Genetic lesions and clinical implications in myeloproliferative neoplasms

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Ronny Nienhold

aus Schmölln, Deutschland

Basel, 2017

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Professor Radek Skoda
Professor Christoph Handschin

Basel, den 21. März 2017

Prof. Dr. Martin Spiess
Dekan der Philosophisch-
Naturwissenschaftlichen Fakultät

Acknowledgements

First of all, I would like to thank Radek Skoda for giving me the opportunity to work as a PhD student in the lab. I'm the most thankful for being able to follow my interests and for getting insight into basic clinical research. Of course, it wasn't always easy, but I enjoyed growing with the tasks and responsibilities that were assigned to me. Furthermore, I want to thank Christoph Handschin and Jürg Schwaller for their support and guidance in the committee meetings.

Then I want to thank Axel and Pontus. The time we spent together on our NGS project was just great. It was a very inspiring period and I learned so many techniques, which I'm still using today. Thank you for everything.

Axel, Beatrice, Radek and Sara, you gave me something I never expected to get, as I am a trained biologist: Insight into clinical routine. This opened up a completely new field to me – and I got fully infected! Thank you for spending your time explaining and introducing me to this whole new world. I hope, I will find a job similar to my current position, which combines biological and clinical research.

Very special thanks to Gabi, Hélène, Hui, Renate and Silvia. All of my research would not have been possible if these five diamonds of the lab wouldn't have prepared this incredible amount of patient samples. Thank you very much for your help and your support!

I also want to thank every other person Adrian, Annalisa, Barbara, Imane, Jakub, Jean, Julian, Lucia, Morgane, Nilabh, Nils, Rao, Shivam, Takafumi and Yukiko for their help, support, trust and honesty. Knowing I would meet you in the lab made it much easier to get up in the morning! I enjoyed working with you and I want to thank each of you for the discussions we had and the suggestions I received from you. This thesis wouldn't have reached that stage without you. Thank you very much! Many thanks also to the colleagues from the neighboring hematology labs for their support!

Furthermore, I would like to thank all the core facilities, which supported me with their expertise during my thesis: The Quantitative Genomics team of the QGF (Christian, Elodie, Fabian, Ina, Katja and Manuel) run all the Illumina-based NGS analyses, which are the basis for most of the projects I was involved in. The Bioinformatics Core Facility (Florian and Robert) supported us in raw data analysis. I also want to acknowledge the support from IT (Antonio, Ilija, Marc, Nikki and Patric), human resources (Heidi and Funda), finances (Gerry) and facility management (Armin, Roy and Yves), Frank, Manuela and Xiomara who keep the DBM running on a daily basis and provided a great framework.

Finally, very special thanks to Oliver, Joel and Geoffroy, who offered us their database to organize our patient samples. Oliver encouraged me to quit my job and start studying. Thank you for everything, Oliver!

Natürlich haben auch Familie und Freunde Anteil am erfolgreichen Ausgang dieses Studiums. Allen voran, Nicole, meine Frau. Ich bedanke mich bei dir für die jahrelange Unterstützung, deine Geduld und natürlich deine Liebe. Wir haben es geschafft! Vielen Dank auch an unsere Familien, die uns im Verlauf des Studiums nicht nur moralisch unterstützt haben. Fabien, Sandrine, Tamara und Thomas vielen lieben Dank dafür, dass ihr uns nicht nur beim Studium unterstützt habt, sondern auch aus der Schweiz unsere neue Heimat gemacht habt. Wer könnte sich mit euch nicht wie daheim fühlen?!

Summary

Myeloproliferative neoplasms (MPN) are a group of chronic hematologic diseases, which are characterized by an over production of mature blood cells. The term MPN summarizes four main hematologic diseases: Chronic myelogenous leukemia (CML), essential thrombocythemia (ET), polycythemia vera (PV) and primary myelofibrosis (PMF). In this group of malignancies, CML patients can be identified by a chromosomal translocation, the Philadelphia chromosome. Patients with ET, PV and PMF share genetic markers, and can only be distinguished by clinical parameters. Currently, it is still unclear, how a specific genetic alteration can lead to different clinical phenotypes.

In the present thesis, we utilized novel NGS technologies and analyzed more than 100 genes in parallel to find secondary somatic mutations with impact on the phenotype of Ph-negative MPN patients. We also analyzed the temporal order of acquisition of these mutations and compared our results between adult MPN patients and rare pediatric cases. Our results show that in adult patients, the number of somatic mutations correlates with adverse survival and transformation to AML. Further, we report cases of patients who carry clones of TP53 mutations for multiple years and only when these clones expand, the MPN phenotypes of these patients progresses to AML. In pediatric MPN cases, we found a different mutational landscape and less genetic mutations per patient.

In the second part of the thesis, we analyzed patients with a very low mutant JAK2 allele burden and found that the MPN clone in these patients expands in the platelet and red cell lineages. Furthermore, we analyzed the genetic alterations of a large combined cohort to uncover instructive effects on the MPN phenotype of less frequently mutated genes. Finally, we show how different genetic alterations may modulate the MPN phenotype and propose a model of disease evolution and risk stratification in correlation to mutational events.

Table of contents

ACKNOWLEDGEMENTS 3

SUMMARY 5

TABLE OF CONTENTS 6

LIST OF ABBREVIATIONS 7

1 INTRODUCTION 8

1.1 HEMATOPOIESIS 8
1.2 CYTOKINE SIGNALING IN NORMAL AND MALIGNANT HEMATOPOIESIS 11
1.3 MYELOPROLIFERATIVE NEOPLASMS AND THEIR CLINICAL CLASSIFICATION 14
1.4 MOLECULAR CHARACTERISTICS OF ET, PV AND PMF 18
1.5 AIM OF THE STUDY 30

2 METHODS 31

2.1 STUDY POPULATION AND SAMPLING INTERVALS 31
2.2 BLOOD SAMPLE PREPARATION 33
2.3 MOLECULAR ANALYSES 36

3 RESULTS 43

3.1 CLONAL EVOLUTION AND CLINICAL CORRELATES OF SOMATIC MUTATIONS IN MPN 43
3.2 PEDIATRIC MPN PATIENTS DISPLAY A DIFFERENT MUTATIONAL PHENOTYPE 63
3.3 PATIENTS WITH LOW JAK2-V617F ALLELE BURDEN 68
3.4 MODULATION OF THE MPN PHENOTYPE BY ADDITIONAL SOMATIC MUTATIONS 79

4 DISCUSSION 90

4.1 MUTATIONAL LANDSCAPE OF MPN 90
4.2 PATTERNS OF MUTATIONAL ACQUISITION AND CLONAL EVOLUTION 93
4.3 LATE CLONAL EXPANSION IN PATIENTS WITH A LOW MUTANT JAK2 BURDEN 95
4.4 PEDIATRIC AND ADULT YOUNG ADULT PATIENTS SHOW DIFFERENT MUTATIONAL LANDSCAPE 97
4.5 SOMATIC MUTATIONS CAUSE DISEASE PROGRESSION AND REDUCED SURVIVAL 99
4.6 IDENTIFYING SOMATIC MUTATIONS AFFECTING THE MPN PHENOTYPE 104
4.7 THE COMPLEXITY OF GENOTYPE - PHENOTYPE CORRELATIONS IN MPN 108
4.8 FUTURE 110

5 DISCLOSURE OF INDIVIDUAL CONTRIBUTIONS 111

6 PUBLICATIONS 112

TABLE OF FIGURES 114

LIST OF TABLES 116

REFERENCES 117

APPENDIX 1: CLONAL EVOLUTION AND CLINICAL CORRELATES OF SOMATIC MUTATIONS IN MYELOPROLIFERATIVE NEOPLASMS 130

APPENDIX 2: MUTATIONAL PROFILE OF CHILDHOOD MYELOPROLIFERATIVE NEOPLASMS 149

List of abbreviations

AML	acute myeloid leukemia
asPCR	allele specific PCR
BFU-E	blast-forming unit erythroid
CFU-G	colony-forming unit granulocyte
CFU-M	colony-forming unit macrophage
CFU-Meg	colony-forming unit megakaryocytes
CLP	common lymphoid progenitor
CMP	common myeloid progenitor
DNA	deoxyribonucleic acid
EP	erythroid progenitor
ER	endoplasmic reticulum
ET	essential thrombocythemia
gDNA	genomic DNA
HSC	hematopoietic stem cell
LOH	loss of heterozygosity
MDS	myelodysplastic syndrome
MEP	megakaryocyte-erythroid progenitor
Mk	megakaryocyte
MkP	megakaryocyte progenitors
MPN	myeloproliferative neoplasms
MPP	multipotent progenitor
NC	normal control
NGS	next-generation sequencing
PBMC	peripheral blood mononuclear cell
PCR	polymerase chain reaction
PMF	primary myelofibrosis
PV	polycythemia vera
WHO	World Health Organization

1 Introduction

1.1 Hematopoiesis

Myeloproliferative neoplasms are a group of hematologic diseases, which are characterized by an overproduction of mature myeloid peripheral blood cells, such as granulocytes, red cells or platelets[1]. MPN are clonal disorders of hematopoietic stem cells (HSC) with an inherent tendency towards leukemic transformation.

During homeostasis, a tightly regulated hierarchical differentiation process called hematopoiesis constantly produces all hematopoietic lineages (Figure 1). HSC, a rare population of 1-2 x 10^5 HSC in human[2], represent the origin of this process. HSC asymmetrically divide giving rise to differentiated progenitors and maintain their own pool by self-renewal. They reside in the bone marrow hematopoietic niche forming a specific microenvironment[3]. The hematopoietic niche supports HSCs by nurturing and controlling quiescence, proliferation and differentiation. In homeostasis, two thirds of HSCs are kept in a quiescent state[4]. At a given time point, only one third of the HSCs actively contribute to hematopoiesis or self-renewal. During hematopoiesis, HSCs give rise to multipotent progenitors (MPPs), which only have limited self-renewal capacity[5]. MPPs give rise to common myeloid progenitors (CMP) and common lymphoid progenitors (CLP) with restricted myeloid or lymphoid lineage formation potential, therefore classified as oligopotent progenitors. The progeny of CMP further segregates between megakaryocyte–erythroid progenitors (MEP) and granulocyte-macrophage progenitors (GMP). The differentiation potential of MEP is limited to progenitors of platelets and red cells. GMPs can give rise to progenitors of granulocytes and macrophages. CLP generate progenitors of natural killer cells, B- and T-cells. In this classical model of hematopoietic hierarchy, it is generally believed that terminally differentiated platelet competent megakaryocytes (Mk) derive from HSC by sequential developmental transitions. According to this hypothesis, HSC differentiate through multipotent progenitors (MPP), common myeloid progenitors (CMP), MK-erythroid progenitors (MEP), and MK progenitor (MkP) stages. This paradigm was recently updated following the identification and characterization of potent stem-like Mk committed progenitors (SL-MkP) within the phenotypic HSC compartment. These SL-MkP are primed and generate MKs without succession through classical intermediate myeloid progenitor stages. In the traditional model of hematopoiesis in Figure 1, a grey dotted line visualizes these recently discovered lineage-restricted progenitors and their potential[5].

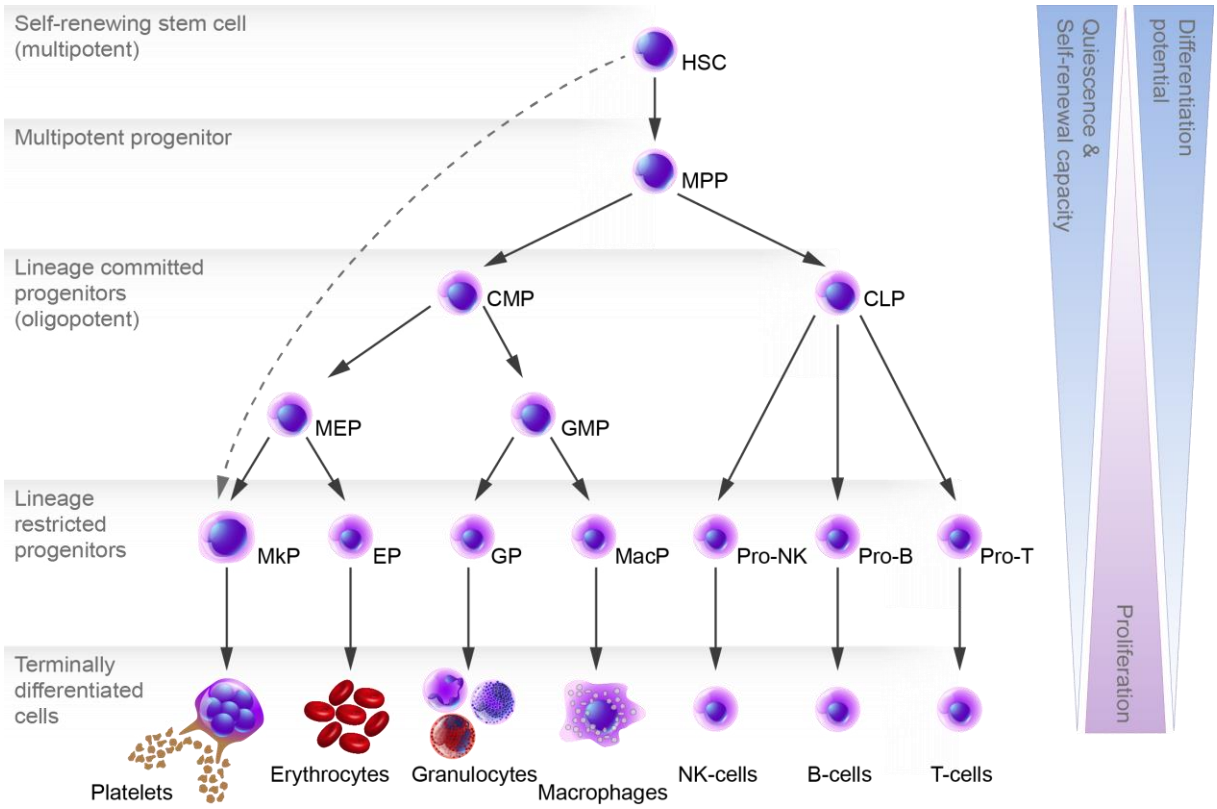


Figure 1 Hierarchy of hematopoiesis

The closer progenitors are to the terminally differentiated peripheral blood cells, to more quiescence, self-renewal and differentiation potential is reduced. In turn, proliferation is increased[4,5]. This results in massive production of mature blood cells in homeostasis: to balance out natural occurring cell death approximately 2.5 x 10^{11} red cells need to be generated per day[6]. The red blood cells represent nearly 95% of the cellular compartment. Platelets, important in hemostasis and wound healing, represent 5% of the blood cells. White blood cells, such granulocytes and monocytes, together with lymphoid cells (natural killer cells (NK-cells), B-cells and T-cells) represent only 0.1 % to 0.2 % of the blood cells. The numbers of the blood cells under normal conditions are summarized in Table 1[7].

In MPN, the numbers of myeloid cells are elevated due to somatic mutations, which affect genes involved in regulation of hematopoiesis. Different subtypes of MPN are defined according to the myeloid lineages that show increased cell counts. Patients are diagnosed with essential thrombocythemia (ET) when only the platelet numbers are increased. A diagnosis of polycythemia vera (PV) requires elevated red cell counts, but may include thrombocytosis and leukocytosis. The third main MPN phenotype, primary myelofibrosis (PMF), is diagnosed when in the bone marrow megakaryocytes are showing increased numbers and atypical morphology.

Blood cells	Cell types	No of cells per liter	Percentage
Red blood cells	Erythrocytes	5 x 10^{12}	94.2 %
Platelets	Platelets	3 x 10^{11}	5.6 %
Leukocytes	Neutrophils	5 x 10^{9}	0.1 %
	Eosinophils	2 x 10^{8}	0.004 %
	Basophils	4 x 10^{7}	0.0008 %
	Monocytes	4 x 10^{8}	0.008 %
Lymphoid cells	Natural Killer cells	1 x 10^{8}	0.002 %
	B-cells	2 x 10^{9}	0.04 %
	T-cells	1 x 10^{9}	0.02 %

Table 1 Cells of the peripheral blood

1.2 Cytokine signaling in normal and malignant hematopoiesis

To maintain the number of blood cells or to adjust them under physiological stress, internal and external regulators control hematopoiesis by balancing proliferation, differentiation and apoptosis. External regulators, like cytokines and growth factors, bind and induce signaling through their cognate cell surface receptors. Ligand binding of receptors is integrated by signaling cascades such as the JAK/STAT, RAS/MAPK and the PI3K/mTOR pathway. These cascades control internal regulators like transcription factors or miRNAs, which regulate hematopoiesis by adjusting expression of target genes at RNA level.

Therefore, in hematopoiesis the cytokine-signaling network is one of the most important systems, which enables cells to receive and concert information. Cytokines, such as thrombopoietin (TPO) and EPO (erythropoietin), are secreted proteins, which bind specific surface receptors. Cytokine receptors are classified in six families, the type I and type II cytokine receptors, the immunoglobulin superfamily, tumor necrosis factor receptor family, IL-17 family and G-protein coupled chemokine receptors (reviewed in [8]). The thrombopoietin receptor (MPL) and the erythropoietin receptor (EPOR) are both members of the type I receptor family and share the structural motifs of this class of receptors. The extracellular domains of type I cytokine receptors contain at least one duplet of cysteine repeats and a WSXWS motif. The intracellular domains of these receptors do not exhibit enzymatic activity and therefore, signaling of TPO and EPO through their cognate receptors MPL and EPOR is dependent on associated Janus kinase 2 (JAK2) enzymes (Figure 2)[9]. The family of JAK kinases consist of four proteins, JAK1, JAK2, JAK3 and TYK2, which all associate with specific motifs at the intra cellular domain of cytokine receptors[10–12]. Upon ligand binding, the JAKs trans-activate each other by trans-phosphorylation and also phosphorylate tyrosine residues at the intracellular domain of the cytokine receptor. These phosphorylated tyrosine residues represent binding sites for Signal Transducers and Activators of Transcription (STAT proteins)[13]. When bound to the receptor, STATs get phosphorylated by JAKs[14]. Upon phosphorylation, STATs dimerize and translocate to the nucleus[15]. In the nucleus, dimers of STATs activate gene expression by binding specific DNA sequences known as Insulin responsive elements (ISRE) or interferon gamma activated sequences (GAS)[16,17]. Cytokine receptors can also activate other signaling networks than the JAK/STAT pathway upon ligand binding[18]. Phosphorylation of the intracellular domain of the cytokine receptor by JAKs provides binding sites for proteins containing SH2 domains[19]. SHC, a protein with such a SH2 domain, binds to phosphorylated tyrosine at the receptor. In collaboration with GRB2 and SOS, SHC builds a scaffold for signaling through the RAS-MAPK-ERK pathway[20–24]. Also PI3K contains a SH2 domain and associates with the intracellular domain of cytokine receptors. The activation of PI3K by

receptor associated Janus kinases subsequently leads to enhanced AKT/mTOR signaling[19,25]. Both pathways, RAS-MAPK-ERK and PI3K-AKT1-mTOR, provide STAT independent signaling routes, which enhance survival and proliferation[22].

In cancer, the cytokine signaling network plays an important role as it is involved in promoting survival, suppressing apoptosis and modulating immune response to malignant cells[26,27]. In MPN, activating mutations in cytokine receptors[28–30], receptor associated JAK2[31–34] and other proteins [35,36] have been shown to stimulate cytokine signaling, even in absence of the receptor ligands. This enhanced signaling leads to constitutive production of mature hematopoietic cells which results in elevated platelet numbers or increased red cell count and represents the clinical features of MPN[1].

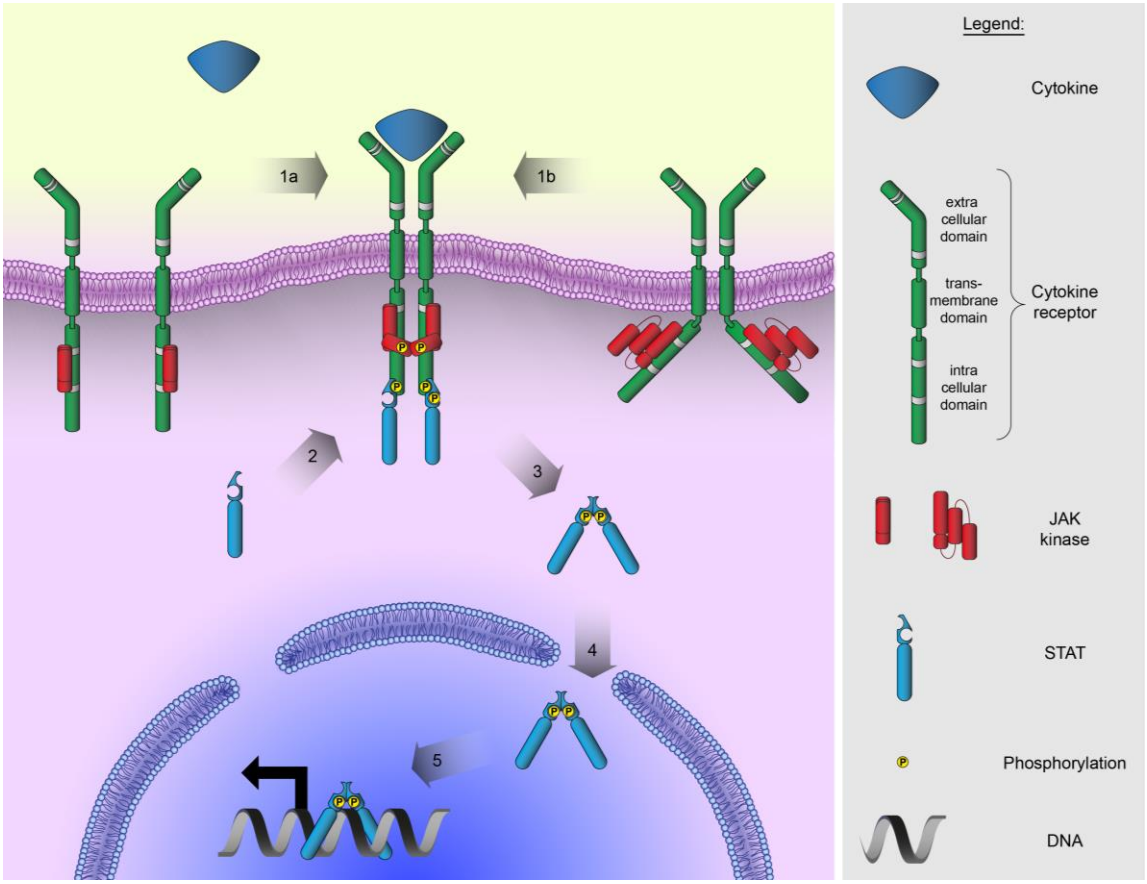


Figure 2 Cytokine-induced JAK-STAT signaling

Schematic drawing of the cytokine-induced JAK-STAT signaling: (1a) Monomeric receptors dimerize upon ligand binding or (1b) receptor multimers change conformation upon ligand binding. The ligand binding leads to phosphorylation of receptor-bound JAKs and the intracellular domains of the receptors. (2) Monomers of STAT associate with the phosphorylated tyrosines of the receptor and also get phosphorylated by JAKs. (3) Phosphorylated STATs form dimers and (4) translocate to the nucleus (5), where they bind to specific DNA motifs and change expression of the nearby genes.

1.3 Myeloproliferative neoplasms and their clinical classification

Since the early 20th century, diseases of the blood were diagnosed based on cell counts in blood smears[37]. When dynamic range and sensitivity of blood-counting technologies were increasing[38], overlapping features in a number of hematologic malignancies were discovered by William Dameshek. He reported that early-diagnosed patients with chronic myeloid leukemia (CML) often presented with signs of erythrocytosis and thrombocytosis, an elevation of red cells and platelets, respectively. Patients with diagnosis of PV frequently shared both characteristics and ET patients were diagnosed by pronounced thrombocytosis. Based on the observation of the overlapping phenotypes, William Dameshek proposed to summarize four diseases in a category, that he called myeloproliferative syndromes (MPS): CML, PV, ET and erythroleukemia (AEL or AML-M6)[39]. Since 1951, the name of the classification was modified from syndromes to myeloproliferative disorders (MPD) and changed into myeloproliferative neoplasms (MPN), but it is still recognized and maintained as an entity by the World Health Organization (WHO). Substantial refinements have been made over time to adapt diagnostic requirements to the current knowledge. Accordingly, the 2016 revision of the WHO classification[1] summarizes seven MPN subcategories: Four so-called classical MPN phenotypes, CML, ET, PV and PMF, and two phenotypes, which are diagnosed by excluding other hematological malignancies[40,41]. These two phenotypes are chronic neutrophilic leukemia (CNL) and chronic eosinophilic leukemia (not otherwise specified, CEL-NOS). The seventh phenotype, MPN unclassifiable (MPN-U), represents a group of patients which frequently shows features of ET, PV and PMF[42] but do not show the minimum criteria for or any of the six other MPN diagnoses.

In MPN diagnosis, molecular markers play an important role. The presence of a chromosomal translocation at position t(9;22)(q34;q11), the Philadelphia (Ph) chromosome, separates between CML and the Ph-negative MPN, ET, PV and PMF[1]. These three MPN phenotypes are distinguished by somatic mutations, blood counts and bone marrow biopsies. Diagnosis of PV requires a somatic mutation in *JAK2*, elevated red cell mass and hypercellularity in the bone marrow with trilineage growth. ET is diagnosed upon elevated platelets and proliferation of megakaryocytes in the bone marrow. Diagnosis of PMF requires megakaryocytic proliferation and atypia accompanied by reticulin and/or collagen fibrosis. The grade of fibrosis, megakaryocyte morphology and the potential presence of leukoerythroblastosis differ between the prefibrotic stage and the overt stage of myelofibrosis. In addition to clinical markers, the diagnosis of ET and PMF requires the presence of a somatic mutation in *JAK2*, *MPL* or *CALR* (Table 2).

	ET	PV	PMF
1st major criterion	Platelet count >450x10^9/L	Hb >165 g/L (males) Hb >160 g/L (females) Or HCT >49 % (males) HCT >48 % (females) Or increased red cell mass	Megakaryocytic proliferation and atypia reticulin fibrosis ≥grade 1, increased age-adjusted bone marrow cellularity and granulocytic proliferation (prePMF) or reticulin and/or collagen fibrosis grade 2 or 3 (overt PMF)
2nd major criterion	Bone marrow biopsy shows proliferation mainly of the megakaryocytic lineage with increased number of enlarged, mature megakaryocytes with hyperlobulated nuclei, no left shift in granulopoiesis or erythropoiesis, rarely fibrosis grade 1	Hypercellular bone marrow with trilineage growth with pleomorphic mature megakaryocytes	Excluding criteria for BCL-ABL+ CML, PV, PMF, MDS or other myeloid neoplasms
3rd major criterion	Presence of mutation in *JAK2*, *CALR* or *MPL*	Presence of *JAK2-V617F* or *JAK2-exon12* mutation	Presence of mutation in JAK2, *CALR* or *MPL*, or another clonal marker
4th major criterion	Excluding criteria for BCL-ABL+ CML, PV, PMF, MDS or other myeloid neoplasms		
Minor criteria	Presence of a clonal marker or reactive thrombocytosis excluded	Subnormal serum erythropoietin level	Anemia not attributed to a comorbid condition, or leukocytosis <11 x 10^9/L, or palpable splenomegaly, or LDH increased to above upper normal limit of institutional reference range, or Leukoerythroblastosis (overt PMF only)
Diagnosis requires	All four major criteria, or the first major criteria 1,2,4 and the minor criterion	All three major criteria or major criteria 1 & 2 and the minor criterion	All three major criteria and at least one minor criterion

Table 2 WHO criteria for the diagnosis of ET, PV and PMF

The data displayed in this table is from reference (1)

A study of the Swedish cancer registry with >9'000 MPN patients has shown that the survival differs between ET, PV and PMF. The relative survival ratio (RSR) of ET and PV patient for 5-year disease duration was comparable, 0.95 and 0.91, respectively. The RSR of PMF patients was significantly reduced for the same period of time. The longer the observation time, the more the survival of MPN patients differed. The RSR of 15 years after diagnosis was 0.66 for patients diagnosed for ET, 0.56 for PV patients and 0.13 for patients with PMF. The RSR of patients with MPN-U is increased compared to PMF but reduced compared to ET or PV[43]. The overall survival of MPN patients with ET, PV, PMF or MPN-U phenotype improved over the last 40 years due to improved prevention of complications such as thromboembolic events[43].

In long-term follow-up, patients with ET and PV may progress into post-ET or post-PV myelofibrosis (MF). In addition, ET, PV and PMF can transform to acute myeloid leukemia (AML) (Figure 3). Transformation of ET into PV has been described in 1-5 % of cases and is lower, when WHO proposed criteria are strictly followed for MPN diagnosis[44]. Transformation from ET or PV to post-ET/post-PV MF requires an increase of fibrosis to grade 2 or 3 along with additional criteria, such as: anemia, loss of need for phlebotomies, increasing splenomegaly or circulating blasts. The frequency of progression into myelofibrosis (MF) is reported to be slightly higher (5-14 %) for PV patients compared to ET patients (5-9 %)[45]. Frequency of transformation from ET to post-ET MF is supposedly lower since prefibrotic MF (preMF) is defined as an own category by WHO criteria. Patients with preMF show a significantly higher frequency (17 %) of transformation into overt MF than ET patients (5 %)[45]. All three classical MPN may progress into AML[46]. In average 5 % of ET and PV patients transform to AML. The frequency of progression to AML is lower for PV patients maintained on phlebotomies (<2 %), compared to PV patients treated with cytoreductive drugs/chemotherapy (up to 16 %). PMF patients have a risk of 16 % to progress to AML during 15 years after MPN diagnosis[47].

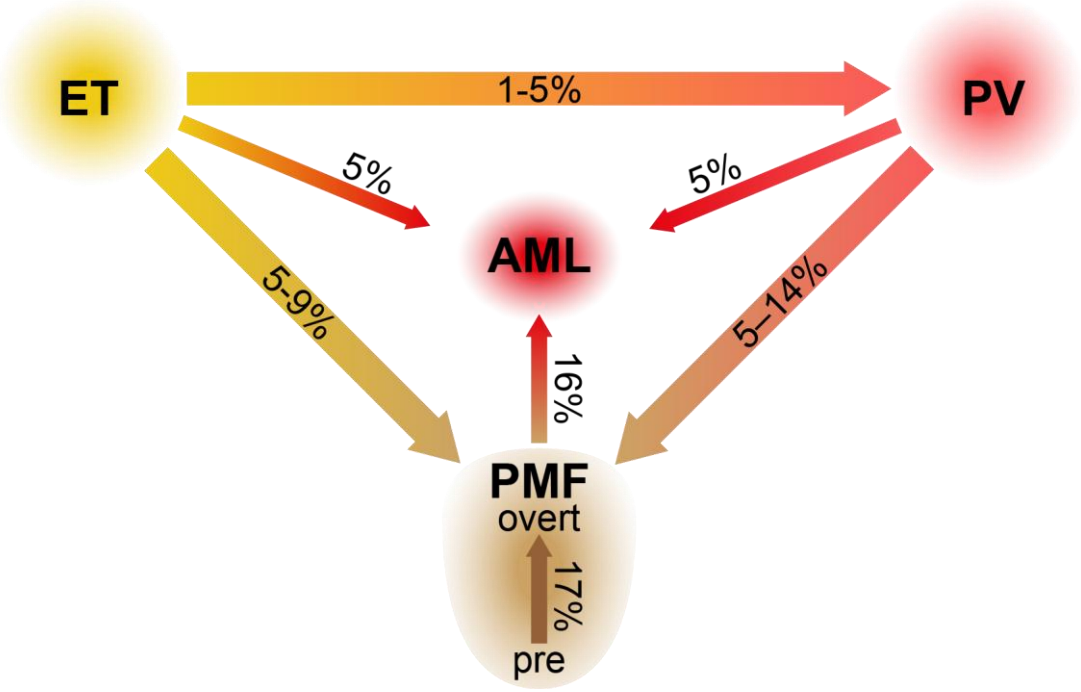


Figure 3 Transformation and progression in MPN

Adapted from data available in [1,44–47]. Values show probabilities for phenotypic transformations within 15 years of follow-up.

1.4 Molecular characteristics of ET, PV and PMF

The diagnosis of ET, PV or PMF, relies on clinical features and the presence of a clonal genetic marker (Table 2, 3rd major criterion). In contrast to CML where the Philadelphia chromosome is the sole genetic requirement for diagnosis, the classical Ph-negative MPN phenotypes share a set of genetic markers in the genes JAK2, MPL and CALR (Table 3)[1]. The *JAK2-V617F* mutation, discovered in 2005 by four labs[31–34], is a single nucleotide point mutation in exon 14 of the *JAK2* gene and is found in ET, PV and PMF patients. The prevalence of varies among the Ph-negative MPN is depending on the disease phenotype: nearly all PV patients (90%) carry this mutation, whereas only 50-60% of the patients diagnosed with ET or PMF are *JAK2-V617F* positive[31–34]. JAK2 exon 12 mutations are exclusively found in PV patients. Another gene recurrently mutated in Ph-negative MPN is the thrombopoietin receptor *MPL*. *MPL* mutations changing the codon of the tryptophan at position 515, are found in about 5% of ET and PMF patients, but are absent in PV patients[29]. The third gene recurrently mutated in ET and PMF is Calreticulin (*CALR*). *CALR* is affected by insertion or deletion mutations, which lead to a frame shift in exon 9. *CALR* mutations are absent in PV patients[35,36]. MPN patients who do not carry any of the previously described mutations in the three genes *JAK2*, *MPL* or *CALR*, are termed “triple negative” patients and represent 5-10% of ET, PV or PMF patients.

Mutation	ET	PV	PMF
JAK2-V617F	50-60%	95%	50-60%
JAK2-exon12	0%	4%	0%
MPL-W515L/K/A	5%	0%	5%
CALR-exon9	30-40%	0%	30-40%
None	5-10%	0-2%	5-10%

Table 3 Genetic markers of MPN

Combined data from references [29,31-34]

1.4.1 Mutations of the MPL receptor in MPN

The thrombopoietin receptor TpoR known as myeloproliferative leukemia protein MPL, is a receptor, which specifically binds the cytokine thrombopoietin (TPO) and signals through the JAK/STAT pathway (Figure 4A). TPO is produced by the liver and specifically binds to the MPL receptor. MPL is expressed by hematopoietic stem and progenitor cells, megakaryocytes and platelets. Upon binding TPO, MPL activates JAK/STAT, MAPK and PI3K-AKT1-mTOR pathways resulting in survival and proliferation. Further, TPO binding induces internalization and degradation of the ligand-receptor complex. The concentration of unbound TPO is therefore regulated by uptake and degradation by MPL expressing cells, mostly platelets and megakaryocytes (Table 1). This mechanism results in an inverse correlation of the concentration of unbound THPO and the number of platelets under homeostasis (Reviewed in [21]).

Several mutations in *MPL* have been found, mostly in MPN, which activate the receptor even in absence of its ligand. They affect the second amino acid of the intracellular domain, tryptophan at position 515[29]. Mutations at this position result in a conformational change of the intracellular domain, which brings associated JAK2 enzymes in spatial proximity. This steric conformation mimics a ligand-bound state and leads to constant phosphorylation and activation of JAK2, MPL and STAT proteins (Figure 4C). Soon after the discovery of the MPL W515L mutation, mouse models have shown the ability of the mutation to induce the MPN phenotypes ET and PMF[29,48].

Other mutations of *MPL* (K39N, P106L) are found in its extracellular domain and the transmembrane domain (S505N, Figure 4B). Recent studies have shown that the P106L mutation in the extracellular domain results in a cell type specific trafficking defect. While this mutant receptor is expressed on the surface of progenitor cells, its surface expression is blocked in adult megakaryocytes[49]. This leads to formation of platelets, which lack the MPL receptor on their surface. Therefore, these platelets are unable to efficiently bind and internalize TPO. The lack of TPO clearance in the blood results in a high concentration of unbound TPO and stimulation of progenitors towards platelet production without activating the receptor[50]. The MPL-S505N mutation induces a conformational change in the transmembrane domain of the receptor, which reduces the distance of associated JAK2 enzymes and results in activation of MPL and the downstream signaling pathways[51]. Both mutations, P106L and S505N, were found in families and their frequency seem to depend on the ethnical background of the studied population[51–54].

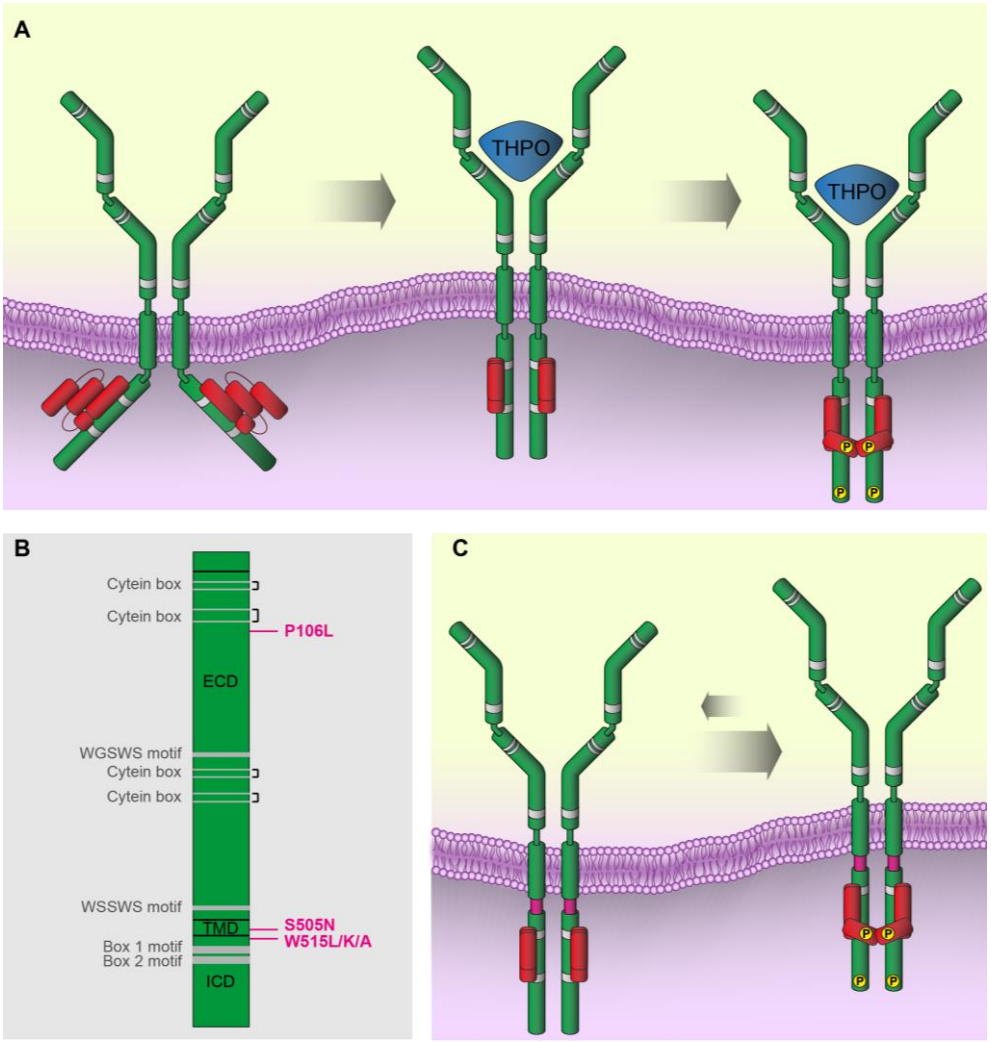


Figure 4 Activation of JAK-STAT signaling by TPO

(A) MPL activation under normal conditions: Pre-formed dimers undergo a change in conformation after binding TPO. Then, the JAK2 kinases get closer to each other so they can phosphorylate each other and the intracellular domains of the MPL receptor. (B) Schematic drawing of the MPL protein: black, names of structural domains; grey, name of specific motifs; pink, common mutations in MPL. (C) Ligand-independent self-activation of MPL W515. The conformational change is induced by the MPL-W515L mutation in absence of the ligand. Receptor bound JAK2 enzymes can activate each other and phosphorylate the receptor for downstream signaling. ECD: extra cellular domain, TMD: transmembrane domain, ICD: intracellular domain.

1.4.2 Activating JAK2 mutations

JAK2 belongs to the family of Janus kinases and specifically binds to cytokine receptors, which are forming dimers in absence of their ligand. Upon ligand binding, the receptor dimers undergo a conformational change, which brings the intracellular domains and the associated JAK2 enzymes in close proximity. This proximity is required for the JAK enzymes to trans-phosphorylate each other and the intracellular domains of the cytokine receptors[9]. MPL and EPOR, but also granulocyte colony-stimulating factor receptor (G-CSF-R) and leptin receptor (LEP-R), follow this mode of action. JAK2 also binds to cytokine receptors, such as the interferon-gamma receptor (IFNGR), which form heterodimers. When bound to these receptors, JAK2 can cooperate with other JAK family members. Upon ligand binding, the monomers receptors associate and the JAKs are trans-activating each other in homo- or heterodimers[55,56]. This transactivation of JAKs enables to translate ligand binding into signaling at enzymatically inactive cytokine receptors.

Structurally, JAK family members consist of a FERM domain, a SH2 domain and two kinase domains at the c-terminus (Figure 5A). The FERM and SH2 domains are required to associate with cytokine receptors at box1 and box 2 motifs[57]. The kinase domain closer to the c-terminus harbors the major catalytic activity for the phosphorylation of targets like STAT. The kinase domain next to the SH2 domain is only 10% as active as the c-terminal kinase domain and therefore is called the pseudokinase domain[58]. The activity of the pseudokinase domain is required for the auto-phosphorylation of serine 523 and tyrosine 570[59,60]. The phosphorylation of these two residues and close contract of the two kinase domains result in reduction of the JAK2 kinase activity (Figure 5, [61]).

Recent studies show that the interface, which generates this auto-inhibitory interaction within a JAK2 molecule is frequently disrupted by somatic mutations[62]. These mutations are the *JAK2-V617F* mutation in exon 14, which is found in all phenotypes of sporadic and familial MPN[31–34], the JAK2-R683G/S mutations in exon 15 of *JAK2*, which were found in Down syndrome-related acute lymphoblastic leukemia (DS-ALL), and the mutations in exon 12 (reviewed in[63]), which are solely found in PV patients. The disruption of the auto-inhibitory interface by these mutations is thought to enable JAK2 kinase activity. Once activated, JAK2 phosphorylates the intracellular domain of receptors even in the absence of their cognate ligand[14].

The *JAK2-V617F* and one of the most common *JAK2-exon12* mutations, JAK2-N542-E543del, have been introduced into mouse models to prove the MPN initiating potential of these JAK2 alterations[64,65]. The first *JAK2-V617F* mouse models were designed by our lab and are based on a human *JAK2-V617F* transgene. A comparative analysis showed that the MPN phenotype

in mice is depending on the expression of *JAK2-V617F* in relation to wild type JAK2. The mice represented an ET phenotype with increased platelet and neutrophil numbers, when *JAK2-V617F* expression was lower than wild type JAK2. Excessive expression of *JAK2-V617F* in relation to wild type JAK2 resulted in a trilineage PV phenotype with thrombocytosis, neutrophilia and increased hemoglobin[64]. Meanwhile, additional mouse models were generated and confirmed the dependence of MPN phenotype on *JAK2-V617F* expression in mice[66]. In the first *JAK2-exon12* mouse model, recently introduced by our group, the expression of mutant JAK2 results in elevated numbers of erythroid progenitors and precursors accompanied with normal counts of platelets and leukocytes. Furthermore, the mice showed increased STAT3 phosphorylation and altered expression of transferrin receptor-1 and erythroferrone, both involved in iron metabolism. Therefore, the *JAK2-exon12* mouse model resembles a PV phenotype similar to the observations in PV *JAK2-exon12* patients[65].

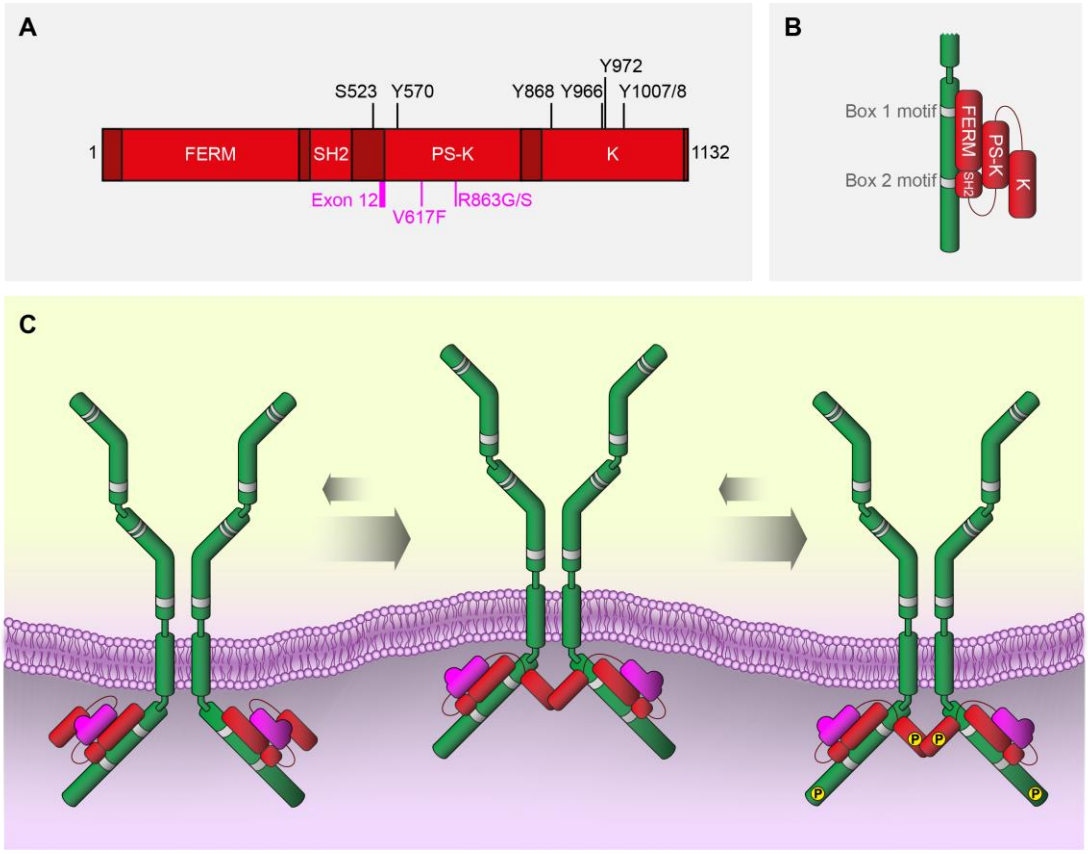


Figure 5 Mutant JAK2 induces cytokine signaling

(A) Schematic drawing of JAK2 kinase. Structural domains in light red: FERM: 4.1, ezrin, radixin, moesin domain, SH2: Src homology 2 domain, PS-K: pseudo-kinase domain, K: kinase domain. Important phosphorylation residues are shown on top of JAK2, most common mutations are shown below. (B) The JAK2 domains FERM and SH2 are binding intracellular domains of cytokine receptors (in green) at specific box motifs. In steady state, the kinase domain is in close contact with the pseudokinase domain and the SH2 linker region. (C) Mutant JAK2 activates cytokine receptors (here: MPL) in absence of their ligand. The above-depicted mutations induce a conformational change of the inhibitory interface (depicted as bulky pink PS-K). The kinase domain, as not associating with PS-K, can get close to the kinase domain of the second JAK enzyme. When the kinase domains are in close proximity, the JAKs trans-phosphorylate each other and the intracellular domains of the receptor and activate downstream signaling pathways.

1.4.3 CALR mutations in MPN

The third gene commonly mutated in MPN is Calreticulin (CALR)[35,36]. CALR protein is located in the endoplasmic reticulum (ER), where it supports the folding of glycoproteins as a chaperone. CALR is also found at the cell surface, however its function in this compartment is less clear[67]. It consists of three domains: the N-terminal domain, the P domain or proline rich region in the center and the C-terminal domain[68,69]. The N-terminal region is a protein-binding domain, which associates with proteins like alpha-integrin, protein disulphide-isomerase (PDI) and other ER proteins. The P-domain interacts with PDI and inherits the chaperone function of CALR and negatively regulates the affinity of the N-terminal domain to its targets. The c-terminal domain mainly consists of negatively charged amino acids, which enable a single CALR protein to bind up to 25 mol of Ca^{2+}. This Ca^{2+} binding capacity regulates the concentration of unbound calcium in the ER and is required to modulate its own function as a chaperone (Figure 6A). The activity of both, N- and P-domain, depend on the amount of Ca^{2+} bound to the C-domain[68,69].

Mutations of *CALR* were found in MPN patients who are negative for mutations in *JAK2* or *MPL*. These mutations, all located in the C-terminal domain, are a diverse group of insertions or deletions. The specific location and length of the deletions and insertions vary within the exon 9 of *CALR*, but nearly all of them result in a -1 frameshift[36]. The two most common mutations are a 52 base pair deletion (CALR type 1 mutation) and a five base pair insertion (CALR type 2 mutation) in exon 9 of *CALR*. In contrast to frameshift mutations in other genes, this -1 frameshift in *CALR* does not immediately lead to a premature stop codon. Instead, from the point of mutation on, the expression of the mutant *CALR* gene results in a new sequence of amino acids. This new sequence represents an accumulation of positively charged amino acids, whereas in wild type CALR protein the majority of amino acids is negatively charged[36].

In the case of mutant CALR protein, the new amino acid sequence of the c-terminal domain seems to reduce the auto-inhibitory function of the p-domain (Figure 6B), resulting in a stronger binding to and constitutive activation of MPL in the absence of its ligand (Figure 6C). This observation was specifically made for the interaction of CALR and MPL, but not for the closely related EPOR receptor, which might explain why mutations of CALR are only found in ET and PMF cases of MPN[70,71]. Recently, a retroviral mouse models has confirmed that expression of the type 1 and type 2 *CALR* mutations result in an ET phenotype, which includes amplification of the megakaryocyte lineage and elevated platelet counts. Mice expressing the CALR type 1 mutation have also been shown to frequently progress to myelofibrosis phenotype including features of anemia, splenomegaly and BM hypocellularity[72].

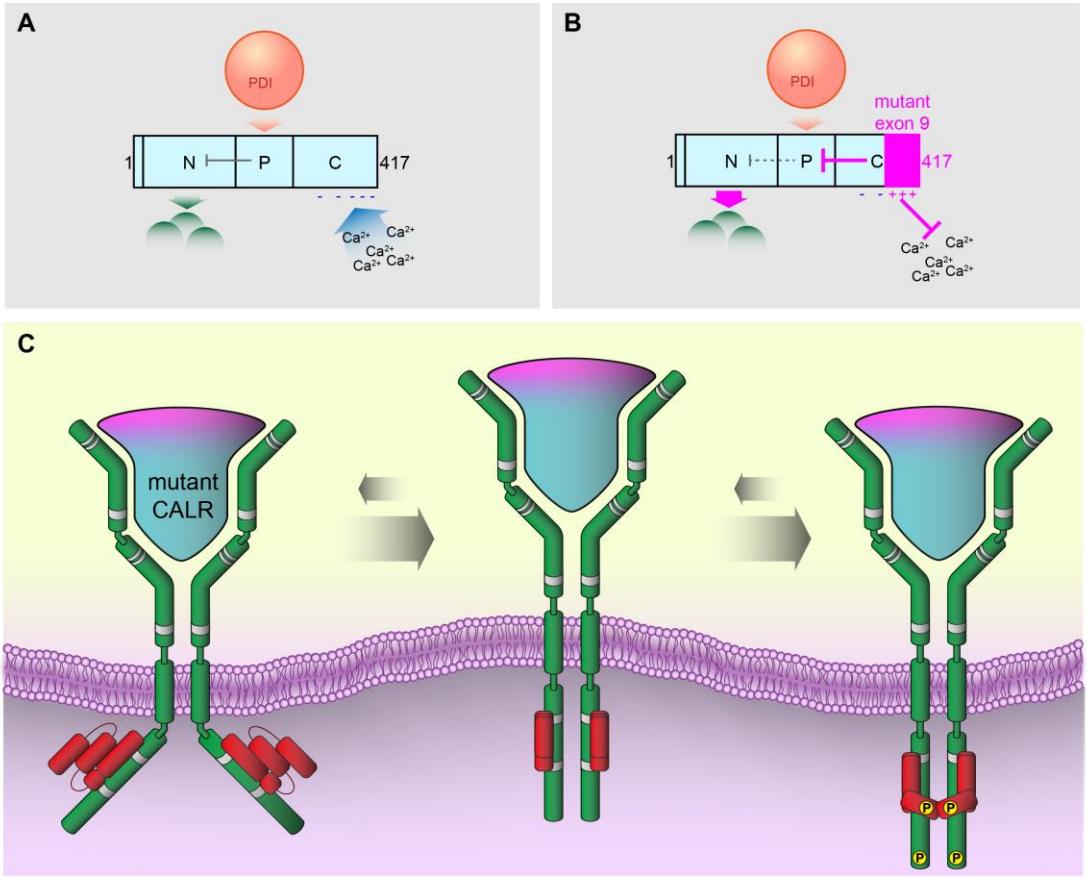


Figure 6 Activation of MPL by mutant CALR

(A) Schematic drawing of wild type *CALR* gene structure. N-domain binds to target proteins (green circles). P-domain recruits partners like PDI (orange sphere) and regulates binding affinity of the N-domain to the target proteins. The C-domain binds Ca2+ ions with its negatively charged amino acids. (B) Schematic drawing of mutant *CALR*. The new sequence in exon 9 (pink square) consists of numerous positively charged amino acids. Ca2+ binding is inhibited and the reducing effect of the P-domain on the affinity of the N-domain to target proteins is hampered (pink T). As a result, the N-domain binds target proteins with higher affinity (pink arrow). (C) Mutant CALR binds MPL receptor and mimics ligand binding in absence of TPO. The receptor associated JAK2 molecules reach critical proximity for trans-phosphorylation and activate downstream signaling by phosphorylation of the MPL intracellular domain.

1.4.4 Additional somatic mutations in hematologic cancers

The recurrent mutations in *MPL*, *JAK2* and *CALR* are recognized as driver mutations, as they were shown to induce MPN in several mouse models[73]. In addition to driver mutations, other somatic mutations have been found, which do not evoke a MPN phenotype on their own. These mutations are referred to as passenger mutations as thought to result from a genome-destabilizing function of the driver mutations[73]. Several of these mutations are also found in other myeloid malignancies including AML, CML or myelodysplastic syndrome (MDS). Most frequently, *DNMT3A*[74,75], *TET2*[76,77], *IDH1/2*[78,79], *EZH2*[80] and *ASXL1*[81,82] are affected. Interestingly, all of these genes are involved in DNA methylation and histone modification (Figure 7)

DNMT3A, TET2, IDH1 and IDH2 regulate the DNA methylation states of GC rich DNA regions, so called CpG islands. DNA (cytosine-5)-methyltransferase 3A (DNMT3A) methylates DNA by transferring a methyl group to cytosine at CpG islands[83]. TET2 is a member of the ten-eleven translocation (TET) family of methylcytosine dioxygenases, which catalyze the conversion of methylcytosine to 5-hydroxymethylcytosine. This catalysis requires α-ketoglutarate, which is generated by isocitrate dehydrogenase (IDH) enzymes. Therefore, mutations in these genes lead to a global change of DNA methylation[84]. *DNMT3A* mutations mostly result in loss of function of the enzyme and are associated with focal DNA hypomethylation. Loss-of-function mutations in *TET2*, *IDH1* and *IDH2* result in global hypermethylation of DNA[85,86]. DNA methylation plays an important role in regulating gene expression, as it blocks the binding of activators of transcription with DNA. Furthermore, methylated DNA is associated with repressive chromatin structure, called heterochromatin[87].

Heterochromatin describes DNA, which is wrapped around histone proteins. This state of tightly packed and transcriptionally inactive DNA is regulated by enzymes, which modify residues of the histone proteins. In hematologic malignancies, mutations in *EZH2*, and later also *SUZ12*, were found frequently. Both proteins belong to the core units of the polycomb repressive complex 2 (PRC2). This complex catalyzes trimethylation of histone 3 lysine 27 (H3K27me3) and silences expression of target genes, by stabilizing the heterochromatin. Mutations in *EZH2*, the enzymatic unit of the PRC2 complex, have been shown to reduce methylation levels of H3K27 and induce expression of genes located in regions of heterochromatin[88].

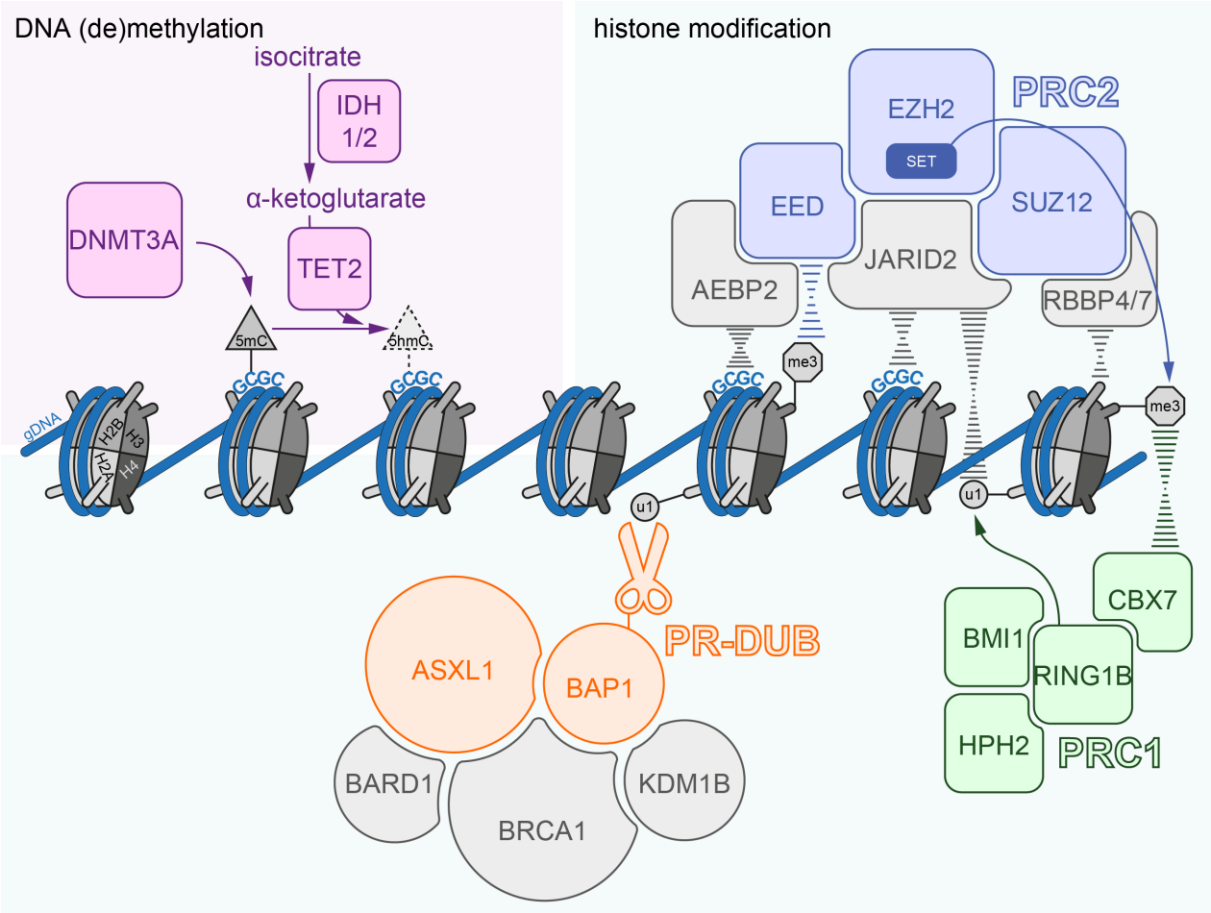


Figure 7 Pathways of DNA methylation and histone modification altered in hematologic malignancies

Enzymes involved in DNA methylation and demethylation are depicted in pink. Core proteins of the polycomb repressive deubiquitinase (PR-DUB) complex and the polycomb repressive complex 1 and 2 (PRC1 & PRC2) are highlighted with individual colors. Optional partners and adapter proteins of these complexes are rendered in grey. Arrows and scissors: enzymatic activity, GCGC: GC-rich DNA regions/CpG islands, 5mC: 5-methylcytosine, 5hmC: 5-hydroxymethylcytosine, me3: trimethylation at H3K27, u1: monoubiquitin at H2AK119.

ASXL1 is part of the polycomb repressive deubiquitinase (PR-DUB) complex, which erases monoubiquitin from histone 2A lysine 199 (H2AK119ub1,[89]). This ubiquitin mark is generated by RING1A/B, which is part of the polycomb repressive complex 1 (PRC1,[90]). H2AK119ub1 is recognized by the PRC2 adapter protein JARID and association of PRC2 with H2AK119ub1 induces generation of repressive H3K27me3 marks[91]. Recent studies show that ASXL1 truncations lead to increased activity of the PR-DUB complex[92] and reduction of global H2AK119ub1 and H3K29me3 levels[93]. In line with these findings, ASXL1 truncations elevate expression of HOXA5 and HOXA9, which are repressed by PRC2 complex activity.

Interestingly, a study showed that DNMT proteins directly interact with EZH2, which underlines the association of DNA methylation and histone methylation[94]. Mutations in *DNMT3A* and *EZH2* have been linked with reduced survival in AML[74,80] and alterations of *IDH1* were associated with progression to AML[78,95]. In MPN, *ASXL1* mutations have been recently associated with the PMF phenotype and reduced survival[96]. In summary, these findings were challenging the notion that passenger mutations are passive and do not impact the primary disease[73].

1.4.5 Sequencing technologies used in MPN research and diagnostics

The described alterations in *DNMT3A*, *TET2*, *IDH1/2*, *EZH2* and *ASXL1* were mainly found by using two different technologies, capillary sequencing and single nucleotide polymorphism (SNP) arrays. Capillary sequencing is based on chain termination technology and is also called "Sanger sequencing"[97]. This method allows screening for small sequence alterations like single base pair mismatches. Duplications and larger deletions were detected by SNP arrays, where DNA hybridizes to immobilized allele-specific oligonucleotide[98]. Therefore, the capillary sequencing was traditionally used for finding sequence alterations in the coding region of a single gene or single loci of multiple genes. The SNP arrays enabled a genome-wide detection of insertions and deletions, but only for known alleles. Even when these two technologies were combined[76,80], cohort analyses remained incomplete, as parallel analysis for multiple genes and their full-length sequence was still missing[99]. The development and commercialization of the so-called next-generation sequencing (NGS) solved these technological constraints as it introduced parallel sequencing of multiple genes with a so far unmatched sensitivity[100].

In this new NGS technology, DNA is fragmented and ligated to adapters. These adapters allow immobilizing individual DNA fragments on a chip (Illumina) or on beads (Ion torrent,[101]). Immobilized fragments then get amplified and analyzed individually. The strength of this technology lies in the amount (or length) of DNA, which can be sequenced, as millions of DNA fragments are analyzed in parallel. Additionally, these NGS technologies allow to a flexible adaption of sensitivity: the smaller the DNA region of interest, the more sensitive the method becomes in detection of mutations. This is important in the field of MPN, as somatic mutations may only affect a low number of blood cells. Therefore, sensitive methods are required to detect small clones in a majority of wild type cells. In the present study, we are using different NGS technologies to screen for somatic mutations, which modulate the MPN phenotype and have so far been undiscovered due to previous technical limitations.

1.5 Aim of the study

Ph-negative MPN are a group of hematologic diseases with three distinct clinical phenotypes. However, it is still unclear what defines ET, PV and PMF on a molecular level, as the disease-initiating clonal markers in *JAK2*, *CALR* and *MPL* are shared between these phenotypes.

Here, we take advantage of the new NGS technologies and analyze more than 100 genes in parallel to find secondary somatic mutations with impact on the phenotype of Ph-negative MPN patients. We also analyze the temporal order of acquisition of these mutations and compare our results between adult MPN patients and rare pediatric cases.

In the second half of the thesis, we analyze patients with a very low mutant JAK2 allele burden to find out if the MPN clone in these patients expands to a different blood cell lineage, which may explain the disease phenotype. Furthermore, we analyze the genetic alterations of a large virtual cohort to discover instructive effects of less frequently mutated genes on the MPN phenotype.

2 Methods

2.1 Study population and sampling intervals

The Basel cohort of sporadic MPN includes patients with diagnosis of ET, PV, PMF and MPN-U. The collection of clinical data and blood samples was performed at the study center in Basel, Switzerland and approved by the local Ethics Committees (Ethik Kommission Beider Basel). Written informed consent was obtained from all patients in accordance with the Declaration of Helsinki. The diagnosis of MPN was established according to the revised criteria of the World Health Organization 2008[(102)] and the update in 2016[(1)]. The clinical data of the Basel sporadic MPN cohort is summarized Table 4 (status: June 2016).

Blood samples were acquired in yearly interval, starting with the first sample at diagnosis. Hair follicles and buccal swabs were collected once per patient. Additional samplings might have been necessary for non-standard analysis, which were proposed based on the patient's individual results in cohort-wide experiments. Patients may refuse the donation of samples and reject from the study at any time. The Basel sporadic MPN cohort is an active cohort and therefore, newly diagnosed patients may join the cohort and fresh blood samples are acquired constantly.

	PV	ET	PMF	MPN-U
Number of patients	132	106	50	6
Female patients	63 (47 %)	71 (66 %)	16 (32 %)	4 (66 %)
Age at diagnosis (range), years	60 (11 - 87)	51 (4 - 83)	63 (21 - 85)	63 (36 - 85)
Follow-up time (range), months	34 (0 - 168)	21 (0 - 164)	21 (0 - 153)	9 (0 - 42)
Hemoglobin (g/L) median (range)	178 (119 - 229)	140 (13 - 182)	121 (12 - 171)	126 (126 - 126)
Platelets (10^9/L) median (range)	474 (90 - 1487)	898 (452 - 1983)	623 (16 - 1677)	1527 (793 - 2261)
Leukocytes (10^9/L) median (range)	11 (3 - 38)	8 (4 - 16)	9 (4 - 26)	13 (9 - 18)
Neutrophils (10^9/L) median (range)	8 (2 - 36)	5 (2 - 15)	6 (2 - 20)	6 (6 - 6)
Transformation to secondary MF	7 (5 %)	3 (3 %)	NA	0
Transformation to AML	8 (6 %)	4 (4 %)	4 (8 %)	0
Death	21 (16 %)	8 (8 %)	11 (22 %)	1 (17 %)

Table 4 Clinical characteristics of the MPN patients at diagnosis

2.2 Blood sample preparation

Blood samples are processed in a standard protocol to allow purification of several cellular subsets and lysates of comparable quality (Figure 8). A usual blood sample comprises five EDTA tubes (S-Monovette 7.5, Cat. No. 01.1601.100, Sarstedt AG & CoKG, Nümbrecht, Germany), with 7.5 mL of whole blood and one tube of 7.5 mL whole blood without anticoagulants ("native blood"). The tube containing coagulating blood is centrifuged at 2300 rcf for 10 minutes. The clear supernatant (serum) is frozen at -80 °C. One tube of whole blood containing citrate as anticoagulant is centrifuged at 2300 rcf for 10 minutes. Here the clear supernatant (platelet-poor plasma, PPP) and the interphase (buffy coat) are frozen at -80 °C. The four remaining tubes of whole blood containing citrate are centrifuged at 100 rcf for 10 minutes. The supernatant of all four tubes (platelet-rich plasma, PRP) is collected for subsequent sephadex column purification of platelets (Sepharose CL-2B, Cat. No. 17-0140-01, VWR international, Radnor, PA, USA). The purified platelets are frozen at -80 °C in western blot buffer (50 mM Tris pH 6.8, 2 % SDS, 1 mg/mL bromphenolblue, 10 % glycerol, 5 % 2-mercaptoethanol) and in TriFast (Trifast Peq Gold FL, Cat. No. 30-2110, Axon Lab AG, Reichenbach an der Fils, Germany) for RNA purification. The remaining red part of the blood from all four tubes is collected and diluted 1:2 in PBS (PBS, pH 7.2, Cat No. 20012068, LuBioscience GmbH, Lucern, Switzerland). This dilution is used to overlay Ficoll (Ficoll Lymphprep, Cat. No. 1114547, Axon Lab AG, Reichenbach an der Fils, Germany). After centrifugation at 100 rcf for 30 mintues, five layers are formed: plasma, peripheral blood derived mononuclear cells (PBMCs), Ficoll Plaque, granulocytes and erythrocytes. PBMCs and granulocytes are isolated separately. Both fractions are red cell lysed (0.15 M NH_4Cl, 0.01 M $KHCO_3$, 0.05 M EDTA, pH8) at room temperature for 10 minutes. The main fraction of the PBMCs are frozen in FBS (Fetal Bovine Serum, Cat. No. F7524-500ml, Sigma Aldrich, St. Louis, MI, USA) including 10 % DMSO (Dimethyl sulfoxide, Cat No. 276855-100ML, Sigma-Aldrich, St. Louis, MI, USA) in liquid nitrogen. A small fraction of cells is frozen at -80 °C in PBS for DNA preparation. The granulocytes are frozen at -80 °C in PBS for DNA preparation and in TriFast (Trifast Peq Gold FL, Cat. No. 30-2110, Axon Lab AG, Reichenbach an der Fils, Germany) for RNA preparation. The fraction of erythrocytes is usually discarded. Buccal swabs are frozen at -80 °C upon arrival. Hair follicles are stored at room temperature until preparation of DNA. For patients with a very low allele burden, a special protocol was developed to isolate reticulocytes for further analysis from the otherwise discarded erythrocyte fraction. The protocol and results of this study are reported in chapter 3.3.

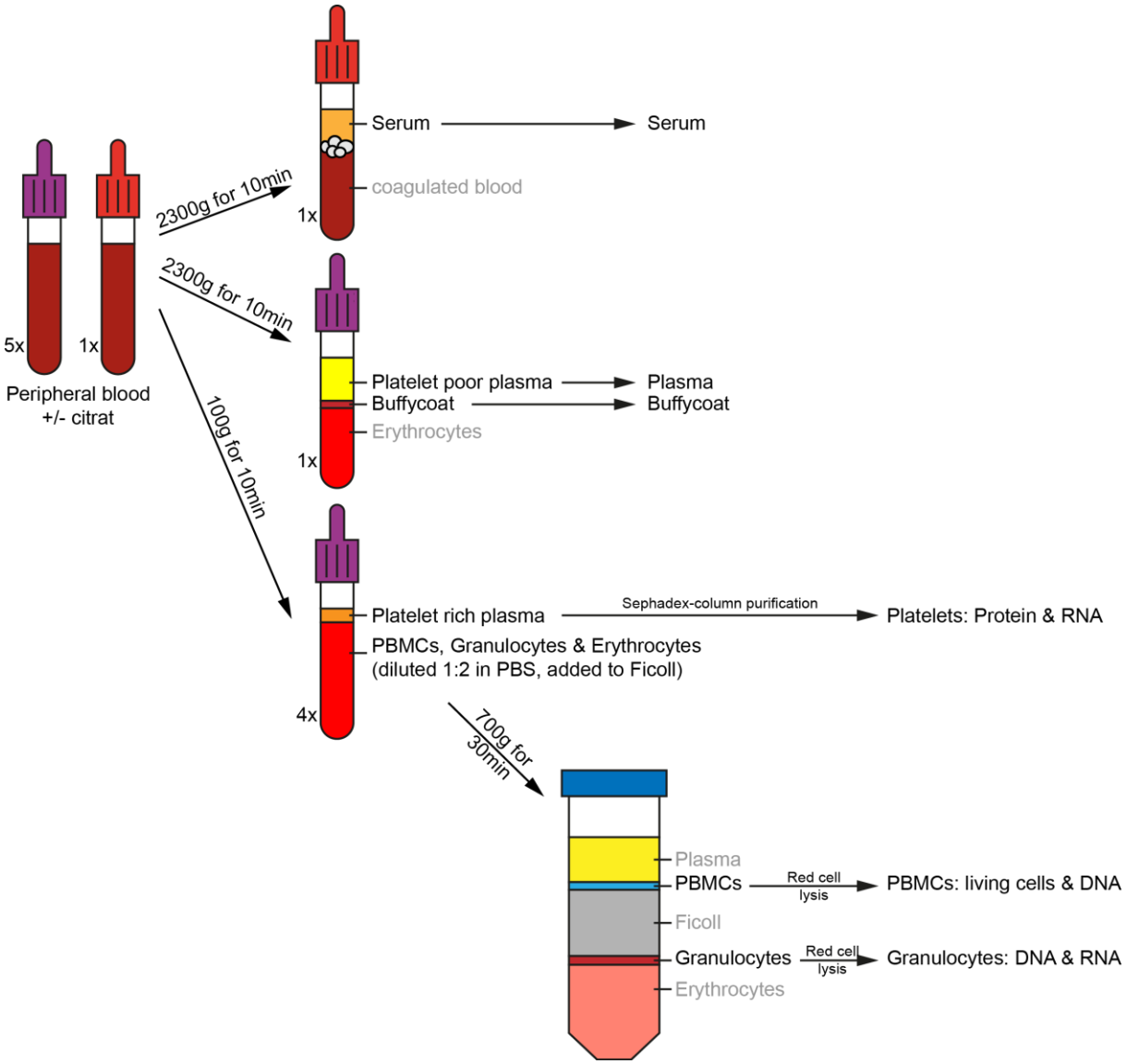


Figure 8 Standard blood sample work-up

2.2.1 Preparation of nucleic acids from different tissues and cell types

2.2.1.1 DNA preparation from granulocytes and buccal swabs

DNA from granulocytes and buccal swabs was prepared using the QIAamp DNA Mini Kit (Cat No. 51306, Qiagen Instruments AG, Hilden, Germany).

2.2.1.2 Preparation of DNA from hair follicles

Hair contains dead keratinocytes, which originate from living keratinocytes located in the follicle of each hair. To generate the hair, living keratinocytes differentiate by a process called cornification, which includes crosslinking of keratin and keratin-associated proteins and the degeneration of the cell nucleus with the contained DNA[(103)]. Therefore, only the follicle contains living cells with high molecular weight DNA, which is required for NGS applications and other genetic analyses. For this reason, single hairs need to be cut as close to the root as possible to prepare a sample. The roots from one individual were collected in a 1.5 mL tube and have been stored at room temperature until lysis (Lysis Buffer T, Cat No. 823, PerkinElmer chemagen Technologie GmbH, Baesweiler, Germany) was started for all samples simultaneously. From lysis to elution, the chemagic BTS Kit special (Cat No. 1046, PerkinElmer chemagen Technologie GmbH, Baesweiler, Germany) was applied to isolate genomic DNA from the hair follicles.

2.2.1.3 Preparation of RNA from granulocytes & platelets

Granulocytes and platelets were isolated from peripheral blood and stored in TriFast as described in chapter 2.2. The preparation of RNA is the same for granulocytes and platelets and follows the protocol of the distributor “PEQLAB_v0815_E”. In brief, 1 mL TriFast was vigorously mixed with 0.2 mL of chloroform (Chloroform, Cat. No. 32211-1L, Sigma Aldrich, St. Louis, MI, USA). After 15 minutes incubation on ice, the samples were centrifuged for 15 minutes at 4 °C and 12’000 rcf. The aqueous phase was transferred to a pre-cooled tube containing 0.5 mL isopropanol (2-Propanol, Cat. No. 59300-1L, Sigma Aldrich, St. Louis, MI, USA) including 1 µL of glycogen (UltraPure Glycogen, Cat. No. 10814-010, Thermo Fisher Scientific Inc, Waltham, MA, USA). After mixing, the samples were incubated 15 minutes on ice and centrifuged for 15 minutes at 4 °C and 12’000 rcf. Then, the supernatant was discarded and 1 mL of 75 % ethanol (Ethanol, Cat. No. 1.00983.1000, VWR international, Radnor, PA, USA) was added to wash the pellet. After brief vortexing, the samples were centrifuged for 15 minutes at 4 °C and 12’000 rcf. All ethanol was removed and the pellet was resolved in 30 µL of RNAse free water.

2.2.1.4 Preparation of RNA from limited number of cells

When RNA was prepared from samples with small cell numbers, the Acturus picopure kit (PicoPure RNA Isolation Kit, Cat. No. KIT0204, Thermo Fisher Scientific Inc, Waltham, MA, USA) was used according to the user manual.

2.3 Molecular analyses

2.3.1 Driver mutation analysis using DNA

Single nucleotide mutation *JAK2-V617F* was characterized by PCR amplification, followed by DNA sequencing. PCR amplification was performed with wild-type JAK2-specific forward primer 5'-GTTTCTTAGTGCATCTTTATTATGGCAGA-3' and reverse primers 5'-6Fam-AAATTACTCTCGTCTCCACAGAA-3' and 5'-6Fam-TTACTCTCGTCTCCACAGAC-3'. The mutation analysis of *CALR-exon9*, -52 bp deletion (type1) and +5 bp insertion (type2) was performed by allele specific PCR as previously reported[36]. The primer sequences were as follows: CALR-intr8-fam-fwd 5'-FAM-GGCAAGGCCCTGAGGTGT-3' and CALR-ex9-rev 5'-GGCCTCAGTCCAGCCCTG-3'. The PCR products were analyzed by fragment analysis with ABI3130xl Genetic Analyzer (Applied Biosystems Inc). The mutant allele burden was calculated by Peak $hight_{mut}$ / (Peak $hight_{mut}$ + Peak $hight_{wt}$) x 100 %.

2.3.2 Quantitative JAK2 analysis using RNA

For the analysis of JAK2 mutations on RNA, the SNaPshot™ Multiplex Kit (Snapshot Multiplex Kit, Cat. No. 4323159, Thermo Fisher Scientific Inc, Waltham, MA, USA) was used. The protocol includes an amplification of the mutated region by PCR, purification of the PCR product with AMPure XP Beads (AMPure XP, Cat No. A63881, Beckman Coulter Inc., Brea, CA, USA) and the generation of mutation specific signal by single base elongation of dedicated primers (Table 5).

Primer function	Primer ID	Primer sequence (5'-3')
Amplification of JAK2 exon 12 from DNA	2075_hJAK2-exon12-F	CAAAGTTCAATGAGTTGACCCC
	2076_hJAK2-exon12-R	TGCTAACATCTAACACAAGGTTGG
Amplification of JAK2 exon 14 from RNA	4336_hJAK2-ex13-fwd	GGCGTACGAAGAGAAGTAG
	4337_hJAK2-ex15-rev	GCCCATGCCAACTGTTTA
Amplification of JAK2 exon 12 from RNA	4416_hJAK2-exon11-fwd	ACTAAATGCTGTCCCCCAAA
	4417_hJAK2-exon13-rev	TACTTCTCTTCGTACGCCTT
Amplification of JAK2 exon 14 from DNA	4586_hJak2_intron13_fwd	AGAATTTTCTGAACTATTTATGG
	4587_hJak2_intron14_rev	ACCTAGCTGTGATCCTGAAACTG
Quantification of *JAK2-V617F*	4338_hJAK2-VF-SNaP-fwd	AAGCATTTGGTTTTAAATTATGGAGTATGT
Quantification of *JAK2-exon12* (P021)	4420_P021del_SNPsht_fwd	AAAGTCTGACAACAAATGGTGTTTCACAAAATCAGA
	4425_P021del_SNPsht_rev	AATCCTTAGGTAAGGCTTTCATTAAATATCAAATCT
Quantification of *JAK2-exon12* (P218)	4426_P216aaTT_SNPsht_rev	TAGGTAAATATCAAATCTTCATTTCTGATT

Table 5 Primers for SNaPshot analysis

2.3.3 Next-generation sequencing by Illumina technology

For the initial mutation screen by Illumina sequencing, 500 ng of genomic DNA was used for the sample preparation protocol. At the beginning, genomic DNA was fragmented by incubation with fragmentase (NEBNext® dsDNA Fragmentase, Cat No. M0348L, New England Biolabs, Ipswich, MA, USA) at 37 °C for 30 minutes. Since 2014, fragmentation was achieved by 5 minutes of Covaris sheering (Covaris E220, Covaris, Woburn, MA, USA). Subsequently, endrepair, adenylation and barcode ligation were performed with the NEXTflex™ Rapid DNA Sequencing Kit (Cat No. 5144-1, Bioo Scientific, Austin, TX, USA). Depending on sample size, NEXTflex™ DNA Barcodes – 48 or NEXTflex™ DNA Barcodes – 96 (Cat No. 514104 or 514105, Bioo Scientific, Austin, TX, USA) were used for indexing of individual samples. Intermediate cleanup steps were conducted using the Agencourt AMPure XP PCR Purification protocol (AMPure XP, Cat No. A63881, Beckman Coulter Inc., Brea, CA, USA).

For each sequencing run, 48 individually barcoded samples were pooled equimolarly prior to enrichment. The most recent version of the “SureSelectXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library” was used to enrich for genes of interest. The system consisted of a SureSelectXT custom panel (Cat No. 5190-xxx, Agilent Technologies Inc., Santa Clara, CA, USA) to enrich for the genes of interest and a Herculase II fusion DNA kit (Cat No. G00677, Agilent Technologies Inc., Santa Clara, CA, USA) to amplify the barcoded and enriched DNA fragments. To order SureSelectXT custom panels, the bait design was generated in SureDesign (https://earray.chem.agilent.com/suredesign/home.htm, last accessed on 24.02.2017). The gene set, which was used to sequence 200 MPN patients, included all exons of 104 genes listed in Table 6. Dynabeads® MyOne™ Streptavidin T1 (Cat No. 65601, Thermo Fisher Scientific Inc, Waltham, MA, USA) were used for intermediate clean-up steps of the target enrichment. The prepared libraries were sequenced on Illumina HiSeq 2500 (Illumina, San Diego, CA, USA). Analysis of the raw reads were performed by the most recent version of CLC genomics Workbench (Qiagen Instruments AG, Hilden, Germany).

AKT1	CBL	ETV6	HOXA9	L3MBTL	PIAS2	PIK3R1	SH2B3
AKT1S1	CEBPA	EVI1	IDH1	MPL	PIAS3	PIK3R2	SOCS1
AKT2	CREBBP	EZH2	IFI30	MYB	PIAS4	PIK3R3	SOCS3
AKT3	CUX1	FLT3	IKZF1	MYBL1	PIK3AP1	PIK3R4	SOSC2
AKTIP	DNMT3A	FOXP1	IL6	MYBL2	PIK3C2A	PIK3R5	STAT3
AML1	EGLN1	GATA1	IL6R	MYC	PIK3C2B	PIK3R6	STAT5
ARNT	EID1	GATA2	IRF4	MYCBP	PIK3C2G	PRMT5	TET2
ARNT2	EID2	GCSF	IRF8	NF1	PIK3C3	PTEN	TNR
ARNTL	EID3	GDF15	JAK2	NFE2	PIK3CA	PTPN11	TP53
ARNTL2	PAS1	GSN	JUNB	NPM1	PIK3CB	PTPRT	TPO
ASXL1	EPO	HIF1A	JUN-D	NRAS	PIK3CD	RBBP5	VHL
BCL2	EPOR	HIF3A	KIF17	P300	PIK3CG	SCF	WT1
BRAF	ERG	HINT1	KRAS	PIAS1	PIK3IP1	SGK2	ZFP36L1
CALR							

Table 6 List of the 104 genes sequenced by NGS

CALR was analyzed by allele-specific PCR

2.3.4 Validation of NGS by ION torrent technology

Variants found by Illumina sequencing were validated by ION torrent sequencing technology. For each variant, primers were generated using the primer design tool of CLC genomics Workbench (Qiagen Instruments AG, Hilden, Germany). Standard PCR method was used to produce amplicons of ~250 bp for each individual variation. Amplicons of multiple mutations were pooled for barcoding. Each so-called “pool” of amplicons was labeled with a different barcode. The decision, which variant was added to which pool, was dependent on the location of the variations (Figure 9). Recurrent variants and variants, which were located in close proximity, must not been blended in the same pool. Otherwise, recurrent mutations of different patients would not be identified in the sequencing analysis. Mutations in close proximity might influence each other’s allele frequency if their amplicons were overlapping. As amplicons are ~250 bp long, the minimal distance between two mutations was set to 500 bp to allow the combination of these mutations in the same pool.

The pools are barcoded using the Ion Xpress™ Fragment Library Kit (Ion Plus Fragment Library Kit, Cat. No. 4471252, Thermo Fisher Scientific Inc, Waltham, MA, USA). The sequencing was performed by Ion PGM (Cat No. 4462921, Thermo Fisher Scientific Inc, Waltham, MA, USA). The acquired sequencing data was analyzed by the Ion Reporter Software (Cat No. 4487118, Thermo Fisher Scientific Inc, Waltham, MA, USA).

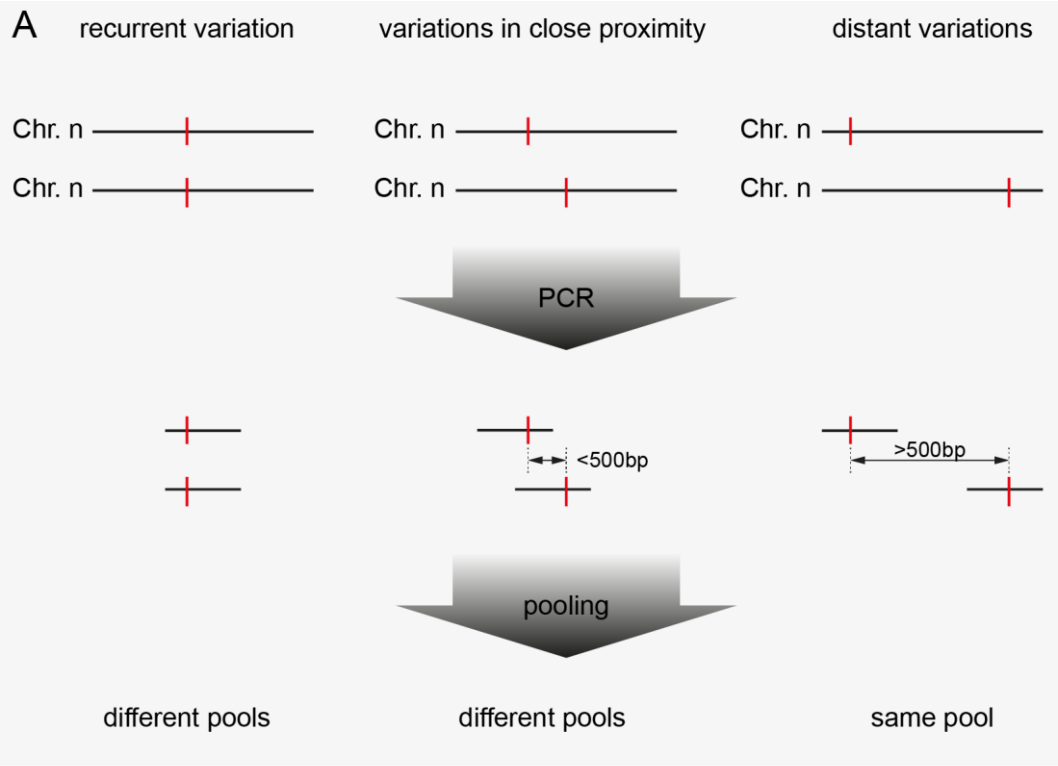


Figure 9 Pooling strategy for validation of mutations

(A) Graphical representation of pooling strategy (B) Formula for automatic assignment of the pool for the individual mutation. Column A requires sorted the numeric chromosomal position; Column B is the output column for the pool the amplicon has to be sorted to.

2.3.5 Analysis for clonal architectures

Patients with more than one validated somatic mutation have been qualified for the analysis of their clonal architecture. The analysis required progenitor cells, which were enriched in the PBMC fraction isolated from peripheral blood (see 2.2).

PBMCs of these patients were thawed from liquid nitrogen and recovered in IMDM (Cat No. 21980032, LuBioscience GmbH, Lucern, Switzerland) including 10 % FBS. After centrifugation at 200 rcf for 5 minutes, the PBMC pellet was reconstituted in 10 mL pure IMDM. The viable cells were counted and 300'000 cells were added to 3 mL of non-selective methylcellulose medium (MethoCult™ H4434 Classic, Cat. No. 04444, Stemcell Technologies, Vancouver, Canada). The cells were vortexed to allow homogenous distribution in the MethoCult. When air bubbles disappeared, the cell/MethoCult mix was placed in three 3 cm dishes (Easy grip tissue culture dish, Cat. No. 353001, Thermo Fisher Scientific Inc, Waltham, MA, USA), 1 mL each. The dishes were incubated at 37 °C and 5 % CO_2 for 14-16 days.

The grown colonies were harvested individually and transferred into 5 % chelex (Chelex 100 Resin, Cat. No. 143-2832, Bio-Rad Laboratories, Hercules, CA, USA) in water containing 0.1 % Tween (Tween 20, Cat. No. 170-6531, Bio-Rad Laboratories, Hercules, CA, USA). The mix was incubated 15 min at 56 °C followed by heat inactivation at 99 °C for 8 minutes. The reaction could be placed at -20 °C for long-time storage.

The chelex-cell mix was used as template for standard capillary sequencing. The sequencing reactions were performed in BigDye version 3.1 (BigDye Terminator v3.1 Cycle Sequencing Kit, Cat. No. 4337455, Thermo Fisher Scientific Inc, Waltham, MA, USA). The sequencing tracks were generated by the Applied Biosystems 3130xl Genetic Analyzer (Thermo Fisher Scientific Inc, Waltham, MA, USA) and analyzed with the CLC genomics Workbench (Qiagen Instruments AG, Hilden, Germany).

3 Results

3.1 Clonal evolution and clinical correlates of somatic mutations in MPN

3.1.1 Cohort definition, sequencing workflow & quality

To define the mutational profile of MPN patients, a patient cohort comprising 200 MPN patients was lined up for NGS analysis. From all patients, germline control material was available to validate if the detected mutations were somatic mutations of the blood system. The clinical data of these patients is summarized in Table 7.

Figure 10 shows the sequencing workflow and Figure 11 displays results from quality control measurements from the individual stages of this workflow. Genomic DNA from granulocytes of all 200 patients was screened in duplicates for mutations by targeted Next-generation sequencing by Illumina technology (Figure 10A). The duplicates of individual patients showed comparable average fold coverage (Figure 11A). In nearly all duplicates, 90 % of the exons were covered at least 20x. In the majority of samples, 90 % of the exons were covered 100x or more (Figure 11B). Based on their low coverage, samples from three patients (P011, P134 and P135) were excluded from further analysis. From 197 patients with high sequencing coverage, 12 patients (6.1 %) had no alterations at all (Figure 10A).

	PV	ET	PMF
Number of patients	94	69	34
Female patients	48 (51 %)	46 (67 %)	9 (26 %)
Age at diagnosis (range), years	58 (18 - 87)	51 (21 - 86)	61 (21 - 85)
Average time of follow-up, months	92	56	49
Hemoglobin (g/L) median (range)	181 (148 - 225)	141 (78 - 225)	126 (90 - 161)
Platelets (10^9/L) median (range)	554 (90 - 1487)	994 (452 - 1983)	635 (16 - 1677)
Leukocytes (10^9/L) median (range)	12 (4 - 39)	9 (5 - 17)	11 (5 - 27)
Neutrophils (10^9/L) median (range)	9 (2 - 36)	6 (3 - 16)	8 (3 - 21)
Transformation to secondary MF	4 (4 %)	1 (1 %)	NA
Transformation to AML	3 (3 %)	2 (3 %)	2 (6 %)

Table 7 Clinical characteristics of MPN patients eligible for NGS analysis

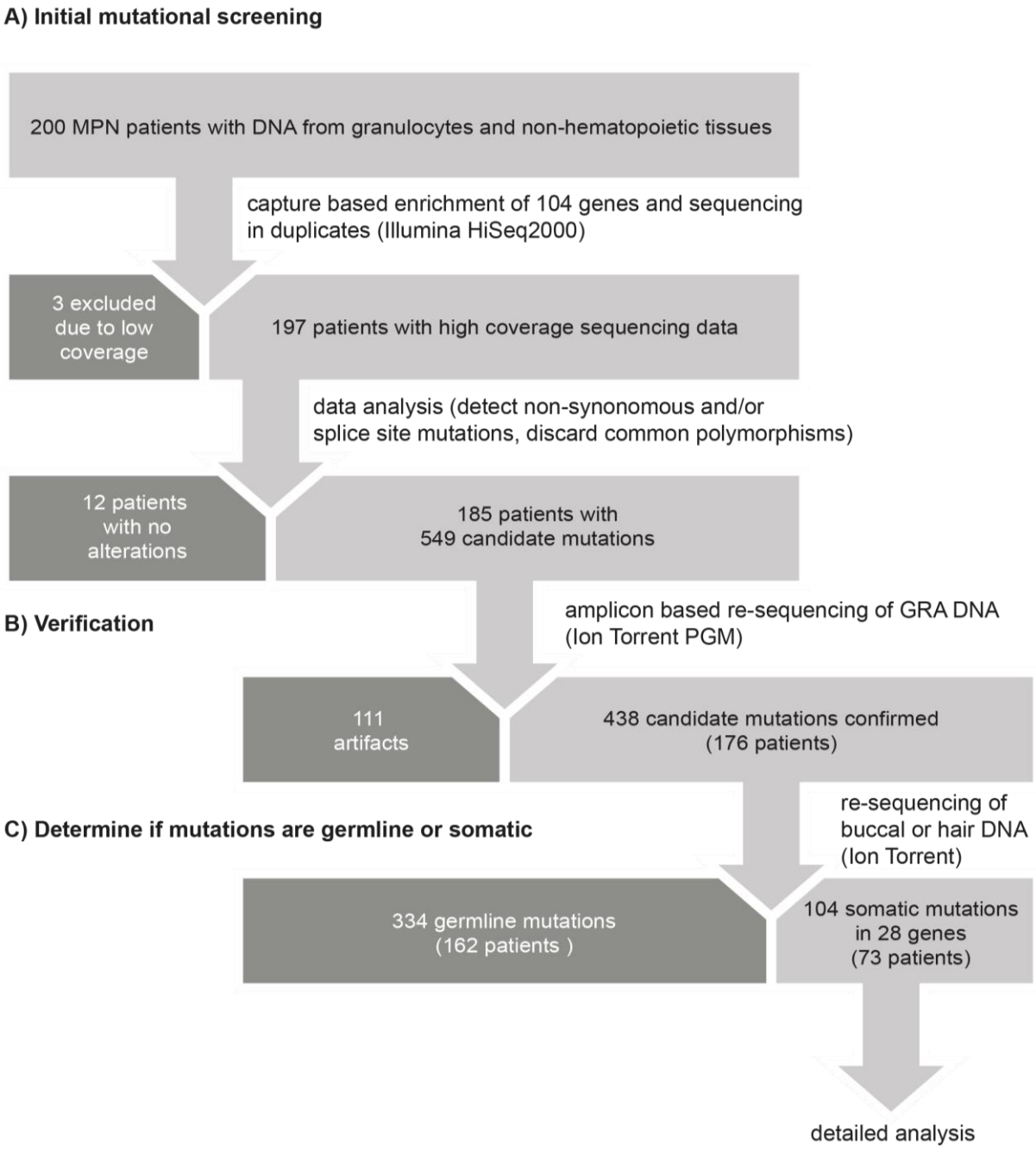
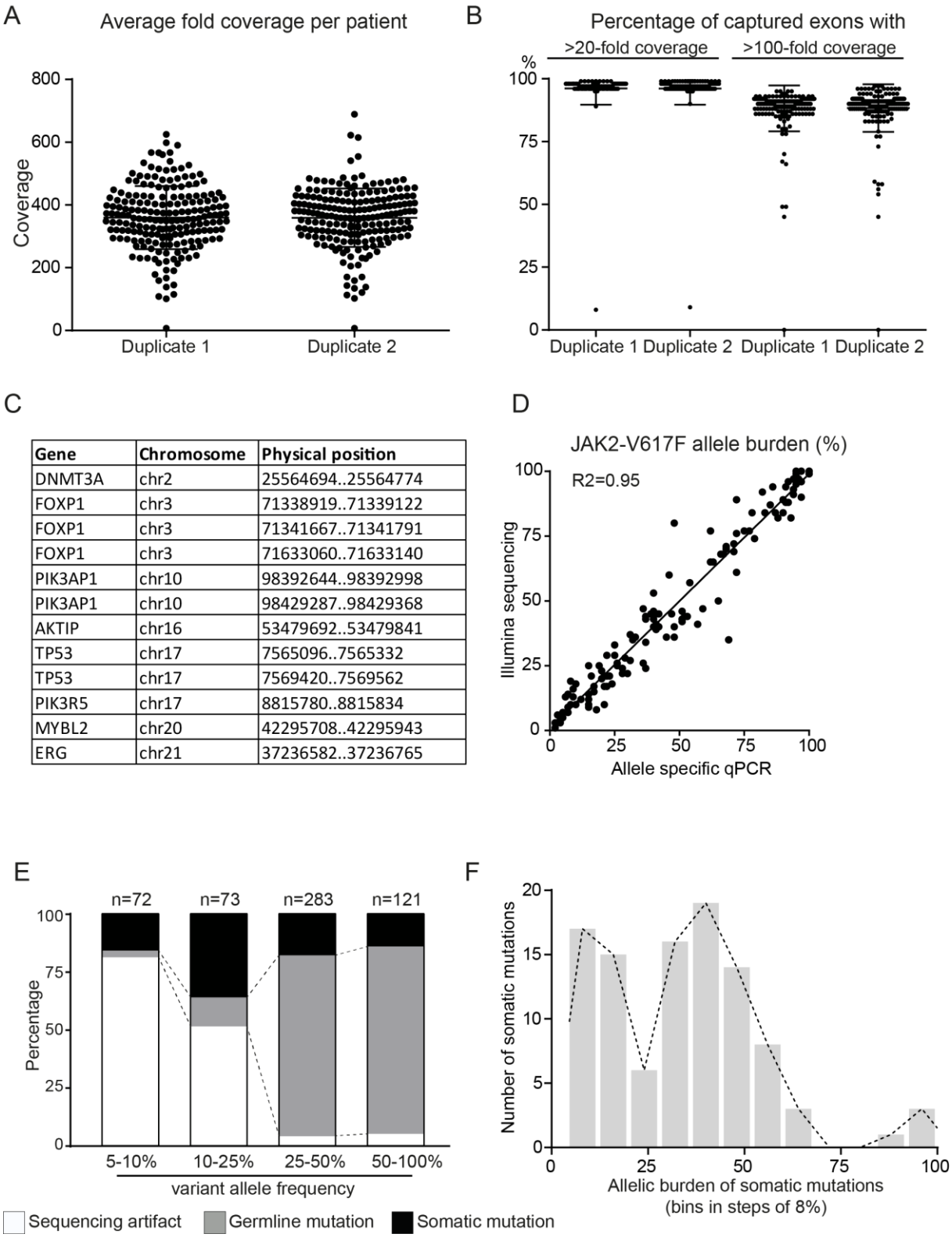


Figure 10 NGS workflow



Gene	Chromosome	Physical position
DNMT3A	chr2	25564694..25564774
FOXP1	chr3	71338919..71339122
FOXP1	chr3	71341667..71341791
FOXP1	chr3	71633060..71633140
PIK3AP1	chr10	98392644..98392998
PIK3AP1	chr10	98429287..98429368
AKTIP	chr16	53479692..53479841
TP53	chr17	7565096..7565332
TP53	chr17	7569420..7569562
PIK3R5	chr17	8815780..8815834
MYBL2	chr20	42295708..42295943
ERG	chr21	37236582..37236765

Figure 11 Quality control measurements of NGS

(A) Average fold coverage as a function of the average number of reads per base in the target region. (B) Percentage of exons per patient covered at least 20 or 100 times. (C) List of targeted regions, which were not covered in the majority of the sequenced samples (D) Correlation between *JAK2-V617F* allele burden determined by Illumina-sequencing (y-axis) and by allele-specific PCR (x-axis). (E) Summary of the validation of variants by Ion Torrent sequencing, sub-divided in columns based on the variant frequency reported by Illumina sequencing. (F) Frequency distribution of confirmed somatic mutations.

Most of the exons from the selected genes were reliably covered in all of the sequenced samples. Regions, which were missed, are listed in Figure 11C. The total size of the commonly missed regions was 1291 bp and represented less than 0.4 % of the targeted regions. To validate the accuracy of the Illumina sequencing method, the variant allele frequency (VAF) of the *JAK2-V617F* mutation was compared to the result acquired by allele specific PCR (asPCR). Both methods showed a strong correlation, which confirm the accuracy of the Illumina sequencing method (Figure 11D).

The Illumina sequencing resulted in 549 candidate mutations, from which 438 could be confirmed by Ion Torrent sequencing (Figure 10B). The validation showed that variants with a frequency between 5 and 10 % are mostly false positive sequencing artifacts (80 %). Roughly, half of the variants with a frequency between 10 and 25 % could be confirmed by Ion Torrent sequencing. Nearly all variants (95 %) with an allele frequency above 25% could be confirmed as either somatic or germline mutation (Figure 11E).

By re-sequencing all 438 confirmed mutations in the germline control DNA, 104 mutations were found to be somatic (Figure 10C). The VAF of all 104 confirmed somatic mutations is shown in a histogram of Figure 11F. The frequency distribution shows three distinct peaks. The first peak at 15 % of allele frequency indicates somatic mutations present in a minority of cells. The second peak at 45 % allele frequency represents somatic mutations, which, if they are heterozygous, are present in nearly all cells of the patient. The last peak at 80 to 100 % of allele frequency marks mutations, which turned homozygous and are present in the great majority of the cells of a patient.

3.1.2 Mutational profile of sporadic MPN

In total, 104 genes were analyzed by NGS (Figure 12). Driver mutations such as *JAK2-V617F* and insertion/deletion mutations in *CALR-exon9* were the most frequent somatic mutations; found in 135 and 29 patients, respectively. Additional somatic mutations were found in 28 of 104 (27 %) genes. In 14 of 28 (50 %) genes, more than one mutation was found (Figure 13A). Epigenetic modifier genes (*TET2*, *ASXL1*, *DNMT3A*, *EZH2* and *IDH1*) represented the most frequently mutated group of genes. Five patients had mutations in *TP53* and four patients carried mutations in *NFE2*. All four mutations found in the transcription factor NFE2 were either deletion or insertion mutations. The genes *CBL*, *CUX1*, *NRAS* and *NF1* were found to be mutated twice in our cohort of MPN patients (Figure 13A).

The 104 somatic mutations, which are not reported as driver mutations, were found in 73 out of 197 (37 %) MPN patients. Only 26 % (18/69) of ET patients carried at least one somatic mutation additional to the driver mutation. In the group of PV and PMF patients, the proportion of patients with additional alterations was slightly larger: 35 of 96 (37 %) PV patients and 12 of 34 (35 %) PMF patients were positive for additional somatic mutations. One PMF patient carried four somatic mutations additional to *JAK2-V617F*.

In 13 % (9/69) of ET patients, 7 % (7/94) of PV patients and 12 % (4/34) PMF patients neither a driver mutation nor any other somatic mutation was found (Figure 13B).

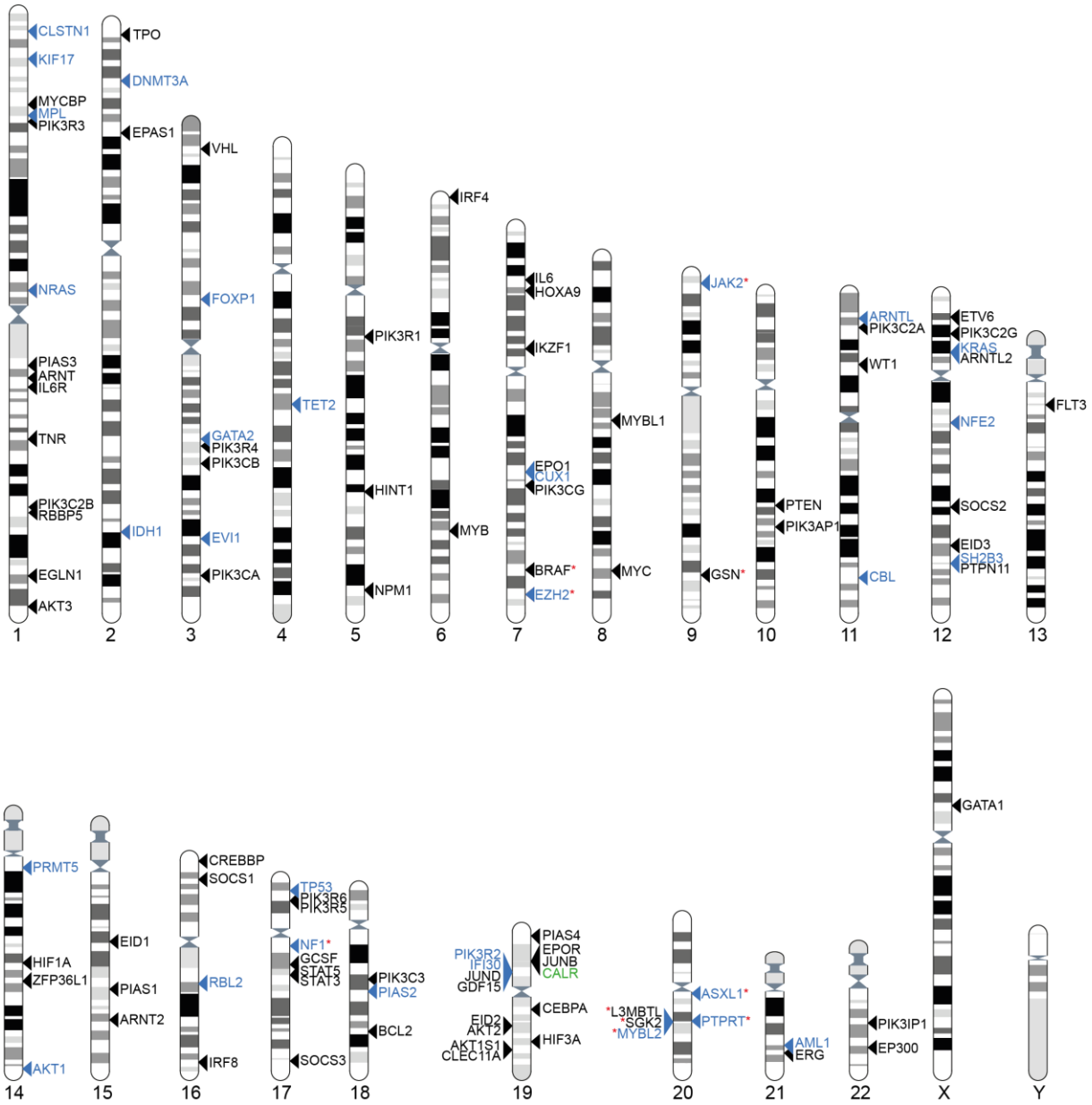


Figure 12 Karyogram and chromosomal location of all sequenced genes

Genes labeled in black have not been found to be mutated. Blue names represent genes with somatic mutations; red asterisks copy number alterations.

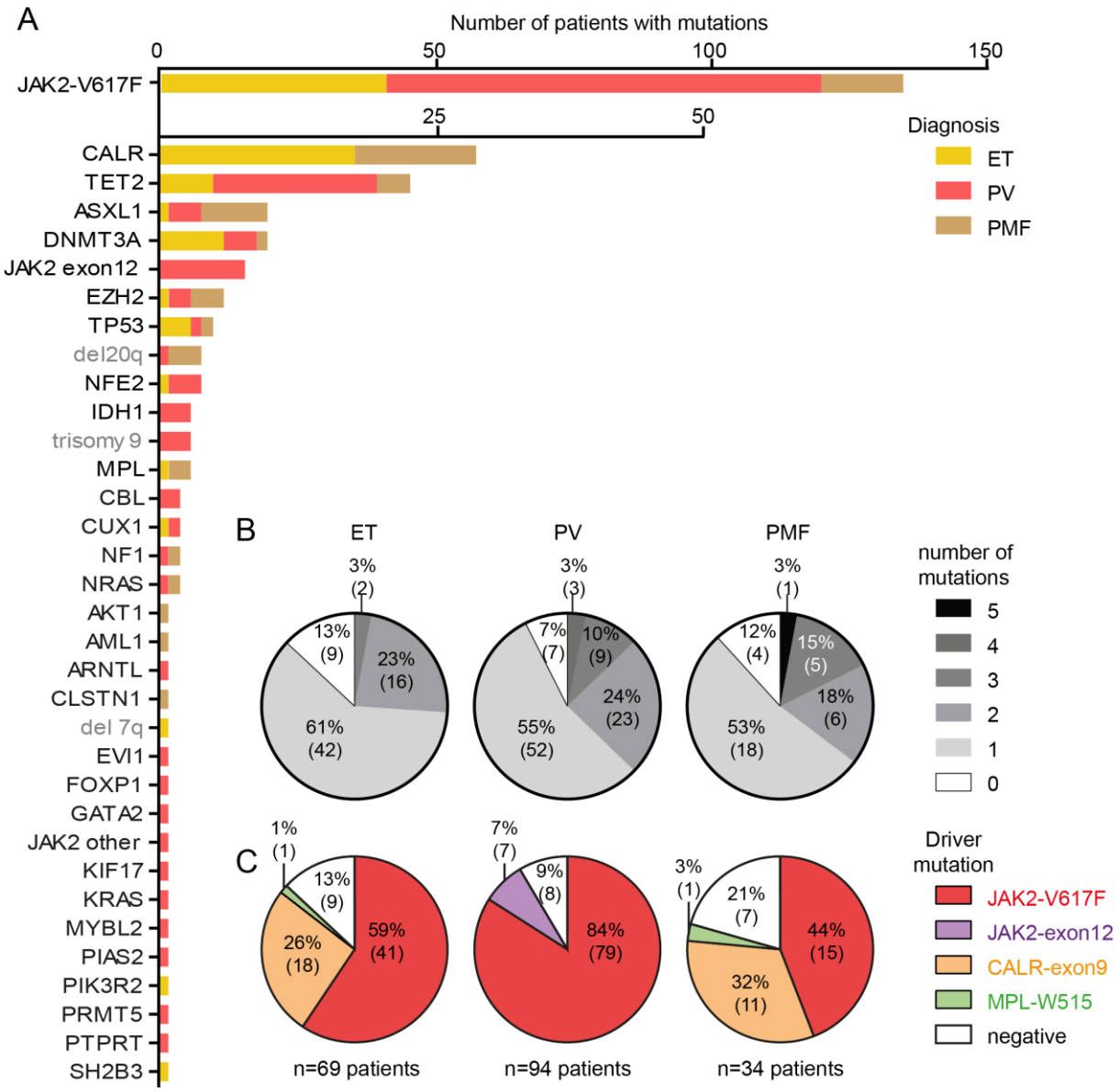


Figure 13 Frequency and distribution of somatic mutations in patients with MPN

(A) Histogram of patients mutated for the listed genes. Numerical aberrations in chromosomes are marked in grey. The colors of the bars indicate the diagnosis of the patients. (B) Distribution of all confirmed somatic mutations among the 197 MPN patients according to their phenotype. The shades of gray indicate the number of somatic mutations per patient. (C) Distribution of the known driver mutations among the same MPN patients. The different colors indicate the individual driver mutation.

In the group of driver mutations, *JAK2-V617F* was found the most frequently. In total, 135 of 197 (69 %) patients of all three MPN phenotypes were positive for *JAK2-V617F*. The insertion deletion mutations in the exon 9 of *CALR* were found in patients with ET and PMF. In total, 18 of 69 (26 %) ET and 11 of 34 (32 %) PMF patient were CALR positive. Mutations in JAK2 exon 12 were present in 7 of 94 (7 %)PV patients. MPL-W515 mutations were found in three patients, one diagnosed for ET and two diagnosed for PMF (Figure 13C).

In addition to somatic mutations, three types of chromosomal aberrations could be detected by targeted NGS. Deletions in the q arm of chromosome 20 (del 20q) were found in one PV patient and three patients with PMF. Three PV patients were affected by a trisomy of chromosome 9 and one ET patient carried a deletion in the q arm of chromosome 7 (del 7q, Figure 13A). Detailed analysis showed that in two out of four patients with del 20q, the ASXL1 gene was included in the deletion (Figure 14).

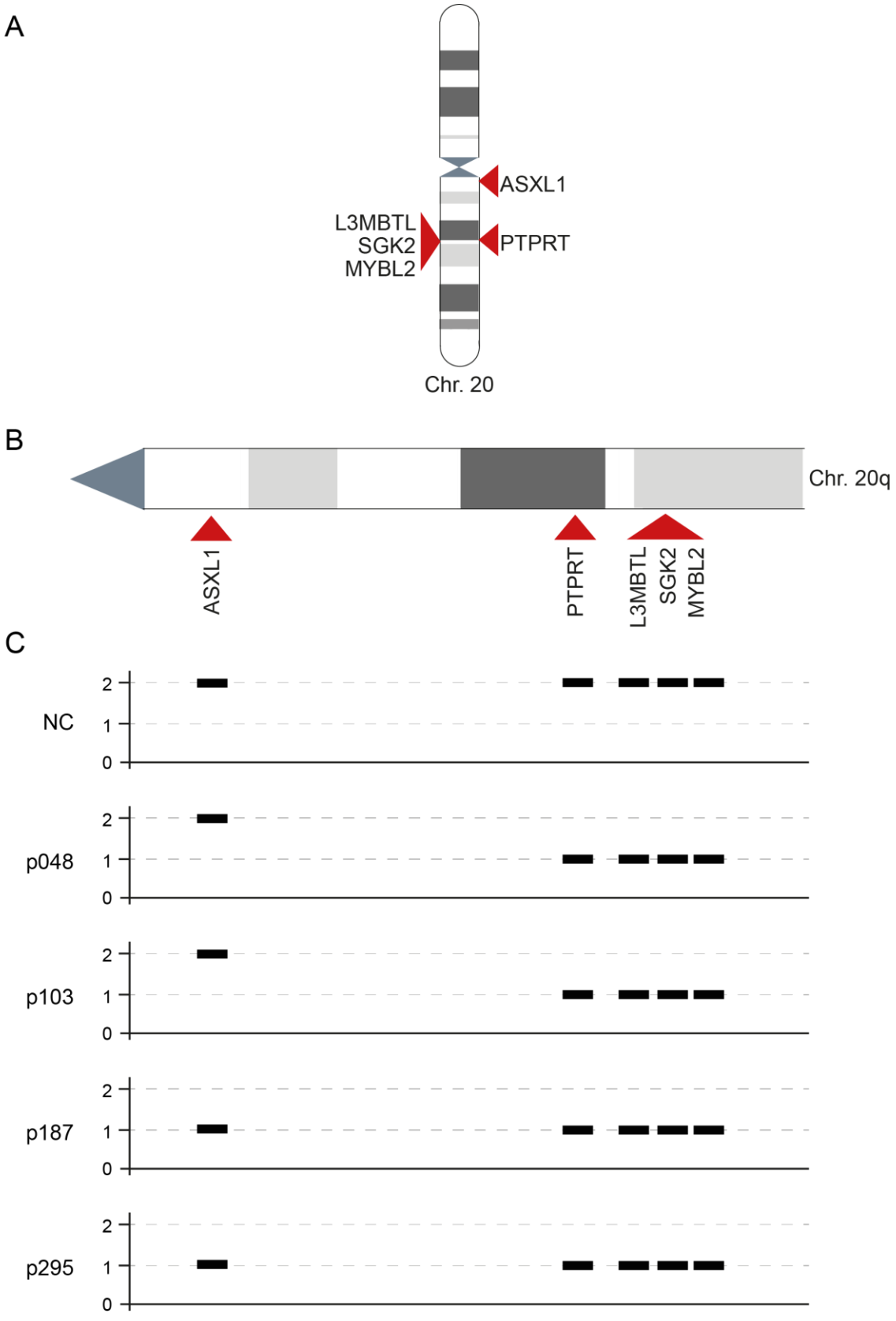


Figure 14 Location of copy number variations on chromosome 20

(A) Schematic view of chromosome 20. Red triangles mark the locations of the genes covered by targeted NGS. (B) Detailed view on q arm of chromosome 20 and the covered genes. (C) Individual graphs for normal controls and four patients indicate the average copy number at the corresponding chromosomal region.

In total, 113 of 197 (57 %) patients showed a single somatic aberration, 64 (32 %) were found to carry more than one. Only 18 patients (9 %) did not present any somatic mutation. The Circos plots in Figure 15 display the co-occurrence of somatic mutations of all patients, who carry at least one somatic mutation. Patients with single mutations usually are positive for a known driver mutation such as *JAK2-V617F*, *JAK2-exon12* or the *CALR-exon9* (Figure 15A-C). Most commonly observed co-occurrences are *DNMT3A* and *TET2* mutations in combination with a driver mutation. Mutations in *IDH1* and *NFE2* are rare, but seem to preferentially appear in *JAK2-V617F* mutated patients (Figure 15A). As mutations in these genes are found less often, larger patient populations are required to validate this pattern of co-occurrence.

Interestingly, we also found two patients who carry two driver mutations, *JAK2-V617F* and *CALR-exon9*. Previous reports indicated that driver mutations would usually be mutually exclusive in a patient[35,36].

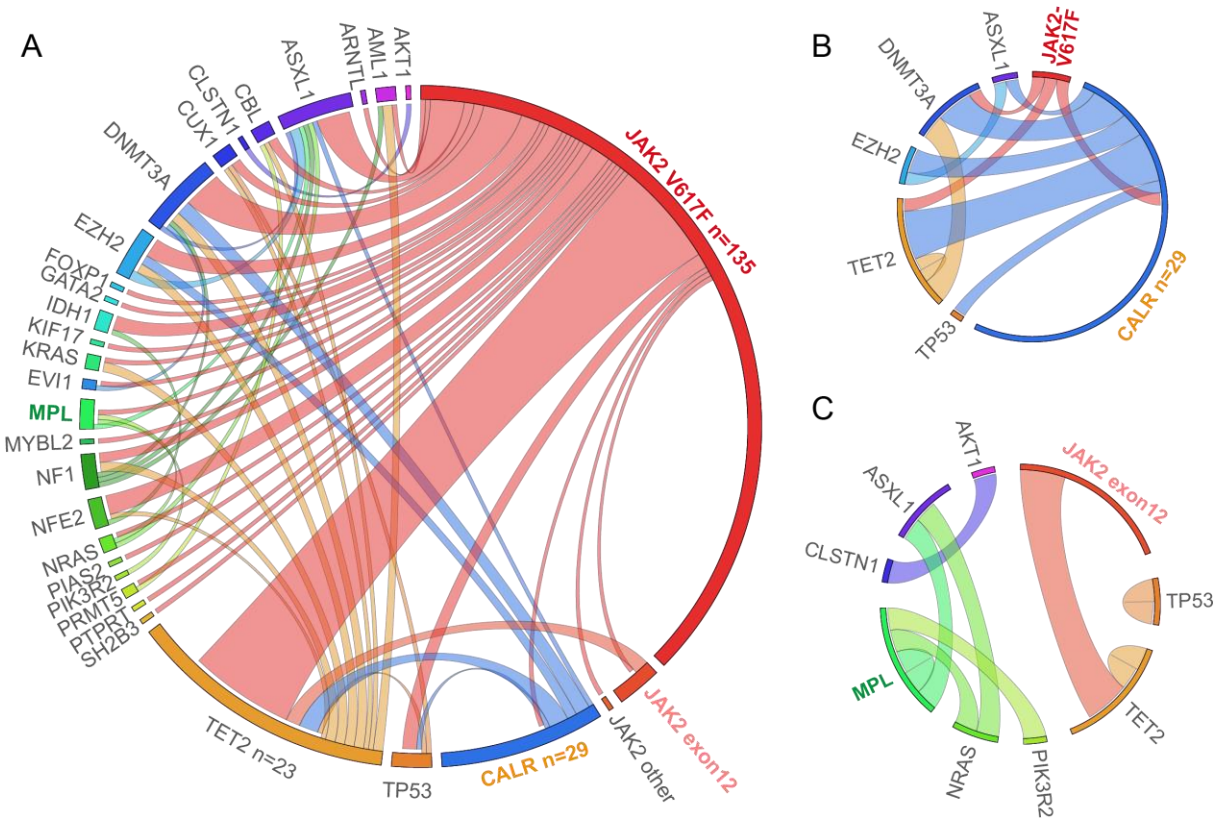


Figure 15 Circos plots showing co-occurrence of somatic mutations

The broader the connection between two genes, the more patients carry mutations in both of the genes. Empty regions of the outer circle, which are not connected to any other gene, represent patients, which are only mutated for this particular gene represented by this part of the circle. (A) Circos plots illustrating co-occurrence of somatic mutations in all patients carrying somatic mutations (B) Circos plots illustrating co-occurrence of somatic mutations in patients with CALR driver mutation (C) Circos plots illustrating co-occurrence of somatic mutations in patients with *JAK2-exon12* driver mutation or no driver mutation at all.

3.1.3 Assessment of clonal evolution in MPN

To combine the picture of co-occurrence of mutations with their temporal order of acquisition, we plated PBMCs in methylcellulose and genotyped the DNA from single grown colonies. The mutations could occur one after another in a sequential order or appear in an extra clone (Figure 16A). By comparing the order of acquisition of mutations in epigenetic modifier genes in relation to *JAK2-V617F*, we found that mutations in *DNMT3A* and *TET2* predominantly occur prior to *JAK2-V617F* or separately, resulting in a bi-clonal pattern. *EZH2* and *ASXL1* mutations appear before and after *JAK2-V617F* or coexisted in separate clones. Only *IDH1* was exclusively acquired after *JAK2-V617F* (Figure 16B).

Patients with three or more somatic mutations are shown in Figure 16C. The patients show various patterns of clonal architecture. In four patients, all mutational events happened in a sequential order, in the four other patients, a branching led to the different subclones. In four cases, mutations in *TET2* were acquired before other mutations such as *JAK2*, *EZH2* or *KRAS*. In two patients with an insertion-deletion mutation in *CALR-exon9*, the *CALR* mutation was acquired before additional somatic mutations, such as *EZH2* and *TET2*.

Two patients in our cohort were found to be mutated for *JAK2-V617F* and an insertion-deletion mutation in *CALR-exon9*. One of these patients, P101, also carries a mutation in *DNMT3A* and *TET2*. The colony analysis of this patient shows that the MPN driver mutations in *JAK2* and *CALR* are present in separate clones. While the *JAK2-V617F* clone disappeared over time, the CALR clone remained stable in size. The clone with mutations in *DNMT3A* and *TET2* expanded during the course of disease (Figure 16C). Another patient, who carried two driver mutations, was positive for *JAK2-V617F* and JAK2-E542-N543-del, a reported driver mutation in exon12 of JAK2. Similar to patient P101, also in this patient the driver mutations were present in individual clones. Our lab has published this third patient in a previous study[77].

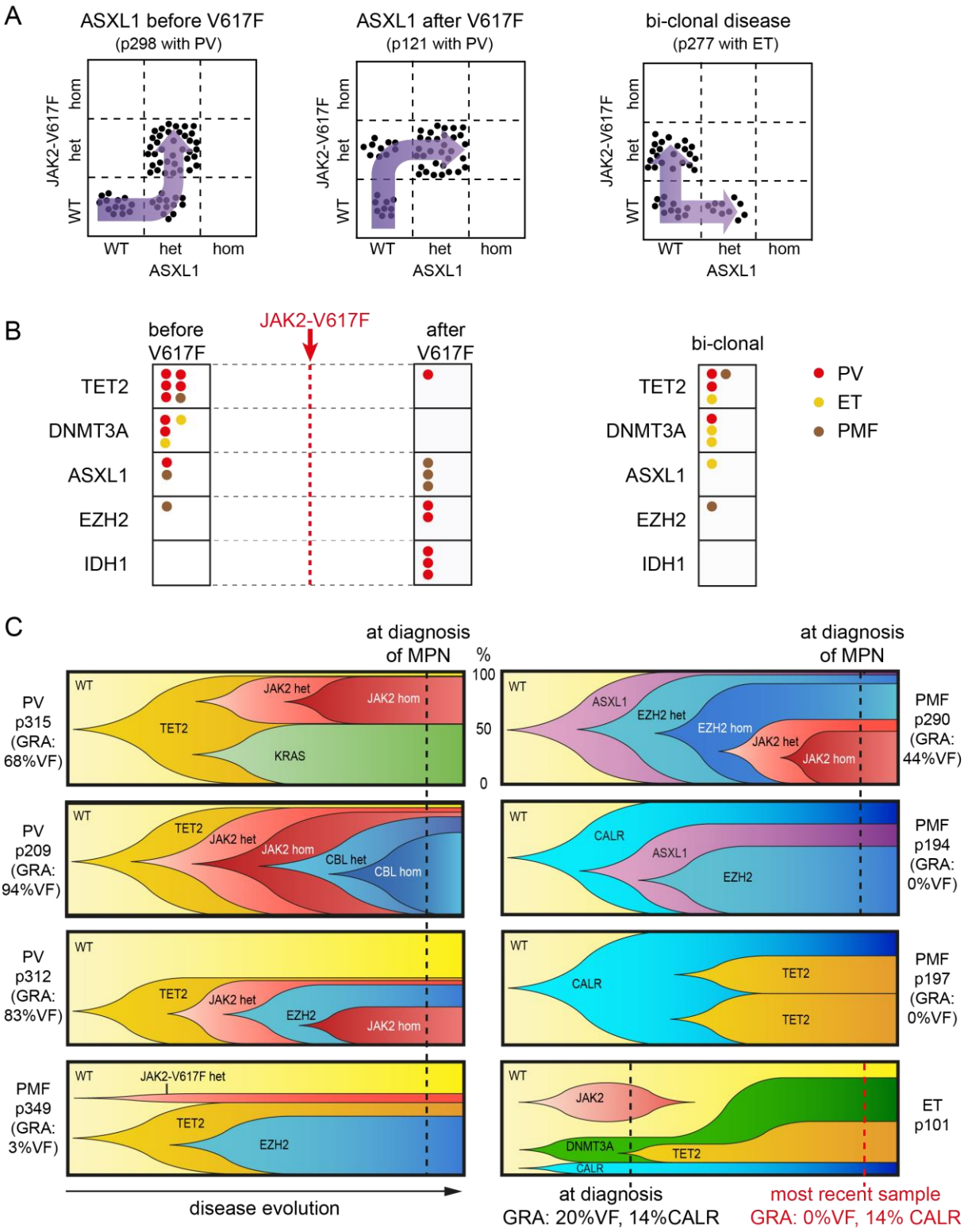


Figure 16 Clonal evolution in MPN patients carrying somatic mutations in epigenetic modifier genes

Figure legend from previous page: Single erythroid or granulocytic colonies (BFU-Es and CFU-G) grown in methylcellulose were individually picked and analyzed for the presence or absence of JAK2 V617F and other somatic mutations. (A) Examples of 3 patients who acquired an ASXL1 mutation before JAK2 V617F (left panel), after JAK2 V617F (middle panel), or in a clone separate from JAK2 V617F (right panel) are shown. Each dot represents a single colony that was genotyped and placed into the corresponding quadrant. (B) Summary of the temporal order of acquisition of mutations in relation to JAK2 V617F. Each dot represents 1 patient analyzed as shown in panel A and placed into the corresponding quadrant. Events in ET patients are depicted in yellow, PV patients in red, and PMF patients in brown. (C) Patterns of clonal evolution in eight MPN patients carrying multiple somatic mutations. Dotted lines denote the time of analysis and the y-axis indicates the percentage of the colonies with or without the corresponding somatic mutations. GRA, granulocytes. %VF, *JAK2-V617F* mutant allele burden in granulocytes purified from peripheral blood. Although the order of events depicted can be deduced from the single-clone analysis (dotted line), the exact timing of the acquisition of the individual mutations and the time needed for the clonal expansion remains unknown and is shown only schematically.

3.1.4 Prognostic value of somatic mutations

The comparison of clinical data with NGS results showed that the number of somatic mutations per patient correlates with increased leukocyte count. Especially patients with two or more somatic mutations or with a mutation in *EZH2* showed an elevated leukocyte count (Figure 17A and B). Patients with somatic mutations in *ASXL1* presented with significantly lower hemoglobin at diagnosis (Figure 17C). The comparison of the NGS result with clinical follow-up data showed that in MPN, the number of somatic mutations per patient correlates with increased risk of transformation to AML and reduced overall survival (Figure 17D and E). Mutations in *TP53*, *TET2* and other epigenetic modifier genes, such as *ASXL1*, *DNMT3A*, *EZH2* and *IDH1*, were analyzed separately for their impact on the risk of transformation to AML and reduced overall survival (Figure 18).

Mutations in *TP53* and *TET2* correlated with transformation into AML and reduced survival, while mutations in other epigenetic modifier genes did not reach significance (Figure 18B-D). Somatic mutations found in the most recent patient samples were re-sequenced and quantified in all previous samples of the same patient. This allowed tracking the mutant allele burden and assessing the clonal size over time for each patient with more than one sample (Figure 18A). From five patients with somatic mutations in *TP53*, serial samples were available from four patients (Figure 18B). Three patients had one somatic mutation in *TP53* (solid lines) and one patient had two somatic mutations in *TP53* (dashed lines, each representing the clone carrying one of the two *TP53* mutations). Four of the five mutations, which could be tracked through previous samples, were present in all of the examined samples. Only one somatic mutation in *TP53* was acquired during the follow-up time of the study. Interestingly, the allele burden of all somatic *TP53* mutations remained very low (<5 %) for multiple years. In four patients, the *TP53* mutant allele burden increased above 50 %, potentially through a combination of loss of heterozygosity and clonal expansion. Notably, the MPN phenotype of these patients transformed to AML, leading to subsequent death only months after the expansion of the clone. The only patient of the cohort with a mutation in *TP53* and an allele burden below 50 % is still alive. This intriguing observation was specific for mutations in *TP53* and could not be found in patients carrying somatic mutations in other genes.

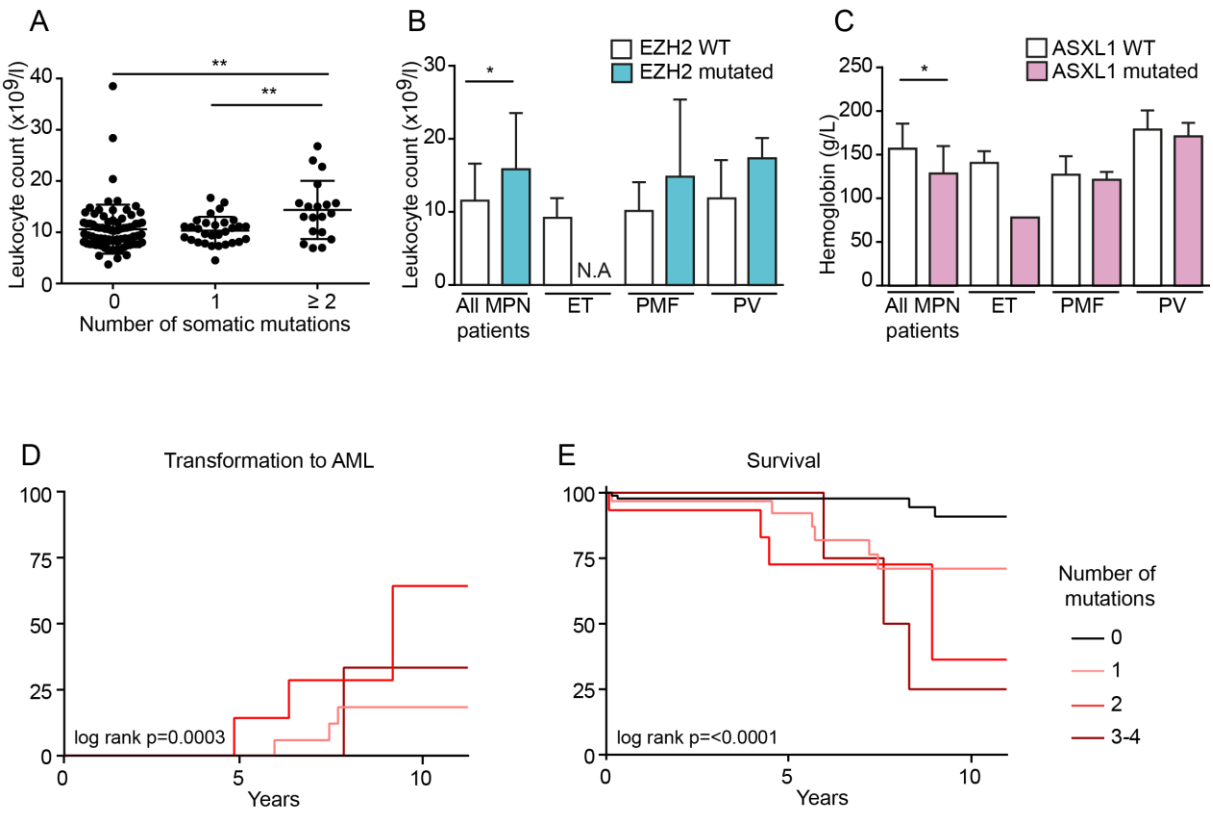


Figure 17 Impact of somatic mutations on blood counts, transformation and survival

(A) Correlations between the number of somatic mutations (excluding driver mutations) and leukocyte count. (B) Correlation between the mutation status of EZH2 and leukocyte count. (C) Correlation between mutational status of ASXL1 and hemoglobin count. Kaplan-Meier curves for correlation of the number of somatic mutations (omitting driver mutations) with transformation to AML (D) and overall survival (E). Symbols: * p<0.05, ** p<0.01, *** p<0.005, **** p<0.001.

Serial samples were available from 12 of 23 patients carrying mutations in *TET2*. In 11 of the 12 patients (92 %), the *TET2* mutation was present in all previous samples. In five of 12 patients with serial samples, the *TET2* clone expanded during follow-up. Only in one patient, the mutant allele burden reached >75 %. All of the cases with stable clonal size displayed a *TET2* mutant allele burden of approximately 50 % from the first to the most recent sample, indicating that clonal expansion happened prior to MPN diagnosis (Figure 18C). Furthermore, colony analysis revealed that *TET2* mutations most often remain heterozygous. Thereby, a *TET2* mutant *allele* burden of 50 % indicates, that most cells of the peripheral blood are carrying a mutant TET2 allele.

Patients (N=11) with somatic mutations in other epigenetic modifier genes seemed to follow the pattern observed in the subgroup of *TET2* mutated patients. In about half of the patients, the mutant clone remained stable at an allele burden of 50 %, whereas the other patients displayed an increased allele burden over time. All mutations found in the most recent sample were present in the previous samples of the patients (Figure 18D).

In total, serial samples were available from 28 patients harboring 38 somatic mutations (Figure 19A). The serial samples covered a total of 133 patient years and during this time only two patients acquired a mutation (Figure 19B and C). P111, the PV patient who gained a *TET2* mutation, was treated with aspirin. The patient who gained the *TP53* mutation (P060), was initially diagnosed with an ET and transformed into AML. This patient was treated with hydroxyurea. All other somatic mutations were found at diagnosis or at the first available sample.

Interestingly, when correlating the number of somatic mutations per patient with their age at diagnosis, patients who did not carry any somatic mutation were diagnosed at younger age (Figure 19D). In total, 18 patients of 197 (9 %) did not carry any mutation: nine ET, seven PV and four PMF patients. Patients with an ET phenotype were diagnosed at median age of 51, PV and PMF patients at 60 and 63 years, respectively (Figure 19E).

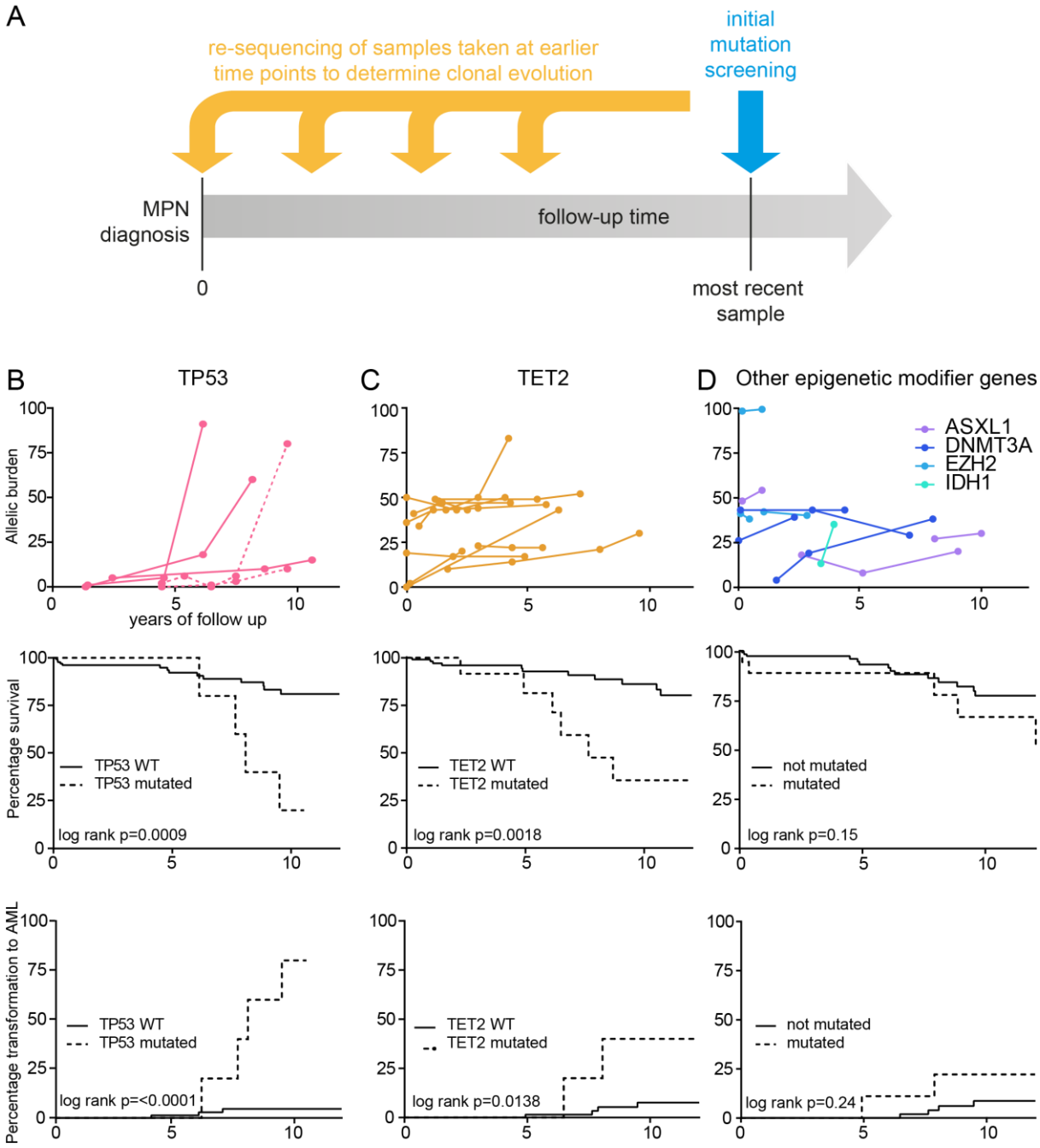


Figure 18 Longitudinal sequencing reveals most mutations are present at diagnosis

(A) Scheme of re-sequencing and quantification of somatic mutations in patient samples acquired prior to the sample used for NGS. (B) Time course of the *TP53* mutant allele burden in serial follow-up samples of four MPN patients with available follow-up samples (upper panel). One patient harbored two distinct *TP53* mutations (dotted lines), only one of which displayed loss of heterozygosity. Survival (middle panel) and transformation to AML (lower panel) is shown below for five patients with mutations in *TP53*. (C) Time course of the *TET2* mutant allele burden in serial follow-up samples of 12 MPN patients (upper panel). Survival (middle panel) and transformation to AML (lower panel) is shown below for 23 patients with mutations in *TET2*. (D) Time course of the mutant allele burden of epigenetic modifiers (*ASXL1*, *DNMT3A*, *EZH2*, and *IDH1*) in serial follow-up samples of 11 MPN patients (upper panel). Survival (middle panel) and transformation to AML (lower panel) is shown below for 29 patients with mutations in *ASXL1*, *DNMT3A*, *EZH2*, or *IDH1*.

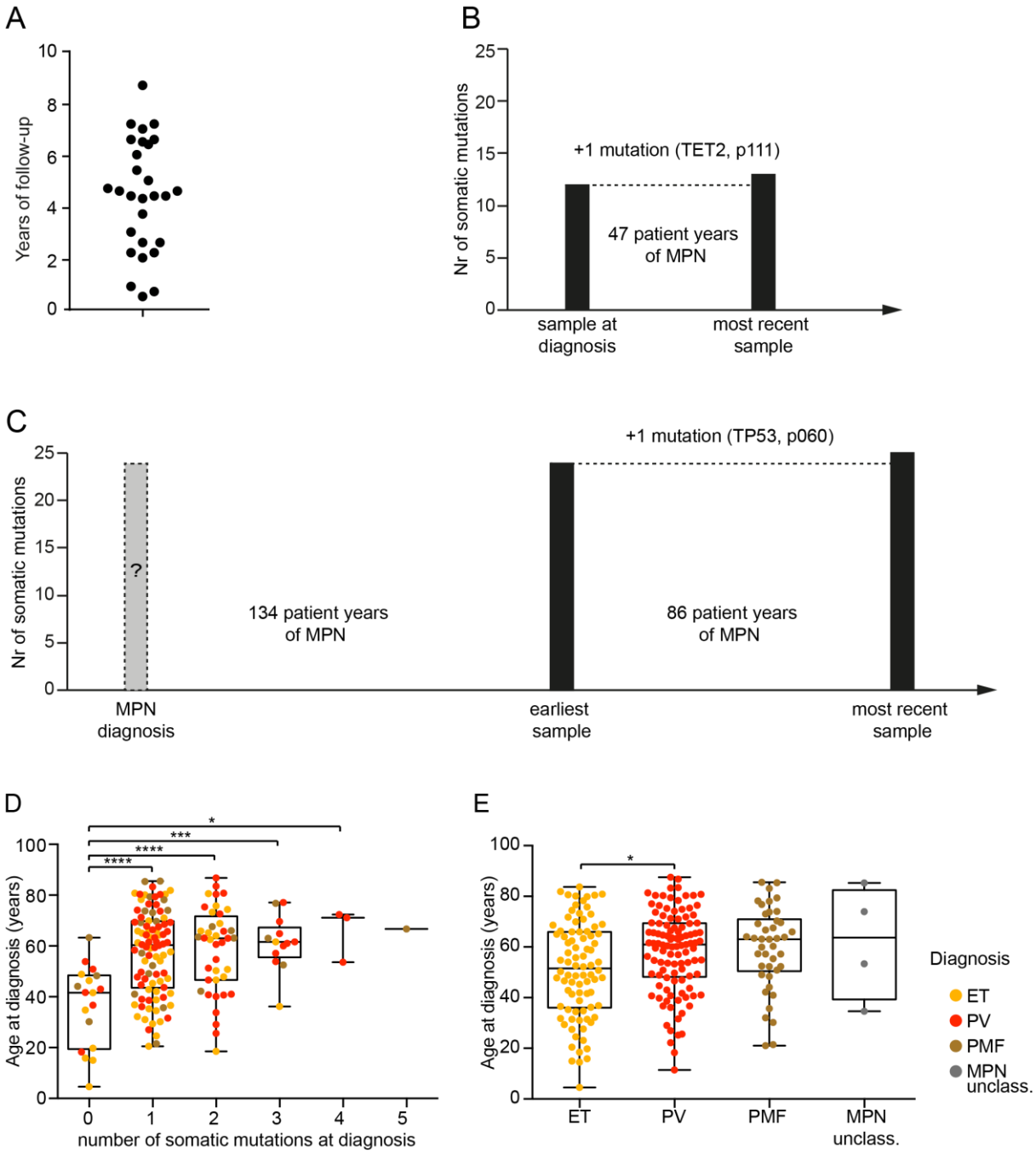


Figure 19 Only two somatic mutations acquired during disease course

(A) Distribution of years of follow-up among the 28 patients shown in panel B & C. (B) Histogram of patients (N=10) with serial samples, starting with the first sample at the date of diagnosis. (C) Histogram of patients (N=18) with serial samples and study entry after the date of diagnosis. (D) Correlation of age at diagnosis with MPN phenotype. (E) Correlation of number of somatic mutations (including driver mutations) with age of diagnosis.

3.2 Pediatric MPN patients display a different mutational phenotype

We next sought to determine the genetic background of pediatric MPN patients in comparison to adult MPN patients. The clinical and laboratory data of 43 patients with pediatric MPN (age ⩽18 years at diagnosis) that were included in this study are summarized in Table 8. Family history of MPN was negative in all children. The WHO 2008 criteria for ET were fulfilled in all 25 cases whose bone marrow histology was available. To establish the diagnosis of ET in the remaining 16 patients without bone marrow examination, we used the proposed revision of the WHO criteria[104] adjusted for age-specific differences in the normal blood counts[105,106]. Elevated platelet counts (>450 × 10^9/L) for at least 12 month of follow-up and absence of signs suggesting a reactive or secondary cause were required for ET diagnosis. Data on spleen size were available for 34 patients with ET and splenomegaly was noted in 14 of them (41 %). There were five hemorrhagic events and one transient ischemic attack observed in five of 41 (12 %) ET patients. Two PV patients were *JAK2-V617F* positive, had hematocrit values >50 % upon follow-up requiring phlebotomies and both had splenomegaly[107].

	ET	**PV**
Number of patients	41	2
Percentage of females	66%	100%
Age at diagnosis median (range)	9 (1-18)	10 (4-17)
Hemoglobin (g/L) of all patients median (range)	128 (80-157)	156 (153-160)
Hemoglobin males only	135 (113-157)	n.a.
Hemoglobin females only	125 (80-146)	156 (153-160)
Platelets (10^9/L) median (range)	1391 (489-4443)	893 (744-1043)
Leukocytes (10^9/L) median (range)	9 (5-17)	19 (10-27)
Splenomegaly	14/34 (41%)	2/2 (100%)
Complications (thrombotic events or hemorrhaging)	5/41 (12%)	n.a.

Table 8 Clinical characteristics of pediatric MPN patients

The DNA of the patients was kindly provided by Maria Lugia Randi (University of Padova, Padova, Italy). We used the same capture-based targeted NGS approach to analyze the same 104 genes in the pediatric patients. The frequencies of the 45 observed sequence alterations are shown in Figure 20A. *JAK2-V617F* (8/43) and *CALR-exon9* (4/43) mutations were found most frequently (Figure 20A). In adults with MPN, the frequency of mutations in genes implicated in epigenetic regulation (*TET2*, *ASXL1*, *DNMT3A*, *EZH2* and *IDH1*) was about 25 %[35,108]. In contrast, we detected mutations in these genes in only four of our 43 pediatric MPN patients (9 %; Figure 20A). We found recurrent mutations in the *IRF8* gene, which encodes an interferon-regulatory transcription factor with a possible role as a leukemia tumor suppressor[109]. Three patients with ET had the same *IRF8-P310A* mutation, which is predicted to be deleterious by all structure prediction algorithms and a fourth patient carried an *IRF8-R228H* mutation, where the predictions were not unanimous. The allele burden of the *IRF8-P310A* mutation was 99, 90 and 67 %, respectively, suggesting that the mutation was homozygous in some or the majority of granulocytes in these patients. Six additional genes were mutated in two different patients each, whereas the other genes were mutated only once (Figure 20A). Two patients with ET carried mutations in the *EPOR* with allele burdens close to 50% suggesting heterozygosity. One *JAK2-V617F* positive ET patient carried an *EPOR-V264G* mutation in the transmembrane domain of EPOR. Based on a model for mouse EPOR[110], this mutation is expected to stabilize a less active dimeric interface for EPOR and predicted to decrease EPOR function. Another ET patient carried an *EPOR-W233G* mutation, which alters the first tryptophan of the conserved WSXWS motif in the extracellular domain of EPOR to GSXWS. In a mouse study, the *Epor-W233G* mutation reduced EPOR surface expression and resulted in a loss of function of the receptor[111]. Thus, both *EPOR* mutations are predicted to reduce or eliminate EPOR function. The fact that both patients have ET and not PV, further argues against a causative role of these mutations[107].

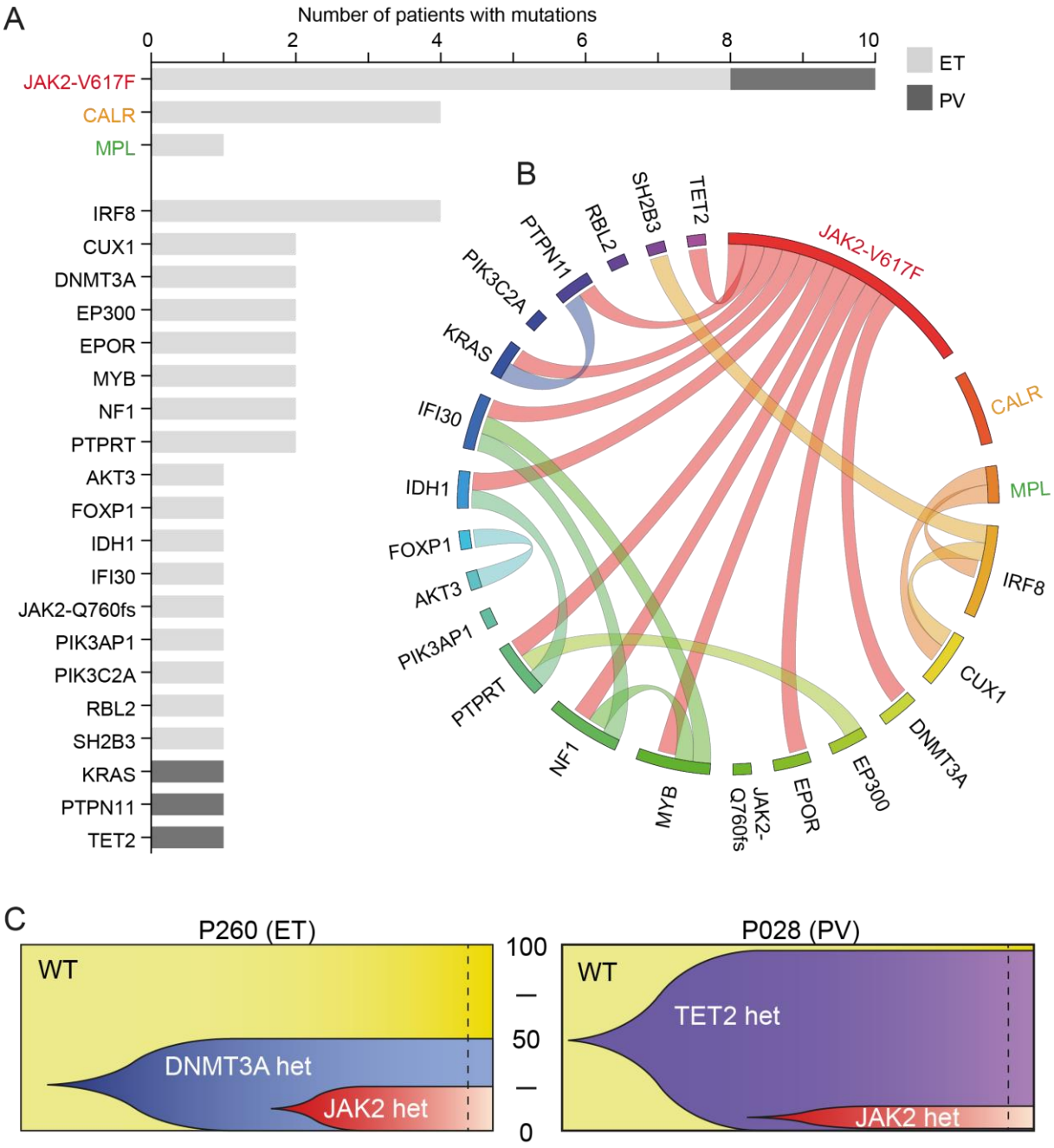


Figure 20 Mutational profile of pediatric MPN

(A) Histogram of patients mutated for the listed genes. ET patients marked in light grey, PV patients marked in dark grey. (B) Circos plot showing the co-occurrence of somatic mutations in pediatric MPN patients (C) Patterns of clonal evolution of two pediatric MPN patients. The dotted line marks the time of diagnosis.

Co-occurrence between mutations in the same patient is shown in Figure 20B. The four patients with *CALR* mutations did not carry additional gene mutations, whereas patients with *JAK2-V617F* or MPL mutations frequently carried other gene mutations. From one pediatric patient (P260) with *JAK2-V617F* and a mutation in *DNMT3A* and the patient (P028) with *JAK2-V617F* and a mutation in *TET2* PBMCs were available. We used these cells to dissect the clonal architecture by genotyping DNA from PBMC-derived single colonies grown in methylcellulose. The *DNMT3A* mutation in this patient had been acquired before *JAK2-V617F* and the clone expanded to account for 50 % of the progenitors. The *TET2* mutation in P028 was also acquired before *JAK2-V617F* and was found in 225 of 226 tested colonies (Figure 20C). Both tested pediatric patients resemble the observation in adult MPN patients, where mutations in *TET2* and *DNMT3A* preferentially also occur early in the development of the MPN clones[(108)].

Figure 21A compares the distribution of the number of mutations per patient in our pediatric cohort with the data from our published MPN cohort of 192 adult patients that was analyzed using the same technologies[(108)]. Mutations in one of the established MPN driver genes *JAK2*, *CALR* or *MPL* were found in a lower percentage of pediatric cases (34 %) than adult MPN patients (90 %; Figure 21A). Conversely, a substantial proportion of pediatric patients who were tested negative for mutations in MPN driver genes, carried mutations in other genes. In addition, a higher percentage of pediatric cases had no detectable mutation in the analyzed genes (32 % versus 8 % in adults; Figure 21A and B). Overall, the mean number of mutations per patient in pediatric MPN was significantly lower than in adult disease (Figure 21C). The subgroup of pediatric patients without detectable mutation showed a trend toward higher platelet counts compared with patients carrying mutations (Figure 21D).

In summary, this study shows that a majority of pediatric patients does not have detectable mutation in any of the genes known to be associated with MPN. Driver mutations in *JAK2*, *MPL* and *CALR* are also the most common hits in this group of patients. Pediatric MPN patients overall also display fewer mutations in genes involved in epigenetic regulation. Two somatic mutations found in *TET2* and *DNMT3A* resemble the order of acquisition observed in adult MPN patients[(108)].

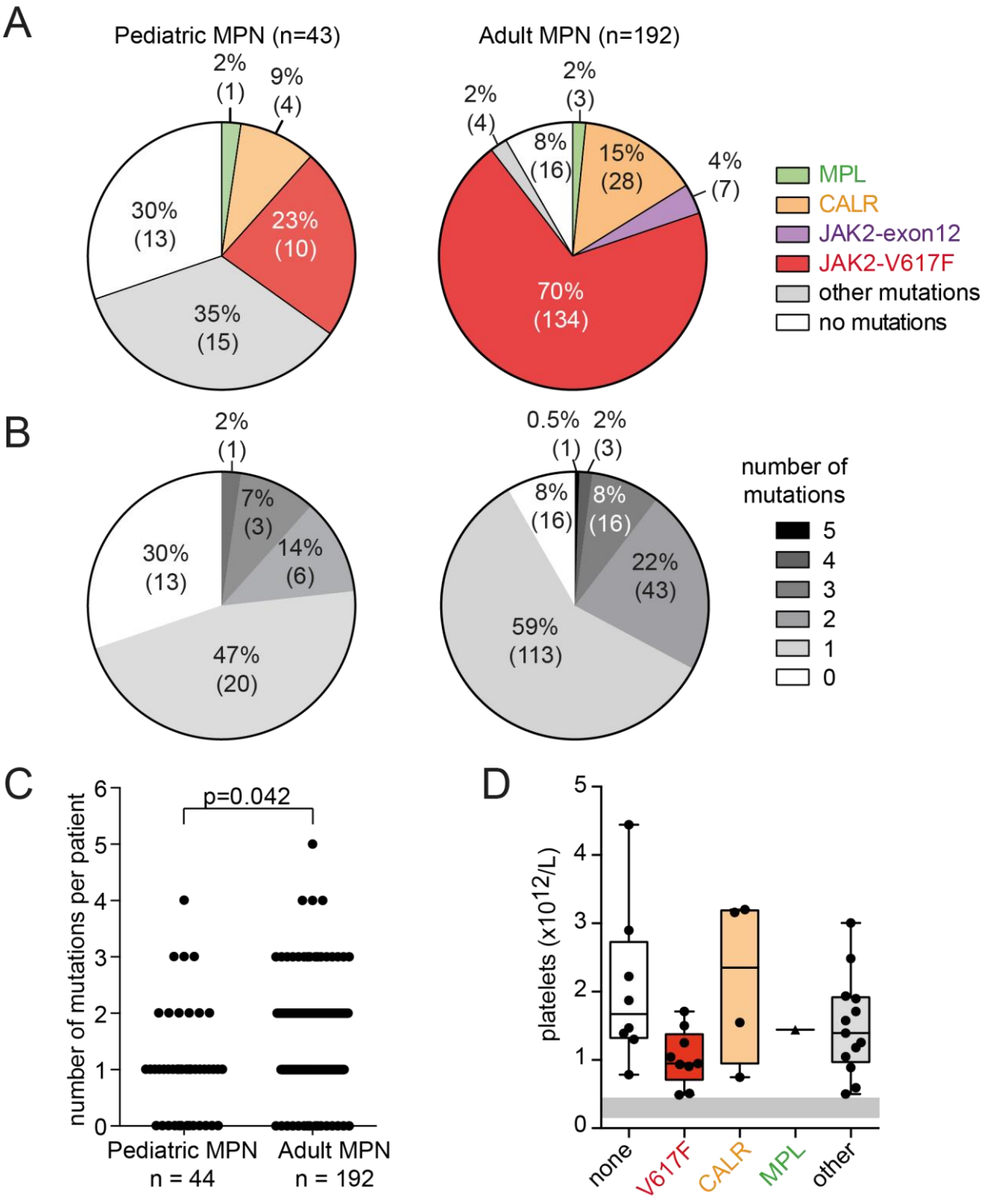


Figure 21 Comparison of mutations in pediatric and adult MPN

(A) Comparison of the distribution of driver mutations in the pediatric cohort (N=43) and a previously analyzed adult MPN cohort (N=192)[108]. The different colors indicate the type of driver mutation. (B) Distribution of somatic mutations among the same pediatric and adult MPN patients. The shades of gray indicate the number of somatic mutations per patient. (C) Comparison of number of mutations per patients for the pediatric and adult MPN cohort. (D) Comparison between platelet count and driver mutation in the pediatric MPN cohort. The gray shaded area indicates the range of normal platelet counts.

3.3 Patients with low *JAK2-V617F* allele burden

The application of NGS technologies to uncover the mutational landscape of MPN, helped to learn more about somatic mutations, but driver mutations remain the most common molecular characteristic in MPN patients. Only 30-40 % of adult MPN patients carry somatic mutations additional to driver mutation. Approximately, 50-60 % of patients are carrying a mutation in *JAK2*, *CALR* or *MPL* as their sole somatic mutation.

For the detection and quantification of driver mutations, granulocytes are enriched from peripheral blood samples of patients with clinical signs of MPN. The DNA of isolated granulocytes is used as template material for allele specific PCR or quantitative sequencing. At the time of MPN diagnosis, the *JAK2-V617F* allele burden can reach a value between >0 % and 100 %. The majority of patients with *JAK2-V617F* mutation and diagnosis of ET has a mutant allele burden <50 % (Figure 22A). Patients with PV show a more heterogeneous frequency distribution. Only in a small proportion of PV patients, the *JAK2-V617F* mutant allele burden remains <20 %. An apparent proportion of patients with diagnosis of PV and a *JAK2-V617F* mutation represent with a mutant allele burden of 80% and above, indicating a loss of heterozygosity in the majority of granulocytes. Both observations were in line with results of previous studies[112,113]. PMF patients carrying the *JAK2-V617F* mutation show a similar heterogeneous distribution as PV patients (Figure 22A).

Studies of *JAK2* negative and MPL negative ET and PMF patients led to the discovery of frame shift mutations in *CALR-exon9*[35,36]. Correlations in ET and PMF patients show, that ET patients represent with a significantly lower *CALR* mutant allele burden than PMF patients. The majority of patients with a *CALR* mutation presented an allele burden of 45 to 55 %. This indicates that nearly all peripheral cells carry a mutant *CALR* allele, as based on our own experience from single colony analysis, loss-of-heterozygosity is rare at the *CALR* locus (Figure 22B). Similar results were obtained from two ET and two PMF patients with a MPL-W515 driver mutation. However, the total number of MPL mutated patients is too low to draw general conclusions from this small group (Figure 22C). Interestingly, PV patients with a *JAK2-exon12* mutation represented an allele burden of <30 %, at time of diagnosis (Figure 22D).

In the next step, we tested for a gender-based skewing in patients positive for *JAK2-V617F* and *CALR* mutations. No significant difference in the mutant allele burden was found when the patients were sub-categorized according to their gender (Figure 22E and F). Furthermore, we tested for correlations of the driver mutation burden and blood counts at diagnosis. In PV patients, the *JAK2-V617F* mutant allele burden at diagnosis positively correlates with hemoglobin and hematocrit and inversely correlates with the number of thrombocytes (Figure 22G-I). The inverse correlation of *JAK2-V617F* burden with platelet count has been observed

before[114]. The correlation analyses were repeated with ET and PMF patients with *JAK2-V617F* or *CALR* mutation, but no correlation was found.

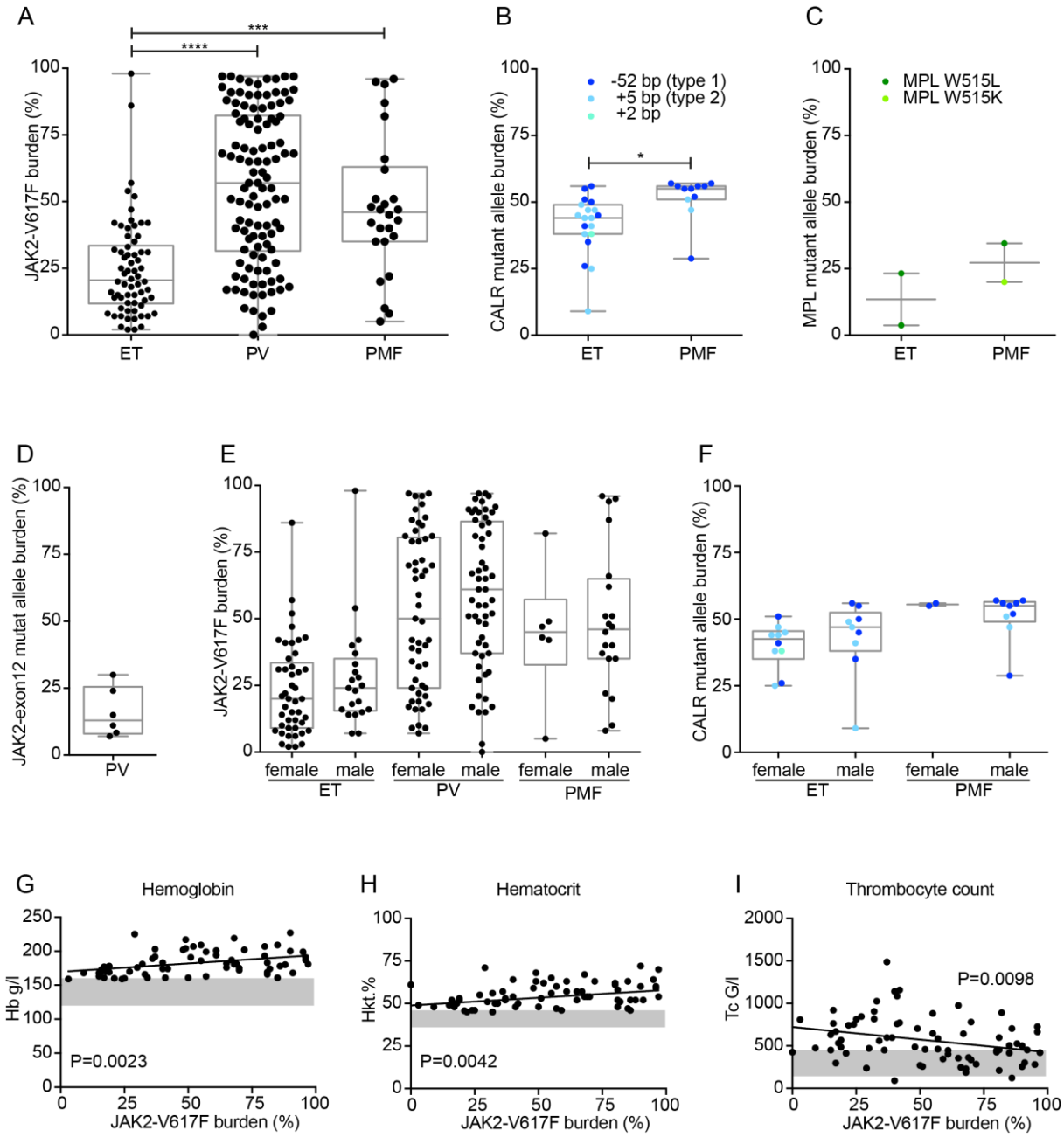


Figure 22 Correlation of driver mutation burden at time of diagnosis

Mutatant allele burden at time of diagnosis of MPN patients mutant for (A) *JAK2-V617F*, (B) insertion-deletion mutations in *CALR-exon9*, (C) mutations in MPL at position W515 and (D) mutations in JAK2-exon12. Correlation for the gender of patients with the mutant allele burden of *JAK2-V617F* (E) and CALR (F) at diagnosis. (G-I) Correlation of clinical blood values with *JAK2-V617F* allele burden of PV patients. The grey area marks the normal range of the specified clinical value

It is generally accepted in the field to use the allele burden of a driver mutation as an indicator of size of the MPN clone: The more mutant alleles are present relative to the wild type alleles, the more cells carry the analyzed mutation and the larger the mutant clone is. This assumption is based on the fact that a normal cell carries two alleles and from standard analysis, homozygous cells can't be differentiated from heterozygous cells. Still, the mutant clone in a patient with an allele burden of 5 % is smaller than the clone in a patient with a mutant allele burden of 25 %. Therefore, the JAK2-mutant allele burden is the marker of molecular response of patient treatments.

Interestingly, the *JAK2-V617F* burden (or: the size of the MPN clone) does not correlate with clinical values in ET and PMF patients at diagnosis. In PV patients, the allele burden of *JAK2-V617F* in granulocytes even inversely correlates with thrombocyte counts. Furthermore, at diagnosis, patients in all MPN phenotypes can be found, which present with a mutant *JAK2* burden <20 % in granulocytes. Hence, we wondered if granulocytes are a reliable indicator for the size of the *JAK2* clone in other lineages of the peripheral blood. We were especially interested in lineages, which are relevant for the diagnosis of MPN, platelets and red cells.

For the patients of the Basel sporadic MPN cohort, RNA of platelets was available from all patients, as the purification of platelets is part of our standard sample workup protocol (Figure 8). To analyze the allele burden in red cells, we decided to enrich reticulocytes, the most advanced red cell progenitors, which still contain RNA. We intended to purify red cell progenitors from frozen PBMC samples, which are also part of our standard sample workup protocol. During multiple pilot experiments, we only managed to enrich only marginal amounts of reticulocytes from the PBMC fraction. Finally, the numbers of reticulocytes purified from frozen PBMCs were not sufficient for subsequent analyses. Presumably, reticulocytes were affected by the treatment with red cell lysis buffer during PBMC preparation. Therefore, we developed an extended purification scheme (Figure 23A), which allowed the parallel isolation of all lineages of the peripheral blood.

Granulocytes, PBMCs and platelets were isolated according to the standard protocol (Figure 8). PBMCs were stained for CD3 (T-cells), CD14 (monocytes), CD19 (B-cells) and CD335 (NK-cells) to isolate Monocytes, NK-, B- and T-cells. Single positive cells were FACS-sorted into individual tubes to prepare cell type specific RNA. In addition to these commonly isolated lineages, we enriched for reticulocytes from the (otherwise discarded) erythrocyte fraction (Figure 23A). Subsequently, enriched reticulocytes were sorted for $CD71^{+}/CD235^{+}/CD42a^{-}/CD45^{-}$ markers to select immature red cells and exclude platelets and white blood cells. Cells from non-red cell lineage origins needed to be excluded, as reticulocytes eject DNA and RNA during their differentiation to mature red cells. Therefore, a contamination of non-red cell-derived RNA would have great impact on cDNA-based analysis of the mutant allele burden.

Figure 23B shows the representative re-analysis of the sorted $CD71^{+}/CD235^{+}/CD42a^{-}/CD45^{-}$ reticulocyte fraction.

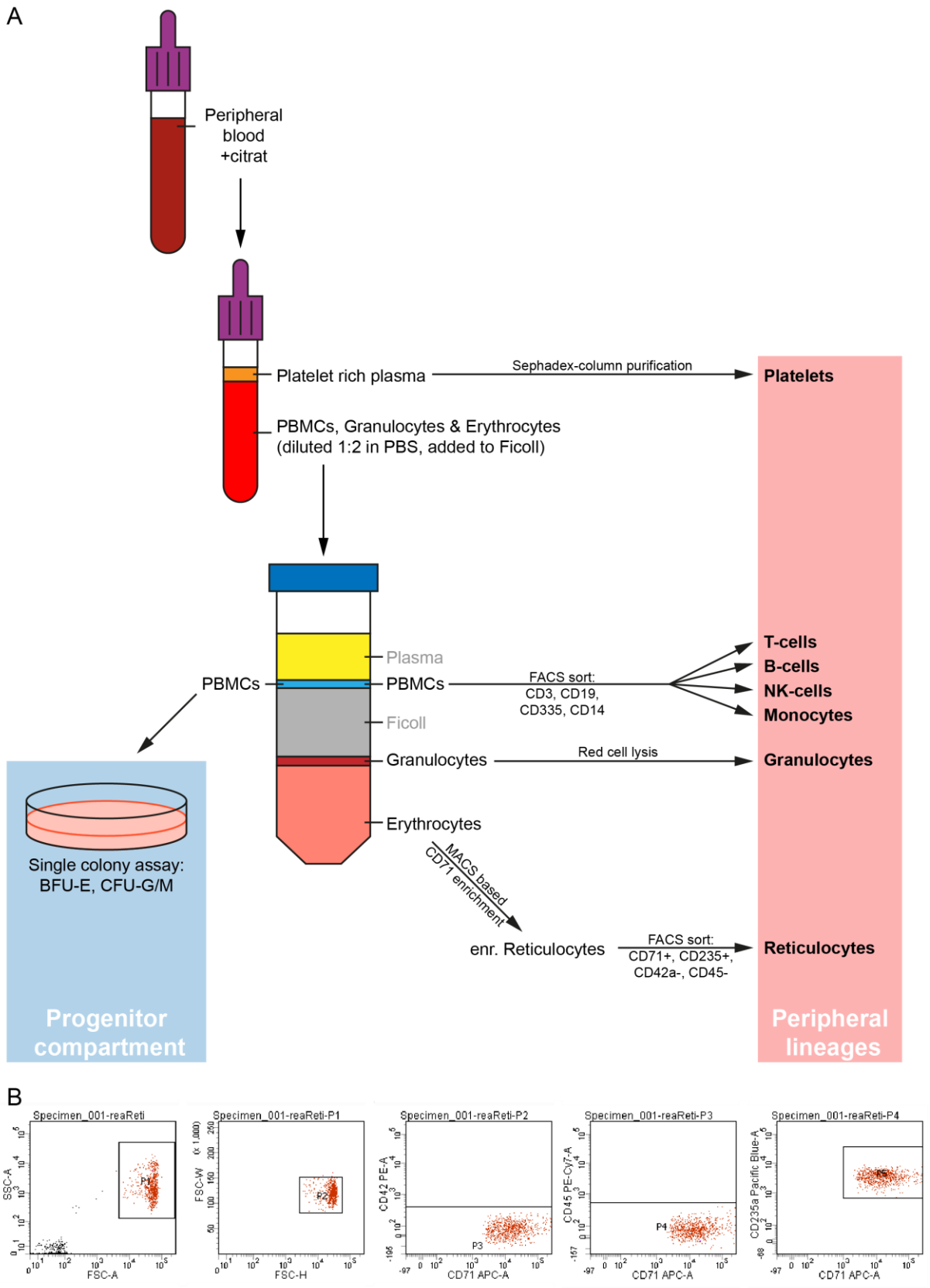


Figure 23 Purification of progenitors and peripheral blood lineages from individual blood samples

(A) Purification scheme (B) FACS analysis of sorted $CD71^{+}/CD235^{+}/CD42a^{-}/CD45^{-}$ reticulocytes

Due to the lack of DNA in platelets and reticulocytes, the driver mutation burden in all peripheral lineages was performed on cDNA using SNaPshot analysis. In granulocytes, the mutant allele burden was quantified in DNA and RNA to show that both, the wild type and the mutant allele, are expressed at comparable levels. To compare results from peripheral lineages of the blood with the progenitor compartment, we also seeded (non-enriched) PBMCs in methylcellulose for colony forming assays and subsequently picked single colonies for driver mutation analysis. As the enrichment of reticulocytes required an extended work up protocol, the complete analysis was limited to freshly donated samples. For this analysis, we selected MPN patients with a *JAK2* mutant allele burden below 20 % in their most recent sample. In total, 13 ET patients and eight PV patients mutated for *JAK2-V617F* were included in this study. We also included two PMF patients with *JAK2-V617F* allele burden <20 % and combined them in the group of ET patients. In addition, two PV patients (P021 and P284) were analyzed, who had a *JAK2-exon12* mutation and a mutant allele burden below 20 %. One of the *JAK2-exon12* PV patients, P021, was also mutated for *JAK2-V617F* and was analyzed for both driver mutations.

The *JAK2-V617F* allele burden in platelets of ET patients showed a significant increase compared to granulocytes. The significant difference in driver mutation burden between granulocytes and platelets confirms the data previously shown by Bellosillo in 2007[(115)]. Interestingly, the reticulocytes of the same ET patients showed a significantly higher allele burden compared to granulocytes. In eight of 13 (62 %) patients, the *JAK2* mutant allele burden in reticulocytes reached >20 %. The two PMF patients followed the trend of ET patients (dotted lines in Figure 24A). A significant increase in allele burden in reticulocytes compared to granulocytes was also observed in nine of 10 (90 %) PV patients (Figure 24B). In PV patients, the median increase from granulocytes to reticulocytes is 3.2 fold and is slightly more pronounced than in ET patients, where a 2.2 fold median increase was measured. Similar values were obtained for the fold increase of allele burden in granulocytes compared to platelets (Figure 24C). For two ET and two PV patients, RNA of granulocytes and platelets was available from previous samples. We measured the mutant allele burden only in these two lineages, as reticulocytes were not collected at earlier time points and it was not possible to enrich them at sufficient numbers from PBMCs of previously obtained samples. The results from these patients show that the observed difference in allele burden of platelets and granulocytes is stable over time (Figure 24D-F).

Patient P021, who carries both, a *JAK2-E542-N543-del* mutation in exon 12 and a *JAK2-V617F* mutation, was reported in a previous study of our lab[(77)]. We showed that both mutations are present in an individual, non-overlapping clone. Interestingly, the allele burden of both mutations, *JAK2-exon12* and *JAK2-V617F*, is significantly increased in platelets compared to granulocytes. The allele burden in the platelets of this patient are 25 % and 35 %

for JAK2-E542-N543-del and *JAK2-V617F*. This indicates that the vast majority of platelets originated from a megakaryocyte, which carried at least one mutant allele of *JAK2*. Furthermore, the data suggests the presence of a homozygous clone and implies that there is only a small population of wild type platelets (Figure 24G).

A case of a patient with diagnosis of PV (P192) caught our attention, as he was the only patient with interferon-alpha treatment in this study. The *JAK2-V617F* allele burden in reticulocytes of this patient was distinct from measurements in granulocytes and platelets. We analyzed previous samples of the patient and found that initially, the *JAK2* mutant allele burden was similar in platelets (18 %) and granulocytes (15 %). Within two years from diagnosis, the allele burden in platelets increased to 28 % and remained stable for the following five years. During the same period of seven years, the mutant allele burden in granulocytes steadily increased to 25 %. At this time point, interferon alpha treatment was started and led to a constant reduction of mutant *JAK2* allele burden in both, granulocytes and platelets. The patient was under treatment of interferon since three years, when the sample was drawn for the analysis of all lineages, including reticulocytes, platelets and granulocytes. The analysis of this sample revealed a mutant allele burden of 8 % in granulocytes, 12 % in platelets and 33 % in reticulocytes (Figure 24H).

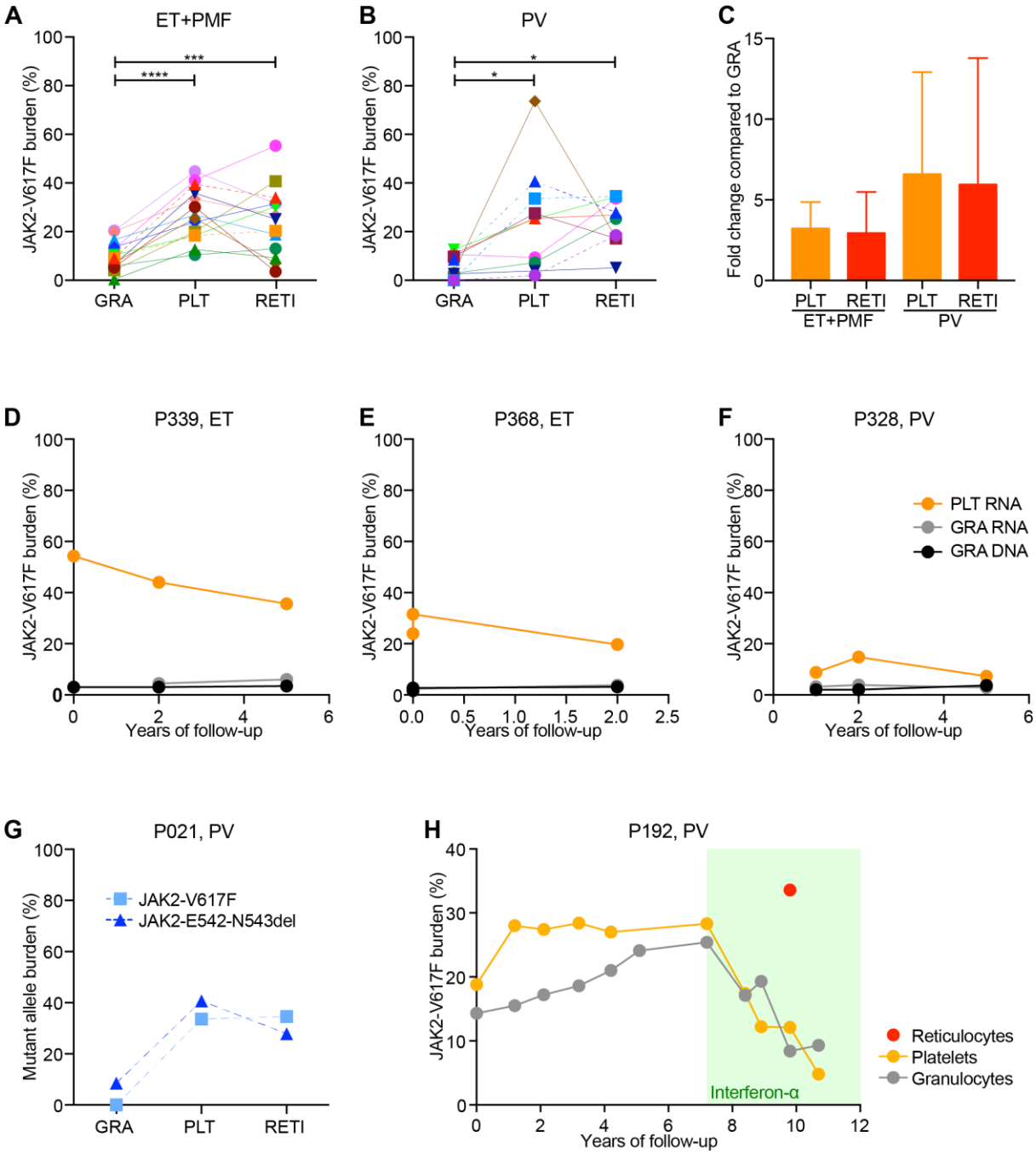


Figure 24 Comprehensive JAK2 analysis in progenitors and peripheral blood lineages

(A) Mutant JAK2 allele burden in myeloid lineages of individual patients diagnosed for ET and PMF. Dotted lines represent the two PMF cases. (B) Mutant JAK2 allele burden in myeloid lineages of individual patients diagnosed for PV. Dotted lines represent patients with *JAK2-exon12* mutation. Dark blue dotted line: P021 *JAK2-V617F*, light blue dotted line: P021 JAK2-exon12, purple dotted line: P284. (C) Average fold difference of mutant JAK2 burden in reticulocytes and platelets compared to granulocytes. (D-F) Longitudinal measurements of JAK2 mutant allele burden in platelets and granulocytes of three individual patients. (G) Representation of allele burden of both driver mutations of P021 in granulocytes, platelets and reticulocytes. (H) Detailed analysis of samples from patient P192, who received interferon-α treatment

Our quantification of the mutant *JAK2* allele in peripheral blood shows a significantly higher allele burden in platelets and red cells compared to granulocytes in a subset of patients. To test if this difference can be found also in the progenitor compartment, we seeded PBMCs in methylcellulose assays and analyzed single colonies for the presence of the driver mutation. PBMC samples were available for five patients with ET, both PMF patients, four PV patients with *JAK2-V617F* mutation and both PV patients with *JAK2-exon12* mutation. Colonies of platelet progenitors, CFU-Meg, required to be fixed and stained for morphologic identification. As this pre-treatment doesn't allow picking and genotyping of CFU-Meg, the platelet progenitors were excluded from single colony analysis. Morphologically distinct colony types, BFU-E, CFU-G and CFU-M were analyzed as progenitors of red cells, granulocytes and monocytes, respectively.

In total, five times more BFU-E colonies than CFU-G colonies were picked from methylcellulose plates. CFU-M colonies represent 5 % of the analyzed colonies and in some patients only a single colony of this phenotype could be found. In the colonies from three of five (60 %) ET patients and three of six (50%) PV patients we found homozygous clones in at least one lineage (Figure 25A+B). Due to the low frequency of double mutant colonies, the homozygous clone of individual patients consisted of only a single colony in two ET and two PV patients. No homozygous colony was found in the two PMF patients (Figure 25A). To test if the progenitors of red cells, granulocytes or monocytes contain a larger fraction of the mutant clone, we pooled the information from all 11 patients for each type of colonies (Figure 25C). In the combined analysis, we found a similar distribution of heterozygous colonies in BFU-E, CFU-G and CFU-M, 5.2 %, 4. 8% and 4.0 % respectively. A homozygous genotype was found in 0.6 % of both, BFU-E and CFU-G colonies. In total, 83 CFU-M colonies were analyzed from 11 patients and none was found to be mutated in both *JAK2* alleles (Figure 25C). In summary, our results show that at the progenitor level, the driver mutation is equally distributed among the analyzed lineage progenitors. At the progenitor level, the red cell lineage doesn't show a larger MPN clone, as found in the mature cells.

When comparing the *JAK2* mutant allele burden in progenitors with the allele burden in their mature counterparts, PV patients showed a significant increase for the red cell lineage, but not the granulocytes (Figure 25D+E). In ET patients, we observed a significant proliferation of the mutant clone from BFU-E progenitor level to reticulocytes. The mutant clone showed a non-significant trend towards expansion during the differentiation from CFU-G to mature granulocytes of individual ET patients (Figure 25F). In contrast to PV patients, the clonal expansion in ET patients was not significantly different between the red cell and the granulocytic lineage (Figure 25G).

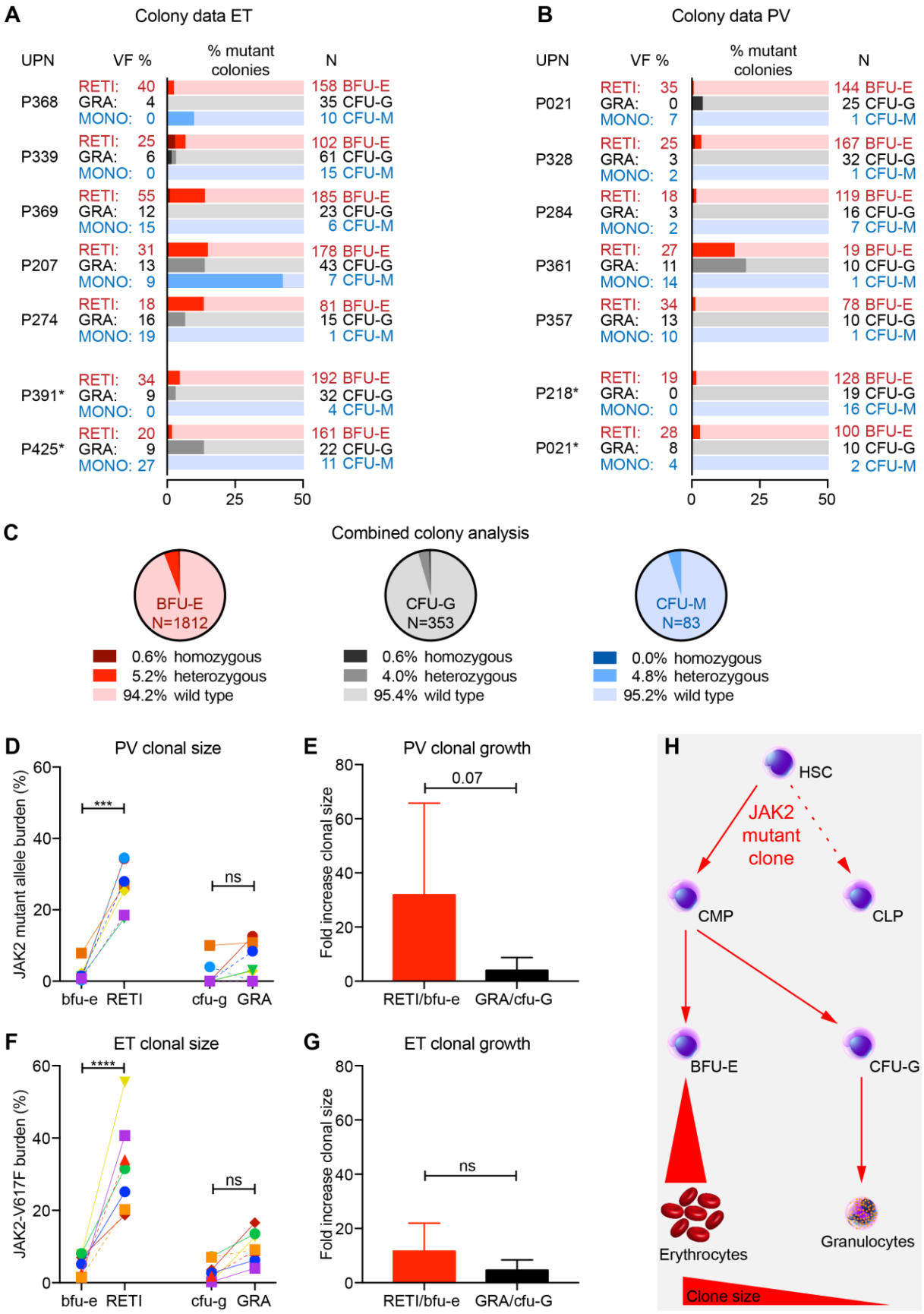


Figure 25 Analysis of JAK2 driver mutations in progenitors of MPN patients

Figure legend from previous page: Analysis of JAK2 genotype in BFU-E, CFU-G and CFU-M colonies in individual patients diagnosed for ET and PMF (A) and PV (B). (C) Combined analysis of JAK2 burden by colony type. (D) Allele burden in progenitors compared to mature cells individual PV patients. (E) Fold clonal expansion in red cell and granulocyte lineage of PV patients. (F) Allele burden in progenitors compared to mature cells individual ET and PMF patients (G) Fold clonal expansion in red cell and granulocyte lineage of ET and PMF patients. (H) Schematic of clonal expansion in MPN during hematopoiesis. BFU-E: blast-forming unit erythroid, CFU-G: colony-forming unit granulocyte, CFU-M: colony-forming unit macrophage

In summary, 12 of 13 (92 %) ET patients and nine of 10 (90 %) of PV patients show a significantly higher allele burden in platelets and red cells compared to granulocytes. The allele burden in these two lineages reached >20 % in nine of 13 (69 %) ET patients and six of 10 (60 %) PV patients. Our results indicate that the majority of platelets and red cells in these patients are part of the *JAK2* driven MPN clone. Therefore, the allele burden in granulocytes as an indicator of clonal size might lead to an underestimation of mutant cells in the periphery. Furthermore, our results might serve as a possible explanation for MPN phenotype in patients with an allele burden in granulocytes at the lower end of detection range of today's standard assays.

MPN are thought to arise from hematopoietic stem cells or the early progenitor level. Previous studies[116] of our lab have shown that only a minority of the *JAK2* mutant MPN clone expands to leukocyte lineages. Data from this analysis shows that the MPN clone equally proliferates within the committed progenitors of the myeloid lineages. Based on the data from colonies, we suggest that the mutant clone expands at very late stages of the red cell development. In contrast, during differentiation of granulocytes, the size of the *JAK2* mutant clone only increases moderately (Figure 25H). It would be interesting to expand our analysis to other progenitor stages and the platelet lineage in future, to learn more about the clonal dynamics in MPN.

3.4 Modulation of the MPN phenotype by additional somatic mutations

3.4.1 Meta-analyses to identify effects of somatic mutations on the MPN phenotype

In parallel to clonal dynamics of MPN driver mutations, we examined how additional somatic mutations may modulate the phenotype of MPN patients. For the most frequent somatic mutations, mouse models have been established to uncover the molecular mechanism of these phenotype modulations. To identify consequences of less frequent mutations, large patient cohorts are required. However, large cohorts are rare due to the low prevalence[117] of MPN.

In our study center in Basel, we have access to 197 sporadic MPN patients with a NGS profile. Since 2014, samples from 82 additional patients were analyzed using the same NGS technology[108]. In total, the Basel cohort of sporadic MPN cohort contains NGS profiles of 279 patients. Furthermore, we included data of 127 patients from the Belgrade sporadic MPN cohort, which we generated by applying our published workflow. Finally, we added the online available data of a MPN cohort of 150 patients, which was analyzed and published by Nangalia et al[35]. In total, 556 patients were summarized in a virtual cohort (Figure 26A). The MPN phenotypes showed a comparable distribution in all three individual cohorts. For the third cohort, we had to consider, that the downloaded data was generated by exome sequencing and not by targeted NGS like the Basel and Belgrade cohorts. Hence, we restricted the analysis to a defined a set of 21 genes, which previously have been reported to be frequently mutated in MPN (Table 9). The exome sequencing used in the Nangalia study covered all of these genes. The gene enrichment panel of the published 197 MPN patients included 19 genes of the 21 reported genes. Later on, our enrichment panel was updated and included all of the genes listed in Table 9. The updated enrichment panel was used for the analysis of 82 new patients of the Basel cohort and 127 patients of the Belgrade cohort.

AEBP2*	DNMT3A	IDH1	KRAS	NRAS	SF3B1	TET2
ASXL1	EED	IDH2*	NF1	RBBP4	SH2B3	TP53
CBL	EZH2	JARID2	NFE2	RBBP7	SUZ12	U2AF1*

Table 9 Frequently mutated genes in MPN

Red: Genes of the PRC2 complex, Green: RAS GTPases summarized in later analysis, Black: genes analyzed individually, asterix: genes, which were not included in the analysis of the first 197 Basel sporadic MPN patients.

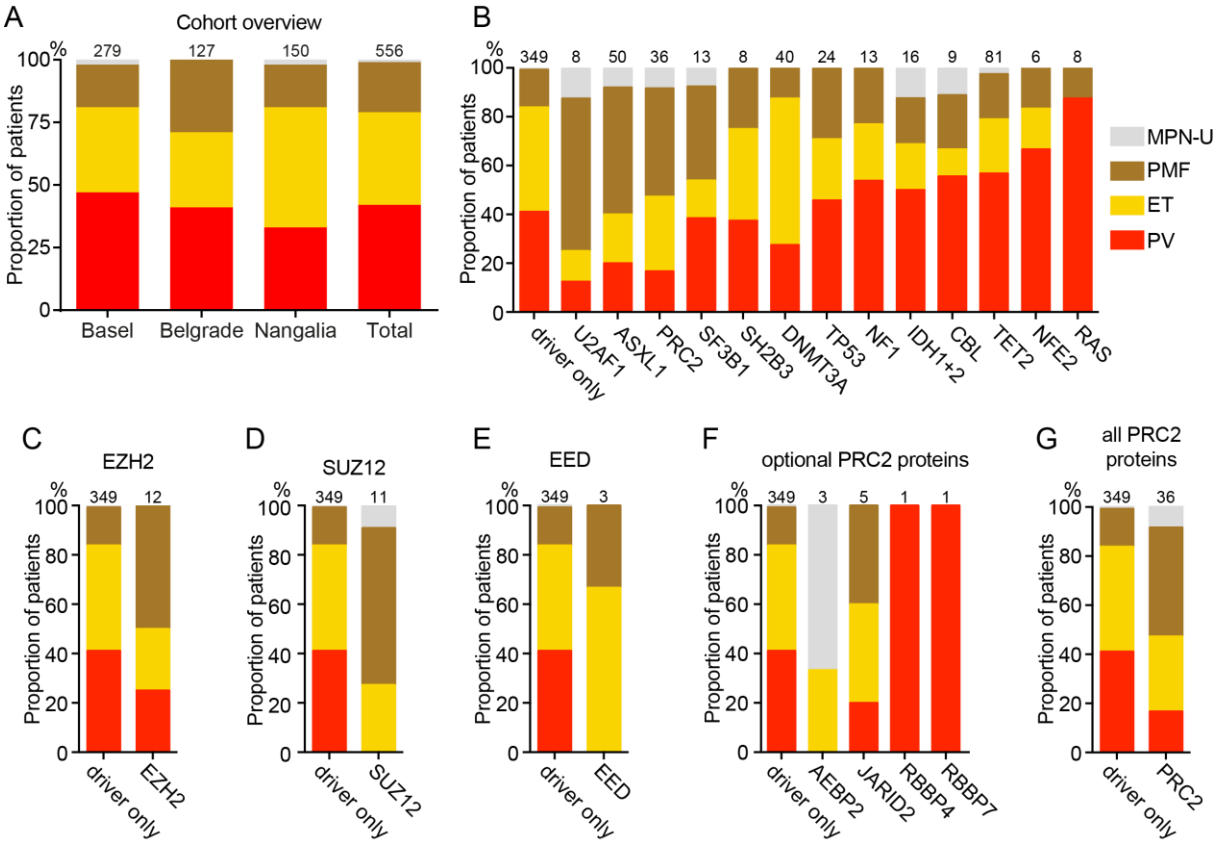


Figure 26 Genotype-phenotype correlations in MPN

(A) Overview of the phenotypes observed in all three individual cohorts and the summarized virtual MPN cohort. (B) Phenotypes of patients, who carry mutations in one of the frequently mutated genes. (C-F) Individual plots of patients mutated for PRC2 complex genes: *EZH2*, *SUZ12*, *EED*, *AEBP2*, *JARID*, *RBBP4* and *RBBP7*. (G) Summary of patients carrying a mutation in at least one PRC2 gene. (A-G) Numbers on top of bars: Number of patients with mutations in the specified genes, driver only: describes patients without any somatic mutation in the predefined gene set.

Subsequently, we split the combined cohort into two groups: one group contained patients, who carry mutations in at least one of the frequently mutated genes. The other group contained patients, who did not show any mutation in the pre-defined group of genes. The great majority of these patients carry only a driver mutation and represent the control group for a normal distribution of MPN phenotypes. Then, we compared the frequency of MPN phenotypes in patients with mutations in at least one of the selected genes, with the control group. Here we assume that a different distribution of MPN phenotypes in patients with mutations in a specific gene would indicate an impact of this gene on generating the MPN phenotype.

Mutations in *U2AF1*, *ASXL1*, PRC2 genes and *SF3B1* were found in eight, 50, 36 and 13 patients respectively. The majority of patients with mutations in *U2AF1* and *ASXl1* was diagnosed for PMF. In addition, five of 13 (38%) patients with mutations in *SF3B1* were diagnosed with PMF. In the control group, 15% of patients were diagnosed with PMF. This data resembles results of previous studies in myeloid malignancies, which were restricted to the analysis of a single gene or smaller gene sets[96,118]. The combination of multiple MPN cohorts uncovered that mutations in genes of the PRC2 complex[119] (*EZH2*, *SUZ12*, *EED*, *AEBP2*, *JARID*, *RBBP4* and *RBBP7*) are primarily found in patients with diagnosis of PMF (Figure 26B). In the past, only mutations in *EZH2* were linked with ET or PMF phenotype and adverse survival[120,121]. Mutations in other PRC2 complex members were only described for *SUZ12*[99] so far. As mutations in other PRC2 genes are less frequently mutated, a correlation with ET and PMF phenotype might so far have been missed due to too low cohort sizes. In our cohort, patients with mutations in *EZH2* (Figure 26C) and *SUZ12* (Figure 26D) are diagnosed with PMF more frequently than patients in the control group. Mutations in *EED* and *JARID* also seem to occur more often in ET or PMF patients. Two mutations in *RBBP4* and *RBBP7* were found in the same triple negative PV patient. Mutations in *AEBP2* were found in one ET patient and one patient with unclassifiable MPN. The patient with diagnosis of MPN-U carries one *AEBP2-A4V* and one frame shift mutation at codon F398 (Figure 26F). In summary, when patients with mutations in PRC2 genes were grouped, about 40% of the cases were diagnosed with PMF, compared to 15% in the group of patients without mutations (Figure 26G). However, our combined cohort only shows digital numbers of patients with mutations in *EED* and PRC2 adapter proteins (AEBP2, JARID, RBBP4 and RBBP7, Figure 26E and F). Therefore, even larger patient cohorts will be required to validate these observations.

Mutations in *SH2B3* were equally distributed among the MPN phenotypes and only mutations in *DNMT3A* mutations were predominantly found in ET patients. *TP53* mutations were previously described to correlate with MPN disease progression and adverse survival[122]. However, patients with mutations in *TP53* show a nearly even distribution of MPN diagnoses with only a small increase of PMF phenotype.

Patients with mutations in *KRAS*, *NRAS*, *NFE2*, *TET2*, *CBL*, *NF1* and *IDH1* or *IDH2* were more often diagnosed as PV than the control cohort of non-mutated patients (Figure 26B). KRAS and NRAS were combined due to similar function and mutations in these genes were mainly found in patients diagnosed for PV. Both genes belong to the group of small GTPases, which regulate proliferation and survival pathways. RAS proteins bind GTP to get activated. GTPase activating proteins (GAPs) bind to RAS proteins to catalyze the hydrolysis of GTP to GDP, which inactivates the RAS proteins. The majority (75%) of *K/NRAS* mutations in our cohort were activating mutations of glycine 12[123]. Other activating mutations in codons 13 or 61 were not observed in our virtual cohort[124]. Mutations in codon 12 of RAS GTPases result in elevated activity, as the mutant proteins do not form the complex with GAPs. Due to the lack of GAP interaction and the low intrinsic activity of RAS proteins to hydrolyze GTP, the mutant RAS proteins remain activated once bound to GTP. This constitutive activation leads to elevated signaling of the MAPK and mTOR pathways, which is inducing proliferation and survival. Activating *RAS* mutations have been described in numerous tumor types and are most common in pancreas cancer (up to 63%[124]). In MPN, *RAS* mutations are found in less than 2% of the cases.

Mutations in *NFE2* have been described at similar frequency in MPN cohorts as mutations in RAS proteins[108,125]. NFE2 is a transcription factor of the human beta globulin and essential for its expression *in vitro*[126]. The most frequent mutations in *NFE2* are frame shift mutations, which lead to premature stop of the protein. The truncated proteins lack the leucine zipper domain of NFE2, which is required for DNA interactions of NFE2 dimers. Mouse studies showed that disruption in the *NFE2* gene prevents shedding of platelets by mature megakaryocytes[127,128]. From three patients with *NFE2* frame shift mutations clinical data at date of diagnosis were available. Contrary to expectations, all patients displayed thrombocyte counts of >700 x 10^9/L, indicating elevated production. Presumably, the constitutive JAK2 signaling in MPN patients surpasses the lack of functional NFE2 protein and results in proliferation of both, the red cell and the megakaryocyte-platelet lineage.

3.4.2 Somatic EZH2 mutations promote megakaryopoiesis, resulting in ET and PMF phenotypes

Among the frequently mutated PRC2 genes in MPN, *EZH2* was reported first[80]. It is the enzymatic core component of the PRC2 and mutations in *EZH2* occur in approximately 3% of the patients in the Basel sporadic MPN cohort[108]. To examine a possible synergistic effect of mutations in *EZH2* on the initiation or progression of MPN, our group generated a mouse model combining the *JAK2-V617F* mutation with a loss-of-function in EZH2. Dr. Takafumi Shimizu crossed mice with a conditional knockout allele of *EZH2*[129] with inducible *JAK2-V617F* mice[64]. The induction of *JAK2-V617F* in mice wild type for *EZH2* results in ET or PV phenotypes, depending on expression levels of *JAK2-V617F* in hematopoietic cells[64]. Loss of *EZH2* in mice with wild type *JAK2* did not lead to significant changes in blood counts or prognosis. Heterozygous loss of *EZH2* in *JAK2-V617F* expressing mice enhances thrombocytosis and neutrophilia and accelerates the transition from PV to PMF. The excision of both *EZH2* alleles in *JAK2-V617F* mice results in an even more pronounced phenotype, leading to PMF, without a preceding PV phase. Limiting dilution transplantation experiments show that loss of *EZH2* on a JAK2-V167F background leads to an expansion of the HSC and progenitor pool, resulting in increased megakaryopoiesis over erythropoiesis[130].

To determine, how the loss of *EZH2* modulates the MPN phenotype, we analyzed the expression pattern of FACS sorted LT-HSC and MEP cells by RNAseq. Principle component analysis revealed that induction of *JAK2-V617F* has significant impact on expression patterns in LT-HSC, whereas homozygous loss of EZH2 mainly abrogates the expression patterns of MEP cells. In MEP cells, expression of *LIN28B* and its target genes, such as *HGMA2*, *IGF2BP3* and *PCOLC2*, were significantly increased upon *EZH2* deletion in mice[130]. The up-regulation of *HMGA2* in myeloid progenitors such as CMP and MEP has been linked in a recent study with increased megakaryopoiesis[131], confirming observations of thrombocytosis in our MPN mouse model. Furthermore, increased expression of LIN28B and HMGA2 has also been shown to increase the stem cell pool by inducing self-renewal in the HSC compartment[132].

To test if the observations from the mouse model resemble the effect of *EZH2* mutations in MPN patients, the expression of *HGMA2*, *IGF2BP3* and *PCOLC2* was measured in RNA from patient derived granulocytes (Figure 27A-C). In total, RNA was available from 21 MPN patients with a mutation in *EZH2*: Five patients were mutant for *EZH2* and an insertion-deletion mutation in *CALR-exon9*, 16 patients were mutant for *EZH2* and *JAK2-V617F*. One CALR mutant patient and four patients mutant for *JAK2-V617F* had a homozygous *EZH2* mutation. *EZH2* mutant patients showed higher *HMGA2* expression compared to their *EZH2* wild type counter parts. The expression level of *HMGA2* correlated with the mutant allele burden of

EZH2 in patients. Our data confirms that reduction of PRC2 activity by *EZH2* mutations leads to overexpression of the *LIN28B-HMGA2* axis in humans, as it has been shown in mice. In the studied mouse model, the combination of *JAK2-V617F* driver mutation and loss of *EZH2* function modulates the MPN phenotype towards megakaryopoiesis. Due to the similar effect on the transcription of key players of megakaryocyte differentiation and proliferation[130], we suggest a similar mode of action in human patients with *EZH2* mutations.

Interestingly, mutations in PRC2 genes other than *EZH2*, such as *SUZ12*, *EED* and *JARID*, frequently occur in patients with ET or PMF diagnosis (Figure 26C-F). The complex requires three core proteins EZH2, EED and SUZ12 to methylate the lysine 27 of histone 3. In addition to these core members, other proteins can associate with the complex, which results in modulated specificity and activity of the PRC2. These optional PRC2 members are AEBP2, JARID2, RBBP4 and RBBP7. AEBP2 and JARID2 are thought to associate with GC-rich promoter regions of PRC2 target genes. In addition, JARID binds to mono-ubiquitinylated lysine 119 of histone H2A[133,134]. RBBP4 and RBBP7 bind to unmodified residues at the N terminus of histones H3 and H4[119]. SUZ12 serves as an adapter between RBBP4/7, JARID and EZH2 and is thought to enhance the activity of EZH2 when bound to these optional complex members[135]. AEBP2 and JARID2 have been shown to associate with EED, which itself specifically binds to H3K27me3[136]. Like SUZ12, EED serves as a linker between EZH2 and optional adapter proteins[119] (Figure 7). In summary, current knowledge suggests a relationship between physical interaction of these proteins and the enzymatic activity of the PRC2 complex. Therefore, we hypothesize that mutations in other PRC2 genes result in similar phenotypes than loss-of-function mutations in *EZH2*. Hence, we scheduled experiments for the expression analysis of *LIN28B* target genes (*HGMA2*, *IGF2BP3* and *PCOLC2*) in patients with mutations in other genes of the PRC2 complex. The experiments were still running while the thesis was being written up.

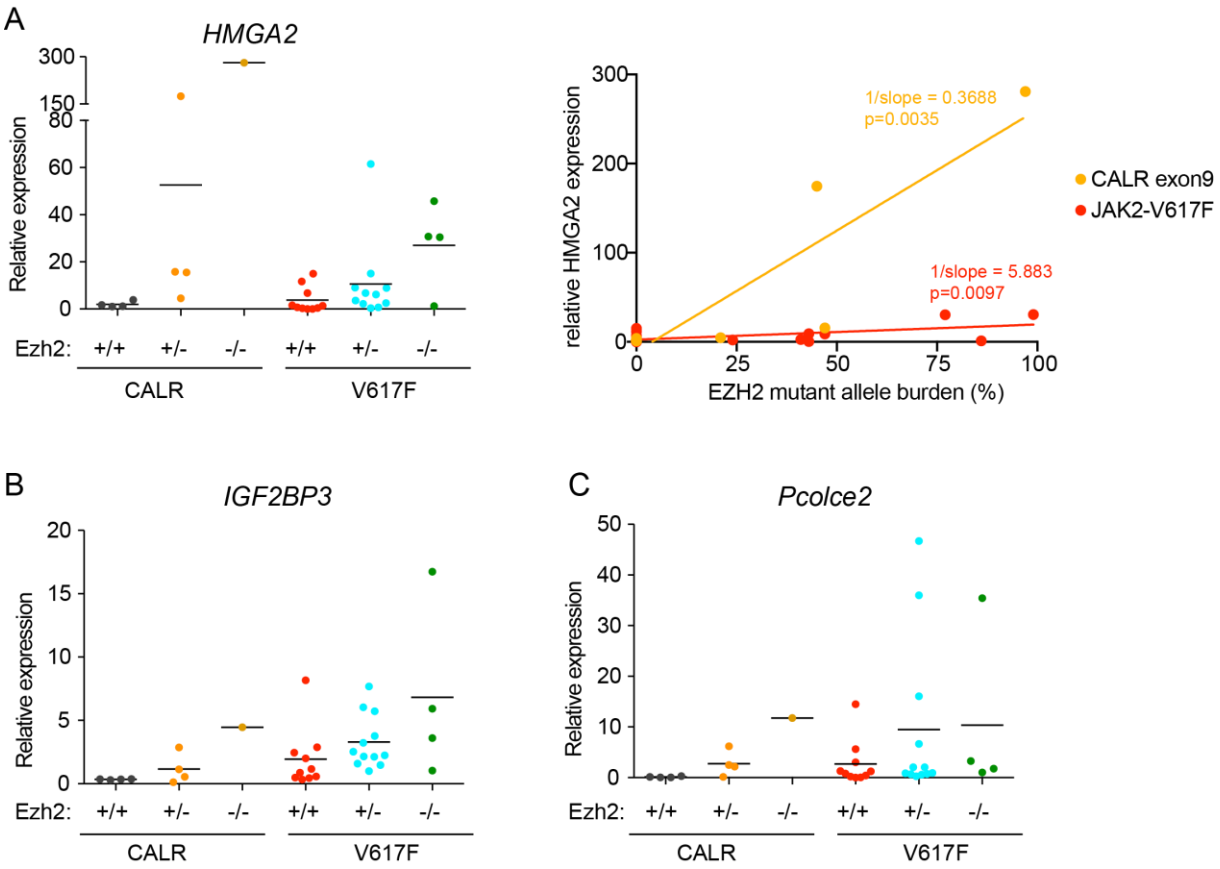


Figure 27 Loss of EZH2 in mice reflects the MPN in human patients

(A) Relative expression of HGMA2 (left panel) and correlation with the mutant allele burden of *EZH2* (right panel). (B) Relative expression of *IGF2BP3* and *PCOLC2* (C). All qPCR analysis were performed using RNA of granulocytes, each dot represents an individual MPN patient.

3.4.3 The type of mutation and additional alterations influence the effect of an individual somatic mutation on the MPN phenotype

Previously, we analyzed which genes are frequently mutated in a specific MPN phenotype. We also compared, if distinct mutations in an individual gene might affect the molecular biology of the MPN phenotype in a different way.

The effect of *IDH1* and *IDH2* gene mutations on the MPN phenotype seems to be the lowest in our preselected group of recurrently mutated genes (Figure 26B). In the group of patients with *IDH* mutations, 50% of the patients were diagnosed for PV, compared to 41% in their non-mutant counterparts (Figure 28A). Interestingly, the frequency of PV diagnosis in patients with *IDH* mutations seems to be dependent on the mutation type. The proportion of PV diagnoses increases when only so-called neomorphic *IDH* mutations are taken into account. These neomorphic mutations are alterations in *IDH1* or *IDH2*, which affect the residues *IDH1-R132*, *IDH2-R140* or *IDH2-R172* and alter the enzymatic activity of the proteins. Endogenous IDH1 and IDH2 proteins catalyze the conversion of isocitrate to α-ketoglutarate. These specific mutations are termed “neomorphic” as mutant proteins produce the novel metabolite 2-hydroxyglutarate instead of α-ketoglutarate[137,138]. These mutations have been shown to be present in 1 % to 2 % of MPN cases[79,139,140], which is comparable to our results. In the combined cohort of 556 patients, eight (1.4%) patients were mutant for either *IDH1-R132* or *IDH2-R140* and PV was diagnosed in six of these eight (75 %) cases. *IDH2-R172* mutations were not found in this virtual cohort. From five patients with other SNVs than the described neomorphic mutations, two patients were diagnosed for PV and PMF and one for ET. In the DNA of three patients, frameshift mutations led to a premature stop codon. These patients were diagnosed with ET, PMF and MPN-U (Figure 28A).

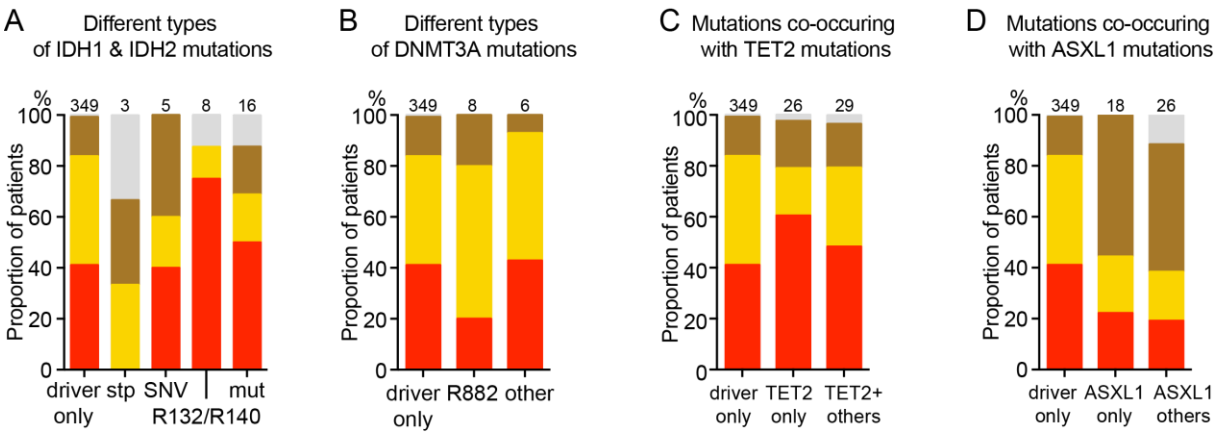
Similar to patients with mutations in *IDH1* and *IDH2*, patients with mutations in *DNMT3A* showed a different ratio of diagnoses based on the observed mutation type. Mutations in *DNMT3A* are most frequently described in the normal aging population[141–143]. In addition, *DNMT3A*, together with *TET2* and *ASXL1*, belongs to the most often mutated genes in MPN[35,108]. The most frequent somatic mutations within *DNMT3A* affect the arginine 882 and convert it to either cysteine or histidine[74,75]. This specific type of mutations has been shown to result in loss-of-function of the mutant protein. Secondary, the mutant protein blocks wild type DNMT3A proteins from forming active tetramers[84]. Mutations in *DNMT3A-R882* therefore result in focal hypomethylation at the DNA. A recent study also showed, that mutant DNMT3A-R882, but not wild type DNMT3A physically interacts with core proteins of the PRC1 complex, such as BMI1, CBX7 and RING1B[144]. In the same study, the interaction of mutant *DNMT3A-R882* correlated with loss of H2K27me3. Activity of the PRC1 complex has been

shown to be required for the full functionality of the PRC2 complex[91]. These results indicate that *DNMT3A-R882*, in addition to its defect in DNA methylation, leads to reduction of PRC2 activity by removing PRC1 from target regions. Interestingly, patients with *DNMT3A-R882* mutations are more frequently diagnosed with ET than patients without somatic mutations (Figure 28B). Other mutations in *DNMT3A*, such as frame shifts, stop mutations or SNVs in other positions of the gene, also show increased frequency of ET diagnoses compared to non-mutant patients, but the effect is less distinct than compared to the group of patients with *DNMT3A-R882* mutations.

Furthermore, we tested if the effect of mutations in a specific gene on the MPN phenotype can be neutralized by co-occurring mutations in other genes. When *TET2* is the only mutated gene, patients show an increased frequency of PV diagnosis compared to patients without mutations. This effect is reversed in the group of patients with other somatic mutations in addition to *TET2* mutations (Figure 28C). A comparable analysis with patients mutated for *ASXL1* showed, that this observation might be gene dependent. Patients who only have somatic mutations in *ASXL1* as well as patients with mutations in *ASXL1* and mutations in other genes show an elevated frequency of PMF and MPN-U diagnosis (Figure 28D). This observation might indicate that the pro-PMF effect of mutant ASXL1 protein might be superior compared to effects from other gene mutations.

When effects of somatic mutations might be overwritten by somatic mutations in other genes, the impact of mutations in an individual gene will mainly be apparent in patients with any other somatic mutation. However, only approximately one third of the MPN patients carry somatic mutations additional to driver mutations. Further, reducing analysis to patients with mutations in a specific gene might result in too low numbers for drawing conclusions about phenotypic effects. Therefore, it might be necessary to analyze all patients of the combined cohort simultaneously using mathematical modeling. This technology assigns a hypothetical phenotypic effect to each mutated gene and validates these effects by successive approximation and comparing the diagnosis and the somatic mutation status of all patients. As an example, we manually analyzed all patients with mutations in CBL to estimate a phenotypic effect of mutations in this gene. All patients with CBL mutations and *JAK2-V617F* were diagnosed for PV. Further, CBL mutations seem not to reverse the hypothetical pro-PV effect in patients with additional mutations in *TET2* and *RAS*. Two patients with co-occurring SF3B1 mutations are both diagnosed with PMF, suggesting a stronger potential pro-PMF effect of *SF3B1* than the effect of *CBL* mutations towards PV. A patient with a deletion in *DNMT3A* and one patient with a *SUZ12* missense mutation additional to *CBL* were diagnosed with ET and MPN-U, respectively (Figure 28E). In summary, we suggest a mild pro-PV effect for mutations in *CBL*, as patients with only *CBL* mutations are all diagnosed with PV and additional somatic mutations seem to be able to overwrite this effect.

In summary, this meta-analysis indicates that the phenotypic effect of a mutation might not only depend on which gene is mutated. In addition, it may play a role where the specific mutation is located within the gene and which other gene mutations are present in the same patient.



E

Patient-ID	Driver mutation	+	Additional somatic mutation(s)			=	MPN diagnosis
Vla007	JAK2-V617F	+	CBL			=	PV
Vla024	JAK2-V617F	+	CBL			=	PV
Vla097	JAK2-V617F	+	CBL			=	PV
P255	JAK2-V617F	+	RAS	+	CBL	=	PV
P209	JAK2-V617F	+	TET2	+	CBL	=	PV
Vla088	tripple neg.	+	SF3B1	+	CBL	=	PMF
P253	JAK2-V617F	+	SF3B1	+	CBL	=	PMF
PD6646	JAK2-V617F	+	DNMT3A	+	CBL	=	ET
P378	JAK2-V617F	+	SUZ12	+	CBL	=	MPN-U

Figure 28 Effect of individual mutations on MPN phenotypes

(A) Patients with different types of mutations in *IDH1* o *IDH2*. stp: frame shift or nonsense mutations leading to premature stop codon, SNV: single nucleotide variants leading to missense mutations (excluding previously described neomorphic mutations), R132/R140: summarizes previously reported neomorphic mutations in *IDH1-R132* or *IDH2-R140* (B) Patients with different types of mutations in *DNMT3A*. R882: *DNMT3A-R882C* or *DNMT3A-R882H* mutations, other: missense, frame shift or nonsense mutations not affecting *DNMT3A* codon 882. (C) Comparison of patients with a mutation in *TET2* only and patients with mutations in *TET2* and additional somatic mutations in genes of the selected gene set. (D) Comparison of patients with a mutation in *ASXL1* only and patients with mutations in *ASXL1* and additional somatic mutations in genes of the selected gene set. (A-D) Numbers on top of bars: Number of patients mutated in the specified genes, driver only: describes patients without any mutation in the predefined gene set. (E) Detailed representation of genetic alterations in patients with mutations in *CBL*. The gene names are colored according to the potential phenotypic effect when affected by mutations.

4 Discussion

4.1 Mutational landscape of MPN

In this thesis, we were using NGS technologies to gain insight in the genomic landscape of MPN, clonal evolution, phenotype development and correlation with clinical outcomes. In the first part of the study, we analyzed 197 patients using a targeted NGS approach and analyzed the coding region of 104 genes in parallel. The most frequent mutations are driver mutations in JAK2 (*JAK2-V617F*, 69%) and *CALR* (15%). We also found *JAK2-exon12* mutations in seven of 197 (3%) cases and MPL-W515 mutations in three of 197 (1.5%) patients. In addition to the driver mutations, we found 108 somatic mutations in 73 patients.

The most frequently mutated genes are *TET2*, *ASXL1*, *EZH2*, *DNMT3A* confirming the results of others[35,36]. At lower frequencies, we found mutations in *TP53*[45,122], *IDH1*[140], *KRAS/NRAS*[124], which were previously linked with progression. In the same year, two studies were published, which were focusing on finding somatic mutations in the normal aging population[142,143] and one study, which showed the mutational landscape in MDS[118]. In Figure 29 the most frequently mutated genes are listed from each of the publications. Interestingly, the mutation patterns are similar, but also reveal differences based on the hematologic status of the analyzed patients. The studies, which focused on the normal aging population, revealed that *DNMT3A*, *TET2* and *ASXL1* are the most commonly mutated genes. In addition, *JAK2* mutations were frequently found in healthy individuals. Mutations of the splicing machinery (*SF3B1*, *SRSF2*, *U2AF1*, and *ZRSR2*) were less frequently found in normal population. In the general populations, the mutation frequency correlated with age and was the highest in individuals with an age of 100 years or older.

In MPN, mutations of *JAK2* and *CALR* are the most frequent. This is not surprising, as mutations in these two genes are requirements for the diagnosis of MPN. However, after excluding driver mutations, the genes, which are most commonly mutated in MPN, are the same as in the aging population: *TET2*, *DNMT3A* and *ASXL1*. In addition to these three epigenetic modifier genes, also *EZH2* is frequently mutated. Mutations of genes involved in splicing were also found in MPN at a lower frequency. In MDS, genes of the splicing machinery and epigenetic modifiers are sharing the top ranks in the list of mutations. Therefore, the mutational landscape of MPN seems to be more similar to the normal aging population than to the pattern observed in MDS (Figure 29). The main differences between the healthy population and MPN patients are that all MPN patients carry driver mutations in the JAK/STAT pathway and somatic mutations, mainly affecting epigenetic modifier genes, at a higher frequency. MDS compared to MPN patients, suffer from more somatic mutations per patient, more frequently carry mutations in genes of the splicing machinery and less often in the JAK/STAT pathway.

MPN compared to MDS have a favorable prognosis[117,145] and this observation might be linked to the type of genes commonly mutated in these diseases. Potentially, mutations disrupting the splicing machinery result in a more severe defect, resulting in lower survival rates in MDS. However, MDS patients also have a higher number of somatic mutations per patient (median: 3 somatic mutations per patient[118]), than MPN patients (median: 1 somatic mutation per patient). Furthermore, we show in our study, that the number of mutations has an impact on survival in MPN patients (Figure 17).

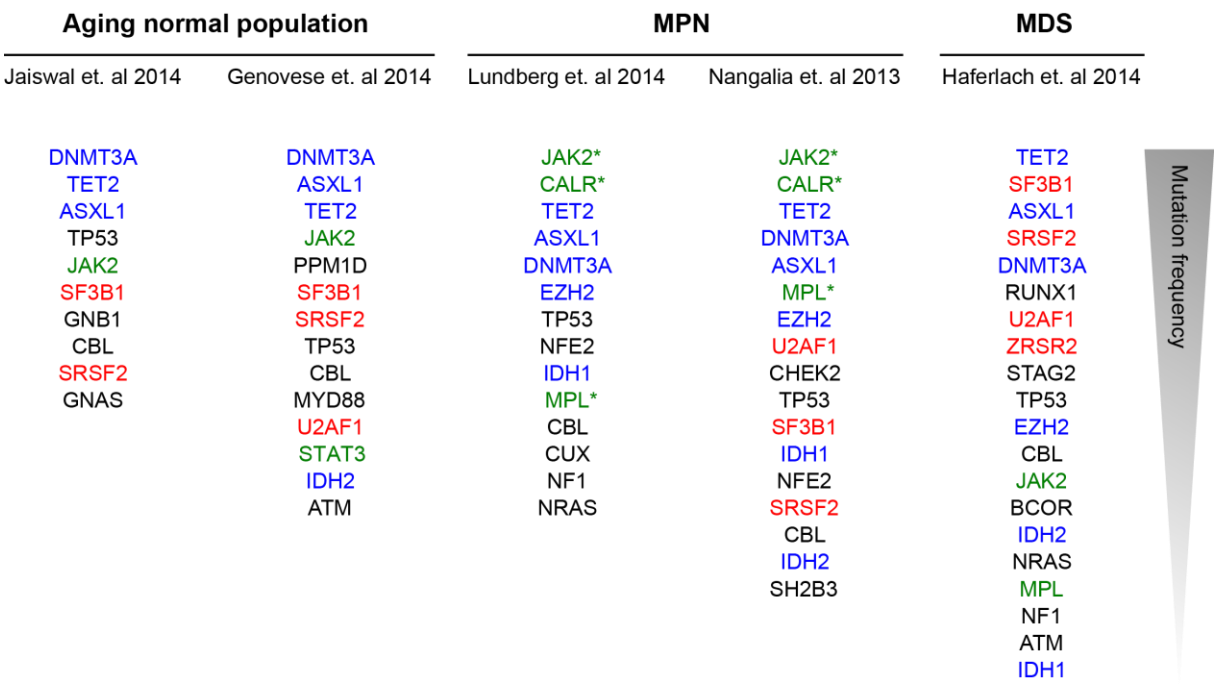


Figure 29 Mutational landscapes of chronic hematologic malignancies

Most frequently mutated genes for the indicated cohorts. The gene names are sorted in descending order by the number of mutated patients. Green: genes of the JAK/STAT pathway, Red: genes of the splicing machinery, blue: genes involved in epigenetic modification, black: genes involved in other pathways. Gene names with asterisk indicate driver mutations. Data extracted from: Jaiswal et. al[(142)], Genovese et. al[(143)], Lundberg et. al[(108)], Nangalia et. al[(35)], Haferlach et. al[(118)]

4.2 Patterns of mutational acquisition and clonal evolution

In our MPN cohort of 197 patients, we found that approximately a third of the patients carries somatic mutations additional to the driver mutation. In patients, we found that patterns of co-occurring mutations were very specific. Nearly each patient had an individual combination of mutations. Recurrent combinations were observed for mutations in *TET2*, *DNMT3A* or *ASXL1* in patients with a driver mutation in *JAK2* or *CALR* (Figure 15). However, this observation might be a consequence of the high frequency of these mutations in MPN patients.

In addition to the co-occurrence analysis, we also determined the clonal architecture of mutations in patients with two or more somatic alterations. The analysis showed that mutations in *TET2* and *DNMT3A* were preferentially acquired before *JAK2-V617F*. Mutations in *EZH2* and *IDH1* frequently were acquired as secondary or tertiary events within the *JAK2-V617F* clone. We also observed biclonal patterns in 30% of analyzed patients, which shows that hematologic cancers, similar to solid tumors, could consist of multiple branches of individual subclones (Figure 16B). Further, our study shows that single cell analysis is required for the correct determination of smaller subclones. Similarly, VAF as sole clonal marker is insufficient to resolve the clonal architecture of MPN. Interestingly, a recent study on patients with *JAK2-V617F* and *TET2* mutations revealed that the order of acquisition has an impact on the proliferation of individual progenitor compartments and clinical features of these double mutant patients[146]. Patients with a *TET2* mutation acquired before *JAK2-V617F* presented at older age of diagnosis, an enlarged CMP pool and a lower risk of thrombosis. In turn, *JAK2* first patients were younger at time of diagnosis and showed an expanded MEP and an increased risk of thrombosis[146]. These findings show that in patients with multiple somatic mutations all subclones may individually contribute to the modulation of the MPN phenotype.

In our study, we combined the analyses of co-occurrence and the clonal architecture with the longitudinal sequencing of the somatic mutations in individual patients. The NGS screen for finding somatic mutations was applied on the most recent sample of each patient. In the study center Basel, blood samples from MPN patients are collected each year. For the patients who carry somatic mutations additional to the known driver mutations, we re-sequenced the somatic mutations in previous samples of the same patients. Surprisingly, 95% of the somatic mutations, which we found in the most recent sample, were also present in all previous samples. Only in two patients, additional mutations were acquired after diagnosis. These results show that the clonal architecture remained stable during disease progression (Figure 19). It remains unclear, if there was a temporary period of genomic instability in these patients. It is also unknown, how much time passes between the acquisition of the first somatic mutation and the diagnosis of MPN. Possibly, MPN patients undergo a “non-symptomatic” phase of disease during which more somatic mutations accumulate, until diagnosis of MPN. This

observation also raises the intriguing question, if there are specific risk factors, which result in the acquisition of somatic mutations in the hematopoietic system. These risk factors could potentially be environmental effects like chronic infections or diseases, which induce cycling and accelerate aging of the hematopoietic stem cell and progenitor pool. Continuous proliferation and stress might leave somatic mutations as genetic scars[147].

In summary, our data confirms the clonal expansion model[141], which hypothesizes the acquisition of pre-leukemic somatic mutations. These pre-leukemic hits are not sufficient to result in a malignant phenotype, as shown by mouse models[148], but increase the risk of developing a hematologic cancer[142,143]. In the Basel sporadic MPN cohort only two patients acquired a somatic mutation post diagnosis, suggesting that the patients undergo a symptom-free period. In this period, a clone with MPN driver mutations may acquire more somatic mutations or gradually expand until the patient gets diagnosed (Figure 30).

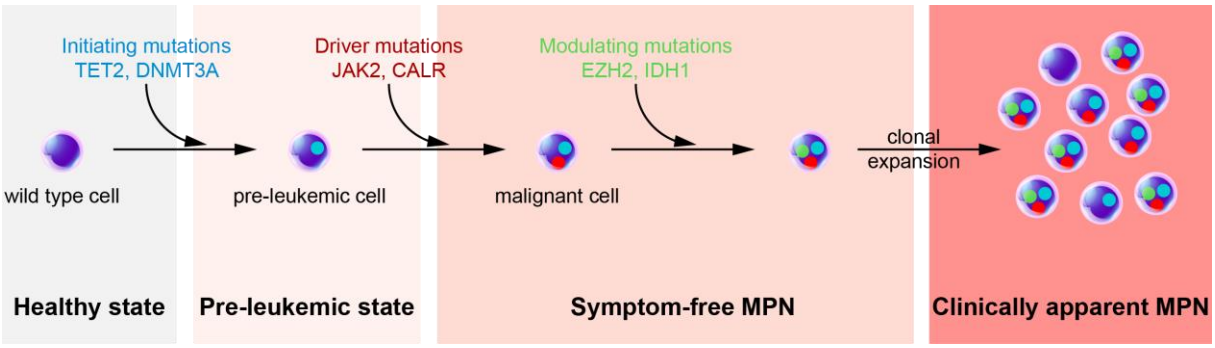


Figure 30 Molecular development to clinically apparent MPN

Model of clonal expansion. Initiating mutations (bright blue circles) induce the acquisition of driver mutations (red circles) and subsequent modulating mutations.

4.3 Late clonal expansion in patients with a low mutant JAK2 burden

Following the model of clonal expansion, one would expect that at diagnosis, MPN patients present with considerably large clones positive for driver mutations. Therefore, the existence of patients who represent with a very low *JAK2-V617F* allele burden, are matter of discussion. With the current PCR based methods, which identify small *JAK2* mutant clones down to 1 % allele burden or smaller, the question was rising, which allele burden is sufficient to classify a patient as mutated for JAK2.

In our set of patients with a *JAK2* mutant allele burden <20 % in granulocytes, nearly all patients showed an increased burden in platelets and/or reticulocytes, independent of the MPN phenotype. In a previous study, ET patients showed a significantly increased JAK-V617F burden platelets of compared to granulocytes[115]. In patients with low allele burden in granulocytes, the number of mutant cells in the peripheral blood might be underestimated. Our analysis shows that the MPN clone does not expand into all lineages of the peripheral blood to the same extent. This observation may also serve as an explanation for the inverse correlation of mutant *JAK2* allele burden and thrombopoiesis[114], which has been published before and is found in our cohort as well (Figure 22I). In pilot experiments (Figure 31), we found that patients with a low mutant *JAK2* allele burden (<20 %) in granulocytes have a higher burden in platelets. When the patients represent with a high mutant *JAK2* allele burden (>75 %) in granulocytes, the burden in platelets frequently is lower. In hospitals, the *JAK2-V617F* burden, measured in granulocytes, serves as an indicator of clonal size and molecular marker for response to treatment[149]. It would be very interesting to analyze the *JAK2* allele burden in all MPN relevant lineages of patients with molecular response to treatment. Potentially, results will show that therapy of MPN does not affect all blood lineages with the same efficiency. Eventually, new insights in dynamics of the MPN clone might help to design the next iteration of drugs used in the treatment of MPN.

Based on the results from peripheral cells, we hypothesized that the difference in allele burden already exists at the progenitor level. Interestingly, the mutant *JAK2* burden is similar in progenitors of red cells compared to granulocyte progenitors. Only in the very late stage of red cell differentiation, between BFU-E level and reticulocytes, the MPN clone expands significantly. Unfortunately, it was not possible to analyze the *JAK2-V617F* burden in CFU-Meg, colonies of platelet progenitors. Therefore, we can only speculate that the observed late expansion in the red cell lineage might also occur in the platelet lineage. In addition, further studies are required to examine if our observations show an excessive clonal expansion specifically at the late stage of red cell differentiation or a block of clonal expansion at the progenitor level.

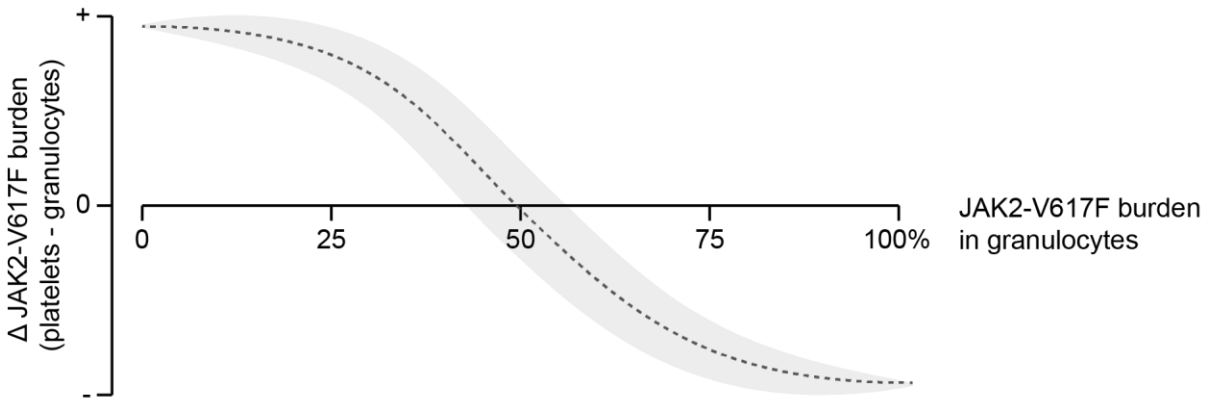


Figure 31 Dynamics of the MPN clone between platelets and granulocytes

Figure shows a schematic of the *JAK*2 burden in platelets relative to the burden in granulocytes; based on results from pilot experiments (data not shown). In patients with a low *JAK2* allele burden in granulocytes (0-25 %), the mutant *JAK2* burden is higher in platelets. When the burden in granulocytes is intermediate (40-60%), the burden in platelets is comparable. In patients with a very high *JAK2* allele burden in granulocytes (75-100%), the burden in platelets is lower.

4.4 Pediatric and adult young adult patients show different mutational landscape

We compared the number of somatic mutation in MPN patients with the age at diagnosis, and found that the number of somatic mutations does not increase with advanced age at diagnosis. Interestingly, we found that patients, who do not carry any driver mutation or any other additional somatic mutation, are frequently diagnosed at younger age (Figure 19D). This is in line with observations from pediatric cases, where patients frequently present with ET phenotype and lack driver as well as additional somatic mutations. When we applied our established NGS method to screen for somatic mutations, we found that pediatric patients presented with less somatic mutations compared to adult cases. Both, driver and somatic mutations, were less frequent in young MPN patients. This lack of molecular markers complicates establishing diagnosis and standard care in pediatric patients, as diagnosis of MPN is otherwise based on blood counts and bone marrow biopsies, which parents frequently refuse to agree on[106,150,151].

In pediatric patients, not only less mutations are observed but also other genes were mutated than reported in adult MPN patients. Mutations in genes like *TET2*, *DNMT3A* and *ASXL1* were found at lower frequency than in our study on adult MPN patients. Further, we found new mutations in IRF8, a transcription factor, which gets activated upon interferon signaling. Interferon receptors belong to the class II cytokine receptors, which also rely on JAK-signaling pathways. The mutations in *IRF8* were unknown and prediction algorithms presented divergent results for the impact of the mutations on the protein. Unfortunately, germline control DNA was not available for most of the pediatric cases. Hence, we could not validate the novel alterations to be somatic or germline mutations. Therefore, we assume that a significant proportion of the reported mutations in our pediatric study may be germline mutations.

Given the fact, that pediatric MPN patients are diagnosed at early age, but lack common somatic mutations, an accumulation of predisposing germline alterations might be an explanation for some of the cases. Genome-wide association studies have shown that the acquisition of the somatic *JAK2-V617F* mutation correlates with a SNP (rs10974944), also known as the 46/1 *JAK2* haplotype[152]. Another study reported three additional SNPs located at the genes *MECOM*, *HBS1L-MYB* and *TERT*, which were correlating with the diagnosis of MPN[153]. A recent genome wide association study of 17’000 individuals of the Estonian population has shown strong correlation of a number of SNPs with blood counts[154]. The top hit, a germline variant of *CEBPA* is associated with basophil counts. Other variants were associated with red cell numbers, mean corpuscular hemoglobin and volume. Interestingly, *HBS1L-MYB* is one of these variants and additionally correlated with elevated platelet numbers

in the Estonian population. These studies show that germline alterations in the normal population are influencing blood counts. Large GWAS analyses are mainly based on material of adult individuals. It might be possible that in pediatric MPN cases, which lack clonal markers, disease might be driven by germline factors, like SNPs modulating blood counts. As SNPs are inherited through the generations, this hypothesis suffers from the fact, that family history requires to be excluded for the diagnosis of sporadic MPN. For all patients analyzed in our study, both, family history and reactive causes were excluded for diagnosis of sporadic pediatric MPN.

4.5 Somatic mutations cause disease progression and reduced survival

In adult patients, we correlated the number of somatic mutations per patient with survival and transformation to AML. We found out that a higher number of somatic mutations associated with adverse outcome and increased risk of transformation. Patients, who only carried a driver mutation showed the most favorable survival. A previous study showed no difference in survival for individuals with *CALR* or *JAK2* driver mutation[155]. In our study, we saw a slight improvement in patients positive for *CALR*, but in our cohort of 197 adult MPN patients, only 29 patients were positive for a *CALR* mutation. Therefore, the discrepancy might be explained by the composition or the size of our cohort.

When searching for an effect of non-driver mutations, we found that alterations in the genes *TET2* and *TP53* increased the risk of transformation and correlated with adverse survival. Mutations in TET2 were observed in previous studies of myeloid malignancies[76,156–158], but were not always found to affect survival or disease progression or in myeloid malignancies. A study with AML patients showed a negative impact of *TET2*[159] and a study in MPN patients found no significant effect of *TET2* mutations. In the AML study, *TET2* variants were classified as somatic mutations, only when they were absent in a paired remission sample of the same patient. The authors of the MPN study classified all mutations as somatic, when they were not reported in the dbSNP database[156]. In our own study, we combined both approaches: first, we annotated reported variants as SNPs and then validated all unknown variants in the germline control DNA. Only variants, which were not present in the germline DNA, were classified as somatic mutations.

We compared patients without mutations in *TET2* with patients with somatic mutations in TET2 and found a significant reduction in survival. When patients with germline *TET2* mutations were compared to the control group, there was no significant difference in survival (Figure 32A). The impact of *TET2* mutations on the survival of MPN patients was reduced when patients with somatic mutations and germline mutations in *TET2* were summarized in one group (Figure 32B). A similar effect was observed for correlations of the *TET2* mutation status with transformation into AML (Figure 32C+D). Therefore, we conclude that currently available databases do not contain enough information yet to replace re-sequencing of variants in control tissue. Especially when the number of somatic mutations will be used for risk stratification, re-sequencing of the mutations in DNA of a different germ layer is recommended.

Furthermore, we found some germline mutations at VAF below 10%, which might indicate that some alleles are more difficult to read than others are. This underlines the importance of germline DNA sequencing and shows that somatic mutation calling based on low VAF is also insufficient to curate NGS data from germline alterations (Figure 11E). Studies without access

to germline control DNA might therefore underestimate the effect of somatic mutations on the disease phenotype.

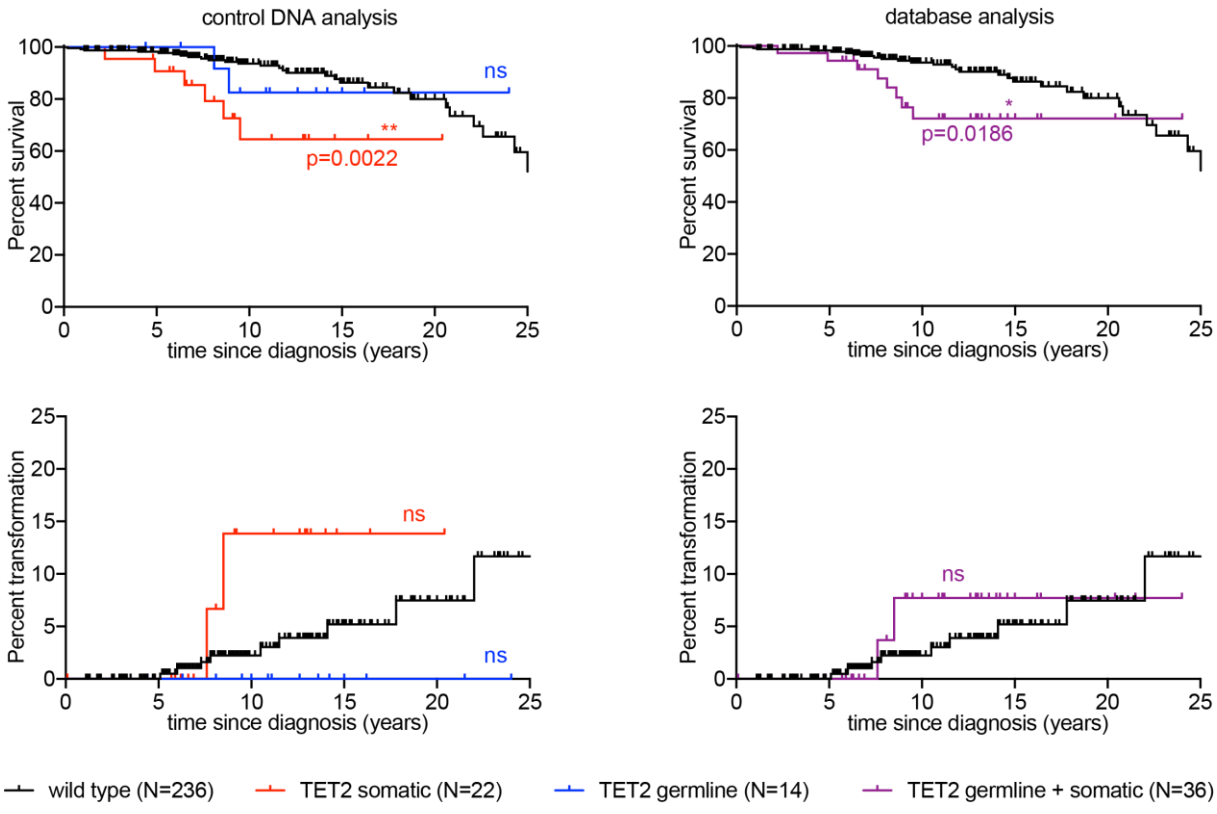


Figure 32 Somatic mutations in TET2 impair survival

Kaplan-Meier curves of patients with mutations in *TET2*. (A) Survival of patients with a somatic mutation, a germline mutation or no mutation in *TET2*. (B) Survival of the same patients with *TET2* mutations, irrespective of if the *TET2* mutation is germline or somatic. (C) Transformation to AML of patients with a somatic mutation, a germline mutation or no mutation in *TET2*. (D) Transformation to AML of the same patients with *TET2* mutations, irrespective of if the *TET2* mutation is germline or somatic. ns: not significant.

The second mutated gene, which was found to correlate with reduced survival and transformation to AML in our cohort, is *TP53*. The gene was found mutated in five patients and four of these developed a post-MPN-AML. Only months after transformation to AML, all four patients died. Re-sequencing of patient samples from earlier time points showed that the size of the clones remained small for years and then expanded rapidly during disease progression (Figure 18B). The final *TP53* allele burden reached >50 % in these patients, which indicates a loss of the wild type allele during the clonal expansion. We therefore suggest that in patients with small *TP53* clones, the loss of the wild type allele in combination with clonal expansion might play a crucial role during the transformation of MPN to AML.

In 2014, a study showed that *TP53* is one of the most common genes found to be mutated in post-MPN-AML patients[160]. The authors also presented results from a mouse model with a full *TP53* knock out and retrovirally introduced *JAK2-V617F*. All mice developed a fully penetrant and lethal disease, which was transplantable to secondary and tertiary recipients. Hence, it could be of interest to screen MPN patients for small *TP53* clones to elucidate, if there are more patients carrying small *TP53* clones. A long-term study would then be required to learn how many of these patients show an expansion of the TP53 clone and subsequently transform to AML. However, to detect mutations at very low allele burden (0.1 – 5 %) sensitive methods need to be developed, as current methods are either not sensitive enough, or only allow to analyze isolated hotspots of mutations. *TP53* mutations mainly occur in the DNA-binding domain, which corresponds to 50 % of the gene sequence, but somatic mutations may also occur at any other position of the gene. Therefore, the combination of a sensitive ultra-deep NGS method with a primer-based gene enrichment might be an appropriate technology to detect unknown TP53 mutations at low allele burden[161,162]

In our initial NGS study of 197 adult MPN patients, mutations in *DNMT3A* and *ASXL1* were found at the similar frequency, in 5% of the analyzed patients. Variants in other genes, like *EZH2* or *IDH1*, were less frequently found. To reach a sufficient number for the correlation of the mutation status survival or risk of transformation, we summarized patients with mutations in these epigenetic modifier genes (Figure 18D). When these four genes were grouped according to their function as epigenetic modifier genes, the survival of mutant patients was not significantly reduced. However, we found that increased number of somatic mutations per patient, independent of the mutated gene, correlates with adverse survival. This implies that individual genes, when they are mutated, do not have the same negative effect on the survival of MPN patients. In a large study focused on MPN patients with PMF diagnosis, the presence of *ASXL1* mutations correlated with adverse survival[96]. This indicates, that our cohort might not contain enough patients to uncover the effect on survival, when the gene is less frequently mutated than *TET2* and the impact of the mutations are smaller than mutations in *TP53*.

In summary, most somatic mutations are acquired before diagnosis of MPN. The first hits, initiating mutations may not be sufficient to induce a leukemic phenotype, but increase the risk of developing a hematologic cancer. MPN patients who only carry driver mutations show a favorable survival compared to patients with additional mutations. Somatic mutations in genes like *ASXL1* and *TP53* indicate high-risk patients, as they increase the risk of transformation from MPN to AML (Figure 33).

Figure 5

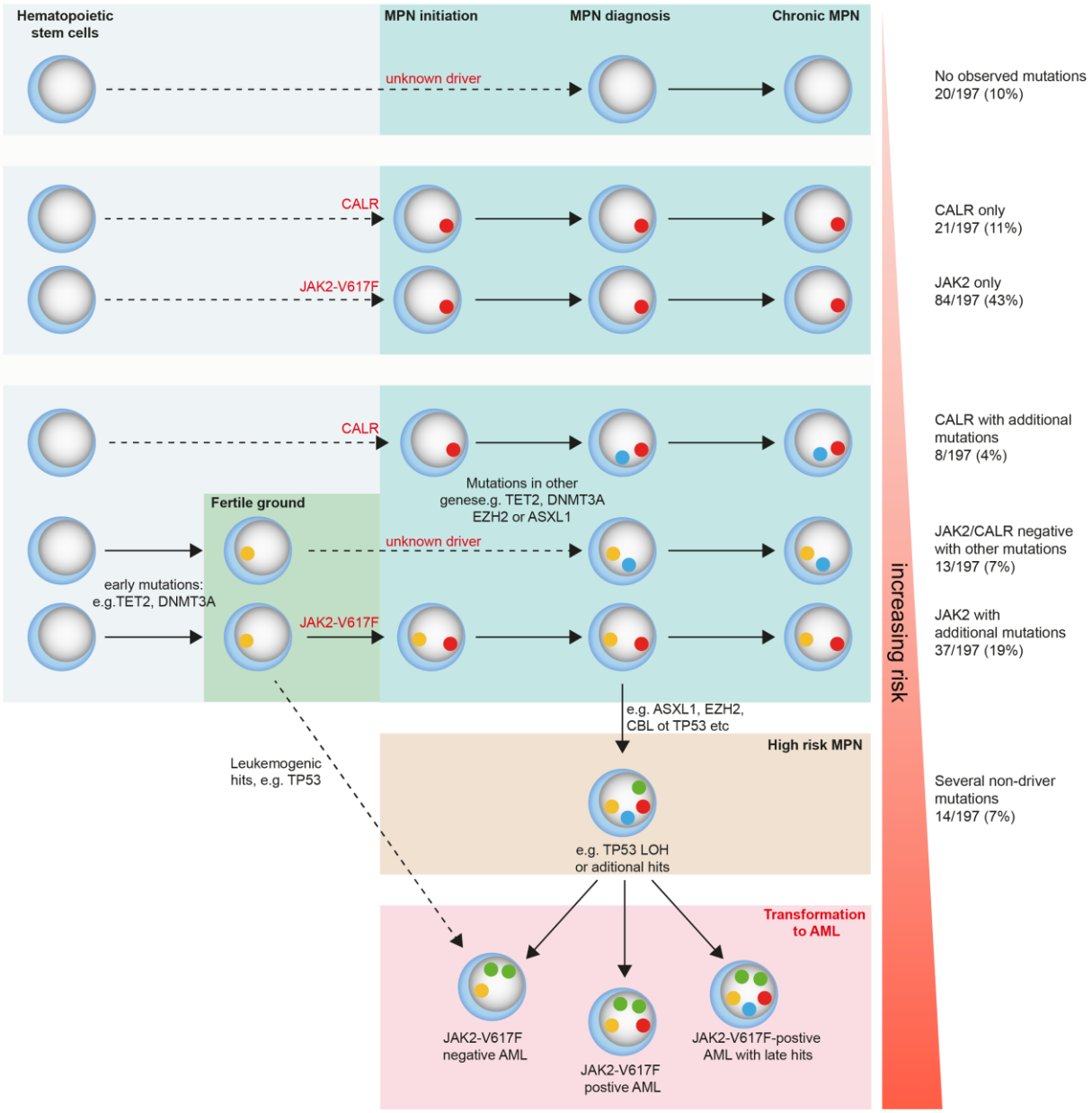


Figure 33 Model of MPN disease evolution and risk stratification in correlation to mutational events

4.6 Identifying somatic mutations affecting the MPN phenotype

Somatic mutations might not only affect the survival of the patient but may also influence the phenotype of MPN in the chronic phase of the disease. Somatic mutations in specific genes may affect the proliferation of individual lineages and thereby promote a distinct phenotype. In patients of the Basel sporadic MPN cohort, we observed that mutations in *ASXL1* and *EZH2* affect the hemoglobin and the neutrophil counts, respectively (Figure 17). In a mouse model, we showed that loss-of *EZH2* promotes proliferation of megakaryocytes, resulting in ET and PMF phenotype. Like for correlations with patient survival, the size of the cohort is critical for genotype-phenotype correlations. We therefore combined three cohorts, which were analyzed with comparable NGS methodologies.

In a combined virtual cohort, we confirmed that somatic ASLX1 mutations preferentially occur in patients with PMF phenotype[(99)]. Interestingly, also mutations in PRC2 genes and *DNMT3A* preferentially were found in patients with ET or PMF phenotype (Figure 26). Within PRC2 genes, the physical interaction of the proteins within the complex might explain, why mutations in *SUZ12* and *EZH2* result in the same MPN phenotype. For *ASXL1* and *DNMT3A*, recent studies showed that mutations in both genes reduce PRC2 activity and result in lowered H3K27me3 mark by affecting PRC1. ASXL1 is part of the complex PR-DUB, which removes H2AK119u1, placed by PRC1. Somatic mutations in *ASXL1* are often frame shift mutations, which disrupt the gene and result in uncontrolled removal of H2AK119u1. H2AK119u1 is recognized by adapter proteins of the PRC2 complex and binding to it enhances the enzymatic activity of PRC2. Therefore, reduced H2AK199u1 results in reduced H3K27me3 marks and increased target gene expression. *DNMT3A* mutations seem to reduce H2AK119u1 marks by interaction with the PRC1 complex. Especially R882 mutations in *DNMT3A* seem to bind PRC1 efficiently and inhibit PRC1 activity. Subsequently, the interaction of DNMT3A and PRC1 leads to enhanced expression of PRC2 target genes[(144)]. Taking all these data together, it seems possible that mutations in *ASXL1*, *DNMT3A* and PRC2 genes might affect the MPN phenotype by similar mechanisms. In mouse models, the alteration of *EZH2* led to reduced PRC2 function in combination with the transcriptional activation of genes, which would otherwise be repressed by the PRC2 histone mark (Figure 34). In a loss-of *EZH2* mouse model, we found that reduction of PR2 activity results in elevated expression of the *LIN28B-HGMA2* axis, which promotes megakaryopoiesis. Enhanced megakaryopoiesis might result in increased platelet counts or stimulate fibrosis in the bone marrow, which are the main diagnostic criteria of the WHO for ET and PMF, respectively.

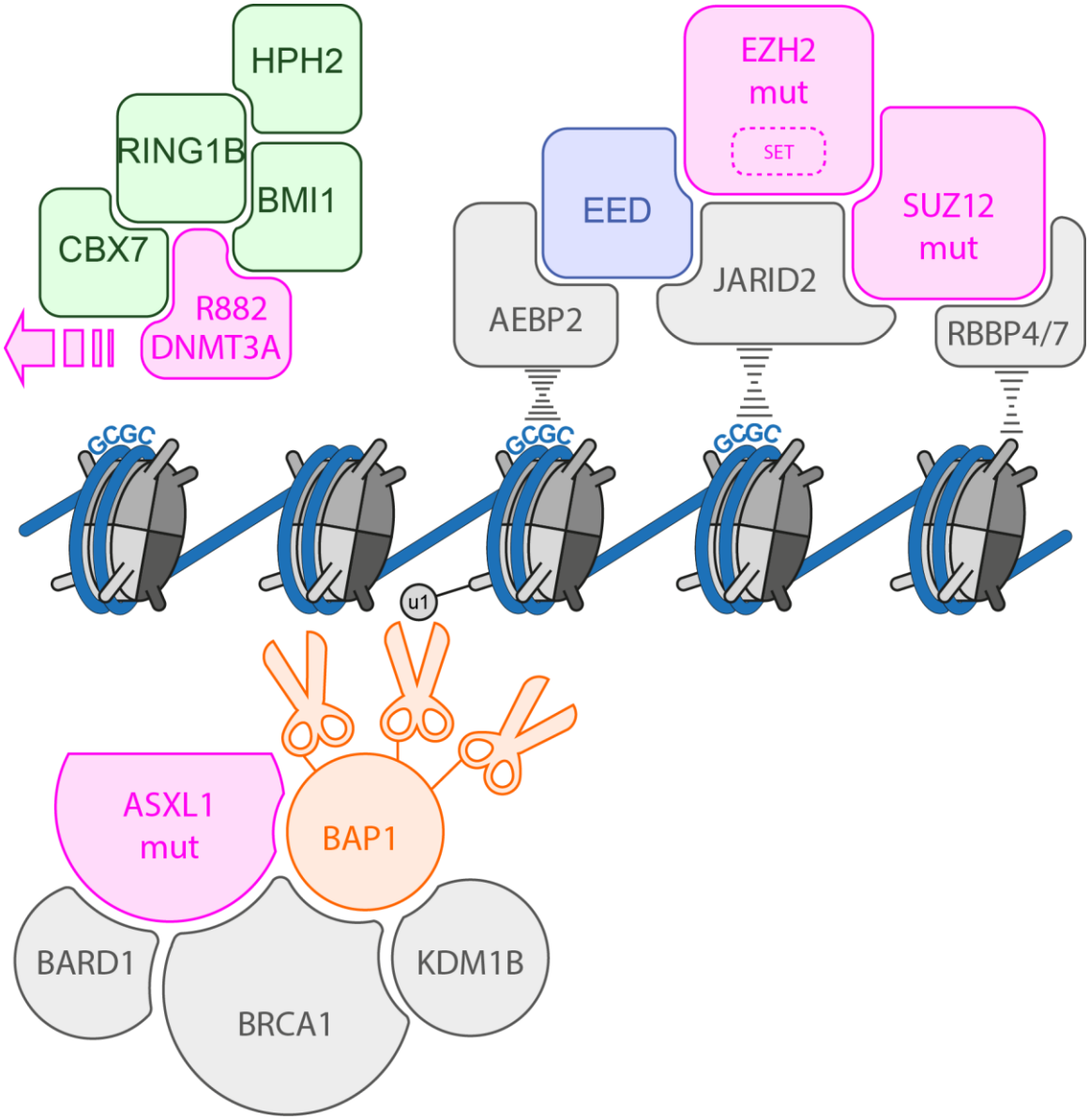


Figure 34 Gene mutations affecting PRC2 activity

Mutations in *ASXL1*, *DNMT3A* and PRC2 genes seem to result in reduction of PRC2 activity either by removing or reduction of supportive histone marks or by affecting the PRC genes directly.

In addition, mutations in genes of the spliceosome complex, *U2AF1* and *SF3B1*, were predominantly found in PMF patients. Spliceosome mutations have been described to be the most frequent somatic mutations in MDS[118,163,164]. MDS are chronic hematologic disorders, which are associated with adverse survival compared to MPN[117]. Within MPN, PMF represents the most advanced phenotype, with the highest risk to transform to AML[46]. Clinically, patients have been observed to share features of both diseases and these patients are diagnosed with MDS/MPN overlap. Patients with MPN driver mutations and additional somatic mutations commonly found in MDS might potentially belong to this more complex group of patients. As a consequence of mutational landscape and diagnosis, these patients might be at higher risk to transform to AML than other MPN patients. Molecularly, spliceosome mutations alter the processing of the transcribed RNA by the recognition of cryptic splice sites or by disturbing endogenous levels of differentially expressed isoforms[165,166]. The *SRSF2-P95S* mutations have been shown to affect splicing of *EZH2* mRNA resulting in non-functional protein[167]. This indicates that also spliceosome mutations might alter the MPN phenotype through reduction of PRC2 activity. However, but it is likely that these mutations also impair the mRNA-splicing of many other genes. This potential systemic effect may explain why these mutations are frequently found in more aggressive chronic hematologic malignancies, like MDS[118,163].

The last mutated gene in our panel, that showed enrichment in PMF patients, is *TP53*. It has been associated with disease progression in our own study and others[45,108]. Therefore, we expected *TP53* mutations to be vastly enriched in patients with PMF diagnosis. Interestingly, the distribution of MPN phenotypes among *TP53* mutated patients is similar to patients without somatic mutations. This might indicate that mutations in *TP53* do not promote a specific MPN phenotype per se and only induce disease progression when the clone lost the wild type allele and expands to a certain size. Potentially, inducing a specific MPN phenotype requires different alterations of the molecular biology than disease progression and not all mutations cover both directions, MPN initiation and progression.

Mutations found in the ET or PMF phenotype frequently are associated with the same specific molecular function, the repression of genes through the PRC2 activity. Genes, which are frequently mutated in PV patients, seem to affect a more diverse set of pathways. In our meta-analysis, we found that mutations in *KRAS*, *NRAS*, *CBL*, *NFE2*, *TET2*, *IDH1* and *IDH2* are more frequent in PV than in other MPN phenotypes. The *RAS* genes frequently contained activating mutations, which are reported to elevate PI3K/AKT and MAPK/ERK signaling and result in enhanced cycling and survival[124]. CBL is a negative regulator of signaling pathways that are induced by receptors at the cell surface[168]. Accordingly, mutations in *CBL* are associated with elevated signaling of these pathways. In patients, mutations in *CBL*, *KRAS*

and *NRAS* might therefore enhance JAK/STAT signaling in addition to the MPN driver mutation. This might correspond to the observations in mouse models, where increased *JAK2-V617F* expression results in PV phenotypes[66].

In our virtual cohort of 556 MPN patients, four of six (66%) patients with *NFE2* mutations were diagnosed as PV. Similar frequencies have been reported in a previous study of 456 MPN patients, where six of eight (75%) patients with *NFE2* mutations were initially diagnosed with PV[125]. Mouse models have shown that gene disruptions in *NFE2* result in a reduced efficacy of platelet production[127,128]. Therefore, in combination with enhanced JAK/STAT signaling, *NFE2* mutations might promote a PV phenotype through lowering the efficiency of platelet production.

Interestingly, mutations in *IDH* and *TET2* mutations are slightly overrepresented in patients with PV. Both genes are involved in DNA demethylation and required to remove the methyl group from cytosine in CpG islands (Figure 7)[83]. Loss-of-function mutations in *TET2* and *IDH* result in a global hypermethylation of DNA. Therefore, mutations in *TET2* and *IDH* display the contrary phenotype as compared to *DNMT3A* mutations, which lead to focal hypomethylation[83]. To some degree, this observation seems to be transferable to the MPN phenotype: *DNMT3A* mutations are slightly more often found in ET patients, whereas mutations in *TET2* and *IDH* are frequently found in PV patients (Figure 28A-C). However, in patients with other somatic mutations in addition to *TET2* mutations, the distribution of MPN phenotypes is comparable to the control group. This indicates that the clonal architecture needs to be considered when trying to understand contribution of single somatic mutations to the phenotype.

To examine the contribution of single somatic mutations to the MPN phenotype, single cell RNAseq could be used to analyze the expression pattern of the patients. However, this analysis requires analyzing the genotype of each single cell in parallel to the transcriptome. Otherwise, transcriptional differences could not be associated with the genotype of single cells. The combination of expression analysis and genotyping is currently not available to date. It might be possible by spiking primers into the library preparation, which are specific for the somatic mutations of individual patient. During amplification of the cDNA of the single cell, mutation specific primers would amplify the regions of interest. In the raw data after NGS analysis, the amplified regions of interest will be detected as PCR duplicates and might be interpreted by a separate genotyping workflow. Reads from endogenous RNA may be analyzed using a standard differential gene expression pipeline.

4.7 The complexity of genotype - phenotype correlations in MPN

During the past five years, NGS technology has enabled to analyze MPN patients in more detail than ever before. Still, one of the most intriguing questions in MPN was and still is how different phenotypes can arise from the same driver mutation. Based on our experience from NGS studies of MPN patients, we suggest, that the phenotype of MPN patients is an integration of various genetic aberrations (Figure 35).

At the first level of these genetic changes, driver mutations in *JAK2*, *CALR* or *MPL* define the subset of MPN phenotypes. Mutations in *CALR* and *MPL* specifically enhance signaling of the MPL receptor and result in either ET or PMF phenotype. As there are no PV patients with *CALR* or *MPL* mutation or ET/PMF patients with *JAK2-exon12* mutation, the effect of these mutations is definitive and non-reversible. The *JAK2-V617F* mutation is found in all MPN phenotypes, which might be due to the universal role of JAK2 in the signal transduction of cytokines and their receptors.

The next level is represented by somatic mutations. The modulating effect of somatic mutations seem to be less pronounced as compared to driver mutations as distinct somatic mutations might enhance or neutralize each other's effect. In this context, clonal architecture might play a role and needs to be analyzed in large scale. Different clonal branches within a patient might add separate individual effects modulating the MPN phenotype. Interestingly, a recent study shows that the order of acquisition of somatic mutations has impact on treatment response and clinical correlates[146].

In another layer of complexity, the type of mutation within one gene might have huge impact, as seen in *IDH1* and *IDH2*, were only specific mutations enable the enzyme to generate a new metabolite[137,138]. Another example is *DNMT3A*, as DNMT3A-R882 mutant protein has been shown to inhibit wild type protein and to associate with the PRC1 complex[144]. Observed frame shift or stop mutations in the *DNMT3A* might not support the exactly same effect due to major changes in the amino acid sequence. The final layer in our current version of this model contains germline alterations. Specific variations in the population have been shown to correlate with elevated numbers of specific blood lineages[154]. In MPN, SNPs might have an impact on the severity of the symptoms and thereby, potentially influence age at diagnosis. In patients with *JAK2-V617F* and no other somatic mutation, germline variants might represent the feather that breaks the balance towards a red cell or megakaryocytic phenotype.

The model doesn't have to end at this layer. As there are approximately 60% of patients, who do not carry somatic mutations additional to the driver mutations, other factors are likely to be uncovered in future. Potential candidates are genetic alterations affecting regulatory elements or miRNAs.

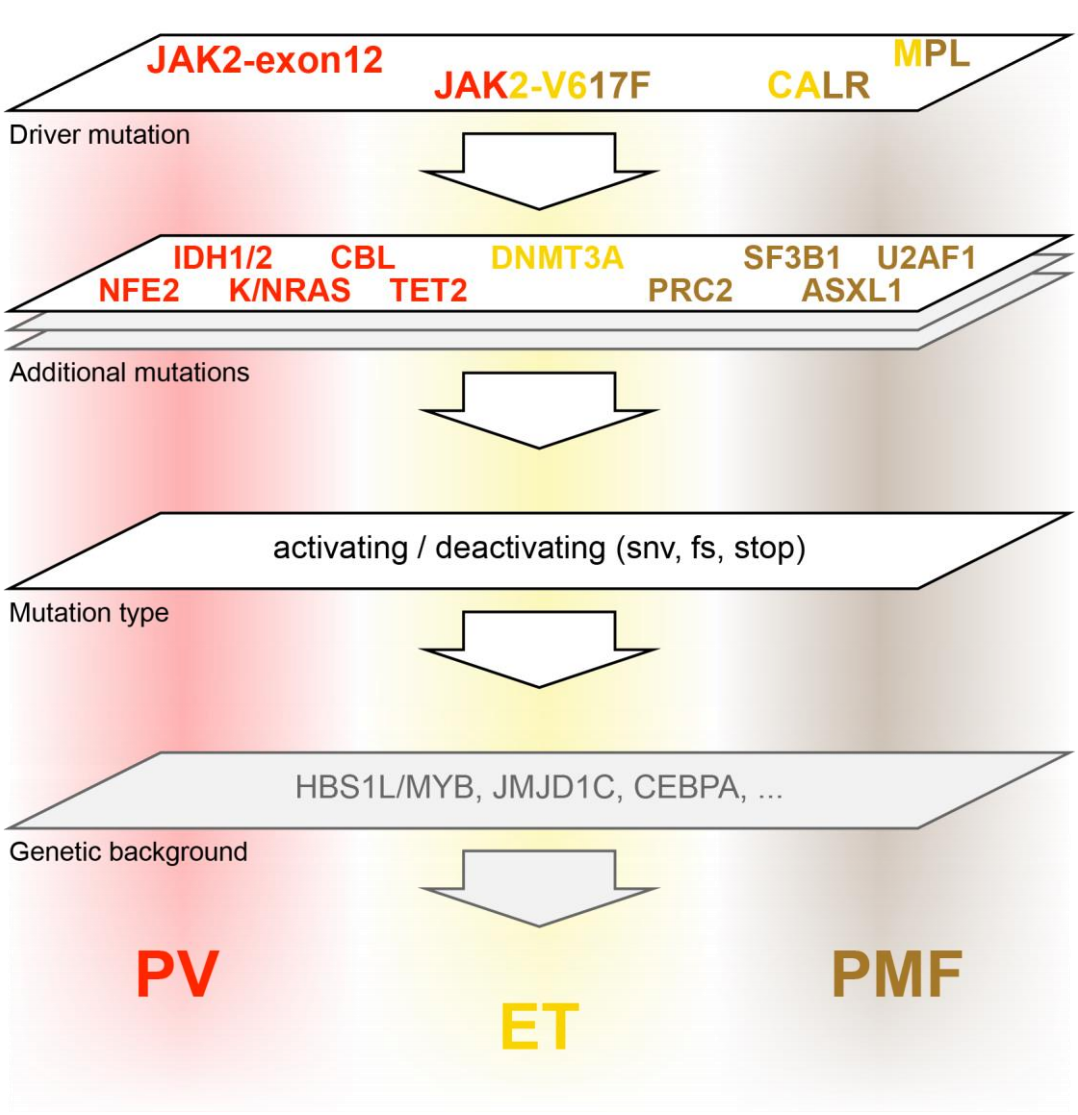


Figure 35 Multiple facets of genotypes generate the MPN phenotype in patients

Proposed multi-layered model of how different genetic alterations might contribute to the MPN phenotypes. The gene names are colored according to the phenotype, which they are suggested to promote in a mutated state. Red: PV, yellow: ET, brown: PMF.

4.8 Future

MPN are representing a rare group of disorders. As research needs to focus on subgroups of patients, local cohorts seem to reach a limit of what conclusions can be made. Currently, research groups are combining their cohorts in order to reach significance in their observations. In future, not only cohorts should be merged, also it will be important to combine NGS results with data from other available knowledge bases. This will help to make better predictions or potentially give insights, which would have been missed when analyzing NGS data as individual, self-contained experiments.

The need of running innovative analysis and complex experiments has been recognized. In hospitals, medical doctors have started work together with biologists and computer scientists to make use of the newest insights and provide detailed diagnoses. The implementation of NGS technologies in the daily routine recently created the fancy terms of precision diagnostics and personalized medicine.

However, as fast as medicine is incorporating new technologies, new challenges arise. BigData, a term which initially was used by giant internet companies, now also finds its way to research and soon also to medicine. The more data is generated by technologies like NGS methods, the more of these analyses infringe the anonymity of the patients. Anonymity is important in research, as the declaration of Helsinki, which sets ethical principles for medical research involving human subjects, clearly states, that study participants must not be affected by disadvantages as a result of the study. Therefore, proper data handling and protection of the study participants' rights will play an important role, when data is combined or shared worldwide. At the same time, the general population started to use gadgets and apps to track health and record achievements in sports. Even whole genome analyses are offered to private persons. Therefore, the accumulation of personal data might also change the way the public values this private information. These recordings might be of interest in future studies, as they might offer information from pre-malignant phases or allow automatic administration of treatment based on on-time measurement, for example heart rate assessed by fitness trackers. It will be interesting to follow which potential can be unleashed from these new developments.

5 Disclosure of individual contributions

Chapter 3.1 is based on a collaboration with Dr. Pontus Lundberg and Dr. med Axel Karow, who shared the first authorship of this study. Both developed the protocol for NGS sequencing by Illumina. Dr. Lundberg analyzed Illumna raw data (Figure 11A-C). Based on this data, I validated all variants in DNA of granulocytes and hair follicles or buccal swabs. Therefore, I developed pooling strategies for amplicon sequencing using the Ion torrent technology (Figure 9). This technology was also used for quantification of VAF from known mutations in earlier patient samples ("longitudinal sequencing", Figure 18). I also supported Dr. med Karow in the high throughput analysis of more than 2'800 colonies (Figure 16). Further, I translated the Illumina sequencing protocol (based on 1.5mL tubes) to 96 well plate preparation. I also was responsible for the graphical representation of clonal architecture analysis (Figure 1) and Circos plots (Figure 15), which represent co-occurrence of somatic mutations, or others (Figure 12, Figure 33). The original publication can be found as appendix 1 at the end of this document.

Chapter 3.2 is based on a collaboration with Dr. Pontus Lundberg and Dr. med Axel Karow. Dr. med Karow and me shared the first authorship of this study. From 43 patients, samples from approximately 20 patients were prepared, sequenced, analyzed and validated in collaboration with Dr. Lundberg and Dr. med Karow. During revision of the study, I added analysis of 23 new patient samples, kindly provided by Dr. med Maria Luicia Randi. The original publication can be found as appendix 2 at the end of this document.

The study presented in chapter 3.3 was prepared by my self.

The cohort analyses presented in chapter 3.4.1 and 3.4.3 were performed by my self. Raw data from the Nangalia cohort was available online[35]. Data from 192 sporadic MPN patients was generated and published in collaboration, as described above. All new patient samples from the study center Basel and Belgrade were sequenced and validated by my self according to the previously published workflow (Figure 10[108]). Chapter 3.4.2 refers to mouse experiments, which I was not involved. However, I supported Dr. Takafumi Shimizu in planning, performing and analyzing qPCR experiments on mouse and human samples to elucidate the molecular effect of EZH loss-of-function mutations. Data from human samples of this study is shown in Figure 27.

All parts of the study were supported by vivid discussions with lab members and helpful comments from many other colleagues.

6 Publications

Lekovic D, Gotic M, Skoda R, Beleslin-Cokic B, Milic N, Mitrovic-Ajtic O, **Nienhold R**, Sefer D, Suboticki T, Buac M, Markovic D, Diklic M, Cokic VP. Bone marrow microvessel density and plasma angiogenic factors in myeloproliferative neoplasms: clinicopathological and molecular correlations. Ann Hematol. 2016 Dec 6.

Pastor V, Hirabayashi S, Karow A, Wehrle J, Kozyra EJ, **Nienhold R**, Ruzaike G, Lebrecht D, Yoshimi A, Niewisch M, Ripperger T, Göhring G, Baumann I, Schwarz S, Strahm B, Flotho C, Skoda RC, Niemeyer CM, Wlodarski MW. Mutational landscape in children with myelodysplastic syndromes is distinct from adults: specific somatic drivers and novel germline variants. Leukemia. 2016 Dec 9. doi: 10.1038/leu.2016.342.

Shimizu T, Kubovcakova L, **Nienhold R**, Zmajkovic J, Meyer SC, Hao-Shen H, Geier F, Dirnhofer S, Guglielmelli P, Vannucchi AM, Feenstra JD, Kralovics R, Orkin SH, Skoda RC.Loss of Ezh2 synergizes with *JAK2-V617F* in initiating myeloproliferative neoplasms and promoting myelofibrosis. J Exp Med. 2016 Jul 25;213(8):1479-96. doi: 10.1084/jem.20151136.

Subotički T, Mitrović Ajtić O, Beleslin-Čokić BB, **Nienhold R**, Diklić M, Djikić D, Leković D, Bulat T, Marković D, Gotić M, Noguchi CT, Schechter AN, Skoda RC, Čokić VP. Angiogenic factors are increased in circulating granulocytes and CD34+ cells of myeloproliferative neoplasms. Mol Carcinog. 2016 Jun 24. doi: 10.1002/mc.22517.

Karow A*, **Nienhold R***, Lundberg P, Peroni E, Putti MC, Randi ML, Skoda RC. Mutational profile of childhood myeloproliferative neoplasms. Leukemia. 2015 Dec;29(12):2407-9. doi: 10.1038/leu.2015.205.

Lundberg P, **Nienhold R**, Ambrosetti A, Cervantes F, Pérez-Encinas MM, Skoda RC. Somatic mutations in calreticulin can be found in pedigrees with familial predisposition to myeloproliferative neoplasms. Blood. 2014 Apr 24;123(17):2744-5. doi: 10.1182/blood-2014-01-550863.

Lundberg P*, Karow A*, **Nienhold R**, Looser R, Hao-Shen H, Nissen I, Girsberger S, Lehmann T, Passweg J, Stern M, Beisel C, Kralovics R, Skoda RC. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. Blood. 2014 Apr 3;123(14):2220-8. doi: 10.1182/blood-2013-11-537167.

Table of figures

Figure 1 Hierarchy of hematopoiesis 9

Figure 2 Cytokine-induced JAK-STAT signaling 13

Figure 3 Transformation and progression in MPN 17

Figure 4 Activation of JAK-STAT signaling by THPO 20

Figure 5 Mutant JAK2 induces cytokine signaling 23

Figure 6 Activation of MPL by mutant CALR 25

Figure 7 Pathways of DNA methylation and histone modification altered in hematologic malignancies 27

Figure 8 Standard blood sample work-up 34

Figure 9 Pooling strategy for validation of mutations 41

Figure 10 NGS workflow 45

Figure 11 Quality control measurements of NGS 46

Figure 12 Karyogram and chromosomal location of all sequenced genes 49

Figure 13 Frequency and distribution of somatic mutations in patients with MPN 50

Figure 14 Location of copy number variations on chromosome 20 52

Figure 15 Circos plots showing co-occurrence of somatic mutations 54

Figure 16 Clonal evolution in MPN patients carrying somatic mutations in epigenetic modifier genes 56

Figure 17 Impact of somatic mutations on blood counts, transformation and survival 59

Figure 18 Longitudinal sequencing reveals most mutations are present at diagnosis 61

Figure 19 Only two somatic mutations acquired during disease course 62

Figure 20 Mutational profile of pediatric MPN 65

Figure 21 Comparison of mutations in pediatric and adult MPN 67

Figure 22 Correlation of driver mutation burden at time of diagnosis 69

Figure 23 Purification of progenitors and peripheral blood lineages from individual blood samples 72

Figure 24 Comprehensive JAK2 analysis in progenitors and peripheral blood lineages 75

Figure 25 Analysis of JAK2 driver mutations in progenitors of MPN patients 77

Figure 26 Genotype-phenotype correlations in MPN 80

Figure 27 Loss of EZH2 in mice reflects the MPN in human patients.....85

Figure 28 Effect of individual mutations on MPN phenotypes 89

Figure 29 Mutational landscapes of chronic hematologic malignancies 92

Figure 30 Molecular development to clinically apparent MPN.....94

Figure 31 Dynamics of the MPN clone between platelets and granulocytes 96

Figure 32 Somatic mutations in TET2 impair survival 100

Figure 33 Model of MPN disease evolution and risk stratification in correlation to mutational events.....103

Figure 34 Gene mutations affecting PRC2 activity.....105

Figure 35 Multiple facets of genotypes generate the MPN phenotype in patients 109

List of tables

Table 1 Cells of the peripheral blood 10

Table 2 WHO criteria for the diagnosis of ET, PV and PMF.....15

Table 3 Genetic markers of MPN 18

Table 4 Clinical characteristics of the MPN patients at diagnosis 32

Table 5 Primers for SNaPshot analysis 37

Table 6 List of the 104 genes sequenced by NGS.....39

Table 7 Clinical characteristics of MPN patients eligible for NGS analysis.....44

Table 8 Clinical characteristics of pediatric MPN patients.....63

Table 9 Frequently mutated genes in MPN.....80

References

1. Arber DA, Orazi A, Hasserjian R, Borowitz MJ, Beau MM Le, Bloomfield CD, et al. The 2016 revision to the World Health Organization classi fi cation of myeloid neoplasms and acute leukemia. Blood. 2016;127(20):2391–406.

2. Abkowitz JL. Evidence that the number of hematopoietic stem cells per animal is conserved in mammals. Blood [Internet]. 2002 Sep 18;100(7):2665–7. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2002-03-0822

3. Morrison SJ, Scadden DT. The bone marrow niche for haematopoietic stem cells. Nature [Internet]. 2014 Jan 15;505(7483):327–34. Available from: http://www.nature.com/doifinder/10.1038/nature12984

4. Passegué E, Wagers AJ, Giuriato S, Anderson WC, Weissman IL. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. J Exp Med [Internet]. 2005 Dec 5;202(11):1599–611. Available from: http://www.jem.org/lookup/doi/10.1084/jem.20050967

5. Yamamoto R, Morita Y, Ooehara J, Hamanaka S, Onodera M, Rudolph KL. Clonal Analysis Unveils Self-Renewing Lineage-Restricted Progenitors Generated Directly from Hematopoietic Stem Cells. Cell [Internet]. 2013;154(5):1112–26. Available from: http://dx.doi.org/10.1016/j.cell.2013.08.007

6. Abkowitz JL, Persik MT, Shelton GH, Ott RL, Kiklevich J V, Catlin SN, et al. Behavior of hematopoietic stem cells in a large animal. Proc Natl Acad Sci U S A. 1995;92(March):2031–5.

7. Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts and PW. Molecular Biology of the Cell. 4th editio. New York: Garland Science; 2002.

8. Hänel K, Cornelissen C, Lüscher B, Baron J. Cytokines and the Skin Barrier. Int J Mol Sci [Internet]. 2013 Mar 26;14(4):6720–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2404683

9. Miura O, Ihle JN. Dimer- and oligomerization of the erythropoietin receptor by disulfide bond formation and significance of the region near the WSXWS motif in intracellular transport. [Internet]. Vol. 306, Archives of biochemistry and biophysics. 1993. p. 200–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8215404

10. Murakami M, Narazaki M, Hibi M, Yawata H, Yasukawa K, Hamaguchi M, et al. Critical cytoplasmic region of the interleukin 6 signal transducer gp130 is conserved in the cytokine receptor family. Proc Natl Acad Sci U S A [Internet]. 1991 Dec 15;88(24):11349–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1662392

11. O'Neal KD, Yu-Lee LY. The proline-rich motif (PRM): a novel feature of the cytokine/hematopoietin receptor superfamily. Lymphokine Cytokine Res [Internet]. 1993 Oct;12(5):309–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8260540

12. Tanner JW, Chen W, Young RL, Longmore GD, Shaw AS. The conserved box 1 motif of cytokine receptors is required for association with JAK kinases. J Biol Chem [Internet]. 1995 Mar 24;270(12):6523–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7896787

13. Chen X, Vinkemeier U, Zhao Y, Jeruzalmi D, Darnell JE, Kuriyan J. Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. Cell. 1998;93(5):827–39.

14. Schindler C, Shuai K, Prezioso VR, Darnell JE. Interferon-dependent tyrosine phosphorylation of a latent cytoplasmic transcription factor. Science [Internet]. 1992 Aug 7;257(5071):809–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1496401

15. Shuai K, Ziemiecki A, Wilks AF, Harpur AG, Sadowski HB, Gilman MZ, et al. Polypeptide signalling to the nucleus through tyrosine phosphorylation of Jak and Stat proteins. Nature [Internet]. 1993 Dec 9;366(6455):580–3. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/7504784

16. Soldaini E, John S, Moro S, Bollenbacher J, Schindler U, Leonard WJ. DNA binding site selection of dimeric and tetrameric Stat5 proteins reveals a large repertoire of divergent tetrameric Stat5a binding sites. Mol Cell Biol [Internet]. 2000 Jan;20(1):389–401. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=85094&tool=pmcentrez&rendertype=abstract

17. Ehret GB, Reichenbach P, Schindler U, Horvath CM, Fritz S, Nabholz M, et al. DNA binding specificity of different STAT proteins: Comparison of in vitro specificity with natural target sites. J Biol Chem. 2001;276(9):6675–88.

18. Okuda K, Foster R, Griffin JD. Signaling domains of the beta c chain of the GM-CSF/IL-3/IL-5 receptor. Ann N Y Acad Sci [Internet]. 1999 Apr 30;872(11):305-12-3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10372132

19. Matsuguchi T, Salgia R, Hallek M, Eder M, Druker B, Ernst TJ, et al. Shc phosphorylation in myeloid cells is regulated by granulocyte macrophage colony-stimulating factor, interleukin-3, and steel factor and is constitutively increased by p210BCR/ABL. J Biol Chem [Internet]. 1994 Feb 18;269(7):5016–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7508932

20. Yamada M, Komatsu N, Okada K, Kato T, Miyazaki H, Miura Y. Thrombopoietin induces tyrosine phosphorylation and activation of mitogen-activated protein kinases in a human thrombopoietin-dependent cell line. Biochem Biophys Res Commun [Internet]. 1995 Dec 5;217(1):230–7. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0006291X85727684

21. Hitchcock IS, Kaushansky K. Thrombopoietin from beginning to end. Br J Haematol [Internet]. 2014 Apr 25;165(2):259–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7961995

22. Okuda K, Sanghera JS, Pelech SL, Kanakura Y, Hallek M, Griffin JD, et al. Granulocyte-macrophage colony-stimulating factor, interleukin-3, and steel factor induce rapid tyrosine phosphorylation of p42 and p44 MAP kinase. Blood [Internet]. 1992 Jun 1;79(11):2880–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10372132

23. Bittorf T, Jaster R, Brock J. Rapid activation of the MAP kinase pathway in hematopoietic cells by erythropoietin, granulocyte-macrophage colony-stimulating factor and interleukin-3. Cell Signal [Internet]. 1994 Mar;6(3):305–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7917788

24. Todokoro K, Sugiyama M, Nishida E, Nakaya K. Activation of mitogen-activated protein kinase cascade through erythropoietin receptor. Biochem Biophys Res Commun [Internet]. 1994 Sep 30;203(3):1912–9. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0006291X84724119

25. Bartalucci N, Guglielmelli P, Vannucchi AM. Rationale for targeting the PI3K/Akt/mTOR pathway in myeloproliferative neoplasms. Clin Lymphoma Myeloma Leuk [Internet]. 2013 Sep;13 Suppl 2(SUPPL. 2):S307-9. Available from: http://dx.doi.org/10.1016/j.clml.2013.07.011

26. Vesely MD, Kershaw MH, Schreiber RD, Smyth MJ. Natural innate and adaptive immunity to cancer. Annu Rev Immunol [Internet]. 2011;29:235–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21219185

27. Fuertes MB, Woo S-R, Burnett B, Fu Y-X, Gajewski TF. Type I interferon response and innate immune sensing of cancer. Trends Immunol [Internet]. 2013 Feb;34(2):67–73. Available from: http://dx.doi.org/10.1016/j.it.2012.10.004

28. Pardanani AD, Levine RL, Lasho T, Pikman Y, Mesa RA, Wadleigh M, et al. MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. Blood [Internet]. 2006 Nov 15;108(10):3472–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16868251

29. Pikman Y, Lee BH, Mercher T, McDowell E, Ebert BL, Gozo M, et al. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. PLoS Med. 2006;3(7):1140–51.

30. Watowich SS, Xie X, Klingmuller U, Kere J, Lindlof M, Berglund S, et al. Erythropoietin receptor mutations associated with familial erythrocytosis cause hypersensitivity to erythropoietin in the heterozygous state. Blood [Internet]. 1999 Oct 1;94(7):2530–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10498627

31. Kralovics R, Passamonti F, Buser AS, Teo S, Tiedt R, Passweg JR, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. N Engl J Med [Internet]. 2005 Apr 28;352(17):1779–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15858187

32. Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S, et al. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet (London, England) [Internet]. 2005;365(9464):1054–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15781101

33. Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJP, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell [Internet]. 2005 Apr;7(4):387–97. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15837627

34. James C, Ugo V, Le Couédic J-P, Staerk J, Delhommeau F, Lacout C, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature [Internet]. 2005 Apr 28;434(7037):1144–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15793561

35. Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC, et al. Somatic CALR Mutations in Myeloproliferative Neoplasms with Nonmutated JAK2. N Engl J Med [Internet]. 2013 Dec 19;369(25):2391–405. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa1312542%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/24325359%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3966280

36. Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD, et al. Somatic Mutations of Calreticulin in Myeloproliferative Neoplasms. N Engl J Med [Internet]. 2013 Dec 19;369(25):2379–90. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa1311347

37. WISEMAN BK. A FUNDAMENTAL, RECIPROCAL RELATIONSHIP BETWEEN MYELOID AND LYMPHOID TISSUES. J Am Med Assoc [Internet]. 1936 Feb 22;106(8):609. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1936.02770080023008

38. VERSO ML. THE EVOLUTION OF BLOOD-COUNTING TECHNIQUES. Med Hist [Internet]. 1964 Apr;8:149–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1033366

39. DAMESHEK W. Some speculations on the myeloproliferative syndromes. Blood [Internet]. 1951 Apr;6(4):372–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14820991

40. Elliott MA, Tefferi A. Chronic neutrophilic leukemia 2016: Update on diagnosis, molecular genetics, prognosis, and management. Am J Hematol [Internet]. 2016 Mar;91(3):341–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26700908

41. Gotlib J. World Health Organization-defined eosinophilic disorders: 2015 update on diagnosis, risk stratification, and management. Am J Hematol. 2015;90(11):1077–89.

42. Gianelli U, Iurlo A, Cattaneo D, Bossi A, Cortinovis I, Augello C, et al. Discrepancies between bone marrow histopathology and clinical phenotype in BCR-ABL1-negative myeloproliferative neoplasms associated with splanchnic vein thrombosis. Leuk Res [Internet]. 2015;39(5):525–9. Available from: http://dx.doi.org/10.1016/j.leukres.2015.03.009

43. Hultcrantz M, Kristinsson SY, Andersson TM-L, Landgren O, Eloranta S, Derolf AR, et al. Patterns of Survival Among Patients With Myeloproliferative Neoplasms Diagnosed in Sweden From 1973 to 2008: A Population-Based Study. J Clin Oncol [Internet]. 2012 Aug 20;30(24):2995–3001. Available from: http://jco.ascopubs.org/cgi/doi/10.1200/JCO.2012.42.1925

44. Barbui T, Thiele J, Carobbio A, Vannucchi AM, Tefferi A. The rate of transformation from JAK2-mutated ET to PV is influenced by an accurate WHO-defined clinico-morphological diagnosis.

Leuk [Internet]. 2015;29(4):992–993 2p. Available from: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=jlh&AN=109780680&site=ehost-live&scope=site

45. Cerquozzi S, Tefferi A. Blast transformation and fibrotic progression in polycythemia vera and essential thrombocythemia: a literature review of incidence and risk factors. Blood Cancer J [Internet]. 2015;5(11):e366. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26565403

46. Bjorkholm M, Derolf AR, Hultcrantz M, Kristinsson SY, Ekstrand C, Goldin LR, et al. Treatment-Related Risk Factors for Transformation to Acute Myeloid Leukemia and Myelodysplastic Syndromes in Myeloproliferative Neoplasms. J Clin Oncol [Internet]. 2011;29(17):2410–5. Available from: http://jco.ascopubs.org/cgi/doi/10.1200/JCO.2011.34.7542

47. Abdulkarim K, Girodon F, Johansson P, Maynadié M, Kutti J, Carli PM, et al. AML transformation in 56 patients with Ph- MPD in two well defined populations. Eur J Haematol. 2009;82(2):106–11.

48. Koppikar P, Abdel-Wahab O, Hedvat C, Marubayashi S, Patel J, Goel A, et al. Efficacy of the JAK2 inhibitor INCB16562 in a murine model of MPLW515L-induced thrombocytosis and myelofibrosis. Blood. 2010;115(14):2919–27.

49. Stockklausner C, Klotter AC, Dickemann N, Kuhlee IN, Duffert CM, Kerber C, et al. The thrombopoietin receptor P106L mutation functionally separates receptor signaling activity from thrombopoietin homeostasis. Blood. 2015;125(7):1159–69.

50. Favale F, Messaoudi K, Varghese LN, Boukour S, Pecquet C, Gryshkova V, et al. An incomplete trafficking defect to the cell-surface leads to paradoxical thrombocytosis for human and murine MPL P106L. Blood [Internet]. 2016 Dec 29;128(26):3146–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25538044

51. Ding J, Komatsu H, Wakita A, Kato-Uranishi M, Ito M, Satoh A, et al. Familial essential thrombocythemia associated with a dominant-positive activating mutation of the c-MPL gene, which encodes for the receptor for thrombopoietin. Blood. 2004;103(11):4198–200.

52. Liu K, Martini M, Rocca B, Amos CI, Teofili L, Giona F, et al. Evidence for a founder effect of the MPL-S505N mutation in eight Italian pedigrees with hereditary thrombocythemia. Haematologica. 2009;94(10):1368–74.

53. El-Harith EHA, Roesl C, Ballmaier M, Germeshausen M, Frye-Boukhriss H, Von Neuhoff N, et al. Familial thrombocytosis caused by the novel germ-line mutation p.Pro106Leu in the MPL gene. Br J Haematol. 2009;144(2):185–94.

54. He X, Chen Z, Jiang Y, Qiu X, Zhao X. Different mutations of the human c-mpl gene indicate distinct haematopoietic diseases. J Hematol Oncol [Internet]. 2013;6(1):11. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3563459&tool=pmcentrez&rendertype=abstract

55. Rane SG, Reddy EP. Janus kinases: components of multiple signaling pathways. Oncogene [Internet]. 2000 Nov 20;19(49):5662–79. Available from: http://www.nature.com/doifinder/10.1038/sj.onc.1203925

56. O'Shea JJ, Schwartz DM, Villarino A V., Gadina M, McInnes IB, Laurence A. The JAK-STAT pathway: impact on human disease and therapeutic intervention. Annu Rev Med [Internet]. 2015;66(1):311–28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25587654%5Cnhttp://www.annualreviews.org/doi/abs/10.1146/annurev-med-051113-024537

57. Richter MF, Duménil G, Uzé G, Fellous M, Pellegrini S. Specific contribution of Tyk2 JH regions to the binding and the expression of the interferon α/β receptor component IFNAR1. J Biol Chem. 1998;273(38):24723–9.

58. Bandaranayake RM, Ungureanu D, Shan Y, Shaw DE, Silvennoinen O, Hubbard SR. Crystal structures of the JAK2 pseudokinase domain and the pathogenic mutant V617F. Nat Struct Mol

Biol [Internet]. 2012 Aug 6;19(8):754–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22820988%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3414675

59. Ungureanu D, Wu J, Pekkala T, Niranjan Y, Young C, Jensen ON, et al. The pseudokinase domain of JAK2 is a dual-specificity protein kinase that negatively regulates cytokine signaling. Nat Struct Mol Biol [Internet]. 2011;18(9):971–6. Available from: http://dx.doi.org/10.1038/nsmb.2099

60. Saharinen P, Silvennoinen O. The pseudokinase domain is required for suppression of basal activity of Jak2 and Jak3 tyrosine kinases and for cytokine-inducible activation of signal transduction. J Biol Chem [Internet]. 2002 Dec 6;277(49):47954–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12351625

61. Shan Y, Gnanasambandan K, Ungureanu D, Kim ET, Hammarén H, Yamashita K, et al. Molecular basis for pseudokinase-dependent autoinhibition of JAK2 tyrosine kinase. Nat Struct Mol Biol [Internet]. 2014 Jun 11;21(7):579–84. Available from: http://dx.doi.org/10.1038/nsmb.2849

62. Silvennoinen O, Hubbard SR. Molecular insights into regulation of JAK2 in myeloproliferative neoplasms. Blood [Internet]. 2015 May 28;125(22):3388–92. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2015-01-621110

63. Scott LM. The JAK2 exon 12 mutations: a comprehensive review. Am J Hematol [Internet]. 2011 Aug;86(8):668–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21674578

64. Tiedt R, Hao-Shen H, Sobas MA, Looser R, Dirnhofer S, Schwaller J, et al. Ratio of mutant *JAK2-V617F* to wild-type Jak2 determines the MPD phenotypes in transgenic mice. Blood [Internet]. 2008 Jan 11;111(8):3931–40. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2007-08-107748

65. Grisouard J, Li S, Kubovcakova L, Rao TN, Meyer SC, Lundberg P, et al. JAK2 exon 12 mutant mice display isolated erythrocytosis and changes in iron metabolism favoring increased erythropoiesis. Blood [Internet]. 2016 Aug 11;128(6):839–51. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2015-12-689216

66. Li J, Kent DG, Chen E, Green AR. Mouse models of myeloproliferative neoplasms: JAK of all grades. Dis Model Mech [Internet]. 2011 May;4(3):311–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21558064

67. Seddiki N, Nato F, Lafaye P, Amoura Z, Piette JC, Mazié JC. Calreticulin, a potential cell surface receptor involved in cell penetration of anti-DNA antibodies. J Immunol [Internet]. 2001 May 15;166(10):6423–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11342668

68. Michalak M, Corbett EF, Mesaeli N, Nakamura K, Opas M. Calreticulin: one protein, one gene, many functions. Biochem J [Internet]. 1999;344 Pt 2:281–92. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1220642&tool=pmcentrez&rendertype=abstract

69. Nakamura K, Zuppini A, Arnaudeau S, Lynch J, Ahsan I, Krause R, et al. Functional specialization of calreticulin domains. J Cell Biol [Internet]. 2001 Sep 3;154(5):961–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11524434

70. Araki M, Yang Y, Masubuchi N, Hironaka Y, Takei H, Morishita S, et al. Activation of the thrombopoietin receptor by mutant calreticulin in CALR-mutant myeloproliferative neoplasms. Blood [Internet]. 2016 Mar 10;127(10):1307–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26817954

71. Chachoua I, Pecquet C, El-Khoury M, Nivarthi H, Albu R-II, Marty C, et al. Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. Blood [Internet]. 2016 Mar 10;127(10):1325–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26668133

72. Marty C, Pecquet C, Nivarthi H, El-Khoury M, Chachoua I, Tulliez M, et al. Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis. Blood. 2016;127(10):1317–24.

73. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature [Internet]. 2009 Apr 9;458(7239):719–24. Available from: http://dx.doi.org/10.1038/nature07943

74. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A mutations in acute myeloid leukemia. N Engl J Med [Internet]. 2010 Dec 16;363(25):2424–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21067377

75. Abdel-Wahab O, Pardanani A, Rampal R, Lasho TL, Levine RL, Tefferi A. DNMT3A mutational analysis in primary myelofibrosis, chronic myelomonocytic leukemia and advanced phases of myeloproliferative neoplasms. Leukemia [Internet]. 2011 Jul;25(7):1219–20. Available from: http://www.nature.com/doifinder/10.1038/leu.2011.82

76. Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Massé A, et al. Mutation in TET2 in myeloid cancers. N Engl J Med [Internet]. 2009 May 28;360(22):2289–301. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19474426

77. Schaub FX, Looser R, Li S, Hao-Shen H, Lehmann T, Tichelli A, et al. Clonal analysis of TET2 and JAK2 mutations suggests that TET2 can be a late event in the progression of myeloproliferative neoplasms. Blood [Internet]. 2010 Mar 11;115(10):2003–7. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2009-09-245381

78. Wagner K, Damm F, Göhring G, Görlich K, Heuser M, Schäfer I, et al. Impact of IDH1 R132 mutations and an IDH1 single nucleotide polymorphism in cytogenetically normal acute myeloid leukemia: SNP rs11554137 is an adverse prognostic factor. J Clin Oncol [Internet]. 2010 May 10;28(14):2356–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20368538

79. Dang L, Jin S, Su SM. IDH mutations in glioma and acute myeloid leukemia. Trends Mol Med [Internet]. 2010;16(9):392–7. Available from: http://dx.doi.org/10.1016/j.molmed.2010.07.002

80. Ernst T, Chase AJ, Score J, Hidalgo-curtis CE, Bryant C, Jones A V, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat Publ Gr [Internet]. 2010;42(8):722–6. Available from: http://dx.doi.org/10.1038/ng.621

81. Gelsi-Boyer V, Trouplin V, Adélaïde J, Bonansea J, Cervera N, Carbuccia N, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. Br J Haematol [Internet]. 2009 Jun;145(6):788–800. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19388938

82. Carbuccia N, Murati A, Trouplin V, Brecqueville M, Adélaïde J, Rey J, et al. Mutations of ASXL1 gene in myeloproliferative neoplasms. Leukemia [Internet]. 2009 Nov;23(11):2183–6. Available from: http://www.nature.com/doifinder/10.1038/leu.2009.141

83. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet [Internet]. 1998 Jul;19(3):219–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9662389

84. Russler-Germain DA, Spencer DH, Young MA, Lamprecht TL, Miller CA, Fulton R, et al. The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers. Cancer Cell [Internet]. 2014 Apr;25(4):442–54. Available from: http://dx.doi.org/10.1016/j.ccr.2014.02.010

85. Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature [Internet]. 2010 Dec 9;468(7325):839–43. Available from: http://www.nature.com/doifinder/10.1038/nature09586

86. McKenney AS, Levine RL. Isocitrate dehydrogenase mutations in leukemia. J Clin Invest [Internet]. 2013 Sep;123(9):3672–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23999441

87. Kass SU, Pruss D, Wolffe AP. How does DNA methylation repress transcription? Trends Genet

[Internet]. 1997 Nov;13(11):444–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9385841

88. Gerstung M, Pellagatti A, Malcovati L, Giagounidis A, Porta MG Della, Jädersten M, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. Nat Commun [Internet]. 2015 Jan 9;6:5901. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25574665%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4338540

89. Scheuermann JC, de Ayala Alonso AG, Oktaba K, Ly-Hartig N, McGinty RK, Fraterman S, et al. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature [Internet]. 2010 May 13;465(7295):243–7. Available from: http://dx.doi.org/10.1038/nature08966

90. Levine SS, Weiss A, Erdjument-Bromage H, Shao Z, Tempst P, Kingston RE. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. Mol Cell Biol [Internet]. 2002 Sep;22(17):6070–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12167701

91. Blackledge NP, Farcas AM, Kondo T, King HW, McGouran JF, Hanssen LLP, et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell [Internet]. 2014 Jun 5;157(6):1445–59. Available from: http://dx.doi.org/10.1016/j.cell.2014.05.004

92. Balasubramani A, Larjo A, Bassein JA, Chang X, Hastie RB, Togher SM, et al. Cancer-associated ASXL1 mutations may act as gain-of-function mutations of the ASXL1-BAP1 complex. Nat Commun [Internet]. 2015 Jun 22;6:7307. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26095772

93. Abdel-Wahab O, Adli M, LaFave LM, Gao J, Hricik T, Shih AH, et al. ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression. Cancer Cell [Internet]. 2012 Aug;22(2):180–93. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1535610812003030

94. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, et al. The Polycomb group protein EZH2 directly controls DNA methylation. Nature [Internet]. 2006 Feb 16;439(7078):871–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16357870

95. Pardanani A, Lasho T, Finke C, Oh ST, Gotlib J, Tefferi A. LNK mutation studies in blast-phase myeloproliferative neoplasms, and in chronic-phase disease with TET2, IDH, JAK2 or MPL mutations. Leukemia [Internet]. 2010 Oct 19;24(10):1713–8. Available from: http://www.nature.com/doifinder/10.1038/leu.2010.163

96. Tefferi A, Guglielmelli P, Lasho TL, Rotunno G, Finke C, Mannarelli C, et al. CALR and ASXL1 mutations-based molecular prognostication in primary myelofibrosis: an international study of 570 patients. Leukemia [Internet]. 2014 Jul;28(7):1494–500. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24496303

97. Sanger F, Nicklen S, Coulson a R. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A [Internet]. 1977 Dec;74(12):5463–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=431765&tool=pmcentrez&rendertype=abstract

98. Baum M, Bielau S, Rittner N, Schmid K, Eggelbusch K, Dahms M, et al. Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. Nucleic Acids Res [Internet]. 2003 Dec 1;31(23):e151. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14627841

99. Brecqueville M, Rey J, Bertucci F, Coppin E, Finetti P, Carbuccia N, et al. Mutation analysis of ASXL1, CBL, DNMT3A, IDH1, IDH2, JAK2, MPL, NF1, SF3B1, SUZ12, and TET2 in myeloproliferative neoplasms. Genes Chromosomes Cancer [Internet]. 2012 Aug;51(8):743–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22489043

100. Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional

genomics. Genomics [Internet]. 2008 Nov;92(5):255–64. Available from: http://dx.doi.org/10.1016/j.ygeno.2008.07.001

101. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics [Internet]. 2012 Jul 24;13(1):341. Available from: http://dx.doi.org/10.1186/1471-2164-13-341%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/22827831%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3431227

102. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood [Internet]. 2009 Jul 30;114(5):937–51. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2009-03-209262

103. Szabo S, Jaeger K, Fischer H, Tschachler E, Parson W, Eckhart L. In situ labeling of DNA reveals interindividual variation in nuclear DNA breakdown in hair and may be useful to predict success of forensic genotyping of hair. Int J Legal Med [Internet]. 2012 Jan 8;126(1):63–70. Available from: http://link.springer.com/10.1007/s00414-011-0566-5

104. Tefferi A, Thiele J, Vannucchi AM, Barbui T. An overview on CALR and CSF3R mutations and a proposal for revision of WHO diagnostic criteria for myeloproliferative neoplasms. Leukemia [Internet]. 2014 Jul 20;28(7):1407–13. Available from: http://dx.doi.org/10.1038/leu.2014.35

105. Randi ML, Putti MC, Scapin M, Pacquola E, Tucci F, Micalizzi C, et al. Pediatric patients with essential thrombocythemia are mostly polyclonal and V617FJAK2 negative. Blood [Internet]. 2006 Nov 15;108(10):3600–2. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2006-04-014746

106. Fu R, Zhang L, Yang R. Paediatric essential thrombocythaemia: clinical and molecular features, diagnosis and treatment. Br J Haematol [Internet]. 2013 Nov;163(3):295–302. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24032343

107. Karow A, Nienhold R, Lundberg P, Peroni E, Putti MC, Randi ML, et al. Mutational profile of childhood myeloproliferative neoplasms. Leukemia [Internet]. 2015 Dec 30;29(12):2407–9. Available from: http://www.nature.com/doifinder/10.1038/leu.2015.205

108. Lundberg P, Karow A, Nienhold R, Looser R, Hao-Shen H, Nissen I, et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. Blood [Internet]. 2014 Apr 3;123(14):2220–8. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2013-11-537167

109. Konieczna I, Horvath E, Wang H, Lindsey S, Saberwal G, Bei L, et al. Constitutive activation of SHP2 in mice cooperates with ICSBP deficiency to accelerate progression to acute myeloid leukemia. J Clin Invest [Internet]. 2008 Mar;118(3):853–67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18246201

110. Seubert N, Royer Y, Staerk J, Kubatzky KF, Moucadel V, Krishnakumar S, et al. Active and inactive orientations of the transmembrane and cytosolic domains of the erythropoietin receptor dimer. Mol Cell [Internet]. 2003 Nov;12(5):1239–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14636581

111. Hilton DJ, Watowich SS, Katz L, Lodish HF. Saturation mutagenesis of the WSXWS motif of the erythropoietin receptor. J Biol Chem [Internet]. 1996 Mar 1;271(9):4699–708. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8617735

112. Malysz J, Crisan D. Correlation of JAK2 V617F mutant allele quantitation with clinical presentation and type of chronic myeloproliferative neoplasm. Ann Clin Lab Sci [Internet]. 2009;39(4):345–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19880761

113. Passamonti F, Rumi E. Clinical relevance of JAK2 (V617F) mutant allele burden. Haematologica [Internet]. 2009 Jan 1;94(1):7–10. Available from: http://www.haematologica.org/cgi/doi/10.3324/haematol.2008.001271

114. Moliterno AR, Williams DM, Rogers O, Spivak JL. Molecular mimicry in the chronic myeloproliferative disorders: reciprocity between quantitative JAK2 V617F and Mpl expression. Blood [Internet]. 2006 Dec 1;108(12):3913–5. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2006-03-008805

115. Bellosillo B, Martínez-Avilés L, Gimeno E, Florensa L, Longarón R, Navarro G, et al. A higher JAK2 V617F-mutated clone is observed in platelets than in granulocytes from essential thrombocythemia patients, but not in patients with polycythemia vera and primary myelofibrosis. Leukemia [Internet]. 2007;21(6):1331–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17361221

116. Li S, Kralovics R, De Libero G, Theocharides A, Gisslinger H, Skoda RC. Li - 2008 - Clonal heterogeneity in polycythemia vera patients with JAK2 exon12 and *JAK2-V617F* mutations-TableS1.pdf. Blood [Internet]. 2008 Apr 1;111(7):3863–6. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2007-09-111971

117. Titmarsh GJ, Duncombe AS, McMullin MF, O'Rorke M, Mesa R, De Vocht F, et al. How common are myeloproliferative neoplasms? A systematic review and meta-analysis. Am J Hematol [Internet]. 2014 Jun;89(6):581–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24971434

118. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia [Internet]. 2014 Feb;28(2):241–7. Available from: http://dx.doi.org/10.1038/leu.2013.336

119. Ciferri C, Lander GC, Maiolica A, Herzog F, Aebersold R, Nogales E. Molecular architecture of human polycomb repressive complex 2. Elife [Internet]. 2012 Oct 30;1(1):e00005. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23110252

120. Guglielmelli P, Lasho TL, Rotunno G, Score J, Mannarelli C, Pancrazzi A, et al. The number of prognostically detrimental mutations and prognosis in primary myelofibrosis : an international study of 797 patients. Leukemia [Internet]. 2014;28(9):1804–10. Available from: http://dx.doi.org/10.1038/leu.2014.76

121. Abdel-Wahab O, Pardanani A, Patel J, Wadleigh M, Lasho T, Heguy A, et al. Concomitant analysis of EZH2 and ASXL1 mutations in myelofibrosis, chronic myelomonocytic leukemia and blast-phase myeloproliferative neoplasms. Leukemia [Internet]. 2011 Jul 1;25(7):1200–2. Available from: http://www.nature.com/doifinder/10.1038/leu.2011.58

122. Abdel-Wahab O, Manshouri T, Patel J, Harris K, Yao J, Hedvat C, et al. Genetic Analysis of Transforming Events That Convert Chronic Myeloproliferative Neoplasms to Leukemias. Cancer Res [Internet]. 2010 Jan 15;70(2):447–52. Available from: http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-09-3783

123. Tuveson D a, Shaw AT, Willis N a, Silver DP, Jackson EL, Chang S, et al. Endogenous oncogenic K-ras(G12D) stimulates proliferation and widespread neoplastic and developmental defects. Cancer Cell [Internet]. 2004 Apr;5(4):375–87. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15093544

124. Prior I a, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. Cancer Res [Internet]. 2012 May 15;72(10):2457–67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22589270

125. Jutzi JS, Bogeska R, Nikoloski G, Schmid CA, Seeger TS, Stegelmann F, et al. MPN patients harbor recurrent truncating mutations in transcription factor NF-E2. J Exp Med [Internet]. 2013 May 6;210(5):1003–19. Available from: http://jem.rupress.org/content/210/5/1003%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/23589569

126. Bean TL, Ney PA. Multiple regions of p45 NF-E2 are required for -globin gene expression in erythroid cells. Nucleic Acids Res [Internet]. 1997 Jun 1;25(12):2509–15. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/25.12.2509

127. Shivdasani RA, Rosenblatt MF, Zucker-Franklin D, Jackson CW, Hunt P, Saris CJ, et al.

Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. Cell [Internet]. 1995 Jun 2;81(5):695–704. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7774011

128. Nagata Y. Proplatelet formation of megakaryocytes is triggered by autocrine-synthesized estradiol. Genes Dev [Internet]. 2003 Dec 1;17(23):2864–9. Available from: http://www.genesdev.org/cgi/doi/10.1101/gad.1128003

129. Neff T, Sinha AU, Kluk MJ, Zhu N, Khattab MH, Stein L, et al. Polycomb repressive complex 2 is required for MLL-AF9 leukemia. Proc Natl Acad Sci U S A [Internet]. 2012 Mar 27;109(13):5028–33. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1202258109

130. Shimizu T, Kubovcakova L, Nienhold R, Zmajkovic J, Meyer SC, Hao-Shen H, et al. Loss of Ezh2 synergizes with JAK2 -V617F in initiating myeloproliferative neoplasms and promoting myelofibrosis. J Exp Med [Internet]. 2016 Jul 25;213(8):1479–96. Available from: http://www.jem.org/lookup/doi/10.1084/jem.20151136

131. Rowe RG, Wang LD, Coma S, Han A, Mathieu R, Pearson DS, et al. Developmental regulation of myeloerythroid progenitor function by the LIN28B – let-7 – Hmga2 axis. J Exp Med [Internet]. 2016 Jul 25;213(8):1497–512. Available from: http://www.jem.org/lookup/doi/10.1084/jem.20151912

132. Copley MR, Babovic S, Benz C, Knapp DJHF, Beer PA, Kent DG, et al. The LIN28B-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. Nat Cell Biol [Internet]. 2013 Aug;15(8):916–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23811688

133. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature [Internet]. 2011 Jan 20;469(7330):343–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3760771&tool=pmcentrez&rendertype=abstract

134. Zhang T, Cooper S, Brockdorff N. The interplay of histone modifications - writers that read. EMBO Rep [Internet]. 2015 Nov;16(11):1467–81. Available from: http://embor.embopress.org/content/embor/16/11/1467.full.pdf

135. Xu C, Bian C, Yang W, Galka M, Ouyang H, Chen C, et al. Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). Proc Natl Acad Sci U S A [Internet]. 2010 Nov 9;107(45):19266–71. Available from: http://www.pnas.org/content/107/45/19266.abstract

136. Xu C, Min J. Structure and function of WD40 domain proteins. Protein Cell [Internet]. 2011 Mar;2(3):202–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21468892

137. Ward PS, Patel J, Wise DR, Abdel-Wahab O, Bennett BD, Coller HA, et al. The Common Feature of Leukemia-Associated IDH1 and IDH2 Mutations Is a Neomorphic Enzyme Activity Converting α-Ketoglutarate to 2-Hydroxyglutarate. Cancer Cell [Internet]. 2010 Mar;17(3):225–34. Available from: http://dx.doi.org/10.1016/j.ccr.2010.01.020

138. Pietrak B, Zhao H, Qi H, Quinn C, Gao E, Boyer JG, et al. A Tale of Two Subunits: How the Neomorphic R132H IDH1 Mutation Enhances Production of αHG. Biochemistry [Internet]. 2011 May 31;50(21):4804–12. Available from: http://pubs.acs.org/doi/abs/10.1021/bi200499m

139. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med [Internet]. 2009 Sep 10;361(11):1058–66. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3201812&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3201812&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.nejm.org/doi/full/10.1056/NEJMoa0903

140. Tefferi A, Lasho TL, Abdel-Wahab O, Guglielmelli P, Patel J, Caramazza D, et al. IDH1 and IDH2 mutation studies in 1473 patients with chronic-, fibrotic- or blast-phase essential thrombocythemia, polycythemia vera or myelofibrosis. Leukemia [Internet]. 2010 Jul

27;24(7):1302–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3035975&tool=pmcentrez&rendertype=abstract

141. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med [Internet]. 2014 Oct 19;20(12):1472–8. Available from: http://dx.doi.org/10.1038/nm.3733

142. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman P V, Mar BG, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. N Engl J Med [Internet]. 2014 Dec 25;371(26):2488–98. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa1408617

143. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose S a., Bakhoum SF, et al. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. N Engl J Med [Internet]. 2014;371(26):2477–87. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25426838%5Cnhttp://www.nejm.org/doi/abs/10.1056/NEJMoa1409405%5Cnhttp://www.nejm.org/doi/abs/10.1056/NEJMoa1409405

144. Koya J, Kataoka K, Sato T, Bando M, Kato Y, Tsuruta-Kishino T, et al. DNMT3A R882 mutants interact with polycomb proteins to block haematopoietic stem and leukaemic cell differentiation. Nat Commun [Internet]. 2016 Mar 24;7:10924. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27010239

145. Malcovati L, Porta MG Della, Pascutto C, Invernizzi R, Boni M, Travaglino E, et al. Prognostic factors and life expectancy in myelodysplastic syndromes classified according to WHO criteria: a basis for clinical decision making. J Clin Oncol [Internet]. 2005 Oct 20;23(30):7594–603. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16186598

146. Ortmann CA, Kent DG, Nangalia J, Silber Y, Wedge DC, Grinfeld J, et al. Effect of Mutation Order on Myeloproliferative Neoplasms. N Engl J Med [Internet]. 2015 Feb 12;372(7):601–12. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa1412098

147. Bielas JH, Heddle J a. Proliferation is necessary for both repair and mutation in transgenic mouse cells. Proc Natl Acad Sci U S A [Internet]. 2000 Oct 10;97(21):11391–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11005832

148. Ko M, An J, Pastor WA, Koralov SB, Rajewsky K, Rao A. TET proteins and 5-methylcytosine oxidation in hematological cancers. Immunol Rev [Internet]. 2015 Jan;263(1):6–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25510268

149. Deininger M, Radich J, Burn TC, Huber R, Paranagama D, Verstovsek S. The effect of long-term ruxolitinib treatment on JAK2p.V617F allele burden in patients with myelofibrosis. Blood [Internet]. 2015 Sep 24;126(13):1551–4. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2015-03-635235

150. Randi ML, Geranio G, Bertozzi I, Micalizzi C, Ramenghi U, Tucci F, et al. Are all cases of paediatric essential thrombocythaemia really myeloproliferative neoplasms? Analysis of a large cohort. Br J Haematol [Internet]. 2015 May;169(4):584–9. Available from: http://doi.wiley.com/10.1111/bjh.13329

151. Kucine N, Chastain KM, Mahler MB, Bussel JB. Primary thrombocytosis in children. Haematologica [Internet]. 2014 Apr 1;99(4):620–8. Available from: http://www.haematologica.org/cgi/doi/10.3324/haematol.2013.092684

152. Kilpivaara O, Mukherjee S, Schram AM, Wadleigh M, Mullally A, Ebert BL, et al. A germline JAK2 SNP is associated with predisposition to the development of JAK2V617F-positive myeloproliferative neoplasms. Nat Genet [Internet]. 2009 Mar 15;41(4):455–9. Available from: http://www.nature.com/doifinder/10.1038/ng.342

153. Tapper W, Jones A V, Kralovics R, Harutyunyan AS, Zoi K, Leung W, et al. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. Nat Commun [Internet]. 2015 Apr 7;6:6691. Available from: http://www.nature.com/doifinder/10.1038/ncomms7691

154. Guo MH, Nandakumar SK, Ulirsch JC, Zekavat SM, Buenrostro JD, Natarajan P, et al. Comprehensive population-based genome sequencing provides insight into hematopoietic regulatory mechanisms. Proc Natl Acad Sci U S A [Internet]. 2016 Dec 28; Available from: http://biorxiv.org/lookup/doi/10.1101/067934

155. Tefferi A, Guglielmelli P, Larson DR, Finke C, Wassie EA, Pieri L, et al. Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis. Blood [Internet]. 2014 Oct 16;124(16):2507–13; quiz 2615. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25037629

156. Tefferi A, Pardanani A, Lim K, Abdel-Wahab O, Lasho TL, Patel J, et al. TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. Leukemia [Internet]. 2009 May 5;23(5):905–11. Available from: http://www.nature.com/doifinder/10.1038/leu.2009.47

157. Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M, et al. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood. 2009;114(1):144–7.

158. Hussein K, Abdel-Wahab O, Lasho TL, Van Dyke DL, Levine RL, Hanson CA, et al. Cytogenetic correlates of TET2 mutations in 199 patients with myeloproliferative neoplasms. Am J Hematol [Internet]. 2010 Jan;85(1):81–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19957346

159. Patel JP, Gönen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med [Internet]. 2012 Mar 22;366(12):1079–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22417203

160. Rampal R, Ahn J, Abdel-Wahab O, Nahas M, Wang K, Lipson D, et al. Genomic and functional analysis of leukemic transformation of myeloproliferative neoplasms. Proc Natl Acad Sci [Internet]. 2014 Dec 16;111(50):E5401–10. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1407792111

161. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci [Internet]. 2011 Jun 7;108(23):9530–5. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1105422108

162. Lange V, Böhme I, Hofmann J, Lang K, Sauter J, Schöne B, et al. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. BMC Genomics [Internet]. 2014;15(1):63. Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-63

163. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature [Internet]. 2011;478(7367):64–9. Available from: http://www.nature.com/nature/journal/v478/n7367/full/nature10496.html#affil-auth%5Cnhttp://dx.doi.org/10.1038/nature10496

164. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood [Internet]. 2013 Nov 21;122(22):3616–27. Available from: http://www.bloodjournal.org/cgi/doi/10.1182/blood-2013-08-518886

165. Inoue D, Bradley RK, Abdel-Wahab O. Spliceosomal gene mutations in myelodysplasia: molecular links to clonal abnormalities of hematopoiesis. Genes Dev [Internet]. 2016;30(9):989–1001. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27151974

166. Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. RNA splicing factors as oncoproteins and tumour suppressors. Nat Rev Cancer [Internet]. 2016 Jul;16(7):413–30. Available from: http://www.nature.com/doifinder/10.1038/nrc.2016.51

167. Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SC-W, Ramakrishnan A, et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell [Internet]. 2015 May;27(5):617–30. Available from: http://dx.doi.org/10.1016/j.ccell.2015.04.006

168. Joazeiro CA. The Tyrosine Kinase Negative Regulator c-Cbl as a RING-Type, E2-Dependent

Ubiquitin-Protein Ligase. Science (80-) [Internet]. 1999 Oct 8;286(5438):309–12. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.286.5438.309

Appendix 1: Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms

MYELOID NEOPLASIA

Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms

Pontus Lundberg,[1] Axel Karow,[1] Ronny Nienhold,[1] Renate Looser,[1] Hui Hao-Shen,[1] Ina Nissen,[2] Sabine Girsberger,[3] Thomas Lehmann,[3] Jakob Passweg,[3] Martin Stern,[1,3] Christian Beisel,[2] Robert Kralovics,[4,5] and Radek C. Skoda[1,3]

[1]Experimental Hematology, Department of Biomedicine, University Hospital of Basel, Basel, Switzerland; [2]Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology, Zurich, Switzerland; [3]Division of Hematology, University Hospital Basel, Basel, Switzerland; [4]CeMM Research Center for Molecular Medicine, Austrian Academy of Sciences, Vienna, Austria; and [5]Division of Hematology and Blood Coagulation, Department of Internal Medicine I, Medical University of Vienna, Vienna, Austria

Key Points

- The total number of somatic mutations was inversely correlated with survival and risk of leukemic transformation in MPN.
- The great majority of somatic mutations were already present at MPN diagnosis, and very few new mutations were detected during follow-up.

Myeloproliferative neoplasms (MPNs) are a group of clonal disorders characterized by aberrant hematopoietic proliferation and an increased tendency toward leukemic transformation. We used targeted next-generation sequencing (NGS) of 104 genes to detect somatic mutations in a cohort of 197 MPN patients and followed clonal evolution and the impact on clinical outcome. Mutations in calreticulin (*CALR*) were detected using a sensitive allele-specific polymerase chain reaction. We observed somatic mutations in 90% of patients, and 37% carried somatic mutations other than *JAK2* V617F and *CALR*. The presence of 2 or more somatic mutations significantly reduced overall survival and increased the risk of transformation into acute myeloid leukemia. In particular, somatic mutations with loss of heterozygosity in *TP53* were strongly associated with leukemic transformation. We used NGS to follow and quantitate somatic mutations in serial samples from MPN patients. Surprisingly, the number of mutations between early and late patient samples did not significantly change, and during a total follow-up of 133 patient years, only 2 new mutations appeared, suggesting that the mutation rate in MPN is rather low. Our data show that comprehensive mutational screening at diagnosis and during follow-up has considerable potential to identify patients at high risk of disease progression.
(*Blood*. 2014;123(14):2220-2228)

Introduction

Myeloproliferative neoplasms (MPNs) are a group of stem cell disorders characterized by aberrant hematopoietic proliferation and an increased tendency toward leukemic transformation. MPNs comprise 3 major subgroups: polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF). An acquired mutation in *JAK2* (*JAK2* V617F) is present in the majority of MPN patients.[1-4] Although *JAK2* mutations have been shown to be the phenotypic drivers in MPN, there is evidence of clonality and mutational events preceding the acquisition of *JAK2* V617F.[5-8] An increasing number of mutations in genes distinct from *JAK2* have been identified in patients with MPN. These include mutations in epigenetic modifiers, such as *TET2*,[8] *DNMT3A*,[9] *ASXL1*,[10] and *EZH2*,[11] and genes involved in hematopoietic signaling (reviewed in Vainchenker et al[12]). Very recently, recurrent mutations in the calreticulin gene (*CALR*) have been reported in ET and PMF by 2 next-generation sequencing (NGS) whole-exome studies.[13,14] In addition, novel recurrent mutations occurring at low frequencies have been also found in *CHEK2*, *SCRIB*, *MIR662*, *BARD1*, *TCF12*, *FAT4*, *DAP3*, and *POLG*.[14,15] Mutations in *TP53*, *TET2*, *SH2B3*, and *IDH1* are more frequently observed in leukemic blasts from transformed MPN patients, suggesting a role for these gene mutations in leukemic transformation.[16-19] However, so far only mutations in *ASXL1* and *NRAS* have been shown to be of prognostic value in patients with PMF.[15,20]

Using targeted NGS to search for mutations in 104 cancer-related genes, we have defined the mutational profile of a cohort of 197 MPN patients and dissected the temporal order of acquisition and clonal architecture of mutational events. We further analyzed the impact of the somatic mutations on clinical outcome. We provide evidence that most somatic mutations were present already at MPN diagnosis. In addition, we show that somatic mutations in *TP53* and *TET2* are associated with decreased overall survival and increased risk for leukemic transformation. Importantly, mutations in *TP53* were present for several years in the chronic MPN phase at a low allelic burden, whereas after loss of the wild-type (WT) *TP53* allele, the clone rapidly expanded, resulting in leukemic transformation.

Submitted November 6, 2013; accepted January 19, 2014. Prepublished online as *Blood* First Edition paper, January 29, 2014; DOI 10.1182/blood-2013-11-537167.

P.L. and A.K. contributed equally to this study.

The online version of this article contains a data supplement.

There is an Inside *Blood* commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

Table 1. Clinical characteristics of the MPN patients at diagnosis

Diagnosis	PV	ET	PMF
Number of patients	94	69	34
% female	51	67	26
Average age at diagnosis (range), y	58 (18-87)	51 (21-86)	61 (21-86)
Average time of follow-up, mo	92	56	49
Hemoglobin (g/L) average (range)	181 (148-225)	141 (78-225)	126 (90-161)
Platelets (10^9/L) average (range)	554 (90-1487)	994 (452-1983)	635 (16-1677)
Leukocytes (10^9/L) average (range)	12 (4-39)	9 (5-17)	11 (5-27)
Neutrophils (10^9/L) average (range)	9 (2-36)	6 (3-16)	8 (3-21)
Transformation to secondary myelofibrosis	4 (4%)	1 (1%)	NA
Transformation to AML	3 (3%)	2 (3%)	2 (6%)

NA, not applicable.

Methods

Patient cohort

The collection of blood samples and clinical data was performed at the study center in Basel, Switzerland and approved by the local Ethics Committees (Ethik Kommission Beider Basel). Written informed consent was obtained from all patients in accordance with the Declaration of Helsinki. The diagnosis of MPN was established according to the revised criteria of the World Health Organization.[21] Table 1 provides clinical data of the patients included in our study.

Illumina library preparation and target region capture

A total of 500 ng of granulocyte DNA derived from the most recent available follow-up samples of the patients was fragmented using Fragmentase (New England Biolabs), resulting in an average fragment size of ~250. The fragmented library was purified using Agencourt AMPure XP beads. Following purification, the library was end-repaired and adenylated (both enzymes from Bioo Scientific), and after each of those steps, the library was purified using Agencourt AMPure XP beads. Finally, patient-specific barcoded adapters were ligated (NEXTflex, Bioo Scientific; 48 different ones in total) and divided into duplicate samples. Subsequently, adaptor-ligated DNA from 48 patients, each assigned with a different barcode, was pooled equimolarly in duplicate tubes.

Bait design and target capture

Capture of target regions was performed using an Agilent SureSelect custom design including the targeted exons ±50 bp of flanking regions with a total size of ~0.44 Mb. Enrichment was performed using the provided Agilent protocol and capture was performed for 72 hours. Postenrichment polymerase chain reaction (PCR) was performed for 10 cycles.

Illumina sequencing and sequencing analysis

Paired-end 100-bp cycle sequencing of the captured libraries was performed using an Illumina HiSeq2000. Demultiplexed samples were mapped and analyzed using the CLC genomics workbench. Mapping was performed using a mismatch cost of 2 and insertion and deletion cost of 3 with a length fraction 0.7 and similarity fraction of 0.8. For mutational calling, the quality-based variant detection was used, using a neighborhood radius of 5, maximum gap and mismatch count of 2, minimum neighborhood phred quality of 25, and minimum central quality of 30. Minimum coverage of called regions was set at 20, and minimum variant frequency was set to 5%. Only nonsynonymous mutations were further pursued, whereas splice-site mutations were determined using the predict splice-site effect module. Average coverage of targeted regions was performed using the coverage analysis module and including only the targeted exons and not the flanking regions. Targets consistently having no coverage are displayed in supplemental Figure 2H, available on the *Blood* Web site.

To assess copy number alterations, the RNA-sequencing analysis module was used, and expression value was calculated using reads per kilobase per million. Statistical analysis was performed on proportions, and as references 5 normal controls were pooled. Genes deviating with >30% in expression value and with a P value of <.01 were considered as candidate regions.

Validation of candidate mutations

Candidate mutations observed in the Illumina screen were validated using the Ion Torrent PGM platform. Amplicons covering the regions of interest were designed with an amplicon length of 150 to 250. Sequencing adapters (IonXpress) were ligated to the amplicons using the IonXpress protocol. Final libraries were sequenced with 200-bp read length on a 318 chip. Mutation calling was performed using the torrent suite variant called using the somatic settings. A mutation was called somatic when the mutant allele burden in buccal DNA was <25% of the value observed in granulocytes. In the great majority of somatic mutations (~90%), no signal was detected in the germline control DNA.

Sanger Sequencing and AS-PCR

For a minority of amplicons, no sequencing coverage was obtained with the Ion Torrent PGM. For these regions, Sanger sequencing for mutation validation was performed according to standard protocols. Allele-specific PCR (AS-PCR) of *CALR* exon 9 was performed as previously reported.[13]

Analyses of patient cell colonies

The colony assays were performed using peripheral blood mononuclear cells from patients as previously published.[7] After 14 days, colonies were picked and analyzed individually for *JAK2* V617F using AS-PCR and for the presence of somatic mutations by Sanger sequencing, respectively. On average, 88 colonies per patient were analyzed. To determine the temporal order of mutation acquisition, at least 2 informative colonies were required. In total, we analyzed 33 patients.

Statistical analysis

Mutational status of genes mutated in ≥5 patients in the cohort was correlated with blood counts at diagnosis (hematocrit, platelets, leukocytes, and neutrophils) in generalized linear models adjusting for patient gender, disease (PV vs ET vs PMF), and age. Survival and transformation curves were estimated using the Kaplan-Meier univariate method and compared by the Mantel-Cox log-rank test. Primary end points were overall survival, defined as time between diagnosis and death by any cause, and transformation to acute myeloid leukemia (AML). Statistical analyses were performed using SPSS (version 20) and GraphPad Prism (version 6).

Results

We characterized a cohort of 200 MPN patients from whom paired granulocyte and nonhematopoietic DNA samples were available (Figure 1). Clinical characteristics of the patients at diagnosis of MPN are summarized in Table 1. Serial blood samples were available

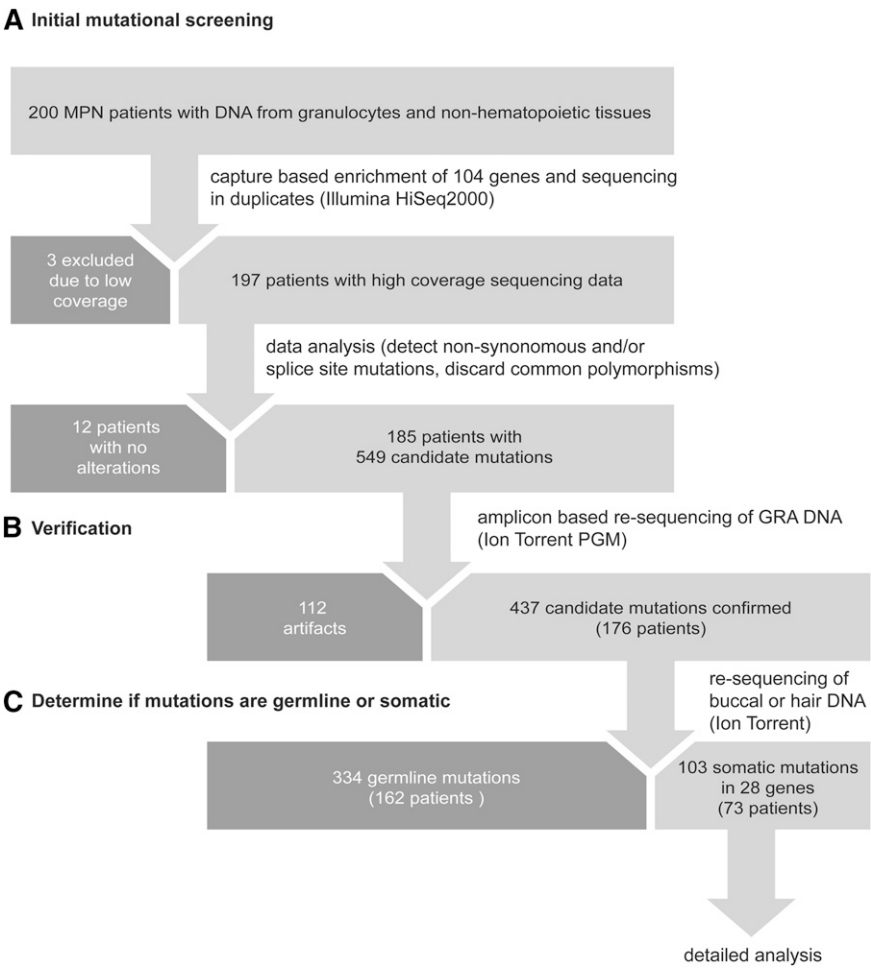


Figure 1. Targeted NGS in MPN: study design and workflow.

for 143 of 200 (72%) patients. To detect the maximal number of candidate mutations, granulocyte DNA from the most recent patient sample was used for initial sequencing. The workflow is summarized in Figure 1. We used the Agilent SureSelect method to capture exons and flanking regions of 104 selected genes with known or possible role in MPN (supplemental Figure 1A). To reduce PCR and sequencing artifacts, all DNA samples were processed and sequenced in duplicates and only sequence alterations that were present in both duplicate samples and displayed a mutant allele burden of >5% were further analyzed. The average exon coverage of Illumina sequencing per patient was 370-fold (supplemental Figure 2A), and only 3 patients had to be excluded due to insufficient coverage (Figure 1A). Resequencing of granulocyte DNA confirmed 437 of the 549 candidate mutations (80%) that were detected in the original screening (Figure 1B). Using DNA derived from buccal mucosa or hair follicles, we found that 334 of the 437 mutations were germline (76%) and 103 were somatic (24%) (Figure 1C). Furthermore, we screened our cohort also for mutation in the *CALR* gene using AS-PCR.[13] Overall, 41 of 94 PV patients (44%), 20 of 69 ET patients (29%), and 12 of 34 PMF patients (35%) carried somatic mutations other than *JAK2* V617F or *CALR*.

Frequency and distribution of mutations in patients with MPN

By NGS, we found that 28 of 104 (27%) of genes analyzed were mutated in at least 1 of the 197 MPN patients (supplemental Table 1). By AS-PCR, in addition, 17 of 69 (25%) ET patients and 11 of 34 (32%) PMF patients carried mutations in *CALR*. After *JAK2* V617F and *CALR*, the most frequently observed mutations affected genes implicated in epigenetic regulation (*TET2*, *ASXL1*, *DNMT3A*, *EZH2*, and *IDH1*) (Figure 2A). We also identified 2 novel somatic mutations in the tumor suppressor *NF1*. Furthermore, we found mutations in *NFE2*, which had only been described in one recent report,[22] and *CUX1*, uncovered previously in an MPN patient transforming to AML.[23] Recurrent somatic mutations were also observed in the genes *TP53*, *CBL*, *MPL*, and *NRAS*. Nonrecurrent mutations were detected in 16 other genes (Figure 2A). By measuring the relative read abundance of targeted regions in patients and normal controls, the NGS approach also detected copy number alterations, for example, deletions on chromosome 20q (Figure 2A and supplemental Figure 3). The distribution of mutations per patient is summarized in Figure 2B and supplemental Figure 2G. Overall, 20 of 197 patients (10%) had no detectable somatic mutation in any of the genes analyzed (9 ET, 7 PV, and 4 PMF). Two or more somatic mutations were found in 65 of 197 (33%) patients. The frequencies of somatic mutations in patients positive for either *CALR* or *JAK2* V617F are depicted in Figure 2C. Circos diagrams show the cooccurrence of all somatic mutations (Figure 2D) and the cooccurrence of events in *CALR*-positive patients (Figure 2E) and patients negative for mutation in both *JAK2* V617F and *CALR* (Figure 2F). In contrast to the recently published studies that reported *JAK2* and *CALR* mutations to be mutually exclusive,[13,14] we observed coexistence of *JAK2* V617F and *CALR* mutations in 1 ET patient (Figure 2E). This coexistence was confirmed in granulocytes from 3 independent time points 1.5 years apart.

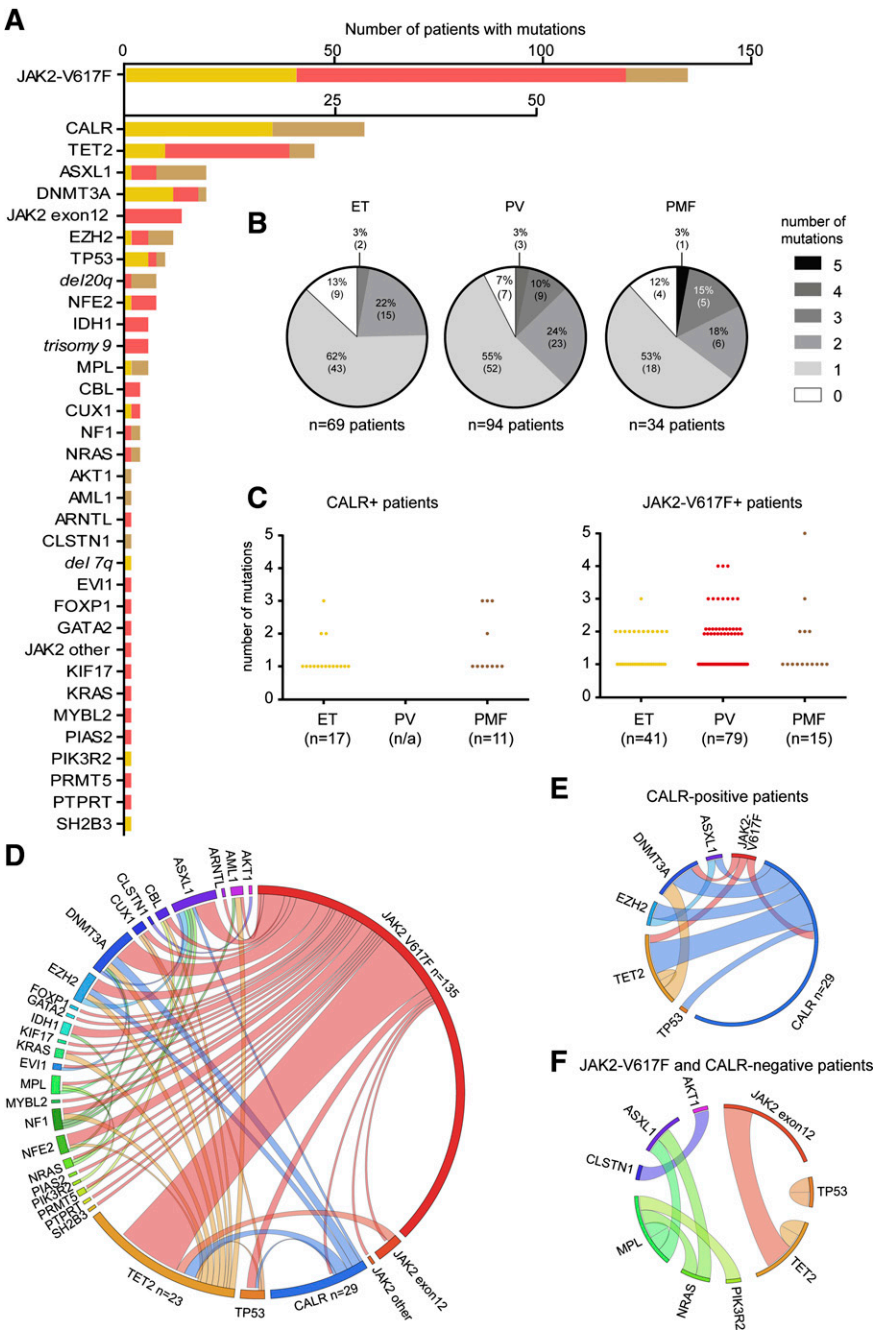


Figure 2. Frequency and distribution of mutations in patients with MPN. (A) Number of patients with mutations in the genes is indicated. ET patients are depicted in yellow, PV patients in red, and PMF patients in light brown. Numeric chromosomal aberrations are marked in italic font. (B) Distribution of somatic mutations among the 197 MPN patients according to phenotype. The shades of gray indicate the number of somatic mutations per patient. (C) Average number of somatic mutations per patient in *CALR*-positive (left panel) and *JAK2*-V617F–positive individuals (left panel) observed in ET, PV, and PMF patients, respectively. (D) Circos plot illustrating cooccurrence of somatic mutations in the same individual. The length of the arc corresponds to the frequency of the mutation, whereas the width of the ribbon corresponds to the relative frequency of co-occurrence of 2 mutations in the same patient. (E) Circos plot showing cooccurrence of somatic mutations in *CALR*-positive patients. (F) Circos plot showing co-occurrence of somatic mutations in patients negative for *JAK2* V617F and mutations in *CALR*.

Serial samples were available from 28 of the 73 patients (38%) carrying somatic mutations other than *JAK2* V617F or *CALR*. To estimate the mutation rate, we determined whether 38 somatic mutations found in the most recent sample were already present in the first patient sample that was available (Figure 3A and supplemental Figure 4). We found that the vast majority of mutations (36/38, 95%) were already detectable in the first sample and only 2 somatic mutations were acquired in a total of 133 patient years of follow-up (supplemental Figure 4). Thus, a patient would have to live ~66 years to acquire 1 mutation in the targeted region.

Clinical correlations and risk stratification

We analyzed the impact of the number of mutations other than *JAK2* V617F on survival and transformation into AML using the log-rank test for trend. We observed that increased number of somatic mutations lead to a significantly reduced overall survival and increased the risk of transformation into AML (Figure 3B). Patients with mutations in *TP53*, *TET2*, or other genes involved in epigenetic regulation (*ASXL1*, *DNMT3A*, *EZH2*, and *IDH1*) were analyzed separately (Figure 3C-E). We had serial samples from 4 out of 5 patients carrying *TP53* mutations. In these 4 patients, the *TP53* mutations were detected at a low allele burden in the first available sample and remained low for several years (Figure 3C). After loss of the WT allele through mitotic recombination or deletion, the hemi- or homozygous *TP53* clone expanded rapidly in 3 out of 4 patients, and these 3 patients transformed to AML (Figure 3C), whereas 1 patient remained stable at a low allelic burden. The fifth patient with *TP53* mutation (from whom serial samples were not available) also

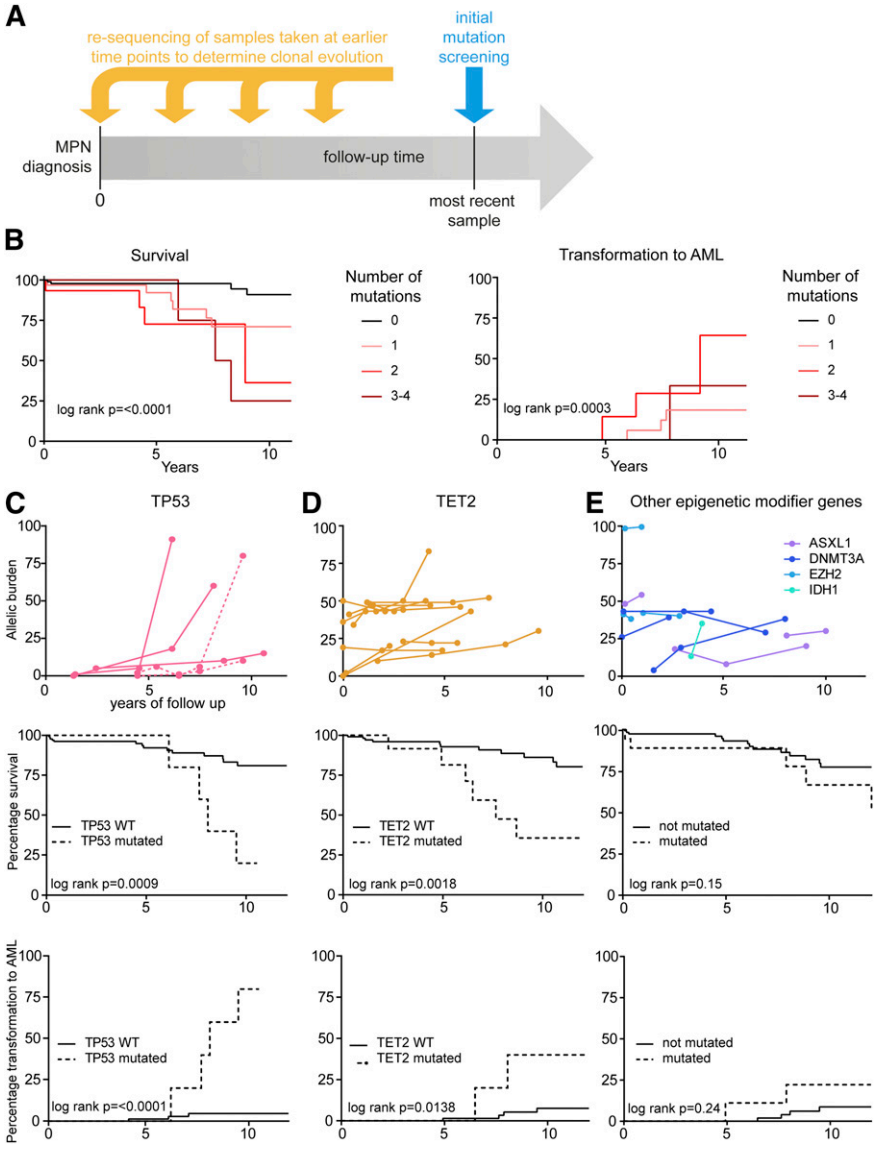


Figure 3. Analysis of sequential samples: clinical correlations and risk stratification. (A) Scheme of resequencing of mutations in serial samples to determine the time of acquisition and clonal evolution. (B) Kaplan-Meier curves for the probabilities of survival (left panel) and transformation into AML (right panel). Numbers indicate the number of somatic mutations per patient omitting *JAK2* V617F and *CALR* mutations. (C) Time course of the TP53 mutant allele burden in serial follow-up samples of 4 MPN patients with available follow-up samples (upper panel). One patient harbored 2 distinct *TP53* mutations (dotted lines), only one of which displayed loss of heterozygosity. Survival (middle panel) and transformation to AML (lower panel) is shown below for 5 patients with mutations in *TP53*. (D) Time course of the *TET2* mutant allele burden in serial follow-up samples of 12 MPN patients (upper panel). Survival (middle panel) and transformation to AML (lower panel) is shown below for 23 patients with mutations in *TET2*. (E) Time course of the mutant allele burden of epigenetic modifiers (*ASXL1*, *DNMT3A*, *EZH2*, and *IDH1*) in serial follow-up samples of 11 MPN patients (upper panel). Survival (middle panel) and transformation to AML (lower panel) is shown below for 29 patients with mutations in *ASXL1*, *DNMT3A*, *EZH2*, or *IDH1*.

transformed to AML. Serial blood samples were also available in 12 out of 23 patients with mutations in *TET2*, and in 11 of these patients, the *TET2* mutation was already present in the initial sample. Patients carrying *TET2* mutations had significantly reduced overall survival and an increased risk of leukemic transformation (Figure 3D). The number of individual patients with mutations in *DNMT3A*, *ASXL1*, *EZH2*, or *IDH1* was low, and when combined as a group, these patients showed no significant differences in the clinical course (Figure 3E). Thus, only 2 patients acquired a mutation during follow-up. One of these patients was treated with hydroxyurea (*TP53* mutation), whereas the second patient was treated with aspirin only (*TET2* mutation).

In addition, we observed correlations between mutation status and blood counts at diagnosis. Patients with an increased number of somatic mutations had a significantly higher leukocyte count (supplemental Figure 5A). Moreover, individuals with *ASXL1* mutations had significantly lower hemoglobin levels than their WT counterparts (supplemental Figure 5B), whereas patients carrying *EZH2* mutations had a significantly increased leukocyte count (supplemental Figure 5C).

Clonal evolution

For clonal analyses, we focused on patients carrying mutations in epigenetic modifier genes. To address the temporal order of acquisition, we genotyped DNA from single colonies grown in methylcellulose and plotted the results for each of the colonies analyzed (Figure 4A). The mutations could be classified as occurring before, after, or in a clone separate from *JAK2* V617F, and 1 example for each of these patterns is shown in Figure 4A. The results from all patients analyzed are summarized in Figure 4B. We found that mutations in *TET2* and *DNMT3A* were predominantly acquired before *JAK2* V617F or coexisted as separate clones (biclonal disease). Mutations in *ASXL1* and *EZH2* occurred before, after, or separate from *JAK2* V617F, whereas in 3 patients, *IDH1* mutation occurred exclusively after *JAK2* V617F (Figure 4B).

For patients with 3 or more somatic mutations, the results from single-colony analyses are shown in Figure 4C. In patients carrying mutations in *JAK2* V617F and epigenetic modifier genes or mutations in *TET2* or *DNMT3A* were predominately acquired as the first event. In 2 patients, *CALR* mutations were acquired first

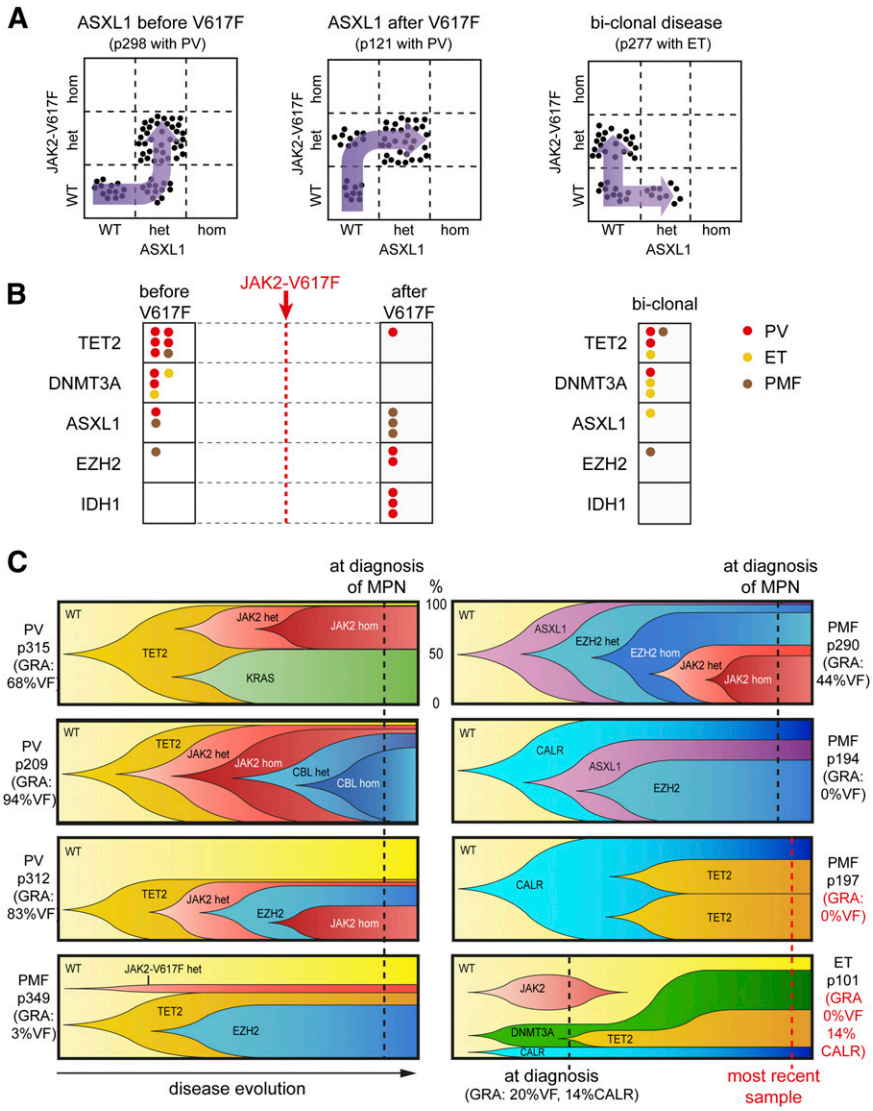


Figure 4. Clonal evolution in MPN patients carrying somatic mutations in epigenetic modifier genes. Single erythroid or granulocytic colonies (BFU-Es and CFU-G) grown in methylcellulose were individually picked and analyzed for the presence or absence of *JAK2* V617F and other somatic mutations. (A) Examples of 3 patients who acquired an *ASXL1* mutation before *JAK2* V617F (left panel), after *JAK2* V617F (middle panel), or in a clone separate from *JAK2* V617F (right panel) are shown. Each dot represents a single colony that was genotyped and placed into the corresponding quadrant. (B) Summary of the temporal order of acquisition of mutations in relation to *JAK2* V617F. Each dot represents 1 patient analyzed as shown in panel A and placed into the corresponding quadrant. Events in ET patients are depicted in yellow, PV patients in red, and PMF patients in brown. (C) Patterns of clonal evolution in 8 MPN patients carrying multiple somatic mutations. Dotted lines denote the time of analysis and the y-axis indicates the percentage of the colonies with or without the corresponding somatic mutations. %VF, *JAK2*-V617F mutant allele burden in purified granulocytes from peripheral blood. Although the order of events depicted can be deduced from the single-clone analysis (dotted line), the exact timing of the acquisition of the individual mutations and the time needed for the clonal expansion remains unknown and is shown only schematically. GRA, granulocytes.

and were present in all colonies examined (p194 and p197), whereas 1 patient (p101) displayed a complex pattern with 3 separate clones at diagnosis with disappearance of the *JAK2* V617F clone during follow-up (Figure 4C). Interestingly, all 7 patients from whom a sample at diagnosis was available already showed a complex mutational pattern in the single-clone analysis (Figure 4C).

Discussion

We used targeted NGS and AS-PCR to assess mutation profiles of 105 genes in a cohort of 197 MPN patients. Our results provide unique insights into the genomic landscape of MPN, its clonal evolution, and correlation with clinical outcomes.

We found that 90% of all MPN patients carried at least 1 somatic mutation. *JAK2* V617F was the most frequent recurrent somatic mutation (69%), followed by *CALR* (15%), *TET2* (12%), *ASXL1* (5%), and *DNMT3A* (5%). These frequencies are similar to those recently reported in an exome study of MPN patients.[14] Mutations with a low allelic burden frequently affected genes considered late events in MPN pathogenesis, such as *TP53*, *IDH1*, and *KRAS/NRAS*.

Our study also examined the longitudinal evolution of mutations in serial samples from patients with MPN using a sensitive NGS approach. Based on the comparison of the sequences in the first-available and the most recent patient samples, we estimated the overall mutation rate in the 104 genes examined to be 1 somatic mutation per 66 patient years (supplemental Figure 4). We also did not observe any de novo *JAK2* V617F mutations in patients that were *JAK2* V617F negative at diagnosis during a follow-up of 116 patient years (supplemental Figure 4). The mutation rate on a cohort basis was then calculated by dividing the age of the patients in years (at the time when the most recent sample was taken) by the number of somatic mutations in the 105 genes found in this sample by NGS. This analysis yielded a mutation rate of 1 somatic mutation per 45 patient years, which is fairly close to the result obtained by the longitudinal analysis (1/66 patient years). These observations do not support the presence of a strong hypermutable state in MPN[24,25] and also question the magnitude of the genomic instability caused by expressing *JAK2* V617F.[26-28] Consistent with the low mutation rate that we observed, a recent exome-based

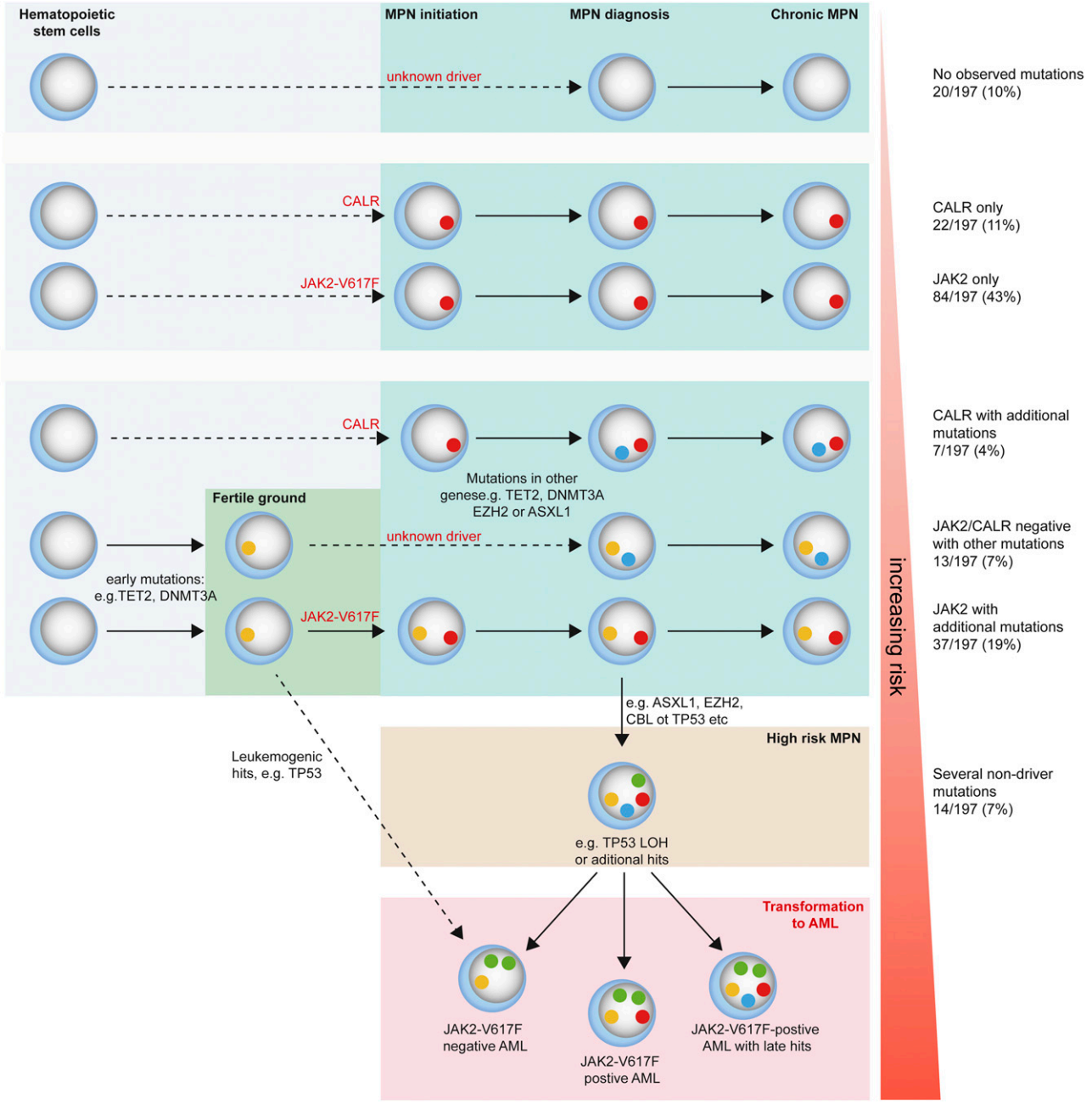


Figure 5. Model of MPN disease evolution and risk stratification in correlation to mutational events.

study detected ~0.2 somatic mutations per Mb in 151 MPN patients, showing that MPN has a low frequency of somatic mutations compared with other malignancies (eg, 0.37 mutations per Mb for AML and ~1 mutation per Mb for multiple myeloma).[14,29]

Our analyses illustrate that one of the strongest predictors of outcome is the number of somatic mutations that occur in addition to *JAK2*, *CALR*, or *MPL* (Figure 3B). Interestingly, in our cohort, the group of patients carrying either no detectable somatic mutation or a mutation in *JAK2*, *CALR*, or *MPL* only had a particularly favorable prognosis. None of these patients showed leukemic transformation, suggesting that most genes with prognostic relevance are part of the gene set that we analyzed. In a study with a similar design in MDS patients, an association of time to AML transformation and number of mutations was found.[30] We observed that mutations in *TP53* and *TET2* were associated with particularly poor outcome (Figure 3C and 3D). *TET2* mutations were recently reported as negative prognostic markers in patients with intermediate-risk AML,[31] and although a previous study in MPN showed no correlation between *TET2* mutational status and survival,[32] other studies found an increased incidence of *TET2* mutation in blasts from patients with leukemic transformation.[16] For *TP53*, we observed that mutations were present in a heterozygous state for an extended period of time during the chronic MPN phase without clonal expansion. However, after loss of the WT allele by either chromosomal deletion or uniparental disomy, the hemi- or homozygous *TP53* clone rapidly expanded, ultimately leading to leukemic transformation (Figure 3C). Thus, patients with *TP53* mutations represent a high-risk group, and screening for *TP53* mutations in MPN patients should be considered.

Because in the chronic phase of MPN the allelic burden of *TP53* mutation was <15%, sensitive methods such as NGS are needed for reliable detection. The prediction of disease progression by *TP53* mutations corresponding to our observations was described previously for patients with low-risk MDS[33] and chronic lymphocytic leukemia.[34] One recent study reported *TP53* mutations to be frequent in leukemic blasts of transformed MPN patients,[17] whereas in chronic MPN from the same cohort, 2 out of 65 patients carried monoallelic *TP53* mutations.

By dissecting the clonal architecture of patients carrying ≥3 distinct somatic mutations, we found diverse patterns, some compatible with a linear acquisition of mutations, but also several cases with an apparent biclonal structure (Figure 4). Overall, such a biclonal pattern was found in 7 of 33 patients (21%), which illustrates that in many patients, the clonal architecture cannot be imputed using allele burden of mutations alone. In general, mutations in *TET2* and *DNMT3A* were early genetic events acquired before *JAK2* V617F, whereas mutations in *ASXL1*, *EZH2*, or *IDH1* were often acquired after *JAK2* V617F. In contrast, mutations in *CALR* appeared to be an early event in the limited number of patients analyzed, consistent with previous reports.[13,14] One patient (p101) showed a complex 3 clonal pattern, and the *CALR* clone was present only in a minority of the colonies (Figure 4C).

Based on our data and previous studies, a model is presented in Figure 5. With the current methodologies, 10% of MPN patients show no detectable somatic mutations in the 105 genes analyzed (top row). In an additional 54% of patients, *JAK2* V617F or *CALR* were the only detected mutations. These 64% of MPN patients in our cohort displayed the most favorable prognosis and the lowest risk of disease progression. In the remaining 36% of MPN patients, we detected combinations of >1 somatic mutation, and in some patients, we can define the stage and order of acquisition. In the case of *JAK2*-V617F–positive MPN, often a somatic mutation occurred before the acquisition of *JAK2* V617F, compatible with providing a "fertile ground" for MPN disease initiation.[35] In contrast, *CALR* mutations appear to be the initiating event that could be followed by mutations in same set of genes as observed in *JAK2*-positive MPN. Patients with multiple mutations formed a high-risk category, with increased risk of transformation and reduced survival. Although based on a limited number of patients, the acquisition of *TP53* appears to be a particularly unfavorable event, and loss of heterozygosity was invariably associated with progression to AML.

Acknowledgments

The authors thank Ferdinand Martius and Sandrine Meyer for contributing patient samples and clinical data and Jürg Schwaller and Claudia Lengerke for helpful comments on the manuscript.

This work was supported by grants from the Gebert Rüf Foundation, the Swiss National Science Foundation (310000-120724/1 and 32003BB_135712/1), the Else Kröner-Fresenius Foundation, and the Swiss Cancer League (KLS-02398-02-2009) (R.C.S.), by a junior scientist scholarship from the University of Basel (A.K.), and by a grant from the Austrian Science Fund (FWF4702-B20) (R.K.).

Authorship

Contribution: P.L. designed and performed research, analyzed data, and wrote the paper; A.K. and R.N. performed research, analyzed data, and wrote the paper; R.L., H.H.S., I.N., and C.B. performed research; S.G., T.L., J.P., and M.S. provided clinical data and analyzed results; R.K. designed research and analyzed data; and R.C.S. designed research, analyzed data, and wrote the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

The current affiliation for T.L. is Kantonsspital St. Gallen, St. Gallen, Switzerland.

Correspondence: Radek C. Skoda, Department of Biomedicine, Experimental Hematology, University Hospital Basel, Hebelstrasse 20, 4031 Basel, Switzerland; e-mail: radek.skoda@unibas.ch.

References

1. Kralovics R, Passamonti F, Buser AS, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med.* 2005; 352(17):1779-1790.
2. Levine RL, Wadleigh M, Cools J, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell.* 2005;7(4):387-397.
3. James C, Ugo V, Le Couédic JP, et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature.* 2005;434(7037):1144-1148.
4. Baxter EJ, Scott LM, Campbell PJ, et al. Cancer Genome Project. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet.* 2005;365(9464):1054-1061.
5. Kralovics R, Teo SS, Li S, et al. Acquisition of the V617F mutation of JAK2 is a late genetic event in a subset of patients with myeloproliferative disorders. *Blood.* 2006;108(4):1377-1380.
6. Schaub FX, Looser R, Li S, et al. Clonal analysis of TET2 and JAK2 mutations suggests that TET2 can be a late event in the progression of myeloproliferative neoplasms. *Blood.* 2010; 115(10):2003-2007.
7. Schaub FX, Jager R, Looser R, et al. Clonal analysis of deletions on chromosome 20q and JAK2-V617F in MPD suggests that del20q acts independently and is not one of the pre-disposing mutations for JAK2-V617F. *Blood.* 2009;113(9): 2022-2027.
8. Delhommeau F, Dupont S, Della Valle V, et al. Mutation in TET2 in myeloid cancers. *N Engl J Med.* 2009;360(22):2289-2301.
9. Abdel-Wahab O, Pardanani A, Rampal R, Lasho TL, Levine RL, Tefferi A. DNMT3A mutational analysis in primary myelofibrosis, chronic myelomonocytic leukemia and advanced phases of myeloproliferative neoplasms. *Leukemia.* 2011; 25(7):1219-1220.
10. Carbuccia N, Murati A, Trouplin V, et al. Mutations of ASXL1 gene in myeloproliferative neoplasms. *Leukemia.* 2009;23(11):2183-2186.
11. Ernst T, Chase AJ, Score J, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet.* 2010; 42(8):722-726.
12. Vainchenker W, Delhommeau F, Constantinescu SN, Bernard OA. New mutations and pathogenesis of myeloproliferative neoplasms. *Blood.* 2011;118(7):1723-1735.
13. Klampfl T, Gisslinger H, Harutyunyan AS, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med.* 2013;369(25):2379-2390.
14. Nangalia J, Massie CE, Baxter EJ, et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med.* 2013; 369(25):2391-2405.
15. Tenedini E, Bernardis I, Artusi V, et al. Targeted cancer exome sequencing reveals recurrent mutations in myeloproliferative neoplasms [published online ahead of print October 22, 2013]. *Leukemia.*
16. Abdel-Wahab O, Manshouri T, Patel J, et al. Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Res.* 2010;70(2):447-452.
17. Harutyunyan A, Klampfl T, Cazzola M, Kralovics R. p53 lesions in leukemic transformation. *N Engl J Med.* 2011;364(5):488-490.
18. Pardanani A, Lasho T, Finke C, Oh ST, Gotlib J, Tefferi A. LNK mutation studies in blast-phase myeloproliferative neoplasms, and in chronic-phase disease with TET2, IDH, JAK2 or MPL mutations. *Leukemia.* 2010;24(10): 1713-1718.
19. Pardanani A, Lasho TL, Finke CM, Mai M, McClure RF, Tefferi A. IDH1 and IDH2 mutation analysis in chronic- and blast-phase myeloproliferative neoplasms. *Leukemia.* 2010;24(6):1146-1151.

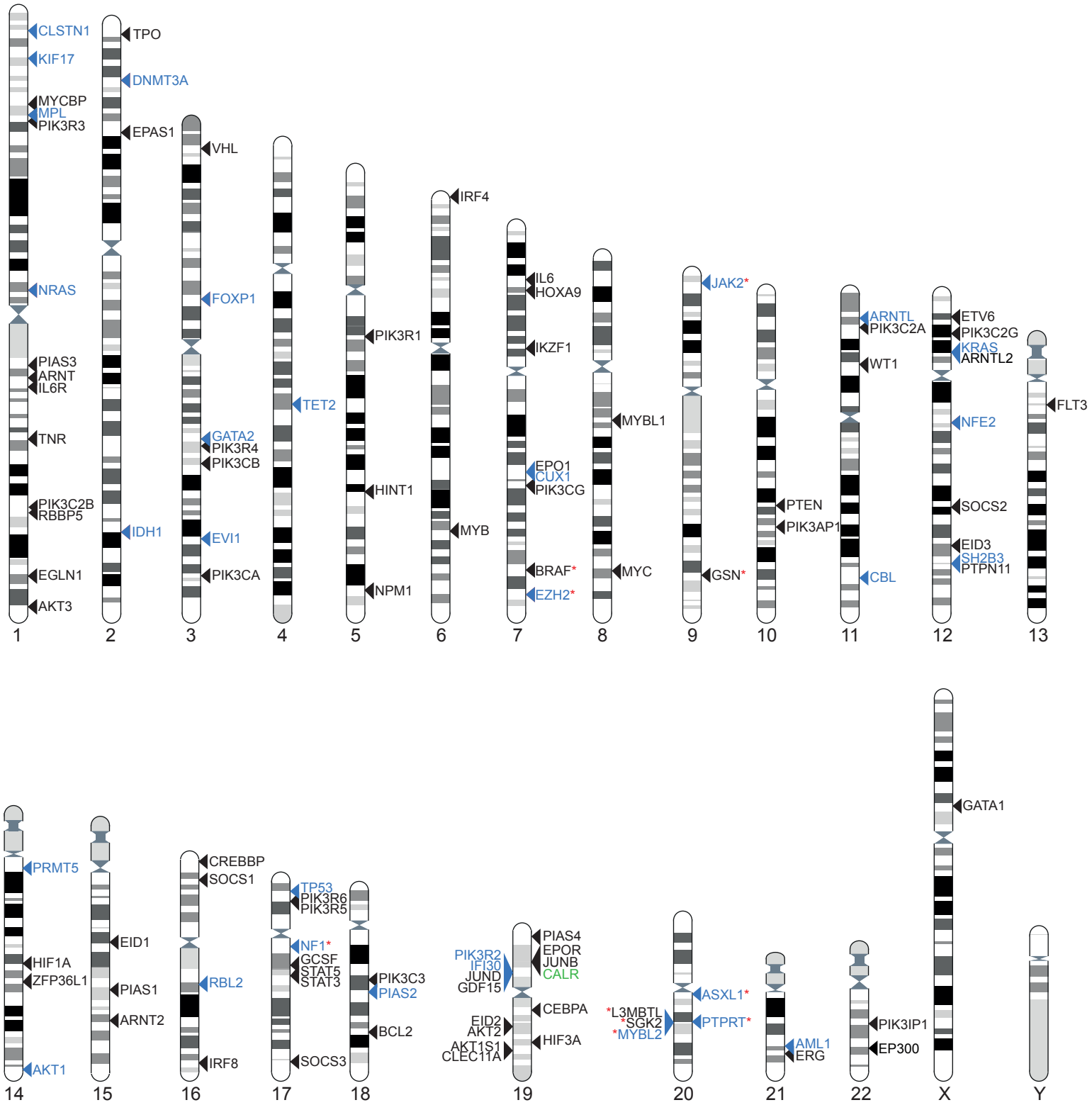
20. Vannucchi AM, Lasho TL, Guglielmelli P, et al. Mutations and prognosis in primary myelofibrosis. *Leukemia.* 2013;27(9):1861-1869.
21. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the World Heatlh Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood.* 2009;114(5):937-951.
22. Jutzi JS, Bogeska R, Nikoloski G, et al. MPN patients harbor recurrent truncating mutations in transcription factor NF-E2. *J Exp Med.* 2013; 210(5):1003-1019.
23. Thoennissen NH, Lasho T, Thoennissen GB, Ogawa S, Tefferi A, Koeffler HP. Novel CUX1 missense mutation in association with 7q- at leukemic transformation of MPN. *Am J Hematol.* 2011;86(8):703-705.
24. Jones AV, Chase A, Silver RT, et al. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet.* 2009;41(4):446-449.
25. Jones AV, Cross NC. Inherited predisposition to myeloproliferative neoplasms. *Ther Adv Hematol.* 2013;4(4):237-253.
26. Marty C, Lacout C, Droin N, et al. A role for reactive oxygen species in JAK2 V617F myeloproliferative neoplasm progression. *Leukemia.* 2013;27(11):2187-2195.
27. Plo I, Nakatake M, Malivert L, et al. JAK2 stimulates homologous recombination and genetic instability: potential implication in the heterogeneity of myeloproliferative disorders. *Blood.* 2008;112(4):1402-1412.
28. Li J, Spensberger D, Ahn JS, et al. JAK2 V617F impairs hematopoietic stem cell function in a conditional knock-in mouse model of JAK2 V617F-positive essential thrombocythemia. *Blood.* 2010;116(9):1528-1538.
29. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* 2013; 499(7457):214-218.
30. Papaemmanuil E, Gerstung M, Malcovati L, et al. Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood.* 2013;122(22):3616-3627, quiz 3699.
31. Patel JP, Gönen M, Figueroa ME, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med.* 2012;366(12):1079-1089.
32. Tefferi A, Pardanani A, Lim KH, et al. TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. *Leukemia.* 2009;23(5): 905-911.
33. Jädersten M, Saft L, Smith A, et al. TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression. *J Clin Oncol.* 2011;29(15):1971-1979.
34. Zenz T, Eichhorst B, Busch R, et al. TP53 mutation and survival in chronic lymphocytic leukemia. *J Clin Oncol.* 2010;28(29): 4473-4479.
35. Busque L, Patel JP, Figueroa ME, et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet.* 2012;44(11):1179-1181.

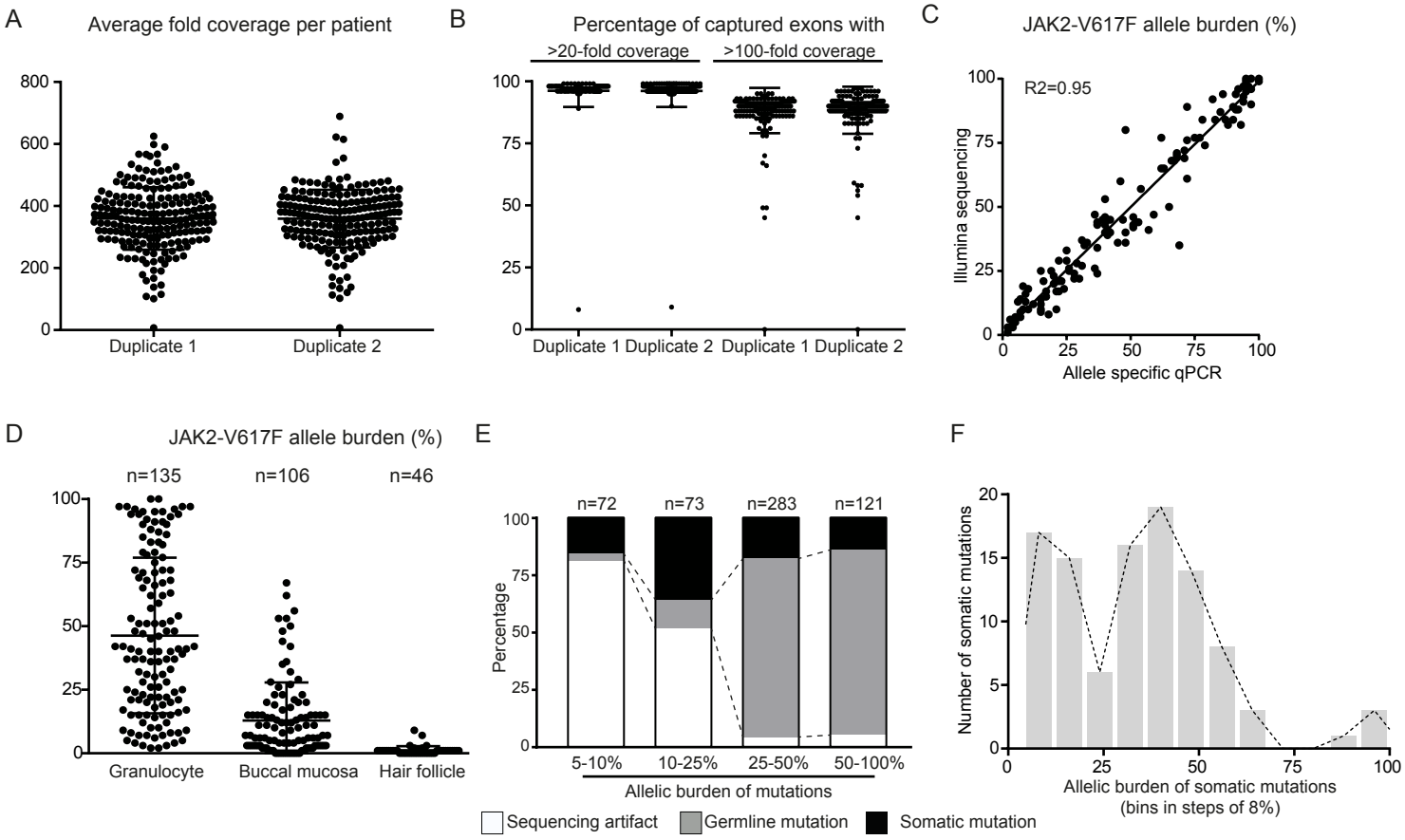
Supplemental Figure 1:

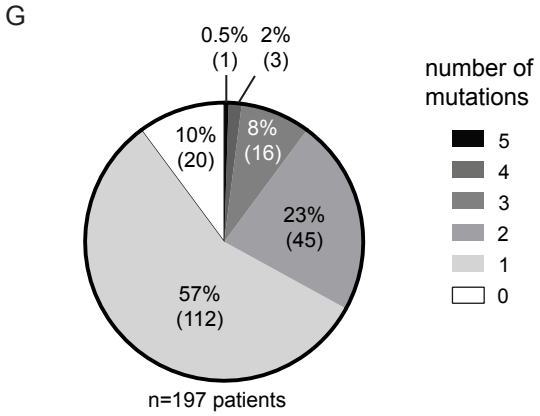
A) List of the 104 genes sequenced by NGS (CALR was analyzed by allele-specific PCR)

AKT1	CBL	ETV6	HOXA9	L3MBTL	PIAS2	PIK3R1	SH2B3
AKT1S1	CEBPA	EVI1	IDH1	MPL	PIAS3	PIK3R2	SOCS1
AKT2	CREBBP	EZH2	IFI30	MYB	PIAS4	PIK3R3	SOCS3
AKT3	CUX1	FLT3	IKZF1	MYBL1	PIK3AP1	PIK3R4	SOSC2
AKTIP	DNMT3A	FOXP1	IL6	MYBL2	PIK3C2A	PIK3R5	STAT3
AML1	EGLN1	GATA1	IL6R	MYC	PIK3C2B	PIK3R6	STAT5
ARNT	EID1	GATA2	IRF4	MYCBP	PIK3C2G	PRMT5	TET2
ARNT2	EID2	GCSF	IRF8	NF1	PIK3C3	PTEN	TNR
ARNTL	EID3	GDF15	JAK2	NFE2	PIK3CA	PTPN11	TP53
ARNTL2	PAS1	GSN	JUNB	NPM1	PIK3CB	PTPRT	TPO
ASXL1	EPO	HIF1A	JUN-D	NRAS	PIK3CD	RBBP5	VHL
BCL2	EPOR	HIF3A	KIF17	p300	PIK3CG	SCF	WT1
BRAF	ERG	HINT1	KRAS	PIAS1	PIK3IP1	SGK2	ZFP36L1
CALR							

B) Chromosomal location of genes with somatic mutations: blue, point mutations; red asterisks, copy number alterations; black, no mutations detected



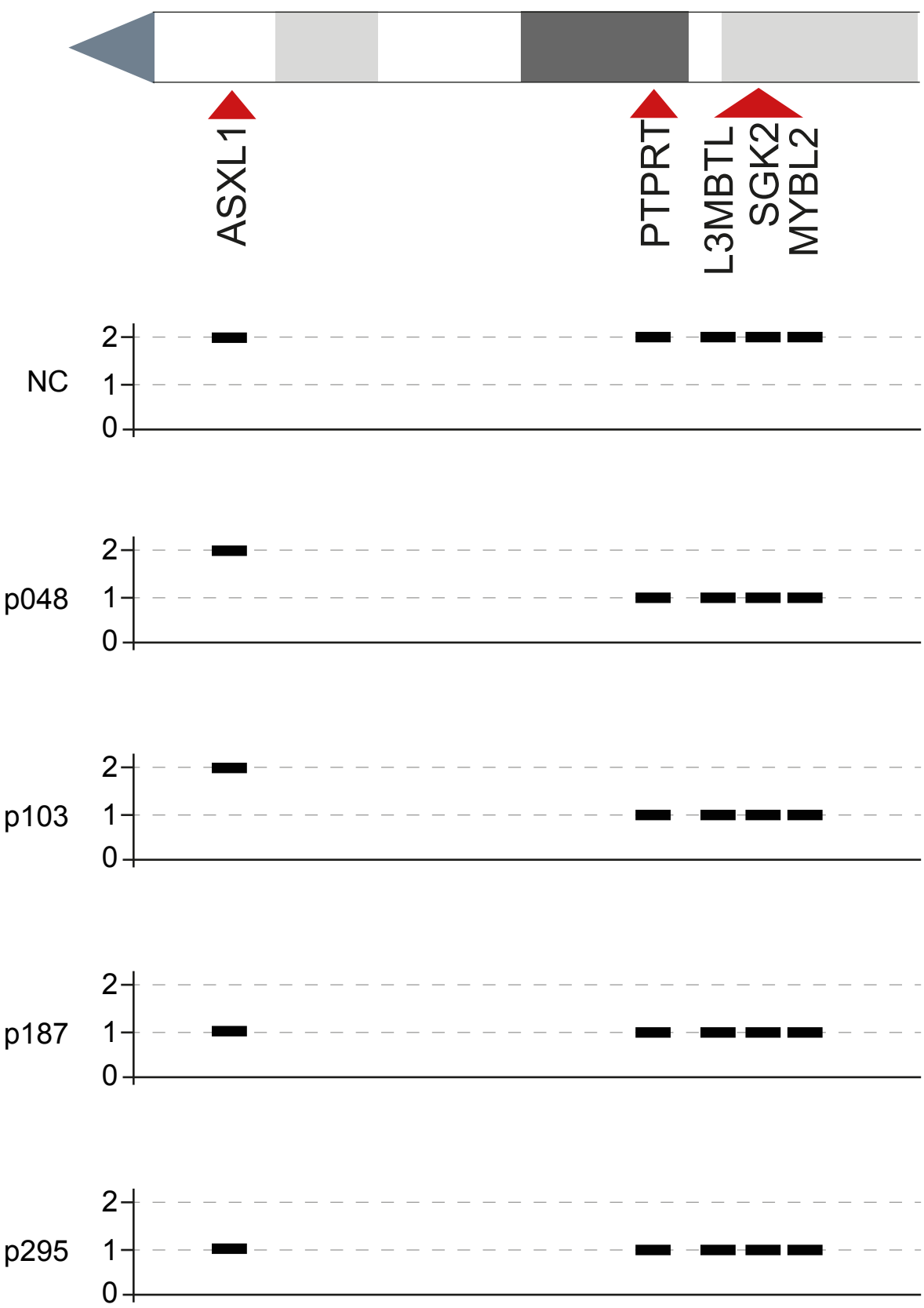




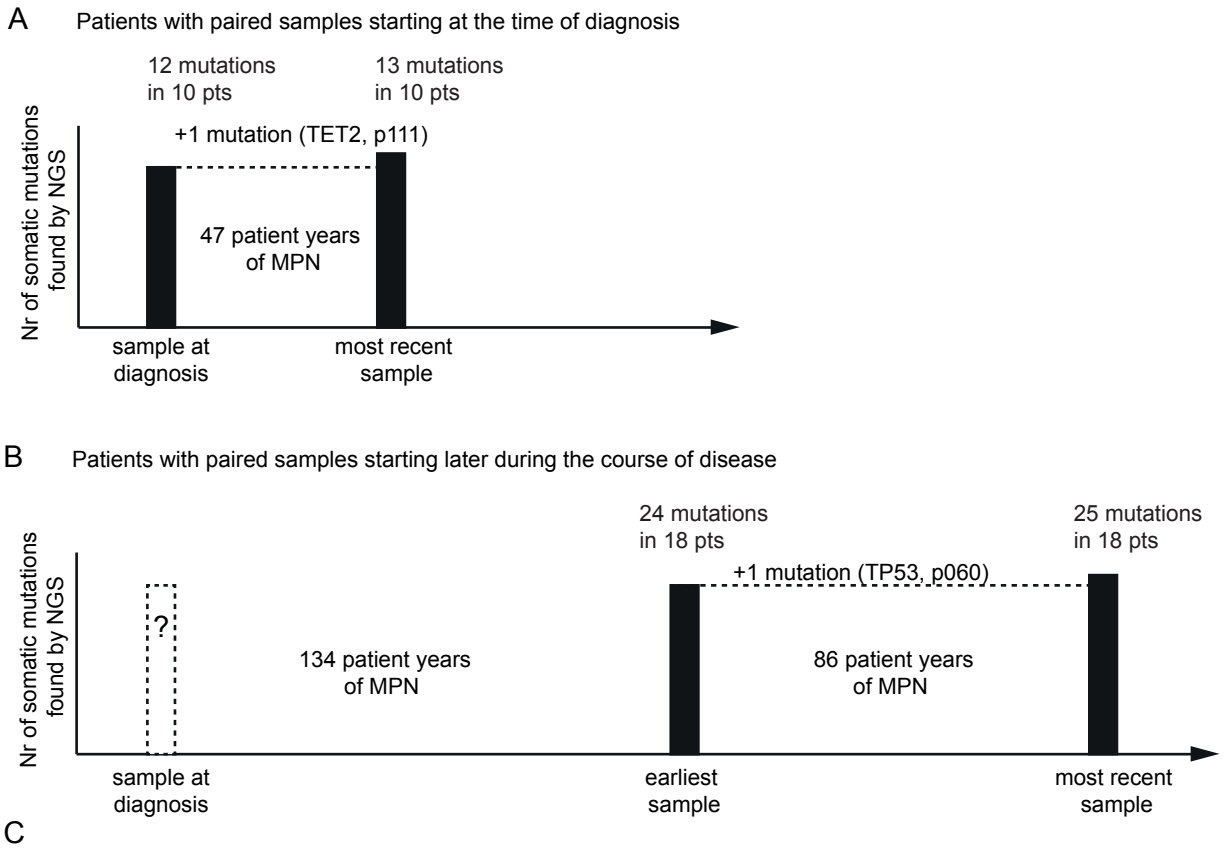
H

Gene	Chromosome	Physical position
DNMT3A	chr2	25564694..25564774
FOXP1	chr3	71338919..71339122
FOXP1	chr3	71341667..71341791
FOXP1	chr3	71633060..71633140
PIK3AP1	chr10	98392644..98392998
PIK3AP1	chr10	98429287..98429368
AKTIP	chr16	53479692..53479841
TP53	chr17	7565096..7565332
TP53	chr17	7569420..7569562
PIK3R5	chr17	8815780..8815834
MYBL2	chr20	42295708..42295943
ERG	chr21	37236582..37236765

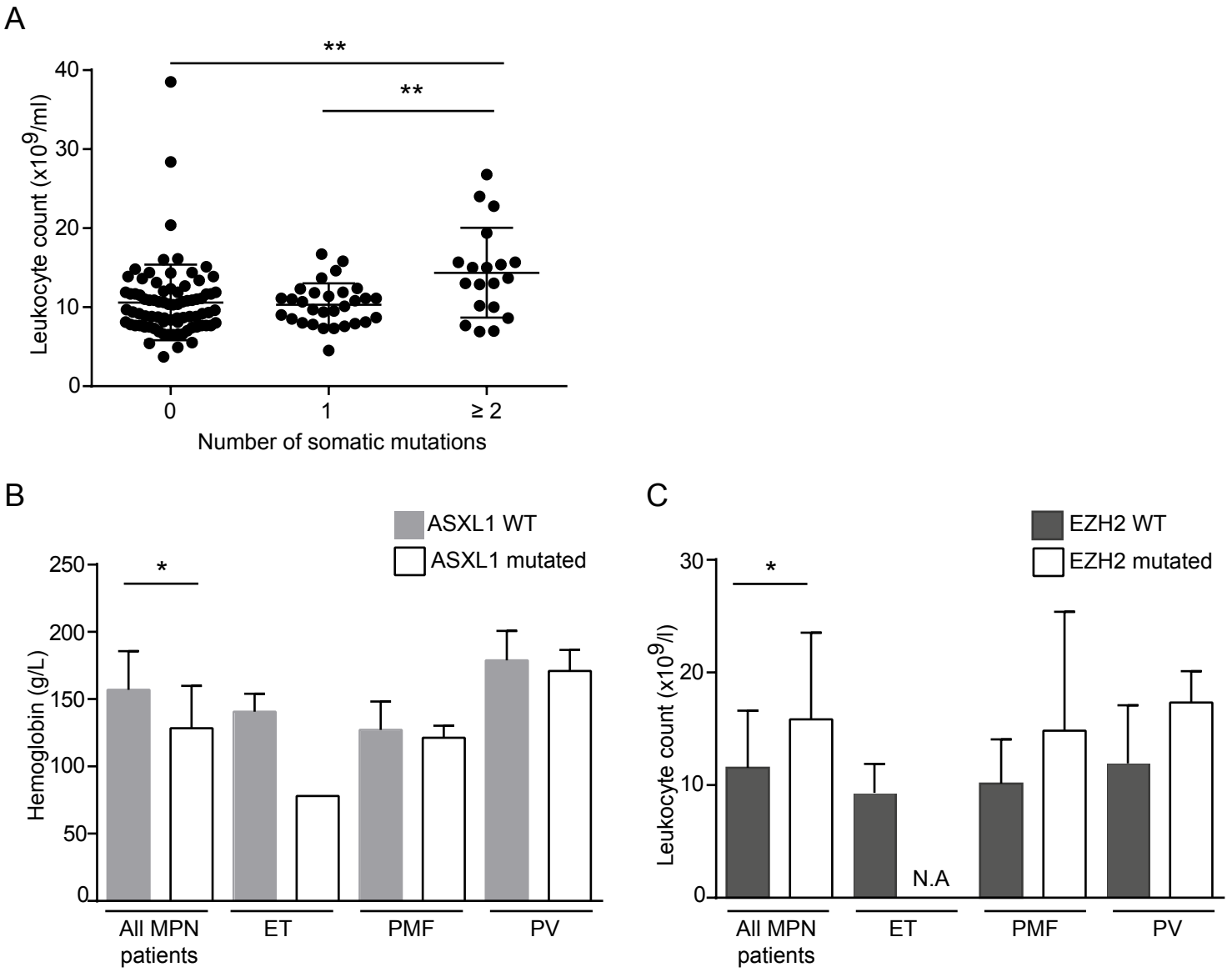
Supplemental Figure 2.
Quality control of the sequencing and validation of candidate mutations. A) Coverage represented by the average number of reads per base of target regions shown for both duplicate samples. No significant difference in the coverage was observed between duplicate samples and approximately 90% of targeted exons reached a coverage of >100-fold. B) Percentage of target regions covered by at least 20 reads (left panel) or at least 100 reads (right panel) shown for the duplicates. C) Correlation between JAK2-V617F allele burden determined by Illumina-sequencing (y-axis) and by allele-specific PCR (x-axis). The correlation was very good (R2=0.95), showing that the NGS approach can be used to quantify mutant allele burden. D) Comparison of JAK2-V617F allele burden in DNA derived from granulocytes, buccal swabs or hair follicles. JAK2-V617F was detectable in DNA from buccal mucosa indicating substantial contamination by hematopoietic cells in some of the samples. However, after re-extraction of buccal swabs taken at other time points and sequencing DNA derived from hair follicles, the germline nature of candidate mutations could be determined in the vast majority of cases. E) Percentage of sequencing artifacts (white), germline alterations (grey) and somatic mutations (black) grouped according to allele burden. Bars represent bins of allelic burden as indicated. F) Frequency distribution of somatic mutations with respect to allele burden. Bars represent bins of allelic burden in 8% steps. Note that two peaks were apparent, one around 15% and another around 40% and only a small minority of somatic mutations showed allele burden >75% that is suggestive of homozygosity. G) Frequency of somatic mutations observed in the MPN cohort H) List of targeted regions with no coverage in the majority of sequenced patients.



Supplemental Figure 3.
Copy number alterations of chromosome 20q. Patients p048 and p103 had deletion of chromosome 20q excluding ASXL1, whereas patients P187 and P295 had a deletion including ASXL1.



Supplemental Figure 4.
Duration of follow-up used for the calculation of mutation rates. A) Patients where the initial sample was taken at the time of diagnosis. B) Patients where the initial patient sample was taken at a time point more than 6 months after diagnosis. C) Scattergram depicting the time between the first and last analyzed sample for patients shown in A and B. Each dot represents one patient.



Supplemental Figure 5.
A) Correlations between the number of mutations and leukocyte count. B) Correlation between mutational status of ASXL1 and hemoglobin count. C) Corelation between the mutation status of EZH2 and leukocyte count.

Supplemental table 1

Affected gene	Chr.	Physical position on chr.	Reference	Variance	Amino acid exchange	UPN	Diagnosis
AKT1	chr 14	105239628	A	G	Met306Thr	P102	PMF
AML1	chr 21	36252877	C	T	Arg135Lys	P191	PMF
ARNTL	chr 11	13399934	C	T	Arg371Trp	P346	PV
ASXL1	chr 20	31021250	C	T	Arg417*	P290	PMF
ASXL1	chr 20	31022288	C	G	Tyr591*	P298	PV
ASXL1	chr 20	31022632	-	A	His706fs	P121	PV
ASXL1	chr 20	31022637	C	T	Gln708*	P277	ET
ASXL1	chr 20	31023162	G	T	Glu883*	P048	PMF
ASXL1	chr 20	31023408	C	T	Arg965*	P202	PMF
ASXL1	chr 20	31023408	C	T	Arg965*	P350	PMF
ASXL1	chr 20	31023437	C	G	Tyr974*	P082	PV
ASXL1	chr 20	31023625	G	A	Trp1037*	P194	PMF
ASXL1	chr 20	31024117	-	A	Gln1201fs	P178	PMF
CALR	chr 19	exon 9	-	+2 insertion		P285	ET
CALR	chr 19	exon 9	-	+5 insertion		P055	ET
CALR	chr 19	exon 9	-	+5 insertion		P080	ET
CALR	chr 19	exon 9	-	+5 insertion		P101	ET
CALR	chr 19	exon 9	-	+5 insertion		P168	ET
CALR	chr 19	exon 9	-	+5 insertion		P230	ET
CALR	chr 19	exon 9	-	+5 insertion		P232	ET
CALR	chr 19	exon 9	-	+5 insertion		P289	ET
CALR	chr 19	exon 9	-	+5 insertion		P340	ET
CALR	chr 19	exon 9	-	+5 insertion		P074	PMF
CALR	chr 19	exon 9	-	+5 insertion		P197	PMF
CALR	chr 19	exon 9	-	+5 insertion		P163	PV
CALR	chr 19	exon 9		-52 deletion		P073	ET
CALR	chr 19	exon 9		-52 deletion		P091	ET
CALR	chr 19	exon 9		-52 deletion		P109	ET
CALR	chr 19	exon 9		-52 deletion		P176	ET
CALR	chr 19	exon 9		-52 deletion		P215	ET
CALR	chr 19	exon 9		-52 deletion		P217	ET
CALR	chr 19	exon 9		-52 deletion		P224	ET
CALR	chr 19	exon 9		-52 deletion		P323	ET
CALR	chr 19	exon 9		-52 deletion		P004	PMF
CALR	chr 19	exon 9		-52 deletion		P194	PMF
CALR	chr 19	exon 9		-52 deletion		P239	PMF
CALR	chr 19	exon 9		-52 deletion		P267	PMF
CALR	chr 19	exon 9		-52 deletion		P276	PMF
CALR	chr 19	exon 9		-52 deletion		P304	PMF
CALR	chr 19	exon 9		-52 deletion		P325	PMF
CALR	chr 19	exon 9		-52 deletion		P336	PMF
CALR	chr 19	exon 9		-52 deletion		P343	PMF
CBL	chr 11	119148991	G	A	Cys404Tyr	P209	PV
CBL	chr 11	119149356	---	ATG	Tyr455delinsTyrAsp	P255	PV
CLSTN1	chr 01	9833363	C	T	Ala61Thr	P102	PMF
CUX1	chr 07	101801854	--	AA	Glu241fs	P019	PV
CUX1	chr 07	101923411	G	A	Arg588Gln	P125	ET
DNMT3A	chr 02	25457209	C	G	Trp893Ser	P325	PMF
DNMT3A	chr 02	25457242	C	T	Arg882His	P101	ET
DNMT3A	chr 02	25457242	C	T	Arg882His	P199	ET
DNMT3A	chr 02	25457243	G	A	Arg659Cys	P018	ET
DNMT3A	chr 02	25457243	G	A	Arg882Cys	P110	PV
DNMT3A	chr 02	25459837	G	A	Gln593*	P181	ET
DNMT3A	chr 02	25467083	G	A	Arg598*	P250	PV
DNMT3A	chr 02	25468129	T	-	Asn516fs	P121	PV
DNMT3A	chr 02	25469932	G	T	Tyr370*	P260	ET
DNMT3A	chr 02	25470480	C	T	Gly332Arg	P091	ET
EZH2	chr 07	148506183	T	-	Glu681fs	P312	PV
EZH2	chr 07	148506443	C	T	Arg646His	P336	PMF
EZH2	chr 07	148508788	C	T	Val582Met	P210	PV
EZH2	chr 07	148512018	A	G	Cys510Arg	P349	ET
EZH2	chr 07	148523590	C	T	Arg249Gln	P290	PMF
EZH2	chr 07	148526857	TA	-	Leu110fs	P194	PMF
FOXP1	chr 03	71021817	C	T	Arg514His	P264	PV
GATA2	chr 03	128205672	G	A	Ala68Val	P284	PV
IDH1	chr 02	209113112	C	T	Arg132His	P033	PV
IDH1	chr 02	209113112	C	T	Arg132His	P052	PV
IDH1	chr 02	209113112	C	T	Arg132His	P280	PV
JAK2	chr 09	5054790	G	A	Gly281Asp	P088	PV
JAK2	chr 09	5070024	ACA	---	His538_Lys539delinsGln	P138	PV
JAK2	chr 09	5070026	A	T	Lys539*	P216	PV
JAK2	chr 09	5070026	A	T	Lys539*	P218	PV
JAK2	chr 09	5070034	AAATGA	------	Arg541_Glu543delinsArg	P021	PV
JAK2	chr 09	5070036	ATGAAG	------	Asn542_Asp544delinsAsn	P166	PV
JAK2	chr 09	5070036	ATGAAG	------	Asn542_Asp544delinsAsn	P307	PV
JAK2	chr 09	5070037	TGAA	-	Asn542fs	P002	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P006	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P009	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P013	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P015	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P017	PMF

JAK2	chr 09	5073770	G	T	Val617Phe	P018	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P019	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P021	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P022	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P023	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P024	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P025	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P033	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P034	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P038	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P046	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P048	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P049	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P052	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P053	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P054	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P062	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P070	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P079	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P082	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P085	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P088	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P089	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P090	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P093	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P099	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P103	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P104	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P110	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P111	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P112	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P113	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P115	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P116	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P119	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P120	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P121	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P125	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P126	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P127	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P136	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P148	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P150	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P153	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P156	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P157	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P162	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P170	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P178	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P180	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P181	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P182	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P187	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P191	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P192	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P193	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P199	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P200	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P203	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P204	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P206	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P207	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P208	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P209	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P210	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P211	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P212	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P213	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P214	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P225	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P231	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P233	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P235	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P238	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P241	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P242	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P243	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P248	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P250	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P254	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P255	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P257	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P258	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P259	ET

JAK2	chr 09	5073770	G	T	Val617Phe	P260	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P264	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P265	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P266	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P268	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P270	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P274	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P277	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P280	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P282	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P284	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P287	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P288	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P290	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P291	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P295	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P297	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P298	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P300	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P303	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P305	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P306	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P310	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P311	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P312	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P315	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P317	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P320	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P324	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P326	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P328	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P329	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P332	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P333	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P337	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P338	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P339	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P342	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P344	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P345	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P346	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P347	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P349	ET
JAK2	chr 09	5073770	G	T	Val617Phe	P351	PMF
JAK2	chr 09	5073770	G	T	Val617Phe	P354	PV
JAK2	chr 09	5073770	G	T	Val617Phe	P355	PV
KIF17	chr 01	20998607	C	T	Arg849Gln	P192	PV
KRAS	chr 12	25398220	A	T	Asp33Glu	P315	PV
MECOM	chr 03	168862979	G	C	putative splicesite	P250	PV
MPL	chr 01	43815008	T	A	Trp515Arg	P319	PMF
MPL	chr 01	43815009	G	T	Trp515Leu	P193	ET
MPL	chr 01	43815009	G	T	Trp515Leu	P202	PMF
MPL	chr 01	43815009	G	A	Trp515Leu	P319	PMF
MPL	chr 01	43815009	G	T	Trp515Leu	P299	ET
MYBL2	chr 20	42320927	G	A	Gly211Ser	P329	PV
NF1	chr 17	29527539	G	C	Ala364Pro	P191	PMF
NF1	chr 17	29557401	G	T	putative splicesite	P121	PV
NFE2	chr 12	54686495	CTCT	----	Glu261fs	P126	PV
NFE2	chr 12	54686495	CTCT	----	Glu261fs	P211	PV
NFE2	chr 12	54686888	-	A	Leu131fs	P333	ET
NFE2	chr 12	54687044	G	-	Pro79fs	P033	PV
NRAS	chr 01	115258747	C	T	Gly12Asp	P202	PMF
NRAS	chr 01	115258748	C	T	Gly12Ser	P148	PV
PIAS2	chr 18	44470777	A	G	Ser89Pro	P254	PV
PIK3R2	chr 19	18272776	G	T	putative splicesite	P299	ET
PRMT5	chr 14	23391763	T	C	Met529Val	P255	PV
PTPRT	chr 20	40735551	C	A	Gly1089Trp	P282	PV
SH2B3	chr 12	111856105	T	G	His52Gln	P339	ET
TET2	chr 04	106155422	A	T	Gln108Leu	P211	PV
TET2	chr 04	106155496	C	-	Pro133fs	P265	PV
TET2	chr 04	106156729	C	T	Arg544*	P209	PV
TET2	chr 04	106157222	C	G	Ser708*	P019	PV
TET2	chr 04	106157275	C	T	Gln726*	P325	PMF
TET2	chr 04	106157346	AATAAAG	-------	Gln749fs	P111	PV
TET2	chr 04	106157527	C	T	Gln810*	P312	PV
TET2	chr 04	106157770	C	T	Gln891*	P136	PV
TET2	chr 04	106157845	C	T	Gln916*	P315	PV
TET2	chr 04	106157961	G	A	Trp954*, Trp954*	P204	ET
TET2	chr 04	106158503	G	A	Cys1135Tyr	P324	ET
TET2	chr 04	106162528	T	C	Tyr1148His	P297	PV
TET2	chr 04	106164038	CT	--	Thr1183fs	P111	PV
TET2	chr 04	106164761	T	C	Leu1231Pro	P191	PMF
TET2	chr 04	106180785	C	G	Cys1292Trp	P349	ET
TET2	chr 04	106180824	CTT	---	Ser1284_Phe1285delinsSer	P101	ET

TET2	chr 04	106182926	T	A	Leu1322Gln	P315	PV
TET2	chr 04	106182958	-	A	Met1333fs	P235	PV
TET2	chr 04	106190774	-	A	Tyr1351fs	P099	PV
TET2	chr 04	106190837	C	T	Thr1372Ile	P021	PV
TET2	chr 04	106190861	A	G	His1380Arg	P306	PV
TET2	chr 04	106193787	G	T	Val1417Phe	P311	ET
TET2	chr 04	106196233	-	G	Gln1544fs	P197	PMF
TET2	chr 04	106196267	C	T	Gln1534*	P138	PV
TET2	chr 04	106196503	TAATCCC	-	Ser1633fs	P197	PMF
TET2	chr 04	106196819	G	T	Val1718Leu	P105	PV
TET2	chr 04	106197080	-	T	Asn1826fs	P191	PMF
TET2	chr 04	106197387	T	A	Met1907Lys	P138	PV
TP53	chr 17	7577121	G	A	Arg273Cys	P017	PMF
TP53	chr 17	7577566	T	C	Asn239Asp	P224	ET
TP53	chr 17	7578203	C	T	Val216Met	P060	ET
TP53	chr 17	7578393	A	T	His179Gln	P019	PV
TP53	chr 17	7578445	A	T	Ile162Asn	P079	ET
TP53	chr 17	7578527	A	G	Cys135Arg	P060	ET

Appendix 2: Mutational profile of childhood myeloproliferative neoplasms

Leukemia (2015), 1–3

www.nature.com/leu

LETTER TO THE EDITOR

Mutational profile of childhood myeloproliferative neoplasms

Leukemia advance online publication, 18 August 2015;
doi:10.1038/leu.2015.205

Myeloproliferative neoplasms (MPN) are a group of stem cell disorders predominantly occurring in elderly,[1] whereas children are affected at much lower frequencies.[2] Therefore, less is known about the mutational spectrum and the biology of childhood MPN.[3,4] Lower incidence of *JAK2*-V617F has been reported in childhood essential thrombocythemia (ET) and polycythemia vera (PV),[5,6] and in recent studies fewer *CALR* mutations were found in children with ET.[7–10]

The clinical and laboratory data of 43 patients with pediatric MPN (age $\leqslant$18 years at diagnosis) that were included in this study are summarized in Table 1. Family history of MPN was negative in all children. The WHO (World Health Organization) 2008 criteria for ET were fulfilled in all 25 cases from whom bone marrow histology was available. To establish the diagnosis of ET in the remaining 16 patients without bone marrow examination, we used the proposed revision of the WHO criteria,[11,12] adjusted for age-specific differences in the normal blood counts.[5,13] Elevated platelet counts ($>450\times10^9$/l) for at least 12 month of follow-up and absence of signs suggesting a reactive or secondary cause were required for ET diagnosis. Data on spleen size were available for 34 patients with ET and splenomegaly was noted in 14 of them (41%). There were 5 hemorrhagic events and 1 transient ischemic attack observed in 5/41 (12%) ET patients. The two PV patients were *JAK2*-V617F positive, had hematocrit values $>$50% upon follow-up requiring phlebotomies and both had splenomegaly.

We used a capture-based targeted next-generation sequencing to simultaneously search for mutations in 104 genes.[14] DNA samples from purified granulocytes were prepared in duplicates from each patient and the exons and flanking regions of 104 selected cancer-related genes were captured using an Agilent SureSelect custom design (Agilent Technologies, Santa Clara, CA, USA). Sequencing was performed with Illumina HiSeq2000 (Illumina, Inc., San Diego, CA, USA) and sequence alterations were analyzed using the CLC genomics workbench (CLC Bio, Aarhus, Denmark). Alterations with an allele burden $>$10% detected in both DNA samples were considered as candidate mutations and validated using PCR based Ion Torrent PGM (Life Technologies, Waltham, MA, USA) resequencing. Known germline polymorphisms were excluded. In three patients DNA from buccal swabs and in one of them also hair root DNA was available to determine the presence of variants in germline DNA. Mutations in *CALR* were screened using a sensitive allele-specific PCR.[15]

The frequencies of the 45 observed sequence alterations are shown in Figure 1a. For detailed information about the individual mutations see Supplementary Table 1. *JAK2*-V617F (8/43) and *CALR* exon 9 (4/43) mutations were found most frequently (Figure 1a). In adults with MPN, the frequency of mutations in genes implicated in epigenetic regulation (*TET2*, *ASXL1*, *DNMT3A*, *EZH2* and *IDH1*) was about 25%.[14,16] In contrast, we detected mutations in these genes in only 4 of our 43 pediatric MPN patients (9%; Figure 1a). We found recurrent mutations in the *IRF8* gene, which encodes an interferon-regulatory transcription factor with a possible role as a leukemia tumor suppressor.[17] Three patients with ET had the same *IRF8*-P310A mutation, which is predicted to be deleterious by all structure prediction algorithms, and a fourth patient carried an *IRF8*-R228H mutation, where the predictions were not unanimous (Supplementary Table 1). The allele burden of the *IRF8*-P310A mutation was 99, 90 and 67%, respectively, suggesting that the mutation was homozygous in some or the majority of granulocytes in these patients. Six additional genes were mutated in two different patients each, whereas the other genes were mutated only once (Figure 1a). Two patients with ET carried mutations in the erythropoietin receptor (*EPOR*) with allele burdens close to 50% suggesting heterozygosity. One *JAK2*-V617F positive ET patient carried an EPOR-V264G mutation in the transmembrane domain of EpoR. Based on a model for mouse EpoR,[18] this mutation is expected to stabilize a less active dimeric interface for EpoR and predicted to result in a reduction of EpoR function. Another ET patient carried an EPOR-W233G mutation, which alters the first tryptophan of the conserved WSXWS motif in the extracellular domain of EpoR to GSXWS. In a mouse study, the *EpoR*-W233G mutation reduced EpoR surface expression and resulted in a loss of function of the receptor.[19] Thus, both *EPOR* mutations are predicted to reduce or eliminate EpoR function. The fact that both patients have ET and not PV, further argues against a causative role of these mutations.

Co-occurrence between mutations in the same patient is shown in Figure 1b. The four patients with *CALR* mutations did not carry additional gene mutations, whereas patients with *JAK2*-V617F or *MPL* mutations frequently carried other gene mutations. Figure 1c compares the distribution of the number of mutations per patient in our pediatric cohort with the data from our published MPN cohort of 192 adult patients that was analyzed using the same technologies.[14] Mutations in one of the established MPN driver genes *JAK2*, *CALR* or *MPL* were found in a lower percentage of pediatric cases (34%) than adult MPN patients (90%; Figure 1c). Conversely, a substantial proportion of pediatric patients who were tested negative for mutations in MPN driver genes, carried

Table 1. Clinical characteristics of the pediatric MPN patients at diagnosis

Diagnosis	*ET*	*PV*
Number of patients	41	2
Percentage of females	66%	100%
Age at diagnosis—median (range)	9 (1–18)	10 (4–17)
Hemoglobin (g/l) of all patients—median (range)	128 (80–157)	156 (153–160)
Hemoglobin males only	135 (113–157)	NA
Hemoglobin females only	125 (80–146)	156 (153–160)
Platelets (10^9/l)—median (range)	1391 (489–4443)	893 (744–1043)
Leukocytes (10^9/l)—median (range)	9 (5–17)	19 (10–27)
Splenomegaly	14/34 (41%)	2/2 (100%)
Complications (thrombotic events or hemorrhaging)	5/41 (12%)	NA

Abbreviation: NA, not applicable.

Accepted article preview online 30 July 2015

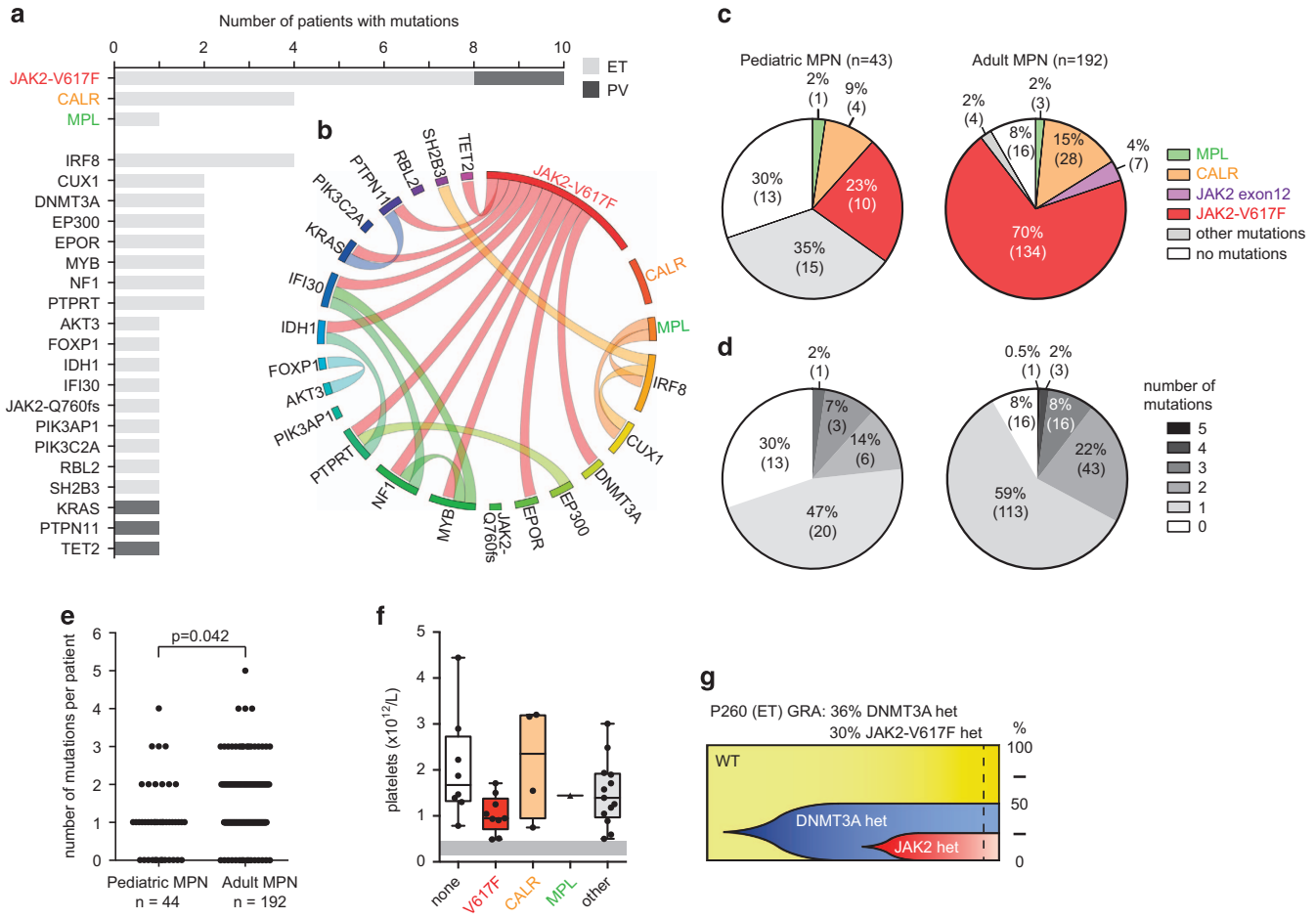


Figure 1. Molecular analysis of pediatric MPN. (**a**) Number of patients with mutations in the genes is indicated. The diagnosis of the patients is indicated in different shades of gray. (**b**) Circos plot illustrating co-occurrence of somatic mutations in the same individual. The length of the arc corresponds to the frequency of the mutation, whereas the width of the ribbon corresponds to the relative frequency of co-occurrence of two mutations in the same patient. (**c**) Comparison of the distribution of driver mutations in the pediatric cohort ($n = 43$) and a previously analyzed adult MPN cohort ($n = 192$).[14] The different colors indicate the type of driver mutation. (**d**) Distribution of somatic mutations among the same pediatric and adult MPN patients. The shades of gray indicate the number of somatic mutations per patient. (**e**) Comparison of number of mutations per patients for the pediatric and adult MPN cohort. (**f**) Comparison between platelet count and driver mutation in the pediatric MPN cohort. The gray shaded area indicates the range of normal platelet counts. (**g**) Pattern of clonal evolution in a patient mutated for *DNMT3A* and *JAK2*-V617F. The text on top of the diagram describes the percentage of affected granulocytes (GRA) based on data acquired from granulocytic DNA. Although the order of events and percentage of mutated progenitors can be deduced from the single clone analysis (dotted line), the exact timing of the acquisition of the individual mutations and the time needed for the clonal expansion remains unknown and is shown only schematically.

mutations in other genes and a higher percentage of pediatric cases had no detectable mutation in the genes analyzed (32% versus 8% in adults; Figures 1c and d). Overall, the mean number of mutations per patient in pediatric MPN was significantly lower than in adult disease (Figure 1e). The subgroup of patients without detectable mutation showed a trend toward higher platelet counts compared with patients carrying mutations (Figure 1f).

From the pediatric patient with *JAK2*-V617F and a mutation in *DNMT3A* frozen peripheral blood mononuclear cells were available and we dissected the clonal architecture by genotyping DNA from peripheral blood mononuclear cell-derived single colonies grown in methylcellulose. The *DNMT3A* mutation in this patient had been acquired before *JAK2*-V617F and the clone expanded to account for 50% of the progenitors (Figure 1f). Similar analysis in adults with MPN showed that *DNMT3A* mutations preferentially also occur early in the development of the MPN clones.[14]

Our study illustrates similarities but also differences in the mutational landscape between pediatric and adult MPN and shows that a larger proportion of pediatric patients have no detectable mutation in any of the genes known to be associated with MPN. Pediatric MPN patients overall also display fewer mutations in genes involved in epigenetic regulation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Stefan Constantinescu for helpful advice on the candidate mutations in *EPOR*. This work was funded by grants 310000-120724/1 and 32003BB_135712/1 from the Swiss National Science Foundation and the Swiss Cancer League (KLS-2950-02-2012) to RCS.

A Karow[1,4], R Nienhold[1,4], P Lundberg[1], E Peroni[2], MC Putti[3], ML Randi[2] and RC Skoda[1]
[1]*Experimental Hematology, Department of Biomedicine, University Hospital of Basel, Basel, Switzerland;*

[2]Department of Medicine–DIMED, University of Padua, Padua, Italy and
[3]Department of Pediatric Oncology and Hematology, University of Padua, Padua, Italy
E-mail: radek.skoda@unibas.ch
[4]These two authors contributed equally to this work.

REFERENCES

1 Moulard O, Mehta J, Fryzek J, Olivares R, Iqbal U, Mesa RA. Epidemiology of myelofibrosis, essential thrombocythemia, and polycythemia vera in the European Union. *Eur J Haematol* 2014; **92**: 289–297.
2 Hasle H. Incidence of essential thrombocythaemia in children. *Br J Haematol* 2000; **110**: 751.
3 Giona F, Teofili L, Moleti ML, Martini M, Palumbo G, Amendola *et al.* Thrombocythemia and polycythemia in patients younger than 20 years at diagnosis: clinical and biologic features, treatment, and long-term outcome. *Blood* 2012; **119**: 2219–2227.
4 Kucine N, Chastain KM, Mahler MB, Bussel JB. Primary thrombocytosis in children. *Haematologica* 2014; **99**: 620–628.
5 Randi ML, Putti MC, Scapin M, Pacquola E, Tucci F, Micalizzi C *et al.* Pediatric patients with essential thrombocythemia are mostly polyclonal and V617FJAK2 negative. *Blood* 2006; **108**: 3600–3602.
6 Teofili L, Giona F, Martini M, Cenci T, Guidi F, Torti L *et al.* Markers of myeloproliferative diseases in childhood polycythemia vera and essential thrombocythemia. *J Clin Oncol* 2007; **25**: 1048–1053.
7 Langabeer SE, Haslam K, McMahon C. CALR mutations are rare in childhood essential thrombocythemia. *Pediatr Blood Cancer* 2014; **61**: 1523.
8 Giona F, Teofili L, Capodimonti S, Laurino M, Martini M, Marzella D *et al.* CALR mutations in patients with essential thrombocythemia diagnosed in childhood and adolescence. *Blood* 2014; **123**: 3677–3679.
9 Langabeer SE, Haslam K, McMahon C. Distinct driver mutation profiles of childhood and adolescent essential thrombocythemia. *Pediatr Blood Cancer* 2014; **62**: 175–176.
10 Randi ML, Geranio G, Bertozzi I, Micalizzi C, Ramenghi U, Tucci F *et al.* Are all cases of paediatric essential thrombocythaemia really myeloproliferative neoplasms? Analysis of a large cohort. *Br J Haematol* 2015; **169**: 584–589.
11 Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit *et al.* The 2008 revision of the WHO classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 2009; **114**: 937–951.
12 Tefferi A, Thiele J, Vannucchi AM, Barbui T. An overview on CALR and CSF3R mutations and a proposal for revision of WHO diagnostic criteria for myeloproliferative neoplasms. *Leukemia* 2014; **28**: 1407–1413.
13 Fu R, Zhang L, Yang R. Paediatric essential thrombocythaemia: clinical and molecular features, diagnosis and treatment. *Br J Haematol* 2013; **163**: 295–302.
14 Lundberg P, Karow A, Nienhold R, Looser R, Hao-Shen H, Nissen I *et al.* Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood* 2014; **123**: 2220–2228.
15 Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD *et al.* Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med* 2013; **369**: 2379–2390.
16 Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC *et al.* Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med* 2013; **369**: 2391–2405.
17 Konieczna I, Horvath E, Wang H, Lindsey S, Saberwal G, Bei L *et al.* Constitutive activation of SHP2 in mice cooperates with ICSBP deficiency to accelerate progression to acute myeloid leukemia. *J Clin Invest* 2008; **118**: 853–867.
18 Seubert N, Royer Y, Staerk J, Kubatzky KF, Moucadel V, Krishnakumar S *et al.* Active and inactive orientations of the transmembrane and cytosolic domains of the erythropoietin receptor dimer. *Mol Cell* 2003; **12**: 1239–1250.
19 Hilton DJ, Watowich SS, Katz L, Lodish HF. Saturation mutagenesis of the WSXWS motif of the erythropoietin receptor. *J Biol Chem* 1996; **271**: 4699–4708.

Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu).