RELATIONSHIP BETWEEN THE GERMLINE POLYMORPHISMS AND SOMATIC MUTATIONS IN THE *JAK2* LOCUS IN CHINESE PATIENTS WITH MYELOPROLIFERATIVE NEOPLASMS

KOH SU PIN

Ph.D

The Hong Kong Polytechnic University

2015

THE HONG KONG POLYTECHNIC UNIVERSITY

Departmental of Health Technology and Informatics

Relationship between the germline polymorphisms and somatic mutations in the *JAK2* locus in Chinese patients with myeloproliferative neoplasms

KOH SU PIN

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

September 2014

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

KOH SU PIN (Name of student)

ABSTRACT

Myeloproliferative neoplasms (MPNs) manifest as a group of haematological malignancies characterised by a somatic mutation (*JAK2* V617F) that causes the bone marrow to produce too many blood cells. This mutation is found in polycythaemia vera (~95%), essential thrombocythaemia and primary myelofibrosis (both ~50%) and is considered as a major genetic factor contributing to the development of these MPNs. Within my study period, there is no genetic association study of MPN in the Hong Kong population. In this study, I aimed at investigating the relationship between germline *JAK2* polymorphisms and MPNs in Hong Kong Chinese to find causal variants that contribute to MPNs.

In part I of this thesis, 19 tag single nucleotide polymorphisms (SNPs) within the *JAK2* locus were analysed in 172 MPN patients and 470 healthy controls. Three of these 19 SNPs defined the reported *JAK2* 46/1 haplotype: rs10974944, rs12343867 and rs12340895. Allele and haplotype frequencies were compared between the patients and controls by logistic regression with adjustment for sex and age. Permutation test was used to correct for multiple comparisons. With significant findings from the 19 SNPs, I then examined 76 additional SNPs across the 148.7-kb region of *JAK2* via imputation with the SNP data from the 1000 Genomes Project. This is the first MPN study that employed SNP imputation in the data analysis.

In single-marker analysis, 15 SNPs showed association with *JAK2* V617F-positive MPNs (n=128). Exhaustive variable-sized sliding-window haplotype analysis identified 184 haplotypes showing significant differences (P<0.05) in frequencies between

patients and controls even after multiple-testing correction. However, single-marker alleles exhibited the strongest association with V617F-positive MPNs and among all MPNs. In local Hong Kong Chinese, the strongest signal came from rs12342421: asymptotic $P=3.76\times10^{-15}$, empirical $P=2.00\times10^{-5}$ for 50,000 permutations, OR=3.55 for the minor allele *C*, and 95% CI, 2.59-4.87. This SNP was also found by conditional logistic regression to contribute an independent effect in significant haplotype windows, and its role was not changed even with the imputation of additional 76 SNPs. In silico analysis suggested that several transcription factors might bind to rs12342421. Further functional validation is necessary to prove its involvement in the pathogenesis of MPNs.

Part II of this study is the first meta-analysis assessing the relationship between *JAK2* polymorphism and the risk of MPNs, within my study period. Significant association was observed between *JAK2* risk alleles and MPNs particularly the V617F-positive cohort, and to a lesser extent the V617F-negative MPNs but the exact mechanism of how the *JAK2* polymorphisms affect MPNs was not explored and thus is still unknown. Further research looking for the true causal variants between *JAK2* polymorphisms and MPNs is demanded. Large sample clinical studies should be carried out to verify more risk factors in order to make early detection and prevention at the gene level possible. However, sample recruitment was beyond the control of the study. International collaboration may increase the power to study and detect rare variants if any.

Based on the results generated from part I and II of this study, we hypothesised that there are *JAK2* disease-causing variants that may remain untyped in the current geno-

typing approach. Based on our association results from case-control study, LocusZoom software identified two recombination hotspots covering a region of 330kb (including 142kb of *JAK2*). Therefore, finer scale deep sequencing analysis was performed to explore the undiscovered variants which may be the real causal variants. The 330kb of *JAK2* gene and the coding exons (all isoforms) of eight MPN-associated genes including 50bp upstream and downstream of the regions were targeted by a custom SeqCap EZ Choice Library from Roche NimbleGen. In total, 121 regions were targeted comprising 401.324 kb of target sequence. Samples from 48 MPN patients and 48 sex- and age-matched controls were sequenced on the Illumina's MiSeq benchtop sequencer. QC filters were applied to remove poor sequencing reads using Illumina internal platforms and GATK. A total of 532 variants were found. NGS and genotyping together with imputation (part I & II of this project) detected strong signal from intronic SNPs within the recombination hotspots. Consistent results from these studies further strengthen our hypothesis that rs12342421 (S8) and other strongly associated polymorphisms, are in strong LD with some untyped causal variants that are predisposing to MPN. Further functional validation is necessary to prove the involvement of such variants in the pathogenesis of MPN. In the long run, future studies on larger sample cohort may give further insight on the pathogenesis of V617F-positive and -negative MPNs.

List of Publications

International journal papers

Published paper:

1. **Su Pin Koh**, Shea Ping Yip, Kwok Kuen Lee, Chi Chung Chan, Sze Man Lau, Chi Shan Kho, Chi Kuen Lau, Shek Ying Lin, Yat Ming Lau, Lap Gate Wong, Ka Leung Au, Kit Fai Wong, Raymond W Chu, Pui Hung Yu, Eudora ED Chow, Kate FS Leung, Wai Chiu Tsoi, Benjamin YM Yung.
   Title: Genetic association between germline *JAK2* polymorphisms and myeloproliferative neoplasms in Hong Kong Chinese population: a case-control study, *BMC Genet*. 2014 Dec 20; 15:147.

Papers in preparation:

2. **Su Pin Koh**, Shea Ping Yip, Kwok Kuen Lee, Chi Chung Chan, Sze Man Lau, Chi Shan Kho, Chi Kuen Lau, Shek Ying Lin, Yat Ming Lau, Lap Gate Wong, Ka Leung Au, Kit Fai Wong, Raymond W Chu, Pui Hung Yu, Eudora ED Chow, Kate FS Leung, Wai Chiu Tsoi, Benjamin YM Yung.
   Title: Association between germline *JAK2* polymorphisms and myeloproliferative neoplasms risk: A Meta-Analysis across populations

3. **Su Pin Koh**, Kwok Kuen Lee, Chi Chung Chan, Sze Man Lau, Chi Shan Kho, Chi Kuen Lau, Shek Ying Lin, Yat Ming Lau, Lap Gate Wong, Ka Leung Au, Kit Fai Wong, Raymond W Chu, Pui Hung Yu, Eudora ED Chow, Kate FS Leung, Wai Chiu Tsoi, Shea Ping Yip, Benjamin YM Yung.
   Title: Identification of SNPs in MPNs using the Next-generation sequencing approach

International conference papers:

4. **Koh SP**, Yip SP, Lee KK, Chan CC, Lau SM, Kho CS, Lau CK, Lau YM, Lin SY, Wong LG, Au KL, Wong KF, Chu W, Yu PH, Chow ED, Leung KFS, Tsoi WC, Yung YM.

   Title: Imputation and meta-analysis: Genetic association between germline *JAK2* polymorphisms and myeloproliferative neoplasms in Chinese and different populations. *ESH International Conference on Myeloproliferative Neoplasms*, October 04 - October 06, 2012 - Vienna, Austria.

5. Yip SP, **Koh SP**, Lee KK, Chan CC, Lau SM, Kho CS, Lau CK, Lau YM, Lin SY, Wong LG, Au KL, Wong KF, Chu W, Yu PH, Chow ED, Leung KFS, Tsoi WC, Yung YM.

   Title: Common germline *JAK2* polymorphisms are associated with myeloproliferative neoplasms with somatic *JAK2* mutation in a Chinese population – A replication study. *Laboratory Medicine Conference*, Taiwan; 4-7 November 2011.

6. **Koh SP**, Yung YM, Yip SP.

   Title: Relationship between germline polymorphisms and somatic mutations in *JAK2* locus in Chinese patients with Myeloproliferative Neoplasms. HTI.*Postgraduate Symposium 2011*, The Hong Kong Polytechnic University.

ACKNOWLEDGEMENTS

I am grateful to work with a lot of inspiring people throughout my PhD study and withal part of my life. Thanks to their guidance and tremendous support, I completed my work.

Quoted from my brother “If science has a pedigree”, my chief supervisor would be my “scientific father” and my co-supervisor would be my “scientific mother”.

As the chief supervisor, Prof. Benjamin Yung oversaw and coordinated all aspects of my PhD project. He edified me the attitude toward good science. Throughout my study period, Prof. Yung has supported me in numerous ways: he allowed me to think wild scientifically; he employed me as a part-time research assistant so that I could write my thesis without financial burden; he gave me opportunity to join discussion with distinguished scientists; he also provided me with references for my various applications. I hope I have made him proud.

As my co-supervisor, Prof. Shea-Ping Yip guided me academically and personally as far as he could. He taught me the most, both in and out of the lab. He supervised my progress and overlooked my personal growth as an independent researcher. Whenever I felt lost and depressed in my research, he motivated and encouraged me. I will always remember the many conversations we had about life, research, and career. I wish you nothing but the best in your future endeavours.

I would like to thank Dr. Lawrence WC Chan, Dr. SC Cesar Wong, and Dr. Parco M Siu for their help and guidance throughout my 5 years here! My heartfelt gratitude extends to Dr. Nancy Tsui for supervising my NGS progress and guiding me the way toward the leading science. She made me realise that charming lady could also be a distinguished scientist!

Of course, I could not survive the long PhD journey without the care, jests and repasts with my lovely lab mates. In particular, I want to thank Dr. Isabella Cheuk for her continuous motivation and encouragement albeit I transgressed the promise that we would graduate at the same time (we can finally go for a trip now!). Thanks also to my best friend Ms. Wanlin Chee for her timely encouragement since I left home for PhD study. I would like to thank Dr. Aimee Pei, Dr. Wang Fengfeng, Dr. Björn Tam, and Dr. Henry Hung for their company during my difficult time. Grateful also to my friends who learned not to ask when I could graduate! I wish you all the very best with your respective pursuits, and look forward to your future successes.

I would like to thank my brother a.k.a. my walking dictionary - Gene C.C. Koh, who will soon start his PhD journey at The University of Cambridge. Not only has he provided me encouragement from a distance, he impressed me with his high level of ebullience and his tenacity toward great science. I am so proud and fortuitous to have such a smart and inspiring brother, a great source of vigour and motivation all the time.

Finally and perhaps most importantly, I am greatly indebted to my family. My education in Hong Kong would not have been possible without the unfaltering supports in various forms from my family. Delighted, I am to dedicate this thesis to my mother Mei Yong Lim and my father, Cho Wah Koh.

Table of Contents

ABBREVIATIONS

5’ UTR	the 5' untranslated region
A	adenine
AML	acute myeloid leukaemia
APS	ammonium persulphate
ARMS	amplification refractory mutation system
AS-PCR	allele-specific polymerase chain reaction
ASXL1	additional Sex Combs-Like 1
Bis	N, N’- methylenebisacrylamide
bp	base pair
C	cytidine
CAE	capillary array electrophoresis
CARL	Calreticulin
CBC	complete blood count
CBL	casitas B-lineage lymphoma proto-oncogene
CE	capillary electrophoresis
CEL	chronic eosinophilic leukaemia
CEL/HES	chronic eosinophilic leukaemia/hypereosinophilic syndrome
CIs	confidence intervals
CLL	chronic lymphocytic leukaemia
CML	chronic myeloid leukaemia
CMPDs	chronic myeloproliferative diseases
CNKI	Chinese National Knowledge Infrastructure
CNL	chronic neutrophilic leukaemia
ddNTP	dideoxynucleotide triphosphate
	[N= adenosine (A), guanosine (G), cytidine (C), thymine (T)]
dHPLC	denaturing high-performance liquid chromatography
dNTP	deoxynucleotide triphosphates
	[N= adenosine (A), guanosine (G), cytidine (C), thymine (T)]
dsDNA	double-stranded DNA
EDTA	ethylene diamine tetraacetic acid
EEC	endogenous erythroid colony
E-M algorithm	algorithm of expectation and maximization

emPCR	emulsion PCR
EMSA	electrophoretic mobility shift assay
EPO	erythropoietin
EPOR	erythropoietin receptor
eQTL	expression quantitative trait loci
ET	essential thrombocythaemia
ExoI	exonuclease I
EZH2	enhancer Of Zeste Homolog 2
FE	fixed-effects
FO	forward outer
Fwt	forward wild-type-specific
G	guanosine
GATK	genome Analysis ToolKit
GWAS	genome-wide association studies
Hb	haemoglobin
HES	hypereosinophilic syndrome
HSC	haematopoietic stem cells
HSESC	Human Subjects Ethics Sub-Committee
Ht	haematocrit
HU	hydroxycarbamide
HWE	Hardy-Weinberg equilibrium
IDH	Isocitrate dehydrogenase
IDT	Integrated DNA Technologies, Inc.
INSL4	Insulin-like 4 (placenta)
JAK	Janus kinase
JAK-STAT	Janus kinase-signal transducer and activator of transcription
JH1	*JAK* homology 1
JH2	*JAK* homology 2
kb	kilobase
LC480	LightCycler® 480
LD	linkage disequilibrium
LM	ligation-mediated
LOH	loss of heterozygosity
MAF	minor allele frequency

MAPK	mitogen-activated protein kinase
MCS	MiSeq Control Software
MF	myelofibrosis
$MgCl_2$	magnesium chloride,
miRNAs	microRNAs
ml	millilitre
mM	millimolar
MPDs	myeloproliferative disorders
MPL	myeloproliferative Leukaemia Virus
MPNs	myeloproliferative neoplasms
MQ	root mean square of the mapping quality
mRNAs	messenger RNAs
MS	multiple sclerosis
N	number of times
ng	nanogram
nm	nanometre
NGS	Next Generation Sequencing
NH-2	amino terminal
nSNVs	non-synonymous single nucleotide variations
OR	odds ratio
ORF	open reading frame
P13	phosphatidylinositol-3 kinase
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
P_{emp}	empirical *P* values
Ph	Philadelphia
phos	phosphorylated
PI3K	phosphatidylinositol 3 kinase
PLT	low platelet count
PMF	primary myelofibrosis
PV	polycythaemia vera
RAS-MAPK	RAS and mitogen-activated protein kinase
RE	random-effects
ReadPosRankSum	read position rank sum test score

RevMan	Review Manager
RFLP	restriction fragment length polymorphism
Rmt	reverse-mutant-specific
RO	reverse outer
ROS	reactive oxygen species
SAP	shrimp alkaline phosphatase
SBS	sequencing by synthesis
SH2	SRC homology 2 domain
SH2B3	SH2B adaptor protein 3 (*LNK*)
SM	specified mastocytosis
SNP array	single nucleotide polymorphism array
SNPs	single nucleotide polymorphisms
ssDNA	single-stranded DNA
SSLD	solid spine of linkage disequilibrium
STATs	signal transducer and activator of transcription
SVT	splanchnic vein thrombosis
T	thymine
TBE	Tris Borate Ethylenediaminetetraacetic acid
TEMED	tetramethvlethvlenediamine
TET2	Ten-Eleven Translocation-2
Tm	melting temperature
Tris-HCl	Tris-hydrochloric acid
tSNP	tagging SNP
TYK2	Tyrosine kinase 2
uMPN	unclassifiable MPNs
U	unit
UPD	uniparental disomy
UPR	unlabelled probe melting curve analysis
UV	ultraviolet
WBC	white blood cell count
WHO	World Health Organisation
µg	microgram
µl	microliter

CHAPTER 1 Introduction

1.1 Normal haematopoiesis

Haematopoiesis is the formation, development, and differentiation of a variety of distinct blood cellular components. It comprises multiple stages, originating from pluripotent stem cells with self-renewal capacity through intermediate progenitors to mature differentiated cells (Shochat et al., 2002; Morrison et al., 1995). All cellular blood components are derived from the earliest blood precursors, called haematopoietic stem cells (HSC) (Borthwick, 2008) which reside in the medulla of the bone marrow. Normal stem cells can divide asymmetrically beside symmetric division under homeostatic pressure to committed progenitor cells with some daughter cells remaining as long-term HSCs to prevent stem cells from depleting (Johnsen et al., 2009; Clarkson et al., 2003). Uniquely, these long-term HSCs produce short-term HSCs and short-term HSCs then develop into two separate lineages, namely the myeloid and lymphoid lineages. Progenitor cells from both lineages can each differentiate into any type of functional mature cells (blood cells) with different functions ranging from immunity to the transportation of oxygen depending on the differentiation pathways undertaken (Schmerer & Evans, 2001). The myeloid lineage produces morphologically and functionally distinct cell types and the lymphoid lineage produces cell types that are responsible for adaptive immunity (Katja Fiedler & Brunner, 2012). This process is balanced against the rate of apoptosis (programmed cell death) and cell cycle time with some progenitor cells eventually become terminally differentiated and stop dividing (Clarkson et al., 2003; Johnsen et al., 2009) (**Figure 1.1**). Biological stresses including bleeding or infection are also known to regulate the haematopoietic system (Vardiman et al., 2009).

There are different sites where haematopoiesis occurs: in adults, the major site of haematopoiesis is bone marrow whereas in the embryo before bone marrow develops, haematopoiesis happens in yolk sac during early embryonic life and; in the foetal liver during the second trimester of pregnancy. **Figure 1.2** depicts the model of haematopoiesis where HSC makes an early decision to commit to either one major pathway, myeloid or lymphoid (Jones, 2010; Katja Fiedler & Brunner, 2012). The progenitor cell then starts the progressive differentiation process to develop the characteristics of its chosen lineage and loses its capability to proliferate and self-renew (Shochat et al., 2002; Jones, 2010; Borthwick, 2008; Katja Fiedler & Brunner, 2012). Mature cells arising from the myeloid lineage include erythrocytes, neutrophils, monocytes/macrophages, eosinophils, basophils, and megakaryocytes/platelets, whereas mature cells arising from lymphoid lineage become either B cells, T cells or natural killer cells (Katja Fiedler & Brunner, 2012). New blood cells will be constantly produced to replace the mature blood cells when they die being too old or damaged. Generally, the lifespans for mature blood cells in the blood circulation are short ranging from few hours to few weeks for white blood cells, 10 days for platelets, and approximately 120 days for erythrocytes (Shah & Zuckerman, 2011; Perry et al., 1959; Katja Fiedler & Brunner, 2012).

Regulation of haematopoiesis is complex and involves internal and external stimuli, which is crucial in balancing the production of a variety of blood cells in need. Internally, transcription factors act as an early intrinsic determinants in programming the fate or lineage of haematopoietic progenitor cells (Orkin & Zon, 2008). Specific interaction between transcription factors and regulatory region of genes define the tran-

scription network of haematopoiesis. Disruption in the network leads to haematopoietic malignancies (Shah & Zuckerman, 2011). In addition, lineage commitment is also programmed by a series of soluble haematopoietic growth factors secreted by stromal cells (e.g. the bone marrow) called cytokines. Cytokines include interleukins, interferons, lymphokines, chemokines, colony-stimulating factors and other haematopoietic hormones. They work as signalling molecules to initiate regulatory pathways by binding to specific cell surface receptors on the progenitor (Shah & Zuckerman, 2011). Cytokines may stimulate distinct biological responses such as proliferation, growth, and differentiation depending on the cell type that it binds. Cytokine-receptor interactions are specific and very much transcriptional and genomic dependent. Cytokine receptors can be classified into specific families according to their signal transduction subunits. The 3 main signalling pathways that regulate transcription of haematopoiesis are the JAK-STAT (Janus kinase-signal transducer and activator of transcription) pathway, the MAPK (mitogen-activated protein kinase) pathway, and the P13 (phosphatidylinositol-3 kinase)/AKT pathway (Shah & Zuckerman, 2011). Other signalling molecules such as Ras, Raf/MEK/ERK, protein kinase C, Src family kinases, TGF-β /SMAD may also act in the regulation of haematopoiesis (Miranda & Johnson, 2007).

MicroRNAs (miRNAs) are important regulators of haematopoiesis. They control gene expression of transcription factors essential for haematopoietic commitment, proliferation, differentiation, and cell death. These small non-coding RNAs can bind to the 3'- or, less commonly 5'-untranslated region or the open reading frame (ORF) of messenger RNAs (mRNAs) resulting in mRNAs degradation or miRNA-mediated translation inhibition (Undi et al., 2013; Shah & Zuckerman, 2011). Overexpression

of miRNAs was associated with haematological malignancies for instance chronic lymphocytic leukaemia (CLL) and MPNs (Undi et al., 2013; Hussein et al., 2009).

Externally, the haematopoietic system is capable of sensing and responding to the changing environmental conditions such as hypoxia (low oxygen concentration from blood loss, high altitudes, and etc) (Kumar & Evans, 2001). The lack of oxygen will trigger haematopoietic system for additional generation of erythrocytes to compensate for the increased respiratory needs.

Haematopoiesis is a regulated ongoing process throughout the lifetime. Failure in its regulation can lead to some common yet severe diseases including leukaemia when the blood cell proliferation is beyond control, or anaemia when there is insufficient of blood cells are made (Schmerer & Evans, 2001).

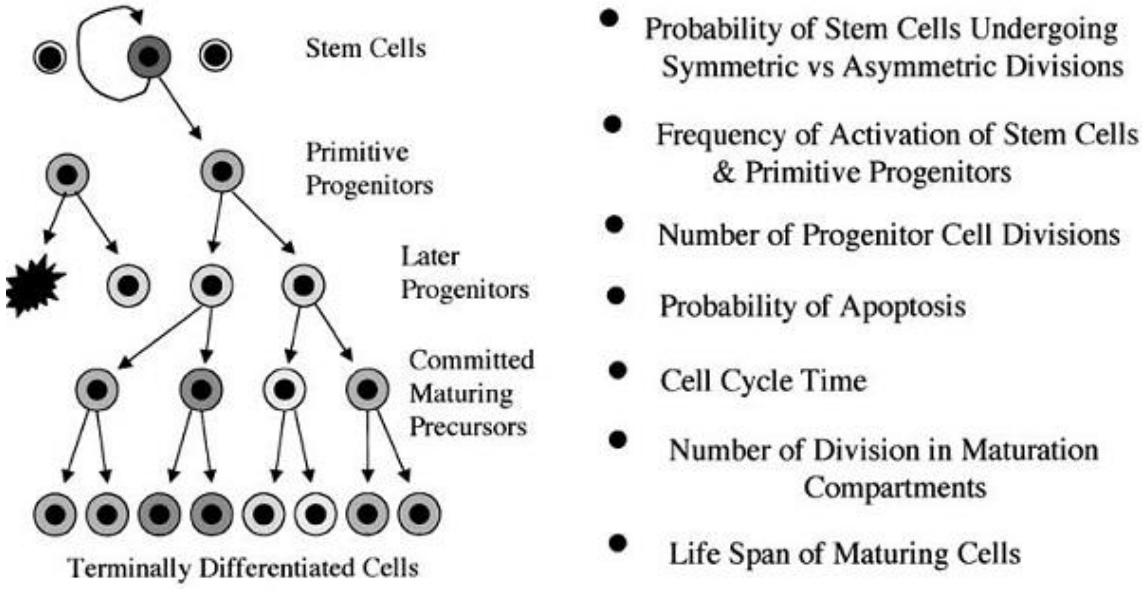


Figure 1. 1 Possible controllable parameters regulating blood cell production.

Adapted from Clarkson et al., 2003.

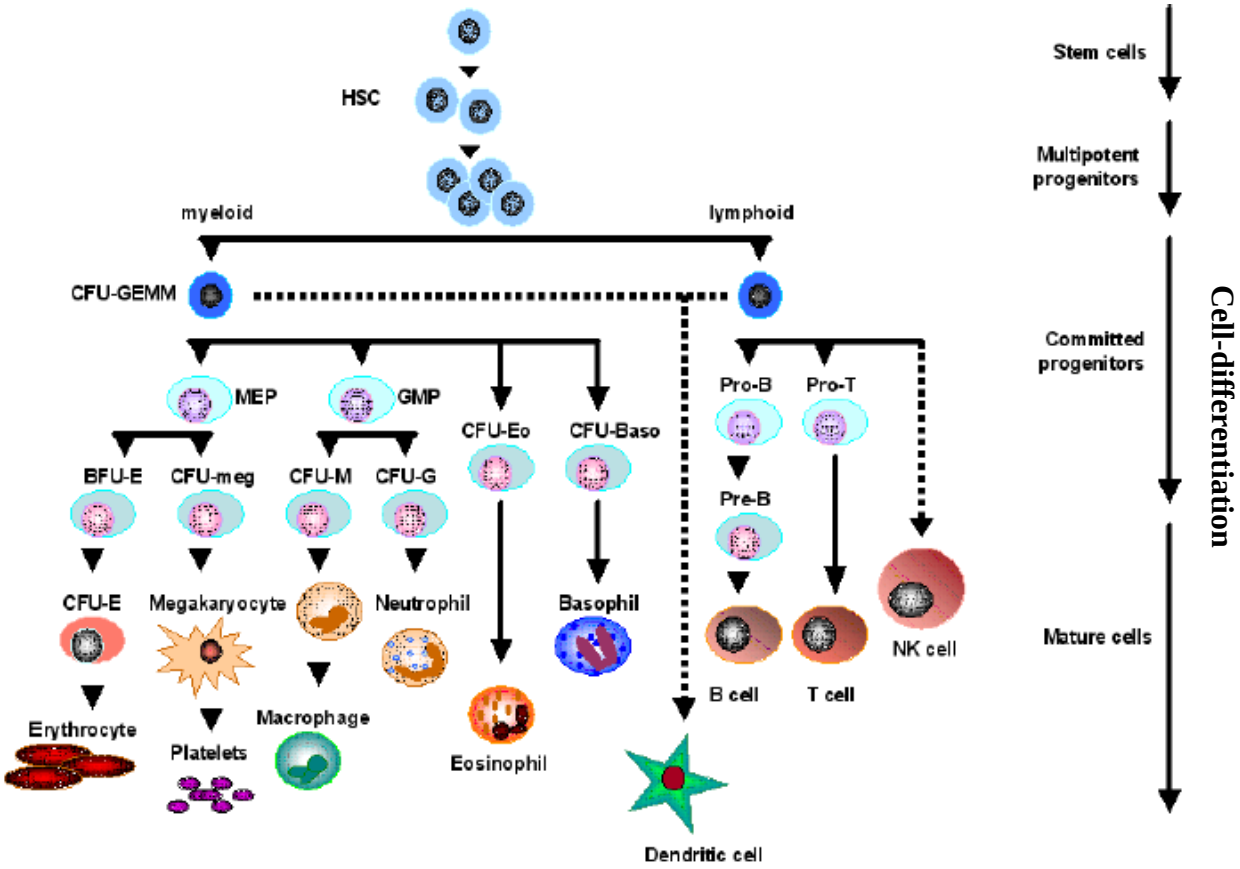


Figure 1. 2 Haematopoiesis.

Every mature blood cell is derived from a multipotent haematopoietic stem cell (HSC). CFU-GEMM; colony forming unit granulocyte erythroid megakaryocyte/ macrophage, MEP; megakaryocyte erythroid progenitor, GMP; granulocyte macrophage progenitor, BFU-E; burst forming unit-erythroid, CFU-E; colony forming unit erythroid, CFU-M; colony forming unit macrophage, CFU-G; colony forming unit granulocyte, CFU-Baso; colony forming unit basophil. Adapted from Jones, A (2010).

1.2 Haematologic malignancies

Haematological malignancies are clonal diseases originated from single cell in the bone marrow and lymph nodes that has undergone disruption in the balance between proliferation, survival, and differentiation. Haematological malignancies are complicated. They account for approximately 7% of all malignancies (Hoffbrand & Moss, 2011a). Haematological malignancies occur due to accumulated genetic alterations or changes that can further activate the (proto) oncogenes (proto-oncogenes are normal genes that can convert into oncogenes if being expressed at higher-than-normal level and under mutation, that can promote uncontrolled growth of cancer) or deactivate the tumour suppresser genes (**Figure 1.3**). Genetic alterations such as chromosomal translocation, chromosomal deletion, or mutations in the DNA sequence can initiate autonomous, proliferating stem cell clone (Kralovics, 2012; Jones, 2010). These alterations are the circumstances resulting from defects in DNA repair and increased production of reactive oxygen species (ROS) (Sallmyr et al., 2008).

Phenotypes of haematological malignancies are related to the genetic alterations encountered. Mainly, they can be classified by the predominant type of blood cells affected - myeloid or lymphoid, and the tissue where the mutated cells arise also determines their disease phenotype (Jones, 2010). For instance, leukaemia (both chronic and acute) involves white blood cells; lymphomas (Hodgkin's and non-Hodgkin's) involve lymphatic system; myeloma involves plasma cells. Their phenotypic traits can also be determined by the cellular pathways (cell growth and survival) they sabotage in a multistep approach with acquisition of more mutations that lead to excess proliferation, and dysregulation of apoptosis (Lichtman, 2008; Hoffbrand & Moss,

2011a). These pathways are deregulated by diverse stimuli both intrinsically and extrinsically (Shah & Zuckerman, 2011).

No single clonal proliferation event is sufficient to cause genetic instability that leads to disease progression but rather the accumulation of clonal events that mess the control of cell cycle, self-renewal, and cell death (Krakow, 2012; Clarkson et al., 2003). The expansion of a stem cell clone will create competition between its progeny and healthy cells for "habitat" in the bone marrow microenvironment. After the first mutation, more mutagenesis occurs in subsequent clonal expansion with selective growth advantage that shapes the cancer genome of an individual resulting in each unique phenotype (**Figure 1.4**). (Kralovics, 2012). Nevertheless, the cancer genome will be constantly reshaped due to genetic instability as the disease progresses (Landau et al., 2014). Clonal expansion of abnormal blood cells derived from a single ancestral cell within the bone marrow not only causes blood cells to accumulate in major organs such as liver and spleen, but also disrupts normal haematopoiesis. These cause abnormal haematopoietic situation for instance anaemia and thrombocytopenia (Jones, 2010).

Today, the aetiology of most haematological malignancies is still unknown considering that most diseases involve cross talk between inheritance and environmental factors. Inherited factors such as genetic predisposition or genetic diseases such as Down's syndrome are greatly associated with acute leukaemia, one of the malignancies. Subtle familial predisposition was also observed in diseases such as acute myeloid leukaemia (AML) and CLL. Environmental factors such as ionising radiation, exposure to chemicals and drugs, infection (during pregnancy, viral and bacterial)

were also reported to increase the risk of developing one or several of these diseases (Hoffbrand & Moss, 2011a).

In this thesis, I will focus on the genetics of myeloproliferative neoplasms (MPNs), one of the phenotypically diverse groups of myeloid malignancies.

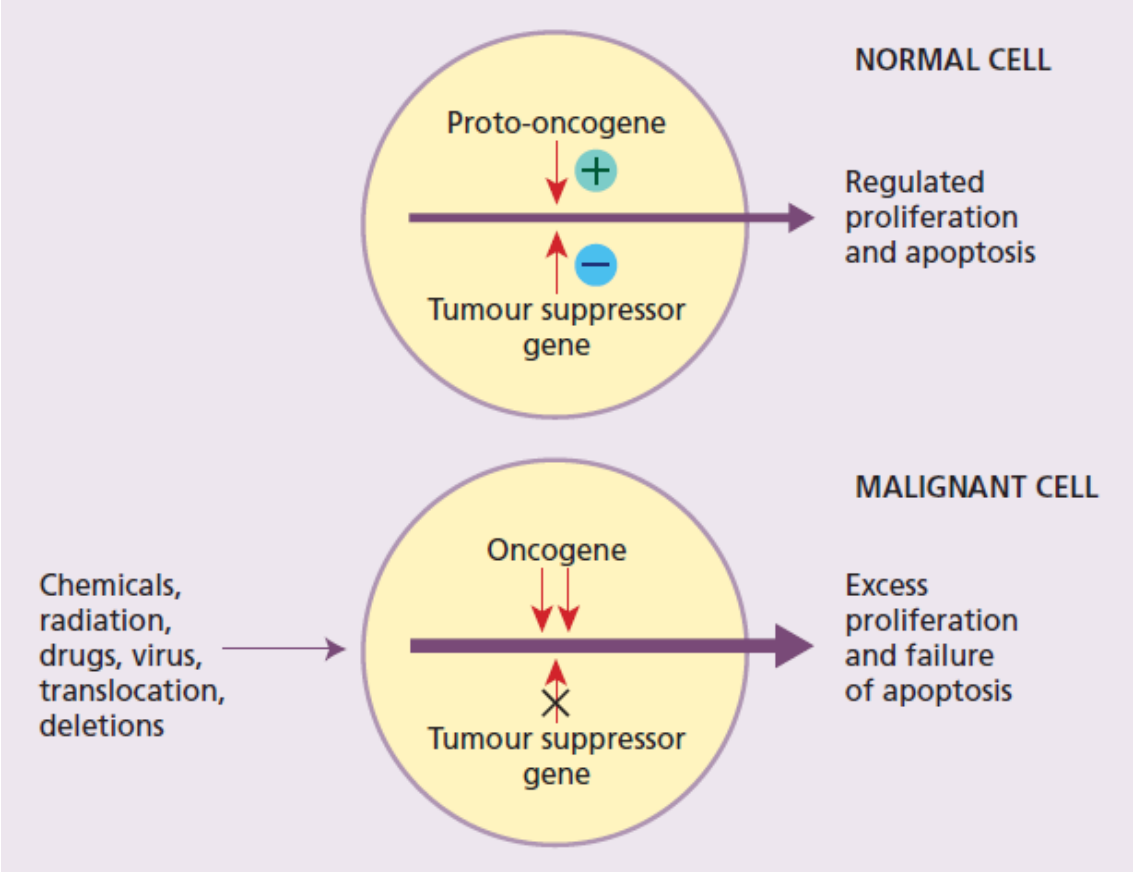


Figure 1. 3 Accumulated genetic alterations lead to haematological malignancies.

Haematological malignancies occur due to proliferation of normal cells depends on a balance between the action of proto-oncogenes and tumour-suppressor genes. In a malignant cell this balance is disturbed leading to uncontrolled cell division. Adapted from Hoffbrand & Moss (2011a).

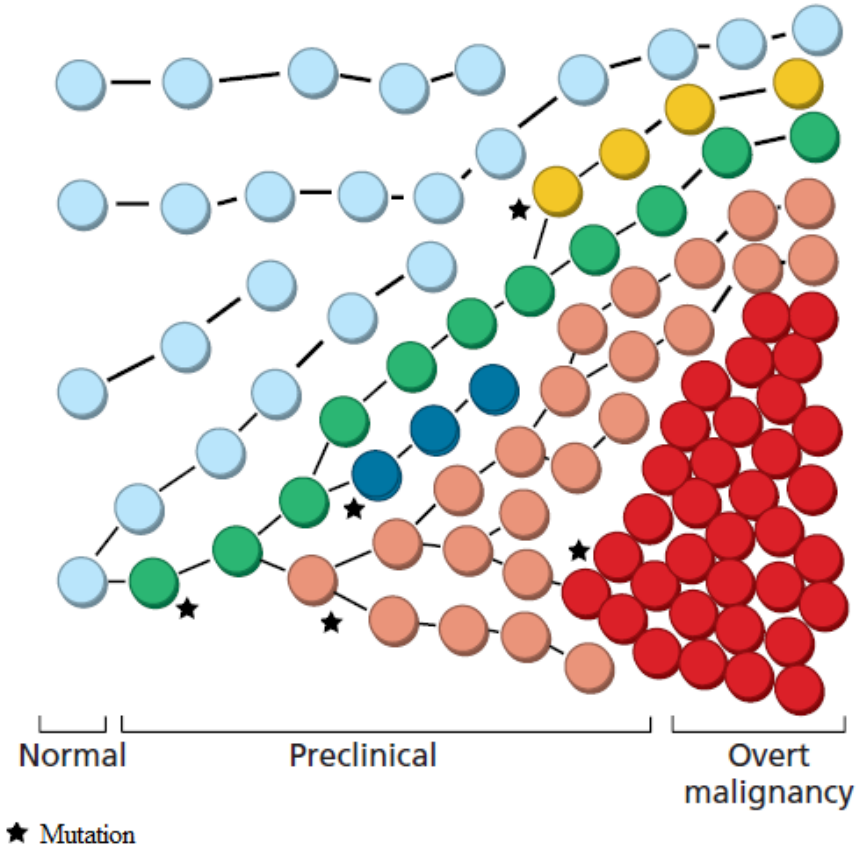


Figure 1. 4 Clonal proliferation in haematological malignancies.

The evolution starts with the acquisition of an advantageous mutation (filled star) that confers survival and growth advantage (e.g., resistance to apoptosis) in one cell (green) during normal haematopoiesis (light blue cells) and forms a local clone. Subsequently, the progeny of this cell (blue, peach, and orange) acquire additional mutations in a linear trend of successive clonal expansions that may lead to the formation of subclones with further aggressive phenotype, or mutations that may be deleterious and cause the clone to die out (blue). The clonal haematological malignancy is characterised by the co-existence of a heterogeneous group of subclones (green and peach) plus a dominant clone (red) that harbour driver mutation to compete for ascendency. All neoplastic cells in the clonal population share overlapping yet distinct sets of mutations (Landau et al., 2014). Adapted and modified from Campbell et al (2007).

1.3 Myeloproliferative Neoplasms (MPNs)

The term myeloproliferative disorders (MPDs) was first described by an American haematologist, William Dameshek in 1951 to represent a group of disorders characterised by pan-myeloid proliferative potential : uncontrolled proliferation and expansion of one or more myeloid cell types with unknown stimulus (Bueso-Ramos & Vardiman, 2011; Kilpivaara & Levine, 2008). William Dameshek grouped chronic myeloid leukaemia (CML), polycythaemia vera (PV), essential thrombocythaemia (ET), and primary myelofibrosis (PMF), and erythroleukaemia into MPDs (Wadleigh & Tefferi, 2010) noting that they share similarities in histology and clinical presentation (Abdel-Wahab & Levine, 2011). Erythroleukaemia and its subtypes were later removed from MPDs to join erythroid leukaemia. In 2001, the myeloid neoplasms classification of World Health Organisation (WHO) group the remaining classic MPDs under the umbrella of chronic myeloproliferative diseases (CMPDs) alongside with chronic neutrophilic leukaemia (CNL), chronic eosinophilic leukaemia/hypereosinophilic syndrome (CEL/HES), and unclassified CMPD (Wadleigh & Tefferi, 2010).

CMPDs comprised few disorders that are individually distinct but biologically similar. Pathologists and scientists have shown that certain genetic alterations targeting related molecules and tyrosine kinases, are the causes to constitutive activation of signalling pathways that leads to the abnormal proliferation of myeloid cell (Wadleigh & Tefferi, 2010). For instance, CML shares pathogenic similarities with some of the MPNs, but it has then evolved to be independent of MPNs and has been studied separately ever since the discovery of *BCR-ABL* fusion gene or the so-called Philadelphia (Ph) chromosome (Rowley, 1973; Campbell & Green, 2010). The sec-

ond most commonly recognised mutation that shed light on *BCR-ABL*-negative MPDs is the mutation (*JAK2* V617F) found in *JAK2* kinase discovered in 2005 (Levine et al., 2005; Jones et al., 2005; Baxter et al., 2005; Kralovics et al., 2005b). This *JAK2* V617F mutation was reported to be detected in most of the *BCR-ABL*-negative MPDs patients. Throughout the years, additional clinical and histologic discoveries have contributed to the research of these diseases. These discoveries marked the revision of 2008 WHO classification and diagnosis algorithms of CMPDs with the incorporation of additional molecular information (Wadleigh & Tefferi, 2010).

1.3.1 Classification of MPNs

CMPDs are clonal haematopoietic disorders originating from the bone marrow with similar biology (Wadleigh & Tefferi, 2010). Genetic abnormalities and histologic features of CMPDs, such as the use of *BCR-ABL* fusion gene in the diagnosis of CML and the most prevalent mutation (*JAK2* V617F) in the diagnosis of *BCR-ABL*-negative MPDs, have taken a big stride in distinguishing *BCR-ABL*-negative MPDs from other disorders because other reactive granulocytic, erythroid, and megakaryocytic hyperplasia always resemble them. To incorporate this new clinical and scientific information into the diagnostic criteria, a revision of WHO classification was made in 2008. As revised, the nomenclature "chronic myeloproliferative disorders" (CMPDs) was changed to "myeloproliferative neoplasms" (MPNs) to better reflect their neoplastic nature rather than reactive (Wadleigh & Tefferi, 2010; Vardiman et al., 2009). MPNs were redefined as a general diagnostic category encompassing a number of phenotypically distinct disorders with a common stem cell-derived clonal heritage attributed by abnormal signal transduction brought about by a range of mutations affecting protein tyrosine kinase or related molecules (Vardiman et al., 2009).

The re-termed MPNs now include *BCR-ABL1*–positive CML, PV, ET, PMF, CNL, CEL/not otherwise specified mastocytosis (SM) and unclassifiable MPNs (uMPN) (Vardiman et al., 2009).

Under MPNs, disorders can be grouped into 'classic' and 'non-classic' subtypes as depicted in **Figure 1.5**. These disorders often involve genetic alterations targeting tyrosine kinases and related molecules. This thesis focuses on the classical MPNs, namely PV, ET, and PMF, three disorders that are clinically and biologically related. They share the *JAK2* V617F mutation in addition to their distinct set of genetic abnormalities and have the potential to transform from one entity into another as the disease progresses (**Figure 1.6**) (Hoffbrand & Moss, 2011b). However, each subtype can still be distinguished among the three if the 2008 WHO diagnostic criteria were followed based on their histologic and molecular characterisation (**Table 1.1**) (Tefferi et al., 2009).

Most of the MPNs cases are asymptomatic and incidentally diagnosed during routine blood tests which reveal an abnormally increased levels of myeloid cells (erythrocytes, granulocytes, and/or platelets), especially in the case of ET and, PV to a lesser degree (Bueso-Ramos & Vardiman, 2011; Cervantes & Hernández-Boluda, 2012). While for the case of PMF, patients are more often diagnosed accompanied by symptoms such as enlarged spleen during routine physical examination or blood test (Barosi, 2011). MPNs have a tendency for transformation but at low and different frequencies to acute myeloid leukaemia (AML) (**Figure 1.6**) (Bueso-Ramos & Vardiman, 2011).

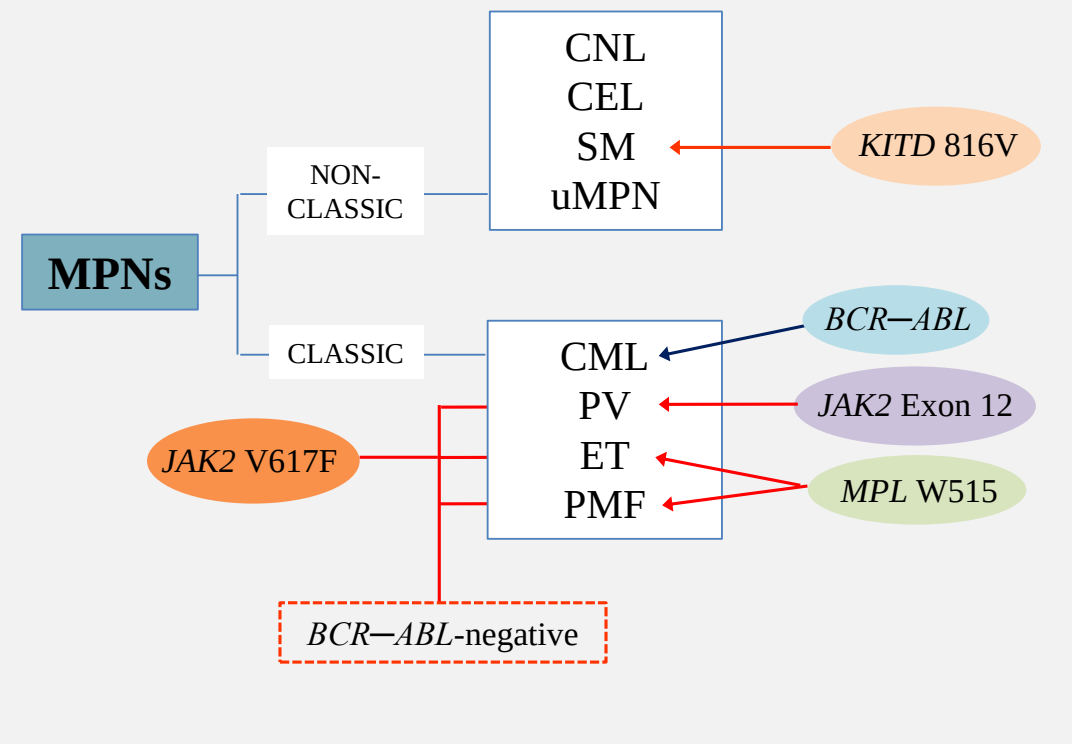


Figure 1. 5 The 2008 World Health Organisation (WHO) classification of MPNs.

MPNs can be categorised into non-classic and classic.subtypes Non-classic subtypes include Chronic Neutrophilic Leukaemia (CNL), Chronic Eosinophilic Leukaemia (CEL), Systemic Mastocytosis (SM), and unclassifiable MPNs (uMPN). Classic Subtypes Include Chronic Myeloid Leukaemia (CML), Polycythaemia Vera (PV), Essential Thrombocythaemia (ET), and Primary Myelofibrosis (PMF). The key genetic abnormalities of the diseases are represented in oval. Modified from Vakil and Tefferi (2011).

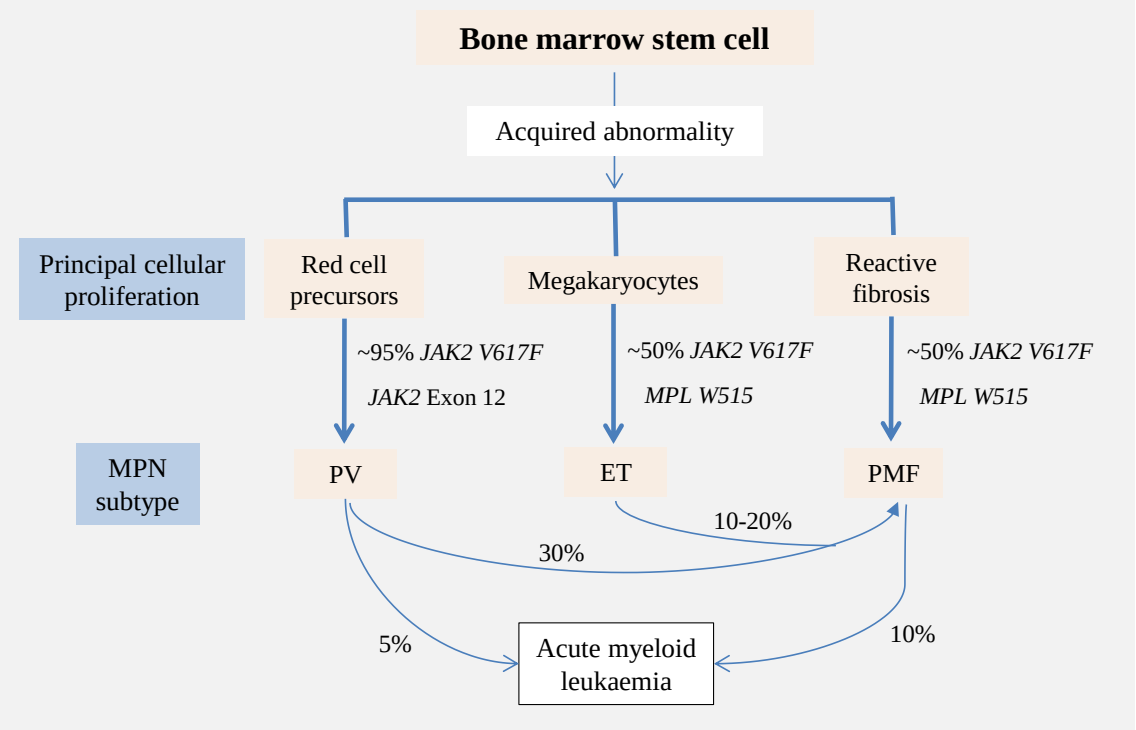


Figure 1. 6 Relationship among the three classical MPNs.

They are phenotypically distinct disorders derived from a common stem cell heritage attributed by abnormal signal transduction brought about by a range of mutations affecting protein tyrosine kinase or related molecules. In some cases, phenotypes of 2 disorders might affect the same patient simultaneously; in other cases, the disease transforms from one entity to another during its course, or to the more severe acute myeloid leukaemia. The three diseases, PV, ET, and PMF, can be characterised by their variable proportions of V617F (see text). Modified from Hoffbrand & Moss Hoffbrand and Moss (2011b).

Table 1.1 The 2008 WHO diagnostic criteria for PV, ET, and PMF

Adapted from Tefferi et al. (2009)

	2008 WHO Diagnostic Criteria		
	PV*	**ET†**	**PMF‡**
Major criteria	1 Hgb >18.5 g/dL (men) >16.5 g/dL (women) or Hgb >17 g/dL (men), or >15 g/dL (women) if associated with a sustained increase of ≥ 2 g/dL from baseline that can not be attributed to correction of iron deficiency or§	1 Platelet count ≥450 × 10^9/L	1 Megakaryocyte proliferation and atypia‖ accompanied by either reticulin and/or collagen fibrosis, or In the absence of reticulin fibrosis, the megakaryocyte changes must be accompanied by increased bone marrow cellularity, granulocytic proliferation and often decreased erythropoiesis (ie, prefibrotic PMF).
	2 Presence of *JAK2*V617F or similar mutation	2 Megakaryocyte proliferation with large and mature morphology. No or little granulocyte or erythroid proliferation	2 Not meeting WHO criteria for CML, PV, MDS, or other myeloid neoplasm
		3 Not meeting WHO criteria for CML, PV, PMF, MDS or other myeloid neoplasm	3 Demonstration of *JAK2*V617F or other clonal marker or no evidence of reactive bone marrow fibrosis
		4 Demonstration of *JAK2*V617F or other clonal marker or no evidence of reactive thrombocytosis	
Minor criteria	1 BM trilineage myeloproliferation		1 Leukoerythroblastosis
	2 Subnormal serum Epo level		2 Increased serum LDH
	3 EEC growth		3 Anemia
			4 Palpable splenomegaly

WHO indicates World Health Organization; PV, polycythemia vera; ET, essential thrombocythemia; PMF, primary myelofibrosis; Hgb, hemoglobin; CML, chronic myelogenous leukemia; MDS, myelodysplastic syndrome; BM, bone marrow; Epo, erythropoietin; LDH, lactate dehydrogenase; EEC, endogenous erythroid colony.

* The diagnosis of PV requires meeting either both major criteria and 1 minor criterion or the first major criterion and 2 minor criteria.

† The diagnosis of ET requires meeting all 4 major criteria.

‡ The diagnosis of PMF requires meeting all 3 major criteria and 2 minor criteria.

§ Or Hgb or hematocrit greater than the 99th percentile of reference range for age, sex, or altitude of residence or red cell mass >25% above the mean normal predicted.

‖ Small to large megakaryocytes with an aberrant nuclear/cytoplasmic ratio and hyperchromatic and irregularly folded nuclei and dense clustering.

1.3.2 Atypical/unclassified MPNs

Non-classical or atypical MPN such as chronic eosinophilic leukaemia (CEL), hypereosinophilic syndrome (HES), chronic neutrophilic leukaemia or unclassifiable MPNs (uMPN) may overlap with MDS/MPN, in which proliferation is accompanied by dysplastic features or ineffective haematopoiesis in other lineages (Orazi & Germing, 2008; Ernst et al.).

The molecular pathogenesis of atypical MPN and MDS/MPN is only partially understood. In many patients, aberrant activation of tyrosine kinase signaling has been found as a consequence of four principal mechanisms: (i) activating tyrosine kinase mutations, e.g. FMS-like tyrosine kinase-3 (*FLT3)* and *JAK2* (Jones et al., 2005), (ii) mutations in downstream signaling components, e.g. *RAS* (Tyner et al., 2009), (iii) mutations in negative regulators, e.g. Casitas B-lineage lymphoma proto-oncogene (*CBL*) (Dunbar et al., 2008; Grand et al., 2009; Sanada et al., 2009) and (iv) constitutively active tyrosine kinase fusion genes arising as a consequence of genomic rearrangements (Ernst et al., 2010a). Collectively, however, these abnormalities account for less than 50% of cases. Thus, many other possible abnormalities remain to be discovered.

1.3.3 Classic BCR-ABL-positive CML

CML is a rare disease with its constant incidence worldwide, affecting approximately 0.6-2/100,000 of the population annually in all countries where statistics are adequate (Rohrbacher & Hasford, 2009). CML is rare below the age of 20 years but occurs at all decades of life, with a median age of onset around 50-60 years. The incidence is

slightly higher in males than in females. Patients are usually in the chronic phase when CML is diagnosed, this chronic phase lasts typically 2-7 years but it may, in rare cases, last more than 15 or even 20 years (Jones, 2010). Most CML cases occur sporadically. The initiating event or events are unknown: there are no known hereditary, familial, geographic, ethnic or economic associations. There may be an increased risk after exposure to the atomic bombs dropped on Hiroshima and Nagasaki but not with lower levels of radiation (Kamada, 2001; Tomonaga, 2001; Goldman, 1997). However, CML can be characterised by a consistent cytogenetic abnormality – a reciprocal translocation between the long arms of chromosomes 22 in the BCR gene and 9 in the abl oncogene, t (9; 22). The result is a shortened chromosome 22, known as the Philadelphia (Ph) chromosome (**Figure 1.7**) (Silver, 2000; Druker, 2008). The *BCR-ABL* fusion gene encodes a chimeric protein with strong tyrosine kinase activity. This constitutively active *BCR-ABL* tyrosine kinase causes CML but how the presence of this oncoprotein leads to the CML phenotype is not fully understood.

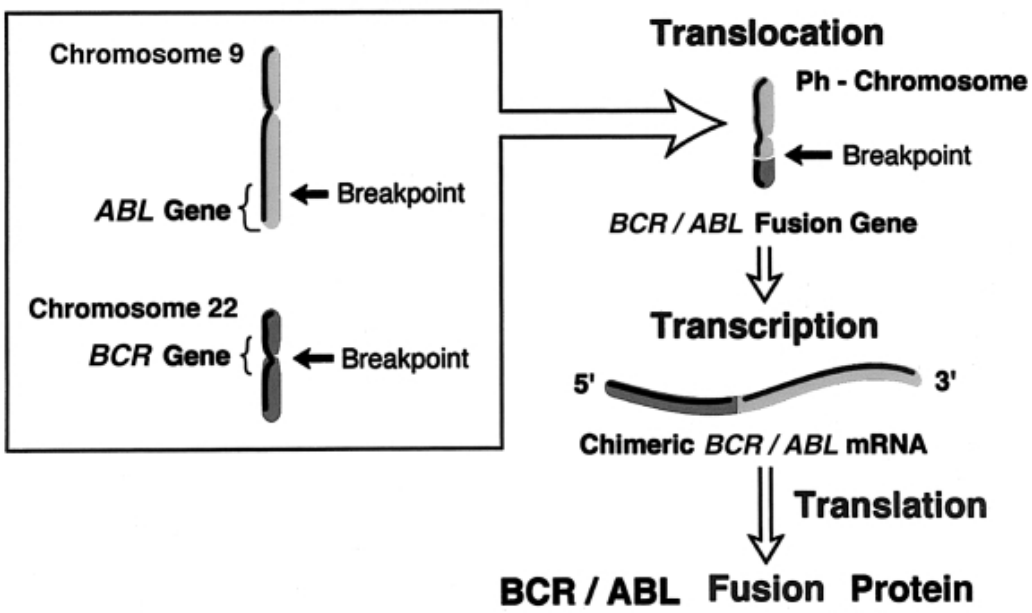


Figure 1. 7 Diagrammatic schema of Ph Chromosome translocation.

Adapted from Silver (2000)

.

1.3.4 Classic BCR-ABL-negative MPNs

In the past, the Philadelphia (Ph) chromosome or *BCR-ABL1* fusion gene was used as a marker to detect and confirm the diagnosis of CML, whereas the prediction and diagnosis of *BCR-ABL*-negative MPNs subtypes were based on their clinical and laboratory features supported by minor contributions from histopathology together with some genetic abnormality (Vardiman et al., 2009). Most Ph negative MPN patients carry an acquired mutation in the Janus Kinase 2 gene in their tumours. Polycythaemia vera (PV), essential thrombocythaemia (ET), and primary myelofibrosis (PMF) form the 3 classic MPNs that possess the same *JAK2* V617F mutation. This shared mutation has conferred some common features to the 3 disorders as listed in **Table 1.2** (Zhan & Spivak, 2009; Campbell & Green, 2006). The dysregulated signal transduction has also made the 3 distinguishable with their primary clinical features: increased red-cell mass (PV), high platelet count (ET), and bone marrow fibrosis (PMF) (**Figure 1.6**) (Tefferi & Gilliland, 2007). All the three disorders originated from multipotent haematopoietic progenitor cells, undergo relatively normal cellular maturation, and have overlaps in their phenotype and genotype presentations (Spivak & Silver, 2008).

This mutation was first discovered in 2005 by five research groups (Baxter et al., 2005; James et al., 2005b; Levine et al., 2005; Kralovics et al., 2005b; Zhao et al., 2005). It was reported in nearly all the patients with PV and in around 50% of those with ET or PMF (**Figure 1.6**) (Tefferi & Gilliland, 2007; Vannucchi et al., 2009a; Olcaydu et al., 2009a; Baxter et al., 2005; Tiedt et al., 2008; Xiao et al., 2008; Chen et al., 2007; Saxena R & SK., 2008). The mutation was also reported in Chinese populations with remotely different rates due to the differences in diagnostic criteria

and assay sensitivity (Lieu et al., 2008; Zhang et al., 2008; Xiao et al., 2008). This mutation is a single point mutation of the *JAK2* kinase gene (1849G>T) on chromosome 9 which results in a valine-to-phenylalanine substitution in codon 617 (encoding *JAK2* V617F, hereafter V617F). This acquired gain of function mutation constitutively activates the proliferation of myeloid cells (Tefferi & Gilliland, 2007; Jones et al., 2009). Comparing with other less frequent *JAK2* abnormalities found in MPNs (**Table 1.3**), only PV, ET, PMF show significant levels of V617F mutation (**Table 1.4**) (Tefferi & Gilliland, 2007).

MPNs are rare disorders with an incidence rate ranging from approximately 0.3 to 2.8 per 100,000 individuals depending on the subtype and geographical area (Jensen et al., 2000). However based on a survey from the North American Association of Central Cancer Registries, MPNs are among the most frequent haematologic neoplasms (Rollison et al., 2008). They estimated an average annual age-adjusted incidence rate of 2.1 per 100, 000 MPNs cases during the period from 2001 to 2003 (Vannucchi et al., 2009b). It is similar to the estimated annual incidence rate of 2.3 cases per 100, 000 in the United States (Goldin et al., 2009). MPNs mainly affecting older adults and modestly reduce patients' lifespan compared with the general population; however these disorders usually directly cause severe and potentially fatal complications to the patients such as blood clots in the arteries or veins and resistance to therapy (Jager et al., 2010; Passamonti et al., 2004; Vannucchi et al., 2009a). Results from a large Swedish study (Landgren et al., 2008) claimed a higher risk of 5.7, 7.4 and 7.5 for MPNs (PV, ET, and unclassified MPNs respectively) in the people of having family history of MPNs. The familial clustering as demonstrated by the Swedish study is in consistent with the hypothesis of the predisposition al-

lele (Vannucchi et al., 2009b). The genetic risk factors underlying these three MPNs will be discussed in the following sections.

Table 1.2 Common features of MPNs

Common features of PV, ET, and PMF
Involvement of a multipotent haematopoietic progenitor cell
Relatively normal cellular maturation
Dominance of the abnormal clone over normal clones
Abnormalities of chromosomes 1, 8, 9, 13 and 20
Marrow hypercellularity and megakaryocyte dysplasia
Haematopoietic growth factor hypersensitivity
Resistance to apoptosis
Growth factor-independent (endogenous) colony formation
Altered production of one or more of the formed elements of the blood
Thrombosis and haemorrhage
Myelofibrosis
Extramedullary haematopoiesis
Transformation but at low and differing frequencies to acute myeloid leukaemia
Expression of *JAK2* V617F, over-expression of *PRV-1* mRNA and impaired expression or mutation of *MPL* but not in all patients

Abbreviations: PV, polycythaemia vera; ET, essential thrombocythaemia; PMF, primary myelofibrosis; *PRV-1*: polycythaemia rubra vera-1; *MPL*: Myeloproliferative Leukaemia Virus

Table 1.3 Prevalence of *JAK2* abnormalities in haematologic neoplasms

Other *JAK2* mutations		
Haematologic neoplasm	Mutation	Reference
PV	***JAK2* exon 12 mutations** F537- K539delinsL H538QK539L K539L N542-E543del	Hellstrom-Lindberg and Cazzola (2008) Alabdulaali. (2009) McLornan et al. (2006) Steven et al. (2008)
AML UMPD	***ETV6-JAK2*** t(9;12)(p24;p13)	
AML UMPD	***PCM1-JAK2*** t(8;9)(p22;p24)	
UMPD	***BCR-JAK2*** t(9;22)(p24;q11.2)	
AML	***JAK2* T875N**	
ALL	***JAK2* rearrangement** t(8;9) t(9;12) *JAK2* R683G and less frequently other R683 point mutations	
Novel rare *JAK2* mutations (Yoo et al., 2009)		
Haematologic neoplasm	Mutation	Reference
PV	V536-I546dup11	Pietra et at. (2008)
PV	F537-I546dup10/F547L	Pietra et at. (2008)
PV	F537-K539delinsL	Scott et al. (2007) Pardanani et al. (2007) Pietra et at. (2008)
PV	H538Q/K539L	Scott et al. (2007)
PV IE	H538-K539delinsL	Williams et al. (2007) Martinez-Aviles et al. (2007) Pietra et at. (2008)
PV IE	K539L	Scott et al. (2007) Martinez-Aviles et al. (2007) Kouroupi et al. (2008)
PV	I540-E543delinsMK	Butcher et al. (2008) Pietra et at. (2008)

Table 1. 3 Prevalence of *JAK2* abnormalities in haematologic neoplasms (Continued)

Novel rare *JAK2* mutations (Yoo et al., 2009)		
Haematologic neoplasm	Mutation	Reference
PV IE	R541-E543delinsK	Williams et al. (2007) Butcher et al. (2008) Pietra et at. (2008)
PV IE	N542-E543del	Martinez-Aviles et al. (2007) Scott et al. (2007) Williams et al. (2007) Pardanani et al. (2007) Colaizzo et al. (2007) Percy et al. (2007) Kouroupi et al. (2008) Pietra et at. (2008)
PV IE	E543-D544del	Wang et al. (2008) Percy et al. (2007) Pietra et at. (2008)
AML without maturation	K607N	Lee et al. (2006)
ALL	L611S	Kratz et al. (2006)
PV	C616Y/V617F	Zhang et al. (2007)
PV	V617F/C618R	Karow et al. (2008) Yoo et al. (2009)
PV	V617F/D620E	Grünebach et al. (2006)
MPS unclassifiable	D620E	Schnittger et al. (2006)
BCR/ABL-negative MPS	E627E	Schnittger et al. (2006)
B cell precursor ALL	ΔIREED	Malinge et al. (2007)
AMKL	T875N	Mercher et al. (2006)

Abbreviations: PV, polycythaemia vera; AML, acute myelogenous leukaemia; UMPD, Unclassified myeloproliferative disorders, ALL, acute lymphoblastic leukaemia; IE, idiopathic erythrocytosis; MPS, myeloproliferative syndrome; AMKL, acute megakaryoblastic leukaemia.

Table 1.4 Worldwide prevalence of V617F in PV, ET, and PMF

Study	Population	Method	PV, n (%)	ET, n (%)	PMF, n (%)
Baxter et al. (2005)	Caucasian	AS-PCR DNA seq	71/73 (97)	29/51 (57)	8/16 (50)
James et al. (2005b)	Caucasian	DNA seq	40/45 (89)	9/21 (43)	3/7 (43)
Jelinek et al. (2005)	Caucasian		25/29 (86)	3/10 (30)	18/19 (95)
Jones et al. (2005)	Caucasian	ARMS	58/72 (81)	24/59 (41)	15/35 (43)
Kralovics et al. (2005b)	Caucasian	DNA seq	83/128 (65)	21/93 (23)	13/23 (57)
Levine et al. (2005)	Caucasian	MS, DNA seq	121/164 (74)	37/115 (32)	16/46 (35)
Zhao et al. (2005)	Caucasian	DNA seq	20/24 (83)		
Tefferi et al. (2006)	Caucasian		58/63 (92)		
Lippert et al. (2006)	France	AS-PCR, RT-PCR	60/62 (97)	45/60 (75)	
Chen et al. (2007)	Chinese (China)		42/57 (74)	40/68 (59)	8/12 (67)
Wong et al. (2008)	Chinese (Hong Kong)			60/95 (63)	
Lieu et al. (2008)	Chinese (Taiwan)		28/33 (85)	29/49 (59)	2/6 (33)
Xiao et al. (2008)	Chinese (China)		109/116 (94)	122/153 (79)	111/142 (78)
Zhang et al. (2008)	Chinese (China)		16/23 (70)	21/45 (47)	3/8 (38)
Basquiera et al. (2009)	Argentine	AS-PCR	40/45 (89)	30/43 (69)	7/15 (47)
Ruiz-Arguelles et al. (2009)	Mexican	ARMS	5/8 (63)	6/17 (35)	1/4 (25)
Shen et al. (2009)	Chinese (China)		35/35 (100)	53/85 (62)	2/3 (66)
Kim et al. (2010)	Korean	AS-PCR	22/24 (92)	12/26 (46)	47/89 (53)
Sazawal et al. (2010)	Indian	PCR-RFLP	28/34 (82)	7/10 (70)	16/31 (52)
Trifa et al. (2010)	Caucasian		61/69 (88)	37/64 (57)	9/15 (60)
Zhang et al. (2010)	Chinese (China)	AS-PCR, DNA seq, MS	73/89 (82)	52/142 (36)	24/47 (51)
Ayad and Nafea (2011)	Egyptian	ARMS	44/54 (81)	15/30 (50)	18/39 (46)
Mahfouz et al. (2011)	Lebanon	RT-PCR	13/13 (100)	28/41 (68)	
Wong et al. (2011)	Chinese (Singapore)			35/102 (34)	
Da Silva et al. (2012)	Brazilian	PCR-RFLP	46/52 (88)	39/81 (47)	8/11 (77)
Karkucak et al. (2012)	Turkish	ARMS	56/70 (80)	33/78 (42)	
Suksomyos et al. (2012)	Thais	AS-PCR, PCR-RFLP	25/31 (81)	29/49 (59)	7/10 (70)
This study	Chinese (Hong Kong)	ARMS	53/61 (87)	63/93 (68)	11/17 (65)

Abbreviations: PV, polycythaemia vera; ET, essential thrombocythaemia; PMF, primary myelofibrosis; AS-PCR, allele-specific polymerase chain reaction; DNA seq, DNA sequencing; ARMS, amplification-refractory mutation sequencing PCR; MS, mass spectrometry-based method; RT-PCR, real-time polymerase chain reaction; PCR-RFLP, polymerase chain reaction -restriction fragment length polymorphism.

1.3.4.1 Polycythaemia vera (PV)

Polycythaemia vera (PV) is the second most common phenotype among the 3 MPNs subtypes. The worldwide incidence of PV varied ranging from 0.7 to 2.8 cases per 100 000 individuals, with a slight male predominance (M: F, 1-2:1) and a median age of 55-60 years (Swerdlow et al., 2008; Johansson, 2006; Publicover & Medd, 2013). PV is recognised by its primary feature: an increased production of red blood cells in addition to its general distinctive features such as increased white cells and platelets (McLornan et al., 2006; Zhao et al., 2005). On top of that, itchiness and splenomegaly are other features of PV. PV is always complicated by thromboembolic phenomena and haemorrhage with the risk of developing myelofibrosis and acute leukaemia (McLornan et al., 2006). PV mainly affects older adults (uncommon in children) between 40 and 60 years old and the patients with this disorder have very sensitive responses in their haematopoietic progenitors to many growth factors and also cytokines (Zhao et al., 2005). PV is relatively indolent, resulting in a moderate reduction of lifespan (Vannucchi et al., 2009a) (median survival exceeds 15 years) (Tefferi, 2008) with increasing mortality rate in PV patients in an age-dependent manner (Cervantes et al., 2008). Criteria for PV classification based on the 2008 revision of the World Health Organization (WHO) are listed in **Table 1.1**.

1.3.4.2 Essential thrombocythaemia (ET)

Among the 3 MPN disorders, essential thrombocythaemia (ET) is the most commonly observed subtype with an annual incidence of 0.6-2.5 per 100,000 individuals. Most cases of ET occurred in patients aged 50-70 years, with both genders nearly equally affected (Johansson, 2006; Mesa et al., 1999; Jensen et al., 2000; Cervantes,

2011; Publicover & Medd, 2013). It is an indolent disorder like PV, normally recognised by a significant increased platelet count but it is clinically asymptomatic with normal life expectancy in most of the ET patients (Vannucchi et al., 2009a). Less than 10% of the patients were observed to transform into myelofibrosis at 10 years (Passamonti et al., 2004; Cervantes et al., 2002). Therefore, normal life expectancy decreases moderately after 10-15 years from diagnosis (McLornan et al., 2006; Vannucchi, 2009; Cervantes et al., 2008). However, detection of this disorder is possible with the thromboembolic events. Reported by McLornan et al. (2006), treatment modalities used can give a small tendency to ET to transform into myelofibrosis and acute leukaemia. Criteria for ET classification based on the 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukaemia (Vardiman et al., 2009) are listed in **Table 1.1**.

1.3.4.3 Primary myelofibrosis (PMF)

Primary myelofibrosis (PMF) is the least common subtype of MPNs with annual incidence estimated at 0.4-1.5 per 100, 000 persons. It is the most severe and also the most difficult to diagnose MPN. PMF cases were often reported nearly equally in both genders aged between 50-70 years (Kutti & Ridell, 2001; Publicover & Medd, 2013; Swerdlow et al., 2008). Bone marrow fibrosis is the primary distinguishing feature of PMF in addition to cytopenia, and splenomegaly (Baxter et al., 2005). PMF has the worst prognosis with a median survival of 5 years. Being the most severe type of MPNs, patients with PV and ET can develop secondary myelofibrosis and eventually transform into acute myeloid leukaemia in their late stage of PMF (McLornan et al., 2006; Vannucchi et al., 2009a). The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukae-

mia was used as the reference for PMF classification in this study (Vardiman et al., 2009) (**Table 1.1**).

1.3 The aetiology of MPNs

A hallmark of the classic MPNs is the overproduction of various mature functional blood cells. Despite knowing that MPNs are often associated with a range of mutations remarkably the V617F mutation, *JAK2* exon 12 mutations, and the aetiology of MPNs remains obscure.

In 2012, there was a critical review investigating environmental, lifestyle, and familial factors associated with MPNs (Anderson et al., 2012). They concluded that Jewish descent and also family with history of MPNs have strongest correlation with MPNs. Najean (1998) and Chaiter (1992) both identified a higher portion of MPNs patients to be Jewish decent compared with reference population in the 1990s. Individuals with familial history of MPNs were announced to have increased risk of developing MPNs compared with those without familial history (Landgren et al., 2008).

Autoimmune conditions particularly the non-malignant Crohn's disease (Barrett et al., 2008) was found to co-exist with MPNs. However, their relationship requires investigation to identify if there are both shared risk genes linking MPNs to Crohn's disease. Surprisingly, a higher incidence of PV was detected among blood donors compared with general population (Kristinsson et al., 2010; Merk et al., 1990). Health-conscious behaviour and cancer screening among blood donors might explain

the increased PV diagnosis among donor population (Merk et al., 1990; Vahidnia et al., 2013).

They also summarized that, higher mortality rates were recorded among poultry workers (Johnson et al., 2010), commercial pressmen (Zoloth et al., 1986), and petroleum refinery workers (Kaplan, 1986) compared with their reference populations. Workers exposed to metals/chemical were more common occupations among MPN patients (Giles et al., 1984; Terreros et al., 1997). Interestingly, occupation such as professional, administrative clerical, sales, transportation/communication, service industries, food and beverage workers, and hairdressers were not correlated with MPNs (Giles et al., 1984).

To date, despite the identification of somatic and germline mutations surrounding *JAK2*, and possibly some potential environmental risk factors, the aetiology of MPNs is still not clear. Larger scale studies may possibly explore more information underlying MPNs.

1.4 Targeted treatment of MPNs

For decades, there have been treatments developed for MPNs patients. However, many new drugs still remain investigational and the safety demonstration has to be ensured in patients to avoid any substantial toxicities induced by *JAK2* inhibition in the treated patients (Kilpivaara & Levine, 2008). Though the considerable devotion to *JAK2* inhibitor development of MPNs (Verstovsek, 2009) has been on its way together with clinical studies, the possible *JAK2* inhibitors resistance in some patients

must not be ignored. Moreover, it was reported that *JAK2* inhibitor therapy in V617F-positive patients might increase their risk of leukaemic transformation (V617F-negative) (Kilpivaara & Levine, 2008; Kralovics et al., 2005b). To date, in vitro studies and also murine models of V617F-induced MPNs suggests that small molecular inhibitors might be useful in blocking *JAK2* kinase activity yet reduce the complications of these disorders (Skoda, 2008). This can be achieved by reducing the cellular mass responsible for the typical clinical complications. However, considering the long life expectancy of MPN patients (Skoda, 2008), more in vivo studies should be performed to improve the therapy or treatment without complications and also to avoid any transformation to more seriously diseases.

As reviewed by Zhan and Spivak (2009), overall survival in the 3 MPNs ranges from a normal life expectancy in ET patients to a median of less than 6 years in many patients with PMF. Disease transformation among the 3 MPNs is another major complication in addition to transformation from MPN to other less common MPNs such as acute myeloid leukaemia/myelodysplastic syndromes as a side effect of frequent use of the drugs that are themselves leukaemogenic in nature as reported (Clark et al., 2007; Najean & Rain, 1997; Berk et al., 1981).

More studies in the development of therapeutic treatments should be carried out to significantly reduce MPNs with the exclusion of unwanted complications. Therapeutic treatment of MPNs is not the focus of this project, therefore this topic will not be discussed into details.

1.5 The Janus Kinase (*JAK2*) gene

JAK2 is a member of the intracellular non-receptor tyrosine kinases consisting of four members, namely Janus kinases (*JAKs*) 1, 2, 3 and tyrosine kinase 2 (*TYK2*). They have seven domains and are located beneath cellular receptors, as the controller of signal transmission downstream (Alabdulaali, 2009).

In mammals, *JAKs* not only mediate normal functions but also play key role in signal transduction pathways. They are mainly involved in the Janus kinases and signal transducers and activators of transcription (JAK-STAT), phosphatidylinositol 3 kinase (PI3K), and RAS and mitogen-activated protein kinase (RAS-MAPK) by associating with cytokine, growth factor, and interferon (Neubauer et al., 1998; Rane & Reddy, 2000). *JAKs* are therefore crucial components in cell growth, survival, proliferation, differentiation, and apoptosis (Rane & Reddy, 2000). In mammals, despite the fact that *JAK3* is primarily expressed in haematopoietic cells, all the other *JAKs* are ubiquitously expressed (Rane & Reddy, 1994; Witthuhn et al., 1994; Schindler et al., 2007). Although not specifically restricted to haematopoietic tissue, the other members of *JAKs* are also indispensable in the development, function, survival, and differentiation of haematopoietic system (Jones, 2010). Disruption in the activity of *JAK* kinase may lead to abnormal cellular response and diseases (Rane & Reddy, 2000). Experiments showed that loss of *JAK1, JAK3,* or *TYK2* destructed lymphoiesis while loss of *JAK2* results in embryonic lethality. Cells cannot survive without definitive erythropoiesis (Neubauer et al., 1998).

JAK2 was first cloned in 1989 (Wilks, 1989) and originally named ‘just another kinase’ and was later renamed Janus kinases after the two-faced Roman God of gates

and passages. It was named so because of its ‘active’ and ‘inactive’ domains (JH1 and JH2) that resembles each other, reflecting the ability of the God Janus to look in two directions at once (McLornan et al., 2006). The structure of *JAK2* is made up of *JAK* homology 1 (JH1) which has a tyrosine kinase catalytic activity (Ihle & Gilliland, 2007b), the *JAK* homology 2 (JH2), a catalytically inactive pseudokinase domain which has similar amino acid sequence to JH1 that exerts an inhibitory effect on basal *JAK* activity by directly inhibiting JH1, a SRC homology 2 domain (SH2), and an amino terminal (NH-2) FERM (4-point-1, Erzin, Radixin, Moesin) homology domain (**Figure 1.8)** (McLornan et al., 2006; Ihle & Gilliland, 2007b; Vainchenker et al., 2008). In 1992, Pritchard and his colleagues successfully mapped the gene on the short arm of chromosome 9p24 (Pritchard et al., 1992). It consists of 140 kb DNA spanning 25 exons and encodes 1132 amino acid-long *JAK*2 protein (Saltzman et al., 1998). Summarised from previous studies by Alabdulaali (2009), *JAK2* works as a signalling molecule for many cytokines and STAT5 is one of the most important pathways activated by *JAK2* in addition to the activation of *BCL-XL* and up-regulation of *BCL2* for cell survival.

JAKs are important because they transmit signals in the inactive cytokine receptors that lack intrinsic catalytic activity. *JAK2*, as an indispensable signal transductor, mediates tyrosine phosphorylation by associating with cytokine receptors. Binding to specific domains of receptors, suppression of activation by pseudokinase (JH2) do-main, and phosphorylation within the activation loop are prerequisite interactions to regulate kinase activity. These interactions activate some downstream signalling pathways by associating with cytokine receptors mediated by JH6 and JH7 domain (Saharinen et al., 2003; Ihle & Gilliland, 2007b). The NH-2 terminal FERM domain

is employed for the interaction and binding of cytokine receptors in addition to its role in trafficking of the cytokine receptor, such as erythropoietin receptor (EPOR) cytoplasmic domain, to the cell surface (McLornan et al., 2006; Constantinescu et al., 2008) (**Figure 1.9)**. The inactive JH2 domain is unique to this *JAK* family and functions by negatively regulating the cytokine receptors in the absence of ligand binding but constitutively activating the kinase in the absence of ligand binding (Baker et al., 2007). Maximal *JAK2* activity can only be achieved with its intact structure (McLornan et al., 2006). Deletion of pseudokinase domain (the JH2) in *JAK2* and *JAK3* has been reported to increase phosphorylation and co-expression of STAT5 (Saharinen & Silvennoinen, 2002). The critical regulatory function of this domain was further demonstrated when majority of patients with PV and approximately half of ET and PMF patients were detected to harbour the V617F activating mutation in the JH2 domain, despite the underlying mechanism being unknown (Silvennoinen et al., 2013).

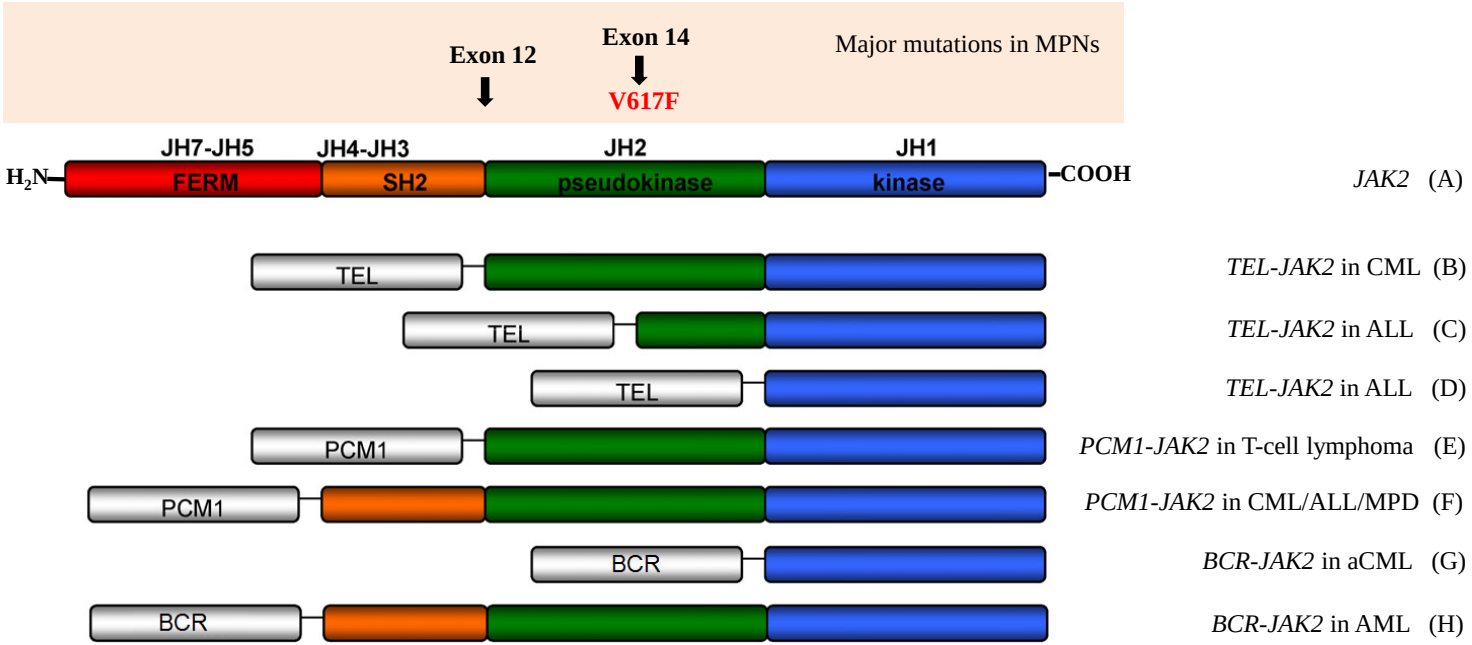


Figure 1. 8 Schematic representation of *JAK2* protein, with its mutations and *JAKs* fusion proteins reported in myeloproliferative disorders.

JAK tyrosine kinase is composed of seven *JAK* homology (JH) domains, JH1 to JH7 (panel A). The common MPN-associated V617F mutation is located in exon 14 within the JH2 domain. A small percentage of V617F-negative PV patients was documented to harbour exon 12 mutation, which lies within the JH2 and SH2 domain (Pietra et al., 2008; Scott et al., 2007). The *TEL-JAK2* fusion proteins reported in CML (panel B) and ALL (panel C and D) contain the intact JH1 domain, but different portions of JH2 (Peeters et al., 1997; Lacronique et al., 1997). The *PCM1-JAK2* fusion proteins contain the entire JH1 and JH2 domains with parts of the SH2 domain reported in cases with T-cell lymphoma (Adelaide et al., 2006) (panel E); not more than JH2 domain in cases of aCML/ALL/MPD (Andreas et al., 2005; Murati et al., 2005; Bousquet et al., 2005) (panel F). The fusion *BCR-JAK2* was identified in a case with aCML (Griesinger et al., 2005) and also AML (Cirmena et al., 2008) (panel I and J). Adopted and modified from Vainchenker et al. (2008)

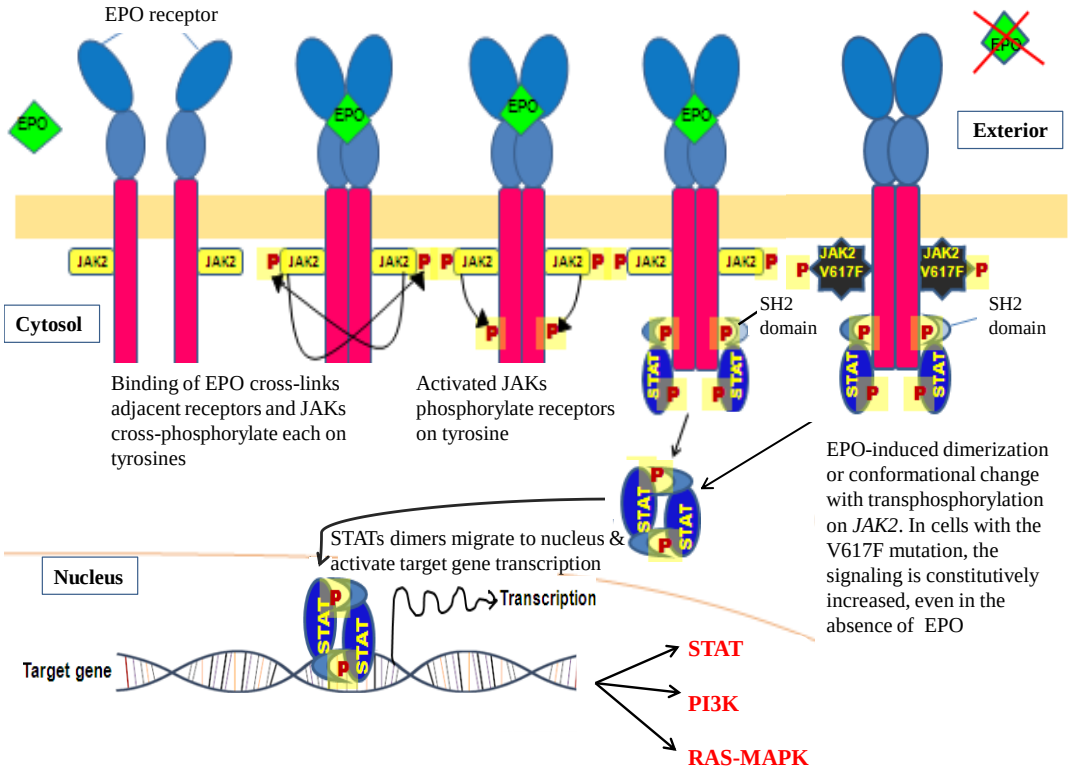


Figure 1. 9 An overview of *JAK2* signalling pathway with the example of growth hormone erythropoietin (EPO).

The EPO receptor is present at the cell surface and forms homodimer in the presence of EPO (ligand) (McLornan et al., 2006; O'Sullivan et al., 2007). Ligand/cytokine (e.g. erythropoietin, thrombopoietin) binding triggers auto (trans)-phosphorylation of *JAKs* bound to the extracellular domain of the cytokine receptors, and activates the *JAK* kinase activity with a conformational change. The activated *JAKs* in turn phosphorylate the cytokine receptors on tyrosines and causes the docking of numerous downstream signalling proteins (McLornan et al., 2006) such as STATs (signal transducer and activator of transcription), MAPK (mitogen-activated protein kinase) and PI3K (phosphatidylinositol 3-kinase) molecules on specific phosphotyrosines. They are then activated. Once activated, these proteins dissociate from receptor, dimerise via their SH2 domains and migrate to nucleus as transcription factors in the modulation of target gene expression. The most right part of the figure shows the hyper-activation of *JAK2*-induced pathway independent of ligand binding. Modified from James et al. (2005a) and Campbell and Green (2006)

1.5.1 The *JAK2* V617F mutation

The history of MPNs dates back actually when an American haematologist, William Dameshek recognised that PV, ET, and PMF might share a pathogenic mechanism (Pesu et al., 2005). The aetiology of MPN is largely unknown, except that radiation and toxic chemicals like benzene were reported to be associated with a higher risk of leukaemia. Almost half of the PMF patients were reported to have non-specific chromosomal abnormalities; while among patients with PV, cytogenetic abnormalities are relatively rare and random but the incidence increases as the disease progresses. In the later stage of PV, patients may eventually develop MDS or acute leukaemia (Rodriguez-Abreu et al., 2007; Hemminki et al., 2009).

In 2005, V617F activating point mutation was postulated to be involved in the aetiology of MPN (Kralovics et al., 2005b; James et al., 2005b; Levine et al., 2005; Baxter et al., 2005; Kilpivaara & Levine, 2008). Kralovics et al. (2005b) concentrated on *JAK2* study based on the loss of heterozygosity (LOH) on *JAK2* gene. James et al. (2005b) investigated the effect of *JAK2* siRNA on endogenous erythroid colony (EEC) formation whereas the Boston group (Levine et al., 2005) studied the structure of tyrosine kinases using high-throughput DNA sequencing approach. On the other hand, Baxter et al. (2005) and Zhao et al. (2005) had the same idea: *JAK2* is a candidate gene in the pathogenesis of PV, ET, and PMF given its pivotal role in haematopoiesis. Consistently, the same conclusion was drawn: V617F mutation was detected in almost all PV patients and nearly 50% of patients with ET and PMF, in their JH2 domain of *JAK2*. Strikingly, the presence of V617F was completely undetected in normal control samples (Pesu et al., 2005). It was exclusively present in only myeloid lineages with a high frequency in PV, ET, PMF (Alabdulaali, 2009) (likely to

arise in the haematopoietic stem cell compartment) (Levine & Wernig, 2006) but not in T cells, buccal cells or hair-follicle cells (Kralovics et al., 2005b; Baxter et al., 2005; McLornan et al., 2006) and has the ability to constitutively activate down-stream STAT, ERK/MAP kinase and P13K/AKT pathways in the absence of ligands/cytokines. It was rarely reported in CML and other rare MPNs.

V617F is a somatically acquired G to T transversion at nucleotide position 1849 on exon 14 which results in a substitution of valine for phenylalanine at codon 617 within the JH2 domain of *JAK2* (**Figure 1.8**). The JH2 domain is believed to be auto-inhibitory. Thus, the valine to phenylalanine substitution at codon 617 disrupts the auto-inhibitory control of *JAK2*. Consequently, *JAK2* gene is constitutively being activated (phosphorylated) even in the absence of haematopoietic growth factors, resulting in a gain-of-function mutation (Ihle & Gilliland, 2007b; Kralovics et al., 2005b; James et al., 2005b). This acquired mutation has conferred some shared features (**Table 1.2**) and distinguishable phenotypic characteristics to the 3 disorders due to the dysregulated signal transduction (Tefferi & Gilliland, 2007) in addition to their primary clinical features: increased red-cell mass (PV), high platelet count (ET), and bone marrow fibrosis (PMF).

Wild-type *JAK2* allele (heterozygous V617F) is often reduced to homozygosity in PV and PMF cases but this is rare in ET. This is resulted from mitotic recombination (generally referred to acquired uniparental disomy; UPD) at chromosome 9 (Scott et al., 2007). Patients with the loss of wild-type *JAK2* allele at this region are always homozygous for the V617F mutation. They carry only one single parental haplotype from one parent that acquired the mutation. This difference suggested that mitotic

recombination that causes V617F homozygosity could be an early genetic event underlying the development of PV among the MPN subtypes. In addition to V617F somatic mutation, inherited MPN predisposition loci were suggested from previous familial clustering studies (Kralovics et al., 2005b; Kilpivaara et al., 2009; Campbell, 2009; Olcaydu et al., 2009a; Jones et al., 2009). In humans, V617F mutation occurs at the stem cell level and is present in haematopoietic stem cell progenitors (Baxter et al., 2005).

To date, in vivo and in vitro assays have been performed to study the relation between V617F and MPNs. In vivo assays of V617F expression in murine bone marrow transplantation assay is by far the most contributing type of study in validating V617F as an oncogene (Kilpivaara & Levine, 2008). The association between the mutation and MPNs has been observed in murine models (Levine & Wernig, 2006). The studies using murine models were in parallel with other human clinical studies, supporting the non-redundant central role of V617F in the pathogenesis of myeloproliferative diseases. It was noted that after V617F expression through a bone marrow transplantation assay, the recipient mice developed PV and eventually myelofibrosis (Levine & Wernig, 2006). Afterwards, Tiedt et al (2008) reported that phenotypic manifestation of MPNs can be determined by V617F allele burden: transgenic mice with low levels of V617F developed an ET-like phenotype while transgenic mice with higher levels of V617F developed a PV-like phenotype. On the whole, V617F mutation was sufficient to cause MPN-like phenotype and the V617F allele burden impacts the phenotype manifestation of MPN subtypes, as suggested by transgenic mice studies.

1.5.1.2 Effect of V617F mutation on patient survival

The acquired somatic mutation V617F is associated with PV, ET, and PMF. At the cellular level, the mutation confers proliferative and survival advantage by making the cells more sensitive to incoming stimuli (Speletas et al., 2007). The haematopoietic progenitor cells thus proliferate under clonal expansion with selective growth advantage that shapes the cancer genome of an individual resulting in each unique phenotype (Kralovics, 2012).

Clinically, V617F-positive patients were generally older individuals displaying higher levels of haematocrit (Ht) and haemoglobin (Hb), high white blood cell count (WBC); low erythropoietin (EPO) levels and low platelet (PLT) count (Speletas et al., 2007; James et al., 2005b; Kittur et al., 2007; Palandri et al., 2009; Campbell et al., 2006). It was also reported that V617F-positive patients encounter more complications, they are more vulnerable to leucocytosis, splenomegaly, and thrombotic events compared with those without the mutation (Speletas et al., 2007; Campbell et al., 2005).

Phenotypically, Kralovics et al (2005b) observed that patients who are homozygous for V617F mutation had a longer disease duration compared with those who are heterozygous. These patients were also had higher incidence rate of developing secondary myelofibrosis which is the most severe disorder among the three. Likewise, Tefferi et al (2006) demonstrated that PV patients with homozygous V617F displayed more advanced clinical presentation or complications such as higher rates for pruritus and also a higher degree of fibrosis. The latter was also noted among PMF cases with V617F compared with cases without mutation (Santos et al., 2011). It was just re-

ported this year in a study investigating PV and ET that thrombosis was correlated with V617F mutation (Payzin et al., 2014). All evidence reinforces the idea that V617F-homozygosity might cause a longer duration and more advanced disease than V617F-heterozygosity. In spite of this, studies of relevant murine models (Shide et al., 2008; Tiedt et al., 2008; Tefferi et al., 2010b) also showed that V617F allele burden might be important in disease phenotype specification. However, the results are as inconsistent as that from human studies.

Disease transformation to myelofibrosis (MF), MDS, or most critically AML, is known to be the long term complications of PV and ET (Campbell & Green, 2006). Campbell et al. (2005) observed six transformation events from V617F-positive ET to PV. Separately, Wolanskyj et al. (2005) showed that ET-to-PV transformation was more likely to happen in patients with V617F mutation. Because of this, ET and PV are being hypothesized a continuum of disease. This also marked the possibility that V617F might serve as a possible factor governing the transformation event of different MPNs alongside with other potential influential factors. Statistically, Passamonti et al. (2010) reported a progression risk from PV to MF in patients with more than 50% of V617F allele burden. These findings affirmed previous observations that increasing V617F allele burden increases the risk of disease progression to MF; however it remains unexplained for the transformation to MDS/AML.

The possible relationship between V617F and overall survival of MPNs patients were controversial. As a case in point, Campbell et al. (2006) reported based on their study consisting of 152 patients that, V617F-positive PMF patients had poorer survival. On the other hand, Tefferi et al. (2005) and Cervantes et al. (2009) found no

prognostic impact of V617F. On top of that, Cervantes et al. (2009) pointed out that the presence of V617F was not associated with survival among the 345 patients. Similarly, Barosi et al. (2007), in their study of 174 patients, found no correlation between V617F mutation and survival but reported V617F as a predictive factor for splenomegaly, need of splenectomy, and leukaemic transformation. While this may be true, leukaemic transformation or patients' overall survival do not seem to be associated with V617F mutant allele (Kittur et al., 2007; Tefferi et al., 2008; Wolanskyj et al., 2005). However, lower V617F allele burden was found to shorten the survival in PMF patients (Guglielmelli et al., 2009).

Adding to the idea, *JAK2* mutational status also affects treatment responses (Campbell et al., 2005). It was observed that V617F-positive ET patients are more sensitive to hydroxyurea (Hydroxycarbamide; HU), an inexpensive drug widely used in treating MPNs to control splenomegaly, leucocytosis, and thrombocytosis but not anagrelide, compared with V617F-negative patients. A low dose of HU can reduce their platelet counts, WBC counts, and haemoglobin concentration in a great extent compared with V617F-negaive patients. Their prevalence of arterial thrombosis was reduced after the treatment without having to receive heavy dose of drugs (Campbell et al., 2005). Moreover, V617F-positive ET patients showed better therapeutic response to HU (Martinez-Trillos et al., 2010). However, documented by Skoda (2008) in a study focusing on *JAK2* inhibitors, cells carrying V617F mutation were more sensitive to the inhibition by lestaurtinib than cells without mutation. Strikingly, cells from V617F-negative MPN patients showed response to lestaurtinib as well. This aroused the curiosity that V617F may not directly cause the constitutive activation of *JAK2* signalling pathway but this hyper active pathway is a rather common event in the pathogenesis in MPN (Skoda, 2008).

Despite the above findings, the conclusion is inconsistent and could not explain the effects of V617F mutation on patients' overall survival.

1.5.2 Additional mutations in MPNs leading to *JAK2* activation

MPNs manifest as a group of blood disorders with complex genetic background. Markedly, the V617F mutation was found in almost all patients with PV but only nearly half of the patients with ET and PMF (Oh et al., 2010a). Adding to the idea, the leukaemic clones of some V617F-positive MPN patients who transformed into AML were found to be V617F-negative (Theocharides et al., 2007). In 2007, Scott and colleagues identified a set of *JAK2* exon 12 mutations in V617F-negative patients with PV (Scott et al., 2007). Various exon 12 mutant alleles induced cytokine-independent/hypersensitive proliferation in erythropoietin receptor-expressing cell lines and constitutively activated the JAK-STAT signaling. Unlike V617F, *JAK2* exon 12 mutations are only observed in V617F-negative PV (Scott et al., 2007; Butcher et al., 2008). This led to more studies on MPNs and the identification of additional signaling mutations as listed below:

- *MPL* Mutations: activating mutations in the thrombopoietic receptor gene *MPL*, which is *JAK2*-mediated, in 5-10% of patients with V617F-negative ET and PMF (Pikman et al., 2006; Rumi et al., 2013);

- *CBL* Mutations: rare mutations in E3 ubiquitin ligase Casitas B-lineage lymphoma proto-oncogene (*CBL*) that are mainly associated with myelofibrosis and indicate a poor prognosis (Velazquez et al., 2002; Grand et al., 2009);

- *SH2B3* Mutations: loss-of-function mutation in the JAK-STAT inhibitory adaptor protein *SH2B3* (*LNK*) with a low frequency (3%) (Oh et al., 2010b);

- Coding Calreticulin Mutations: recent exome sequencing identified recurrent Calreticulin (*CARL*) mutations in 60-84% of MPN patients with wild type *JAK2* and *MPL*. The mutant proteins result in an activated JAK/STAT pathway (Nangalia et al., 2013; Klampfl et al., 2013).

1.5.3 Aditional mutations in other pathways in MPNs

Mutations affecting the epigenome and spliceosome were also reported in MPN patients (Martinez-Aviles et al., 2012). These mutations were suggested to be the primary events leading to MPN or they can be acquired as secondary events causing a leukemic transformation. The fact that these mutations were primarily found in myelofibrosis, has suggested that myelofibrosis is a rapidly progressive type of MPN. The most frequent mutations include:

- *TET2* Mutations: being the most prevalent epigenetic mutation in MPNs, loss-of-function mutations in Ten-Eleven Translocation-2 (*TET2*) was found in 4% of ET, 10-16 % of PV, and 8-15% of PMF. Mutations in *TET 2* affect its active DNA demethylation (Delhommeau et al., 2009);

- *DNMT3A* mutations: mutations in DNA (cytosine-5-)-methyltransferase 3 alpha (*DNMT3A*), methyltransferase that adds methyl groups to cytosine in CpG island to effect gene expression, affect the state of DNA methylation (Abdel-Wahab et al., 2011b);

- *ASXL1* mutations: heterozygous, loss of function mutations in the polycomb associated protein Additional Sex Combs-Like 1 (*ASXL1*), a tumour suppressor gene, lead to *ASXL1* haploinsufficiency and promote progression to myelofibrosis or a leukemic transformation (Gelsi-Boyer et al., 2009);

- *IDH-1/2* Mutations: gain-of-functions mutations in Isocitrate dehydrogenase 1 and 2 (*IDH1* and *IDH2*) produce 2-hydroxyglutarate which in turn inhibits *TET2* (2-oxyglutarate dependent enzyme) and lead to the progression to myelofibrosis or a leukemic transformation (Figueroa et al., 2010);

- *EZH2* mutations: loss-of-function mutations in Enhancer Of Zeste Homolog 2 (*EZH2*), proven and putative chromatin modifiers, promote progression to myelofibrosis or a leukemic transformation; they are also associated with a poor prognosis in PMF (Guglielmelli et al., 2011).

1.6 Approaches used to study the germline polymorphisms underlying MPNs

The mechanism of the identical somatic mutation in the *JAK2* tyrosine kinase gene (V617F) observed in the three phenotypically distinct disorders, namely PV, ET, and PMF, remains a major hurdle in the pathogenesis of MPNs. In 2008, World Health Organisation included V617F as one of the diagnostic criteria for this group of MPNs (Tefferi & Vardiman, 2008). Subsequently, it was demonstrated by the Swedish familial clustering study that, there is a higher risk of developing MPNs among the first-degree relatives of MPN patients (Landgren et al., 2008). These epidemiologic

data led to the hypothesis that additional genetic events might contribute to the underlying pathogenesis of MPNs

In 2009, the *JAK2* germline haplotype block (or '46/1') was reported to be associated with the predisposition of the development of V617F-positive MPNs almost simultaneously by three independent groups (Jones et al., 2009; Olcaydu et al., 2009a; Kilpivaara et al., 2009). This MPN-specific haplotype block is intronic and located between exons 12 and 13 of *JAK2* and can be tagged by single nucleotide polymorphisms (SNP) for instance rs10974944, rs12343868, and rs12340895 (Olcaydu et al., 2009a; Jones et al., 2009; Kilpivaara et al., 2009). The three studies all demonstrated that the GG/GC genotypes with the dominant G allele are predisposing factors to a threefold to four-fold increased risk of developing clinical MPN disease (Olcaydu et al., 2009a; Abdel-Wahab; Jones et al., 2009; Trifa et al.). The haplotype tagging SNP (tSNP) rs10974944 resides in the same haplotype block with V617F mutation. It means that the risk allele might act on disease predisposition along the same strand of DNA with the V617F mutation. It was confirmed by allele-specific PCR that the V617F mutation was acquired in cis with the risk allele which concluded that the *JAK2* haplotype block increases the risk of developing V617F mutation in the same chromosome (Kilpivaara et al., 2009).

Despite knowing their location is distinct from *JAK2* promoter and 5' exons, the mechanism or interaction between this *JAK2* haplotype block and the V617F mutation leading to MPNs remains another hurdle. There are currently two most possible mechanisms (Abdel-Wahab; Olcaydu et al., 2009b; Kilpivaara et al., 2009):

1) The "hyper-mutability" hypothesis. The haplotype confers increased chances of acquiring mutations at the *JAK2* locus, and those mutations have selective growth advantage specifically for V617F, ultimately leading to clonal disorder. Evidence fortifying this hypothesis is the preferential acquisition of *JAK2* mutation in independent clones of 46/1 allele among MPN patients. However, no mechanism has been proposed to explain the genetic instability at *JAK2* locus which the haplotype offers essential features favouring mutations even at a genetic distance (Campbell, 2009). Moreover, the haplotype was recently also detected not only in some V617F-negative MPNs (Tefferi et al., 2010b; Pardanani et al., 2010) but also the non-malignant Crohn's disease (Barrett et al., 2008).

2) The second hypothesis is the "fertile ground" hypothesis. This hypothesis suggests the haplotype might offer a selective growth advantage to the cells where mutations such as V617F and exon 12 preferentially occur albeit these mutations can arise freely on different haplotypes. Similarly, this hypothesis is also not proven. Even without proposed mechanism, the increased risk of *JAK2* exon 12 and *MPL* mutations by the haplotype favour the fertile ground hypothesis over the hypermutability hypothesis (Olcaydu et al., 2009b). Nevertheless, there might be additional disease-associated functional variants that are linked with the haplotype that are yet to be identified.

As mentioned, environmental factors including ionising radiation, exposure to chemicals and drugs, infection (during pregnancy, viral and bacterial) increased the risk of developing haematologic malignancy (Hoffbrand & Moss, 2011a). Adding to the

findings of germline genetic underlying the pathogenesis of MPNs, there might be other potential environmental factors or unidentified genetic events that makes the carriers more vulnerable to V617F mutation (Abdel-Wahab & Levine, 2011). Recent epidemiologic data from three counties in eastern Pennsylvania where concerns have been expressed in the exposure to hazardous material has suggested a link between environmental factors and MPNs V617F analysis helped to identify a cluster of PV cases not identified by traditional clinical testing. The incidence of PV is significantly higher in these hazardous sites than the other areas in the United States (Seaman et al., 2009). On top of that, five patients from Denmark were diagnosed with both MPNs and multiple sclerosis (MS). This unusual connection between the 2 diseases has also brought attention to the possibility of yet identified predisposing factors especially environmental factors or genetic alterations (Thorsteinsdottir et al., 2013).

Nevertheless, there is currently no identification of the involvement of aetiologic environmental factor. The initiating genetic/environmental events responsible for the development of MPNs are still not totally understood.

1.6.1 Case-control association study

Traditionally, identification of disease genes that are transmitted within the families begins with linkage analysis. Linkage analysis aims to identity the genomic region of the disease genes and to establish a linkage between the genes, chiefly in simple Mendelian inheritance (single gene inheritance) (Silverman & Palmer, 2000). This can best be illustrated with the identification of cystic fibrosis gene in the late 1980s (Rommens et al., 1989). However, this powerful method has its own limitations. One major limitation is the difficulty in fine mapping of disease genes underlying a com-

plex disorder. The main limitation of linkage analysis, however, is solvable with association studies.

Association analysis is more powerful than linkage analysis when it comes to localizing the disease genes or genetic determinants of complex disorders and those with weak effect. The smaller sample size required also makes it a better method to detect association than to detect linkage (Risch & Merikangas, 1996). They also pointed out that, by using candidate gene approach in association study even at whole genome scale, association study is still a more powerful method in the analysis of complex disorders.

Case-control association study or population candidate gene association study is genetic association study that aims to identify association between genetic determinants and a trait. Genetic polymorphism is the most commonly used genetic determinant (Silverman & Palmer, 2000). The trait can be a disease (MPNs in this case), a quantitative characteristic (e.g., height and weight) or a discrete attribute. The usual way of performing case-control study is to compare possible risk factors between a group of unrelated affected individual (cases) and a group of unrelated unaffected individuals (controls) (Silverman & Palmer, 2000).

Genetic polymorphisms examined in such study are usually selected from candidate genes that are usually reported to be associated with the disease. Then any difference in the allele or genotype frequencies between cases and controls will be analysed with proper statistics. Statistical significant differences in allele or genotype frequencies between the two groups suggest either that (1) the polymorphism is the causal

variant that directly affects the phenotype of the trait; (2) the polymorphism is not directly involved in the trait, it is in linkage disequilibrium (LD; the degree of an allele of one SNP on the same chromosome that is co-inherited with another allele of another SNP within a population) with the real causal variant or a disease susceptibility gene, or (3) the finding could be an artefact due to confounding factors for instance population stratification, selection bias or genotyping errors (Silverman & Palmer, 2000).

However, no single analysis is perfect, so as association studies. To ensure the detection of a genuine association, few factors should be carefully taken when performing case–control association studies. Potential factors such as selection criteria of case and controls, selection of candidate genes, functional significance of polymorphisms chosen for study, and statistical analysis will be discussed in the next session.

Following experiments, appropriate statistical tests should be used to defend against spurious results in a case-control study. Stringency of the statistical test used should always be met before any significance can be concluded. Before a statistical test can be regarded a real significance, multiple comparisons must be corrected. There are no fixed methods to apply. The simplest and most conservative approach is the Bonferroni correction which adjusts the alpha value from $\alpha = 0.05$ to $\alpha = (0.05/k)$ where k is the number of tests being conducted. (Bush & Moore, 2012). A genome-wide association studies (GWAS) examines 500,000 to one million SNPs. Therefore, for a typical study investigating 500,000 SNPs, the significance threshold will be 1×10^{-7}(Bush & Moore, 2012). There thresholds vary among different studies, however *P* value around the 5×10^{-7} or 10^{-8} are usually used (Risch & Merikangas, 1996;

Dudbridge & Gusnanto, 2008). Another commonly used correction method is called permutation. It is more computationally intensive but straightforward method to generate the empirical distribution of test statistics from the input dataset when null hypothesis holds true. Permutation breaks the genotype-phenotype relationship among the cases and controls, and randomly swaps the case and control status for a predefined number of times (N) to produce an empirical distribution with N resolution(Bush & Moore, 2012). As an illustration, a permutation with N = 1500 generates an empirical *P* value within $1/1500^{th}$ of a decimal place.

Finally, a particular genetic association should be replicable in independent studies using independent samples. Replication following the study design in additional independent samples is the gold standard to validate any genetic association (Bush & Moore, 2012). Functional assessment is then the next step towards understanding the responsible mechanisms behind the observed associations.

1.6.2 Amplification refractory mutation system (ARMS)

Amplification refractory mutation system (ARMS)-PCR was used for V617F detection due to the limited restriction enzyme cutting sites for this mutation. ARMS-PCR detection of V617F provides an advantage to the small amounts of mutant DNA in a wild type background based on its high sensitivity. HotStarTaq Plus DNA polymerases were used to reduce nonspecific primer annealing (Baxter et al., 2005; Steensma, 2006).

First described by Newton et al. (Newton et al., 1989a) in his study of α-1 antitrypsin deficiency, amplification refractory mutation system (ARMS) successfully analysed the single nucleotide differences in patients' DNA. Since then, ARMS has been widely applied in the diagnosis of many genetic disorders (Old, 1996; Old et al., 1990; Newton et al., 1989b). It is a simple and rapid method of mutation detection. One prerequisite for successful ARMS is that, the system must be highly sensitive and specific so that during PCR, DNA polymerase extends these primers only when there is perfect annealing of oligonucleotide primers at their 3' ends of the sequences (Newton et al., 1989a; Steensma, 2006). The special mechanism permits the detection of a single base change under optimised PCR conditions. ARMS can also detect point mutations, restriction fragment length polymorphisms, deletion or insertion of a sequence. Allele-specific priming of the PCR reaction is the principle underlying ARMS, taking advantage of the use of 4 primers (2 outer primers and 2 inner primers) to specifically amplify the wild type and mutant sequences including a positive control band in a single reaction mixture (Jones et al., 2005).

This method has been applied in the detection of V617F mutation in MPNs. Several groups have generated different sets of ARMS primers for the detection of V617F (G>T conversion) mutation (Baxter et al., 2005; Jones et al., 2005; McClure et al., 2006) based on the design of primers that match only a specific DNA point mutation (T) but do not match the wild type allele (G). Another set of primer is used for the matching of only wild type allele (G) so that polymorphic alleles can be distinguished by the specific primer sets (Steensma, 2006). The particular target nucleotide in the DNA sequence can be detected easily by electrophoresis with its compatible staining procedure. In the case of V617F mutation (G→T) detection, the mutation-specific primer is designed based on the point mutation (T) with its 3' terminal nu-

cleotide (A) complementary to the point mutation (T). The 3' nucleotide of the mutation specific-primer will form a mismatch (A-G) with the wild type DNA sequence, prevent the extension of the oligonucleotide primer and thus halt the amplification. However, not all 3' terminal mismatches completely prevent the extension of the oligonucleotide primer, therefore a deliberate mismatch is introduced in the 3rd nucleotide from the 3' end to increase the allelic specificity with better allele discrimination (Ye et al., 2001). A second primer set or common primer set is required to generate the allele-specific product and also 2 control primers (outer primers). The outer primers are required to ensure the reaction is working correctly and is not affected PCR failure but absence of the point mutation that the probe is targeting at (Steensma, 2006).

1.6.3 Single Nucleotide Polymorphism (SNP)

A single nucleotide polymorphism (SNP) is a DNA sequence variations most frequently occurring in the human genome due to a change in a single base. SNPs are widely used as molecular markers to pinpoint a disease on the human genome because of their high frequency and binary variation pattern (Kim & Misra, 2007; Tang et al., 2008). It is found in more than 1% of the population and over 80% sequence variation in the 0.1% difference in 2 unrelated individuals. On average, SNPs can be found every 300bp throughout the human genome (Ke et al., 2008; Salisbury et al., 2003; Nelson et al., 2004). SNPs do not necessary cause disease even though they may be associated with certain diseases. SNPs can be used to potentially predict an individual's risk for certain diseases due to their genetic location near to a disease causal gene. SNPs residing on coding region or regulatory region of a gene may also

directly cause a disease by affecting certain genes from functioning properly (Bethesda, 2014).

SNPs that are physically close to each other on the same chromosome are often linked. They have great probability of being transmitted together to the next generation as a block. The pattern of alleles on a block forms a haplotype. The SNPs in a block are said to be in linkage disequilibrium. It is the co-inheritance of particular combinations of two or more alleles on the same chromosome at frequencies different than would be expected by chance in a population (Reich et al., 2001; 2003). The neighbouring loci within a haplotype block are unlikely to be separated by recombination due to their small genomic distance, unless in a disease setting where new mutations arise or recombination events happen in subsequent generations (Gabriel et al., 2002). Nevertheless, disease markers located close to the disease causal variant are less likely to be separated by a recombination event. Therefore, in a well-designed genetic association study, haplotypes can be used for disease marker mapping and for tracing human population history (Strachan & Read, 1999; Strachan & Read, 1999 ; Sabatine et al., 2006; Alberts et al., 2002). LD analyses are often performed with the Haploview software (Barrett et al., 2005).

Previous studies suggested that inherited SNPs within *JAK2* or some inherited factors predispose to MPNs and are in association with other MPN subtypes (Jones et al., 2009). Methods such as restriction fragment length polymorphisms, denaturing high performance chromatography (Yip et al., 2003), allele-specific PCR (Wu et al., 2005) (depending on the logistic arrangement for instrument use and the cost) (Zha et al., 2009) were widely used to determine the associated factors and to find out the actual

causative gene. However, it is still debatable whether it is the only causative linker to the pathogenesis or as observed, a predisposing mutation, or a genetic modifier in these disorders pathogenesis (Ihle & Gilliland, 2007b; Alabdulaali, 2009).

1.6.4 SNP genotyping

Molecular genetic-based methods are now an integral component of the diagnostic workup for haematological malignancies. Molecular genetics has moved from research-based applications and the simple amplification and detection of an abnormal gene to highly sensitive and reproducible monitoring of minimal residual disease (Erber, 2010; Ward & Kinniburgh, 2000).

Genotyping allows researchers to determine the genetic make-up of an individual, namely genotype (Kim & Misra, 2007). Case-control genetic studies utilising SNPs aim to identify SNPs that affect diseases by comparing the genotypes between patient group and control group. This is done by genotyping. Then, the statistically significant difference in genotypes between the 2 groups can then be used to predict the disease risk associated with a particular genotype. Ultimately, this information is then used to find the related functional proteins for earlier prevention and better treatment (Kim & Misra, 2007).

Genotyping is generally carried out in two steps. The first step is to generate DNA products with the specific allele for the SNPs under study. Allele is a nucleotide (which has alternative forms) residing at a specific location of a gene. The second step then detects the products to determine the genotype. Generally, polymerase

chain reaction (PCR) is the amplification step to generate thousands of copies of the same sequence encompassing the SNP-containing region (Kim & Misra, 2007). It utilises repeated cycles and heating and cooling as well as a thermostable DNA polymerase to replicate DNA. The heat physically separates the double strands of DNA and replication occurs at the cooler temperatures. The cycles of heating and cooling are repeated over and over, allowing primers to bind to the target DNA and to newly synthesized sequences.

1.6.4.1 Restriction fragment length polymorphism (RFLP)

Restriction fragment length polymorphism (RFLP) is the standard genotyping method used in this study. RFLP is a popular technique for genotyping (Rasmussen, 2012). It was first used as a tool for genetic analysis in 1974 (Botstein et al., 1980). RFLP refers to a difference in homologous DNA sequence based on the different lengths of DNA fragments after enzymatic digestion with specific restriction endonuclease. These enzymes are able to recognize specific sequences and structures of DNA and thus can cleave (digest) at a specific region of the sequence. RFLP genotyping method enables distinction between two or more DNA samples taking advantage of the different restriction endonuclease digestion sites.

PCR-RFLP genotyping involves two steps. The first step is the amplification of DNA product of study subjects by PCR. The PCR-amplified product, the ‘target’ is then cut into fragments of defined lengths according to their specific endonuclease cutting sites by compatible restriction enzymes. The resulting fragments can then be separated and recognised by gel electrophoresis based on their defined lengths to reveal the genotypic differences among individuals. DNA restriction enzymes recognise specif-

ic DNA sequences can catalyse endonucleotic cleavages or digestion. This creates fragments of different lengths. A RFLP is said to be existed when the length of enzymatic cut fragments varies among individuals resulting from different genetic makeup among individuals.(Botstein et al., 1980; Clark & Thein, 2004).

This inexpensive and simple method appears to be superior in genotyping studies (Simsek et al., 2001) and have the ability to assay for individuals, thus facilitates large population studies (Botstein et al., 1980). The design of PCR-RFLP (primers and choice of enzymes) is not difficult and can be performed using public available programs such as SNP Cutter (http://bioapp.psych.uic.edu/SNP_cutter.htm) or commercial software that is user-friendly such as OLIGO (Molecular Biology Insights, Inc., USA). The use of PCR-RFLP has also been demonstrated in haematological malignancies. For instance, the InVivoScribe Technologies (San Diego, CA) employed restriction enzyme digestion to produce its V617F Activating Mutation Assay kit marketed "for research use only". Other than commercial kit, Baxter et al.(2005) and Antonioli et al. (2008) had successfully used *Bsa*XI restriction enzyme for RFLP in their research studies of V617F.

On the other hand, RFLP has a few disadvantages. Restriction enzymes require specific conditions to digest the specific recognition site. Besides that, there might be several restriction enzymes that can target on the same recognition site (Rasmussen, 2012) and can affect the sensitivity of the assay. Hence, careful design and optimisation of the experimental conditions are critical for a success. To increase the specificity of the primers, websites such as PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) can be a useful tool to detect un-

wanted mispriming events so that to improve primer design. In some instances, a single or two base mismatches are introduced in the amplification primer (short strand of nucleic acid which is complementary to the template strand that serves as a starting point for DNA replication) to introduce recognition site if the recognition sites are altered by nearby mutations. An internal digestion control site is also recommended to ensure the enzymatic digestion is optimal and complete. The internal digestion control site should be a non-polymorphic site, or otherwise manual alteration of the sequence is a way to create an artificial recognition site (Rasmussen, 2012; Clark & Thein, 2004). Primer design is within user's control; however optimisation of the experimental conditions is time-consuming. In fact, the complete enzymatic digestion itself can take 16 hours, after the PCR amplification of the products. On top of that, some restriction enzymes are expensive.

Despite that PCR-RFLP is an easy-to-design analysis; it is technically demanding yet it cannot seem to fit the high-throughput platform. Other genotyping methods that can be amended for automation and high-throughput analysis, for instance, the next-generation sequencing platforms, offer more for the field (Rasmussen, 2012).

1.6.4.2 High-resolution melting curve analysis using unlabelled probe

A less demanding high-resolution melting analysis using unlabelled probe is a good alternative to RFLP genotyping analysis. Compared with the expensive fluorescent labelled oligonucleotide probes, unlabelled probes are much cheaper and equally sensitive with the labelled probes (Erali et al., 2008; Montgomery et al., 2007). This is a rapid yet sensitive detection method of known sequence variations including SNPs, small insertions and deletions.

Key to the method is the use of asymmetric PCR (usually 1:5 to 1:10 primer ratio) to generate single-stranded DNA (ssDNA) of the target strand excessively so that the probe can hybridize to the target strands (Erali et al., 2008). Genotyping is accomplished by monitoring the melting of probe-target duplexes post-PCR. When the PCR is complete, the probe and a saturating dsDNA dye are added to the ssDNA target for high-resolution melting analysis. In addition to a pair of PCR primers, oligonucleotides, a relatively short strand (13-25 nucleotides) of single-stranded DNA molecules (Dias & Stein, 2002), are required. For optimal reaction, oligonucleotides between 20 and 40bp with *T*m from 50 to 85°C are designed to serve as hybridisation probes in the presence of double-stranded DNA (dsDNA) specific fluorescent dyes such as SYBR Green I, LCGreen, SYTO9 or EvaGreen. Specific design of the oligonucleotides enables a significant shift in the melting temperature (*T*m) of the probe-amplicon duplex upon a single base variation within the probe. The visible shift of *T*ms is able to differentiate homozygous and heterozygous variants (Montgomery et al., 2007). Longer probes yield better signal intensity than shorter probes. In my work, *T*ms from 60 to 75°C have been successfully used as long as the probe *T*m is lower than that of a PCR extension to avoid overlapping of the melting profiles. The signals for probe are stronger with shorter PCR amplicons ranging from 100bp to 200bp(Erali et al., 2008). To prevent the probe from extending during asymmetric PCR, a phosphate group or a poly (A)/(T) tail is added to the 3' end of the unlabelled probe.

After the PCR amplification, melting curve analysis assesses the dissociation characteristics of DNA as the strands convert from double-stranded DNA (dsDNA) into single-stranded DNA (ssDNA) with increasing temperature in the presence of dsDNA fluorescent dye (Smith et al., 2009; Ririe et al., 1997). When the temperature in-

creases, the dsDNA denatures (melts) and releases fluorescent dye with a resultant decrease in the fluorescent signal. The temperature at which dsDNA melts is determined by factors such as nucleotide sequence, length and GC/AT ratio (Smith et al., 2009).

The reaction can be performed with LightCycler® 480 Real-time PCR System (Roche). Probe/ssDNA amplicon duplexes were generated by heating and cooling of the samples. Melting curves can then be analysed with LightCycler® 480 Software compatible with the platform used. The method is diagrammed in **Figure 1.10**.

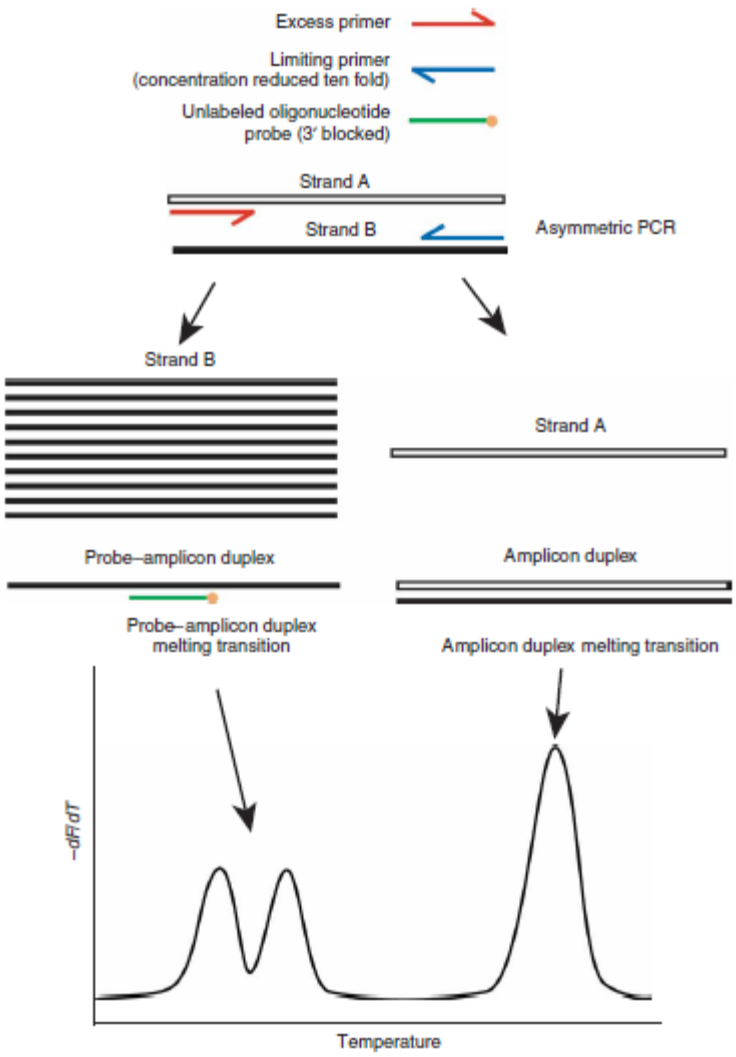


Figure 1. 10 High-resolution melting curve analysis using unlabelled probe.

In an asymmetric PCR, three oligonucleotides are needed: excess primer (red), limiting primer (blue), and unlabelled probe (green). Excess primer that is complimentary to strand A is used to produce strand B in excess. Limiting primer that is complimentary to strand B is used to produce limited copies of strand A. Excessive strand B increases the signal of the probe-amplicon duplex melting transition together with the amplicon melting transition on the same melting profile. This single profile enables variant scanning and genotyping concurrently Peaks displayed are the negative derivative of both melting transitions after background subtraction and normalisation of fluorescence with respect to temperature (-dF/dT) against temperature (Montgomery et al., 2007).

1.6.4.3 DNA Sequencing by Capillary Electrophoresis

DNA sequencing techniques are widely used in many fields including genetics, biotechnology, molecular technology, forensic sciences, archaeology, and others (Franca et al., 2002). Sequencing technology enables the capture of the most detailed information about any genomic locus. In 1975, (Sanger & Coulson) introduced dideoxynucleotide chain termination coupled with size separation by gel electrophoresis. It is a catalysed enzymatic reaction that polymerizes the DNA fragments complementary to the unknown template DNA of interest.

To start synthesising material for sequencing, abundant copies of the targeted sequence are purified and amplified by PCR. Then to start sequencing DNA, a primer is designed to anneal to a specific known region on the template DNA that is complementary to this primer. In the presence of DNA polymerases I and free nucleotides, deoxynucleotide triphosphates (dNTP) polymerize onto the DNA to yield a growing chain. This catalytic polymerisation continues until a terminator or dideoxynucleotide triphosphate (ddNTP) is added into the growing chain. The ddNTPs are modified fluorescent nucleotide which do not poses the hydroxyl group that is required to terminate the chain elongation. There are four different terminators (A, C, G and T) each with a different fluophores. The reaction occurs in from the 5’ end towards 3’, so all the growing chains/fragments have the same 5’-end but a different ddNTP-3’-end. The random and irreversible termination creates abundant fragments of different lengths. Once the reaction is complete, the mixture of fragments will be sorted and visualised by capillary electrophoresis, which is unlike the original way, it was sorted by molecular weight using gel electrophoresis. The unknown sequence can then be read by laser (Franca et al., 2002; Karger & Guttman, 2009). The sche-

matic representation of DNA Sequencing by Capillary Electrophoresis is depicted in **Figure 1.11**.

Over the years, the field of genomics is evolving rapidly from labour intensive slab gel electrophoresis to automated multicapillary electrophoresis systems. To better fit the high-throughput demands, sequencing platforms have evolved from being able to determine only a few hundred nucleotides to a few sequences at once, to 96-format capillary sequencer and more rarely the 384-format capillary sequencer or, to the 'next-generation' sequencing platforms or nanopore and single molecule sequencing; the field has more to offer now.

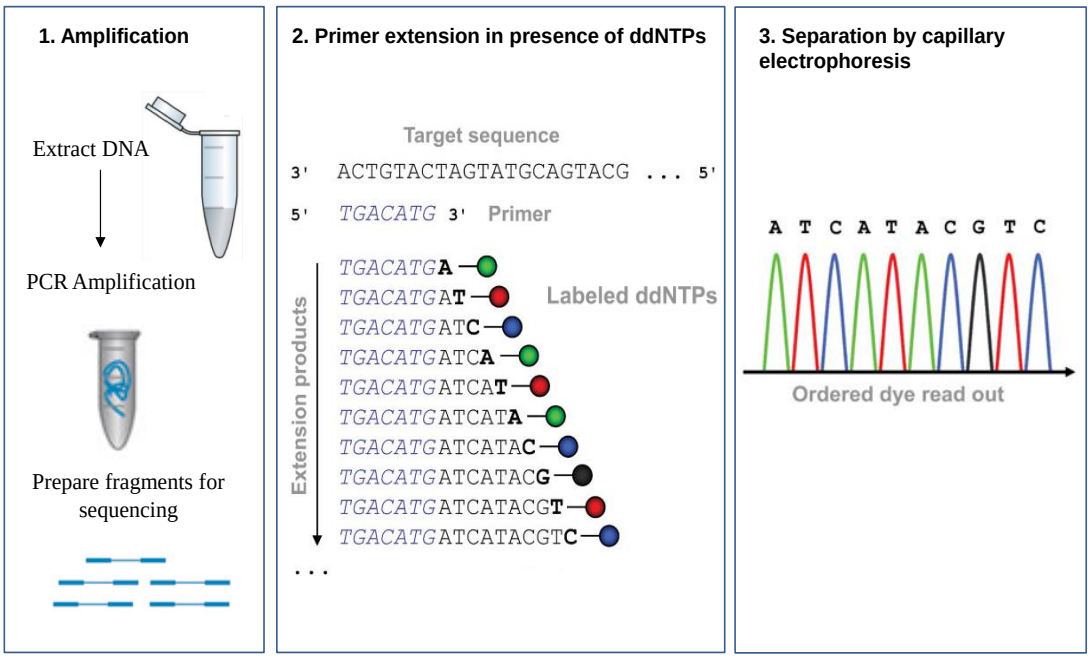


Figure 1. 11 Schematic representation of DNA Sequencing by Capillary Electrophoresis.

Input DNA is PCR amplified into desired fragments for sequencing. Sequencing reaction is performed in the presence of DNA polymerases I, free nucleotides, and a primer to yield a growing chain. The addition of dideoxynucleotide triphosphate (ddNTP) into the growing chain causes the extension to be irreversibly terminated. This creates abundant fragments of different lengths. Once the reaction is complete, the mixture of fragments will be sorted and visualised by capillary electrophoresis. Modified from Kircher and Kelso (2010).

1.6.5 Targeted Next Generation Sequencing (NGS)

The technology is rapidly evolving that the multicapillary format has to be replaced to meet higher demands. The sequencing limit of single reaction of Sanger DNA sequencing coordinating with multicapillary electrophoresis is only 1000 nucleotides. Therefore, microchips have replaced capillary electrophoretic separation to offer higher-throughput sequencing in time-efficient and cost-effective way and possibly to offer multi-integrated pipelines (Hert et al., 2008). Massively parallel sequencing platforms have been widely used. Cyclic array sequencing, sequencing by hybridisation or ligation, bridge PCR-based sequencing, nanopore sequencing and the single molecular approach are all the accelerators behind biological and biomedical research nowadays. Illumina, 454 Life Sciences, Applied Biosystems, Oxford Nanopore Technologies, Pacific Biosciences are among the major companies supporting these NGS platforms. They enable more comprehensive analysis of the omics studies such as genomics and transcriptomics to be more affordable and possibly a routine (Karger & Guttman, 2009).

Despite the tremendous advances in NGS, the cost to sequence the whole genome is still significant (Metzker, 2010). Another concern arises from whole genome sequencing might be the massive data generated that can complicate data processing and analysis. In studies that are target-oriented, with limited budget and time, targeted sequencing seems to be a better alternative. Traditional genotyping methods such as RFLP and melting curve analysis seem to fit in these goals. However, they are useful in investigating known variants but not in discovering unknown genetic variants. Fine mapping or targeted NGS is able to zoom into the susceptibility region after an initial positive signal at higher quality and lower cost than whole genome se-

quencing. Based on the positive signal from genetic association study and meta-analysis, the whole *JAK2* gene together with some other major MPN-associated genes were targeted for a refined investigation. Such targeted deep sequencing is relatively less expensive and can serve as a discovery tool for novel variants underlying the disease. Oligonucleotide-selective sequencing, multiplex PCR amplification, selective target circulation, and hybrid-capture are the current commonly used methods (Shen et al., 2013).

The sequencing process can be generally grouped into 3 major parts: template preparation (library preparation), sequencing and imaging, and data analysis (Metzker, 2010). In this study, targeted sequencing approach was applied. The SeqCap EZ capture method from Roche NimbleGen, Inc was used for targeted enrichment. This in-solution target enrichment is a custom-capture technology that is compatible with Illumina MiSeq sequencing platform and has its standard protocol prepared for Illumina library preparation. The workflow is diagrammed in **Figure 1.12.** All information and illustrations are adapted from the illumina (http://www.illumina.com/) and Roche NimbleGen website (http://www.nimblegen.com/products/seqcap/ez/choice/).

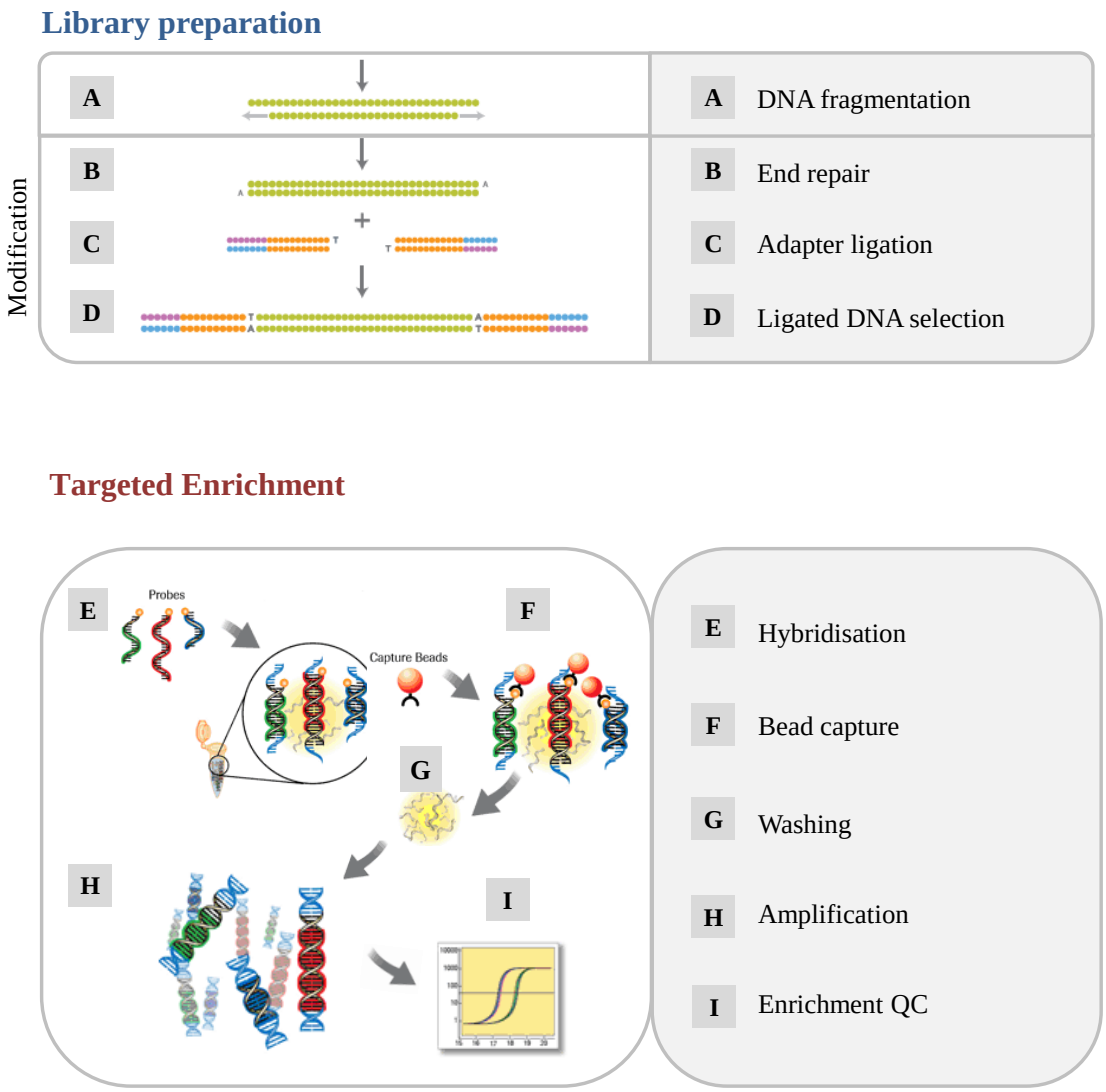
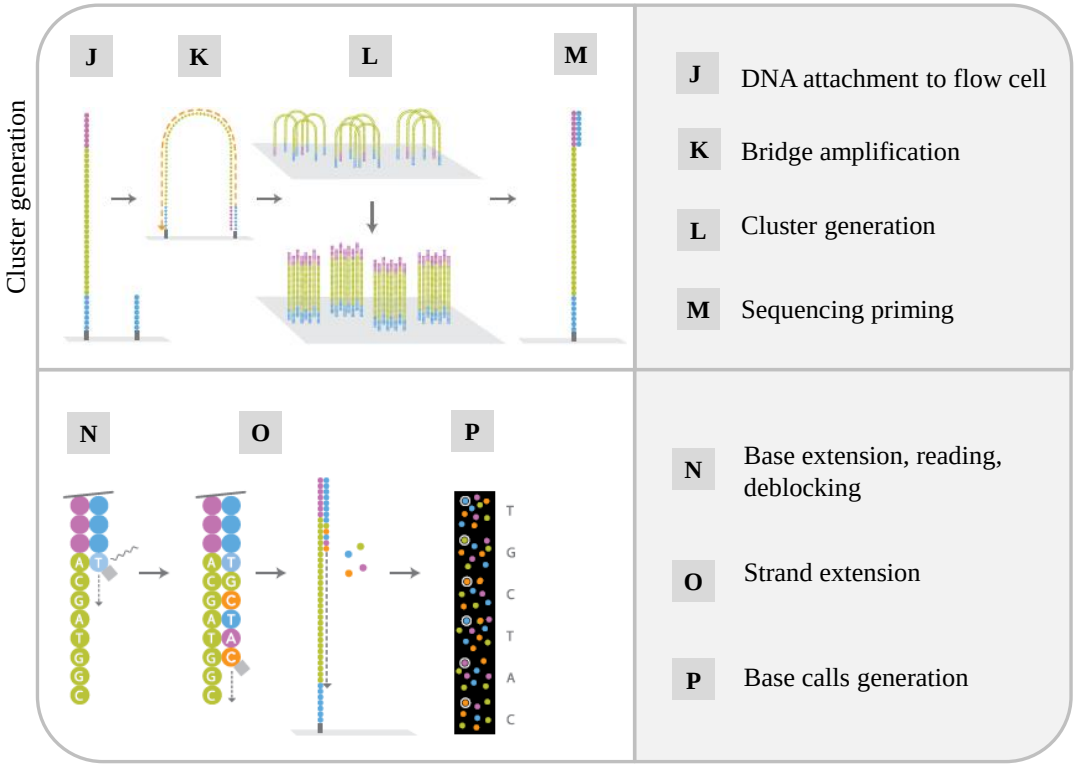


Figure 1.12 Schematic representation of the next generation sequencing workflow.

Sequencing



Data analysis

Figure 1.13 Schematic representation of the next generation sequencing workflow. (continued)

1.6.5.1 Library preparation

The general processing steps for NGS technologies are similar which includes fragmentation, modification, and quantification (**Figure 1.12**). The specific technical details especially the modification part (the most diverse part) are the distinguishing features among the protocols. Library preparation aims to prepare and amplify a set of nucleic acids (DNA in this study) for high throughput sequencing environment.

Library preparation starts with fragmentation of the long-range genomic DNA. Fragmentation of genomic DNA is a crucial step in library preparation and can be performed by physical, enzymatic, and chemical approach (Knierim et al., 2011; Head et al., 2014). After fragmentation, samples are modified by ligating sequence specific adapters and indexes to the fragmented DNA. For this study, the NEBNext® dsDNA Fragmentase (New England Biolabs) was used to generate DNA fragments around 300bp length, with the recommended incubation time. NEBNext® dsDNA Fragmentase contains two enzymes to work in pair: one randomly generates nicks in the dsDNA while the other recognises the nicks and cuts the opposite strands to produce DNA fragments with short overhangs, 5′-phosphates and 3′-OH-groups that are ready for end repair and adapter ligation. DNA are repaired and converted to DNA having 3´ dA-tails by the NEBNext End Repair dA-tailing Module. Then, sequence specific adapters are ligated to the fragmented DNA to prime the amplification by generating template clusters that can bind to the flow cell of sequencer (Lynch, 2002; Lundin, 2012).

Once constructed, multiplex PCR is performed using different barcoded PCR primers (index) to amplify different ligated products to produce sufficient starting material

for subsequent processes. The addition of index or barcode to the library template facilitates multiplexing of up to 96 samples per sequencing run and helps to identify the samples easily. Many indexed libraries with different number of regions can be combined in single run. The PCR products are then selected by the desired insert size (excluding the adapter sequences with constant length) depending on the NGS instrumentation and the sequencing application used. For this exome sequencing, I employed 2 × 150 paired-end reads. The insert sized approximately 350 bases in length so to avoid overlapping read during sequencing. In addition, this length enables the capture of most exons because more than 80% of human exomes size below 200 bases in length. The essence in controlling the desired length of the library insert is the time used for fragmentation reaction (Sakharkar et al., 2004).

The pooled library can be targeted enriched for the region of interest. After the size selection, purification, quality check, and quantification, the pooled libraries are ready for sequencing on a MiSeq system (Grada & Weinbrecht, 2013; Lynch, 2002).

1.6.5.2 Benchtop sequencing platform

The Illumina MiSeq (San Diego, CA.) was used in this thesis. It is a revolution of benchtop sequencing platform. It uses a reversible sequencing by synthesis (SBS) approach for a high throughput of sequencing (**Figure 1.12**). The SBS approach uses bridge amplification technology to form template clusters for sequencing. Once binding to the flow cell, the library fragments undergo bridge amplification and form new template clusters on a flow cell for sequencing (Grada & Weinbrecht, 2013; Lynch, 2002). Upon cluster formation, fluorescently labelled dideoxynucleotide triphosphate (ddNTP) is incorporated into the growing strand of DNA. The fluorescently labelled ddNTP terminates the polymerisation and the fluorescence emitted from parallel reactions is captured and imaged for base calling. Simultaneously, enzymatic cleavage allows incorporation of the next nucleotide. These processes repeat with the incorporation of known ddNTP in a sequential order. As the strand grows upon ddNTP incorporation, the digital images are recorded and 'read' as sequence (Grada & Weinbrecht, 2013; Lynch, 2002). A summary of key features of Illumina MiSeq is shown in **Table 1.5.**

MiSeq's integration of cluster generation and SBS technology enables the generation of highly accurate data up to 8 GB within days. It enables a very fast turnaround time of the process with exceptional data quality through a short and simple instrument workflow (Liu et al., 2012).

Table 1.5 Characteristics of MiSeq sequencing platform.

Product name	MiSeq
Description	Focused power, speed and simplicity for targeted and small-genome sequencing.
Key applications	Small genome, amplicon, and targeted gene panel sequencing.
Flow cells processed per run	1
Output range	0.3-15 Gb
Run time	5-55 hours
Reads per flow cell*	25 Million†
Maximum read length	2 × 300 bp

* Clusters passing filter.

† For MiSeq V3 Kits only.

Adapted from http://systems.illumina.com/systems/miseq.ilmn

1.6.5.3 Data analysis

Once sequencing is complete, raw sequence data are processed following typical analysis stages. The general workflow for NGS data analysis starts with pre-processing the raw data to remove adapters and low-quality reads. When the raw data are ‘clean’, the reads can be de novo assembled or mapped to a reference genome, and the compiled sequences can be analysed. Depending on one’s interest, different NGS data analyses can be performed using various bioinformatics pipelines. For instance, genetic variant calling is used to detect SNPs or indels (i.e., insertion and deletion of bases). With special interest, this study aims to find novel or existing somatic and germline mutations that may contribute to the development of MPNs among V617F-positive patients. Then, functional annotation and advanced analysis can be conducted to investigate the relationship between the variants and disease. Nevertheless, it is likely that the future will employ a combination of cytogenetics, expression, and oligonucleotide arrays for a full diagnostic picture and hence, our understanding of the pathogenesis of haematological malignancies will be enhanced (Kearney & Horsley, 2005).

1.7 Identification of research gap

The *JAK2* gene is the common susceptibility gene to MPN not only in germline but also in somatic cells of V617F-positive MPN patients (Kilpivaara et al., 2009; Olcaydu et al., 2009a; Jones et al., 2009). It was reported that disease susceptibility loci may harbour somatic mutations contributing to the disease pathogenesis. MPN genetic studies were all initially done in cohorts of Caucasians (Olcaydu et al., 2009a) until August 2009, one study on Chinese MPN patients was published. Nevertheless, Zhang et al. (2010) observed no significant genotype-phenotype association among the 3 MPNs subtypes in Chinese patients however concluded that V617F is still a major molecular factor in Chinese MPN patients. In 2011, it was finally reported association between *JAK2* haplotype and an increased risk of acquiring the V617F somatic mutation in Chinese Han (Hu et al., 2011). Until March 2012, Tian et al. (2012) investigated the relationship between V617F and *JAK2* haplotype with their establishment of allele-specific multiplex PCR method. The haplotype was detected in V617F-positive and V617F-negative MPN with frequencies of 69.33% and 43.09% patients respectively. However, genotyping was not performed. Following the first genetic association study in Chinese population, second paper was published in October 2012, positing that *JAK2* haplotype is a risk factor in Chinese patients with MPNs (Zhang et al., 2012a). Later on, Wang et al. (2013) reaffirmed this association by concluding that *JAK2* haplotype predisposed to acquisition of V617F mutation in Chinese population. Apart from Chinese population, studies in Taiwanese and Japanese also revealed specific alleles/additional locus predisposing to ET/MPNs respectively. Nonetheless, these studies only focused on the V617F-positive group (Ohyashiki et al., 2012; Hsiao et al., 2011).

Other polymorphic loci might have to be included in the future studies on Chinese MPN patients in addition to those reported haplotypes. This is understood because different populations have different haplotype histories and thus different association with the disease. This group of hospital-based researchers showed more interest in the targeted therapy rather than the mechanism behind this mysterious disease. Therefore, there is a desire to study the genotype-phenotype association in MPN Chinese patients.

The association signals arose from genetic association studies are inconsistent. The germline *JAK2* haplotype (*JAK2* 46/1 or GGCC haplotype) was implicated to be associated with V617F-positive and/or in V617F negative MPNs, being the results reported contradictory. Moreover, these studies were conducted mainly in Caucasians, only a few in Asians with small sample size. Therefore, meta-analysis was conducted to affirm the genetic association in MPNs, stratifying for V617F mutation status and also ethnicity. Nonetheless, the mechanism stays obscure. With my data from the association study and meta-analysis, there might be some other causal variants that are not detected with conventional genotyping platforms such as RLFP and melting curve analysis. Therefore, targeted deep sequencing might serve as a useful tool to discover the causal variants that are in LD with the SNPs identified in this study.

1.8 Aims of this study

The general aim of this study is to investigate the relationship between *JAK2* germline polymorphisms and MPNs in the Hong Kong Chinese population and to find the causal variants associated with MPNs. My research focused on:

1. Detection of V617F in Hong Kong Chinese population with MPNs and also healthy control.

2. Identification of risk alleles that predispose to MPNs in Hong Kong Chinese population.

3. Meta-analysis of *JAK2* SNPs to detect the association in V617F-negative MPNs patients.

4. Targeted deep sequencing to examine any nucleotide variation on the whole *JAK2* gene and also the exons of some MPN-associated genes.

Part I Genetic association study of germline

Polymorphisms in *JAK2* gene

CHAPTER 2 Materials and methods

2.1 Chemicals and reagents

All solutions and buffers used in subsequent experiments were prepared using reverse osmosis water (Millipore, Bedford, MA, USA) and sterilised by autoclaving at 121^{o}C for 15 minutes. UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen Life Technologies, Carlsbad, CA) was used for PCR, RFLP, unlabelled probe melting curve analysis, and sequencing reaction.

2.1.1 DNA extraction

FlexiGene DNA kit from Qiagen (Hilden, Germany) was used to extract nucleic acid from cases and controls. The solutions required to work with the extraction kit were listed as follows:

Absolute ethanol

70% ethanol was prepared from AnalaR-grade absolute ethanol (Riedel-de Haën, Seelze, Germany)

Isopropanol

AnalaR grade (Riedel-de Haën, Seelze, Germany)

2.1.2 Adjustment of DNA concentration

I×TE buffer

- 10 mM Tris-hydrochloric acid (Tris-HCl) from Sigma (St. Louis, USA)
- 1 mM Na_2EDTA from BDH (Poole, UK)

2.1.3 Polymerase chain reaction (PCR)

Unless otherwise stated, all PCR reactions were performed using **HotStar Taq Plus**, **1×PCR buffer** (which contains KCl and $(NH_4)_2SO_4$), and **magnesium chloride**, ($MgCl_2$) were purchased from Qiagen (Hilden, Germany). **Deoxyribonucleoside triphosphates** (dNTPs): dATP, dGTP, dCTP and dTTP each at 100 mM were purchased from GE Healthcare (Piscataway, USA). For stocking solution, 100 mM of all dNTPs were equally diluted to 20mM and for working solution; 20mM was further diluted to 2mM. **Primers** were ordered from Integrated DNA Technologies, Inc. (IDT, Commercial Park, Coralville, IA, Iowa United States).

2.1.3 Electrophoresis reagents

2.1.3.1 Agarose gel

Agarose gel was prepared using either Seakem LE agarose (Cambrex Bio Science, Rockland, ME) or Biowest regular agarose G-10 (Biowest, Spain). **Ethidium bromide** from Sigma (St. Louis, USA) was used as stain for visualisation. **1 kb Plus DNA Ladder** from Invitrogen Life Technologies (Carlsbad, CA) was used as a reference for gel electrophoresis of PCR products. Loading gel and TBE buffer needed for electrophoresis were prepared as below:

6× loading dye

- 30% Glycerol from AFAX Chemicals (Auburn, Australia)
- I× TE buffer
- 0.05% Bromophenol blue from Sigma (St. Louis, USA)

10× Tris Borate Ethylenediaminetetraacetic acid (TBE) buffer

- 890 mM Tris base from Sigma (St. Louis, USA)
- 890 mM Boric Acid (Riedil-deHaën, Seelze, Germany)
- 25 mM disodium ethylenediaminetetraaceticacid, Na_2EDTA, pH 8.3 (BDH, Poole, UK)

It was then diluted to 0.5× solution for agarose gel electrophoresis.

2.1.3.2 Polyacrylamide gel (PAGE)

Polyacrylamide gel (PAGE) was prepared using 40% acrylamide-bis solution, 25% Ammonium persulphate (APS), 10× TBE (Tris Borate Ethylenediaminetetraaceticac-id) buffer, and Tetramethvlethvlenediamine (TEMED).

40% acrylamide-bis (19:1) solution

One litre of the solution was prepared by mixing the reagents below and made up to 1 litre with purifier water and filtered:

- 76g acrylamide from Acros Organic (Thermo Fisher Scientific, Geel, Belgium)
- 4g bis from Acros Organic (Thermo Fisher Scientific, Geel, Belgium)

10× Tris Borate Ethylenediaminetetraacetic acid (TBE) buffer

- 890 mM Tris base from Sigma (St. Louis, USA)
- 890 mM Boric Acid (Riedil-deHaën, Seelze, Germany)
- 25 mM disodium ethylenediaminetetraaceticacid, Na_2EDTA, pH 8.3 (BDH, Poole, UK)

Ammonium persulphate (APS)

APS was purchased from Sigma (St. Louis, USA)

Tetramethylethylenediamine (TEMED)

TEMED was purchased from Bio-Rad (Hercules, CA, USA)

Coating solution for glass plate

Glass plates were pre-coated with a salinizing solution from Sigma (St. Louis, USA).that contained γ-methacryloxypropyltrimethoxysilane (silane), 10% acetic acid, and absolute ethanol.

SYBR® Green I Nucleic Acid Gel Stain

It was obtained from Invitrogen Life Technologies (Carlsbad, CA) to visualize gel bands.

2.1.4 Restriction fragment length polymorphism (RFLP)

Restriction enzymes were purchased from either MBI Fermentas (Vilnius, Lithuania) or New England Biolabs (Beverly, MA, USA).

2.1.5 High-resolution melting curve analysis

All the 3’ blocked probes were ordered from Integrated DNA Technologies, Inc. (IDT). SYTO 9 green fluorescent nucleic acid saturated dye was purchased from Invitrogen Life Technologies (Carlsbad, CA).

2.1.6 DNA sequencing

2.1.6.1 Purification of PCR products

Exonuclease I (ExoI)

New England Biolabs (Beverly, MA, USA)

Shrimp alkaline phosphatase (SAP)

SAP is supplied with 10× SAP dilution Buffer. from GE Healthcare (Piscataway, USA).

2.1.6.2 Cycle sequencing reactions

BigDye® Terminator Cycle Sequencing Kit (version 1.1)

Applied Biosystems (Foster City, CA, USA)

2.1.6.3 Precipitation of PCR products

Sodium acetate 3M

Sigma-Aldrich (St. Louis, USA)

Absolute ethanol

Preparation of 95% and 70% ethanol was done with AnalaR-grade absolute ethanol from Sigma-Aldrich (St. Louis, USA).

Hi-Di™ Formamide

DNA template was re-suspended in Hi-Di™ Formamide from Applied Biosystems, (Foster City, CA, USA).

2.2 Laboratory equipment

Measurement of DNA concentration

NanoDrop ND1000 UV/Vis spectrophotometer (Thermo Scientific, Wilmington, USA)

PCR

96-well GeneAmp 9700 PCR system from Applied Biosystems, (Foster City, CA, USA)

Gel Visualisation

Chemi Genius^2 BIO imaging system (SYNGENE*;* Frederick, MD, USA)

Melting curve analysis

LightCycler® 480 (LC480) from Roche (Basel, Switzerland)

Direct Sequencing

ABI PRISM 3130 Genetic Analyser from Applied Biosystems, (Foster City, CA, USA)

Centrifugal vacuum concentrator

Savant DNA 110 SpeedVac®concentrator centrifuge

2.3 Methods

Study workflow is summarised in **Figure 2.1**.

2.3.1 Sample size and power calculation

The sample size required was calculated using Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html) (Purcell et al., 2003). It was estimated that a sample size of 128 cases and 460 controls would have 80% power, based on the data from previous studies: the average MPN prevalence of 0.00002, minor allele frequency (MAF) of 0.1, genotypic relative risk of 2.5 for Aa (heterozygous) and 5.0 for AA (homozygous) (**Table 2.1**).

Table 2.1 Parameters used for sample size estimation

Parameters	Values
High risk allele frequency (A)	0.1
Prevalence of MPNs	2.2/100,000 =0.000022
Genotype relative risk (Aa)	2.5
Genotype relative risk (AA)	5.0
D-prime	0.80
Number of tag SNPs for this study	18
Number of cases	150
Control : case ratio	2:1
Type 1 error	0.05/18= 0.002777
User-defined Power	90%
Sample size estimated for experimental: control	**150:300**

2.3.2 Subject recruitment

MPN patients

This study was approved by the Human Subjects Ethics Sub-Committee of the University (reference numbers: 20090801001 and 20111118001) and Research Ethics Committees of the hospitals, according to the guideline of the Declaration of Helsinki. The Research Ethics Committees of the hospitals under Hospital Authority included the following: Kowloon West Cluster Clinical Research Ethics Committee (reference number: KW/EX/09-076); Research Ethics Committee, Kowloon Central / Kowloon East Clusters (KC/KE-09-0120/FR-3); Joint The Chinese University of Hong Kong–New Territories East Cluster Clinical Research Ethics Committee (reference number: CRE-2009.423); and Ethics Committee, Hong Kong Easter Cluster (HKEC-2009-069). One hundred ninety-two unrelated Hong Kong Chinese patients of MPNs were recruited as the cases between August 2010 and April 2011. The following patients were selected: follow-up MPNs patients with existing records kept track by collaborating hospitals (cases) and taken care of by co-investigators in hospitals. Patients were excluded if they did no fulfil the inclusion criteria based on WHO criteria (2008) (Vardiman et al., 2009). Clinical data of these patients were also gathered retrospectively. After the compliance with patients and a written informed consent obtained, blood and saliva were collected.

Control

A corresponding group of anonymous Chinese blood donors were collected from the Hong Kong Red Cross Blood Transfusion Service as control subjects. They were matched for sex- and age (±1) as much as possible to exclude any potential confounding factor that might restrict the study of this multifactorial disease. Only blood samples were collected and processed appropriately for controls because of the low prevalence of MPNs and UPD in the population (Goldin et al., 2009).

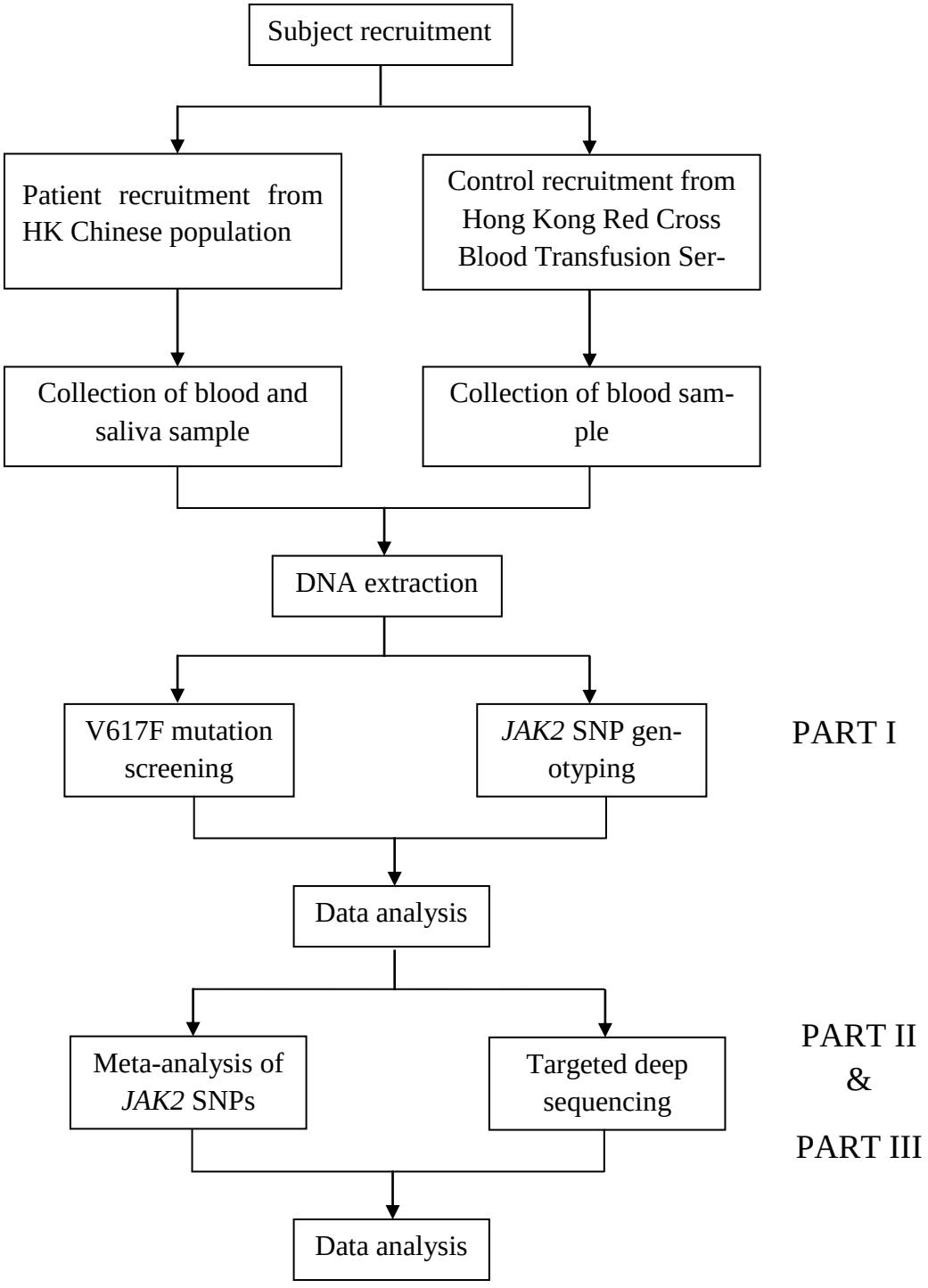


Figure 2.1 The workflow of genetic association study in this project.

2.3.3 DNA extraction from peripheral blood sample

Patients' peripheral blood samples (9mL) were drawn into ethylene diamine tetraacetic acid (EDTA)-anticoagulated tubes provided by hospitals. Complete blood count (CBC) was performed using the automated haematology analyser CELL-DYN 3200 multi-parameter haematology analyser (Abbott Diagnostics) for the sake of consistency despite the CBC done in hospitals. The blood tubes were centrifuged at 3500rpm for 15 minutes at 4^0C. Plasma was aspirated from the aqueous layer and stored at -80^0C while DNA was extracted from buffy coat using FlexiGene DNA extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and quantified by NanoDrop ND1000 UV/Vis spectrophotometer (Thermo Scientific, Wilmington, USA).

2.3.4 DNA extraction from saliva

Saliva samples from cases were collected using the Oragene DNA self-collection kit (DNA Genotek) according to the manufacturer's instructions and used for SNP genotyping. Approximately 2mL saliva was collected from each patient. Salivary DNA was then extracted following manufacturer's instructions and quantified by NanoDrop ND1000 UV/Vis spectrophotometer (Thermo Scientific, Wilmington, USA).

2.3.5 Selection of SNPs

JAK2 germline variants and *JAK2* risk haplotypes-tagging SNPs (tSNPs) that are associated with the development of MPNs were the selection pool. *JAK2* tSNPs were selected from genotype data from International HapMap Project

(http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap24_B36/) using the Tagger software (de Bakker et al., 2005). Tag SNPs were selected from a 148.7-kb region encompassing the *JAK2* locus and its potential regulatory regions (3 kb upstream and downstream of *JAK2*) with MAF≥ 0. 1 and pairwise tagging algorithm, $r^2 \geq 0.8$, based on HapMap CHB database (release #24/phase II) (2003). In line with previous studies that have suggested the *JAK2* 46/1 haplotype as a strong predisposing risk factor for MPN development, the risk-haplotype tSNPs (rs10974944, rs12343867, and rs12340895; i.e. S9, S12 and S13) were force-included (Jones et al., 2009; Olcaydu et al., 2009a; Kilpivaara et al., 2009; Trifa et al., 2010). Nineteen SNPs were studied in this project. The SNPs were called S1, S2, …., and S19 in the sequential order from the 5' end to the 3' end of the *JAK2* sense strand for ease of discussion.

2.3.6 Design of primers and probes

Flanking sequence of selected tSNPs were obtained from the 1000 Genomes Project data (Genomes Project et al., 2012) for primer design. Primers for both RFLP and unlabelled probe melting curve analysis were designed using OLIGO version 6.62 software (Molecular Biology Insights, Inc., USA) (**Table 2.2-3**). RFLP is the standard genotyping method used in this study. For its primer design, the polymorphic sequence or the SNP rs number was input into WatCut SNP-RFLP analysis (http://watcut.uwaterloo.ca/template.php?act=snp_new) to search for restriction enzyme cleavage sites. Upon knowing the available cleavage sites, the sequence was then input into NEBcutter V2.0 (http://tools.neb.com/NEBcutter2/index.php) for appropriate restriction enzymes for RFLP. The expected cutting patterns were also recorded from the website. Unlabelled probe melting curve analysis was the alternative to RFLP when genotyping could not be performed with RFLP under the circum-

stances in which there was no restriction enzyme cutting site, or restriction enzyme was too costly.

All primers were designed to have similar melting temperature (*T*m, °C) while the formation of primer-dimer and hairpin structures were avoided. *T*m of the primers was adjusted by manipulating the length and GC content of the primers, with the inspection by an in-house equation:

$$69.3 + (0.41 \times \text{GC \%}) - (650 \div \text{primer length in bp})$$

While for melting curve analysis, *T*m was predicted by the *T*m Utility™ software (version 1.3, Idaho technology, Utah, USA). *T*m was estimated for PCR primers to be a reference during PCR optimisation. While for the probes, *T*m difference between the probes with perfectly matched allele and the mismatched allele must be greater than 4°C for clear genotype calling. Specificity of the primers was checked with PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) to detect unwanted mispriming events against the reference assembly sequence of the human genome.

2.3.7 Polymerase chain reaction (PCR)

Nineteen primer pairs (**Table 2.2**) were designed to amplify the 19 *JAK2* tSNPs. These genetic markers are all intronic SNPs except rs3808850 (S1) which is 5' upstream of the *JAK2* gene. Conventional PCR was used to amplify target regions of SNPs genotyped by RFLP; asymmetric PCR was used to amplify target regions of SNPs genotyped by unlabelled probe melting curve analysis.

2.3.7.1 Conventional PCR

The PCR reaction was performed separately for each primer pair in a reaction volume of 10µL containing the following:

10 ng genomic DNA
2.5 or 3.5 mM $MgCl_2$
1×PCR buffer
0.3 µM of each primer
0.2 mM of each dNTP
0.3 unit of HotStarTaq Plus DNA polymerase

Amplification was performed in 96-well plates with a GeneAmp 9700 PCR system or Veriti® Thermal Cycler both from Applied Biosystems, (Foster City, CA, USA). Typical cycling conditions for PCR with HotStarTaq Plus DNA polymerase were:

Step	**Temp (°C)**	**Time**	
1 Initialisation step	95	5 mins	
2 Denaturation step	95	30 sec	30 or 35 cycles
3 Annealing step	60-63	30 sec	
4 Extension step	72	30 sec	
5 Final extension step	72	5 mins	
6 Hold	15	∞	

Precise PCR conditions for each SNP can be found in **Table 2.2**.

2.3.7.2 Asymmetric PCR

The target regions of SNPs to be genotyped by unlabelled probe melting curve analysis were amplified separately by asymmetric PCR. This method generate single-

stranded DNA (ssDNA) product with the concentration of limiting and excess primers in the ratio of either 1:10 or 1:15. A phosphate group or a poly (A)/(T) tail was added to the 3' end of the unlabelled probe to prevent probe from extending. Asymmetric PCR was performed in a 10-µL reaction mixture containing the following:

Amplification was performed in 96-well plates with a GeneAmp 9700 PCR system or Veriti® Thermal Cycler both from Applied Biosystems, (Foster City, CA, USA). The PCR reaction was performed separately for each primer pair in a reaction volume of 10µL containing the following:

10 ng genomic DNA

2.5 or 3.5 mM $MgCl_2$

0.2 µM excess primer with a limiting: excess primer ratio of 1:10 or 1:15

0.2 mM of each dNTP

1×PCR buffer

0.3 unit of HotStarTaq Plus DNA polymerase

Amplification was performed in 96-well plates with a GeneAmp 9700 PCR system or Veriti® Thermal Cycler both from Applied Biosystems, (Foster City, CA, USA). The cycling conditions for asymmetric PCR were:

Step	Temp (°C)	Time	
1 Initialisation step	95	5 mins	
2 Denaturation step	95	30 sec	55 cycles
3 Annealing step	50-56	20 sec	
4 Extension step	72	20 sec	
5 Final extension step	72	5 mins	
6 Hold	15	∞	

Precise PCR conditions for each SNP can be found in Table 2.3**.**

Table 2.2 PCR conditions for RFLP

SNP [a]	Restriction enzyme	Primer sequences (5'>3') [b]	PCR conditions				
			Mg^{2+} (mM)	Primer (mM)	T_m (°C)	Cycles	Size (bp)
rs3808850 (S1)	XmnI	F: $(T)_{30}$ GAA TTC CTT CTT TCC TGC AAA CAA AAA CTG R: $(T)_{32}$ CAC ATT TCC ATT TCC ACA GTT GTG AAG C	2.5	0.3	62	35	201
rs2149555 (S4)	BseGI	F: ACA TGC TTT CAA AGA GAG GCC AG R: $(T)_{20}$ ATT TTC CAT GCC GTA ATG TAT GCT AAT C	2.5	0.3	62	35	273
rs1536798 (S5)	XceI	F: $(T)_9$ ATG GGA CTG TAT TTG GAC TTG GCT R: $(T)_{27}$ TGC ACA ATT CCA ACA ACA TGT CAG TAT AA	2.5	0.3	60	35	251
rs10815148 (S6)	SspI	F: $(T)_{16}$ GAT ACA TCA TGT TTC TTG CGG AAT ATT ATT CAT G R: TGA CAG TGA AAA CAA AAG GTT ACT AAC ACA TAA TAC	2.5	0.3	60	35	204
rs2149556 (S7)	BseLI	F: $(T)_{20}$ GTC TAA ATG GAA TGA TAC TGT ATG TAC TCT TTT A R: GTA AAT GTA TCC CTA TAA TCC CCA AAG TG	2.5	0.3	60	35	237
rs10974944 (S9)	BclI	F: CTG TTC AAG GGT CAA CTG TAG TAC ATA AGA R: $(T)_{30}$ CTG GTT TTG ATC AGG GAT AGT CTC ACT T	2.5	0.3	63	30	208
rs10119004 (S10)	HphI	F: $(A)_{12}$ AAG TAA ATA CAT CCT CAG TAA AAC AAC ATA TAG AA R: CAA GCA CCA CAC AAT ATT AGG AGA GTA	2.5	0.3	60	35	244
rs12343867 (S12)	TaiI	F: $(T)_{26}$ GGT TGA ACA TAA CGT TGG AAT AAC TG R: ATA CTT TTA GTA GTC TCT GTG AAC ACC T	2.5	0.3	60	35	207
rs12340895 (S13)	BslI	F: $(T)_3$ TCA TGA CTT GCC TTA TTA TGG TAG TCT R: $(T)_{10}$ TAT ATA AAC AAT TTT CTT GAA TGT AAC CTT TGT G	2.5	0.3	60	35	122
rs12343065 (S14)	TasI	F: $(T)_{10}$ ACG ATA GTG GTA AGC TCT TTC TCG R: (T) AAA AAG GGC AAG CAA ACA AAA ACC AAG	2.5	0.3	60	35	253
rs3824432 (S16)	PagI	F: GGT TCA TCA TTG TTA GTA TGT TTG TCG CAA R: $(T)_{20}$ GAC TTA AGC CTA TTC ATG AGC CAA AAT CTA	2.5	0.3	60	32	230
rs7847294 (S17)	SspI	F: $(T)_{29}$ TCC CAA TAT TAT ACA CCC CTT TTC GTC CAA C R: GAG AGC ATC CCA AAG CCT GAT AGA ATA AG	2.5	0.3	66	35	260
rs3780378 (S18)	BseGI	F: $(T)_{20}$ GCC TGG GAG GAA CGT CGC A R: CAG GAG TCG GAA CGC CAA GAG	2.5	0.3	60	35	386
rs10815162 (S19)	Bst1107I	F: $(T)_{28}$ TCC ACT AAG CCA AAA CGT TCC CTT GTA TA R: $(T)_8$ TTA AAG CTA TAA AAG AAA GAA CCT GTA TAC CAG AAG	2.5	0.3	60	35	172

Abbreviations: SNP, single nucleotide polymorphism; RFLP, restriction fragment length polymorphism; and T_m, annealing temperature for PCR.

[a] The SNPs are shown in sequential order from the 5' end to the 3' end of the *JAK2* gene (sense strand).

[b] F indicate forward primers, R reverse primers. A few primers have a poly-T tail at the 5' end to enhance the size difference of restricted fragments for easy calling of genotypes.

Table 2.3 PCR conditions for unlabelled probe melting curve analysis

SNP [a]	Primer sequences (5'>3') [b]	PCR conditions				
		Mg^{2+} (mM)	Primer (mM)	T_m (°C)	Cycles	Size (bp)
rs7849191 (S2)	F: GTT GTA TTA GTC ACT TCC TGG R: AAA TTA AGG GGT AGA AAA TGG TAA C rP: TTA AGG AGA TGA ACA GAA GTA GAA GCA-phos	2.5	0.2 0.02	55	55	115
rs7046736 (s3)	F: AGC CAA TTC GAG TCA CTT ATT C R: CAG AAT TCC ATA GGC ACT CAG rP: TTA GAA AAT GCG TTG ATG CTT TGC TAC GAA AAA- phos	3.5	0.2 0.013	55	55	119
rs12342421 (S8)	F: TAG ATC CTT ACT TCA ATA CTG GGT R: ATA TTA AGC ATA GAC TAA ATC AAC TCT TT fP: TCT CTT GGC TAG GAT GTG GTT TAT GTT GAC- phos	2.5	0.02 0.2	55	55	240
rs10974947 (S11)	F: GAT TAA AAA AAT CAA TTC CAA ACT A R: AAA TGT AAT TGT AGA GGA GC fP: GTT TAG CAT TAT GTT AGG AGT GTT ATT AC*T AAA AAA AT* [c]	3.5	0.02 0.2	50	55	207
rs7857730 (S15)	F: TTT TGA GAC ATA ATT TTA AGT GAA TAT AC R: ATT TTA AGC TCA CGG AAC TAT G fP: GTA ATT TTG AGT TAA TGT TCT TTT GCT TTT T- phos	3.5	0.013 0.2	56	55	130

Abbreviations: SNP, single nucleotide polymorphism; UP, unlabelled probe (melting analysis); and T_m, annealing temperature for PCR.

[a] The SNPs are shown in sequential order from the 5' end to the 3' end of the *JAK2* gene (sense strand).

[b] F indicate forward primers, R reverse primers, and fP / rP forward / reverse unlabelled probes. Note that all probes are phosphorylated (phos) at the 3' end to prevent extension by DNA with one exception (Probe of rs10974947; see below).

[c] A poly(A)/(T) tail is added to the 3' end of the probe of rs10974947 to prevent extension by DNA. In fact, the addition of poly (A)/(T) tail was later adapted in our lab.

2.3.7.3 Amplification refractory mutation system (ARMS)

Amplification refractory mutation system was used to detect V617F mutation in cases and controls. In this project, the method from Jones et al. (2005) was adapted and modified. Amplification was performed in 96-well plates with a GeneAmp 9700 PCR system or Veriti® Thermal Cycler both from Applied Biosystems, (Foster City, CA, USA). The PCR reaction was performed in a reaction volume of 15µL containing the following:

30 ng genomic DNA
3.5 mM $MgCl_2$
1× PCR buffer
0.12 mM of each dNTP
0.3 unit of HotStarTaq Plus DNA polymerase
0.3 µM each of the forward and reverse outer primer
0.5 µM of the forward wild type primer
0.6 µM of the reverse mutant primers

The primers were adapted from Jones et al. (2005) as shown below:

PCR primers:

forward outer (**FO**),5'-TCCTCAGAACGTTGATGGCAG-3';
reverse outer (**RO**), 5'-ATTGCTTTCCTTTTTCACAAGAT-3';
forward wild-type-specific (**Fwt**), 5'-GCATTTGGTTTTAAATTATGGAGTATaTG-3';
reverse-mutant-specific (**Rmt**), 5'-GTTTTACTTACTCTCGTCTCCACAaAA-3'.

These four primers can specifically amplify the sequences of wild type and mutant with Fwt or Rmt, and generate a positive control band with the FO and RO. The underlined bases represent mutant/wild type-specific bases. Mismatches (lowercase) were introduced to maximize discrimination of the 2 alleles (wild type and mutant).

Amplification was performed in 96-well plates with a GeneAmp 9700 PCR system or Veriti® Thermal Cycler both from Applied Biosystems, (Foster City, CA, USA). The cycling conditions were modified as follows:

Step	Temp (°C)	Time	
1 Initialisation step	95	5 mins	
2 Denaturation step	95	30 sec	40 cycles
3 Annealing step	58	45 sec	
4 Extension step	72	35 sec	
5 Final extension step	72	5 mins	
6 Hold	15	∞	

Positive, negative, and blank controls were included in each assay. PCR products were pre-stained with SYBR Green I and visualised by electrophoresis on 5% poly-acrylamide gels at 130 V for 90 minutes. Sequencing was performed in representative samples to confirm the mutation status (BigDye® Terminator Cycle Sequencing Kit version 1.1).

2.3.8 Electrophoresis

2.3.8.1 Agarose gel

Agarose gel electrophoresis is the benchmark technique used to isolate DNA fragments according to their sizes. It is also the standard visualisation method for RFLP. Based on the electric field used in this method, DNA molecules migrate according to their sizes which are reverse proportional to their molecular weights. The migration rate also depends on the electric field applied and the charges on the molecules (Westermeier, 2011). Upon gel migration, the bands are visualised with dyes that fluoresce under ultraviolet (UV) light, such as ethidium bromide, a DNA-intercalating stain (McMurray & van Holde, 1986).

Agarose gels were used to estimate the size and examine the specificity of PCR products prior to genotyping. Gel percentage was determined by the fragment sizes of nucleic acid. Gels containing 1.5 % (w/v) of agarose in 1x TBE buffer were used to separate nucleic acid. Prior to gel loading, DNA samples were mixed with 1/5 volume of 6× loading dye. As a molecular weight standard, 1 kb Plus DNA Ladder from Invitrogen Life Technologies (Carlsbad, CA) was used. The electrophoresis was performed in 0.5× TBE (wells were ensured completely covered) at constant voltage adjusted according to sizes of the products and gels. Once the electrophoretic run is complete, gels were stained with ethidium bromide (Sigma, St. Louis, USA). The ethidium bromide -DNA complexes were then visualised by UV light using Chemi Genius^2 BIO imaging system (SYNGENE*;* Frederick, MD, USA).

2.3.8.2 Polyacrylamide gel electrophoresis (PAGE)

Non-denaturing polyacrylamide gels are used to separate and purify fragments of double-stranded DNA molecules according to their sizes. PAGE is formed by the chemical co-polymerisation of acrylamide monomers with a cross-linker, usually N, N'- methylenebisacrylamide (Bis). The chemical polymerisation occurs in the presence of free radicals provided by ammonium persulfate (APS) as catalyst. This process is usually accelerated by a tertiary amino group such as tetramethvlethvlenediamine (TEMED). The pore size is determined by %T, the total (w/v) concentration of acrylamide and cross-linker, and %C, the (w/w) percentage of cross-linker relative to the total amount of acrylamide included in %T. The pore size decreases with increasing %T. When the cross-linking increases, the pore size follows a parabolic function: the pores are large at both high and low cross-linking. However, 4% cross-linking yields the minimum pore size (Westermeier, 2011).

PAGE electrophoresis was used to separate DNA fragments after RFLP genotyping. Gel plates were washed thoroughly before cleaning with alcohol for twice and left air-dried. The plate facing gel matrix was equally treated with silane (Sigma, St. Louis, USA) and wiped dry. This step eased the removal of the gel from the custom-made 96-well mould without tearing the gel. Gel solution with different concentrations according to fragment sizes were prepared with the components listed in Section 2.1.3.2. The gel solution was gently poured onto the custom-made 96-well mould and layered with the silanised glass plate for polymerisation. After polymerisation, the glass plate was removed from the mould and secured into the electrophoresis system containing 0.5× TBE buffer (wells were ensured completely covered). PAGE was generally performed in a volume of 20µl as follows: RFLP-digested products (15µL) were stained with a 5µL mixed solution (6× loading dye and 100×

SYBR® Green I Nucleic Acid Gel Stain prepared in 9:1 ratio) for 30 minutes. 1 kb Plus DNA Ladder from Invitrogen Life Technologies (Carlsbad, CA) was also pre-stained similarly. PAGE was performed in 0.5× TBE buffer at constant voltage adjusted according to sizes of the DNA fragments. Once the electrophoretic run was complete, gels were visualised by UV transilluminator with short band pass filter form Chemi Genius^2 BIO imaging system (SYNGENE*;* Frederick, MD, USA).

2.3.9 DNA Sequencing by Capillary Electrophoresis

To confirm the V617F mutation detected by ARMS and also genotypes obtained by RFLP and unlabelled probe melting analysis, direct DNA sequencing was carried out with ABI PRISM 3130 Genetic Analyser from Applied Biosystems, (Foster City, CA, USA). Representative dummy samples with different genotypes were sequenced to attest the genotypes obtained before genotyping real studied subjects. Direct sequencing was also the ultimate alternative when certain samples gave ambiguous genotypes with RFLP or melting curve analysis.

The sequencing was performed on specific PCR products amplified with the same primer pairs for each SNP. The specificity of PCR products were examined by agarose gel. The sequencing reaction was performed with BigDye® Terminator Cycle Sequencing Kit (version 1.1, Applied Biosystems) as described below.

2.3.9.1 Enzymatic pre-treatment of PCR product

To prepare PCR products for fluorescent-based sequencing, each product was first incubated with a combination of two hydrolytic enzymes: exonuclease I (Exo I, New

England Biolabs) which digests excess primers, and shrimp alkaline phosphatase (SAP, GE Healthcare) which removes 5' phosphate group from nucleotides, thereby maintaining them as unincorporated nucleotides in PCR reactions to prepare templates for the sequencing reaction.

First, a mixture of Exo I-SAP was prepared with 1 unit of Exo I (20U/ µl) and 1 unit of SAP (5U/ µl) and made up to 0.40µL. PCR product (6µL) was then added into the Exo I-SAP mixture (0.40µL) and made up to 7µL with 10× SAP buffer. This combination was incubated at 37°C for 30 minutes for purification, and heated to 80°C for 20 minutes to inactivate the enzymes. The incubation was performed in GeneAmp 9700 PCR system or Veriti® Thermal Cycler both from Applied Biosystems, (Foster City, CA, USA). The preparation was as follow:

Exo I- shrimp alkaline phosphatase (SAP) mix*:

Shrimp alkaline phosphatase (5U/ µl)	0.20µl
Exo I (20U/ µl)	0.05µl
Water	0.15µl
	0.40µl
Specific PCR products	6µl
10× SAP buffer	0.6µl
Exo-AP mix*	0.4µl
	7µl

PCR process

Step	Temp (°C)	duration
1	37°C	30min
2	80°C	20min
3	15°C	∞

2.3.9.2 Cycle sequencing reaction

After purification and enzyme inactivation, cycle sequencing was conducted. A reaction mix containing of BigDye® Terminator v 1.1, single primer (forward or reverse) at a final concentration of 0.01µM, UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen Life Technologies, Carlsbad, CA) was added separately to the PCR template for cycle sequencing reaction. The reaction mix for typical cycle sequencing reaction was as follows:

Cycle sequencing mix

PCR template	6µl
BigDye® Terminator v 1.1	2.0µl
0.01µM primer (1/10 dilution)	1.6µl
UltraPure™ DNase/RNase-Free Distilled Water	0.4µl
	10µl

Cycle sequencing conditions were as follows:

Step	**Temp (°C)**	**duration**	
1	96°C	1min	1 cycle
2	**96°C**	**10s**	
3	**50°C (Tm)**	**30s**	**37 cycle**
4	**60°C**	**6min**	
5	15°C	∞	

2.3.9.3 Removal of unincorporated dye terminators

Ethanol/sodium acetate precipitation method was used to remove excess dye-terminators. Products from cycle sequencing were transferred to microcentrifuge tube (600uL) containing a mixture of 95% ethanol (Sigma-Aldrich, St. Louis, USA),

3M sodium acetate (Sigma-Aldrich, St. Louis, USA), and UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen Life Technologies, Carlsbad, CA).

Precipitation mixture

95% ethanol	62.5µl
3M sodium acetate (pH 4.6)	3µl
MilliQ	24.5µl
DNA template	10µl
Total	**100µl**

The mixture was incubated in dark at -20°C for 15 minutes and centrifuged at 14000 *g* (12000 rpm) at 4°C for 20 minutes. The samples were incubated not more than 15 minutes to avoid very small fragments from going back into the solution because it may reduce signal strength at the start of the sequencing reaction. The supernatant was carefully discarded and 300µl of 70% ethanol was added immediately to the pellet to obviate salt and free nucleotides from sticking to the tube when ethanol evaporates. The resultant mixture was centrifuged at 14000 *g* at 4°C for another 10 minutes. The supernatant was carefully discarded and any remaining liquid residue was dried using Savant DNA 110 SpeedVac®concentrator centrifuge at a medium drying rate for approximately 15 minutes. After drying, DNA pellet was re-suspended in 15µl of Hi-Di™ Formamide from Applied Biosystems, (Foster City, CA, USA) and incubated at 4°C for another 15 minutes.

2.3.9.4 Sequencing analysis

The DNA sequences of the samples were read by capillary gel electrophoresis on ABI PRISM 3130 Genetic Analyser following the manufacturer's instructions. DNA Sequencing Analysis Software (version 5.2 Patch 2) from Applied Biosystems was used to process and visualize the sequencing data. An example of a successful sequencing for a SNP genotyped is depicted in **Figure 2.2**.

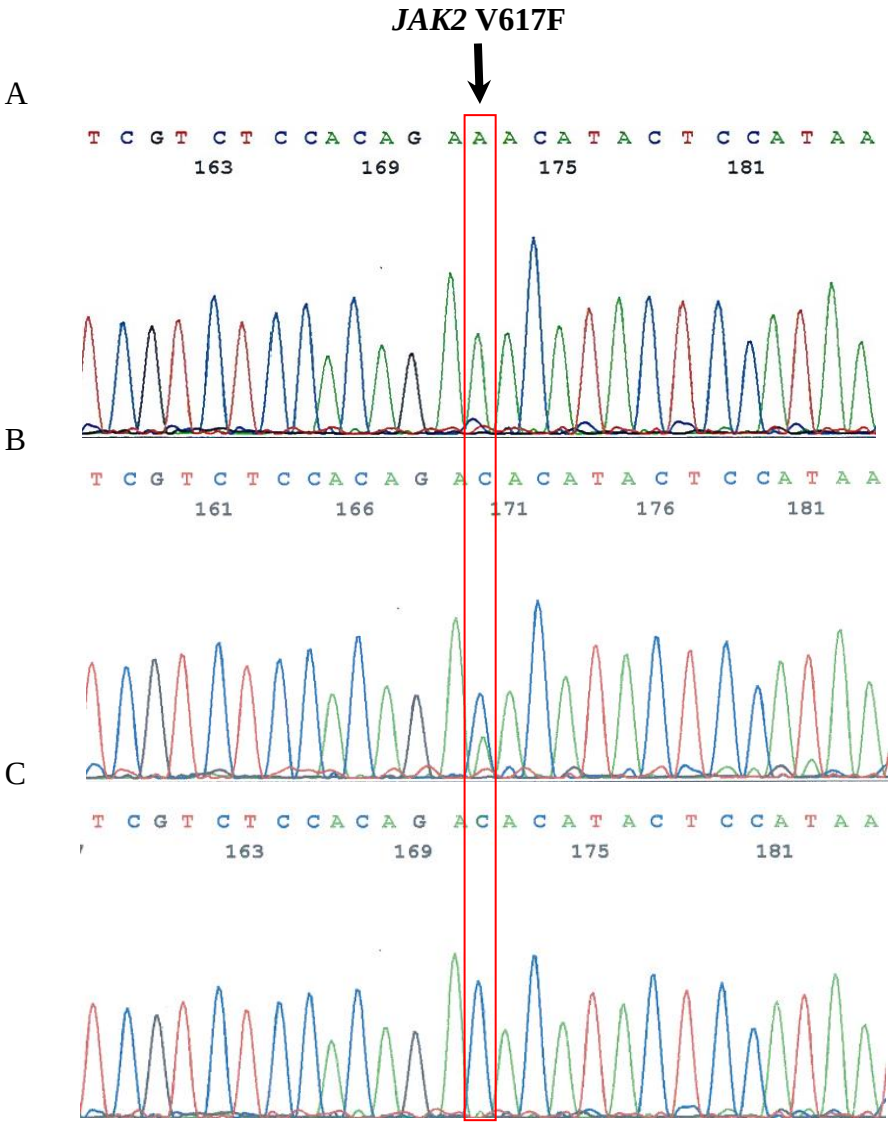


Figure 2. 2 Validation of *JAK2* 1849 G>T variant by Sanger sequencing.

Sequence chromatograms corresponding to the reverse strand of sequenced PCR products for the *JAK2* 1849 G>T mutation (A) Homozygous mutant genotype TT (B) Heterozygous genotype GT (C) Homozygous wild type genotype GG

2.3.10 SNP genotyping

As covered in Section 1.6.5, two common methods were used for genotyping: RFLP (Section 1.6.5.1) and unlabelled probe melting curve analysis (Section 1.6.5.2).

2.3.10.1 Restriction fragment length polymorphism (RFLP)

Fourteen SNPs were genotyped by RFLP analysis (**Table 2.2**). Prior to enzymatic digestion, DNA fragments containing the SNP site of interest and internal digestion control site (a non-polymorphic site) were amplified with primers designed with OLIGO version 6.62 software (Molecular Biology Insights, Inc., USA). PCR was performed in a 10-µL reaction mixture (Section 2.2.7.1) in 96-well plates with a GeneAmp 9700 PCR system or Veriti® Thermal Cycler both from Applied Biosystems, following the conditions for each SNP (**Table 2.2)**. Specific restriction enzymes were added to the PCR products according to the manufacturer's recommendations. The reaction mixture of PCR products and restriction enzymes (MBI Fermentas or New England Biolabs) were incubated at different reaction temperatures as recommended by the manufacturer for 12 to 16 hrs. The reactions were stopped by inactivating at 80°C for 20 minutes. Digested products were pre-stained with SYBR Green I and then separated by electrophoresis in polyacrylamide gels of appropriate concentration (Yiu et al., 2013; Zhou et al., 2004). In all cases, UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen Life Technologies, Carlsbad, CA) was used in the reactions. A gel picture of RFLP performed on one of the SNPs is shown in **Figure 2.3**.

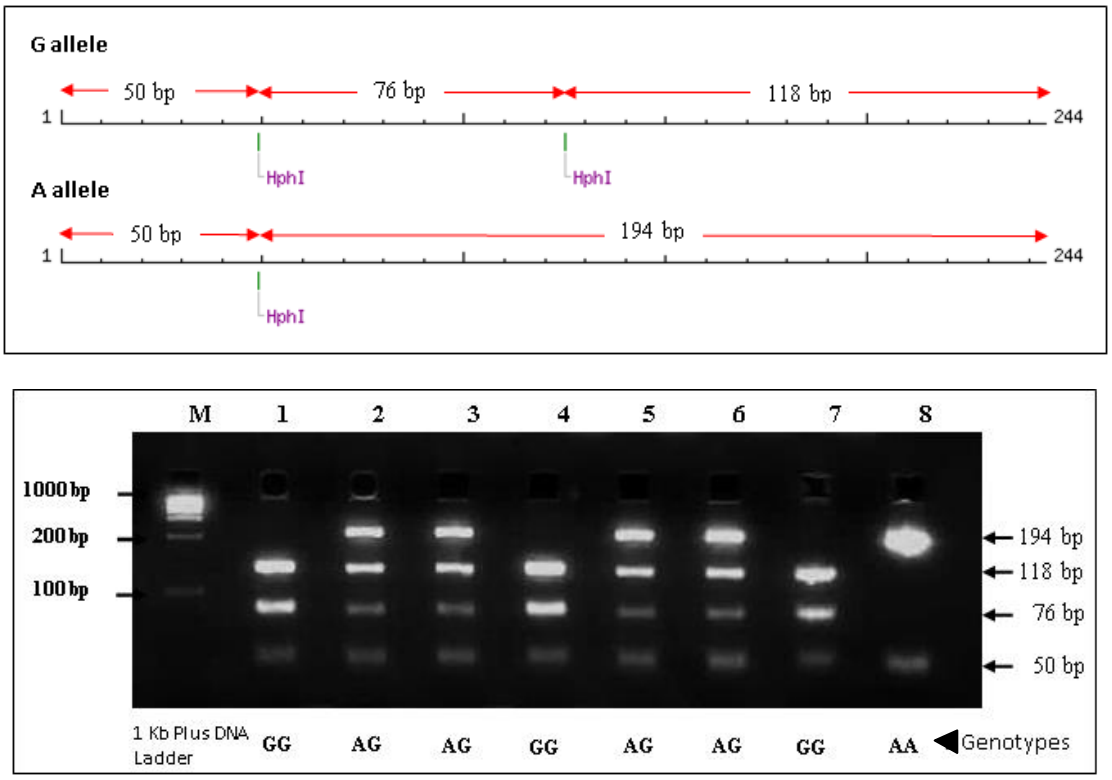


Figure 2. 3 SNP genotyping by restriction fragment length polymorphism.

For illustration, rs10119004 (S10) is used as an example. A PCR fragment of 244 bp is amplified to encompass the SNP site (see **Table 2.2**). **Upper panel:** Restriction patterns for the two alleles of rs10119004 (S10) upon restriction digestion by the restriction enzyme HphI. **Lower panel:** Electrophoresis banding patterns on 8% polyacrylamide gel and stained with SYBR Green I. The DNA ladder is the 1 kb Plus DNA Ladder (lane M) from Invitrogen Life Technologies.

2.3.10.2 Unlabelled probe melting curve analysis (UPR)

Unlabelled probe melting curve analysis was used to genotype the SNPs when RFLP failed to genotype the SNPs due to the following reasons: there was no restriction site in the sequence, or the restriction enzyme was too expensive. in this study, five SNPs were genotyped by UPR using the saturating dye SYTO 9 green fluorescent nucleic acid stain (final concentration of 2 μM; Invitrogen Life Technologies (Carlsbad, CA) (Yiu et al., 2013; Zhou et al., 2004). This method uses asymmetric PCR to generate single-stranded DNA (ssDNA) product. A phosphate group or a poly (A)/ (T) tail is added to the 3' end of the unlabelled probe to prevent probe extension. Asymmetric PCR was performed in a 10-μL reaction mixture (Section 2.2.7.2) in 96-well plates with a GeneAmp 9700 PCR system or Veriti® Thermal Cycler both from Applied Biosystems, (Foster City, CA, USA) following the conditions for each SNP (**Table 2.3)**.

After PCR, the probe and a saturated dsDNA dye were added to the ssDNA target for high-resolution melting analysis. A final 10-μL reaction mixture containing 9.1 μl of PCR product, 0.6 μM 3'-blocked probe (synthesised by IDT) and 2 μM SYTO 9 green fluorescent nucleic acid stain (Invitrogen) was prepared in 96-well white plates, and subjected to melting in LightCycler® 480 Real-time PCR System (Roche). Probe/ssDNA amplicon duplexes were generated by heating samples to 95°C for 30 seconds, then cooling to 50°C for 30 seconds. The melting data were collected between 50°C and 95°C with a slope of 0.11°C/s at 5 acquisitions per °C, using the "melting-curves" analysis mode. Samples were again cooled to 40°C for 10 seconds and the melting curves were analysed with LightCycler® 480 Software (version 1.5,

Roche). An example of the melting curve obtained for *JAK2* rs7849191 (S2) is depicted in **Figure 2.4**.

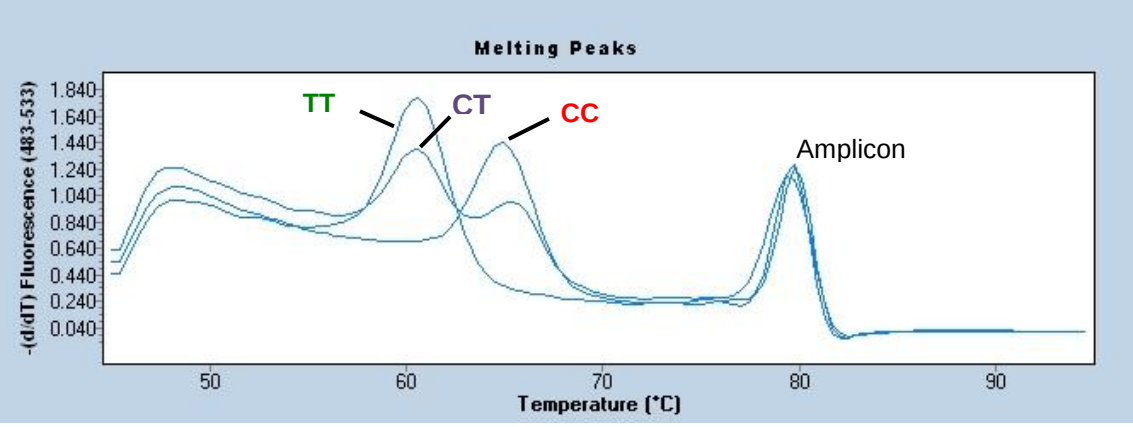


Figure 2. 4 SNP genotyping by unlabelled probe melting analysis.

For illustration, rs7849191 (S2) is used as an example. A PCR fragment of 115 bp (see **Table 2.3**) is amplified to encompass the SNP site. The unlabelled probe is designed to match the C allele in this example. The probe-amplicon duplex of the homozygous genotype CC has a higher melting temperature than the probe-amplicon duplex of the homozygous genotype TT. Two peaks are obtained for the heterozygous genotype CT.

2.4 Statistical analysis for genetic association study

Prior to data analysis, genotypes were rested for Hardy-Weinberg equilibrium (HWE) by Fisher's exact test using PLINK (version 1.07; http://pngu.mgh.harvard.edu/,purcell/plink/index.shtml) (Purcell et al., 2007) for cases and controls, separately. In large populations with random mating, the genotype distribution for an individual marker should abide the principle of HWE, assuming the absence of genetic drift, selection, migration, or mutation affecting allele frequencies. Deviation from HWE among control subjects in genetic association study may suggest genotyping errors because there should be no bias of the control subjects in a well-designed study population (Xu et al., 2002; Hosking et al., 2004; Zintzaras, 2010; Hardy, 1908). Therefore, a default *P* value of 0.001 (Purcell et al., 2007) was used as the significance threshold to safeguard the validity of results generated in this study.

PLINK (version.1.07; http://pngu.mgh.harvard.edu/,purcell/plink/index.shtml) (Purcell et al., 2007) was used as the main toolset for statistical analysis of SNP data in this study.

2.3.1 Gender and age adjustment

Cases and controls were matched for sex and age as much as possible, however they were not perfectly matched.

Therefore, single-marker analysis was conducted between cases and controls with logistic regression adjusted for sex and age (age at diagnosis for MPN patients) as

covariates to avoid any confounding factor. The respective asymptotic *P* value adjusted for age and sex was denoted as P_{asym}.

Unpaired t-test was used to calculate the difference of age and sex distribution between cases and controls.

2.3.2 Single marker analysis of *JAK2* SNPs

For single marker analysis, PLINK (version 1.07; http://pngu.mgh.harvard.edu/,purcell/plink/index.shtml) (Purcell et al., 2007) was used for statistical analysis for all the 19 directly genotyped SNPs and the 76 imputed SNPs, which were not directly genotyped in our initial study, and also the haplotype association tests.

Cases and controls were compared for their allele and genotype frequencies to obtain the corresponding *P* values for each individual SNP. Association tests were performed for five genetic models (allelic, genotypic, additive, dominant and recessive) via chi-square or trend tests provided by PLINK. Correction for multiple comparisons was done with a type of re-sampling randomisation method namely permutation. The correction was achieved by generating empirical *P* values (P_{emp}) based on 50,000 permutations, i.e., randomly assigning the case/control status to each studied subject for 50,000 times. The P_{emp} from permutation test was generated based on the best result among allelic, dominant, and recessive models. In each permutation, the affection status (either cases or control) of the study subjects was randomly shuffled to remove any genotype/ phenotype association but retain the correlation among

genotypes and among phenotypes based on the LD information within each individual. This process was done with the constraint that the total numbers of cases and controls, and genotypes of each subject must remain unchanged for all the SNPs selected for correction. A test statistic was calculated for the permuted data based on the original data set in which the original test was counted as one of the permuted tests, and recomputed on each permuted data. The P_{emp} is the fraction of the permuted data sets that give the minimum *P* value or of the permuted test statistic greater than or equal to (i.e. more significant than or as significant as) that obtained from the original data set association tests. According the PLINK, the empirical *P* value is calculated as (R+1)/ (N+1) where R is the number of times the distinct permuted test statistic is greater than observed value of the statistic; N is the total number of permutations (N = 50 000 in this study). Permutation test is widely considered the gold standard for accurately correcting multiple comparisons in GWAS because it provides unbiased type I error control and high power, and estimates the probability of the null hypothesis. It is widely applied to genetic association studies because it is less conservative than the Bonferroni procedure, which assumes the markers to be independent of each other (Purcell et al., 2007; Lunetta, 2008).

2.3.2 Haplotype analysis of *JAK2* SNPs

As in single marker analysis, PLINK (version 1.07; http://pngu.mgh.harvard.edu/,purcell/plink/index.shtml) (Purcell et al., 2007) was used to search for MPN-associated haplotypes. Haplotype analysis was also conducted between cases and controls with logistic regression adjusted for sex and age (age at diagnosis for MPN patients) as covariates as mentioned above; the respective asymptotic *P* value was denoted as P_{asym}. Haplotype frequencies were estimated using

the standard algorithm of expectation and maximisation (E-M algorithm). Logistic regression was performed for haplotype association analysis using the exhaustic variable-sized sliding window method, in which a 'window' consisting of neighbouring SNPs are analysed together in a stepwise fashion across the target region (Li et al., 2007). Sliding windows of all possible sizes (1 to n SNPs per window, where n is the number of *JAK2* SNPs examined) were exhaustively examined to search for the MPN-associated haplotypes. For a given window size, the test was performed for the entire sliding windows by shifting one SNP at a time towards the 3' end of the gene. For haplotype analysis in a case control association analysis, a single case-control omnibus test of (H -1) degree of freedom was executed to jointly evaluate the significance of the haplotypic effects at that position, with covariate adjustment of gender and age; where H is the number of haplotypes in the sliding window being examined. The odds ratio (OR) of a particular haplotype was calculated with respect to the reference haplotype, i.e. to all other haplotypes of the same haplotype window. The reference group of haplotypes was not fixed for different haplotypes being considered.

Correction for multiple comparisons was achieved by generating empirical *P* values (P_{emp}) based on 50,000 permutations, i.e., swapping of the case-control status 50,000 times. Haplotypes were defined by a variable-sized sliding-window approach based on all possible sizes of SNPs spanning the whole genomic region. This optimum marker combination showed no bias towards either single SNP or haplotype. Subsequently, the contribution of individual SNPs to a significant haplotype association with MPNs was examined by conditional logistic regression analysis. Haploview version 4.2 (Barrett et al., 2005) was used to generate the linkage disequilibrium (LD) map of the *JAK2* gene based on an algorithm called solid spine of linkage disequilibrium (SSLD) (Barrett et al., 2005). SSLD defines a haplotype block upon discovering

a ‘spine’ of strong LD within a block: the first SNP and the last SNP of the block are in strong LD consisting all intermediate SNPs that are not necessarily in strong LD with each other (Barrett et al., 2005).

2.3.3 Imputation

SNPs imputation was performed using Beagle version 3.2 (Browning & Browning, 2009). The genotype data of the 1000 Genomes Project (phase 1) based on 97 CHB subjects were used as the reference panel. A quality control check was manually done by removing some of the known genotypes of the 19 directly genotyped SNPs, followed by imputation process done with Beagle version 3.2. The post-imputation results were merged with the original data to check for the imputation accuracy based on the known missing genotypes.

2.3.4 In silico analysis

In silico analysis was performed to obtain information about the SNPs. Databases, programs and online tools used in this study are summarised in **Table 2.4**.

Table 2. 4 Programs and databases used in this study

Program/ Database	Application	Source /URL
Genetic Power Calculator	Sample size and power calculation	http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html
HapMap (phase II , release 24)	Selection of tSNPs based on the interface with the Tagger package	http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap24_B36/
WatCut SNP-RFLP analysis	Searching of restriction enzyme cleavage sites of SNPs for PCR-RFLP genotyping	http://watcut.uwaterloo.ca/template.php?act=snp_new
NEBcutter (version 2.0)	Searching of appropriate restriction enzymes for PCR-RFLP	http://tools.neb.com/NEBcutter2/
OLIGO version 6.62	For designing primers and probes	Molecular Biology Insights, Inc., USA
Tm Utility™ (version 1.3)	Prediction of the T_m for unlabelled probes	From Idaho Technology's Web site at www.idahotech.com
Primer-BLAST	Checking of the specificity of primers against the genome	http://www.ncbi.nlm.nih.gov/tools/primer-blast/
Chemi Genius^2 BIO imaging system	Visualisation and image capture of gels	SYNGENE; Frederick, MD, USA
LightCycler® 480 Software (Version 1.5)	Processing and analysing of melting curves generated by LightCycler® 480	Roche, Basel, Switzerland
DNA Sequencing Analysis Software (version 5.2 Patch 2)	Processing and visualisation of sequencing data	Applied Biosystems, Foster City, CA, USA
PLINK (version 1.07)	Statistical analysis for genetic association data: HWE testing and association testing (single marker and haplotype analysis)	http://pngu.mgh.harvard.edu/~purcell/plink/
BEAGLE (version 3.2)	Imputation	http://faculty.washington.edu/browning/beagle/beagle.html
The 1000 Genomes Database	Obtaining flanking sequence of SNPs	http://www.1000genomes.org/

Table 2.4 Programs and databases used in this study (Continued)

Program/ Database	Application	Source/ URL
The 1000 Genomes Database	Downloading genotype data from the 1000 Genomes Project (phase 1) of 97 CHB individuals as the reference panel for imputation	http://www.1000genomes.org/
Haploview (Version 4.2)	LD maps construction	http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview
SNPnexus	Functional annotation of SNPs	http://snp-nexus.org/index.html
SNP Function Prediction (FuncPred)	Functional annotation of SNPs	http://manticore.niehs.nih.gov/snpfunc.htm
F-SNP	Functional annotation of SNPs	http://compbio.cs.queensu.ca/F-SNP/
eQTL resources @ the Pritchard lab	Expression quantitative trait locus *(*eQTL*)* analysis	http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/
seeQTL	Expression quantitative trait locus *(*eQTL*)* analysis	http://www.bios.unc.edu/research/genomic_software/seeQTL/
UCSC Genome Browser	Expression quantitative trait locus *(*eQTL*)* analysis	http://genome.ucsc.edu/cgi-bin/hgTracks?org=human
MatInspector	Prediction of putative transcription factor binding sites	www.genomatix.de/matinspector.html
Cochrane RevMan software (version 5.3.3)	Meta-analysis	http://tech.cochrane.org/revman/download
Institute of Human Genetics-Genetic statistics	HWE testing	http://ihg.gsf.de/cgi-bin/hw/hwa1.pl

Table 2.4 Programs and databases used in this study (Continued)

Program/ Database	Application	Source/ URL
z to P Calculator	Calculation of *P* values from Z value for |z|≥5	http://vassarstats.net/tabs.html#z
GraphPad Software - P value Calculator	Calculation of *P* values from Z value	http://graphpad.com/quickcalcs/PValue1.cfm
LocusZoom - Plot with Your Data	Regional visualisation of association signal relative to genomic position, local linkage disequilibrium (LD) and recombination patterns and the positions of genes in the region	http://csg.sph.umich.edu/locuszoom/
ANNOVAR	Functional annotation of genetic variants from high-throughput sequencing data	http://www.openbioinformatics.org/annovar/
False Discovery Rate Calculator for 2x2 Contingency Tables	Estimating false discovery rates for contingency tables	http://research.microsoft.com/en-us/um/redmond/projects/MSCompBio/FalseDiscoveryRate/
MutationTaster software	Variant detection for potential pathogenic effects	http://www.mutationtaster.org/ChrPos.html

CHAPTER 3 Detection of V617F in MPNs

3.1 Introduction

Most Ph- negative MPN patients carry an acquired mutation in the *JAK2* gene in their tumours. It is believed that the acquired single point somatic gain-of-function mutation of the *JAK2* kinase gene (1849G>T) on chromosome 9 (Tefferi & Gilliland, 2007) (Jones et al., 2009), constitutively activates the proliferation of myeloid cells by a valine-to-phenylalanine substitution in codon 617 (encoding V617F). Comparing with other mutations found in MPNs (**Table 1.3**), only PV, ET, PMF show significant levels of V617F mutation (Tefferi & Gilliland, 2007).

The V617F mutation was reported to occur in nearly all patients with PV and in around half of those with ET and PMF. Nonetheless, the worldwide prevalence of V617F mutation reported was different ranges from 63% to 100% for PV, 23% to 79% for ET, and 25% to 95% for PMF (Lippert et al., 2006; Nelson & Steensma, 2006; Sazawal et al., 2010; Jones et al., 2005; Suksomyos et al., 2012). The diagnostic criteria and assay sensitivity applied in individual study, or inadequate sample size could be the reasons accounting for the different rates in each individual study.

3.1.1 Research aim

When this project started in 2009, there were only a few V617F detection studies done in Asian populations but none in Hong Kong Chinese population (Lieu et al., 2008; Zhang et al., 2008; Xiao et al., 2008; Chen et al., 2007). Subsequently, more studies were done in Chinese populations. The trend of this V617F mutation occur-

rence ratio was reported with slightly different rates (**Table 2**) due to the differences in diagnostic criteria and assay sensitivity (Wong et al., 2011; Shen et al., 2009).

However, the same conclusion was drawn as in Caucasian population: V617F is indeed a major molecular pathogenesis in Chinese MPN patients. Still, little information was available from Hong Kong population (**Table 1.4**). Therefore, the aim of this part of study was to:

1. Investigate the involvement of V617F mutation in local Hong Kong Chinese population with MPNs and also healthy control
2. Determine the prevalence of V617F mutation in local Hong Kong Chinese population with MPNs

3.2 Methodology

Peripheral blood samples were obtained from affected and unaffected individuals. To be consistent, all recruited MPNs cases were screened for V617F mutation by amplification refractory mutation system modified from Jones et al. (2005). ARMS products were analysed by electrophoresis on 5% polyacrylamide gels. Details are provided in Section 2.2.7.3.

3.2.1 Patient samples

DNA samples extracted from patients' peripheral blood were used for V617F detection. DNA was extracted with FlexiGene DNA Kit (Qiagen) according to the manu-

facturer's instructions and quantified by NanoDrop ND1000 UV/Vis spectrophotometer (Thermo Scientific, Wilmington, USA).

3.2.2 Control samples

DNA samples from controls were extracted from peripheral blood and were used for both V617F detection and genotyping. DNA extraction was performed with FlexiGene DNA Kit (Qiagen) according to the manufacturer's instructions and quantified by NanoDrop ND1000 UV/Vis spectrophotometer (Thermo Scientific, Wilmington, USA).

3.2.3 V617F mutation detection by amplification refractory mutation system (ARMS)

DNA samples were evaluated for the presence of guanine to thymine transversion in *JAK2* exon 14 (encoding V617F) with the application of Amplification Refractory Mutation System (ARMS) modified from Jones et al. (2005). The PCR products – control band (463bp) were directly sequenced in both directions on ABI PRISM 3130 Genetic Analyzer with BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) to make sure that the primers work correctly prior to real sample screening.

3.3 Results

3.3.1 Summary of MPNs patients in Hong Kong

In this case-control studied cohort, participants were Chinese MPNs patients diagnosed according to WHO 2008 criteria recruited from seven local hospitals and healthy blood donors from Hong Kong Red Cross Blood Transfusion Service. With written informed consent, 192 MPNs patients were recruited from 7 local hospitals. Of these, 172 patients with clear diagnosis of MPNs were included in this project: 61 with PV, 93 with ET, 17 with PMF, and 1 with unclassified MPN, and 86 males (50.0%) and 86 females (50.0%). Their mean age was 57 years (ranges: 15-88 years). For the control group, 470 healthy controls consisting of 236 males (50.2%) and 234 females (49.8%) were recruited. The mean age of the control subjects was 51 years (ranges: 16-75 years).

3.3.2 The distribution of V617F mutation in PV, ET, and PMF

All cases and controls were screened for *V617F* mutation by ARMS adapted from Jones et al. (2005). Sequencing reaction was performed on representative samples to confirm the genotypes detected by ARMS. The results of ARMS were visualised by polyacrylamide gels as displayed in **Figure 3.1**. Overall, 128 (74.4%) MPN patients were positive and 44 (25.6%) negative for V617F. The prevalence of V617F in our cohort was 87% (53/61) in PV, 68% (63/93) in ET, 65% (11/17) in PMF, and 100% (1/1) in unclassified MPN (**Table 3.1**).

Fisher's exact test suggested no significant difference in sex ratio between the two groups (P>0.27). Age differed significantly between V617F-positive MPN cases and

healthy controls ($P<0.0001$) whereas there was no difference in age between V617F-negative MPNs and controls ($P=0.7342$). However, there was still statistically significant difference in age between *all MPN cases* (both V617F-positive and -negative) and controls ($P<0.0001$). On the whole, the mutation frequency did not differ by sex and age in the patient group.

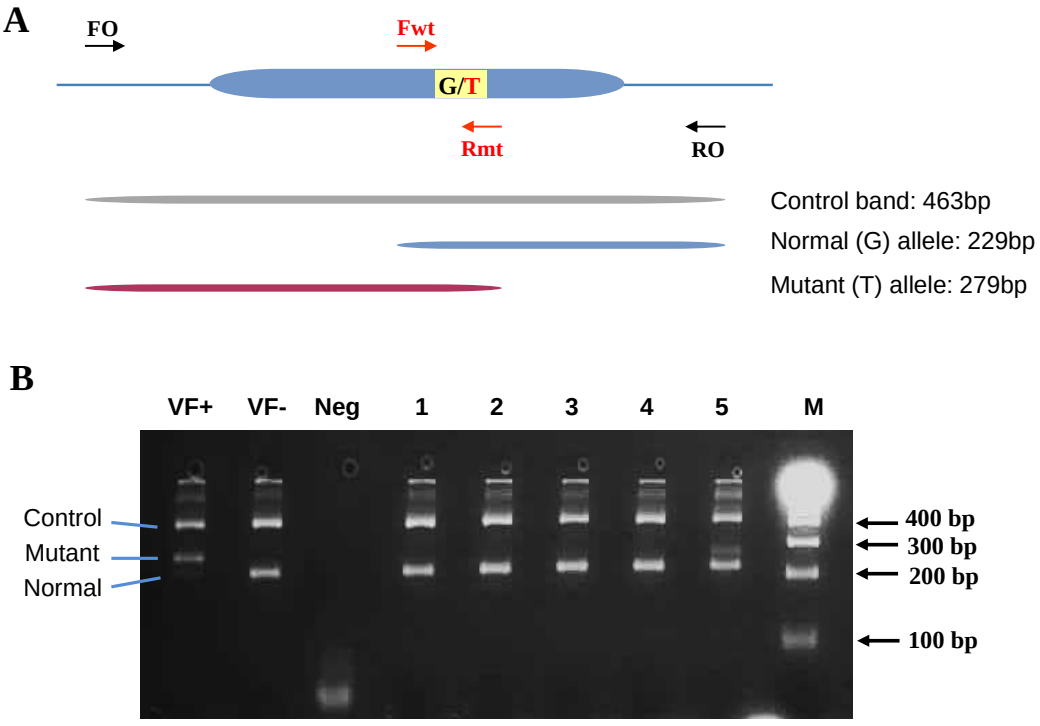


Figure 3.1 Representative results of V617F mutation by ARMS.

(A) Schematic representation of ARMS assay. As control band for successful amplification of the sequence flanking V617F mutation, a pair of outer primers (FO and RO) should generate a 463-bp product in all samples. Primers Fwt and RO should generate a wild-type-specific product of 229-bp while primers FO and Rmt should generate a mutant-specific product of 279-bp. (B) Representative results of V617F detected by ARMS in this project as displayed on 5% polyacrylamide gel. Lanes 1-4 show a normal genotype without the mutant band; track 5 show a mutant band for the V617F mutation; VF+ and VF- represent V617F-positive and negative control bands; Neg represents template free control; and M is the 1 kb Plus DNA Ladder from Invitrogen Life Technologies (Carlsbad, CA).(Jones et al., 2005) Sequencing results confirming the genotypes in ARMS can be found in **Figure 2.2**.

Table 3.1 Prevalence of V617F mutation in MPN patients and controls.

MPN subtype	N	Age at diagnosis, Y[a] (range)	V617F-positive N (%)	V617F-positive N,Gender; Y[a] (range)	V617F-negative N (%)	V617F-negative N,Gender; Y[a] (range)
PV	61	56 (21-86)	53 (87)	30M, 23F; 57 (21-86)	8 (13)	8M, 0F; 54 (41-69)
ET	93	56 (15-88)	63 (68)	24M, 39F; 58 (25-88)	30 (32)	15M, 15F; 52 (15-85)
PMF	17	59 (31-79)	11 (65)	5M, 6F; 65 (48-79)	6 (35)	3M, 3F; 48 (31-70)
uMPN	1	62	1 (100)	62	-	-
Total	172	57 (15-88)	128 (74)	60M, 68F; 58 (21-88)	44 (26)	26M, 18F; 51 (15-85)
Control	470	51 (16-75)[b]	0 (0)	-	470 (100)	236M, 234F; 51 (16-75)

Abbreviations: V617F, *JAK2* V617F mutation; MPN, myeloproliferative neoplasms; N, number; Y, years.

[a] Median age in terms of years

[b] Median age (years) of controls at the age of blood donation. Control subjects were matched for sex- and age (±1) as much as possible.

3.4 Discussion

In 2005, identification of single recurrent MPN-associated dysregulation of *JAK2* kinase activity within the pseudokinase domain in a number of MPNs patients (Ihle & Gilliland, 2007a), has widened our understanding in the pathogenesis of *BCR-ABL*-negative myeloproliferative neoplasms (MPNs), namely PV, ET, and PMF.

The V617F mutation is highly associated with MPNs in addition to other less common *JAK2* mutations. In 2008, this mutation was included as a standard diagnostic criterion for MPNs (Vannucchi et al., 2009a). Since the detection of V617F mutation indicates the diagnosis of these disorders, screening of V617F could be helpful in early detection of the disorders. Therefore, despite the V617F mutation screening done by hospitals, in house V617F screening with ARMS was still performed on MPN patients for the sake of consistency.

The results from this project corroborate the worldwide prevalence of V617F mutation in the majority of patients with PV and nearly half of the patients with ET and PMF. The prevalence of V617F mutation was higher in patients with PV than in patients with ET and PMF. Altogether, the data in our Hong Kong Chinese population suggested that both genders can be equally affected with MPNs in their late fifth decade of life, (**Table 3.1**).

3.4.1 V617F mutation is found in MPN patients but not in controls

To study the association between *JAK2* germline polymorphisms and MPNs in Hong Kong Chinese population, V617F mutation was first screened in MPN cases and healthy controls.

In this project, the V617F mutation was detected in the majority of PV patients (87%) and approximately half of those with ET (68%) and PMF (65%). The prevalence of V617F mutation in this project agreed with previous reports although different results had been reported. The worldwide prevalence of V617F mutation reported for PV was the highest ranging from 63% to 100%; while the average prevalence of V617F in ET and PMF was approximately 50% ranging from 23-79% for ET; and 25-95% for PMF (**Table 1.4**). The discrepancies in the reported findings of V617F mutation may be a consequence of the different diagnostic criteria and methods used to identify mutations in individual study (Shen et al., 2009). Over the years, various methodologies including direct DNA sequencing, pyrosequencing, allele-specific PCR, ARMS, and PCR-RFLP screening had been used to study the V617F mutation (**Table 1.4**) (Veneri et al., 2009; Steensma, 2006). These techniques differ in terms of sensitivity and performance characteristics and thus may lead to the discrepancies in V617F prevalence. The diagnostic criteria applied in individual study might also affect the discrepancies of V617F prevalence. Among others, small sample size could be another reason, such as that in a Mexican study with 4 PMF, 8 PV, and 17 ET patients (Ruiz-Arguelles et al., 2009). Inadequate sample size can lead to unreliable evidence: false positive or false negative results (Lurati Buse et al., 2012; Hall, 2011).

The low annual incidence of MPN (around 2 per 100,000 people) suggested that general population should be considerably healthy (Goldin et al., 2009; Vannucchi et al., 2009b). Among the MPN patients, V617F mutation was reported to attack majority patients with PV and nearly half of ET and PMF. With these data in mind, V617F mutation was hypothesised to be absent in the control population. In this project, the mutation was not detected in the 470 healthy controls as hypothesised. Although the mutation does not affect all MPN patients, its occurrence among healthy individuals might mark an early molecular onset prior to the development of MPN (Kralovics et al., 2005b). Studies conducted by Sidon et al. (2006) and Xu et al. (2007), reported V617F mutation in a minority of the healthy individuals. In 2 larger studies conducted by Nielsen et al. (2011; 2014), 18 of 10507 (0.2%) and 63 of 49488 (0.1%) individuals without overt signs of MPNs were tested positive for V617F mutation. Among the V617F-positive healthy individuals, some were diagnosed with MPNs based on their haematological parameters together with an increase in their V617F mutation burden during a follow up several years later. This suggested that V617F mutation might happen prior to the occurrence of MPNs. The group also pointed out an association between V617F mutation burden level and development of MPNs. As examined in their study cohort, V617F mutation burden level was associated with MPN's progression rate in the order of increasing severity of MPNs from MPN-free to ET to PV and lastly to PMF (Nielsen et al., 2014). Thus, if this mutation is detected at a higher sensitivity, it can aid the early detection of MPNs. Nonetheless, their interpretation was based only on Caucasian population hence studies in different races may reveal more information regarding the role of V617F in the development of MPNs.

Limitation of this project should be considered. The ARMS detection method used to study V617F mutation in this study is a highly sensitive qualitative detection method. However this method does not allow quantitative detection of the somatic mutation (Jones et al., 2005). Therefore, allelic burden of V617F mutation could not be quantified in this project. In addition to this, clinical status was not assessed and correlated with mutational status. Thus, clinical impact could not be assessed in this study.

In summary, the distribution of V617F of our Hong Kong MPN patients (PV, ET and PMF) in this project is similar to those reported in other studies around the globe (Jones et al., 2005; Baxter et al., 2005; James et al., 2005b; Kralovics et al., 2005a). On the other hand, the diagnosis of V617F-negative MPNs has to be dependent on haematological parameters and other molecular criteria. Future studies on larger sample cohort may give further insight on the pathogenesis of V617F-positive and -negative MPNs. Nonetheless, the results from my part I study corroborate previous reports on the prevalence of V617F in other populations and thus provide a justification for my part II study.

CHAPTER 4 Genetic association study of germline polymorphisms in *JAK2* gene

4.1 Introduction

4.1.1 Genetic association studies in MPNs

It has been more than ten years since the first discovery of the key mutation in MPNs and that was when MPNs no longer orphan disorders. Despite updates from scientists, haematologists, and pathologists; there is still no answer to how one mutation can interact with *JAK2* germline polymorphisms and lead to 3 related yet phenotypically distinct disorders. This intrigues our interest in finding the causal variant and understanding the mechanism underlying the three MPNs.

In 2009, few germline polymorphisms and haplotypes have been identified to be associated with the pathogenesis of MPNs. Olcaydu et al. (2009a) hypothesised that a certain SNP combination may contribute to different haplotypes that are susceptible to somatic mutagenesis such as V617F mutation. These SNPs were speculated to predispose MPN patients to V617F mutation and were preferentially acquired in cis with the V617F allele. This was supported by a strong association found between the *JAK2* mutated malignant clone and a particular haplotype defined by nearby single nucleotide polymorphisms (SNPs) in patients that were V617F-positive (Kilpivaara et al., 2009; Olcaydu et al., 2009a; Jones et al., 2009; Goldin et al., 2009). However, research is still ongoing to understand the interaction between this single point mutation (single disease allele) and 3 phenotypically distinct MPNs, with the in vivo evidence of V617F-induced haematopoiesis.

When this project started in 2009, there were only a few genetic association studies, all in Caucasian population (Pardanani et al., 2008; Jones et al., 2009; Kilpivaara et al., 2009; Olcaydu et al., 2009a). To date, most of the studies on the relationship between the germline polymorphisms and somatic mutations in the *JAK2* locus (predisposition alleles study) were still done in Caucasian population, only a few in the Chinese population (Hsiao et al., 2011; Hu et al., 2011; Wang et al., 2013) reporting the same association. As a consequence, little information is available regarding the disease in Chinese population as well as other ethnic groups. Therefore, it is important to understand the correlation among/interrelationship between these predisposition alleles or V617F with the *JAK2* haplotype(s) in other populations to understand the function underlying this mutation.

4.1.2 Research aim

The major aim of this part of my study was to investigate the association of the reported *JAK2* polymorphism in Hong Kong Chinese population. Based on this idea, the following parts were investigated:

1. To identify whether this polymorphism predisposes to *JAK2* mutation in MPN is restricted to Caucasian population or a general phenomenon in other populations.

2. To analyse the possible association of *other* variants spanning the *JAK2* gene including, but not limited to, the reported haplotype in the Chinese population of Hong Kong.

4.2 Methodology

4.2.1 Patient samples

To avoid the complication from loss of heterozygosity resulting from somatic isodisomy (UPD) in clonal myeloid cells, DNA from patients' saliva samples was used for SNP genotyping (Ng et al., 2006). Saliva samples from cases were collected using the Oragene DNA self-collection kit (DNA Genotek) according to the manufacturer's instructions (Section 2.2.4) and quantified by NanoDrop ND1000 UV/Vis spectrophotometer (Thermo Scientific, Wilmington, USA).

4.2.2 Control samples

The same peripheral blood DNA samples from controls used for V617F mutation detection were used for SNP genotyping. Controls used in this study were not perfectly paired with cases; however they were matched for sex and age as much as possible.

4.2.3 SNP selection and genotyping

In this study, I attempted to identify *JAK2* germline variants that are associated with the development of MPNs in local Hong Kong Chinese population, in addition to the reported *JAK2* risk haplotypes (rs10974944, rs12343867 and rs12340895, i.e. S9, S12 and S13). For ease of discussion, the SNPs were called S1, S2, …., and so on in the sequential order from the 5' end to the 3' end of the *JAK2* sense strand.

Selection of tSNPs was done as described in Section 2.2.5. The tSNPs were selected based on the genotype data from HapMap CHB database (release #24/phase II) (http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap24_B36/) using the Tagger software (de Bakker et al., 2005). *JAK2* region of 148.7kb including its potential regulatory regions was subjected to tSNP pick with parameters of minor allele frequency (MAF) ≥0. 1 and pairwise tagging algorithm, $r^2 \geq 0.8$. With evidence suggesting *JAK2* haplotype to be a major risk factor for MPN development, I forced-included the risk-haplotype tSNPs (rs10974944, rs12343867, and rs12340895) (Jones et al., 2009; Olcaydu et al., 2009a; Kilpivaara et al., 2009; Trifa et al., 2010).

Two methods were used for genotyping the SNPs depending on the logistic arrangement for instrument use and the cost (Section 2.2.10): 14 SNPs were genotyped by RFLP and 5 SNPs were genotyped by unlabelled probe melting analysis (Zhou et al., 2004; Yiu et al., 2013). Primers were designed using OLIGO version 6.62 software (Molecular Biology Insights, Inc., USA). Details of primer sequences and reaction conditions are given in **Table 2.2** and **Table 2.3**. PCR reactions were performed on the 96-well plates with a GeneAmp 9700 PCR system or Veriti® Thermal Cycler both from Applied Biosystems, (Foster City, CA, USA) following the conditions for each SNP (**Table 2.2-3)** with the in-house established protocol (Section 2.2.7). All reactions were subjected to optimisation and PCR products were sequenced to confirm the specially designed primers work correctly prior to real sample genotyping.

4.2.3.1 Genotyping by RFLP

Fourteen SNPs were genotyped by RFLP analysis (**Table 2.2**). Polymerase chain reaction (PCR) was performed as described in Section 2.2.7. Following amplification,

the PCR products were digested with specific restriction enzymes overnight according to enzymes' working temperatures. The digested RFLP-PCR products were pre-stained with SYBR Green I and visualised on polyacrylamide gels of appropriate concentration to confirm the size of the digested fragments.

4.2.3.2 Genotyping by unlabelled probe melting curve analysis

The remaining five SNPs were genotyped by unlabelled probe melting analysis using the saturating dsDNA dye (Invitrogen) as described in Section 2.2.10.2. The asymmetric PCR (Section 2.2.7.2) generated single-stranded DNA (ssDNA) product with a 3' end blocked unlabelled probe (**Table 2.3**). After PCR, the probe and SYTO 9 green fluorescent nucleic acid stain were added to the ssDNA target for high-resolution melting analysis. The reaction was performed in 96-well white plates, and subjected to melting in LightCycler® 480 Real-time PCR System (Roche).

4.2.3.3 Imputation of genotypes for 76 *JAK2* SNPs

Genotypes of 76 additional SNPs within the 148.7-kb region under study were imputed by Beagle version 3.2 (Browning & Browning, 2009). One of the imputed SNP rs4495487, was recently reported to contribute to MPN development in the Japanese population (Ohyashiki et al., 2012). The genotype data of the 1000 Genomes Project (phase 1) based on 97 CHB subjects were used as the reference panel. I manually conducted a quality control check by removing some of the known genotypes of the 19 directly genotyped SNPs, and imputed them with Beagle version 3.2. The post-imputation results were merged with the original data to check for the imputation accuracy.

4.2.5 Statistical analysis

Statistical analyses were conducted using the computer package PLINK (version 1.07; http://pngu.mgh.harvard.edu/,purcell/plink/index.shtml) (Purcell et al., 2007). Testing for deviation from Hardy-Weinberg equilibrium (HWE) was performed using Fisher's exact test with PLINK (version 1.07; http://pngu.mgh.harvard.edu/,purcell/plink/index.shtml) (Purcell et al., 2007) prior to data analysis. Single-marker and haplotype analyses were conducted between cases and controls with logistic regression adjusted for sex and age (age at diagnosis for MPN patients) as covariates; the respective asymptotic *P* value was denoted as P_{asym}. Correction for multiple comparisons was achieved by generating empirical *P* values (P_{emp}) after 50,000 permutations. Haplotypes were defined by variable-sized sliding-window approach. This approach considers all possible sizes of SNPs from the input and thus shows no bias towards either single SNP or haplotype. Subsequently, I studied the contribution of individual SNPs to significant haplotype association with disease by conditional logistic regression analysis. Haploview version 4.2 (Barrett et al., 2005) was used to generate the linkage disequilibrium (LD) map of the *JAK2* gene based on an algorithm called solid spine of linkage disequilibrium (SSLD) (Barrett et al., 2005).

The potential impact of these genetic markers was predicted using several web-based SNP prediction tools as listed in **Table 2.4,** such as SNPnexus (http://snpnexus.org/index.html) (Dayem Ullah et al., 2012), SNP Function Prediction (FuncPred**)** **(**http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm) (Xu & Taylor, 2009), F-SNP (http://compbio.cs.queensu.ca/F-SNP/) (Lee & Shatkay, 2008), and MatInspector (www.genomatix.de/matinspector.html) (Cartharius et al., 2005). These databases integrate bioinformatics tools and are updated to be synchronised with data-

bases to predict SNPs that may have biological function from regulatory potential to effect on protein function.

To identify regulatory potential of these SNPs and to link them to any specific genes that may be causal in the pathogenesis of MPNs, analysis of expression quantitative trait loci (eQTL) was conducted across the *JAK2* gene (142.8 kb) with several eQTL online tools as listed in **Table 2.4** including eQTL resources @ the Pritchard lab (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/) (Pritchard), seeQTL (http://www.bios.unc.edu/research/genomic_software/seeQTL/) (Xia et al., 2012), and UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgTracks?org=human) (Kent et al., 2002).

Particularly interested in transcription factor binding sites, we inspected the 40 bases encompassing S8 (both major and minor alleles were considered) in the centre using the program MatInspector (www.genomatix.de/matinspector.html) (Cartharius et al., 2005). This software locates matches in the input DNA sequences for putative transcription factor binding sites. The parameters were set to include only vertebrate as the source for transcription factor from both strands of the DNA. For the input, optimised matrix similarity was selected as a threshold to minimize false positive while for the output, core similarity and matrix similarity below 0.80 were excluded.

4.3 Results

In total, 19 tSNPs were selected, capturing the genetic information of 95 SNPs in the study region (148.7 kb) with a mean r^2 of 0.96. All of the SNPs are intronic except

rs3808850 (5' upstream). As explained in the section of Methodology, *JAK2* risk-haplotype tSNPs (rs10974944, rs12343867 and rs12340895, i.e. S9, S12 and S13) were forced to be included.

The genotype distribution of all SNPs in the control group was in HWE (Fisher's exact test greater than default *P* value of 0.001). In general, LD among the 19 SNPs in the combined group of V617F-positive MPN cases and healthy controls was not strong except for those tagging for the *JAK2* risk-haplotype (**Figure 4.1**). The same applied to the LD measures (r^2) for the combined group of V617F-negative MPN cases and healthy controls (**Figure 4.2**).

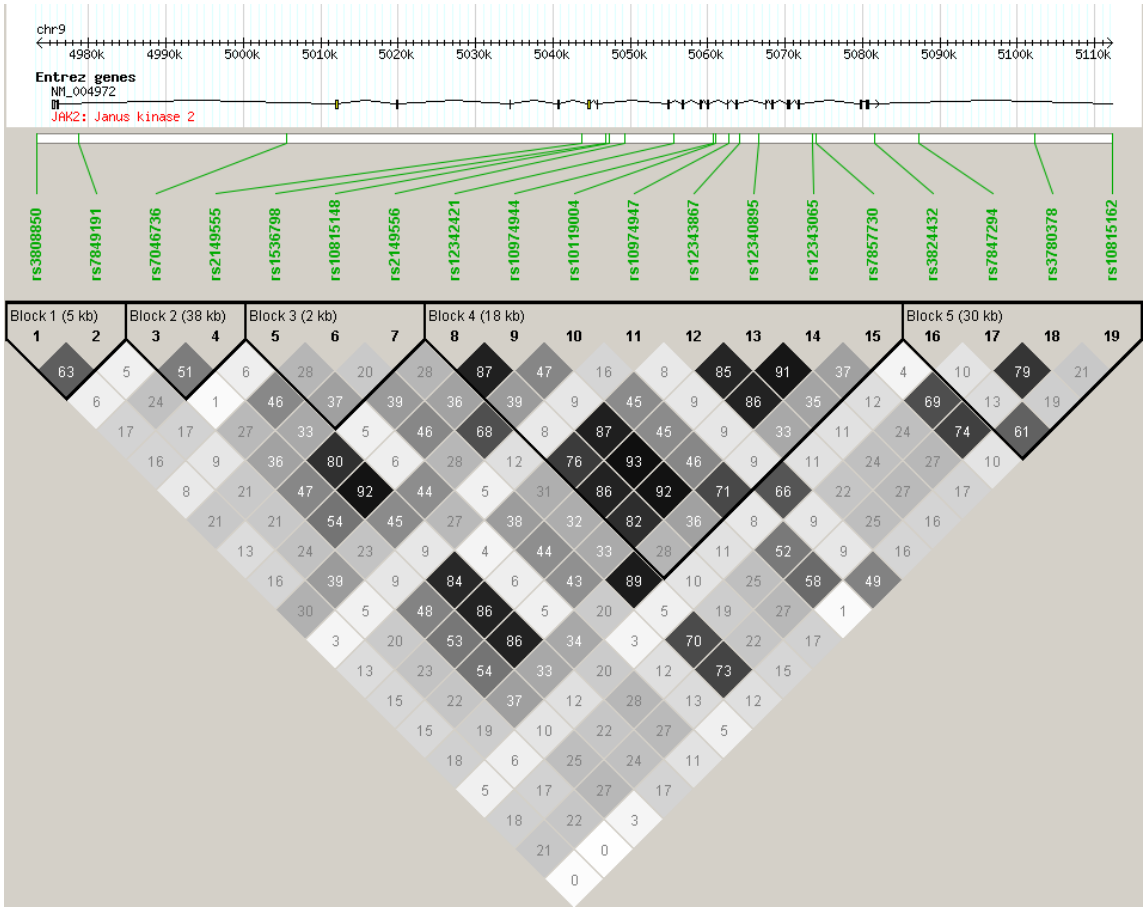


Figure 4.1 Linkage disequilibrium pattern for 19 *JAK2* SNPs for V617F-positive MPN cases and healthy controls.

Linkage disequilibrium plots were generated utilising the Haploview software. The values in the box indicate r^2 between the pairs. Haplotype blocks were defined by solid spine of linkage disequilibrium (SSLD).

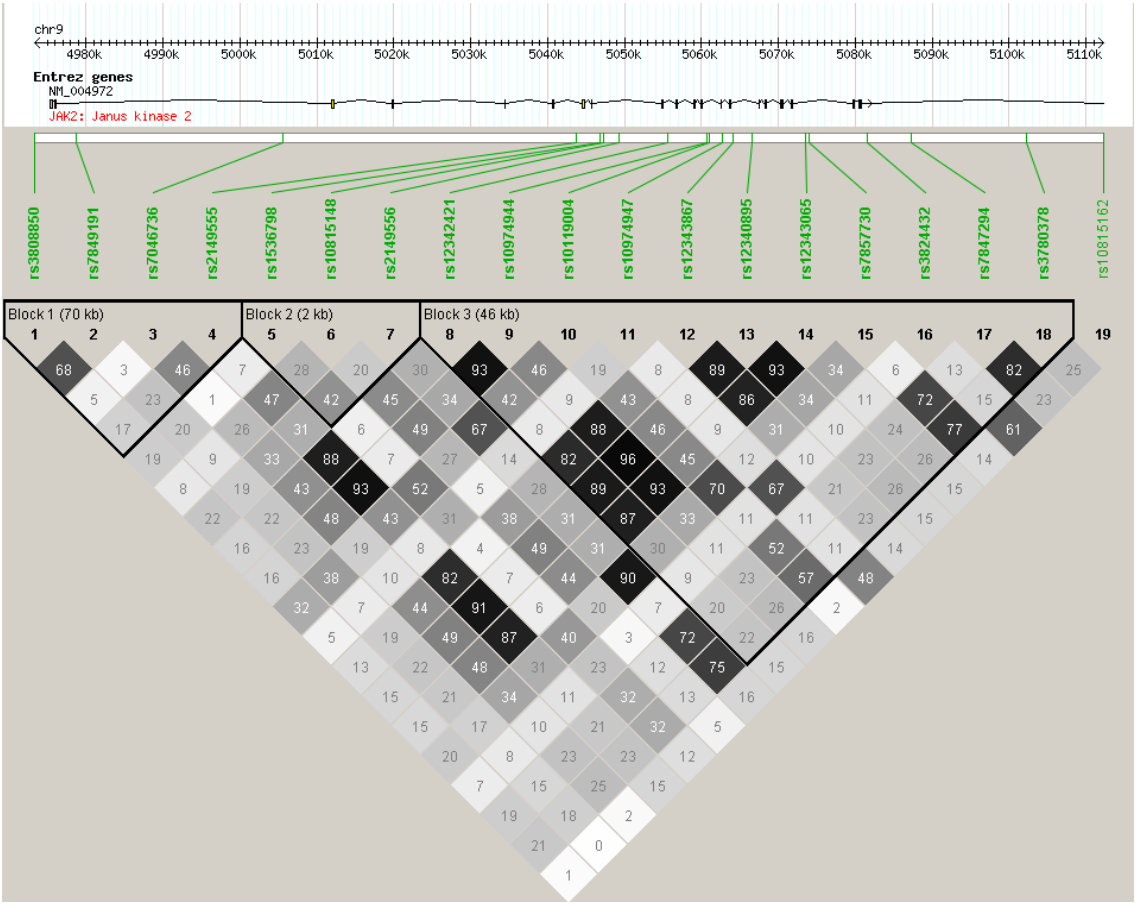


Figure 4.2 Linkage disequilibrium pattern for 19 *JAK2* SNPs for V617F-negative MPN cases and healthy controls.

Linkage disequilibrium plots were generated utilising the Haploview software. The values in the box indicate r^2 between the pairs. Haplotype blocks were defined by solid spine of linkage disequilibrium (SSLD).

4.3.1 Single marker analysis

As mentioned in Section 3.3.2, age did not differ between V617F-negative MPNs and controls (P=0.7342). However, when being analysed separately, there was still a significant difference between the patient and the control groups in terms of age (P<0.0001). Therefore, age was adjusted by logistic regression to minimise its influence in the analysis. To be consistent with previous studies, gender was also adjusted in the analyses although the proportions of males and females did not differ significantly between cases and controls. Among the five genetic models tested (genotypic, additive, allelic, dominant and recessive) for the 19 directly genotyped SNPs, data from the allelic model gave the most significant results. Therefore, I increased the stringency of allelic test by comparing the 19 SNPs between V617F-positive MPNs and controls with covariate adjustment for sex and age, and correction for multiple comparisons by 50,000 permutations. All of the 19 SNPs were associated with V617F-positive MPNs before permutation except rs1536798 (S5; P_{asym}=0.0765) and rs10974947 (S11; P_{asym}=0.1414) while 4 other SNPs did not survive after 50,000 permutations with P_{emp}> 0.05 (**Table 4.1**). Similar findings to those reported in European populations were detected: the minor alleles of the *JAK2* risk-haplotype tSNPs (allele G of rs12340895 (S13), allele G of rs10974944 (S9) and allele C of rs12343867 (S12)) were strongly associated with V617F-positive MPNs with a descending order of odds ratios (ORs; 3.27, 2.87, and 2.60, respectively, with $P_{asym}\leq 3.80\times10^{-9}$). All 3 SNPs statistically survived the 50,000 permutations with $P_{emp}=2.00\times10^{-5}$; note that the lowest P_{emp} value achievable with 50,000 permutations is 2×10^{-5}. These results suggested that S9, S12, and S13 were strongly associated with V617F-positive MPNs. *JAK2* SNP rs12342421 (S8) was identified as the most significantly associated SNP ($P_{asym}=3.76\times10^{-15}$ and $P_{emp}=2.00\times10^{-5}$) among the 19

candidate SNPs in Hong Kong Chinese population. The corresponding OR for the minor allele C was 3.55 with 95% CI ranging from 2.59 to 4.87.

Following the allelic association detected in the group of V617F-positive MPNs and healthy controls, V617F-negative MPN patients were also tested for the same 19 SNPs. Overall, comparison of V617F-negative MPNs and controls generated *insignificant* disease risk ($0.01<P_{asym}<0.46$) with ORs ranging from 0.63 to 1.78 (**Table 4.2**). None of the SNPs survived after 50,000 permutations with rs12342421 (S8) still being the strongest SNP ($P_{emp}=0.0832$). Likewise, haplotype analysis of V617F-negative MPNs yielded no significance ($P_{emp}\geq 0.2298$; data not shown). Interestingly, a comparison of the SNP allele frequencies between V617F-positive and V617F-negative patients also did not reveal any significant difference except for rs12342421 (S8; $P_{asym}=0.0031$ and $P_{emp}=0.0303$) and rs12340895 (S13; $P_{asym}=0.0075$ and $P_{emp}=0.0380$).

Table 4.1 Allelic association tests for 19 genotyped tag SNPs of the *JAK2* gene in V617F-positive MPNs

		Alleles [b]		Genotype Counts (11/12/22)[c]		Minor allele (1) freq.			Allelic Test [d]	
SNP rs [a]		1	2	Cases	Controls	Cases	Controls	OR (95% CI) [c]	P_{asym}	P_{emp}
rs3808850	(S1)	T	A	10/53/65	66/230/174	0.2852	0.3851	0.61 (0.45-0.84)	0.0022	0.0213
rs7849191	(S2)	T	C	8/40/80	26/233/170	0.2188	0.3904	0.43 (0.30-0.60)	6.35×10^{-7}	4.00×10^{-5}
rs7046736	(S3)	A	C	40/69/19	65/225/180	0.5820	0.3777	2.53 (1.85-3.46)	5.92×10^{-9}	2.00×10^{-5}
rs2149555	(S4)	T	C	20/85/23	41/194/235	0.4883	0.2936	2.51 (1.82-3.48)	2.01×10^{-8}	2.00×10^{-5}
rs1536798	(S5)	A	C	28/57/43	59/228/183	0.4414	0.3681	1.30 (0.97-1.74)	0.0765	0.4500
rs10815148	(S6)	A	T	9/59/60	20/167/283	0.3008	0.2202	1.60 (1.15-2.23)	0.0057	0.0509
rs2149556	(S7)	C	T	8/63/57	86/245/139	0.3086	0.4436	0.51 (0.37-0.71)	5.23×10^{-5}	0.0005
rs12342421	(S8)	C	G	52/54/22	43/197/230	0.6172	0.3011	3.55 (2.59-4.87)	3.76×10^{-15}	2.00×10^{-5}
rs10974944	(S9)	G	C	29/76/23	40/198/232	0.5234	0.2957	2.87 (2.08-3.96)	1.50×10^{-10}	2.00×10^{-5}
rs10119004	(S10)	G	A	10/71/47	121/248/101	0.3555	0.5213	0.46 (0.33-0.63)	1.65×10^{-6}	4.00×10^{-5}
rs10974947	(S11)	A	G	1/31/96	15/129/326	0.1289	0.1691	0.73 (0.48-1.11)	0.1414	0.6676
rs12343867	(S12)	C	T	22/80/23	39/186/245	0.4844	0.2809	2.60 (1.89-3.58)	3.80×10^{-9}	2.00×10^{-5}
rs12340895	(S13)	G	C	40/65/23	41/200/229	0.5664	0.3000	3.27 (2.37-4.51)	4.68×10^{-13}	2.00×10^{-5}
rs12343065	(S14)	T	C	28/77/23	41/201/228	0.5195	0.3011	2.80 (2.03-3.87)	3.80×10^{-10}	2.00×10^{-5}
rs7857730	(S15)	G	T	10/61/57	89/245/136	0.3164	0.4500	0.53 (0.38-0.73)	0.0001	0.0012
rs3824432	(S16)	A	G	1/37/90	26/148/296	0.1523	0.2128	0.67 (0.45-0.98)	0.0382	0.2711
rs7847294	(S17)	A	C	2/55/71	63/240/167	0.2305	0.3894	0.39 (0.27-0.56)	3.74×10^{-7}	2.00×10^{-5}
rs3780378	(S18)	C	T	8/58/62	84/239/147	0.2891	0.4330	0.49 (0.36-0.68)	2.25×10^{-5}	0.0002
rs10815162	(S19)	C	G	2/43/83	40/182/248	0.1836	0.2787	0.59 (0.41-0.84)	0.0037	0.0336

Abbreviations: SNP, single nucleotide polymorphism; OR, odds ratio; P_{asym}, asymptotic *P* value; P_{emp}, empirical *P* value.

[a] The SNPs are listed in sequential order from the 5' end to the 3' end of the sense strand of the *JAK2* gene. They are also designated S1 to S19 for the sake of easy reference and discussion.

[b] Alleles 1 and 2 represent the minor and major alleles of that SNP respectively. There are 128 cases and 470 controls.

[c] Calculated for minor allele (allele 1) with major allele (allele 2) as the reference allele.

[d] Allele frequencies were compared by logistic regression with adjustment for sex and age to give the P_{asym} value. Multiple comparisons were corrected by 50,000 permutations to give the P_{emp} value.

Table 4.2 Allelic association tests for 19 genotyped tag SNPs of the *JAK2* gene in V617F-negative MPNs

SNP rs [a]		Alleles [b] 1	2	Genotype Counts (11/12/22)[c] Cases	Controls	Minor allele (1) freq. Cases	Controls	OR (95% CI) [c]	Allelic Test [d] P_{asym}	P_{emp}
rs3808850	(S1)	T	A	3/21/20	66/230/174	0.3068	0.3851	0.68 (0.42-1.11)	0.1253	0.6068
rs7849191	(S2)	T	C	3/20/21	67/233/170	0.2955	0.3904	0.63 (0.39-1.04)	0.0674	0.4081
rs7046736	(S3)	A	C	6/31/7	65/225/180	0.4886	0.3777	1.60 (1.01-2.53)	0.0434	0.2935
rs2149555	(S4)	T	C	5/27/12	41/194/235	0.4205	0.2936	1.75 (1.11-2.75)	0.0156	0.1267
rs1536798	(S5)	A	C	5/27/12	59/228/183	0.4205	0.3681	1.27 (0.80-2.01)	0.2980	0.9225
rs10815148	(S6)	A	T	0/22/22	20/167/283	0.2500	0.2202	1.17 (0.69-1.98)	0.5666	0.9981
rs2149556	(S7)	C	T	2/26/16	86/245/139	0.3409	0.4436	0.63 (0.39-1.01)	0.0573	0.3565
rs12342421	(S8)	C	G	7/25/12	43/197/230	0.4432	0.3011	1.85 (1.18-2.90)	0.0078	0.0621
rs10974944	(S9)	G	C	5/27/12	40/198/232	0.4205	0.2957	1.74 (1.10-2.76)	0.0181	0.1328
rs10119004	(S10)	G	A	13/26/5	101/248/121	0.5909	0.4787	1.62 (1.02-2.57)	0.0441	0.2812
rs10974947	(S11)	A	G	0/15/29	15/129/326	0.1705	0.1691	1.02 (0.57-1.83)	0.9061	1.0000
rs12343867	(S12)	C	T	4/27/13	39/186/245	0.3977	0.2809	1.69 (1.07-2.66)	0.0246	0.1822
rs12340895	(S13)	G	C	4/29/11	41/200/229	0.4205	0.3000	1.72 (1.09-2.74)	0.0191	0.1560
rs12343065	(S14)	T	C	5/27/12	41/201/228	0.4205	0.3011	1.71 (1.08-2.70)	0.0222	0.1737
rs7857730	(S15)	G	T	1/29/14	89/245/136	0.3523	0.4500	0.64 (0.40-1.03)	0.0730	0.4060
rs3824432	(S16)	A	G	0/15/29	26/148/296	0.1705	0.2128	0.78 (0.44-1.38)	0.4091	0.9731
rs7847294	(S17)	A	C	1/26/17	63/240/167	0.3182	0.3894	0.70 (0.43-1.15)	0.1598	0.7130
rs3780378	(S18)	C	T	2/27/15	84/239/147	0.3523	0.4330	0.69 (0.43-1.11)	0.1190	0.6211
rs10815162	(S19)	C	G	1/18/25	40/182/248	0.2273	0.2787	0.78 (0.47-1.31)	0.3464	0.9494

Abbreviation: SNP, single nucleotide polymorphism; OR, odds ratio; P_{asym}, asymptotic *P* value; P_{emp}, empirical *P* value.

[a] The SNPs are listed in sequential order from the 5' end to the 3' end of the sense strand of the *JAK2* gene. They are also designated S1 to S19 for the sake of easy reference and discussion.

[b] Alleles 1 and 2 represent the minor and major alleles of that SNP respectively. There are 44 cases and 470 controls.

[c] Calculated for minor allele (allele 1) with major allele (allele 2) as the reference allele.

[d] Allele frequencies were compared between cases and controls by logistic regression adjusted for sex and age to give the P_{asym} value. Multiple comparisons were corrected by 50,000 permutations to give the P_{emp} value.

4.3.2 Haplotype analysis

Haplotype analysis was carried out to compare V617F-positive MPNs and controls with adjustment for sex and age. Exhaustive variable-sized sliding-window haplotype analysis was done on the 19 genotyped SNPs using PLINK (Purcell et al., 2007). One hundred ninety windows were examined with 1 to 19 SNPs per window. PLINK identified 184 haplotype windows (96.8%) showing significant differences (P_{emp}<0.05) in frequencies between patients and controls even after 50,000 permutations (**Table 4.3**). Third column from the right of **Table 4.3** showed haplotype window with the most significant omnibus test among all the sliding haplotype windows of a given size. We examined such most significant haplotype windows for all possible window sizes, and noted that all these most significant haplotype windows *always* included rs12342421 (S8) as a constituent SNP. Above all, the 1-SNP window rs12342421 (S8) itself achieved the strongest association with V617F-positive MPNs (P_{asym}=3.76×10^{-15} and P_{emp}=2.00×10^{-5}) (top row in **Table 4.3**). These results were similar to the haplotype blocks generated from Haploview (**Figure 4.1**).

According to data obtained from the 1000 Genomes Project, SNP rs12342421 (S8) is in perfect LD (**Figure 4.3**) with *JAK2* risk-haplotype tSNPs (rs10974944, rs12343867 and rs12340895, i.e. S9, S12 and S13) for the CHB population, and in very strong LD (r^2≥0.94; **Figure 4.3B**) with these three SNPs in the European (CEU) population. The LD was moderately strong (r^2≥0.75; **Figure 4.1**) for the corresponding pairs of SNPs in my study cohort. However, on the basis of SSLD, rs12342421 was not in the same LD block with *JAK2* risk-haplotype tSNPs in the CEU population (**Figure 4.3B**). Overall, a higher degree of correlation was observed among these

few SNP pairs in the 1000 Genomes Project data of CHB and CEU populations (**Figure 4.3A**) when compared with study cohort in this project.

Table 4.3 Exhaustive haplotype analyses for variable-sized sliding windows across 19 genotyped *JAK2* SNPs for V617F-positive MPNs

		SW with Omnibus Test $P_{emp} < .05$			Most Significant Omnibus Test		
SNPs, No.	SWs, No.	SWs, No.	First SW	Last SW	SW	P_{asym}	P_{emp}
1	19	14 [a]	S1	S18	S8 [c]	3.76×10^{-15}	2.00×10^{-5}
2	18	17 [b]	S1…S2	S18…S19	S8…S9	2.13×10^{-14}	2.00×10^{-5}
3	17	17	S1…S3	S17…S19	S8…S10	6.33×10^{-14}	2.00×10^{-5}
4	16	16	S1…S4	S16…S19	S7…S10	8.25×10^{-13}	2.00×10^{-5}
5	15	15	S1…S5	S15…S19	S8…S12	4.45×10^{-12}	2.00×10^{-5}
6	14	14	S1…S6	S14…S19	S8…S13	2.75×10^{-12}	2.00×10^{-5}
7	13	13	S1…S7	S13…S19	S8…S14	2.21×10^{-12}	2.00×10^{-5}
8	12	12	S1…S8	S12…S19	S8…S15	9.00×10^{-12}	2.00×10^{-5}
9	11	11	S1…S9	S11…S19	S6…S14	2.38×10^{-11}	2.00×10^{-5}
10	10	10	S1…S10	S10…S19	S6…S15	9.09×10^{-12}	2.00×10^{-5}
11	9	9	S1…S11	S9…S19	S6…S16	3.03×10^{-11}	2.00×10^{-5}
12	8	8	S1…S12	S8…S19	S6…S17	6.69×10^{-11}	2.00×10^{-5}
13	7	7	S1…S13	S7…S19	S6…S18	1.21×10^{-10}	2.00×10^{-5}
14	6	6	S1…S14	S6…S19	S6…S19	2.60×10^{-10}	2.00×10^{-5}
15	5	5	S1…S15	S5…S19	S4…S18	3.02×10^{-9}	2.00×10^{-5}
16	4	4	S1…S16	S4…S19	S4…S19	1.97×10^{-9}	2.00×10^{-5}
17	3	3	S1…S17	S3…S19	S3…S19	2.74×10^{-9}	2.00×10^{-5}
18	2	2	S1…S18	S2…S19	S2…S19	2.96×10^{-8}	2.00×10^{-5}
19	1	1	S1…S19	S1…S19	S1…S19	6.72×10^{-8}	2.00×10^{-5}

Abbreviations: SNP, single nucleotide polymorphism; SW, sliding window; P_{asym}, asymptotic *P* value; P_{emp}, empirical *P* value.

The SW is shown as Sx…Sy, where Sx is the first SNP and Sy is the last SNP of the SW for *JAK2* gene. Please refer to Table 1 for the identity of the SNP concerned. Each sliding window was tested by an omnibus test adjusted for sex and age (implemented in PLINK). Multiple comparisons were corrected by running 50,000 permutations to give the P_{emp} value. The smallest P_{emp} value generated after permutation is the same for all fixed-size SWs (2×10^{-5}); note that the lowest P_{emp} value achievable with 50,000 permutations is 2×10^{-5}. The most significant results for each fixed-size SW is shown in the three rightmost columns. Note that, among all the 190 SWs tested, S8 always appears in the most significant SW.

[a] Of the nineteen SNPs tested, five (S5, S6, S11, S16, and S19) did not give $P_{emp} < 0.05$.

[b] All the SWs gives $P_{emp} < 0.05$ except S5…S6.

[c] Of all the 190 SWs tested, S8 (i.e. rs12342421) alone gives the most significant result for association with *V617F*-positive MPNs.

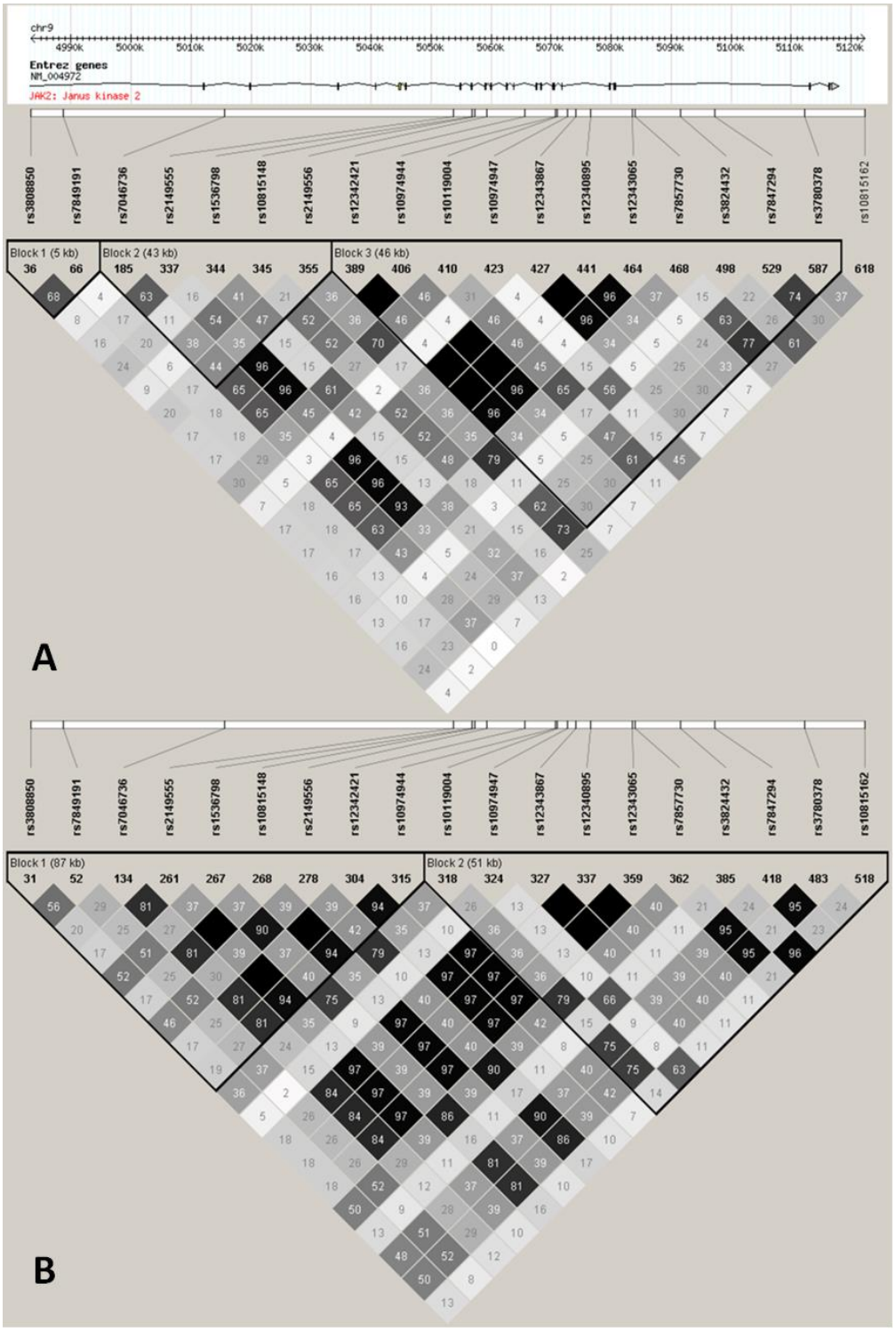


Figure 4.3 Haploview-generated linkage disequilibrium (LD) map of 19 *JAK2* SNPs in Han Chinese in Beijing (A) and Caucasians of European ancestry (B) based on the 1000 Genomes Project data.

LD plots were generated utilising the Haploview software. The values in the boxes indicate the r2 values between the respective pairs of SNPs and the empty boxes represent those with r2=1.0. Haplotype blocks are defined by solid spine of linkage disequilibrium(Barrett et al., 2005).

4.3.3 Conditional logistic regression

Based on the results from PLINK, I tested the individual effect of the strongest MPN-associated SNP (rs12342421, i.e. S8) and the risk-haplotype tSNPs (rs10974944, rs12343867 and rs12340895, i.e. S9, S12 and S13) in the corresponding sliding window. The shortest and most significant sliding haplotype window containing these four SNPs was the 6-SNP window S8…S13 (P_{asym}=2.75×10^{-12}; **Table 4.3**), which was therefore selected for conditional logistic regression analysis. Conditional analysis testing for independent effect of one SNP at a time suggested that *only* rs12342421 (S8) contributed independently to the significant association between the 6-SNP window and *V617F*-positive MPNs (*P*=0.0005 for omnibus test of independent effect, **Table 4.4**). Logically, controlling for all the single SNPs except rs12342421 (S8) yielded a reduced but still statistically significant *P* value of ≤0.0072) while controlling for rs12342421 (S8) demolished the significance (*P*=0.4360) (**Table 4.4**). This means that when rs12342421 (S8) was combined with other SNPs, its effect become less significant.

Our data suggested that *JAK2* germline polymorphisms, especially rs12342421 (S8), are very significantly associated with *V617F*-positive MPNs in Hong Kong Chinese population.

Table 4.4 Conditional haplotype-based test: independent effects of individual *JAK2* SNPs on the 6-SNP sliding window S8…S13 [a]

		Conditional haplotype-based association test, *P* value	
Sx [b]		Independent effect of Sx [c]	Controlling for Sx [d]
rs12342421	(S8)	0.0005	0.4360
rs10974944	(S9)	– [e]	0.0072
rs10119004	(S10)	0.4700	2.84×10^{-7}
rs10974947	(S11)	0.2480	1.79×10^{-14}
rs12343867	(S12)	0.7970	0.0019
rs12340895	(S13)	– [e]	0.0072

[a] This table shows the individual effects of the constituent single nucleotide polymorphisms (SNPs) on the shortest and most significant sliding window that contains the most impressive SNP in our study (rs12342421, i.e. S8) and the risk-haplotype tSNPs (rs10974944, rs12343867 and rs12340895, i.e. S9, S12 and S13). Conditional logistic regression was performed with adjustment for sex and age. The shortest and most significant sliding window carrying these four SNPs is S8…S13 (see Table 2). The conditional omnibus test invoked by the "--chap" command of PLINK gives a *P* value of 1.34×10^{-14} (based on likelihood ratio test).

Note that this *P* value is similar, but not identical, to the *P* value of 2.75×10^{-12} (based on Wald test, Table 2) generated by the omnibus test of logistic regression invoked by the "--logistic" command of PLINK in the sliding-window approach.

[b] Sx indicates the SNP tested for an independent effect one at a time by the conditional haplotype-based analysis of the sliding window S8…S13. Please refer to Table 1 for the identity of the SNPs concerned.

[c] Omnibus *P* value for the effect of Sx that is independent of the other SNPs in the sliding window S8…S13.

[d] Omnibus *P* value for the sliding window S8…S13 when Sx is controlled for.

[e] Not a valid comparison due to identical alternate and null models.

4.3.4 Genetic association of genotyped and imputed SNPs

With the significant findings achieved from the 19 SNPs, I further performed imputation with Beagle for 76 additional *JAK2* SNPs (selected from Tagger with MAF 0.01) to better examine the 148.7-kb region encompassing *JAK2* locus. Manual quality control check on Beagle indicated an accuracy of >95% in imputing the missing (removed) genotypes. Consistent trends were identified when all 95 SNPs (19 directly genotyped and 76 imputed) were analysed together by logistic regression adjusted for sex and age: single-marker analysis gave the strongest association signal for rs12342421 (S8) as in my initial study with only 19 SNPs. Of these 95 SNPs, 67 showed association exceeding the significance of 8×10^{-8} (P_{asym}). The strongest association was detected for rs12342421 (S8; $P_{asym}=3.76\times10^{-15}$, $P_{emp}=2.00\times10^{-5}$ and OR=3.55), while SNPs of high LD showed similar levels of association. **Table 4.5** lists the top 20 SNPs among the 95 *JAK2* SNPs analysed.

To have an overall picture, I examined the LD map (**Figure 4.4**) for all 95 SNPs (19 directly genotyped and 76 imputed). Data showed that rs12342421 (SNP no. 43 in **Figure 4.4**) also tagged ($r^2=0.83$) for rs4495487 (SNP no. 49 in **Figure 4.4**) that was reported to be the additional variant contributing to MPN predisposition in Japanese population (Ohyashiki et al., 2012). All the SNPs within this haplotype block showed very strong extent of LD (r^2 close to 1; bottom panel of **Figure 4.4**).

To narrow the disease-associated region and to identify MPN-predisposing variants or haplotypes unbiasedly, these 95 SNPs were examined in the same way as the 19 genotyped SNPs using the exhaustive haplotype approach with age and sex adjusted as covariates (**Table 4.6**). Of the 1634 haplotype windows formed among these 95

SNPs after multiple comparison correction by 50,000 permutations, SNP rs12342421 (S8) again survived as the most significantly associated 1-SNP window (S8 itself): $P_{asym}=3.76\times10^{-15}$ and $P_{emp}=2.00\times10^{-5}$ (**Table 4.6**). Similarly as in the sliding windows for the 19 directly genotyped SNPs, the remaining most significantly associated haplotypes cantered in rs12342421 (S8), with its adjacent SNPs. The SNP rs12342421 (S8) was obviously important because almost all the statistically significant haplotypes contained this SNP.

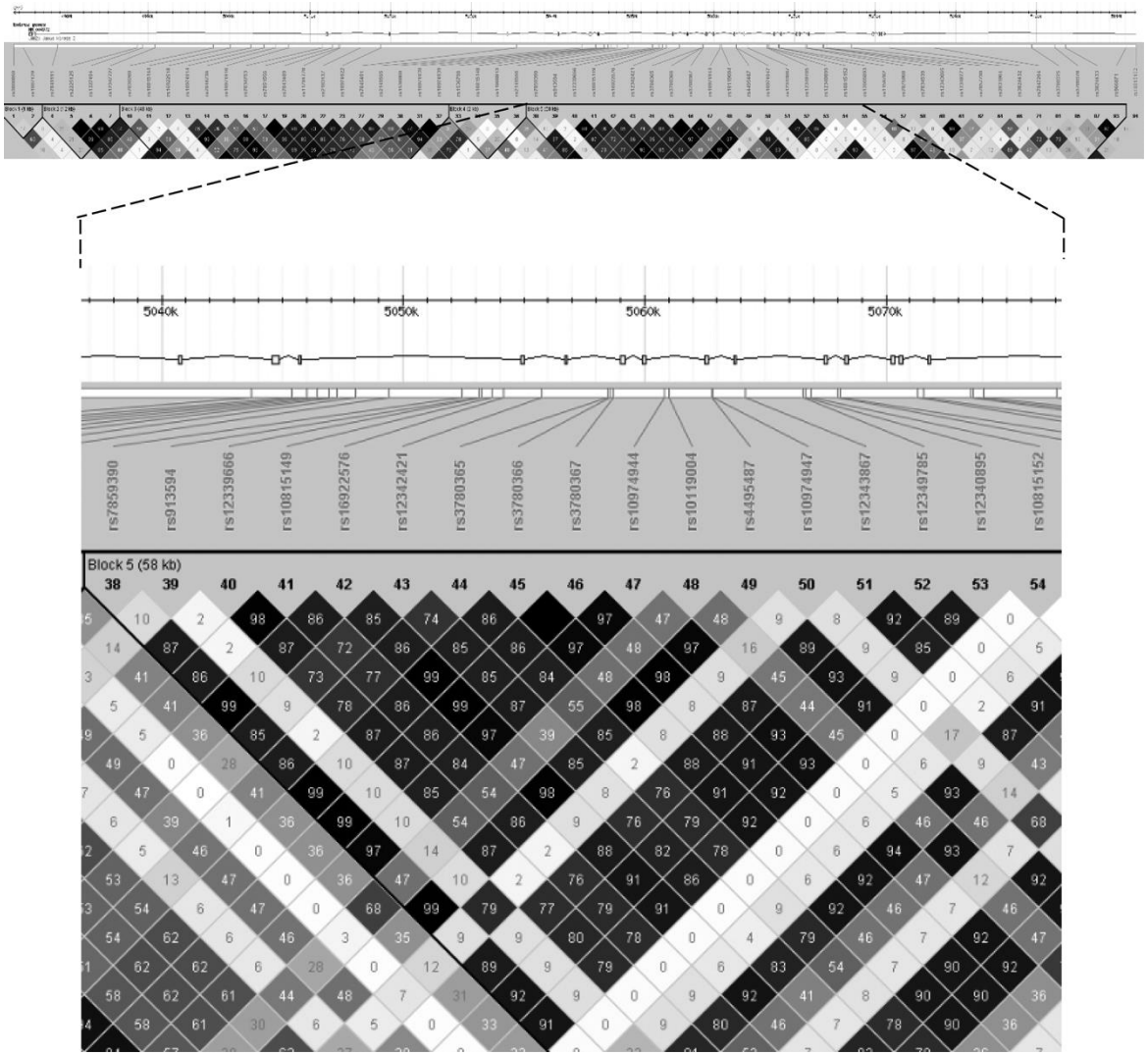


Figure 4.4 Linkage disequilibrium pattern for 95 *JAK2* SNPs for V617F-positive MPN cases and healthy controls.

Linkage disequilibrium plots were generated utilising the Haploview software. The values in the box indicate r^2 between the pairs and the empty boxes represent those with r^2=1.0. Haplotype blocks were defined by solid spine of linkage disequilibrium (SSLD).

Table 4.5 Logistic regression tests: Top 20 SNPs among 95 genotyped/imputed *JAK2* SNPs in *V617F*-positive MPNs

SNP [a]		Alleles [b]		Minor Allele Freq.			Allelic Test [d]	
		1	2	Cases	Controls	OR (95% CI) [c]	P_{asym}	P_{emp}
rs12342421	(S8) [e]	C	G	0.6172	0.3011	3.55 (2.59-4.87)	3.76×10^{-15}	2.00×10^{-5}
rs12347727		G	A	0.5508	0.2734	3.33 (2.43-4.56)	7.72×10^{-14}	2.00×10^{-5}
rs2225125		G	A	0.5508	0.2755	3.30 (2.41-4.53)	1.18×10^{-13}	2.00×10^{-5}
rs1327494		G	A	0.5508	0.2755	3.30 (2.41-4.53)	1.18×10^{-13}	2.00×10^{-5}
rs11794778		T	G	0.5508	0.2798	3.26 (2.38-4.47)	2.38×10^{-13}	2.00×10^{-5}
rs12340895	(S13) [e]	G	C	0.5664	0.3000	3.27 (2.37-4.51)	4.68×10^{-13}	2.00×10^{-5}
rs10974914		A	G	0.5547	0.3032	3.18 (2.30-4.40)	2.46×10^{-12}	2.00×10^{-5}
rs10974916		A	G	0.5547	0.3043	3.17 (2.29-4.38)	2.94×10^{-12}	2.00×10^{-5}
rs2183137		G	A	0.5508	0.2989	3.11 (2.26-4.28)	3.02×10^{-12}	2.00×10^{-5}
rs7851556		T	C	0.5547	0.3064	3.14 (2.27-4.34)	4.79×10^{-12}	2.00×10^{-5}
rs7043489		C	A	0.5547	0.3064	3.14 (2.27-4.34)	4.79×10^{-12}	2.00×10^{-5}
rs11794708		A	G	0.4258	0.1702	2.67 (2.02-3.53)	5.21×10^{-12}	2.00×10^{-5}
rs10974921		A	T	0.4258	0.1702	2.67 (2.02-3.53)	5.21×10^{-12}	2.00×10^{-5}
rs7030260		A	C	0.5547	0.3202	3.05 (2.20-4.23)	2.00×10^{-11}	2.00×10^{-5}
rs10974922		T	C	0.5547	0.3213	3.04 (2.20-4.22)	2.25×10^{-11}	2.00×10^{-5}
rs12349785		C	G	0.5195	0.2830	2.95 (2.14-4.06)	3.35×10^{-11}	2.00×10^{-5}
rs966871		T	A	0.5391	0.3021	2.86 (2.09-3.92)	5.80×10^{-11}	2.00×10^{-5}
rs3824433		T	C	0.5312	0.3000	2.88 (2.09-3.95)	7.73×10^{-11}	2.00×10^{-5}
rs1159782		C	T	0.5195	0.2936	2.89 (2.09-3.99)	1.16×10^{-10}	2.00×10^{-5}
rs10974944	(S9) [e]	G	C	0.5234	0.2957	2.87 (2.08-3.96)	1.50×10^{-10}	2.00×10^{-5}

Abbreviations: SNP, single nucleotide polymorphism; OR, odds ratio; P_{asym}, asymptotic *P* value; P_{emp}, empirical *P* value.

[a] The SNPs are listed in ascending order in terms of their P_{asym} among the top 20 most significantly associated *JAK2* SNPs in V617F-positive MPN patients. Association was tested by logistic regression with adjustment for sex and age.

[b] Alleles 1 and 2 represent the minor and major alleles of that SNP respectively. There are 128 cases and 470 controls.

[c] Calculated for minor allele (allele 1) with major allele (allele 2) as the reference allele.

[d] Allele frequencies were calculated by logistic regression with sex and age as covariates to give the P_{asym} value. Multiple comparisons were corrected by 50,000 permutations to give the P_{emp} value.

[e] These three SNPs (S8, S9 and S13) were directly genotyped in this study while the rest were imputed by Beagle v3.2 (Browning & Browning, 2009).

Table 4.6 Summary of exhaustive haplotype analyses based on age- and sex-adjusted omnibus tests for sliding windows of up to 19 SNPs per window for 95 genotyped/imputed JAK2 SNPs for V617F-positive MPNs a

		SW with Omnibus Test $P_{emp} < 0.05$			Most Significant Omnibus Test		
No.	No. of SWs	No. of SWs	First SW [b]	Last SW [b]	SW	P_{asym}	P_{emp}
1	95	74	m5	m95	m43* [c]	3.76×10^{-15}	2.00×10^{-5}
2	94	88	m3...m4*	m94*...m95	m42...m43*	8.76×10^{-15}	2.00×10^{-5}
3	93	92	m1*...m3	m93...m95	m41...m43*	1.77×10^{-14}	2.00×10^{-5}
4	92	91	m1*...m4*	m92...m95	m40...m43*	1.71×10^{-14}	2.00×10^{-5}
5	91	90	m1*...m5	m91...m95	m42...m46	5.55×10^{-14}	2.00×10^{-5}
6	90	89	m1*...m6	m90...m95	m38...m43*	7.19×10^{-14}	2.00×10^{-5}
7	89	88	m1*...m7	m89...m95	m37...m43*	7.20×10^{-14}	2.00×10^{-5}
8	88	86	m1*...m8	m87...m94*	m42...m49	2.09×10^{-14}	2.00×10^{-5}
9	87	86	m1*...m9	m87...m95	m37...m45	5.84×10^{-13}	2.00×10^{-5}
10	86	85	m1*...m10	m86...m95	m37...m46	5.84×10^{-13}	2.00×10^{-5}
11	85	84	m1*...m11	m85*...m95	m37...m47*	1.65×10^{-12}	2.00×10^{-5}
12	84	82	m1*...m12	m84...m95	m37...m48*	2.05×10^{-12}	2.00×10^{-5}
13	83	81	m1*...m13	m83...m95	m37...m49	2.05×10^{-12}	2.00×10^{-5}
14	82	81	m1*...m14*	m82...m95	m40...m53*	6.73×10^{-12}	2.00×10^{-5}
15	81	80	m1*...m15	m81...m95	m38...m52	9.20×10^{-12}	2.00×10^{-5}
16	80	79	m1*...m16	m80...m95	m38...m53*	7.44×10^{-12}	2.00×10^{-5}
17	79	78	m1*...m17	m79...m95	m37...m53*	1.01×10^{-11}	2.00×10^{-5}
18	78	77	m1*...m18	m78...m95	m38...m55	3.21×10^{-11}	2.00×10^{-5}
19	77	76	m1*...m19	m77...m95	m39...m57	1.37×10^{-11}	2.00×10^{-5}

Abbreviations: SNP, single nucleotide polymorphism; SW, sliding window; P_{asym}, asymptotic *P* value; P_{emp}, empirical *P* value.

[a] Note that this SNP identity is not the same as in **Table 4.1**. Please refer to **Table 4.7** (see below) for the identity of the SNPs concerned.

The 95 genotyped/imputed *JAK2* SNPs were only tested for up to 19-SNP per SW as we fixed the comparison similar with that from the 19 genotyped SNPs for *V617F*-positive MPNs. The SW is shown as mx…my, where mx is the first SNP and my is the last SNP of the SW for the *JAK2* gene. Multiple comparisons were corrected by running 50,000 permutations to give the P_{emp} value. The smallest P_{emp} value generated after permutation is the same for all fixed-size SWs (2×10^{-5}). The most significant results for each fixed-size SW is shown in the three rightmost columns. Note that, among all the 1634 SWs tested, m43 (see footnote c) always appears in the most significant SW.

[b] Within each set of SWs with a given size (SNPs per SW), there are 1-2 SNPs or SWs that are *not* significant (P_{emp}>0.05) and interrupt the consecutive significant SWs. Of particular note, there are 18 non-significant interrupting SWs for 1-SNP SW set, and 4 non-significant interrupting SWs for 2-SNP SW set.

[c] m43* is the same rs12342421 (S8). Of the 1634 SWs tested, m43* alone gives the most significant result for association with *V617F*-positive MPNs among 95 *JAK2* SNPs.

Table 4.7 SNP Identity of the 95 genotyped/ imputed *JAK2* SNPs

No.	SNP		bp	No.	SNP		bp	No.	SNP		bp
m1*	rs3808850	(S1)	4973311	m33*	rs1536798	(S5)	5046931	m65	rs2274649		5080934
m2	rs1887429		4974549	m34*	rs10815148	(S6)	5047284	m66*	rs3824432	(S16)	5081675
m3	rs2274471		4975879	m35	rs11998913		5048048	m67	rs12340866		5084185
m4*	rs7849191	(S2)	4978761	m36*	rs2149556	(S7)	5049440	m68	rs3780370		5085167
m5	rs2225125		4988639	m37	rs1571437		5050334	m69	rs10974960		5085842
m6	rs1327494		4989303	m38	rs7859390		5052473	m70	rs7847141		5087171
m7	rs12347727		4990811	m39	rs913594		5053199	m71*	rs7847294	(S17)	5087281
m8	rs4372063		4993338	m40	rs12339666		5053296	m72	rs3780372		5087544
m9	rs10115312		4993973	m41	rs10815149		5053701	m73	rs3780373		5088223
m10	rs7030260		4998070	m42	rs16922576		5054193	m74	rs10121077		5088411
m11	rs10815144		5000192	m43*	rs12342421	(S8)	5055750	m75	rs3780374		5089677
m12	rs16922518		5002696	m44	rs3780365		5058520	m76	rs7870694		5090628
m13	rs10974914		5004332	m45	rs3780366		5058596	m77	rs10974963		5091305
m14*	rs7046736	(S3)	5005732	m46	rs3780367		5058755	m78	rs2104685		5096023
m15	rs10974916		5007350	m47*	rs10974944	(S9)	5060831	m79	rs4593605		5097278
m16	rs7034753		5011514	m48*	rs10119004	(S10)	5061049	m80	rs10815157		5098771
m17	rs7851556		5012807	m49	rs4495487		5062798	m81	rs3780375		5099431
m18	rs11794708		5013441	m50*	rs10974947	(S11)	5062846	m82	rs11793659		5099707
m19	rs7043489		5013604	m51*	rs12343867	(S12)	5064189	m83	rs17425637		5100000
m20	rs11794778		5013794	m52	rs12349785		5066613	m84	rs3780377		5100899
m21	rs10974921		5014427	m53*	rs12340895	(S13)	5066691	m85*	rs3780378	(S18)	5102288
m22	rs2183137		5016293	m54	rs10815152		5066946	m86	rs3780379		5102519
m23	rs10974922		5018813	m55	rs12005893		5068046	m87	rs3824433		5103577
m24	rs7023146		5030163	m56	rs1159782		5068117	m88	rs884132		5104522
m25	rs7043371		5030203	m57	rs7875908		5071334	m89	rs3780381		5104523
m26	rs7037207		5033156	m58	rs7034539		5071585	m90	rs17425819		5104773
m27	rs7045491		5035658	m59	rs1410779		5073173	m91	rs10815160		5106616
m28	rs1328917		5039065	m60*	rs12343065	(S14)	5073533	m92	rs11788963		5110157
m29*	rs2149555	(S4)	5043743	m61	rs12348771		5073634	m93	rs966871		5111070
m30	rs1536800		5045434	m62*	rs7857730	(S15)	5074049	m94*	rs10815162	(S19)	5112291
m31	rs10974938		5046037	m63	rs6476939		5074837	m95	rs10974969		5115336
m32	rs10974939		5046482	m64	rs2031904		5077087				

bp = base position

4.4 Discussion

Back in 2009 when this project was just started, studies on the relationship between the germline polymorphisms and somatic mutations in the *JAK2* locus (predisposition alleles study) were only available in Caucasian population (Pardanani et al., 2008; Jones et al., 2009; Kilpivaara et al., 2009; Olcaydu et al., 2009a). Little information is available regarding the disease in Chinese population as well as in other ethnic groups. Therefore, I employed a case-control study design to explore the described genetic susceptibility to MPNs in the Hong Kong Chinese population.

To date, few genetic association studies of MPNs had been reported in Asian populations (Hsiao et al., 2011; Hu et al., 2011; Wang et al., 2013). In line with previous reports from Caucasian populations, data from Asian populations also pointed that *JAK2* haplotype poses a higher risk of developing V617F-positive MPNs (Hu et al., 2011; Zhang et al., 2012a). In this case-control study conducted in Hong Kong population, age differed significantly between cases (V617F-positive MNPs in particular) and controls, analyses were thus conducted with adjustment for age to avoid confounding issue. Despite similar sex ratios in cases and controls, we also adjusted for sex to be in line with other existing studies for easy comparison (Tefferi et al., 2010b; Zhang et al., 2012a). To avoid missing any potential causal variant in the region, we investigated not only the risk-haplotype-tagging SNPs but also a total of 95 SNPs in two stages with an increased sample size. In the first stage, 19 tSNPs of the *JAK2* locus were genotyped. By genotyping these 19 SNPs, the reported *JAK2* haplotype was also studied because the risk-haplotype tSNPs (S9, S12 and S13) were force-included in my study design. In the second stage, I carried out genotype imputation

on additional 76 *JAK2* SNPs (95 in total), and carefully examined both datasets by both single-marker analysis and haplotype analysis.

4.4.1 *JAK2* SNPs are associated with V617F-positive MPNs

The distribution of V617F mutation in Hong Kong MPN patients (PV, ET and PMF) is similar to those in other studies (Jones et al., 2005; Baxter et al., 2005; James et al., 2005b; Kralovics et al., 2005a). This justified that my findings were comparable to those in other populations. Taken together, results from this study corroborate the findings that *JAK2* variants are predisposing factors towards MPNs development dependent on V617F in Hong Kong Chinese, especially rs12342421 (S8). Conceivably, the failure to detect, in my study, the association between V617F-negative group and controls as reported elsewhere (Pardanani et al., 2010; Tefferi et al., 2010b) can be ascribed to the small sample size of the cases (n=44). Larger sample size would probably be needed to detect positive association for V617F-negative MPNs. Meta-analysis (part II of this study) can be a way to solve the problem of insufficient sample size. In addition to meta-analysis, international collaboration may have larger power to study and detect rare variants if any. This aids the identification of causal variants for genetic susceptibility to the development of MPNs.

4.4.2 V617F-associated MPNs are associated more significantly with single marker

In addition to single-marker analysis, haplotypic effects of the SNPs over *JAK2* locus were also examined. I adopted an unbiased manner: exhaustive variable-sized sliding-window strategy. This approach offered the best capture of either single markers

or haplotypes that are most significantly associated with MPNs, it detected the optimum marker combination of all possible sizes among the SNPs (Guo et al., 2009). This approach comprehensively examined all 190 haplotype windows for the 19 directly genotyped SNPs and identified 97% of all 190 haplotype windows (184 haplotype windows; P_{emp}<0.05, **Table 4.3**) to be significant associated with V617F-positive MPNs even after 50,000 permutations. However, single-marker analyses of both the 19 SNPs and the 76 imputed SNPs showed that, despite the strong association of the risk-haplotype tSNPs (rs10974944, rs12343867 and rs12340895, i.e., S9, S12 and S13) with V617F-positive MPNs (**Table 4.1**), the single SNP rs12342421 (S8, also tagging the risk haplotype), was found to be more significantly associated with MPNs than the haplotypes (**Table 4.1** vs **Table 4.3**, and **Table 4.5** vs **Table 4.6.** Conditional logistic regression analysis further demonstrated that this single SNP contributed an independent effect to the most significant association between haplotypes and MPNs. The C allele was enriched in V617F-positive MPN patients when compared with controls.

4.4.3 Relationship between single marker rs12342421 and MPNs

As mentioned in the results, the LD pattern indicates that our significant association with MPNs in Hong Kong Chinese is compatible with the results in other populations. However, based on SSLD, rs12342421 (S8) was not in the same LD block with *JAK2* risk-haplotype tSNPs in the CEU population (**Figure 4.3B**) although it is still in strong LD (r^2 close to 1) with *JAK2* risk-haplotype tSNPs. This may explain the stronger association observed in Hong Kong population for this single SNP rather than the *JAK2* risk-haplotype tSNPs.

Inevitably, differences observed between the LD patterns of rs12342421 (S8) and *JAK2* risk-haplotype tSNPs might be caused by errors in the phasing process. Intriguingly, the LD differences were mainly observed in the group of V617F-positive MPN patients, *but not* the control subjects. It has been shown that there can be extensive variation in the extent of LD between cases and controls in a region of genetic association. The variation in LD patterns observed in our cases (V617F-positive MPN patients) and controls suggests that the region surrounding rs12342421 (S8) is associated with V617F-positive MPNs – consistent with the main finding of this study.

In fact, S8 is in LD with both the original risk-haplotype tSNPs and also rs4495487, the MPN-associated SNP reported in Japanese population (Ohyashiki et al., 2012). However, rs4495487 did not survive the analysis as the top 20 SNPs among 95 genotyped/imputed *JAK2* SNPs (**Table 4.5**), thus it was excluded for further analysis. In fact, only rs12342421 (S8), rs12340895 (S13) and rs10974944 (S9) survived as the top 20 most significantly associated SNPs with *V617F*-positive MPNs in the *genotyped/imputed* data: S8 ranked the 1st, S13 the 6th and S9 the 20th (**Table 4.5**).

From the genotyped data, the shortest and most significant sliding haplotype window containing the original tSNPs (from the 2009 papers) were selected for further analysis. When conditional analysis for the independent effect of one SNP at a time was used, *only* rs12342421 (S8) contributed an independent effect to the significant association between the 6-SNP window and V617F-positive MPNs (P=0.0005, **Table 4.4**). When I controlled for rs12342421 (S8) from the window, the effect was demolished (*P*=0.4360) (**Table 4.4**). When I controlled for all the other SNPs, the effects

were still statistically significant (*P* value of ≤0.0072). This means that the original tSNPs from the 2009 papers (rs10974944, rs12343867 and rs12340895) did not explain all the association signals. Analyses showed that rs12342421 (S8) has stronger association when it was not combined with other SNPs, i.e. as a single marker rather than in a haplotype. Results from this study therefore hypothesised that the risk effect of single marker rs12342421 (S8) was stronger than the reported haplotypes, in the development of V617F-positive MPN in Hong Kong Chinese.

Within my study period, this is the first imputation study in genetic association studies of MPNs. Being an essential component in genetic association study, imputation enabled us to test many untyped markers for associations with MPNs and hence increased the chance to identify causal variants. Although the causal variant was still not found in this study, data generated from imputation and conditional logistic regression suggested that rs12342421 (S8) contributed an independent effect to the most significant association between *JAK2* risk-haplotype and MPNs.

4.4.4 *JAK2* SNPs do not directly affect normal function of genes

Next, I asked whether these strongly associated *JAK2* SNPs affect normal function of genes. The potential functions of the SNPs were examined using several web-based tools as listed in **Table 2.4**: SNPnexus (http://snp-nexus.org/index.html) (Dayem Ullah et al., 2012), SNP Function Prediction (FuncPred) **(**http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm) (Xu & Taylor, 2009), and F-SNP (http://compbio.cs.queensu.ca/F-SNP/) (Lee & Shatkay, 2008). None of the 19 SNPs was predicted to be functional except a weak prediction by FuncPred: 5'-upstream SNP (rs3808850 (S1)) and two intronic SNPs (rs3780378 (S18) and rs7849191 (S2))

may be involved in transcription factor binding. Nonetheless, the genetic association of these three SNPs with V617F-positive MPNs was marginal. A possible explanation for the lack of functional annotation of these highly associated markers is that these SNPs have not been extensively studied and hence have not been curated in the database. Alternatively, they may be located near a functional variant that has yet to be genotyped. Subsequent experiments are therefore required to investigate the functional impact of these SNPs. Functional studies and studies investigating additional polymorphisms may be able to clarify this. Not as expected, rs12342421 (S8) was not predicted to be functional.

I also identified regulatory potential of these SNPs and to link them to any specific genes that may be causal in the pathogenesis of MPNs, through analysis of expression quantitative trait loci (eQTL) across the *JAK2* gene (142.8 kb). Several eQTL online tools were used: eQTL resources @ the Pritchard lab (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/) (Pritchard), seeQTL (http://www.bios.unc.edu/research/genomic_software/seeQTL/) (Xia et al., 2012), and UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgTracks?org=human) (Kent et al., 2002). eQTL analysis of the *JAK2* gene detected neither cis- nor trans-regulatory regions. However, two eQTLs were detected when I expanded the region to 500 kb encompassing the *JAK2* gene: rs1053889 a cis-eQTL for HSS00330739, and rs7871570 an exon-cis-QTL for CDC37L1 (cell division cycle 37-like 1).1. From literature, HSS00330739 correlates with inflammation gene (Puig et al., 2011) while CDC37L1 facilitates the protein folding associated with HSP90 (Scholz et al., 2001). HSP90 has been shown to be upregulated in various cancers. It was discovered that CDC37 and HSP90 are required for type I and II interferon pathways. The inhibition of HSP90 degrades *JAK2* and thus blocks the interferon pathways through

STAT 1 that has not been completely understood (Shang & Tomasi, 2006). Nevertheless, no evidence of mechanism was shown in the context of MPN's pathogenesis. In addition, the eQTLs were not in LD with the *JAK2* gene and were also located beyond the two recombination hotspots encompassing the *JAK2* gene (Olcaydu et al., 2009a). All these circumstantial findings argue that the causal variants driving the disease development are unlikely the SNPs or haplotypes reported here. Owing to limited eQTL studies on different tissues or cell types, eQTL studies might provide only limited knowledge for linking regulatory variants to specific genes in various tissues or cell types. There might be some other eQTLs that have not been curated, leading to the limited information (Flutre et al., 2013). In silico analysis suggest that these *JAK2* SNPs did not directly regulate the expression of specific genes. It could be that these SNPs do not *directly* regulate the expression level of mRNAs or proteins, but rather *indirectly* by their interference with other factors.

Since the SNP and eQTL in silico analyses suggested no/ indirect function of the SNPs, I focused on the centre 40 bases of S8 (both major and minor alleles were considered) that was shown to be the most significantly associated SNP with MPNs. By filtering out data output with core similarity and matrix similarity below 0.80, four transcriptional factors were predicted in this 40-bases S8 region using the program MatInspector (www.genomatix.de/matinspector.html) (Cartharius et al., 2005). The transcriptional factors are *RFX4* (regulatory factor X, 4), *FOXQ1* (forkhead box Q1), *RUNX2* (Runt-related transcription factor) or *Cbfal* (Core-binding factor al) or *AML3* (acute myeloid leukemia), and (*Brn-5*),-homeodomain factor Pou6f1.

RFX4 was downregulated in gene expression analysis of bone marrow stromal cells co-cultured with CD34+ cells that revealed the upregulation of metabolism related pathways (Civini et al., 2013).

Brn-5 was implicated to be involved in T-cell specification in an expression analysis for transcription factor dynamics in T-cel (David-Fung et al., 2009). There was no evidence on the involvement of T-cell in MPN's development except the regulation of T-cell inflammation by *SOCS1* and *SOCS3* through JAK-STAT pathways (Tamiya et al., 2011) which are hyper-activated in MPNs. Future work to study their interaction would be enlightening to understand how lymphoid lineage is correlated with MPN development on myeloid lineage.

FOXQ1 was recently shown to be a novel target of the Wnt pathway and also one of the most over-expressed genes in colorectal cancer (Christensen et al., 2013). The Wnt pathway is important in haematopoiesis and it has just been indicated to be involved in the epigenetic regulation of MPN with aberrant methylation (Bennemann et al., 2012). The mechanism is not completely understood but epigenetic event may serve to complement genetic events in the pathogenesis in MPNs.

RUNX2 is known to be expressed in haematopoietic stem cell compartment that is associated with acute myeloid leukaemia at high level (Kuo et al., 2009), and in multiple myeloma (Colla et al., 2005). Depending on cell context, *RUNX2* can promote skeletal development, suppress cell proliferation, and function as a tumour suppressor or oncogene (He et al., 2011). Animal study showed that mice with doxycycline-induced *RUNX2*-II isoform showed evidence of myeloproliferative disorder and ob-

struction of lymphocyte development (He et al., 2011). Furthermore, transforming growth factor-β (*TGF-β*), one of the regulators of *RUNX2*, (Alarcon-Riquelme, 2004) was shown to inhibit the growth of human myeloid leukaemia cells by regulating the expression and activities of G2 checkpoint kinases (Hu et al., 2007). All these evidence support the biological plausibility of our finding that S8 may be an important variant in disease pathogenesis. However, its involvement in the pathogenesis of MPNs remains unclear until further experimental validation.

4.5 Conclusion

The *JAK2* germline polymorphisms or risk alleles were correlated with MPN patients with V617F mutation in Hong Kong Chinese population. Individual *JAK2* SNP rs12342421 (S8) is independently associated with predisposition to the development of V617F-positive MPN by 3.55 fold for the minor allele C. No significant correlation was found between V617F-negative MPN patients and the *JAK2* risk alleles.

Part II Meta-analysis

CHAPTER 5 Meta-analysis of *JAK2* SNPs

5.1 Introduction

There has been evidence suggesting that *JAK2* 46/1 haplotype contributed to the development of V617F-positive MPNs, but the findings for V617F-negative MPNs are inconsistent and less convincing. While most of the studies detected no association between the risk-haplotype and V617F-negative MPNs (Zhang et al., 2012a; Hsiao et al., 2011; Olcaydu et al., 2009a; Jones et al., 2009; Trifa et al., 2010), a positive correlation with V617F-negative MPN patients was reported in a few studies with slightly larger sample size than ours (n>53) (Pardanani et al., 2010; Tefferi et al., 2010b). The cumulative contrary results are possibly the reflection of inadequate statistical power. Moreover, these studies were conducted mainly in Caucasians, only a few in Asian with small sample size.

China is a multi-ethnic country. It will be inspiring if the relationship between V617F and *JAK2* haplotype is revealed in China's different ethnic groups. Unfortunately, the results of the genetic association studies on the role of the *JAK2* polymorphisms in MPNs have generated substantial controversy. Moreover, there was insufficiency of such studies in the Chinese population. Apart from Chinese population, studies in Taiwanese and Japanese also revealed specific alleles/additional locus predisposing to ET/MPNs respectively. Nonetheless, these studies only focused on the V617F-positive group (Ohyashiki et al., 2012; Hsiao et al., 2011).

In this study, *JAK2* polymorphisms were not associated with V617F-negative MPNs. The failure to detect a genetic effect between *JAK2* polymorphisms and V617F-

negative MPNs as reported elsewhere (Pardanani et al., 2010; Tefferi et al., 2010b) in Hong Kong population could be attributed to the small sample size (n=44) in this study (Button et al., 2013), or there is no genuine genetic association. Thus, given the sufficient data from publications, a meta-analysis was performed across populations from different racial descent to reconcile the conflicting findings and to validate the consistency of V617F-genetic effects in Asian populations. Such analysis of combined data from individual studies is proved to enhance the power through increasing sample size, ethnicities and races (Walker et al., 2008).

5.1.1 Gene variants and frequency

There exist 3 commonly employed tagging single nucleotide polymorphisms (tSNPs) of the *JAK2* haplotype- the diallelic SNPs (46/1 allele; not-46/1 allele): rs10974944 (**G**; C), rs12343867 (**C**; T), and rs12340895 (**G**; C) (Jones et al., 2009; Olcaydu et al., 2011). The **G**, **C**, and **G** alleles are in complete linkage disequilibrium (LD) with the 46/1 haplotype and are associated with increased occurrence of V617F-positive myeloproliferative neoplasms in Caucasians (Trifa et al., 2010; Pardanani et al., 2010). The overall frequencies of the **G**, **C**, and **G** alleles are all 0.20 among CHB population included to date in the 1000 Genomes Project. The three SNPs have similar allele frequencies in European and Asian populations however a slightly lower of that in African populations: 0.16 (**G**), 0.16 (**C**), 0.13 (**G**) from the 1000 Genomes Project. The importance of *JAK2* locus in MPNs susceptibility in the Europeans prompted the investigation of its genetic architecture in Chinese population since there could exist substantial differences in linkage disequilibrium (LD) structure between different populations. Genetics replication studies in Asian, therefore, serve to validate the reported findings in European populations.

Although the lower allele frequencies in African lower the power to detect an association, the African population, having the high genetic variability and generally lower LD, is suitable for LD examination and identification of potential functional sequence variants (Bye et al., 2012). Such studies in population with shorter LD distance could increase the efficiency of causal variant identification (Teo et al., 2010). Unfortunately, there was no such study done in African and only little information is available from African with MPNs: the insignificant different V617F frequencies between African Americans and Caucasian Americans by The Washington DC VA Medical Center (Mobarek et al., 2008).

5.1.2 Research aim

The aim of this part was to conduct a comprehensive review of previous findings and to study the effect of ethnicity and V617F mutational status in genetic association between *JAK2* gene and MPNs.

5.2 Methodology

Data from current genetic association study (Chapter 4) were included in this meta-analysis. Details of genotyping can be found in Section 2.2.10. The relevant studies were retrieved in duplicate by two reviewers from PubMed, Scopus and Chinese National Knowledge Infrastructure (CNKI) databases by two separate reviewers. Meta-analysis was performed using Cochrane RevMan software version 5.3.3 (Review Manager (RevMan) [Computer program]. Version 5.3.3. Copenhagen: The Nordic Cochrane Centre).

5.2.1 Phenotype definition

Healthy controls (n=470) recruited from Hong Kong Red Cross Blood Transfusion Service were classified as controls. MPN patients diagnosed according to WHO 2008 criteria were classified as cases or controls depending on the category of analysis.

5.2.2 Search strategy

Studies evaluating the genetic polymorphisms of the *JAK2* gene in MPNs patients were included. The search strategy used in this study was developed with the help of an experienced librarian. Three databases were used for comprehensive search including PubMed, Scopus and China National Knowledge Infrastructure (CNKI, a Chinese database) electronic databases for publication from 1 January 2005 to 4 September 2014 without language restriction. The same four sets of searching criteria and keywords were used for literature search except for CNKI; only keyword set 4 (SNP rs-numbers) were used to search for relevant studies based on the fact that this set of keywords included all the SNPs of interest. Reference lists of included studies and review articles were also checked for additional articles that are eligible to be included in the meta-analysis. Investigators were contacted by email for any clarification of information if necessary, up to 3 emails to follow up. The keywords for our search strategy can be found at **Table 5.1**.

Table 5.1 Keywords used in literature search

	Keywords	**Databases**
Set 1	(Allele* OR Haplotype*) AND (Myeloproliferative Neoplasm* OR MPN* OR polycythaemia vera OR essential thrombocythaemia OR primary myelofibrosis OR Myeloproliferative disease* OR Myeloproliferative disorder*)	PubMed, Scopus
Set 2	(Single nucleotide polymorphism* OR SNP*) AND (Myeloproliferative Neoplasm* MPN* OR polycythaemia vera OR essential thrombocythaemia OR primary myelofibrosis OR Myeloproliferative disease* OR Myeloproliferative disorder*)	PubMed, Scopus
Set 3	(*JAK2* OR Janus Kinase 2) AND Haplotype* AND (Myeloproliferative Neoplasm* OR MPN* OR polycythaemia vera OR essential thrombocythaemia OR primary myelofibrosis OR Myeloproliferative disease* OR Myeloproliferative disorder*)	PubMed, Scopus
Set 4	rs12342421 OR rs12343867 OR rs10974944 OR rs12340895 OR rs7046736 OR rs3808850 OR rs10974947	PubMed, Scopus, CNKI

5.2.3 Study selection

Study selection was performed by SP. Koh and SY. Lee, with disagreements resolved through discussion and by the opinion of a third reviewer (SP. Yip). Full-text publications were retrieved for relevant studies with the following inclusion and exclusion criteria listed in **Table 5.2**: 1) Diagnosis of MPNs; 2) at least 10 patients were included; 3) case-control study; 4) studies concerning the association of the *JAK2* gene polymorphisms with MPNs; 5) information of genotype or allele frequencies was provided; 6) MPN patients with splanchnic vein thrombosis (SVT) subdivided into Budd-Chiari syndrome and portal vein thrombosis were not excluded. Studies with available genotype data were evaluated for Hardy-Weinberg equilibrium (HWE) and studies that showed a deviation from exact test of HWE in the controls were excluded.

Table 5.2 Inclusion and exclusion criteria used in this study

	Inclusions	*Exclusions*
Participants	Myeloproliferative neoplasms (MPNs) subtypes • Polycythaemia vera • Essential thrombocythaemia • Primary myelofibrosis Common complications in MPNs • Splanchnic vein thrombosis (SVT) subdivided into Budd-Chiari syndrome and portal vein thrombosis	Diseases other than MPN subtypes
Types of studies	Case-control studies • Retrospective Abstracts or unpublished work with sufficient information	Study without control sample Study with less than 10 patients Study with overlapping data Abstracts or unpublished work with insufficient information Study with deviation from HWE in the control group
Outcomes to be measured	*JAK2* polymorphism data: • Allelic count • Genotypic count	Non-*JAK2* polymorphism data
Period covered	1 January 2005 to 4 September 2014	

5.2.4 Data abstraction

Data extraction was performed with the help of SY. Lee, supplemented with a review by SP. Yip according to the inclusion criteria. The following data were collected from each study:

First author's name, study design, classification of MPNs and controls, the origin of the study, ethnicity, *JAK2* SNPs genotyped, genotyping method, and remarks if any. *JAK2* polymorphisms were included in this meta-analysis only when their association with MPNs were assessed at a minimum of 2 individual studies. Extracted data were compared with disagreement resolved by mutual consensus.

5.2.5 Statistical analysis

The meta-analysis examined the association between *JAK2* polymorphisms and MPNs by comparing Odds ratio (ORs) of the risk alleles for cases verses controls and V617F-positive cases verses V617F-negative cases. Hardy-Weinberg equilibrium (HWE) of all the SNPs with sufficient genotype data (allelic count) was assessed by Fisher's exact test using an online freeware (http://ihg.gsf.de/cgi-bin/hw/hwa1.pl) due to a lack of raw data. Studies that deviated from exact test of HWE distribution of genotypes in the controls were excluded (exact test $P<0.001$).All analyses were performed with Cochrane RevMan software (Review Manager (RevMan) [Computer program]. Version 5.3.3. Copenhagen: The Nordic Cochrane Centre). The Odds ratio (ORs) and 95% confidence intervals (CIs) were calculated for the associations tested. Then, based on the individual ORs, a pooled OR was estimated according to the per-

cent weight gained from each individual study using fixed-effects (FE) (Mantel-Haenszel) or random-effects (RE) (DerSimonian and Laird) models.

In meta-analysis containing 10 or more studies, the percentage of heterogeneity or the variability across studies were evaluated by Cochran's Q test quantified by I^2 (Higgins & Green). This statistic to quantify inconsistency across studies is

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$$

This step aimed to eliminate the unreliable estimates from studies with low statistical power. *P* value of heterogeneity <0.10 or I^2 >50% indicates significant heterogeneity whereas I^2= 0 indicated no heterogeneity (Higgins & Green). The FE model was used as default model and also in the situation when less than 10 studies were included in meta-analysis. If heterogeneity existed in single analysis with more than 10 studies, the analysis was done with FE model first and repeated in RE model. Sensitivity analysis was then performed by removing one individual study each time and re-analysing the remainder to identify any single source of heterogeneity responsible for the combined results. Additionally, the heaviest study (in terms of sample size) was omitted and the remainder was re-analysed. Individual study composed of only single MPN subtype was also omitted and analysis was re-conducted on the remainder.

In addition, potential source of heterogeneity among studies was also studied by stratifying the eligible studies into MPN subgroup (PV or ET or PMF). This aimed to check if the V617F loss of heterozygosity (LOH), especially in PV patients, affects the magnitude of predisposition to MPNs. Stratification by ethnicity was also con-

ducted by classifying the subjects into Caucasian group and Asian group. The Z-test of interaction was used to measure the difference between two effect estimates. This test, based on natural logarithm of ORs, generates z-values for the calculation of *P* values using online freewares (http://vassarstats.net/tabs.html#z & http://graphpad.com/quickcalcs/PValue1.cfm). Meanwhile, the presence of any publication bias was assessed with funnel plot. The asymmetry of funnel plots based on natural logarithm of the OR was measured.

5.3 Results

5.3.1 Study identification and selection

Literature search from PubMed, Scopus, and CNKI returned 1030 potentially relevant studies using 4 sets of predefined keywords within 1 January 2005 to 4 September 2014 (**Table 5.1**). After deduplication, 970 records were screened by their title, abstract, and relevance. Of the 970 records, 270 records were retrieved in full for eligibility assessment. From first preliminary screening, 37 studies were selected for data retrieval while 233 records were excluded for the following reasons: duplicates, studies with overlapped samples/data, non-*JAK2* association studies, review articles. One additional study was identified through manual searching from the reference list (Hu et al., 2011). When identical data were described in more than one publication, only the study with larger sample size was included in the meta-analysis.

Further selection based on inclusion and exclusion criteria identified 23 studies (including current study) to be included in meta-analysis. In the final selection, 15 studies were excluded because of the unmet inclusion criteria: three studies do not provide genotype data (Tian et al., 2012; Zhang et al., 2010; Olcaydu et al., 2009b), 1

article (Colaizzo et al., 2011) and 1 thesis (European patients) (Gao, 2011) lacked a clear MPN definition and control group, 3 studies with incompatible SNPs (Kim et al., 2010; Lambert et al., 2009; Wang et al., 2011), 3 studies contained overlapped data (Patnaik et al.; Rumi et al., 2011; Pietra et al., 2012), 3 studies do not have a control group (Alvarez & Frank, 2004; Tefferi et al., 2010a; Martinez-Aviles et al., 2007), and 1 study deviated from HWE distribution in the control group (Pagliarini-e-Silva et al., 2013). HWE in controls was checked for all selected publications except for 2 studies (Jones et al., 2010; Jones et al., 2009) without genotype frequencies (allele frequencies as reference), HWE was assumed. The flow of study identification and selection was summarised in **Figure 5.1**.

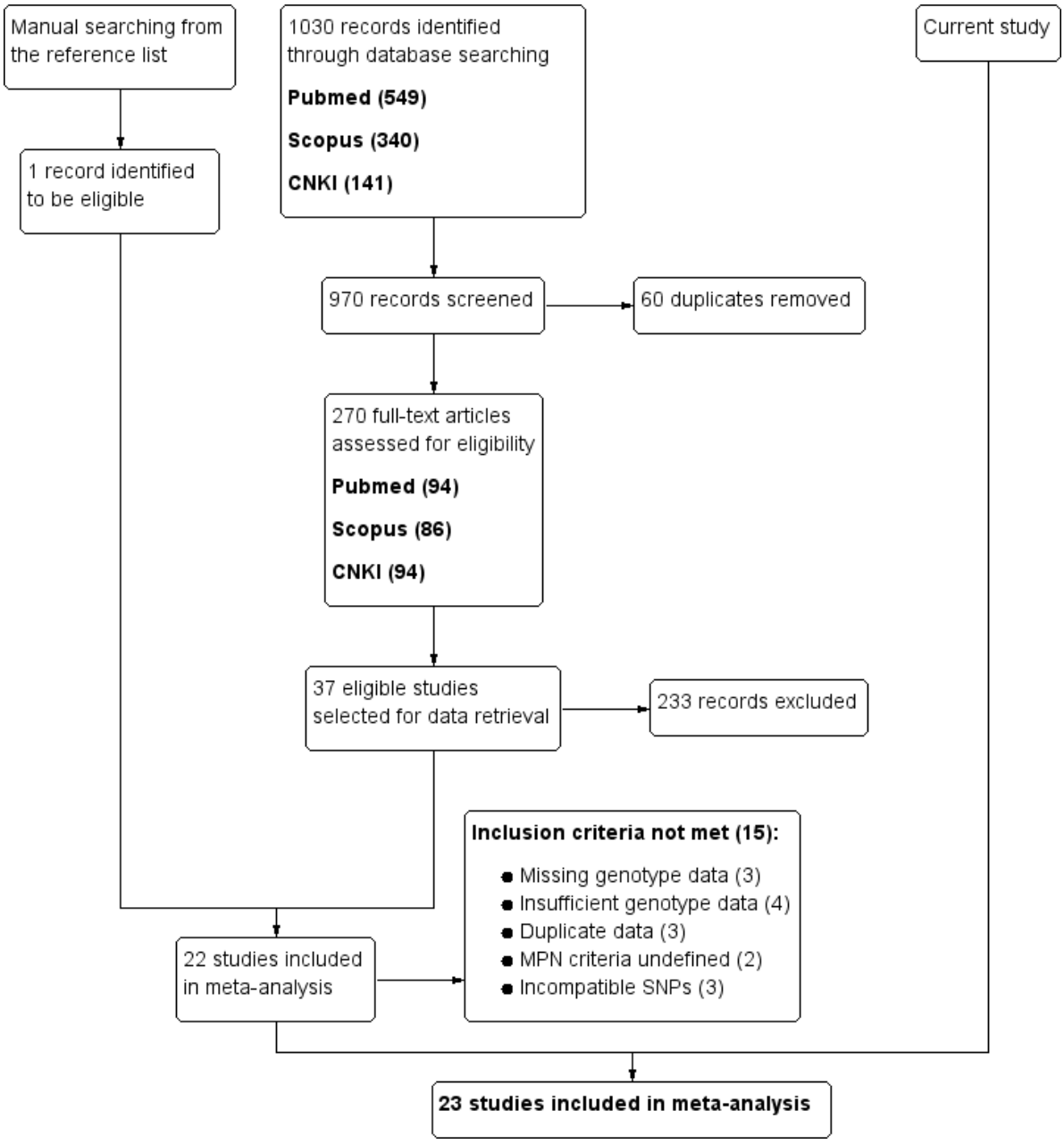


Figure 5.1 Flow chart of the study selection process.

5.3.2 Study characteristics

A review of the 15 excluded studies revealed the population distribution of 12 in Caucasian populations and 3 in Asian; Korean (Kim et al., 2010) and Chinese (Tian et al., 2012; Zhang et al., 2010). Four studies (Tefferi et al., 2010a; Olcaydu et al., 2009b; Alvarez-Larran et al., 2012; Martinez-Aviles et al., 2012) were conducted in a population-based cohort design, 1 in retrospective design (Colaizzo et al., 2011), and otherwise case-control design. Study size ranged from 34 to 1473 patients. Eleven of the 15 studies screened for the V617F mutation in their MPN cohorts; Olcaydu et al (2009b) and Wang et al (2011) investigated the germline association in MPNs development irrespective of V617F mutational status while Patnaik (2010) focused on only V617F-negative MPN patients. Among the excluded studies, 10 studies included control group: healthy local (Olcaydu et al., 2009b; Gao, 2011; Kim et al., 2010; Tian et al., 2012; Patnaik et al., 2010; Pietra et al., 2012; Zhang et al., 2010), haematological normal controls (Lambert et al., 2009), multiple primary care clinics and outpatient practices (Wang et al., 2011), and a combination of healthy subjects and V617F-negative secondary erythrocytosis and leucocytosis (Rumi et al., 2011). Genotype frequencies were unavailable in most of these studies however 3 studies were eligible for minor allele frequency comparison (**Table 5.3**). Studies from Spain (Alvarez-Larran et al., 2012; Martinez-Aviles et al., 2012) observed a lower minor allele frequency (MAF) when V617F was excluded, compared with MPNs irrespective of V617F status. Wang et al. (2011) reported a higher allele frequency in PV patients. This could be explained by the V617F loss of heterozygosity (LOH) in PV patients while MPN-risk alleles were preferentially acquired in cis with V617F. However, in a study by Lambert et al. (2009), both controls and ET patients showed comparable allele frequencies. They hypothesised that the leukaemic blasts are negative in many cases of acute myeloid leukaemia transformed from V617F-positive ET.

The reduced V617F-positive clones might explain the slight decrease of minor allele frequency.

Table 5.4 summarizes the study characteristics of MPNs cases and controls in the *JAK2* genetic association study. All studies were conducted in a population based case-control manner: 1 in Taiwan (Hsiao et al., 2011), 2 in Japan (Ohyashiki et al., 2012; Tanaka et al., 2013), 2 in China (Hu et al., 2011; Zhang et al., 2012b; Wang et al., 2013), current study based in Hong Kong; all otherwise Caucasian populations. Healthy local individuals were used as controls in most of the studies while database controls were included in 7 studies (Pardanani et al., 2008; Kilpivaara et al., 2009; Pardanani et al., 2010; Jones et al., 2009; Jones et al., 2010; Guglielmelli et al., 2010; Villani et al., 2012). Other control populations were non-MPNs patients with peripheral vein thrombosis (PVT) (Kouroupi et al., 2011) and individuals with normal hemogram or a haematologic reactive condition (Olcaydu et al., 2011). For studies in which both local and database controls were used, a combination of both was analysed provided no overlapping of controls in the analysis. Sample size ranged from 32 (Spasovski et al., 2013) to 542 (Olcaydu et al., 2011) MPN patients. Three studies (Kouroupi et al., 2011; Smalberg et al., 2011; Villani et al., 2012) with splanchnic vein thrombosis (SVT) associated-MPNs were also included in the meta-analysis. *JAK2* risk-haplotype tSNPs rs12343867, rs10974944, and rs12340895 were the top 3 most frequently assessed SNPs; while Taqman SNP genotyping assay was the predominant genotyping method among the studies (9 out of 23 studies). World Health Organisation (WHO) criteria were used as the major MPN diagnosis. Publications by Jones et al. (2010; 2009) were partly overlapped in the patient sub-cohorts therefore analyses were conducted with exclusion of the overlapped sub-population; this was also applied for the IRCCS-based studies (Olcaydu et al., 2011; Villani et al., 2012;

Guglielmelli et al., 2010) and the WTCCC controls. Of note, multiple publications from the same group were included in 2 situations: Pardanani et al. (2008; 2010) investigated different *JAK2* polymorphisms while Olcaydu et al. (2009a; 2011) investigated different populations in their studies.

Table 5.3 Allele frequencies of *JAK2* polymorphisms from three excluded studies

Study	SNP			Cases		VF-pos Group	Cases		Non-cases	
		MA	Group	n	MAF		n	MAF	n	MAF
Alvarez-Larran et al. (2012)	rs12343867 /rs12340895		PV	26	0.462	-	-	-	-	-
			ET	36	0.417	-	-	-	-	-
Wang et al. (2011)	rs11999802	G	PV	34	0.583	-	-	-	3278	0.241
	rs3780373	C	PV	34	0.621	-	-	-	3278	0.284
	rs12347727	G	PV	34	0.581	-	-	-	3278	0.266
	rs10974993	T	PV	34	0.603	-	-	-	3278	0.284
	rs7030260	A	PV	34	0.583	-	-	-	3278	0.271
	rs10118930	T	PV	34	0.603	-	-	-	3278	0.284
	rs7047795	T	PV	34	0.603	-	-	-	3278	0.284
	rs7851556	T	PV	34	0.586	-	-	-	3278	0.271
	rs10815149	C	PV	34	0.586	-	-	-	3278	0.273
	rs10491651	G	PV	34	0.569	-	-	-	3278	0.273
	rs16922786	G	PV	34	0.569	-	-	-	3278	0.273
	rs3780381	C	PV	34	0.567	-	-	-	3278	0.278
	rs7862042	A	PV	34	0.567	-	-	-	3278	0.279
	rs6476948	T	PV	34	0.569	-	-	-	3278	0.274
Lambert et al. (2009)	rs2230724	A	ET	111	0.599	ET	43	0.640	114	0.540

JAK2, Janus kinase 2; SNP, single nucleotide polymorphisms; VF-pos, V617F-positive; MA, Minor allele; MAF, minor allele frequency; SVT, splanchnic vein thrombosis; PV, polycythemia vera; ET, essential thrombocythemia; MPN, Myeloproliferative Neoplasms (classic: PV, ET, MF).

Table 5. 4 Baseline characteristics of included studies

No	Study/ Ethnicity of the study	MPN criteria	Cases Classification SVT/MPN/PV/ET/MF/ O	VF Classification SVT/MPN/PV/ET/MF/ uMPN	Control	*JAK2* SNPs meta-analysed	Genotyping method	Remarks*
1	Pardanani et al., 2008 Caucasian-USA	WHO	0/0/84/37/58/0 (179)	0/0/79/19/28/0 (126)	CEU HapMap 24,2008	rs3808850 rs7046736 rs10815148 rs12342421	GenomeLab SNP-stream genotyping system	Cases: MPNs; V617F was set as covariate
2	Jones et al., 2009 Caucasian-USA, UK, Greece, Germany	NA	0/775/192/357/41/247 (775)	0/0/203/224/41/124 (454)	188 UK controls 108 GR controls 1500 UK WTCCC controls	rs12340895	Pyrosequencing	Cases: MPNs, PV, PMF, uMPN; ET cases were excluded for analysis to avoid duplicate data with Jones 2010; NA for HWE. Of the 247 other MPNs, only 124 uMPN cases were included in meta-analysis
3	Kilpivaara et al., 2009 Caucasian-USA	Medical records	0/324/0/0/0/0 (324)	0/245/0/0/0/0 (245)	2999 WTCCC European	rs10974944	Taqman	Cases: MPNs, ET MPN diagnosis was referred to author's reference
4	Olcaydu et al., 2009 Caucasian-Austria	WHO	0/333/0/0/0/0 (333)	0/213/0/0/0/0 (213)	99 non-MPN local	rs10974944 rs12343867 rs3780367 rs1159782	Taqman	Cases: MPNs; Olcaydu- based in Vienna
5	Andrikovics et al., 2010 Caucasian-Hungary	NA	0/312/153/131/28/339 (312)	0/251/153/78/20/0 (251)	331 local	rs12343867	Melting curve analysis	Cases: MPNs, PV, ET, PMF
6	Guglielmelli et al., 2010 Caucasian-Italy	2008 WHO; BM	0/0/0/0/202/0 (202)	0/0/0/0/133/0 (133)	235 local	rs12343867	SNP RT-PCR assay	IRCCS study; Cases: PMF; only local controls were used to avoid duplicate WTCCC data for rs12343867

Table 5. 4 Baseline characteristics of included studies (continued)

No	Study/ Ethnicity of the study	MPN criteria	Cases Classification SVT/MPN/PV/ET/MF/ O	VF Classification SVT/MPN/PV/ET/MF/ uMPN	Control	*JAK2* SNPs meta-analysed	Genotyping method	Remarks*
7	Jones et al., 2010 Caucasian-UK, Germany, Italy, Greece	PV Study Group or WHO	0/0/0/751/0/0 (751)	0/0/0/404/0/0 (404)	1814 KORA controls & 1200 InCHIANTI controls	rs12340895	Pyrosequencing	Cases: ET; Duplicate GR and UK controls with Jones 2009 were excluded from analysis; NA for HWE
8	Pardanani et al., 2010 Caucasian-USA	2008 WHO	0/0/0/226/0/0 (226)	0/0/0/118/0/0 (118)	1500 UK WTCCC controls	rs12343867	TaqMan	Cases: ET ; only WTCCC controls were used to avoid duplicate data with Tefferi et al., 2010 for rs12343867
9	Tefferi et al., 2010 Caucasian-USA	2008 WHO	0/0/0/0/130/0 (130)	0/0/0/0/77/0 (77)	57 local	rs12343867	TaqMan	Cases: PMF
10	Trifa et al., 2010 Caucasian-Romania	WHO	0/0/69/65/15/0 (149)	0/107/0/0/0/0 (107)	150 local	rs10974944	PCR-RFLP & sequencing	Cases: MPNs
11	Hsiao et al., 2011 Asian-Taiwan	Medical reports	0/0/0/61/0/0 (61)	0/0/0/34/0/0 (34)	106 local	rs10974944 rs12340895 rs12343867	PCR-RFLP	Cases: ET
12	Hu et al., 2011 Asian-China	2008 WHO	0/0/125/87/0/0 (212)	0/0/81/45/0/0 (126)	213 local	rs12343867	HRM	Cases: PV, ET; Article in Chinese
13	Kouroupi et al., 2011 Caucasian-France	PVSG/ WHO	170/0/31/0/0/0 (201)	75/0/31/0/0/0 (106)	58 non-MPN patients with PVT	rs10974944	Taqman allelic discrimination assays	Cases: Only confirmed PV cases. MPN classification was referred to author's reference

Table 5. 4 Baseline characteristics of included studies (continued)

No	Study/ Ethnicity of the study	MPN criteria	Cases Classification SVT/MPN/PV/ET/MF/ O	VF Classification SVT/MPN/PV/ET/MF/uMPN	Control	*JAK2* SNPs meta-analysed	Genotyping method	Remarks*
14	Olcaydu et al., 2011 Caucasian-Italy	2008 WHO	0/772/258/400/111/3 (772)	0/542/0/0/0/0 (542)	203 Italian controls with a normal hemogram (n=43) or a haematologic reactive condition (n=160)	rs10974944	TaqMan	Cases: only sporadic MPNs including 683 PV, ET, PMF, Post-PV myelofibrosis, post-ET myelofibrosis, uMPN (one of the IRCCS studies but different SNP: rs10974944)
15	Smalberg et al., 2011 Caucasian-Italy	Bone marrow biopsy, red cell mass measurement	199/66/0/0/0/133 (199)	54/54/0/0/0/0 (54)	100 local	rs12343867	Taqman	Cases: MPN-associated SVT from (EN-Vie) study cohort. MPN classification was referred to author's reference
16	Ohyashiki et al., 2012 Asian-Japan	2008 WHO	0/138/33/96/9/0 (138)	0/95/33/57/5/0 (95)	107 local	rs10974944rs12343867rs4495487	AS-PCR	Cases: PV, only genotype data from V617F-positive PV group was available; rs4495487 was excluded for its deviation from HWE *P*=0.002782 (Exact test) in the control group
17	Villani et al., 2012 Caucasian-Italy	2008 WHO	108/0/0/32/55/21 (108)	76/0/0/21/42/0 (76)	56 local	rs12343867	AS-PCR or RT-PCR	IRCCS study; Cases: SVT-associated ET, SVT-associated uMPN (suspected PMF duplicate data as in Guglielmelli et al., 2010 was excluded); Only local controls were used to avoid duplicate WTCCC data for rs12343867

Table 5. 4 Baseline characteristics of included studies (continued)

No	Study/ Ethnicity of the study	MPN criteria	Cases Classification SVT/MPN/PV/ET/MF/ O	VF Classification SVT/MPN/PV/ET/MF/u MPN	Control	*JAK2* SNPs meta-analysed	Genotyping method	Remarks*
18	Zhang et al., 2012 Asian-China	2008 WHO	0/225/77/70/36/42 (225)	0/129/68/38/23/0 (129)	226 local	rs12340895	HRM	Cases: MPNs; Slight HWE deviation p=0. 030712 (Exact test) in control group
19	Wang et al., 2013 Asian-China	WHO	0/635/125/263/206/0 (635)	0/396/NA/155/99/0 (396)	360 local	rs12340895 rs10974944	Taqman	Cases: MPNs, PV, ET, PMF
20	Spasovski et al., 2013 European-Serbia	WHO	0/51/19/26/6/0 (51)	0/32/0/0/0 (32)	12 local	rs12343867	Sequencing	Cases: MPNs, PV, ET, PMF
21	Zerjavic et al., 2013 European-Slovenia	standard diagnostic procedures, which also included the V617F mutation screening	0/226/57/149/20/0 (226)	0/130/47/68/15/0 (130)	459 local	rs12342421 rs12343867 rs10974944	AS-PCR	Cases: MPNs
22	Tanaka et al., 2013 Asian-Japan	WHO	0/108/19/61/10/18 (108)	0/59/19/36/3/1 (59)	104 local	rs10974944	TaqMan SNP genotyping assays, direct sequencing	Cases: MPNs, PV, ET, PMF

Table 5. 4 Baseline characteristics of included studies (continued)

No	Study/ Ethnicity of the study	MPN criteria	Cases Classification SVT/MPN/PV/ET/MF/ O	VF Classification SVT/MPN/PV/ET/MF/ uMPN	Control	*JAK2* SNPs meta-analysed	Genotyping method	Remarks*
23	Current study, 2012 Asian-Hong Kong	WHO	0/171/NA/NA/NA/1 (172)	0/128/NA/NA/NA/1 (137)	470 local	rs10974944 rs12343867 rs12340895 rs7046736 rs3808850 rs10815148 rs12342421 rs3780367 rs1159782	RFLP, UPR-melting	Data from some of the SNPs were imputed. Cases: MPNs, PV, ET, PMF

All studies were conducted in case-control design. No HWE deviation in the control group unless specified. MPNs include a mixture of patient population consisting of PV, ET, and PMF.

*Cases under remarks represent MPN cases used in current meta-analysis.

MPN indicates Myeloproliferative Neoplasms (classic: PV, ET, PMF); SVT, splanchnic vein thrombosis; PV, polycythaemia vera; ET, essential thrombocythaemia; PMF, primary myelofibrosis; O, others; VF, V617F mutation; *JAK2*, Janus kinase 2; SNPs, single nucleotide polymorphisms; WHO, world health organisation; NA, not available; HWE, Hardy-Weinberg equilibrium; BM, bone marrow biopsy; RT-PCR, Real time-Polymerase Chain Reaction; RFLP, Restriction fragment length polymorphism; HRM, High-resolution melting; PVT, peripheral vein thrombosis; uMPN, unclassified MPN; CML, chronic myeloid leukaemia; EN-Vie, European Network for Vascular Disease of the Liver; AS-PCR, allele specific-polymerase chain reaction; UPR, unlabelled probe.

5.3.3 Overall meta-analysis

The data pool composed of a maximum of 11 studies for 1 SNP comparison. The genotype distribution in control group of all studies were in HWE except a slight deviation (*P*=0. 037) in the study by Zhang et al. (2012b). However, this study was included considering the subtle heterogeneity (*P*=0. 037) it caused and MPN studies scarcity in Chinese population. Data were summarised in the sequential order from the 5' end to the 3' end of the *JAK2* sense strand as presented in **Table 5.5-5.8**.

Table 5.5 summaries the association for 6 SNPs (rs3808850, rs7046736, rs10815148, rs12342421, rs3780367, and rs1159782) in which only 2 studies were available for meta-analysis. The minor alleles in rs7046736, rs10815148, rs12342421, rs3780367, and rs1159782 all showed risk effects towards the development of V617F-positive MPN (**Table 5.5**). However, rs3808850 (OR=1.01; 95%CI, 0.81-1.26; *P*=0.92), SNP located on *JAK2* 5' upstream, showed no association with MPNs regardless of V617F mutation.

Table 5.5 Association of rs3808850, rs7046736, rs10815148, rs12342421, rs3780367 and MPNs

JAK2 snp, MAF	No. of study (case/control)	FE OR [95% CI], *P*	het: I^2, *P*	Remarks
rs3808850, allele T=0.48				
1. Combined-MPNs and controls	2 (351;530)	1. 01 [0.81-1. 26], *P*=0.92	I^2=98%, *P*<0.00001	EUR: MAF(A):0.32 CEU: MAF(A) 0.34 CHB: MAF(T) 0.48
rs7046736, allele A=0.28				
1. Combined-MPNs and controls	2 (342;530)	2.21 [1.78-2.75], *P*<0.00001	I^2=0%, *P*=0. 35	
rs10815148, allele A=0.13				
1. Combined-MPNs and controls	2 (347;527)	1.70 [1.35-2.16], *P*<0.0001	I^2=79%, *P*=0. 03	
rs12342421, allele C=0.20				
1. Combined-MPNs and controls	3 (568;986)	2.33 [1.99-2.74], *P*<0.00001	I^2=92%, *P*<0.00001	
2. V617F-pos MPNs and controls	2 (258; 929)	2.81 [2.30-3.43], *P*<0.00001	I^2=87%, *P*=0. 006	
3. V617F-neg MPNs and controls	2 (140; 929)	1.15 [0.88-1.51], *P*=0.30	I^2=85%, *P*=0. 01	
rs3780367, allele G=0.20				
1. V617F-pos MPNs and controls	2 (341; 569)	2.70 [2.16-3.38], *P*<0.00001	I^2=29%, *P*=0. 21	
rs1159782, allele C=0.20				
1. V617F-pos MPNs and controls	2 (341; 569)	2. 67 [2.13-3.33], *P*<0.00001	I^2=0%, *P*=0.79	

MAF indicates minor allele frequency for Han Chinese in Bejing; FE, fixed-effects model; OR, odds ratio; 95 % CI, 95% confidence interval; *P*, *P* value; het, heterogeneity under FE model; I^2, heterogeneity statistical tests; MPNs, Myeloproliferative Neoplasms (classic: PV, ET, MF); V617F-pos, V617F-positive; PV, polycythaemia vera; ET, essential thrombocythaemia; PMF, primary myelofibrosis; V617F-neg, V617F-negative.

5.3.4 Stratification analysis

Since most studies focused on the *JAK2* risk-haplotype tSNPs (rs10974944 (S9), rs12343867 (S12), and rs12340895 (S13)), subgroup analyses were performed on the 3 tSNPs. Depending on data availability, specific data for *JAK2* SNPs were stratified on the basis of V617F status: V617F-positive, V617F-negative and the combined group of V617F-positive and -negative patients. The groups were: V617F-positive MPN patients verses controls, V617F-negative MPN patients verses controls, overall MPN patients verses controls and V617F-positive MPN patients verses V617F-negative MPN patients.

Our results showed that V617F strengthened the magnitude of association between the risk alleles and MPNs for all the *JAK2* risk-haplotype tSNPs.

5.3.4.1 Association of *JAK2* rs10974944 and MPNs

Overall, there was evidence of an association between the risk allele (G) of rs10974944 (S9) and V617F-positive and -negative MPNs when all the eligible studies were pooled into the meta-analysis. The minor allele G was significantly associated with V617F-positive MPNs (OR 2.82; 95% CI: 2.57-3.10, $P<0.00001$; $I^2= 26\%$, $P=0.20$) and in lesser extent in that of V617F-negative MPNs (OR 1.27; 95% CI: 1.12-1.44, $P=0.0001$; $I^2= 26\%$, $P=0.22$) when compared with controls. When V617F-positive and -negative patients were combined as overall MPNs and compared with controls, the OR was in between that of V617F-positive and V617F-negative groups (OR 2.24; 95% CI: 2.06-2.43, $P<0.00001$; $I^2= 49\%$, $P=0.03$) (**Figure 5.2 A-C; Table 5.6**). To know if the association is different between V617F-positive and -negative MPNs, V617F-positive MPN patients were compared with V617F-negative MPN

patients. The result indicated a significant difference between V617F-positive MPNs and V617F-negative MPNs (P<0.00001) (**Figure 5.3**).

Analyses were further stratified by MPN subtype (PV, ET or PMF) verses controls among eligible studies. Under the comparisons of V617F-positive verses controls, significant risks were found for PV and ET: the OR in V617F-positive PV subgroup (OR 4.07; 95% CI: 3.03-5.46, P<0.00001; I^2= 0%, P=0.45) was stronger by almost 1.5-fold than that of ET subgroup (OR 2.30; 95% CI: 1.95-2.73, P<0.00001; I^2= 0%, P=0.98) On the contrary, marginal association (ORs near 1) was detected in sub-group analysis of V617F-negative verses controls (**Table 5.6**). PMF cases could not be appraised individually because of a lack of specific data.

In the stratified analysis by ethnicity, increased risks were found among Caucasians in the sub-analysis of V617F-positive MPNs verses controls, and V617F-negative MPNs verses controls. However, Z-test of interaction indicated the differences to be insignificant (P>0.05). Nonetheless, when V617F-positive and -negative patients were combined as overall MPNs and compared with controls; the pooled OR significantly increased in Caucasian population (Z-test, P=0.1119) (**Table 5.6**). Overall, test of heterogeneity suggested insignificant inconsistency among the studies.

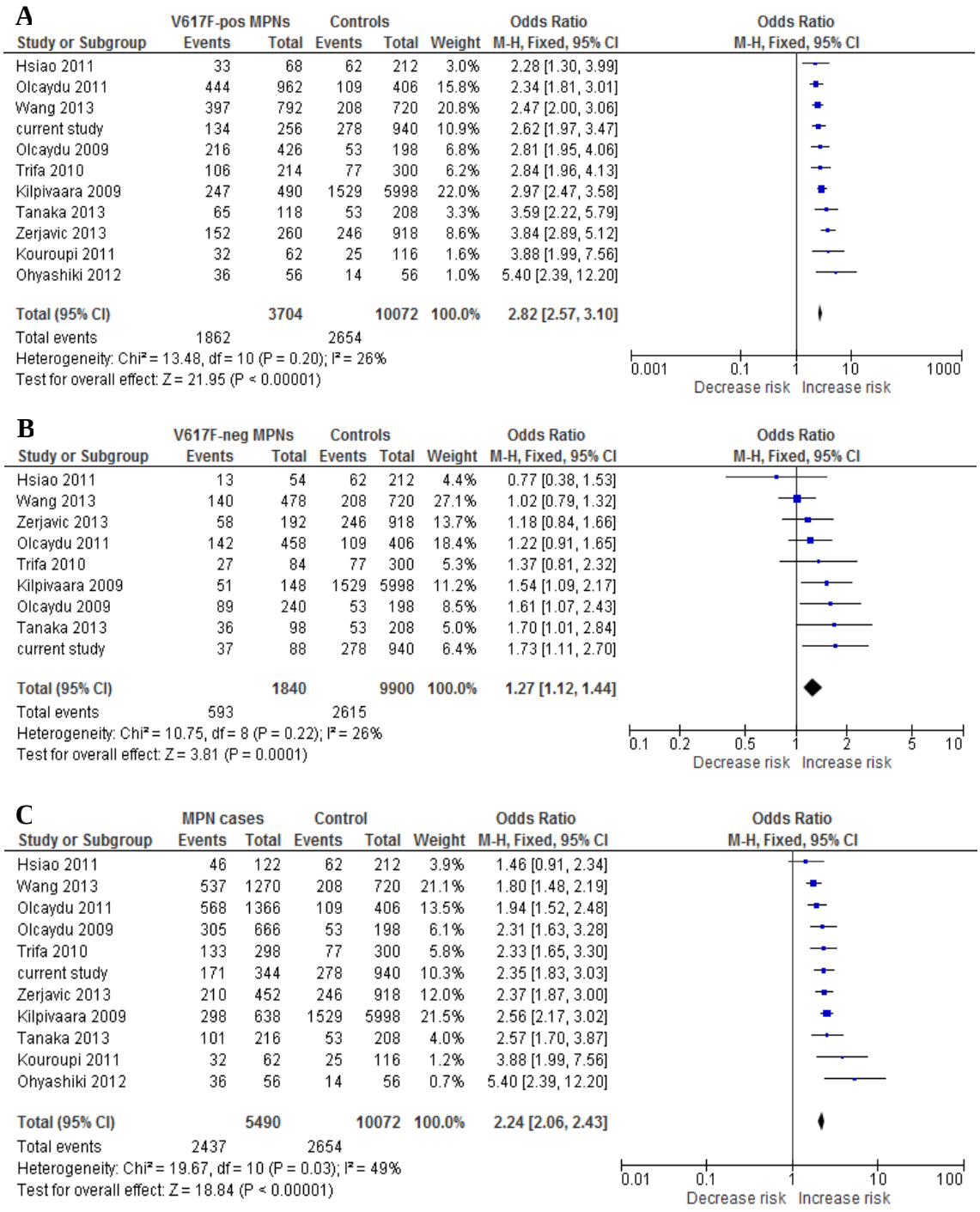


Figure 5.2 Forest plots of *JAK2* rs10974944 and MPN risk under fixed-effects model.

A. V617F-positive MPN cases verses controls, B. V617F-negative MPN cases verses controls, C. Overall MPN cases verses controls

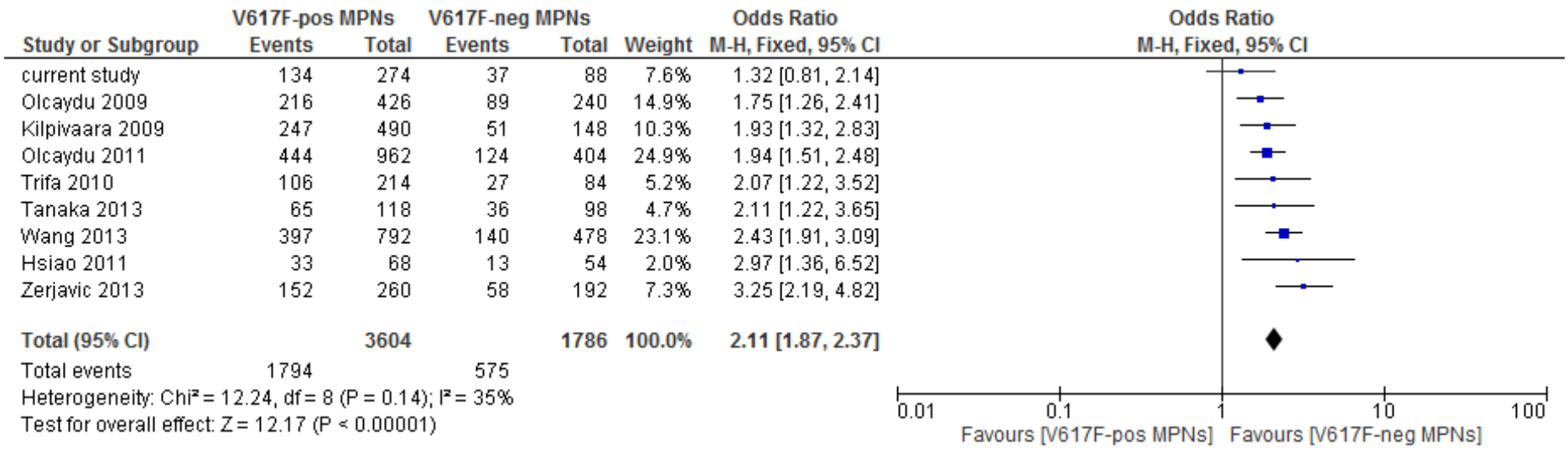


Figure 5.3 Forest plot of *JAK2* rs10974944 and MPN risk: MPN V617F-positive MPNs patients verses V617F-negative.

Table 5.6 Meta-analysis of *JAK2* rs10974944 on MPNs risk

***JAK2* snp, MAF**	**Study (no. of case/control)**	**FE OR [95% CI], *P***	**Test of interaction, Z**	**het: I^2, *P***	**Remarks**
rs10974944, allele G=0.20					
1. V617F-pos MPNs and controls [1,3]	11 (1852;5036)	2.82 [2.57, 3.10], *P*<0.00001		I^2= 26%, P=0.2000	
Studies with only PV/ET/PMF removed [1]	8 (1759;4844)	2.80 [2.54-3.08], *P*<0.00001	[1] *Z*=0.1038; *P*=0.9173	I^2= 27%, P=0.2100	1 ET; 2 PV studies
Asians [2]	4 (586;964)	2.58 [2.20-3.03], *P*<0.00001		I^2= 14%, P=0.3200	
Caucasians [2]	7 (1266;4072)	2.96 [2.64-3.31], *P*<0.00001	[2] *Z*=1.3742; *P*=0.1694	I^2= 24%, *P*=0.2400	
PV	4 (131;660)	4.07 [3.03-5.46], *P*<0.00001		I^2= 0%, *P*=0.4500	
ET	5 (358; 4039)	2.30 [1.95-2.73], *P*<0.00001		I^2= 0%, *P*=0.9800	
The heaviest study removed [3]	10 (1607;2037)	2.78 [2.50-3.09], *P*<0.00001	[3] *Z*=0.1979; *P*=0.8431	I^2= 32%, *P*=0.1600	Kilpivaara 2009
2. V617F-neg MPNs and controls [4, 6]	9 (920; 4950)	1.27 [1.12-1.44], *P*=0.0001		I^2= 26%, *P*=0.2200	
Studies with only PV/ET/PMF removed [4]	8 (893; 4844)	1.30 [1.14-1.47], *P*<0.0001	[4] *Z*=-0.2560; *P*=0.7980	I^2=18%, *P*=0.2900	1 ET study
Asians [5]	4 (359; 1040)	1.18 [0.97-1.43], *P*=0.1000		I^2= 60%, *P*=0.0600	
Caucasians [5]	5 (561;3910)	1.35 [1.15-1. 58], *P*=0.0003	[5] *Z*=1.0519; *P*=0.2928	I^2= 0%, *P*=0.6900	
ET	4 (190;1040)	1.18 [0.93-1.50], *P*=0.1800		I^2= 18%, *P*=0.3000	
The heaviest study removed [6]	8 (681;4590)	1.37 [1.19-1.58], *P*<0.0001	[6] *Z*=-0.7843; *P*=0.4329	I^2= 0%, *P*=0.4600	Wang 2013

Table 5.6 Meta-analysis of *JAK2* rs10974944 on MPNs risk (continued)

***JAK2* snp, MAF**	**Study (no. of case/control)**	**FE OR [95% CI], *P***	**Test of interaction, Z**	**het: I^2, *P***	**Remarks**
3.Combined-MPNs and controls [7, 9]	11 (2745; 5036)	2.24 [2.06-2.43], P<0.00001		I^2= 49%, P=0.0300	
Studies with only PV/ET/PMF removed [7]	8 (2625; 4844)	2.23 [2.04-2.43], P<0.00001	[7] Z=0.0729; *P*=0.9419	I^2= 26%, P=0.2200	1 ET; 2 PV studies
Asians [8]	5 (1004; 1068)	2.05 [1.79-2.35], P<0.00001		I^2= 65%, P=0.0200	
Caucasians [8]	6 (1741; 3968)	2.36 [2.12-2.63], P<0.00001	[8] Z=1.5898; *P*=0.1119	I^2= 10%, P=0.3500	
The heaviest study removed [9]	10 (2426; 2037)	2.15 [1.95-2.37], P<0.00001	[9] Z=0.6289; *P*=0.5294	I^2= 46%, P=0.0600	Kilpivaara 2009
4. V617F-pos and V617F-neg MPNs [10]	9 (1802; 893)	2. 11 [1.87-2.37], *P*<0.00001		I^2= 35%, *P*=0.1400	
The heaviest study removed [10]	6 (1321; 691)	2. 16 [1.89-2.48], P<0.00001	[10] Z=-0.2547; *P*=0.7990	I^2= 40%, *P*=0. 1100	Olcaydu 2011
Asians [11]	4 (626; 359)	2. 19 [1.80-2.66], *P*<0.00001		I^2= 45%, *P*=0.1400	
Caucasians [11]	5 (1176; 534)	2.06 [1.76-2.39], *P*<0.00001	[11] *Z*=-0.4835; *P*=0.6287	I^2= 38%, *P*=0.1600	

MAF indicates minor allele frequency for Han Chinese in Bejing; FE, fixed-effects model; OR, odds ratio; 95 % CI, 95% confidence interval; *P*, *P* value; het, heterogeneity under FE model; I^2, heterogeneity statistical tests; MPNs, Myeloproliferative Neoplasms (classic: PV, ET, MF); V617F-pos, V617F-positive; PV, polycythaemia vera; ET, essential thrombocythaemia; PMF, primary myelofibrosis; V617F-neg, V617F-negative.

[1, 2, 3…11] Z-test comparing ORs between two subgroups with the same number.

5.3.4.2 Association of *JAK2* rs2343867 and MPNs

Overall, there was evidence of an association between the risk allele (C) of rs12343867 (S12) and V617F-positive and -negative MPNs and also the combined group of V617F-positive and -negative MPNs when all the eligible studies were pooled into the meta-analysis. The minor allele C was significantly associated with V617F-positive MPNs (OR 2.42; 95% CI: 2.18-2.70, P<0.00001; I^2= 69%, P=0.0002) and in lesser extent with V617F-negative MPNs (OR 1.52; 95% CI: 1.33-1.73, P<0.00001; I^2= 76%, P<0.00001) when compared with controls. When V617F-positive and -negative patients were combined as overall MPNs and compared with controls, the OR was in between that of V617F-positive and V617F-negative groups (**Figure 5.4 A-C**). To know if the association is different between V617F-positive and -negative MPNs, V617F-positive MPN patients were compared with V617F-negative MPN patients. The result indicated a significant difference between wild type V617F and mutated V617F (P<0.00001) (**Table 5.7; Figure 5.6**). Significant between-study heterogeneity (I^2> 50%) was observed in the analyses, thus both FE and RE models were applied in analysis since more than 10 studies were included. The results in FE and RE models were only differed in which FE model provided slightly narrower 95% CI; otherwise the effects were similar (**Figure 5.5 A-C**). This suggests high-stability of the results in this meta-analysis.

Analyses were further stratified by MPN subtype (PV, ET or PMF) verses controls among eligible studies. In the comparisons of V617F-positive verses controls, similar significant associations were observed for ET (OR 2.45; 95% CI: 2.09-2.88, P<0.00001; I^2= 0%, P=0.45) and PMF (OR 2.51; 95% CI: 1.75-3.61, P<0.00001; I^2= 0%, P=0.65), as in the analysis without subtype stratification. Meanwhile an in-

creased risk was detected in PV (OR 2.97; 95% CI: 2.46-3.60, $P<0.00001$; $I^2=65\%$, $P=0.03$). Significant heterogeneity ($I^2=65\%$) was observed in PV therefore analyses were repeated with the removal of one outlier study as indicated by funnel plot. Upon removal of outlier study, the OR increased by more than 1 fold (**Table 5.7**).

When studies were stratified for ethnicity, disease risks were significantly higher (Z-test: $P<0.05$) among Asians in all subgroups: V617F-positive MPNs verses controls, V617F-negative MPNs verses controls, and combined MPNs verses controls (**Table 5.7**). Overall, test of heterogeneity suggested insignificant inconsistency among the studies.

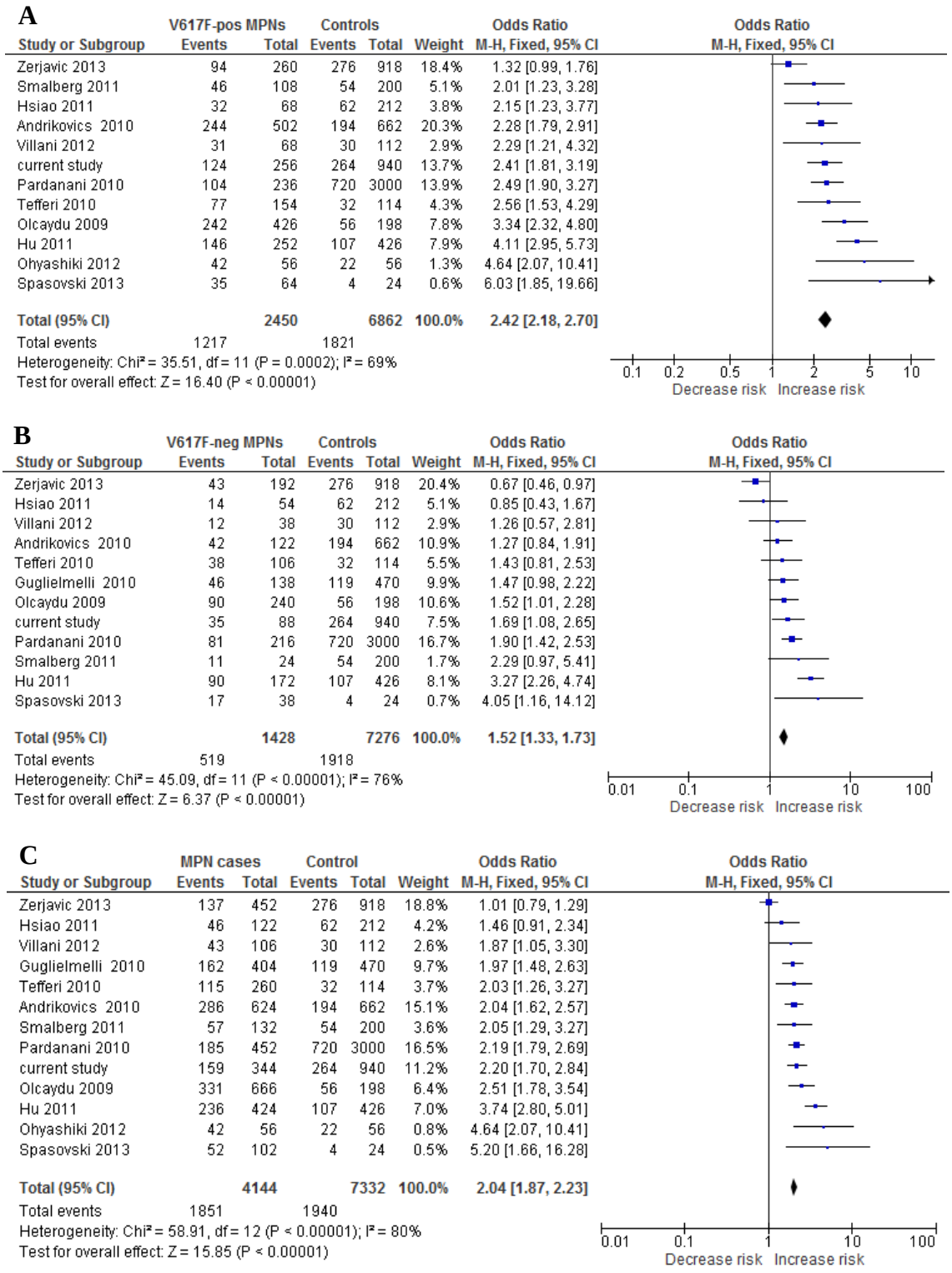


Figure 5.4 Forest plots of *JAK2* rs12343867 and MPN risk under fixed-effects model.

A. V617F-positive MPN cases verses controls, B. V617F-negative MPN cases verses controls, C. Overall MPN cases verses controls

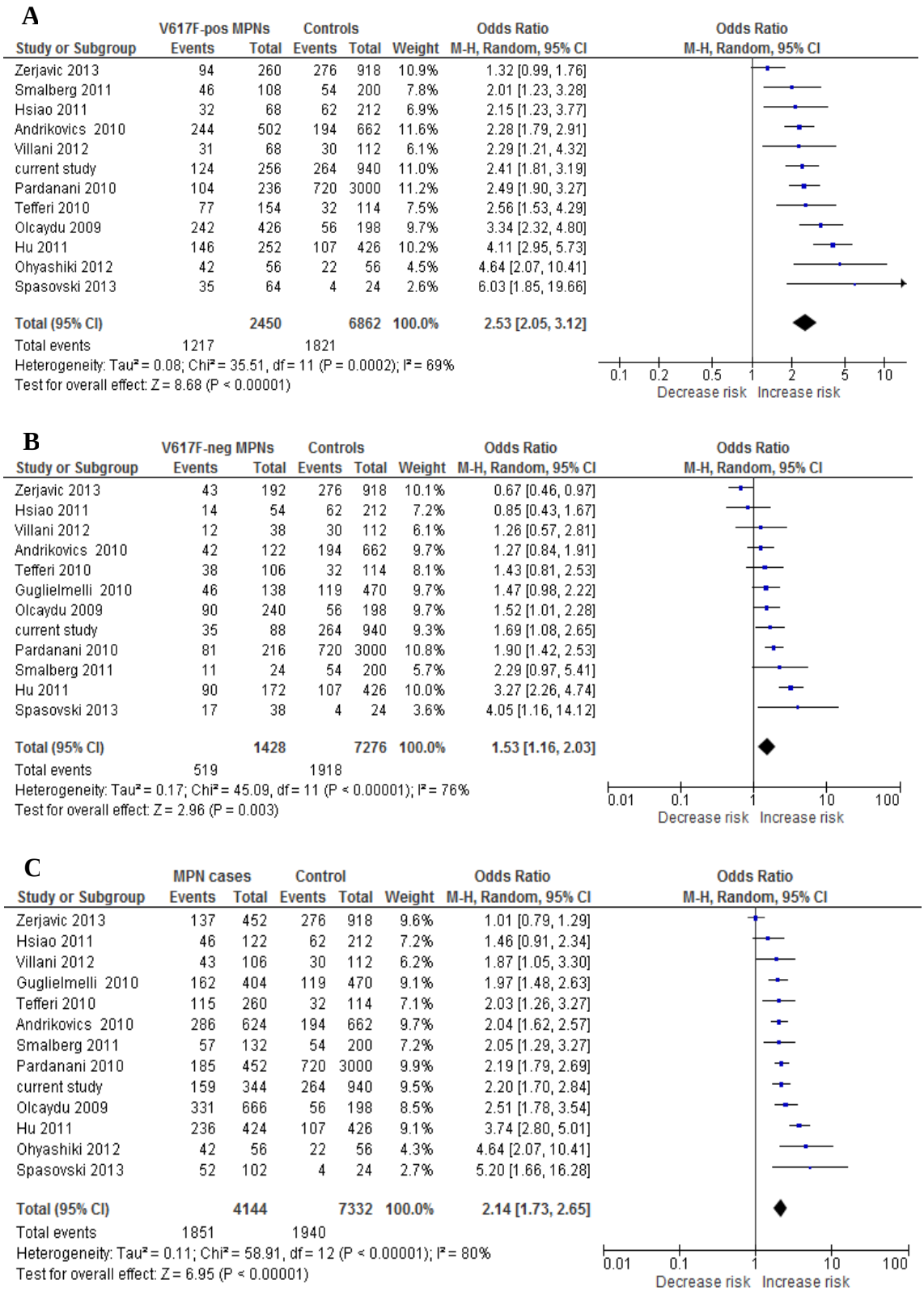


Figure 5.5 Forest plots of *JAK2* rs12343867 and MPN risk under random-effects model.

A. V617F-positive MPN cases verses controls, B. V617F-negative MPN cases verses controls, C. Overall MPN cases verses controls

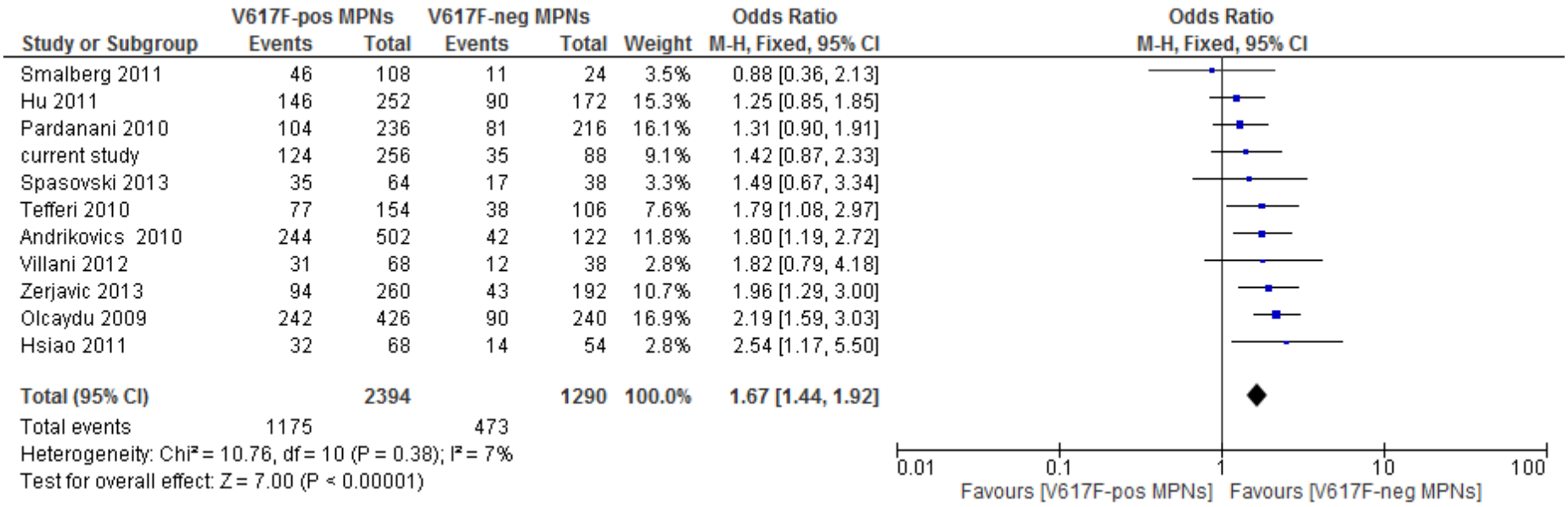


Figure 5.6 Forest plot of *JAK2* rs12343867 and MPN risk: MPN V617F-positive MPNs patients verses V617F-negative.

Table 5.7 Meta-analysis of *JAK2* rs2343867 on MPNs risk.

***JAK2* snp, MAF**	**Study (no. of case/control)**	**FE OR [95% CI], *P***	**Test of interaction, Z**	**het: I^2, *P***	**Remarks**
***JAK2* rs12343867, allele C=0.20**					
1. V617F-pos MPNs and controls [1, 2, 6]	12 (1225; 3431)	2.42 [2.18-2.70], *P*<0.00001 2.53 [2.05-3.12], *P*<0.0001*		I^2= 69%, *P*=0. 0002	
Outliers removed (n=2) [1]	10 (969; 2982)	2.52 [2.23-2.85], *P*<0.00001	[1] *Z*=-0.4876; *P*=0.6258	I^2= 0%, *P*=0.4800	Hu 2011; Zerjavic 2013
Studies with only PV/ET/PMF removed [2]	8 (781; 1213)	2.70 [2.35-3.10], P<0.00001	[2] *Z*=-1.2263; *P*=0.2201	I^2= *65%*, *P*=0.0200	1 PV; 2 ET; 1PMF studies
Asians [3, 5]	4 (316; 817)	2.97 [2.45-3.60], *P*<0.00001	[3] *Z*=-2.4823; *P*=0.0131	I^2= 63%, *P*=0.0400	
Caucasians [3, 4]	8 (909; 2614)	2.22 [1.96-2.52], *P*<0.00001	[4] *Z*=-1.3096; *P*=0.1903	I^2= 67%, *P*=0.0030	
Caucasians: Outlier removed (n=1) [4, 5]	7(779, 2155)	2.52 [2.19-2.91], *P*<0.00001	[5] *Z*=1.3462; *P*=0.1782	I^2= 0%, *P*=0.4300	Zerjavic 2013
PV [6]	4 (313; 1042)	2.97 [2.46-3.60], *P*<0.00001		I^2= 65%, *P*=0.0300	
PV: Outlier removed (n=1) [6]	3 (160; 711)	3.71 [2.85-4.83], *P*<0.00001	[6] *Z*=-1.3404; *P*=0.1801	I^2= 15%, *P*=0.3100	Andrikovics 2010
ET	6 (359; 2676)	2.45 [2.09-2.88], *P*<0.00001		I^2= 0%, *P*=0.5600	
PMF	3 (108; 858)	2.51 [1.75-3.61], *P*<0.00001		I^2= 0%, *P*=0.6500	
The heaviest study removed [7]	11 (974; 3100)	2.46 [2.19-2.77], *P*<0.00001 2.58 [2.03-3.30], *P*<0.00001*	[7] *Z*=-0.2022; *P*=0.8398	I^2= 72%, *P*=0. 0001	Andrikovics 2010

Table 5.7 Meta-analysis of *JAK2* rs2343867 on MPNs risk. (continued)

***JAK2* snp, MAF**	**Study (no. of case/control)**	**FE OR [95% CI], *P***	**Test of interaction, Z**	**het: I^2, *P***	**Remarks**
***JAK2* rs12343867, allele C=0.20**					
2. V617F-neg MPNs and controls [8, 9 12]	12 (714; 3638)	1.52 [1.33-1.73], *P*<0.00001 1.63 [1.30-2.05], *P*<0.0001*		I^2= 76%, *P*<0.00001	
Outliers removed (n=2) [8]	10 (532; 2966)	1.56 [1.34-1.82], *P*<0.00001	[8] *Z*=-0.2523; *P*=0.8008	I^2= *5%*, *P*=0.3900	Hu 2011; Zejavic2013
Studies with only PV/ET/PMF removed [9]	8 (457; 1740)	1.49 [1.26-1.75], *P*<0.00001	[9] *Z*=0.1857; *P*=0.8527	I^2= 82%, *P*<0.00001	2 ET; 2PMF studies
Asians [10, 12]	3 (157; 789)	2.10 [1.62-2.71], *P*<0.00001	[10] *Z*=-2.8154; *P*=0.0049	I^2= 85%, *P*=0. 0001	
Caucasians [10, 11]	9 (557; 2849)	1.37 [1.18-1.59], *P*<0.0001	[11]*Z*=-1.4123; *P*=0.1579	I^2= 67%, *P*=0. 0020	
Caucasians: Outlier removed (n=1) [11, 12]	8 (461; 2390)	1.61 [1.36-1.90], *P*<0.00001	[12] *Z*=-1.6974; *P*=0.0896	I^2= 0%, *P*=0.5300	Zerjavic2013
ET	7 (324; 2733)	1.53 [1.28-1.83], *P*<0.00001		I^2= 9%, *P*=0.3600	
PMF	5 (149; 1149)	1.43 [1.07-1.92], *P*=0.02		I^2= 0%, *P*=0.8000	
The heaviest study removed [13]	11 (618; 3179)	1.74 [1. 51-1.99], *P*<0.00001 1.69 [1. 35-2.12], *P*<0.00001*	[13] *Z*=-1.3900; *P*=0.1645	I^2= 55%, *P*=0.0100	Zerjavic2013

Table 5.7 Meta-analysis of *JAK2* rs2343867 on MPNs risk. (continued)

***JAK2* snp, MAF**	**Study (no. of case/control)**	**FE OR [95% CI], *P***	**Test of interaction, Z**	**het: I^2, *P***	**Remarks**
***JAK2* rs12343867, allele C=0.20**					
3. Combined-MPNs and controls [14, 15, 18]	13 (2072; 3666)	2.04 [1.87-2.23], *P*<0.00001 2.14[1.73-2.65], *P*<0.00001*		I^2= 80%, *P*<0.00001	
Outliers removed (n=2) [14]	11 (1634; 2994)	2.14 [1.93-2.37], P<0.00001	[14] *Z*=-0.6935; *P*=0.488	I^2= 0%, *P=0. 44*	Hu 2011; Zerjavic 2013
Studies with only PV/ET/PMF removed [15]	8 (1425; 1740)	2.02 [1.81-2.25], *P*<0.00001 2.18[1.56-3.03], *P*<0.00001*	[15] *Z*=0.1380; *P*=0.8902	I^2= 87%, *P*<0.00001	1 PV; 2 ET; 2PMF studies
Asians [16, 18]	4 (473; 817)	2.62 [2.21-3.11], *P*<0.00001	[16] *Z*=-3.3019; *P*=0.0010	I^2= 80%, *P*=0.002	
Caucasians [16, 17]	9 (1599; 2849)	1.87 [1.68-2. 07], *P*<0.00001	[17] *Z*=-3.3019; *P*=0.0010	I^2= 76%, *P*<0.00001	
Caucasians: Outlier removed (n=1) [17, 18]	8 (1373; 2390)	2.14 [1.91-2. 40], *P*<0.00001	[18] *Z*=1.9305; *P*=0.0535	I^2= 0%, *P*=0.79	Zerjavic 2013
The heaviest study removed [19]	12 (1846; 3207)	2.28 [2.07-2.51], *P*<0.00001 2.29 [1.97-2. 66], *P*<0.00001*	[19] Z=-1.6702; *P*=0.0949	I^2= *52%*, *P=0.02*	Zerjavic 2013
4. V617F-pos and V617F-neg MPNs [20]	11 (1197; 645)	1.67 [1.44-1.92], *P*<0.00001 1.66 [1.43-1.93], *P*<0.00001*		I^2= 7%, *P*=0.28	
Asian population [21]	3 (288; 157)	1.44 [1.09-1.92], *P*=0.01		I^2= 22%, *P*=0.24	
Caucasian population [21]	8 (909; 488)	1.75 [1.48-2.06], *P*<0.00001	[21] *Z*=1.1657; *P*=0.2437	I^2= 0%, *P*=0.44	
The heaviest study removed [20]	10 (984; 525)	1.56 [1.33-1.83], *P*<0.00001	[20] *Z*=0.6217; *P*=0.5341	I^2= 0%, *P*=0.61	Olcaydu 2009

MAF indicates minor allele frequency for Han Chinese in Bejing; FE, fixed-effects model; OR, odds ratio; 95 % CI, 95% confidence interval; *P*, *P* value; het, heterogeneity under FE model; I^2, heterogeneity statistical tests; MPNs, Myeloproliferative Neoplasms (classic: PV, ET, MF); V617F-pos, V617F-positive; PV, polycythaemia vera; ET, essential thrombocythaemia; PMF, primary myelofibrosis; V617F-neg, V617F-negative.

* Effect estimate in random model

[1, 2, 3…21] Z- test comparing ORs (based on FE model) between two subgroups.

5.3.4.3 Association of *JAK2* rs12340895 and MPNs

As for rs12340895 (S13), significant between-study heterogeneity (I^2> 50%) was observed however only FE was applied in the analyses considering the meta-analysis containing less than 10 studies. Overall, the risk allele (G) of rs12340895 (S13) was found to be significantly associated with V617F-positive MPNs and the combined group of V617F-positve and -negative MPNs except for V617F-negative MPNs when compare with controls (**Table 5.8**). The minor allele G was significantly associated with V617F-positive MPNs (OR 2.47; 95% CI: 2.26-2.70, P<0.00001; I^2=93%, P<0.00001) and in lesser extent with the combined group of V617F-positive and -negative MPNs when compared with controls (OR 1.90; 95% CI: 1.76-2.05, P<0.00001; I^2=95%, P<0.00001). However, no significant association was found for analyses under the subgroup of V617F-negative MPNs patients verses controls, except for the subtype ET (OR 1.11; 95% CI: 0.97-1.26, P=0.12; I^2= 45%, P=0.12) (**Figure 5.7**). The comparison between V617F-positive MPNs and V617F-negative MPNs indicated an almost 2-fold difference between wild type V617F and mutated V617F (P<0.00001) (**Table 5.8; Figure 5.9**).

Analyses were further stratified by MPN subtype (PV, ET or PMF) verses controls among eligible studies. In the comparisons of V617F-positive verses controls, significant associations were detected for PV (OR 4.11; 95% CI: 3.45-4.89, P<0.00001; I^2=55%, P=0.11), ET (OR 1.84; 95% CI: 1.63-2.07, P<0.00001; I^2= 71%, P=0.007) and PMF (OR 2.46; 95% CI: 1.95-3.11, P<0.00001; I^2=6%, P=0.36). Significant heterogeneity (I^2>50%) was observed among ET and PV therefore analyses were repeated by removing the outlier study as indicated by funnel plot. Upon removal of outlier study, significant risk was found increased by 1-fold for PV subgroup. While for ET subgroup, similar OR with PMF subgroup was found (OR 2.43; 95% CI:

2.01-2.94, $P<0.00001$; $I^2=0\%$, $P=0.96$). Similar results were concluded for the ET- and PMF-subgroup analyses (**Table 5.8**).

When studies were stratified for ethnicity, disease risks were significantly higher (Z-test: $P<0.05$) among Asian populations in all subgroups: V617F-positive MPNs verses controls, V617F-negative MPNs verses controls, and combined MPNs verses controls (**Table 5.8**).

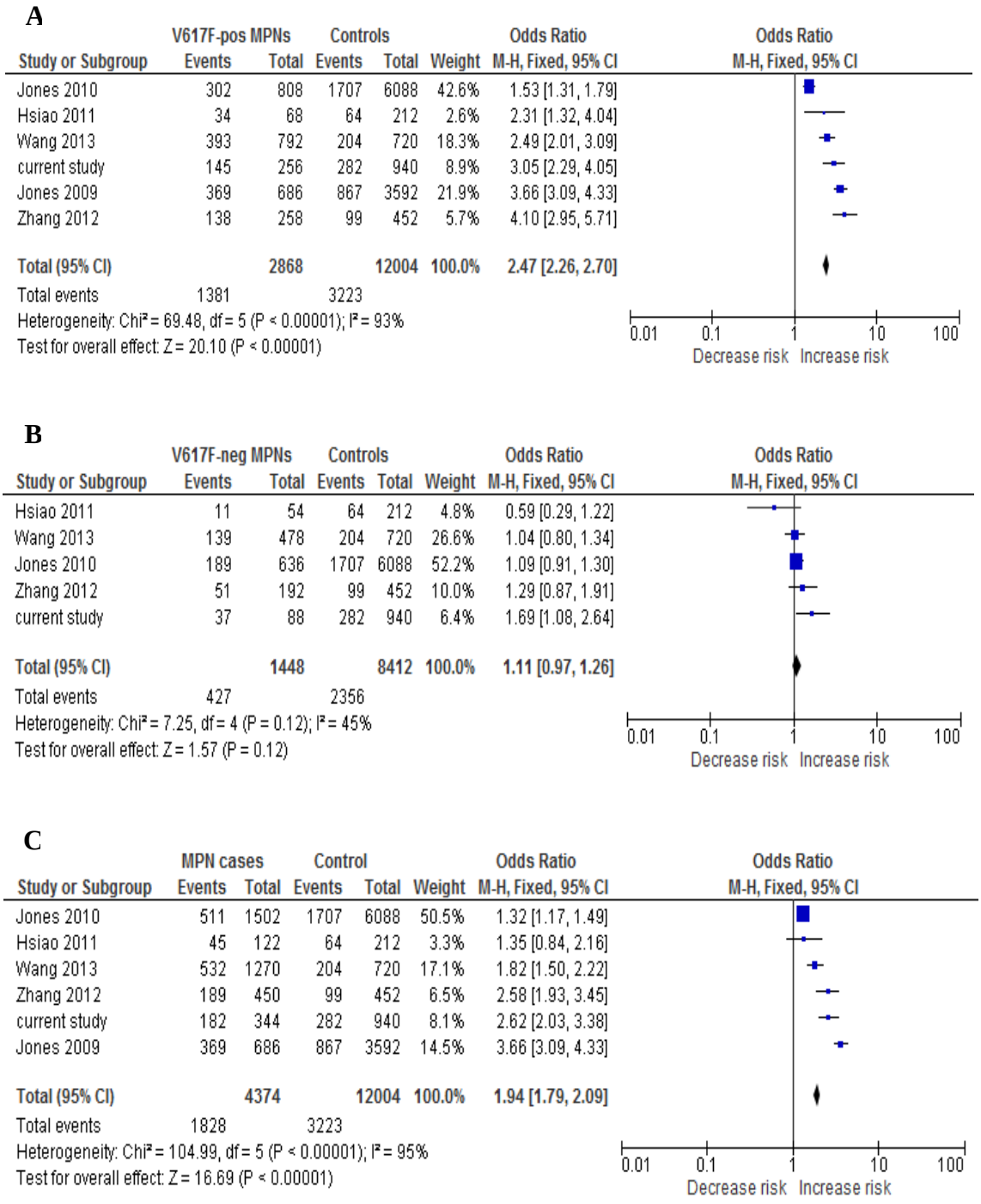


Figure 5.7 Meta-analysis for the association between rs12340895 and MPN risk under fixed-effects model.

A. V617F-positive MPN cases verses controls, B. V617F-negative MPN cases verses controls, C. Overall MPN cases verses controls

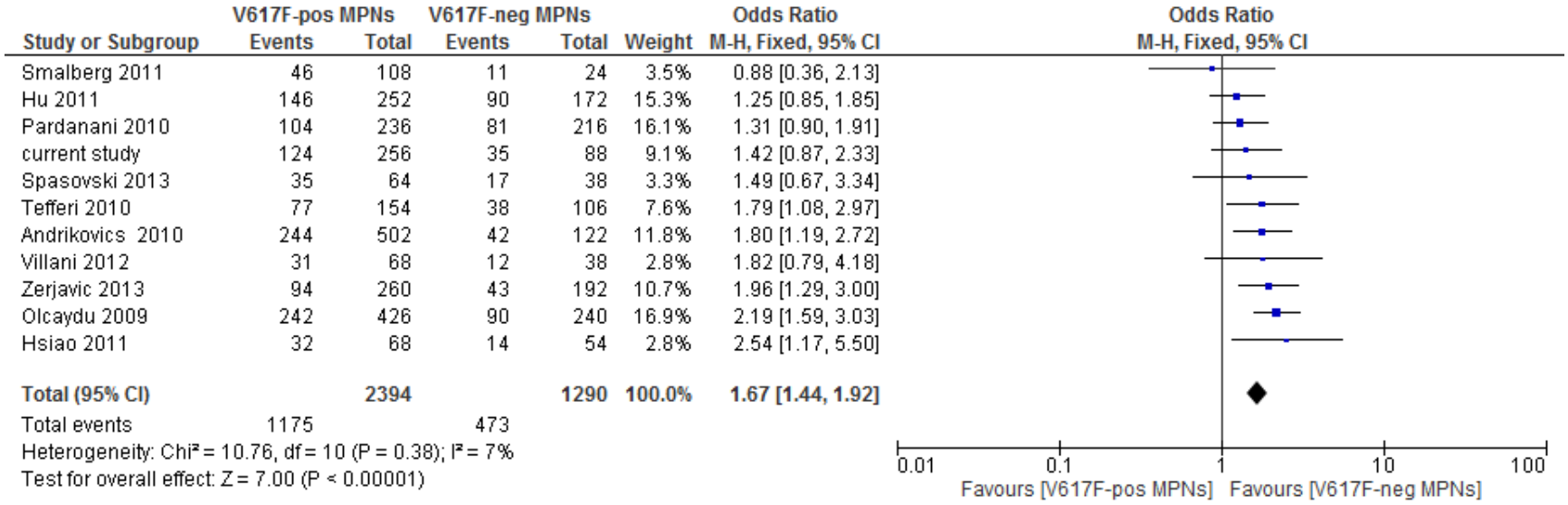


Figure 5.8 Forest plot of *JAK2* rs12340895 and MPN risk: MPN V617F-positive MPNs patients verses V617F-negative.

Table 5.8 Meta-analysis of *JAK2* rs12340895 on MPNs risk.

***JAK2* snp, MAF**	**Study (no. of case/control)**	**FE OR [95% CI], *P***	**Test of interaction, Z**	**het: I^2, *P***	**Remarks**
***JAK2* rs12340895, allele G=0.20**					
1. V617F-pos MPNs and controls [1, 2, 6]	6 (1434; 6002)	2.47 [2. 26-2.70], *P*<0.00001		I^2= 93%, P<0.00001	
Three outliers removed [1]	3 (558; 936)	2.64 [2. 24-3.11], P<0.00001	[1] *Z*=-0.6990; *P*=0.4846	*I^2= 0%, P=0. 48*	Jones 2009; Jones 2010; Zhang 2012
Studies with only PV/ET/PMF removed [2]	4 (996; 2852)	3.21 [2.87-3.60], *P*<0.00001	[2] *Z*=-3.5656; *P*=0.0004	I^2= 70%, P=0.02	2 ET studies
Asians [3]	4 (687; 1162)	2.88 [2.48-3.33], *P*<0.00001		I^2= 56%, P=0. 08	
Caucasians [3]	2 (747; 4840)	2.25 [2.02-2.52], *P*<0.00001	[3] *Z*=-2.6435; *P*=0.0082	I^2= 98%, *P*<0.00001	
PV [4]	3 (305; 2492)	4.11 [3.45-4. 89], *P*<0.00001		I^2= 55%, *P*=0. 11	
PV: Outlier removed (n=1) [4]	2 (237; 2266)	3.46 [3.05-3.91], *P*<0.00001	[4] *Z*=1.5759; *P*=0.1150	I^2= 0%, *P*=0. 45	Zhang 2012
ET [5]	5 (694; 4206)	1.84 [1.63-2.07], *P*<0.00001		I^2= 71%, *P*=0. 007	
ET: Outlier removed (n=1) [5]	4 (290; 1162)	2.43 [2.01-2.94], *P*<0.00001	[5] *Z*=-2.4275; *P*=0.0152	I^2= 0%, *P*=0. 96	Jones 2010
PMF	4 (170; 2852)	2.46 [1.95-3.11], *P*<0.00001		I^2= 6%, *P*=0. 36	
The heaviest study removed [6]	5 (918; 2958)	3.06 [2.72-3.44], *P*<0.00001	[6] *Z*=-2.8501; *P*=0.0044	I^2= 56%, *P*=0.06	1 ET study
2. V617F-neg MPNs and controls [7, 9]	5 (724; 4206)	1.11 [0.97-1.26], *P*=0.12		I^2= 45%, P=0. 12	
Studies with only PV/ET/PMF removed [7]	3 (379; 1056)	1.19 [0.98-1.45], *P*=0.07	[7] *Z*=-0.5791; *P*=0.5625	I^2= 46%, P=0.16	2 ET studies
Asians [8]	4 (406; 1162)	1.13 [0.94-1. 37], *P*=0.18		I^2= 58%, P=0. 07	
Caucasians [8]	1 (318; 3044)	1.09 [0.91-1. 30], *P*=0.37	[8] *Z*=-0.2723; *P*=0.7854	*Not applicable*	
ET	5 (373; 4206)	1.26 [1.07-1. 49], *P*=0.005		I^2= 35%, *P*=0. 19	
PMF	3 (126; 1056)	1.10 [0.81-1.49], *P*=0.55		I^2= 0%, *P*=0. 49	
The heaviest study removed [9]	5 (406; 1162)	1.13 [0.94-1.37], *P*=0.18	[9] *Z*=-0.1526; *P*=0.8787	I^2= 58%, *P*=0. 07	Jones 2010

Table 5.8 Meta-analysis of *JAK2* rs12340895 on MPNs risk (continued)

***JAK2* snp, MAF**	**Study (no. of case/control)**	**FE OR [95% CI], *P***	**Test of interaction, Z**	**het: I^2, *P***	**Remarks**
***JAK2* rs12340895, allele G=0.20**					
3. Combined-MPNs and controls [10, 12]	6 (2187; 6002)	1.90 [1.76-2.05], *P*<0.00001		I^2= 95%, *P*<0.00001	
Studies with only PV/ET/PMF removed [10]	4 (1375; 2852)	2.65 [2.38-2. 94], *P*<0.00001	[10] *Z*=-5.0045; *P*~0	I^2= 89%, *P*<0.00001	2 ET studies
Asians [11]	4 (1093; 1162)	2.10 [1.84-2.40], *P*<0.00001		I^2= 71%, *P*=0.0200	
Caucasians [11]	2 (1094; 4840)	1.85[1.68-2.03], *P*<0.00001	[11]*Z*=-1.5232; *P*=0.1277	I^2= 99%, *P*<0.00001	
The heaviest study removed [12]	5 (1436; 2958)	2.56 [2.31-2.84], *P*<0.00001	[12]*Z*=-4.5518; *P*=5.413E-06	I^2= 89%, *P*<0.00001	Jones 2010
4. V617F-pos and V617F-neg MPNs [13, 15]	5 (1091; 724)	1.99 [1.72-2.29], *P*<0.00001		I^2= 80%, *P*=0.0006	
Outlier removed (n=1) [13]	4 (687; 406)	2.51 [2.09-3.02], *P*<0.00001	[13]*Z*=-1.9517; *P*=0.0510	I^2= 31%, *P*=0.2300	Jones 2010
Asian population [14]	4 (687; 406)	2.51 [2.09-3.02], *P*<0.00001		I^2= 31%, *P*=0.2300	
Caucasian population [14]	1 (404; 318)	1.41 [1.13-1. 76], *P*=0.002	[14] *Z*=3.9244; *P*=0.0001	*Not applicable*	
The heaviest study removed [15]	4 (687; 406)	2.51 [2.09-3.02], *P*<0.00001	[15]*Z*=-1.9517; *P*=0.0510	I^2= 31%, *P*=0.2300	Jones 2010

MAF indicates minor allele frequency for Han Chinese in Bejing; FE, fixed-effects model; OR, odds ratio; 95 % CI, 95% confidence interval; *P*, *P* value; het, heterogeneity under FE model; I^2, heterogeneity statistical tests; MPNs, Myeloproliferative Neoplasms (classic: PV, ET, MF); V617F-pos, V617F-positive; PV, polycythaemia vera; ET, essential thrombocythaemia; PMF, primary myelofibrosis; V617F-neg, V617F-negative.

[1, 2, 3…15] Z-test comparing ORs between two subgroups

5.3.6 Sensitivity analyses and publication bias

To determine whether modification of the inclusion criteria affected the final effect pooled ORs, sensitivity analyses were conducted. First, the magnitude of effect was re-estimated with the removal of the most influential study (indicated by percent weights given to each study). Then, funnel plots were generated to graphically estimate the publication bias among the analyses. The asymmetry of funnel plots based on natural logarithm of the OR was measured. Consequently, studies represented by dots scattered outside were removed to determine the influence of individual study on the pooled effect estimate. Considering substantial heterogeneity in rs12343867 and rs12340895, sensitivity analyses were performed with the removal of individual outlier study (indicated by funnel plot) to test the robustness and accuracy of this meta-analysis. Heterogeneity was completely otherwise almost completely diminished as indicated by the I^2 upon the removal of skewed studies.

5.3.6.1 rs10974944

When the heaviest study as indicated by sample size was removed, similar pooled ORs were obtained before and after the removal of the heaviest study as indicated by *Z*-test (**Table 5.6**), suggesting stability of the results. Funnel plots of the studies did not reveal any evidence of obvious asymmetry (data not shown). The removal of MPN subtype (PV, ET or PMF) groups did not reveal significant difference on the summary effect estimates (**Table 5.6**).

5.3.6.2 rs12343867

Omission of the heaviest study as indicated by sample size revealed that the pooled ORs remained similar after the removal of the heaviest study (indicated by *Z*-test in **Table 5.7**), resembling high stability of the results.

Considering substantial heterogeneity detected in subgroup analyses of rs12343867, sensitivity analyses were conducted to identify single source of heterogeneity on the pooled ORs by sequentially removing each eligible study. Omission of one individual study at a time did not reveal single study that heavily influenced the pooled effect estimates. However, funnel plots detected outliers in the subgroup analyses; and heterogeneity was completely or almost completely diminished when outliers were omitted (**Figure 5.9**). Nevertheless the summary estimates were not significantly affected in the subgroup analyses as indicated by *Z*-test (**Table 5.7**). The removal of MPN subtype (PV, ET or PMF) groups also did not reveal significant difference on the summary effect estimates (**Table 5.7**).

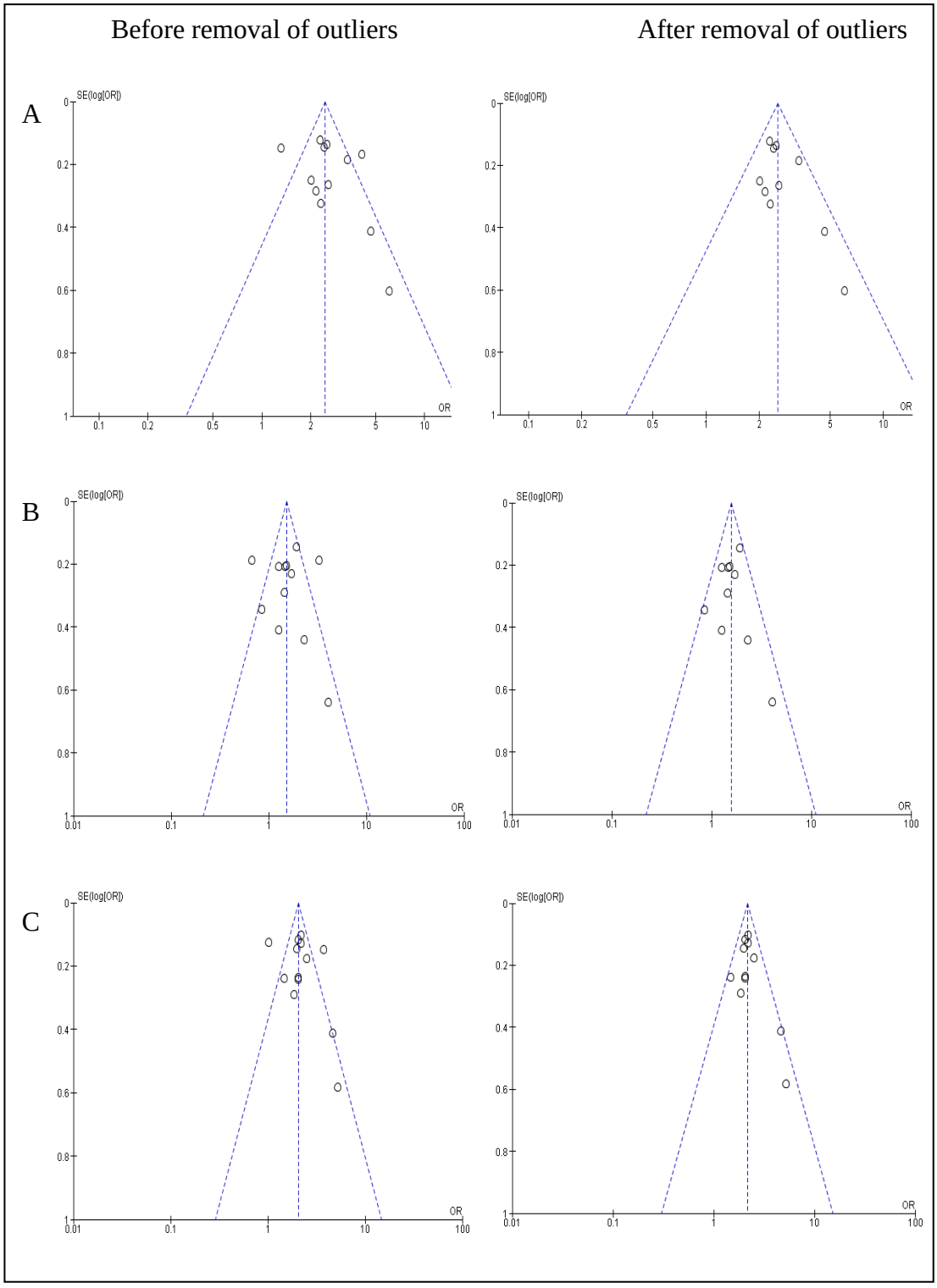


Figure 5.9 Funnel plots for the association between rs12343867 and MPN risk.

A. V617F-positive MPN cases verses controls, B. V617F-negative MPN cases verses controls, C. Overall MPN cases verses controls

5.3.6.3 rs12340895

Omission of the heaviest study as indicated by sample size revealed that the pooled ORs increased significantly (Z-test: $P<0.05$) among V617F-positive and the combined MPNs when the removed study was the same ET subgroup (Jones et al., 2010) (**Table 5.8**). While for the V617F-negative MPNs verses controls subgroup, the pooled OR was still not significant: 1. 13 (95% CI: 0.94-1.37, $P=0.18$). Overall, the trend of MPNs predisposition among the subgroups was consistent.

To identify single source of existing heterogeneity on the pooled ORs, outlier dataset as indicated by funnel plot was removed. Funnel plots revealed the existence of 3 outliers in the V617F-positive MPNs subgroup. Heterogeneity was completely diminished when outliers were omitted (**Table 5.8, Figure 5.10**). However, the pooled ORs before and after the removal of outlier datasets were not significantly different. The removal of MPN subtype (PV, ET or PMF) groups revealed an increase in the pooled ORs among V617F-positive MPNs, V617F-negative MPN and overall MPNs (**Table 5.8**) compared with the results before stratification.

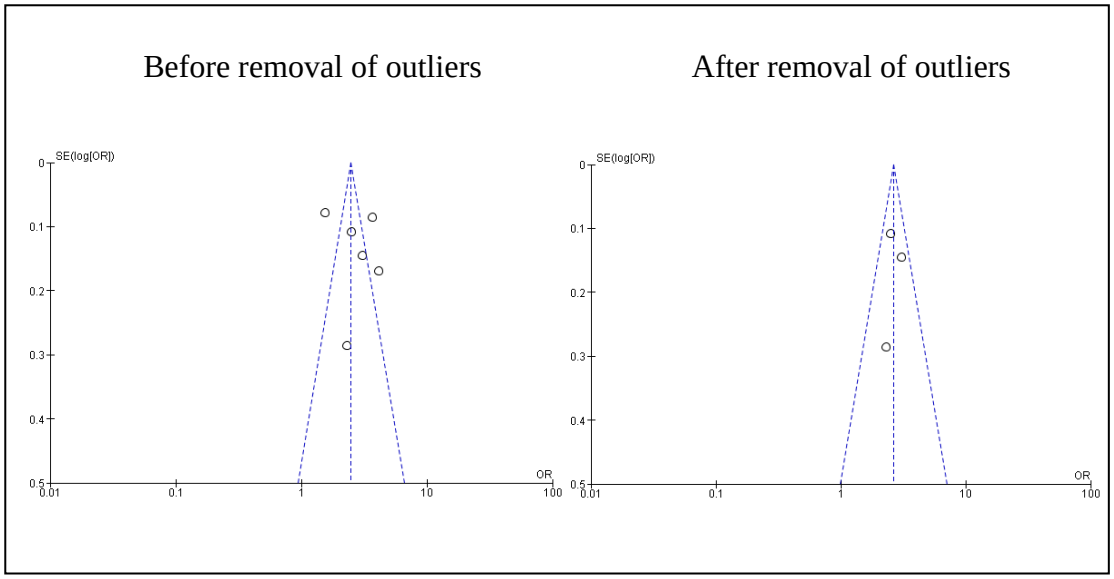


Figure 5.9 Funnel plots for the association between rs12340895 and MPN risk: V617F-positive MPN cases verses controls.

5.4 Discussion

Genetic association study of *JAK2* polymorphisms (part I) of this project demonstrated the *JAK2* haplotype tSNPs (rs10974944 (S9), rs12343867 (S12), and rs12340895 (S13)) are risk factors in developing V617F-positive MPNs. However, the literature on the relationship between *JAK2* germline polymorphisms (*JAK2* 46/1 or GGCC haplotype) and V617F-negative MPNs is replete with small studies that report contradictory findings. Inadequate statistical power or small sample size, variations in the study designs and diagnostic criteria are few factors that could lead to inconsistent findings. The lack of studies of ethnicity backgrounds may also cause deviation to the final effect estimates and fail to give further insights on this issue. Moreover, the susceptibility loci associated with complex diseases are usually modest-risk alleles with complicated allelic architectures (Pritchard & Cox, 2002; Manolio et al., 2009). Greater statistical power (large sample size) is needed to detect those loci. Within my study period, no clear consensus has been reached and there is no systematic review. Large sample size and unbiased genetic studies of predisposition polymorphisms across populations could provide insight to the aetiology of MPNs. Therefore, meta-analysis was conducted to give further insights on the genetic predisposition to MPNs. In this study, 23 studies were included, dating back from 1 January 2005 to 4 September 2014. Studies were retrieved from PubMed, Scopus and Chinese National Knowledge Infrastructure (CNKI) databases without language restriction. Data were meta-analysed by Cochrane RevMan software. Individual studies were pooled into single meta-analysis to reduce the risk of random error, and to statistically detect the real effect/ association between *JAK2* haplotype with MPNs across populations by increasing the analytical power.

5.4.1 Association between *JAK2* polymorphisms and MPNs

Meta-analysis for 6 SNPs (rs3808850, rs7046736, rs10815148, rs12342421, rs3780367, and rs1159782) in which only 2 studies were available for meta-analysis detected an association with MPN for 5 SNPs except rs3808850, SNP located on *JAK2* 5' upstream (OR 1.03, 95%CI, 0.83-1.28; P=0.77). This probably was due to the contrary minor allele in Han Chinese (T, MAF 0.48) and Caucasian (A, MAF 0.32). Among the 6 SNPs, the highest pooled OR was detected for rs12342421 when V617F-positive MPN cases were compared with control. Nevertheless, significant heterogeneity was observed. Though significant associations were observed, the results should be interpreted cautiously in these analyses with small sample sizes.

Meta-analysis of *JAK2* haplotype tSNPs rs10974944 (S9) and rs12343867 (S12) revealed that both V617F-negative and -positive MPNs were significantly associated with *JAK2* risk polymorphisms. Results showed that the disease association for the 2 SNPs was ≥1 fold higher in V617F-positive MPNs than V167F-negative MPNs. Thus, V617F-negative MPN and V617F-positive MPN patients might share similar genetic background although wild-type *JAK2* seems to dilute the magnitude of disease predisposition. A high degree of correlation (r^2>0.8), was observed among the *JAK2* SNPs as presented in their pairwise LD (**Figure 4.1** and **4.2**). This could explain the commensurate trends in meta-analyses for S9 and S12.

In the meta-analysis for rs10974944 (S9) and rs12343867 (S12), significant association can be found across the entire group of V617F-positive, V617F-negative, and the combined V617F-positive and -negative MPNs except in two for the stratification analysis for rs10974944 (S9) under V617F-negative subgroup: Asian population and

ET subtype analysis. However, the effect difference between Asian and Caucasian populations was not significant (Z-test, P=0.2928; **Table 5.6**). As proposed earlier, wild-type *JAK2* seems to dilute the effect magnitude of disease predisposition. The sample size for Asian subgroup (cases=359) was smaller than in Caucasian population (cases=561), thus suggesting that a larger sample size with stronger statistical power is demanded to draw a fair conclusion. As for rs12343867 (S12), a Slovenian study, Zerjavic et al. (2013), was detected to be the single source of heterogeneity among studies of Caucasian population, analyses were therefore repeated by removing this study. Removal of the outlier did not affect the result. The substantial heterogeneity introduced by this study might be explained. First, different genotyping methods and errors may introduce heterogeneity. The AS-PCR genotyping method used in their study was well-established and widely used in MPN research therefore genotyping method was not the cause. Based on a report of short tandem repeat loci by Pajnic et al. (2014), the genetic make-up between Slovenian and other European population is not expected to differ significantly. The results suggested that the MPN diagnostic criteria used which were not WHO criteria might contribute to the outlier effect. Besides, environmental factors, lifestyle and other unknown factors may also be the source of heterogeneity. Overall, the association detected in Asian and Caucasian population was not significantly different as suggested by test of interaction (Z-test).

Conversely, no association was detected between rs12340895 (S13) and V617F-negative MPNs. However, when stratified by subtype of MPN, marginal association between rs12340895 (S13) and risk of developing V617F-negative ET was found. Moderate between-study variation was detected ($I^2 \geq 45\%$; **Table 5.11**) in the subgroup analysis however neither single source nor combined-studies were found to

contribute to the heterogeneity. In fact, when the analyses were stratified for V617F-negativity, only 5 studies were included in the analyses of S13, compared with 9 and 12 studies for S9 and S12 respectively. Small number of study pool leads to an underpowered test and affects the stability of the pooled outcomes, thus the results should be interpreted with caution (Higgins & Green). Moreover, effect of the causal genetic variants may be masked by the presence of other unidentified causal variants involved in MPNs' pathogenesis. Thus, more studies are demanded. Since MPNs are a group of complex disorder involving not only various genetic and epigenetic alterations (Abdel-Wahab, 2011) but also environmental, and ethnic factors (Anderson et al., 2012), it is likely that other risk factors of MPNs also interact in different pathways.

5.4.2 Insignificant effect of heterogeneity and publication bias

In sensitivity analysis, outliers that have the largest influence on overall effect were removed owing to the absence of single source of heterogeneity. Results showed that the pooled ORs before and after the removal of outliers were not different significantly, indicating the high reliability of the results. Since significant association was detected among Asian and Caucasian V617F-negative MPN patients for rs10974944 (S9) and rs12343867 (S12), chances are low that the conflicting results of genetic studies of *JAK2* can be explained by ethnicity. Studies with a relatively small sample size (smallest 34 cases and 106 controls) in a single analysis might be underpowered for a fair conclusion. Inconsistent association in V617F-negative MPNs indicate that there may be differences in the effect magnitudes of the genetic contribution to MPN susceptibility by V617F mutation. Apart from *JAK2* polymorphisms, SNPs and haplotypes from other genes should also be considered such as the MPN-related

Myeloproliferative Leukaemia Virus (*MPL*), Ten-Eleven Translocation-2 (*TET2*), Additional Sex Combs-Like 1 (*ASXL1*), among others.

5.4.2 Strengths and limitations of the study

The strengths of our study could be summarised as follows. Within my study period, it is the first study to examine the relationship between *JAK2* polymorphisms and MPNs across populations, and on the basis of a comprehensive literature search by means of various searching approaches. In this meta-analysis, publication bias was assessed and the potential source of heterogeneity was identified. In addition, the stratified analysis by ethnicity in this study covered the potential problem of different polymorphisms in different ethnics. Complex disease like MPNs might be associated with different polymorphisms in different ethnic groups. This meta-analysis aimed at solving the contradictory findings of *JAK2* polymorphisms and V617F-negative MPNs. Although outliers were identified in the analysis, the removal of those studies did not materially alter the effect estimates. The results are therefore robust to show that *JAK2* polymorphisms were associated with MPNs regardless of V617F status however; with a one-fold difference in the magnitude of MPNs predisposition. Results posited that a somehow similar genetic background was shared between *JAK2*-positive and *JAK2*-negative MPNs however the V617F mutation interacts differently in the pathogenesis or, different pathways were involved thus giving rise to the 3 clinically distinct diseases.

Certain limitations of this meta-analysis should also be acknowledged. First, age and genders were not investigated in this study because the specific data was not available. However, it was widely accepted that age and genders are not confounding vari-

ables in the genetic studies of MPNs (Pardanani et al., 2010; Tefferi et al., 2010b; Ohyashiki et al., 2012; Zhang et al., 2012a). Second, the definition of MPNs was not precisely the same in all studies, because the original authors used different criteria (WHO, bone marrow aspiration, PV study group, or medical reports). Third, the patients included in some studies could not be classified into individual MPN subtype when the original authors did not indicate the MPN subtype for those participants. These may have reduced the comparability of stratified analysis in this study. Fourth, in the stratified analysis by ethnicity, inadequate eligible studies made it impossible to access the subgroups within populations, such as Taiwan and China for Asian populations. In addition, this meta-analysis did not include SNPs from other genes and thus any risk effect estimated was only restricted to *JAK2*. It is important to analyse multiple SNPs and their interaction to find more reliable prognostic or predictive biomarkers, because it is hard to understand the complex MPNs by using only one SNP or based on one gene. Additional polymorphisms and functional studies may be required to evaluate the possibility of gene-gene or SNP-SNP interactions.

5.5 Conclusion

This meta-analysis quantitatively assessed the association between *JAK2* polymorphisms and MPNs risk based on currently available data. Within my PhD period, this is the first meta-analysis assessing the relationship between *JAK2* polymorphism and the risk of MPNs. Significant association was observed between *JAK2* risk alleles and MPNs particularly the V617F-positive cohort, and to a lesser extent the V617F-negative MPNs but the exact mechanism of how the *JAK2* polymorphisms affect MPNs is unknown. Further exploration for the true causal variant between *JAK2* polymorphisms and MPNs is demanded. Large sample clinical studies should be carried

out to verify more risk factors in order to make early detection and prevention at the gene level possible. However, sample recruitment was beyond the control of this study. International collaboration may increase the power to study and detect rare variants if any. Finer scale deep sequencing analysis may serve as an alternative to larger study in exploring the undiscovered variants which may be the real disease variants.

Part III Next-generation Sequencing

CHAPTER 6 Next-generation sequencing

6.1 Introduction

The clonal haematopoietic disorders myeloproliferative neoplasms (MPNs) include three classic yet different phenotypes namely polycythaemia vera (PV), essential thrombocytosis (ET) and primary myelofibrosis (PMF). This group of disorders is primarily associated with the V617F mutation in almost all patients with PV and nearly half of the patients with ET and PMF (Oh et al., 2010b) as well as other mutations such as *JAK2* exon 12, thrombopoietic receptor gene (*MPL*), E3 ubiquitin ligase Casitas B-lineage lymphoma proto-oncogene (*CBL*), and mutations that affect the epigenome of MPN patients for instance the loss-of-function mutations in Ten-Eleven Translocation-2 (*TET2*) which affects the state of DNA methylation, and loss-of-function mutations in Enhancer Of Zeste Homolog 2 (*EZH2*) which alters the putative chromatin structure, among others (Martinez-Aviles et al., 2012). Nonetheless, most of these common genetic alterations were not detected in the majority of V617F-negative MPNs cases. Moreover, the leukaemic clones of some V617F-positive MPN patients who transformed into AML were found to be V617F-negative (Theocharides et al., 2007). It is believed there are more unknown and undiscovered mutations accounting for the complexity of the disease.

Results from part I and part II of this study suggested a strong association between *JAK2* polymorphisms and MPNs particularly the single SNP rs12342421 (S8, *JAK2* haplotype tSNP) rather than the haplotypes. However, rs12342421 (S8) and other strongly associated polymorphisms were non-functional, suggesting that the SNPs or haplotypes reported here are unlikely the causal variants driving the development of

MPN risk. There may be some untyped variants in strong linkage disequilibrium (LD) with these markers that are causing the disease. Therefore, the strongly associated SNPs might be in LD with other variants that are not within the studied region. Using LocusZoom online free tool (https://statgen.sph.umich.edu/locuszoom/genform.php?type=yourdata), two recombination hotspots (the spikes in **Figure 6.1**) covering 330-kb region (including 142kb of *JAK2*) with high conservation were discovered, based on the *P* values (PLINK data of the 95 SNPs) obtained from part I. Any correlation between alleles within this region is unlikely to be destroyed. Genes that typically stay together during recombination are said to be linked.

It is also possible that current functional annotation platforms are underpowered to detect either direct or indirect putative functions of the associated SNPs. MPNs are a group of complex disorder involving various factors including genetic, epigenetic (Abdel-Wahab, 2011), environmental, and ethnic factors (Anderson et al., 2012), it is possible that there exist other risk factors for MPNs involving different pathways. In this study, Illumina MiSeq was used to sequence a region of 330kb encompassing *JAK2* gene for 48 case-control pairs (96 samples) by Capture Sequencing-Target Enrichment Method. With a desired coverage of 200×, the platform enables a total of 600kb sequencing read. Therefore, exomes from other genes that have previously been implicated as being involved in the development of MPNs were included to fill up the remaining ~270kb capacity. We used targeted NGS to search for mutations in eight MPN-associated genes, according to their mutational frequencies reported in MPNs. These include genes involved in intracellular signalling, epigenetic modifiers, and leukaemic transformation, such as *MPL*, *TET2*, *ASXL1*, *DNMT3A*, *EZH2*, *LNK*, *IDH1*, and *CBL* (**Table 6.1**).

The genomic regions of interest were directly selected by 'targeted enrichment' strategy using the SeqCap EZ capture from Roche NimbleGen, Inc. This is a custom-capture technology that is compatible with Illumina MiSeq - a benchtop sequencing platform. This in-solution target enrichment process has its standard protocol prepared for Illumina library preparation. The complete exome (protein-coding sequence) of the eight MPN-associated genes (**Table 6.1**) rather than the whole genome were enriched and subjected to quality control. Sequencing was performed upon selective recovery of the enriched targeted regions. This strategy is more cost-effective and less laborious compared with whole-genome sequencing (Mertes et al., 2011; Summerer, 2009). The mechanism is diagrammed in **Figure 1.12**.

6.1.2 Research aims

The aims of this study are:

1. To conduct a preliminary study that will screen for potential genetic variants in the 330kb-region on chromosome 9
2. To discover any mutation or causal variants associated with V617F-positive MPN

The fundamental aim of this study was to expand this protocol for additional MPN-associated genes screening involving larger cohorts of samples.

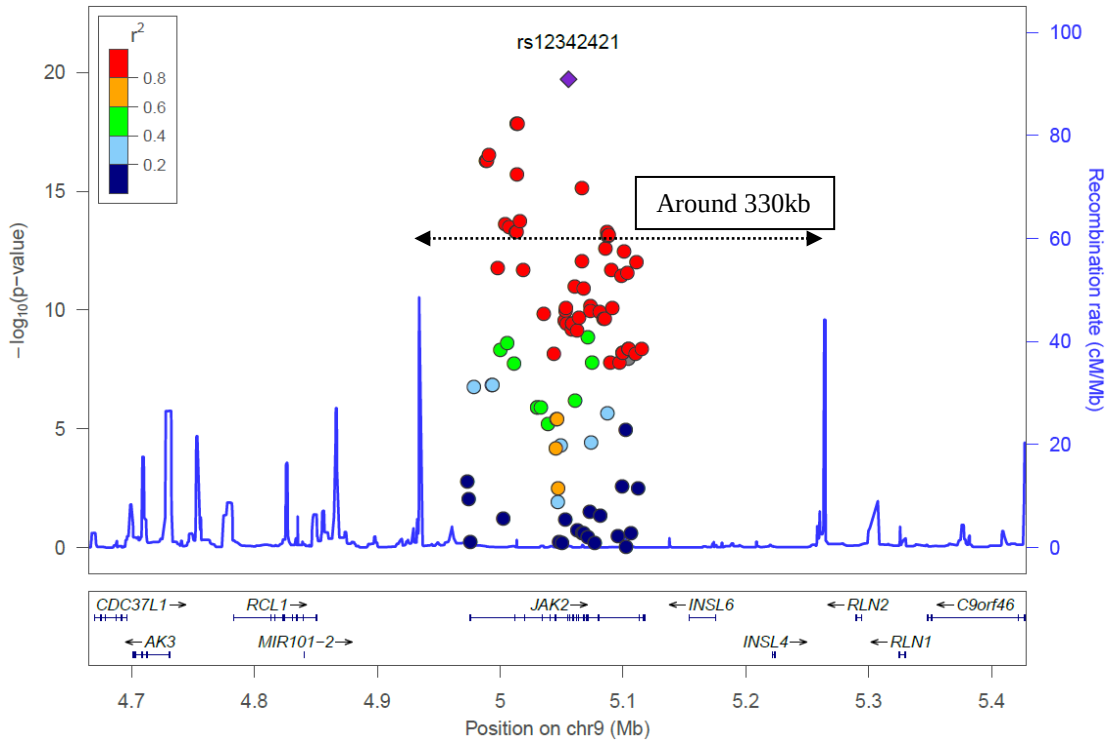


Figure 6. 1 LocusZoom plot showing the MPN-associated region surrounding the *JAK2* gene.

Table 6.1 Genes selected for deep sequencing

No	Gene	Localisation	Function	Comment	Consequence	Frequencies in		
						PV	ET	PMF
Genes of founding mutations								
1	*JAK2*	9p24	Tyrosine kinase, signalling	Gain of function	Activated JAK-STAT signalling	95-99%	50-70%	40-60%
	Exon 12 *JAK2*	9p24	Tyrosine kinase, signalling	Gain of function	Activated JAK-STAT signalling	1-4%	Rare	Rare
2	*MPL*	1p34	Receptor, signalling	Gain of function	Activated JAK-STAT signalling	Rare	1-5%	5-11%
Genes involved in intracellular signalling								
3	*LNK* (SH2B3)	12q24	Adaptor, negative feedback regulation	Loss of function	Activated JAK-STAT signalling	Rare/ Not Reported	5%	<5%
4	*CBL*	11q23	Adaptor, E3 ubiquitin ligase, signal regulation	Loss of function	Dysregulated cytokine-mediated signalling, poor prognosis in myelofibrosis	Rare	Rare	5-10%
5	*TET2*	4q24	DNA hydroxymethylation	Loss of function	Dysregulated gene transcription	7-16%	4-11%	8-19%
6	*DNMT3A*	2p23	Epigenetic regulator	Coexisting *JAK2* V617F, *TET2*, *ASXL1*, and *IDH 1/2* mutations	Poor outcome	5-7%	3%	7-15%
Genes involved in epigenetic regulation and leukaemic progression								
7	*ASXL1*	20q11.21	Chromatin modifications	Loss of function	Not well understood	2-5%	5-8%	13-40%
8	*EZH2*	7q35	Chromatin methylation	Loss of function	Poor prognosis in PMF	3-5%	Rare	6-13%
9	*IDH1*	2q33.3	Metabolism (oxidative decarboxylation of isocitrate to α-ketoglutarate)	Neomorphic enzyme	Epigenetic dysregulation, worse survival	2%	1%	4%

Abbreviation; *JAK2*, Janus kinase 2; *MPL*, Myeloproliferative leukaemia virus oncogene; *LNK*, SH2B adaptor protein 3; *CBL*, Casitas B-cell lymphoma; *IDH1*, Isocitrate dehydrogenase 1; *TET2*, Ten-Eleven-Translocation 2, *DNMT3A*, DNA (cytosine-5-)-methyltransferase 3 alpha, *ASXL1*, Additional sex combs like transcriptional regulator 1; *EZH2*, Enhancer of zeste homolog; PV, polycythaemia vera; ET, essential thrombocythaemia; PMF, primary myelofibrosis.

Information was obtained and compiled from Pasquier et al. (2014), Scott et al. (2007), (Abdel-Wahab et al., 2011b), (Abdel-Wahab et al., 2010), (Abdel-Wahab et al., 2011a; Ernst et al., 2010b), (Vannucchi et al., 2013), Vainchenker et al. (2011), Nguyen and Gotlib (2012).

6.2 Methodology

The sequencing process started with nucleic acid extraction to prepare the starting material for sequencing (library preparation). Once sequencing library was established, sequencing could be performed. Finally, base calls were analysed with bioinformatics tools. The workflow is shown in **Figure 1.12**.

6.2.1 Subjects and DNA samples

This study targeted sequenced 48 case-control pairs from the sample pool. The selected pool of MPN patients includes 16 PV patients, 28 ET patients, and 4 PMF patients. Genomic DNA from buffy coat was isolated from EDTA blood following standard protocols by FlexiGene DNA extraction kit (Qiagen, Hilden, Germany) (Section 2.2.3). The diagnoses of all patients were made based on WHO criteria (2008) (Vardiman et al., 2009) and clinical data.

6.2.2 Sequence Capture and MiSeq sequencing

Next-generation sequencing platform

Illumina (San Diego, CA, USA) MiSeq Desktop Sequencer was chosen as the targeted deep sequencing platform. To be cost-effective and time-efficient, targeted enrichment method was used to target, capture, and sequence only the genomic region of interest.

NimbleGen SeqCap EZ Choice Library Design

The custom-capture technology, SeqCap EZ capture from Roche NimbleGen, Inc. was used to enrich the exomes of *JAK2* gene and other MPN-associated regions (**Table 6.1**). Based on GRCh37/hg19 human reference genome, a set of DNA probe capturing genomic coordinates of coding exons in all isoforms of the selected genes were identified and custom designed by NimbleGen. The percent of target bases covered was 87.9%.

6.2.2.1 Library preparation

To prepare the starting material for deep sequencing, genomic DNA from the 48 MPN patients and 48 sex- and age-matched controls were quantified following standard protocols using Qubit® dsDNA BR Assay Kits from Invitrogen Life Technologies (Carlsbad, CA). Overall library preparation was conducted using combination of kits from different companies depending on performance and cost.

Fragmentation

After optimisation, each gDNA of 500ng was fragmented (20 minutes) to approximately 300-400bp length using the NEBNext® dsDNA Fragmentase (New England Biolabs). After fragmentation, DNA was purified using QIAquick PCR purification kit from Qiagen (Hilden, Germany).

Modification

As depicted in **Figure 1.12**, fragmented DNA was end-repaired and ligated with adapters using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (New Eng-

land Biolabs). Standard protocols were followed. Next, the adapter-ligated DNA was size selected using AMPure XP beads from Beckman Coulter following recommended conditions based on approximate insert size (300-400bp). The selected adapter-ligated DNA fragments of the 96 samples were PCR amplified using individually assigned index primer from NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1 and 2). Amplification was performed in Axygen™ 8-Strip PCR Tubes with Veriti® Thermal Cycler from Applied Biosystems, (Foster City, CA, USA). PCR products were purified by QIAquick PCR purification kit (Qiagen). Size of the purified PCR products were then confirmed with DNA 12000 kit High-Sensitivity chips on an Agilent Bioanalyzer 2100 (Agilent Technologies, Waldbronn, Germany) following the manufacturer's protocols. Next, the library was quantified using GeneRead Library Quant Kit for Illumina (Qiagen) following the manufacturer's protocols on LightCycler® 480 (LC480) from Roche. After quantification, equal amounts (by mass) of each amplified sample library were pooled to be "Multiplex DNA Sample Library Pool" with a final mass of 1.25µg for subsequent enrichment using NimbleGen SeqCap EZ Choice Library kit. To obtain equal numbers of sequencing reads, accurate quantification and handling of the amplified DNA library were important to pool equal amounts of each individual sample.

Targeted Enrichment

The targeted regions of all selected genes were enriched using NimbleGen SeqCap EZ Choice Library kit following manufacturer's protocols. This method is an in-solution approach that enabled enrichment of all the 96 samples at ease because it that does not require additional equipment and thus it is more cost-effective and is readily scalable. The solution-based targeted enrichment method was processed in

single tubes as depicted in **Figure 1.12**. The workflow included (H) hybridising the sample and SeqCap EZ libraries using biotinylated DNA complementary probes, (G) washing unbound fragments and (F) recovering captured multiplex DNA samples using streptavidin-labelled magnetic beads, (H) amplifying captured multiplex DNA samples using LM-PCR. After amplification, the amplified captured DNA was cleaned up using AMPure XP beads (Beckman Coulter), and size of the purified amplified captured DNA was checked using DNA 12000 kit High-Sensitivity chips on an Agilent Bioanalyzer 2100. Lastly, before sequencing, (I) the library was quantified using GeneRead Library Quant Kit for Illumina (Qiagen) and enrichment was measured using qPCR on LightCycler® 480 from Roche.

6.2.2.2 MiSeq Sequencing

Once sequencing library was established, samples (48 MPN cases and 48 paired-controls) were sequenced on Illumina's MiSeq benchtop sequencer (San Diego, CA.) using the MiSeq Reagent Kit v2 with 2 × 150 paired-end reads, multiplexing 24, single indexed samples per run.

6.2.3 Data analysis

Analysis of the sequencing data can be separated into 2 parts. Primary analysis includes image capture, image analysis, and base calling. The secondary analysis includes quality control, mapping and variant calling. Then, functional annotation analysis was performed.

6.2.3.1 Sequence Alignment and Variant Calling

Upon completion of the sequencing reaction, Illumina's Real Time Analysis (RTA) primary analysis software would generate base calls necessary for secondary analysis. Secondary analysis was automatically performed with the pre-installed MiSeq Reporter Software (MSR; version 2.3.32) once all the primary analysis files were generated. The resequencing workflow was used in this study to identify any variants by aligning and comparing the base reads against reference genome (GRCh37/hg19). As an overview, the MSR workflow demultiplexed the indexed samples, generated FASTQ file, aligned the base reads to hg19 reference genome and generated a BAM (Binary Alignment/Map i.e. tab-delimited text file containing sequence alignment data) file format, and identified variants in VCF (variant call format) file. Data encoding was performed using Sanger/Illumina 1.9 and was evaluated with internal quality control check measured by the Phred quality score (Q score) (Ewing & Green, 1998). Q score is an indication of the probability of base calling error. Q score of 30 (Q30) indicates a calling error probability of 1 in 1000 times; that is, a 99.9% accuracy in base calling. Thus, the higher the Q score assigned by sequencing system, the higher accuracy the base calling is. To generate accurate conclusions, base calling was monitored by a Q score of Q30 or above.

By default, pre-processed BAM files were aligned to reference genome (GRCh37/hg19) and subjected to haplotype calling and variant filtration by Genome Analysis Toolkit (GATK, version 2.8-1-g932cd3a) (McKenna et al., 2010). I then used Integrative Genomics Viewer (IGV) (Robinson et al., 2011) to visualise the sequence alignment and confirm the identified variants.

6.2.3.2 Variant analysis

Functional annotation

Variant annotation was performed with VCF as the input file. The ANNOVAR (Wang et al., 2010) Annotate VCF was used to identify and annotate the VCF with functional information with respect to genes.

The resulting SNPs were filtered by the following filters:

QD (Variant Confidence/Quality by Depth) <2.0
FS (Phred-scaled *P* value based on Fisher's exact test) >60.0
MQ (root mean square of the mapping quality of reads across all samples) <40.0
HaplotypeScore >13
MappingQualityRankSum (mapping quality rank sum test score) <-12.5
ReadPosRankSum (read position rank sum test score) <-8.0

Only variants that passed the filter were selected. Visual inspection of the variants was done using IGV.

Identification of MPN-associated variants

To identify whether the variants were case- or control-specific, variants from cases and controls were sorted and compared. For each variant, I calculated the absolute count of alternate alleles (all non-reference alleles). The difference between cases and controls were then calculated using Fisher's exact test and then filtered with natural positive false discovery rate (pFDR) method that automatically accounts for multiple testing (Storey, 2002; Carlson et al., 2009). *P* value of 0.05 implies that 5% of all tests will result in false positives. Based on this *P* value, pFDR approach generates an adjusted *P* value called *q*-values. The *q*-values were calculated using an online freeware False Discovery Rate Calculator (http://research.microsoft.com/en-

us/um/redmond/projects/MSCompBio/FalseDiscoveryRate/default.aspx). The *q*-value of 0.01 implies that 1% of the significant tests will result in false positives. This narrows the list and is useful in ranking the list for a cut-off point. Fisher's exact test generates more accurate statistics for comparison with sample size less than 30 (Stommel, 2014). The selected variants (P <0.01 and q <0.01) were again predicted for biological functions using SNPnexus (http://snp-nexus.org/index.html) (Dayem Ullah et al., 2012) and SNP Function Prediction (FuncPred**)** (http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm) (Xu & Taylor, 2009). These tools annotate variants for various categories by integrating functional information of SNPs obtained from different databases. Prediction for functional effect of SNPs include population data, the effect of non-synonymous coding SNPs on protein function, regulatory elements, conservation, phenotype & disease association, and structural variations. The results were cross-checked with that generated from ANNOVAR server.

Identification of MPN-associated variants within the two recombination hotspots

To discover any 'hidden markers' within the two recombination hotspots on chromosome 9 (**Figure 6.1**), I performed an independent analysis for the 330-kb region (including 142kb of *JAK2*). The same statistical procedure above was used. The filtered variants were cross-checked with the SNP list filtered by Fisher's exact test (P <0.01) and q <0.01. Functional annotation was also performed for the selected variants.

Novel variants

Any post-filtered variants without SNP rs number were manually cross-checked with databases including UCSC Genome Browser (http://genome.ucsc.edu/cgi-

bin/hgTracks?org=human) (Kent et al., 2002) and 1000 Genomes Project (Genomes Project et al., 2012) to avoid missing any known variant. After SNP ID annotation, any novel variant was identified. The newly detected variants were analysed using MutationTaster software (http://www.mutationtaster.org/ChrPos.html) (Schwarz et al., 2014) for potential pathogenic effects.

6.4 Results

Targeted resequencing of *JAK2* and eight MPN-associated genes in 96 cases and controls

Samples from 48 MPN patients and 48 sex- and age-matched controls were sequenced on the Illumina's MiSeq benchtop sequencer (San Diego, CA; average output 4.5-5.1 Gb). All samples were enriched, pooled and sequenced in 24-multiplex. Targeted enrichment was performed using the NimbleGen SeqCap EZ choice sequence capture approach.

Clinical characteristics of subjects

The selected pool of MPN patients (16 PV patients, 28 ET patients, and 4 PMF patients) were with high allele burden of V617F mutation, based on mutant band intensity from the gel (Section 3.3.2). In addition to V617F-positivity, the 48 samples were of high nucleic acid quality and quantity as determined by NanoDrop ND1000 UV/Vis spectrophotometer. Their mean age was 59 years (ranges: 21-88 years). As paired control, 48 controls that were perfectly matched for sex and as much as possible for age were selected from the control pool. Their mean age was 55 years (ranges: 21-77 years). Age was not completely matched because of a lack of blood donation by older persons.

6.4.1 Detection of variants across the targeted regions

The 330-kb region of 2 recombination hotspots encompassing *JAK2* gene and all coding exons (all isoforms) of 8 MPN-associated genes (**Table 6.1**) were targeted by a custom SeqCap EZ Choice Library from Roche NimbleGen. For the 8 MPN-associated genes, 50bp upstream and downstream of the regions were also included. Transcription binding sites are typically 10 base pairs long (it can vary between 5 to 30 nucleotide long) (Stewart et al., 2012), hence targeting the 50bp-flanking sequence avoids missing any potential regulatory region. In total, 121 regions comprising a total length of 401.324 kb were targeted. The final design covered approximately 87.9% of the requested target regions. The mean and median of the post-filtered alignment were 71.5% and 76.25% respectively (for an individual, percentage of reads aligned ranges from 34.2 to 99%). The mean depth of variant coverage per sample was 65.9, which is above the minimum requirement (30×) for accurate detection of heterozygous variants (Ku et al., 2012). **Table 6.2** shows the basic statistics on raw data generated from MiSeq. Variant calling identified 532 variants that passed the standard quality control filters (**Table 6.3)**.

6.4.2 Identification of MPN-associated variants

The 532 variants were filtered based on Fisher's exact test (P <0.01) followed by pFDR (q <0.01). After filtering, 200 known variants remained. All of the 200 variants are located in chromosome 9, i.e., within the two recombination hotspots on chromosome 9. Odds ratio between cases and controls for these variants were plotted along chromosomal region as depicted in **Figure 6.2**. Among the 200 variants, *JAK2* SNPs constituted 45.5% (91), and variants with the highest odds ratio were also found to be residing within the *JAK2* gene region. Using SNPnexus (Dayem Ullah et

al., 2012) and SNP Function Prediction (FuncPred) (Xu & Taylor, 2009), the variants were classified into insertions and deletions, mostly in introns (55.5%), intergenic (41%), and some in 5' untranslated region (UTR) (0.5%), 5' upstream (1%), and 3' downstream region (1%). Only two variants were identified in the coding regions of *JAK2* and *INSL4* (insulin-like 4 (placenta)). The results were cross-checked with that generated from ANNOVAR server and compiled in **Table 6.4** according to different types of genetic variants. Some important features were also documented.

6.4.3 Identification of MPN-associated variants within the two recombination hotspots

The 330-kb region within the recombination hotspots were examined using Fisher's exact test (P <0.01) followed by pFDR (q <0.01). After filtering, 202 variants remained: 199 known variants and 3 novel variants. This cut-off point (q <0.01) yielded very much similar results when Fisher's exact test (P <0.01) was applied to the variant pool. Potential functional variants were summarised in **Table 6.5**. Based on the pFDR approach (q <0.01), a total of 179 variants were found to be overlapping with the 200 Fisher-filtered variants. From the post-filtered variant pool, 17 variants were found within the 18-kb *JAK2* haplotype region flanking 5065750-5084049 (GRCh37/hg19) on chromosome 9.

From the Fisher-filtered and pFDR-filtered variants, 3 novel variants were identified. Only one *JAK2* variant (chr9: 5062501, TAAAAAAAAAAAAAAAAAA/ T) was annotated as intronic polymorphism by MutationTaster software. This deletion was not predicted to be deleterious, however can possibly affect the protein features and

bring changes to splice site. H3K36me3, Histone, Histone 3 Lysine 36 Tri-Methylation were predicted to be the regulatory features associated with this deletion.

Table 6. 2 Basic MiSeq sequencing quality control metrics for the 96 samples

Sample ID	PF_Reads	GC (%)	PF_UQ_Reads aligned (%)	Mean Coverage	Coverage >30× (%)	Library size
MPN-006	1293059	44	62.5	34.4	56.5	650960
MPN-034	3357826	44	71.1	149.3	83.8	741710
MPN-035	331043	44	78.9	17.4	16.8	390055
MPN-037	1381909	44	64.1	38.0	58.8	644444
MPN-040	1333696	43	85.4	150.7	84.9	764613
MPN-042	649529	44	73.8	32.5	55.2	552562
MPN-046	853134	44	58.2	22.5	32.1	586634
MPN-047	599932	44	75.5	28.7	48.5	365303
MPN-053	2374209	44	57.7	55.4	72.8	826762
MPN-054	2006759	42	84.5	189.2	85.6	674749
MPN-055	748024	42	84.6	89.9	82.7	414045
MPN-056	1162577	44	60.5	30.3	48.9	654429
MPN-066	1122835	43	41.5	17.6	15.2	211537
MPN-070	1057581	44	65.5	30.6	49.3	675574
MPN-071	1158589	43	83.3	132.9	84.9	678353
MPN-072	698231	44	78.1	36.2	59.3	663446
MPN-077	1746594	44	73.9	88.4	77.8	663396
MPN-078	922607	43	84.9	115.0	82.4	814440
MPN-079	1661016	44	71.9	75.6	74.5	680272
MPN-091	1254890	43	86.3	151.6	85.5	758075
MPN-093	287654	45	78.8	12.9	7.0	262663
MPN-094	985915	43	86.0	113.3	83.1	562598
MPN-095	364179	45	81.0	17.3	18.0	331173
MPN-096	1183750	43	47.9	26.6	45.2	484191
MPN-107	1348008	44	71.7	60.9	69.6	575361
MPN-117	4072077	44	46.3	115.2	80.5	703511
MPN-120	2864654	44	42.6	82.1	75.8	652052
MPN-121	981957	43	81.3	110.5	82.4	675528
MPN-122	1154534	43	83.3	139.1	84.3	818451
MPN-123	3470930	45	42.8	96.0	78.1	793912
MPN-124	1140741	44	99.0	26.7	42.8	616668
MPN-128	506533	45	70.1	21.7	30.2	519552
MPN-129	851156	43	87.7	113.2	82.7	784417
MPN-131	883802	43	86.0	110.3	81.9	723791
MPN-134	361482	45	80.2	17.5	20.4	459534
MPN-135	187808	45	77.5	8.7	1.2	362612
MPN-138	1190637	43	83.4	136.2	84.9	711069
MPN-139	382134	44	63.2	15.2	12.3	346440
MPN-140	420478	45	83.4	21.7	30.3	524238
MPN-148	253719	44	70.2	10.6	2.0	323171
MPN-158	2283933	45	43.7	71.9	71.5	915752
MPN-162	1612369	44	42.6	47.7	64.2	600978

Sample ID	PF_Reads	GC (%)	PF_UQ_Reads aligned (%)	Mean Coverage	Coverage >30× (%)	Library size
MPN-164	1034685	44	66.9	46.7	62.0	586269
MPN-166	870876	44	59.1	24.3	37.3	603104
MPN-174	2474976	44	58.0	87.2	76.7	618879
MPN-175	1669335	44	64.6	44.7	68.4	709122
MPN-176	956389	44	77.8	48.4	63.0	557019
MPN-177	1320855	43	72.2	41.2	63.1	776394
MC-019	488477	44	74.3	21.5	30.4	340413
MC-022	1335527	44	75.9	138.2	84.0	754515
MC-045	867889	44	82.5	50.6	70.3	676692
MC-054	571976	44	80.1	30.2	50.4	449970
MC-059	2589290	45	35.3	35.6	54.9	625834
MC-061	377163	45	81.5	19.6	25.5	608396
MC-065	844540	45	76.6	47.0	58.1	691483
MC-083	1151400	45	73.4	60.4	68.1	628117
MC-123	587272	44	74.5	28.1	47.0	371265
MC-130	531847	44	82.5	30.6	50.5	759156
MC-131	3183808	44	34.9	49.9	67.1	843414
MC-132	1741689	44	98.3	45.6	63.4	628164
MC-144	428445	45	82.8	21.8	30.5	533985
MC-183	1929725	43	67.2	55.1	74.1	981510
MC-197	348357	44	69.3	15.2	10.2	277554
MC-198	758334	45	77.3	38.2	52.2	690875
MC-200	2087276	44	39.4	38.8	59.2	833823
MC-204	1178518	44	61.2	32.7	49.3	545935
MC-209	470357	45	81.5	23.5	33.9	577780
MC-213	979764	44	77.7	48.3	64.9	461479
MC-217	2122920	45	34.2	32.4	50.0	569492
MC-240	518044	45	82.6	28.2	45.6	711363
MC-255	1756919	45	81.1	101.3	77.0	785908
MC-256	1255401	45	77.1	67.0	68.8	834973
MC-257	1524064	43	77.0	155.7	85.8	696975
MC-259	1066477	43	73.4	117.2	84.4	575851
MC-260	1355804	42	79.4	128.4	85.8	392490
MC-269	797781	44	73.3	38.3	53.7	476438
MC-278	1241752	45	79.5	65.5	68.4	890940
MC-294	376478	44	71.4	17.6	18.7	323919
MC-343	1354580	43	85.3	161.2	84.0	791419
MC-344	1375311	45	78.0	72.4	71.9	969600
MC-345	643192	44	75.8	35.9	51.6	616500
MC-348	1702013	44	64.8	46.0	62.8	655527
MC-349	1920576	44	45.3	40.2	58.6	1096474
MC-350	1647910	44	55.9	36.7	59.6	499198
MC-365	1421719	43	79.8	143.1	84.8	619288
MC-370	1374536	44	77.0	132.0	76.9	588317

Sample ID	PF_Reads	GC (%)	PF_UQ_Reads aligned (%)	Mean Coverage	Coverage >30× (%)	Library size
MC-378	1304802	43	85.0	137.5	82.8	565167
MC-389	1262836	43	83.9	143.1	84.6	759250
MC-393	1469924	45	36.0	22.7	30.7	601090
MC-423	336642	45	82.2	17.6	19.3	460400
MC-424	1154497	42	80.8	124.4	84.8	526709
MC-428	726378	43	82.3	41.3	65.1	585653
MC-436	2183427	43	61.3	63.6	77.7	791541
MC-437	718792	44	74.6	36.7	58.4	439918
MC-443	1631106	42	84.8	168.7	86.5	612137
MC-471	1394464	43	86.3	166.5	86.0	797884

Abbreviation; PF_Reads, The number of reads that pass the vendor's filter; GC (%), the overall % GC of all bases in all sequences; PF_UQ_Reads aligned, post-filtered unique reads aligned/ total post-filtered reads; Coverage >30×, The percentage of ALL target bases achieving 30× or greater coverage

Mean Coverage here indicates the mean coverage of targets that received at least coverage depth = 2 at one base.

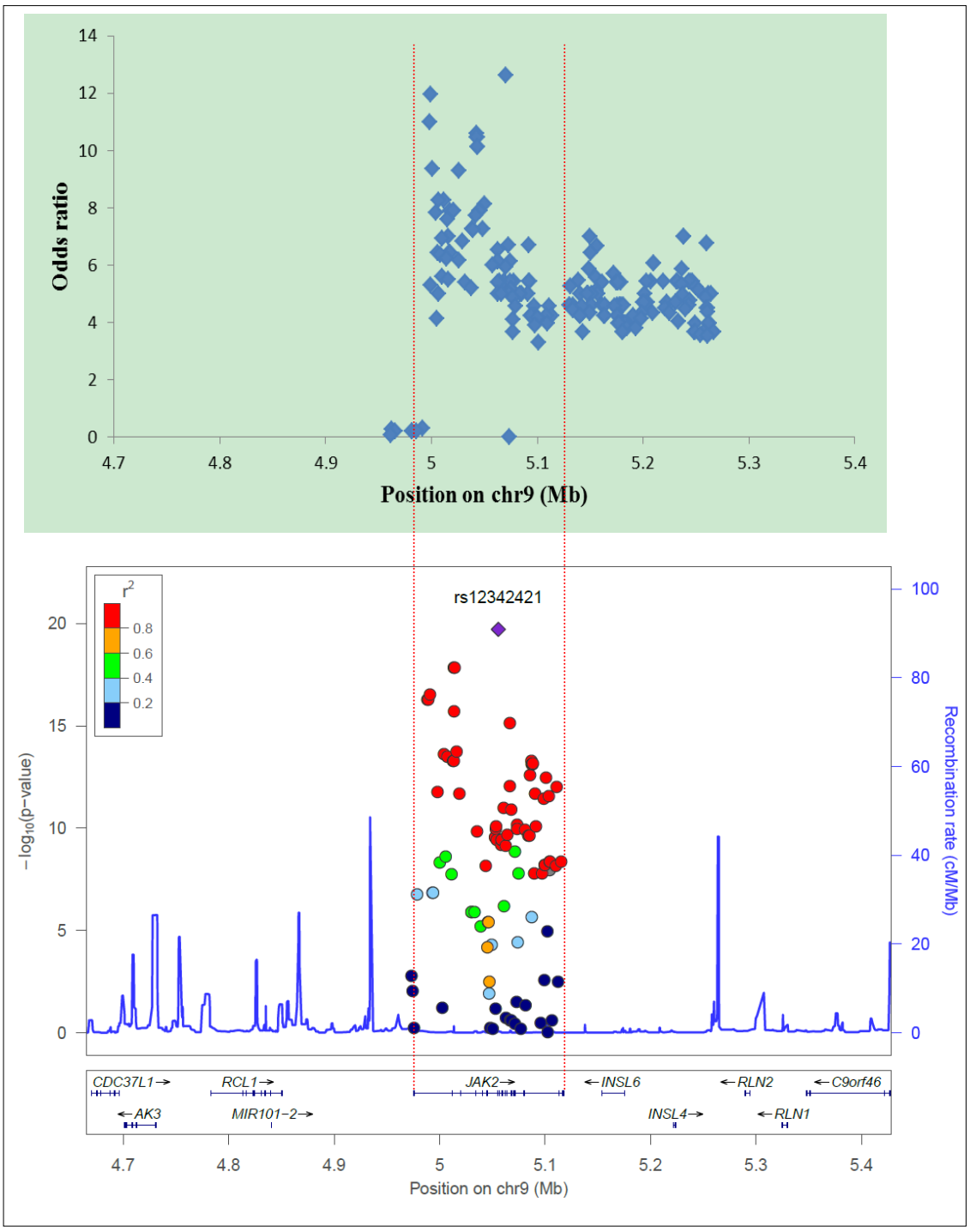


Figure 6. 2 Distribution of post-filtered variants across chromosome 9.

Upper panel: Variants with highest odds ratio were found distributed within the *JAK2* gene region. Lower panel: LocusZoom plot showing the MPN-associated region surrounding the *JAK2* gene.

Table 6. 3 All variants identified by ANNOVAR software

#CHROM	POS	ID	REF	ALT
chr1	43819016	rs192016153	G	A
chr1	43819085	.	G	A
chr2	25456647	.	CAAA	C
chr2	25466888	rs2289093	G	T
chr2	25469502	rs2276598	C	T
chr2	25469913	rs2276599	C	T
chr2	25536827	rs41284843	G	A
chr2	209100914	rs16840781	C	T
chr2	209101311	rs6730955	C	A
chr2	209101905	rs57383668	GA	G
chr2	209106885	.	TAA	T
chr2	209106897	rs112594278	A	G
chr2	209120640	rs1446325	C	T
chr2	209120959	rs369485172	T	G
chr2	209120963	.	T	G
chr2	209120967	rs10207062	T	G
chr4	106067592	rs75973876	G	T
chr4	106158216	rs3796927	G	A
chr4	106159644	rs6847204	T	A
chr4	106160133	rs56185013	G	A
chr4	106160365	rs7670522	A	C
chr4	106163095	.	CTT	C
chr4	106181026	.	A	C
chr4	106196951	rs2454206	A	G
chr4	106199328	.	CAT	C
chr4	106201014	rs11944001	G	A
chr7	148504854	rs55877618	A	AGACTT
chr7	148508833	rs2072407	A	G
chr7	148525904	rs2302427	C	G
chr7	148543693	rs3214332	TA	T
chr7	148543753	rs28723387	C	T
chr7	148580809	rs80052686	G	C
chr7	148581173	.	C	T
chr7	148581436	.	G	T
chr9	4945967	.	G	GTTTGT
chr9	4945971	rs10758668	G	T
chr9	4946096	.	CT	C,CTTT
chr9	4947510	rs4742054	C	T
chr9	4947714	rs4742055	A	T
chr9	4947863	rs7027180	C	G
chr9	4951232	rs55774436	G	C
chr9	4953916	rs4742057	G	A
chr9	4954167	rs6476929	C	T

#CHROM	POS	ID	REF	ALT
chr9	4954246	rs6476930	A	T
chr9	4954276	rs4740805	T	A
chr9	4955072	rs7864359	G	A
chr9	4955495	rs7038475	C	A
chr9	4955496	rs7037557	G	A
chr9	4955592	.	A	G
chr9	4955900	.	CT	C
chr9	4956047	rs202098666	TA	T
chr9	4956697	rs12235675	C	G
chr9	4957298	rs10974887	T	C
chr9	4957480	rs113101535	TA	T
chr9	4957650	rs719673	G	A
chr9	4959074	rs75768081	T	A
chr9	4961260	rs1327500	A	G
chr9	4961314	rs10974890	C	T
chr9	4961487	rs1327499	T	C
chr9	4961616	rs10119552	T	G
chr9	4961636	rs7867937	G	A
chr9	4962247	rs10815141	C	T
chr9	4962383	rs142891372	GTT	G,GT
chr9	4963966	rs368522530	T	C
chr9	4965713	rs2031754	C	A
chr9	4965715	rs1041215	C	G
chr9	4967539	rs1327497	T	A
chr9	4968331	rs10974896	T	C
chr9	4969094	rs77469874	T	C
chr9	4969106	rs28668598	G	A
chr9	4969787	rs7872943	C	T
chr9	4970600	rs369017005	AC	A
chr9	4970715	rs10974897	G	C
chr9	4970844	rs1887426	C	A
chr9	4972060	rs142154231	T	A
chr9	4974779	.	CA	C
chr9	4976186	rs10124001	C	T
chr9	4977563	rs62541511	G	T
chr9	4978610	rs372031520	G	C
chr9	4978816	rs4742061	C	T
chr9	4979730	rs1887427	A	G
chr9	4980756	rs77594484	GT	G
chr9	4980929	rs141192940	T	TTTTTA
chr9	4981191	rs2031753	A	C
chr9	4981354	rs2031752	C	G
chr9	4981602	rs10758669	C	A
chr9	4981684	rs57473427	TA	T
chr9	4981866	rs36051895	G	T

#CHROM	POS	ID	REF	ALT
chr9	4984530	rs1887428	G	C
chr9	4985316	rs375852577	T	C
chr9	4985542	rs2274472	T	C
chr9	4985879	rs2274471	A	G
chr9	4987734	rs60356569	TA	T
chr9	4988341	rs10156475	C	T
chr9	4988683	rs7849067	C	G
chr9	4991379	rs10758670	G	A
chr9	4995125	rs7026646	G	A
chr9	4996403	rs193172174	G	T
chr9	4996928	.	CTTT	C,CT,CTT
chr9	4996997	rs10815142	G	A
chr9	4997918	.	AT	A
chr9	4998639	rs2225125	A	G
chr9	4998741	.	CAA	C,CA
chr9	4998855	rs2209773	G	C
chr9	4999303	rs1327494	A	G
chr9	5000811	rs12347727	A	G
chr9	5001204	rs58346692	G	A
chr9	5001205	rs57291545	A	T
chr9	5001625	rs201169908	AT	A
chr9	5002011	rs4587378	C	T
chr9	5002470	.	C	A
chr9	5003338	rs4372063	G	A
chr9	5003851	rs55689419	AG	A
chr9	5003973	rs10115312	G	T
chr9	5004607	rs191229511	G	C
chr9	5004918	.	CTT	C,CT
chr9	5005034	rs59384377	A	T
chr9	5005040	rs7863708	T	C
chr9	5006345	rs62541529	G	C
chr9	5006743	rs11999928	T	G
chr9	5007138	rs72699568	C	T
chr9	5007418	rs191499683	G	C
chr9	5007725	.	CT	C
chr9	5007948	.	G	T
chr9	5008070	rs7030260	C	A
chr9	5010091	rs10974910	G	A
chr9	5010192	rs10815144	G	A
chr9	5010381	rs112538722	TA	T
chr9	5010471	rs10974911	C	A
chr9	5011795	rs62541531	C	T
chr9	5013266	.	T	A
chr9	5013726	rs183362981	G	A
chr9	5014164	rs200845162	CT	C

#CHROM	POS	ID	REF	ALT
chr9	5014184	rs10974913	C	G
chr9	5015143	rs7045000	G	A
chr9	5015538	rs35167433	CT	C
chr9	5015732	rs7046736	C	A
chr9	5015901	rs10815146	T	A
chr9	5016145	rs10815147	C	T
chr9	5016455	rs4527935	A	T
chr9	5017350	rs10974916	G	A
chr9	5017384	rs62541532	C	A
chr9	5018041	rs12352214	G	A
chr9	5020529	rs7030315	A	G
chr9	5021140	rs12346093	T	A
chr9	5021191	rs12345127	A	T
chr9	5021514	rs7034753	A	G
chr9	5021738	rs10974919	G	A
chr9	5022332	rs150829797	G	T
chr9	5025746	rs10124627	T	G
chr9	5026293	rs2183137	A	G
chr9	5026407	rs143015863	C	T
chr9	5029077	rs60221565	T	G
chr9	5032053	rs200998878	TC	T
chr9	5034454	rs149390964	C	A
chr9	5037694	rs11999076	C	T
chr9	5038164	rs62541538	G	A
chr9	5038180	rs62541539	C	T
chr9	5038596	rs58788809	ATTT	A,AT,ATT
chr9	5039652	rs1576271	A	C
chr9	5040163	rs7023146	G	A
chr9	5040203	rs7043371	T	A
chr9	5040817	rs62541541	G	C
chr9	5042046	rs7032785	C	T
chr9	5042082	rs62541543	G	A
chr9	5042114	.	G	A
chr9	5042981	rs7033052	G	C
chr9	5042987	rs7033053	G	A
chr9	5043007	rs12551254	C	T
chr9	5043152	rs7025540	T	C
chr9	5043156	rs7037207	C	T
chr9	5043688	rs59668095	C	G
chr9	5043767	rs58345215	A	T
chr9	5045328	rs7046077	C	G
chr9	5045695	rs7046511	C	T
chr9	5045916	rs12339774	C	T
chr9	5046881	rs7039518	T	C
chr9	5046935	rs10121491	T	C

#CHROM	POS	ID	REF	ALT
chr9	5048213	rs12343374	C	T
chr9	5048814	rs7028112	G	A
chr9	5049065	rs1328917	G	T
chr9	5050307	rs10974935	C	T
chr9	5052405	.	C	G
chr9	5055015	rs1536799	C	T
chr9	5055434	rs1536800	C	T
chr9	5056037	rs10974938	A	G
chr9	5057284	rs10815148	T	A
chr9	5057568	rs12338854	G	C
chr9	5059440	rs2149556	C	T
chr9	5062473	rs7859390	T	A
chr9	5062500	.	TAAAAAAAAAAAAAAAAAA	T
chr9	5062610	rs57204002	T	C
chr9	5062816	rs61284219	C	A
chr9	5062844	rs57707737	A	T
chr9	5062932	rs57299248	C	G
chr9	5063884	rs150038496	A	G
chr9	5064193	rs16922576	T	C
chr9	5065553	.	A	C
chr9	5065750	rs12342421	G	C
chr9	5067659	rs148586907	T	C
chr9	5068174	rs201280728	CAA	C,CA
chr9	5068596	rs3780366	T	C
chr9	5069837	rs7869668	G	A
chr9	5070227	rs141632941	ATTATC	A
chr9	5070401	rs67006981	A	G
chr9	5071049	rs10119004	G	A
chr9	5072798	rs4495487	T	C
chr9	5073289	rs10283730	G	A
chr9	5073770	rs77375493	G	T
chr9	5074189	rs12343867	T	C
chr9	5074466	rs113659315	TA	T
chr9	5075603	rs10283563	C	T
chr9	5075628	rs10283564	C	G
chr9	5075747	rs150334355	A	G
chr9	5076613	rs12349785	G	C
chr9	5076691	rs12340895	C	G
chr9	5076945	rs35463922	T	TA
chr9	5078117	rs1159782	T	C
chr9	5079828	rs10974952	A	T
chr9	5081138	rs12554699	G	A
chr9	5081780	rs2230724	G	A
chr9	5082106	rs2149560	G	T
chr9	5082333	rs7865719	A	G

#CHROM	POS	ID	REF	ALT
chr9	5082671	rs57839907	T	A
chr9	5082839	rs73395330	C	T
chr9	5083533	rs12343065	C	T
chr9	5083786	rs61194629	A	AT
chr9	5083827	.	T	C
chr9	5084049	rs7857730	G	T
chr9	5084837	rs6476939	T	A
chr9	5085417	rs62543863	C	T
chr9	5085859	rs62543876	G	C
chr9	5087238	rs2031905	G	A
chr9	5088542	rs59331584	CA	C
chr9	5089566	.	G	A
chr9	5089568	.	C	A
chr9	5089570	.	G	A
chr9	5090641	rs10974955	G	A
chr9	5090966	rs146385350	ATTG	A
chr9	5092263	rs11788790	G	T
chr9	5092466	rs11788834	G	A
chr9	5093188	rs3780368	G	C
chr9	5094034	rs10117459	G	A
chr9	5094185	rs12340866	G	A
chr9	5095095	rs3780369	A	C
chr9	5095167	rs3780370	A	C
chr9	5095191	rs3780371	T	A
chr9	5095538	rs10815154	G	A
chr9	5096728	rs7857081	G	A
chr9	5097171	rs7847141	A	T
chr9	5097544	rs3780372	C	T
chr9	5098223	rs3780373	T	C
chr9	5098411	rs10121077	G	A
chr9	5100501	rs7851455	G	C
chr9	5101061	.	T	A
chr9	5101305	rs10974963	G	A
chr9	5102910	rs7032780	G	C
chr9	5103450	rs7022333	A	G
chr9	5103807	rs146051909	C	A
chr9	5103942	rs11531774	C	T
chr9	5105679	rs2094622	C	A
chr9	5106610	rs7019858	C	T
chr9	5106775	rs12340894	C	A
chr9	5107278	rs4593605	C	A
chr9	5107394	rs1328918	A	G
chr9	5108771	rs10815157	G	C
chr9	5109208	rs12344629	C	T
chr9	5109707	rs11793659	A	G

#CHROM	POS	ID	REF	ALT
chr9	5110000	rs17425637	C	T
chr9	5110899	rs3780377	T	C
chr9	5111358	rs138377711	AC	A
chr9	5112288	rs3780378	C	T
chr9	5112844	rs10815158	G	A
chr9	5113424	rs373525228	TAAAAAAAAA	T
chr9	5113452	rs9987451	C	T
chr9	5113577	rs3824433	C	T
chr9	5115213	rs10815159	C	T
chr9	5115860	rs7470337	C	G
chr9	5116369	rs10758672	T	C
chr9	5116616	rs10815160	T	G
chr9	5117673	rs148147368	TAATA	T
chr9	5118421	.	G	C
chr9	5119256	.	T	C
chr9	5119932	.	AAT	A
chr9	5119939	rs10217652	A	G
chr9	5120803	rs10815161	T	C
chr9	5123888	.	CT	C
chr9	5124523	.	C	T
chr9	5124570	rs147906148	T	C
chr9	5125336	rs10974969	G	A
chr9	5125469	rs11793886	A	G
chr9	5125741	rs10114315	G	A
chr9	5125979	rs368802437	T	A
chr9	5126200	.	T	C
chr9	5126843	rs144726392	A	G
chr9	5127451	rs150306172	A	G
chr9	5128080	.	TTGTGTG	T,TTG,TTGTG,TTGTGTGTG
chr9	5128107	.	T	C
chr9	5130146	rs10974972	G	A
chr9	5130363	rs10123980	T	C
chr9	5130710	rs12349735	A	C
chr9	5130726	.	AT	A
chr9	5130792	rs4342662	G	A
chr9	5131312	rs10815164	T	A
chr9	5131435	.	ATT	A,AT
chr9	5133609	rs6476941	C	T
chr9	5133665	rs7035251	T	A
chr9	5134065	rs7038687	T	C
chr9	5134213	rs11789744	G	A
chr9	5138730	rs5896113	T	TA
chr9	5138917	rs111396397	AT	A
chr9	5140115	rs1410780	T	C
chr9	5140313	rs1571438	A	C

#CHROM	POS	ID	REF	ALT
chr9	5140671	rs10974980	C	G
chr9	5140715	rs35146492	A	AC
chr9	5140841	rs10115962	T	C
chr9	5140963	rs10115335	A	T
chr9	5142216	rs12344116	G	A
chr9	5143291	rs62541899	C	T
chr9	5143663	rs58643904	CT	C
chr9	5144167	.	C	A
chr9	5145887	rs2381195	T	C
chr9	5147539	rs10121316	A	G
chr9	5147817	rs7850484	G	A
chr9	5148278	rs10118930	C	T
chr9	5149138	rs10815165	T	C
chr9	5149564	rs62541910	A	G
chr9	5149657	rs10974983	G	T
chr9	5149708	rs59001032	AT	A
chr9	5150058	rs10815167	A	G
chr9	5150256	rs35036287	TTA	T
chr9	5150800	rs12684720	T	C
chr9	5151222	rs60001190	C	CT
chr9	5151275	rs62541915	A	T
chr9	5151644	rs188719843	G	A
chr9	5152855	.	A	G
chr9	5152861	.	T	C
chr9	5153141	rs57869658	GT	G
chr9	5155261	rs2209774	C	T
chr9	5155477	.	TA	T
chr9	5155529	.	TA	T
chr9	5155608	rs7019912	A	T
chr9	5155686	rs10815169	G	A
chr9	5155758	rs11787793	G	A
chr9	5156141	rs58456759	G	GA
chr9	5156542	.	T	TA
chr9	5156637	.	T	C
chr9	5156671	.	T	C
chr9	5157277	rs61461940	T	C
chr9	5158837	.	C	T
chr9	5159318	rs62541944	G	T
chr9	5159359	rs147765596	AT	A
chr9	5160176	.	CAA	C
chr9	5161453	rs72701654	T	C
chr9	5161836	rs12352022	T	G
chr9	5161853	rs141972979	G	A
chr9	5162828	rs12353218	T	C
chr9	5163633	rs34772339	A	ACTT

#CHROM	POS	ID	REF	ALT
chr9	5163645	rs10758674	G	A
chr9	5163848	rs35123699	CA	C
chr9	5166273	rs11790841	G	C
chr9	5166314	rs200663498	T	TCC
chr9	5166364	.	G	C
chr9	5166548	rs1328919	C	T
chr9	5167921	rs2381211	T	C
chr9	5168585	rs2381212	C	A
chr9	5168871	rs2381213	G	A
chr9	5168927	.	ATT	A,AT,ATTT
chr9	5168995	rs2149553	G	A
chr9	5169804	rs111573904	A	T
chr9	5169814	rs7858245	T	A
chr9	5170080	rs7849717	G	A
chr9	5170229	.	CA	C
chr9	5170556	.	CTT	C,CT,CTTT
chr9	5170645	rs7850425	G	T
chr9	5172540	rs11791281	G	A
chr9	5172567	rs10815171	G	A
chr9	5172907	rs11791350	G	T
chr9	5173399	rs7036034	C	T
chr9	5173406	.	A	G
chr9	5173776	rs200177613	TA	T
chr9	5174782	rs7040806	C	T
chr9	5175288	rs7871515	C	T
chr9	5175621	rs62541958	T	C
chr9	5175687	rs62541959	G	C
chr9	5175722	rs188687177	T	C
chr9	5176147	rs12340303	C	T
chr9	5176215	rs12340333	C	G
chr9	5176352	rs12340379	C	T
chr9	5176719	rs34357993	CA	C
chr9	5177462	rs1581927	C	T
chr9	5177533	rs1590803	T	C
chr9	5177872	rs12341844	C	T
chr9	5177937	rs60460912	G	A
chr9	5178579	rs12343038	C	T
chr9	5179002	rs7030853	C	T
chr9	5179044	.	CA	C,CAA
chr9	5179920	rs11795305	G	C
chr9	5180065	rs113657238	T	C
chr9	5180068	rs62543572	G	C
chr9	5180728	.	C	T
chr9	5180961	rs7029244	A	C
chr9	5180996	rs7032616	T	C

#CHROM	POS	ID	REF	ALT
chr9	5181467	rs7047795	C	T
chr9	5181610	rs7036536	T	C
chr9	5182159	rs10974993	C	T
chr9	5182301	rs58848917	TA	T
chr9	5182316	rs142715161	G	A
chr9	5183028	rs9657576	T	G
chr9	5184449	rs7856912	G	A
chr9	5186332	rs7038681	C	T
chr9	5188078	rs10758677	A	G
chr9	5190444	rs1322222	A	G
chr9	5192874	rs10122037	C	T
chr9	5193002	rs192100388	A	C
chr9	5193017	rs78521709	C	T
chr9	5196262	rs2038588	C	T
chr9	5198335	rs11790596	T	C
chr9	5198781	rs7036761	G	T
chr9	5199750	rs7869015	C	T
chr9	5200714	rs16922786	A	G
chr9	5201515	rs143523391	A	AAC
chr9	5203054	rs10491650	T	C
chr9	5203866	rs4560849	C	G
chr9	5206584	.	G	T
chr9	5206667	.	A	G
chr9	5206672	.	G	A
chr9	5206800	rs113683988	T	C
chr9	5206829	rs112603351	G	A
chr9	5207294	rs72701680	A	T
chr9	5209433	rs1590268	T	A
chr9	5209797	rs7858422	G	T
chr9	5209804	rs7874495	A	G
chr9	5209906	rs138768170	G	T
chr9	5212254	rs10815172	T	A
chr9	5215064	rs4008433	T	G
chr9	5218713	rs58870312	TA	T
chr9	5220255	.	T	C
chr9	5220397	rs10975009	T	G
chr9	5221884	rs74609285	C	A
chr9	5222208	rs4425810	A	G
chr9	5224676	rs60768043	T	G
chr9	5226045	.	A	G
chr9	5226459	rs190723365	T	G
chr9	5226484	rs7036833	G	A
chr9	5228050	.	G	GGT
chr9	5230444	rs2093448	T	G
chr9	5230447	rs2104166	T	C

#CHROM	POS	ID	REF	ALT
chr9	5231712	rs12720	T	A
chr9	5232409	rs3780383	G	C
chr9	5233054	rs1555476	T	C
chr9	5236822	rs7027871	C	G
chr9	5236924	rs6476952	T	C
chr9	5237442	rs7848312	T	C
chr9	5237479	rs7859361	C	T
chr9	5238460	rs10118434	T	C
chr9	5238588	rs1575283	C	T
chr9	5239325	.	A	AAC
chr9	5239778	.	CTT	C,CT
chr9	5239797	.	GTTT	G,GT,GTT
chr9	5243511	rs10120337	G	A
chr9	5243736	rs10118267	T	C
chr9	5244058	rs7023639	A	C
chr9	5244708	rs10283473	C	A
chr9	5246403	rs10975024	A	T
chr9	5246714	rs59673540	C	CA
chr9	5248768	rs11790680	C	T
chr9	5248827	rs10975027	G	A
chr9	5249364	rs1853221	A	G
chr9	5250273	rs11792629	G	T
chr9	5250918	rs35533149	AAAG	A
chr9	5252789	rs11506668	C	T
chr9	5252803	rs138616994	CTT	C
chr9	5253031	.	G	A
chr9	5253819	rs369869008	G	A
chr9	5254224	rs12349113	T	A
chr9	5254927	rs4500135	G	A
chr9	5255920	rs147084794	G	A
chr9	5256278	.	C	T
chr9	5257271	rs36049801	GT	G
chr9	5257334	rs77010723	C	T
chr9	5257430	rs11506293	A	G
chr9	5259615	rs148390822	G	A
chr9	5259620	rs12350079	C	T
chr9	5259667	rs111881603	C	T
chr9	5259847	rs7019418	G	C
chr9	5260079	rs1830610	C	T
chr9	5260940	rs7859286	G	T
chr9	5261238	rs6476956	A	G
chr9	5261324	rs6476957	A	G
chr9	5261440	rs7035456	G	A
chr9	5261536	.	A	G
chr9	5261794	rs7025005	T	C

#CHROM	POS	ID	REF	ALT
chr9	5262349	rs12351715	A	G
chr9	5262607	rs2381215	C	T
chr9	5264425	rs1322223	C	T
chr9	5266202	rs1555477	C	T
chr9	5267043	rs1575284	G	T
chr9	5271598	rs10758678	C	T
chr9	5272381	rs12683801	T	C
chr9	5274464	rs10975037	A	G
chr9	5274545	rs12684195	A	C
chr9	5274648	rs12684204	A	G
chr9	5275005	rs7869831	C	G
chr11	119077003	rs7108857	C	T
chr11	119144791	.	CTT	C,CT
chr11	119145705	rs3842642	G	GT
chr11	119146659	rs2511854	G	C
chr11	119148573	rs2510152	G	T
chr11	119155618	rs2298650	G	T
chr11	119172536	rs1047417	A	G
chr11	119175075	rs2509671	C	A
chr11	119175340	.	AT	A
chr11	119175422	rs2510145	T	C
chr11	119177938	rs11217234	A	G
chr12	111856673	rs78894077	C	T
chr12	111856738	rs7973120	T	A
chr12	111884608	rs3184504	T	C
chr12	111885351	rs111340708	ATGGGG	A
chr12	111886967	rs11065904	T	A
chr12	111887659	rs739496	A	G
chr12	111887974	rs11065905	G	A
chr20	30954295	rs2295454	A	G
chr20	30957801	rs9798511	C	T
chr20	30958216	.	CA	C
chr20	30958768	.	T	TA
chr20	31016314	rs3818190	A	G
chr20	31022959	rs6058694	T	C
chr20	31024274	rs4911231	T	C
chr20	31025163	rs2295764	A	G
chr20	31025231	rs2295763	T	C
chr20	31025535	rs2295762	A	G

Abbreviation; CHROM #, chromosome number; POS, start position; ID, variant identity; REF. the reference nucleotides; ALT, the observed nucleotides; chr, chromosome

A dot '.' under the column ID represents novel polymorphism

Table 6. 4 Summary of post-filtered genetic variants

Gene	Seq. Region	Predicted function						Total (%)
		missense	synonymous	indels	TFBS	Splicing (ESE or ESS)	coding	
RCL1/ JAK2	**intergenic**			1	1			5 (2.5)
JAK2	**5' UTR**				1			1 (0.5)
JAK2	**intron 2**			1				24 (12)
JAK2	**intron 3**							3 (1.5)
JAK2	**intron 4**			1				11 (5.5)
JAK2	**intron 5**							4 (2)
JAK2	**intron 8**			1				7 (3.5)
JAK2	**intron 9**							1 (0.5)
JAK2	**intron 10**							1 (0.5)
JAK2	**intron 12**			1				2 (1)
JAK2	**intron 13**							2 (1)
JAK2	**exon 14**	1					1	1 (0.5)
JAK2	**intron 14**			2				7 (3.5)
JAK2	**intron 15**							1 (0.5)
JAK2	**intron 16**							1 (0.5)
JAK2	**intron 18**							1 (0.5)
JAK2	**intron 19**							1 (0.5)
JAK2	**intron 22**			1	4	5		20 (10)
JAK2/ INSL6	**intergenic**			5				29 (14.5)
INSL6	**downstream**			1				2 (1)
INSL6	**intronic**							23 (11.5)
INSL6/ INSL4	**intergenic**							17 (8.5)
INSL4	**upstream**				2			2 (1)
INSL4	**exonic**		1			1	1	1 (0.5)
INSL4	**intronic**				1			2 (1)
INSL4/ RLN2	**intergenic**			1				31 (15.5)

This table summarises the 200 post-filtered variants according to different variant type. These variants were selected based on Fisher's exact test (P <0.01) and pFDR (q <0.01). All the variants were located on chromosome 9 spanning *JAK2* gene. Since each SNP is sometimes predicted to have more than one function (can sometimes have more than one predicted function) or none, thus an unequal total numbers of SNPs at the last column were expected.

Seq. Region, sequenced region; indel, insertion and deletion; TFBS, transcription factor binding site; ESE, exonic splicing enhancer; ESS, Exonic splicing silencer; 5' UTR, 5' untranslated region; *RCL1*, RNA terminal phosphate cyclase-like 1; *JAK2*, Janus kinase 2; *INSL6*, insulin-like 6; *INSL4*, insulin-like 4; *RLN2*, relaxin 2.

* The missense mutation in the coding region of exon 14 is the driver mutation in MPN, *JAK2* V617F.

Table 6. 5 MPN-associated functional variants within the 330kb recombination hotspots

SNP ID	Position	Region	Nearby Gene	Allele	Conservation [a]	Predicted function [b]	Reported disease association [c]
rs10758669	4971602	intergenic	*JAK2*/*RCL1*	A/C	0	TFBS	N/A
rs2274472	4975542	UTR5	*JAK2*	T/C	0	TFBS; RegPotential (0.3721)	• Acute Myeloid Leukaemia • Polycythemia vera induced Budd-Chiari syndrome • Chronic Myelomonocytic Leukemia • Crohn's disease • Leukemia • Myeloproliferative neoplasms • Ulcerative colitis • Venous thromboembolism (PMID: 16293597, 16762626, 17307838, 17440677, 17577920, 17440984, 19181784, 19287384, 19773259, 19734476, 19847199, 19915573, 18587394, 20228799)
rs10156475	4978341	intronic	*JAK2*	C/T	0.558	TFBS	
rs1327494	4989303	intronic	*JAK2*	A/G	0.981		
rs77375493*	5073770	coding	*JAK2*	G/T	RS_Score (460)[†]	SIFT (0) - damaging; Polyphen (0.987) -propably damaging;	
rs11788790	5082263	intronic	*JAK2*	G/T	0.003	RegPotential (0.3056)	
rs11788834	5082466	intronic	*JAK2*	A/G	0.015	Splicing (ESE or ESS); RegPotential (0.0499)	
rs10117459	5084034	intronic	*JAK2*	A/G	1	Splicing(ESE or ESS)	
rs12340866	5084185	intronic	*JAK2*	A/G	0	RegPotential (0.2374)	
rs3780370	5085167	intronic	*JAK2*	A/C	1	Splicing (ESE or ESS)	
rs3780371	5085191	intronic	*JAK2*	A/T	1	Splicing (ESE or ESS)	
rs10815154	5085538	intronic	*JAK2*	A/G	0.076	RegPotential (0.1593)	
rs7857081	5086728	intronic	*JAK2*	A/G	0.015	RegPotential (0.2760)	
rs7847141	5087171	intronic	*JAK2*	A/T	0.076	RegPotential (0.1834)	
rs3780372	5087544	intronic	*JAK2*	C/T	0.011	RegPotential (0.2440)	
rs3780373	5088223	intronic	*JAK2*	C/T	0.993	Splicing (ESE or ESS); RegPotential (0.1527)	
rs10121077	5088411	intronic	*JAK2*	A/G	0.018	RegPotential (0.0914)	

SNP ID	Position	Region	Nearby Gene	Allele	Conservation	Predicted function	Reported disease association
rs7851455	5090501	intronic	*JAK2*	C/G	0.007	RegPotential (0.1714)	(continued)
rs10815157	5098771	intronic	*JAK2*	C/G	0.974	TFBS; RegPotential (0.2537)	
rs11793659	5099707	intronic	*JAK2*	A/G	0.005	TFBS	
rs17425637	5100000	intronic	*JAK2*	C/T	0.968	TFBS; RegPotential (0.2410)	
rs3780377	5100899	intronic	*JAK2*	C/T	0	TFBS; RegPotential (0.2835)	
rs3824433	5103577	intronic	*JAK2*	C/T	0.163	TFBS; RegPotential (0.0304)	
rs10815164	5121312	intergenic	*JAK2*||*INSL6*	A/T	0.991		N/A
rs10758677	5178078	intergenic	*INSL6*||*INSL4*	A/G	0	TFBS	N/A
rs1322222	5180444	intergenic	*INSL6*||*INSL4*	A/G	0	TFBS	N/A
rs2093448	5220444	upstream	*INSL4*	G/T	0	TFBS	N/A
rs2104166	5220447	upstream	*INSL4*	C/T	0	TFBS	N/A

SNP ID, single nucleotide polymorphisms identity; *JAK2*, Janus kinase 2; *RCL1*, RNA terminal phosphate cyclase-like 1; TFBS, transcription factor binding site; N/A, Not Available; RegPotential, Regulatory potential; SIFT, **S**orting **I**ntolerant **F**rom **T**olerant; PolyPhen, **Poly**morphism **Phen**otyping; RS_score, Rejected Substitutions score for the conserved element as determined from GERP++ (Genomic Evolutionary Rate Profiling) package; ESE, exonic splicing enhancer; ESS, Exonic splicing silencer; PMID, unique identifier number used in PubMed; *INSL6*, insulin-like 6; *INSL4*, insulin-like 4. * rs77375493 is the SNP accession number for *JAK2* V617F mutation. [†] RS_score ranges from 0-1000, the larger the score, the more conserved the site.

[a] Conservation score: vertebrate Multiz Alignment and Conservation score(17 Species) downloaded from UCSC genome bioinformatics web site (http://genome.ucsc.edu/). The score ranges between 0 and 1: the larger the score, the more conserved the site.

[b] Online tools were used for functional prediction of variants, including and SNP Function Prediction (FuncPred), and SNPnexus. These tools incorporate databases/ tools such as SIFT (the lower the score [0-1], the more damaging it is) and PolyPhen (the higher the score [0-1], th emore damaging it is) to predict amino acid changes that affect protein function**.** Splicing (ESE or ESS) identified SNPs that are located at 2 base pair of intron-exon junction region that may act as ESE or ESS to disrupt splicing activity and cause alternative splicing. RegPotential scores predict variants that may regulate gene transcription (based on RP score ≥ 0).

[c] Databases were assessed for disease association, including COSMIC (Catalogue of Somatic Mutations in Cancer), and GAD (Genetic Association Database).

6.5 Discussion

Based on the results generated from part I and II of this study, we believed that some untyped variants that reside within the vicinity of *JAK2* are the disease-causing factors. Using LocusZoom, I further analysed the genotyping data and identified two recombination hotspots covering a region of 330kb (including 142kb of *JAK2*). With great interest on *JAK2*, the whole 330-kb region was sequenced alongside the exome of the other genes reported to be involved in the development of MPNs (**Table 6.1)**. Soon after the start of this NGS project, the search for driver mutations in MPN has made a breakthrough with the discovery of somatic mutations in the endoplasmic reticulum chaperone - Calreticulin (*CALR*). Besides guarding proper folding of newly synthesized glycoproteins and calcium homeostasis, *CALR* is also found to be involved in various biologic processes outside the endoplasmic reticulum. Patients harbouring a *CALR* mutation were reported to have a lower risk of thrombosis and thus they have longer overall survival compared with those harbouring a *JAK2* mutation (Klampfl et al., 2013). However, *CALR* was not included as the targeted gene although it was reported to be the second major driver mutation in MPNs. This was because *CALR* mutation was only detected in V617F- and MPL-negative ET and MF (Nangalia et al., 2013; Klampfl et al., 2013); while in this study, the number for such cases was small (**Table 3.1**). We have inadequate V617F-negative MPN cases and the sample recruitment was beyond our control. Most importantly, the stronger association detected between *JAK2* polymorphisms and V617F-positive MPNs urged me to start with V617F-positive MPN cases for this study.

It is important to analyse multiple genes/ SNPs and their interaction for more reliable prognostic or predictive biomarkers, because it is hard to understand complex disease

by studying only one gene. We expect to sequence many individuals to discover, screen, or validate genetic variation within this region among Hong Kong Chinese. Also, if we can pool samples and obtain high sequence coverage during a single run, we could identify rarer variants that are missed, or too expensive to identify, using Sanger sequencing. Existing technologies, e.g. Sanger sequencing, has low-throughput, i.e., time consuming & labour intensive to scan through a long genomic region in a large sample cohort. Therefore, deep sequencing was performed to identify and elucidate additional genetic defects causing MPNs that could serve as disease biomarkers.

In this study, the comparison could only be carried out in a qualitative manner because paired tissue samples were not available. Therefore I sought to identify the genetic variation underlying MPNs by comparing case with control samples. First, with the use of reference genome, any known and unknown variants along the sequenced region could be identified. Second, by taking normal healthy controls as reference, the disease-specific variants can be identified.

6.5.1 Identification of variants across the targeted regions

In this study, I used targeted sequencing method to reveal the genetic variation background of 9 genes in 48 V617F-positive MPN patients and 48 paired-controls. The final sequencing design covered approximately 88% of the targeted region. For each individual from the 96 samples, there were between 1% and 65.8% of reads that did not map on the reference genome. This constitutes an average of 28.5% of the unmapped reads. Research has shown that an overall mean sequencing coverage read depth of >30×; i.e., each position is sequenced at an average of more than 30 times,

enables detection of heterozygous variants with more than 99% accuracy (Ku et al., 2012; Bentley et al., 2008). The default filter identified 532 variants from the targeted region which were not classified into disease- or normal group and therefore serve as the variant profiles for the targeted regions. To determine MPN-specific variants, these 532 variants were further analysed.

6.5.2 Identification of MPN-associated variants

To distinguish case/control-specific variants from common variants, I considered the detection of alternate allele counts between case and control group. Qualitatively, I counted and grouped all the alleles other than reference allele as alternate allele. The difference between cases and controls harbouring alternate alleles (any alleles other than the ancestor allele) were then calculated. A cut-off point of pFDR ($q < 0.0.1$) based on Fisher's exact test ($P < 0.01$) identified 200 known variants residing within the two recombination hotspots. *JAK2* SNPs constituted 45.5% from the variant pool. Odds ratio between cases and controls for these variants also showed that the top 27 variants with highest ORs were *JAK2* variants. In addition, a cross-check with **Table 4.5** revealed that 11 of these variants were among the top 20 SNPs for logistic regression tests in *V617F*-positive MPNs, including *JAK2* SNP rs12342421 (S8), the most significantly associated SNP with MPNs.

To explore functional variants underlying MPNs, I used ANNOVAR server (Wang et al., 2010), SNPnexus (Dayem Ullah et al., 2012) and SNP Function Prediction (FuncPred) (Xu & Taylor, 2009) softwares to annotate biological functions. Out of the 200 variants, only rs77375493 and rs12720 were identified as coding SNPs of *JAK2* and *INSL4* (insulin-like 4 (placenta)) respectively. The rs number rs77375493

is the SNP accession number for V617F mutation, which is known to cause substitution of valine for phenylalanine at codon 617 within the JH2 domain of *JAK2* and disrupts the auto-inhibitory control of *JAK2* in the JAK2 STAT pathway (Ihle & Gilliland, 2007b; Kralovics et al., 2005b; James et al., 2005b). On the other hand, despite little knowledge about the biological function of *INSL4*, it was suggested to be an important polypeptide that regulates proper physiologic processes such as cellular proliferation and differentiation through membrane receptor tyrosine kinases (Shabanpoor et al., 2009; Koman et al., 1996). Nonetheless, it has not been reported to be associated with haematologic malignancies.

6.5.3 Identification of MPN-associated variants within the two recombination hotspots

Considering each gene separately and independently of each other and considering the 330-kb region as potential haplotype, I focused on the region within the two recombination hotspots arising from my genotyping results. Since Fisher's exact test could not pick up any significant variants from regions other than chromosome 9, I examined the 330-kb region using the pFDR (*q*-value) on top of Fisher's exact test. The *q*-value was used to estimate the false discovery rate among the set of variants within the region. Using pFDR approach, 202 variants were selected and most of them are known variants, only 3 are novel variants. Results generated based on pFDR cut-off point ($q < 0.01$) agreed with that generated based on Fisher's exact test ($P < 0.01$), with 179 overlapping variants residing within the recombination hotspots within the vicinity of *JAK2*. Despite the discovery of additional genetic alterations in MPNs, *JAK2* mutations are still the primary disease driver. Although more and more

additional genetic alterations were documented underlying MPNs, they were reported in lower prevalence (**Table 6.1**).

Both NGS and genotyping together with imputation (part I & II of this project) detected strong signal from intronic SNPs. Consistent results from these studies further strengthen our hypothesis that rs12342421 (S8) and other strongly associated polymorphisms that were non-functional, are in strong LD with some untyped causal variants underlying MPN. It is thus not surprising to find that all of the MPN-associated variants are located within chromosome 9.

6.5.4 Strengths and limitations of the study

The strengths of this part of my study could be summarised as follows. Unlike most of the NGS studies, I applied NGS with case-control pool as a mean for genetic association study. With the use of multiplexing approach, massive parallel sequencing of pooled samples can be done at lower costs and within a shorter time. Rather than just documenting all the variants residing within the sequencing region, disease-associated variants could be identified by comparing the genetic variation between cased and controls. With the high accuracy and sensitivity from NGS approach, the single marker S8 consistently appear to be the most significantly associated SNP. If we can identify disease causing variants from these intronic SNPs, we may discover the buried treasure that may help unravel the mystery within our genes (Cooper, 2010).

Although there is no gold standard for NGS study, certain limitation of this NGS study should also be acknowledged. First, there were no paired tissue samples/ disease-control samples from each individual. Therefore, I could not classify the variants into germinal or somatic. Besides, the sample size (48 cases and 48 controls) in this project might be underpowered to detect some rare variants. Nevertheless, results generated from this study may serve as preliminary data for a larger-scale NGS study. Careful design of the study and parameters, and sample recruitment are essential in a genetic association study using NGS approach to avoid identifying pool of variants with unknown significance. Carefully planned NGS study with larger study cohort, paired-sample collection, and careful selection of the sequencing panel might provide better insight towards unravelling the genetic variation underlying MPNs.

As the sequencing platforms advanced dramatically, they require the bioinformatics approaches to be improved, due to the rapid generation of an enormous amount of data and the current lack of appropriate data processing tools and standardisation. To date, researchers can rely on publicly available online tools and powerful bioinformatics software packages to analyse massive NGS data (Grada & Weinbrecht, 2013; Gogol-Doring & Chen, 2012). This limiting factor has raised concern so much so that it has become the concern of editorials and reviewers (Byrd et al., 2014). Therefore, the results should be interpreted carefully.

6.5 Conclusion

This project helps to guide the protocol for future larger-scale studies. Our NGS generated consistent results with that generated from genotyping: *JAK2* SNPs are the most significantly MPN-associated factors among the genes/variants studied in this

thesis. Larger-scale NGS of MPNs patients with paired -samples is likely to help us understand the genetic variation underlying each individual and thus provide early diagnosis and targeted treatment.

CHAPTER 7 Overall Discussion and Conclusion

7.1 Prevalence of *V617F* in Hong Kong Chinese population

Prior to genetic association study, the involvement of V617F in Chinese MPN patients were confirmed. The results from part I of study corroborate previous reports on the prevalence of V617F in other populations. The distribution of V617F of our Hong Kong MPN patients (PV, ET and PMF) is similar to worldwide reported statistics (Jones et al., 2005; Baxter et al., 2005; James et al., 2005b; Kralovics et al., 2005a). This finding justified part II study of this project.

7.2 Single marker predisposition to *V617F*-positive MPNs

The results arising from this part agreed with that reported in European populations. Confirming previous genetic association studies is the gold standard in genetic association studies to eliminate false positive and support the true associations (Konig, 2011; Ku et al., 2010). Replication in different cohorts increased the confidence before declaring a real association and before any further functional annotation can be given to the identified markers. Our study excluded the possibility of population-specific genetic architecture, for instance, epistasis (Greene et al., 2009). Moreover, replication across populations enables the identification of any common or unique risk alleles (Ku et al., 2010). In this study, we identified the risk effect of single marker rs12342421 (S8), rather than the reported haplotypes, in the development of V617F-positive MPN in Hong Kong Chinese. Replication and thus validation are necessary to identify casual variants for complex diseases at genetic level.

This study is the first to perform genotype imputation in genetic association studies of MPNs. Although the causal variant was still not found in this study, imputation together with conditional logistic regression further strengthens my confidence to conclude that rs12342421 (S8) contributed an independent effect to the most significant association between *JAK2* risk-haplotype and MPNs.

7.3 Meta-analysis showed the inherited predisposition in V617F-negative MPNs

There have been contradictory results reported for the association between germline *JAK2* haplotype and V617F-positive or V617F-negative MPNs. Our genotyping data showed that *JAK2* germline polymorphisms were associated with V617F-positive MPN patients in Hong Kong Chinese population.

In this study, single marker rs12342421 (S8), rather than the reported haplotypes, was identified to be the risk effect in the development of V617F-positive MPN in Hong Kong Chinese, by 3.55 fold for the minor allele C. Nonetheless, no significant correlation was found between V617F-negative MPN patients and the *JAK2* risk alleles. Inadequate statistical power is possibly the cause for the cumulative contrary results. Moreover, these studies were conducted mainly in Caucasians, only a few in Asian with small sample size. This urged us to conduct a comprehensive review of previous findings in confirming the germline association in MPNs irrespective of V617F mutational status and populations. Meta-analysis of overall population and subgroup concludes that *JAK2* polymorphisms are involved in the development of MPNs, irrespective of the V617F mutation status. A routine screening of such

germline polymorphisms may serve as a better surrogate for V617F in the diagnosis of MPNs. However, more effort is needed to generate a more comprehensive profile of disease-specific variants before such diagnostic tests are made available.

7.4 Next Generation Sequencing

Based on the results generated from part I and II of this study, a hypothesis was generated: some untyped *JAK2* variants are the disease-causing factors. Using LocusZoom, two recombination hotspots were identified covering a region of 330kb (including 142kb of *JAK2*). Therefore, targeted sequencing was performed to examine every possible variant within the region. All of the MPN cases selected were screened positive for the *JAK2* mutation. The reasons why this NGS study focused on only V617F-positive patients were because: 1) previous chapters showed that *JAK2* polymorphisms were strongly associated with V617F-positive MPNs; 2) beyond our control, sample size and a lack of paired-samples were the limiting factors in this NGS project. Therefore, with such constraints and also as a follow-up study of previous chapter, this part of NGS study examined only V617F-positive MPN cases and controls.

In this study, the genetic variation underlying MPNs were qualitatively analysed by comparing case with control samples. First, any known and unknown variant could be identified by comparing to a reference genome. Second, disease-specific variants could be identified when normal healthy controls were used as reference. With this approach, pFDR (q <0.0.1) based on Fisher's exact test (P <0.01) identified 200 known variants residing within the two recombination hotspots, and 45.5% from the variant pool were *JAK2* variants. The strongest signal (difference between cases and

controls) was again being detected in the *JAK2* gene and 11 of these variants were among the top 20 SNPs for logistic regression tests in *V617F*-positive MPNs, including *JAK2* SNP rs12342421 (S8), the most significantly associated SNP with MPNs (**Table 4.5**).

Functional annotation detected rs77375493 and rs12720 as coding SNPs of *JAK2* and *INSL4* (insulin-like 4 (placenta)) respectively. The SNP rs77375493 represents V617F mutation, which is known to disrupt the auto-inhibitory control of *JAK2* in the JAK2 STAT pathway (Ihle & Gilliland, 2007b; Kralovics et al., 2005b; James et al., 2005b). On the other hand, *INSL4* has not been reported to be associated with haematologic malignancies, despite little knowledge about its biological function of *INSL4.*

In this project, more than half of the identified variants reside within introns. This may be due to the fact that introns constitute more than 90% of our genes (most mutations were found in exons) (Seo et al., 2013). Although not directly involved in genomic functions, introns have the potential to regulate transcription and participate in genome organisation, depending on the intron life span phase (Chorev & Carmel, 2012). Introns can affect transcription when the process is coupled with splicing. In fact, splicing is a major property of introns throughout the intron life span phases. In contrary to the "bad" effects of introns, they are indispensable in the survival of some higher eukaryotes (Lynch, 2002). This implies that the conservation level of intron measures its importance in the genome. The higher conservative the intron is, the more important it is in the genome. Therefore, intron that is highly conservative

might also be functional of any type. **Table 6.5** documents a list of potential SNPs that may modulate the activity of *JAK2* gene.

Our NGS data help to revise protocol for a large-scale NGS study. If paired samples are available, future large-scale NGS study could possibly make individual disease profiling possible. This impacts on the management of individual MPN risk and provide early diagnosis and targeted treatment.

7.5 Conclusion

Polycythemia vera, essential thrombocythemia and primary myelofibrosis are clinically distinct MPNs associated with a spectrum of mutations and germline polymorphisms. This study of genetic variation used SNPs to tag genetic loci associated with MPNs to understand the differential predisposition to MPNs. We identified a subset of statistically significant SNPs. However, our knowledge of the SNP-to-function studies is inadequate. It is therefore important to find the potential biological impacts of such SNPs to identify the factors underlying susceptibility to MPNs.

In conclusion, NGS and genotyping together with imputation (part I & II of this project) detected strong signal from intronic SNPs. Consistent results from these studies further strengthen our hypothesis that rs12342421 (S8) and other strongly associated polymorphisms, are in strong LD with some untyped causal variants that are predisposing to MPN. It is thus not surprising to find that all of the MPN-associated variants are located within chromosome 9. However, functional investigation of these

SNPs is required to fully translate the underlying mechanism before they can be considered as disease-causing variants.

7.6 Future direction

To maintain the value of genetic association studies, potential functional mechanisms underlying these polymorphisms should be explored. In this study, we statistically identified the genetic polymorphisms underlying MPNs using a case-control study approach. Future studies should be focused on the follow up study based on the findings from this work. If a disease-causing SNP profile can be identified, early prevention of the disease is no longer a dream. Perhaps, the population can benefit from early diagnosis and also effective treatment.

7.6.1 Sanger sequencing validation of NGS SNP calls

Validation of suspected variants using another method is critical prior to accurate interpretation of the variants in the development of disease (Motoike et al., 2014). As NGS may be cost-ineffective for routine screening or SNP call validation, the flagged regions can be assessed by Sanger sequencing as described under Section 2.3.9. The Sanger sequencing results can then be compared with NGS data. Based on the initial positive NGS results, variants documented in **Table 6.5** (MPN-associated functional variants) are subjected to further validation.

7.6.2 From SNPs to functional studies

Upon variant call validation, functional studies are necessary to prove the involvement of SNPs in the development of a disease. Without functional analyses, any predicted functional variants are of unknown significance and may not be recognised to be clinically relevant. Follow up study can be conducted based on the data generated from our genotyping and NGS. **Table 6.5** documents a list of potential functional SNPs. Functional impact on splicing and transcription are the major property of SNPs selected from my NGS data. By employing cell lines (or more advanced level, the transgenic mice with modification), we may explore the mechanism for the association and interaction underneath the pathogenesis of MPNs.

7.6.2.1 Functional studies for novel variants

In this study, most of the identified SNPs were intronic SNPs with unknown biological impact. As most of the selected SNPs are predicted to have regulatory potential, we would therefore like to examine their ability to alter DNA-binding (Remes-Lenicov et al., 2007). Chromatin immunoprecipitation has always been a very powerful technique study the interaction between DNA and protein. This method is very effective and useful, but that requires a specific antibody directed against the protein of interest. This does not help to identify new, previously unknown proteins. We therefore modified a strategy to identify potential functional SNP.

Since there is no target DNA elements reported to be associated with the variants, we will perform a pull down assay to capture DNA: protein complex. Mass spectrometric analysis will then be used to identify the captured protein, if any. To capture DNA-protein complexes that are specific to a target sequence containing the SNP

allele, a biotinylated probe (short sequence) will be designed to include both the major and minor alleles in the middle. These biotinylated probes will be used to capture any protein that binds on the sequence by its ligand avidin in the pull down assay. The probes and avidin will then undergo overnight binding process on a rotation wheel at 4°C. Meanwhile, a source of transcription factor is required. MPN-relevant cell lines will be used for cell lysate preparation. V617F-positive cell lines such as MB-02 and UKE-1 will be used. They are acute myeloid leukaemia (AML) cell lines with histories of MPD/MDS (Quentmeier et al., 2006). Cell lysates will be pre-cleared with plain beads without probe for 2 hours on rotation wheel to wash away proteins that bind non-specifically to beads. Then the complex will be incubated with pre-cleared cell extract overnight on a rotation wheel at 4°C. After that, non-DNA-binding proteins are washed away; the remaining sequence-specific proteins will be eluted for SDS-PAGE followed by identification by mass spectrometric experiment.

Once we have identified some proteins, we will use electrophoretic mobility shift assay (EMSA) to identify the specific DNA-binding elements using the probes again. The reaction products are then analysed on a non-denaturing PAGE. Probes that do not bind any proteins or those that show same intensity for both common and rare variant are likely to bind outside the SNP location, will be rejected for further analysis. Then, luciferase reporter assay or supershift EMSA will be performed. Reporter assay can test the SNP effect on the transcriptional activity of a promoter while supershift EMSA can examine the ability of the remaining SNPs to alter specific protein of interest by using antibody.

7.6.2.2 Functional studies for known variants

Similar strategy proposed above can be used to study the functional impact for known variants. Since these variants were predicted to be involved in certain biological events, we may skip the pull down assay and mass spectrometric analysis for DNA elements/ protein identification and proceed directly from cell culture and molecular analyses for studying gene regulation and DNA: protein interaction.

Supplementary information

Supplementary Table 1 Information of MPN-associated genes examined using NGS

Targeted Gene	Chr	Start[1]	End[1]	PCR product size (bp)
MPL	chr1	43803375	43803679	305
MPL	chr1	43803425	43803648	224
MPL	chr1	43803637	43803985	349
MPL	chr1	43803720	43803952	233
MPL	chr1	43804113	43804470	358
MPL	chr1	43804163	43804441	279
MPL	chr1	43804842	43805322	481
MPL	chr1	43804892	43805290	399
MPL	chr1	43805540	43805878	339
MPL	chr1	43805585	43805847	263
MPL	chr1	43805960	43806268	309
MPL	chr1	43806008	43806234	227
MPL	chr1	43812016	43812376	361
MPL	chr1	43812066	43812350	285
MPL	chr1	43812330	43812688	359
MPL	chr1	43812413	43812655	243
MPL	chr1	43814414	43814759	346
MPL	chr1	43814464	43814723	260
MPL	chr1	43814838	43815112	275
MPL	chr1	43814884	43815080	197
MPL	chr1	43817788	43818055	268
MPL	chr1	43817837	43818024	188
MPL	chr1	43818094	43818977	884
MPL	chr1	43818139	43820185	2047
MPL	chr1	43819549	43820213	665
DNMT3A	chr2	25455746	25456639	894
DNMT3A	chr2	25455796	25457339	1544
DNMT3A	chr2	25456671	25457097	427
DNMT3A	chr2	25457101	25457373	273
DNMT3A	chr2	25458481	25458778	298
DNMT3A	chr2	25458526	25458744	219
DNMT3A	chr2	25459705	25459956	252
DNMT3A	chr2	25459755	25459924	170
DNMT3A	chr2	25461949	25462432	484
DNMT3A	chr2	25461953	25462473	521
DNMT3A	chr2	25463075	25463399	325
DNMT3A	chr2	25463121	25463369	249
DNMT3A	chr2	25463409	25463686	278
DNMT3A	chr2	25463459	25463649	191
DNMT3A	chr2	25464331	25464663	333
DNMT3A	chr2	25464381	25464626	246
DNMT3A	chr2	25466669	25466938	270
DNMT3A	chr2	25466717	25466901	185
DNMT3A	chr2	25466887	25467289	403
DNMT3A	chr2	25466974	25467257	284
DNMT3A	chr2	25467313	25467597	285
DNMT3A	chr2	25467359	25467571	213
DNMT3A	chr2	25468025	25468285	261
DNMT3A	chr2	25468072	25468251	180
DNMT3A	chr2	25468794	25469258	465
DNMT3A	chr2	25468839	25469228	390
DNMT3A	chr2	25469394	25469731	338
DNMT3A	chr2	25469439	25469695	257
DNMT3A	chr2	25469823	25470108	286
DNMT3A	chr2	25469870	25470077	208
DNMT3A	chr2	25470360	25470695	336
DNMT3A	chr2	25470410	25470668	259
DNMT3A	chr2	25470811	25471200	390
DNMT3A	chr2	25470856	25471171	316
DNMT3A	chr2	25472429	25472669	241
DNMT3A	chr2	25472476	25472643	168
DNMT3A	chr2	25474963	25475260	298
DNMT3A	chr2	25475013	25475234	222
DNMT3A	chr2	25497735	25498039	305
DNMT3A	chr2	25497760	25498006	247
DNMT3A	chr2	25498272	25498495	224
DNMT3A	chr2	25498319	25498462	144
DNMT3A	chr2	25504226	25505665	1440
DNMT3A	chr2	25504271	25505630	1360
DNMT3A	chr2	25522913	25523201	289
DNMT3A	chr2	25522958	25523162	205
DNMT3A	chr2	25536682	25537113	432
DNMT3A	chr2	25536732	25537080	349

DNMT3A	chr2	25564600	25564780	181
DNMT3A	chr2	25564645	25564834	190
DNMT3A	chr2	25565204	25565542	339
DNMT3A	chr2	25565249	25565509	261
DNMT3A	chr2	209100939	209101943	1005
DNMT3A	chr2	209103745	209104007	263
DNMT3A	chr2	209104537	209104777	241
DNMT3A	chr2	209106668	209106919	252
DNMT3A	chr2	209108101	209108378	278
DNMT3A	chr2	209109993	209110198	206
DNMT3A	chr2	209113043	209113434	392
DNMT3A	chr2	209116104	209116341	238
DNMT3A	chr2	209118560	209119096	537
DNMT3A	chr2	209119613	209119856	244
DNMT3A	chr2	209119907	209120205	299
DNMT3A	chr2	209120577	209120968	392
IDH1	chr2	209100893	209101977	1085
IDH1	chr2	209103699	209104038	340
IDH1	chr2	209104492	209104819	328
IDH1	chr2	209106621	209106947	327
IDH1	chr2	209108051	209108404	354
IDH1	chr2	209109943	209110181	239
IDH1	chr2	209112993	209113458	466
IDH1	chr2	209116098	209116367	270
IDH1	chr2	209118510	209119119	610
IDH1	chr2	209119563	209119887	325
IDH1	chr2	209119835	209120237	403
IDH1	chr2	209120722	209120954	233
TET2	chr4	106066934	106067328	395
TET2	chr4	106066982	106067293	312
TET2	chr4	106067355	106068209	855
TET2	chr4	106067400	106068186	787
TET2	chr4	106111422	106111742	321
TET2	chr4	106111467	106111712	246
TET2	chr4	106155004	106159476	4473
TET2	chr4	106155004	106164134	9131
TET2	chr4	106159539	106159631	93
TET2	chr4	106159644	106159887	244
TET2	chr4	106159904	106160200	297
TET2	chr4	106160274	106160395	122
TET2	chr4	106160454	106160576	123
TET2	chr4	106160599	106160779	181
TET2	chr4	106161084	106161258	175
TET2	chr4	106161344	106163434	2091
TET2	chr4	106163444	106164167	724
TET2	chr4	106164627	106165007	381
TET2	chr4	106164677	106164985	309
TET2	chr4	106180681	106181009	329
TET2	chr4	106180726	106180976	251
TET2	chr4	106182816	106183090	275
TET2	chr4	106182866	106183055	190
TET2	chr4	106190672	106190953	282
TET2	chr4	106190717	106190954	238
TET2	chr4	106193661	106194152	492
TET2	chr4	106193671	106194125	455
TET2	chr4	106196106	106198500	2395
TET2	chr4	106196155	106201023	4869
TET2	chr4	106198533	106199317	785
TET2	chr4	106199350	106201037	1688
EZH2	chr7	148504368	148504885	518
EZH2	chr7	148504414	148504848	435
EZH2	chr7	148506067	148506326	260
EZH2	chr7	148506113	148506297	185
EZH2	chr7	148506267	148506568	302
EZH2	chr7	148506352	148506532	181
EZH2	chr7	148507328	148507590	263
EZH2	chr7	148507375	148507556	182
EZH2	chr7	148508619	148508897	279
EZH2	chr7	148508667	148508862	196
EZH2	chr7	148510952	148511310	359
EZH2	chr7	148511001	148511279	279
EZH2	chr7	148511909	148512217	309
EZH2	chr7	148511956	148512181	226
EZH2	chr7	148512500	148512724	225
EZH2	chr7	148512548	148512688	141
EZH2	chr7	148513676	148513957	282
EZH2	chr7	148513726	148513920	195

EZH2	chr7	148514215	148514565	351
EZH2	chr7	148514264	148514533	270
EZH2	chr7	148514869	148515294	426
EZH2	chr7	148514919	148515259	341
EZH2	chr7	148516593	148516861	269
EZH2	chr7	148516638	148516829	192
EZH2	chr7	148523448	148523802	355
EZH2	chr7	148523496	148523774	279
EZH2	chr7	148524156	148524432	277
EZH2	chr7	148524206	148524408	203
EZH2	chr7	148525737	148526047	311
EZH2	chr7	148525782	148526022	241
EZH2	chr7	148526725	148527014	290
EZH2	chr7	148526770	148526990	221
EZH2	chr7	148529631	148529922	292
EZH2	chr7	148529676	148529892	217
EZH2	chr7	148543462	148543774	313
EZH2	chr7	148543512	148543740	229
EZH2	chr7	148544175	148544483	309
EZH2	chr7	148544224	148544447	224
EZH2	chr7	148580193	148581515	1323
EZH2	chr7	148580239	148581491	1253
JAK2	chr9	4945009	4945177	169
JAK2	chr9	4945285	4945360	76
JAK2	chr9	4945829	4945978	150
JAK2	chr9	4945990	4946106	117
JAK2	chr9	4946380	4946647	268
JAK2	chr9	4946930	4947390	461
JAK2	chr9	4947485	4947719	235
JAK2	chr9	4953825	4953965	141
JAK2	chr9	4954105	4954526	422
JAK2	chr9	4954537	4954848	312
JAK2	chr9	4954870	4955199	330
JAK2	chr9	4955489	4955905	417
JAK2	chr9	4955913	4956058	146
JAK2	chr9	4956060	4956479	420
JAK2	chr9	4957115	4958686	1572
JAK2	chr9	4958750	4958830	81
JAK2	chr9	4958980	4959106	127
JAK2	chr9	4959110	4960167	1058
JAK2	chr9	4960235	4960535	301
JAK2	chr9	4960825	4961347	523
JAK2	chr9	4961382	4961546	165
JAK2	chr9	4961620	4961761	142
JAK2	chr9	4961774	4961934	161
JAK2	chr9	4962030	4962371	342
JAK2	chr9	4962395	4963113	719
JAK2	chr9	4963140	4964067	928
JAK2	chr9	4964069	4966621	2553
JAK2	chr9	4966770	4968384	1615
JAK2	chr9	4968645	4968918	274
JAK2	chr9	4968950	4969471	522
JAK2	chr9	4969490	4970326	837
JAK2	chr9	4970595	4972063	1469
JAK2	chr9	4972265	4973964	1700
JAK2	chr9	4974022	4974264	243
JAK2	chr9	4974350	4974511	162
JAK2	chr9	4974795	4975090	296
JAK2	chr9	4975370	4980929	5560
JAK2	chr9	4981279	4987473	6195
JAK2	chr9	4987755	4996231	8477
JAK2	chr9	4996565	4996936	372
JAK2	chr9	4997315	4997541	227
JAK2	chr9	4997590	4998266	677
JAK2	chr9	4998545	4998743	199
JAK2	chr9	4998760	4998846	87
JAK2	chr9	4999110	4999534	425
JAK2	chr9	4999725	5001635	1911
JAK2	chr9	5001655	5003458	1804
JAK2	chr9	5003510	5003940	431
JAK2	chr9	5004020	5004137	118
JAK2	chr9	5004170	5004392	223
JAK2	chr9	5004690	5004925	236
JAK2	chr9	5005310	5005831	522
JAK2	chr9	5006415	5006560	146
JAK2	chr9	5006590	5006695	106
JAK2	chr9	5006699	5006794	96
JAK2	chr9	5006817	5007102	286

JAK2	chr9	5007105	5007666	562
JAK2	chr9	5008015	5008910	896
JAK2	chr9	5008914	5009445	532
JAK2	chr9	5009449	5009780	332
JAK2	chr9	5009960	5010041	82
JAK2	chr9	5010050	5010337	288
JAK2	chr9	5010608	5011055	448
JAK2	chr9	5011060	5011255	196
JAK2	chr9	5011460	5013777	2318
JAK2	chr9	5013780	5014166	387
JAK2	chr9	5014290	5015525	1236
JAK2	chr9	5015640	5016518	879
JAK2	chr9	5016529	5017511	983
JAK2	chr9	5017554	5017811	258
JAK2	chr9	5017840	5017949	110
JAK2	chr9	5017950	5018141	192
JAK2	chr9	5018145	5018325	181
JAK2	chr9	5018630	5018815	186
JAK2	chr9	5018879	5019246	368
JAK2	chr9	5019299	5021629	2331
JAK2	chr9	5021920	5023040	1121
JAK2	chr9	5023070	5023236	167
JAK2	chr9	5023310	5023983	674
JAK2	chr9	5024260	5025468	1209
JAK2	chr9	5025809	5027388	1580
JAK2	chr9	5027510	5027719	210
JAK2	chr9	5027785	5028188	404
JAK2	chr9	5028260	5028343	84
JAK2	chr9	5028370	5028599	230
JAK2	chr9	5028600	5028700	101
JAK2	chr9	5028710	5028778	69
JAK2	chr9	5028795	5029097	303
JAK2	chr9	5029100	5029196	97
JAK2	chr9	5029210	5029374	165
JAK2	chr9	5029455	5030484	1030
JAK2	chr9	5030485	5031814	1330
JAK2	chr9	5032045	5032121	77
JAK2	chr9	5037855	5038479	625
JAK2	chr9	5038605	5039362	758
JAK2	chr9	5039445	5039612	168
JAK2	chr9	5039615	5039707	93
JAK2	chr9	5039710	5040244	535
JAK2	chr9	5040295	5040408	114
JAK2	chr9	5040410	5042132	1723
JAK2	chr9	5042425	5043318	894
JAK2	chr9	5043385	5043534	150
JAK2	chr9	5043629	5043749	121
JAK2	chr9	5043825	5043952	128
JAK2	chr9	5043955	5045434	1480
JAK2	chr9	5045510	5045693	184
JAK2	chr9	5045813	5045927	115
JAK2	chr9	5045935	5046144	210
JAK2	chr9	5046150	5046249	100
JAK2	chr9	5046260	5046435	176
JAK2	chr9	5046466	5046536	71
JAK2	chr9	5046563	5046630	68
JAK2	chr9	5046675	5047656	982
JAK2	chr9	5047889	5048146	258
JAK2	chr9	5048430	5050298	1869
JAK2	chr9	5050555	5051975	1421
JAK2	chr9	5051985	5052731	747
JAK2	chr9	5052869	5053215	347
JAK2	chr9	5053325	5054524	1200
JAK2	chr9	5054530	5057400	2871
JAK2	chr9	5057430	5057585	156
JAK2	chr9	5057900	5058143	244
JAK2	chr9	5058250	5058323	74
JAK2	chr9	5058590	5058756	167
JAK2	chr9	5058959	5060066	1108
JAK2	chr9	5060080	5060590	511
JAK2	chr9	5060700	5060865	166
JAK2	chr9	5060903	5061084	182
JAK2	chr9	5061165	5061424	260
JAK2	chr9	5061623	5061807	185
JAK2	chr9	5061845	5061918	74
JAK2	chr9	5061930	5062508	579
JAK2	chr9	5062525	5062904	380
JAK2	chr9	5062910	5063880	971

JAK2	chr9	5064140	5064267	128
JAK2	chr9	5064540	5066984	2445
JAK2	chr9	5067015	5067883	869
JAK2	chr9	5068185	5070307	2123
JAK2	chr9	5070310	5070562	253
JAK2	chr9	5070625	5074417	3793
JAK2	chr9	5074463	5074644	182
JAK2	chr9	5074645	5075109	465
JAK2	chr9	5075240	5075342	103
JAK2	chr9	5075365	5075500	136
JAK2	chr9	5075870	5076139	270
JAK2	chr9	5076159	5076289	131
JAK2	chr9	5076365	5076939	575
JAK2	chr9	5076945	5077149	205
JAK2	chr9	5077230	5079655	2426
JAK2	chr9	5079830	5080890	1061
JAK2	chr9	5081180	5082064	885
JAK2	chr9	5082980	5084497	1518
JAK2	chr9	5084550	5087129	2580
JAK2	chr9	5087465	5088613	1149
JAK2	chr9	5088620	5088752	133
JAK2	chr9	5088759	5089262	504
JAK2	chr9	5089580	5091027	1448
JAK2	chr9	5091045	5093233	2189
JAK2	chr9	5093315	5093793	479
JAK2	chr9	5093794	5095014	1221
JAK2	chr9	5095038	5096614	1577
JAK2	chr9	5096618	5098705	2088
JAK2	chr9	5098983	5100945	1963
JAK2	chr9	5101259	5101340	82
JAK2	chr9	5103640	5103722	83
JAK2	chr9	5106005	5106094	90
JAK2	chr9	5107030	5107139	110
JAK2	chr9	5107157	5108240	1084
JAK2	chr9	5108265	5109360	1096
JAK2	chr9	5109395	5110217	823
JAK2	chr9	5110235	5113202	2968
JAK2	chr9	5113210	5113426	217
JAK2	chr9	5113440	5114942	1503
JAK2	chr9	5116065	5116998	934
JAK2	chr9	5116999	5117209	211
JAK2	chr9	5117219	5117675	457
JAK2	chr9	5117685	5118106	422
JAK2	chr9	5118164	5119862	1699
JAK2	chr9	5119880	5120010	131
JAK2	chr9	5120047	5123439	3393
JAK2	chr9	5123600	5123765	166
JAK2	chr9	5124005	5124234	230
JAK2	chr9	5124240	5124351	112
JAK2	chr9	5124772	5128082	3311
JAK2	chr9	5128110	5130706	2597
JAK2	chr9	5131015	5131445	431
JAK2	chr9	5131720	5132549	830
JAK2	chr9	5132560	5133466	907
JAK2	chr9	5134035	5134140	106
JAK2	chr9	5138735	5140744	2010
JAK2	chr9	5142075	5142154	80
JAK2	chr9	5142400	5142587	188
JAK2	chr9	5142998	5143072	75
JAK2	chr9	5143310	5143388	79
JAK2	chr9	5143440	5143670	231
JAK2	chr9	5143940	5144025	86
JAK2	chr9	5144295	5144541	247
JAK2	chr9	5144610	5144687	78
JAK2	chr9	5145030	5145118	89
JAK2	chr9	5145245	5145370	126
JAK2	chr9	5145399	5145474	76
JAK2	chr9	5145545	5145636	92
JAK2	chr9	5145665	5145870	206
JAK2	chr9	5145941	5146033	93
JAK2	chr9	5146129	5147674	1546
JAK2	chr9	5147725	5149598	1874
JAK2	chr9	5149600	5149695	96
JAK2	chr9	5149736	5149804	69
JAK2	chr9	5149933	5150088	156
JAK2	chr9	5150111	5150286	176
JAK2	chr9	5150479	5150594	116
JAK2	chr9	5150595	5150822	228

JAK2	chr9	5150925	5151177	253
JAK2	chr9	5151195	5152777	1583
JAK2	chr9	5153290	5153374	85
JAK2	chr9	5155475	5155537	63
JAK2	chr9	5155930	5156879	950
JAK2	chr9	5157025	5157922	898
JAK2	chr9	5158170	5158440	271
JAK2	chr9	5158480	5158557	78
JAK2	chr9	5158591	5158660	70
JAK2	chr9	5158695	5159030	336
JAK2	chr9	5159031	5159107	77
JAK2	chr9	5159145	5159217	73
JAK2	chr9	5159219	5159338	120
JAK2	chr9	5159370	5159448	79
JAK2	chr9	5159450	5159544	95
JAK2	chr9	5159564	5159737	174
JAK2	chr9	5159770	5159898	129
JAK2	chr9	5160340	5160540	201
JAK2	chr9	5160585	5161146	562
JAK2	chr9	5161220	5161327	108
JAK2	chr9	5161358	5161426	69
JAK2	chr9	5161463	5161567	105
JAK2	chr9	5161599	5161674	76
JAK2	chr9	5161675	5161746	72
JAK2	chr9	5162075	5164683	2609
JAK2	chr9	5165236	5165349	114
JAK2	chr9	5165350	5165542	193
JAK2	chr9	5165585	5166210	626
JAK2	chr9	5166220	5166501	282
JAK2	chr9	5166550	5168034	1485
JAK2	chr9	5168085	5168341	257
JAK2	chr9	5168395	5168495	101
JAK2	chr9	5168540	5168650	111
JAK2	chr9	5168724	5168853	130
JAK2	chr9	5169212	5169339	128
JAK2	chr9	5169557	5169639	83
JAK2	chr9	5170329	5170456	128
JAK2	chr9	5171380	5171453	74
JAK2	chr9	5171645	5171738	94
JAK2	chr9	5172129	5172261	133
JAK2	chr9	5172352	5172426	75
JAK2	chr9	5173554	5173624	71
JAK2	chr9	5173785	5174948	1164
JAK2	chr9	5175220	5175640	421
JAK2	chr9	5175885	5175992	108
JAK2	chr9	5176015	5177878	1864
JAK2	chr9	5178155	5178603	449
JAK2	chr9	5178750	5178832	83
JAK2	chr9	5179975	5180071	97
JAK2	chr9	5180484	5180616	133
JAK2	chr9	5180840	5181028	189
JAK2	chr9	5181090	5183779	2690
JAK2	chr9	5183780	5187242	3463
JAK2	chr9	5187245	5187619	375
JAK2	chr9	5187920	5189344	1425
JAK2	chr9	5189360	5196407	7048
JAK2	chr9	5196480	5196592	113
JAK2	chr9	5196665	5196757	93
JAK2	chr9	5197045	5198320	1276
JAK2	chr9	5198333	5198691	359
JAK2	chr9	5198699	5198822	124
JAK2	chr9	5198855	5200538	1684
JAK2	chr9	5200555	5204008	3454
JAK2	chr9	5204125	5204223	99
JAK2	chr9	5204339	5206367	2029
JAK2	chr9	5206454	5206605	152
JAK2	chr9	5207105	5207205	101
JAK2	chr9	5207230	5207342	113
JAK2	chr9	5207420	5207507	88
JAK2	chr9	5207710	5207812	103
JAK2	chr9	5207825	5207955	131
JAK2	chr9	5207959	5208078	120
JAK2	chr9	5208120	5208196	77
JAK2	chr9	5208199	5208274	76
JAK2	chr9	5208313	5208698	386
JAK2	chr9	5208719	5208792	74
JAK2	chr9	5208895	5209010	116
JAK2	chr9	5209060	5209244	185

JAK2	chr9	5209362	5209464	103
JAK2	chr9	5209511	5209580	70
JAK2	chr9	5209595	5209696	102
JAK2	chr9	5209760	5210003	244
JAK2	chr9	5210115	5212816	2702
JAK2	chr9	5212820	5213223	404
JAK2	chr9	5213225	5214093	869
JAK2	chr9	5214327	5214918	592
JAK2	chr9	5215985	5216529	545
JAK2	chr9	5216785	5220291	3507
JAK2	chr9	5220355	5220437	83
JAK2	chr9	5220710	5220886	177
JAK2	chr9	5220910	5222995	2086
JAK2	chr9	5223104	5223307	204
JAK2	chr9	5223505	5223579	75
JAK2	chr9	5223604	5223685	82
JAK2	chr9	5223910	5224044	135
JAK2	chr9	5224390	5224564	175
JAK2	chr9	5224670	5224747	78
JAK2	chr9	5224803	5224888	86
JAK2	chr9	5225149	5225278	130
JAK2	chr9	5226175	5226255	81
JAK2	chr9	5227665	5227766	102
JAK2	chr9	5227890	5228057	168
JAK2	chr9	5228090	5228792	703
JAK2	chr9	5228825	5228990	166
JAK2	chr9	5229015	5229523	509
JAK2	chr9	5229545	5229757	213
JAK2	chr9	5229818	5229888	71
JAK2	chr9	5229905	5230001	97
JAK2	chr9	5230030	5230204	175
JAK2	chr9	5230244	5230991	748
JAK2	chr9	5230999	5234205	3207
JAK2	chr9	5234235	5236354	2120
JAK2	chr9	5236365	5236715	351
JAK2	chr9	5236718	5238162	1445
JAK2	chr9	5238164	5238754	591
JAK2	chr9	5238765	5239326	562
JAK2	chr9	5239360	5239787	428
JAK2	chr9	5243264	5245263	2000
JAK2	chr9	5245264	5246424	1161
JAK2	chr9	5247400	5247833	434
JAK2	chr9	5247835	5248205	371
JAK2	chr9	5248234	5248494	261
JAK2	chr9	5248610	5248833	224
JAK2	chr9	5248890	5248971	82
JAK2	chr9	5249015	5249121	107
JAK2	chr9	5249285	5249358	74
JAK2	chr9	5249438	5249513	76
JAK2	chr9	5249549	5249656	108
JAK2	chr9	5250318	5250398	81
JAK2	chr9	5250725	5250798	74
JAK2	chr9	5250920	5251026	107
JAK2	chr9	5251225	5251323	99
JAK2	chr9	5251420	5251895	476
JAK2	chr9	5251940	5252428	489
JAK2	chr9	5252500	5252675	176
JAK2	chr9	5252848	5252932	85
JAK2	chr9	5253040	5253143	104
JAK2	chr9	5253235	5253487	253
JAK2	chr9	5253500	5253588	89
JAK2	chr9	5253630	5253819	190
JAK2	chr9	5253963	5254048	86
JAK2	chr9	5254120	5254207	88
JAK2	chr9	5256197	5256276	80
JAK2	chr9	5256385	5256874	490
JAK2	chr9	5256930	5257250	321
JAK2	chr9	5257295	5257518	224
JAK2	chr9	5257574	5258339	766
JAK2	chr9	5258340	5258451	112
JAK2	chr9	5258505	5258784	280
JAK2	chr9	5258838	5258914	77
JAK2	chr9	5259384	5259502	119
JAK2	chr9	5259910	5260155	246
JAK2	chr9	5260168	5260272	105
JAK2	chr9	5260290	5260520	231
JAK2	chr9	5260555	5260734	180
JAK2	chr9	5260815	5260971	157

JAK2	chr9	5261020	5261561	542
JAK2	chr9	5261580	5261695	116
JAK2	chr9	5261700	5261805	106
JAK2	chr9	5261845	5262460	616
JAK2	chr9	5262489	5262612	124
JAK2	chr9	5262615	5263569	955
JAK2	chr9	5263599	5263965	367
JAK2	chr9	5264085	5264531	447
JAK2	chr9	5264569	5267171	2603
JAK2	chr9	5267175	5269167	1993
JAK2	chr9	5269170	5269845	676
JAK2	chr9	5269850	5270558	709
JAK2	chr9	5270560	5270658	99
JAK2	chr9	5270859	5271054	196
JAK2	chr9	5271115	5271201	87
JAK2	chr9	5271300	5271425	126
JAK2	chr9	5271905	5275036	3132
CBL	chr11	119076655	119077009	355
CBL	chr11	119076702	119077372	671
CBL	chr11	119077022	119077406	385
CBL	chr11	119103060	119103487	428
CBL	chr11	119103108	119103455	348
CBL	chr11	119142395	119142667	273
CBL	chr11	119142395	119142641	247
CBL	chr11	119144480	119144801	322
CBL	chr11	119144528	119144784	257
CBL	chr11	119145467	119145747	281
CBL	chr11	119145492	119145713	222
CBL	chr11	119146607	119146922	316
CBL	chr11	119146657	119146894	238
CBL	chr11	119148369	119148642	274
CBL	chr11	119148417	119148604	188
CBL	chr11	119148776	119149090	315
CBL	chr11	119148826	119149057	232
CBL	chr11	119149125	119149503	379
CBL	chr11	119149170	119149473	304
CBL	chr11	119155579	119156332	754
CBL	chr11	119155629	119156326	698
CBL	chr11	119158464	119158740	277
CBL	chr11	119158512	119158706	195
CBL	chr11	119167531	119167821	291
CBL	chr11	119167578	119167794	217
CBL	chr11	119167998	119168281	284
CBL	chr11	119168044	119168241	198
CBL	chr11	119168968	119169337	370
CBL	chr11	119169018	119169300	283
CBL	chr11	119170105	119170856	752
CBL	chr11	119170155	119178909	8755
CBL	chr11	119170860	119175350	4491
CBL	chr11	119175625	119177632	2008
CBL	chr11	119177655	119178944	1290
LNK	chr12	111843657	111844157	501
LNK	chr12	111843702	111844131	430
LNK	chr12	111855827	111856756	930
LNK	chr12	111855873	111856731	859
LNK	chr12	111884457	111885108	652
LNK	chr12	111884507	111885073	567
LNK	chr12	111885002	111885363	362
LNK	chr12	111885084	111885398	315
LNK	chr12	111885364	111885433	70
LNK	chr12	111885364	111885718	355
LNK	chr12	111885410	111885681	272
LNK	chr12	111885650	111887740	2091
LNK	chr12	111885737	111889477	3741
LNK	chr12	111888017	111889509	1493
ASXL1	chr20	30946097	30946685	589
ASXL1	chr20	30946137	30946707	571
ASXL1	chr20	30954087	30954349	263
ASXL1	chr20	30954137	30954319	183
ASXL1	chr20	30955460	30955606	147
ASXL1	chr20	30955480	30955582	103
ASXL1	chr20	30956715	30957062	348
ASXL1	chr20	30956765	30960402	3638
ASXL1	chr20	30957345	30957525	181
ASXL1	chr20	30957810	30958224	415
ASXL1	chr20	30958225	30958680	456
ASXL1	chr20	30958930	30959112	183

ASXL1	chr20	30959675	30959972	298
ASXL1	chr20	30959975	30960403	429
ASXL1	chr20	31015836	31016311	476
ASXL1	chr20	31015881	31016275	395
ASXL1	chr20	31017044	31017316	273
ASXL1	chr20	31017091	31017284	194
ASXL1	chr20	31017606	31017943	338
ASXL1	chr20	31017654	31017906	253
ASXL1	chr20	31019074	31019560	487
ASXL1	chr20	31019074	31019532	459
ASXL1	chr20	31020586	31020870	285
ASXL1	chr20	31020633	31020838	206
ASXL1	chr20	31020987	31021801	815
ASXL1	chr20	31021037	31021770	734
ASXL1	chr20	31022139	31023114	976
ASXL1	chr20	31022185	31027172	4988
ASXL1	chr20	31023120	31025981	2862
ASXL1	chr20	31026000	31027208	1209

Abbreviation; Chr, Chromosome; *JAK2*, Janus kinase 2; *MPL*, Myeloproliferative leukaemia virus oncogene; *LNK*, SH2B adaptor protein 3; *CBL*, Casitas B-cell lymphoma; *IDH1*, Isocitrate dehydrogenase 1; *TET2*, Ten-Eleven-Translocation 2, *DNMT3A*, DNA (cytosine-5-)-methyltransferase 3 alpha, *ASXL1*, Additional sex combs like transcriptional regulator 1; *EZH2*, Enhancer of zeste homolog; PV, polycythaemia vera; ET, essential thrombocythaemia; PMF, primary myelofibrosis. Selected regions included 50bp upstream and downstream of the targeted regions.

[1] Start and End refer to locations in UCSC hg19 **(GRCh37)** assembly, the Homo sapiens genome deposited February 2009.

References

2003. The International HapMap Project. *Nature,* 426**,** 789-96.

ABDEL-WAHAB, O. Genetics of the myeloproliferative neoplasms. *Curr Opin Hematol,* 18**,** 117-23.

ABDEL-WAHAB, O. 2011. Genetics of the myeloproliferative neoplasms. *Current opinion in hematology,* 18**,** 117-23.

ABDEL-WAHAB, O. & LEVINE, R. 2011. Genetics of the Myeloproliferative Neoplasms. *In:* VERSTOVSEK, S. & TEFFERI, A. (eds.) *Myeloproliferative Neoplasms.* Humana Press.

ABDEL-WAHAB, O.,MANSHOURI, T.,PATEL, J.,HARRIS, K.,YAO, J.,HEDVAT, C.,HEGUY, A.,BUESO-RAMOS, C.,KANTARJIAN, H.,LEVINE, R. L. & VERSTOVSEK, S. 2010. Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Res,* 70**,** 447-52.

ABDEL-WAHAB, O.,PARDANANI, A.,PATEL, J.,WADLEIGH, M.,LASHO, T.,HEGUY, A.,BERAN, M.,GILLILAND, D. G.,LEVINE, R. L. & TEFFERI, A. 2011a. Concomitant analysis of EZH2 and ASXL1 mutations in myelofibrosis, chronic myelomonocytic leukemia and blast-phase myeloproliferative neoplasms. *Leukemia,* 25**,** 1200-2.

ABDEL-WAHAB, O.,PARDANANI, A.,RAMPAL, R.,LASHO, T. L.,LEVINE, R. L. & TEFFERI, A. 2011b. DNMT3A mutational analysis in primary myelofibrosis, chronic myelomonocytic leukemia and advanced phases of myeloproliferative neoplasms. *Leukemia,* 25**,** 1219-20.

ADELAIDE, J.,PEROT, C.,GELSI-BOYER, V.,PAUTAS, C.,MURATI, A.,COPIE-BERGMAN, C.,IMBERT, M.,CHAFFANET, M.,BIRNBAUM, D. & MOZZICONACCI, M. J. 2006. A t(8;9) translocation with PCM1-JAK2 fusion in a patient with T-cell lymphoma. *Leukemia,* 20**,** 536-537.

ALABDULAALI, M. K. 2009. The role of JAK2 abnormalities in hematologic neoplasms. *Hematology Reviews,* 1**,** 6.

ALARCON-RIQUELME, M. E. 2004. Role of RUNX in autoimmune diseases linking rheumatoid arthritis, psoriasis and lupus. *Arthritis research & therapy,* 6**,** 169-73.

ALBERTS, B.,JOHNSON, A.,LEWIS, J.,RAFF, M.,ROBERTS, K. & WALTER, P. 2002. Studying Gene Expression and Function. *Molecular Biology of the Cell.* 4th edition ed. New York: Garland Science.

ALVAREZ-LARRAN, A.,ANGONA, A.,MARTINEZ-AVILES, L.,BELLOSILLO, B. & BESSES, C. 2012. Influence of JAK2 46/1 haplotype in the natural evolution of JAK2V617F allele burden in patients with myeloproliferative neoplasms. *Leuk Res,* 36**,** 324-6.

ALVAREZ, J. V. & FRANK, D. A. 2004. Genome-wide analysis of STAT target genes: elucidating the mechanism of STAT-mediated oncogenesis. *Cancer Biol Ther,* 3**,** 1045-50.

ANDERSON, L. A.,DUNCOMBE, A. S.,HUGHES, M.,MILLS, M. E.,WILSON, J. C. & MCMULLIN, M. F. 2012. Environmental, lifestyle, and familial/ethnic factors associated with myeloproliferative neoplasms. *American Journal of Hematology,* 87**,** 175-182.

ANDREAS, R.,CHRISTOPH, W.,ANN, W.,CLAUDIA, S.,ILONA, B.,BRIGITTE, S.,UTE, B.,NICK, T.,SHILANI, A.,JOHN, A. Y.,DANNY, V.,HELEN, F. B.,PETER, C. T.,AISLING, O. D.,FABIO, B.,CORNELIA, R.,HANS-

JOCHEM, K.,ANDREAS, H.,RÜDIGER, H.,ANDREW, C. & NICHOLAS, C. P. C. 2005. The t(8;9)(p22;p24) Is a Recurrent Abnormality in Chronic and Acute Leukemia that Fuses PCM1 to JAK2. *Cancer Research,* 65**,** 2662-2667.

ANTONIOLI, E.,GUGLIELMELLI, P.,POLI, G.,BOGANI, C.,PANCRAZZI, A.,LONGO, G.,PONZIANI, V.,TOZZI, L.,PIERI, L.,SANTINI, V.,BOSI, A. & VANNUCCHI, A. M. 2008. Influence of JAK2V617F allele burden on phenotype in essential thrombocythemia. *Haematologica,* 93**,** 41-8.

AYAD, M. W. & NAFEA, D. 2011. Acquired mutation of the tyrosine kinase JAK2V617F in Egyptian patients with myeloid disorders. *Genetic testing and molecular biomarkers,* 15**,** 17-21.

BAKER, S. J.,RANE, S. G. & REDDY, E. P. 2007. Hematopoietic cytokine receptor signaling. *Oncogene,* 26**,** 6724-37.

BAROSI, G. 2011. Conventional and Investigational Therapy for Primary Myelofibrosis. *In:* VERSTOVSEK, S. & TEFFERI, A. (eds.) *Myeloproliferative Neoplasms.* Humana Press.

BAROSI, G.,BERGAMASCHI, G.,MARCHETTI, M.,VANNUCCHI, A. M.,GUGLIELMELLI, P.,ANTONIOLI, E.,MASSA, M.,ROSTI, V.,CAMPANELLI, R.,VILLANI, L.,VIARENGO, G.,GATTONI, E.,GERLI, G.,SPECCHIA, G.,TINELLI, C.,RAMBALDI, A. & BARBUI, T. 2007. JAK2 V617F mutational status predicts progression to large splenomegaly and leukemic transformation in primary myelofibrosis. *Blood,* 110**,** 4030-6.

BARRETT, J. C.,FRY, B.,MALLER, J. & DALY, M. J. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics,* 21**,** 263-5.

BARRETT, J. C.,HANSOUL, S.,NICOLAE, D. L.,CHO, J. H.,DUERR, R. H.,RIOUX, J. D.,BRANT, S. R.,SILVERBERG, M. S.,TAYLOR, K. D.,BARMADA, M. M.,BITTON, A.,DASSOPOULOS, T.,DATTA, L. W.,GREEN, T.,GRIFFITHS, A. M.,KISTNER, E. O.,MURTHA, M. T.,REGUEIRO, M. D.,ROTTER, J. I.,SCHUMM, L. P.,STEINHART, A. H.,TARGAN, S. R.,XAVIER, R. J.,LIBIOULLE, C.,SANDOR, C.,LATHROP, M.,BELAICHE, J.,DEWIT, O.,GUT, I.,HEATH, S.,LAUKENS, D.,MNI, M.,RUTGEERTS, P.,VAN GOSSUM, A.,ZELENIKA, D.,FRANCHIMONT, D.,HUGOT, J. P.,DE VOS, M.,VERMEIRE, S.,LOUIS, E.,CARDON, L. R.,ANDERSON, C. A.,DRUMMOND, H.,NIMMO, E.,AHMAD, T.,PRESCOTT, N. J.,ONNIE, C. M.,FISHER, S. A.,MARCHINI, J.,GHORI, J.,BUMPSTEAD, S.,GWILLIAM, R.,TREMELLING, M.,DELOUKAS, P.,MANSFIELD, J.,JEWELL, D.,SATSANGI, J.,MATHEW, C. G.,PARKES, M.,GEORGES, M. & DALY, M. J. 2008. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet,* 40**,** 955-62.

BASQUIERA, A. L.,SORIA, N. W.,RYSER, R.,SALGUERO, M.,MOIRAGHI, B.,SACKMANN, F.,STURICH, A. G.,BORELLO, A.,BERRETTA, A.,BONAFE, M.,BARRAL, J. M.,PALAZZO, E. D. & GARCIA, J. J. 2009. Clinical significance of V617F mutation of the JAK2 gene in patients with chronic myeloproliferative disorders. *Hematology,* 14**,** 323-30.

BAXTER, E. J.,SCOTT, L. M.,CAMPBELL, P. J.,EAST, C.,FOUROUCLAS, N.,SWANTON, S.,VASSILIOU, G. S.,BENCH, A. J.,BOYD, E. M.,CURTIN, N.,SCOTT, M. A.,ERBER, W. N. & GREEN, A. R. 2005. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet,* 365**,** 1054-61.

BENNEMANN, K.,GALM, O.,WILOP, S.,SCHUBERT, C.,BRUMMENDORF, T. H. & JOST, E. 2012. Epigenetic dysregulation of secreted frizzled-related

proteins in myeloproliferative neoplasms complements the JAK2V617F-mutation. *Clinical epigenetics,* 4**,** 12.

BENTLEY, D. R.,BALASUBRAMANIAN, S.,SWERDLOW, H. P.,SMITH, G. P.,MILTON, J.,BROWN, C. G.,HALL, K. P.,EVERS, D. J.,BARNES, C. L.,BIGNELL, H. R.,BOUTELL, J. M.,BRYANT, J.,CARTER, R. J.,KEIRA CHEETHAM, R.,COX, A. J.,ELLIS, D. J.,FLATBUSH, M. R.,GORMLEY, N. A.,HUMPHRAY, S. J.,IRVING, L. J.,KARBELASHVILI, M. S.,KIRK, S. M.,LI, H.,LIU, X.,MAISINGER, K. S.,MURRAY, L. J.,OBRADOVIC, B.,OST, T.,PARKINSON, M. L.,PRATT, M. R.,RASOLONJATOVO, I. M.,REED, M. T.,RIGATTI, R.,RODIGHIERO, C.,ROSS, M. T.,SABOT, A.,SANKAR, S. V.,SCALLY, A.,SCHROTH, G. P.,SMITH, M. E.,SMITH, V. P.,SPIRIDOU, A.,TORRANCE, P. E.,TZONEV, S. S.,VERMAAS, E. H.,WALTER, K.,WU, X.,ZHANG, L.,ALAM, M. D.,ANASTASI, C.,ANIEBO, I. C.,BAILEY, D. M.,BANCARZ, I. R.,BANERJEE, S.,BARBOUR, S. G.,BAYBAYAN, P. A.,BENOIT, V. A.,BENSON, K. F.,BEVIS, C.,BLACK, P. J.,BOODHUN, A.,BRENNAN, J. S.,BRIDGHAM, J. A.,BROWN, R. C.,BROWN, A. A.,BUERMANN, D. H.,BUNDU, A. A.,BURROWS, J. C.,CARTER, N. P.,CASTILLO, N.,CHIARA, E. C. M.,CHANG, S.,NEIL COOLEY, R.,CRAKE, N. R.,DADA, O. O.,DIAKOUMAKOS, K. D.,DOMINGUEZ-FERNANDEZ, B.,EARNSHAW, D. J.,EGBUJOR, U. C.,ELMORE, D. W.,ETCHIN, S. S.,EWAN, M. R.,FEDURCO, M.,FRASER, L. J.,FUENTES FAJARDO, K. V.,SCOTT FUREY, W.,GEORGE, D.,GIETZEN, K. J.,GODDARD, C. P.,GOLDA, G. S.,GRANIERI, P. A.,GREEN, D. E.,GUSTAFSON, D. L.,HANSEN, N. F.,HARNISH, K.,HAUDENSCHILD, C. D.,HEYER, N. I.,HIMS, M. M.,HO, J. T.,HORGAN, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456**,** 53-9.

BERK, P. D.,GOLDBERG, J. D.,SILVERSTEIN, M. N.,WEINFELD, A.,DONOVAN, P. B.,ELLIS, J. T.,LANDAW, S. A.,LASZLO, J.,NAJEAN, Y.,PISCIOTTA, A. V. & WASSERMAN, L. R. 1981. Increased incidence of acute leukemia in polycythemia vera associated with chlorambucil therapy. *N Engl J Med,* 304**,** 441-7.

BETHESDA. 2014. *National Library of Medicine (US). Genetics Home Reference [Internet]* [Online]. Available: http://ghr.nlm.nih.gov/handbook/genomicresearch/snp [Accessed 10 July 2014].

BORTHWICK, L. 2008. Human haematopoietic stem cells identified.

BOTSTEIN, D.,WHITE, R. L.,SKOLNICK, M. & DAVIS, R. W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet,* 32**,** 314-31.

BOUSQUET, M.,QUELEN, C.,DE MAS, V.,DUCHAYNE, E.,ROQUEFEUIL, B.,DELSOL, G.,LAURENT, G.,DASTUGUE, N. & BROUSSET, P. 2005. The t(8;9)(p22;p24) translocation in atypical chronic myeloid leukaemia yields a new PCM1-JAK2 fusion gene. *Oncogene,* 24**,** 7248-52.

BROWNING, B. L. & BROWNING, S. R. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet,* 84**,** 210-23.

BUESO-RAMOS, C. & VARDIMAN, J. 2011. Diagnosis and Classification of the BCR-ABL1-Negative Myeloproliferative Neoplasms. *In:* VERSTOVSEK, S. & TEFFERI, A. (eds.) *Myeloproliferative Neoplasms.* Humana Press.

BUSH, W. S. & MOORE, J. H. 2012. Chapter 11: Genome-wide association studies. *PLoS computational biology,* 8**,** e1002822.

BUTCHER, C. M.,HAHN, U.,TO, L. B.,GECZ, J.,WILKINS, E. J.,SCOTT, H. S.,BARDY, P. G. & D'ANDREA, R. J. 2008. Two novel JAK2 exon 12 mutations in JAK2V617F-negative polycythaemia vera patients. *Leukemia,* 22**,** 870-3.

BUTTON, K. S.,IOANNIDIS, J. P. A.,MOKRYSZ, C.,NOSEK, B. A.,FLINT, J.,ROBINSON, E. S. J. & MUNAFO, M. R. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci,* 14**,** 365-376.

BYE, H.,PRESCOTT, N. J.,LEWIS, C. M.,MATEJCIC, M.,MOODLEY, L.,ROBERTSON, B.,RENSBURG, C. V.,PARKER, M. I. & MATHEW, C. G. 2012. Distinct genetic association at the PLCE1 locus with oesophageal squamous cell carcinoma in the South African population. *Carcinogenesis*.

BYRD, A. L.,PEREZ-ROGERS, J. F.,MANIMARAN, S.,CASTRO-NALLAR, E.,TOMA, I.,MCCAFFREY, T.,SIEGEL, M.,BENSON, G.,CRANDALL, K. A. & JOHNSON, W. E. 2014. Clinical PathoScope: rapid alignment and filtration for accurate pathogen identification in clinical samples using unassembled sequencing data. *BMC Bioinformatics,* 15**,** 262.

CAMPBELL, P. J. 2009. Somatic and germline genetics at the JAK2 locus. *Nat Genet,* 41**,** 385-6.

CAMPBELL, P. J.,BENCH, A. J. & GREEN, A. R. 2007. The Molecular Basis of Leukaemia and Lymphoma. *Postgraduate Haematology*. Blackwell Publishing Ltd.

CAMPBELL, P. J. & GREEN, A. R. 2006. The myeloproliferative disorders. *N Engl J Med,* 355**,** 2452-66.

CAMPBELL, P. J. & GREEN, A. R. 2010. Myeloproliferative Neoplasms. *Postgraduate Haematology*. Wiley-Blackwell.

CAMPBELL, P. J.,GRIESSHAMMER, M.,DOHNER, K.,DOHNER, H.,KUSEC, R.,HASSELBALCH, H. C.,LARSEN, T. S.,PALLISGAARD, N.,GIRAUDIER, S.,LE BOUSSE-KERDILES, M. C.,DESTERKE, C.,GUERTON, B.,DUPRIEZ, B.,BORDESSOULE, D.,FENAUX, P.,KILADJIAN, J. J.,VIALLARD, J. F.,BRIERE, J.,HARRISON, C. N.,GREEN, A. R. & REILLY, J. T. 2006. V617F mutation in JAK2 is associated with poorer survival in idiopathic myelofibrosis. *Blood,* 107**,** 2098-100.

CAMPBELL, P. J.,SCOTT, L. M.,BUCK, G.,WHEATLEY, K.,EAST, C. L.,MARSDEN, J. T.,DUFFY, A.,BOYD, E. M.,BENCH, A. J.,SCOTT, M. A.,VASSILIOU, G. S.,MILLIGAN, D. W.,SMITH, S. R.,ERBER, W. N.,BAREFORD, D.,WILKINS, B. S.,REILLY, J. T.,HARRISON, C. N. & GREEN, A. R. 2005. Definition of subtypes of essential thrombocythaemia and relation to polycythaemia vera based on JAK2 V617F mutation status: a prospective study. *Lancet,* 366**,** 1945-53.

CARLSON, J. M.,HECKERMAN, D. & SHANI, G. 2009. Estimating false discovery rates for contingency tables.

CARTHARIUS, K.,FRECH, K.,GROTE, K.,KLOCKE, B.,HALTMEIER, M.,KLINGENHOFF, A.,FRISCH, M.,BAYERLEIN, M. & WERNER, T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics,* 21**,** 2933-2942.

CERVANTES, F. 2011. Management of Essential Thrombocythemia. *ASH Education Program Book,* 2011**,** 215-221.

CERVANTES, F.,ALVAREZ-LARRÁN, A.,TALARN, C.,GÓMEZ, M. & MONTSERRAT, E. 2002. Myelofibrosis with myeloid metaplasia following essential thrombocythaemia: actuarial probability, presenting characteristics and evolution in a series of 195 patients. *British Journal of Haematology,* 118**,** 786-790.

CERVANTES, F.,DUPRIEZ, B.,PEREIRA, A.,PASSAMONTI, F.,REILLY, J. T.,MORRA, E.,VANNUCCHI, A. M.,MESA, R. A.,DEMORY, J. L.,BAROSI, G.,RUMI, E. & TEFFERI, A. 2009. New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. *Blood,* 113**,** 2895-901.

CERVANTES, F. & HERNÁNDEZ-BOLUDA, J.-C. 2012. Patient's Information and Examinations Needed Before Planning Therapy in the Myeloproliferative Neoplasms. *In:* BARBUI, T. & TEFFERI, A. (eds.) *Myeloproliferative Neoplasms.* Springer Berlin Heidelberg.

CERVANTES, F.,PASSAMONTI, F. & BAROSI, G. 2008. Life expectancy and prognostic factors in the classic BCR/ABL-negative myeloproliferative disorders. *Leukemia,* 22**,** 905-14.

CHAITER, Y.,BRENNER, B.,AGHAI, E. & TATARSKY, I. 1992. High incidence of myeloproliferative disorders in Ashkenazi Jews in northern Israel. *Leukemia & lymphoma,* 7**,** 251-5.

CHEN, S.,FEI, H.,ZHANG, R.,XUE, Y.,PAN, J.,WU, Y. & CENG, J. 2007. Analysis of JAK2(V617F) mutation in Chinese patients with myeloproliferative disorders. *Am J Hematol,* 82**,** 458-9.

CHOREV, M. & CARMEL, L. 2012. The function of introns. *Front Genet,* 3**,** 55.

CHRISTENSEN, J.,BENTZ, S.,SENGSTAG, T.,SHASTRI, V. P. & ANDERLE, P. 2013. FOXQ1, a novel target of the Wnt pathway and a new marker for activation of Wnt signaling in solid tumors. *PloS one,* 8**,** e60051.

CIRMENA, G.,ALIANO, S.,FUGAZZA, G.,BRUZZONE, R.,GARUTI, A.,BOCCIARDI, R.,BACIGALUPO, A.,RAVAZZOLO, R.,BALLESTRERO, A. & SESSAREGO, M. 2008. A BCR-JAK2 fusion gene as the result of a t(9;22)(p24;q11) in a patient with acute myeloid leukemia. *Cancer Genetics and Cytogenetics,* 183**,** 105-108.

CIVINI, S.,JIN, P.,REN, J.,SABATINO, M.,CASTIELLO, L.,JIN, J.,WANG, H.,ZHAO, Y.,MARINCOLA, F. & STRONCEK, D. 2013. Leukemia cells induce changes in human bone marrow stromal cells. *Journal of translational medicine,* 11**,** 298.

CLARK, B. E. & THEIN, S. L. 2004. Molecular diagnosis of haemoglobin disorders. *Clinical and laboratory haematology,* 26**,** 159-76.

CLARK, M. P.,GEORGE, K. M.,BOOKLAND, R. G.,CHEN, J.,LAUGHLIN, S. K.,THAKUR, K. D.,LEE, W.,DAVIS, J. R.,CABRERA, E. J.,BRUGEL, T. A.,VANRENS, J. C.,LAUFERSWEILER, M. J.,MAIER, J. A.,SABAT, M. P.,GOLEBIOWSKI, A.,EASWARAN, V.,WEBSTER, M. E.,DE, B. & ZHANG, G. 2007. Development of new pyrrolopyrimidine-based inhibitors of Janus kinase 3 (JAK3). *Bioorg Med Chem Lett,* 17**,** 1250-3.

CLARKSON, B.,STRIFE, A.,WISNIEWSKI, D.,LAMBEK, C. L. & LIU, C. 2003. Chronic myelogenous leukemia as a paradigm of early cancer and possible curative strategies. *Leukemia,* 17**,** 1211-1262.

COLAIZZO, D.,AMITRANO, L.,TISCIA, G. L.,GRANDONE, E.,GUARDASCIONE, M. A. & MARGAGLIONE, M. 2007. A new JAK2 gene mutation in patients with polycythemia vera and splanchnic vein thrombosis. *Blood,* 110**,** 2768-9.

COLAIZZO, D.,TISCIA, G. L.,BAFUNNO, V.,AMITRANO, L.,VERGURA, P.,LUPONE, M. R.,GRANDONE, E.,GUARDASCIONE, M. A. & MARGAGLIONE, M. 2011. Sex modulation of the occurrence of jak2 v617f mutation in patients with splanchnic venous thrombosis. *Thromb Res,* 128**,** 233-6.

COLLA, S.,MORANDI, F.,LAZZARETTI, M.,RIZZATO, R.,LUNGHI, P.,BONOMINI, S.,MANCINI, C.,PEDRAZZONI, M.,CRUGNOLA, M.,RIZZOLI, V. & GIULIANI, N. 2005. Human myeloma cells express the bone regulating gene Runx2/Cbfa1 and produce osteopontin that is involved in angiogenesis in multiple myeloma patients. *Leukemia,* 19**,** 2166-76.

CONSTANTINESCU, S. N.,GIRARDOT, M. & PECQUET, C. 2008. Mining for JAK-STAT mutations in cancer. *Trends Biochem Sci,* 33**,** 122-31.

COOPER, D. N. 2010. Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. *Hum Genomics,* 4**,** 284-8.

DA SILVA, R. R.,DOMINGUES HATZLHOFER, B. L.,MACHADO, C. G.,LIMA, A. S.,DE ALBUQUERQUE, D. M.,DOS SANTOS, M. N.,FERTRIN, K. Y.,COSTA, F. F.,ARAUJO ADA, S. & BEZERRA, M. A. 2012. JAK2 V617F mutation prevalence in myeloproliferative neoplasms in Pernambuco, Brazil. *Genetic testing and molecular biomarkers,* 16**,** 802-5.

DAVID-FUNG, E. S.,BUTLER, R.,BUZI, G.,YUI, M. A.,DIAMOND, R. A.,ANDERSON, M. K.,ROWEN, L. & ROTHENBERG, E. V. 2009. Transcription factor expression dynamics of early T-lymphocyte specification and commitment. *Developmental biology,* 325**,** 444-67.

DAYEM ULLAH, A. Z.,LEMOINE, N. R. & CHELALA, C. 2012. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic acids research,* 40**,** W65-70.

DE BAKKER, P. I.,YELENSKY, R.,PE'ER, I.,GABRIEL, S. B.,DALY, M. J. & ALTSHULER, D. 2005. Efficiency and power in genetic association studies. *Nature genetics,* 37**,** 1217-23.

DELHOMMEAU, F.,DUPONT, S.,DELLA VALLE, V.,JAMES, C.,TRANNOY, S.,MASSE, A.,KOSMIDER, O.,LE COUEDIC, J. P.,ROBERT, F.,ALBERDI, A.,LECLUSE, Y.,PLO, I.,DREYFUS, F. J.,MARZAC, C.,CASADEVALL, N.,LACOMBE, C.,ROMANA, S. P.,DESSEN, P.,SOULIER, J.,VIGUIE, F.,FONTENAY, M.,VAINCHENKER, W. & BERNARD, O. A. 2009. Mutation in TET2 in myeloid cancers. *N Engl J Med,* 360**,** 2289-301.

DIAS, N. & STEIN, C. A. 2002. Antisense Oligonucleotides: Basic Concepts and Mechanisms. *Molecular Cancer Therapeutics,* 1**,** 347-355.

DRUKER, B. J. 2008. Translation of the Philadelphia chromosome into therapy for CML. *Blood,* 112**,** 4808-4817.

DUDBRIDGE, F. & GUSNANTO, A. 2008. Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology,* 32**,** 227-234.

DUNBAR, A. J.,GONDEK, L. P.,O'KEEFE, C. L.,MAKISHIMA, H.,RATAUL, M. S.,SZPURKA, H.,SEKERES, M. A.,WANG, X. F.,MCDEVITT, M. A. & MACIEJEWSKI, J. P. 2008. 250K single nucleotide polymorphism array karyotyping identifies acquired uniparental disomy and homozygous mutations, including novel missense substitutions of c-Cbl, in myeloid malignancies. *Cancer Res,* 68**,** 10349-57.

ERALI, M.,PALAIS, R. & WITTWER, C. 2008. SNP genotyping by unlabeled probe melting analysis. *Methods in molecular biology,* 429**,** 199-206.

ERBER, W. 2010. *Diagnostic Techniques in Hematological Malignancies,* New York, Cambridge University Press.

ERNST, T.,CHASE, A.,ZOI, K.,WAGHORN, K.,HIDALGO-CURTIS, C.,SCORE, J.,JONES, A.,GRAND, F.,REITER, A.,HOCHHAUS, A. & CROSS, N. C. 2010a. Transcription factor mutations in myelodysplastic/myeloproliferative neoplasms. *Haematologica,* 95**,** 1473-80.

ERNST, T.,CHASE, A.,ZOI, K.,WAGHORN, K.,HIDALGO-CURTIS, C.,SCORE, J.,JONES, A.,GRAND, F.,REITER, A.,HOCHHAUS, A. & CROSS, N. C. P. Transcription factor mutations in myelodysplastic/myeloproliferative neoplasms. *Haematologica,* 95**,** 1473-1480.

ERNST, T.,CHASE, A. J.,SCORE, J.,HIDALGO-CURTIS, C. E.,BRYANT, C.,JONES, A. V.,WAGHORN, K.,ZOI, K.,ROSS, F. M.,REITER, A.,HOCHHAUS, A.,DREXLER, H. G.,DUNCOMBE, A.,CERVANTES, F.,OSCIER, D.,BOULTWOOD, J.,GRAND, F. H. & CROSS, N. C. 2010b. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet,* 42**,** 722-6.

EWING, B. & GREEN, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res,* 8**,** 186-94.

FIGUEROA, M. E.,ABDEL-WAHAB, O.,LU, C.,WARD, P. S.,PATEL, J.,SHIH, A.,LI, Y.,BHAGWAT, N.,VASANTHAKUMAR, A.,FERNANDEZ, H. F.,TALLMAN, M. S.,SUN, Z.,WOLNIAK, K.,PEETERS, J. K.,LIU, W.,CHOE, S. E.,FANTIN, V. R.,PAIETTA, E.,LOWENBERG, B.,LICHT, J. D.,GODLEY, L. A.,DELWEL, R.,VALK, P. J.,THOMPSON, C. B.,LEVINE, R. L. & MELNICK, A. 2010. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell,* 18**,** 553-67.

FLUTRE, T.,WEN, X.,PRITCHARD, J. & STEPHENS, M. 2013. A statistical framework for joint eQTL analysis in multiple tissues. *PLoS genetics,* 9**,** e1003486.

FRANCA, L. T.,CARRILHO, E. & KIST, T. B. 2002. A review of DNA sequencing techniques. *Quarterly reviews of biophysics,* 35**,** 169-200.

GABRIEL, S. B.,SCHAFFNER, S. F.,NGUYEN, H.,MOORE, J. M.,ROY, J.,BLUMENSTIEL, B.,HIGGINS, J.,DEFELICE, M.,LOCHNER, A.,FAGGART, M.,LIU-CORDERO, S. N.,ROTIMI, C.,ADEYEMO, A.,COOPER, R.,WARD, R.,LANDER, E. S.,DALY, M. J. & ALTSHULER, D. 2002. The structure of haplotype blocks in the human genome. *Science,* 296**,** 2225-9.

GAO, R. 2011. *Association Analysis between JAK2V617F and Flanking SNPs in the Western Population and Development of a New Method for Detection of JAK2V617F* PhD, Jilin University China.

GELSI-BOYER, V.,TROUPLIN, V.,ADELAIDE, J.,BONANSEA, J.,CERVERA, N.,CARBUCCIA, N.,LAGARDE, A.,PREBET, T.,NEZRI, M.,SAINTY, D.,OLSCHWANG, S.,XERRI, L.,CHAFFANET, M.,MOZZICONACCI, M. J.,VEY, N. & BIRNBAUM, D. 2009. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol,* 145**,** 788-800.

GENOMES PROJECT, C.,ABECASIS, G. R.,AUTON, A.,BROOKS, L. D.,DEPRISTO, M. A.,DURBIN, R. M.,HANDSAKER, R. E.,KANG, H. M.,MARTH, G. T. & MCVEAN, G. A. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature,* 491**,** 56-65.

GILES, G. G.,LICKISS, J. N.,BAIKIE, M. J.,LOWENTHAL, R. M. & PANTON, J. 1984. Myeloproliferative and lymphoproliferative disorders in Tasmania, 1972-80: occupational and familial aspects. *Journal of the National Cancer Institute,* 72**,** 1233-40.

GOGOL-DORING, A. & CHEN, W. 2012. An overview of the analysis of next generation sequencing data. *Methods in molecular biology,* 802**,** 249-57.

GOLDIN, L. R.,BJORKHOLM, M.,KRISTINSSON, S. Y.,SAMUELSSON, J. & LANDGREN, O. 2009. Germline and somatic JAK2 mutations and susceptibility to chronic myeloproliferative neoplasms. *Genome Med,* 1**,** 55.

GOLDMAN, J. 1997. ABC of clinical haematology. Chronic myeloid leukaemia. *BMJ,* 314**,** 657-60.

GRADA, A. & WEINBRECHT, K. 2013. Next-Generation Sequencing: Methodology and Application. *The Journal of investigative dermatology,* 133**,** e11.

GRAND, F. H.,HIDALGO-CURTIS, C. E.,ERNST, T.,ZOI, K.,ZOI, C.,MCGUIRE, C.,KREIL, S.,JONES, A.,SCORE, J.,METZGEROTH, G.,OSCIER, D.,HALL, A.,BRANDTS, C.,SERVE, H.,REITER, A.,CHASE, A. J. & CROSS, N. C. 2009. Frequent CBL mutations associated with 11q acquired uniparental disomy in myeloproliferative neoplasms. *Blood,* 113**,** 6182-92.

GREENE, C. S.,PENROD, N. M.,WILLIAMS, S. M. & MOORE, J. H. 2009. Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS One,* 4**,** e5639.

GRIESINGER, F.,HENNIG, H.,HILLMER, F.,PODLESCHNY, M.,STEFFENS, R.,PIES, A.,WÖRMANN, B.,HAASE, D. & BOHLANDER, S. K. 2005. A BCR–JAK2 fusion gene as the result of a t(9;22)(p24;q11.2) translocation in a patient with a clinically typical chronic myeloid leukemia. *Genes, Chromosomes and Cancer,* 44**,** 329-333.

GRUNEBACH, F.,BROSS-BACH, U.,KANZ, L. & BROSSART, P. 2006. Detection of a new JAK2 D620E mutation in addition to V617F in a patient with polycythemia vera. *Leukemia,* 20**,** 2210-1.

GUGLIELMELLI, P.,BAROSI, G.,SPECCHIA, G.,RAMBALDI, A.,LO COCO, F.,ANTONIOLI, E.,PIERI, L.,PANCRAZZI, A.,PONZIANI, V.,DELAINI, F.,LONGO, G.,AMMATUNA, E.,LISO, V.,BOSI, A.,BARBUI, T. & VANNUCCHI, A. M. 2009. Identification of patients with poorer survival in primary myelofibrosis based on the burden of JAK2V617F mutated allele. *Blood,* 114**,** 1477-83.

GUGLIELMELLI, P.,BIAMONTE, F.,SCORE, J.,HIDALGO-CURTIS, C.,CERVANTES, F.,MAFFIOLI, M.,FANELLI, T.,ERNST, T.,WINKELMAN, N.,JONES, A. V.,ZOI, K.,REITER, A.,DUNCOMBE, A.,VILLANI, L.,BOSI, A.,BAROSI, G.,CROSS, N. C. & VANNUCCHI, A. M. 2011. EZH2 mutational status predicts poor survival in myelofibrosis. *Blood,* 118**,** 5227-34.

GUGLIELMELLI, P.,BIAMONTE, F.,SPOLVERINI, A.,PIERI, L.,ISGRO, A.,ANTONIOLI, E.,PANCRAZZI, A.,BOSI, A.,BAROSI, G. & VANNUCCHI, A. M. 2010. Frequency and clinical correlates of JAK2 46/1 (GGCC) haplotype in primary myelofibrosis. *Leukemia,* 24**,** 1533-7.

GUO, Y.,LI, J.,BONHAM, A. J.,WANG, Y. & DENG, H. 2009. Gains in power for exhaustive analyses of haplotypes using variable-sized sliding window strategy: a comparison of association-mapping strategies. *European journal of human genetics : EJHG,* 17**,** 785-92.

HALL, J. C. 2011. How to dissect surgical journals: IX – Sample size. *ANZ Journal of Surgery,* 81**,** 294-297.

HARDY, G. H. 1908. Mendelian Proportions in a Mixed Population. *Science,* 28**,** 49-50.

HE, N.,XIAO, Z.,YIN, T.,STUBBS, J.,LI, L. & QUARLES, L. D. 2011. Inducible expression of Runx2 results in multiorgan abnormalities in mice. *Journal of cellular biochemistry,* 112**,** 653-65.

HEAD, S. R.,KOMORI, H. K.,LAMERE, S. A.,WHISENANT, T.,VAN NIEUWERBURGH, F.,SALOMON, D. R. & ORDOUKHANIAN, P. 2014. Library construction for next-generation sequencing: overviews and challenges. *BioTechniques,* 56**,** 61-4, 66, 68, passim.

HELLSTROM-LINDBERG, E. & CAZZOLA, M. 2008. The role of JAK2 mutations in RARS and other MDS. *Hematology Am Soc Hematol Educ Program***,** 52-9.

HEMMINKI, K.,SUNDQUIST, J. & BERMEJO, J. L. 2009. Associated cancers in parents and offspring of polycythaemia vera and myelofibrosis patients. *Br J Haematol,* 147**,** 526-30.

HERT, D. G.,FREDLAKE, C. P. & BARRON, A. E. 2008. DNA sequencing by microchip electrophoresis using mixtures of high- and low-molar mass poly(N,N-dimethylacrylamide) matrices. *Electrophoresis,* 29**,** 4663-4668.

HIGGINS, J. & GREEN, S. E. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011] ed.: The Cochrane Collaboration, 2011.

HOFFBRAND, V. & MOSS, P. 2011a. The aetiology and genetics of haematological malignancies. *Essential Haematology.* 6th ed.: Wiley-Blackwell.

HOFFBRAND, V. & MOSS, P. 2011b. The non-leukaemic myeloproliferative neoplasms. *Essential Haematology.* 6th ed.: Wiley-Blackwell.

HOSKING, L.,LUMSDEN, S.,LEWIS, K.,YEO, A.,MCCARTHY, L.,BANSAL, A.,RILEY, J.,PURVIS, I. & XU, C.-F. 2004. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *Eur J Hum Genet,* 12**,** 395-399.

HSIAO, H. H.,LIU, Y. C.,TSAI, H. J.,LEE, C. P.,HSU, J. F. & LIN, S. F. 2011. JAK2V617F mutation is associated with special alleles in essential thrombocythemia. *Leuk Lymphoma,* 52**,** 478-82.

HU, T. T.,ZHANG, X. J. & GUAN, M. 2011. [The association between jak2 46/1 haplotype and the susceptibility of PV and ET in chinese Han population]. *zhonghua yixue jianyan zazhi,* 34**,** 717-721.

HU, X.,CUI, D.,MOSCINSKI, L. C.,ZHANG, X.,MACCACHERO, V. & ZUCKERMAN, K. S. 2007. TGFbeta regulates the expression and activities of G2 checkpoint kinases in human myeloid leukemia cells. *Cytokine,* 37**,** 155-62.

HUSSEIN, K.,DRALLE, W.,THEOPHILE, K.,KREIPE, H. & BOCK, O. 2009. Megakaryocytic expression of miRNA 10a, 17-5p, 20a and 126 in Philadelphia chromosome-negative myeloproliferative neoplasm. *Ann Hematol,* 88**,** 325-32.

IHLE, J. N. & GILLILAND, D. G. 2007a. Jak2: normal function and role in hematopoietic disorders. *Current Opinion in Genetics & Development,* 17**,** 8-14.

IHLE, J. N. & GILLILAND, D. G. 2007b. Jak2: normal function and role in hematopoietic disorders. *Curr Opin Genet Dev,* 17**,** 8-14.

JAGER, R.,GISSLINGER, H.,PASSAMONTI, F.,RUMI, E.,BERG, T.,GISSLINGER, B.,PIETRA, D.,HARUTYUNYAN, A.,KLAMPFL, T.,OLCAYDU, D.,CAZZOLA, M. & KRALOVICS, R. 2010. Deletions of the transcription factor Ikaros in myeloproliferative neoplasms. *Leukemia,* 24**,** 1290-8.

JAMES, C.,UGO, V.,CASADEVALL, N.,CONSTANTINESCU, S. N. & VAINCHENKER, W. 2005a. A JAK2 mutation in myeloproliferative

disorders: pathogenesis and therapeutic and scientific prospects. *Trends Mol Med,* 11**,** 546-54.

JAMES, C.,UGO, V.,LE COUEDIC, J. P.,STAERK, J.,DELHOMMEAU, F.,LACOUT, C.,GARCON, L.,RASLOVA, H.,BERGER, R.,BENNACEUR-GRISCELLI, A.,VILLEVAL, J. L.,CONSTANTINESCU, S. N.,CASADEVALL, N. & VAINCHENKER, W. 2005b. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature,* 434**,** 1144-8.

JELINEK, J.,OKI, Y.,GHARIBYAN, V.,BUESO-RAMOS, C.,PRCHAL, J. T.,VERSTOVSEK, S.,BERAN, M.,ESTEY, E.,KANTARJIAN, H. M. & ISSA, J. P. 2005. JAK2 mutation 1849G>T is rare in acute leukemias but can be found in CMML, Philadelphia chromosome-negative CML, and megakaryocytic leukemia. *Blood,* 106**,** 3370-3.

JENSEN, M. K.,DE NULLY BROWN, P.,NIELSEN, O. J. & HASSELBALCH, H. C. 2000. Incidence, clinical features and outcome of essential thrombocythaemia in a well defined geographical area. *European journal of haematology,* 65**,** 132-9.

JOHANSSON, P. 2006. Epidemiology of the myeloproliferative disorders polycythemia vera and essential thrombocythemia. *Semin Thromb Hemost,* 32**,** 171-3.

JOHNSEN, H. E.,KJELDSEN, M. K.,URUP, T.,FOGD, K.,PILGAARD, L.,BOEGSTED, M.,NYEGAARD, M.,CHRISTIANSEN, I.,BUKH, A. & DYBKAER, K. 2009. Cancer stem cells and the cellular hierarchy in haematological malignancies. *Eur J Cancer,* 45 Suppl 1**,** 194-201.

JOHNSON, E. S.,ZHOU, Y.,LILLIAN YAU, C.,PRABHAKAR, D.,NDETAN, H.,SINGH, K. & PREACELY, N. 2010. Mortality from malignant diseases-update of the Baltimore union poultry cohort. *Cancer causes & control : CCC,* 21**,** 215-21.

JONES, A. 2010. *The molecular pathogenesis of myeloproliferative neoplasms.* Doctor of Philosophy, University of Southampton.

JONES, A. V.,CAMPBELL, P. J.,BEER, P. A.,SCHNITTGER, S.,VANNUCCHI, A. M.,ZOI, K.,PERCY, M. J.,MCMULLIN, M. F.,SCOTT, L. M.,TAPPER, W.,SILVER, R. T.,OSCIER, D.,HARRISON, C. N.,GRALLERT, H.,KISIALIOU, A.,STRIKE, P.,CHASE, A. J.,GREEN, A. R. & CROSS, N. C. 2010. The JAK2 46/1 haplotype predisposes to MPL-mutated myeloproliferative neoplasms. *Blood,* 115**,** 4517-23.

JONES, A. V.,CHASE, A.,SILVER, R. T.,OSCIER, D.,ZOI, K.,WANG, Y. L.,CARIO, H.,PAHL, H. L.,COLLINS, A.,REITER, A.,GRAND, F. & CROSS, N. C. 2009. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet,* 41**,** 446-9.

JONES, A. V.,KREIL, S.,ZOI, K.,WAGHORN, K.,CURTIS, C.,ZHANG, L.,SCORE, J.,SEEAR, R.,CHASE, A. J.,GRAND, F. H.,WHITE, H.,ZOI, C.,LOUKOPOULOS, D.,TERPOS, E.,VERVESSOU, E. C.,SCHULTHEIS, B.,EMIG, M.,ERNST, T.,LENGFELDER, E.,HEHLMANN, R.,HOCHHAUS, A.,OSCIER, D.,SILVER, R. T.,REITER, A. & CROSS, N. C. 2005. Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders. *Blood,* 106**,** 2162-8.

KAMADA, N. 2001. [Radiation-related CML]. *Nippon Rinsho,* 59**,** 2329-35.

KAPLAN, S. D. 1986. Update of a mortality study of workers in petroleum refineries. *Journal of occupational medicine. : official publication of the Industrial Medical Association,* 28**,** 514-6.

KARGER, B. L. & GUTTMAN, A. 2009. DNA sequencing by CE. *Electrophoresis,* 30 Suppl 1**,** S196-202.

KARKUCAK, M.,YAKUT, T.,OZKOCAMAN, V.,OZKALEMKAS, F.,ALI, R.,BAYRAM, M.,GORUKMEZ, O. & OCAKOGLU, G. 2012. Evaluation of the JAK2-V617F gene mutation in Turkish patients with essential thrombocythemia and polycythemia vera. *Molecular biology reports,* 39**,** 8663-7.

KAROW, A.,WALLER, C.,REIMANN, C.,NIEMEYER, C. M. & KRATZ, C. P. 2008. JAK2 mutations other than V617F: a novel mutation and mini review. *Leuk Res,* 32**,** 365-6.

KATJA FIEDLER & BRUNNER, C. 2012. Mechanisms Controlling Hematopoiesis

In: LAWRIE, D. C. (ed.) *Hematology - Science and Practice.* InTech.

KE, X.,TAYLOR, M. S. & CARDON, L. R. 2008. Singleton SNPs in the human genome and implications for genome-wide association studies. *Eur J Hum Genet,* 16**,** 506-515.

KEARNEY, L. & HORSLEY, S. W. 2005. Molecular cytogenetics in haematological malignancy: current technology and future prospects. *Chromosoma,* 114**,** 286-94.

KENT, W. J.,SUGNET, C. W.,FUREY, T. S.,ROSKIN, K. M.,PRINGLE, T. H.,ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome research,* 12**,** 996-1006.

KILPIVAARA, O. & LEVINE, R. L. 2008. JAK2 and MPL mutations in myeloproliferative neoplasms: discovery and science. *Leukemia,* 22**,** 1813-7.

KILPIVAARA, O.,MUKHERJEE, S.,SCHRAM, A. M.,WADLEIGH, M.,MULLALLY, A.,EBERT, B. L.,BASS, A.,MARUBAYASHI, S.,HEGUY, A.,GARCIA-MANERO, G.,KANTARJIAN, H.,OFFIT, K.,STONE, R. M.,GILLILAND, D. G.,KLEIN, R. J. & LEVINE, R. L. 2009. A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms. *Nat Genet,* 41**,** 455-9.

KIM, J. T.,CHO, Y. G.,CHOI, S. I.,LEE, Y. J.,KIM, H. R.,JANG, S. J.,MOON, D. S.,PARK, Y. J. & PARK, G. 2010. [JAK2 V617F and exon 12 genetic variations in Korean patients with BCR/ABL1-negative myeloproliferative neoplasms]. *Korean J Lab Med,* 30**,** 567-74.

KIM, S. & MISRA, A. 2007. SNP genotyping: technologies and biomedical applications. *Annu Rev Biomed Eng,* 9**,** 289-320.

KIRCHER, M. & KELSO, J. 2010. High-throughput DNA sequencing--concepts and limitations. *BioEssays : news and reviews in molecular, cellular and developmental biology,* 32**,** 524-36.

KITTUR, J.,KNUDSON, R. A.,LASHO, T. L.,FINKE, C. M.,GANGAT, N.,WOLANSKYJ, A. P.,LI, C. Y.,WU, W.,KETTERLING, R. P.,PARDANANI, A. & TEFFERI, A. 2007. Clinical correlates of JAK2V617F allele burden in essential thrombocythemia. *Cancer,* 109**,** 2279-84.

KLAMPFL, T.,GISSLINGER, H.,HARUTYUNYAN, A. S.,NIVARTHI, H.,RUMI, E.,MILOSEVIC, J. D.,THEM, N. C.,BERG, T.,GISSLINGER, B.,PIETRA, D.,CHEN, D.,VLADIMER, G. I.,BAGIENSKI, K.,MILANESI, C.,CASETTI, I. C.,SANT'ANTONIO, E.,FERRETTI, V.,ELENA, C.,SCHISCHLIK, F.,CLEARY, C.,SIX, M.,SCHALLING, M.,SCHONEGGER, A.,BOCK, C.,MALCOVATI, L.,PASCUTTO, C.,SUPERTI-FURGA, G.,CAZZOLA, M. & KRALOVICS, R. 2013.

Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med,* 369**,** 2379-90.

KNIERIM, E.,LUCKE, B.,SCHWARZ, J. M.,SCHUELKE, M. & SEELOW, D. 2011. Systematic Comparison of Three Methods for Fragmentation of Long-Range PCR Products for Next Generation Sequencing. *PLoS One,* 6**,** e28240.

KOMAN, A.,CAZAUBON, S.,COURAUD, P. O.,ULLRICH, A. & STROSBERG, A. D. 1996. Molecular characterization and in vitro biological activity of placentin, a new member of the insulin gene family. *J Biol Chem,* 271**,** 20238-41.

KONIG, I. R. 2011. Validation in genetic association studies. *Briefings in Bioinformatics,* 12**,** 253-8.

KOUROUPI, E.,KILADJIAN, J. J.,CHOMIENNE, C.,DOSQUET, C.,BELLUCCI, S.,VALLA, D. & CASSINAT, B. 2011. The JAK2 46/1 haplotype in splanchnic vein thrombosis. *Blood,* 117**,** 5777-8.

KOUROUPI, E.,ZOI, K.,PARQUET, N.,ZOI, C.,KILADJIAN, J. J.,GRIGORAKI, V.,VAINCHENKER, W.,LELLOUCHE, F.,MARZAC, C.,SCHLAGETER, M. H.,DOSQUET, C.,SCOTT, L. M.,FENAUX, P.,LOUKOPOULOS, D.,CHOMIENNE, C. & CASSINAT, B. 2008. Mutations in exon 12 of JAK2 are mainly found in JAK2 V617F-negative polycythaemia vera patients. *Br J Haematol,* 142**,** 676-9.

KRAKOW, E. F. 2012. Chapter 179. Hematologic Malignancies. *In:* FRASER, G. (ed.) *Principles and Practice of Hospital Medicine.*

KRALOVICS, R. 2012. Update on the Biology of Myeloproliferative Neoplasms. *In:* BARBUI, T. & TEFFERI, A. (eds.) *Myeloproliferative Neoplasms.* Springer Berlin Heidelberg.

KRALOVICS, R.,PASSAMONTI, F.,BUSER, A. S.,TEO, S. S.,TIEDT, R.,PASSWEG, J. R.,TICHELLI, A.,CAZZOLA, M. & SKODA, R. C. 2005a. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *The New England journal of medicine,* 352**,** 1779-90.

KRALOVICS, R.,PASSAMONTI, F.,BUSER, A. S.,TEO, S. S.,TIEDT, R.,PASSWEG, J. R.,TICHELLI, A.,CAZZOLA, M. & SKODA, R. C. 2005b. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med,* 352**,** 1779-90.

KRATZ, C. P.,BOLL, S.,KONTNY, U.,SCHRAPPE, M.,NIEMEYER, C. M. & STANULLA, M. 2006. Mutational screen reveals a novel JAK2 mutation, L611S, in a child with acute lymphoblastic leukemia. *Leukemia,* 20**,** 381-3.

KRISTINSSON, S. Y.,LANDGREN, O.,SAMUELSSON, J.,BJORKHOLM, M. & GOLDIN, L. R. 2010. Autoimmunity and the risk of myeloproliferative neoplasms. *Haematologica,* 95**,** 1216-20.

KU, C. S.,COOPER, D. N.,POLYCHRONAKOS, C.,NAIDOO, N.,WU, M. & SOONG, R. 2012. Exome sequencing: dual role as a discovery and diagnostic tool. *Ann Neurol,* 71**,** 5-14.

KU, C. S.,LOY, E. Y.,PAWITAN, Y. & CHIA, K. S. 2010. The pursuit of genome-wide association studies: where are we now? *Journal of human genetics,* 55**,** 195-206.

KUMAR, R. & EVANS, T. 2001. Haematopoiesis. *eLS.* John Wiley & Sons, Ltd.

KUO, Y. H.,ZAIDI, S. K.,GORNOSTAEVA, S.,KOMORI, T.,STEIN, G. S. & CASTILLA, L. H. 2009. Runx2 induces acute myeloid leukemia in cooperation with Cbfbeta-SMMHC in mice. *Blood,* 113**,** 3323-32.

KUTTI, J. & RIDELL, B. 2001. Epidemiology of the myeloproliferative disorders: essential thrombocythaemia, polycythaemia vera and idiopathic myelofibrosis. *Pathol Biol (Paris),* 49**,** 164-6.

LACRONIQUE, V.,BOUREUX, A.,DELLA VALLE, V.,POIREL, H.,QUANG, C. T.,MAUCHAUFFÉ, M.,BERTHOU, C.,LESSARD, M.,BERGER, R.,GHYSDAEL, J. & BERNARD, O. A. 1997. A TEL-JAK2 Fusion Protein with Constitutive Kinase Activity in Human Leukemia. *Science,* 278**,** 1309-1312.

LAMBERT, J. R.,EVERINGTON, T.,LINCH, D. C. & GALE, R. E. 2009. In essential thrombocythemia, multiple JAK2-V617F clones are present in most mutant-positive patients: a new disease paradigm. *Blood,* 114**,** 3018-23.

LANDAU, D. A.,CARTER, S. L.,GETZ, G. & WU, C. J. 2014. Clonal evolution in hematological malignancies and therapeutic implications. *Leukemia,* 28**,** 34-43.

LANDGREN, O.,GOLDIN, L. R.,KRISTINSSON, S. Y.,HELGADOTTIR, E. A.,SAMUELSSON, J. & BJORKHOLM, M. 2008. Increased risks of polycythemia vera, essential thrombocythemia, and myelofibrosis among 24,577 first-degree relatives of 11,039 patients with myeloproliferative neoplasms in Sweden. *Blood,* 112**,** 2199-204.

LEE, J. W.,KIM, Y. G.,SOUNG, Y. H.,HAN, K. J.,KIM, S. Y.,RHIM, H. S.,MIN, W. S.,NAM, S. W.,PARK, W. S.,LEE, J. Y.,YOO, N. J. & LEE, S. H. 2006. The JAK2 V617F mutation in de novo acute myelogenous leukemias. *Oncogene,* 25**,** 1434-6.

LEE, P. H. & SHATKAY, H. 2008. F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic acids research,* 36**,** D820-4.

LEVINE, R. L.,WADLEIGH, M.,COOLS, J.,EBERT, B. L.,WERNIG, G.,HUNTLY, B. J.,BOGGON, T. J.,WLODARSKA, I.,CLARK, J. J.,MOORE, S.,ADELSPERGER, J.,KOO, S.,LEE, J. C.,GABRIEL, S.,MERCHER, T.,D'ANDREA, A.,FROHLING, S.,DOHNER, K.,MARYNEN, P.,VANDENBERGHE, P.,MESA, R. A.,TEFFERI, A.,GRIFFIN, J. D.,ECK, M. J.,SELLERS, W. R.,MEYERSON, M.,GOLUB, T. R.,LEE, S. J. & GILLILAND, D. G. 2005. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer cell,* 7**,** 387-97.

LEVINE, R. L. & WERNIG, G. 2006. Role of JAK-STAT signaling in the pathogenesis of myeloproliferative disorders. *Hematology Am Soc Hematol Educ Program***,** 233-9, 510.

LI, Y.,SUNG, W. K. & LIU, J. J. 2007. Association mapping via regularized regression analysis of single-nucleotide-polymorphism haplotypes in variable-sized sliding windows. *American journal of human genetics,* 80**,** 705-15.

LICHTMAN, M. A. 2008. Battling the hematological malignancies: the 200 years' war. *The oncologist,* 13**,** 126-38.

LIEU, C. H.,WU, H. S.,HON, Y. C.,TSAI, W. H.,YANG, C. F.,WANG, C. C.,LIN, Y. C.,SHIH, C. H. & HSU, H. C. 2008. Prevalence of the JAK2-V617F mutation in Taiwanese patients with chronic myeloproliferative disorders. *Intern Med J,* 38**,** 422-6.

LIPPERT, E.,BOISSINOT, M.,KRALOVICS, R.,GIRODON, F.,DOBO, I.,PRALORAN, V.,BOIRET-DUPRE, N.,SKODA, R. C. & HERMOUET, S. 2006. The JAK2-V617F mutation is frequently present at diagnosis in patients with essential thrombocythemia and polycythemia vera. *Blood,* 108**,** 1865-7.

LIU, L.,LI, Y.,LI, S.,HU, N.,HE, Y.,PONG, R.,LIN, D.,LU, L. & LAW, M. 2012. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology,* 2012**,** 11.

LUNDIN, S. 2012. *Methods to Prepare DNA for Efficient Massive Sequencing.* Royal Institute of Technology (KTH).

LUNETTA, K. L. 2008. Genetic Association Studies. *Circulation,* 118**,** 96-101.

LURATI BUSE, G. A.,BOTTO, F. & DEVEREAUX, P. J. 2012. Revisiting sample size: are big trials the answer? *The Journal of bone and joint surgery. American volume,* 94 Suppl 1**,** 75-9.

LYNCH, M. 2002. Intron evolution as a population-genetic process. *Proc Natl Acad Sci U S A,* 99**,** 6118-23.

MAHFOUZ, R. A.,HOTEIT, R.,SALEM, Z.,BAZARBACHI, A.,MUGHARBEL, A.,FARHAT, F.,ZIYADEH, A.,IBRAHIM, A. & TAHER, A. 2011. JAK2 V617F gene mutation in the laboratory work-up of myeloproliferative disorders: experience of a major referral center in Lebanon. *Genetic testing and molecular biomarkers,* 15**,** 263-5.

MALINGE, S.,BEN-ABDELALI, R.,SETTEGRANA, C.,RADFORD-WEISS, I.,DEBRE, M.,BELDJORD, K.,MACINTYRE, E. A.,VILLEVAL, J. L.,VAINCHENKER, W.,BERGER, R.,BERNARD, O. A.,DELABESSE, E. & PENARD-LACRONIQUE, V. 2007. Novel activating JAK2 mutation in a patient with Down syndrome and B-cell precursor acute lymphoblastic leukemia. *Blood,* 109**,** 2202-4.

MANOLIO, T. A.,COLLINS, F. S.,COX, N. J.,GOLDSTEIN, D. B.,HINDORFF, L. A.,HUNTER, D. J.,MCCARTHY, M. I.,RAMOS, E. M.,CARDON, L. R.,CHAKRAVARTI, A.,CHO, J. H.,GUTTMACHER, A. E.,KONG, A.,KRUGLYAK, L.,MARDIS, E.,ROTIMI, C. N.,SLATKIN, M.,VALLE, D.,WHITTEMORE, A. S.,BOEHNKE, M.,CLARK, A. G.,EICHLER, E. E.,GIBSON, G.,HAINES, J. L.,MACKAY, T. F.,MCCARROLL, S. A. & VISSCHER, P. M. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-53.

MARTINEZ-AVILES, L.,BESSES, C.,ALVAREZ-LARRAN, A.,CERVANTES, F.,HERNANDEZ-BOLUDA, J. C. & BELLOSILLO, B. 2007. JAK2 exon 12 mutations in polycythemia vera or idiopathic erythrocytosis. *Haematologica,* 92**,** 1717-8.

MARTINEZ-AVILES, L.,BESSES, C.,ALVAREZ-LARRAN, A.,TORRES, E.,SERRANO, S. & BELLOSILLO, B. 2012. TET2, ASXL1, IDH1, IDH2, and c-CBL genes in JAK2- and MPL-negative myeloproliferative neoplasms. *Ann Hematol,* 91**,** 533-41.

MARTINEZ-TRILLOS, A.,GAYA, A.,MAFFIOLI, M.,ARELLANO-RODRIGO, E.,CALVO, X.,DIAZ-BEYA, M. & CERVANTES, F. 2010. Efficacy and tolerability of hydroxyurea in the treatment of the hyperproliferative manifestations of myelofibrosis: results in 40 patients. *Ann Hematol,* 89**,** 1233-7.

MCCLURE, R.,MAI, M. & LASHO, T. 2006. Validation of two clinically useful assays for evaluation of JAK2 V617F mutation in chronic myeloproliferative disorders. *Leukemia,* 20**,** 168-71.

MCKENNA, A.,HANNA, M.,BANKS, E.,SIVACHENKO, A.,CIBULSKIS, K.,KERNYTSKY, A.,GARIMELLA, K.,ALTSHULER, D.,GABRIEL, S.,DALY, M. & DEPRISTO, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res,* 20**,** 1297-303.

MCLORNAN, D.,PERCY, M. & MCMULLIN, M. F. 2006. JAK2 V617F: a single mutation in the myeloproliferative group of disorders. *Ulster Med J,* 75**,** 112-9.

MCMURRAY, C. T. & VAN HOLDE, K. E. 1986. Binding of ethidium bromide causes dissociation of the nucleosome core particle. *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 8472-6.

MERCHER, T.,WERNIG, G.,MOORE, S. A.,LEVINE, R. L.,GU, T. L.,FROHLING, S.,CULLEN, D.,POLAKIEWICZ, R. D.,BERNARD, O. A.,BOGGON, T. J.,LEE, B. H. & GILLILAND, D. G. 2006. JAK2T875N is a novel activating mutation that results in myeloproliferative disease with features of megakaryoblastic leukemia in a murine bone marrow transplantation model. *Blood,* 108**,** 2770-9.

MERK, K.,MATTSSON, B.,MATTSSON, A.,HOLM, G.,GULLBRING, B. & BJORKHOLM, M. 1990. The incidence of cancer among blood donors. *International journal of epidemiology,* 19**,** 505-9.

MERTES, F.,ELSHARAWY, A.,SAUER, S.,VAN HELVOORT, J. M. L. M.,VAN DER ZAAG, P. J.,FRANKE, A.,NILSSON, M.,LEHRACH, H. & BROOKES, A. J. 2011. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in Functional Genomics,* 10**,** 374-386.

MESA, R. A.,SILVERSTEIN, M. N.,JACOBSEN, S. J.,WOLLAN, P. C. & TEFFERI, A. 1999. Population-based incidence and survival figures in essential thrombocythemia and agnogenic myeloid metaplasia: an Olmsted County Study, 1976-1995. *Am J Hematol,* 61**,** 10-5.

METZKER, M. L. 2010. Sequencing technologies — the next generation. *Nature Reviews Genetics,* 11**,** 31-46.

MIRANDA, M. B. & JOHNSON, D. E. 2007. Signal transduction pathways that contribute to myeloid differentiation. *Leukemia,* 21**,** 1363-77.

MOBAREK, D. A.,ASCENSAO, J. L.,STRONCEK, D.,BENNETT, M.,ADAMS, S.,KVANLI, M. & SCHECHTER, G. P. Comparison of JAK2V617F prevalence in African American and caucasian patients with myeloproliferative disease (MPD). 2008 American Society of Clinical Oncology Annual Meeting, 2008 Chicago. American Society of Clinical Oncology.

MONTGOMERY, J.,WITTWER, C. T.,PALAIS, R. & ZHOU, L. 2007. Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. *Nat Protoc,* 2**,** 59-66.

MORRISON, S. J.,UCHIDA, N. & WEISSMAN, I. L. 1995. The Biology of Hematopoietic Stem Cells. *Annual Review of Cell and Developmental Biology,* 11**,** 35-71.

MOTOIKE, I. N.,MATSUMOTO, M.,DANJOH, I.,KATSUOKA, F.,KOJIMA, K.,NARIAI, N.,SATO, Y.,YAMAGUCHI-KABATA, Y.,ITO, S.,KUDO, H.,NISHIJIMA, I.,NISHIKAWA, S.,PAN, X.,SAITO, R.,SAITO, S.,SAITO, T.,SHIROTA, M.,TSUDA, K.,YOKOZAWA, J.,IGARASHI, K.,MINEGISHI, N.,TANABE, O.,FUSE, N.,NAGASAKI, M.,KINOSHITA, K.,YASUDA, J. & YAMAMOTO, M. 2014. Validation of multiple single nucleotide variation calls by additional exome analysis with a semiconductor sequencer to supplement data of whole-genome sequencing of a human population. *BMC Genomics,* 15**,** 673.

MURATI, A.,GELSI-BOYER, V.,ADELAIDE, J.,PEROT, C.,TALMANT, P.,GIRAUDIER, S.,LODE, L.,LETESSIER, A.,DELAVAL, B.,BRUNEL, V.,IMBERT, M.,GARAND, R.,XERRI, L.,BIRNBAUM, D.,MOZZICONACCI, M. J. & CHAFFANET, M. 2005. PCM1-JAK2 fusion in myeloproliferative disorders and acute erythroid leukemia with t(8;9) translocation. *Leukemia,* 19**,** 1692-6.

NAJEAN, Y. & RAIN, J. D. 1997. Treatment of polycythemia vera: the use of hydroxyurea and pipobroman in 292 patients under the age of 65 years. *Blood,* 90**,** 3370-7.

NAJEAN, Y.,RAIN, J. D. & BILLOTEY, C. 1998. Epidemiological data in polycythaemia vera: a study of 842 cases. *Hematology and cell therapy,* 40**,** 159-65.

NANGALIA, J.,MASSIE, C. E.,BAXTER, E. J.,NICE, F. L.,GUNDEM, G.,WEDGE, D. C.,AVEZOV, E.,LI, J.,KOLLMANN, K.,KENT, D. G.,AZIZ, A.,GODFREY, A. L.,HINTON, J.,MARTINCORENA, I.,VAN LOO, P.,JONES, A. V.,GUGLIELMELLI, P.,TARPEY, P.,HARDING, H. P.,FITZPATRICK, J. D.,GOUDIE, C. T.,ORTMANN, C. A.,LOUGHRAN, S. J.,RAINE, K.,JONES, D. R.,BUTLER, A. P.,TEAGUE, J. W.,O'MEARA, S.,MCLAREN, S.,BIANCHI, M.,SILBER, Y.,DIMITROPOULOU, D.,BLOXHAM, D.,MUDIE, L.,MADDISON, M.,ROBINSON, B.,KEOHANE, C.,MACLEAN, C.,HILL, K.,ORCHARD, K.,TAURO, S.,DU, M. Q.,GREAVES, M.,BOWEN, D.,HUNTLY, B. J.,HARRISON, C. N.,CROSS, N. C.,RON, D.,VANNUCCHI, A. M.,PAPAEMMANUIL, E.,CAMPBELL, P. J. & GREEN, A. R. 2013. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med,* 369**,** 2391-405.

NELSON, M. & STEENSMA, D. 2006. JAK2 V617F in myeloid disorders: What do we know now, and where are we headed? *Leukemia & Lymphoma,* 47**,** 177-194.

NELSON, M. R.,MARNELLOS, G.,KAMMERER, S.,HOYAL, C. R.,SHI, M. M.,CANTOR, C. R. & BRAUN, A. 2004. Large-scale validation of single nucleotide polymorphisms in gene regions. *Genome Res,* 14**,** 1664-8.

NEUBAUER, H.,CUMANO, A.,MULLER, M.,WU, H.,HUFFSTADT, U. & PFEFFER, K. 1998. Jak2 deficiency defines an essential developmental checkpoint in definitive hematopoiesis. *Cell,* 93**,** 397-409.

NEWTON, C. R.,GRAHAM, A.,HEPTINSTALL, L. E.,POWELL, S. J.,SUMMERS, C.,KALSHEKER, N.,SMITH, J. C. & MARKHAM, A. F. 1989a. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res,* 17**,** 2503-16.

NEWTON, C. R.,HEPTINSTALL, L. E.,SUMMERS, C.,SUPER, M.,SCHWARZ, M.,ANWAR, R.,GRAHAM, A.,SMITH, J. C. & MARKHAM, A. F. 1989b. Amplification refractory mutation system for prenatal diagnosis and carrier assessment in cystic fibrosis. *Lancet,* 2**,** 1481-3.

NG, D. P.,KOH, D.,CHOO, S. & CHIA, K. S. 2006. Saliva as a viable alternative source of human genomic DNA in genetic epidemiology. *Clin Chim Acta,* 367**,** 81-5.

NGUYEN, H. M. & GOTLIB, J. 2012. Insights into the molecular genetics of myeloproliferative neoplasms. *Am Soc Clin Oncol Educ Book***,** 411-8.

NIELSEN, C.,BIRGENS, H. S.,NORDESTGAARD, B. G.,KJAER, L. & BOJESEN, S. E. 2011. The JAK2 V617F somatic mutation, mortality and cancer risk in the general population. *Haematologica,* 96**,** 450-3.

NIELSEN, C.,BOJESEN, S. E.,NORDESTGAARD, B. G.,KOFOED, K. F. & BIRGENS, H. S. 2014. JAK2V617F somatic mutation in the general population: myeloproliferative neoplasm development and progression rate. *Haematologica*.

O'SULLIVAN, L. A.,LIONGUE, C.,LEWIS, R. S.,STEPHENSON, S. E. & WARD, A. C. 2007. Cytokine receptor signaling through the Jak-Stat-Socs pathway in disease. *Mol Immunol,* 44**,** 2497-506.

OH, S. T.,SIMONDS, E. F.,JONES, C.,HALE, M. B.,GOLTSEV, Y.,GIBBS, K. D., JR.,MERKER, J. D.,ZEHNDER, J. L.,NOLAN, G. P. & GOTLIB, J. 2010a. Novel mutations in the inhibitory adaptor protein LNK drive JAK-STAT signaling in patients with myeloproliferative neoplasms. *Blood,* 116**,** 988-92.

OH, S. T.,SIMONDS, E. F.,JONES, C.,HALE, M. B.,GOLTSEV, Y.,GIBBS, K. D.,MERKER, J. D.,ZEHNDER, J. L.,NOLAN, G. P. & GOTLIB, J. 2010b. Novel mutations in the inhibitory adaptor protein LNK drive JAK-STAT signaling in patients with myeloproliferative neoplasms. *Blood,* 116**,** 988-992.

OHYASHIKI, J. H.,YONETA, M.,HISATOMI, H.,IWABUCHI, T.,UMEZU, T. & OHYASHIKI, K. 2012. The C allele of JAK2 rs4495487 is an additional candidate locus that contributes to myeloproliferative neoplasm predisposition in the Japanese population. *BMC Med Genet,* 13**,** 6.

OLCAYDU, D.,HARUTYUNYAN, A.,JAGER, R.,BERG, T.,GISSLINGER, B.,PABINGER, I.,GISSLINGER, H. & KRALOVICS, R. 2009a. A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat Genet,* 41**,** 450-4.

OLCAYDU, D.,RUMI, E.,HARUTYUNYAN, A.,PASSAMONTI, F.,PIETRA, D.,PASCUTTO, C.,BERG, T.,JAGER, R.,HAMMOND, E.,CAZZOLA, M. & KRALOVICS, R. 2011. The role of the JAK2 GGCC haplotype and the TET2 gene in familial myeloproliferative neoplasms. *Haematologica,* 96**,** 367-74.

OLCAYDU, D.,SKODA, R. C.,LOOSER, R.,LI, S.,CAZZOLA, M.,PIETRA, D.,PASSAMONTI, F.,LIPPERT, E.,CARILLO, S.,GIRODON, F.,VANNUCCHI, A.,READING, N. S.,PRCHAL, J. T.,AY, C.,PABINGER, I.,GISSLINGER, H. & KRALOVICS, R. 2009b. The 'GGCC' haplotype of JAK2 confers susceptibility to JAK2 exon 12 mutation-positive polycythemia vera. *Leukemia,* 23**,** 1924-6.

OLD, J. 1996. Haemoglobinopathies. *Prenat Diagn,* 16**,** 1181-6.

OLD, J. M.,VARAWALLA, N. Y. & WEATHERALL, D. J. 1990. Rapid detection and prenatal diagnosis of beta-thalassaemia: studies in Indian and Cypriot populations in the UK. *Lancet,* 336**,** 834-7.

ORAZI, A. & GERMING, U. 2008. The myelodysplastic/myeloproliferative neoplasms: myeloproliferative diseases with dysplastic features. *Leukemia,* 22**,** 1308-19.

ORKIN, S. H. & ZON, L. I. 2008. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell,* 132**,** 631-44.

PAGLIARINI-E-SILVA, S.,SANTOS, B. C.,PEREIRA, E. M.,FERREIRA, M. E.,BARALDI, E. C.,SELL, A. M. & VISENTAINER, J. E. 2013. Evaluation of the association between the JAK2 46/1 haplotype and chronic myeloproliferative neoplasms in a Brazilian population. *Clinics (Sao Paulo),* 68**,** 5-9.

PAJNIC, I. Z.,PODOVSOVNIK AXELSSON, E. & BALAZIC, J. 2014. Slovenian population data for five new European Standard Set short tandem repeat loci and SE33 locus. *Croat Med J,* 55**,** 14-8.

PALANDRI, F.,OTTAVIANI, E.,SALMI, F.,CATANI, L.,POLVERELLI, N.,FIACCHINI, M.,MARTINELLI, G.,BACCARANI, M. & VIANELLI, N. 2009. JAK2 V617F mutation in essential thrombocythemia: correlation with clinical characteristics, response to therapy and long-term outcome in a cohort of 275 patients. *Leuk Lymphoma,* 50**,** 247-53.

PARDANANI, A.,FRIDLEY, B. L.,LASHO, T. L.,GILLILAND, D. G. & TEFFERI, A. 2008. Host genetic variation contributes to phenotypic diversity in myeloproliferative disorders. *Blood,* 111**,** 2785-9.

PARDANANI, A.,LASHO, T. L.,FINKE, C.,HANSON, C. A. & TEFFERI, A. 2007. Prevalence and clinicopathologic correlates of JAK2 exon 12 mutations in JAK2V617F-negative polycythemia vera. *Leukemia,* 21**,** 1960-3.

PARDANANI, A.,LASHO, T. L.,FINKE, C. M.,GANGAT, N.,WOLANSKYJ, A. P.,HANSON, C. A. & TEFFERI, A. 2010. The JAK2 46/1 haplotype confers susceptibility to essential thrombocythemia regardless of JAK2V617F mutational status-clinical correlates in a study of 226 consecutive patients. *Leukemia,* 24**,** 110-4.

PASQUIER, F.,CABAGNOLS, X.,SECARDIN, L.,PLO, I. & VAINCHENKER, W. 2014. Myeloproliferative neoplasms: JAK2 signaling pathway as a central target for therapy. *Clin Lymphoma Myeloma Leuk,* 14 Suppl**,** S23-35.

PASSAMONTI, F.,RUMI, E.,PIETRA, D.,ELENA, C.,BOVERI, E.,ARCAINI, L.,RONCORONI, E.,ASTORI, C.,MERLI, M.,BOGGI, S.,PASCUTTO, C.,LAZZARINO, M. & CAZZOLA, M. 2010. A prospective study of 338 patients with polycythemia vera: the impact of JAK2 (V617F) allele burden and leukocytosis on fibrotic or leukemic disease transformation and vascular complications. *Leukemia,* 24**,** 1574-9.

PASSAMONTI, F.,RUMI, E.,PUNGOLINO, E.,MALABARBA, L.,BERTAZZONI, P.,VALENTINI, M.,ORLANDI, E.,ARCAINI, L.,BRUSAMOLINO, E.,PASCUTTO, C.,CAZZOLA, M.,MORRA, E. & LAZZARINO, M. 2004. Life expectancy and prognostic factors for survival in patients with polycythemia vera and essential thrombocythemia. *Am J Med,* 117**,** 755-61.

PATNAIK, M. M.,LASHO, T. L.,FINKE, C. M.,GANGAT, N.,CARAMAZZA, D.,SIRAGUSA, S.,HANSON, C. A.,PARDANANI, A. & TEFFERI, A. MPL mutation effect on JAK2 46/1 haplotype frequency in JAK2V617F-negative myeloproliferative neoplasms. *Leukemia,* 24**,** 859-60.

PATNAIK, M. M.,LASHO, T. L.,FINKE, C. M.,GANGAT, N.,CARAMAZZA, D.,SIRAGUSA, S.,HANSON, C. A.,PARDANANI, A. & TEFFERI, A. 2010. MPL mutation effect on JAK2 46/1 haplotype frequency in JAK2V617F-negative myeloproliferative neoplasms. *Leukemia,* 24**,** 859-60.

PAYZIN, K. B.,SAVASOGLU, K.,ALACACIOGLU, I.,OZDEMIRKIRAN, F.,MUTLU, B. B.,BENER, S.,CALLI, A. O.,KUCUKZEYBEK, B. B. & HAKSUN, S. 2014. JAK2 V617F Mutation Status of 232 Patients Diagnosed With Chronic Myeloproliferative Neoplasms. *Clinical lymphoma, myeloma & leukemia*.

PEETERS, P.,RAYNAUD, S. D.,COOLS, J.,WLODARSKA, I.,GROSGEORGE, J.,PHILIP, P.,MONPOUX, F.,VAN ROMPAEY, L.,BAENS, M.,VAN DEN BERGHE, H. & MARYNEN, P. 1997. Fusion of TEL, the ETS-variant gene 6 (ETV6), to the receptor-associated kinase JAK2 as a result of t(9;12) in a lymphoid and t(9;15;12) in a myeloid leukemia. *Blood,* 90**,** 2535-40.

PERCY, M. J.,SCOTT, L. M.,ERBER, W. N.,HARRISON, C. N.,REILLY, J. T.,JONES, F. G.,GREEN, A. R. & MCMULLIN, M. F. 2007. The frequency of JAK2 exon 12 mutations in idiopathic erythrocytosis patients with low serum erythropoietin levels. *Haematologica,* 92**,** 1607-14.

PERRY, S.,CRADDOCK, C. G.,JR,PAUL, G. & LAWRENCE, J. S. 1959. LYmphocyte production and turnover. *A.M.A. Archives of Internal Medicine,* 103**,** 224-230.

PESU, M.,O'SHEA, J.,HENNIGHAUSEN, L. & SILVENNOINEN, O. 2005. Identification of an acquired mutation in Jak2 provides molecular insights into the pathogenesis of myeloproliferative disorders. *Mol Interv,* 5**,** 211-5.

PIETRA, D.,CASETTI, I.,DA VIA, M. C.,ELENA, C.,MILANESI, C. & RUMI, E. 2012. JAK2 GGCC haplotype in MPL mutated myeloproliferative neoplasms. *Am J Hematol,* 87**,** 746-7.

PIETRA, D.,LI, S.,BRISCI, A.,PASSAMONTI, F.,RUMI, E.,THEOCHARIDES, A.,FERRARI, M.,GISSLINGER, H.,KRALOVICS, R.,CREMONESI, L.,SKODA, R. & CAZZOLA, M. 2008. Somatic mutations of JAK2 exon 12 in patients with JAK2 (V617F)-negative myeloproliferative disorders. *Blood,* 111**,** 1686-9.

PIKMAN, Y.,LEE, B. H.,MERCHER, T.,MCDOWELL, E.,EBERT, B. L.,GOZO, M.,CUKER, A.,WERNIG, G.,MOORE, S.,GALINSKY, I.,DEANGELO, D. J.,CLARK, J. J.,LEE, S. J.,GOLUB, T. R.,WADLEIGH, M.,GILLILAND, D. G. & LEVINE, R. L. 2006. <italic>MPLW515L</italic> Is a Novel Somatic Activating Mutation in Myelofibrosis with Myeloid Metaplasia. *PLoS Med,* 3**,** e270.

PRITCHARD, J. *eQTL resources @ the Pritchard lab* [Online]. Available: http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/.

PRITCHARD, J. K. & COX, N. J. 2002. The allelic architecture of human disease genes: common disease–common variant…or not? *Human Molecular Genetics,* 11**,** 2417-2423.

PRITCHARD, M. A.,BAKER, E.,CALLEN, D. F.,SUTHERLAND, G. R. & WILKS, A. F. 1992. Two members of the JAK family of protein tyrosine kinases map to chromosomes 1p31.3 and 9p24. *Mamm Genome,* 3**,** 36-8.

PUBLICOVER, A. & MEDD, P. 2013. Myeloproliferative neoplasms. *Clinical Medicine,* 13**,** 188-192.

PUIG, O.,YUAN, J.,STEPANIANTS, S.,ZIEBA, R.,ZYCBAND, E.,MORRIS, M.,COULTER, S.,YU, X.,MENKE, J.,WOODS, J.,CHEN, F.,RAMEY, D. R.,HE, X.,O'NEILL, E. A.,HAILMAN, E.,JOHNS, D. G.,HUBBARD, B. K.,YEE LUM, P.,WRIGHT, S. D.,DESOUZA, M. M.,PLUMP, A. & REISER, V. 2011. A gene expression signature that classifies human atherosclerotic plaque by relative inflammation status. *Circ Cardiovasc Genet,* 4**,** 595-604.

PURCELL, S.,CHERNY, S. S. & SHAM, P. C. 2003. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics,* 19**,** 149-50.

PURCELL, S.,NEALE, B.,TODD-BROWN, K.,THOMAS, L.,FERREIRA, M. A.,BENDER, D.,MALLER, J.,SKLAR, P.,DE BAKKER, P. I.,DALY, M. J. & SHAM, P. C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet,* 81**,** 559-75.

QUENTMEIER, H.,MACLEOD, R. A. F.,ZABORSKI, M. & DREXLER, H. G. 2006. JAK2 V617F tyrosine kinase mutation in cell lines derived from myeloproliferative disorders. *Leukemia,* 20**,** 471-476.

RANE, S. G. & REDDY, E. P. 1994. JAK3: a novel JAK kinase associated with terminal differentiation of hematopoietic cells. *Oncogene,* 9**,** 2415-23.

RANE, S. G. & REDDY, E. P. 2000. Janus kinases: components of multiple signaling pathways. *Oncogene,* 19**,** 5662-79.

RASMUSSEN, H. B. 2012. Restriction Fragment Length Polymorphism Analysis of PCR-Amplified Fragments (PCR-RFLP) and Gel Electrophoresis - Valuable Tool for Genotyping and Genetic Fingerprinting. *In:* MAGDELDIN, S. (ed.) *Gel Electrophoresis - Principles and Basics.* InTech.

REICH, D. E.,CARGILL, M.,BOLK, S.,IRELAND, J.,SABETI, P. C.,RICHTER, D. J.,LAVERY, T.,KOUYOUMJIAN, R.,FARHADIAN, S. F.,WARD, R. &

LANDER, E. S. 2001. Linkage disequilibrium in the human genome. *Nature,* 411**,** 199-204.

REMES-LENICOV, F.,JACOBSEN, K. X.,ROGAEVA, A.,CZESAK, M.,HADJIGHASEM, M.,DAIGLE, M. & ALBERT, P. R. 2007. Identification of Novel Transcriptional Regulators in the Nervous System. *In:* CHATTOPADHYAY, A. (ed.) *Serotonin Receptors in Neurobiology.* Boca Raton (FL).

REVIEW MANAGER (REVMAN) [COMPUTER PROGRAM]. VERSION 5.3.3. COPENHAGEN: THE NORDIC COCHRANE CENTRE, T. C. C., 2011.

RIRIE, K.,RASMUSSEN, R. & WITTWER, C. 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem,* 245**,** 154 - 160.

RISCH, N. & MERIKANGAS, K. 1996. The Future of Genetic Studies of Complex Human Diseases. *Science,* 273**,** 1516-1517.

ROBINSON, J. T.,THORVALDSDOTTIR, H.,WINCKLER, W.,GUTTMAN, M.,LANDER, E. S.,GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nat Biotechnol,* 29**,** 24-6.

RODRIGUEZ-ABREU, D.,BORDONI, A. & ZUCCA, E. 2007. Epidemiology of hematological malignancies. *Annals of Oncology,* 18**,** i3-i8.

ROHRBACHER, M. & HASFORD, J. 2009. Epidemiology of chronic myeloid leukaemia (CML). *Best Pract Res Clin Haematol,* 22**,** 295-302.

ROLLISON, D. E.,HOWLADER, N.,SMITH, M. T.,STROM, S. S.,MERRITT, W. D.,RIES, L. A.,EDWARDS, B. K. & LIST, A. F. 2008. Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs. *Blood,* 112**,** 45-52.

ROMMENS, J. M.,IANNUZZI, M. C.,KEREM, B.,DRUMM, M. L.,MELMER, G.,DEAN, M.,ROZMAHEL, R.,COLE, J. L.,KENNEDY, D.,HIDAKA, N. & ET AL. 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science,* 245**,** 1059-65.

ROWLEY, J. D. 1973. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. *Nature,* 243**,** 290-293.

RUIZ-ARGUELLES, G. J.,GARCES-EISELE, J.,ORTIZ-LOPEZ, R.,RIVAS-LLAMAS, R.,GOMEZ-ALMAGUER, D. & RUIZ-DELGADO, G. J. 2009. Molecular characterization of chronic myeloproliferative neoplasias in Mexico. *Hematology,* 14**,** 261-5.

RUMI, E.,PASSAMONTI, F.,ELENA, C.,PIETRA, D.,ARCAINI, L.,ASTORI, C.,ZIBELLINI, S.,BOVERI, E.,PASCUTTO, C. & LAZZARINO, M. 2011. Increased risk of lymphoid neoplasm in patients with myeloproliferative neoplasm: a study of 1,915 patients. *Haematologica,* 96**,** 454-8.

RUMI, E.,PIETRA, D.,GUGLIELMELLI, P.,BORDONI, R.,CASETTI, I.,MILANESI, C.,SANT'ANTONIO, E.,FERRETTI, V.,PANCRAZZI, A.,ROTUNNO, G.,SEVERGNINI, M.,PIETRELLI, A.,ASTORI, C.,FUGAZZA, E.,PASCUTTO, C.,BOVERI, E.,PASSAMONTI, F.,DE BELLIS, G.,VANNUCCHI, A.,CAZZOLA, M. & ASSOCIAZIONE ITALIANA PER LA RICERCA SUL CANCRO GRUPPO ITALIANO MALATTIE, M. 2013. Acquired copy-neutral loss of heterozygosity of chromosome 1p as a molecular event associated with marrow fibrosis in MPL-mutated myeloproliferative neoplasms. *Blood,* 121**,** 4388-95.

SABATINE, M. S.,SEIDMAN, J. G. & SEIDMAN, C. E. 2006. Cardiovascular Genomics. *Circulation,* 113**,** e450-e455.

SAHARINEN, P. & SILVENNOINEN, O. 2002. The Pseudokinase Domain Is Required for Suppression of Basal Activity of Jak2 and Jak3 Tyrosine Kinases and for Cytokine-inducible Activation of Signal Transduction. *Journal of Biological Chemistry,* 277**,** 47954-47963.

SAHARINEN, P.,VIHINEN, M. & SILVENNOINEN, O. 2003. Autoinhibition of Jak2 tyrosine kinase is dependent on specific regions in its pseudokinase domain. *Mol Biol Cell,* 14**,** 1448-59.

SAKHARKAR, M. K.,CHOW, V. T. & KANGUEANE, P. 2004. Distributions of exons and introns in the human genome. *In Silico Biol,* 4**,** 387-93.

SALISBURY, B. A.,PUNGLIYA, M.,CHOI, J. Y.,JIANG, R.,SUN, X. J. & STEPHENS, J. C. 2003. SNP and haplotype variation in the human genome. *Mutat Res,* 526**,** 53-61.

SALLMYR, A.,FAN, J. & RASSOOL, F. V. 2008. Genomic instability in myeloid malignancies: increased reactive oxygen species (ROS), DNA double strand breaks (DSBs) and error-prone repair. *Cancer Lett,* 270**,** 1-9.

SALTZMAN, A.,STONE, M.,FRANKS, C.,SEARFOSS, G.,MUNRO, R.,JAYE, M. & IVASHCHENKO, Y. 1998. Cloning and characterization of human Jak-2 kinase: high mRNA expression in immune cells and muscle tissue. *Biochem Biophys Res Commun,* 246**,** 627-33.

SANADA, M.,SUZUKI, T.,SHIH, L. Y.,OTSU, M.,KATO, M.,YAMAZAKI, S.,TAMURA, A.,HONDA, H.,SAKATA-YANAGIMOTO, M.,KUMANO, K.,ODA, H.,YAMAGATA, T.,TAKITA, J.,GOTOH, N.,NAKAZAKI, K.,KAWAMATA, N.,ONODERA, M.,NOBUYOSHI, M.,HAYASHI, Y.,HARADA, H.,KUROKAWA, M.,CHIBA, S.,MORI, H.,OZAWA, K.,OMINE, M.,HIRAI, H.,NAKAUCHI, H.,KOEFFLER, H. P. & OGAWA, S. 2009. Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. *Nature,* 460**,** 904-8.

SANGER, F. & COULSON, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol,* 94**,** 441-8.

SANTOS, L. C. D.,RIBEIRO, J. C. D. C.,SILVA, N. P.,CERUTTI, J.,SILVA, M. R. R. D. & CHAUFFAILLE, M. D. L. L. F. 2011. Cytogenetics, JAK2 and MPL mutations in polycythemia vera, primary myelofibrosis and essential thrombocythemia. *Revista Brasileira de Hematologia e Hemoterapia,* 33**,** 417-424.

SAXENA R & SK., G. 2008. Myeloproliferative Neoplasms: The Era of JAK2 Mutations and More... *Indian Journal of Medical & Paediatric Oncology,* 29.

SAZAWAL, S.,BAJAJ, J.,CHIKKARA, S.,JAIN, S.,BHARGAVA, R.,MAHAPATRA, M. & SAXENA, R. 2010. Prevalence of JAK2 V617F mutation in Indian patients with chronic myeloproliferative disorders. *The Indian journal of medical research,* 132**,** 423-7.

SCHINDLER, C.,LEVY, D. E. & DECKER, T. 2007. JAK-STAT Signaling: From Interferons to Cytokines. *Journal of Biological Chemistry,* 282**,** 20059-20063.

SCHMERER, M. & EVANS, T. 2001. Haematopoiesis. *eLS.* John Wiley & Sons, Ltd.

SCHNITTGER, S.,BACHER, U.,KERN, W.,SCHRODER, M.,HAFERLACH, T. & SCHOCH, C. 2006. Report on two novel nucleotide exchanges in the JAK2 pseudokinase domain: D620E and E627E. *Leukemia,* 20**,** 2195-7.

SCHOLZ, G. M.,CARTLEDGE, K. & HALL, N. E. 2001. Identification and characterization of Harc, a novel Hsp90-associating relative of Cdc37. *J Biol Chem,* 276**,** 30971-9.

SCHWARZ, J. M.,COOPER, D. N.,SCHUELKE, M. & SEELOW, D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Meth,* 11**,** 361-362.

SCOTT, L. M.,TONG, W.,LEVINE, R. L.,SCOTT, M. A.,BEER, P. A.,STRATTON, M. R.,FUTREAL, P. A.,ERBER, W. N.,MCMULLIN, M. F.,HARRISON, C. N.,WARREN, A. J.,GILLILAND, D. G.,LODISH, H. F. & GREEN, A. R. 2007. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *N Engl J Med,* 356**,** 459-68.

SEAMAN, V.,JUMAAN, A.,YANNI, E.,LEWIS, B.,NEYER, J.,RODA, P.,XU, M. & HOFFMAN, R. 2009. Use of molecular testing to identify a cluster of patients with polycythemia vera in eastern Pennsylvania. *Cancer Epidemiol Biomarkers Prev,* 18**,** 534-40.

SEO, S.,TAKAYAMA, K.,UNO, K.,OHI, K.,HASHIMOTO, R.,NISHIZAWA, D.,IKEDA, K.,OZAKI, N.,NABESHIMA, T.,MIYAMOTO, Y. & NITTA, A. 2013. Functional Analysis of Deep Intronic SNP rs13438494 in Intron 24 of <italic>PCLO</italic> Gene. *PLoS ONE,* 8**,** e76960.

SHABANPOOR, F.,SEPAROVIC, F. & WADE, J. D. 2009. The human insulin superfamily of polypeptide hormones. *Vitam Horm,* 80**,** 1-31.

SHAH, B. D. & ZUCKERMAN, K. S. 2011. Normal and Malignant Hematopoiesis. *Advances in Malignant Hematology.* Wiley-Blackwell.

SHANG, L. & TOMASI, T. B. 2006. The heat shock protein 90-CDC37 chaperone complex is required for signaling by types I and II interferons. *The Journal of biological chemistry,* 281**,** 1876-84.

SHEN, P.,WANG, W.,CHI, A.-K.,FAN, Y.,DAVIS, R. & SCHARFE, C. 2013. Multiplex target capture with double-stranded DNA probes. *Genome Medicine,* 5**,** 50.

SHEN, Y. M.,CHAO, H. Y.,ZHANG, R.,FENG, Y. F.,CEN, J. N.,YAO, L.,SHEN, H. J.,ZHU, Z. L. & XUE, Y. Q. 2009. Quantitative assay for Janus kinase 2 (JAK2) mutation in Chinese patients with chronic myeloproliferative disorders. *J Int Med Res,* 37**,** 37-46.

SHIDE, K.,SHIMODA, H. K.,KUMANO, T.,KARUBE, K.,KAMEDA, T.,TAKENAKA, K.,OKU, S.,ABE, H.,KATAYOSE, K. S.,KUBUKI, Y.,KUSUMOTO, K.,HASUIKE, S.,TAHARA, Y.,NAGATA, K.,MATSUDA, T.,OHSHIMA, K.,HARADA, M. & SHIMODA, K. 2008. Development of ET, primary myelofibrosis and PV in mice expressing JAK2 V617F. *Leukemia,* 22**,** 87-95.

SHOCHAT, E.,STEMMER, S. & SEGEL, L. 2002. Human haematopoiesis in steady state and following intense perturbations. *Bulletin of Mathematical Biology,* 64**,** 861-886.

SIDON, P.,EL HOUSNI, H.,DESSARS, B. & HEIMANN, P. 2006. The JAK2V617F mutation is detectable at very low level in peripheral blood of healthy donors. *Leukemia,* 20**,** 1622.

SILVENNOINEN, O.,UNGUREANU, D.,NIRANJAN, Y.,HAMMAREN, H.,BANDARANAYAKE, R. & HUBBARD, S. R. 2013. New insights into the structure and function of the pseudokinase domain in JAK2. *Biochem Soc Trans,* 41**,** 1002-7.

SILVER, R. 2000. Chronic Myeloid Leukemia. *In:* BAST RC JR, K. D., POLLOCK RE, ET AL. (ed.) *Holland-Frei Cancer Medicine.* 5th ed.: BC Decker Inc.

SILVERMAN, E. K. & PALMER, L. J. 2000. Case-Control Association Studies for the Genetics of Complex Respiratory Diseases. *American Journal of Respiratory Cell and Molecular Biology,* 22**,** 645-648.

SIMSEK, M.,TANIRA, M. O.,AL-BALOUSHI, K. A.,AL-BARWANI, H. S.,LAWATIA, K. M. & BAYOUMI, R. A. 2001. A precaution in the detection of heterozygotes by sequencing: comparison of automated DNA sequencing and PCR-restriction fragment length polymorphism methods. *Clin Chem,* 47**,** 134-7.

SKODA, R. C. 2008. Can we control JAK? *Blood,* 111**,** 5419-20.

SMALBERG, J. H.,KOEHLER, E.,DARWISH MURAD, S.,PLESSIER, A.,SEIJO, S.,TREBICKA, J.,PRIMIGNANI, M.,DE MAAT, M. P.,GARCIA-PAGAN, J. C.,VALLA, D. C.,JANSSEN, H. L. & LEEBEEK, F. W. 2011. The JAK2 46/1 haplotype in Budd-Chiari syndrome and portal vein thrombosis. *Blood,* 117**,** 3968-73.

SMITH, E.,JONES, M. & DREW, P. 2009. Quantitation of DNA methylation by melt curve analysis. *BMC Cancer,* 9**,** 123.

SPASOVSKI, V.,TOSIC, N.,NIKCEVIC, G.,STOJILJKOVIC, M.,ZUKIC, B.,RADMILOVIC, M.,KARAN-DJURASEVIC, T.,SRZENTIC, S.,COLOVIC, M. & PAVLOVIC, S. 2013. The influence of novel transcriptional regulatory element in intron 14 on the expression of Janus kinase 2 gene in myeloproliferative neoplasms. *J Appl Genet,* 54**,** 21-6.

SPELETAS, M.,KATODRITOU, E.,DAIOU, C.,MANDALA, E.,PAPADAKIS, E.,KIOUMI, A.,RITIS, K. & KORANTZIS, I. 2007. Correlations of JAK2-V617F mutation with clinical and laboratory findings in patients with myeloproliferative disorders. *Leukemia research,* 31**,** 1053-62.

SPIVAK, J. L. & SILVER, R. T. 2008. The revised World Health Organization diagnostic criteria for polycythemia vera, essential thrombocytosis, and primary myelofibrosis: an alternative proposal. *Blood,* 112**,** 231-9.

STEENSMA, D. P. 2006. JAK2 V617F in myeloid disorders: molecular diagnostic techniques and their clinical utility: a paper from the 2005 William Beaumont Hospital Symposium on Molecular Pathology. *J Mol Diagn,* 8**,** 397-411; quiz 526.

STEVEN, P.,RUPP, J.,HUTTMANN, G.,KOOP, N.,LENSING, C.,LAQUA, H. & GEBERT, A. 2008. Experimental induction and three-dimensional two-photon imaging of conjunctiva-associated lymphoid tissue. *Invest Ophthalmol Vis Sci,* 49**,** 1512-7.

STEWART, A. J.,HANNENHALLI, S. & PLOTKIN, J. B. 2012. Why Transcription Factor Binding Sites Are Ten Nucleotides Long. *Genetics,* 192**,** 973-985.

STOMMEL, M. D., KATHERINE J. 2014. *Statistics for advanced practice nurses and health professionals,* New York, NY Springer Publishing Company.

STOREY, J. D. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 64**,** 479-498.

STRACHAN, T. & READ, A. 1999. Genetic mapping of mendelian characters. *Human Molecular Genetics.* 2nd edition ed. New York Wiley-Liss.

STRACHAN, T. & READ, A. 1999 Genetic mapping of complex characters. *Human Molecular Genetics.* 2nd edition ed. New York: Wiley-Liss.

SUKSOMYOS, N.,CHANPRASERT, S.,MAUNPASITPORN, C. & ROJNUCKARIN, P. 2012. Prevalence of jak2v617f mutation and its clinical correlation in thais with myeloproliferative neoplasm *International Journal of Biological & Medical Research,* 3**,** 1801-1805.

SUMMERER, D. 2009. Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics,* 94**,** 363-368.

SWERDLOW, S. H.,CAMPO, E.,HARRIS, N. L.,JAFFE, E. S.,PILERI, S. A.,STEIN, H.,THIELE, J. & VARDIMAN, J. W. 2008. *WHO Classification*

of Tumours of Haematopoietic and Lymphoid Tissues, France, Lyon : International Agency for Research on Cancer.

TAMIYA, T.,KASHIWAGI, I.,TAKAHASHI, R.,YASUKAWA, H. & YOSHIMURA, A. 2011. Suppressors of Cytokine Signaling (SOCS) Proteins and JAK/STAT Pathways: Regulation of T-Cell Inflammation by SOCS1 and SOCS3. *Arteriosclerosis, Thrombosis, and Vascular Biology,* 31**,** 980-985.

TANAKA, M.,YUJIRI, T.,ITO, S.,OKAYAMA, N.,TAKAHASHI, T.,SHINOHARA, K.,AZUNO, Y.,NAWATA, R.,HINODA, Y. & TANIZAWA, Y. 2013. JAK2 46/1 haplotype is associated with JAK2 V617F-positive myeloproliferative neoplasms in Japanese patients. *International journal of hematology,* 97**,** 409-13.

TANG, W. C.,YAP, M. K. & YIP, S. P. 2008. A review of current approaches to identifying human genes involved in myopia. *Clinical and Experimental Optometry,* 91**,** 4-22.

TEFFERI, A. 2008. Essential thrombocythemia, polycythemia vera, and myelofibrosis: current management and the prospect of targeted therapy. *Am J Hematol,* 83**,** 491-7.

TEFFERI, A. & GILLILAND, D. G. 2007. Oncogenes in myeloproliferative disorders. *Cell Cycle,* 6**,** 550-66.

TEFFERI, A.,LASHO, T. L.,ABDEL-WAHAB, O.,GUGLIELMELLI, P.,PATEL, J.,CARAMAZZA, D.,PIERI, L.,FINKE, C. M.,KILPIVAARA, O.,WADLEIGH, M.,MAI, M.,MCCLURE, R. F.,GILLILAND, D. G.,LEVINE, R. L.,PARDANANI, A. & VANNUCCHI, A. M. 2010a. IDH1 and IDH2 mutation studies in 1473 patients with chronic-, fibrotic- or blast-phase essential thrombocythemia, polycythemia vera or myelofibrosis. *Leukemia,* 24**,** 1302-9.

TEFFERI, A.,LASHO, T. L.,HUANG, J.,FINKE, C.,MESA, R. A.,LI, C. Y.,WU, W.,HANSON, C. A. & PARDANANI, A. 2008. Low JAK2V617F allele burden in primary myelofibrosis, compared to either a higher allele burden or unmutated status, is associated with inferior overall and leukemia-free survival. *Leukemia,* 22**,** 756-61.

TEFFERI, A.,LASHO, T. L.,PATNAIK, M. M.,FINKE, C. M.,HUSSEIN, K.,HOGAN, W. J.,ELLIOTT, M. A.,LITZOW, M. R.,HANSON, C. A. & PARDANANI, A. 2010b. JAK2 germline genetic variation affects disease susceptibility in primary myelofibrosis regardless of V617F mutational status: nullizygosity for the JAK2 46/1 haplotype is associated with inferior survival. *Leukemia,* 24**,** 105-9.

TEFFERI, A.,LASHO, T. L.,SCHWAGER, S. M.,STEENSMA, D. P.,MESA, R. A.,LI, C. Y.,WADLEIGH, M. & GARY GILLILAND, D. 2005. The JAK2(V617F) tyrosine kinase mutation in myelofibrosis with myeloid metaplasia: lineage specificity and clinical correlates. *Br J Haematol,* 131**,** 320-8.

TEFFERI, A.,LASHO, T. L.,SCHWAGER, S. M.,STRAND, J. S.,ELLIOTT, M.,MESA, R.,LI, C. Y.,WADLEIGH, M.,LEE, S. J. & GILLILAND, D. G. 2006. The clinical phenotype of wild-type, heterozygous, and homozygous JAK2V617F in polycythemia vera. *Cancer,* 106**,** 631-5.

TEFFERI, A.,THIELE, J. & VARDIMAN, J. W. 2009. The 2008 World Health Organization classification system for myeloproliferative neoplasms. *Cancer,* 115**,** 3842-3847.

TEFFERI, A. & VARDIMAN, J. W. 2008. Classification and diagnosis of myeloproliferative neoplasms: the 2008 World Health Organization criteria and point-of-care diagnostic algorithms. *Leukemia,* 22**,** 14-22.

TEO, Y. Y.,SMALL, K. S. & KWIATKOWSKI, D. P. 2010. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet,* 11**,** 149-60.

TERREROS, M. C.,APEZTEGUIA, M.,SLAVUTSKY, I. R. & GUIMAREY, L. M. 1997. Exposure to occupational and environmental risk factors in hematologic disorders. *Neoplasia,* 14**,** 133–136.

THEOCHARIDES, A.,BOISSINOT, M.,GIRODON, F.,GARAND, R.,TEO, S.-S.,LIPPERT, E.,TALMANT, P.,TICHELLI, A.,HERMOUET, S. & SKODA, R. C. 2007. *Leukemic blasts in transformed JAK2-V617F–positive myeloproliferative disorders are frequently negative for the JAK2-V617F mutation*.

THORSTEINSDOTTIR, S.,BJERRUM, O. W. & HASSELBALCH, H. C. 2013. Myeloproliferative neoplasms in five multiple sclerosis patients. *Leukemia Research Reports,* 2**,** 61-63.

TIAN, Z. Q.,ZHU, P.,LIU, H. X.,CHEN, Y.,WANG, F.,ZHANG, Y.,TENG, W.,GONG, Y. W.,XIA, J. Y.,BAI, D. C.,LIU, Q. & WU, X. Q. 2012. [Relationship between V617F mutation and 46/1 haplotype in JAK2 gene in patients with chronic myeloproliferative diseases and frequencies of 46/1 haplotype in different Chinese nationalities]. *Zhongguo Shi Yan Xue Ye Xue Za Zhi,* 20**,** 362-7.

TIEDT, R.,HAO-SHEN, H.,SOBAS, M. A.,LOOSER, R.,DIRNHOFER, S.,SCHWALLER, J. & SKODA, R. C. 2008. Ratio of mutant JAK2-V617F to wild-type Jak2 determines the MPD phenotypes in transgenic mice. *Blood,* 111**,** 3931-40.

TOMONAGA, M. 2001. [New classification of leukemia]. *Rinsho Byori,* Suppl 115**,** 45-53.

TRIFA, A. P.,CUCUIANU, A.,PETROV, L.,URIAN, L.,MILITARU, M. S.,DIMA, D.,POP, I. V. & POPP, R. A. The G allele of the JAK2 rs10974944 SNP, part of JAK2 46/1 haplotype, is strongly associated with JAK2 V617F-positive myeloproliferative neoplasms. *Ann Hematol,* 89**,** 979-83.

TRIFA, A. P.,CUCUIANU, A.,PETROV, L.,URIAN, L.,MILITARU, M. S.,DIMA, D.,POP, I. V. & POPP, R. A. 2010. The G allele of the JAK2 rs10974944 SNP, part of JAK2 46/1 haplotype, is strongly associated with JAK2 V617F-positive myeloproliferative neoplasms. *Ann Hematol,* 89**,** 979-83.

TYNER, J. W.,ERICKSON, H.,DEININGER, M. W.,WILLIS, S. G.,EIDE, C. A.,LEVINE, R. L.,HEINRICH, M. C.,GATTERMANN, N.,GILLILAND, D. G.,DRUKER, B. J. & LORIAUX, M. M. 2009. High-throughput sequencing screen reveals novel, transforming RAS mutations in myeloid leukemia patients. *Blood,* 113**,** 1749-55.

UNDI, R. B.,KANDI, R. & GUTTI, R. K. 2013. MicroRNAs as Haematopoiesis Regulators. *Advances in hematology,* 2013**,** 695754.

VAHIDNIA, F.,HIRSCHLER, N. V.,AGAPOVA, M.,CHINN, A.,BUSCH, M. P. & CUSTER, B. 2013. Cancer Incidence and Mortality in a Cohort of US Blood Donors: A 20-Year Study. *Journal of cancer epidemiology,* 2013**,** 814842.

VAINCHENKER, W.,DELHOMMEAU, F.,CONSTANTINESCU, S. N. & BERNARD, O. A. 2011. New mutations and pathogenesis of myeloproliferative neoplasms. *Blood,* 118**,** 1723-35.

VAINCHENKER, W.,DUSA, A. & CONSTANTINESCU, S. N. 2008. JAKs in pathology: role of Janus kinases in hematopoietic malignancies and immunodeficiencies. *Semin Cell Dev Biol,* 19**,** 385-93.

VAKIL, E. & TEFFERI, A. 2011. BCR-ABL1--negative myeloproliferative neoplasms: a review of molecular biology, diagnosis, and treatment. *Clinical lymphoma, myeloma & leukemia,* 11 Suppl 1**,** S37-45.

VANNUCCHI, A. M. 2009. Insights into the pathogenesis and management of thrombosis in polycythemia vera and essential thrombocythemia. *Intern Emerg Med*.

VANNUCCHI, A. M.,GUGLIELMELLI, P. & TEFFERI, A. 2009a. Advances in understanding and management of myeloproliferative neoplasms. *CA Cancer J Clin,* 59**,** 171-91.

VANNUCCHI, A. M.,LASHO, T. L.,GUGLIELMELLI, P.,BIAMONTE, F.,PARDANANI, A.,PEREIRA, A.,FINKE, C.,SCORE, J.,GANGAT, N.,MANNARELLI, C.,KETTERLING, R. P.,ROTUNNO, G.,KNUDSON, R. A.,SUSINI, M. C.,LABORDE, R. R.,SPOLVERINI, A.,PANCRAZZI, A.,PIERI, L.,MANFREDINI, R.,TAGLIAFICO, E.,ZINI, R.,JONES, A.,ZOI, K.,REITER, A.,DUNCOMBE, A.,PIETRA, D.,RUMI, E.,CERVANTES, F.,BAROSI, G.,CAZZOLA, M.,CROSS, N. C. & TEFFERI, A. 2013. Mutations and prognosis in primary myelofibrosis. *Leukemia,* 27**,** 1861-9.

VANNUCCHI, A. M.,MASALA, G.,ANTONIOLI, E.,CHIARA SUSINI, M.,GUGLIELMELLI, P.,PIERI, L.,MAGGI, L.,CAINI, S.,PALLI, D.,BOGANI, C.,PONZIANI, V.,PANCRAZZI, A.,ANNUNZIATO, F. & BOSI, A. 2009b. Increased risk of lymphoid neoplasms in patients with Philadelphia chromosome-negative myeloproliferative neoplasms. *Cancer Epidemiol Biomarkers Prev,* 18**,** 2068-73.

VARDIMAN, J. W.,THIELE, J.,ARBER, D. A.,BRUNNING, R. D.,BOROWITZ, M. J.,PORWIT, A.,HARRIS, N. L.,LE BEAU, M. M.,HELLSTROM-LINDBERG, E.,TEFFERI, A. & BLOOMFIELD, C. D. 2009. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood,* 114**,** 937-51.

VELAZQUEZ, L.,CHENG, A. M.,FLEMING, H. E.,FURLONGER, C.,VESELY, S.,BERNSTEIN, A.,PAIGE, C. J. & PAWSON, T. 2002. Cytokine signaling and hematopoietic homeostasis are disrupted in Lnk-deficient mice. *J Exp Med,* 195**,** 1599-611.

VENERI, D.,CAPUZZO, E.,DE MATTEIS, G.,FRANCHINI, M.,BARITONO, E.,BENATI, M.,SOLERO, G. P.,AMBROSETTI, A.,QUARESMINI, G. & PIZZOLO, G. 2009. Comparison of JAK2V617F mutation assessment employing different molecular diagnostic techniques. *Blood transfusion = Trasfusione del sangue,* 7**,** 204-9.

VERSTOVSEK, S. 2009. Therapeutic potential of JAK2 inhibitors. *Hematology Am Soc Hematol Educ Program***,** 636-42.

VILLANI, L.,BERGAMASCHI, G.,PRIMIGNANI, M.,ROSTI, V.,CAROLEI, A.,POLETTO, V.,CATARSI, P.,SPOLVERINI, A.,VANNUCCHI, A. M. & BAROSI, G. 2012. JAK2 46/1 haplotype predisposes to splanchnic vein thrombosis-associated BCR-ABL negative classic myeloproliferative neoplasms. *Leuk Res,* 36**,** e7-9.

WADLEIGH, M. & TEFFERI, A. 2010. Classification and diagnosis of myeloproliferative neoplasms according to the 2008 World Health Organization criteria. *International journal of hematology,* 91**,** 174-9.

WALKER, E.,HERNANDEZ, A. V. & KATTAN, M. W. 2008. Meta-analysis: Its strengths and limitations. *Cleve Clin J Med,* 75**,** 431-9.

WANG, J.,XU, Z.,LIU, L.,GALE, R. P.,CROSS, N. C.,JONES, A. V.,QIN, T.,AI, X.,XU, J.,ZHANG, T.,SUN, X.,LI, Q.,ZHANG, P.,ZHANG, Y. & XIAO, Z.

2013. JAK2V617F allele burden, JAK2 46/1 haplotype and clinical features of Chinese with myeloproliferative neoplasms. *Leukemia,* 27**,** 1763-7.

WANG, K.,LI, M. & HAKONARSON, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res,* 38**,** e164.

WANG, K.,SWIERCZEK, S.,HICKMAN, K.,HAKONARSON, H. & PRCHAL, J. T. 2011. Convergent mechanisms of somatic mutations in polycythemia vera. *Discov Med,* 12**,** 25-32.

WANG, Y. L.,VANDRIS, K.,JONES, A.,CROSS, N. C.,CHRISTOS, P.,ADRIANO, F. & SILVER, R. T. 2008. JAK2 Mutations are present in all cases of polycythemia vera. *Leukemia,* 22**,** 1289.

WARD, P. & KINNIBURGH, A. 2000. Molecular Genetic Methods in Diagnosis and Treatment. *In:* FAGUET, G. (ed.) *Hematologic Malignancies : Methods and Techniques.* Human press.

WESTERMEIER, R. 2011. Electrophoresis in gels. *Methods of biochemical analysis,* 54**,** 365-77.

WILKS, A. F. 1989. Two putative protein-tyrosine kinases identified by application of the polymerase chain reaction. *Proc Natl Acad Sci U S A,* 86**,** 1603-7.

WILLIAMS, D. M.,KIM, A. H.,ROGERS, O.,SPIVAK, J. L. & MOLITERNO, A. R. 2007. Phenotypic variations and new mutations in JAK2 V617F-negative polycythemia vera, erythrocytosis, and idiopathic myelofibrosis. *Exp Hematol,* 35**,** 1641-6.

WITTHUHN, B. A.,SILVENNOINEN, O.,MIURA, O.,LAI, K. S.,CWIK, C.,LIU, E. T. & IHLE, J. N. 1994. Involvement of the Jak-3 Janus kinase in signalling by interleukins 2 and 4 in lymphoid and myeloid cells. *Nature,* 370**,** 153-7.

WOLANSKYJ, A. P.,LASHO, T. L.,SCHWAGER, S. M.,MCCLURE, R. F.,WADLEIGH, M.,LEE, S. J.,GILLILAND, D. G. & TEFFERI, A. 2005. JAK2 mutation in essential thrombocythaemia: clinical associations and long-term prognostic relevance. *British Journal of Haematology,* 131**,** 208-13.

WONG, G. C.,KAM, G. L. & KOAY, E. S. 2011. JAK2 mutations in Asian patients with essential thrombocythaemia. *Intern Med J,* 41**,** 191-6.

WONG, R. S.,CHENG, C. K.,CHAN, N. P.,CHENG, S. H.,WONG, W. S.,LAU, K. M.,CHENG, G. & NG, M. H. 2008. JAK2 V617F mutation is associated with increased risk of thrombosis in Chinese patients with essential thrombocythaemia. *Br J Haematol,* 141**,** 902-4.

WU, W. M.,TSAI, H. J.,PANG, J. H.,WANG, H. S.,HONG, H. S. & LEE, Y. S. 2005. Touchdown thermocycling program enables a robust single nucleotide polymorphism typing method based on allele-specific real-time polymerase chain reaction. *Anal Biochem,* 339**,** 290-6.

XIA, K.,SHABALIN, A. A.,HUANG, S.,MADAR, V.,ZHOU, Y. H.,WANG, W.,ZOU, F.,SUN, W.,SULLIVAN, P. F. & WRIGHT, F. A. 2012. seeQTL: a searchable database for human eQTLs. *Bioinformatics,* 28**,** 451-2.

XIAO, Z.,ZHANG, Y.,LI, L.,NIE, L.,YANG, L. & XU, S. 2008. The Janus kinase 2 (JAK2) V617F mutation in Chinese patients with chronic myeloproliferative disorders. *Haematologica,* 93**,** 787-8.

XU, J.,TURNER, A.,LITTLE, J.,BLEECKER, E. R. & MEYERS, D. A. 2002. Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: Hint for genotyping error? *Human Genetics,* 111**,** 573-574.

XU, X.,ZHANG, Q.,LUO, J.,XING, S.,LI, Q.,KRANTZ, S. B.,FU, X. & ZHAO, Z. J. 2007. JAK2(V617F): Prevalence in a large Chinese hospital population. *Blood,* 109**,** 339-42.

XU, Z. & TAYLOR, J. A. 2009. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic acids research,* 37**,** W600-5.

YE, S.,DHILLON, S.,KE, X.,COLLINS, A. R. & DAY, I. N. 2001. An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res,* 29**,** E88-8.

YIP, S. P.,PUN, S. F.,LEUNG, K. H. & LEE, S. Y. 2003. Rapid, Simultaneous Genotyping of Five Common Southeast Asian {beta}-Thalassemia Mutations by Multiplex Minisequencing and Denaturing HPLC. *Clin Chem,* 49**,** 1656-1659.

YIU, W. C.,YAP, M. K.,FUNG, W. Y.,NG, P. W. & YIP, S. P. 2013. Genetic susceptibility to refractive error: association of vasoactive intestinal peptide receptor 2 (VIPR2) with high myopia in Chinese. *PloS one,* 8**,** e61805.

YOO, J. H.,PARK, T. S.,MAENG, H. Y.,SUN, Y. K.,KIM, Y. A.,KIE, J. H.,CHO, E. H.,SONG, J.,LEE, K. A.,SUH, B. & CHOI, J. R. 2009. JAK2 V617F/C618R mutation in a patient with polycythemia vera: a case study and review of the literature. *Cancer Genet Cytogenet,* 189**,** 43-7.

ZERJAVIC, K.,ZAGRADISNIK, B.,LOKAR, L.,KRASEVAC, M. G. & VOKAC, N. K. 2013. The association of the JAK2 46/1 haplotype with non-splanchnic venous thrombosis. *Thrombosis research,* 132**,** e86-93.

ZHA, Y.,LEUNG, K. H.,LO, K. K.,FUNG, W. Y.,NG, P. W.,SHI, M.-G.,YAP, M. K. H. & YIP, S. P. 2009. TGFB1 as a Susceptibility Gene for High Myopia: A Replication Study With New Findings. *Arch Ophthalmol,* 127**,** 541-548.

ZHAN, H. & SPIVAK, J. L. 2009. The diagnosis and management of polycythemia vera, essential thrombocythemia, and primary myelofibrosis in the JAK2 V617F era. *Clin Adv Hematol Oncol,* 7**,** 334-42.

ZHANG, S.,QIU, H.,FISCHER, B. S.,LI, W.,DUAN, L.,SUN, X.,XU, W. & LI, J. 2008. JAK2 V617F patients with essential thrombocythemia present with clinical features of polycythemia vera. *Leuk Lymphoma,* 49**,** 696-9.

ZHANG, S. J.,LI, J. Y.,LI, W. D.,SONG, J. H.,XU, W. & QIU, H. X. 2007. The investigation of JAK2 mutation in Chinese myeloproliferative diseases-identification of a novel C616Y point mutation in a PV patient. *Int J Lab Hematol,* 29**,** 71-2.

ZHANG, S. J.,QIU, H. X.,LI, J. Y.,SHI, J. Y. & XU, W. 2010. The analysis of JAK2 and MPL mutations and JAK2 single nucleotide polymorphisms in MPN patients by MassARRAY assay. *Int J Lab Hematol,* 32**,** 381-6.

ZHANG, X.,HU, T.,WU, Z.,KANG, Z.,LIU, W. & GUAN, M. 2012a. The JAK2 46/1 haplotype is a risk factor for myeloproliferative neoplasms in Chinese patients. *International Journal of Hematology***,** 1-6.

ZHANG, X.,HU, T.,WU, Z.,KANG, Z.,LIU, W. & GUAN, M. 2012b. The JAK2 46/1 haplotype is a risk factor for myeloproliferative neoplasms in Chinese patients. *Int J Hematol,* 96**,** 611-6.

ZHAO, R.,XING, S.,LI, Z.,FU, X.,LI, Q.,KRANTZ, S. B. & ZHAO, Z. J. 2005. Identification of an acquired JAK2 mutation in polycythemia vera. *J Biol Chem,* 280**,** 22788-92.

ZHOU, L.,MYERS, A. N.,VANDERSTEEN, J. G.,WANG, L. & WITTWER, C. T. 2004. Closed-tube genotyping with unlabeled oligonucleotide probes and a saturating DNA dye. *Clinical chemistry,* 50**,** 1328-35.

ZINTZARAS, E. 2010. Impact of Hardy-Weinberg equilibrium deviation on allele-based risk effect of genetic association studies and meta-analysis. *European journal of epidemiology,* 25**,** 553-60.

ZOLOTH, S. R.,MICHAELS, D. M.,VILLALBI, J. R. & LACHER, M. 1986. Patterns of mortality among commercial pressmen. *Journal of the National Cancer Institute,* 76**,** 1047-51.