UNIVERSITÀ
DI PAVIA

Department of Molecular Medicine

Genomic profiling and genotype-phenotype correlations in myeloproliferative neoplasms

Ilaria Carola Casetti

PhD in Translational Medicine
Ciclo XXXIII – A.A. 2019-2020

Table of contents

Abstract

Myeloproliferative neoplasms (MPNs) are disorders of the stem cell, due to acquired mutations causing a clonal proliferation of one or more hemopoietic progenitor in the bone marrow. According to the World Health Organization (WHO) 2016 classification of myeloid neoplasms, Ph-negative MPNs include polycythemia vera (PV), essential thrombocythemia (ET), primary myelofibrosis (PMF), and prefibrotic myelofibrosis (prePMF). A somatic mutation in *JAK2*, *CALR* or *MPL* genes is found in the great majority of patients with MPNs. These driver mutations constitutively activate the JAK/STAT pathway, resulting in increased phosphorylation of its substrates and leading to increased cytokine responsiveness of myeloid cells. MPNs may progress to more aggressive diseases. This evolution is typically associated with the acquisition of somatic variants in genes involved in different pathways, including DNA methylation and regulation of chromatin structure, transcriptional regulators, signaling pathway and splicing factors. The phenotype of MPNs seems to be related to the combination of driver and subclonal variants and their order of acquisition but the molecular and clinical correlations are not yet completely understood.

This work is a genotype-phenotype study aimed to correlate biological and clinical features in a cohort of 509 MPN patients, diagnosed with ET, prePMF and PMF at the UOC Ematologia, Fondazione IRCCS Policlinico San Matteo, between 1985 and 2019. DNA sequence variants were studied through a NGS approach using the Illumina Nextera Rapid Capture Custom Enrichment Kit and HiSeq2500 platform. The panel targeted the coding sequence of 81 genes known to be involved in myeloid neoplasms.

Overall, 589 additional somatic variants were detected. Compared to ET and prePMF, PMF showed a larger proportion of patients carrying at least one additional somatic variant, a higher average number of variants per patient and a greater involvement of high molecular risk genes. ET, prePMF and PMF showed different mutational landscapes: in ET the most commonly involved pathway was DNA methylation genes, while in prePMF RNA splicing genes were often affected,

together with DNA methylation; in PMF chromatin structure, DNA methylation and RNA splicing were the most recurrent mutated pathways. This finding suggests that prePMF and PMF share molecular features with MDS, since RNA splicing is often involved in MDS development. No significative association between driver mutation and additional variants was found, except for mutations in the spliceosome genes, which did not occur in *CALR*- mutated patients. The correlations of the additional variants with the clinical picture highlighted a negative impact of high-risk genes on OS and on the progression of the disease.

In conclusion, these data suggest that not only clinical and histopathological criteria, but also distinct mutation patterns might differentiate ET, prePMF and PMF. In the era of precision medicine, an accurate diagnosis and prognostic stratification will be useful in the choice of a proper treatment for each patient.

Introduction

Myeloproliferative neoplasms (MPNs) are acquired diseases that occur in the hematopoietic stem cell. Neoplastic transformation causes clonal proliferation of one or more hemopoietic progenitors in the bone marrow[1]. The adjective "myeloproliferative" was used for the first time by William Dameshek in 1951 to describe, among a broad spectrum of diseases with common pathogenesis and partially similar clinical findings, four different clinical entities: polycythemia vera (PV), essential thrombocythemia (ET), primary myelofibrosis (PMF) and chronic myeloid leukemia (CML)[2].

The molecular basis of CML was discovered in the 1980s, when balanced translocation t(9;22)(q34.1;q11.2), which results in the formation of the Philadelphia (Ph) chromosome, containing the *BCR-ABL1* rearrangement, was identified. This fusion gene encodes for an abnormal and constitutively activated tyrosine-kinase with enhanced enzymatic activity that is sufficient to cause the disease: in CML, *BCR-ABL1* is found in all myeloid lineages and in some lymphoid and endothelial cells. The diagnosis requires therefore the detection of the Ph chromosome and/or *BCR-ABL1* in the appropriate clinical and laboratory settings[3].

After this discovery, myeloproliferative neoplasms have been commonly split into Ph-positive MPNs (those carrying Ph chromosome and/or *BCR-ABL1* rearrangement*)* and Ph-negative MPNs *(*PV, ET and PMF), whose molecular basis was unexplained for a very long time.

In the most recent WHO 2016, classification chronic eosinophilic leukemia not otherwise specified (NOS), chronic neutrophilic leukemia and unclassifiable myeloproliferative neoplasms were included in MPNs (Table 1)[4]

Table 1 WHO classification of myeloid neoplasms

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

The last WHO reclassification differentiates “true” essential thrombocythemia (ET) from prefibrotic/early primary myelofibrosis (prePMF), previously considered comparable to ET, but actually different because of morphologic findings in the bone marrow biopsy, including the lack of reticulin fibrosis at onset. This distinction has also prognostic implications.[5]

The genetic lesion associated with chronic neutrophilic leukemia was discovered in 2013, when a few recurrent mutations in the *CSF3R* gene (colony-stimulating factor 3) were identified.[6]

The molecular basis of Ph negative MPNs, unknown for a very long time, was clarified in the 21st century, when driver mutations in *JAK2*, *MPL* e *CALR* genes were identified.[7] (Figure 1)

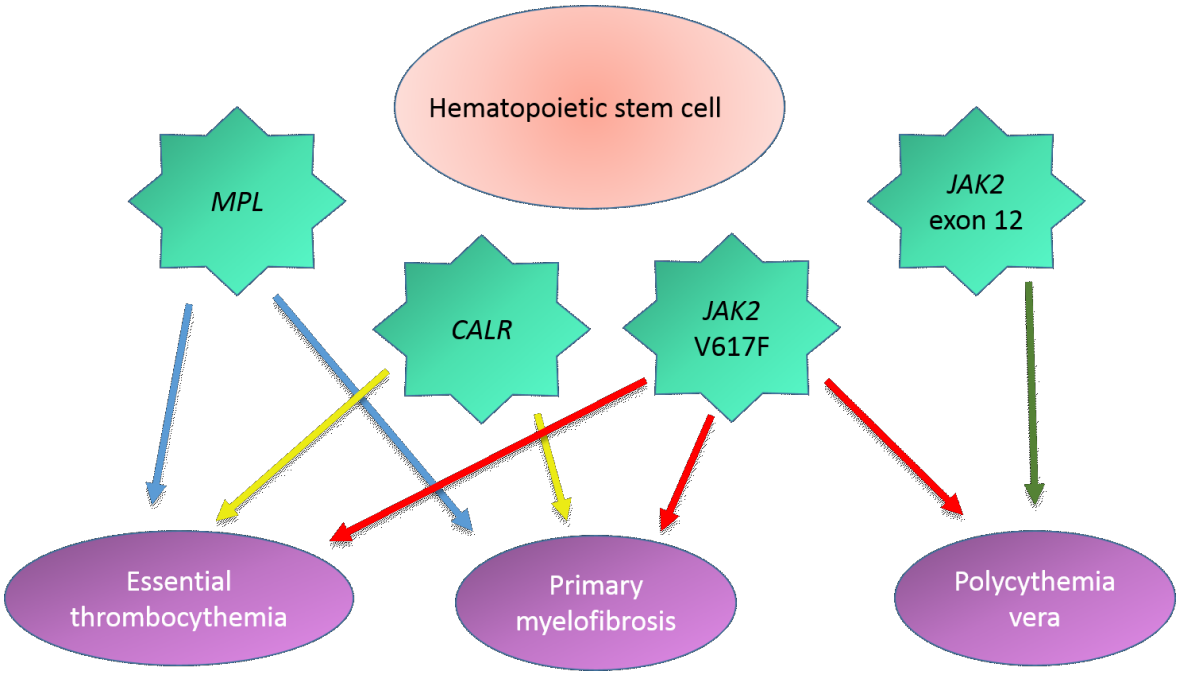

Figure 1. Pathogenetic model of MPNs

MPNs are rare diseases and their incidence increases with age. Incidence of PV is between 0.4 e 2.8 cases/100.000 people per year, median age at diagnosis is 60 years and it is more common in males; incidence of ET is estimated between 0.38 e 1.7 cases/100.000 people per year, with a peak of incidence among young women in their 30s; incidence of PMF is between 0.1 e 1 cases/100.000 people per year, increases with age and shows no differences in males and females.[8]

From a clinical point of view, MPNs show alteration of the blood count and organomegaly due to extramedullar hemopoiesis. Major complications, having a great impact on morbidity and mortality, are vascular events, both thrombotic and

haemorrhagic, and potential evolution into acute myeloid leukemia, also called blastic phase.

Wide population studies reported a prevalence of thrombotic events of 34-39% at diagnosis in patients affected with PV, 10-29% in patients with ET. Prevalence of major thrombosis during the follow up ranger from 8 and 19% in PV patients and from 8 and 31% in patients with ET.[9]

Blastic evolution at 20 years follow up is less than 10% in patients with PV and around 5% in patients with ET[5]; patients affected by PMF have an increased risk of evolution with significant differences among prognostic categories.[10]

Survival of patients with MPNs is therefore reduced if compared to general population. In ET median survival is 19.8 years, in PV it is 13.7 years, while PMF patients have a median survival of 5.9 years.[11] Prognostic stratification has a major role when making therapeutic decisions. Treatment success is still limited, so accurate prognostic scores are needed to identify patients who have high risk of progression and who therefore would be the best candidates for intensive therapies, such as stem cell transplantation, which is potentially curative but at high risk of fatal complications. On the other hand, patients with a low/moderate risk could benefit from targeted therapies, maybe not curative but much less aggressive.

Molecular basis of Myeloproliferative Neoplasms: driver mutations

The introduction of massive parallel sequencing technologies has enabled the identification of new gene mutations, as well as studies on their coexistence and interaction in a patient. Their order of onset can be estimated through the variant allele frequency (VAF), which is the proportion of a mutant allele in a tumor sample. The sooner the mutation is acquired, the higher will be the percentage of mutated cells and, as a consequence, the VAF (VAF=50% if all cells of a clone are heterozygous for that mutation).

Based on the effect that mutations have on a cell, they can be distinguished into 2 main categories: driver mutations and passenger mutations. Driver mutations are genome abnormalities that cause a selective advantage, leading to cell proliferation and formation of a mutated clone. They can be divided into 2 subgroups: (i) founding or initiating mutations, which are sufficient to give rise to the initial clone of a malignancy and will be present in almost all cells of the clone (VAF = 40-50%), and (ii) subclonal or cooperative mutations, which occur in an already established clone, generating a subclone of cells carrying both the founding and the newly acquired mutation (VAF < 50%). On the contrary, passenger mutations, having no effects on the fitness of the cell, do not have pathophysiological significance and they can occur whether before neoplastic transformation of the cell or during subclonal evolution.

In MPNs somatic mutations of *JAK2*, *MPL* and *CALR* behave as founding driver mutations, causing the myeloproliferative phenotype. All of them cooperate in the JAK-STAT signal transduction pathway that promotes cell proliferation and differentiation. A unifying model of the pathophysiology of myeloproliferative neoplasms suggests that the founding driver mutation activates the pathway in megakaryocytes, resulting in thrombocytosis initially and in bone marrow fibrosis in the long term (Figure 2).[12]

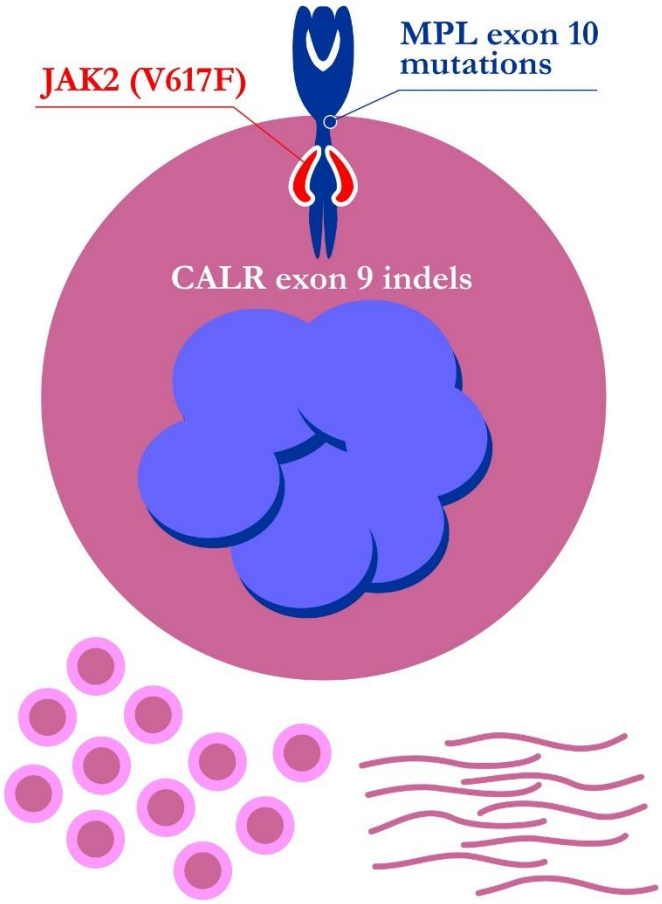

Figure 2. Driver mutation model in MPNs

The *JAK2* (V617F) mutation

Our knowledge of the genetic basis of MPNs began in 2005, when a unique base substitution in Janus Kinase 2 gene *JAK2*(V617F) was found in patients with PV, ET and PMF.[13-16]

JAK2, located of on chromosome 9p*,* encodes for a kinase protein. In physiological conditions, the binding of the ligand induces a conformational change of the kinase receptor and therefore the activation of the cellular metabolic pathways. (Figure 3). The V617F mutation causes an increase in JAK2 activity and a proliferation of myeloid progenitors in the bone marrow with different phenotypes. In PV erythroid lineage is predominant, in ET megakaryocytes are primarily involved, while in PMF both megakaryocytes and myeloid lineage proliferate.

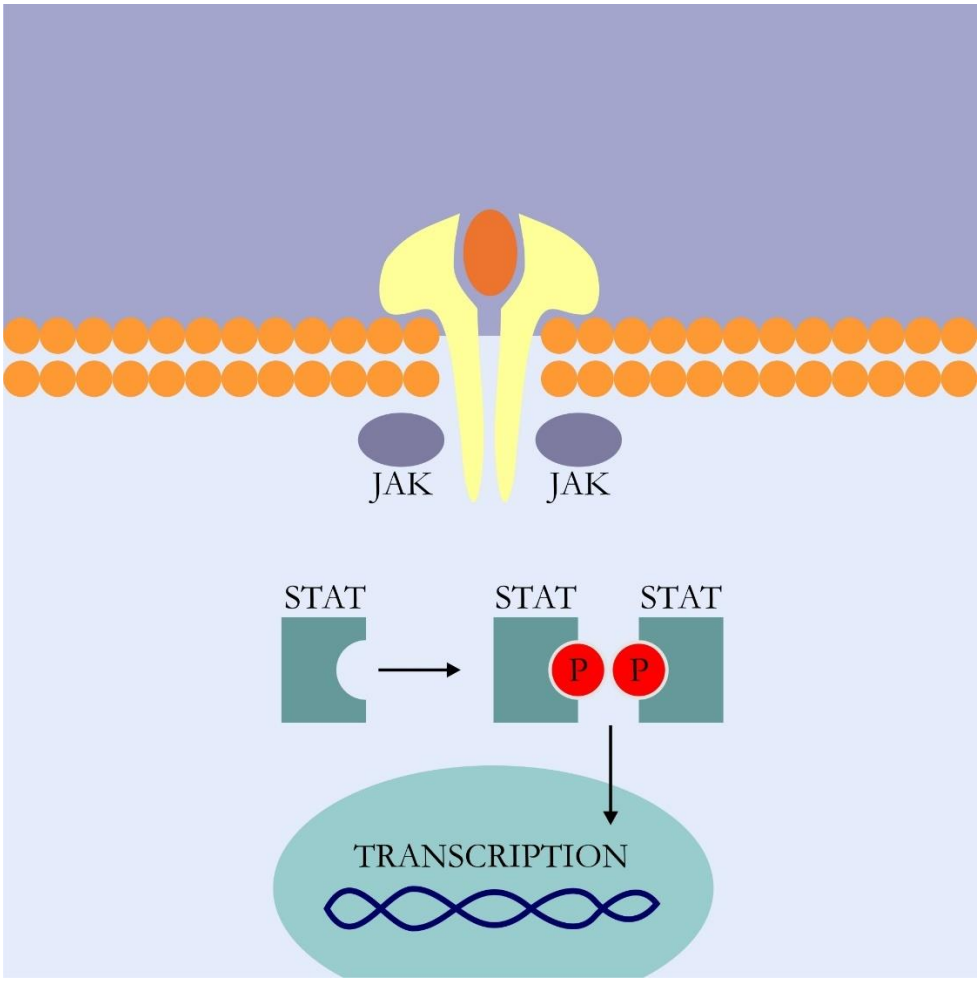

Figure 3. Janus kinase and cytokines receptors

Two different lesions in *JAK2* gene have been identified:

-transversion G-T in exon 14[14,16]

- loss of heterozigosity (LOH) in chromosome 9[15]

9pLOH is actually the most common chromosomic abnormality in patients with PV[17] (a third of the cases) and is also quite represented in ET patients. Studying people carrying this chromosomic abnormality, Kralovics's research group identified a 6.2 Mbp region that was shared by the whole cohort. This region includes *JAK2* gene. The sequencing of this region allowed the identification of a G-->T transversion, causing phenylalanine to be substituted for valine at position 617 of *JAK2* (V617F). Patients showing 9pLOH were demonstrated to be homozygous for *JAK2*(V617F), on the other hand people without 9pLOH were heterozygous and did not show this mutation.

JAK2(V617F) involves the pseudo-kinase domain JH2 in JAK2 protein; this domain plays a crucial role in the inhibition process of kinase activity of JH1: In physiological conditions valine in position 617 stabilizes the kinase domain in an inactive conformation.[18,19] (Figure 4)

Figure 4. Structure and domains of Janus kinase

JAK2(V617F) is present in 95% of the PV cases, but is shown in only 50-60% of ET or PMF patients. In the remaining part of *JAK2*(V617F)-negative PV patients point mutations or deletions in exon 12 of *JAK2* gene were identified[20]. Also, those genetic abnormalities cause an increase in the kinase activity of the mutated protein. The clinical picture is dominated by erythrocythosis but prognosis and natural history are very similar to other MPNs showing the classical *JAK2* mutation[21].

MPL mutations

Thrombopoietin (TPO) is the main megakaryopoiesis and platelet production regulator. It acts on both myeloid stem cells and on the platelets. Platelets carry thrombopoietin receptor, c-Mpl. In regular conditions TPO plasmatic concentration is inversely proportional to platelet count: this mechanism allows the maintenance of the correct count. In 2006, a year later than the discovery of the *JAK2*V617F mutation, a further step was made in the knowledge of the molecular basis of MPNs. It was shown that a small portion of patients with ET (3-5%) and PMF (8-11%) carries somatic mutations in exon 10 of the *MPL* gene, which encodes for thrombopoietin receptor[22,23]. *MPL* (myeloproliferative leukemia virus) is located on 1p34 chromosome and has 12 exons. Many mutations have already been described and the most common ones are W515L and W515K[24,25]. These abnormalities, all of whom gain of function, cause a modification of the transmembrane domain of MPL receptor; this leads to the proliferation of mature myeloid cells, due to a constitutive activation of the JAK-STAT pathway[22]. Mutations on *MPL*W515 are restricted to ET and PMF and can be also found in MDS/MPN with ring sideroblasts and thrombocytosis (MDS/MPN-RS-T, according to WHO 2016 revision)[26]. An activating mutation of TPO gene (*MPL*S505N) was discovered in familial cases of hereditary thrombocytosis and, rarely, in sporadic cases of MPNs[24,27].

CALR mutations

The molecular basis of 30-40% of ET or PMF with non-mutated *JAK2* and *MPL* was unrevealed in December 2013, when two research groups identified recurrent somatic mutations in the gene encoding for calreticulin (*CALR*)[28,29]. *CALR* is therefore the most common mutated gene after *JAK2* in MPNs and more than 60 mutations have already been described. *CALR* gene is located on 19p and is made of 9 exons. Calreticulin mutations are mutually exclusive with mutations in both *JAK2* and *MPL* and are found in exon 9.

Calreticulin is a calcium binding protein, usually located in the endoplasmic reticulum, and it is involved in the homeostasis of calcium in the cytosol and in the endoplasmic reticulum. It is a chaperonin: it plays a crucial role in the regulation of the three dimensional folding of the proteins, including molecules of the class I major histocompatibility complex (MHC). It is also involved in cellular proliferation and death[30,31]. All *CALR* mutations are insertions or deletions resulting in a frameshift, and cluster in the last exon (exon 9) of the gene. Thus far, more than 60 different types of mutations in *CALR* have been detected, but a 52-bp deletion (c.1092_1143del; L367fs*46, type 1 mutation) and a 5-bp insertion (c.1154_1155insTTGTC; K385fs*47, type 2) are the most frequent types (Figure 5), overall being found in more than 80% of all patients with mutant *CALR.* The C-terminal region of wild-type calreticulin includes a negatively charged calcium-binding domain and the endoplasmic reticulum retention motif (KDEL amino acid sequence) at the end. *CALR* mutations generate a novel C-terminus of the mutated protein, in which the negatively charged amino acids are replaced by neutral and positively charged amino-acids. In addition, the endoplasmic reticulum retention motif is lost in the mutant variants. This suggests that both impaired calcium-binding activity and cellular dislocation may play a role in the abnormal proliferation of cells expressing a mutant calreticulin. In vitro experiments in interleukin 3–dependent Ba/F3 cells showed that overexpression of type 1 *CALR* mutation (the 52-bp deletion) led to interleukin 3–independent growth and hypersensitivity to interleukin 3, and that JAK-STAT signalling was involved in these abnormal processes[28].

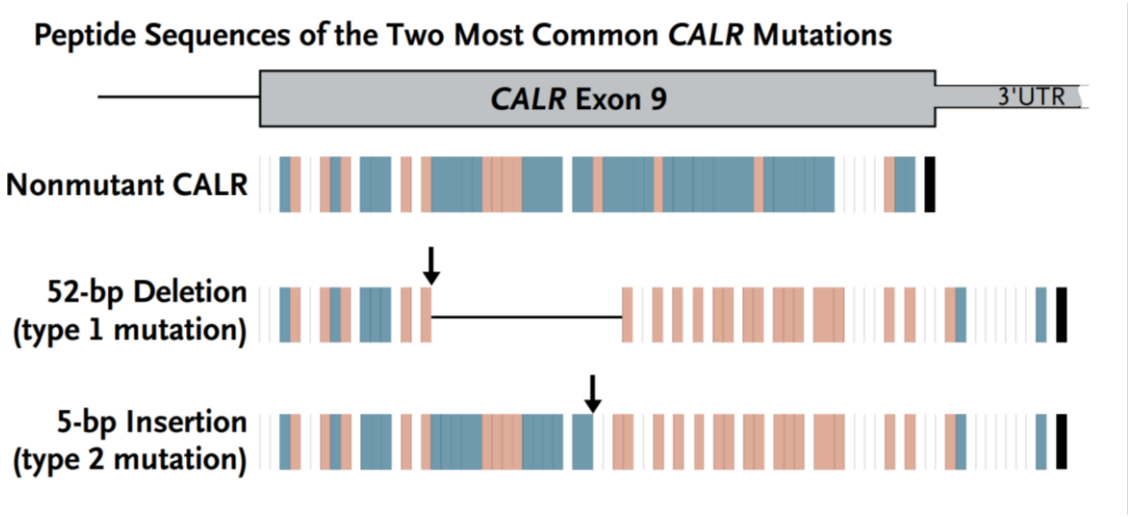

Figure 5. Peptidic sequence of wt CALR and the two most common mutated forms (Klampfl, NEJM 2013)

Even if calreticulin is usually located in the endoplasmic reticulum, it can also be found on the cell surface of neoplastic cells. In this location it acts as a regulator of adhesion process and cell migration; it also has an immunoregulatory function and promotes phagocytosis[32,33].

CALR mutations have been reported in other myeloid neoplasms at a very low frequency. In particular, mutations have been described in patients affected by MDS/MPN with ring sideroblasts and thrombocytosis (MDS/MPN-RS-T) and occasionally in cases of Atypical chronic myeloid leukemia (aCML) and Chronic myelomonocytic leukemia (CMML). There are no reported *CALR* mutated lymphoid neoplasms and the mutation has never been found in healthy subjects[28,29]. All those myeloid neoplasms have a thrombocytosic phenotype.

A study published in 2014 showed that MPNs are characterized by an upregulation of the genes involved in the JAK-STAT pathway, regardless of the clinical phenotype or the mutational status[34].

Recently the pathogenetic mechanism correlating *CALR* mutations, clonal proliferation and the development of thrombocytosis have been clarified. Thrombopoietin receptor (MPL) has a crucial role. Mutated calreticulin interacts with MPL, inducing a constitutive activation of JAK/STAT pathway; moreover, c-mpl and

TPO deficient mice do not develop thrombocytosis, even when carrying *CALR* mutations[35].

Mutant calreticulin activates specifically MPL receptor through the modified C terminus, causing its activation regardless of the presence of TPO[36]. A study conducted in Japan confirmed that mutant calreticulin, in contrast with wild-type protein, can activate MPL and the JAK/STAT pathway, leading to megakaryocyte proliferation independently of TPO[37]. Interaction between mutant calreticulin and MPL takes place in the endoplasmic reticulum, where calreticulin also acts as a chaperoning in the folding process of proteins. The complex MPL-calreticulin is then exported to the cellular surface, where MPL turns into its dimeric form, starting the signalling and promoting megakaryocyte proliferation.

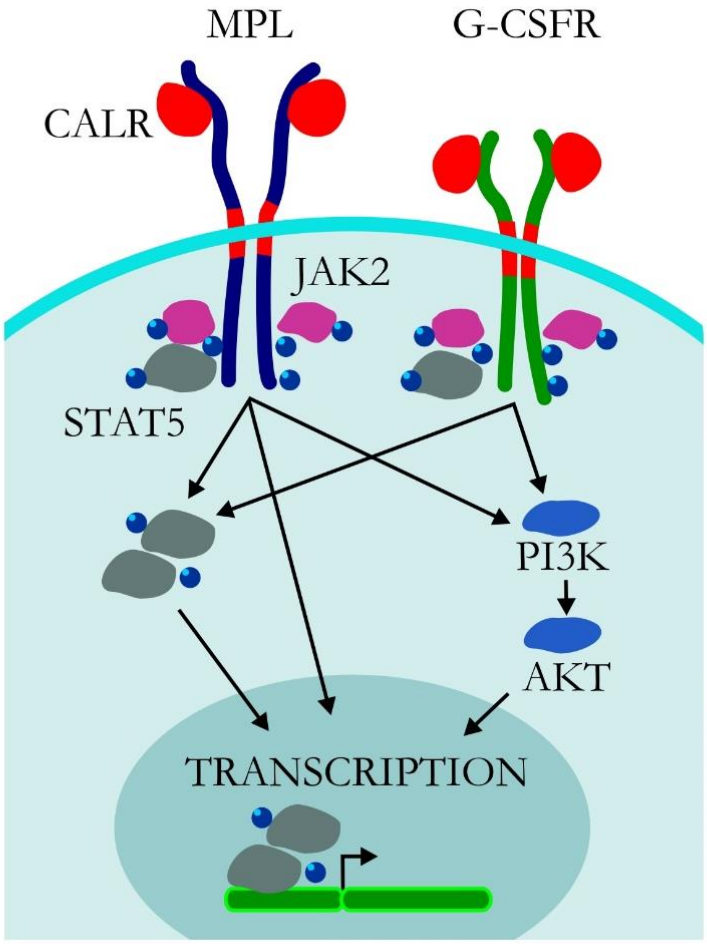

Figure 6. CALR mutants mainly activate MPL, explaining thrombocytosis

Most ET and PMF patients carrying *CALR* mutations have a 40-50% allelic burden. This shows that hemopoiesis is completely clonal and suggest that MPL activation can determine proliferative advantage in the mutant stem cell; anyway, only megakaryocytes are stimulated by MPL activation, leading to thrombocytosis, which is the typical clinical phenotype in *CALR* mutated MPNs in their initial phase[38]. Some of those patients develop bone marrow fibrosis and myeloid metaplasia later[39].

Patients with type 1 mutation show more often a myelofibrotic phenotype and a significantly higher risk of myelofibrotic transformation in ET, compared to type 2

mutation. This difference between *CALR* variants is probably due to the markedly impaired calcium binding activity of type 1 mutant calreticulin[40].

From a clinical point of view, several studies have confirmed that *CALR* mutated patients are preferentially male and show higher platelet count and lower hemoglobin and leukocyte count compared with *JAK2*- and *MPL*-mutated patients. Patients carrying the *CALR* mutation have a lower risk of thrombosis than *JAK2*- and *MPL*-mutated patients, despite significantly higher platelet count[41].

No difference in OS was observed between patients with *JAK2* (V617F)-mutated ET and those carrying a *CALR* mutation[39], while other studies on MPN patients showed a better survival in *CALR* PMF compared to *JAK2, MPL* mutated and wild type PMF (so called triple negative PMF)[42,43]. *CALR* mutations in PMF are associated with younger age, higher platelet count and cluster in lower risk prognostic category according the DIPSS score. *CALR*-mutated patients are also less likely to be anemic, require transfusions or show leukocytosis. Triple-negative patients displayed a greater trend to develop blastic phase[42].

Patients with *CALR* mutated PMF are generally younger, have a lower risk of developing anemia, thrombocytopenia, and marked leukocytosis compared with other subtypes. They also have a lower risk of thrombosis compared with patients carrying *JAK2*(V617F). At the opposite, triple-negative patients had higher incidence of leukemic transformation compared with either *CALR*-mutant or *JAK2*-mutant patients. Median overall survival was 17.7 years in *CALR*-mutant, 9.2 years in *JAK2*-mutant, 9.1 years in *MPL*-mutant, and 3.2 years in triple-negative patients. *CALR* mutated patients have a better OS compared to *JAK2* positive or triple negative subjects, regardless of the current prognostic scores[43].

Subclonal mutations

The molecular landscape of the Ph- MPNs is much more complex than was initially assumed based on the discovery of the three main mutations. Somatic mutations in *JAK2*, *MPL* e *CALR* are defined as driver founding mutations, because they give a selective advantage to a hematopoietic progenitor, leading to the formation of a mutant clone. Driver mutation are sufficient to originate a myeloproliferative phenotype, but they are not necessarily the first genetic hit in MPN pathogenesis. Several additional mutations, which are usually acquired in variably sized sub-clones of hematopoietic progenitors have been identified. However, these mutations are not restricted to MPNs and are even more frequent in other myeloid neoplasms, such as myelodisplastic syndromes and acute myeloid leukemia. Subclonal mutations are usually associated with disease progression and can help in identifying patients with poorer prognosis, in terms of reduced overall survival and/or increased risk of leukemic transformation[44].

Most common mutations target epigenetic regulators of DNA (such as *DNMT3A*, *TET2*, IDH1 and IDH2), genes involved in the maintenance of chromatin structure (*ASXL1* and *EZH2*), spliceosome genes (such as *SF3B1*, *U2AF1* and *SRSF2*) or transcription factors and tumor suppressor genes (eg *RUNX1*, *TP53*)[7]. Approximately 5-30% of patients with chronic MPNs carry subclonal mutations (Table 2). In particular, most PMF patients carry 3 or more somatic mutations, while ET and PV patients usually carry only the mutation in MPN driver gene[44].

Non-driver mutations may appear at any time-point in the clinical history of the neoplasia. Mutations in some genes, e.g.*TET2*, might occur earlier than the driver mutations; most occur at the same time or later than the driver mutation.

The order of mutations seems to influence clinical phenotype, the biology of stem and progenitor cells, and clonal evolution in patients with MPNs. This partially explains heterogeneity of the MPNs, even when they share the same driver mutation. By genotyping hematopoietic colonies or by means of NGS, Ortmann and colleagues showed that, as compared with patients in whom the *TET2* mutation was acquired first ("*TET2*-first patients"), patients in whom the *JAK2* mutation was

acquired first ("*JAK2*-first-patients") were younger and had a greater likelihood of presenting with PV than with ET, an increased risk of thrombosis, larger homozygous subclones, and an increased sensitivity of *JAK2*-mutant progenitors to Ruxolitinib in vitro[45].

Table 2 Somatic mutations in MPNs

Gene and function	Location	Mutation	Protein function	Frequency in MPNs	Consequences
Signaling driver					
JAK2	9p24	*JAK2*V617F *JAK2* exon12	Tyrosine Kinase	95% PV, 50-60% ET and PMF 3% PV	Increased RBC, PLT and WBC Increased RBC
MPL	1p34	*MPL*515L/K	TPO-R	3% ET, 3-5% PMF	Increased PLT
CALR	19p13	Indel exon 9	Chaperon	25-30% ET and PMF	Increased PLT
Other signaling					
LNK	12q24	Missense (loss of function) Deletion	JAK2 regulator	1% ET 2% PMF	Progression
CBL	11q23;3	Missense (loss of function)		4% PMF	Progression to leukemia
NRAS	1p13	Missense (activation)	ERK/MAPK signaling	Rare in PMF	Progression to leukemia

Gene and function	Location	Mutation	Protein function	Frequency in MPNs	Consequences
NF1	17q11	Missense deletion	ERK/MAPK signaling	Rare in PMF	Progression to leukemia
FLT3	13q12	FLT3-ITD	Cytokine receptor	<3%MPN	Progression to leukemia
DNA methylation					
DNMT3A	2p23	Missense	DNA methylase	5-10% MPN	Initiation
IDH1	2q33	Missense	Enzyme	1-3% PMF	Progression
IDH2	15q26		Enzyme	1-3% PMF	Progression
TET2	4q24		Dioxygenase	10-20% MPN	Initiation
Histone modification					
EZH2	7q35-36	Missense, indel	Methyltransferase	5-10% PMF	Progression
ASXL1	20q11	Nonsense, indel	Chromatin-binding protein	25% PMF, 1-3% ET and PV	Progression
Transcription factors					
TP53	17p13	Missense, indel	DNA repair, apoptosis	<5% MPN	Progression to leukemia
CUX1	7q22	Deletion	Regulation of TP3	<3% MPN	Progression to leukemia

Gene and function	Location	Mutation	Protein function	Frequency in MPNs	Consequences
IKZF1	7p12	Deletion, indel	Transcription factor	<3% MPN	Progression to leukemia
ETV6	12p13	Missense, indel	Transcription factor	<3% MPN	Progression to leukemia
RUNX1	21q22	Nonsense, missense, indel	Transcription factor	<3% MPN	Progression to leukemia
RNA splicing					
SRSF2	17q25	Missense	RNA splicing factor	<2% ET, 10-15% PMF	Progression
SF3B1	2q33	Missense	RNA splicing factor	<3% ET	Phenotypic change (anemia)
U2AF1	21.q22	Missense	RNA splicing factor	10-15% PMF	Phenotypic change (anemia)

Signalling proteins

LNK/SH2B adaptor protein 3 (SH2B3) gene encodes for an adaptor protein, which negatively regulates JAK2 activity and belongs to a protein family whose members share several structural motifs, including a proline-rich N-terminal dimerization domain (DD), a central Pleckstrin homology (PH) domain, a Src homology 2 (SH2) domains, and a conserved C-terminal tyrosine residue. LNK binds to MPL via its SH2 domain and colocalizes to the plasma membrane via its PH domain. Upon cytokine stimulation with TPO, LNK binds strongly to JAK2 and inhibits downstream STAT activation, thereby providing critical negative feedback regulation[46].

LNK mutations have been found in many different types of hemopathies, including *JAK2*-negative erythrocytosis[47]. *LNK* mutations can be both germ line and somatic mutations associated with *JAK2* V617F or *CALR* mutations, especially during disease progression[48]. Thus, rather than being primary drivers, mutations in *LNK*, either germ line or acquired, may cooperate with acquired driver mutations in *JAK2*, *CALR*, or *MPL* to determine the disease phenotype in MPNs[49].

DNA Methylation

The TET and DNMT3 protein families play a central role in DNA methylation at cytosine guanine dinucleotides (CpG), an important epigenetic regulatory mechanism controlling the expression of genes. In particular, DNMT3A (DNA methyltransferase 3A) carries out de novo methylation of cytosine, while TET2 demethylates DNA by converting 5-methylcytosine to 5-hydroxymethylcytosine.

All *TET2* mutations are loss of function point mutations or deletions and they are found in 5-17% of MPNs, usually on 1 allele. As already described, they may be either the first molecular event in *JAK2* V617F mutated MPNs, leading to a clone dominance, or a secondary event occurring after the acquisition of the driver mutation[45]. Secondary *TET2* mutations can also be associated with disease progression, more particularly when the mutation is homozygous[50].

DNMT3A mutations were first discovered in acute myeloid leukemia and are less frequent in MPNs than *TET2* (5-10% of cases). The most common mutation is R882H.

Both *TET2* and *DNMT3A* mutations are able to increase the self-renewal ability of HSCs both in human and in mice[51]. Even if the mechanism is not completely understood, TET2 and DNMT3A, by their modulation of CpG methylation, control the expression of genes involved in HSC properties and differentiation.

Interestingly, *TET2* and *DNMT3A* are the 2 most frequently mutated genes associated with clonal hematopoiesis during aging[52,53], suggesting that they confer a clonal advantage at the level of HSCs rather than drive the overproduction of erythroid and/or megakaryocyte cells that is characteristic of MPNs. On the other hand, there is no clear evidence that these mutations induce gene instability. However, by inducing a clonal hematopoiesis and thus increasing replication of mutated stem cells, they may indirectly induce secondary mutations. They may also favour leukemogenesis by progressively modifying the equilibrium between self-renewal and differentiation, but their role is not clear yet.

Histone modification

Post-translational modification of histones also influences gene expression. A key role is played by an inhibitor of transcription, known as Polycomb Repressive Coex 2 (PRC2). ASXL1 protein (Addition of Sex Combs Like 1) is a mediator of PRC2 action: it binds to the chromatin and recruits the PRC2 complex to specific loci through its direct interaction with EZH2. It can be also involved in the PRC1 complex by its association with the deubiquitinating enzyme (DUB) BRCA-1–associated protein (BAP1), a critical tumor suppressor in solid tumor.

Mutations in *ASXL1* are loss-of-function mutations due to either focal deletion or nonsense mutation or insertion/deletion leading to frameshift. They have been described in 2-10% of PV and ET patients and in about 25% of PMF patients[54]; in this last case *ASXL1* has an unfavourable prognostic significance, being associated

with more severe anemia, shorter survival and higher risk of leukemic evolution, independently on the International Prognostic Scoring System classification[55].

Some genes codifying for enzymes interacting with PRC2 complex, such as *EZH2* (Histone-lysine N metyltransferase EZH2) or *IDH1* and *IDH2* (Isocitrate dehydrogenase 1 and 2), can be affected by loss of function mutations in MPN patients in 3 to 13% of the cases.

EZH2 encodes for 1 of the 2 H3K27 methyltransferases composing the PRC2 complex and is involved in the repressive H3K27 trimethylation. Loss-of-function mutations and cytogenetic lesions in *EZH2* and other PCR2 members have been described in MPNs as well as across all myeloid malignancies. *EZH2* mutations are found in 5-10% of PMF cases and are associated with an increased risk of leukemic evolution[56].

IDH1 and IDH2 are also involved in regulating histone methylation. These enzymes physiologically catalyze the conversion of isocitrate to alpha-ketoglutarate. *IDH1* and *IDH2* mutations have been detected in a small percentage of MPN in a chronic phase (1-4%), but have a greater incidence (up to 20%) in MPNs in blastic phase[57].

RNA splicing

Mutations in spliceosome genes (e.g. *SF3B1*, *SRSF2* and *U2AF1*) have been detected in several hematopoietic malignancies, more frequently in MDS. These mutations result in altered mRNA splicing and are specifically associated with the presence of ring sideroblasts in MDS[58]. *SF3B1* gene encodes for subunit 1 of the splicing factor 3b, which, together with splicing factor 3a and a 12S RNA unit, forms the U2 small nuclear ribonucleo-proteins complex (U2 snRNP), a core component of RNA splicing machinery. The close relationship between *SF3B1* mutation and ring sideroblasts makes *SF3B1* the first gene to be associated with this specific morphological feature characteristic of MDS/MPN-RS-T, a subtype of MDS/MPN in which mutations of both *SF3B1* and *JAK2*, *MPL*, or *CALR* can be found.

SRSF2 encodes for a member of the serine/arginine-rich splicing factor family that binds to exonic splicing enhancer (ESE) sequences in pre-mRNAs. The great

majority of mutations involve proline 95 (P95) and do not cause loss of function, but determine a preferential recognition of the CCNG ESE motifs, whereas the wild-type sequence recognizes both the CCNG and GGNG ESE motifs. This alters splicing of several pre-mRNA leading to numerous functionally important misspliced events. In MPNs, *SRSF2* mutations are associated with a poor prognosis, with reduced overall survival and increased risk of leukemic transformation (19% of the cases for *SRSF2*)[59].

Tumor suppressors (*TP53*)

A strong association between alterations of chromosome 17 and leukemic evolution has been described. *TP53*, mapping 17p.13.1 and encoding for an essential protein in DNA repair process and in apoptosis, is often affected in several neoplastic diseases, both hematological and solid.

TP53 is a tumor suppressor protein, involved in several cellular stress responses by inducing cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. Various mutations and cytogenetic lesions targeting *TP53* tumor suppressor function are found in 40-50% of secondary leukemia of MPNs. These mutations usually include missense mutations or deletions on the gene itself or amplification of chromosome 1 targeting *MDM4,* a *TP53* transcriptional inhibitor. Heterozygous *TP53*-mutated clone can be found early in MPN but the disease evolution to secondary acute leukemia is associated with its clonal dominance and transition from heterozygosity to homozygosity for the mutation[60]. This mode of transformation mimics therapy-related leukemia and is quite different from sporadic leukemia. Using targeted next-generation sequencing (NGS) of 104 genes in a cohort of 197 MPN patients, Lundberg and colleagues demonstrated that loss of heterozygosity in *TP53* are strongly associated with leukemic transformation. Surprisingly, the great majority of somatic mutations were already detectable at MPN diagnosis at a low allele burden, and very few new ones were detected during follow-up, as the number of mutations between early and late patient samples did not significantly change, suggesting that the mutation rate in MPNs is rather low[50]. *TP53* mutations are apparently more frequent in post-PV or -ET AML than on post-MF leukemia. The role of *TP53* mutations has been validated later in an acute myeloid leukemia mouse model *JAK2*V617F mutated and *TP53* knock out[61].

Other secondary leukemias can also be associated with mutations or deletions in other genes encoding for signalling proteins (*CBL*, *LNK*) or transcription factors (*RUNX1, IKZF1* and *CUX1).*

To summarize, mutations in the driver genes result in a ligand-independent activation of the JAK/STAT pathway: the activated STAT protein dimerizes,

translocates to the nucleus and mediates gene transcription. Additional subclonal mutations may affect genes involved in epigenetic mechanisms or in the splicing machinery. Methylation and hydroxymethylation of cytosines in DNA and methylation of histones are the most common mechanisms of achieving transcriptional repression. Gene expression is regulated also by post-translational modification of histones and regulation of chromatin structure. Therefore, all mutations involving genes acting in one step of the pathway may lead to a myeloproliferative phenotype or to disease progression.

MPNs in WHO 2016 classification of myeloid neoplasms

The last revision of the WHO diagnostic criteria became necessary due to the recent discoveries of molecular markers associated to chronic myeloid neoplasms and to acute myeloid leukemias. Diagnostic criteria have been modified and some new entities were created. In the new MPN classification a distinction between prefibrotic myelofibrosis and fibrotic myelofibrosis (overt-PMF) was introduced. The first one used to be considered as ET and actually, from a clinical and therapeutic point of view, patients with prefibrotic PMF are very close to ET ones. Some studies, on the other hand, suggested a difference in prognosis between prefibrotic PMF (prePMF) and ET[62,63]. An Italian work confirmed that these two pathological conditions should be considered separately; prefibrotic PMF shows an OS significantly shorter than ET and a greater trend to evolve into overt fibrotic MF or to blastic phase[5].

To integrate the new discoveries and to correctly differentiate ET from prefibrotic PMF a revision of the WHO criteria was suggested[64]; the reclassification of myeloid neoplasms was published in April 2016 and is reported hereafter[4].

Polycythemia Vera	
Major Criteria	Hemoglobin >16.5 g/dL in men , Hemoglobin >16.0 g/dL in women or Hct >49% in men, >48% in women or increased red cell mass (RCM)
	BM biopsy showing hypercellularity for age with trilineage growth (panmyelosis) including prominent erythroid, granulocytic, and megakaryocytic proliferation with pleomorphic, mature megakaryocytes (differences in size)
	Presence of *JAK2*V617F or *JAK2* exon 12 mutation
Minor Criterion	Subnormal serum erythropoietin level

Diagnosis of PV requires meeting either all 3 major criteria, or the first 2 major criteria and the minor criterion.

Essential Thrombocythemia	
Major Criteria	Platelet count ≥450 × 10^9/L
	BM biopsy showing proliferation mainly of the megakaryocyte lineage with increased numbers of enlarged, mature megakaryocytes with hyperlobulated nuclei. No significant increase or left shift in neutrophil granulopoiesis or erythropoiesis and very rarely minor (grade 1) increase in reticulin fibers
	Not meeting WHO criteria for *BCR-ABL1*$^+$ CML, PV, PMF, myelodysplastic syndromes, or other myeloid neoplasms
	Presence of *JAK2*, *CALR*, or *MPL* mutation
Minor Criterion	Presence of a clonal marker or absence of evidence for reactive thrombocytosis

Diagnosis of ET requires meeting all 4 major criteria or the first 3 major criteria and the minor criterion.

Prefibrotic PMF	
Major Criteria	Megakaryocytic proliferation and atypia, without reticulin fibrosis >grade 1, accompanied by increased age-adjusted BM cellularity, granulocytic proliferation, and often decreased erythropoiesis
	Not meeting the WHO criteria for *BCR-ABL1*$^{+}$ CML, PV, ET, myelodysplastic syndromes, or other myeloid neoplasms
	Presence of *JAK2*, *CALR*, or *MPL* mutation or in the absence of these mutations, presence of another clonal marker, or absence of minor reactive BM reticulin fibrosis
Minor Criteria	Anemia not attributed to a comorbid condition
	Leukocytosis ≥11 × 10^9/L
	Palpable splenomegaly
	LDH increased to above upper normal limit of institutional reference range

Diagnosis of prePMF requires meeting all 3 major criteria, and at least 1 minor criterion, confirmed in 2 consecutive determinations.

Overt PMF	
Major Criteria	Presence of megakaryocytic proliferation and atypia, accompanied by either reticulin and/or collagen fibrosis grades 2 or 3
	Not meeting WHO criteria for ET, PV, *BCR-ABL1*$^{+}$ CML, myelodysplastic syndromes, or other myeloid neoplasms
	Presence of *JAK2*, *CALR*, or *MPL* mutation or in the absence of these mutations, presence of another clonal marker or absence of reactive myelofibrosis
Minor Criteria	Anemia not attributed to a comorbid condition
	Leukocytosis ≥11 × 10^9/L
	Palpable splenomegaly
	LDH increased to above upper normal limit of institutional reference range
	Leukoerythroblastosis

Diagnosis of overt PMF requires meeting all 3 major criteria, and at least 1 minor criterion.

Aim of the work

The year 2005, with the discovery of the *JAK2* V617F point mutation, was the turning point in the knowledge on the molecular mechanisms of myeloproliferative neoplasms (MPNs). This discovery, together with the subsequent identification of mutations in exon 12 of *JAK2* and in codon 515 of *MPL*, led the World Health Organization (WHO) to revise the diagnostic criteria for myeloid malignancies in 2008, and focused the attention of researchers on altered signaling of the JAK/STAT pathway, which was found to play a key role in the pathogenesis of MPNs.

The more recently discovered mutations of the *calreticulin* gene (*CALR*), present in most *JAK2*- and *MPL*-negative patients, are now included in the 2016 revision of the WHO diagnostic criteria, and have generated novel and more complex hypotheses on pathogenic mechanisms. A wide variety of animal models have clearly demonstrated that mutations in *JAK2*, *MPL* and *CALR* are “driver” or “phenotypic” mutations, as their expression in animals recapitulates the disease phenotype, which can be treated by *JAK2* inhibitors. At the same time, however, the landscape of mutations in MPNs has become more complex as a result of the emergence of additional mutations, which are usually expressed in a subclone of the hematopoietic compartment and affect genes involved in DNA regulation and transcription. These mutations, which may be also found in myelodysplastic syndromes and in some cases of acute myeloid leukemia, suggest a hierarchy of mutational events that should be considered for a more complete definition of the disease mechanisms and, more importantly, for a personalized therapeutic approach.

This work is a genotype-phenotype study aimed to correlate driver mutations, subclonal variants, clinical data and diagnosis in a cohort of 509 MPN patients, in order to define a possible diagnostic and prognostic role of these variants.

To investigate the subclonal variants we used a next generation sequencing approach (NGS), targeting the whole coding sequence of 81 genes commonly involved in myeloid neoplasms. In order to obtain a reliable list of somatic subclonal variants, as well as potentially germline ones, genomic DNA of both tumor and

normal tissues were analyzed in each patient. Germline mutations were excluded from the analysis and we focused on somatic additional mutations.

Since PV has a very specific molecular characterization, we focused on patients having an ET, prefibrotic PMF or overt PMF diagnosis according to WHO 2016 classification.

Somatic subclonal variants were correlated with WHO 2016 diagnosis and driver mutational status, considering their average number, as well as the affected genes and pathways. In addition, based on the relative variant allele frequency of the different found variants, the mutational timing of driver and subclonal lesions was considered, in order to hypothesize models of clonal evolution. We then analyzed correlations between clinical data and additional somatic mutations.

All the clinical data were collected at the Hematology Department, Fondazione IRCCS Policlinico San Matteo. All the experimental procedures were performed in the laboratories of the U.O.C. Ematologia, Dipartimento di Scienze Mediche e Malattie Infettive, Fondazione IRCCS Policlinico San Matteo of Pavia.

Patients and Methods

Study population

This study was approved by the local institutional Ethics Committee; the procedures followed were in accordance with the Declaration of Helsinki of 1975, as revised in 2000, and samples were obtained after patients provided written informed consent.

Inclusion criteria were (i) a diagnosis of ET, pre PMF or overt PMF according to 2016 WHO criteria[4], (ii) a peripheral blood sampling (PB) at diagnosis or, as an alternative, before therapy administration, for which the molecular characterization of the driver mutation had been already performed, and the residual DNA sample was suitable for NGS molecular studies.

Overall, 509 patients diagnosed at the UOC Ematologia, Fondazione IRCCS Policlinico San Matteo and University of Pavia, between 1985 and 2019, were identified in our databases according to the inclusion criteria. Our study cohort included 182 patients with diagnosis of overt PMF, 91 patients with diagnosis of prefibrotic PMF and 236 subjects with an ET diagnosis, whose clinical and molecular data, including general information at the diagnosis (sex, age and blood parameters) and mutational profile of driver genes, are summarized in Table 3.

General data included: date of birth (age), sex

Clinical parameters at the time of evaluation included: date of diagnosis, diagnosis according WHO 2016 criteria, driver mutation, complete blood count, LDH, spleen size, circulating CD34+ cells

Table 3. Demographic, clinical and mutational status in the study population

	Total (n=509)	**ET (n=236)**	**PMF (n=182)**	**prePMF (n=91)**
Age at evaluation (years), median (IQR)	59 (45-68)	54 (42-66)	62 (54-70)	58 (42-68)
Sex, n (%)				
M	239 (47%)	90 (38%)	105 (58%)	44 (48%)
F	270 (53%)	146 (62%)	77 (42%)	47 (52%)
PLT ($x10^9$/l), median (IQR)	593 (355-792)	664 (569-838)	244 (139-486)	768 (592-990)
WBC ($x10^9$/l), median (IQR)	8.7 (7.0-11.5)	8.4 (7.2-10.5)	9.0 (5.9-13.9)	9.8 (7.7-11.9)
HB (g/dl), median (IQR)	13.2 (11.2-14.6)	14.2 (13.2-15.1)	10.5 (9.3-12.2)	13.2 (12.2-14.5)
LDH (mU/ml), median (IQR)	257 (195-383)	201 (173-252)	415 (309-660)	272 (229-350)
Circulating CD34+ cells (per mcl), median (IQR)	6.8 (3.6-24.1)	3.9 (2.7-6.3)	38.4 (11.5-137.5)	6.4 (4.7-14.8)
Spleen size (cm from the lower costal margin, median (IQR)	0 (0-3)	0 (0-0)	5 (1-10)	0 (0-0)
Driver mutation				
CALR	127 (25%)	45 (19.1%)	49 (26.9%)	33 (36.3%)
JAK2	309 (60.7%)	149 (63.1%)	111 (61.0%)	49 (53.8%)
MPL	28 (5.5%)	12 (5.1%)	12 (6.6%)	4 (4.4%)
Triple negative (TN)	45 (8.8%)	30 (12.7%)	10 (5.5%)	5 (5.5%)

Samples collection and cell separation

In this study we used samples collected from peripheral blood. In 472 out of 509 (92,5%) cases, for each patient both circulating granulocytes (polymorphonuclear cells, PMN), as tumor sample, and T-lymphocytes (T-Ly), as individual-matching control sample, were isolated. In 13 patients (2,5%), mononuclear cells (MNC) were obtained instead of T-Ly, while in the remaining 24 cases (5%), no internal normal controls were available.

PMNs were separated from MNC by standard density gradient centrifugation using Lympholyte-H Separation Media (CEDARLANE Laboratories Ltd, Burlington, Canada): after the dilution of PB with an equal volume of Phosphate Buffer Saline Dulbecco's (PBS) modified medium (EuroClone Life Science Division, Pero, Italy), the sample was gently layered over Lympholyte-H. After centrifugation (600g for 30 minutes, at room temperature, without brake), PMN sedimented in the pellet with red blood cells (RBC), while MNC formed a well-defined layer at the interface between medium and Lympholyte-H. PMNs were finally isolated through osmotic lysis of RBC using 9 volumes of a hypotonic solution (NH_4Cl 1.55 M, $NaHCO_3$ 0.12 M and EDTA 0.25 M at pH 7.4).

The layer of MNC was collected, washed and used for T-Ly isolation. CD3+ T-Ly were obtained through positive immunomagnetic selection using the CD3 MicroBeads antybody and MACS technology (Miltenyi Biotec, Bergisch Gladbach, Germany), according to manufacturer's procedure.

DNA extraction and driver mutation analysis

DNA was extracted using the Gentra Puregene Blood Kit (Qiagen, Germantown, MD, USA), according to manufacturer's procedure. DNA was quantified using NanoDrop 2000c Spectophotometer (Thermo Fisher Scientific, Eugene, OR) and verified using the Qubit dsDNA BR Assay kit on the Qubit2.0 Fluorometer (all Invitrogen, by Thermo Fisher Scientific, Eugene, OR).

Granulocyte *JAK2* (V617F) mutation burden was assessed using a quantitative PCR-based allelic discrimination assay. PCR was performed on a RotorGene

6000TM real-time analyzer on a 100-well Gene Disk (Corbett Life Sciences, Mortlake, NSW, Australia) in two separate tubes for normal and mutated alleles. In all, 40 nanograms of genomic DNA were amplified in a 40-cycle PCR at an annealing temperature of 55 °C. All reactions were carried out in a final volume of 12 µl containing 1 × Brilliant SYBR Green QPCR master mix (Stratagene, Cedar Creek, TX, USA) and 300 nM of both forward and reverse primers. Serial dilutions, starting at 80 ng/µl and ending at the 0.4 ng/µl, of both a wild type and a fully mutated DNA were used to construct standard curves from which the wild type and V617F quantities were calculated for each sample. Results were expressed as percentage of V617F alleles among total *JAK2* alleles (WT+V617F).[65]

MPL mutation scanning was performed on granulocyte and T lymphocyte gDNA using both high resolution melting (HRM) analysis and direct sequencing. [66]

Granulocytes' DNA from peripheral blood was used for molecular analysis. *CALR* exon 9 mutations were assessed with Sanger sequencing. The evaluation of *CALR*-mutant burden was done comparing the height of the mutated and wild type peaks.[67]

Targeted gene sequencing: library preparation, run and data analysis

A core panel of 81 genes was selected on the basis of prior implication in the pathogenesis of myeloid disease by recurrent somatic mutation, recurrent mutation or aberrations in common cancers, or candidate gene mapping within regions of common copy number alterations (Table 4).

On the basis of the total target size, an enrichment approach was defined to capture the coding exons and splice sites of the selected genes. The free web-based tool Illumina DesignStudio Sequencing Assay Designer was used for the initial design against the human GRCh37/h19 reference genome of our probe set, which was then verified and optimized by the Illumina Concierge Services (Illumina, San Diego, CA, USA). The final project consisted of 1422 targets and 2838 probes, with a cumulative target length of 235.9 kb.

Table 4 List of the 81 genes sequenced by NGS, with corresponding genomic coordinates according to human GRCh37/h19 reference genome and pathway. Orange: MPN phenotype driver genes; blue: PMF high risk molecular genes (HMR genes)

Target Gene	Chr	Start	Stop	Pathway
ABCG2	4	89,011,416	89,152,474	ABC transporters
ABL1	9	133,589,268	133,763,062	signaling
AEBP2	12	19,592,608	19,675,173	chromatin structure
ASXL1	20	30,946,147	31,027,122	chromatin structure
ATM	11	108,093,559	108,239,826	genetic cancer susceptibility
ATRX	X	76,760,356	77,041,719	chromatin structure
BCOR	X	39,910,499	40,036,582	chromatin structure
BCORL1	X	129,139,164	129,192,058	chromatin structure
BRAF	7	140,433,813	140,624,564	signaling
CALR	19	13,049,414	13,055,304	JAK/STAT
CBL	11	119,076,986	119,178,859	signaling
CBLB	3	105,377,109	105,587,887	signaling
CDKN2A	9	21,967,751	21,994,490	tumor suppressor gene
CEBPA	19	33,790,840	33,793,430	transcriptional regulation
CHEK2	22	29,083,731	29,137,822	tumor suppressor gene
CREBBP	16	3,775,056	3,930,121	transcriptional regulation
CSF1R	5	149,432,854	149,492,935	signaling
CSF3R	1	36,931,644	36,948,915	signaling
CUX1	7	101,459,184	101,927,250	transcriptional regulation
DDB1	11	61,066,919	61,100,684	chromatin structure
DNMT3A	2	25,455,830	25,565,459	DNA methylation
EED	11	85,955,806	85,989,785	chromatin structure
EGFR	7	55,086,725	55,275,031	signaling
EP300	22	41,488,614	41,576,081	chromatin structure

ETV6	12	11,802,788	12,048,325	transcriptional regulation
EZH2	7	148,504,464	148,581,441	chromatin structure
FBXW7	4	153,242,410	153,456,393	signaling
FLT3	13	28,577,411	28,674,729	signaling
GATA1	X	48,644,982	48,652,717	transcriptional regulation
GATA2	3	128,198,265	128,212,030	transcriptional regulation
GNAS	20	57,414,795	57,486,250	genetic cancer susceptibility
HRAS	11	532,242	535,550	signaling
IDH1	2	209,100,953	209,119,806	DNA methylation
IDH2	15	90,627,212	90,645,708	DNA methylation
IKZF1	7	50,344,378	50,472,798	transcriptional regulation
IKZF2	2	213,864,411	214,016,333	transcriptional regulation
JAK2	9	4,985,245	5,128,183	JAK/STAT
JAK3	19	17,935,593	17,958,841	signaling
JARID2	6	15,246,206	15,522,273	chromatin structure
KDM6A	X	44,732,423	44,971,845	chromatin structure
KIT	4	55,524,095	55,606,881	signaling
KMT2A	11	118,307,205	118,397,539	chromatin structure
KRAS	12	25,358,180	25,403,854	signaling
LMO2	11	33,880,123	33,913,836	chromatin structure
MDM2	12	69,201,971	69,239,320	tumor suppressor gene
MPL	1	43,803,475	43,820,135	JAK/STAT
MYD88	3	38,179,969	38,184,512	signaling
NF1	17	29,421,945	29,704,695	genetic cancer susceptibility
NFE2	12	54,685,891	54,694,821	transcriptional regulation

NOTCH1	9	139,388,896	139,440,238	signaling
NPM1	5	170,814,708	170,837,888	transcriptional regulation
NRAS	1	115,247,085	115,259,515	signaling
PDGFA	7	536,897	559,481	signaling
PDS5B	13	33,160,564	33,352,158	cohesin complex
PHF6	X	133,507,342	133,562,822	chromatin structure
PIK3CA	3	178,866,311	178,952,497	signaling
PIM1	6	37,137,922	37,143,204	signaling
PRPF8	17	1,553,923	1,588,176	splicing
PTEN	10	89,623,195	89,728,532	tumor suppressor gene
PTPN11	12	112,856,536	112,947,717	signaling
RAD21	8	117,858,173	117,887,105	cohesin complex
RIT1	1	155,867,599	155,881,193	signaling
RUNX1	21	36,160,098	36,421,595	transcriptional regulation
RUNX2	6	45,296,054	45,518,819	transcriptional regulation
SETBP1	18	42,260,138	42,648,475	genetic cancer susceptibility
SF3B1	2	198,256,698	198,299,771	splicing
SH2B3	12	111,843,752	111,889,427	signaling
SMC1A	X	53,401,070	53,449,618	cohesin complex
SMC3	10	112,327,449	112,364,392	cohesin complex
SOCS3	17	76,352,859	76,356,158	signaling
SRSF2	17	74,730,197	74,733,493	splicing
STAG1	3	136,055,999	136,471,245	cohesin complex
STAG2	X	123,094,475	123,236,505	cohesin complex
STAT3	17	40,465,343	40,540,513	signaling
SUZ12	17	30,264,044	30,328,057	chromatin structure
TET2	4	106,067,032	106,200,960	DNA methylation
TP53	17	7,571,720	7,590,868	tumor suppressor gene

U2AF1	21	44,513,066	44,527,688	splicing
U2AF2	19	56,165,416	56,186,082	splicing
WT1	11	32,409,322	32,457,081	tumor suppressor gene
ZRSR2	X	15,808,574	15,841,382	splicing

The Nextera Rapid Capture Custom Enrichment kit (Illumina, San Diego, CA, USA) was used for library preparation and custom target enrichment, according to the manufacturer's procedure.

Library preparation

50 ng of genomic DNA were enzymatically fragmented to simultaneously get DNA fragment of about 300 bp in length and tag them with Read 1 and Read 2 Sequencing Primers. The magnetically purified tagmented DNA was then amplified using a 10-cycle PCR program to add Index 1 (i7), Index 2 (i5), and sequencing adapters required for cluster amplification. The libraries were quantified using the Qubit fluorimetric method and, after magnetical purification, 500 ng of each library coded with the same i5 index was by-row pooled in a 12-plex library pool (n=8) of a total DNA mass of 6000 ng (Figure 15B). Biotinylated probes hybridized to the target regions were magnetically captured after binding with streptavidin beads for 12-plex libraries enrichment and removal of nonspecific sequences (Figure 15C-E). A second round of these 2 steps was performed to ensure a highly specific hybridization of the captured target regions.

The enriched 12-plex libraries were purified and amplified using a 12-cycle PCR program before the final cleaning procedure. The DNA libraries were quantified using the Qubit fluorimetric method and the obtained concentrations converted from ng/µl to nM applying the following formula, in which a 400-bp library size was assumed:

$$Concentration\ in\ nM = \frac{Concentration\ in\ ng/\mu l}{400x660g/mol} x10\text{^}6$$

Finally, the 8 12-plex pools were normalized at the same lowest concentration and combined into a single 96-sample pool.

Sequencing reactions

The HiSeq Rapid Cluster PE kit v.2 (Illumina, San Diego, CA, USA) was used to denaturate and dilute libraries for on-board clustering on the HiSeq 2500 using a rapid flow cell. Ten µl of 2 nM of pooled library were incubate for 5 minutes with 10 µl of NaOH 0.1N to denaturate the library into single strands, then 980 of prechilled HT1 (provided with the kit) were added to prepare a 20 pM library.

The HiSeq Rapid SBS kit version 2 (Illumina, San Diego, CA, USA) was used for a dual-indexed paired-end 2x150-cycle rapid run sequencing of 8 pM libraries on the HiSeq 2500 platform.

An average sequencing coverage of 550x was obtained.

Data analysis

Data analysis was performed by bioinformatic algorithms and it was composed of three steps: primary, secondary and tertiary analysis. In the primary step, raw data were converted to sequences data: the images generated by fluorescent signals were translated in nucleotide bases and, at the end, in nucleotide sequences (reads). Primary analysis also provided a quality control values for each base. This analysis was directly made by Illumina on-instrument software that generated the FASTQ file. The secondary analysis consisted in the alignment of reads to a reference genome (hg19) using BWA (Burrows-Wheeler Alignment), a mapping tool for short reads of DNA, and BAM file was created from the alignment of the FASTQ file to the reference genome. The third step was the variant calling stage, consisting in the determination of genomic locations that differs from the reference genome sequence. Somatic variants were detected using GATK best practice for somatic short variant discovery (SNVs and Indels).[68] This workflow, that usually relies on Mutect2 for variant calling, integrated Scalpel[69], a second variant caller specific for Indels calling. Data were finally delivered as Variant Call Format (VCF). Variant

allele frequency (VAF, was calculated for each variant identified as number of variant reads divided by total reads.

The detected variants were then filtered through a previously defined pipeline that considered different parameters: quality of the reads, transcripts on which the variants mapped, nucleotide positions and the effect of the variants on aminoacid changes.[70] Specifically, variants with a coverage lower than 30X and less than 10 supporting reads were filtered. Variants that had been reported in more than 80% of patients of our internal database with a VAF in PMN included between 45% and 55% or higher than 90% were annotated as germline variants. Concerning the transcripts, the filter retained only one of them: it could be either the transcript reported on NCBI website as “primary” or “variant 1” or the first transcript on which the variant mapped (e.g. if the variant mapped on transcript variant 2 and 3, the latter was filtered). Non-exonic variants and synonymous variants were also filtered while functionally annotated variants were then filtered on the basis of information coming from public databases of polymorphisms (e.g. dbSNP, 1000 Genomes) with a frequency >=1% in population.

Downstream to variant detection, somatic variants were annotated using information retrieved from peer-reviewed literature, public databases (e.g. COSMIC, dbSNP, gnomAD, 1000 Genomes, ClinVar, IARC TP53 database, and UniProt) and in silico variant effect predictors (SIFT, PolyPhen-2, PROVEAN, Mutation Taster, DANN, PaPI) as well as the inclusion of mutated aminoacids in a conserved/functional protein domain.

All the bioinformatics analyses were performed by enGenome srl (Pavia, Italy).

In cancer genetics, the comparison between sequencing data obtained in the neoplastic tissue and those obtained in an individual-matched control tissue is used to discriminate somatically acquired variants from polymorphic and/or pathogenic inherited ones. Thus, in our study, we used circulating granulocytes and, whenever available, T-lymphocytes (T-Ly, 472 cases) or mononuclear cells (MNC, 13 cases) as neoplastic and internal control samples, respectively. Nevertheless, T-Ly, and especially MNC, may be contaminated by PMN. As MPN phenotype driver gene mutations are, by definition, somatically acquired, in each patient we manually

compared the variant allele frequency (VAF) of driver mutations in the 2 different cell types to estimate the degree of contamination, which was then considered for the definitive annotation of the called variants.

The possible oncogenic variants and the calls filtered by the informatic pipeline were manually revised based on the comparison of the VAF in PMN and control tissue and of information retrieved from the main MPN variant dataset present in the literature.[71,72] For filtered variants, known pathogenicity and mutation type (nonsense and frame-shift mutations) were also considered.

SRSF2 mutation analysis

Because of the extremely high GC content in the region where the P95 codon maps, the coverage in this *SRSF2* mutational hotspot was too low for a reliable variant calling using our NGS approach. Thus, *SRSF2* exon 1 was analyzed by directly Sanger sequencing in all patients.

PMN genomic DNA (30 ng) was amplified using the following primers: forward 5'-CAACCTGACCTACCGCACCTC-3' and reverse 5'-GCCCGCACCACGTGCTT-CG-3', that generated an amplicon of 402 bp (Garza*, et al* 2015). PCR reactions were performed in a final volume of 50 mcl containing 0.25 U of Hot StarTaq DNA polymerase together with 1x Buffer and 1x Q-solution (Qiagen, Hilden, Germany), 0.8 mM dNTPs, and 300 nM of each primer. The thermal protocol was as follows: 95°C for 15 minutes (1 cycle); 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds (40 cycles), 72°C for 10 minutes (1 cycle). Amplification reactions were carried out on an AB9700 thermal-cycler (Applied Biosystems, Foster City, CA).

Then, 8 µl of the PCR product were visualized on 2% agarose gel electrophoresis and 5 µl were purified by incubation with 2 µl of Illustra ExoProStar 1-Step Enzymatic PCR and Sequence Reaction Clean-up kit (GE Healthcare Life Sciences, Little Chalfont, UK) at 37°C for 15 minutes and 80°C for 15 minutes.

Two µl of the purified product were sequenced in a final reaction mix of 20 µl, using a unique internal forward primer (5'-ATCCCGCGGGACCGCTACAC-3') and the ABI

PRISM BigDye Terminator, version 3.1, Ready Reaction Cycle sequencing kit (Life technologies, a Thermo Fisher brand, Darmstadt, Germany), according to manufacturer's protocol.

The sequencing products were purified using the BigDye Xterminator Purification Kit (Life technologies, a Thermo Fisher brand, Darmstadt, Germany) by adding to each sample 90 µl of SAM solution and 20 µl of X-terminator. The samples were mixed vigorously for 30 minutes, centrifuged for 2 minutes at 1000g at RT and run on an ABI PRISM 3500 (Life technologies, a Thermo Fisher brand, Darmstadt, Germany). The results were finally evaluated using Chromas Lite and CLC sequencing viewer software (freely available online) and the VAF were estimated by comparing the height of the mutated and wild type peaks of chromatograms.[67]

Statistical analysis

Quantitative variables have been summarized as median and interquartile range (IQR). Qualitative variables were described as counts and percentages of each category.

Exact binomial 95% Confidence Intervals of VAF were calculated to estimate in each patient the timing of additional mutations respect to the driver one.

Associations between two qualitative variables were tested via Fisher's exact test. Mann-Whitney test and Kruskall-Wallis test were used to compare quantitative variables among two or more groups of patients, respectively (Post-hoc comparisons were adjusted according to Bonferroni's correction).

Logistic regression models were carried out to evaluate the association between clinical features and mutational status (for the more frequently mutated genes), adjusting for treatment status and time elapsed between diagnosis and NGS evaluation.

The association between mutational status and diagnosis, adjusting for time elapsed between diagnosis and NGS evaluation, was estimated via multinomial logistic regression models.

Proportional hazard Cox models and Fine&Gray models (competing-risk approach) were applied to estimate the impact of mutational status on Overall Survival (OS) and on Cumulative Incidence of Leukemia transformation, respectively. Both for Overall Survival and for Cumulative incidence of Leukemia, data were left-truncated, to take into account the time elapsed between diagnosis and NGS evaluation.

All statistical analyses were performed with Stata 16 software (release 16, StataCorp, College Station, TX, USA).

Results

Table 5 Clinical characteristics of the cohort

	Total (n=509)	ET (n=236)	PMF (n=182)	prePMF (n=91)	p-value	ET vs PMF	ET vs prePMF	PMF vs prePMF
Age at blood collection Median (IQR), years	59 (45-68)	54 (42-66)	62 (54-70)	58 (42-68)	<0.001	<0.001	0.281	0.014
Age>=60, n (%)	235 (46%)	89 (38%)	107 (59%)	39 (43%)	<0.001	<0.001	>0.90	0.044
NGS within 6 months from the diagnosis, n (%)	363 (71.3%)	196 (83.1%)	83 (45.6%)	84 (92.3%)	<0.001	<0.001	0.104	<0.001
Sex, n (%)					<0.001	<0.001	0.310	0.472
M	239 (47%)	90 (38%)	105 (58%)	44 (48%)				
F	270 (53%)	146 (62%)	77 (42%)	47 (52%)				
PLT, median (IQR)	593 (355-792)	664 (569-838)	244 (139-486)	768 (592-990)	<0.001	<0.001	0.263	<0.001
PLT pathologic, n (%)	403/508 (79.3%)	221/236 (93.6%)	98/182 (53.9%)	84/90 (93.3%)	<0.001	<0.001	>0.90	<0.001
WBC, median (IQR)	8.7 (7.0-11.5)	8.4 (7.2-10.5)	9.0 (5.9-13.9)	9.8 (7.7-11.9)	0.038	0.370	0.017	0.181
WBC pathologic, n (%)	211/508 (41.5%)	69/236 (29.2%)	98/182 (53.9%)	44/90 (48.9%)	<0.001	<0.001	0.003	>0.90
Hb, median (IQR)	13.2 (11.2-14.6)	14.2 (13.2-15.1)	10.5 (9.3-12.2)	13.2 (12.2-14.5)	<0.001	<0.001	<0.001	<0.001
Hb low, n (%)	187/508 (36.8%)	16/236 (6.8%)	144/182 (79.1%)	27/90 (30.0%)	<0.001	<0.001	<0.001	<0.001
Spleen, median cm (IQR)	0 (0-3)	0 (0-0)	5 (1-10)	0 (0-0)	<0.001	<0.001	0.047	<0.001
Spleen >0, n (%)	176/499 (35%)	15/235 (6.4%)	142/178 (79.8%)	19/86 (22.1%)	<0.001	<0.001	<0.001	<0.001
LDH, median (IQR)	257 (195-383)	201 (173-252)	415 (309-660)	272 (229-350)	<0.001	<0.001	<0.001	<0.001
LDH >220, n (%)	275/420 (65.5%)	76/196 (38.8%)	131/141 (92.9%)	68/83 (81.9%)	<0.001	<0.001	<0.001	0.046
CD34+, median (IQR)	6.8 (3.6-24.1)	3.9 (2.7-6.3)	38.4 (11.5-137.5)	6.4 (4.7-14.8)	<0.001	<0.001	<0.001	<0.001
CD34+>12, n (%)	139/411 (33.8%)	5/178 (2.8%)	112/156 (71.8%)	22/77 (28.6%)	<0.001	<0.001	<0.001	<0.001

Characteristics of the patient cohort

Demographic and clinical details of the patient cohort is summarized in Table 5.

Median age at blood collection was 59 years. Patients with ET or prePMF diagnosis were significantly younger compared to PMF patients (p value <0.001 ET vs PMF, 0.014 PMF vs prePMF), while we did not observe a statistically significant difference between ET and prePMF.

The median time from diagnosis to blood collection was 7,7 months in patients having a PMF and 0 months in ET and prePMF patients. 71,3% of the patients had NGS performed within 6 months from the diagnosis (46% of PMF, 92,3% of prePMF, 83,1% of ET). All the statistical analysis performed in the next chapters were adjusted accordingly.

Platelet count was significantly higher in ET and prePMF compared to PMF (p value <0.001 ET vs PMF and <0.001 prePMF vs PMF), while there were no statistically significant differences regarding WBC count in any of the groups. Hemoglobin level was significantly lower in PMF and prePMF compared to ET (p value <0.001 ET vs PMF and <0.001 ET vs prePMF) and PMF patients were significantly more often anemic also compared to prePMF (p value <0.001 PMF vs prePMF).

LDH serum level was significantly higher in PMF compared to both ET and prePMF (p value <0.001 ET vs PMF and <0.001 PMF vs prePMF) and was higher in prePMF compared to ET (p value <0.001 ET vs prePMF). CD34+ circulating cells were significantly higher in PMF when compared both to ET and prePMF (p value <0.001 ET vs PMF and <0.001 PMF vs prePMF) and were significantly higher in prePMF than in ET (p value <0.001 ET vs prePMF).

Considering the driver mutational status (Table 6), 309 patients carried the *JAK2*V617F mutation (61% of PMF patients, 54% of prePMF, 63% of ET), 127 patients had a *CALR* mutation (27% of PMF patients, 36% of prePMF, 19% of ET),

28 had an *MPL* exon 10 mutation (7% of the PMF, 4% of prePMF, 5% of ET) and 45 patients were triple negative (TN, 5% of PMF, 6% of prePMF, 13% of ET).

Table 6 Driver mutations in the cohort

	ET n %	**PMF n %**	**prePMF n %**	**Total n %**
JAK2 V617F- mutated patients n (%)	149 (63%)	111 (61%)	49 (54%)	**309 (61%)**
CALR - mutated patients n (%)	45 (19%)	49 (27%)	33 (36%)	**127 (25%)**
MPL - mutated patients n (%)	12 (5%)	12 (7%)	4 (4%)	**28 (5%)**
Triple negative patients n (%)	30 (13%)	10 (5%)	5 (6%)	**45 (9%)**

Patients with additional somatic mutations

Patients carrying at least one additional somatic mutation were 278 (54.6% of the total); most PMF patients (157 out of 182, 86.3%) showed at least one additional mutation, compared to 34.8% of ET (82 out of 236) and 42.9% of prePMF (39 out of 91). The difference between PMF and both ET and prePMF was statistically significant ($p<0.001$ for both), while ET and prePMF did not show any significant difference.

ET patients had a significantly lower risk of carrying an additional somatic mutation (relative risk reduction, RRR 0.07) than PMF patients. PrePMF had a lower incidence of additional mutations compared to PMF (RRR 0.1), but had a similar risk to ET (RRR 1.36). We found a statistically significant difference between PMF and both ET and prePMF (p value <0.001 ET vs PMF and p value <0.001 prePMF vs PMF). There was no statistically significant difference between ET and prePMF (Table 7).

Table 7 Relative risk of having at least 1 additional mutation according to diagnosis

	ET vs PMF			prePMF vs PMF			prePMF vs ET		
	RRR	**95%CI**	**p**	**RRR**	**95%CI**	**p**	**RRR**	**95%CI**	**p**
At least 1 additional mutation	0.07	0.04-0.13	<0.001	0.10	0.05-0.19	<0.001	1.36	0.83-2.24	0.221

Considering the number of additional somatic mutations (Figure 7), we found that most ET (65.3%) and prePMF (57.1%) patients did not show any, while only 13.7% of PMF patients carried no additional mutations. On the other hand, more than a third of the PMF (36.3%) had 3 or more additional mutations, compared to 5.1% in ET and 7.7% in prePMF (p <0.001, no statistically significant difference between ET and prePMF).

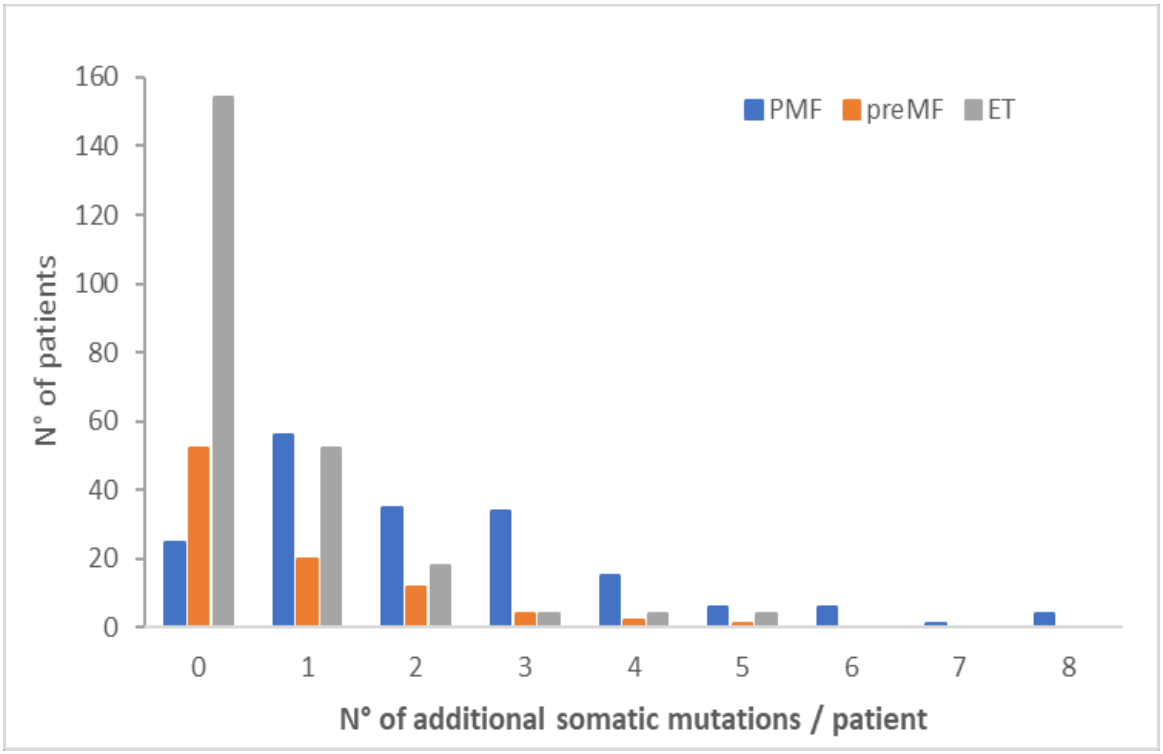

Figure 7. Number of additional somatic mutations per patient considering clinical diagnosis

Categorizing patients according to the driver mutation, independently of the diagnosis, we detected at least one additional somatic variant in 182 out of 309 *JAK2* mutants (59%), in 59 out of 127 *CALR* mutants (46.5%), in 23 out of 28 *MPL* mutated patients (82%) and in 14 out of 45 triple negative individuals (31%).

The number of patients with at least one additional somatic mutation was significantly different among the 4 molecular groups (p value 0.002), with *MPL* and TN patients showing the highest and lowest percentage respectively (Figure 8).

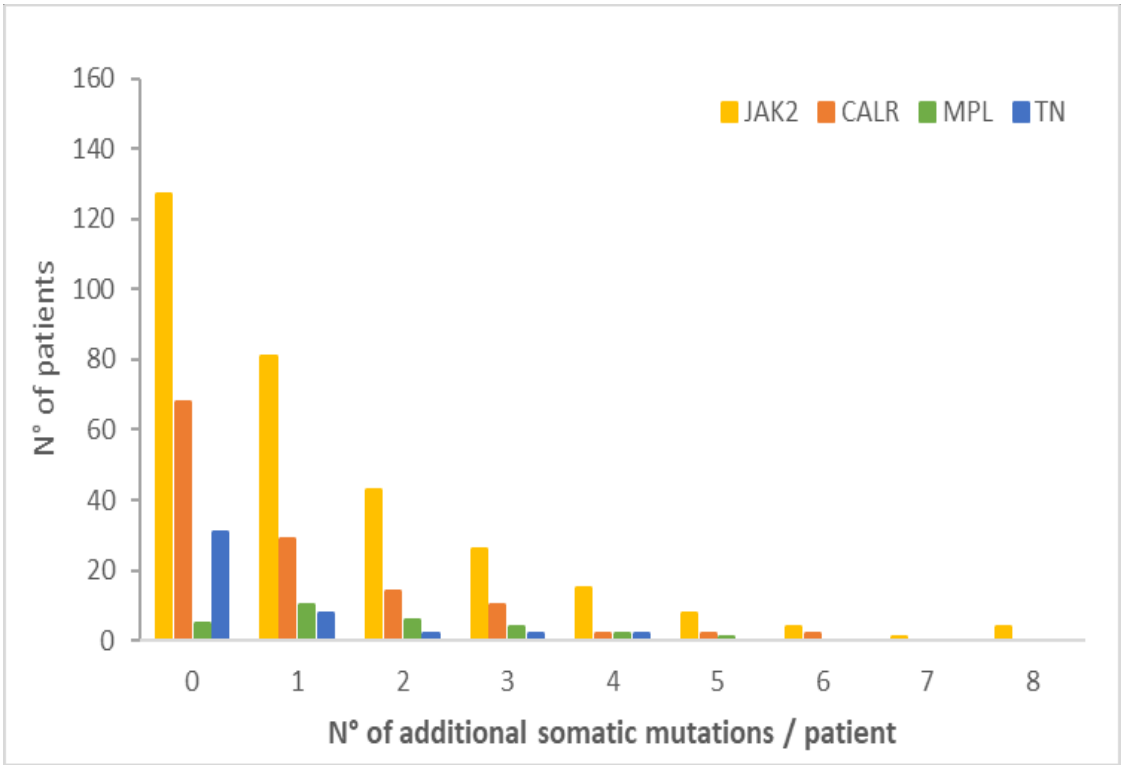

Figure 8. Number of additional somatic mutations per patient considering driver mutations

The median number of mutations was significantly higher in *JAK2* mutated patients compared to *CALR* (p 0.036) and in the *MPL* group compared to *CALR* (0.003), while we did not find a significant difference between *JAK2* and *MPL* groups. TN patients had a significantly lower median number of mutations compared to *MPL* and *JAK2* patients (p <0.001 and p 0.002 respectively), but there was no difference compared to *CALR* patients.

Mutated genes and pathways

Overall, we detected 598 somatic variants in 59 genes out of 81 genes targeted in the panel. Specifically, we found 393 variants in 50 genes in PMF, 69 variants in 24 genes in prePMF and 136 variants in 34 genes in ET.

As pictured in Figure 9, where genes were sorted according to their mutation frequency in PMF, the most frequently involved genes in PMF were *ASXL1*, *TET2*, *EZH2*, *SRSF2* and *U2AF1*; in prePMF the most common were *SRSF2*, *DNMT3A*,

TET2, *ASXL1* and *SF3B1*. In ET patients the most frequently mutated gene was *TET2*, followed by *DNMT3A*, *ASXL1*, *NF1* and *SF3B1*.

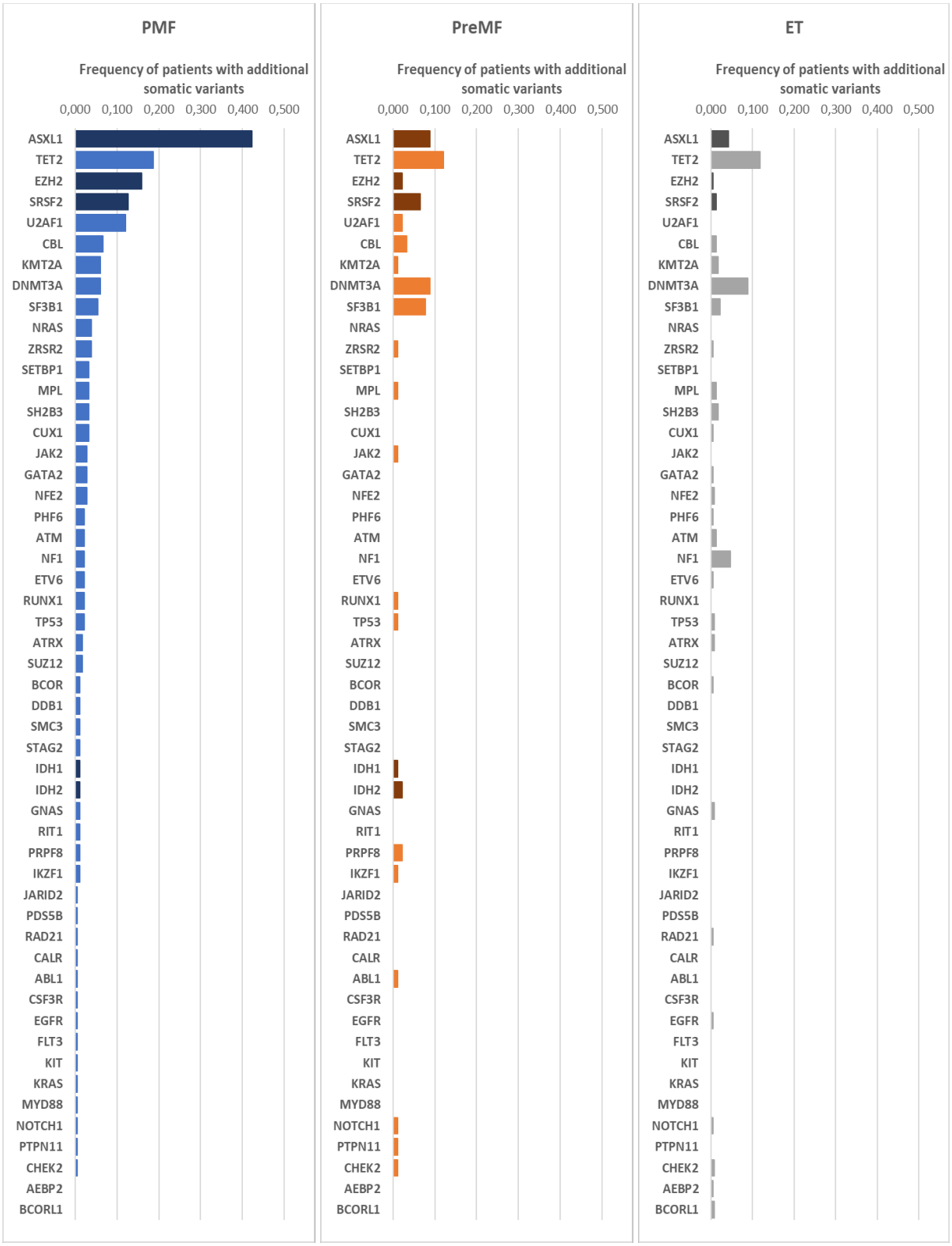

Figure 9. Frequency of additional variants according to clinical diagnosis. Genes are sorted in descending order accordingly to their mutation frequency in PMF. The high molecular risk (HMR) genes are highlighted as darker bars

A similar frequency of *TET2* mutation was observed in ET, prePMF and PMF (12%, 12% and 19% respectively). *DNMT3A* was more often involved in prePMF and ET (9% in both) than in PMF (6%), but the difference was not statistically significative.

We found a higher prevalence of *SRSF2* in PMF (13%) than ET (1%), with a statistically significant difference ($p<0.001$). There was no significant difference between PMF (13%) and prePMF (7%). PrePMF had a higher prevalence of *SRSF2* mutations than ET and the difference was borderline (p 0.048).

ASXL1 showed a significantly higher variant frequency in PMF (42%) compared to prePMF and ET (9% and 4% respectively); we did not observe a statistically significant difference between ET and prePMF.

Focusing on the high molecular risk genes (*SRSF2*, *ASXL1*, *IDH1*/*2*, *EZH2*)[73], we observed that 73% of PMF patients showed at least 1 variant, while 21% of prePMF and only 6% of ET patients carried at least 1 HMR mutation (Table 8). Interestingly, we did not detect any mutation in *IDH1*/*2* in ET patients.

Table 8 Frequency of HRM variants according to clinical diagnosis

	PMF (n=182)	**preMF (n=91)**	**ET (n=236)**
ASXL1	77	8	10
EZH2	29	2	1
SRSF2	23	6	3
IDH1	2	1	0
IDH2	2	2	0
Total	**133 (73%)**	**19 (21%)**	**14 (6%)**

Categorizing patients based on the driver mutation, we observed that the most frequently involved genes in *JAK2* mutants were *ASXL1*, *TET2*, *DNMT3A*, *SRSF2*

and *U2AF1*; in *CALR* mutants *ASXL1*, *EZH2*, *TET2*, *DNMT3A* and *KMT2A*. In *MPL* patients the most frequently mutated genes were *SRSF2, DNMT3A, ASXL1*, *TET2* and *IDH2*; notably, in this population we also detected additional noncanonical variants in *MPL* and *JAK2* (4 non canonical *MPL* and 1 mutation *JAK2*). Triple negative subjects had generally a lower frequency of additional mutations, with *TET2, EZH2* and *DNMT3A* being more frequently involved, followed by *ASXL1, MPL* and *SRSF2*; we found 4 noncanonical *MPL* mutations and 1 mutation in *JAK2* (11% of TN patients). Frequency of additional mutations according to the driver mutation are showed in Figure 10, where genes are sorted in descending order accordingly to their frequency in *JAK2* mutants.

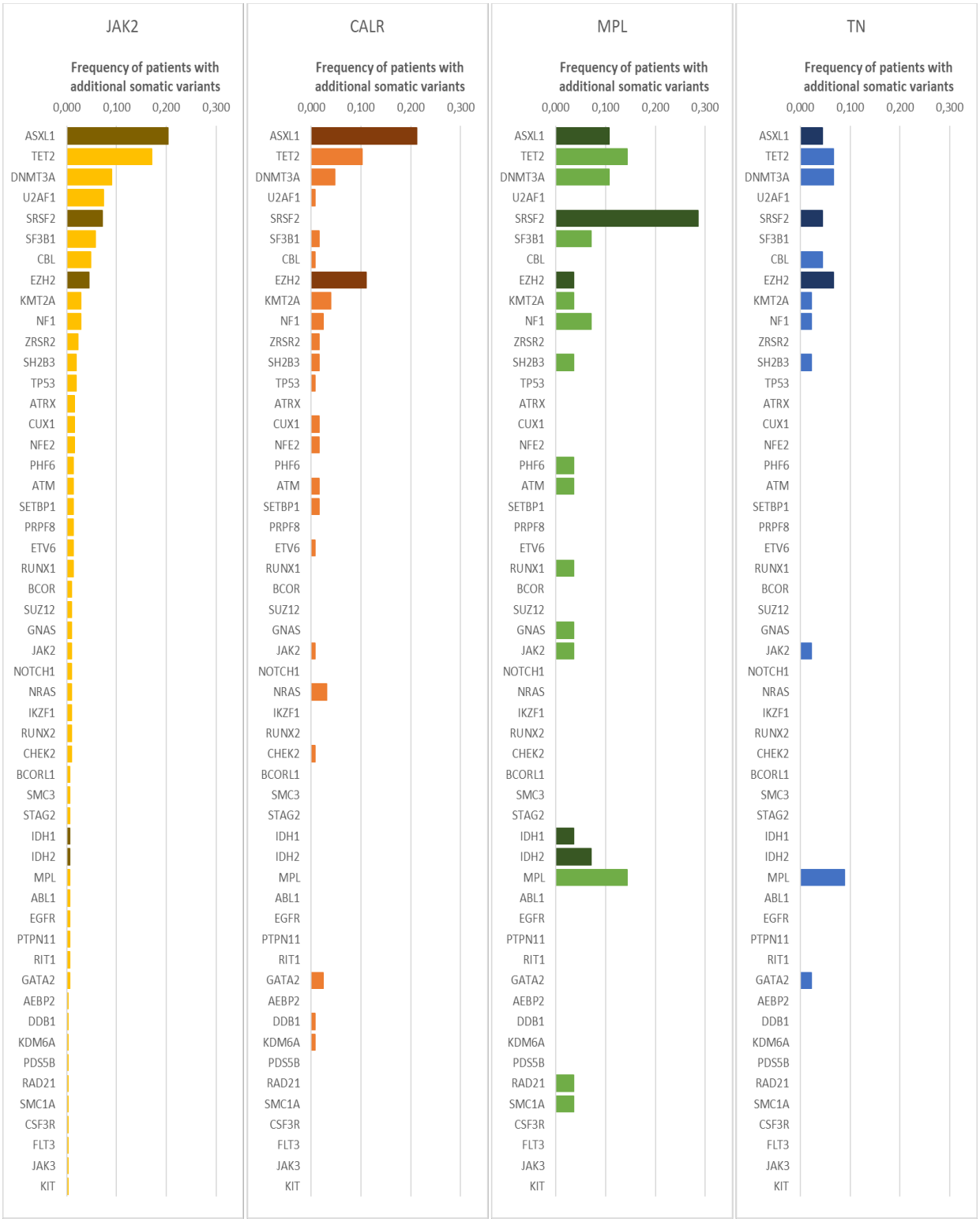

Figure 10. Frequency of additional variants according to the driver mutation. Genes are sorted in descending order accordingly to their mutation frequency in PMF. The high molecular risk (HMR) genes are highlighted as darker bars

Focusing on the high molecular risk genes[73], we observed that *MPL* mutated patients had the highest frequency of at least 1 HMR variant (54%), while TN had the lowest (16%). *JAK2* and *CALR* mutants had similar prevalence of at least 1 HMR variants (33% and 32% respectively). Interestingly, we did not detect any mutation in *SRSF2* nor in *IDH1*/*2* in *CALR* mutated patients (Table 9).

Table 9 Frequency of HRM variants according to the driver mutation

	JAK2 (n=309)	**CALR (n=127)**	**MPL (n=28)**	**TN (n=45)**
ASXL1	63	27	3	2
EZH2	14	14	1	3
SRSF2	22	0	8	2
IDH1	2	0	1	0
IDH2	2	0	2	0
Total	**103 (33%)**	**41 (32%)**	**15 (54%)**	**7 (16%)**

Additional somatic mutations were then classified according to the cellular pathway in which they are involved in the cell cycle. The mutated pathways were significantly different among the 3 WHO entities. We observed that the most frequently affected pathway in patients carrying at least one additional mutation was chromatin structure regulation (54%) in PMF, followed by the splicing machinery (33,5%) and the DNA methylation proteins (25%). In ET, the most involved pathway was the DNA methylation (18%), as well as in prePMF (21%), but we also found a high prevalence of mutations in the splicing machinery in the latter (17,6%). (Figure 11)

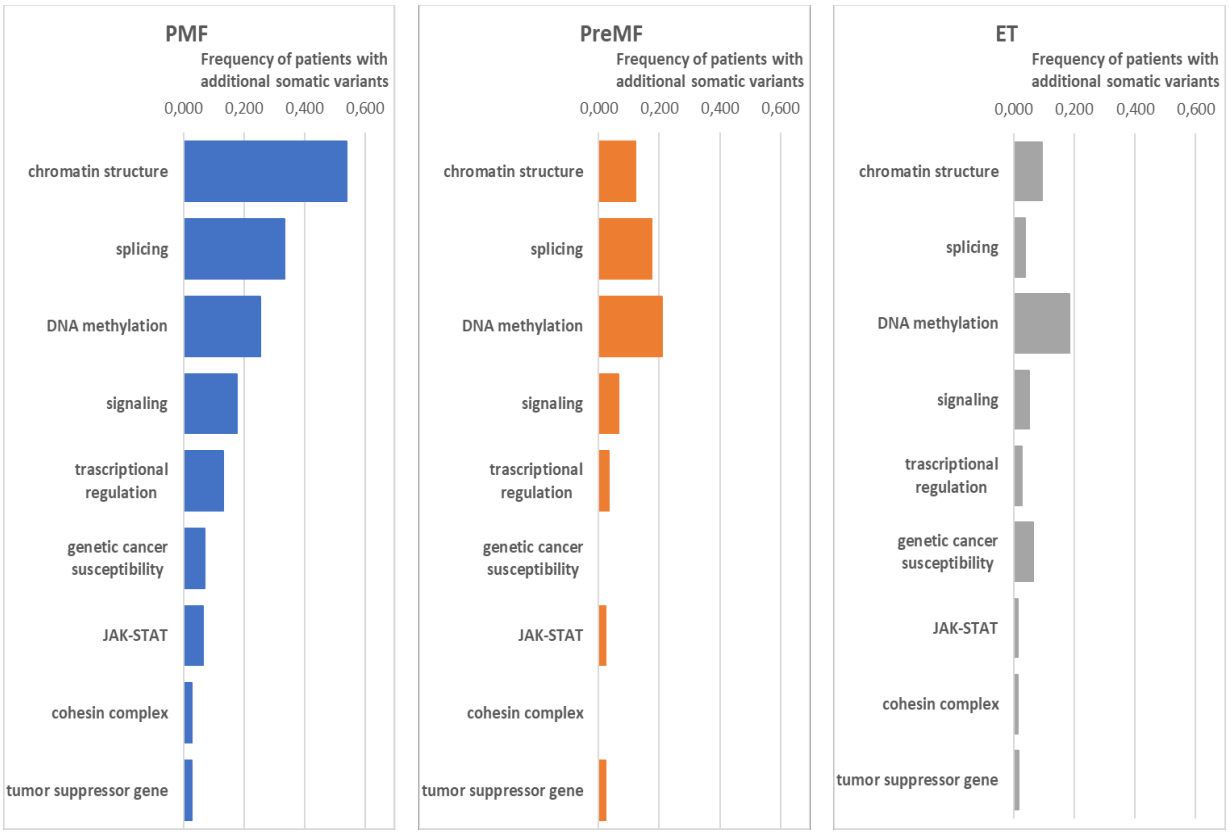

Figure 11. Frequency of patients having variants in the different pathways

In detail, chromatin structure was significantly more affected in PMF than in ET and prePMF ($p < 0.001$ for both) and the risk of finding a mutation in that pathway was 13 folds higher in PMF than ET (RRR 13.8, IC 7.76-24.598). We observed a significant difference in the splicing machinery genes. This pathway was significantly higher in PMF than ET ($p < 0.001$) and prePMF (p 0.007) and was significantly higher in prePMF than in ET ($p < 0.001$). Compared to ET, patients with a PMF had a 13 folds higher risk of having a mutation in genes of the splicing (RRR 13.507, IC 6.286-29.022), while prePMF had a 5 folds higher risk (RRR 5.307, IC 2.249-12.525). Genes involved in transcriptional regulation were more frequently affected in PMF than in ET ($p<0.001$) and we observed a trend towards a greater involvement in PMF compared to prePMF (p 0.02). We did not find statistically significant differences in the DNA methylation pathway. (Table 10)

Table 10. Pathways of additional somatic variants according to clinical diagnosis

	Totale (n=509)	**ET (n=236)**	**PMF (n=182)**	**prePMF (n=91)**	**ET vs PMF**	**ET vs pre**	**PMF vs pre**
JAK-STAT (excluding driver mut)	17 (3.3%)	3 (1.3%)	12 (6.6%)	2 (2.2%)	NS	NS	NS
Chromatin structure	131 (25.7%)	22 (9.3%)	98 (53.9%)	11 (12.1%)	<0.001	NS	<0.001
DNA methylation	108 (21.2%)	43 (18.2%)	46 (25.3%)	19 (20.9%)	NS	NS	NS
Splicing	86 (16.9%)	9 (3.8%)	61 (33.5%)	16 (17.6%)	<0.001	<0.001	0.007
Signaling	50 (9.8%)	12 (5.1%)	32 (17.6%)	6 (6.6%)	<0.001	NS	NS
Transcriptional regulation	33 (6.5%)	6 (2.5%)	24 (13.2%)	3 (3.3%)	<0.001	NS	0.020
Genetic cancer susceptibility	28 (5.5%)	15 (6.4%)	13 (7.1%)	0 (0.0%)	NS	NS	NS
Tumor suppressor gene	11 (2.2%)	4 (1.7%)	5 (2.8%)	2 (2.2%)	NS	NS	NS
Cohesin complex	8 (1.6%)	3 (1.3%)	5 (2.8%)	0 (0.0%)	NS	NS	NS

Categorizing patients based on the driver mutation, genes regulating chromatin structure were the most frequently affected in *CALR* mutants; in *JAK2* mutants the most frequently affected pathways were those regarding DNA methylation,

chromatin structure and splicing; *MPL* mutants showed more often involvement of the DNA methylation and the splicing regulating genes, while TN had more often additional mutations in genes responsible for DNA methylation, followed by genes involved in chromatin structure regulation and signaling (Figure 12).

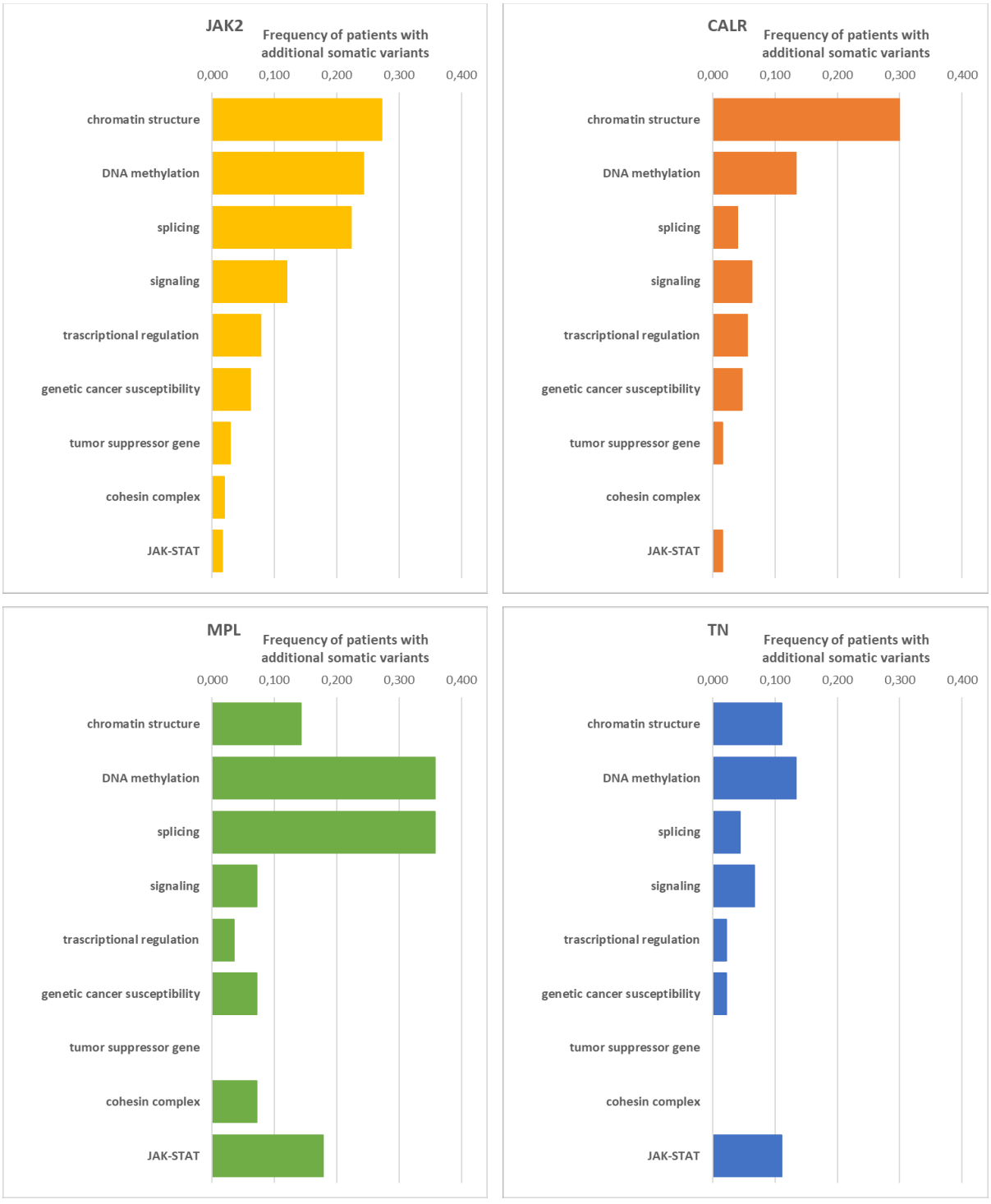

Figure 12. Frequency of patients with additional somatic variants according to the driver mutation

We observed that there was a statistically significant difference in the splicing pathway and in the JAK2-STAT pathways only. Genes regulating the splicing were less frequently affected in *CALR* or TN mutated patients, while there was no

difference between *MPL* and *JAK2* mutants (p <0.001). Additional mutations in the JAK-STAT pathway were more frequently detected in patients with *MPL* exon 10 mutated MPN (Table 11).

Table 11 Pathways of additional somatic variants according to the driver mutation

	CALR (n=127)	**JAK2 (n=309)**	**MPL (n=28)**	**TN (n=45)**	**CALR vs JAK2**	**CALR vs MPL**	**CALR vs TN**	**JAK2 vs MPL**	**JAK2 vs TN**	**MPL vs TN**
JAK-STAT excluding driver mut	2 (1.6%)	5 (1.6%)	5 (17.9%)	5 (11.1%)	NS	0.013	NS	0.003	0.025	NS
Chromatin structure	38 (29.9%)	84 (27.2%)	4 (14.3%)	5 (11.1%)	NS	NS	NS	NS	NS	NS
DNA methylation	17 (13.4%)	75 (24.3%)	10 (35.7%)	6 (13.3%)	NS	NS	NS	NS	NS	NS
Splicing	5 (3.9%)	69 (22.3%)	10 (35.7%)	2 (4.4%)	<0.001	<0.001	NS	NS	0.016	0.005
Signaling	8 (6.3%)	37 (12.0%)	2 (7.1%)	3 (6.7%)	-	-	-	-	-	-
Transcriptional regulation	7 (5.5%)	24 (7.8%)	1 (3.6%)	1 (2.2%)	NS	NS	NS	NS	NS	NS
Genetic cancer susceptibility	6 (4.7%)	19 (6.2%)	2 (7.1%)	1 (2.2%)	NS	-	-	-	-	-
Tumor suppressor gene	2 (1.6%)	9 (2.9%)	0 (0.0%)	0 (0.0%)	-	-	-	-	-	-
Cohesin complex	0 (0.0%)	6 (1.9%)	2 (7.1%)	0 (0.0%)	-	-	-	-	-	-

Timing of additional somatic mutations

We analysed the timing of additional somatic variants acquisition in relation to the driver mutation, to define a possible temporal hierarchy and consequently a model for the clonal evolution of ET, PMF and prePMF.

Based on the notion that the VAF is a direct measure of the mutant clone size and therefore of the precocity of the acquisition of that mutation (that is, the higher is the VAF of a certain mutation, the earlier that mutation occurred during disease evolution), in each patient we compared the VAF of additional somatic variants with the allele burden of the co-occurring driver mutation to define the temporal order of their acquisition. Somatic variants with a VAF lower than the allele burden of the driver mutation were classified as “post-driver”, because it is likely that they were acquired after the driver mutation. On the contrary, somatic variants with a VAF higher than that of the driver mutation were classified as “pre-driver”. Additional variants with VAF similar to the driver mutation burden were classified as “synchro to driver”. All the additional mutations and their timing are showed in Table 12.

We observed that mutations in the most common affected genes (*ASXL1*, *TET2*, *EZH2*, *SRSF2*) were more frequently acquired after the driver mutation; the exception was *DNMT3A*, being more frequently mutated at the same time as the driver. Overall, the great majority of the additional variants was acquired simultaneously or after the driver mutation.

Categorizing patients according to their clinical diagnosis, we found that, in ET patients, 17 out of 136 total additional mutations (12,5%) were acquired before the driver mutation, 51 (37.5%) were synchronous and 68 (50%) were acquired later. In PMF, only 36 out of 393 (9%) occurred earlier than the driver mutation, 128 (32%) were synchronous and 232 (59%) were acquired afterwards. In prePMF, we found a 15% (10 out of 69) of early acquired somatic mutations, 42% of synchronous (29) and 44% (30) were later events.

Categorizing according to the driver mutation, we observed that in *CALR* mutants most mutations occurred later than the driver (67 out of 117, 57%) or were synchronous (37 out of 117, 32%). In *JAK2* mutants, 48 out of 408 mutations (12%) occurred earlier than the driver, 147 (36%) simultaneously and 213 (52%) later. In

MPL mutants, 2 out of 47 (4%) were acquired earlier, 12 (45%) were synchronous and 24 (51%) occurred later.

Table 12 Timing of additional mutations

Gene	Pre driver	Synchro to driver	Post driver	Number of mutations
ASXL1	9 (7,8%)	31 (27%)	75 (65,2%)	115
TET2	4 (4,6%)	34 (39,1%)	49 (56,3%)	87
DNMT3A	7 (15,9%)	20 (45,5%)	17 (38,6%)	44
EZH2	5 (12,8%)	8 (20,5%)	26 (66,7%)	39
SRSF2	2 (6,3%)	11 (34,4%)	19 (59,4%)	32
U2AF1	3 (12,5%)	17 (70,8%)	4 (16,7%)	24
SF3B1	4 (17,4%)	14 (60,9%)	5 (21,7%)	23
CBL	4 (21,1%)	2 (10,5%)	13 (68,4%)	19
NF1	1 (5,9%)	5 (29,4%)	11 (64,7%)	17
KMT2A	0 (0%)	2 (12,5%)	14 (87,5%)	16
SH2B3	1 (9,1%)	3 (27,3%)	7 (63,6%)	11
ZRSR2	1 (9,1%)	3 (27,3%)	7 (63,6%)	11
MPL	0 (0%)	3 (30%)	7 (70%)	10
NFE2	2 (20%)	1 (10%)	7 (70%)	10
CUX1	2 (25%)	4 (50%)	2 (25%)	8
GATA2	2 (25%)	2 (25%)	4 (50%)	8
ATM	1 (14,3%)	1 (14,3%)	5 (71,4%)	7
NRAS	0 (0%)	3 (42,9%)	4 (57,1%)	7
SETBP1	1 (14,3%)	2 (28,6%)	4 (57,1%)	7

TP53	1 (14,3%)	4 (57,1%)	2 (28,6%)	7
JAK2	0 (0%)	3 (50%)	3 (50%)	6
ATRX	1 (20%)	0 (0%)	4 (80%)	5
ETV6	0 (0%)	2 (40%)	3 (60%)	5
IDH2	0 (0%)	0 (0%)	5 (100%)	5
PHF6	1 (20%)	3 (60%)	1 (20%)	5
RUNX1	0 (0%)	4 (80%)	1 (20%)	5
CHEK2	1 (25%)	0 (0%)	3 (75%)	4
GNAS	1 (25%)	2 (50%)	1 (25%)	4
PRPF8	2 (50%)	0 (0%)	2 (50%)	4
BCOR	0 (0%)	2 (66,7%)	1 (33,3%)	3
IDH1	0 (0%)	2 (66,7%)	1 (33,3%)	3
IKZF1	0 (0%)	1 (33,3%)	2 (66,7%)	3
NOTCH1	2 (66,7%)	0 (0%)	1 (33,3%)	3
RUNX2	1 (33,3%)	1 (33,3%)	1 (33,3%)	3
SUZ12	0 (0%)	0 (0%)	3 (100%)	3
ABL1	1 (50%)	1 (50%)	0 (0%)	2
BCORL1	0 (0%)	2 (100%)	0 (0%)	2
DDB1	0 (0%)	1 (50%)	1 (50%)	2
EGFR	0 (0%)	1 (50%)	1 (50%)	2
KDM6A	1 (50%)	0 (0%)	1 (50%)	2
PTPN11	1 (50%)	1 (50%)	0 (0%)	2

RAD21	0 (0%)	0 (0%)	2 (100%)	2
RIT1	0 (0%)	0 (0%)	2 (100%)	2
SMC1A	0 (0%)	2 (100%)	0 (0%)	2
SMC3	0 (0%)	1 (50%)	1 (50%)	2
STAG2	0 (0%)	1 (50%)	1 (50%)	2
AEBP2	1 (100%)	0 (0%)	0 (0%)	1
BRAF	0 (0%)	0 (0%)	1 (100%)	1
CALR	0 (0%)	1 (100%)	0 (0%)	1
CSF3R	0 (0%)	1 (100%)	0 (0%)	1
FLT3	0 (0%)	0 (0%)	1 (100%)	1
JAK3	0 (0%)	0 (0%)	1 (100%)	1
JARID2	0 (0%)	0 (0%)	1 (100%)	1
KIT	0 (0%)	0 (0%)	1 (100%)	1
KRAS	0 (0%)	0 (0%)	1 (100%)	1
MYD88	0 (0%)	1 (100%)	0 (0%)	1
PDS5B	0 (0%)	1 (100%)	0 (0%)	1
PIK3CA	0 (0%)	0 (0%)	1 (100%)	1
STAT3	0 (0%)	1 (100%)	0 (0%)	1

Genotype-phenotype correlations

We finally analysed correlations of additional mutations and clinical parameters (age, gender, alterations in full blood count, splenomegaly, increased LDH, increased CD34+ circulating cells), overall survival (OS) and risk of evolution.

We focused on the most common mutations, defined as mutations detected in at least 10 patients (frequency>5%): *ASXL1*, *DNMT3A*, *TET2*, *EZH2*, *SRSF2*. All the statistical analysis was adjusted considering the possible treatment.

Mutations in *ASXL1* were positively correlated with age and with male gender: people >=60 years had a significantly higher risk of acquiring an additional variant (OR 2.54) than people younger than 60. Females had a reduced risk of having a mutation in *ASXL1* compared to males.

ASXL1 mutations were also associated with splenomegaly and anemia; a pathologic white blood cell count and increased LDH were more frequent but the association was statistically borderline, while the platelet count was more often normal ($p<0.001$). CD34+ circulating cells were significantly higher in *ASXL1* mutated patients. (Table 13)

Table 13. Clinical association of ASXL1 mutation

***ASXL1* mutation**	Association adjusted according to time to NGS and treatment
Age (≥ vs <60)	OR=2.5; 95%CI: 1.6-4.2 p<0.001
Gender (F vs M)	OR=0.35; 95%CI: 0.2-0.6 p<0.001
Platelet count (pathologic vs normal)	OR=0.25; 95%CI: 0.2-0.4 p<0.001
WBC (pathologic vs normal)	OR=2.3; 95%CI: 1.4-3.7 p<0.001
Hb (low vs normal)	OR=5.6; 95%CI: 3.3-9.7 p<0.001
Palpable spleen (>0 vs 0)	OR=8.3; 95%CI: 4.8-14.7 p<0.001
LDH (> 220 mU/ml vs ≤220 mU/ml)	OR=3.5; 95%CI: 1.7-7.1 p=0.001
CD34+ circulating cells (>12/mcl vs ≤12/mcl)	OR=7.2; 95%CI: 4-12.7 p<0.001

TET2 was associated with age, but not with gender. We did not find any correlation with clinical parameters (Table 14)

Table 14. Clinical association with TET2 mutations

***TET2* mutation**	Association adjusted according to time to NGS and treatment
Age (≥ vs <60)	OR=3.8; 95%CI: 2.1-6.7 p p<0.001
Gender (F vs M)	OR=1.6; 95%CI: 0.9-2.7 p NS
Platelet count (pathologic vs normal)	OR=1.1; 95%CI: 0.6-2.1 p NS
WBC (pathologic vs normal)	OR=1; 95%CI: 0.6-1.7 p NS
Hb (low vs normal)	OR=1; 95%CI: 0.5-1.7 p NS
Palpable spleen (>0 vs 0)	OR=1.5; 95%CI: 0.8-2.6 p NS

LDH (> 220 mU/ml vs ≤220 mU/ml)	OR=1.4; 95%CI: 0.7-2.6 p NS
CD34+ circulating cells (>12/mcl vs ≤12/mcl)	OR=1.9; 95%CI: 1-3.5 p NS

DNMT3A was not associated with age or gender and did not find any correlation with clinical parameters (Table 15).

Table 15. Clinical associations with DNMT3A mutations

***DNMT3A* mutation**	Association adjusted according to time to NGS and treatment
Age (≥ vs <60)	OR=1.4; 95%CI: 0.7-2.7 p NS
Gender (F vs M)	OR=1.3; 95%CI: 0.7-2.6 p NS
Platelet count (pathologic vs normal)	OR=0.9; 95%CI: 0.4-2.2 p NS
WBC (pathologic vs normal)	OR=0.8; 95%CI: 0.4-1.6 p NS
Hb (low vs normal)	OR=0.7; 95%CI: 0.3-1.4 p NS
Palpable spleen (>0 vs 0)	OR=0.7; 95%CI: 0.3-1.6 p NS
LDH (> 220 mU/ml vs ≤220 mU/ml)	OR=1.9; 95%CI: 0.8-4.3 p NS
CD34+ circulating cells (>12/mcl vs ≤12/mcl)	OR=1; 95%CI: 0.4-2.2 p NS

EZH2 was not associated with age or gender; we found a correlation with splenomegaly and increased CD34+ circulating cells, but no significant correlation with blood parameters, even if a trend of association with anemia was found (Table 16).

Table 16. Clinical associations with EZH2 mutations

***EZH2* mutation**	Association adjusted according to time to NGS and treatment
Age (≥ vs <60)	OR=2.4; 95%CI: 1.1-5.4 p NS
Gender (F vs M)	OR=0.5; 95%CI: 0.1-1.1 p NS
Platelet count (pathologic vs normal)	OR=0.3; 95%CI: 0.1-0.6 p NS
WBC (pathologic vs normal)	OR=2.6; 95%CI: 1.2-5.6 p NS
Hb (low vs normal)	OR=3.7; 95%CI: 1.6-8.3 p NS
Palpable spleen (>0 vs 0)	OR=6.7; 95%CI: 2.8-16.1 p<0.001
LDH (> 220 mU/ml vs ≤220 mU/ml)	OR=4.1; 95%CI: 1.2-14.2 p NS
CD34+ circulating cells (>12/mcl vs ≤12/mcl)	OR=5.6; 95%CI: 2.3-13 p<0.001

SRSF2 was not associated with gender and the association with age was borderline; we did not find any correlation with clinical parameters, but there was a trend to a higher frequency of anemia in *SRSF2* mutated patients (Table 17)

Table 17. Clinical associations with SRSF2 mutations

***SRSF2* mutation**	Association adjusted according to time to NGS and treatment
Age (≥ vs <60)	OR=5.2; 95%CI: 1.9-14 p=0.001
Gender (F vs M)	OR=0.2; 95%CI: 0.1-0.6 p NS
Platelet count (pathologic vs normal)	OR=0.3; 95%CI: 0.1-0.6 p NS
WBC (pathologic vs normal)	OR=1.4; 95%CI: 0.7-3 p NS
Hb (low vs normal)	OR=3.8; 95%CI: 1.6-9.3 p NS

Palpable spleen (>0 vs 0)	OR=2; 95%CI: 0.9-4.5 p NS
LDH (> 220 mU/ml vs ≤220 mU/ml)	OR=3.6; 95%CI: 1-12.4 p NS
CD34+ circulating cells (>12/mcl vs ≤12/mcl)	OR=1.7; 95%CI: 0.7-4.1 p NS

In the end, we analyzed the impact of the most common additional mutation and of the involved pathways on progression to blastic phase and on the overall survival (OS).

We observed a negative role on evolution in patients carrying additional mutations in *ASXL1*, *EZH2* and *SRSF2,* while mutations in *TET2* or *DNMT3A* apparently had no significant impact on progression. (Table 18)

Table 18. Impact of additional mutations in disease evolution to blastic phase

	With left-truncation
ASXL1	sHR=4.5 (1.7-11.8) p=0.003
TET2	sHR=1.4 (0.4-4.8) p=0.600
DNMT3A	sHR=1.6 (0.4-6.9) p=0.516
EZH2	sHR=4.8 (1.6-14.4) p=0.005
SRSF2	sHR=5.2 (1.7-15.7) p=0.003

We found that patients having mutations in genes involved in the chromatin structure regulation, in the splicing machinery and in the signaling pathway had significantly increased risk of progression to blastic phase (Table 19)

Table 19. Impact of mutated pathway in evolution to blastic phase

	With left-truncation
JAK-STAT	sHR=5.45 (1.3-23) p=0.021
Chromatin structure	sHR=4.8 (1.8-12.7) p=0.002
DNA methylation	sHR=2.8 (1.1-7.4) p=0.035
Splicing	sHR=6.3 (2.4-16.4) p<0.001
Signaling	sHR=6.6 (2.4-17.9) p<0.001
Transcriptional regulation	sHR=3.5 (0.95-12.6) p=0.061
Genetic cancer susceptibility	na

Considering the OS, we observed that variants in *ASXL1* and *SRSF2* had a negative impact on prognosis, while the mutations in chromatin regulation, in the JAK2-STAT pathway, in DNA methylation, splicing and signaling pathways and transcriptional regulation were associated to reduced OS (Table 20 and 21)

Table 20. Impact of additional mutations on OS

	With left-truncation
ASXL1	HR=6.2 (4.3-9.0) p<0.001
TET2	HR=1.8 (1.1-2.8) p=0.012
DNMT3A	HR=1.0 (0.6-2.0) p=0.898
EZH2	HR=1.9 (1.0-3.5) p=0.036
SRSF2	HR=5.1 (3.1-8.5) p<0.001

Table 21 Impact of mutated pathways on OS

	With left-truncation
JAK-STAT	HR=6.8 (3.4-13.4) p<0.001
Chromatin structure	HR=4.6 (3.2-6.6) p<0.001
DNA methylation	HR=1.8 (1.2-2.7) p=0.004
Splicing	HR=6.6 (4.5-9.6) p<0.001
Signaling	HR=2.9 (1.8-4.8) p<0.001
Transcriptional regulation	HR=4.4 (2.7-7.4) p<0.001
Genetic cancer susceptibility	HR=1.3 (0.6-2.7) p=0.461

Discussion

The three classic Philadelphia-negative myeloproliferative neoplasms (MPNs), polycythemia vera (PV), essential thrombocythemia (ET) and primary myelofibrosis (PMF) are related clonal hematologic disorders caused by the expansion of a single somatically mutated hematopoietic stem cell (HSC), resulting in the accumulation of one or more myeloid cell lineages and an increased risk to evolve in acute myeloid leukemia (AML), called blastic phase.

The year 2005, with the discovery of the *JAK2* V617F point mutation, was the turning point in knowledge on the molecular mechanisms of MPNs (at that time still called "diseases")[13-16]. This initial discovery led to a progressive understanding of the genetic landscape of MPNs, with the identification of *MPL* exon 10 mutations[22,23] and *CALR* exon 9 indels[28,29].

Dysregulation of *JAK2* signaling has emerged as a central pathogenetic mechanism of MPNs, so that the World Health Organization (WHO) recently included all the mutations known so far in the 2016 revision of the diagnostic criteria[4]. The MPN-restricted driver mutations (those in *JAK2*, *CALR*, and *MPL*) abnormally activate the cytokine receptor/JAK2 pathway and their downstream effectors, especially the STATs. The most frequent mutation, *JAK2* V617F, activates the 3 main myeloid cytokine receptors (erythropoietin receptor, granulocyte colony-stimulating factor receptor, and MPL) whereas *CALR* or *MPL* mutants are restricted to MPL activation. This explains why *JAK2* V617F is associated with PV, ET, and PMF, whereas *CALR* and *MPL* mutants are found in ET and PMF.[44] Although mutations in *JAK2*, *MPL* and *CALR* are sufficient to lead to MPN development, the genetic landscape of MPNs is not as simple as initially seemed. The first suggestion of their clonal complexity came with the finding that *JAK2*-mutated MPNs may transform into *JAK2*-unmutated AML, supporting the existence of other mutant subclones[74].

Exome sequencing studies showed that, in addition the so-called driver mutations, multiple somatic acquired mutations may also be present. These additional variants affect functionality of DNA methylation pathway (*TET2*, *DNMT3A*, *IDH1*/*2*) and regulation of chromatin structure (*ASXL1*, *EZH2*), tumor suppressor genes (*TP53*)

and transcriptional regulators (*ETV6*, *RUNX1*), signaling pathway (*NRAS*, *CBL*, *FLT3*, *LNK*) and splicing factors (*SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2*). Mutations in these genes, which are often found in other myeloid neoplasms, such as myelodysplastic syndromes (MDS) and AML, are usually associated with progression of the disease. Nowadays, it is evident that both heterogeneity of classical MPNs and prognosis are determined by a specific genomic landscape, that is, type of MPN driver mutations, additional mutations, and their order of acquisition, as recently reviewed[44,75]. It is indeed well established that, even if MPNs have many clinical and biological aspects in common, including the constitutive activation of JAK/STAT signaling, they also exhibit a wide range of distinct clinical phenotypes, according to the driver mutations and subclonal variants acquired.[12]

PV is a condition almost exclusively associated with *JAK2* mutations, its main feature is erythrocytosis and it may progress to secondary myelofibrosis because of a dominant clone of cells that are homozygous for the *JAK2* mutation, while the occurrence of subclonal mutations may lead to transformation into AML. ET is mainly characterized by thrombocytosis and by an increased risk of vascular events and may progress to more aggressive myeloid neoplasms; *JAK2* V617F-mutant ET has an increased risk of thrombosis, and may progress to PV while *CALR*-mutant ET shows a lower risk of thrombosis and no evolution to a polycythemic phenotype. ET may progress to secondary myelofibrosis, associated with occurrence of a dominant clone homozygous for *JAK2* or *MPL* (and less frequently *CALR*) mutations, while the acquisition of subclonal mutations may cause leukemic transformation.

PMF is a more heterogeneous disorder in terms of both its clinical and biological features. The last WHO revision[4] split PMF into two entities: prefibrotic PMF (prePMF) and overt PMF. The first one is clinically similar to ET but has different histopathogical features and prognosis. Overt PMF has the worst prognosis and shows the highest risk of progression to AML within MPNs, with some differences among the different molecular groups. In detail, *CALR*-mutant PMF is associated with longer survival compared with other genotypes (*JAK2* V617F *and MPL*-mutant).

On the other hand, triple-negative (TN) PMF is an aggressive disease associated with a short OS due to the high risk of leukemic transformation.

The application of the new criteria led to a better distinction of MPN phenotype, helping in making the correct diagnosis and reflecting the different outcomes of these pathologies. For example, the reclassification according to the new revised criteria of patients previously diagnosed with ET or PMF with grade of fibrosis 0 or 1 or ET, showed significantly different clinical courses and outcomes between prePMF and ET, with the former displaying shorter OS, a higher risk of evolution to overt PMF and a trend towards a higher risk of leukemic evolution, when compared to ET patients[76]. Similarly, in a parallel study it was confirmed that, compared with pre-PMF, overt PMF has more severe clinical features and a worst prognosis characterized by a significantly shortened median survival and a greater propensity to disease progression[77].

In this study, we used targeted sequencing of 81 genes commonly involved in myeloid malignancies to characterize subclonal variants in 509 patients diagnosed with MPNs (ET, prePMF and PMF) and we made a genotype-phenotype study to correlate driver mutations, subclonal variants and clinical data, to better define the diagnostic, prognostic, and potentially therapeutic role of these variants in MPNs.

Overall, by the screening of 236 patients diagnosed with ET, 91 with prePMF and 182 with overt PMF, we detected 598 additional somatic variants.

In this work, we showed that not only clinical and istopathological criteria, but also distinct mutation patterns might differentiate prePMF from overt PMF and ET.

Firstly, compared with prePMF and ET, PMF showed a larger proportion of patients with at least one additional somatic variant (86.3% versus 43% and 35% respectively), a higher average number of somatic variants per patient compared to prePMF and ET and a greater involvement of high molecular risk genes in PMF. These findings support that overt PMF is not a more advanced stage of prePMF, but is a different clinical entity associated with higher tendency in variants accumulation and, consequently, in disease progression, as previously described[77].

Secondly, ET, prePMF and overt PMF showed different mutational landscapes. In ET the most commonly mutated genes belong to the DNA methylation regulation and the most frequently involved gene was *TET2*. In PMF, chromatin structure, RNA splicing and DNA methylation were the most recurrently involved pathway (54%, 33% and 25% of the detected additional somatic respectively) and *ASXL1* and *TET2* were the two most frequently mutated genes. In preMF most of the subclonal variants impaired DNA methylation (21%) and RNA splicing (18%), and *TET2* and *SF3B1* were the most frequently mutated genes.

RNA splicing and DNA methylation are the most recurrently altered pathways in MDS, with mutations in *SF3B1*, *TET2*, and *SRSF2* being identified in 24%, 22%, and 14%, respectively, of MDS patients[78]. Abnormalities in the spliceosome may affect exon utilization or activate alternative splicing sites, creating novel protein isoforms, inducing ectopic expression on inappropriate isoforms or changing gene expression[79]. The consequence of these aberrations is the development of clonal dominance of mutated stem cells. Mutations in genes involved in RNA splicing are usually funding events (VAF=40-50%)[80] and they are mutually exclusive[78]. DNA methylation is a key mechanism of the epigenetic regulation and a hypermethylation profile in the promoter region of tumor suppressor genes, leading to gene silencing, is recurrent in MDS pathogenesis. Thus, PMF and prePMF share some molecular signatures with MDS, reinforcing the hypothesis that myeloid neoplasms are part of a continuum spectrum of diseases[44].

Thirdly, we did not find any association between the driver mutation and additional somatic mutations, except for the mutations in the spliceosome genes (*SF3B1*, SRSF3, U2AF1). We found a significant lower rate of splicing factors alterations in *CALR*-positive patients in comparison with *JAK2* or *MPL* exon 10 mutants. In detail, we did not detect any *SRSF2* mutation and we detect only two *SF3B1* mutations in *CALR* mutants. That supports the finding of mutual exclusion of *CALR* and *SRSF2*, already suggested in the study by Grinfeld et al in 2018[81]. On the other hand, there was a high prevalence of *SRSF2* mutations in *MPL* mutated patients.

Lastly, correlating additional variants with the clinical picture, we confirmed that mutations in *ASXL1*, *SRSF2* and *EZH2* have an impact, both on the blood

parameters and on the progression of the disease, as already highlighted in previous studies[73,82]. More specifically, *ASXL1* is associated with anemia and splenomegaly, a significantly higher risk of leukemic evolution and shorter OS, *SRSF2* has a significant impact on progression and reduced OS, while *EZH2* only shows a negative impact on OS.

Generally, any mutation in genes involved in DNA methylation, RNA splicing, chromatin structure, transcriptional regulation and signaling pathways has a negative impact on OS, but the most relevant role in disease progression is played by mutations in proteins involved in the epigenetic regulation, in the spliceosome and in the signaling pathway.

Conclusion

In this work we suggest that not only clinical criteria, but also distinct mutation patterns might differentiate ET, prePMF and overt PMF. The integration of molecular data and genotype-phenotype taking account of the 2016 WHO classification shall provide fundamental insights on the molecular basis of myeloid neoplasms.

The genomic profiling and the clinical correlations may provide useful tools for better disease definition, more accurate diagnosis, direction of therapy, and refined prognostication. The clinical course of MPNs is profoundly influenced not only by the founding driver mutations but also by subclonal events and the timing of the acquisition plays a role in disease initiation and progression. Next step will be analyzing the co-mutations in our cohort to obtain a better understanding of the mutation hierarchy and its clinical impact.

The current challenge is to develop a prognostic model that accounts for both clinical and molecular parameters, including relevant subclonal mutations. An accurate prognostic tool will be useful also for predicting response to different treatments and to refer accurately selected patients to targeted therapies or to allogeneic stem cell transplantation and thus reach a precision medicine approach in real life.

Acknowledgments

Associazione Italiana per la Ricerca sul Cancro (AIRC, Milan, Italy) through the project "Actionable targets in clonal progression and systemic spreading of myeloid neoplasms”, MYNERVA project (Project Code: 21267) to E. Rumi.

Bibliography

1. Spivak, J.L. (2002). Polycythemia vera: myths, mechanisms, and management. Blood *100*, 4272-4290. 10.1182/blood-2001-12-0349.

2. Dameshek W. Some speculations on the myeloproliferative syndromes [editorial]. Blood. 1951;6(4):372-375. (2016). Blood *127*, 663. 10.1182/blood-2015-12-686402.

3. Bartram, C.R., de Klein, A., Hagemeijer, A., van Agthoven, T., Geurts van Kessel, A., Bootsma, D., Grosveld, G., Ferguson-Smith, M.A., Davies, T., and Stone, M. (1983). Translocation of c-ab1 oncogene correlates with the presence of a Philadelphia chromosome in chronic myelocytic leukaemia. Nature *306*, 277-280. 10.1038/306277a0.

4. Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M., and Vardiman, J.W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood *127*, 2391-2405. 10.1182/blood-2016-03-643544.

5. Barbui, T., Thiele, J., Passamonti, F., Rumi, E., Boveri, E., Ruggeri, M., Rodeghiero, F., d'Amore, E.S., Randi, M.L., Bertozzi, I., et al. (2011). Survival and disease progression in essential thrombocythemia are significantly influenced by accurate morphologic diagnosis: an international study. J Clin Oncol *29*, 3179-3184. 10.1200/JCO.2010.34.5298.

6. Maxson, J.E., Gotlib, J., Pollyea, D.A., Fleischman, A.G., Agarwal, A., Eide, C.A., Bottomly, D., Wilmot, B., McWeeney, S.K., Tognon, C.E., et al. (2013). Oncogenic CSF3R mutations in chronic neutrophilic leukemia and atypical CML. N Engl J Med *368*, 1781-1790. 10.1056/NEJMoa1214514.

7. Nangalia, J., and Green, T.R. (2014). The evolving genomic landscape of myeloproliferative neoplasms. Hematology Am Soc Hematol Educ Program *2014*, 287-296. 10.1182/asheducation-2014.1.287.

8. Moulard, O., Mehta, J., Fryzek, J., Olivares, R., Iqbal, U., and Mesa, R.A. (2014). Epidemiology of myelofibrosis, essential thrombocythemia, and polycythemia vera in the European Union. Eur J Haematol *92*, 289-297. 10.1111/ejh.12256.

9. Tefferi, A., and Elliott, M. (2007). Thrombosis in myeloproliferative disorders: prevalence, prognostic factors, and the role of leukocytes and JAK2V617F. Semin Thromb Hemost *33*, 313-320. 10.1055/s-2007-976165.

10. Cervantes, F., Dupriez, B., Passamonti, F., Vannucchi, A.M., Morra, E., Reilly, J.T., Demory, J.L., Rumi, E., Guglielmelli, P., Roncoroni, E., et al. (2012). Improving survival trends in primary myelofibrosis: an international study. J Clin Oncol *30*, 2981-2987. 10.1200/JCO.2012.42.0240.

11. Tefferi, A., Guglielmelli, P., Larson, D.R., Finke, C., Wassie, E.A., Pieri, L., Gangat, N., Fjerza, R., Belachew, A.A., Lasho, T.L., et al. (2014). Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis. Blood *124*, 2507-2513; quiz 2615. 10.1182/blood-2014-05-579136.

12. Cazzola, M., and Kralovics, R. (2014). From Janus kinase 2 to calreticulin: the clinically relevant genomic landscape of myeloproliferative neoplasms. Blood *123*, 3714-3719. 10.1182/blood-2014-03-530865.

13. James, C., Ugo, V., Le Couédic, J.P., Staerk, J., Delhommeau, F., Lacout, C., Garçon, L., Raslova, H., Berger, R., Bennaceur-Griscelli, A., et al. (2005). A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature *434*, 1144-1148. 10.1038/nature03546.

14. Levine, R.L., Wadleigh, M., Cools, J., Ebert, B.L., Wernig, G., Huntly, B.J., Boggon, T.J., Wlodarska, I., Clark, J.J., Moore, S., et al. (2005). Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell *7*, 387-397. 10.1016/j.ccr.2005.03.023.

15. Kralovics, R., Passamonti, F., Buser, A.S., Teo, S.S., Tiedt, R., Passweg, J.R., Tichelli, A., Cazzola, M., and Skoda, R.C. (2005). A gain-of-function mutation of JAK2 in myeloproliferative disorders. N Engl J Med *352*, 1779-1790. 10.1056/NEJMoa051113.

16. Baxter, E.J., Scott, L.M., Campbell, P.J., East, C., Fourouclas, N., Swanton, S., Vassiliou, G.S., Bench, A.J., Boyd, E.M., Curtin, N., et al. (2005). Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet *365*, 1054-1061. 10.1016/S0140-6736(05)71142-9.

17. Kralovics, R., Guan, Y., and Prchal, J.T. (2002). Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera. Exp Hematol *30*, 229-236. 10.1016/s0301-472x(01)00789-5.

18. Lindauer, K., Loerting, T., Liedl, K.R., and Kroemer, R.T. (2001). Prediction of the structure of human Janus kinase 2 (JAK2) comprising the two carboxy-terminal domains reveals a mechanism for autoregulation. Protein Eng *14*, 27-37. 10.1093/protein/14.1.27.

19. Bandaranayake, R.M., Ungureanu, D., Shan, Y., Shaw, D.E., Silvennoinen, O., and Hubbard, S.R. (2012). Crystal structures of the JAK2 pseudokinase domain and the pathogenic mutant V617F. Nat Struct Mol Biol *19*, 754-759. 10.1038/nsmb.2348.

20. Scott, L.M., Tong, W., Levine, R.L., Scott, M.A., Beer, P.A., Stratton, M.R., Futreal, P.A., Erber, W.N., McMullin, M.F., Harrison, C.N., et al. (2007).

JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. N Engl J Med *356*, 459-468. 10.1056/NEJMoa065202.

21. Passamonti, F., Elena, C., Schnittger, S., Skoda, R.C., Green, A.R., Girodon, F., Kiladjian, J.J., McMullin, M.F., Ruggeri, M., Besses, C., et al. (2011). Molecular and clinical features of the myeloproliferative neoplasm associated with JAK2 exon 12 mutations. Blood *117*, 2813-2816. 10.1182/blood-2010-11-316810.

22. Pikman, Y., Lee, B.H., Mercher, T., McDowell, E., Ebert, B.L., Gozo, M., Cuker, A., Wernig, G., Moore, S., Galinsky, I., et al. (2006). MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. PLoS Med *3*, e270. 10.1371/journal.pmed.0030270.

23. Pardanani, A.D., Levine, R.L., Lasho, T., Pikman, Y., Mesa, R.A., Wadleigh, M., Steensma, D.P., Elliott, M.A., Wolanskyj, A.P., Hogan, W.J., et al. (2006). MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. Blood *108*, 3472-3476. 10.1182/blood-2006-04-018879.

24. Beer, P.A., Campbell, P.J., Scott, L.M., Bench, A.J., Erber, W.N., Bareford, D., Wilkins, B.S., Reilly, J.T., Hasselbalch, H.C., Bowman, R., et al. (2008). MPL mutations in myeloproliferative disorders: analysis of the PT-1 cohort. Blood *112*, 141-149. 10.1182/blood-2008-01-131664.

25. Chaligné, R., Tonetti, C., Besancenot, R., Roy, L., Marty, C., Mossuz, P., Kiladjian, J.J., Socié, G., Bordessoule, D., Le Bousse-Kerdilès, M.C., et al. (2008). New mutations of MPL in primitive myelofibrosis: only the MPL W515 mutations promote a G1/S-phase transition. Leukemia *22*, 1557-1566. 10.1038/leu.2008.137.

26. Malcovati, L., Della Porta, M.G., Pietra, D., Boveri, E., Pellagatti, A., Gallì, A., Travaglino, E., Brisci, A., Rumi, E., Passamonti, F., et al. (2009).

Molecular and clinical features of refractory anemia with ringed sideroblasts associated with marked thrombocytosis. Blood *114*, 3538-3545. 10.1182/blood-2009-05-222331.

27. Ding, J., Komatsu, H., Wakita, A., Kato-Uranishi, M., Ito, M., Satoh, A., Tsuboi, K., Nitta, M., Miyazaki, H., Iida, S., and Ueda, R. (2004). Familial essential thrombocythemia associated with a dominant-positive activating mutation of the c-MPL gene, which encodes for the receptor for thrombopoietin. Blood *103*, 4198-4200. 10.1182/blood-2003-10-3471.

28. Klampfl, T., Gisslinger, H., Harutyunyan, A.S., Nivarthi, H., Rumi, E., Milosevic, J.D., Them, N.C., Berg, T., Gisslinger, B., Pietra, D., et al. (2013). Somatic mutations of calreticulin in myeloproliferative neoplasms. N Engl J Med *369*, 2379-2390. 10.1056/NEJMoa1311347.

29. Nangalia, J., Massie, C.E., Baxter, E.J., Nice, F.L., Gundem, G., Wedge, D.C., Avezov, E., Li, J., Kollmann, K., Kent, D.G., et al. (2013). Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. N Engl J Med *369*, 2391-2405. 10.1056/NEJMoa1312542.

30. Michalak, M., Groenendyk, J., Szabo, E., Gold, L.I., and Opas, M. (2009). Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. Biochem J *417*, 651-666. 10.1042/BJ20081847.

31. Michalak, M., Corbett, E.F., Mesaeli, N., Nakamura, K., and Opas, M. (1999). Calreticulin: one protein, one gene, many functions. Biochem J *344 Pt 2*, 281-292.

32. Gold, L.I., Eggleton, P., Sweetwyne, M.T., Van Duyn, L.B., Greives, M.R., Naylor, S.M., Michalak, M., and Murphy-Ullrich, J.E. (2010). Calreticulin: non-endoplasmic reticulum functions in physiology and disease. FASEB J *24*, 665-683. 10.1096/fj.09-145482.

33. Raghavan, M., Wijeyesakere, S.J., Peters, L.R., and Del Cid, N. (2013). Calreticulin in the immune system: ins and outs. Trends Immunol *34*, 13-21. 10.1016/j.it.2012.08.002.

34. Rampal, R., Al-Shahrour, F., Abdel-Wahab, O., Patel, J.P., Brunel, J.P., Mermel, C.H., Bass, A.J., Pretz, J., Ahn, J., Hricik, T., et al. (2014). Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. Blood *123*, e123-133. 10.1182/blood-2014-02-554634.

35. Marty, C., Pecquet, C., Nivarthi, H., El-Khoury, M., Chachoua, I., Tulliez, M., Villeval, J.L., Raslova, H., Kralovics, R., Constantinescu, S.N., et al. (2016). Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis. Blood *127*, 1317-1324. 10.1182/blood-2015-11-679571.

36. Chachoua, I., Pecquet, C., El-Khoury, M., Nivarthi, H., Albu, R.I., Marty, C., Gryshkova, V., Defour, J.P., Vertenoeil, G., Ngo, A., et al. (2016). Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. Blood *127*, 1325-1335. 10.1182/blood-2015-11-681932.

37. Araki, M., Yang, Y., Masubuchi, N., Hironaka, Y., Takei, H., Morishita, S., Mizukami, Y., Kan, S., Shirane, S., Edahiro, Y., et al. (2016). Activation of the thrombopoietin receptor by mutant calreticulin in CALR-mutant myeloproliferative neoplasms. Blood *127*, 1307-1316. 10.1182/blood-2015-09-671172.

38. Cazzola, M. (2016). Mutant calreticulin: when a chaperone becomes intrusive. Blood *127*, 1219-1221. 10.1182/blood-2016-01-694182.

39. Rumi, E., Pietra, D., Ferretti, V., Klampfl, T., Harutyunyan, A.S., Milosevic, J.D., Them, N.C., Berg, T., Elena, C., Casetti, I.C., et al. (2014). JAK2 or

CALR mutation status defines subtypes of essential thrombocythemia with substantially different clinical course and outcomes. Blood *123*, 1544-1551. 10.1182/blood-2013-11-539098.

40. Pietra, D., Rumi, E., Ferretti, V.V., Di Buduo, C.A., Milanesi, C., Cavalloni, C., Sant'Antonio, E., Abbonante, V., Moccia, F., Casetti, I.C., et al. (2016). Differential clinical effects of different mutation subtypes in CALR-mutant myeloproliferative neoplasms. Leukemia *30*, 431-438. 10.1038/leu.2015.277.

41. Rotunno, G., Mannarelli, C., Guglielmelli, P., Pacilli, A., Pancrazzi, A., Pieri, L., Fanelli, T., Bosi, A., Vannucchi, A.M., and Investigators, A.I.p.I.R.s.C.G.I.M.M. (2014). Impact of calreticulin mutations on clinical and hematological phenotype and outcome in essential thrombocythemia. Blood *123*, 1552-1555. 10.1182/blood-2013-11-538983.

42. Tefferi, A., Lasho, T.L., Finke, C.M., Knudson, R.A., Ketterling, R., Hanson, C.H., Maffioli, M., Caramazza, D., Passamonti, F., and Pardanani, A. (2014). CALR vs JAK2 vs MPL-mutated or triple-negative myelofibrosis: clinical, cytogenetic and molecular comparisons. Leukemia *28*, 1472-1477. 10.1038/leu.2014.3.

43. Rumi, E., Pietra, D., Pascutto, C., Guglielmelli, P., Martínez-Trillos, A., Casetti, I., Colomer, D., Pieri, L., Pratcorona, M., Rotunno, G., et al. (2014). Clinical effect of driver mutations of JAK2, CALR, or MPL in primary myelofibrosis. Blood *124*, 1062-1069. 10.1182/blood-2014-05-578435.

44. Vainchenker, W., and Kralovics, R. (2017). Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. Blood *129*, 667-679. 10.1182/blood-2016-10-695940.

45. Ortmann, C.A., Kent, D.G., Nangalia, J., Silber, Y., Wedge, D.C., Grinfeld, J., Baxter, E.J., Massie, C.E., Papaemmanuil, E., Menon, S., et al. (2015).

Effect of mutation order on myeloproliferative neoplasms. N Engl J Med *372*, 601-612. 10.1056/NEJMoa1412098.

46. Maslah, N., Cassinat, B., Verger, E., Kiladjian, J.J., and Velazquez, L. (2017). The role of LNK/SH2B3 genetic alterations in myeloproliferative neoplasms and other hematological disorders. Leukemia *31*, 1661-1670. 10.1038/leu.2017.139.

47. Lasho, T.L., Pardanani, A., and Tefferi, A. (2010). LNK mutations in JAK2 mutation-negative erythrocytosis. N Engl J Med *363*, 1189-1190. 10.1056/NEJMc1006966.

48. Pardanani, A., Lasho, T., Finke, C., Oh, S.T., Gotlib, J., and Tefferi, A. (2010). LNK mutation studies in blast-phase myeloproliferative neoplasms, and in chronic-phase disease with TET2, IDH, JAK2 or MPL mutations. Leukemia *24*, 1713-1718. 10.1038/leu.2010.163.

49. Rumi, E., Harutyunyan, A.S., Pietra, D., Feenstra, J.D., Cavalloni, C., Roncoroni, E., Casetti, I., Bellini, M., Milanesi, C., Renna, M.C., et al. (2016). LNK mutations in familial myeloproliferative neoplasms. Blood *128*, 144-145. 10.1182/blood-2016-04-711150.

50. Lundberg, P., Karow, A., Nienhold, R., Looser, R., Hao-Shen, H., Nissen, I., Girsberger, S., Lehmann, T., Passweg, J., Stern, M., et al. (2014). Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. Blood *123*, 2220-2228. 10.1182/blood-2013-11-537167.

51. Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2011). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet *44*, 23-31. 10.1038/ng.1009.

52. Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med *371*, 2477-2487. 10.1056/NEJMoa1409405.

53. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med *371*, 2488-2498. 10.1056/NEJMoa1408617.

54. Carbuccia, N., Murati, A., Trouplin, V., Brecqueville, M., Adélaïde, J., Rey, J., Vainchenker, W., Bernard, O.A., Chaffanet, M., Vey, N., et al. (2009). Mutations of ASXL1 gene in myeloproliferative neoplasms. Leukemia *23*, 2183-2186. 10.1038/leu.2009.141.

55. Vannucchi, A.M., Lasho, T.L., Guglielmelli, P., Biamonte, F., Pardanani, A., Pereira, A., Finke, C., Score, J., Gangat, N., Mannarelli, C., et al. (2013). Mutations and prognosis in primary myelofibrosis. Leukemia *27*, 1861-1869. 10.1038/leu.2013.119.

56. Guglielmelli, P., Biamonte, F., Score, J., Hidalgo-Curtis, C., Cervantes, F., Maffioli, M., Fanelli, T., Ernst, T., Winkelman, N., Jones, A.V., et al. (2011). EZH2 mutational status predicts poor survival in myelofibrosis. Blood *118*, 5227-5234. 10.1182/blood-2011-06-363424.

57. Pardanani, A., Lasho, T.L., Finke, C.M., Mai, M., McClure, R.F., and Tefferi, A. (2010). IDH1 and IDH2 mutation analysis in chronic- and blast-phase myeloproliferative neoplasms. Leukemia *24*, 1146-1151. 10.1038/leu.2010.77.

58. Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J.S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., et al. (2011). Somatic SF3B1 mutation in

myelodysplasia with ring sideroblasts. N Engl J Med *365*, 1384-1395. 10.1056/NEJMoa1103283.

59. Zhang, S.J., Rampal, R., Manshouri, T., Patel, J., Mensah, N., Kayserian, A., Hricik, T., Heguy, A., Hedvat, C., Gönen, M., et al. (2012). Genetic analysis of patients with leukemic transformation of myeloproliferative neoplasms shows recurrent SRSF2 mutations that are associated with adverse outcome. Blood *119*, 4480-4485. 10.1182/blood-2011-11-390252.

60. Harutyunyan, A., Klampfl, T., Cazzola, M., and Kralovics, R. (2011). p53 lesions in leukemic transformation. N Engl J Med *364*, 488-490. 10.1056/NEJMc1012718.

61. Rampal, R., Ahn, J., Abdel-Wahab, O., Nahas, M., Wang, K., Lipson, D., Otto, G.A., Yelensky, R., Hricik, T., McKenney, A.S., et al. (2014). Genomic and functional analysis of leukemic transformation of myeloproliferative neoplasms. Proc Natl Acad Sci U S A *111*, E5401-5410. 10.1073/pnas.1407792111.

62. Barosi, G. (2014). Essential thrombocythemia vs. early/prefibrotic myelofibrosis: why does it matter. Best Pract Res Clin Haematol *27*, 129-140. 10.1016/j.beha.2014.07.004.

63. Alvarez-Larrán, A., Ancochea, A., García, M., Climent, F., García-Pallarols, F., Angona, A., Senín, A., Barranco, C., Martínez-Avilés, L., Serrano, S., et al. (2014). WHO-histological criteria for myeloproliferative neoplasms: reproducibility, diagnostic accuracy and correlation with gene mutations and clinical outcomes. Br J Haematol *166*, 911-919. 10.1111/bjh.12990.

64. Barbui, T., Thiele, J., Vannucchi, A.M., and Tefferi, A. (2015). Rationale for revision and proposed changes of the WHO diagnostic criteria for polycythemia vera, essential thrombocythemia and primary myelofibrosis. Blood Cancer J *5*, e337. 10.1038/bcj.2015.64.

65. Passamonti, F., Rumi, E., Pietra, D., Elena, C., Boveri, E., Arcaini, L., Roncoroni, E., Astori, C., Merli, M., Boggi, S., et al. (2010). A prospective study of 338 patients with polycythemia vera: the impact of JAK2 (V617F) allele burden and leukocytosis on fibrotic or leukemic disease transformation and vascular complications. Leukemia *24*, 1574-1579. 10.1038/leu.2010.148.

66. Pietra, D., Brisci, A., Rumi, E., Boggi, S., Elena, C., Pietrelli, A., Bordoni, R., Ferrari, M., Passamonti, F., De Bellis, G., et al. (2011). Deep sequencing reveals double mutations in cis of MPL exon 10 in myeloproliferative neoplasms. Haematologica *96*, 607-611. 10.3324/haematol.2010.034793.

67. Cavalloni, C., Rumi, E., Ferretti, V.V., Pietra, D., Roncoroni, E., Bellini, M., Ciboddo, M., Casetti, I.C., Landini, B., Fugazza, E., et al. (2017). Sequential evaluation of CALR mutant burden in patients with myeloproliferative neoplasms. Oncotarget *8*, 33416-33421. 10.18632/oncotarget.16797.

68. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet *43*, 491-498. 10.1038/ng.806.

69. Fang, H., Bergmann, E.A., Arora, K., Vacic, V., Zody, M.C., Iossifov, I., O'Rawe, J.A., Wu, Y., Jimenez Barron, L.T., Rosenbaum, J., et al. (2016). Indel variant analysis of short-read sequencing data with Scalpel. Nat Protoc *11*, 2529-2548. 10.1038/nprot.2016.150.

70. Malcovati, L., Gallì, A., Travaglino, E., Ambaglio, I., Rizzo, E., Molteni, E., Elena, C., Ferretti, V.V., Catricalà, S., Bono, E., et al. (2017). Clinical significance of somatic mutation in unexplained blood cytopenia. Blood *129*, 3371-3378. 10.1182/blood-2017-01-763425.

71. Hirsch, P., Zhang, Y., Tang, R., Joulin, V., Boutroux, H., Pronier, E., Moatti, H., Flandrin, P., Marzac, C., Bories, D., et al. (2016). Genetic hierarchy and temporal variegation in the clonal history of acute myeloid leukaemia. Nat Commun *7*, 12475. 10.1038/ncomms12475.

72. Milosevic Feenstra, J.D., Nivarthi, H., Gisslinger, H., Leroy, E., Rumi, E., Chachoua, I., Bagienski, K., Kubesova, B., Pietra, D., Gisslinger, B., et al. (2016). Whole-exome sequencing identifies novel MPL and JAK2 mutations in triple-negative myeloproliferative neoplasms. Blood *127*, 325-332. 10.1182/blood-2015-07-661835.

73. Guglielmelli, P., Lasho, T.L., Rotunno, G., Mudireddy, M., Mannarelli, C., Nicolosi, M., Pacilli, A., Pardanani, A., Rumi, E., Rosti, V., et al. (2018). MIPSS70: Mutation-Enhanced International Prognostic Score System for Transplantation-Age Patients With Primary Myelofibrosis. J Clin Oncol *36*, 310-318. 10.1200/JCO.2017.76.4886.

74. Campbell, P.J., Baxter, E.J., Beer, P.A., Scott, L.M., Bench, A.J., Huntly, B.J., Erber, W.N., Kusec, R., Larsen, T.S., Giraudier, S., et al. (2006). Mutation of JAK2 in the myeloproliferative disorders: timing, clonality studies, cytogenetic associations, and role in leukemic transformation. Blood *108*, 3548-3555. 10.1182/blood-2005-12-013748.

75. Nangalia, J., Grinfeld, J., and Green, A.R. (2016). Pathogenesis of Myeloproliferative Disorders. Annu Rev Pathol *11*, 101-126. 10.1146/annurev-pathol-012615-044454.

76. Rumi, E., Boveri, E., Bellini, M., Pietra, D., Ferretti, V.V., Sant'Antonio, E., Cavalloni, C., Casetti, I.C., Roncoroni, E., Ciboddo, M., et al. (2017). Clinical course and outcome of essential thrombocythemia and prefibrotic myelofibrosis according to the revised WHO 2016 diagnostic criteria. Oncotarget *8*, 101735-101744. 10.18632/oncotarget.21594.

77. Guglielmelli, P., Pacilli, A., Rotunno, G., Rumi, E., Rosti, V., Delaini, F., Maffioli, M., Fanelli, T., Pancrazzi, A., Pietra, D., et al. (2017). Presentation and outcome of patients with 2016 WHO diagnosis of prefibrotic and overt primary myelofibrosis. Blood *129*, 3227-3236. 10.1182/blood-2017-01-761999.

78. Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A., et al. (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood *122*, 3616-3627; quiz 3699. 10.1182/blood-2013-08-518886.

79. Lindsley, R.C., and Ebert, B.L. (2013). The biology and clinical impact of genetic lesions in myeloid malignancies. Blood *122*, 3741-3748. 10.1182/blood-2013-06-460295.

80. Cazzola, M., Della Porta, M.G., and Malcovati, L. (2013). The genetic basis of myelodysplasia and its clinical relevance. Blood *122*, 4021-4034. 10.1182/blood-2013-09-381665.

81. Grinfeld, J., Nangalia, J., Baxter, E.J., Wedge, D.C., Angelopoulos, N., Cantrill, R., Godfrey, A.L., Papaemmanuil, E., Gundem, G., MacLean, C., et al. (2018). Classification and Personalized Prognosis in Myeloproliferative Neoplasms. N Engl J Med *379*, 1416-1430. 10.1056/NEJMoa1716614.

82. Tefferi, A., Guglielmelli, P., Nicolosi, M., Mannelli, F., Mudireddy, M., Bartalucci, N., Finke, C.M., Lasho, T.L., Hanson, C.A., Ketterling, R.P., et al. (2018). GIPSS: genetically inspired prognostic scoring system for primary myelofibrosis. Leukemia *32*, 1631-1642. 10.1038/s41375-018-0107-z.