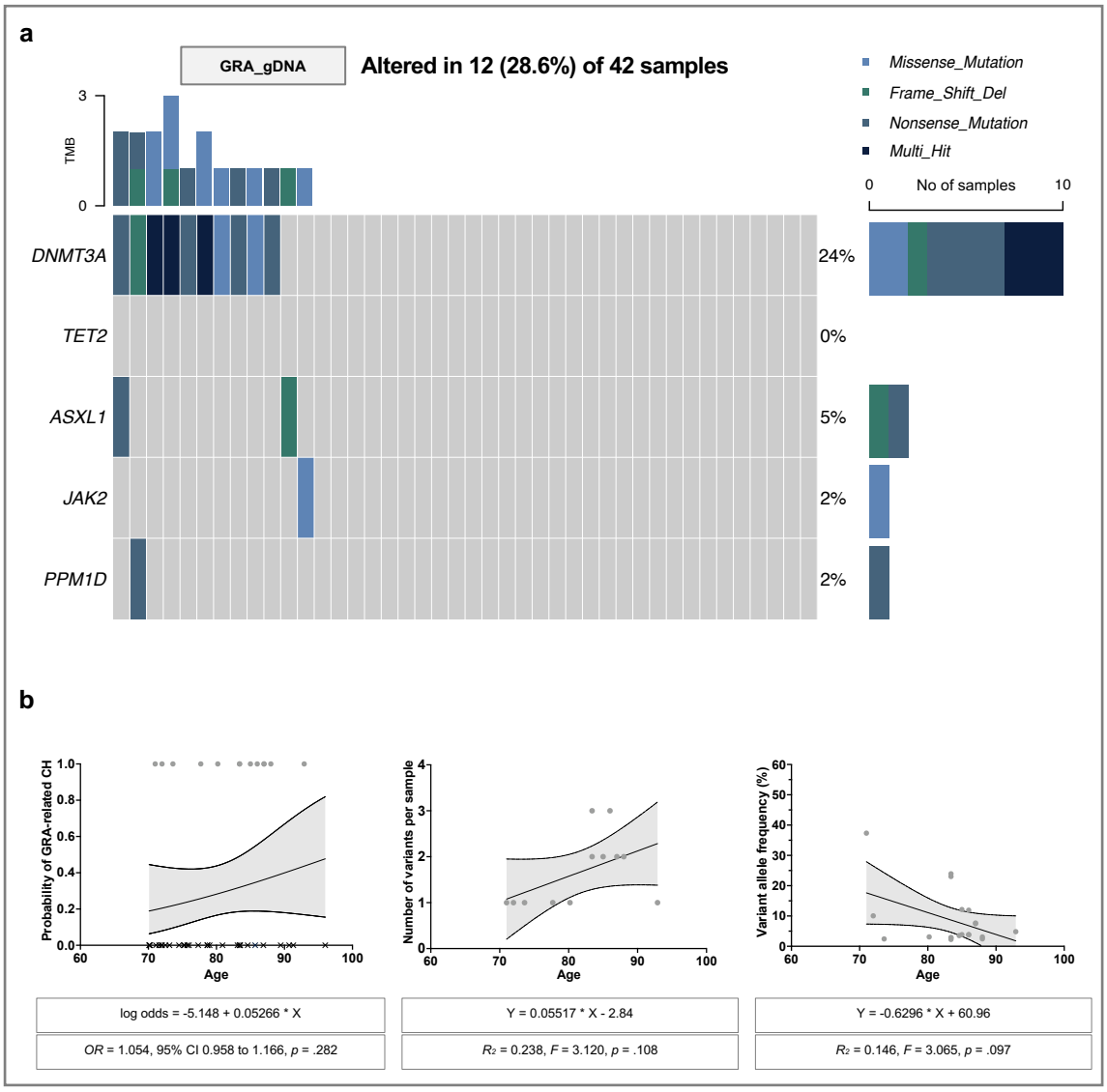


***Figure 5-4. Results of targeted next-generation sequencing via analysis of granulocyte gDNA only. a**, Oncoplot showing the detected CHIP variants and variant classification. **b**, Regression analysis of the impact of age on (left panel) presence of CH (simple logistic regression; p = .282), (middle panel) number of variants per sample (simple linear regression; p = .108), and (right panel) variant allele frequency of the detected variants (simple linear regression; p = .097). Estimated probability (curve fitted to data) is represented with the continuous line, while 95% CI for the fitted distribution is shown with the shaded area within error bands. CH, clonal haematopoiesis; CHIP, clonal haematopoiesis of indeterminate potential; GRA, granulocyte; TMB, tumour mutation burden.*

Next, I explored whether study of matched platelet cDNA could provide additional information on CH. The results of the combined analysis of the NGS sequencing data are shown in Figure 5-5 to Figure 5-8. Remarkably, platelet analysis substantially enhanced the list of detected variants, including a significant number of platelet-restricted clones, and increasing the number of individuals with CH from 12 (28.6%) to 26 (61.9%) (Figure 5-5). Figure 5-6 shows the per gene and across gene distribution of the detected variants, their co-mutation profile, and VAF distribution per gene.

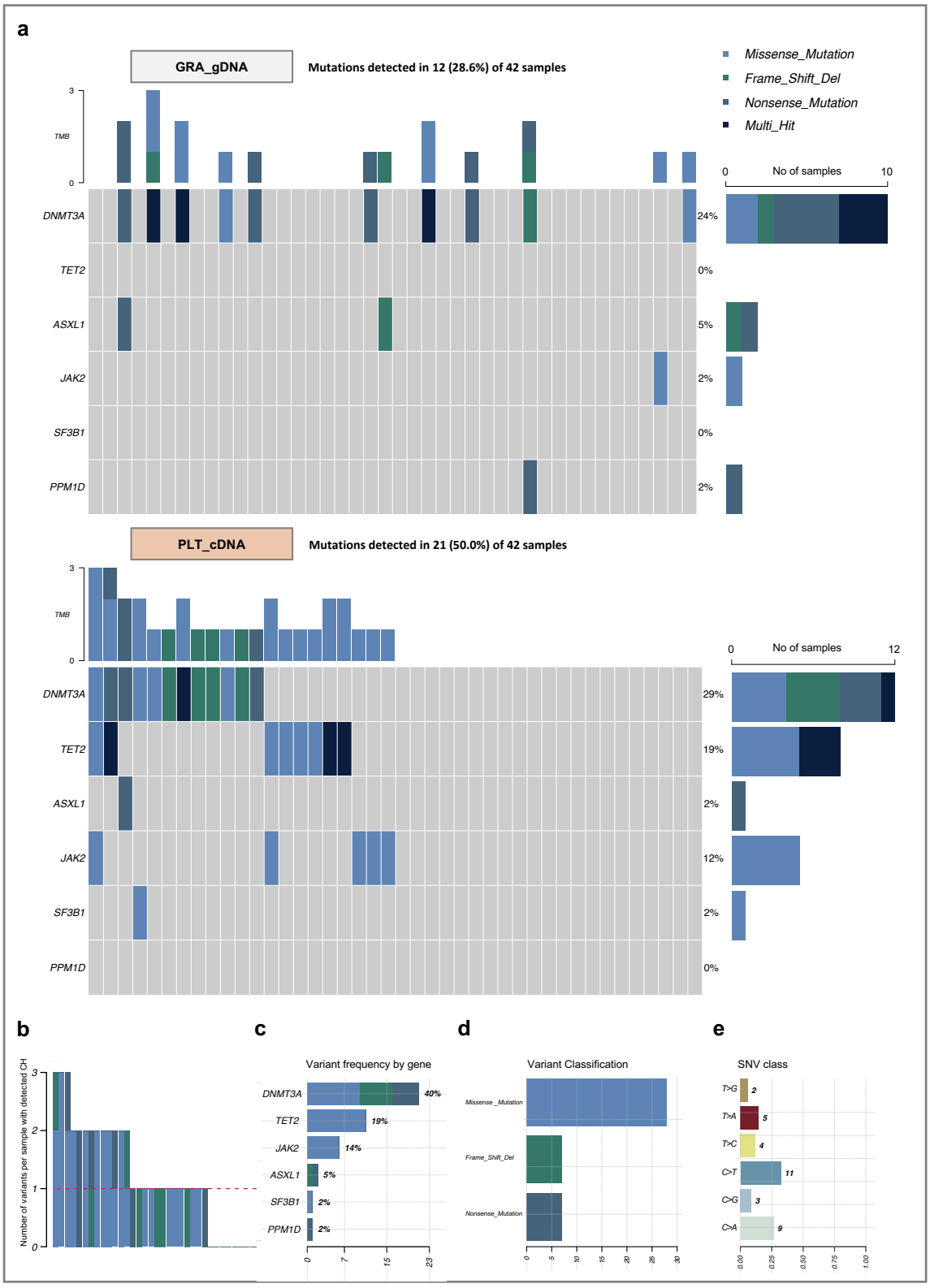


Figure 5-5. Combined analysis of the granulocyte gDNA and matched platelet cDNA NGS data: effect on the detection of CH-related clones. ***a****, Oncoplots showing the distribution and variant classification of the NGS-detected CHIP variants per gene (in horizontal orientation) in (upper panel) granulocyte gDNA and*

(lower panel) platelet cDNA for each individual (in the vertical orientation). ***b****, Number of variants per sample with CH. The red grid line corresponds to the median number of variants (median = 1).* ***c****, Gene distribution,* ***d****, variant classification, and* ***e****, SNV class of the detected variants. CH, clonal haematopoiesis; CHIP, clonal haematopoiesis of indeterminate potential; GRA, granulocyte; NGS, next-generation sequencing; PLT, platelet; SNV, single-nucleotide variant; TMB, tumour mutation burden.*

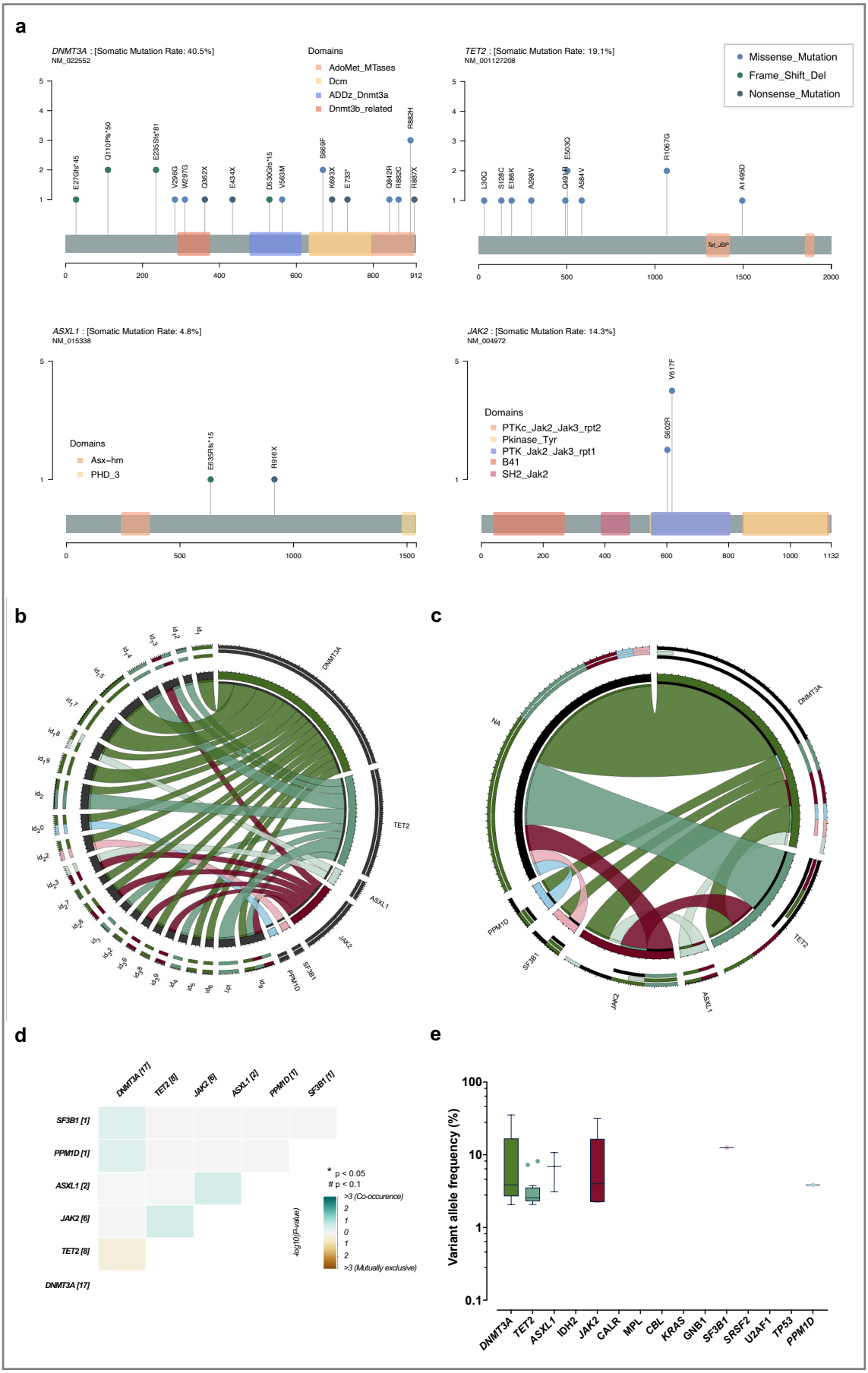

a
DNMT3A : [Somatic Mutation Rate: 40.5%]
NM_022552
Domains
AdoMet_MTases
Dcm
ADDz_Dnmt3a
Dnmt3b_related
TET2 : [Somatic Mutation Rate: 19.1%]
NM_001127208
Missense_Mutation
Frame_Shift_Del
Nonsense_Mutation
Tet_JBP
ASXL1 : [Somatic Mutation Rate: 4.8%]
NM_015338
Domains
Asx−hm
PHD_3
JAK2 : [Somatic Mutation Rate: 14.3%]
NM_004972
Domains
PTKc_Jak2_Jak3_rpt2
Pkinase_Tyr
PTK_Jak2_Jak3_rpt1
B41
SH2_Jak2
b
c
d
DNMT3A [17]
TET2 [8]
JAK2 [6]
ASXL1 [2]
PPM1D [1]
SF3B1 [1]
* p < 0.05
# p < 0.1
>3 (Co-occurence)
>3 (Mutually exclusive)
-log10(Pvalue)
e
Variant allele frequency (%)
DNMT3A
TET2
ASXL1
IDH2
JAK2
CALR
MPL
CBL
KRAS
GNB1
SF3B1
SRSF2
U2AF1
TP53
PPM1D

Figure 5-6. Combined analysis of the granulocyte gDNA and matched platelet cDNA NGS data: Summary of the characteristics of the studied population. ***a****, Lollipop plots showing the within gene distribution of the detected variants for DNMT3A, TET2, ASXL1, and JAK2.* ***b****, Circos plot visualising the distribution of CH variants per gene for each individual.* ***c****, Circos plot, showing the co-mutational profile of the study population. Segment length corresponds to number of individuals with detected variant(s) in the respective gene. Multiple variants in the same gene are not considered. Ribbons thickness corresponds to the frequency of co-occurrence of two gene variants in the same individual.* ***d****, Co-mutational spectrum taking into consideration multiple variants per gene when present in the same sample (pairwise Fisher's exact test).* ***e****, Tukey box and whisker plot, showing the distribution of variant allele frequency of the CH variants grouped by gene. CH, clonal haematopoiesis; NGS, next-generation sequencing.*

Notably, looking into the lineage distribution of detected CH variants, granulocyte-related variants counted for less than half of the detected variants, while study of platelets was able to detect approximately 75% of the variants, with 55% of them being platelet-restricted. (Figure 5-7a,b). Median VAF for the all CH variants was calculated 3.9% (95% CI, 2.2% to 32.3%) (Figure 5-7c), while there was no difference in average VAF between granulocyte-related and platelet-related CH (Figure 5-7d). Comparison of VAF between granulocyte-restricted, platelet-restricted, and non-lineage-restricted CH, showed significantly increased average VAF in the cases of non-lineage-restricted CH (p = .038), while no significant difference was detected between granulocyte-restricted and platelet-restricted clones (Figure 5-7e,f). Within the cases of non-lineage-restricted CH (n = 9) there was also no difference in the VAF measured in the granulocyte gDNA and the platelet-derived cDNA.

The detected variants were distributed across *DNMT3A*, *TET2*, *ASXL1*, *JAK2*, *SF3B1*, and *PPM1D* genes (Figure 5-5 to Figure 5-8). In more detail, *DNMT3A* accounted for 31 variant calls, consisting of 18 variants detected in 17 individuals, and the following lineage distribution: 8 granulocyte-restricted, 7 platelet-restricted, and 8 non-lineage-restricted. *TET2* accounted for 11 cases, consisting of 9 variants detected in 8 individuals. Interestingly, *TET2*-related CHIP was not detected in the granulocyte gDNA of any of the

sequenced samples, as all cases of *TET2*-related CH in this dataset were platelet-restricted (Figure 5-8). *ASXL1* accounted for 3 variant calls, consisting of 2 variants detected in 2 individuals, one being detected in granulocyte gDNA only and the other being detected in both granulocyte- and platelet-derived DNA. Two *JAK2* variants were found in 6 individuals, showing granulocyte restriction in one case and platelet restriction in the remaining of the cases. An *SF3B1* mutation (NM_012433.4(SF3B1):c.1998 p.K666N) was found in one case, showing platelet restriction, while a *PPM1D* mutation (NM_003620.4(PPM1D):c.1270 p.E424X) was found in another case, showing granulocyte restriction.

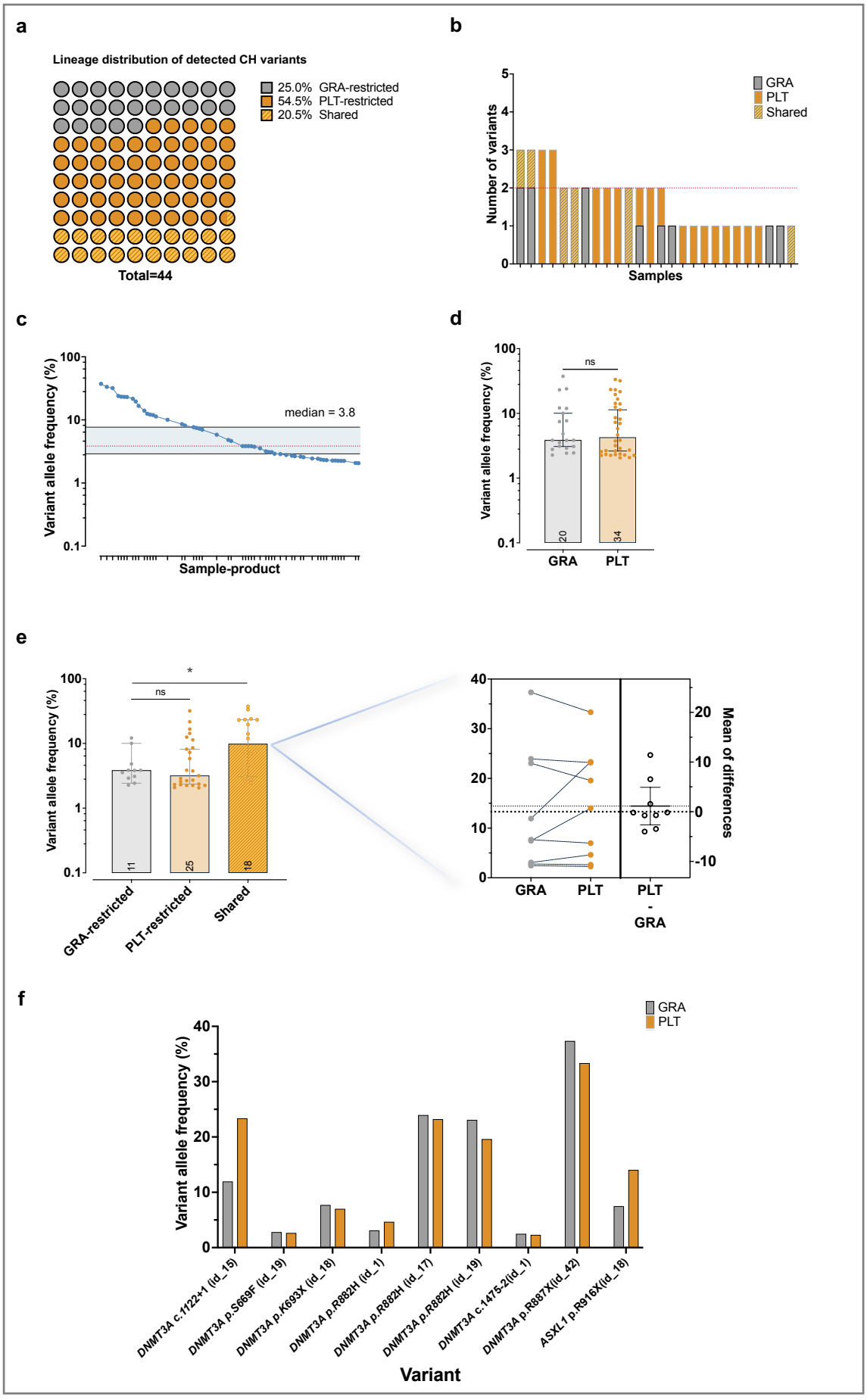

a
Lineage distribution of detected CH variants
25.0% GRA-restricted
54.5% PLT-restricted
20.5% Shared
Total=44
b
GRA
PLT
Shared
Number of variants
Samples
c
Variant allele frequency (%)
median = 3.8
Sample-product
d
ns
Variant allele frequency (%)
20
34
GRA
PLT
e
*
ns
Variant allele frequency (%)
11
25
18
GRA-restricted
PLT-restricted
Shared
Mean of differences
GRA
PLT
PLT - GRA
f
GRA
PLT
Variant allele frequency (%)
DNMT3A c.1122+1 (id_15)
DNMT3A p.S669F (id_19)
DNMT3A p.K693X (id_18)
DNMT3A p.R882H (id_1)
DNMT3A p.R882H (id_17)
DNMT3A p.R882H (id_19)
DNMT3A c.1475-2(id_1)
DNMT3A p.R887X(id_42)
ASXL1 p.R916X(id_18)
Variant

Figure 5-7. Combined analysis of the granulocyte gDNA and matched platelet cDNA NGS data: Lineage distribution of detected variants and VAF comparison. ***a****, Lineage distribution of the detected variants, and* ***b****, their distribution in each screened individual.* ***c****, Scatter plot showing the VAF of all detected variants included in the analysis (VAF > 2%) ranking from highest to lowest. The red grid line corresponds to the median VAF, while the shaded pale blue area depicts the 95% CI of median.* ***d****, Comparison of VAF between granulocyte-related CH and platelet-related CH showed no significant difference (Mann-Whitney test; p = .518). Sample size is shown at the bottom of the bars.* ***e****, (left panel) Comparison of VAF between granulocyte-restricted CH, platelet-restricted CH, and non-lineage-restricted CH showed significantly increased average VAF in the cases of non-lineage-restricted CH (Kruskal Wallis test; p = .038), while no significant difference was detected between granulocyte-restricted and platelet-restricted clones. Sample size is shown at the bottom of the bars. Focused comparison of VAF between granulocyte gDNA and matched platelet cDNA for the non-lineage-restricted CH revealed no significant difference (paired t test; p = .506).* ***f****, Bar plot showing the variants that were detected in both granulocyte- and platelet-derived DNA and their corresponding VAFs. CH, clonal haematopoiesis; GRA, granulocyte; NGS, next-generation sequencing; PLT, platelet; VAF, variant allele frequency.*

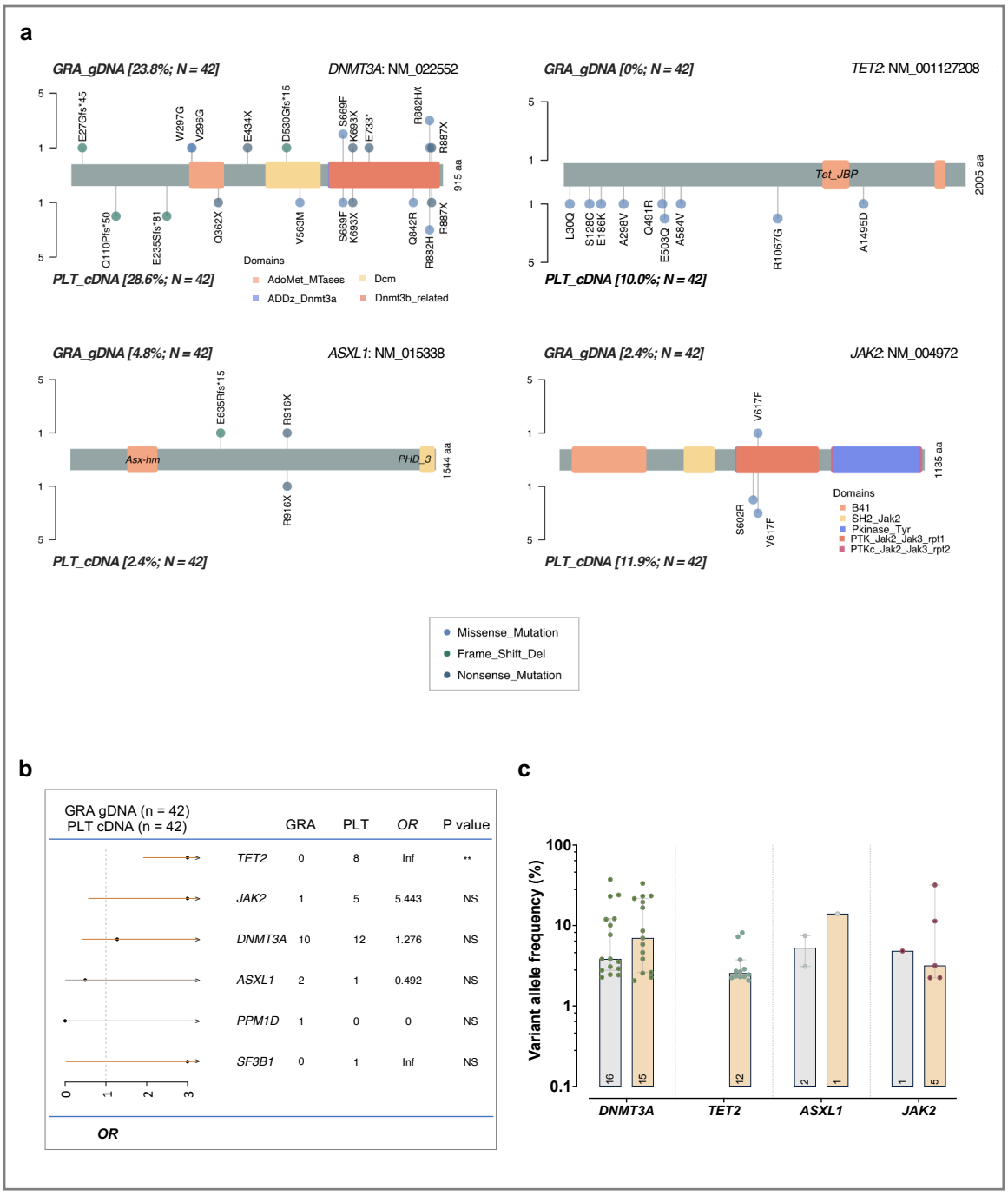


Figure 5-8. Combined analysis of the granulocyte gDNA and matched platelet cDNA NGS data: Lineage distribution of detected variants across the targeted genes. ***a****, Lollipop plots depicting lineage distribution of the detected variants for DNMT3A, TET2, ASXL1, and JAK2.* ***b****, Forest plot showing that in the present dataset, there was evidence of lineage bias (towards platelets) for presence of TET2-related CH.* ***c****, Lineage distribution of the detected variants grouped by gene. **, p ≤ .01; CH, clonal haematopoiesis; GRA, granulocytes; NGS, next-generation sequencing; PLT, platelet.*

Next, given the strong association between age and CH, I aimed to explore the relationship between age and specifically platelet-related and platelet-restricted CH. Regression analysis to assess age as the predictor for the probability for presence of CH (Figure 5-9a), number of variants in individuals with CH (Figure 5-9b), or VAF of the CH-related variants (Figure 5-9c) revealed no significant relationship for any of the analyses.

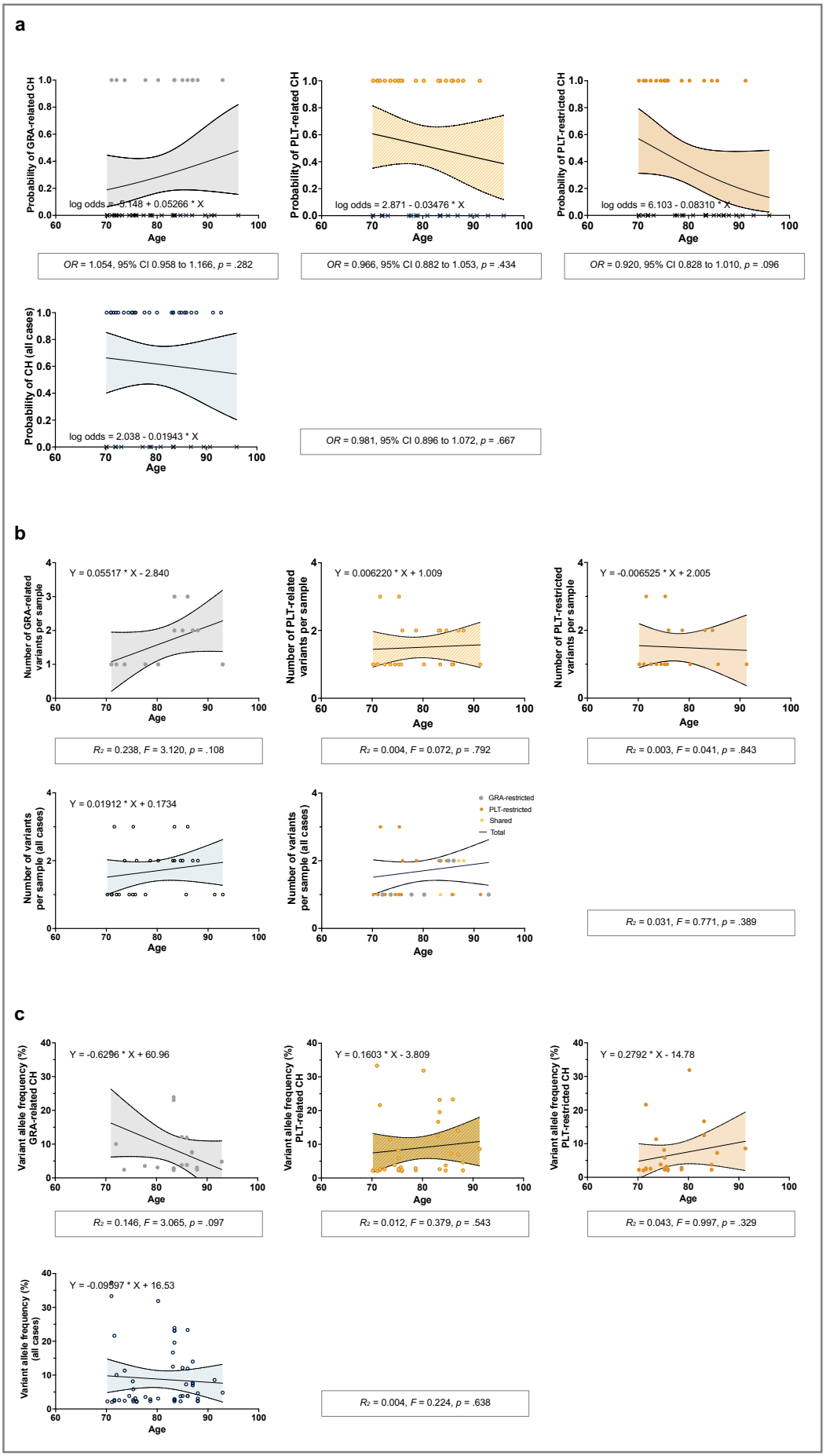

a
Probability of GRA-related CH
log odds = -5.148 + 0.05266 * X
Age
OR = 1.054, 95% CI 0.958 to 1.166, p = .282
Probability of PLT-related CH
log odds = 2.871 - 0.03476 * X
Age
OR = 0.966, 95% CI 0.882 to 1.053, p = .434
Probability of PLT-restricted CH
log odds = 6.103 - 0.08310 * X
Age
OR = 0.920, 95% CI 0.828 to 1.010, p = .096
Probability of CH (all cases)
log odds = 2.038 - 0.01943 * X
Age
OR = 0.981, 95% CI 0.896 to 1.072, p = .667
b
Number of GRA-related variants per sample
Y = 0.05517 * X - 2.840
Age
R2 = 0.238, F = 3.120, p = .108
Number of PLT-related variants per sample
Y = 0.006220 * X + 1.009
Age
R2 = 0.004, F = 0.072, p = .792
Number of PLT-restricted variants per sample
Y = -0.006525 * X + 2.005
Age
R2 = 0.003, F = 0.041, p = .843
Number of variants per sample (all cases)
Y = 0.01912 * X + 0.1734
Age
Number of variants per sample (all cases)
GRA-restricted
PLT-restricted
Shared
Total
Age
R2 = 0.031, F = 0.771, p = .389
c
Variant allele frequency (%) GRA-related CH
Y = -0.6296 * X + 60.96
Age
R2 = 0.146, F = 3.065, p = .097
Variant allele frequency (%) PLT-related CH
Y = 0.1603 * X - 3.809
Age
R2 = 0.012, F = 0.379, p = .543
Variant allele frequency (%) PLT-restricted CH
Y = 0.2792 * X - 14.78
Age
R2 = 0.043, F = 0.997, p = .329
Variant allele frequency (%) (all cases)
Y = -0.09597 * X + 16.53
Age
R2 = 0.004, F = 0.224, p = .638

Figure 5-9. Relationship between age and probability for presence of CH, number of variants in individuals with CH, or VAF of the CH-related variants. ***a****, Simple logistic regression analysis of age (as the predictor) and the probability of presence of CH (as the predicted outcome) for (left panel) granulocyte-related CH, (middle panel) platelet-related CH, (right panel) platelet-restricted CH, and (second row) all CH variants. No significant odds ratios were calculated for the present database.* ***b****, Simple linear regression analysis of age (as the predictor) and the number of CH-related variants per individual (as the predicted outcome) for (left panel) granulocyte-related CH, (middle panel) platelet-related CH, (right panel) platelet-restricted CH, and (second row) all CH variants. Based on the present database, no significant relationship was calculated for any of the comparisons.* ***C****, Simple linear regression analysis of age (as the predictor) and VAF of the CH-related variants (as the predicted outcome) for (left panel) granulocyte-related CH, (middle panel) platelet-related CH, (right panel) platelet-restricted CH, and (second row) all CH variants. Based on the present database, no significant relationship was calculated for any of the comparisons. In all plots estimated probability (curve fitted to data) is represented with the continuous line, while 95% CI for the fitted distribution is shown with the shaded area within error bands. CH, clonal haematopoiesis; GRA, granulocyte; PLT, platelet; VAF, variant allele frequency.*

5.4 Discussion

Following the identification of platelet-biased *JAK2V617F*-related CH (presented in Chapter 4), this part of my project aimed to explore whether platelet-biased CH represents a broader phenomenon involving additional genes that have previously been associated with CH and haematological cancer.

Our findings strongly support a substantial contribution of platelet studies in the detection of CH related to the wider spectrum of CH driver genes. Based on the present dataset, platelet-restricted clones accounted for more than half of the detected variants. Importantly, the variants that were included in the analysis had a VAF of ≥ 2%, a threshold that has been associated with further clonal expansion and CHIP evolution(238). Besides, platelets were compared to granulocytes rather than unfractionated total white blood cells, which would be expected to have decreased VAFs (relative to granulocytes), as they comprise a mixture of long-lived lymphocytes together with short-lived, high-turnover

granulocytes. Notably, as it has recently been reported in a pre-print that circulating platelets may contain Mk-derived gDNA fragments(308), I visually inspected the platelet-restricted variants that should have been captured by exon-exon junction cDNA-specific probes and confirmed that they corresponded to cDNA templates. Noteworthy, a cut-off VAF of ≥ 2% reduces the possibility of picking variants closer to the NGS error rate, although visual inspection of the detected variants is required and is planned as the next step in the NGS data curation. Moreover, independent validation of the detected variants through additional methodologies is required.

In a study on therapy-related CH, several correlations were identified with various blood count parameters(57); however, it is unclear whether CH has a causative association with these changes or these should be attributed to exposure to radio-/chemotherapy. Moreover, Buscarlet et al.(41) have previously reported that *TET2*-associated CH is associated with mild decrease in neutrophils, which, however, remain within the normal range. This finding was not confirmed in our database (data not shown), but notably, all cases of *TET2*-related CH (n = 8) were platelet-lineage restricted. Moreover, given that most of the blood samples were collected during hospitalisation of the study participants, it would not be safe to draw conclusions on the association of CH to blood parameters of the NGS-screened individuals.

Notably, *TET2*-driven CH has been causatively linked with atherosclerotic CVD(80), and (alongside CH related to *DNMT3A*, *IDH1/2*, *TP53*, and spliceosome gene mutations) it has also been associated with increased risk of AML in healthy individuals(104, 294). However, the clinical significance of platelet-biased clones (and whether they conform to the sequalae of granulocyte-related CH) is unclear and should be explored in further studies. It is also expected that (much alike our current understanding of granulocyte-detected CH(37, 238, 294)) different genes or variants would have a differential impact on

HSC fitness advantage and clonal evolution. Furthermore, the molecular landscape of platelet-restricted CH may differ from this of CH related to granulocytes. Besides, the spectrum of candidate driver genes in granulocyte-related CH is only partially identified(30). Notably, Guo et al.(27) have previously shown that Mks from patients with MPNs carry unique somatic mutations and reported that increased VAF correlated with increased fibrosis. Although these variants included mostly chromatin remodelling genes (such as *BOD1L1, SETD2*, and *KANSL1*), which should be considered as secondary to the multiple rounds of endoreplication during megakaryopoiesis, unique somatic variants were also frequent in cell signalling genes (such as *JAK1*, *JAK2*, *SH2B3*, *BRAF*, *RAF1*, *TLR4*, and *FGFR1*), some of them in high VAFs reaching 100%, being suggestive of a Mk-biased HSPC origin. The present work did not aim to assess clinical associations, while the targeted panel approach does not allow a thorough investigation of the molecular landscape of platelet-biased CH.

Exploratory data analysis showed that, in the present dataset, the trend over increased probability of CH as age increases was confirmed for granulocyte-related CH (though non-significant). Interestingly, for platelet-related CH there was an inverse trend with an increasing magnitude for platelet-restricted CH (which, however, did not reach significance in either analysis). Regarding the number of detected variants per study participant, granulocyte-related CH had a trend towards increasing numbers with age (though non-significant), while in platelet-related CH age seemed to have no effect on the number of detected variants per individual. Lastly, allelic burden of detected variants showed a trend towards decreasing VAFs with age for granulocyte-related CH, while in platelet-related CH age seemed to have no effect on VAF levels. Notably, these relationships were assessed in the context of exploratory data analysis and were based on simple regression models. The study design and sample size were not powered to infer

conclusive results on these relationships, which can also be affected by further NGS data curation. Confirmation of the observed patterns through sequencing of additional samples in extended cohorts is required.

Overall, in this proof-of-concept project, I aimed to assess the presence of platelet-biased/restricted CH, as a surrogate of presence of Mk-biased stem cells in human haematopoiesis. Targeted NGS showed presence of platelet-biased/restricted clones in genes that are commonly affected in CH and MPN, with significant contribution to the overall prevalence of CH. This is a novel finding that has direct implications for the study of CH and broader application in the early detection and monitoring of myeloid neoplasms.

Chapter 6 Conclusions and future directions

6.1 Summary of findings

This work highlights the importance of genetic lineage tracing in CH and MPNs and provides insight into their pathogenesis. It provides novel findings that advance our understanding of the pathophysiology of these conditions with direct clinical implications regarding their detection and monitoring, while it opens a potential for early preventive treatment strategies.

6.1.1 Genetic lineage tracing in monozygotic twins with adult-onset myeloproliferative neoplasms

In the first part of the project, I described a remarkable case of monozygotic twins presenting with *CALR* mutation-positive MPNs in their late 30s, with a clinical phenotype of PMF. Through lineage tracing of their *CALR* mutation by fragment analysis, I initially showed that mutation was somatically acquired, as it was detected in the granulocytes (derived by the MPN clone) of the twins but not in their T cells and dermal fibroblasts (as the germline DNA). Subsequently, lineage tracing of the MPN clone by WGS revealed a common clonal origin of the *CALR*-mutant MPN clone, which occurred *in utero* followed by twin-to-twin transplacental transmission and subsequent similar disease latency. Importantly, bioinformatic analysis was orthogonally validated by single-cell genotyping. Furthermore, index sorting and single-colony genotyping revealed HSCs as the likely MPN-propagating cell. Lastly, neonatal blood spot analysis confirmed *in utero* origin of the *JAK2V617F* mutation in a patient presenting with polycythaemia vera (aged 34 years),

providing strong evidence that the findings of the twin study may be a widespread phenomenon in MPN.

Altogether, the results of this paradigm-shifting study shed light in the pathogenesis and evolutionary dynamics of MPNs, suggesting a new model of prolonged latency between the emergence of the premalignant clone and the development of overt disease. Further studies in the evolutionary dynamics during this phase will improve early detection of MPNs. Moreover, early detection, in combination with available robust polygenic risk scores(295) and potentially disease-modifying treatment options,(204) can possibly result in early intervention strategies, aiming to modify clonal evolution and alter the clinical course of the disease, preventing disease clinical manifestations and complications. Notably, the new model suggested by these findings may have a broader application in cancer evolution, given the plethora of mutations under selection in histologically normal tissues,(44, 309-312) and further supported by emerging evidence from WGS studies in solid tumours.(313, 314).

6.1.2 Mutational analysis of platelet lineage cells in clonal haematopoiesis and myeloproliferative neoplasms

In the next part of my work, I focused on the study of myeloid lineage against the study of platelets as a surrogate of the Mk lineage. In view of this I developed and optimised a blood product isolation method, alongside a *JAK2V617F* ddPCR assay for analysis of either gDNA or cDNA templates, and a hybridisation, targeted, myeloid-gene NGS assay panel for simultaneous analysis of both gDNA and cDNA for genes commonly involved in CH and MPN.

I initially targeted *JAK2V6176F* and showed that the combined study of platelets alongside granulocytes can significantly increase the sensitivity for detection of *JAK2V617F* CH,

attributed to the presence of platelet-biased CH, with broad range age distribution. Importantly, study of granulocytes alone would have missed *JAK2*-related CH in 8.6% of normal individuals in the studied population, half of them accounting for clones with *JAK2V617F* VAF of more than 2% (CHIP). Besides, given the association of CH with age, it is possible that additional individuals in the cohort who were not found to have *JAK2V617F*-CH are likely to develop CH/CHIP in future.

This is the first study showing presence of platelet-restricted *JAK2V617F*-associated CH in normal individuals. Given that a key complication of CH is vascular disease as well as the central role of both platelets and *JAK2V617F* mutation in thrombosis, platelet-biased clones may have significant clinical sequelae. Furthermore, as the *JAK2V617F* mutation accounts for the vast majority of MPNs, platelet-restricted CH may also have significant implications for the subsequent development of MPN.

As the next step, I extended my study to test for presence of platelet-biased clones involving other driver genes commonly found in CH and MPN. Using a custom-made targeted myeloid-gene panel that enabled the simultaneous analysis of both granulocyte gDNA and matched platelet cDNA, I was able to detect CH detected in platelets only, but not the corresponding granulocytes. Notably, based on this dataset, platelet-restricted clones accounted for approximately 75% of the detected variants, related to the wider spectrum of known CH driver genes.

Overall, the CH studies confirmed for the first time the presence of platelet-based CH in humans, with profound effect in the detection of CH and potentially significant role in the CH-associated clinical sequalae. They highlight the significance of platelet-derived information in the study of CH and clonal haematological disorders, that is currently not addressed by the conventional genomic-based testing for their diagnosis and monitoring.

6.2 Implications of this work

This project has resulted in novel findings that have a significant contribution in the field of CH and MPN and overall understanding of the haematopoietic trajectory.

The twin study provided strong evidence of long latency in the evolutionary trajectory of premalignant clones in MPNs, which is further supported by emerging data on the evolutionary landscape of MPN(184, 185) and solid tumours(30, 279). Collectively, these results have established a new model in the pathogenesis of MPNs, which has revealed a prolonged pre-disease phase that has direct implications in early detection and could also result in preventive intervention strategies aiming to modify clonal evolution and drastically alter the clinical course of the disease.

The CH studies have resulted in the detection of platelet-biased CH in humans, that has also been a novel finding. Platelet-biased clones carried mutations in *JAK2* as well as other genes commonly involved in CH and MPNs and had a significant contribution in the overall prevalence of CH. This finding has direct implication for CH detection and early diagnosis of CH-associated conditions. Given that a key complication of CH is vascular disease as well as the central role of both platelets and *JAK2V617F* mutation in thrombosis, platelet-biased clones may have significant clinical sequelae. Importantly, the presence of platelet-restricted CH has considerable implications for genomic-based approaches for early detection and monitoring of disease, which, in haematology, are primarily based on isolation of nucleic acid from leukocytes and would fail to detect emerging platelet-biased clones.

Besides, given the association of CH with haematological malignancy, it is reasonable to assume that platelet-biased CH might hold an increased risk of progression to MPNs, that are typically characterised by increased platelet counts. Clonal haematopoiesis plays an

important role in the premalignant state and predictive tools for calculation of its risk to progression to haematological malignancies are becoming available(109). Further understanding of the factors determining the clinical outcome of CH can result in screening and precision preventive treatment strategies.

6.3 Thesis limitations

A significant limitation of the CH study is the limited sample size, as this was aimed to be a proof-of-concept study. Therefore, it is not possible to draw a definitive conclusion on the prevalence of platelet-biased/restricted CH in normal individuals or specific subgroups. However, the findings clearly show that the study of platelets alongside the conventional study of granulocytes significantly increase the sensitivity in the detection of CH. Therefore, it is important that the findings of this study are further confirmed and expanded in larger cohorts of general population or specific subgroups of normal individuals and patients.

Furthermore, this study was designed to assess the presence of platelet-biased variants in genes that have previously been associated with CH and MPNs, using a targeted NGS panel. This does not allow for a comprehensive description of the molecular landscape of platelet-biased CH, which might be different to the known mutational spectrum in granulocytes. Besides, even in the cases of granulocyte-related CH, the majority of the detected clones have no known driver mutations(30). Moreover, given the novelty of the CH study findings, the detected platelet-restricted variants should be independently validated by additional assays.

It should also be mentioned that this study did not aim to test longevity and longitudinal dynamics of the detected clones. Notably, the variants that were included in the analysis had a VAF ≥ 2%, a threshold that has been associated with further clonal expansion and CHIP evolution(238). However, it is still unclear whether the detected clones are transient or established clones with potential clinical consequences. Longitudinal cohort studies would enable monitoring of such clones and description of their evolutionary trajectory.

Lastly, the design of our study did not allow for phenotypic associations. As presence of platelet-biased CH is a novel finding, its clinical relevance is unclear, and it should be investigated in well-designed epidemiological studies, confirmed by further mechanistic studies.

6.4 Future directions

This thesis has enabled an increased understanding of the pathogenesis of MPNs. Its results suggested a new model of increased latency from the acquisition of the driver mutation to the development of full-blown disease, providing a potential for early detection and precision prevention and/or treatment. Further twin studies could enhance our understanding of the evolutionary trajectory of premalignant clones in MPNs and other myeloid neoplasms. Deeper understanding of the evolutionary dynamics of such clones would improve early detection, while further development of treatment agents with potential disease-modifying effect would allow implementation of early preventive treatment strategies.

Furthermore, the novel findings on platelet-biased CH should be further confirmed in larger cohorts, that would enable a more detailed description of its characteristics. Genetic

association and longitudinal studies can unveil the prevalence and clinical consequences of platelet-biased/restricted CH. Ultimately, this would allow the integration of platelet-biased/restricted CH into robust risk assessment and stratification tools, which would allow the identification of individuals at increased risk of complications and implementation of precision medicine strategies. Therefore, these studies should be carefully designed to include a comprehensive assessment of additional (classical and novel) risk factors for the expected clinical sequelae, including but not limited to CVD and cancer.

Apart from association studies, mechanistic studies are also necessary for understanding the biology of platelet-biased/restricted CH and its phenotypic associations. Mouse models mimicking CH have been in development by Mead group, while the Psaila group have recently developed a human bone marrow organoid(315), that can also be used in the study of the biology of platelet-biased CH. Using a *GP1ba-Cre* transgenic mouse model (which allows highly specific deletion of floxed genes in the *Mk* lineage and delineation of platelet and leukocyte functions)(316), we could model platelet-restricted CH (crossing the *Cre* with *JAK2*- or *TET2*-mutant mice) to assess gene expression and clonal dynamics, systemic inflammation profiling, platelet function, risk of malignancy as well as tumour microenvironment, cardiovascular phenotype, and drug screening. Notably, while mouse models represent an overall useful tool in the study of haematopoiesis, the function of mouse and human genes are not always fully conserved. Moreover, mice average lifespan is only approximately 2.5 years, that restrains the study of slow-developing CH. Additional models are therefore necessary, such as non-human primate models, to study CH in longer-lived and more representative models. Besides, although mouse models carrying mutations in *TET2* and/or *DNMT3A* develop CH-associated phenotypes, these mutations have interestingly not been reported as recurring in aged mice(317). Additional use of an

organoid model of platelet-restricted CH can provide an indication of intrinsic Mk-biased HSC clonal dynamics and dissect the interactions of clones with the bone marrow niche.

The present work assessed prevalence of the platelet-biased CH, but longitudinal studies are important to assess longevity of the detected platelet-biased clones. Previous epidemiological studies in CH required large human cohorts in order to be able to reach conclusive results(37, 44, 238, 294), but to my knowledge no such cohorts with stored platelet material are available and prospective collection will require a considerable expenditure of resources. However, the results of this thesis suggest that platelet-biased clones significantly contribute to CH and understanding of their role in normal haematopoiesis and disease is important.

Study of pre-leukaemic CH clones in AML have provided evidence that CH originates within functional HSCs and do not inhibit HSC multilineage potential(318). Single-cell multiomic HSC studies of the platelet-biased clones will provide insights into the biology of platelet-biased CH and the characterisation of the HSCs that initiate and drive the expansion of these clones.

Further studies should also explore the molecular landscape of platelet-biased CH and whether this is conserved between granulocyte-related and platelet-related CH. Extensive characterisation of the transcriptomic landscape of platelets through RNA sequencing can provide insights into Mk/platelet gene expression alterations in individuals carrying such clones.

Notably, in view of understanding the biology of these clones, apart from the cell intrinsic alterations, it is also vital to investigate the role of the aging bone marrow niche on their development and clonal expansion.

6.5 Conclusion

This work highlights the importance of genetic lineage tracing in CH and MPNs and provides insight into their molecular pathogenesis. It introduces a novel model of MPN pathogenesis, involving a prolonged disease latency between the acquisition of the premalignant MPN clone and the development of established disease, thus opening the potential for early diagnosis and preventive precision treatment strategies. It also describes for the first time presence of platelet-biased CH with direct clinical implications in the detection of CH and a broader importance for the diagnosis and monitoring of haematological disorders.

Appendix A Supplementary Material and Methods

A.1 Blood product isolation protocol

A.1.1 Equipment and Reagents

A.1.1.1 Reagents

1. List of reagents:

Table A-1. Summary of reagents used in the blood cell isolation protocol

Reagent	Manufacturer	Cat. No. / ID	Application
CD3 MicroBeads human	Miltenyi Biotec	130-050-101	T cell isolation only
CD34 MicroBead kit UltraPure human	Miltenyi Biotec	130-100-453	HSPC isolation only
DMSO	MilliporeSigma (Merck)	67-68-5	Freezing media
EDTA	Corning Inc.	46-034-CI	Wash Buffer
Ficoll-Paque Premium	GE Healthcare	17-5442-03	
FBS	MilliporeSigma (Merck)	F9665-500ML	
PBS, pH 7.4	Life Technologies (Thermo Fisher Scientific Inc.)	10010-056	
RBC Lysis Buffer (10X)	eBioscience (Thermo Fisher Scientific Inc.)	00-4300-54	
Trypan Blue solution	MilliporeSigma (Merck)	T8154-100ML	
Water, cell culture grade	MilliporeSigma (Merck)	W3500-6X500ML	

Note: “Freezing Medium”: 90% FBS, 10% DMSO; “PLT Freezing Medium”: 0.9% (w/v) NaCl (normal saline), 27% DMSO; “Wash Buffer”: PBS pH 7.4, 0.5% FBS, 2 mM EDTA. DMSO, dimethyl sulfoxide; EDTA, ethylenediaminetetraacetic acid; FBS, fetal bovine serum; HSPC, haematopoietic stem/progenitor cell; PBS, phosphate-buffered saline; PLT, platelet; RBC, red blood cell.

2. Plastics:

Falcon tubes (15ml, 50ml); Eppendorf tubes; Serological Pipettes; Sterile filter tips; Sterile plastic Pasteur pipettes; Bio disposal bottle; Barcoded cryovials (FluidX, #68-1002-11; AltemisLab, #202076-1), 70-µm cell strainers (Greiner Bio-One Ltd, #542070) [*primarily for marrow aspirates and clumpy / old samples*].

A.1.1.2 Equipment

Centrifuge; Haemocytometer; Liquid Nitrogen tank; MACS Separator (Miltenyi Biotec); Pipettor; Tabletop centrifuge; Tissue culture hood; -80ºC freezer.

A.1.2 Method

Skip to step 10 if plasma (PLA) and platelets (PLT) not required.

A.1.2.1 Plasma isolation

1. Centrifuge blood in their collection vials at 200 g, 10 min, room temperature, deceleration off.
2. Remove approximately 2/3 of the (platelet-rich) plasma layer (upper layer of the centrifuged sample) and put in an Eppendorf tube.
3. Take 500 µl of the (platelet-rich) plasma in the Eppendorf tube, put in another Eppendorf tube and centrifuge at 1250 g, 10 min, room temperature.
4. Collect supernatant PLA and aliquot in two PLA Eppendorf tubes (containing 200 µl of PLA each).

Label Eppendorf tube, put on ice (or dry ice), then store in -80ºC freezer making a note of the label number and where it was stored in the sample processing sheet.

5. If PLT are required continue to the next step. Keep the rest of the sample aside for MNC, Granulocyte, T cell and/or HSPC isolation later.

Label format: AJM-PLA-XXXXXX

A.1.2.2 Platelet isolation

Following on from step 2, after centrifugation and isolation of the plasma:

6. Divide remaining plasma into 1 ml Eppendorf tube aliquots (ideally three to four PLT aliquots).
7. Add 265 µl of saline 27% DMSO per 1 ml aliquot (adjust for plasma volume if required[#]) to achieve a 6% final DMSO concentration.

   # $V_x = 0.06 * V_P / 0.2262$,

   V_x = required volume of saline, V_P = existing volume of plasma.
8. Centrifuge at 1250 g, 10 min, room temperature, and remove supernatant.
9. Label and store PLT pellet at -80°C. It can be kept on dry ice as you process the sample.

Label format: AJM-PTL-XXXXXX

A.1.2.3 MNC, GRA, T cell and/or HSPC isolation

After the plasma / platelets have been removed or if not required, start with whole blood.

MNC isolation

10. If bone marrow (or a clumpy blood sample) pass through a 70-µm nylon cell strainer into a 50 ml tube, note the volume.
11. Rinse out the sample tube with Wash Buffer (and pass through 70-µm nylon cell strainer if used in step 10) to ensure that sample is diluted 1:1 compared to initial volume.
12. Gently mix the diluted sample.
13. Use half the volume of Ficoll as you have diluted blood, e.g. layer 10 ml of Ficoll and then 20 ml diluted bone marrow or blood on top in a 50-ml falcon tube.

Note: Take care not to disturb the interface between the two phases.

Use the “gravity” setting on the pipette boy and hold the falcon tube at a near-horizontal angle.

14. Centrifuge at 400 g, 30 min, room temperature, deceleration off.
15. Collect the MNCs in the inter-phase (white cloudy ring, “buffy coat”) into a fresh 50-ml falcon tube, make up to 30 to 50 ml with wash buffer. Pool MNC samples if several tubes have been used for Ficoll step.
16. Take aliquot of this solution to count MNC number with trypan blue. Place aliquot in clean Eppendorf tube and add Trypan Blue solution to it, mix well, and load to the haemocytometer.

    Typically, 1:2, e.g., 10 µl cells / 10µl Trypan Blue (but could do 1:10 or more),

    Calculation for cell number:

    Concentration of viable cells per ml = (average count from each of the 4 sets of 16 corner squares) * dilution * 10^4

    Total number viable cells = (viable cells / ml) * (suspension vol)

If MNC <u>pellet</u> is required, before proceeding to step 17, proceed with “MNC pellet isolation”.

17. Centrifuge at 300 g, 10 min, 4°C, and remove the supernatant (do not pour, use stripette instead).
18. Aim to freeze 10 vials of cells max. Aim for 5 ×10^6 to 30 ×10^6 cells per vial. Resuspend cell pellet with Freezing Medium (90% FBS, 10% DMSO).

For samples with more than 500 ×10^6 cells, freeze 10 vials of MNCs at 20 ×10^6 each and perform $CD34^+$ selection for the remaining 300 ×10^6 cells.

19. Transfer into barcoded cryovials (1 ml per vial), place into a cool cell, and immediately transfer to -80°C overnight before transferring to LN2 tanks. Make a note of the barcodes used in the sample processing sheet.

Label format: XXXXXXXXXX

MNC pellet isolation

Following on from step 16 of the "MNC, GRA and T cell isolation", after counting MNCs:

17. Gently mix the MNC solution, remove a volume corresponding to 10 $\times 10^6$ cells, and put in an Eppendorf tube.

18. Centrifuge at 500 g, 5 min, 4°C, remove the supernatant, and label.

Label format: AJM-MNC-XXXXXX

GRA pellet isolation

Following on from step 15 of the "MNC, GRA and T cell isolation", after the buffy coat has been removed:

16. Remove the remainder of the Ficoll to within 0.5 to 1 cm of the red cell layer (containing granulocytes) with a stripette.
17. Resuspend the red cell layer by gently flicking the tube and make up to 45 ml with chilled, unfiltered, 1× RBC Lysis Buffer
    (typically, RBC Lysis Buffer should be at least 3 times the volume of the red cell layer volume).
18. Incubate for 10 min in room temperature, mixing by inversion every 5 min. Lysed product has a dark red appearance.

19. Centrifuge at 500 g for 5 min, 4°C.
20. Discard supernatant

    (if red cell pellet is loose, remove supernatant using a stripette to within 1 cm of the cells).
21. If needed, repeat steps 17-20 once more.
22. Resuspend granulocytes in 1 ml Wash Buffer and transfer to an Eppendorf tube.
23. Centrifuge at 500 g for 5 min, 4°C, remove the supernatant, and label.

Label format: AJM-GRA-XXXXXX

T cell isolation ($CD3^+$ cell selection)

Following on from step 16 of the "MNC, GRA and T cell isolation", after counting MNCs:

(work fast, keep the cells cooled, use pre-cooled solutions to prevent capping of antibodies on the cell surface and non-specific cell labelling).

Collect 10 $\times 10^6$ MNCs for normal donor, PV and ET samples and 20 $\times 10^6$ MNCs for MF and samples.

If MNC number is over 300 $\times 10^6$ just use the remainder for this process.

17. Centrifuge at 300 g, 10 min, 4°C and remove supernatant.
18. Resuspend cell pellet in 80 µl of MACS buffer per 10 $\times 10^6$ MNCs

    (i.e., for a higher number of cells, scale up all reagent volumes and total volumes accordingly).
19. Add 20 µl of CD3 MicroBeads per 10 $\times 10^6$ MNCs and mix well. If more cells used, scale up accordingly.

20. Incubate for 15 min in the refrigerator (4-8°C).

21. Wash cells by adding 1 to 2 ml of Wash Buffer per 10 $\times 10^6$ MNCs.

22. Centrifuge at 300 g, 10 mins, 4°C and remove supernatant.

23. Resuspend up to 100 $\times 10^6$ cells in 500 µl Wash Buffer

    (for higher cell numbers, scale up buffer volume accordingly).

24. Place a LS column in the magnetic field of a suitable MACS Separator.

25. Prepare column by rinsing with 3 ml of Wash Buffer. Allow all liquid to pass the column.

26. Apply cell suspension onto the column.

27. If less than 10 vials of MNCs have been frozen down, then collect flow through (negative fraction) and store as CD3 negative MNCs. Wash column with 3 × 3 ml for LS column. Add each 3 ml of buffer sequentially, waiting each time for column reservoir to be empty.

28. Remove column from the separator and place it on a suitable collection tube (15-ml conical tube).

29. Pipette 5 ml of Wash Buffer onto the column. Immediately flush out fraction with the magnetically labelled cells by firmly applying the plunger supplied with the column.

30. Centrifuge at 500 g, 5 min, 4°C, and aspirate supernatant.

    Resuspend in 1 ml of Wash Buffer and transfer to labelled Eppendorf tube.

31. Centrifuge at 500 g, 5 min, 4°C, and remove as much liquid as possible without disturbing the pellet.

The pellet might be difficult to visualise.

32. Store T cell pellet at -80°C for later extraction and make note of location in processing sheet.

Label format: AJM-TCL-XXXXXX

HSPC isolation (CD34+ cell selection)

Following on from step 16 of the "MNC, GRA and T cell isolation", after counting MNCs:

(work fast, keep the cells cooled, use pre-cooled solutions to prevent capping of antibodies on the cell surface and non-specific cell labelling). **If MNC number is over 300 $\times 10^6$ just use the remainder for this process. Maximum cells per LS colum is 2 $\times 10^9$ total cells and the max number of labelled cells is 10^8).**

17. Centrifuge at 300 g, 10 min, and aspirate supernatant completely.

18. Resuspend cell pellet in 300 µL of buffer for up to 10^8 total cells.

Note: For a higher number of cells, scale up all reagent volumes and total volumes accordingly.

19. Add 100 µL of FcR Blocking Reagent for up to 10^8 total cells.

20. Add 100 µL of CD34 MicroBeads UltraPure (mix well before) for up to 10^8 total cells.

21. Mix well and incubate for 30 min in the refrigerator (2-8°C).

22. Wash cells by adding 5 to 10 ml of buffer for up to 10^8 cells and centrifuge at 300 g, 10 min, 4°C.

23. Aspirate supernatant completely and resuspend up to 10^8 cells in 500 µL of buffer.

Note: For higher cell numbers, scale up buffer volume accordingly.

24. Place a LS column in the magnetic field of a suitable MACS Separator.

25. Prepare column by rinsing with 3 ml of Wash Buffer. Allow all liquid to pass the column.

26. Apply cell suspension onto the column.

27. Wash column with 3 x 3 ml for LS column. Add each 3 ml of buffer sequentially, waiting each time for column reservoir to be empty.

28. Remove column from the separator and place it on a suitable collection tube (15-ml conical tube).

29. Pipette 5 ml of Wash Buffer onto the column. Immediately flush out fraction with the magnetically labelled cells by firmly applying the plunger supplied with the column.

30. Take aliquot to count MNCs with trypan blue.

Typically, 1:2, e.g., 10 µl cells / 10 µl Trypan Blue (but could do 1:10 or more)

Calculation for cell number:

Concentration of viable cells per ml = (average count from each of the 4 sets of 16 corner squares) * dilution * 10^4

Total number viable cells = (viable cells per millilitre) * (suspension vol)

31. Centrifuge at 350 g for 10 min and remove the supernatant (do not pour, use stripette instead).

32. Aim to freeze 10 vials of cells max. Aim for 0.5 $\times 10^6$ to 1 $\times 10^6$ cells per vial.

Resuspend cell pellet with Freezing Medium (90% FBS, 10% DMSO).

33. Transfer into barcoded cryovials (1 ml per vial), place into a cool cell immediately transfer to -80^{o}C overnight before transferring to LN2 tanks. Make a note of the barcodes used in the sample processing sheet.

Label format: XXXXXXXXXX

Appendix B Supplementary Tables

B.1 Catalogue of the genes targeted in the extended next-generation sequencing myeloid-gene panel

Gene	RefSeq_accession_number
ABL1	NM_005157; NM_007313
ANKRD26	NM_001256053; NM_014915
ASXL1	NM_015338
ATRX	NM_000489
BCOR	NM_001123383; NM_017745
BCORL1	NM_021946; NM_001184772
BRAF	NM_004333
BRCC3	NM_024332
CALR	NM_004343
CBL	NM_005188
CBLB	NM_170662
CBLC	NM_012116
CDKN2A	NM_001195132; NM_000077; NM_058195
CEBPA	NM_004364
CHEK2	NM_001005735; NM_007194
CSF3R	NM_156039
CUX1	NM_001202543; NM_181552
DDX41	NM_016222
DNMT3A	NM_022552; NM_175629
ETV6	NM_001987
EZH2	NM_004456
FBXW7	NM_033632; NM_018315
FLT3	NM_004119
GATA1	NM_002049
GATA2	NM_032638; NM_001145661
GNAS	NM_000516; NM_001077490
HRAS	NM_005343; NM_001130442
IDH1	NM_005896; NM_001282386; NM_001282387
IDH2	NM_002168
IKZF1	NM_001220765; NM_006060
JAK2	NM_004972; NM_001322194; NM_001322195; NM_001322196
JAK3	NM_000215
KDM6A	NM_021140
KIT	NM_000222

KRAS	NM_033360
KMT2A	NM_001197104; NM_005933
MPL	NM_005373
MYD88	NM_001172567
NF1	NM_000267
NFE2	NM_001136023; NM_001261461; NM_006163
NOTCH1	NM_017617
NPM1	NM_002520; NM_001355006
NRAS	NM_002524
PDGFRA	NM_006206
PHF6	NM_032458; NM_001015877
PPM1D	NM_003620
PTEN	NM_000314
PTPN11	NM_002834
RAD21	NM_006265
RUNX1	NM_001754
SETBP1	NM_015559
SETD2	NM_014159
SETDB1	NM_001145415
SF3B1	NM_012433
SH2B3	NM_005475
SMC1A	NM_006306
SMC3	NM_005445
SRSF2	NM_001195427; NM_003016
STAG1	NM_005862
STAG2	NM_001042749; NM_006603; NM_001282418
STAT3	NM_139276
STAT5B	NM_012448
TET2	NM_001127208
TP53	NM_000546; NM_001126112
U2AF1	NM_001025203
WT1	NM_024426
ZRSR2	NM_005089

Note: The list is alphabetised by gene name. Gene names are provided in italics in accordance with the current HUGO Gene Nomenclature Committee (HGNC) guidelines(173).

B.2 Summary of multidimensional mutation clustering

(a) Summary of mutation assignments

Cluster_no	CCF_in_Twin_A	CCF_in_Twin_B	Estimated_no_of_mutations	No_of_mutations_assigned
1	0.000	0.315	452.348	452
2	0.359	0.000	425.794	426
3	0.898	1.003	9.859	10

(b) Individual mutation assignments

Most_likely_cluster	Chr	Mutation_start	Twin_A_Subclonal_Fraction	Twin_B_Subclonal_Fraction
1	1	104033383	0.000	0.300
2	1	106288464	0.254	0.000
1	1	110190300	0.000	0.674
1	1	110738023	0.000	0.434
2	1	111854251	0.167	0.000
2	1	112723366	0.444	0.000
1	1	114307529	0.000	0.271
1	1	12823937	0.000	0.271
1	1	143898748	0.000	0.643
1	1	147277076	0.000	0.362
2	1	152011365	0.291	0.000
2	1	152568644	0.444	0.000
2	1	155442289	0.342	0.044
2	1	155643901	0.838	0.000
1	1	157062690	0.000	0.251
1	1	158329018	0.000	0.643
1	1	159742120	0.000	0.334
2	1	164325871	0.392	0.000
1	1	165976313	0.000	0.181
1	1	166242296	0.000	0.322
2	1	172874706	0.296	0.000
2	1	173897024	0.272	0.000
1	1	175165067	0.000	0.241
1	1	176214397	0.000	0.470
1	1	177128801	0.000	0.478
2	1	17778871	0.190	0.000
2	1	178136491	0.747	0.000
2	1	185587702	0.644	0.000
2	1	185717034	0.533	0.000
1	1	187326523	0.000	0.217
2	1	188790583	0.491	0.000
1	1	190010459	0.000	0.507
2	1	19011134	0.508	0.000
2	1	190678569	0.503	0.000
2	1	191963370	0.471	0.000
1	1	194805882	0.000	0.329

1	1	1948293	0.000	0.592
2	1	196771008	0.667	0.000
2	1	199078846	0.368	0.101
1	1	211364253	0.000	0.247
1	1	21195179	0.000	0.931
1	1	216808720	0.000	0.310
2	1	218758531	0.410	0.000
1	1	228759153	0.000	0.724
1	1	231661954	0.000	0.310
1	1	239287292	0.000	0.140
1	1	239790569	0.000	0.354
3	1	54128162	0.966	0.909
1	1	57203495	0.000	0.322
2	1	58846527	0.500	0.000
1	1	68690432	0.000	0.869
2	1	81244570	0.339	0.000
2	1	82141981	0.267	0.043
1	1	90494264	0.000	0.661
2	1	91064439	1.191	0.000
2	1	91598893	0.410	0.000
2	1	93828702	0.654	0.310
1	2	103642003	0.000	0.194
1	2	104572208	0.000	0.078
2	2	105464070	0.291	0.000
1	2	110152812	0.000	0.350
1	2	112972342	0.000	0.212
2	2	113030007	0.632	0.000
1	2	113360949	0.000	0.000
1	2	114347202	0.000	0.000
2	2	11568288	0.182	0.263
1	2	122468457	0.000	0.000
2	2	122548088	0.333	0.496
2	2	124139117	0.288	0.000
2	2	12436887	0.381	0.207
1	2	124707645	0.000	0.000
1	2	125662877	0.000	0.217
2	2	126802154	0.453	0.000
2	2	13111603	0.427	0.000
2	2	135703624	0.308	0.217
1	2	136809746	0.000	0.601
1	2	139215973	0.000	0.651
2	2	139236448	0.356	0.290
1	2	139897503	0.000	0.000
2	2	140606025	0.522	0.000
2	2	142582383	0.140	0.362
1	2	142622241	0.000	0.142

2	2	145569408	0.381	0.217
2	2	145603291	0.200	0.371
1	2	147580489	0.000	0.000
2	2	148242718	0.444	0.496
1	2	14870316	0.000	0.000
2	2	153181357	0.533	0.383
1	2	156749565	0.000	0.231
1	2	158735231	0.000	0.000
2	2	163100502	0.395	0.054
2	2	167809460	0.281	0.416
2	2	174496556	0.471	0.000
2	2	175727129	0.556	0.000
2	2	17632861	0.054	0.000
1	2	17816780	0.000	0.189
2	2	181542157	0.222	0.000
1	2	181754253	0.000	0.290
2	2	184966	0.519	0.000
1	2	186809029	0.000	0.000
1	2	187978655	0.000	0.434
1	2	190310023	0.000	0.290
2	2	193562677	0.533	0.117
2	2	19486986	0.494	0.000
2	2	196497946	0.344	0.000
1	2	197588962	0.000	0.354
2	2	199505098	0.296	0.449
2	2	199984230	0.317	0.000
2	2	200975338	0.333	0.434
1	2	202941148	0.000	0.000
1	2	214886582	0.000	0.695
1	2	215762883	0.000	0.062
1	2	21701231	0.000	0.000
2	2	217649629	0.533	0.620
1	2	22021980	0.000	0.000
1	2	220248897	0.000	0.455
2	2	221924308	0.406	0.248
1	2	222991555	0.000	0.197
1	2	22585137	0.000	0.465
1	2	22585139	0.000	0.571
1	2	226321782	0.000	0.000
2	2	227692889	0.325	0.000
1	2	22994571	0.000	0.462
2	2	235048421	0.421	0.395
2	2	235142193	0.133	0.347
3	2	236024598	1.259	0.597
2	2	236165626	0.417	0.494
1	2	238340846	0.000	0.265

1	2	238788248	0.000	0.000
1	2	24019295	0.000	0.000
2	2	24266260	0.533	0.176
2	2	25819904	0.351	0.698
1	2	27314362	0.000	0.300
1	2	28287298	0.000	0.000
1	2	29298533	0.000	0.000
2	2	2982012	0.444	0.000
2	2	29907671	0.184	0.400
2	2	31671794	0.320	0.000
1	2	33204005	0.000	0.000
1	2	34609950	0.000	0.229
2	2	35209815	0.222	0.261
1	2	39543295	0.000	0.000
1	2	41851787	0.000	0.338
2	2	5026216	0.454	0.000
1	2	5809965	0.000	0.000
1	2	76343534	0.000	0.000
1	2	77138659	0.000	0.000
1	2	77305966	0.000	0.000
2	2	78265814	0.356	0.000
1	2	7937210	0.000	0.106
1	2	79741961	0.000	0.000
1	2	84976628	0.000	0.606
2	2	90026936	0.866	0.362
2	2	98022727	0.364	0.302
1	3	101873401	0.000	0.000
2	3	103475498	0.808	0.253
2	3	107699193	0.571	0.511
1	3	107880323	0.000	0.000
1	3	111393544	0.000	0.371
2	3	118298534	0.296	0.000
1	3	120551312	0.000	0.494
1	3	124318634	0.000	0.000
2	3	1259968	0.390	0.308
2	3	126102405	0.356	0.368
1	3	127184194	0.000	0.064
1	3	146968300	0.000	0.000
1	3	148123392	0.000	0.465
1	3	150872501	0.000	0.352
1	3	152516100	0.000	0.000
2	3	152974761	0.392	0.111
2	3	153068733	0.491	0.339
2	3	153706759	0.340	0.217
1	3	157178549	0.000	0.390
2	3	160115640	0.578	0.310

1	3	169022850	0.000	0.225
1	3	182077482	0.000	0.334
3	3	187366417	1.333	0.000
1	3	192580552	0.000	0.000
1	3	197965767	0.000	0.000
2	3	26440532	0.267	0.000
2	3	29763656	0.216	0.339
2	3	32019778	0.342	0.420
2	3	32526653	0.610	0.000
1	3	39756857	0.000	0.229
1	3	50879771	0.000	0.000
1	3	52975648	0.000	0.140
1	3	54092594	0.000	0.000
1	3	57095586	0.000	0.140
1	3	59401686	0.000	0.000
2	3	59893891	0.242	0.000
1	3	62217141	0.000	0.000
1	3	6268058	0.000	0.310
2	3	65074311	0.702	0.310
2	3	72405852	0.566	0.434
1	3	74463202	0.000	0.000
2	3	74980380	0.229	0.376
2	3	76086971	0.305	0.597
2	3	80858231	0.508	0.326
1	3	8179865	0.000	0.661
1	3	8330619	0.000	0.000
1	3	84248079	0.000	0.241
2	3	86921756	0.218	0.133
1	3	8749340	0.000	0.170
1	3	8766880	0.000	0.724
1	3	8800977	0.000	0.181
1	3	89748810	0.000	0.449
2	3	90430217	0.821	0.496
1	3	95223077	0.000	0.526
3	3	9646697	0.948	0.000
1	3	96820605	0.000	0.207
2	4	105234449	0.190	0.000
2	4	106394908	0.444	0.414
1	4	108394132	0.000	0.424
2	4	109876682	0.533	0.000
2	4	111359412	0.600	0.300
1	4	115259241	0.000	0.000
2	4	119161162	0.747	0.000
2	4	119833655	0.389	0.000
1	4	120599980	0.000	0.000
2	4	122600220	0.519	0.000

1	4	125617334	0.000	0.000
1	4	126647636	0.000	0.000
1	4	127771889	0.000	0.000
2	4	134850961	0.381	0.434
1	4	140154007	0.000	0.000
1	4	144879760	0.000	0.000
1	4	147616580	0.000	0.000
1	4	152098840	0.000	0.000
2	4	160621868	0.410	0.554
2	4	161252551	0.163	0.521
2	4	161288198	0.397	0.362
1	4	161716379	0.000	0.059
2	4	162816440	0.549	0.280
1	4	166613505	0.000	0.000
3	4	16908350	0.789	0.651
1	4	174170616	0.000	0.167
1	4	179639868	0.000	0.171
1	4	180135661	0.000	0.261
2	4	180265089	0.455	0.056
2	4	183453115	0.089	0.390
1	4	184638048	0.000	0.000
2	4	187661068	0.099	0.155
2	4	188628950	0.516	0.334
2	4	22103283	0.284	0.150
2	4	26343555	0.432	0.404
1	4	26875683	0.000	0.070
2	4	26935882	0.144	0.217
2	4	3011910	0.381	0.271
2	4	31169536	0.323	0.163
1	4	33615886	0.000	0.170
2	4	35852584	0.316	0.383
1	4	37418639	0.000	0.271
1	4	39429436	0.000	0.434
2	4	44413264	0.296	0.000
1	4	46668719	0.000	0.402
1	4	47220483	0.000	0.000
1	4	52190355	0.000	0.235
2	4	52545084	0.491	0.630
2	4	5718308	0.310	0.000
2	4	61895466	0.615	0.000
1	4	64216296	0.000	0.352
1	4	6451943	0.000	0.339
1	4	64665477	0.000	0.668
2	4	69630424	0.373	0.000
1	4	70520668	0.000	0.258
1	4	71648848	0.000	0.000

2	4	74668604	0.505	0.034
1	4	76047967	0.000	0.000
2	4	82531664	0.237	0.000
1	4	86440642	0.000	0.000
1	4	88387262	0.000	0.000
2	5	100034257	0.381	0.000
2	5	100071988	0.274	0.265
2	5	100155088	0.471	0.000
2	5	100295778	0.649	0.000
1	5	105561137	0.000	0.323
2	5	105941453	0.296	0.286
1	5	106685060	0.116	0.189
2	5	10937820	0.360	0.053
1	5	110479443	0.000	0.000
2	5	111449354	0.323	0.679
1	5	11361065	0.000	0.000
2	5	115666098	0.308	0.177
2	5	116208129	0.276	0.300
2	5	116797618	0.340	0.000
1	5	122619069	0.000	0.592
2	5	123223704	0.278	0.217
2	5	12387040	0.471	0.000
2	5	131362688	0.310	0.000
2	5	131622186	0.437	0.000
1	5	137416371	0.000	0.000
1	5	140836906	0.000	0.407
1	5	147046232	0.000	0.814
2	5	147611565	0.058	0.000
2	5	152460283	0.260	0.395
2	5	155157242	0.547	0.973
1	5	155317926	0.000	0.000
1	5	157005759	0.000	1.025
1	5	159980907	0.000	0.000
1	5	161395724	0.000	0.532
1	5	161830310	0.000	0.323
2	5	162912266	0.267	0.461
2	5	163087087	0.333	0.362
2	5	166031221	0.284	0.000
1	5	175182485	0.000	0.217
1	5	178484296	0.000	0.334
3	5	18474168	0.889	0.000
1	5	19703201	0.000	0.000
2	5	2114089	0.561	0.223
1	5	30228903	0.000	0.159
1	5	32183802	0.000	0.362
2	5	33779372	0.342	0.189

1	5	35163531	0.000	0.000
2	5	35831884	0.909	0.370
2	5	36636178	0.516	0.543
1	5	41232152	0.000	0.322
1	5	41663190	0.000	0.000
1	5	42222760	0.000	0.197
1	5	45194560	0.000	0.296
2	5	4579966	0.545	0.000
1	5	53160317	0.000	0.310
2	5	56387833	0.368	0.271
1	5	5721155	0.000	0.407
2	5	63548792	0.577	0.000
2	5	6643035	0.421	0.000
2	5	69055388	0.372	0.263
1	5	72404389	0.000	0.000
1	5	7425190	0.000	0.000
2	5	75285945	0.513	0.078
2	5	7817145	0.327	0.000
2	5	79269736	0.533	0.271
2	5	81377269	0.424	0.310
1	5	85129366	0.000	0.235
2	5	85688616	0.741	0.000
2	5	91742536	0.593	0.000
1	5	9500811	0.000	0.000
2	5	95231337	0.392	0.000
2	5	95701747	0.715	0.280
2	5	97765568	0.201	0.000
1	6	101437584	0.000	0.094
1	6	102787939	0.000	0.000
2	6	103570525	0.381	0.362
1	6	105574693	0.000	0.629
1	6	110560428	0.000	0.362
2	6	111252036	0.485	0.000
1	6	115939145	0.000	0.620
2	6	117652832	0.181	0.255
2	6	123368975	0.092	0.000
2	6	124380080	0.800	0.000
2	6	126389883	0.086	0.519
2	6	131059163	0.351	0.000
2	6	131399570	0.406	0.372
1	6	132482937	0.000	0.064
2	6	145547033	0.410	0.062
1	6	153868755	0.000	0.329
1	6	155942473	0.000	0.000
1	6	1850433	0.000	1.067
1	6	18553216	0.000	0.000

1	6	20182475	0.000	0.223
1	6	23725753	0.000	0.643
1	6	23820823	0.000	0.000
1	6	33355096	0.000	0.197
1	6	40322738	0.000	0.103
2	6	40934220	0.638	0.000
1	6	45795560	0.000	0.362
2	6	46776061	0.232	0.000
2	6	46995372	0.406	0.827
1	6	49245413	0.000	0.000
1	6	49866189	0.000	0.280
2	6	60497665	0.208	0.000
2	6	62533854	0.585	0.445
2	6	70161977	0.558	0.000
2	6	71671466	0.067	0.322
2	6	7238595	0.415	0.000
1	6	73054903	0.000	0.362
1	6	78332113	0.000	0.526
1	6	80558206	0.000	0.543
2	6	9330828	0.267	0.109
2	6	93696846	0.310	0.290
2	6	95778674	0.274	0.326
1	7	10181486	0.000	0.000
2	7	102065090	0.301	0.452
1	7	108865841	0.000	0.255
2	7	110615318	0.464	0.000
1	7	111235913	0.000	0.318
1	7	116155279	0.000	0.000
2	7	122016458	0.381	0.043
2	7	124142825	0.323	0.000
1	7	134178896	0.000	0.322
2	7	137296579	0.417	0.386
1	7	137739877	0.000	0.000
2	7	139519359	0.211	0.000
1	7	142637901	0.000	0.099
2	7	153528906	0.424	0.758
2	7	154611485	0.305	0.383
2	7	156931430	0.184	0.000
1	7	20905885	0.000	0.266
1	7	26002719	0.000	0.161
1	7	27125974	0.000	0.000
1	7	30124532	0.000	0.192
2	7	30831221	0.261	0.000
1	7	33443171	0.000	0.000
1	7	33920177	0.000	0.661
2	7	41732610	0.434	0.000

2	7	43378810	0.503	0.362
1	7	43433707	0.000	0.000
1	7	4358841	0.000	0.558
1	7	44053423	0.000	0.164
2	7	44090851	0.261	0.000
1	7	45053757	0.000	0.000
1	7	48916465	0.000	0.000
2	7	49269510	0.296	0.000
2	7	49269512	0.296	0.000
2	7	4979367	0.308	0.255
2	7	53091502	0.627	0.000
2	7	54587046	0.978	0.494
2	7	60485855	0.229	0.000
1	7	65590326	0.000	0.300
1	7	67521209	0.000	0.371
2	7	69630257	0.305	0.472
2	7	7056369	0.352	0.259
3	7	71219697	0.770	0.000
1	7	71503989	0.000	0.000
1	7	7383076	0.000	0.496
1	7	75784737	0.000	0.000
1	7	86130259	0.000	0.000
2	7	88505688	0.372	0.000
2	7	90565456	0.232	0.465
2	7	9227657	0.571	0.302
2	7	95205017	0.547	0.461
2	7	95298520	0.314	0.528
2	7	96055652	0.296	0.000
2	7	98561959	0.421	0.181
2	8	104362862	0.444	0.606
2	8	108043942	0.356	0.000
1	8	111647834	0.000	0.101
1	8	114563854	0.000	0.535
1	8	116551195	0.000	0.451
2	8	117615713	0.480	0.388
1	8	122234933	0.000	0.000
2	8	124975564	0.065	0.478
2	8	129273738	0.281	0.000
2	8	129372882	0.567	0.049
2	8	129733477	0.261	1.022
1	8	131552936	0.000	0.000
1	8	133562727	0.000	0.418
2	8	135261370	0.549	0.543
1	8	136497219	0.000	0.225
2	8	17691704	0.296	0.000
2	8	19550828	0.480	0.000

1	8	19798556	0.000	0.310
1	8	20863668	0.000	0.148
1	8	25151260	0.000	0.247
2	8	25892225	0.325	0.000
1	8	26940513	0.000	0.000
2	8	30978684	0.424	0.000
1	8	32400294	0.000	0.231
1	8	32524703	0.000	0.148
1	8	3391860	0.000	0.000
1	8	34194743	0.000	0.517
2	8	36379501	0.351	0.326
2	8	36570758	0.444	0.000
1	8	41188472	0.000	0.362
1	8	43877252	0.000	0.310
1	8	46315152	0.000	0.352
2	8	48680993	0.711	0.159
1	8	49102414	0.000	0.000
1	8	4997791	0.000	0.248
1	8	50264084	0.000	0.362
2	8	51711248	0.388	0.207
1	8	53427578	0.000	0.348
1	8	57557945	0.000	0.000
1	8	5826762	0.000	0.225
1	8	58506466	0.000	0.000
1	8	59792693	0.000	0.163
2	8	63967261	0.327	0.585
1	8	64469993	0.000	0.418
1	8	924933	0.000	0.090
1	8	98657310	0.000	0.000
2	9	101191229	0.218	1.048
1	9	10229559	0.000	0.000
1	9	102339974	0.000	0.000
2	9	102487319	0.690	0.290
1	9	102993672	0.000	0.434
1	9	103368304	0.000	0.560
1	9	104257777	0.000	0.163
1	9	104331598	0.000	0.000
1	9	105298502	0.000	0.000
2	9	111554985	0.274	0.189
2	9	118247951	0.086	0.000
1	9	118261593	0.000	0.117
2	9	120700338	0.604	0.000
2	9	122872477	0.491	0.139
1	9	124851965	0.000	0.000
2	9	125539853	0.254	0.000
1	9	12626664	0.000	0.000

2	9	127663224	0.174	0.000
1	9	12964481	0.000	0.000
2	9	131190221	0.698	0.255
1	9	131860557	0.000	0.000
2	9	14104419	0.519	0.323
2	9	1822815	0.511	0.630
2	9	19348792	0.410	0.114
1	9	21096138	0.000	0.965
1	9	23036773	0.000	0.819
2	9	23699353	0.250	0.000
1	9	24723875	0.000	0.372
1	9	24924904	0.000	0.000
2	9	28150104	0.235	0.217
1	9	38292964	0.000	0.000
2	9	395970	0.404	0.420
1	9	42290039	0.000	0.136
2	9	4316936	0.622	0.000
2	9	50071	0.143	0.000
1	9	5583708	0.000	0.338
1	9	595757	0.000	0.374
2	9	61566797	0.401	0.000
2	9	73087756	0.610	0.000
1	9	73452437	0.000	0.310
1	9	74317541	0.000	0.068
2	9	77004114	0.808	0.000
2	9	83484424	0.322	0.724
2	9	84858874	0.267	0.452
1	9	88848472	0.000	0.000
1	9	9208774	0.000	0.000
1	9	95722226	0.000	0.489
1	9	98676266	0.000	0.869
1	10	102661031	0.000	0.370
2	10	10849842	0.410	0.362
2	10	109091857	0.485	0.278
1	10	10933457	0.000	0.000
1	10	110609195	0.000	0.000
2	10	11168124	0.625	0.000
2	10	111830308	0.267	0.374
2	10	113814958	0.638	0.000
1	10	115355253	0.000	0.514
2	10	115409272	1.091	0.374
1	10	116513252	0.000	1.357
2	10	116526803	0.400	0.507
1	10	118845653	0.000	0.286
3	10	120224348	2.256	0.000
2	10	12067169	0.356	0.000

2	10	121613197	0.519	0.000
2	10	124588678	0.278	0.159
1	10	131096200	0.000	0.334
1	10	132055519	0.000	0.235
1	10	14501740	0.000	0.286
1	10	19776842	0.000	0.452
2	10	23617158	0.464	0.543
2	10	23958376	0.281	0.472
1	10	24984824	0.000	0.000
2	10	27864409	0.460	0.286
1	10	30158617	0.000	0.483
1	10	33849532	0.000	0.000
2	10	34892061	0.213	0.000
1	10	37826153	0.000	0.505
2	10	4183898	0.406	0.000
1	10	42358812	0.000	0.000
1	10	46311814	0.000	0.000
2	10	47221434	0.370	0.302
2	10	54174376	0.403	0.334
1	10	54568022	0.000	0.280
2	10	56140794	0.267	0.000
2	10	56950504	0.556	0.278
2	10	60147699	0.278	0.047
2	10	62705771	0.747	0.380
2	10	64210270	0.491	0.318
1	10	65525504	0.000	0.000
2	10	67571470	0.403	0.152
2	10	70021310	0.320	1.018
1	10	73104873	0.000	0.475
1	10	75129152	0.000	0.000
2	10	75905930	0.523	0.150
2	10	78393818	0.344	0.235
2	10	83492892	0.544	0.000
1	10	86750662	0.000	0.000
1	10	90700494	0.000	0.575
2	10	9239651	0.325	0.225
1	10	92770872	0.000	0.000
2	10	93187731	0.741	0.724
1	10	93234495	0.000	0.000
2	10	94576898	0.508	0.434
2	10	94950682	0.186	0.193
1	11	103587114	0.000	0.204
2	11	105917523	0.083	0.000
1	11	107071946	0.000	0.579
1	11	11456484	0.000	0.283
1	11	114708783	0.000	0.248

1	11	114958122	0.000	0.290
1	11	117410900	0.000	0.000
2	11	118973448	0.421	0.000
2	11	121731425	0.356	0.066
1	11	122025753	0.000	0.639
1	11	122979808	0.000	0.000
1	11	124291010	0.000	0.251
1	11	132785635	0.000	1.062
1	11	133338345	0.000	0.374
1	11	133761282	0.000	0.620
2	11	17634218	0.640	0.235
2	11	18214790	0.680	0.000
1	11	19063121	0.000	0.000
1	11	20315541	0.000	0.620
1	11	20325176	0.000	0.000
2	11	25629284	0.533	0.000
2	11	29268317	0.348	0.000
2	11	30414572	0.434	0.255
1	11	36738318	0.000	0.181
2	11	40970112	0.271	0.000
2	11	42181489	0.444	0.000
2	11	4222159	0.148	0.000
1	11	42391773	0.000	0.383
2	11	42992114	0.276	0.109
1	11	44879657	0.000	0.414
2	11	49443145	0.457	0.271
2	11	59547928	0.148	0.000
2	11	62907973	0.696	0.310
2	11	6802600	0.600	0.526
2	11	69424352	0.430	0.181
2	11	70172132	0.556	0.000
2	11	71917009	0.406	0.000
2	11	78991748	0.696	0.255
1	11	81625515	0.000	0.121
1	11	82375060	0.000	0.704
1	11	83760647	0.000	0.646
2	11	83859358	0.242	0.000
1	11	8836985	0.000	0.263
1	11	88442713	0.000	0.056
2	11	89364256	0.314	0.056
1	11	91384045	0.000	0.461
2	11	92625315	0.296	0.000
1	11	93116767	0.000	0.334
2	11	93572399	0.583	0.117
2	11	98979001	0.355	0.000
2	11	99123208	0.441	0.000

1	12	100396583	0.000	0.114
2	12	101478301	0.317	0.000
1	12	105124962	0.000	0.247
1	12	109600533	0.000	0.000
2	12	111703138	0.356	0.517
2	12	112201473	0.978	0.000
1	12	113992218	0.000	0.362
1	12	114026760	0.000	0.000
1	12	115283072	0.000	0.557
1	12	126210035	0.000	0.000
1	12	128451756	0.000	0.000
1	12	129158414	0.000	0.101
2	12	129194036	0.260	0.743
2	12	129699976	0.533	0.000
2	12	130201343	0.571	0.000
2	12	132658754	0.387	0.000
1	12	18037838	0.000	0.114
1	12	20532321	0.000	0.334
2	12	27135862	0.333	0.229
1	12	32986809	0.000	0.111
2	12	37544249	0.533	0.000
1	12	38746830	0.000	0.000
2	12	39556427	0.474	0.000
2	12	45801804	0.271	0.399
2	12	47093368	0.444	0.226
2	12	54193118	0.274	0.124
2	12	61096059	0.390	0.310
2	12	6117683	0.107	0.514
1	12	61470230	0.000	0.000
1	12	6307820	0.000	0.000
2	12	67241412	0.303	0.000
1	12	6789295	0.000	0.217
1	12	77504158	0.000	0.145
1	12	78382902	0.000	0.814
1	12	79065377	0.000	0.404
2	12	80465550	0.209	0.041
1	12	80780558	0.000	0.263
1	12	8205043	0.000	0.894
1	12	84484910	0.000	0.000
1	12	87232193	0.000	0.698
1	12	87810848	0.000	0.000
1	12	88071791	0.000	0.000
1	12	90780665	0.000	0.911
1	12	95719552	0.000	0.213
2	12	96021004	0.222	0.931
1	12	96233025	0.000	0.746

2	12	99195605	0.356	0.000
1	13	106932614	0.000	0.217
1	13	109688927	0.000	0.000
2	13	112177050	0.588	0.241
1	13	112351281	0.000	0.000
2	13	19458900	0.200	0.181
2	13	21198801	0.296	0.000
2	13	21198803	0.505	0.277
2	13	22486341	0.667	0.202
2	13	24407230	0.424	0.000
2	13	35458160	0.800	0.106
2	13	44807574	0.290	0.000
2	13	51185115	0.274	0.000
1	13	51338069	0.000	0.181
2	13	54356322	0.222	0.000
2	13	57950093	0.923	0.000
2	13	60127004	0.889	0.181
2	13	61979712	0.208	0.070
1	13	64005202	0.000	0.000
1	13	68771491	0.000	0.000
1	13	81289794	0.000	0.310
2	13	86387551	0.333	0.633
1	13	95505939	0.000	0.000
2	14	81424362	0.598	0.383
1	14	87098315	0.000	0.668
2	14	87211954	0.390	0.000
1	14	94965692	0.000	0.271
1	14	96647102	0.000	0.000
1	14	99920248	0.000	0.000
1	15	25046210	0.000	0.000
2	15	25170693	0.248	0.000
1	15	26017195	0.000	0.000
1	15	45848202	0.000	0.000
1	15	48080087	0.000	0.302
1	15	48928081	0.000	0.000
2	15	56330772	0.406	0.310
1	15	60120205	0.000	0.174
1	15	60352764	0.000	0.281
2	15	60956234	0.327	0.343
2	15	63420915	0.325	0.582
1	15	72502854	0.000	0.691
1	15	77267388	0.000	0.420
1	15	80681168	0.000	0.226
2	15	80950475	0.497	0.045
1	15	83132180	0.000	0.000
2	15	83158851	0.344	0.161

1	15	87983261	0.000	0.000
1	15	95004101	0.000	0.000
2	15	96972773	0.390	0.207
2	15	99025033	0.323	0.326
1	16	1550273	0.000	0.000
1	16	18537225	0.000	0.000
1	16	20424714	0.000	0.000
2	16	22306291	0.600	0.000
1	16	22677228	0.000	0.283
2	16	25681378	0.232	0.290
2	16	3192939	0.311	0.475
2	16	34085953	0.946	0.350
2	16	35232723	0.570	0.000
2	16	47099927	0.325	0.000
2	16	47823047	0.174	0.000
2	16	48066801	0.622	0.592
1	16	50321074	0.000	0.062
2	16	51823583	0.213	0.760
2	16	52276870	0.261	0.000
1	16	5534803	0.000	0.434
1	16	57821035	0.000	0.402
2	16	61886296	0.230	0.000
2	16	64432341	0.232	0.000
2	16	64614439	0.427	0.452
1	16	64699930	0.000	0.000
2	16	64919655	0.174	0.391
2	16	66602481	0.250	0.046
1	16	67105628	0.000	0.465
2	16	67608941	0.081	0.000
1	16	69148663	0.000	0.000
1	16	70802883	0.000	0.395
2	16	71030715	0.254	0.345
2	16	7113874	0.195	0.395
2	16	72362589	0.505	0.434
2	16	7450654	0.857	0.286
1	16	77256710	0.000	0.000
1	16	77493285	0.000	0.362
2	16	77781037	0.370	0.287
1	16	82066056	0.000	0.000
1	16	82874338	0.000	0.229
2	16	83070380	0.667	0.109
3	16	85313584	0.912	0.981
2	16	86284090	0.194	0.334
1	16	9532674	0.000	0.000
1	17	10363095	0.000	0.212
1	17	10978272	0.000	0.241

1	17	14670774	0.000	0.000
2	17	14728731	0.267	0.000
1	17	30755275	0.000	0.000
1	17	31863521	0.000	0.000
2	17	32918315	0.686	0.225
2	17	34590397	0.722	0.307
1	17	34811921	0.000	0.060
2	17	34849723	0.247	0.362
1	17	40381174	0.000	0.000
1	17	41094327	0.000	0.000
2	17	41287213	0.348	0.822
1	17	41488518	0.000	0.347
1	17	41662731	0.000	0.271
1	17	4511744	0.000	0.248
2	17	48780087	0.567	0.543
1	17	4976602	0.000	0.000
1	17	52148573	0.000	0.000
2	17	52202053	0.092	0.000
1	17	59529075	0.000	0.000
1	17	63728030	0.000	0.000
1	17	64448232	0.000	0.000
2	17	72065475	0.566	0.000
2	17	73729002	0.296	0.000
1	17	74236263	0.000	0.000
2	17	78164605	0.333	0.420
2	17	8080028	0.364	0.388
2	17	82853402	0.427	0.322
2	17	9019997	0.310	0.000
1	17	9021059	0.000	0.271
1	18	14791710	0.000	0.000
1	18	15293364	0.000	0.112
2	18	27867692	0.577	0.000
2	18	29896712	0.351	0.000
2	18	50652311	0.242	0.339
2	18	51531692	0.296	0.197
1	18	524695	0.000	0.000
2	18	52758063	0.718	0.395
1	18	56544081	0.000	0.319
2	18	57470379	0.404	0.000
1	18	58043724	0.000	0.452
1	18	61155354	0.000	0.271
1	18	61706959	0.000	0.217
2	18	63454630	0.271	0.000
1	18	63887009	0.000	0.334
1	18	67455918	0.000	0.000
2	18	68909930	0.325	0.345

2	18	70340966	0.496	0.263
1	18	74237595	0.000	0.370
2	18	75784106	0.496	0.724
1	18	75984267	0.000	0.000
1	18	8101586	0.000	0.000
1	18	924915	0.000	0.445
2	18	9601220	0.491	0.566
2	19	10065329	0.571	0.494
3	19	12943750	0.687	0.000
2	19	16234332	0.424	0.193
1	19	19549515	0.000	0.901
1	19	22624713	0.000	0.136
2	19	22883084	0.982	0.000
2	19	22892036	0.649	0.246
1	19	27496424	0.000	0.651
2	19	27753864	0.333	0.672
1	19	28110861	0.000	0.457
2	19	28326785	0.821	0.197
1	19	31656627	0.000	0.062
2	19	3661315	0.410	0.343
1	19	38958665	0.000	0.519
2	19	40522348	0.250	0.452
1	19	41492820	0.000	0.000
2	19	48040435	0.415	0.300
1	19	53377680	0.000	0.524
2	19	7346680	0.325	0.000
1	19	8086471	0.000	0.987
1	20	12995529	0.000	0.414
2	20	13915926	0.323	0.543
1	20	14336990	0.000	0.334
1	20	21629260	0.000	0.383
1	20	22689135	0.000	0.000
2	20	28497717	0.125	0.000
2	20	39766967	0.260	0.434
1	20	39854921	0.000	0.000
2	20	44696207	0.430	0.000
2	20	45737606	0.213	0.293
1	20	48547263	0.000	0.045
1	20	50715766	0.000	0.404
1	20	50895064	0.000	0.000
1	20	52629042	0.000	0.223
2	20	53076039	0.513	0.000
2	20	55145960	0.533	0.271
1	20	64283610	0.000	0.000
2	21	10536357	0.362	0.000
1	21	24993236	0.000	0.079

2	21	25235168	0.667	0.132
2	21	26121394	0.397	0.378
2	21	27842645	0.286	0.000
2	21	28932587	0.260	0.899
2	21	31309695	0.186	0.449
2	21	32999596	0.267	0.000
2	21	35666262	0.496	0.475
1	21	36182754	0.000	0.000
2	21	37443516	0.063	0.157
1	21	40180743	0.000	0.000
1	21	40558302	0.000	0.000
2	21	42078747	0.119	0.259
1	21	46037210	0.000	0.511
1	22	11356944	0.075	0.000
2	22	20582589	1.143	0.000
1	22	32657960	0.000	0.326
2	22	33353441	0.260	0.286
1	22	34793591	0.000	0.000
2	22	37322215	0.360	0.000
1	22	37961231	0.000	0.000
1	22	42966759	0.000	0.170
2	22	44313831	0.545	0.452
2	22	45732457	0.278	0.703
1	22	48568894	0.000	0.302

Note: Results from multidimensional Bayesian Dirichlet Process-based mutation clustering. Columns *CCF_in_Twin_A* and *CCF_in_Twin_B* show for each cluster in each twin the proportion of tumour cells that the cluster represents. CCF, cancer cell fraction.

B.3 Additional candidate cancer genes

Gene			
ABL2	*EP300*	*MKL1*	*RAD21*
AFF3	*ERCC2*	*MLL*	*RB1*
APC	*ETV6*	*MLL2*	*RET*
AR	*EZH2*	*MLL3*	*RUNX1*
ARID2	*FBXW7*	*MLL5*	*SDHAF2*
ASXL1	*FLT3*	*MPL*	*SF1*
ATE1	*FOXP1*	*MYC*	*SF3A1*
ATM	*GATA2*	*MYD88*	*SF3B1*
ATRX	*GNAS*	*MYO5A*	*SH2B3*
BCOR	*GNB1*	*NF1*	*SRSF2*
BRAF	*IDH1*	*NOTCH1*	*STAG2*
CALR	*IDH2*	*NPM1*	*STAT3*
CBL	*IKZF1*	*NRAS*	*TET2*
CBLB	*JAK2*	*PABPC1*	*TMEM127*
CDKN2A	*KDM5A*	*PHF6*	*TP53*
CEBPA	*KDM5C*	*PPARG*	*TP63*
CHEK2	*KDM6A*	*PPM1D*	*U2AF1*
CREBBP	*KIT*	*PRPF40B*	*WT1*
CUX1	*KMT2C*	*PTEN*	*ZRSR2*
DNM2	*KRAS*	*PTPN11*	
DNMT3A	*LZTR1*	*PTPRD*	

Note: Additional gene list for putative driver mutation analyses. Curated list based on literature support for genes associated with blood cancer risk of biology(274-276). Gene names are listed in alphabetical order and provided in italics in accordance with the current HUGO Gene Nomenclature Committee (HGNC) guidelines(173).

B.4 Catalogue of all somatic substitutions and indels in twin A and twin B

Individual	Chr	Pos	Ref	Alt	Mutation_ type	Annotation	Gene	Existing_ variation
Twin_A	1	7048875	C	T	Substitution	intron_variant	*CAMTA1*	rs190299624
Twin_A	1	7276191	T	TA	Insertion	intron_variant	*CAMTA1*	rs34532883
Twin_A	1	7932734	TAC	T	Deletion	intron_variant	*TNFRSF9*	rs3841802
Twin_A	1	17778871	G	A	Substitution	intron_variant	*ACTL8*	-
Twin_A	1	19011134	G	A	Substitution	regulatory_region_variant	-	-
Twin_A	1	24201443	G	A	Substitution	intron_variant	*LINC02800*	-
Twin_A	1	25634524	A	G	Substitution	intron_variant	*MAN1C1*	-
Twin_A	1	26355029	G	A	Substitution	upstream_gene_variant	*CRYBG2*	-
Twin_A	1	32905284	G	GT	Insertion	upstream_gene_variant	*TMEM54*	rs201477428
Twin_A	1	36908778	GGT	G	Deletion	intron_variant	*GRIK3*	rs56228690
Twin_A	1	53327841	CA	C	Deletion	frameshift_variant	*LRP8*	rs761955852
Twin_A	1	54128162	C	CT	Insertion	downstream_gene_variant	*CDCP2*	rs35156752
Twin_A	1	58795504	G	GA	Insertion	intron_variant	*LINC01135*	rs773735239
Twin_A	1	58846527	T	C	Substitution	intron_variant	*LINC01135*	rs1310251048
Twin_A	1	62922910	C	A	Substitution	intergenic_variant	-	-
Twin_A	1	81244570	C	T	Substitution	intergenic_variant	-	-
Twin_A	1	82141981	T	TA	Insertion	intergenic_variant	-	rs1036875026
Twin_A	1	91064439	G	T	Substitution	intergenic_variant	-	rs556744289
Twin_A	1	91598893	G	A	Substitution	upstream_gene_variant	*AL714022.1*	-
Twin_A	1	93828702	C	CT	Insertion	intron_variant	*BCAR3*	rs969811086
Twin_A	1	106288464	G	T	Substitution	intergenic_variant	-	rs183684643
Twin_A	1	111854251	A	G	Substitution	intron_variant	*KCND3*	-
Twin_A	1	112723366	C	A	Substitution	intron_variant	*TAFA3*	-
Twin_A	1	117227167	C	T	Substitution	intergenic_variant	-	rs1249405708
Twin_A	1	122502993	T	A	Substitution	intergenic_variant	-	rs1271741426
Twin_A	1	123009960	C	A	Substitution	intergenic_variant	-	rs1355262045
Twin_A	1	144518342	G	A	Substitution	intergenic_variant	-	rs1289463903
Twin_A	1	152011365	A	G	Substitution	intron_variant	*AL450992.1*	-
Twin_A	1	152568644	A	G	Substitution	upstream_gene_variant	*LCE3E*	-
Twin_A	1	155136742	C	CA	Insertion	intron_variant	*SLC50A1*	rs879631277
Twin_A	1	155442289	T	TA	Insertion	intron_variant	*ASH1L*	rs1016155747
Twin_A	1	155643901	C	T	Substitution	intron_variant	*AL353807.3*	-
Twin_A	1	164325871	G	A	Substitution	intergenic_variant	-	-
Twin_A	1	167888410	A	AT	Insertion	intron_variant	*ADCY10*	rs930197268
Twin_A	1	172874706	T	C	Substitution	intron_variant	*AL031599.1*	-
Twin_A	1	173897024	T	A	Substitution	3_prime_UTR_variant	*ZBTB37*	rs1000471968
Twin_A	1	178136491	C	T	Substitution	intron_variant	*RASAL2*	-
Twin_A	1	182688357	GCT	G	Deletion	upstream_gene_variant	*RGS8*	rs57313959
Twin_A	1	182732230	TTC	T	Deletion	upstream_gene_variant	*AL353778.1*	rs1191757083
Twin_A	1	185587702	C	A	Substitution	intron_variant	*LINC01350*	-
Twin_A	1	185717034	C	T	Substitution	regulatory_region_variant	-	rs74766303
Twin_A	1	188790583	C	T	Substitution	intergenic_variant	-	-

Twin_A	1	190678569	TGTTG	T	Deletion	intron_variant	*LINC01720*	-
Twin_A	1	191116119	T	TTA	Insertion	intron_variant	*AL138927.1*	rs386354456
Twin_A	1	191963370	T	TC	Insertion	intron_variant	*LINC02770*	-
Twin_A	1	196771008	G	T	Substitution	upstream_gene_variant	*CFHR3*	-
Twin_A	1	199078846	G	GA	Insertion	downstream_gene_variant	*LINC01221*	rs914403440
Twin_A	1	213790163	G	T	Substitution	intron_variant	*AC096639.1*	-
Twin_A	1	218758531	G	A	Substitution	intergenic_variant	-	-
Twin_A	1	237523679	G	A	Substitution	intron_variant	*RYR2*	rs1165384474
Twin_A	1	238480115	T	C	Substitution	downstream_gene_variant	*LINC01139*	rs1384492382
Twin_A	1	240415525	T	TATA	Insertion	intron_variant	*FMN2*	rs202218606
Twin_A	1	243567256	GC	G	Deletion	intron_variant	*AKT3*	-
Twin_A	2	184966	A	T	Substitution	intergenic_variant	-	-
Twin_A	2	2982012	G	A	Substitution	intron_variant	*LINC01250*	rs1038718094
Twin_A	2	5026216	TA	T	Deletion	intergenic_variant	-	rs1002511716
Twin_A	2	11568288	G	A	Substitution	intron_variant	*GREB1*	rs1225502656
Twin_A	2	12436887	T	G	Substitution	intron_variant	*MIR3681HG*	-
Twin_A	2	13111603	C	A	Substitution	intron_variant	*AC093912.1*	-
Twin_A	2	17632861	A	C	Substitution	intron_variant	*VSNL1*	-
Twin_A	2	19486986	C	CT	Insertion	intron_variant	*LINC01808*	rs796988189
Twin_A	2	24266260	C	T	Substitution	intron_variant	*ITSN2*	-
Twin_A	2	25819904	C	CTTAT	Insertion	intron_variant	*ASXL2*	rs879625913
Twin_A	2	29907671	C	G	Substitution	downstream_gene_variant	*AC106870.1*	rs1275974026
Twin_A	2	31671794	C	T	Substitution	intergenic_variant	-	rs936962348
Twin_A	2	32351096	CT	C	Deletion	intergenic_variant	-	rs747843331
Twin_A	2	35209815	A	G	Substitution	intergenic_variant	-	-
Twin_A	2	38570879	C	CA	Insertion	intron_variant	*HNRNPLL*	rs796515789
Twin_A	2	49012790	C	T	Substitution	intron_variant	*FSHR*	rs1051686975
Twin_A	2	54291307	G	A	Substitution	intron_variant	*ACYP2*	-
Twin_A	2	62581105	A	G	Substitution	intron_variant	*AC092155.1*	rs1357993052
Twin_A	2	64592352	T	TA	Insertion	3_prime_UTR_variant	*AFTPH*	rs879897290
Twin_A	2	66088937	C	T	Substitution	downstream_gene_variant	*AC007389.1*	-
Twin_A	2	78204244	C	A	Substitution	intron_variant	*AC012494.1*	-
Twin_A	2	78265814	T	C	Substitution	intron_variant	*AC012494.1*	-
Twin_A	2	79075253	T	TA	Insertion	intergenic_variant	-	rs200270320
Twin_A	2	83704515	C	T	Substitution	intergenic_variant	-	-
Twin_A	2	86143605	A	AT	Insertion	downstream_gene_variant	*PTCD3*	rs952364218
Twin_A	2	86786422	C	T	Substitution	intron_variant	*CD8A*	-
Twin_A	2	87046122	A	C	Substitution	intron_variant	*ANAPC1P2*	rs79742084
Twin_A	2	89100821	C	G	Substitution	upstream_gene_variant	*IGKV1*	-
Twin_A	2	90026936	G	A	Substitution	downstream_gene_variant	*IGKV6D*	rs568946974
Twin_A	2	93865725	G	C	Substitution	intergenic_variant	-	-
Twin_A	2	98022727	T	TTG	Insertion	regulatory_region_variant	-	rs371216469
Twin_A	2	105464070	C	CA	Insertion	intergenic_variant	-	rs1161295349
Twin_A	2	107747999	TAC	T	Deletion	intron_variant	*LINC01885*	rs146173650
Twin_A	2	113030007	C	T	Substitution	intron_variant	*IL36B*	rs771673800
Twin_A	2	115188061	AAG	A	Deletion	intron_variant	*DPP10*	rs59672278

Twin_A	2	118358987	C	T	Substitution	intergenic_variant	-	-
Twin_A	2	119522301	G	GA	Insertion	intron_variant	*SCTR*	rs797004076
Twin_A	2	122548088	A	G	Substitution	intron_variant	*AC011246.1*	rs893268258
Twin_A	2	124139117	G	GT	Insertion	downstream_gene_variant	*CNTNAP5*	rs1052761072
Twin_A	2	126462653	G	A	Substitution	intergenic_variant	-	rs147150064
Twin_A	2	126802154	C	T	Substitution	regulatory_region_variant	-	rs1390212792
Twin_A	2	130631605	G	T	Substitution	intron_variant	*POTEJ*	-
Twin_A	2	135703624	G	GT	Insertion	intron_variant	*R3HDM1*	rs1194873094
Twin_A	2	139236448	C	A	Substitution	intergenic_variant	-	-
Twin_A	2	140606025	AAG	A	Deletion	intron_variant	*LRP1B*	rs541895855
Twin_A	2	142582383	T	C	Substitution	regulatory_region_variant	-	-
Twin_A	2	145569408	G	A	Substitution	non_coding_transcript_ exon_variant	*AC092484.1*	-
Twin_A	2	145603291	G	A	Substitution	non_coding_transcript_ exon_variant	*AC079163.1*	rs373802799
Twin_A	2	148242718	A	T	Substitution	intron_variant	*MBD5*	-
Twin_A	2	153181357	C	CA	Insertion	intron_variant	*AC011901.1*	rs34588886
Twin_A	2	153730111	G	GA	Insertion	intergenic_variant	-	rs929649211
Twin_A	2	154639539	G	A	Substitution	intergenic_variant	-	-
Twin_A	2	162949464	C	G	Substitution	intergenic_variant	-	-
Twin_A	2	163100502	C	A	Substitution	intergenic_variant	-	-
Twin_A	2	167809460	T	G	Substitution	intergenic_variant	-	-
Twin_A	2	174496556	C	A	Substitution	intergenic_variant	-	-
Twin_A	2	175727129	G	T	Substitution	intergenic_variant	-	rs182157918
Twin_A	2	181542157	CAGTG	C	Deletion	intron_variant	*CERKL*	-
Twin_A	2	183736872	T	C	Substitution	intergenic_variant	-	-
Twin_A	2	193562677	CGT	C	Deletion	intergenic_variant	-	rs59053823
Twin_A	2	194108461	C	CT	Insertion	intergenic_variant	-	rs879718049
Twin_A	2	196497946	T	C	Substitution	intron_variant	*HECW2*	-
Twin_A	2	199505098	A	G	Substitution	intergenic_variant	-	-
Twin_A	2	199984230	A	G	Substitution	intron_variant	*MAIP1*	-
Twin_A	2	200975338	A	T	Substitution	downstream_gene_variant	*FAM126B*	-
Twin_A	2	202920157	C	CT	Insertion	intron_variant	*CARF*	rs969806621
Twin_A	2	217649629	G	C	Substitution	intron_variant	*DIRC3*	-
Twin_A	2	218749467	C	CT	Insertion	intron_variant	*TTLL4*	rs368058156
Twin_A	2	221924308	A	G	Substitution	intergenic_variant	-	-
Twin_A	2	227692889	T	C	Substitution	intron_variant	*SLC19A3*	-
Twin_A	2	228021121	C	T	Substitution	intron_variant	*SPHKAP*	rs531054237
Twin_A	2	234919271	G	GT	Insertion	intergenic_variant	-	rs893811865
Twin_A	2	235048421	C	A	Substitution	intron_variant	*SH3BP4*	-
Twin_A	2	235142193	T	C	Substitution	downstream_gene_variant	*CEP19P1*	rs953123210
Twin_A	2	236024598	A	G	Substitution	intron_variant	*AGAP1*	-
Twin_A	2	236165626	A	C	Substitution	3_prime_UTR_variant	*GBX2*	-
Twin_A	2	240285993	G	A	Substitution	upstream_gene_variant	*U3*	rs79159698
Twin_A	2	240487775	C	T	Substitution	intron_variant	*ANKMY1*	-
Twin_A	3	1259968	A	G	Substitution	intron_variant	*CNTN6*	-
Twin_A	3	6904723	G	GT	Insertion	intron_variant	*GRM7*	rs143641576
Twin_A	3	9646697	T	A	Substitution	upstream_gene_variant	*MTMR14*	rs1001727401

Twin_A	3	26440532	A	T	Substitution	intergenic_variant	-	-
Twin_A	3	29763656	A	C	Substitution	intron_variant	*RBMS3*	-
Twin_A	3	32019778	T	C	Substitution	intron_variant	*OSBPL10*	rs545152631
Twin_A	3	32526653	CCA	C	Deletion	3_prime_UTR_variant	*DYNC1LI1*	rs3836420
Twin_A	3	40432373	TGA	T	Deletion	downstream_gene_variant	*ENTPD3*	rs142976010
Twin_A	3	42455599	G	T	Substitution	intergenic_variant	-	-
Twin_A	3	59893891	T	C	Substitution	intron_variant	*FHIT*	-
Twin_A	3	65074311	C	T	Substitution	intergenic_variant	-	-
Twin_A	3	71644657	A	T	Substitution	intron_variant	*AC097634.4*	rs11925350
Twin_A	3	72405852	G	A	Substitution	intron_variant	*RYBP*	-
Twin_A	3	74296285	G	A	Substitution	intron_variant	*CNTN3*	rs74970959
Twin_A	3	74299076	G	A	Substitution	intron_variant	*CNTN3*	rs149006907
Twin_A	3	74980380	T	A	Substitution	regulatory_region_variant	-	-
Twin_A	3	75692610	CTTA	C	Deletion	intron_variant	*ZNF717*	rs140744949
Twin_A	3	76086971	T	C	Substitution	intron_variant	*ROBO2*	-
Twin_A	3	80858231	C	T	Substitution	intergenic_variant	-	rs982669611
Twin_A	3	86921756	A	G	Substitution	intron_variant	*VGLL3*	-
Twin_A	3	90430217	C	G	Substitution	intergenic_variant	-	-
Twin_A	3	90965981	A	G	Substitution	intergenic_variant	-	rs1304555355
Twin_A	3	95430870	C	T	Substitution	intergenic_variant	-	rs562186321
Twin_A	3	103475498	G	A	Substitution	intergenic_variant	-	-
Twin_A	3	107699193	C	T	Substitution	intron_variant	*BBX*	-
Twin_A	3	117666302	C	CT	Insertion	intergenic_variant	-	rs560342962
Twin_A	3	118298534	T	C	Substitution	intron_variant	*AC068633.1*	-
Twin_A	3	124027883	G	A	Substitution	intergenic_variant	-	-
Twin_A	3	125726439	G	A	Substitution	downstream_gene_variant	*OR7E93P*	rs147522636
Twin_A	3	126102405	G	A	Substitution	downstream_gene_variant	*ALDH1L1*	rs1348872806
Twin_A	3	147257030	A	G	Substitution	intron_variant	*AC092957.1*	-
Twin_A	3	148087688	G	GT	Insertion	intron_variant	*LINC02032*	rs35840438
Twin_A	3	152974761	T	C	Substitution	intergenic_variant	-	rs137865769
Twin_A	3	153068733	G	A	Substitution	downstream_gene_variant	*HMGN2P13*	-
Twin_A	3	153706759	T	G	Substitution	intron_variant	*LINC02006*	-
Twin_A	3	156297775	C	T	Substitution	intron_variant	*KCNAB1*	rs370869760
Twin_A	3	160115640	G	A	Substitution	intron_variant	*IL12A-AS1*	rs985125306
Twin_A	3	165572607	C	T	Substitution	intron_variant	*LINC01322*	-
Twin_A	3	168769424	G	A	Substitution	intron_variant	*EGFEM1P*	rs536413401
Twin_A	3	168769557	T	C	Substitution	intron_variant	*EGFEM1P*	rs575522209
Twin_A	3	183534087	TTACTTTTAAAGTACTTTAAAAGTACTTTACTTTTAAAGTACTTTG	T	Deletion	intron_variant	*KLHL6*	rs1193720974
Twin_A	3	187366417	G	A	Substitution	upstream_gene_variant	*RTP4*	-
Twin_A	4	3011910	G	A	Substitution	intron_variant	*GRK4*	rs1056281673
Twin_A	4	5718308	T	A	Substitution	intron_variant	*EVC*	-
Twin_A	4	15955763	G	C	Substitution	downstream_gene_variant	*FGFBP2*	-
Twin_A	4	17001308	C	T	Substitution	intron_variant	*AC106894.1*	rs1437750527
Twin_A	4	17415320	G	A	Substitution	intergenic_variant	-	-

Twin_A	4	22103283	A	G	Substitution	intron_variant	*AC096719.1*	-
Twin_A	4	26343555	G	GT	Insertion	intron_variant	*RBPJ*	rs150586776
Twin_A	4	26935882	A	G	Substitution	intron_variant	*STIM2*	rs1392350220
Twin_A	4	31169536	C	A	Substitution	upstream_gene_variant	*LINC02497*	-
Twin_A	4	35852584	CAG	C	Deletion	intergenic_variant	-	rs150417183
Twin_A	4	38452605	G	A	Substitution	non_coding_transcript_ exon_variant	*LINC01258*	rs527412993
Twin_A	4	44413264	T	G	Substitution	intron_variant	*KCTD8*	-
Twin_A	4	52545084	CCTCT	C	Deletion	intron_variant	*AC097522.2*	rs748919064
Twin_A	4	54682436	C	CT	Insertion	intron_variant	*KIT*	rs368264394
Twin_A	4	55961258	A	G	Substitution	intron_variant	*CEP135*	-
Twin_A	4	61895466	AAAT	A	Deletion	upstream_gene_variant	*ADGRL3*	rs10603633
Twin_A	4	63566747	C	T	Substitution	intergenic_variant	-	-
Twin_A	4	66771456	G	C	Substitution	intergenic_variant	-	-
Twin_A	4	69630424	C	T	Substitution	intron_variant	*UGT2A1*	-
Twin_A	4	74248292	T	C	Substitution	intron_variant	*MTHFD2L*	-
Twin_A	4	74655308	C	T	Substitution	regulatory_region_variant	-	rs781899546
Twin_A	4	74668604	C	CT	Insertion	regulatory_region_variant	-	rs1418001328
Twin_A	4	76432023	G	T	Substitution	upstream_gene_variant	*SHROOM3*	-
Twin_A	4	82531664	A	G	Substitution	intron_variant	*TMEM150C*	-
Twin_A	4	105234449	TA	T	Deletion	frameshift_variant	*TET2*	-
Twin_A	4	106394908	AAT	A	Deletion	intergenic_variant	-	rs140334876
Twin_A	4	109876682	A	G	Substitution	downstream_gene_variant	*LRIT3*	-
Twin_A	4	110522189	C	CT	Insertion	intron_variant	*ENPEP*	rs539673220
Twin_A	4	111359412	C	T	Substitution	intergenic_variant	-	-
Twin_A	4	113822564	A	AT	Insertion	regulatory_region_variant	-	rs34730469
Twin_A	4	115289705	A	T	Substitution	intergenic_variant	-	rs201585293
Twin_A	4	119833655	A	T	Substitution	intergenic_variant	-	-
Twin_A	4	122600220	C	T	Substitution	intergenic_variant	-	-
Twin_A	4	124416683	A	AT	Insertion	intergenic_variant	-	rs112541914
Twin_A	4	134850961	G	A	Substitution	intergenic_variant	-	-
Twin_A	4	145648023	AT	A	Deletion	upstream_gene_variant	*MMAA*	rs771723312
Twin_A	4	160621868	A	G	Substitution	intergenic_variant	-	-
Twin_A	4	161252551	T	C	Substitution	regulatory_region_variant	-	-
Twin_A	4	161288198	G	GA	Insertion	intergenic_variant	-	rs969053461
Twin_A	4	162816440	T	C	Substitution	intron_variant	*AC021134.1*	-
Twin_A	4	180265089	G	A	Substitution	intergenic_variant	-	-
Twin_A	4	180667100	G	A	Substitution	regulatory_region_variant	-	rs563546983
Twin_A	4	182387149	G	A	Substitution	intron_variant	*TENM3*	rs983423020
Twin_A	4	183453115	T	C	Substitution	downstream_gene_variant	*CDKN2AIP*	-
Twin_A	4	185461063	T	C	Substitution	missense_variant	*CCDC110*	rs773624975
Twin_A	4	186034311	C	T	Substitution	intergenic_variant	-	rs199807699
Twin_A	4	187661068	G	T	Substitution	downstream_gene_variant	*AC097521.1*	-
Twin_A	4	188628950	T	TA	Insertion	intron_variant	*LINC01060*	-
Twin_A	4	105259633 *	G	A	Substitution	missense_variant	*TET2*	rs1456238164
Twin_A	5	2114089	AT	A	Deletion	intron_variant	*AC124852.1*	rs374364105
Twin_A	5	4579966	C	T	Substitution	intron_variant	*AC106799.2*	-

Twin_A	5	6643035	G	GT	Insertion	intron_variant	*SRD5A1*	rs35377021
Twin_A	5	7817145	G	GA	Insertion	intron_variant	*ADCY2*	rs57771304
Twin_A	5	7987372	C	T	Substitution	intergenic_variant	-	rs1320304875
Twin_A	5	10921181	G	GT	Insertion	intergenic_variant	-	rs60156777
Twin_A	5	10937820	A	G	Substitution	intergenic_variant	-	rs561589454
Twin_A	5	12387040	G	A	Substitution	intergenic_variant	-	-
Twin_A	5	18474168	C	T	Substitution	intergenic_variant	-	-
Twin_A	5	24399357	G	A	Substitution	intergenic_variant	-	rs868374455
Twin_A	5	33779372	G	A	Substitution	intron_variant	*ADAMTS12*	-
Twin_A	5	35831884	C	T	Substitution	upstream_gene_variant	*AC137810.1*	-
Twin_A	5	36636178	G	T	Substitution	intron_variant	*SLC1A3*	-
Twin_A	5	41753270	TAC	T	Deletion	intron_variant	*OXCT1*	rs36095905
Twin_A	5	52121218	T	G	Substitution	intergenic_variant	-	rs111904173
Twin_A	5	56387833	C	T	Substitution	intergenic_variant	-	-
Twin_A	5	58833324	TCA	T	Deletion	intron_variant	*RAB3C*	rs571940681
Twin_A	5	58833331	C	T	Substitution	intron_variant	*RAB3C*	rs199895867
Twin_A	5	63548792	G	C	Substitution	intergenic_variant	-	-
Twin_A	5	69055388	A	T	Substitution	intergenic_variant	-	-
Twin_A	5	73740730	C	T	Substitution	intron_variant	*ARHGEF28*	rs934596518
Twin_A	5	75285945	G	T	Substitution	intergenic_variant	-	-
Twin_A	5	78363812	A	AT	Insertion	upstream_gene_variant	*SCAMP1*	rs566623870
Twin_A	5	79269736	A	AT	Insertion	intron_variant	*JMY*	rs905517032
Twin_A	5	80583714	T	TGC	Insertion	intergenic_variant	-	rs113455475
Twin_A	5	81377269	G	A	Substitution	intron_variant	*ACOT12*	-
Twin_A	5	82451636	A	G	Substitution	intergenic_variant	-	-
Twin_A	5	85688616	C	T	Substitution	intergenic_variant	-	-
Twin_A	5	86606508	G	GT	Insertion	intergenic_variant	-	rs1353717748
Twin_A	5	91742536	G	C	Substitution	intron_variant	*AC093298.2*	-
Twin_A	5	93374643	C	T	Substitution	intron_variant	*NR2F1*	rs529601636
Twin_A	5	95231337	GAT	G	Deletion	intron_variant	*MCTP1*	rs138755290
Twin_A	5	95701747	A	G	Substitution	upstream_gene_variant	*SPATA9*	-
Twin_A	5	97765568	G	A	Substitution	intergenic_variant	-	-
Twin_A	5	100034257	G	A	Substitution	intergenic_variant	-	-
Twin_A	5	100071988	T	A	Substitution	upstream_gene_variant	*AC092278.1*	rs78484001
Twin_A	5	100155088	G	GT	Insertion	downstream_gene_variant	*RNU6*	rs5869895
Twin_A	5	100295778	C	T	Substitution	intergenic_variant	-	-
Twin_A	5	105654988	C	T	Substitution	intergenic_variant	-	rs1173103088
Twin_A	5	105941453	T	G	Substitution	intergenic_variant	-	-
Twin_A	5	111449354	A	G	Substitution	intron_variant	*CAMK4*	-
Twin_A	5	115666098	C	T	Substitution	regulatory_region_variant	-	-
Twin_A	5	116208129	T	C	Substitution	intron_variant	*COMMD10*	-
Twin_A	5	116797618	T	TA	Insertion	intergenic_variant	-	rs370022009
Twin_A	5	123223704	C	A	Substitution	intergenic_variant	-	-
Twin_A	5	130388901	G	A	Substitution	upstream_gene_variant	*RNU7*	rs1308696583
Twin_A	5	131362688	C	T	Substitution	intron_variant	*CDC42SE2*	-
Twin_A	5	131622186	T	TA	Insertion	intron_variant	*RAPGEF6*	rs368148920

Twin_A	5	142866780	T	G	Substitution	intron_variant	*ARHGAP26*	-
Twin_A	5	147611565	T	C	Substitution	intron_variant	*JAKMIP2*	-
Twin_A	5	152460283	G	A	Substitution	intron_variant	*AC008571.2*	-
Twin_A	5	155157242	C	T	Substitution	intron_variant	*AC010476.2*	rs950323196
Twin_A	5	162912266	C	A	Substitution	intron_variant	*AC113414.1*	-
Twin_A	5	163087087	G	GA	Insertion	intron_variant	*AC113414.1*	rs979728237
Twin_A	5	166031221	GTAGA	G	Deletion	upstream_gene_variant	*RPL7P20*	-
Twin_A	6	7238595	C	CT	Insertion	intron_variant	*RREB1*	rs796560101
Twin_A	6	8925748	C	T	Substitution	intron_variant	*HULC*	-
Twin_A	6	9330828	G	A	Substitution	intergenic_variant	-	-
Twin_A	6	15203913	A	AT	Insertion	intergenic_variant	-	rs1241556169
Twin_A	6	16644979	C	T	Substitution	intron_variant	*ATXN1*	-
Twin_A	6	19952193	G	A	Substitution	intergenic_variant	-	rs779233914
Twin_A	6	31570888	C	T	Substitution	upstream_gene_variant	*LTA*	rs956044849
Twin_A	6	36883002	G	T	Substitution	upstream_gene_variant	*C6orf89*	-
Twin_A	6	40639879	ACT	A	Deletion	intergenic_variant	-	rs35021402
Twin_A	6	40934220	G	A	Substitution	intergenic_variant	-	-
Twin_A	6	46776061	TA	T	Deletion	intergenic_variant	-	-
Twin_A	6	46995372	G	GT	Insertion	downstream_gene_variant	*ADGRF1*	rs142966227
Twin_A	6	58029106	G	A	Substitution	intron_variant	*AL445250.1*	rs1539609
Twin_A	6	60497665	T	C	Substitution	intron_variant	*AC244258.1*	-
Twin_A	6	62533854	G	A	Substitution	intergenic_variant	-	rs1050425018
Twin_A	6	70161977	C	A	Substitution	intron_variant	*COL19A1*	rs754784425
Twin_A	6	71671466	A	T	Substitution	intergenic_variant	-	-
Twin_A	6	80070589	C	T	Substitution	upstream_gene_variant	*AL591135.1*	rs113123867
Twin_A	6	90318383	C	CT	Insertion	upstream_gene_variant	*AL132996.1*	rs1230991775
Twin_A	6	93048770	T	TA	Insertion	regulatory_region_variant	-	rs1214397059
Twin_A	6	93696846	GTC	G	Deletion	intergenic_variant	-	rs142217537
Twin_A	6	95778674	T	A	Substitution	intergenic_variant	-	-
Twin_A	6	97233067	G	GA	Insertion	intron_variant	*MMS22L*	rs573093148
Twin_A	6	103570525	T	TTA	Insertion	intergenic_variant	-	rs533182054
Twin_A	6	111252036	C	T	Substitution	intron_variant	*AL360227.1*	-
Twin_A	6	117652832	G	A	Substitution	intergenic_variant	-	-
Twin_A	6	119799146	C	CT	Insertion	intergenic_variant	-	rs1218766914
Twin_A	6	119917789	C	G	Substitution	intergenic_variant	-	-
Twin_A	6	123368975	C	G	Substitution	intron_variant	*TRDN*	-
Twin_A	6	124380080	A	T	Substitution	intron_variant	*NKAIN2*	-
Twin_A	6	126389883	T	C	Substitution	intron_variant	*AL356534.1*	-
Twin_A	6	131059163	G	A	Substitution	intron_variant	*EPB41L2*	-
Twin_A	6	131399570	C	G	Substitution	intergenic_variant	-	-
Twin_A	6	134069108	C	T	Substitution	intergenic_variant	-	rs187112667
Twin_A	6	140986451	A	C	Substitution	intergenic_variant	-	-
Twin_A	6	145547033	T	G	Substitution	intron_variant	*EPM2A*	-
Twin_A	6	159660393	C	T	Substitution	intron_variant	*AL078604.4*	rs769079326
Twin_A	6	159881330	C	CAT	Insertion	intergenic_variant	-	rs10672852
Twin_A	6	166093829	G	T	Substitution	upstream_gene_variant	*AL627443.3*	-

Twin_A	6	166187121	G	GA	Insertion	intergenic_variant	-	-
Twin_A	7	4979367	T	C	Substitution	intron_variant	*RNF216P1*	rs868499198
Twin_A	7	5000741	A	G	Substitution	downstream_gene_variant	*RNF216P1*	rs1338617888
Twin_A	7	7056369	G	A	Substitution	intron_variant	*AC079804.3*	-
Twin_A	7	9227657	G	A	Substitution	regulatory_region_variant	-	rs999966239
Twin_A	7	21817695	A	AT	Insertion	intron_variant	*DNAH11*	rs201247500
Twin_A	7	30831221	C	T	Substitution	intron_variant	*MINDY4*	-
Twin_A	7	36071829	A	G	Substitution	intergenic_variant	-	rs757946495
Twin_A	7	38945518	G	GA	Insertion	intergenic_variant	-	rs796386392
Twin_A	7	41732610	A	G	Substitution	intron_variant	*INHBA*	-
Twin_A	7	43378810	C	CA	Insertion	intron_variant	*HECW1*	rs202007736
Twin_A	7	44090851	C	T	Substitution	intergenic_variant	-	-
Twin_A	7	44510030	T	A	Substitution	downstream_gene_variant	*NPC1L1*	-
Twin_A	7	49269510	AT	A	Deletion	intergenic_variant	-	rs1265301921
Twin_A	7	49269512	A	C	Substitution	intergenic_variant	-	rs984581632
Twin_A	7	53068163	G	A	Substitution	intergenic_variant	-	rs935736922
Twin_A	7	53091502	C	T	Substitution	intergenic_variant	-	-
Twin_A	7	53364914	G	A	Substitution	upstream_gene_variant	*RNU1*	rs34714639
Twin_A	7	54587046	C	A	Substitution	intergenic_variant	-	-
Twin_A	7	62529040	C	T	Substitution	intergenic_variant	-	rs555323358
Twin_A	7	64713934	GTATT	G	Deletion	downstream_gene_variant	*ZNF107*	rs369110012
Twin_A	7	67857057	G	A	Substitution	intergenic_variant	-	rs1319269938
Twin_A	7	68076463	CAT	C	Deletion	intergenic_variant	-	rs139100940
Twin_A	7	69630257	A	G	Substitution	intron_variant	*AUTS2*	-
Twin_A	7	73627805	C	T	Substitution	upstream_gene_variant	*MLXIPL*	-
Twin_A	7	81046412	CT	C	Deletion	intergenic_variant	-	rs1381425510
Twin_A	7	88505688	G	A	Substitution	intron_variant	*AC002069.2*	-
Twin_A	7	89031211	GTA	G	Deletion	intron_variant	*ZNF804B*	-
Twin_A	7	90565456	G	GT	Insertion	intron_variant	*CDK14*	rs373870738
Twin_A	7	95205017	CCA	C	Deletion	intron_variant	*PPP1R9A*	rs143843788
Twin_A	7	95298520	G	C	Substitution	3_prime_UTR_variant	*PON1*	-
Twin_A	7	96055652	T	TA	Insertion	intron_variant	*DYNC1I1*	rs11374590
Twin_A	7	98561959	C	A	Substitution	intergenic_variant	-	-
Twin_A	7	102065090	G	GT	Insertion	intron_variant	*CUX1*	rs562493293
Twin_A	7	109185877	G	A	Substitution	intergenic_variant	-	rs1228545900
Twin_A	7	110615318	C	T	Substitution	intergenic_variant	-	-
Twin_A	7	122016458	C	A	Substitution	intron_variant	*PTPRZ1*	-
Twin_A	7	124142825	C	T	Substitution	intron_variant	*AC006148.1*	-
Twin_A	7	128429252	C	CT	Insertion	downstream_gene_variant	*AC010655.1*	rs56056515
Twin_A	7	129776973	A	G	Substitution	upstream_gene_variant	*MIR96*	rs1401666012
Twin_A	7	137296579	A	G	Substitution	intron_variant	*PTN*	-
Twin_A	7	139519359	T	TA	Insertion	upstream_gene_variant	*CLEC2L*	rs1220935636
Twin_A	7	149998259	A	G	Substitution	upstream_gene_variant	*AC092681.2*	rs2969815
Twin_A	7	153528906	T	C	Substitution	intergenic_variant	-	-
Twin_A	7	154611485	A	G	Substitution	intron_variant	*DPP6*	-
Twin_A	8	503037	C	T	Substitution	intron_variant	*TDRP*	rs1166917437

Twin_A	8	13876611	G	GA	Insertion	regulatory_region_variant	-	rs34968549
Twin_A	8	17691704	G	GT	Insertion	intron_variant	*MTUS1*	rs200014137
Twin_A	8	19550828	C	T	Substitution	intron_variant	*CSGALNACT1*	-
Twin_A	8	25892225	T	C	Substitution	intron_variant	*EBF2*	-
Twin_A	8	30978684	T	C	Substitution	downstream_gene_variant	*AC008066.1*	-
Twin_A	8	36379501	CAT	C	Deletion	intron_variant	*AC090809.1*	rs148082032
Twin_A	8	36570758	T	A	Substitution	intron_variant	*AC090809.1*	-
Twin_A	8	39656195	G	GA	Insertion	intron_variant	*ADAM18*	rs201971418
Twin_A	8	48680993	A	G	Substitution	intron_variant	*AC022915.2*	-
Twin_A	8	51711248	T	C	Substitution	intron_variant	*PXDNL*	COSV62492491
Twin_A	8	53511516	G	GA	Insertion	downstream_gene_variant	*AC022034.1*	rs535409810
Twin_A	8	61457436	G	A	Substitution	intron_variant	*CLVS1*	-
Twin_A	8	63967261	A	AT	Insertion	intron_variant	*LINC01414*	rs5891933
Twin_A	8	64128898	G	A	Substitution	intron_variant	*LINC01414*	rs545104948
Twin_A	8	65916408	G	A	Substitution	intergenic_variant	-	-
Twin_A	8	68253460	T	TA	Insertion	regulatory_region_variant	-	rs1000407340
Twin_A	8	72453042	C	T	Substitution	intergenic_variant	-	rs1051744174
Twin_A	8	77781391	C	CT	Insertion	intergenic_variant	-	rs1198479285
Twin_A	8	83358311	GGCTACAGACAAACCTTCAATGGA	G	Deletion	intergenic_variant	-	-
Twin_A	8	86318421	G	GA	Insertion	intron_variant	*SLC7A13*	rs34723665
Twin_A	8	88356215	C	A	Substitution	intron_variant	*AC090578.1*	-
Twin_A	8	95037819	G	GA	Insertion	intron_variant	*NDUFAF6*	rs59511473
Twin_A	8	104362862	A	G	Substitution	intron_variant	*DPYS*	rs555618444
Twin_A	8	108043942	A	T	Substitution	intron_variant	*RSPO2*	-
Twin_A	8	115536798	G	GT	Insertion	intron_variant	*TRPS1*	rs900943605
Twin_A	8	117615713	ATCT	A	Deletion	regulatory_region_variant	-	-
Twin_A	8	124975564	C	G	Substitution	intron_variant	*ZNF572*	-
Twin_A	8	129273738	G	GT	Insertion	intron_variant	*CCDC26*	rs891751618
Twin_A	8	129372882	C	CA	Insertion	intron_variant	*CCDC26*	rs560030895
Twin_A	8	129733477	G	A	Substitution	downstream_gene_variant	*MTRF1LP2*	-
Twin_A	8	132428552	CAG	C	Deletion	intron_variant	*KCNQ3*	rs201318711
Twin_A	8	134856654	C	G	Substitution	intron_variant	*AC103764.1*	rs1321197388
Twin_A	8	135261370	G	C	Substitution	intron_variant	*LINC01591*	-
Twin_A	8	137128263	C	T	Substitution	intergenic_variant	-	rs1053069745
Twin_A	8	138139847	C	T	Substitution	intron_variant	*FAM135B*	rs947110030
Twin_A	8	143051626	G	A	Substitution	intron_variant	*C8orf31*	-
Twin_A	9	50071	AT	A	Deletion	intron_variant	*WASHC1*	-
Twin_A	9	395970	T	G	Substitution	intron_variant	*DOCK8*	-
Twin_A	9	1822815	C	T	Substitution	intergenic_variant	-	-
Twin_A	9	4316936	C	T	Substitution	intron_variant	*GLIS3*	-
Twin_A	9	5368234	T	C	Substitution	intron_variant	*PLGRKT*	-
Twin_A	9	14104419	G	GT	Insertion	intron_variant	*NFIB*	rs945961609
Twin_A	9	17180226	GT	G	Deletion	intron_variant	*CNTLN*	rs1264390576
Twin_A	9	19348792	G	GA	Insertion	intron_variant	*DENND4C*	rs367658826
Twin_A	9	23699353	A	G	Substitution	intron_variant	*ELAVL2*	rs568285627

Twin_A	9	24066549	G	A	Substitution	upstream_gene_variant	*AL513317.1*	rs895109680
Twin_A	9	28150104	C	A	Substitution	intron_variant	*LINGO2*	-
Twin_A	9	31556284	G	A	Substitution	intergenic_variant	-	rs12347583
Twin_A	9	39080916	C	T	Substitution	intron_variant	*CNTNAP3*	rs369395719
Twin_A	9	41655135	C	T	Substitution	upstream_gene_variant	*AL591926.2*	rs1226839471
Twin_A	9	45155466	G	C	Substitution	intergenic_variant	-	rs1476645657
Twin_A	9	61566797	T	C	Substitution	intergenic_variant	-	-
Twin_A	9	73087756	A	G	Substitution	intergenic_variant	-	-
Twin_A	9	77004114	C	A	Substitution	intergenic_variant	-	-
Twin_A	9	81050746	G	A	Substitution	intergenic_variant	-	rs764741538
Twin_A	9	82371224	C	T	Substitution	intron_variant	*AL162726.3*	-
Twin_A	9	83484424	C	CT	Insertion	intron_variant	*FRMD3*	-
Twin_A	9	84858874	G	GA	Insertion	intron_variant	*NTRK2*	rs965433677
Twin_A	9	87748283	T	TA	Insertion	regulatory_region_variant	-	rs143725130
Twin_A	9	88768716	G	A	Substitution	intron_variant	*AL390791.1*	rs548341913
Twin_A	9	96382496	C	CTA	Insertion	intron_variant	*SLC35D2*	rs142809659
Twin_A	9	101191229	G	A	Substitution	intron_variant	*PLPPR1*	-
Twin_A	9	101784275	G	A	Substitution	intergenic_variant	-	-
Twin_A	9	102487319	C	T	Substitution	intergenic_variant	-	rs1189198082
Twin_A	9	111554985	G	T	Substitution	3_prime_UTR_variant	*ZNF483*	rs1039810650
Twin_A	9	118247951	T	C	Substitution	intergenic_variant	-	-
Twin_A	9	120700338	T	A	Substitution	intron_variant	*MEGF9*	-
Twin_A	9	122872477	T	C	Substitution	downstream_gene_variant	*RC3H2*	rs996687988
Twin_A	9	125539853	T	TA	Insertion	intron_variant	*MAPKAP1*	rs201724260
Twin_A	9	127663224	C	G	Substitution	missense_variant	*STXBP1*	-
Twin_A	9	131190221	A	T	Substitution	intron_variant	*NUP214*	-
Twin_A	9	134828811	CCA	C	Deletion	intron_variant	*COL5A1*	rs906680873
Twin_A	9	136468856	C	CT	Insertion	intron_variant	*SEC16A*	-
Twin_A	10	4183898	A	G	Substitution	regulatory_region_variant	-	-
Twin_A	10	9239651	C	CT	Insertion	intergenic_variant	-	rs963757968
Twin_A	10	10849842	T	A	Substitution	intron_variant	*CELF2*	rs1365457744
Twin_A	10	11168124	TTC	T	Deletion	intron_variant	*CELF2*	-
Twin_A	10	15299120	C	T	Substitution	intron_variant	*FAM171A1*	rs558867054
Twin_A	10	18976904	C	T	Substitution	intergenic_variant	-	rs181225799
Twin_A	10	23617158	C	CAT	Insertion	intergenic_variant	-	rs144728449
Twin_A	10	23958376	G	A	Substitution	intron_variant	*KIAA1217*	-
Twin_A	10	27864409	G	GA	Insertion	intron_variant	*ARMC4*	rs998440262
Twin_A	10	34892061	G	T	Substitution	intergenic_variant	-	-
Twin_A	10	47221434	C	T	Substitution	intergenic_variant	-	-
Twin_A	10	54174376	C	CA	Insertion	intron_variant	*PCDH15*	rs112450699
Twin_A	10	56140794	C	CAT	Insertion	intergenic_variant	-	rs999676086
Twin_A	10	56950504	A	T	Substitution	intergenic_variant	-	-
Twin_A	10	60147699	C	G	Substitution	intron_variant	*ANK3*	-
Twin_A	10	62705771	AAT	A	Deletion	intron_variant	*AC067751.1*	rs10573238
Twin_A	10	64210270	G	T	Substitution	intron_variant	*AC013287.1*	rs75029261
Twin_A	10	67571470	TA	T	Deletion	intron_variant	*CTNNA3*	-

Twin_A	10	70021310	G	GA	Insertion	upstream_gene_variant	*LINC02636*	rs944814046
Twin_A	10	75905930	G	GT	Insertion	intron_variant	*LRMDA*	rs5786198
Twin_A	10	78393818	G	A	Substitution	intron_variant	*AC012560.1*	-
Twin_A	10	83188578	G	A	Substitution	intergenic_variant	-	rs778523013
Twin_A	10	83492892	C	T	Substitution	intergenic_variant	-	-
Twin_A	10	93187731	G	A	Substitution	intergenic_variant	-	-
Twin_A	10	94576898	A	AT	Insertion	intron_variant	*HELLS*	rs540746312
Twin_A	10	94950682	A	G	Substitution	intron_variant	*CYP2C9*	rs907386662
Twin_A	10	103273122	G	A	Substitution	upstream_gene_variant	*INA*	rs1286474689
Twin_A	10	109091857	G	A	Substitution	intergenic_variant	-	rs954171718
Twin_A	10	111830308	A	G	Substitution	intron_variant	*AL136119.1*	rs1430607514
Twin_A	10	113814958	A	G	Substitution	intergenic_variant	-	-
Twin_A	10	115169017	CTT	C	Deletion	intron_variant	*ATRNL1*	rs71010011
Twin_A	10	115409272	G	A	Substitution	intron_variant	*ATRNL1*	-
Twin_A	10	116526803	T	C	Substitution	intergenic_variant	-	-
Twin_A	10	120224348	T	TA	Insertion	intergenic_variant	-	-
Twin_A	10	121613197	T	C	Substitution	downstream_gene_variant	*AC009988.1*	-
Twin_A	10	122518538	G	A	Substitution	downstream_gene_variant	*HTRA1*	rs568255913
Twin_A	10	124140841	T	C	Substitution	intergenic_variant	-	rs1455378017
Twin_A	10	124588678	T	G	Substitution	intron_variant	*LHPP*	-
Twin_A	11	6802600	C	T	Substitution	upstream_gene_variant	*OR6A2*	rs745932107
Twin_A	11	12635917	G	A	Substitution	intron_variant	*AC107881.1*	rs1470450176
Twin_A	11	17634218	C	T	Substitution	stop_gained	*OTOG*	rs866476223
Twin_A	11	18214790	G	A	Substitution	downstream_gene_variant	*SLC25A51P4*	-
Twin_A	11	25629284	G	T	Substitution	upstream_gene_variant	*AC100770.1*	-
Twin_A	11	27431935	C	T	Substitution	intron_variant	*LGR4*	rs187542892
Twin_A	11	29268317	G	GTA	Insertion	downstream_gene_variant	*AC090791.1*	rs57529313
Twin_A	11	30414572	A	AT	Insertion	intron_variant	*MPPED2*	rs574457624
Twin_A	11	34947336	TAC	T	Deletion	intron_variant	*PDHX*	rs201136678
Twin_A	11	40970112	A	T	Substitution	intron_variant	*LRRC4C*	-
Twin_A	11	42181489	A	AT	Insertion	downstream_gene_variant	*LINC02740*	rs1268779757
Twin_A	11	42992114	T	A	Substitution	intergenic_variant	-	-
Twin_A	11	49443145	TATC	T	Deletion	intergenic_variant	-	rs1327126227
Twin_A	11	59547928	A	G	Substitution	intron_variant	*AP003778.1*	-
Twin_A	11	62907973	C	T	Substitution	downstream_gene_variant	*M1*	-
Twin_A	11	69424352	G	A	Substitution	upstream_gene_variant	*AP000439.2*	-
Twin_A	11	70172132	A	G	Substitution	intron_variant	*ANO1*	rs11235475
Twin_A	11	71917009	GTTA	G	Deletion	downstream_gene_variant	*AP002495.1*	rs550657033
Twin_A	11	72630002	T	A	Substitution	intron_variant	*PDE2A*	rs956005801
Twin_A	11	78991748	C	T	Substitution	intron_variant	*TENM4*	-
Twin_A	11	83859358	A	G	Substitution	intron_variant	*DLG2*	-
Twin_A	11	83984865	G	A	Substitution	intron_variant	*DLG2*	rs2514172
Twin_A	11	87084313	C	CT	Insertion	intron_variant	*TMEM135*	rs5793234
Twin_A	11	89364256	TTC	T	Deletion	intron_variant	*NOX4*	-
Twin_A	11	91531418	GTC	G	Deletion	intergenic_variant	-	rs57259378
Twin_A	11	92625315	G	A	Substitution	intron_variant	*FAT3*	rs1044839433

Twin_A	11	93542010	C	T	Substitution	intron_variant	*SMCO4*	rs1238441458
Twin_A	11	93572399	A	T	Substitution	regulatory_region_variant	-	-
Twin_A	11	96395662	G	GA	Insertion	downstream_gene_variant	*JRKL*	rs1197147759
Twin_A	11	98979001	G	GCA	Insertion	intergenic_variant	-	rs5793962
Twin_A	11	99123208	G	GT	Insertion	downstream_gene_variant	*RN7SKP53*	rs370959291
Twin_A	11	105917523	A	G	Substitution	intron_variant	*GRIA4*	-
Twin_A	11	112108171	G	T	Substitution	intron_variant	*AP002884.2*	-
Twin_A	11	118973448	G	T	Substitution	intron_variant	*FOXR1*	-
Twin_A	11	121731425	T	C	Substitution	intron_variant	*AP001977.1*	-
Twin_A	11	130155838	CA	C	Deletion	upstream_gene_variant	*ST14*	rs71061368
Twin_A	11	133629356	A	T	Substitution	intergenic_variant	-	-
Twin_A	12	8992708	A	AT	Insertion	intron_variant	*KLRG1*	rs1011939413
Twin_A	12	14002642	GTA	G	Deletion	intergenic_variant	-	rs767637242
Twin_A	12	27135862	C	A	Substitution	intron_variant	*AC092747.1*	-
Twin_A	12	33793906	A	G	Substitution	intergenic_variant	-	rs1160899620
Twin_A	12	37544249	GGT	G	Deletion	intergenic_variant	-	rs139856566
Twin_A	12	39556427	T	A	Substitution	intron_variant	*ABCD2*	-
Twin_A	12	45801804	CAG	C	Deletion	intron_variant	*ARID2*	rs200185178
Twin_A	12	47093368	A	AT	Insertion	intron_variant	*PCED1B*	rs142397807
Twin_A	12	54193118	A	G	Substitution	upstream_gene_variant	*SMUG1*	-
Twin_A	12	55987166	G	GT	Insertion	intron_variant	*RAB5B*	rs34223637
Twin_A	12	61094238	T	C	Substitution	intergenic_variant	-	rs795318
Twin_A	12	61096059	C	T	Substitution	intergenic_variant	-	-
Twin_A	12	63499491	A	T	Substitution	intergenic_variant	-	rs367633111
Twin_A	12	67241412	G	GA	Insertion	intergenic_variant	-	rs570864048
Twin_A	12	80465550	CA	C	Deletion	intron_variant	*PTPRQ*	rs1565723030
Twin_A	12	96021004	T	A	Substitution	intron_variant	*LTA4H*	-
Twin_A	12	99195605	T	C	Substitution	intron_variant	*ANKS1B*	-
Twin_A	12	99570645	C	CA	Insertion	intron_variant	*ANKS1B*	rs34453262
Twin_A	12	100385137	A	T	Substitution	intron_variant	*SLC17A8*	rs564312355
Twin_A	12	101478301	G	A	Substitution	intron_variant	*SPIC*	-
Twin_A	12	111703138	A	G	Substitution	intron_variant	*ACAD10*	-
Twin_A	12	112201473	G	GA	Insertion	intron_variant	*HECTD4*	rs59104653
Twin_A	12	117866839	T	TA	Insertion	intron_variant	*KSR2*	rs34463256
Twin_A	12	120238552	T	A	Substitution	intron_variant	*PXN*	-
Twin_A	12	129194036	A	G	Substitution	intron_variant	*TMEM132D*	-
Twin_A	12	129699976	C	T	Substitution	missense_variant	*TMEM132D*	COSV70606773
Twin_A	12	130201343	C	T	Substitution	intergenic_variant	-	rs752475595
Twin_A	12	131372998	C	T	Substitution	downstream_gene_variant	*AC140118.1*	rs1044089851
Twin_A	12	131449745	T	TTATA TATAT ATA	Insertion	intron_variant	*AC073578.2*	rs60625980
Twin_A	12	131913217	G	T	Substitution	synonymous_variant	*ULK1*	COSV58854885
Twin_A	12	132658754	CG	C	Deletion	intron_variant	*POLE*	-
Twin_A	13	16312671	T	G	Substitution	intergenic_variant	-	rs1452678559
Twin_A	13	19458900	T	C	Substitution	intron_variant	*TPTE2*	-
Twin_A	13	21198801	T	A	Substitution	regulatory_region_variant	-	-
Twin_A	13	21198803	G	T	Substitution	regulatory_region_variant	-	-

Twin_A	13	22486341	G	A	Substitution	intergenic_variant	-	-
Twin_A	13	24407230	C	T	Substitution	upstream_gene_variant	*TPTE2P6*	-
Twin_A	13	27747684	T	C	Substitution	downstream_gene_variant	*POLR1D*	rs530116991
Twin_A	13	33789922	C	T	Substitution	regulatory_region_variant	-	-
Twin_A	13	35458160	C	G	Substitution	intron_variant	*NBEA*	-
Twin_A	13	40248330	G	A	Substitution	intron_variant	*LINC00598*	rs570712758
Twin_A	13	44807574	G	A	Substitution	intron_variant	*LINC00330*	-
Twin_A	13	51185115	A	T	Substitution	intron_variant	*C13orf42*	rs982603544
Twin_A	13	54356322	T	A	Substitution	regulatory_region_variant	-	-
Twin_A	13	57950093	C	T	Substitution	intergenic_variant	-	-
Twin_A	13	60127004	G	A	Substitution	intron_variant	*DIAPH3*	-
Twin_A	13	61979712	AT	A	Deletion	intergenic_variant	-	-
Twin_A	13	86387551	T	TA	Insertion	intergenic_variant	-	rs578195336
Twin_A	13	95408931	G	T	Substitution	intergenic_variant	-	-
Twin_A	13	97673189	G	A	Substitution	downstream_gene_variant	*AL359502.1*	rs1006927199
Twin_A	13	100032713	C	CA	Insertion	upstream_gene_variant	*NDUFA12P1*	rs914603334
Twin_A	13	110038016	G	GCA	Insertion	intron_variant	*AL390755.1*	rs34470615
Twin_A	13	112177050	A	T	Substitution	intergenic_variant	-	-
Twin_A	13	112393758	C	T	Substitution	intron_variant	*SPACA7*	-
Twin_A	14	27812525	AAG	A	Deletion	intron_variant	*AL390334.1*	rs368944583
Twin_A	14	27980040	G	A	Substitution	intergenic_variant	-	-
Twin_A	14	29729980	G	A	Substitution	intron_variant	*PRKD1*	rs536812458
Twin_A	14	30110595	C	CT	Insertion	intron_variant	*PRKD1*	rs1412349193
Twin_A	14	37141508	T	TA	Insertion	intron_variant	*SLC25A21*	rs533050099
Twin_A	14	39514017	ATATC	A	Deletion	downstream_gene_variant	*AL049828.1*	rs78432958
Twin_A	14	58577605	T	A	Substitution	intergenic_variant	-	-
Twin_A	14	59907903	T	C	Substitution	intergenic_variant	-	-
Twin_A	14	59942776	C	CT	Insertion	intron_variant	*LRRC9*	rs1440781193
Twin_A	14	61869166	A	G	Substitution	intron_variant	*SYT16*	-
Twin_A	14	81424362	TG	T	Deletion	intron_variant	*STON2*	-
Twin_A	14	87211954	G	GA	Insertion	regulatory_region_variant	-	-
Twin_A	15	17735761	T	G	Substitution	intergenic_variant	-	rs1425932497
Twin_A	15	24574554	C	A	Substitution	non_coding_transcript_ exon_variant	*PWRN1*	-
Twin_A	15	25170693	A	G	Substitution	upstream_gene_variant	*SNORD115*	-
Twin_A	15	27699074	C	T	Substitution	intergenic_variant	-	rs979345355
Twin_A	15	27798227	G	A	Substitution	intron_variant	*OCA2*	rs575593376
Twin_A	15	56330772	G	GA	Insertion	intron_variant	*TEX9*	rs202192979
Twin_A	15	58918915	C	CT	Insertion	intron_variant	*SLTM*	rs566389959
Twin_A	15	60956234	G	GT	Insertion	intron_variant	*RORA*	rs796500717
Twin_A	15	63420915	CAT	C	Deletion	intron_variant	*LINC02568*	rs142303229
Twin_A	15	69253960	G	A	Substitution	intron_variant	*GLCE*	-
Twin_A	15	80950475	T	A	Substitution	3_prime_UTR_variant	*CEMIP*	-
Twin_A	15	83158851	C	CT	Insertion	intron_variant	*HDGFL3*	rs554737989
Twin_A	15	96972773	A	G	Substitution	intergenic_variant	-	-
Twin_A	15	99025033	T	A	Substitution	intron_variant	*AC036108.1*	-
Twin_A	16	3192939	A	AT	Insertion	intron_variant	*AJ003147.2*	-

Twin_A	16	7113874	T	C	Substitution	intron_variant	*RBFOX1*	-
Twin_A	16	7450654	G	A	Substitution	intron_variant	*RBFOX1*	rs146577787
Twin_A	16	7806761	G	A	Substitution	intergenic_variant	-	rs752635066
Twin_A	16	8247620	G	A	Substitution	intergenic_variant	-	rs942245726
Twin_A	16	8372617	C	T	Substitution	intron_variant	*AC018767.1*	rs909464601
Twin_A	16	11338624	C	CT	Insertion	intron_variant	*RMI2*	rs377660326
Twin_A	16	13671261	G	A	Substitution	intergenic_variant	-	rs544425045
Twin_A	16	22306291	C	T	Substitution	intron_variant	*POLR3E*	-
Twin_A	16	25681378	C	T	Substitution	intergenic_variant	-	-
Twin_A	16	34085953	C	T	Substitution	regulatory_region_variant	-	rs759751179
Twin_A	16	35232723	C	A	Substitution	intron_variant	*AC023824.3*	-
Twin_A	16	35736091	G	A	Substitution	downstream_gene_variant	*RNA5SP407*	rs1168416095
Twin_A	16	47099927	C	T	Substitution	intron_variant	*NETO2*	-
Twin_A	16	47823047	C	A	Substitution	intron_variant	*LINC02133*	-
Twin_A	16	48066801	G	T	Substitution	intergenic_variant	-	-
Twin_A	16	49303993	C	T	Substitution	intron_variant	*AC007614.1*	-
Twin_A	16	49413774	A	C	Substitution	regulatory_region_variant	-	-
Twin_A	16	51823583	C	T	Substitution	intergenic_variant	-	-
Twin_A	16	52276870	G	A	Substitution	intron_variant	*CASC22*	-
Twin_A	16	52638073	C	T	Substitution	intron_variant	*CASC16*	rs995891657
Twin_A	16	61886296	T	TA	Insertion	intron_variant	*CDH8*	rs535315018
Twin_A	16	64432341	G	GT	Insertion	intron_variant	*AC092131.1*	rs901796829
Twin_A	16	64614439	G	A	Substitution	intergenic_variant	-	-
Twin_A	16	64919655	T	A	Substitution	intergenic_variant	-	-
Twin_A	16	66602481	T	TA	Insertion	upstream_gene_variant	*CMTM3*	rs145668319
Twin_A	16	67608941	A	G	Substitution	intron_variant	*CTCF*	rs892105462
Twin_A	16	71030715	G	C	Substitution	intron_variant	*HYDIN*	rs12102892
Twin_A	16	72362589	G	T	Substitution	intron_variant	*LINC01572*	-
Twin_A	16	74358477	G	GTGTG TGTGT A	Insertion	intron_variant	*AC009053.1*	rs1355695440
Twin_A	16	77781037	CAGG	C	Deletion	regulatory_region_variant	-	rs145741815
Twin_A	16	83070380	C	T	Substitution	intron_variant	*CDH13*	-
Twin_A	16	85313584	C	A	Substitution	intron_variant	*GSE1*	rs1483404740
Twin_A	16	86284090	G	A	Substitution	intron_variant	*LINC01081*	-
Twin_A	17	578020	G	C	Substitution	intron_variant	*VPS53*	-
Twin_A	17	2793153	AT	A	Deletion	upstream_gene_variant	*RAP1GAP2*	rs200252910
Twin_A	17	8040550	GA	G	Deletion	intron_variant	*ALOX15B*	rs57248868
Twin_A	17	8080028	C	T	Substitution	non_coding_transcript_ exon_variant	*AC129492.1*	-
Twin_A	17	9019997	A	G	Substitution	upstream_gene_variant	*NTN1*	-
Twin_A	17	9546387	TA	T	Deletion	intron_variant	*STX8*	rs56901611
Twin_A	17	14728731	G	A	Substitution	intron_variant	*AC013248.1*	-
Twin_A	17	15719138	C	CA	Insertion	3_prime_UTR_variant	*ZNF286A*	rs372793188
Twin_A	17	32918315	T	G	Substitution	regulatory_region_variant	-	-
Twin_A	17	34590397	T	C	Substitution	intron_variant	*TMEM132E*	rs1390883369
Twin_A	17	34849723	CACA	C	Deletion	intron_variant	*AC022903.1*	rs952016458
Twin_A	17	41287213	A	G	Substitution	regulatory_region_variant	-	-
Twin_A	17	48780087	G	GA	Insertion	intron_variant	*TTLL6*	rs371091415

Twin_A	17	63184014	C	CA	Insertion	intron_variant	*TANC2*	rs570141267
Twin_A	17	64353316	C	T	Substitution	intron_variant	*PECAM1*	rs879198491
Twin_A	17	66569297	G	A	Substitution	intron_variant	*PRKCA*	rs540966408
Twin_A	17	72065475	G	A	Substitution	intron_variant	*AC007461.1*	-
Twin_A	17	72109757	C	T	Substitution	intron_variant, non_coding_transcript_variant	*SOX9-AS1*	-
Twin_A	17	73729002	C	G	Substitution	intergenic_variant	-	-
Twin_A	17	78164605	G	A	Substitution	upstream_gene_variant	*SYNGR2*	-
Twin_A	17	79118078	G	A	Substitution	intron_variant	*RBFOX3*	rs560513487
Twin_A	17	82853402	C	CT	Insertion	intron_variant	*TBCD*	-
Twin_A	18	9601220	A	G	Substitution	intron_variant	*PPP4R1*	rs982592192
Twin_A	18	14227264	CAA	C	Deletion	intron_variant	*ANKRD20A5P*	rs200661335
Twin_A	18	20189631	G	A	Substitution	intergenic_variant	-	rs1337042912
Twin_A	18	20830406	G	A	Substitution	intergenic_variant	-	-
Twin_A	18	27867692	T	C	Substitution	intergenic_variant	-	-
Twin_A	18	29896712	C	T	Substitution	intergenic_variant	-	-
Twin_A	18	32161626	G	T	Substitution	intron_variant	*GAREM1*	-
Twin_A	18	32191812	T	C	Substitution	intron_variant	*MEP1B*	-
Twin_A	18	33837694	C	A	Substitution	regulatory_region_variant	-	-
Twin_A	18	34514791	G	A	Substitution	intron_variant	*DTNA*	-
Twin_A	18	41122265	A	G	Substitution	intergenic_variant	-	-
Twin_A	18	50652311	G	A	Substitution	intron_variant	*MAPK4*	-
Twin_A	18	51531692	G	A	Substitution	intron_variant	*LINC01630*	-
Twin_A	18	52758063	C	T	Substitution	intron_variant	*DCC*	rs75634824
Twin_A	18	53231840	T	C	Substitution	intron_variant	*DCC*	rs1236652956
Twin_A	18	56463814	T	TA	Insertion	intergenic_variant	-	rs139733893
Twin_A	18	57470379	T	C	Substitution	intron_variant	*ONECUT2*	-
Twin_A	18	63454630	A	AT	Insertion	regulatory_region_variant	-	rs900204433
Twin_A	18	68909930	A	T	Substitution	intron_variant	*CCDC102B*	-
Twin_A	18	70340966	C	G	Substitution	upstream_gene_variant	*LIVAR*	-
Twin_A	18	75784106	G	A	Substitution	intergenic_variant	-	-
Twin_A	18	76850613	G	GT	Insertion	intron_variant	*ZNF236*	-
Twin_A	19	3661315	C	A	Substitution	intron_variant	*PIP5K1C*	-
Twin_A	19	7346680	T	C	Substitution	upstream_gene_variant	*AC119396.1*	-
Twin_A	19	10065329	CAA	C	Deletion	intron_variant	*C3P1*	-
Twin_A	19	12943750	AGCAGAGGCTTAAGGAGGAGGAAGAAGACAAGAAACGCAAAGAGGAGGAG	A	Deletion	frameshift_variant	*CALR*	rs1555760738
Twin_A	19	14455965	T	TA	Insertion	intron_variant	*PKN1*	rs540413655
Twin_A	19	15471572	G	C	Substitution	intron_variant	*PGLYRP2*	-
Twin_A	19	16234332	C	T	Substitution	intron_variant	*AP1M1*	rs1266106157
Twin_A	19	22883084	C	T	Substitution	regulatory_region_variant	-	-
Twin_A	19	22892036	CT	C	Deletion	intergenic_variant	-	rs59650600
Twin_A	19	27404653	C	G	Substitution	intergenic_variant	-	-
Twin_A	19	27753864	G	T	Substitution	downstream_gene_variant	*AC006504.1*	-
Twin_A	19	28326785	C	T	Substitution	intron_variant	*AC005580.1*	rs1436772262

Twin_A	19	28434046	C	T	Substitution	upstream_gene_variant	*AC005307.1*	rs1029855214
Twin_A	19	40522348	A	T	Substitution	intron_variant	*SPTBN4*	rs1434639441
Twin_A	19	44493750	T	TCTCTC TTG	Insertion	intron_variant	*ZNF180*	rs1258057463
Twin_A	19	48040435	C	A	Substitution	intron_variant	*CABP5*	-
Twin_A	19	54956200	CAAGG	C	Deletion	intron_variant	*NLRP7*	rs752034667
Twin_A	19	55462076	G	GGA	Insertion	upstream_gene_variant	*ISOC2*	-
Twin_A	20	4322280	G	GT	Insertion	regulatory_region_variant	-	rs921737367
Twin_A	20	13915926	C	G	Substitution	intron_variant	*SEL1L2*	rs1392345509
Twin_A	20	24530811	G	GT	Insertion	intron_variant	*SYNDIG1*	rs374193810
Twin_A	20	28628245	A	G	Substitution	intergenic_variant	-	rs879967418
Twin_A	20	39766967	C	T	Substitution	intron_variant	*AL118523.1*	rs1468797562
Twin_A	20	42620323	G	C	Substitution	intron_variant	*PTPRT*	-
Twin_A	20	44696207	C	A	Substitution	upstream_gene_variant	*AL139352.1*	-
Twin_A	20	45737606	A	T	Substitution	intergenic_variant	-	-
Twin_A	20	51168664	C	A	Substitution	intergenic_variant	-	-
Twin_A	20	53076039	C	T	Substitution	intron_variant	*TSHZ2*	rs559693937
Twin_A	20	55145960	T	C	Substitution	intergenic_variant	-	-
Twin_A	20	55476594	G	A	Substitution	intron_variant	*LINC01440*	rs910371781
Twin_A	20	57038323	C	T	Substitution	intergenic_variant	-	-
Twin_A	21	10536357	C	T	Substitution	intron_variant	*TPTE*	-
Twin_A	21	18807398	GCA	G	Deletion	intergenic_variant	-	rs56088107
Twin_A	21	24361606	G	GT	Insertion	intron_variant	*LINC01684*	rs564109768
Twin_A	21	25056551	G	A	Substitution	intron_variant	*AP000235.1*	rs529745996
Twin_A	21	25235168	GTTT	G	Deletion	intron_variant	*AP001341.1*	-
Twin_A	21	26121394	C	CT	Insertion	intron_variant	*APP*	rs147930861
Twin_A	21	27842645	T	C	Substitution	intergenic_variant	-	-
Twin_A	21	28932587	T	C	Substitution	synonymous_variant	*LTN1*	-
Twin_A	21	31309695	A	G	Substitution	intron_variant	*TIAM1*	-
Twin_A	21	32999596	A	T	Substitution	intron_variant	*AP000282.1*	rs1265368759
Twin_A	21	35313715	G	A	Substitution	intron_variant	*RUNX1*	rs1050527217
Twin_A	21	35666262	C	T	Substitution	intron_variant	*RUNX1*	-
Twin_A	21	37443516	T	G	Substitution	intron_variant	*DYRK1A*	rs575428175
Twin_A	21	41859981	G	T	Substitution	intron_variant	*PRDM15*	-
Twin_A	21	42078747	A	G	Substitution	intron_variant	*UMODL1*	-
Twin_A	22	12521506	A	T	Substitution	intergenic_variant	-	rs1476432335
Twin_A	22	15748369	C	T	Substitution	downstream_gene_variant	*GRAMD4P2*	rs529661114
Twin_A	22	16418819	G	A	Substitution	intergenic_variant	-	rs73387994
Twin_A	22	20582589	G	A	Substitution	intron_variant	*MED15*	rs765884898
Twin_A	22	30118881	G	A	Substitution	intron_variant	*HORMAD2*	-
Twin_A	22	33353441	A	G	Substitution	intron_variant	*LARGE1*	rs1007665809
Twin_A	22	34641670	G	A	Substitution	non_coding_transcript_ exon_variant	*AL021877.2*	rs765266284
Twin_A	22	37322215	G	A	Substitution	intergenic_variant	-	rs898967364
Twin_A	22	44313831	C	T	Substitution	upstream_gene_variant	*SHISAL1*	-
Twin_A	22	45732457	C	CA	Insertion	intron_variant	*ATXN10*	rs971974596
Twin_A	22	50478755	C	G	Substitution	upstream_gene_variant	*SBF1*	-
Twin_A	X	1688073	G	A	Substitution	intergenic_variant	-	rs1402629241

Twin_A	X	3810086	TAC	T	Deletion	intergenic_variant	-	rs768564761
Twin_A	X	9793750	C	CT	Insertion	intron_variant	*SHROOM2*	rs368389307
Twin_A	X	10753565	T	TG	Insertion	intron_variant	*MID1*	rs745425008
Twin_A	X	20062481	T	TA	Insertion	intron_variant	*MAP7D2*	rs141341947
Twin_A	X	21515041	T	TA	Insertion	intron_variant	*CNKSR2*	rs111880925
Twin_A	X	26568200	C	T	Substitution	intron_variant	*VENTXP1*	rs760685569
Twin_A	X	35432699	T	C	Substitution	intergenic_variant	-	-
Twin_A	X	41283486	G	T	Substitution	intergenic_variant	-	-
Twin_A	X	43355581	AAG	A	Deletion	intron_variant	*AL023574.1*	rs747346757
Twin_A	X	48123449	G	A	Substitution	downstream_gene_variant	*SSX6P*	rs1311655959
Twin_A	X	53838712	T	TTTTTTC	Insertion	intergenic_variant	-	rs1491090843
Twin_A	X	56963616	T	TCA	Insertion	intron_variant	*SPIN3*	rs367803469
Twin_A	X	64210261	C	CA	Insertion	upstream_gene_variant	*AMER1*	rs1361525685
Twin_A	X	66930421	A	T	Substitution	intergenic_variant	-	-
Twin_A	X	69456838	G	GT	Insertion	intergenic_variant	-	rs367604349
Twin_A	X	73938408	AAAAT	A	Deletion	downstream_gene_variant	*FTX*	rs1467359046
Twin_A	X	74173149	C	CA	Insertion	upstream_gene_variant	*ATP5MDP1*	rs1215712048
Twin_A	X	80107643	A	T	Substitution	intergenic_variant	-	-
Twin_A	X	102759593	AAC	A	Deletion	intergenic_variant	-	rs4009667
Twin_A	X	105950823	CAG	C	Deletion	intron_variant	*NRK*	rs746466025
Twin_A	X	113523343	C	G	Substitution	downstream_gene_variant	*AC233289.1*	-
Twin_A	X	114062259	G	GT	Insertion	upstream_gene_variant	*XACT*	rs1357977704
Twin_A	X	115465393	T	C	Substitution	intergenic_variant	-	-
Twin_A	X	119227169	C	CA	Insertion	upstream_gene_variant	*AC004835.2*	rs11420910
Twin_A	X	130938304	T	C	Substitution	intergenic_variant	-	-
Twin_A	X	142708225	C	A	Substitution	intron_variant	*AL031073.2*	-
Twin_A	X	144608023	C	T	Substitution	intergenic_variant	-	-
Twin_A	Y	12123479	G	GA	Insertion	intron_variant	*GYG2P1*	rs113188804
Twin_A	Y	20863824	C	T	Substitution	downstream_gene_variant	*GAPDHP17*	-
Twin_B	1	1948293	G	A	Substitution	intron_variant	*CFAP74*	-
Twin_B	1	5360980	A	G	Substitution	intergenic_variant	-	-
Twin_B	1	7752725	A	G	Substitution	intron_variant	*CAMTA1*	rs1558304741
Twin_B	1	8034377	T	C	Substitution	intron_variant	*AL034417.2*	-
Twin_B	1	12823937	C	T	Substitution	downstream_gene_variant	*PRAMEF11*	-
Twin_B	1	16874999	G	T	Substitution	downstream_gene_variant	*BX284668.2*	rs6586499
Twin_B	1	21195179	C	T	Substitution	TF_binding_site_variant	*AL023574.1*	rs1040263005
Twin_B	1	28903853	G	GT	Insertion	intron_variant	*EPB41*	rs776964978
Twin_B	1	30198944	G	A	Substitution	intergenic_variant	-	rs959100631
Twin_B	1	46567799	T	A	Substitution	downstream_gene_variant	*MKNK1*	-
Twin_B	1	46733432	G	A	Substitution	intergenic_variant	-	rs910046833
Twin_B	1	46925759	G	C	Substitution	downstream_gene_variant	*CYP4A11*	-
Twin_B	1	47112037	C	T	Substitution	intron_variant	*CYP4Z1*	rs764020558
Twin_B	1	49651607	C	T	Substitution	intron_variant	*AGBL4*	-
Twin_B	1	52446089	C	T	Substitution	downstream_gene_variant	*TUT4*	-
Twin_B	1	55508033	CA	C	Deletion	intron_variant	*AL603840.1*	-
Twin_B	1	57203495	C	T	Substitution	intron_variant	*DAB1*	-

Twin_B	1	68690432	C	T	Substitution	intron_variant	*AL033530.1*	rs1263200065
Twin_B	1	90494264	G	A	Substitution	intergenic_variant	-	-
Twin_B	1	104033383	A	G	Substitution	intergenic_variant	-	-
Twin_B	1	110190300	G	C	Substitution	intron_variant	*SLC6A17*	-
Twin_B	1	110738023	A	G	Substitution	intergenic_variant	-	-
Twin_B	1	114307529	G	C	Substitution	intergenic_variant	-	-
Twin_B	1	115892616	C	T	Substitution	intergenic_variant	-	rs1302223823
Twin_B	1	143898748	G	A	Substitution	downstream_gene_variant	*H3*	-
Twin_B	1	147277076	C	A	Substitution	downstream_gene_variant	*CHD1L*	-
Twin_B	1	157062690	T	C	Substitution	upstream_gene_variant	*SMU1P1*	-
Twin_B	1	158329018	G	A	Substitution	splice_region_variant	*CD1B*	-
Twin_B	1	159742120	G	T	Substitution	intergenic_variant	-	-
Twin_B	1	165976313	A	G	Substitution	intergenic_variant	-	-
Twin_B	1	166242296	C	T	Substitution	intron_variant	*AL596087.2*	-
Twin_B	1	166972081	G	A	Substitution	intron_variant	*ILDR2*	rs747200728
Twin_B	1	170852137	G	C	Substitution	intron_variant	*BX284613.2*	-
Twin_B	1	175165067	G	T	Substitution	intron_variant	*KIAA0040*	-
Twin_B	1	176214397	G	A	Substitution	intron_variant	*AL591043.2*	-
Twin_B	1	177128801	C	A	Substitution	intron_variant	*ASTN1*	-
Twin_B	1	187326523	G	A	Substitution	intron_variant	*LINC01036*	-
Twin_B	1	190010459	C	T	Substitution	intergenic_variant	-	rs574701171
Twin_B	1	192061801	C	A	Substitution	intergenic_variant	-	-
Twin_B	1	194370180	G	A	Substitution	intergenic_variant	-	rs568526830
Twin_B	1	194805882	A	T	Substitution	intron_variant	*AL353072.2*	rs529354380
Twin_B	1	197898293	C	A	Substitution	upstream_gene_variant	*C1orf53*	-
Twin_B	1	211364253	C	T	Substitution	downstream_gene_variant	*TRAF5*	-
Twin_B	1	213739818	C	T	Substitution	intron_variant	*AC096639.1*	rs751851426
Twin_B	1	216808720	G	A	Substitution	intron_variant	*ESRRG*	-
Twin_B	1	226114304	T	TTA	Insertion	intergenic_variant	-	rs1263797763
Twin_B	1	228759153	C	T	Substitution	intergenic_variant	-	-
Twin_B	1	231661954	G	T	Substitution	intron_variant	*DISC1*	-
Twin_B	1	237523679	G	A	Substitution	intron_variant	*RYR2*	rs1165384474
Twin_B	1	239287292	C	T	Substitution	intergenic_variant	-	-
Twin_B	1	239790569	T	G	Substitution	intron_variant	*CHRM3*	-
Twin_B	2	3059701	C	T	Substitution	intron_variant	*LINC01250*	rs902349418
Twin_B	2	5809965	G	A	Substitution	downstream_gene_variant	*LINC01810*	-
Twin_B	2	7937210	A	T	Substitution	intron_variant	*LINC00298*	-
Twin_B	2	14870316	G	A	Substitution	intron_variant	*AC068286.2*	-
Twin_B	2	17702798	C	A	Substitution	downstream_gene_variant	*SMC6*	rs956379181
Twin_B	2	17816780	A	G	Substitution	missense_variant	*MSGN1*	-
Twin_B	2	21701231	C	T	Substitution	intron_variant	*AC018742.1*	-
Twin_B	2	22021980	T	C	Substitution	intron_variant	*AC096570.1*	-
Twin_B	2	22585137	GT	G	Deletion	intergenic_variant	-	-
Twin_B	2	22585139	T	A	Substitution	intergenic_variant	-	-
Twin_B	2	22619775	C	T	Substitution	intergenic_variant	-	rs540314768
Twin_B	2	22994571	A	G	Substitution	intergenic_variant	-	rs1455047191

Twin_B	2	24019295	T	C	Substitution	downstream_gene_variant	*MFSD2B*	-
Twin_B	2	27314362	G	A	Substitution	downstream_gene_variant	*MPV17*	-
Twin_B	2	28287298	T	A	Substitution	intron_variant	*BABAM2*	-
Twin_B	2	29298533	A	G	Substitution	intron_variant	*ALK*	-
Twin_B	2	30708062	G	A	Substitution	intergenic_variant	-	rs1306752997
Twin_B	2	33204005	TG	T	Deletion	intron_variant	*LTBP1*	-
Twin_B	2	34609950	A	G	Substitution	intron_variant	*LINC01320*	-
Twin_B	2	35621513	C	T	Substitution	intergenic_variant	-	rs185192135
Twin_B	2	39543295	C	T	Substitution	intron_variant	*MAP4K3*	-
Twin_B	2	41851787	C	T	Substitution	regulatory_region_variant	-	-
Twin_B	2	52199227	T	C	Substitution	intron_variant	*AC007402.1*	-
Twin_B	2	55637956	C	T	Substitution	intron_variant	*PNPT1*	rs931753562
Twin_B	2	55778474	G	A	Substitution	intergenic_variant	-	rs1269019834
Twin_B	2	59742006	A	T	Substitution	intron_variant	*AC007100.1*	-
Twin_B	2	65862748	C	T	Substitution	intron_variant	*AC007389.1*	-
Twin_B	2	67673061	AGC	A	Deletion	intron_variant	*AC007422.1*	rs142178881
Twin_B	2	72017331	C	T	Substitution	intergenic_variant	-	-
Twin_B	2	76343534	G	A	Substitution	intron_variant	*AC073091.1*	-
Twin_B	2	77138659	C	T	Substitution	intron_variant	*LRRTM4*	-
Twin_B	2	77305966	T	C	Substitution	intron_variant	*LRRTM4*	rs1354282051
Twin_B	2	79741961	T	A	Substitution	intron_variant	*CTNNA2*	-
Twin_B	2	84976628	T	A	Substitution	intron_variant	*KCMF1*	rs1459851336
Twin_B	2	89011497	G	A	Substitution	upstream_gene_variant	*IGKV1*	rs1327184549
Twin_B	2	92889972	CAT	C	Deletion	intergenic_variant	-	rs1381310181
Twin_B	2	94087365	T	C	Substitution	intergenic_variant	-	rs1304472866
Twin_B	2	100572501	T	C	Substitution	downstream_gene_variant	*PDCL3*	-
Twin_B	2	103642003	G	A	Substitution	intergenic_variant	-	-
Twin_B	2	104572208	AC	A	Deletion	intergenic_variant	-	rs1234997111
Twin_B	2	110152812	G	A	Substitution	intron_variant	*NPHP1*	-
Twin_B	2	112972342	G	A	Substitution	regulatory_region_variant	-	-
Twin_B	2	113360949	C	G	Substitution	intron_variant	*AC016745.1*	-
Twin_B	2	114347202	C	G	Substitution	intergenic_variant	-	-
Twin_B	2	122468457	C	T	Substitution	intron_variant	*AC011246.1*	-
Twin_B	2	124707645	G	T	Substitution	intron_variant	*CNTNAP5*	-
Twin_B	2	125662877	G	T	Substitution	upstream_gene_variant	*AC097499.2*	-
Twin_B	2	130970780	G	C	Substitution	intron_variant	*ARHGEF4*	-
Twin_B	2	136809746	AG	A	Deletion	intron_variant	*THSD7B*	-
Twin_B	2	139199861	C	T	Substitution	intergenic_variant	-	rs9646664
Twin_B	2	139215973	T	A	Substitution	intergenic_variant	-	-
Twin_B	2	139897503	C	CTT	Insertion	intergenic_variant	-	-
Twin_B	2	140219327	G	T	Substitution	downstream_gene_variant	*MTND1P27*	-
Twin_B	2	142622241	G	A	Substitution	intergenic_variant	-	-
Twin_B	2	147580489	T	C	Substitution	intergenic_variant	-	-
Twin_B	2	156625930	G	A	Substitution	intergenic_variant	-	rs919742480
Twin_B	2	156749565	A	G	Substitution	intergenic_variant	-	-

Twin_B	2	157345619	GTGGAAAAGAAGCTGCTGGGATCTACTTACC	G	Deletion	intergenic_variant	-	-
Twin_B	2	158735231	G	A	Substitution	intron_variant	*AC005042.1*	-
Twin_B	2	166646939	T	TC	Insertion	intergenic_variant	-	rs397801435
Twin_B	2	168435246	A	T	Substitution	intergenic_variant	-	rs13393085
Twin_B	2	181754253	G	A	Substitution	intergenic_variant	-	rs928244754
Twin_B	2	186809029	C	A	Substitution	intergenic_variant	-	-
Twin_B	2	187978655	C	T	Substitution	intron_variant	*LINC01090*	-
Twin_B	2	190310023	C	T	Substitution	intron_variant	*HIBCH*	rs1380735746
Twin_B	2	191398351	T	G	Substitution	intron_variant	*MYO1B*	-
Twin_B	2	197588962	G	A	Substitution	intron_variant	*RFTN2*	-
Twin_B	2	202594928	C	T	Substitution	intergenic_variant	-	rs987671953
Twin_B	2	202941148	T	C	Substitution	intron_variant	*CARF*	-
Twin_B	2	214886582	T	G	Substitution	intron_variant	*SNHG31*	-
Twin_B	2	215762883	A	C	Substitution	intron_variant	*LINC00607*	-
Twin_B	2	220248897	TTCTC	T	Deletion	intron_variant	*AC019211.1*	-
Twin_B	2	222991555	T	C	Substitution	intergenic_variant	-	-
Twin_B	2	226321782	C	T	Substitution	intergenic_variant	-	-
Twin_B	2	228241318	G	A	Substitution	intergenic_variant	-	rs746719263
Twin_B	2	236024598	A	G	Substitution	intron_variant	*AGAP1*	-
Twin_B	2	238340846	T	G	Substitution	intron_variant	*TRAF3IP1*	rs200969651
Twin_B	2	238788248	G	A	Substitution	downstream_gene_variant	*AC145625.1*	-
Twin_B	3	4096954	C	T	Substitution	intron_variant	*SUMF1*	rs993902730
Twin_B	3	6268058	G	C	Substitution	intron_variant	*AC026167.1*	-
Twin_B	3	8179865	C	T	Substitution	intron_variant	*LMCD1*	-
Twin_B	3	8330619	T	A	Substitution	intron_variant	*LMCD1*	-
Twin_B	3	8749340	G	C	Substitution	downstream_gene_variant	*CAV3*	-
Twin_B	3	8766880	C	G	Substitution	downstream_gene_variant	*OXTR*	-
Twin_B	3	8800977	A	G	Substitution	intron_variant	*RAD18*	-
Twin_B	3	9646697	T	A	Substitution	regulatory_region_variant	-	rs1001727401
Twin_B	3	39756857	A	T	Substitution	intergenic_variant	-	rs950684554
Twin_B	3	50879771	C	G	Substitution	intron_variant	*DOCK3*	-
Twin_B	3	52975648	C	A	Substitution	intron_variant	*SFMBT1*	-
Twin_B	3	54092594	G	A	Substitution	intron_variant	*AC115282.1*	-
Twin_B	3	57095586	C	T	Substitution	3_prime_UTR_variant	*IL17RD*	rs1428340204
Twin_B	3	59401686	C	T	Substitution	intron_variant	*AC126121.3*	-
Twin_B	3	62217141	G	T	Substitution	downstream_gene_variant	*PTPRG*	-
Twin_B	3	66880474	C	T	Substitution	intergenic_variant	-	rs1357891267
Twin_B	3	74463202	G	A	Substitution	intron_variant	*CNTN3*	-
Twin_B	3	80315771	G	C	Substitution	intergenic_variant	-	-
Twin_B	3	84248079	T	A	Substitution	intergenic_variant	-	rs543628716
Twin_B	3	84322937	C	A	Substitution	intergenic_variant	-	rs912367414
Twin_B	3	89748810	A	C	Substitution	intergenic_variant	-	-
Twin_B	3	95223077	C	T	Substitution	intergenic_variant	-	-
Twin_B	3	96820605	G	A	Substitution	intron_variant	*EPHA6*	-
Twin_B	3	99124717	G	C	Substitution	intergenic_variant	-	rs868051607

Twin_B	3	101873401	C	A	Substitution	regulatory_region_variant	-	-
Twin_B	3	107880323	C	T	Substitution	intron_variant	*LINC00635*	-
Twin_B	3	111393544	T	G	Substitution	intron_variant	*CD96*	-
Twin_B	3	120551312	G	T	Substitution	intergenic_variant	-	-
Twin_B	3	124318634	C	T	Substitution	intron_variant	*KALRN*	-
Twin_B	3	127184194	T	A	Substitution	intron_variant	*AC112482.2*	-
Twin_B	3	145553273	C	T	Substitution	intergenic_variant	-	rs181820859
Twin_B	3	146968300	C	A	Substitution	intron_variant	*AC092957.1*	-
Twin_B	3	148123392	C	T	Substitution	intron_variant	*AC092958.2*	rs893487280
Twin_B	3	150872501	C	T	Substitution	intron_variant	*MINDY4B*	-
Twin_B	3	152516100	G	A	Substitution	intergenic_variant	-	-
Twin_B	3	157178549	C	G	Substitution	downstream_gene_variant	*AC092944.1*	-
Twin_B	3	169022850	C	A	Substitution	intron_variant	*LINC01997*	-
Twin_B	3	179003062	C	T	Substitution	intron_variant	*ZMAT3*	-
Twin_B	3	182077482	T	A	Substitution	intergenic_variant	-	-
Twin_B	3	187366417	G	A	Substitution	upstream_gene_variant	*RTP4*	-
Twin_B	3	192580552	T	A	Substitution	intron_variant	*FGF12*	-
Twin_B	3	197965767	A	G	Substitution	intron_variant	*LMLN*	-
Twin_B	4	1811861	C	T	Substitution	3_prime_UTR_variant	*LETM1*	-
Twin_B	4	6451943	C	A	Substitution	intron_variant	*PPP2R2C*	-
Twin_B	4	12745680	C	T	Substitution	intergenic_variant	-	rs1006183911
Twin_B	4	16908350	G	T	Substitution	intergenic_variant	-	rs750655685
Twin_B	4	26875683	C	G	Substitution	intron_variant	*STIM2*	-
Twin_B	4	27652841	G	A	Substitution	intergenic_variant	-	rs1037907678
Twin_B	4	33615886	T	A	Substitution	intron_variant	*AC079772.1*	-
Twin_B	4	37418639	A	T	Substitution	intron_variant	*NWD2*	rs543621535
Twin_B	4	39429436	G	A	Substitution	intron_variant	*KLB*	-
Twin_B	4	46668719	G	C	Substitution	intergenic_variant	-	-
Twin_B	4	47220483	T	C	Substitution	intron_variant	*GABRB1*	-
Twin_B	4	49731501	T	C	Substitution	intergenic_variant	-	rs1458775311
Twin_B	4	52190355	A	G	Substitution	intergenic_variant	-	-
Twin_B	4	57634212	TG	T	Deletion	intron_variant	*AC093725.2*	-
Twin_B	4	57634214	TG	T	Deletion	intron_variant	*AC093725.2*	-
Twin_B	4	60517882	C	T	Substitution	intergenic_variant	-	-
Twin_B	4	64216296	A	G	Substitution	intergenic_variant	-	rs562121918
Twin_B	4	64665477	T	A	Substitution	intergenic_variant	-	-
Twin_B	4	67380380	G	A	Substitution	intergenic_variant	-	rs551895523
Twin_B	4	70520668	A	T	Substitution	intron_variant	*AMTN*	-
Twin_B	4	71648848	A	G	Substitution	intergenic_variant	-	rs1040824596
Twin_B	4	76047967	G	A	Substitution	intron_variant	*ART3*	-
Twin_B	4	86440642	A	G	Substitution	intron_variant	*MAPK10*	-
Twin_B	4	88387262	T	C	Substitution	upstream_gene_variant	*HERC6*	-
Twin_B	4	108394132	G	A	Substitution	intergenic_variant	-	-
Twin_B	4	115259241	A	T	Substitution	intergenic_variant	-	-
Twin_B	4	117969239	T	C	Substitution	intergenic_variant	-	rs71627746
Twin_B	4	119161162	A	C	Substitution	intron_variant	*MYOZ2*	-

Twin_B	4	120599980	G	A	Substitution	intergenic_variant	-	-
Twin_B	4	125617334	A	T	Substitution	intergenic_variant	-	-
Twin_B	4	126647636	A	G	Substitution	intron_variant	*AC097528.1*	-
Twin_B	4	127771889	A	G	Substitution	intron_variant	*SLC25A31*	rs1201508233
Twin_B	4	140154007	T	G	Substitution	5_prime_UTR_variant	*MAML3*	-
Twin_B	4	144879760	C	T	Substitution	downstream_gene_variant	*AC109811.1*	-
Twin_B	4	147616580	G	C	Substitution	intron_variant	*AC093835.1*	-
Twin_B	4	152098840	G	A	Substitution	intron_variant	*LINC02273*	-
Twin_B	4	161716379	T	C	Substitution	intron_variant	*FSTL5*	-
Twin_B	4	166613505	T	C	Substitution	intergenic_variant	-	-
Twin_B	4	168686621	G	A	Substitution	intron_variant	*PALLD*	rs1030969924
Twin_B	4	174170616	G	A	Substitution	intron_variant	*LINC02268*	rs1241611683
Twin_B	4	179639868	C	T	Substitution	intergenic_variant	-	-
Twin_B	4	180135661	C	A	Substitution	intergenic_variant	-	-
Twin_B	4	184638048	T	C	Substitution	intron_variant	*CASP3*	-
Twin_B	4	187716917	C	T	Substitution	downstream_gene_variant	*AC097521.1*	rs1306869395
Twin_B	5	4311163	C	T	Substitution	intergenic_variant	-	rs755225209
Twin_B	5	4490226	G	A	Substitution	intron_variant	*AC106799.2*	-
Twin_B	5	5721155	C	A	Substitution	intergenic_variant	-	rs1404060316
Twin_B	5	7425190	G	A	Substitution	intron_variant	*ADCY2*	-
Twin_B	5	9500811	T	C	Substitution	intron_variant	*SEMA5A*	-
Twin_B	5	11361065	G	A	Substitution	downstream_gene_variant	*CTNND2*	-
Twin_B	5	18474168	C	T	Substitution	intergenic_variant	-	-
Twin_B	5	19703201	C	T	Substitution	intron_variant	*CDH18*	-
Twin_B	5	29927998	G	A	Substitution	intergenic_variant	-	rs910971117
Twin_B	5	30228903	G	A	Substitution	intergenic_variant	-	-
Twin_B	5	32183802	G	A	Substitution	intergenic_variant	-	rs1321151704
Twin_B	5	35163531	G	A	Substitution	intron_variant	*PRLR*	-
Twin_B	5	41232152	T	G	Substitution	intron_variant	*C6*	-
Twin_B	5	41663190	G	A	Substitution	intergenic_variant	-	-
Twin_B	5	42222760	G	A	Substitution	intergenic_variant	-	-
Twin_B	5	45194560	C	T	Substitution	intergenic_variant	-	rs550304760
Twin_B	5	47101412	G	A	Substitution	intergenic_variant	-	rs1255895881
Twin_B	5	50007872	T	C	Substitution	intergenic_variant	-	-
Twin_B	5	53160317	A	G	Substitution	intergenic_variant	-	-
Twin_B	5	59291742	T	G	Substitution	intron_variant	*PDE4D*	rs57801368
Twin_B	5	63614795	GAC	G	Deletion	intergenic_variant	-	rs1396639575
Twin_B	5	72404389	A	G	Substitution	intergenic_variant	-	-
Twin_B	5	84595108	T	C	Substitution	intergenic_variant	-	rs1003756160
Twin_B	5	85129366	T	C	Substitution	intergenic_variant	-	-
Twin_B	5	93483038	G	A	Substitution	intron_variant	*NR2F1*	rs918359530
Twin_B	5	105561137	C	T	Substitution	intergenic_variant	-	-
Twin_B	5	106685060	G	A	Substitution	intergenic_variant	-	-
Twin_B	5	110479443	C	T	Substitution	intron_variant	*TMEM232*	-
Twin_B	5	119572129	C	T	Substitution	intron_variant	*HSD17B4*	-
Twin_B	5	122619069	G	A	Substitution	intergenic_variant	-	-

Twin_B	5	130265666	C	T	Substitution	intergenic_variant	-	rs62391477
Twin_B	5	134868896	C	T	Substitution	intergenic_variant	-	rs910097048
Twin_B	5	137416371	G	C	Substitution	intron_variant	*SPOCK1*	-
Twin_B	5	140836906	A	G	Substitution	3_prime_UTR_variant	*PCDHA7*	-
Twin_B	5	147046232	G	T	Substitution	intron_variant	*PPP2R2B*	-
Twin_B	5	155317926	G	T	Substitution	intergenic_variant	-	-
Twin_B	5	157005759	C	A	Substitution	intergenic_variant	-	-
Twin_B	5	158293742	G	A	Substitution	intergenic_variant	-	-
Twin_B	5	159980907	T	C	Substitution	intergenic_variant	-	-
Twin_B	5	161395724	G	A	Substitution	intron_variant	*GABRB2*	-
Twin_B	5	161830310	G	T	Substitution	intergenic_variant	-	-
Twin_B	5	166256348	G	A	Substitution	intergenic_variant	-	-
Twin_B	5	175182485	G	A	Substitution	intergenic_variant	-	-
Twin_B	5	178484296	T	G	Substitution	intron_variant	*COL23A1*	-
Twin_B	6	1850433	CT	C	Deletion	intron_variant	*GMDS*	-
Twin_B	6	18553216	G	T	Substitution	intron_variant	*MIR548A1HG*	-
Twin_B	6	20182475	T	G	Substitution	intron_variant	*MBOAT1*	-
Twin_B	6	23725753	C	T	Substitution	intergenic_variant	-	-
Twin_B	6	23820823	AG	A	Deletion	intergenic_variant	-	-
Twin_B	6	32530993	T	C	Substitution	regulatory_region_variant	-	rs74714255
Twin_B	6	33355096	C	T	Substitution	intergenic_variant	-	-
Twin_B	6	40322738	T	C	Substitution	intergenic_variant	-	-
Twin_B	6	45795560	T	G	Substitution	intergenic_variant	-	-
Twin_B	6	49245413	G	A	Substitution	intergenic_variant	-	-
Twin_B	6	49866189	C	T	Substitution	intron_variant	*CRISP1*	-
Twin_B	6	50314470	C	T	Substitution	intergenic_variant	-	rs192304872
Twin_B	6	68164502	G	A	Substitution	intergenic_variant	-	rs539171573
Twin_B	6	73054903	C	A	Substitution	intron_variant	*KCNQ5*	-
Twin_B	6	77608472	C	G	Substitution	intergenic_variant	-	rs145374562
Twin_B	6	78332113	A	G	Substitution	intergenic_variant	-	-
Twin_B	6	80558206	T	C	Substitution	upstream_gene_variant	*AL590824.1*	-
Twin_B	6	95406340	C	T	Substitution	intergenic_variant	-	rs548629996
Twin_B	6	101437584	G	A	Substitution	intron_variant	*GRIK2*	rs1368442237
Twin_B	6	102787939	T	A	Substitution	intergenic_variant	-	-
Twin_B	6	105574693	A	C	Substitution	intron_variant	*AL096794.1*	-
Twin_B	6	110560428	G	A	Substitution	downstream_gene_variant	*AL512430.3*	-
Twin_B	6	115939145	A	G	Substitution	3_prime_UTR_variant	*FRK*	-
Twin_B	6	132482937	C	T	Substitution	intron_variant	*STX7*	-
Twin_B	6	133738256	C	A	Substitution	intron_variant	*TARID*	-
Twin_B	6	153868755	C	A	Substitution	intergenic_variant	-	-
Twin_B	6	155942473	C	T	Substitution	intron_variant	*AL589693.1*	rs1029874863
Twin_B	6	157796066	T	G	Substitution	intron_variant	*SNX9*	-
Twin_B	6	160228462	G	T	Substitution	intron_variant	*SLC22A2*	-
Twin_B	6	160512492	C	T	Substitution	intron_variant	*LPAL2*	rs573849624
Twin_B	6	163903570	G	A	Substitution	intergenic_variant	-	rs138646256
Twin_B	6	164337205	G	A	Substitution	intergenic_variant	-	rs1395517073

Twin_B	6	170147589	C	A	Substitution	downstream_gene_variant	*AL596442.4*	-
Twin_B	7	4358841	C	T	Substitution	intergenic_variant	-	rs746784235
Twin_B	7	7383076	A	T	Substitution	intron_variant	*COL28A1*	-
Twin_B	7	9012014	G	A	Substitution	intergenic_variant	-	-
Twin_B	7	10181486	C	T	Substitution	intergenic_variant	-	-
Twin_B	7	20905885	C	G	Substitution	intron_variant	*LINC01162*	-
Twin_B	7	26002719	G	A	Substitution	intergenic_variant	-	-
Twin_B	7	27125974	G	A	Substitution	downstream_gene_variant	*AC004080.6*	-
Twin_B	7	30124532	G	T	Substitution	downstream_gene_variant	*AC007036.5*	-
Twin_B	7	33443171	A	T	Substitution	intron_variant	*BBS9*	-
Twin_B	7	33920177	G	A	Substitution	intron_variant	*BMPER*	-
Twin_B	7	34052098	C	T	Substitution	intron_variant	*BMPER*	-
Twin_B	7	43433707	A	C	Substitution	intron_variant	*HECW1*	-
Twin_B	7	44053423	G	A	Substitution	downstream_gene_variant	*DBNL*	rs1254859423
Twin_B	7	45053757	A	G	Substitution	intron_variant	*CCM2*	-
Twin_B	7	45393430	C	G	Substitution	upstream_gene_variant	*ELK1P1*	-
Twin_B	7	48916465	T	C	Substitution	upstream_gene_variant	*CDC14C*	rs1245635730
Twin_B	7	54587046	C	A	Substitution	intergenic_variant	-	-
Twin_B	7	57036330	G	A	Substitution	intergenic_variant	-	rs368641980
Twin_B	7	60485855	A	C	Substitution	intergenic_variant	-	rs1335884497
Twin_B	7	60938781	CA	C	Deletion	intergenic_variant	-	rs1250869720
Twin_B	7	61436516	T	A	Substitution	intergenic_variant	-	rs1169588355
Twin_B	7	61970818	T	C	Substitution	intergenic_variant	-	rs1283251044
Twin_B	7	62882059	A	C	Substitution	intergenic_variant	-	rs112905112
Twin_B	7	62882071	C	A	Substitution	intergenic_variant	-	rs1395947656
Twin_B	7	65034946	G	A	Substitution	upstream_gene_variant	*AC073210.2*	rs954732530
Twin_B	7	65590326	T	C	Substitution	intergenic_variant	-	-
Twin_B	7	67521209	C	T	Substitution	intergenic_variant	-	rs183903994
Twin_B	7	71219697	C	T	Substitution	intron_variant	*GALNT17*	-
Twin_B	7	71503989	G	A	Substitution	intron_variant	*GALNT17*	-
Twin_B	7	73980208	CAA	C	Deletion	intergenic_variant	-	rs60681274
Twin_B	7	75784737	G	A	Substitution	intron_variant	*CCL26*	-
Twin_B	7	86130259	T	G	Substitution	intergenic_variant	-	-
Twin_B	7	102788991	TATG	T	Deletion	downstream_gene_variant	*FAM185A*	rs1163997473
Twin_B	7	108865841	G	A	Substitution	intergenic_variant	-	-
Twin_B	7	111235913	G	A	Substitution	intron_variant	*IMMP2L*	-
Twin_B	7	116155279	G	C	Substitution	intron_variant	*TFEC*	-
Twin_B	7	134178896	G	A	Substitution	intron_variant	*LRGUK*	-
Twin_B	7	137739877	G	T	Substitution	intron_variant	*DGKI*	rs190071670
Twin_B	7	142637901	C	A	Substitution	downstream_gene_variant	*TRBV21*	-
Twin_B	7	146668441	T	TGG	Insertion	intron_variant	*CNTNAP2*	rs35693921
Twin_B	7	150085019	G	A	Substitution	intergenic_variant	-	rs1461424202
Twin_B	7	152410761	T	C	Substitution	intron_variant	*KMT2C*	rs80138121
Twin_B	7	156931430	C	T	Substitution	intergenic_variant	-	-
Twin_B	8	924933	C	T	Substitution	intron_variant	*DLGAP2*	-
Twin_B	8	3391860	G	T	Substitution	intron_variant	*CSMD1*	-

Twin_B	8	4997791	G	A	Substitution	regulatory_region_variant	-	-
Twin_B	8	5826762	C	T	Substitution	intergenic_variant	-	-
Twin_B	8	19798556	G	T	Substitution	intergenic_variant	-	-
Twin_B	8	20863668	T	A	Substitution	intergenic_variant	-	-
Twin_B	8	25151260	C	A	Substitution	intergenic_variant	-	-
Twin_B	8	26940513	G	A	Substitution	intergenic_variant	-	-
Twin_B	8	32400294	G	T	Substitution	intron_variant	*NRG1*	-
Twin_B	8	32524703	G	A	Substitution	intron_variant	*NRG1*	-
Twin_B	8	34194743	G	C	Substitution	intron_variant	*AC087855.2*	-
Twin_B	8	41188472	A	T	Substitution	intergenic_variant	-	-
Twin_B	8	43877252	G	GT	Insertion	intergenic_variant	-	-
Twin_B	8	44701946	T	G	Substitution	intergenic_variant	-	rs1325351437
Twin_B	8	46315152	G	A	Substitution	intergenic_variant	-	-
Twin_B	8	49102414	A	C	Substitution	intron_variant	*AC044893.1*	-
Twin_B	8	50264084	A	G	Substitution	intron_variant	*SNTG1*	-
Twin_B	8	51178687	G	A	Substitution	intergenic_variant	-	rs1207366775
Twin_B	8	53427578	TAC	T	Deletion	intron_variant	*AC022034.3*	-
Twin_B	8	57557945	G	A	Substitution	intron_variant	*AC090796.1*	-
Twin_B	8	58506466	G	T	Substitution	downstream_gene_variant	*PPIAP85*	rs973571307
Twin_B	8	59792693	T	A	Substitution	intergenic_variant	-	-
Twin_B	8	64469993	C	A	Substitution	intergenic_variant	-	-
Twin_B	8	69896186	C	A	Substitution	intergenic_variant	-	rs1204844083
Twin_B	8	70358021	AGCCTG	A	Deletion	downstream_gene_variant	*RNY3P14*	-
Twin_B	8	71477385	T	C	Substitution	intron_variant	*EYA1*	-
Twin_B	8	78483436	T	C	Substitution	intron_variant	*PKIA*	-
Twin_B	8	81785469	A	G	Substitution	intergenic_variant	-	rs1439987215
Twin_B	8	82359769	T	C	Substitution	intron_variant	*AC060765.2*	-
Twin_B	8	84130068	C	T	Substitution	intron_variant	*AC015522.1*	rs114417960
Twin_B	8	87159687	A	T	Substitution	intron_variant	*CNBD1*	-
Twin_B	8	87159688	G	A	Substitution	intron_variant	*CNBD1*	-
Twin_B	8	98496864	T	C	Substitution	intron_variant	*STK3*	rs1015493993
Twin_B	8	98657310	C	T	Substitution	intron_variant	*STK3*	-
Twin_B	8	100119116	G	A	Substitution	intron_variant	*RGS22*	-
Twin_B	8	109438926	A	G	Substitution	missense_variant	*PKHD1L1*	-
Twin_B	8	109534244	C	T	Substitution	3_prime_UTR_variant	*PKHD1L1*	rs539684370
Twin_B	8	111647834	G	C	Substitution	intron_variant	*LINC02237*	-
Twin_B	8	111905799	G	A	Substitution	intergenic_variant	-	rs1453724199
Twin_B	8	113516282	C	A	Substitution	intergenic_variant	-	-
Twin_B	8	114563854	C	T	Substitution	intergenic_variant	-	-
Twin_B	8	116551195	C	T	Substitution	intergenic_variant	-	-
Twin_B	8	122234933	G	A	Substitution	intergenic_variant	-	rs1423584842
Twin_B	8	126381321	C	T	Substitution	intergenic_variant	-	rs1433139292
Twin_B	8	131552936	T	A	Substitution	intergenic_variant	-	-
Twin_B	8	133562727	G	A	Substitution	downstream_gene_variant	*ST3GAL1*	-
Twin_B	8	136497219	A	G	Substitution	intron_variant	*LINC02055*	-
Twin_B	8	139066773	G	A	Substitution	intergenic_variant	-	-

Twin_B	9	595757	G	A	Substitution	intron_variant	*KANK1*	rs1028227944
Twin_B	9	5583708	G	A	Substitution	intron_variant	*AL162253.2*	-
Twin_B	9	9208774	T	C	Substitution	intron_variant	*PTPRD*	-
Twin_B	9	10229559	G	A	Substitution	intron_variant	*PTPRD*	-
Twin_B	9	12626664	C	T	Substitution	intergenic_variant	-	-
Twin_B	9	12964481	A	T	Substitution	intron_variant	*AL161449.2*	rs1563767337
Twin_B	9	21096138	A	G	Substitution	intergenic_variant	-	rs896126975
Twin_B	9	23036773	C	T	Substitution	intron_variant	*AL391117.1*	-
Twin_B	9	24723875	G	T	Substitution	intergenic_variant	-	-
Twin_B	9	24924904	G	A	Substitution	intergenic_variant	-	rs193241464
Twin_B	9	38292964	C	T	Substitution	intergenic_variant	-	-
Twin_B	9	38995507	T	C	Substitution	intron_variant	*AL953883.1*	rs1422022710
Twin_B	9	42290039	T	A	Substitution	intron_variant	*AL445584.2*	rs1211321020
Twin_B	9	42760705	C	T	Substitution	intron_variant	*BX664718.2*	rs56025544
Twin_B	9	42945900	G	A	Substitution	intron_variant	*ANKRD20A7P*	-
Twin_B	9	65672213	T	C	Substitution	intron_variant	*CBWD5*	rs796503817
Twin_B	9	66779265	C	T	Substitution	intergenic_variant	-	rs1271105675
Twin_B	9	67205531	TG	T	Deletion	intron_variant	*AL162233.1*	rs368975415
Twin_B	9	73452437	GTTTC	G	Deletion	intergenic_variant	-	-
Twin_B	9	74317541	T	C	Substitution	intergenic_variant	-	-
Twin_B	9	77105186	C	T	Substitution	intergenic_variant	-	rs538603299
Twin_B	9	87267784	G	A	Substitution	intergenic_variant	-	rs1328259639
Twin_B	9	88848472	C	A	Substitution	intergenic_variant	-	-
Twin_B	9	95722226	G	T	Substitution	intergenic_variant	-	-
Twin_B	9	98676266	A	T	Substitution	intron_variant	*GABBR2*	-
Twin_B	9	102339974	A	G	Substitution	intergenic_variant	-	-
Twin_B	9	102993672	C	A	Substitution	upstream_gene_variant	*CYLC2*	-
Twin_B	9	103368304	G	A	Substitution	intron_variant	*AL390962.1*	-
Twin_B	9	104257777	G	T	Substitution	intergenic_variant	-	-
Twin_B	9	104331598	T	G	Substitution	upstream_gene_variant	*AL512646.1*	-
Twin_B	9	105298502	A	G	Substitution	intron_variant	*SLC44A1*	-
Twin_B	9	118261593	G	A	Substitution	intergenic_variant	-	rs1353495123
Twin_B	9	124851965	C	T	Substitution	upstream_gene_variant	*WDR38*	rs962035631
Twin_B	9	127439569	C	T	Substitution	intron_variant	*ZNF79*	-
Twin_B	9	130429019	C	T	Substitution	intron_variant	*HMCN2*	rs554518691
Twin_B	9	131860557	T	A	Substitution	missense_variant	*MED27*	-
Twin_B	9	134007274	C	T	Substitution	intergenic_variant	-	-
Twin_B	9	135090386	T	G	Substitution	intron_variant	*OLFM1*	rs56337838
Twin_B	10	1176537	C	T	Substitution	downstream_gene_variant	*ADARB2*	rs757910551
Twin_B	10	9492286	A	C	Substitution	intergenic_variant	-	rs148903298
Twin_B	10	10933457	T	A	Substitution	downstream_gene_variant	*LINC00710*	-
Twin_B	10	12067169	A	G	Substitution	upstream_gene_variant	*DHTKD1*	rs1055077923
Twin_B	10	14501740	C	A	Substitution	intergenic_variant	-	-
Twin_B	10	19776842	A	T	Substitution	intron_variant	*MALRD1*	-
Twin_B	10	24984824	G	A	Substitution	missense_variant	*ENKUR*	rs200773323
Twin_B	10	30158617	C	A	Substitution	intergenic_variant	-	-

Twin_B	10	33849532	T	C	Substitution	intergenic_variant	-	-
Twin_B	10	37826153	C	T	Substitution	downstream_gene_variant	*ZNF248*	-
Twin_B	10	42358812	A	G	Substitution	intron_variant	*BX322639.1*	-
Twin_B	10	46311814	C	A	Substitution	downstream_gene_variant	*ANTXRL*	-
Twin_B	10	54568022	G	T	Substitution	intron_variant	*PCDH15*	-
Twin_B	10	65525504	T	C	Substitution	intergenic_variant	-	-
Twin_B	10	73104873	G	A	Substitution	upstream_gene_variant	*AL731563.3*	-
Twin_B	10	75129152	A	T	Substitution	downstream_gene_variant	*RPS26P42*	rs1477868012
Twin_B	10	82739562	G	A	Substitution	intron_variant	*NRG3*	-
Twin_B	10	86750662	C	A	Substitution	non_coding_transcript_exon_variant	*AC067750.1*	-
Twin_B	10	88654107	TA	T	Deletion	intergenic_variant	-	rs200426640
Twin_B	10	90700494	C	T	Substitution	intergenic_variant	-	-
Twin_B	10	92770872	T	A	Substitution	intergenic_variant	-	-
Twin_B	10	93234495	G	T	Substitution	intergenic_variant	-	rs184258432
Twin_B	10	102661031	C	T	Substitution	downstream_gene_variant	*TRIM8*	-
Twin_B	10	103273122	G	A	Substitution	TF_binding_site_variant	*INA*	rs1286474689
Twin_B	10	110609195	T	A	Substitution	downstream_gene_variant	*SMC3*	-
Twin_B	10	115355253	A	G	Substitution	intron_variant	*ATRNL1*	-
Twin_B	10	116513252	G	A	Substitution	intergenic_variant	-	-
Twin_B	10	118845653	C	T	Substitution	intron_variant	*AL157388.1*	-
Twin_B	10	121236516	G	A	Substitution	intergenic_variant	-	rs75201857
Twin_B	10	125904313	C	T	Substitution	intron_variant	*FANK1*	rs571397776
Twin_B	10	131096200	G	A	Substitution	downstream_gene_variant	*TCERG1L*	-
Twin_B	10	132055519	G	A	Substitution	intron_variant	*JAKMIP3*	rs11813033
Twin_B	11	2172588	G	A	Substitution	upstream_gene_variant	*MIR4686*	rs970284747
Twin_B	11	4222159	G	A	Substitution	intergenic_variant	-	rs78524524
Twin_B	11	4861410	G	A	Substitution	intron_variant	*MMP26*	-
Twin_B	11	5952018	A	G	Substitution	downstream_gene_variant	*OR56A3*	-
Twin_B	11	8836985	G	A	Substitution	intron_variant	*DENND2B*	-
Twin_B	11	11456484	C	T	Substitution	intron_variant	*GALNT18*	-
Twin_B	11	19063121	G	A	Substitution	upstream_gene_variant	*MRGPRX2*	-
Twin_B	11	20315541	A	G	Substitution	intergenic_variant	-	-
Twin_B	11	20325176	G	T	Substitution	intergenic_variant	-	-
Twin_B	11	36738318	A	G	Substitution	intergenic_variant	-	rs1297968115
Twin_B	11	36973840	G	A	Substitution	intergenic_variant	-	rs983687999
Twin_B	11	42300900	G	A	Substitution	intergenic_variant	-	rs915445215
Twin_B	11	42391773	A	G	Substitution	intergenic_variant	-	-
Twin_B	11	44879657	AG	A	Deletion	intron_variant	*TSPAN18*	-
Twin_B	11	54940328	C	A	Substitution	intergenic_variant	-	-
Twin_B	11	55783372	G	A	Substitution	upstream_gene_variant	*OR5D15P*	rs566667825
Twin_B	11	81625515	T	C	Substitution	intergenic_variant	-	rs1018593916
Twin_B	11	82290505	C	T	Substitution	intron_variant	*MIR4300HG*	rs181088482
Twin_B	11	82375060	A	T	Substitution	intron_variant	*MIR4300HG*	-
Twin_B	11	83760647	C	A	Substitution	intron_variant	*DLG2*	-
Twin_B	11	88442713	G	T	Substitution	intergenic_variant	-	-
Twin_B	11	91384045	G	T	Substitution	downstream_gene_variant	*AP002791.1*	-

Twin_B	11	93116767	C	A	Substitution	intergenic_variant	-	-
Twin_B	11	93907782	C	T	Substitution	intergenic_variant	-	rs1382214357
Twin_B	11	103587114	C	G	Substitution	intergenic_variant	-	rs1026010882
Twin_B	11	107071946	A	G	Substitution	intergenic_variant	-	-
Twin_B	11	112616847	C	T	Substitution	intron_variant	*LINC02763*	rs920020081
Twin_B	11	114708783	G	A	Substitution	downstream_gene_variant	*NXPE2*	-
Twin_B	11	114958122	G	T	Substitution	intergenic_variant	-	-
Twin_B	11	117410900	C	A	Substitution	downstream_gene_variant	*CEP164*	-
Twin_B	11	122025753	C	T	Substitution	downstream_gene_variant	*MIR100HG*	-
Twin_B	11	122979808	C	A	Substitution	intron_variant	*BSX*	-
Twin_B	11	124291010	G	A	Substitution	intergenic_variant	-	-
Twin_B	11	128554716	C	T	Substitution	intron_variant	*ETS1*	rs78814353
Twin_B	11	132785635	G	A	Substitution	intron_variant	*OPCML*	-
Twin_B	11	133338345	C	T	Substitution	intron_variant	*OPCML*	rs907145809
Twin_B	11	133761282	T	A	Substitution	intergenic_variant	-	-
Twin_B	12	6117683	C	T	Substitution	downstream_gene_variant	*VWF*	rs546633664
Twin_B	12	6307820	C	T	Substitution	regulatory_region_variant	-	rs1256264888
Twin_B	12	6520513	G	A	Substitution	downstream_gene_variant	*NCAPD2*	-
Twin_B	12	6789295	C	T	Substitution	intron_variant	*CD4*	-
Twin_B	12	8205043	GC	G	Deletion	downstream_gene_variant	*FAM66C*	rs144267844
Twin_B	12	9683004	G	A	Substitution	downstream_gene_variant	*CLEC2D*	rs188982524
Twin_B	12	11785599	A	G	Substitution	intron_variant	*ETV6*	rs778627490
Twin_B	12	18037838	C	T	Substitution	intergenic_variant	-	-
Twin_B	12	20532321	G	A	Substitution	intron_variant	*PDE3A*	-
Twin_B	12	24620263	G	A	Substitution	intergenic_variant	-	-
Twin_B	12	25350278	C	T	Substitution	intergenic_variant	-	rs944670622
Twin_B	12	26030102	G	A	Substitution	intron_variant	*RASSF8*	rs771217533
Twin_B	12	32986809	G	A	Substitution	non_coding_transcript_exon_variant	*AC087311.1*	-
Twin_B	12	34148556	C	T	Substitution	upstream_gene_variant	*AC140847.1*	rs1252261441
Twin_B	12	35988466	T	C	Substitution	intergenic_variant	-	rs1164988538
Twin_B	12	38746830	T	C	Substitution	intron_variant	*CPNE8*	-
Twin_B	12	43641305	G	A	Substitution	intergenic_variant	-	rs568390024
Twin_B	12	46321777	C	T	Substitution	intergenic_variant	-	rs150303645
Twin_B	12	50558428	A	T	Substitution	intron_variant	*DIP2B*	-
Twin_B	12	61470230	G	C	Substitution	intergenic_variant	-	rs1421114192
Twin_B	12	68411586	C	A	Substitution	intron_variant	*LINC02384*	-
Twin_B	12	77504158	G	A	Substitution	intron_variant	*NAV3*	rs991324887
Twin_B	12	78382902	C	T	Substitution	intron_variant	*AC128707.1*	-
Twin_B	12	79065377	C	T	Substitution	intron_variant	*SYT1*	rs993033305
Twin_B	12	80780558	A	T	Substitution	intergenic_variant	-	-
Twin_B	12	83300749	C	T	Substitution	intergenic_variant	-	rs538003003
Twin_B	12	84484910	G	A	Substitution	intergenic_variant	-	-
Twin_B	12	86386833	C	A	Substitution	intron_variant	*MGAT4C*	-
Twin_B	12	87232193	G	A	Substitution	intergenic_variant	-	-
Twin_B	12	87810848	T	G	Substitution	intergenic_variant	-	-
Twin_B	12	88071791	G	A	Substitution	missense_variant	*CEP290*	rs761572735

Twin_B	12	90780665	C	T	Substitution	intron_variant	*AC112481.2*	-
Twin_B	12	95719552	T	C	Substitution	intron_variant	*NTN4*	COSV59220221
Twin_B	12	96233025	A	G	Substitution	intron_variant	*ELK3*	-
Twin_B	12	100396583	T	A	Substitution	intron_variant	*SLC17A8*	rs1220972791
Twin_B	12	105124962	C	T	Substitution	intron_variant	*WASHC4*	-
Twin_B	12	109600533	C	G	Substitution	downstream_gene_variant	*MVK*	-
Twin_B	12	113992218	G	A	Substitution	intergenic_variant	-	-
Twin_B	12	114026760	G	A	Substitution	intergenic_variant	-	rs142779521
Twin_B	12	115283072	C	A	Substitution	intergenic_variant	-	-
Twin_B	12	116627058	G	C	Substitution	intergenic_variant	-	-
Twin_B	12	121968151	A	T	Substitution	missense_variant	*CFAP251*	-
Twin_B	12	126210035	C	T	Substitution	intron_variant	*AC007368.1*	-
Twin_B	12	128451756	C	T	Substitution	intron_variant	*TMEM132C*	-
Twin_B	12	129158414	A	G	Substitution	intron_variant	*TMEM132D*	-
Twin_B	12	129545572	G	A	Substitution	intron_variant	*TMEM132D*	rs548980444
Twin_B	12	132821132	C	T	Substitution	downstream_gene_variant	*CHFR*	rs981825385
Twin_B	13	16017183	C	T	Substitution	intergenic_variant	-	rs1346805079
Twin_B	13	16661736	C	A	Substitution	intergenic_variant	-	rs1306983302
Twin_B	13	17304667	T	A	Substitution	intergenic_variant	-	rs1456450575
Twin_B	13	17466593	C	T	Substitution	intergenic_variant	-	rs1444264382
Twin_B	13	34712768	G	A	Substitution	intergenic_variant	-	rs956638263
Twin_B	13	37492154	T	A	Substitution	intergenic_variant	-	rs5028161
Twin_B	13	51338069	C	T	Substitution	intron_variant	*SERPINE3*	rs943522065
Twin_B	13	57247644	C	T	Substitution	intergenic_variant	-	rs1015436925
Twin_B	13	64005202	A	C	Substitution	intron_variant	*LINC00355*	-
Twin_B	13	66744717	G	A	Substitution	intron_variant	*PCDH9*	rs1436271609
Twin_B	13	68771491	A	G	Substitution	intergenic_variant	-	-
Twin_B	13	81289794	T	C	Substitution	intron_variant	*AL353633.1*	rs1481888041
Twin_B	13	92237367	G	A	Substitution	downstream_gene_variant	*AL157815.1*	rs1265338638
Twin_B	13	95505939	C	G	Substitution	intron_variant	*CLDN10*	-
Twin_B	13	106932614	C	T	Substitution	intergenic_variant	-	-
Twin_B	13	109688927	C	T	Substitution	intron_variant	*AL163541.1*	-
Twin_B	13	112351281	G	A	Substitution	intergenic_variant	-	rs115643630
Twin_B	13	112451541	C	T	Substitution	intergenic_variant	-	rs1471601068
Twin_B	14	25383375	A	AT	Insertion	intergenic_variant	-	rs1443265250
Twin_B	14	28045753	A	G	Substitution	intergenic_variant	-	rs1373127559
Twin_B	14	28160448	C	G	Substitution	intergenic_variant	-	-
Twin_B	14	28481339	C	T	Substitution	intergenic_variant	-	rs987212689
Twin_B	14	28933071	A	T	Substitution	downstream_gene_variant	*LINC02327*	-
Twin_B	14	34755189	T	C	Substitution	downstream_gene_variant	*BAZ1A*	-
Twin_B	14	39719161	C	T	Substitution	intron_variant	*AL049828.1*	-
Twin_B	14	39760915	GC	G	Deletion	intron_variant	*AL049828.1*	-
Twin_B	14	40646136	G	C	Substitution	intron_variant	*AL390800.1*	-
Twin_B	14	42163357	C	T	Substitution	intergenic_variant	-	-
Twin_B	14	44806590	C	A	Substitution	intergenic_variant	-	-
Twin_B	14	45687961	C	A	Substitution	intergenic_variant	-	-

Twin_B	14	49208635	C	T	Substitution	intergenic_variant	-	-
Twin_B	14	52546562	G	A	Substitution	intron_variant	*TXNDC16*	-
Twin_B	14	55273174	G	GTT	Insertion	intron_variant	*FBXO34*	-
Twin_B	14	62950522	C	T	Substitution	missense_variant	*KCNH5*	rs587777164, CM1310653, COSV59758318, COSV59758899
Twin_B	14	76286201	T	G	Substitution	downstream_gene_variant	*AC016526.3*	-
Twin_B	14	87098315	C	T	Substitution	intergenic_variant	-	-
Twin_B	14	94965692	G	C	Substitution	upstream_gene_variant	*LINC02279*	-
Twin_B	14	96647102	C	T	Substitution	intergenic_variant	-	-
Twin_B	14	98254213	A	T	Substitution	intergenic_variant	-	rs200791071
Twin_B	14	99920248	A	G	Substitution	downstream_gene_variant	*EML1*	-
Twin_B	15	17031425	C	T	Substitution	intergenic_variant	-	rs988560630
Twin_B	15	24913380	C	T	Substitution	downstream_gene_variant	*AC090983.2*	rs995136908
Twin_B	15	25046210	A	G	Substitution	downstream_gene_variant	*SNHG14*	-
Twin_B	15	26017195	A	G	Substitution	intron_variant	*LINC02346*	-
Twin_B	15	40140978	A	G	Substitution	intergenic_variant	-	-
Twin_B	15	45848202	CAAA	C	Deletion	intron_variant	*AC068722.1*	rs1232293898
Twin_B	15	48080087	A	G	Substitution	intergenic_variant	-	-
Twin_B	15	48928081	AG	A	Deletion	intron_variant	*SHC4*	-
Twin_B	15	60120205	G	A	Substitution	regulatory_region_variant	-	-
Twin_B	15	60352764	T	C	Substitution	downstream_gene_variant	*ANXA2*	rs1242885715
Twin_B	15	71202412	A	G	Substitution	intron_variant	*THSD4*	-
Twin_B	15	72502854	C	T	Substitution	intron_variant	*ARIH1*	rs1435288541
Twin_B	15	77267388	A	G	Substitution	intron_variant	*PEAK1*	-
Twin_B	15	79562906	G	A	Substitution	upstream_gene_variant	*AC023968.2*	-
Twin_B	15	80681168	T	A	Substitution	intron_variant	*ABHD17C*	-
Twin_B	15	83132180	G	C	Substitution	3_prime_UTR_variant	*HDGFL3*	-
Twin_B	15	87983261	T	A	Substitution	intron_variant	*NTRK3*	-
Twin_B	15	95004101	T	C	Substitution	intron_variant	*AC087633.2*	-
Twin_B	16	737082	C	G	Substitution	downstream_gene_variant	*CIAO3*	-
Twin_B	16	1550273	G	A	Substitution	downstream_gene_variant	*AL031719.2*	-
Twin_B	16	5534803	T	C	Substitution	intron_variant	*RBFOX1*	-
Twin_B	16	9532674	G	T	Substitution	intergenic_variant	-	rs1341344135
Twin_B	16	18537225	T	TAG	Insertion	downstream_gene_variant	*AC136618.2*	-
Twin_B	16	19362955	T	G	Substitution	intron_variant	*AC130456.2*	rs140765053
Twin_B	16	20424714	G	A	Substitution	TF_binding_site_variant	*ACSM5*	-
Twin_B	16	22677228	G	A	Substitution	intergenic_variant	-	-
Twin_B	16	32008073	G	A	Substitution	downstream_gene_variant	*AC142381.3*	-
Twin_B	16	35787574	T	C	Substitution	upstream_gene_variant	*AC106785.2*	rs1243395959
Twin_B	16	36133454	G	A	Substitution	intergenic_variant	-	-
Twin_B	16	50321074	G	A	Substitution	downstream_gene_variant	*ADCY7*	rs372755442
Twin_B	16	51336582	G	A	Substitution	intron_variant	*AC087564.1*	rs1461056097
Twin_B	16	54256124	G	C	Substitution	intron_variant	*LINC02169*	-
Twin_B	16	57821035	G	A	Substitution	downstream_gene_variant	*AC092118.1*	-
Twin_B	16	64699930	G	A	Substitution	intergenic_variant	-	-
Twin_B	16	65902588	C	T	Substitution	intergenic_variant	-	rs115844235

Twin_B	16	67105628	T	A	Substitution	downstream_gene_variant	*CBFB*	rs1349050758
Twin_B	16	67608941	A	G	Substitution	intron_variant	*CTCF*	rs892105462
Twin_B	16	68548732	C	G	Substitution	intron_variant	*ZFP90*	rs564497116
Twin_B	16	69148663	G	A	Substitution	intron_variant	*UTP4*	rs1469117705
Twin_B	16	70802883	G	T	Substitution	3_prime_UTR_variant	*HYDIN*	-
Twin_B	16	73552032	A	T	Substitution	intron_variant	*ZFHX3*	rs562962066
Twin_B	16	77256710	G	T	Substitution	intron_variant	*ADAMTS18*	-
Twin_B	16	77493285	T	C	Substitution	intergenic_variant	-	-
Twin_B	16	82066056	G	T	Substitution	intron_variant	*HSD17B2*	-
Twin_B	16	82874338	C	T	Substitution	intron_variant	*CDH13*	-
Twin_B	16	85313584	C	A	Substitution	downstream_gene_variant	*AC092275.1*	rs1483404740
Twin_B	17	4511744	T	G	Substitution	intron_variant	*SPNS2*	-
Twin_B	17	4976602	G	T	Substitution	downstream_gene_variant	*AC004771.1*	-
Twin_B	17	9021059	G	T	Substitution	regulatory_region_variant	-	-
Twin_B	17	10363095	A	T	Substitution	intron_variant	*MYH13*	-
Twin_B	17	10978272	C	T	Substitution	intergenic_variant	-	-
Twin_B	17	14670774	C	A	Substitution	intergenic_variant	-	-
Twin_B	17	18385438	G	A	Substitution	downstream_gene_variant	*AL353997.5*	rs1682277
Twin_B	17	27983490	G	A	Substitution	intron_variant	*LINC01992*	rs778201958
Twin_B	17	30755275	A	C	Substitution	intron_variant	*SUZ12P1*	-
Twin_B	17	31863521	T	A	Substitution	splice_region_variant	*UTP6*	rs78181996
Twin_B	17	34811921	G	A	Substitution	intergenic_variant	-	rs1011328968
Twin_B	17	40381174	C	T	Substitution	non_coding_transcript_ exon_variant	*AC080112.2*	-
Twin_B	17	41094327	C	T	Substitution	downstream_gene_variant	*AC100808.1*	-
Twin_B	17	41488518	C	T	Substitution	intron_variant	*KRT36*	-
Twin_B	17	41662731	C	G	Substitution	TF_binding_site_variant	*AC130686.1*	-
Twin_B	17	52148573	C	T	Substitution	intron_variant	*CA10*	-
Twin_B	17	52202053	T	G	Substitution	intergenic_variant	-	-
Twin_B	17	56525449	C	T	Substitution	intergenic_variant	-	-
Twin_B	17	59529075	G	A	Substitution	upstream_gene_variant	*LINC01476*	-
Twin_B	17	63728030	A	G	Substitution	intron_variant	*STRADA*	-
Twin_B	17	64448232	C	A	Substitution	upstream_gene_variant	*MILR1*	-
Twin_B	17	74236263	A	G	Substitution	intron_variant	*TTYH2*	rs1373649197
Twin_B	17	74482732	T	C	Substitution	intron_variant	*CD300A*	rs9900660
Twin_B	17	79561251	G	A	Substitution	intergenic_variant	-	rs924318212
Twin_B	18	524695	A	T	Substitution	intergenic_variant	-	-
Twin_B	18	924915	C	G	Substitution	downstream_gene_variant	*LINC01904*	-
Twin_B	18	8101586	A	C	Substitution	intron_variant	*PTPRM*	-
Twin_B	18	14702966	G	A	Substitution	intron_variant	*MODIFIER*	rs1201652554
Twin_B	18	14791710	G	A	Substitution	intron_variant	*ANKRD30B*	-
Twin_B	18	15293364	T	C	Substitution	intergenic_variant	-	-
Twin_B	18	20715870	G	A	Substitution	intergenic_variant	-	rs1457573477
Twin_B	18	27402273	G	A	Substitution	downstream_gene_variant	*AC068408.1*	rs1054519887
Twin_B	18	29688544	T	C	Substitution	intergenic_variant	-	rs147589821
Twin_B	18	29768107	G	A	Substitution	intergenic_variant	-	rs1280023452
Twin_B	18	30770852	G	A	Substitution	intergenic_variant	-	rs554280292

Twin_B	18	32293060	C	T	Substitution	downstream_gene_variant	*AC015563.2*	-
Twin_B	18	33117524	G	A	Substitution	intron_variant	*CCDC178*	rs138071520
Twin_B	18	33505461	T	C	Substitution	intergenic_variant	-	-
Twin_B	18	43720938	CA	C	Deletion	intergenic_variant	-	-
Twin_B	18	52406485	T	C	Substitution	intron_variant	*DCC*	-
Twin_B	18	56544081	T	C	Substitution	intergenic_variant	-	rs1341586653
Twin_B	18	58043724	G	A	Substitution	regulatory_region_variant	-	-
Twin_B	18	61155354	C	T	Substitution	intergenic_variant	-	-
Twin_B	18	61706959	G	A	Substitution	intron_variant	*AC105094.2*	-
Twin_B	18	62562373	C	A	Substitution	downstream_gene_variant	*ZCCHC2*	-
Twin_B	18	63887009	AG	A	Deletion	intron_variant	*SERPINB2*	-
Twin_B	18	67455918	G	A	Substitution	intergenic_variant	-	-
Twin_B	18	74237595	C	T	Substitution	intron_variant	*AC090398.1*	-
Twin_B	18	75984267	C	G	Substitution	intergenic_variant	-	-
Twin_B	19	8086471	T	A	Substitution	downstream_gene_variant	*FBN3*	rs113759974
Twin_B	19	8183739	C	A	Substitution	intergenic_variant	-	rs946588717
Twin_B	19	12943750	AGCAGAGGCTTAAGGAGGAGGAAGAAGACAAGAAACGCAAAGAGGAGGAG	A	Deletion	frameshift_variant	*CALR*	rs1555760738
Twin_B	19	16774565	G	A	Substitution	intron_variant	*NWD1*	rs577537244
Twin_B	19	17247444	C	A	Substitution	downstream_gene_variant	*USHBP1*	-
Twin_B	19	19549515	C	T	Substitution	downstream_gene_variant	*CILP2*	-
Twin_B	19	22564447	G	A	Substitution	intergenic_variant	-	rs146882932
Twin_B	19	22624713	TG	T	Deletion	downstream_gene_variant	*AC011467.2*	rs1285085532
Twin_B	19	27496424	T	C	Substitution	intergenic_variant	-	rs1568343244
Twin_B	19	28110861	A	C	Substitution	intron_variant	*AC010511.1*	-
Twin_B	19	31656627	G	A	Substitution	upstream_gene_variant	*RNA5SP471*	-
Twin_B	19	31931918	G	C	Substitution	intron_variant	*LINC01837*	-
Twin_B	19	38958665	C	G	Substitution	intron_variant	*FBXO17*	-
Twin_B	19	41492820	G	A	Substitution	downstream_gene_variant	*PCAT19*	-
Twin_B	19	47920579	G	A	Substitution	TF_binding_site_variant	-	rs367708441
Twin_B	19	53377680	A	G	Substitution	downstream_gene_variant	*ZNF525*	-
Twin_B	19	55277604	C	G	Substitution	downstream_gene_variant	*HSPBP1*	-
Twin_B	20	173236	G	A	Substitution	intergenic_variant	-	rs1034386228
Twin_B	20	1061518	G	A	Substitution	intron_variant	*AL110114.1*	-
Twin_B	20	12995529	T	C	Substitution	intron_variant	*LINC01723*	rs1438512520
Twin_B	20	14336990	T	C	Substitution	intron_variant	*FLRT3*	-
Twin_B	20	21629260	C	T	Substitution	intron_variant	*LINC01726*	rs540343176
Twin_B	20	22689135	A	T	Substitution	intron_variant	*AL158175.1*	-
Twin_B	20	28497717	AG	A	Deletion	intergenic_variant	-	rs1434446279
Twin_B	20	29525288	A	AC	Insertion	intergenic_variant	-	rs1188254666
Twin_B	20	30212896	T	G	Substitution	downstream_gene_variant	*CFTRP3*	-
Twin_B	20	30755746	C	T	Substitution	intergenic_variant	-	-
Twin_B	20	39854921	G	A	Substitution	intergenic_variant	-	rs1346211420
Twin_B	20	48547263	G	A	Substitution	intergenic_variant	-	-

Twin_B	20	50715766	T	A	Substitution	intergenic_variant	-	-
Twin_B	20	50895064	C	T	Substitution	intron_variant	*ADNP*	-
Twin_B	20	52629042	C	G	Substitution	intron_variant	*LINC01524*	-
Twin_B	20	64283610	C	G	Substitution	intron_variant	*PCMTD2*	rs1428268334
Twin_B	21	8020747	T	A	Substitution	intergenic_variant	-	rs1443418402
Twin_B	21	9928024	G	A	Substitution	intergenic_variant	-	rs1342179447
Twin_B	21	14455201	C	A	Substitution	intergenic_variant	-	-
Twin_B	21	24993236	A	G	Substitution	intron_variant	*LINC01692*	-
Twin_B	21	36182754	A	G	Substitution	intron_variant	*DOP1B*	-
Twin_B	21	40180743	A	T	Substitution	intron_variant	*DSCAM*	-
Twin_B	21	40558302	G	T	Substitution	intron_variant	*DSCAM*	-
Twin_B	21	46037210	G	A	Substitution	non_coding_transcript_ exon_variant	*AP001476.1*	-
Twin_B	22	11356944	CT	C	Deletion	intergenic_variant	-	rs1322963034
Twin_B	22	16847213	CTTAA	C	Deletion	downstream_gene_variant	*AC007064.4*	-
Twin_B	22	28259506	G	A	Substitution	intron_variant	*TTC28*	-
Twin_B	22	30635421	T	C	Substitution	TF_binding_site_variant	-	-
Twin_B	22	32657960	T	G	Substitution	intron_variant	*SYN3*	-
Twin_B	22	34618472	G	A	Substitution	intergenic_variant	-	rs867876389
Twin_B	22	34793591	T	C	Substitution	intron_variant	*LINC02885*	-
Twin_B	22	36709756	C	T	Substitution	intron_variant	*AL049749.1*	rs567131721
Twin_B	22	37961231	A	G	Substitution	intron_variant	*POLR2F*	-
Twin_B	22	42966759	C	A	Substitution	intron_variant	*PACSIN2*	-
Twin_B	22	43588922	C	T	Substitution	intron_variant	*EFCAB6*	rs567321760
Twin_B	22	48568894	G	C	Substitution	intron_variant	*TAFA5*	-
Twin_B	X	1604020	G	T	Substitution	downstream_gene_variant	*AKAP17A*	rs560737268
Twin_B	X	12284898	G	A	Substitution	intron_variant	*FRMPD4*	COSV66220435
Twin_B	X	17501537	C	A	Substitution	intron_variant	*NHS*	-
Twin_B	X	19248964	C	A	Substitution	intergenic_variant	-	-
Twin_B	X	24411967	C	T	Substitution	intergenic_variant	-	-
Twin_B	X	35556794	T	C	Substitution	intergenic_variant	-	-
Twin_B	X	35907670	T	C	Substitution	intergenic_variant	-	-
Twin_B	X	62714059	C	T	Substitution	intergenic_variant	-	rs143334744
Twin_B	X	70098264	T	A	Substitution	intergenic_variant	-	-
Twin_B	X	73335262	T	C	Substitution	intergenic_variant	-	-
Twin_B	X	78738428	T	A	Substitution	intergenic_variant	-	-
Twin_B	X	80869548	G	C	Substitution	intergenic_variant	-	-
Twin_B	X	87003014	A	G	Substitution	intergenic_variant	-	-
Twin_B	X	92949595	G	A	Substitution	intergenic_variant	-	-
Twin_B	X	107445958	C	T	Substitution	intergenic_variant	-	rs1441643081
Twin_B	X	108369773	T	A	Substitution	intron_variant	*COL4A6*	-
Twin_B	X	113433231	T	G	Substitution	intergenic_variant	-	-
Twin_B	X	115105220	C	T	Substitution	intergenic_variant	-	-
Twin_B	X	117525429	C	A	Substitution	intergenic_variant	-	-

Note: Gene annotation not provided for intergenic variants, regulatory region variants and TF binding site variants. Catalogue corresponds to the high-confidence call set prior to vafCorrect genotyping and filtering

(see Chapter 2 Materials and Methods/2.17 Bioinformatic analysis/2.17.3 Whole genome sequencing analysis/2.17.3.2 Variant calling). Gene names are provided in italics in accordance with the current HUGO Gene Nomenclature Committee (HGNC) guidelines(173). *An additional variant from the myeloid-gene panel is also included for Twin A (marked with an asterisk). This is a variant detected in the targeted myeloid-gene panel sequencing but with no high-confidence support from whole genome sequencing data; therefore, it was excluded from the downstream analysis.

B.5 Catalogue of reagents

Reagent	Manufacturer	Cat. No. / ID	Application
Agencourt AMPure XP Magnetic Beads	Beckman Coulter Inc.	A63881	Twin study
Agilent DNA 7500 Kit	Agilent	5067-1506	Twin study, CH
Agilent DNA 7500 Reagents	Agilent	5067-1507	Twin study, CH
Agilent dsDNA 905 Reagent kit (1-500bp)	Agilent	DNF-905-K0500	Twin study, CH
Agilent High Sensitivity D1000 Reagents	Agilent	5067-5585	Twin study, CH
Agilent High Sensitivity D1000 ScreenTape	Agilent	5067-5584	Twin study, CH
Agilent High Sensitivity D5000 Reagents	Agilent	5067-5593	Twin study, CH
Agilent High Sensitivity D5000 ScreenTape	Agilent	5067-5592	CH / NGS
Automated Droplet Generation Oil for Probes	Bio-Rad	186-4110	Twin study, CH
ddPCR 96-Well Plates (semi-skirted)	Bio-Rad	12001925	Twin study, CH
ddPCR Buffer Control for probes	Bio-Rad	186-3052	Twin study, CH
ddPCR supermix for probes (No dUTP)	Bio-Rad	186-3025	Twin study, CH
Citrate-dextrose solution (ACD)	MilliporeSigma^	C3821-50ML	CH
CD3 MicroBeads human	Miltenyi Biotec	130-050-101	Twin study, CH / cell isolation
CD34 MicroBead kit UltraPure human	Miltenyi Biotec	130-100-453	Twin study, CH / cell isolation
CD42b Mouse anti-Human, Unlabeled, Clone: HIP1	BD Biosciences§	555471	CH
CD45 Purified Mouse anti-Human, Clone: HI30	BD Biosciences§	555480	CH
DG32 Automated Droplet Generator Cartridges	Bio-Rad	186-4108	Twin study, CH
DMSO	MilliporeSigma^	67-68-5	Twin study, CH / cell isolation
DNase I, Bovine Pancreas	MilliporeSigma^	260913-25MU	Twin study, CH
DNeasy Blood & Tissue Kit	QIAGEN	69506	Twin study, CH
Dynabeads™ Pan Mouse IgG	Invitrogen#	11041	CH
EDTA	Corning Inc.	46-034-CI	Twin study, CH / cell isolation, flow cytometry
EvoScript Universal cDNA Master	Roche (F. Hoffmann-La Roche AG)	07912439001	CH
FcR Blocking Reagent, human	Miltenyi Biotec	130-059-901	Twin study, CH/ purity studies
Ficoll-Paque Premium	GE Healthcare	17-5442-03	Twin study, CH / cell isolation
FBS	MilliporeSigma^	F9665-500ML	Twin study, CH
IMDM (1X)	Gibco#	21056-023	Twin study, CH
Monarch DNA Gel Extraction Kit	New England Biolabs	T1020L	Twin study
NextSeq 500/550 Mid Output Kit v2.5 (150 cycles)	Illumina, Inc.	20024904	CH
PBS, pH 7.4	Life Technologies#	10010-056	Twin study, CH
Prostaglandin I2 (sodium salt)	Chemical Chemical Company	CAYM18220-1	CH
QIAamp DNA Micro kit	QIAGEN	56304	Twin study
Qubit 1X dsDNA HS Assay Kit	Invitrogen#	Q22321	Twin study, CH
Qubit dsDNA BR Assay Kit	Invitrogen#	Q32853	Twin study, CH
RBC Lysis Buffer (10X)	eBioscience#	00-4300-54	Twin study, CH / cell isolation
REPLI-g Midi	QIAGEN	150045	Twin study
REPLI-g Mini	QIAGEN	150025	Twin study
RNeasy Mini Kit	QIAGEN	74106	CH
TaqMan Fast Advanced Master Mix	Applied Biosystems#	4444557	CH / PLT purity studies
TaqMan Gene Expression Assay	Applied Biosystems#	4331182	CH / PLT purity studies
TaqMan Gene Expression Assay (*APOC1*)	Applied Biosystems#	4331182 / Hs00155790_m1	CH / PLT purity studies

TaqMan Gene Expression Assay (*CD3G*)	Applied Biosystems#	4331182 / Hs00962186_m1	CH / PLT purity studies
TaqMan Gene Expression Assay (*CD79A*)	Applied Biosystems#	4331182 / Hs00233566_m1	CH / PLT purity studies
TaqMan Gene Expression Assay (*CSF3R*)	Applied Biosystems#	4331182 / Hs01114420_m1	CH / PLT purity studies
TaqMan Gene Expression Assay (*ELANE*)	Applied Biosystems#	4331182 / Hs00236952_m1	CH / PLT purity studies
TaqMan Gene Expression Assay (*GAPDH*)	Applied Biosystems#	4331182 / Hs02786624_g1	CH / PLT purity studies
TaqMan Gene Expression Assay (*KLF1*)	Applied Biosystems#	4331182 / Hs00610592_m1	CH / PLT purity studies
TaqMan Gene Expression Assay (*PF4*)	Applied Biosystems#	4331182 / Hs00427220_g1	CH / PLT purity studies
TaqMan Gene Expression Assay (*SELP*)	Applied Biosystems#	4331182 / Hs00927900_m1	CH / PLT purity studies
TaqMan Gene Expression Assay (*TGFB1*)	Applied Biosystems#	4331182 / Hs00998133_m1	CH / PLT purity studies
TaqMan Gene Expression Assay (*TMOD*)	Applied Biosystems#	4331182 / Hs00268513_m1	CH / PLT purity studies
Richard-Allan Scientific Three-Step Stain Set	Thermo Fisher Scientific	22-050-272	Twin study, CH / PLT purity studies
Twist Binding and Purification Beads	Twist Bioscience	100983	CH / NGS
Twist Custom Panel	Twist Bioscience	101001 (ID: TE-91729386)	CH / NGS
Twist Equinox Library amp mix	Twist Bioscience	104107; 1041078	CH / NGS
Twist Hybridization Reagents	Twist Bioscience	100930	CH / NGS
Twist Library Preparation Kit 1	Twist Bioscience	100572	CH / NGS
Twist Library Preparation Kit 2	Twist Bioscience	100573	CH / NGS
Twist Universal Adapter System – TruSeq Compatible	Twist Bioscience	101308	CH / NGS
Twist Universal Blockers	Twist Bioscience	100578	CH / NGS
Twist Wash Buffers	Twist Bioscience	100985	CH / NGS
Trypan Blue solution	MilliporeSigma^	T8154-100ML	Twin study, CH / cell isolation
Water, cell culture grade	MilliporeSigma	W3500-6X500ML	Twin study, CH

Note: The list is alphabetised by reagent name. Flow cytometry antibodies are provided separately in Table 2-3 (Chapter 2 Materials and Methods/2.15 Flow cytometry and cell sorting). ^Merck; §Becton, Dickinson and Company; #Thermo Fisher Scientific Inc.; Cat. No., catalog number; CH, clonal haematopoiesis; DNA, Deoxyribonucleic acid; ddPCR, droplet digital PCR; DMSO, dimethyl sulfoxide; EDTA, ethylenediaminetetraacetic acid; FBS, fetal bovine serum; FcR, Fc receptor; HSPC, haematopoietic stem/progenitor cell; IMDM, Iscove's Modified Dulbecco's Medium, NGS, next-generation sequencing; PBS, phosphate-buffered saline; PLT, platelet; RBC, red blood cell.

Bibliography

1. Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. Nature. 2018;553(7689):418-26.

2. Psaila B, Mead AJ. Single-cell approaches reveal novel cellular pathways for megakaryocyte and erythroid differentiation. Blood. 2019;133(13):1427-35.

3. Notta F, Zandi S, Takayama N, Dobson S, Gan OI, Wilson G, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science. 2016;351(6269):aab2116.

4. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol. 2017;19(4):271-81.

5. Tusi BK, Wolock SL, Weinreb C, Hwang Y, Hidalgo D, Zilionis R, et al. Population snapshots predict early haematopoietic and erythroid hierarchies. Nature. 2018;555(7694):54-60.

6. Adolfsson J, Månsson R, Buza-Vidas N, Hultquist A, Liuba K, Jensen CT, et al. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell. 2005;121(2):295-306.

7. Benz C, Copley MR, Kent DG, Wohrer S, Cortes A, Aghaeepour N, et al. Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. Cell Stem Cell. 2012;10(3):273-83.

8. Carrelha J, Meng Y, Kettyle LM, Luis TC, Norfo R, Alcolea V, et al. Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. Nature. 2018;554(7690):106-11.

9. Rodriguez-Fraticelli AE, Wolock SL, Weinreb CS, Panero R, Patel SH, Jankovic M, et al. Clonal analysis of lineage fate in native haematopoiesis. Nature. 2018;553(7687):212-6.

10. Sanjuan-Pla A, Macaulay IC, Jensen CT, Woll PS, Luis TC, Mead A, et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nature. 2013;502(7470):232-6.

11. Shin JY, Hu W, Naramura M, Park CY. High c-Kit expression identifies hematopoietic stem cells with impaired self-renewal and megakaryocytic bias. J Exp Med. 2014;211(2):217-31.

12. Yamamoto R, Morita Y, Ooehara J, Hamanaka S, Onodera M, Rudolph KL, et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell. 2013;154(5):1112-26.

13. Roch A, Trachsel V, Lutolf MP. Brief Report: Single-Cell Analysis Reveals Cell Division-Independent Emergence of Megakaryocytes From Phenotypic Hematopoietic Stem Cells. Stem Cells. 2015;33(10):3152-7.

14. Mandel J, Casari M, Stepanyan M, Martyanov A, Deppermann C. Beyond Hemostasis: Platelet Innate Immune Interactions and Thromboinflammation. Int J Mol Sci. 2022;23(7).

15. Haas S, Hansson J, Klimmeck D, Loeffler D, Velten L, Uckelmann H, et al. Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors. Cell Stem Cell. 2015;17(4):422-34.

16. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153(6):1194-217.

17. Yang D, de Haan G. Inflammation and Aging of Hematopoietic Stem Cells in Their Niche. Cells. 2021;10(8).

18. Groarke EM, Young NS. Aging and Hematopoiesis. Clin Geriatr Med. 2019;35(3):285-93.

19. Pang WW, Price EA, Sahoo D, Beerman I, Maloney WJ, Rossi DJ, et al. Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. Proc Natl Acad Sci U S A. 2011;108(50):20012-7.

20. Kuranda K, Vargaftig J, de la Rochere P, Dosquet C, Charron D, Bardin F, et al. Age-related changes in human hematopoietic stem/progenitor cells. Aging Cell. 2011;10(3):542-6.

21. Dorshkind K, Montecino-Rodriguez E, Signer RA. The ageing immune system: is it ever too old to become young again? Nat Rev Immunol. 2009;9(1):57-62.

22. Grover A, Sanjuan-Pla A, Thongjuea S, Carrelha J, Giustacchini A, Gambardella A, et al. Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. Nat Commun. 2016;7:11075.

23. Yamamoto R, Wilkinson AC, Ooehara J, Lan X, Lai C-Y, Nakauchi Y, et al. Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment. Cell stem cell. 2018;22(4):600-7.e4.

24. Signer RA, Montecino-Rodriguez E, Witte ON, McLaughlin J, Dorshkind K. Age-related defects in B lymphopoiesis underlie the myeloid dominance of adult leukemia. Blood. 2007;110(6):1831-9.

25. Akunuru S, Geiger H. Aging, Clonality, and Rejuvenation of Hematopoietic Stem Cells. Trends Mol Med. 2016;22(8):701-12.

26. Frisch BJ, Hoffman CM, Latchney SE, LaMere MW, Myers J, Ashton J, et al. Aged marrow macrophages expand platelet-biased hematopoietic stem cells via Interleukin1B. JCI Insight. 2019;5(10):e124213.

27. Guo BB, Allcock RJ, Mirzai B, Malherbe JA, Choudry FA, Frontini M, et al. Megakaryocytes in Myeloproliferative Neoplasms Have Unique Somatic Mutations. Am J Pathol. 2017;187(7):1512-22.

28. Psaila B, Wang G, Rodriguez-Meira A, Li R, Heuston EF, Murphy L, et al. Single-Cell Analyses Reveal Megakaryocyte-Biased Hematopoiesis in Myelofibrosis and Identify Mutant Clone-Specific Targets. Mol Cell. 2020;78(3):477-92.e8.

29. Osorio FG, Rosendahl Huber A, Oka R, Verheul M, Patel SH, Hasaart K, et al. Somatic Mutations Reveal Lineage Relationships and Age-Related Mutagenesis in Human Hematopoiesis. Cell Rep. 2018;25(9):2308-16.e4.

30. Mitchell E, Spencer Chapman M, Williams N, Dawson KJ, Mende N, Calderbank EF, et al. Clonal dynamics of haematopoiesis across the human lifespan. Nature. 2022;606(7913):343-50.

31. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014;371(26):2477-87.

32. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371(26):2488-98.

33. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nature medicine. 2014;20(12):1472-8.

34. Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood. 2017;130(6):742-52.

35. Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood. 2015;126(1):9-16.

36. Busque L, Buscarlet M, Mollica L, Levine RL. Concise Review: Age-Related Clonal Hematopoiesis: Stem Cells Tempting the Devil. Stem Cells. 2018;36(9):1287-94.

37. Robertson NA, Latorre-Crespo E, Terradas-Terradas M, Lemos-Portela J, Purcell AC, Livesey BJ, et al. Longitudinal dynamics of clonal hematopoiesis identifies gene-specific fitness effects. Nat Med. 2022;28(7):1439-46.

38. van Deuren RC, Andersson-Assarsson JC, Kristensson FM, Steehouwer M, Sjöholm K, Svensson P-A, et al. Expansion of mutation-driven haematopoietic clones is associated with insulin resistance and low HDL-cholesterol in individuals with obesity. bioRxiv. 2021;2021.05.12.443095.

39. von Beck K, von Beck T, Ferrell PB, Jr., Bick AG, Kishtagari A. Lymphoid clonal hematopoiesis: implications for malignancy, immunity, and treatment. Blood Cancer J. 2023;13(1):5.

40. Niroula A, Sekar A, Murakami MA, Trinder M, Agrawal M, Wong WJ, et al. Distinction of lymphoid and myeloid clonal hematopoiesis. Nat Med. 2021;27(11):1921-7.

41. Buscarlet M, Provost S, Zada YF, Barhdadi A, Bourgoin V, Lépine G, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. Blood. 2017;130(6):753-62.

42. Arends CM, Galan-Sousa J, Hoyer K, Chan W, Jäger M, Yoshida K, et al. Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. Leukemia. 2018;32(9):1908-19.

43. Ostrander EL, Koh WK, Mallaney C, Kramer AC, Wilson WC, Zhang B, et al. The GNAS(R201C) mutation associated with clonal hematopoiesis supports transplantable hematopoietic stem cell activity. Exp Hematol. 2018;57:14-20.

44. Watson CJ, Papula AL, Poon GYP, Wong WH, Young AL, Druley TE, et al. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. Science. 2020;367(6485):1449-54.

45. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nature Communications. 2016;7(1):12484.

46. Buscarlet M, Provost S, Zada YF, Bourgoin V, Mollica L, Dubé MP, et al. Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. Blood. 2018;132(3):277-80.

47. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat Genet. 2012;44(6):642-50.

48. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. 2012;44(6):651-8.

49. Forsberg LA, Rasi C, Razzaghian HR, Pakalapati G, Waite L, Thilbeault KS, et al. Age-related somatic structural changes in the nuclear genome of human blood cells. Am J Hum Genet. 2012;90(2):217-28.

50. Acuna-Hidalgo R, Sengul H, Steehouwer M, van de Vorst M, Vermeulen SH, Kiemeney LALM, et al. Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. Am J Hum Genet. 2017;101(1):50-64.

51. McKerrell T, Park N, Moreno T, Grove CS, Ponstingl H, Stephens J, et al. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. Cell Rep. 2015;10(8):1239-45.

52. Acuna-Hidalgo R, Sengul H, Steehouwer M, van de Vorst M, Vermeulen SH, Kiemeney L, et al. Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. Am J Hum Genet. 2017;101(1):50-64.

53. Kar SP, Quiros PM, Gu M, Jiang T, Mitchell J, Langdon R, et al. Genome-wide analyses of 200,453 individuals yield new insights into the causes and consequences of clonal hematopoiesis. Nat Genet. 2022;54(8):1155-66.

54. Bolton KL, Ptashkin RN, Gao T, Braunstein L, Devlin SM, Kelly D, et al. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. Nat Genet. 2020;52(11):1219-26.

55. Weber-Lassalle K, Ernst C, Reuss A, Möllenhoff K, Baumann K, Jackisch C, et al. Clonal Hematopoiesis-Associated Gene Mutations in a Clinical Cohort of 448 Patients With Ovarian Cancer. J Natl Cancer Inst. 2022;114(4):565-70.

56. Kwan TT, Oza AM, Tinker AV, Ray-Coquard I, Oaknin A, Aghajanian C, et al. Preexisting TP53-Variant Clonal Hematopoiesis and Risk of Secondary Myeloid Neoplasms in Patients With High-grade Ovarian Cancer Treated With Rucaparib. JAMA Oncol. 2021;7(12):1772-81.

57. Coombs CC, Zehir A, Devlin SM, Kishtagari A, Syed A, Jonsson P, et al. Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. Cell Stem Cell. 2017;21(3):374-82.e4.

58. Abegunde SO, Buckstein R, Wells RA, Rauh MJ. An inflammatory environment containing TNFα favors Tet2-mutant clonal hematopoiesis. Exp Hematol. 2018;59:60-5.

59. van Kamp H, Landegent JE, Jansen RP, Willemze R, Fibbe WE. Clonal hematopoiesis in patients with acquired aplastic anemia. Blood. 1991;78(12):3209-14.

60. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. N Engl J Med. 2017;377(2):111-21.

61. Jaiswal S, Ebert BL. Clonal hematopoiesis in human aging and disease. Science. 2019;366(6465).

62. Zekavat SM, Lin SH, Bick AG, Liu A, Paruchuri K, Wang C, et al. Hematopoietic mosaic chromosomal alterations increase the risk for diverse types of infection. Nat Med. 2021;27(6):1012-24.

63. Kessler MD, Damask A, O'Keeffe S, Banerjee N, Li D, Watanabe K, et al. Common and rare variant associations with clonal haematopoiesis phenotypes. Nature. 2022;612(7939):301-9.

64. Lin AE, Rauch PJ, Jaiswal S, Ebert BL. Clonal Hematopoiesis: Confluence of Malignant and Nonmalignant Diseases. Annu Rev Cancer Biol. 2022;6(1):187-200.

65. Severson EA, Riedlinger GM, Connelly CF, Vergilio JA, Goldfinger M, Ramkissoon S, et al. Detection of clonal hematopoiesis of indeterminate potential in clinical sequencing of solid tumor specimens. Blood. 2018;131(22):2501-5.

66. Chan HT, Nagayama S, Chin YM, Otaki M, Hayashi R, Kiyotani K, et al. Clinical significance of clonal hematopoiesis in the interpretation of blood liquid biopsy. Mol Oncol. 2020;14(8):1719-30.

67. Antonarakis ES, Kuang Z, Tukachinsky H, Parachoniak C, Kelly AD, Gjoerup O, et al. Prevalence of inferred clonal hematopoiesis (CH) detected on comprehensive

genomic profiling (CGP) of solid tumor tissue or circulating tumor DNA (ctDNA). J Clin Oncol. 2021;39(15_suppl):3009.

68. Zhang Y, Yao Y, Xu Y, Li L, Gong Y, Zhang K, et al. Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. Nat Commun. 2021;12(1):11.

69. Marshall CH, Holler AE, Tsai H-L, Gondek L, Luo J, Antonarakis ES. Clonal hematopoiesis in prostate cancer inferred from somatic tumor profiling. J Clin Oncol. 2021;39(15_suppl):e17001.

70. Arends CM, Dimitriou S, Stahler A, Hablesreiter R, Strzelecka PM, Stein CM, et al. Clonal hematopoiesis is associated with improved survival in patients with metastatic colorectal cancer from the FIRE-3 trial. Blood. 2022;139(10):1593-7.

71. Coombs CC, Gillis NK, Tan X, Berg JS, Ball M, Balasis ME, et al. Identification of Clonal Hematopoiesis Mutations in Solid Tumor Patients Undergoing Unpaired Next-Generation Sequencing Assays. Clin Cancer Res. 2018;24(23):5918-24.

72. Li Z, Huang W, Yin JC, Na C, Wu X, Shao Y, et al. Comprehensive next-generation profiling of clonal hematopoiesis in cancer patients using paired tumor-blood sequencing for guiding personalized therapies. Clin Transl Med. 2020;10(7):e222.

73. Miller PG, Qiao D, Rojas-Quintero J, Honigberg MC, Sperling AS, Gibson CJ, et al. Association of clonal hematopoiesis with chronic obstructive pulmonary disease. Blood. 2022;139(3):357-68.

74. Kim PG, Niroula A, Shkolnik V, McConkey M, Lin AE, Słabicki M, et al. Dnmt3a-mutated clonal hematopoiesis promotes osteoporosis. J Exp Med. 2021;218(12):e20211872.

75. Vlasschaert C, McNaughton AJM, Chong M, Cook EK, Hopman W, Kestenbaum B, et al. Association of Clonal Hematopoiesis of Indeterminate Potential with Worse Kidney Function and Anemia in Two Cohorts of Patients with Advanced Chronic Kidney Disease. J Am Soc Nephrol. 2022;33(5):985-95.

76. Bouzid H, Belk JA, Jan M, Qi Y, Sarnowski C, Wirth S, et al. Clonal hematopoiesis is associated with protection from Alzheimer's disease. Nat Med. 2023;29(7):1662-70.

77. Dorsheimer L, Assmus B, Rasper T, Ortmann CA, Ecke A, Abou-El-Ardat K, et al. Association of Mutations Contributing to Clonal Hematopoiesis With Prognosis in Chronic Ischemic Heart Failure. JAMA Cardiol. 2019;4(1):25-33.

78. Pascual-Figal DA, Bayes-Genis A, Diez-Diez M, Hernandez-Vicente A, Vazquez-Andres D, de la Barrera J, et al. Clonal Hematopoiesis and Risk of Progression of Heart Failure With Reduced Left Ventricular Ejection Fraction. J Am Coll Cardiol. 2021;77(14):1747-59.

79. Assmus B, Cremer S, Kirschbaum K, Culmann D, Kiefer K, Dorsheimer L, et al. Clonal haematopoiesis in chronic ischaemic heart failure: prognostic role of clone size for DNMT3A- and TET2-driver gene mutations. Eur Heart J. 2021;42(3):257-65.

80. Fuster JJ, MacLauchlan S, Zuriaga MA, Polackal MN, Ostriker AC, Chakraborty R, et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. Science. 2017;355(6327):842-7.

81. Nakao T, Natarajan P. Clonal hematopoiesis, multi-omics and coronary artery disease. Nat Cardiovasc Res. 2022;1(11):965-7.

82. Natarajan P. Genomic Aging, Clonal Hematopoiesis, and Cardiovascular Disease. Arterioscler Thromb Vasc Biol. 2023;43(1):3-14.

83. Bick AG, Weinstock JS, Nandakumar SK, Fulco CP, Bao EL, Zekavat SM, et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. Nature. 2020;586(7831):763-8.

84. Abplanalp WT, Mas-Peiro S, Cremer S, John D, Dimmeler S, Zeiher AM. Association of Clonal Hematopoiesis of Indeterminate Potential With Inflammatory Gene Expression in Patients With Severe Degenerative Aortic Valve Stenosis or Chronic Postischemic Heart Failure. JAMA Cardiol. 2020;5(10):1170-5.

85. Fidler TP, Xue C, Yalcinkaya M, Hardaway B, Abramowicz S, Xiao T, et al. The AIM2 inflammasome exacerbates atherosclerosis in clonal haematopoiesis. Nature. 2021;592(7853):296-301.

86. Moon I, Kong MG, Ji YS, Kim SH, Park SK, Suh J, et al. Clinical, Mutational, and Transcriptomic Characteristics in Elderly Korean Individuals With Clonal Hematopoiesis Driver Mutations. Ann Lab Med. 2023;43(2):145-52.

87. Wolach O, Sellar RS, Martinod K, Cherpokova D, McConkey M, Chappell RJ, et al. Increased neutrophil extracellular trap formation promotes thrombosis in myeloproliferative neoplasms. Sci Transl Med. 2018;10(436).

88. Shi K, Zhao W, Chen Y, Ho WT, Yang P, Zhao ZJ. Cardiac hypertrophy associated with myeloproliferative neoplasms in JAK2V617F transgenic mice. J Hematol Oncol. 2014;7:25.

89. Landolfi R, Marchioli R, Kutti J, Gisslinger H, Tognoni G, Patrono C, et al. Efficacy and safety of low-dose aspirin in polycythemia vera. N Engl J Med. 2004;350(2):114-24.

90. Kearon C, Ageno W, Cannegieter SC, Cosmi B, Geersing GJ, Kyrle PA. Categorization of patients as having provoked or unprovoked venous thromboembolism: guidance from the SSC of ISTH. J Thromb Haemost. 2016;14(7):1480-3.

91. Chang WT, Chang CL, Ho CH, Hong CS, Wang JJ, Chen ZC. Long-Term Effects of Unprovoked Venous Thromboembolism on Mortality and Major Cardiovascular Events. J Am Heart Assoc. 2017;6(5):e005466.

92. Carrier M, Le Gal G, Wells PS, Fergusson D, Ramsay T, Rodger MA. Systematic Review: The Trousseau Syndrome Revisited: Should We Screen Extensively for Cancer in Patients with Venous Thromboembolism? Ann Intern Med. 2008;149(5):323-33.

93. Carrier M, Lazo-Langner A, Shivakumar S, Tagalakis V, Zarychanski R, Solymoss S, et al. Screening for Occult Cancer in Unprovoked Venous Thromboembolism. N Engl J Med. 2015;373(8):697-704.

94. Sørensen HT, Mellemkjær L, Olsen JH, Baron JA. Prognosis of Cancers Associated with Venous Thromboembolism. N Engl J Med. 2000;343(25):1846-50.

95. Bick AG, Pirruccello JP, Griffin GK, Gupta N, Gabriel S, Saleheen D, et al. Genetic Interleukin 6 Signaling Deficiency Attenuates Cardiovascular Risk in Clonal Hematopoiesis. Circulation. 2020;141(2):124-31.

96. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. N Engl J Med. 2017;377(12):1119-31.

97. Svensson EC, Madar A, Campbell CD, He Y, Sultan M, Healey ML, et al. TET2-Driven Clonal Hematopoiesis and Response to Canakinumab: An Exploratory Analysis of the CANTOS Randomized Clinical Trial. JAMA Cardiol. 2022;7(5):521-8.

98. Challen GA, Sun D, Jeong M, Luo M, Jelinek J, Berg JS, et al. Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet. 2011;44(1):23-31.

99. Jeong M, Park HJ, Celik H, Ostrander EL, Reyes JM, Guzman A, et al. Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. Cell Rep. 2018;23(1):1-10.

100. Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell. 2011;20(1):11-24.

101. Quivoron C, Couronné L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell. 2011;20(1):25-38.

102. Steensma DP. Clinical consequences of clonal hematopoiesis of indeterminate potential. Blood Adv. 2018;2(22):3404-10.

103. Abelson S, Collord G, Ng SWK, Weissbrod O, Mendelson Cohen N, Niemeyer E, et al. Prediction of acute myeloid leukaemia risk in healthy individuals. Nature. 2018;559(7714):400-4.

104. Desai P, Mencia-Trinchant N, Savenkov O, Simon MS, Cheang G, Lee S, et al. Somatic mutations precede acute myeloid leukemia years before diagnosis. Nat Med. 2018;24(7):1015-23.

105. Warren JT, Link DC. Clonal hematopoiesis and risk for hematologic malignancy. Blood. 2020;136(14):1599-605.

106. Gibson CJ, Lindsley RC, Tchekmedyian V, Mar BG, Shi J, Jaiswal S, et al. Clonal Hematopoiesis Associated With Adverse Outcomes After Autologous Stem-Cell Transplantation for Lymphoma. J Clin Oncol. 2017;35(14):1598-605.

107. van Zeventer IA, de Graaf AO, Salzbrunn JB, Nolte IM, Kamphuis P, Dinmohamed A, et al. Evolutionary landscape of clonal hematopoiesis in 3,359 individuals from the general population. 2023;41(6):1017-31.e4.

108. Malcovati L, Gallì A, Travaglino E, Ambaglio I, Rizzo E, Molteni E, et al. Clinical significance of somatic mutation in unexplained blood cytopenia. Blood. 2017;129(25):3371-8.

109. Weeks LD, Niroula A, Neuberg DS, Wong WJ, Lindsley RC, Luskin MR, et al. Prediction of Risk for Myeloid Malignancy in Clonal Hematopoiesis. Blood. 2022;140(Supplement 1):2229-31.

110. Rothenberg-Thurley M, Amler S, Goerlich D, Köhnke T, Konstandin NP, Schneider S, et al. Persistence of pre-leukemic clones during first remission and risk of relapse in acute myeloid leukemia. Leukemia. 2017; doi: 10.1038/leu.2017.350. Epub ahead of print.

111. Morita K, Kantarjian HM, Wang F, Yan Y, Bueso-Ramos C, Sasaki K, et al. Clearance of Somatic Mutations at Remission and the Risk of Relapse in Acute Myeloid Leukemia. J Clin Oncol. 2018;36(18):1788-97.

112. Hirsch P, Tang R, Abermil N, Flandrin P, Moatti H, Favale F, et al. Precision and prognostic value of clone-specific minimal residual disease in acute myeloid leukemia. Haematologica. 2017;102(7):1227-37.

113. Berger G, Kroeze LI, Koorenhof-Scheele TN, de Graaf AO, Yoshida K, Ueno H, et al. Early detection and evolution of preleukemic clones in therapy-related myeloid neoplasms following autologous SCT. Blood. 2018;131(16):1846-57.

114. Chitre S, Stölzel F, Cuthill K, Streetly M, Graham C, Dill C, et al. Clonal hematopoiesis in patients with multiple myeloma undergoing autologous stem cell transplantation. Leukemia. 2018;32(9):2020-4.

115. Gibson CJ, Kennedy JA, Nikiforow S, Kuo FC, Alyea EP, Ho V, et al. Donor-engrafted CHIP is common among stem cell transplant recipients with unexplained cytopenias. Blood. 2017;130(1):91-4.

116. Frick M, Chan W, Arends CM, Hablesreiter R, Halik A, Heuser M, et al. Role of Donor Clonal Hematopoiesis in Allogeneic Hematopoietic Stem-Cell Transplantation. J Clin Oncol. 2019;37(5):375-85.

117. Cook EK, Luo M, Rauh MJ. Clonal hematopoiesis and inflammation: Partners in leukemogenesis and comorbidity. Exp Hematol. 2020;83:85-94.

118. Cook EK, Izukawa T, Young S, Rosen G, Jamali M, Zhang L, et al. Comorbid and inflammatory characteristics of genetic subtypes of clonal hematopoiesis. Blood Adv. 2019;3(16):2482-6.

119. Trowbridge JJ, Starczynowski DT. Innate immune pathways and inflammation in hematopoietic aging, clonal hematopoiesis, and MDS. J Exp Med. 2021;218(7).

120. Yura Y, Sano S, Walsh K. Clonal Hematopoiesis: A New Step Linking Inflammation to Heart Failure. JACC Basic Transl Sci. 2020;5(2):196-207.

121. Cobo I, Tanaka T, Glass CK, Yeang C. Clonal hematopoiesis driven by DNMT3A and TET2 mutations: role in monocyte and macrophage biology and atherosclerotic cardiovascular disease. Curr Opin Hematol. 2022;29(1):1-7.

122. Cull AH, Snetsinger B, Buckstein R, Wells RA, Rauh MJ. Tet2 restrains inflammatory gene expression in macrophages. Exp Hematol. 2017;55:56-70.e13.

123. Pietras EM, Mirantes-Barbeito C, Fong S, Loeffler D, Kovtonyuk LV, Zhang S, et al. Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the expense of self-renewal. Nat Cell Biol. 2016;18(6):607-18.

124. Caiado F, Pietras EM, Manz MG. Inflammation as a regulator of hematopoietic stem cell function in disease, aging, and clonal selection. J Exp Med. 2021;218(7) :e20201541.

125. Muto T, Walker CS, Choi K, Hueneman K, Smith MA, Gul Z, et al. Adaptive response to inflammation contributes to sustained myelopoiesis and confers a competitive advantage in myelodysplastic syndrome HSCs. Nat Immunol. 2020;21(5):535-45.

126. Luque Paz D, Kralovics R, Skoda RC. Genetic basis and molecular profiling in myeloproliferative neoplasms. Blood. 2023;141(16):1909-21.

127. Hormaechea-Agulla D, Matatall KA, Le DT, Kain B, Long X, Kus P, et al. Chronic infection drives Dnmt3a-loss-of-function clonal hematopoiesis via IFNγ signaling. Cell Stem Cell. 2021;28(8):1428-42.e6.

128. Mendez Luque LF, Blackmon AL, Ramanathan G, Fleischman AG. Key Role of Inflammation in Myeloproliferative Neoplasms: Instigator of Disease Initiation, Progression. and Symptoms. Curr Hematol Malig Rep. 2019;14(3):145-53.

129. Langlais BT, Geyer H, Scherber R, Mesa RA, Dueck AC. Quality of life and symptom burden among myeloproliferative neoplasm patients: do symptoms impact quality of life? Leuk Lymphoma. 2019;60(2):402-8.

130. Tremblay D, Mesa R. Addressing symptom burden in myeloproliferative neoplasms. Best Pract Res Clin Haematol. 2022;35(2):101372.

131. DAMESHEK W. Editorial: Some Speculations on the Myeloproliferative Syndromes. Blood. 1951;6(4):372-5.

132. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood. 2009;114(5):937-51.

133. Tefferi A, Thiele J, Vardiman JW. The 2008 World Health Organization classification system for myeloproliferative neoplasms: order out of chaos. Cancer. 2009;115(17):3842-7.

134. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016;127(20):2391-405.

135. Khoury JD, Solary E, Abla O, Akkari Y, Alaggio R, Apperley JF, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. Leukemia. 2022;36(7):1703-19.

136. Mead AJ, Mullally A. Myeloproliferative neoplasm stem cells. Blood. 2017;129(12):1607-16.

137. Vainchenker W, Kralovics R. Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. Blood. 2017;129(6):667-79.

138. Ma X, Vanasse G, Cartmel B, Wang Y, Selinger HA. Prevalence of polycythemia vera and essential thrombocythemia. Am J Hematol. 2008;83(5):359-62.

139. Johansson P. Epidemiology of the myeloproliferative disorders polycythemia vera and essential thrombocythemia. Semin Thromb Hemost. 2006;32(3):171-3.

140. Meier B, Burton JH. Myeloproliferative Disorders. Hematol Oncol Clin North Am. 2017;31(6):1029-44.

141. Moulard O, Mehta J, Fryzek J, Olivares R, Iqbal U, Mesa RA. Epidemiology of myelofibrosis, essential thrombocythemia, and polycythemia vera in the European Union. Eur J Haematol. 2014;92(4):289-97.

142. Titmarsh GJ, Duncombe AS, McMullin MF, O'Rorke M, Mesa R, De Vocht F, et al. How common are myeloproliferative neoplasms? A systematic review and meta-analysis. Am J Hematol. 2014;89(6):581-7.

143. Mehta J, Wang H, Iqbal SU, Mesa R. Epidemiology of myeloproliferative neoplasms in the United States. Leuk Lymphoma. 2014;55(3):595-600.

144. Accurso V, Santoro M, Mancuso S, Napolitano M, Carlisi M, Mattana M, et al. The Essential Thrombocythemia in 2020: What We Know and Where We Still Have to Dig Deep. Clin Med Insights Blood Disord. 2020;13:2634853520978210.

145. Palandri F, Mora B, Gangat N, Catani L. Is there a gender effect in polycythemia vera? Ann Hematol. 2021;100(1):11-25.

146. Verstovsek S, Yu J, Scherber RM, Verma S, Dieyi C, Chen CC, et al. Changes in the incidence and overall survival of patients with myeloproliferative neoplasms between 2002 and 2016 in the United States. Leuk Lymphoma. 2022;63(3):694-702.

147. Hultcrantz M, Ravn Landtblom A, Andréasson B, Samuelsson J, Dickman PW, Kristinsson SY, et al. Incidence of myeloproliferative neoplasms - trends by subgroup and age in a population-based study in Sweden. J Intern Med. 2020;287(4):448-54.

148. Baumeister J, Chatain N, Sofias AM, Lammers T, Koschmieder S. Progression of Myeloproliferative Neoplasms (MPN): Diagnostic and Therapeutic Perspectives. Cells. 2021;10(12):3551.

149. Cervantes F, Tassies D, Salgado C, Rovira M, Pereira A, Rozman C. Acute transformation in nonleukemic chronic myeloproliferative disorders: actuarial probability and main characteristics in a series of 218 patients. Acta Haematol. 1991;85(3):124-7.

150. Noor SJ, Tan W, Wilding GE, Ford LA, Barcos M, Sait SN, et al. Myeloid blastic transformation of myeloproliferative neoplasms--a review of 112 cases. Leuk Res. 2011;35(5):608-13.

151. Tefferi A, Barbui T. Polycythemia vera and essential thrombocythemia: 2021 update on diagnosis, risk-stratification and management. Am J Hematol. 2020;95(12):1599-613.

152. Luque Paz D, Jouanneau-Courville R, Riou J, Ianotto JC, Boyer F, Chauveau A, et al. Leukemic evolution of polycythemia vera and essential thrombocythemia: genomic profiles predict time to transformation. Blood Adv. 2020;4(19):4887-97.

153. Cerquozzi S, Tefferi A. Blast transformation and fibrotic progression in polycythemia vera and essential thrombocythemia: a literature review of incidence and risk factors. Blood Cancer J. 2015;5(11):e366.

154. Gangat N, Caramazza D, Vaidya R, George G, Begna K, Schwager S, et al. DIPSS plus: a refined Dynamic International Prognostic Scoring System for primary myelofibrosis that incorporates prognostic information from karyotype, platelet count, and transfusion status. J Clin Oncol. 2011;29(4):392-7.

155. Mesa RA, Li CY, Ketterling RP, Schroeder GS, Knudson RA, Tefferi A. Leukemic transformation in myelofibrosis with myeloid metaplasia: a single-institution experience with 91 cases. Blood. 2005;105(3):973-7.

156. Kennedy JA, Atenafu EG, Messner HA, Craddock KJ, Brandwein JM, Lipton JH, et al. Treatment outcomes following leukemic transformation in Philadelphia-negative myeloproliferative neoplasms. Blood. 2013;121(14):2725-33.

157. Thepot S, Itzykson R, Seegers V, Raffoux E, Quesnel B, Chait Y, et al. Treatment of progression of Philadelphia-negative myeloproliferative neoplasms to myelodysplastic syndrome or acute myeloid leukemia by azacitidine: a report on 54 cases on the behalf of the Groupe Francophone des Myelodysplasies (GFM). Blood. 2010;116(19):3735-42.

158. Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet. 2005;365(9464):1054-61.

159. Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. N Engl J Med. 2005;352(17):1779-90.

160. Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell. 2005;7(4):387-97.

161. James C, Ugo V, Le Couédic JP, Staerk J, Delhommeau F, Lacout C, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature. 2005;434(7037):1144-8.

162. Scott LM, Beer PA, Bench AJ, Erber WN, Green AR. Prevalance of JAK2 V617F and exon 12 mutations in polycythaemia vera. Br J Haematol. 2007;139(3):511-2.

163. Butcher CM, Hahn U, To LB, Gecz J, Wilkins EJ, Scott HS, et al. Two novel JAK2 exon 12 mutations in JAK2V617F-negative polycythaemia vera patients. Leukemia. 2008; 22(4):870-3.

164. Pikman Y, Lee BH, Mercher T, McDowell E, Ebert BL, Gozo M, et al. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. PLoS Med. 2006;3(7):e270.

165. Pardanani AD, Levine RL, Lasho T, Pikman Y, Mesa RA, Wadleigh M, et al. MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. Blood. 2006;108(10):3472-6.

166. Constantinescu SN, Vainchenker W, Levy G, Papadopoulos N. Functional Consequences of Mutations in Myeloproliferative Neoplasms. Hemasphere. 2021;5(6):e578.

167. Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. N Engl J Med. 2013;369(25):2379-90.

168. Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. N Engl J Med. 2013;369(25):2391-405.

169. Sangkhae V, Etheridge SL, Kaushansky K, Hitchcock IS. The thrombopoietin receptor, MPL, is critical for development of a JAK2V617F-induced myeloproliferative neoplasm. Blood. 2014;124(26):3956-63.

170. Vannucchi AM, Guglielmelli P. Molecular prognostication in Ph-negative MPNs in 2022. Hematology. 2022;2022(1):225-34.

171. Tefferi A, Lasho TL, Finke CM, Knudson RA, Ketterling R, Hanson CH, et al. CALR vs JAK2 vs MPL-mutated or triple-negative myelofibrosis: clinical, cytogenetic and molecular comparisons. Leukemia. 2014;28(7):1472-7.

172. Alimam S, Villiers W, Dillon R, Simpson M, Runglall M, Smith A, et al. Patients with triple-negative, JAK2V617F- and CALR-mutated essential thrombocythemia share a unique gene expression signature. Blood Adv. 2021;5(4):1059-68.

173. Bruford EA, Braschi B, Denny P, Jones TEM, Seal RL, Tweedie S. Guidelines for human gene nomenclature. Nat Genet. 2020;52(8):754-8.

174. Wolman IJ. Parallel responses to chemotherapy in identical twin infants with concordant leukemia. J Pediatr. 1962;60:91-5.

175. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. Blood. 2003;102(7):2321-33.

176. Andersen MA, Bjerrum OW, Ranjan A, Skov V, Kruse TA, Thomassen M, et al. Myeloproliferative Neoplasms in Danish Twins. Acta Haematol. 2018;139(3):195-8.

177. Ma Y, Dobbins SE, Sherborne AL, Chubb D, Galbiati M, Cazzaniga G, et al. Developmental timing of mutations revealed by whole-genome sequencing of twins with acute lymphoblastic leukemia. Proc Natl Acad Sci U S A. 2013;110(18):7429-33.

178. Broadfield ZJ, Hain RD, Harrison CJ, Reza Jalali G, McKinley M, Michalová K, et al. Complex chromosomal abnormalities in utero, 5 years before leukaemia. Br J Haematol. 2004;126(3):307-12.

179. Wiemels JL, Xiao Z, Buffler PA, Maia AT, Ma X, Dicks BM, et al. In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. Blood. 2002;99(10):3801-5.

180. Valdés-Mas R, Gutiérrez-Abril J, Pitiot AS, Santamaría I, Puente DA, Muñiz Lobato S, et al. Transplacental transfer of essential thrombocythemia in monozygotic twins. Blood. 2016;128(14):1894-6.

181. Fabre MA, McKerrell T, Zwiebel M, Vijayabaskar MS, Park N, Wells PM, et al. Concordance for clonal hematopoiesis is limited in elderly twins. Blood. 2020;135(4):269-73.

182. Hansen JW, Pedersen DA, Larsen LA, Husby S, Clemmensen SB, Hjelmborg J, et al. Clonal hematopoiesis in elderly twins: concordance, discordance, and mortality. Blood. 2020;135(4):261-8.

183. Molenaar RJ, Pleyer C, Radivoyevitch T, Sidana S, Godley A, Advani AS, et al. Risk of developing chronic myeloid neoplasms in well-differentiated thyroid cancer patients treated with radioactive iodine. Leukemia. 2018;32(4):952-9.

184. Van Egeren D, Escabi J, Nguyen M, Liu S, Reilly CR, Patel S, et al. Reconstructing the Lineage Histories and Differentiation Trajectories of Individual Cancer Cells in Myeloproliferative Neoplasms. Cell Stem Cell. 2021;28(3):514-23.e9.

185. Williams N, Lee J, Mitchell E, Moore L, Baxter EJ, Hewinson J, et al. Life histories of myeloproliferative neoplasms inferred from phylogenies. Nature. 2022;602(7895):162-8.

186. Sousos N, Ní Leathlobhair M, Simoglou Karali C, Louka E, Bienz N, Royston D, et al. In utero origin of myelofibrosis presenting in adult monozygotic twins. Nat Med. 2022;28(6):1207-11.

187. Waksal JA, Mascarenhas J. Novel Therapies in Myelofibrosis: Beyond JAK Inhibitors. Curr Hematol Malig Rep. 2022;17(5):140-54.

188. How J, Garcia JS, Mullally A. Biology and therapeutic targeting of molecular mechanisms in MPN. Blood. 2022;141(16):1922-33.

189. Waksal JA, Tremblay DA. Contemporary and future strategies in polycythemia vera. Best Pract Res Clin Haematol. 2022;35(2):101370.

190. Godfrey AL, Green A, Harrison CN. Essential thrombocythemia: challenges in clinical practice and future prospects. Blood. 2023;141(16):1943-53.

191. Ballen KK, Woolfrey AE, Zhu X, Ahn KW, Wirk B, Arora M, et al. Allogeneic hematopoietic cell transplantation for advanced polycythemia vera and essential thrombocythemia. Biol Blood Marrow Transplant. 2012;18(9):1446-54.

192. Ballen KK, Shrestha S, Sobocinski KA, Zhang MJ, Bashey A, Bolwell BJ, et al. Outcome of transplantation for myelofibrosis. Biol Blood Marrow Transplant. 2010;16(3):358-67.

193. Kröger NM, Deeg JH, Olavarria E, Niederwieser D, Bacigalupo A, Barbui T, et al. Indication and management of allogeneic stem cell transplantation in primary myelofibrosis: a consensus process by an EBMT/ELN international working group. Leukemia. 2015;29(11):2126-33.

194. Savani M, Dulery R, Bazarbachi AH, Mohty R, Brissot E, Malard F, et al. Allogeneic haematopoietic cell transplantation for myelofibrosis: a real-life perspective. Br J Haematol. 2021;195(4):495-506.

195. Bacigalupo A, Innocenti I, Rossi E, Sora F, Galli E, Autore F, et al. Allogeneic Hemopoietic Stem Cell Transplantation for Myelofibrosis: 2021. Front Immunol. 2021;12:637512.

196. Ali H, Bacigalupo A. 2021 Update on allogeneic hematopoietic stem cell transplant for myelofibrosis: A review of current data and applications on risk stratification and management. Am J Hematol. 2021;96(11):1532-8.

197. McLornan DP, Szydlo R, Robin M, van Biezen A, Koster L, Blok HJP, et al. Outcome of patients with Myelofibrosis relapsing after allogeneic stem cell transplant: a retrospective study by the Chronic Malignancies Working Party of EBMT. Br J Haematol. 2018;182(3):418-22.

198. Wolfe HR, Horwitz ME, Rein LAM. The Use of Allogeneic Hematopoietic Stem Cell Transplantation in Primary Myelofibrosis. J Pers Med. 2022;12(4).

199. Mascarenhas J, Hoffman R. Ruxolitinib: the first FDA approved therapy for the treatment of myelofibrosis. Clin Cancer Res. 2012;18(11):3008-14.

200. Blair HA. Fedratinib: First Approval. Drugs. 2019;79(15):1719-25.

201. Lamb YN. Pacritinib: First Approval. Drugs. 2022;82(7):831-8.

202. Kiladjian JJ. Long-term treatment with interferon alfa for myeloproliferative neoplasms. Lancet Haematol. 2017;4(4):e150-1.

203. Yacoub A, Mascarenhas J, Kosiorek H, Prchal JT, Berenzon D, Baer MR, et al. Pegylated interferon alfa-2a for polycythemia vera or essential thrombocythemia resistant or intolerant to hydroxyurea. Blood. 2019;134(18):1498-509.

204. Abu-Zeinah G, Krichevsky S, Cruz T, Hoberman G, Jaber D, Savage N, et al. Interferon-alpha for treating polycythemia vera yields improved myelofibrosis-free and overall survival. Leukemia. 2021;35(9):2592-601.

205. Mosca M, Hermange G, Tisserand A, Noble R, Marzac C, Marty C, et al. Inferring the dynamics of mutated hematopoietic stem and progenitor cells induced by IFNα in myeloproliferative neoplasms. Blood. 2021;138(22):2231-43.

206. Kiladjian JJ, Klade C, Georgiev P, Krochmalczyk D, Gercheva-Kyuchukova L, Egyed M, et al. Long-term outcomes of polycythemia vera patients treated with ropeginterferon Alfa-2b. Leukemia. 2022;36(5):1408-11.

207. Verger E, Soret-Dulphy J, Maslah N, Roy L, Rey J, Ghrieb Z, et al. Ropeginterferon alpha-2b targets JAK2V617F-positive polycythemia vera cells in vitro and in vivo. Blood Cancer J. 2018;8(10):94.

208. Kiladjian JJ, Giraudier S, Cassinat B. Interferon-alpha for the therapy of myeloproliferative neoplasms: targeting the malignant clone. Leukemia. 2016;30(4):776-81.

209. Gisslinger H, Klade C, Georgiev P, Krochmalczyk D, Gercheva-Kyuchukova L, Egyed M, et al. Ropeginterferon alfa-2b versus standard therapy for polycythaemia vera (PROUD-PV and CONTINUATION-PV): a randomised, non-inferiority, phase 3 trial and its extension study. Lancet Haematol. 2020;7(3):e196-208.

210. Kiladjian JJ, Barbui T. From leeches to interferon: should cytoreduction be prescribed for all patients with polycythemia vera? Leukemia. 2020;34(11):2837-9.

211. Harrison C, Nangalia J, Boucher RH, Jackson A, Yap C, O'Sullivan J, et al. Ruxolitinib Versus Best Available Therapy for PV Intolerant or Resistant to Hydroxycarbamide in a Randomized Trial. Blood. 2022;140(Supplement 1):1781-3.

212. Verstovsek S, Kantarjian HM, Estrov Z, Cortes JE, Thomas DA, Kadia T, et al. Long-term outcomes of 107 patients with myelofibrosis receiving JAK1/JAK2 inhibitor ruxolitinib: survival advantage in comparison to matched historical controls. Blood. 2012;120(6):1202-9.

213. Tefferi A, Litzow MR, Pardanani A. Long-term outcome of treatment with ruxolitinib in myelofibrosis. N Engl J Med. 2011;365(15):1455-7.

214. Passamonti F, Maffioli M, Cervantes F, Vannucchi AM, Morra E, Barbui T, et al. Impact of ruxolitinib on the natural history of primary myelofibrosis: a comparison of the DIPSS and the COMFORT-2 cohorts. Blood. 2014;123(12):1833-5.

215. Verstovsek S, Gotlib J, Mesa RA, Vannucchi AM, Kiladjian JJ, Cervantes F, et al. Long-term survival in patients treated with ruxolitinib for myelofibrosis: COMFORT-I and -II pooled analyses. J Hematol Oncol. 2017;10(1):156.

216. Verstovsek S, Parasuraman S, Yu J, Shah A, Kumar S, Xi A, et al. Real-world survival of US patients with intermediate- to high-risk myelofibrosis: impact of ruxolitinib approval. Ann Hematol. 2022;101(1):131-7.

217. Thomas JW, Jamy O, Shah MV, Vachhani P, Go RS, Goyal G. Risk of mortality and second malignancies in primary myelofibrosis before and after ruxolitinib approval. Leuk Res. 2022;112:106770.

218. Guglielmelli P, Ghirardi A, Carobbio A, Masciulli A, Maccari C, Mora B, et al. Impact of ruxolitinib on survival of patients with myelofibrosis in the real world: update of the ERNEST Study. Blood Adv. 2022;6(2):373-5.

219. Tefferi A, Barraco D, Lasho TL, Shah S, Begna KH, Al-Kali A, et al. Momelotinib therapy for myelofibrosis: a 7-year follow-up. Blood Cancer J. 2018;8(3):29.

220. Barosi G, Gale RP. Does ruxolitinib really prolong survival in individuals with myelofibrosis? The never-ending story. Blood Adv. 2022;6(7):2331-3.

221. Verstovsek S, Mesa RA, Gotlib J, Levy RS, Gupta V, DiPersio JF, et al. A double-blind, placebo-controlled trial of ruxolitinib for myelofibrosis. N Engl J Med. 2012;366(9):799-807.

222. Harrison C, Kiladjian JJ, Al-Ali HK, Gisslinger H, Waltzman R, Stalbovskaya V, et al. JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. N Engl J Med. 2012;366(9):787-98.

223. Verstovsek S, Kiladjian J-J, Vannucchi AM, Mesa RA, Scherber R, Hamer-Maansson JE, et al. Does Early Intervention in Myelofibrosis Impact Outcomes? a Pooled Analysis of the Comfort I and II Studies. Blood. 2021;138(Supplement 1):1505.

224. Palandri F, Sabattini E, Maffioli M. Treating early-stage myelofibrosis. Ann Hematol. 2019;98(2):241-53.

225. Vannucchi AM, Kantarjian HM, Kiladjian JJ, Gotlib J, Cervantes F, Mesa RA, et al. A pooled analysis of overall survival in COMFORT-I and COMFORT-II, 2 randomized phase III trials of ruxolitinib for the treatment of myelofibrosis. Haematologica. 2015;100(9):1139-45.

226. Barbui T, Thiele J, Gisslinger H, Kvasnicka HM, Vannucchi AM, Guglielmelli P, et al. The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion. Blood Cancer J. 2018;8(2):15.

227. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310(20):2191-4.

228. Valeri CR, Ragno G, Khuri S. Freezing human platelets with 6 percent dimethyl sulfoxide with removal of the supernatant solution before freezing and storage at -80 degrees C without postthaw processing. Transfusion. 2005;45(12):1890-8.

229. Slichter SJ, Jones M, Ransom J, Gettinger I, Jones MK, Christoffel T, et al. Review of in vivo studies of dimethyl sulfoxide cryopreserved platelets. Transfus Med Rev. 2014;28(4):212-25.

230. Yılmaz S, Çetinkaya RA, Eker İ, Ünlü A, Uyanık M, Tapan S, et al. Freezing of Apheresis Platelet Concentrates in 6% Dimethyl Sulfoxide: The First Preliminary Study in Turkey. Turk J Haematol. 2016;33(1):28-33.

231. Pinheiro LB, Coleman VA, Hindson CM, Herrmann J, Hindson BJ, Bhat S, et al. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. Anal Chem. 2012;84(2):1003-11.

232. Rowlands V, Rutkowski AJ, Meuser E, Carr TH, Harrington EA, Barrett JC. Optimisation of robust singleplex and multiplex droplet digital PCR assays for high confidence mutation detection in circulating tumour DNA. Sci Rep. 2019;9(1):12620.

233. Cordua S, Kjaer L, Skov V, Pallisgaard N, Hasselbalch HC, Ellervik C. Prevalence and phenotypes of JAK2 V617F and calreticulin mutations in a Danish general population. Blood. 2019;134(5):469-79.

234. Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, et al. Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res. 2016;44(W1):W147-53.

235. Silveira DRA, Quek L, Santos IS, Corby A, Coelho-Silva JL, Pereira-Martins DA, et al. Integrating clinical features with genetic factors enhances survival prediction for adults with acute myeloid leukemia. Blood Adv. 2020;4(10):2339-50.

236. NextSeq 500 System Guide. San Diego, California 92122 U.S.A.: illumina; October 2021. Report No.: 15046563 v07.

237. Bench AJ, White HE, Foroni L, Godfrey AL, Gerrard G, Akiki S, et al. Molecular diagnosis of the myeloproliferative neoplasms: UK guidelines for the detection of JAK2 V617F and other relevant mutations. Br J Haematol. 2013;160(1):25-34.

238. Uddin MM, Zhou Y, Bick AG, Burugula BB, Jaiswal S, Desai P, et al. Longitudinal profiling of clonal hematopoiesis provides insight into clonal dynamics. Immun Ageing. 2022;19(1):23.

239. Samorodnitsky E, Jewell BM, Hagopian R, Miya J, Wing MR, Lyon E, et al. Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. Hum Mutat. 2015;36(9):903-14.

240. Hung SS, Meissner B, Chavez EA, Ben-Neriah S, Ennishi D, Jones MR, et al. Assessment of Capture and Amplicon-Based Approaches for the Development of a Targeted Next-Generation Sequencing Pipeline to Personalize Lymphoma Management. J Mol Diagn. 2018;20(2):203-14.

241. Library Preparation Protocol - Enzymatic Fragmentation and Twist Universal Adapter System. Twist Bioscience; 07 March 2022. Report No.: 001065 rev 4.0.

242. Library Preparation Protocol - Twist Target Enrichment Standard Hybridization v1 Protocol. Twist Bioscience; 02 February 2021. Report No.: 001085 rev 4.0.

243. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.

244. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. 2016;44(11):e108.

245. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic SNVs and Indels with Mutect2. bioRxiv. 2019;861054.

246. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009;25(21):2865-71.

247. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

248. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812-4.

249. Schischlik F, Jäger R, Rosebrock F, Hug E, Schuster M, Holly R, et al. Mutational landscape of the transcriptome offers putative targets for immunotherapy of myeloproliferative neoplasms. Blood. 2019;134(2):199-210.

250. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17(1):3.

251. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

252. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

253. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.

254. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.

255. Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). Bioinformatics. 2023;39(1):btac830.

256. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29(1):24-6.

257. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47(D1):D941-d7.

258. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28(11):1747-56.

259. Posit team. RStudio: Integrated Development Environment for R. Boston, MA: Posit Software, PBC; 2023.

260. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.

261. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639-45.

262. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods. 2018;15(8):591-4.

263. Cooke DP, Wedge DC, Lunter G. A unified haplotype-based method for accurate and comprehensive variant calling. Nat Biotechnol. 2021;39(7):885-92.

264. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell. 2012;149(5):994-1007.

265. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. Brief Bioinform. 2013;14(2):144-61.

266. Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. Nat Commun. 2014;5:2997.

267. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell. 2017;32(2):169-84.e7.

268. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32(8):1220-2.

269. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15(6):R84.

270. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. Nat Methods. 2015;12(10):966-8.

271. Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, et al. AnnotSV: an integrated tool for structural variations annotation. Bioinformatics. 2018;34(20):3572-4.

272. Bergstrom EN, Huang MN, Mahto U, Barnes M, Stratton MR, Rozen SG, et al. SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. BMC Genomics. 2019;20(1):685.

273. Python Software Foundation. Python Language Reference, version 3.8. Available at http://www.python.org.

274. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016;374(23):2209-21.

275. Nangalia J, Green AR. Myeloproliferative neoplasms: from origins to outcomes. Hematology Am Soc Hematol Educ Program. 2017;2017(1):470-9.

276. Pich O, Reyes-Salazar I, Gonzalez-Perez A, Lopez-Bigas N. Discovering the drivers of clonal hematopoiesis. bioRxiv. 2020;2020.10.22.350140.

277. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.

278. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47(D1):D886-94.

279. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. Nature. 2020;578(7793):122-8.

280. Rustad EH, Yellapantula V, Leongamornlert D, Bolli N, Ledergor G, Nadeu F, et al. Timing the initiation of multiple myeloma. Nat Commun. 2020;11(1):1917.

281. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10(1):33.

282. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. Journal of Statistical Software. 2017;76(1):1-32.

283. Gronau QF, Singmann H, Wagenmakers E-J. bridgesampling: An R Package for Estimating Normalizing Constants. Journal of Statistical Software. 2020;92(10):1-29.

284. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21.

285. Danial-Farran N, Brownstein Z, Gulsuner S, Tammer L, Khayat M, Aleme O, et al. Genetics of hearing loss in the Arab population of Northern Israel. Eur J Hum Genet. 2018;26(12):1840-7.

286. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43(Database issue):D805-11.

287. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nat Protoc. 2016;11(1):1-9.

288. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7.20.

289. Hong D, Gupta R, Ancliff P, Atzberger A, Brown J, Soneji S, et al. Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. Science. 2008;319(5861):336-9.

290. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. Cell. 2018;173(3):611-23.e17.

291. Greaves M, Hughes W. Cancer cell transmission via the placenta. Evol Med Public Health. 2018;2018(1):106-15.

292. Van Egeren D, Escabi J, Nguyen M, Liu S, Reilly CR, Patel S, et al. Reconstructing the Lineage Histories and Differentiation Trajectories of Individual Cancer Cells in Myeloproliferative Neoplasms. Cell Stem Cell. 2021;28(3):514-23.e9.

293. Hermange G, Rakotonirainy A, Bentriou M, Tisserand A, El-Khoury M, Girodon F, et al. Inferring the initiation and development of myeloproliferative neoplasms. Proc Natl Acad Sci U S A. 2022;119(37):e2120374119.

294. Fabre MA, de Almeida JG, Fiorillo E, Mitchell E, Damaskou A, Rak J, et al. The longitudinal dynamics and natural history of clonal haematopoiesis. Nature. 2022;606(7913):335-42.

295. Bao EL, Nandakumar SK, Liao X, Bick AG, Karjalainen J, Tabaka M, et al. Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. Nature. 2020;586(7831):769-75.

296. Ueda K, Ikeda K, Ikezoe T, Harada-Shirado K, Ogawa K, Hashimoto Y, et al. Hmga2 collaborates with JAK2V617F in the development of myeloproliferative neoplasms. Blood Adv. 2017;1(15):1001-15.

297. Michejda M. Which stem cells should be used for transplantation? Fetal Diagn Ther. 2004;19(1):2-8.

298. Billingham RE, Brent L, Medawar PB. ‘Actively Acquired Tolerance’ of Foreign Cells. Nature. 1953;172(4379):603-6.

299. Palmer E. Negative selection — clearing out the bad apples from the T-cell repertoire. Nat Rev Immunol. 2003;3(5):383-91.

300. Takahama Y. Journey through the thymus: stromal guides for T-cell development and selection. Nat Rev Immunol. 2006;6(2):127-35.

301. Santore MT, Roybal JL, Flake AW. Prenatal Stem Cell Transplantation and Gene Therapy. Fetal Surgery. 2009;36(2):451-71.

302. Giles C. The platelet count and mean platelet volume. Br J Haematol. 1981;48(1):31-7.

303. Zipursky A, Bow E, Seshadri RS, Brown EJ. Leukocyte density and volume in normal subjects and in patients with acute lymphoblastic leukemia. Blood. 1976;48(3):361-71.

304. Cohn CS, Williams S. Cryopreserved platelets: the thaw begins … (Article, p. 2794). Transfusion. 2019;59(9):2759-62.

305. Nielsen C, Bojesen SE, Nordestgaard BG, Kofoed KF, Birgens HS. JAK2V617F somatic mutation in the general population: myeloproliferative neoplasm development and progression rate. Haematologica. 2014;99(9):1448-55.

306. Cordua S, Kjaer L, Skov V, Pallisgaard N, Kefala M, Gjerdrum LMR, et al. Early detection of myeloproliferative neoplasms in a Danish general population study. Leukemia. 2021;35(9):2706-9.

307. Koboldt DC. Best practices for variant calling in clinical sequencing. Genome Med. 2020;12(1):91.

308. Moss J, Ben-Ami R, Shai E, Kalish Y, Klochender A, Cann G, et al. Megakaryocyte and erythroblast DNA in plasma and platelets. bioRxiv. 2022;2022.10.03.510502.

309. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, et al. Somatic mutant clones colonize the human esophagus with age. Science. 2018;362(6417):911-7.

310. Yokoyama A, Kakiuchi N, Yoshizato T, Nannya Y, Suzuki H, Takeuchi Y, et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. Nature. 2019;565(7739):312-7.

311. Brunner SF, Roberts ND, Wylie LA, Moore L, Aitken SJ, Davies SE, et al. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. Nature. 2019;574(7779):538-42.

312. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015;348(6237):880-6.

313. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, et al. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. Cell. 2018;173(3):611-23 e17.

314. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. Nature. 2020;578(7793):122-8.

315. Khan AO, Rodriguez-Romera A, Reyat JS, Olijnik AA, Colombo M, Wang G, et al. Human Bone Marrow Organoids for Disease Modeling, Discovery, and Validation of Therapeutic Targets in Hematologic Malignancies. Cancer Discov. 2023;13(2):364-85.

316. Nagy Z, Vögtle T, Geer MJ, Mori J, Heising S, Di Nunzio G, et al. The Gp1ba-Cre transgenic mouse: a new model to delineate platelet and leukocyte functions. Blood. 2019;133(4):331-43.

317. Luis TC, Wilkinson AC, Beerman I, Jaiswal S, Shlush LI. Biological implications of clonal hematopoiesis. Exp Hematol. 2019;77:1-5.

318. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature. 2014;506(7488):328-33.