UNIVERSITY OF
MODENA AND REGGIO EMILIA

International Doctorate School
in Clinical and Experimental Medicine
Curriculum: Translational Medicine
XXXIV Cycle

A gene expression-based computational model to improve risk stratification in myelofibrosis

Ph.D. candidate
Sara Castellano

Supervisor
Prof. Enrico Tagliafico

Ph.D. Program Coordinator
Prof. Giuseppe Biagini

Academic year 2021-2022

Abstract

Primary myelofibrosis (PMF), together with polycythemia vera (PV) and essential thrombocythemia (ET), belongs to the group of related hematologic cancers named classic Philadelphia-negative myeloproliferative neoplasms (MPNs). PV and ET can evolve to myelofibrosis giving rise to post-PV (PPV-MF) and post-ET (PET-MF) myelofibrosis, which are both defined as secondary myelofibrosis (SMF). Despite the differences, PMF and SMF patients are currently managed in the same way, and risk stratification is based mainly on clinical features and the presence of driver mutations. None of the existing models for MF (e.g. DIPSS, MIPSS70) integrates transcriptomic data. On the other hand, interest has grown in the last few years concerning the ability of gene expression profiles (GEPs) to provide valuable prognostic information. Several studies demonstrated that GEP can improve risk classification in other hematologic malignancies. Therefore, there is a need to better characterize the transcriptomic profile of myelofibrosis in order to add more robustness to the current scoring systems.

The main scope of this project was to identify a molecular signature and to build a robust classification model able to distinguish “high risk” MF patients with inferior overall survival from “low risk” ones.

We analyzed the gene expression profiles of granulocytes isolated from 114 patients with MF. Cox regression analysis led to the identification of a list of 832 survival-related transcripts characterizing patients who are at high risk for death. Nearest shrunken centroids, subsequent iterations, and k-fold cross-validation were used to build, optimize and validate a classification model, obtaining a final model based on 273 transcripts.

Classification of the 114 samples of our dataset with this model resulted in 54 high-risk and 60 low-risk samples. High-risk patients displayed an inferior

overall survival and leukemia-free survival. In addition, we observed significant enrichment, within the high-risk group, of clinical and molecular detrimental features included in contemporary prognostic models. Strikingly, several patients belonging to the low and intermediate-1 categories of existing prognostic scores were classified as high-risk with our model. These patients were deceased or leukemia transformed earlier than the prognostic class reference median survival. Moreover, our model showed good performance particularly in distinguishing high-risk and low-risk patients within DIPSS and MIPSS70 intermediate categories. It is noteworthy that intermediate-risk classes represent the most challenging patients' categories, for whom determining the optimal therapeutic strategy is more difficult. Additionally, to assess if our model was able to improve the prognostic power of current scoring systems, we designed two new combined models by integrating information from our gene expression-based classification within two existing scores (DIPSS and MIPSS70). The Akaike information criterion (AIC) score was used to compare models for prediction of survival. It turned out that both our new combined models showed better AIC values than DIPSS and MIPSS70 alone.

Overall, these results demonstrate that GEPs in MF patients correlate with their molecular and clinical features, particularly their survival. Thus suggesting that the evaluation of granulocytes' gene expression profiles can improve prognostication, particularly with the identification of MF patients' subgroups characterized by poor prognosis, allowing these patients to be directed towards the most appropriate therapeutic option. These results should be validated in an independent dataset to confirm their predictive power.

Sommario

La mielofibrosi primaria (PMF), la policitemia vera (PV) e la trombocitemia essenziale (ET), appartengono al gruppo di tumori ematologici correlati denominati neoplasie mieloproliferative classiche Philadelphia negative (MPNs). PV e ET possono evolvere in mielofibrosi (MF), dando origine a MF post-PV (PPV-MF) e post-ET (PET-MF), che sono entrambe definite come MF secondarie (SMF). Nonostante le differenze, i pazienti con PMF e SMF sono gestiti nello stesso modo e la stratificazione del rischio si basa perlopiù su caratteristiche cliniche e mutazioni driver. Nessuno dei modelli esistenti per MF (es. DIPSS e MIPSS70) integra dati trascrittomici. D'altra parte, negli ultimi anni l'interesse verso la capacità dei profili di espressione genica (GEP) di fornire informazioni prognostiche è aumentato. Diversi studi hanno dimostrato che i GEP possono migliorare la classificazione del rischio in altre neoplasie ematologiche. Pertanto, si evidenzia la necessità di migliorare la caratterizzazione trascrittomica della MF al fine di aggiungere robustezza ai sistemi di scoring attualmente usati. Lo scopo principale di questo progetto era quello di identificare una firma molecolare e di costruire un modello di classificazione robusto per distinguere pazienti con MF ad alto rischio, con una sopravvivenza globale inferiore, da pazienti a basso rischio. Abbiamo analizzato i GEP di granulociti isolati da 114 pazienti con MF. La regressione di Cox ha portato all'identificazione di 832 trascritti correlati con la sopravvivenza, caratterizzanti i pazienti ad alto rischio di morte. Il metodo "nearest shrunken centroids", iterazioni successive e la convalida incrociata k-fold sono stati utilizzati per costruire, ottimizzare e convalidare un modello di classificazione, ottenendo un modello finale basato su 273 trascritti.

La classificazione dei 114 campioni ha prodotto 54 campioni ad alto rischio (HR) e 60 campioni a basso rischio (LR). I pazienti HR hanno mostrato sopravvivenza globale e libera da leucemia inferiore. È stato osservato un arricchimento, nel gruppo HR, di caratteristiche cliniche e molecolari dannose incluse nei modelli prognostici attuali. Diversi pazienti a rischio basso e intermedio-1 secondo gli score prognostici attuali, ma HR secondo il nostro modello, sono deceduti o hanno subito una trasformazione leucemica in un tempo inferiore rispetto alla sopravvivenza mediana di riferimento. Inoltre, il nostro modello ha mostrato buone prestazioni nel distinguere pazienti HR e LR nelle categorie intermedie di DIPSS e MIPSS70. È interessante notare che le classi di rischio intermedio rappresentano le categorie di pazienti più complesse, per le quali è più difficile determinare la strategia terapeutica ottimale. In aggiunta, per valutare se il nostro modello è in grado di migliorare il potere prognostico dei sistemi di stratificazione del rischio attuali, abbiamo progettato due nuovi modelli combinati, integrando le informazioni provenienti dalla nostra classificazione basata su GEP all'interno di due score esistenti (DIPSS e MIPSS70). Per confrontare i modelli è stato usato il criterio d'informazione di Akaike (AIC). È risultato che entrambi i nuovi modelli hanno mostrato valori di AIC migliori rispetto a DIPSS e MIPSS70 da soli.

Nel complesso questi risultati dimostrano che i GEP nei pazienti con MF correlano con le loro caratteristiche molecolari e cliniche, in particolare con la loro sopravvivenza. Suggerendo, così, che la valutazione dei GEP dei granulociti può migliorare l'inquadramento prognostico attraverso l’identificazione di sottogruppi di pazienti caratterizzati da prognosi sfavorevole, che possono essere indirizzati verso l’opzione terapeutica più appropriata. Questi risultati dovrebbero essere validati in un dataset indipendente per confermarne il potere predittivo.

Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike information criterion
BM	bone marrow
DIPSS	Dynamic International Prognostic Scoring System
EMH	Extramedullary hematopoiesis
ET	essential thrombocythemia
GEP	gene expression profile
HMR	high molecular risk
HR	high-risk
LFS	leukemia-free survival
LR	low-risk
MF	myelofibrosis
MIPSS70	mutation-enhanced international prognostic score system for transplant-age patients
MPNs	myeloproliferative neoplasms
OS	overall survival
PCA	principal component analysis
PMF	primary myelofibrosis
PPV-MF	post-PV myelofibrosis
PV	polycythemia vera
SMF	secondary myelofibrosis

1 Introduction

1.1 Myeloproliferative neoplasms

1.1.1 Definition and classification

Myeloproliferative neoplasms (MPNs) are a related group of rare, yet potentially life-threatening, hematologic cancers characterized by an excessive proliferation of terminally differentiated myeloid cells. The World Health Organization (WHO) classification system for hematopoietic tumors include acute myeloid leukemia (AML) and related neoplasms, myelodysplastic syndromes (MDS), myeloproliferative neoplasms, MDW/MPN overlap, mastocytosis, eosinophilia-associated myeloid/lymphoid neoplasms with recurrent mutations, and myeloid neoplasms with germline predisposition in major categories of myeloid malignancies. Specifically, MPNs are classified as either BCR-ABL1-positive or negative, based on the presence or absence of the BCR-ABL1 fusion gene. This is a frequent acquired somatic mutation that consists in a reciprocal chromosomal translocation – t(9;22) (q34;q11) – causing the fusion of the ABL1 gene from chromosome 9 with the BCR gene on chromosome 22, with consequent production of the BRC-ABL1 fusion protein (NOWELL and HUNGERFORD, 1960). As this fusion gene is located in the so-called Philadelphia (Ph) chromosome, BCR-ABL1-negative MPNs are also referred to as classic Philadelphia-negative (Ph-negative) MPNs (Figure 1) and they include essential thrombocythemia (ET), polycythemia vera (PV), and primary myelofibrosis (PMF). These three disorders are classified as rare cancers with an incidence rate lower than 6 per 1000.000 (but it can vary widely among different countries),

and they usually affect the older population (Rumi and Cazzola, 2017). Specifically, PMF is a condition that is diagnosed without any preceding myeloproliferative neoplasm, but myelofibrosis can also be secondary to an antecedent myeloproliferative neoplasm. Indeed, PV and ET can progress to post-PV (PPV-MF) and post-ET (PET-MF) myelofibrosis, which are both defined as secondary myelofibrosis (SMF).

Instead, the most common disorder belonging to the BCR-ABL1-positive category of MPNs is chronic myelogenous leukemia (CML) (Arber *et al.*, 2016).

To better understand the definition and classification of these pathologies, a brief historical introduction is given in the following paragraph. It underscores the main discoveries made by physicians and scientists of different nationalities since the second half of the nineteenth century.

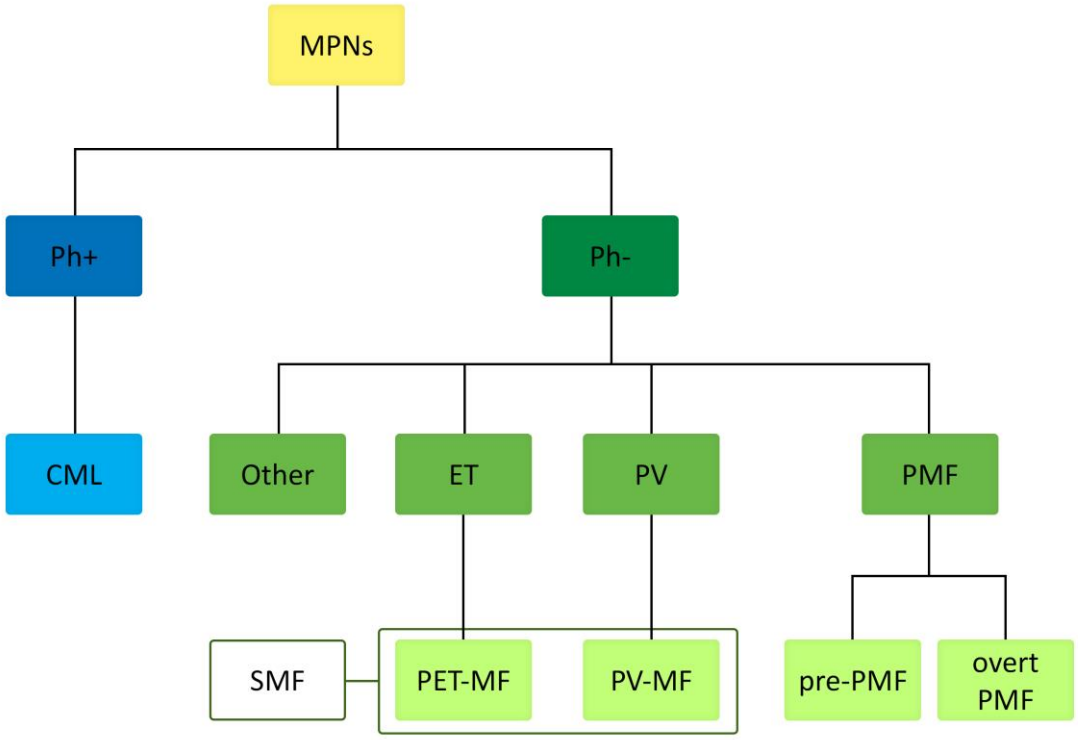

Figure 1. Classic myeloproliferative neoplasms (MPNs) according to 2016 World Health Organization (WHO) classification. Ph+: Philadelphia positive; Ph-: Philadelphia negative; CML: chronic myeloid leukemia; ET: essential thrombocytemia; PV: polycythemia vera; PMF: primary myelofibrosis; pre-PMF: prefibrotic/early stage PMF; overt PMF; overt fibrotic stage PMF; PET-MF: post-ET myelofibrosis; PPV-MF: post-PV myelofibrosis; SMF: secondary

myelofibrosis; Other: Chronic neutrophilic leukemia (CNL), Chronic eosinophilic leukemia not otherwise specified (CEL-NOS), MPN unclassifiable (MPN-U).

1.1.2 Historical prelude

The first description of primary myelofibrosis (PMF) dates back to 1879, when the German surgeon Gustav Heuck highlighted the presence of bone marrow fibrosis, osteosclerosis, and extramedullary hematopoiesis in PMF, thus highlighting the morphological features that allow distinguishing it from CML (Heuck, 1879), which was described in 1845 by the English pathologist John Hughes Bennett (Bennett, 1845). In 1892 a French physician, Louis Henri Vaquez, first described PV (Vaquez, 1892). ET was described for the first time in the following century, in 1934, by two Austrian pathologists, Amil Epstein and Alfred Goedel (Epstein and Goedel, 1934). The term "myeloproliferative disorders (MPD)" was coined by William Dameshek in 1951 to bring out the clinical and morphologic similarities between CML, PV, ET, and PMF (Dameshek, 1951). Philip Fialkow, an American physician-scientist, established the four classic MPN as clonal stem cell diseases after a series of laboratory studies conducted between 1967 and 1981 (Fialkow, Gartler and Yoshida, 1967; Adamson *et al.*, 1976; Jacobson, Salo and Fialkow, 1978; Fialkow *et al.*, 1981). In the second half of the 20th century, several important discoveries in CML were made. In 1960 the Philadelphia chromosome and its association with CML were discovered by two American scientists, Peter Nowell and David Hungerford (Hungerford and Nowell, 1960; NOWELL and HUNGERFORD, 1960). In 1972 the American geneticist Janet Rowley characterized the Ph chromosome as a reciprocal translocation between chromosomes 9 and 22 (ROWLEY, 1973). The BCR-ABL transcript (Shtivelman *et al.*, 1985; Stam *et al.*, 1985) and its fusion protein product (Ben-Neriah *et al.*, 1986) were identified in 1985-6 and retroviral infection of hematopoietic stem cells with this product was shown to induce CML-like disease in mice in 1990 (Daley, Van Etten and Baltimore, 1990). These findings allowed distinguishing CML from other MPD. Furthermore, the understanding of the molecular pathogenesis of classic Ph-negative MPD received a great boost with the discovery of the JAK2 mutation (JAK2V617F) in 2005, which was found to be present in more than 95% of PV patients and about

60% of ET and PMF patients (Baxter *et al.*, 2005; James *et al.*, 2005; Kralovics *et al.*, 2005; Levine *et al.*, 2005). This finding provided evidence of clonality of the proliferative process that characterizes these malignancies and consequently influenced disease classification. Indeed, with the revision of the WHO classification of myeloid neoplasms in 2008, the term "myeloproliferative disorders" was replaced with "myeloproliferative neoplasms", highlighting the neoplastic origin of myeloproliferation (Vardiman *et al.*, 2009). The latest revision of the WHO classification of myeloid neoplasms was released in 2016 to improve disease definition and diagnostic criteria (Arber *et al.*, 2016). Although the WHO category of MPN still includes CML, the term "MPN" is now practically used to refer to PV, ET, and PMF.

1.1.3 Classic Philadelphia-negative myeloproliferative neoplasms

Classic MPNs share a defective blood cell production (hematopoiesis) in the bone marrow, which results in the production of too many or too few blood cells. PV, ET and PMF are included in this category as they share manifold common features; howbeit they are currently described as heterogeneous disorders. As already mentioned at the end of the previous section, clonal hematopoiesis is a feature of MPNs and it is due to the alteration of one or a few hematopoietic stem cells, which exhibit a competitive advantage over normal hematopoietic progenitors. This leads to the creation of a neoplastic clone that is responsible for bone marrow hypercellularity (Spivak, 2003).

The three clinical entities also share the presence of recurrent gene mutations that support clonal hematopoiesis, i.e. driver mutations in JAK, MPL, and CALR genes (Levine *et al.*, 2005; Pikman *et al.*, 2006; Nangalia *et al.*, 2013), which are discussed in more detail in section 1.1.7.

Further elements that characterize all MPNs subtypes are spontaneous transformation into AML, extramedullary hematopoiesis (EMH) in various degrees, development of marrow fibrosis, and elevated risk of thrombotic and hemorrhagic events (Delhommeau *et al.*, 2006). Moreover, PV patients can evolve into ET, and vice versa, PV and ET can progress to post-PV or post-PV

myelofibrosis, and PMF can evolve into PV, so each type of MPN is capable of evolving into another type (Spivak, 2017).
Taken together, these considerations motivated some scientists to regard MPNs as a biological continuum with shades determined by physiological or genetic modifiers (Campbell *et al.*, 2005). Notwithstanding, the different natural history, clinical manifestations, and prognosis that characterize PV, ET, and PMF suggest that they must be considered as three distinct entities (Vannucchi, Guglielmelli and Tefferi, 2009).

1.1.4 Polycythemia Vera

PV involves elevated red-cell counts (erythrocytosis) and it may also be associated with elevated white blood cell counts (leukocytosis), elevated platelet count (thrombocytosis), and/or enlarged spleen (splenomegaly) (Finazzi and Barbui, 2007; Arber *et al.*, 2016). This disease accounts for 45% of the classical MPNs, representing the most common MPN (Rollison *et al.*, 2008). The incidence rate for PV in the European population is 0.68-2.6 per 100.000 per year and it is higher for males than for females. The median age at diagnosis is 65 years (Moulard *et al.*, 2014; Rumi and Cazzola, 2017). The median survival for patients with a PV diagnosis is approximately 15-20 years and 10-20% of them will have their disease progress to MF, while 2-7% will progress to acute myeloid leukemia (AML) (Passamonti, Maffioli, *et al.*, 2011; Tefferi *et al.*, 2013). Advanced age, abnormal karyotype, and leukocytosis represent risk factors for leukemic transformation, while the JAK2V617F allele burden of >50% has been shown to be associated with fibrotic transformation (Vannucchi *et al.*, 2007; Tefferi and Barbui, 2020).
One of the major diagnostic criteria in PV is erythrocytosis, and since it is generally associated with reduced endogenous levels of erythropoietin (EPO) (Rumi and Cazzola, 2017), EPO hypersensitivity of PV erythroid precursors is the main responsible for the increase in red blood cell count. This has been demonstrated by in vitro studies where PV progenitor cells, unlike normal progenitor cells, showed the ability to form erythroid colonies in the absence of exogenous EPO (Prchal and Axelrad, 1974). Nevertheless, neither genetic alterations in the EPO receptor (EPO-R) nor abnormalities in the expression of

the EPO-R and its ability to bind EPO have been found in PV patients, even when EPO serum levels are normal or subnormal (Hess *et al.*, 1994; Green, 1996). This phenomenon can be explained by the presence of the JAK2V617 mutation, which has been observed to induce cytokine hypersensitivity or independence in IL-3 dependent cell lines, mimicking the EPO hypersensitivity or independence observed in PV erythroid progenitors (James *et al.*, 2005). Since about 95% of PV patients harbor one or more JAK2 mutations, they are essential for PV diagnosis, as well as playing a pivotal role in the pathogenesis of PV (Rumi and Cazzola, 2017).

Although many PV patients may be asymptomatic at diagnosis, blood hyperviscosity due to increased red cell mass leads to an increased risk of arterial or venous thrombosis and/or bleeding (Marchioli *et al.*, 2013; Tefferi and Barbui, 2020), particularly in patients with more than 60 years and with thrombosis history (Vannucchi *et al.*, 2007; Tefferi and Barbui, 2020). Patients frequently develop symptoms such as headache, blurred vision, loss of weight, pruritus, mucous membrane bleeding, and erythromelalgia. The disease can progress if hematopoietic stem cells migrate from the marrow to secondary hematopoietic sites (e.g. liver and spleen) leading to hepatomegaly or splenomegaly (Spivak, 2017).

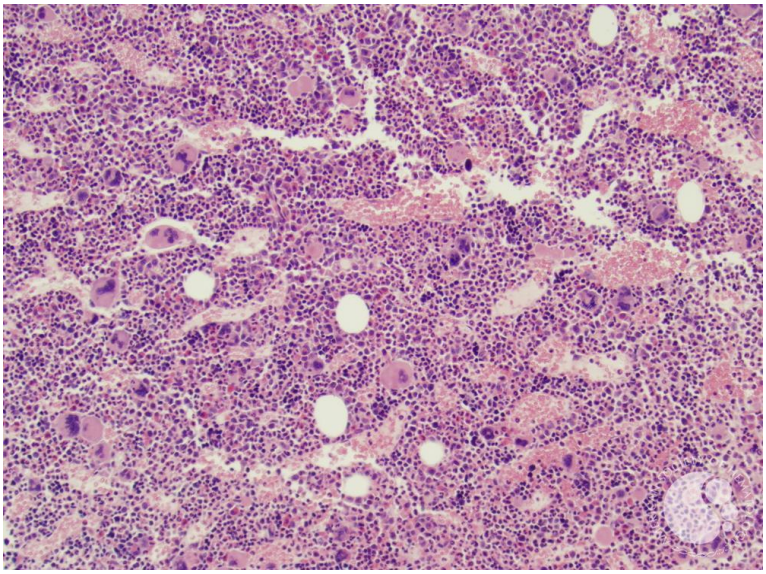

Figure 2. Trephine core biopsy from a PV patient in polycythemic phase. The image shows a hypercellular bone marrow with panmyelosis (proliferation of the erythroid, granulocytic, and megakaryocytic lineages). Megakaryocytes are increased and include frequent hyperlobated

forms. This image was originally published in ASH Image Bank. Elizabeth L. Courville, MD. Polycythemia vera (PV), polycythemic phase, core biopsy 1. ASH Image Bank. 2015; #00060161. © the American Society of Hematology.

1.1.5 Essential Thrombocythemia

ET is defined by increased platelets in the blood (thrombocytosis) and proliferation of enlarged, mature megakaryocytes, which are responsible for platelets production (Tefferi and Barbui, 2015; Arber *et al.*, 2016). The main complications that can occur in patients with ET are thrombosis and bleeding caused by the overproduction of platelets, often associated with abnormalities in platelet morphology and functions (Balduini *et al.*, 1991). ET represents approximately 25% of the classical MPNs (Rollison *et al.*, 2008), it tends to be more common in females than males and it is associated with the most favorable prognosis (median survival ranging between 18 and 20 years) (Wolanskyj *et al.*, 2006; Radaelli *et al.*, 2008). The incidence rate of ET stands between 0.38 and 1.7 per 100.000 per year and the median age at diagnosis is 68 (Iland *et al.*, 1983; Rumi and Cazzola, 2017). One of the major diagnostic criteria for ET Is thrombocytosis. Bone marrow of ET patients shows a normal or even slightly reduced cellularity and enrichment in enlarged, mature megakaryocytes with hyperlobulated nuclei in the marrow cavity (Iland *et al.*, 1983; Thiele *et al.*, 2009; Buhr *et al.*, 2012). ET shares many characteristics with pre-PMF, but the two diseases have different prognoses. For this reason, it is very important to diagnose them correctly. This aspect is explored in more detail in the following section (1.1.6).

Thrombopoietin (TPO) and other thrombotic cytokines have been suggested to have a potential pathogenic role in the increase of megakaryocytes. Nevertheless, TPO serum levels in ET patients are usually normal (Harrison *et al.*, 1999). A possible explanation of this phenomenon came from molecular studies, which demonstrated that JAK2, CALR, and MPL mutations directly or indirectly activate the JAK/STAT signaling pathway downstream to the MPL receptor, making a fundamental contribution to the development of the disease (Rumi and Cazzola, 2017). Indeed, these mutations occur frequently in ET patients: JAK2 in 55%, CALR in 25%, and MPL in 3% of ET patients. Still, almost 17% of ET patients do not harbor any of these mutations (Tefferi and Barbui, 2020).

At diagnosis, most ET patients have no symptoms, but as the disease progress, they may present erythromelalgia, peripheral paresthesia, major thrombosis, headache, blurred vision, and light-headedness (Michiels *et al.*, 2006). These symptoms are related to thrombocytosis and the abnormal interaction between the endothelium and platelets. Major hemorrhagic events may occur, despite the increased number of circulating platelets, principally in the gastrointestinal tract. This is probably due to the depletion and sequestration of the Von Willebrand Factor (VWF) from the plasma (Cortelazzo *et al.*, 1995; Elliott and Tefferi, 2005). Only a small portion of ET patients (20-50%) present with splenomegaly and an even smaller fraction of patients (15-20%) present with hepatomegaly (Finazzi and Harrison, 2005).

Advanced age, leukocytosis, and thrombosis represent risk factors for survival in ET; older age (greater than 60 years), JAK/MPL mutated status, and a previous history of thrombosis are associated with the risk of thrombosis in ET (Barbui *et al.*, 2012); and extreme thrombocytosis (platelets >1000x10^9/L) is a risk factor for bleeding (Tefferi and Barbui, 2020). Moreover, it has been shown that ET patients harboring a JAK2V617F mutation are more likely to progress to PV or MF; CALR-mutant ET patients have a lower risk of thrombosis and a higher risk of fibrotic progression; while mutations in the MPL gene are associated with a higher risk of fibrotic progression (Rumi and Cazzola, 2017; Tefferi and Barbui, 2020).

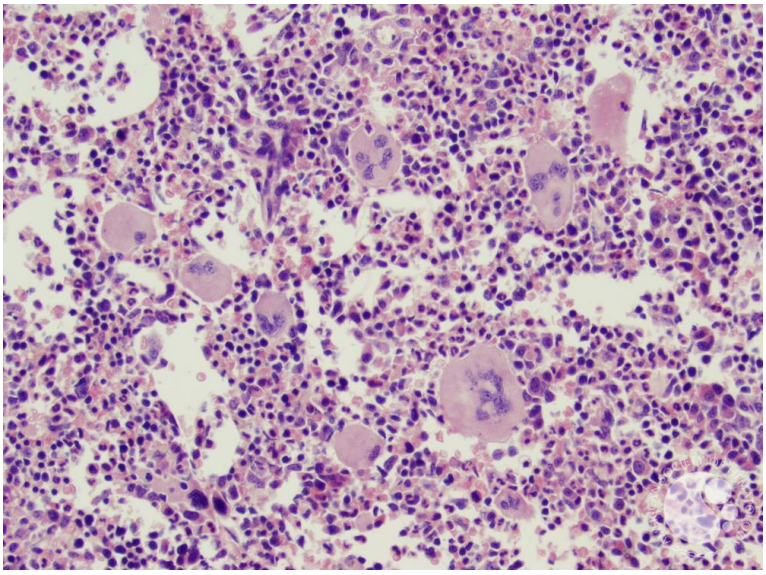

Figure 3. Trephine core biopsy from an ET patient. The image shows a hypercellular bone marrow with increased megakaryocytes. The megakaryocytes are dispersed throughout the marrow and include frequent large forms with abundant cytoplasm and deeply lobated nuclei. This image was originally published in ASH Image Bank. Elizabeth L. Courville, MD. Essential thrombocythemia. ASH Image Bank. 2015; #00060081. © the American Society of Hematology.

1.1.6 Primary Myelofibrosis

While PV and ET are chronic-phase MPNs, myelofibrosis represents the advanced disease that is diagnosed either initially, without pre-existing conditions, as primary MF (PMF), or after the diagnosis of ET or PV as post-ET MF or post-PV MF, respectively (Grinfeld *et al.*, 2018). Indeed, about 15% of ET or PV patients progress to a PMF-like phenotype over time (Tefferi, 2021). Patients with PMF have a worse prognosis and a shorter life expectancy than patients with PV or ET (Geyer and Mesa, 2014), indeed the median survival in PMF is about 6 years (Vannucchi, Guglielmelli and Tefferi, 2009; Ayalew Tefferi, Guglielmelli, Larson, *et al.*, 2014). However, as detailed in section 1.1.10, the prognosis varies depending on the sub-category in which the patient is classified by different risk stratification models that take into account various factors. Approximately 20% of patients with PMF undergo leukemic progression, which is a possible cause of death, but many patients also die of comorbid conditions including cardiovascular events and infection or bleeding as a result of cytopenias (Ayalew Tefferi, Mudireddy, *et al.*, 2018).

PMF is less common than ET and PV, with an incidence rate of 0.1-1 per 100.000 per year, and males are more frequently affected than females. The median age at clinical presentation is 70 (Moulard *et al.*, 2014; Rumi and Cazzola, 2017).
MF is characterized mainly by progressive bone marrow fibrosis or scarring, hepatosplenomegaly and debilitating symptoms (e.g., fatigue, fever, night sweats) (Abdel-Wahab and Levine, 2009; Tonkin *et al.*, 2012), and it can include also severe anemia, cachexia, splenic infarct, bone pain, pruritus, bleeding and thrombosis, as well as portal hypertension and non-hepatosplenic EMH (Tefferi, 2021). Scarring of the bone marrow leads to an impairment of the patient's ability to produce blood cells. PMF involves stem cell-derived clonal myeloproliferation, in particular the megakaryocyte lineage, giving rise to hyperplastic and dysplastic megakaryocytes in the bone marrow accompanied by reactive bone marrow fibrosis, angiogenesis, osteosclerosis, abnormal cytokine production, and EMH (Thiele *et al.*, 1991; Tefferi, 2005). Peripheral blood cell counts can change through the course of the disease, going from a thrombocytotic state to a severe cytopenia. It is frequently to observe an increase in white blood cells (WBCs), erythrocytes, and/or platelets at the beginning, while severe anemia, leukopoenia, and thrombocytopenia can occur at a later stage (Tefferi, 2021). Immature erythroblasts and leukocytes may be present in peripheral blood smears at different percentages, along with teardrop-shaped erythrocytes, morphologically altered platelets, and circulating megakaryocytes. An accelerated phase can be recognized with the presence of 10-19% of immature blasts in the peripheral blood, whereas if the percentage of blasts exceeds 20% it indicates a leukemic transformation (Tefferi, 2021).

According to the revised 2016 WHO classification of myeloid neoplasms PMF is further sub-classified into "prefibrotic" (pre-PMF) and "overtly fibrotic" PMF (overt PMF). The main criteria used to distinguish the two diseases are bone marrow morphology, fibrosis grade, and peripheral blood leukoerythroblastosis. Indeed, BM biopsy of pre-PMF patients shows megakaryocytic proliferation and atypia in addition to increased bone marrow cellularity and granulocytic proliferation. Whereas, in overt PMF reticulin and/or collagen fibrosis are usually combined with megakaryocytic proliferation and atypia (Leiva *et al.*, 2017; Rumi and Cazzola, 2017; Tefferi, 2021).

In a “real-life” study by Guglielmelli et al. pre-PMF and overt PMF were shown to be distinct diseases in terms of patterns of presentation, survival, and disease progression. The authors observed that patients with pre-PMF were generally females of younger age and presented higher leukocyte, hemoglobin, and platelet levels compared to patients with overt PMF. On the other hand, pre-PMF patients less frequently had peripheral blood blasts, symptoms, and extensive splenomegaly. While no differences were shown regarding driver mutations' profile, high molecular risk (HMR) mutations were enriched among overt PMF. Eventually, significantly shorter progression-free survival was observed in patients with overt PMF. Although overall survival and leukemia-free survival were longer in pre-PMF than overt PMF, they were significantly reduced in both categories compared with patients with ET. (Guglielmelli and Vannucchi, 2016; Guglielmelli *et al.*, 2017; Mudireddy *et al.*, 2018).

Even though there has been much debate over the past few years about the existence of pre-PMF as a separate entity and its differentiation from ET (Barbui *et al.*, 2013), significant differences were found in the occurrence of bleeding, rate of death, progression to overt myelofibrosis, and transformation to leukemia in the largest multicenter study (Barbui *et al.*, 2011). Furthermore, megakaryocytes were found to display abnormal maturation in pre-PMF whereas they appear enlarged and mature in ET (Barbui *et al.*, 2013; Arber *et al.*, 2016). Therefore, as both ET and pre-PMF patients frequently present with thrombocytosis (Barosi *et al.*, 2012), BM biopsy represents the most important criterion for distinguishing ET from pre-PMF. Guglielmelli et al. hypothesized that the pre-PMF and overt PMF represent a continuum where disease presentation and progression are affected by unknown individual characteristics and/or germline or somatic gene variants (Guglielmelli *et al.*, 2017).

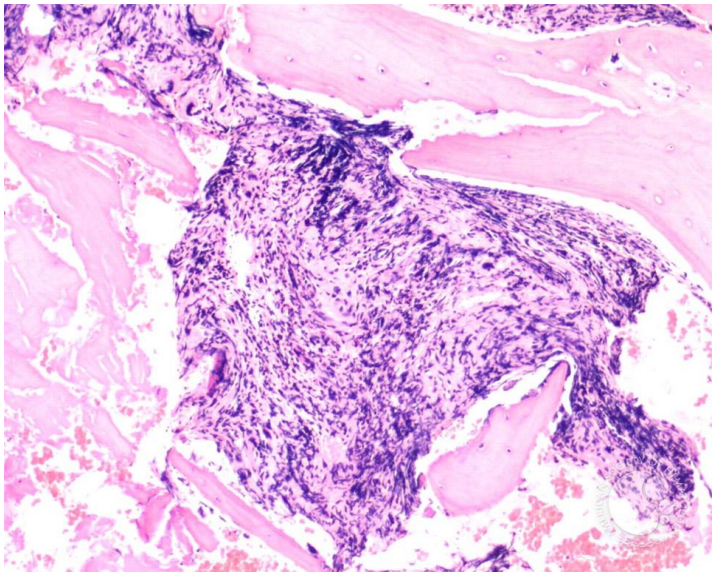

Figure 4. Bone marrow biopsy from a MF patient. The image shows extensive fibrosis with clustered megakaryocytes noted in the center of the specimen. Other hematopoietic elements are not noted. This image was originally published in ASH Image Bank. John Lazarchick. Marrow fibrosis in primary myelofibrosis - 1. ASH Image Bank. 2009; #00004099. © the American Society of Hematology.

1.1.7 Genomics

During the past decades, several genetic aberrations have been identified and characterized in MPN patients, helping to delineate a complex genomic landscape that probably contributes to the great heterogeneity in diagnostic features and outcomes of this group of diseases. Somatic mutations are classified into "driver" and "other" mutations: the former are directly correlated with phenotype and prognosis of MPNs, while the latter might contribute to disease progression and leukemic transformation (Tefferi, 2021). The first driver mutation to be discovered, in 2005, is defined as JAK2V617F as it is located in the JAK2 (Janus Kinase 2) gene and causes the release of inhibition and hyperactivation of JAK-STAT signaling (Baxter *et al.*, 2005; James *et al.*, 2005; Kralovics *et al.*, 2005; Levine *et al.*, 2005). This is the most frequent mutation, occurring in almost all PV patients (over 90%) (James *et al.*, 2006; Tefferi, Lasho, *et al.*, 2006) and 50-60% of patients with ET (Antonioli *et al.*, 2005; Campbell *et al.*, 2005; Wolanskyj *et al.*, 2005; Kittur *et al.*, 2007) or PMF (Tefferi, 2016; Gangat and Tefferi, 2020). The following year, in 2006, a somatic mutation in the MPL gene was described

in JAK2V617F-negative PMF patients. The MPLW515L mutation is absent in PV, while it is found in 3-4% of ET and 6-7% of PMF patients (Pikman *et al.*, 2006). In 2007, other JAK2 mutations affecting exon 12 were identified in PV patients without JAK2V617F (Scott *et al.*, 2007). Most recently, in 2013, another driver mutation has been described in CALR. Mutated CALR results in a multi-functional Ca^{2++} binding protein chaperon with a missing KDEL (lysine, aspartic acid, glutamic acid, and leucine) ER (endoplasmic reticulum) retention motif in the C-terminal. In the same way as MPL, the CALR mutation is absent in PV and is expressed by 20-25% of patients with ET and PMF (Klampfl *et al.*, 2013; Nangalia *et al.*, 2013).

The main contribution of JAK2 and MPL mutations in MPN pathogenesis is probably the direct activation of the JAK-STAT pathway, while mutations in CALR might indirectly activate the same pathway, but they could also influence platelet production as suggested by mouse models (Marty *et al.*, 2014).

Driver mutations have been extensively studied and correlations with phenotype in MPNs have been identified. In general, mutant JAK2 is more frequent in elderly patients and is associated with leukocytosis, lower platelet levels, and increased risk of thrombosis in ET.

Association with phenotype and CALR mutations is different in ET and PMF, but younger age and higher platelet count are common. Additionally, male sex and lower leukocyte count are associated with mutant CALR in ET, while in PMF there is an association with leukocytosis and lower frequencies of anemia (Barbui *et al.*, 2012; A Tefferi, Thiele, *et al.*, 2014). Mutations in CALR are further classified into type 1 and type 2: the former concerns a 52-bp deletion, while the latter is a 5-bp insertion. About 80% of patients with mutant CALR harbor one of two variant types, the remaining mutations are structurally similar to type 1 and type 2 variants, consequently, they are operationally classified into “type 1-like” and “type 2-like” variants (A Tefferi, Lasho, *et al.*, 2014; Ayalew Tefferi, Lasho, Tischer, *et al.*, 2014; Ayalew Tefferi, Wassie, Guglielmelli, *et al.*, 2014). The order of mutations acquisition was suggested to be a further determinant of the phenotype in MPN (Ortmann *et al.*, 2015).

Patients with ET or PMF who do not show any one of the three driver mutations represent about 10-15% of the total and are defined as “triple-negative” (Ayalew Tefferi, Guglielmelli, Larson, *et al.*, 2014). These patients often harbor mutations

in other genes, which are relevant to epigenetic (e.g., ASXL1, DNMT3A, EZH2, IDH1, IDH2, TET2), RNA splicing (e.g., SF3B1, SRSF2, U2AF1), or transcriptional regulation (e.g., CUX1, IKZF1, NF-E2, TP53) (Vannucchi *et al.*, 2013; A Tefferi, Finke, *et al.*, 2014). Although the pathogenic contribution of these other mutations is not yet fully understood, it would appear that their cooperation with the driver mutations might facilitate disease progression (Rampal, Ahn, *et al.*, 2014; Chen *et al.*, 2015). Furthermore, diagnostic features and outcomes in patients with MPN show strong heterogeneity, probably as a consequence of this complex genetic landscape (Grinfeld *et al.*, 2018).

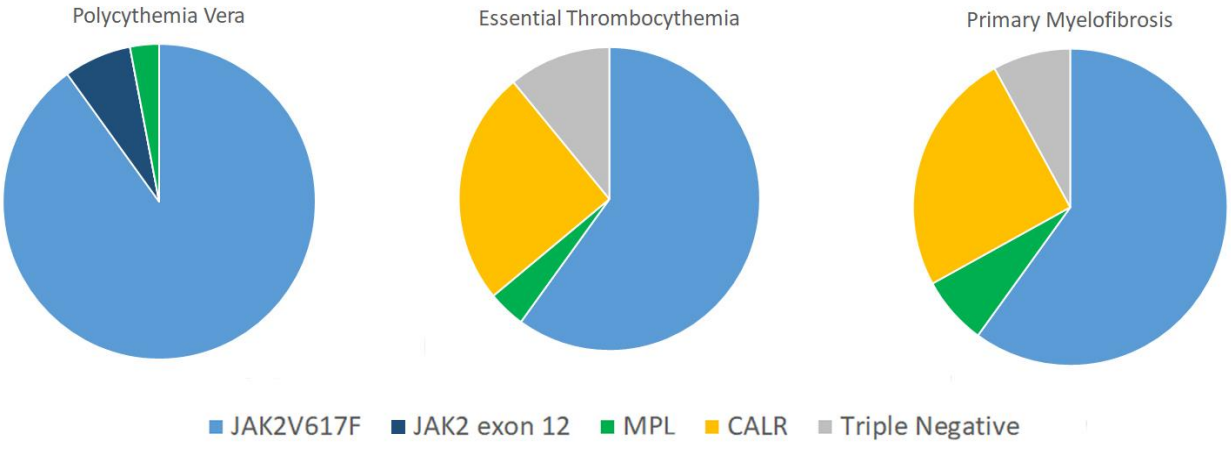

Figure 5. Frequency of JAK2, CALR, and MPL mutations in PV, ET, and PMF. Adapted from Nangalia and Green, 2014

1.1.8 Myelofibrosis biology

Among the three diseases included in the classic Philadelphia negative MPNs, myelofibrosis is the one associated with the worst prognosis. It is defined either as primary MF (PMF) if known underlying malignant processes are absent or as secondary MF (SMF) if it arises from the progression of ET or PV. Fibrosis is a hallmark of the disease, but its genesis remains elusive. Nevertheless, different putative mechanisms have been hypothesized to be implicated in the evolution of fibrosis. They are discussed hereafter.

Murine models and *in vitro* studies allowed discovering the fundamental role of abnormal megakaryocytes in myelofibrosis. CD34+ hematopoietic progenitor cells isolated from MF patients showed an increased propensity to generate megakaryocytes and impaired apoptosis due to overexpression of anti-apoptotic proteins, which together promote megakaryocytic hyperplasia (Wang *et al.*, 2002; Ciurea *et al.*, 2007). Resulting aberrant immature megakaryocytes secrete a plethora of pro-inflammatory cytokines and growth factors, causing fibrosis. In particular, an increased expression of osteocalcin, TGF-beta, platelet-derived growth factor (PDGF), and vascular endothelial growth factor (VEGF) was hypothesized as responsible for the genesis of fibrosis (Vannucchi *et al.*, 2002). Furthermore, it has been observed an impact of aberrant megakaryocytes also on hematopoietic stem cells (HSC), with which they reside in close proximity. Production of the chemokine CXCL-4 by megakaryocytes influences HCS cell cycle activity (Bruns *et al.*, 2014).

Another peculiar feature shared by cancer cells, which certainly plays a role in contributing to MF pathogenesis, is the ability to activate proliferative signaling pathways to sustain chronic growth. Specifically, the JAK/STAT pathway has been shown to be frequently altered in MF, with over 90% of PV patients and 60% of PMF and ET patients harboring the JAK2V617F mutation, as already detailed in section 1.1.7. This somatic gain of function mutation causes a release of inhibition and a consequent hyperactivation of JAK/STAT signaling (Baxter *et al.*, 2005; James *et al.*, 2005; Kralovics *et al.*, 2005; Levine *et al.*, 2005). However, the contribution of the constitutive activation of JAK/STAT to the pathogenesis of MF has been widely acknowledged and confirmed by gene expression profiling studies even in the absence of the JAK2V617F mutation (Rampal, Al-Shahrour, *et al.*, 2014). JAK2V617F-negative patients with PV were found to harbor JAK2 exon 12 mutations, which may occur also in post-PV MF (Scott *et al.*, 2007). These patients had similar thrombotic complications and fibrotic/leukemic transformation rates compared to JAK2V617F-positive PV patients, even if they showed differences in other clinical features such as age, hemoglobin, white count, and platelet count (Passamonti, Elena, *et al.*, 2011; Ayalew Tefferi, Lavu, *et al.*, 2018).

Other mutations that indirectly affect the activation of the JAK/STAT pathway are MPLW515K/L mutations, which has been identified in 7% of PMF patients (Pikman *et al.*, 2006); and CALR mutations, discovered in 2013 in 23% of PMF patients with JAK2 or MPL wild type (Klampfl *et al.*, 2013; Nangalia *et al.*, 2013). MPL encodes the thrombopoietin (TPO) receptor, which plays an important role in the expansion and regulation of megakaryocytes and self-renewal of hematopoietic stem cells (Kaushansky *et al.*, 1994). CALR encodes a chaperone of the endoplasmic reticulum which, when mutated, lacks the KDEL retrieval sequence and the calcium-binding sites. It binds to the thrombopoietin (TPO) receptor, activating it and causing the phosphorylation of JAK2 with consequent constitutive activation of JAK/STAT signaling (Araki *et al.*, 2016; Chachoua *et al.*, 2016; Elf *et al.*, 2016; Marty *et al.*, 2016; How, Hobbs and Mullally, 2019). In addition, other downstream signaling pathways, such as MAPK and PI3K/AKT, are activated (Khan *et al.*, 2013), and the significant excess of pro-inflammatory cytokines leads to NF-kB pathway hyperactivation (Fisher *et al.*, 2017).

Other biological processes, besides abnormal JAK/STAT signaling, are involved in the induction and promotion of myelofibrosis. Indeed, it is well known that gene expression can be influenced also by DNA methylation, histone modifications, and micro RNAs (miRNA). A targeted sequencing study revealed mutations in epigenetic regulators in 81% of patients with PMF (Tefferi *et al.*, 2016). Among these non-driver mutations, the most frequent are additional sex comb-like 1 (ASXL1) mutations, which cause repression of known leukemogenic target genes by losing polycomb repressive complex 2 (PRC2)-mediated histone H3 lysine 27 (H3K27) tri-methylation (Abdel-Wahab *et al.*, 2012). Different types of mutation can occur in ASXL1 (frameshift, non-sense, or missense) but they all show an equipotent adverse prognostic impact in PMF (A Tefferi, Guglielmelli, *et al.*, 2014; A Tefferi *et al.*, 2018). A CRISPR/Cas9 approach demonstrated that DNA methyltransferase 3 (DNMT3A) mutations ease disease progression by loss of activation of enhancers and aberrant inflammatory signaling (Jacquelin *et al.*, 2018); nevertheless, mutations in this gene are found only in 7% of patients in PMF (Abdel-Wahab *et al.*, 2011). Another infrequent mutated (5-9%) gene in PMF is enhancer of zeste homolog 2 (EZH2), a histone methyltransferase that catalyzes histone H3 methylation at lysine 27 (H3K27) repressing transcription

(Guglielmelli, Biamonte, *et al.*, 2011). EZH2 has been shown to have a tumor suppressor function in PMF, as its deletion in JAK2V617F murine models resulted in a decreased erythropoiesis and an increased megakaryopoiesis (Shimizu *et al.*, 2016; Yang *et al.*, 2016). SRSF2 and U2AF1 are splicing factors that are frequently mutated in PMF patients (18% and 16% respectively) and although the precise mechanisms by which mutations in these genes contribute to the pathogenesis of PMF are still unclear, they are known to be involved in altered pre-mRNA splicing (Lasho *et al.*, 2012; Ayalew Tefferi, Finke, Lasho, *et al.*, 2018). IDH mutations are rare in PMF (4%) but cooperation with JAK2V617F in the transformation to blast phase was hypothesized (Tefferi *et al.*, 2012). The aberrant miRNA expression in MF has been underlined by several studies and specific miRNA signatures have been found to distinguish between MF granulocytes and those of healthy individuals (Guglielmelli *et al.*, 2007; Hussein *et al.*, 2009; Paola Guglielmelli *et al.*, 2015).

Another important element involved in MF pathogenesis is the bone marrow microenvironment. It is acknowledged that the osteoblastic and vascular niches together compose the stem cell niche, where HSCs reside. A disrupted interaction within these connected niches, which are physiologically characterized by continuous cross-talk, is frequently observed in MF, affecting the regulation of HSC self-renewal and differentiation (Calvi *et al.*, 2003; Yin and Li, 2006; Varricchio, Mancini and Migliaccio, 2009). Indeed, it has been observed that secretion of IL-1 beta by JAK2 mutant HSCs favors their own expansion over normal HSCs by inducing apoptosis of mesenchymal cells, which are a unique element of the HSC niche that support the survival of normal HSC. Malignant HSC proliferation is further promoted by JAK2 mutated endothelial cells. In addition, mesenchymal cells are activated by fibrotic cytokines released by aberrant mutant megakaryocytes, causing fibrosis (Bock *et al.*, 2008; Lin, Kaushansky and Zhan, 2016). Disruption of the bone marrow leads to the development of anemia and HSCs mobilization with the consequent onset of EMH, primary in the spleen and liver. In turn, EMH causes the development of splenomegaly and hepatomegaly, which are involved in the occurrence of constitutional symptoms in PMF patients.

The overall picture explaining the biology of myelofibrosis remains, therefore, puzzling. Multiple elements seem to be involved and different mechanisms have been hypothesized, with contributions from aberrant megakaryocytes, constitutive activation of JAK/STAT and other pathways, epigenetic dysregulation, bone marrow microenvironment, and abnormal pro-inflammatory cytokine production.

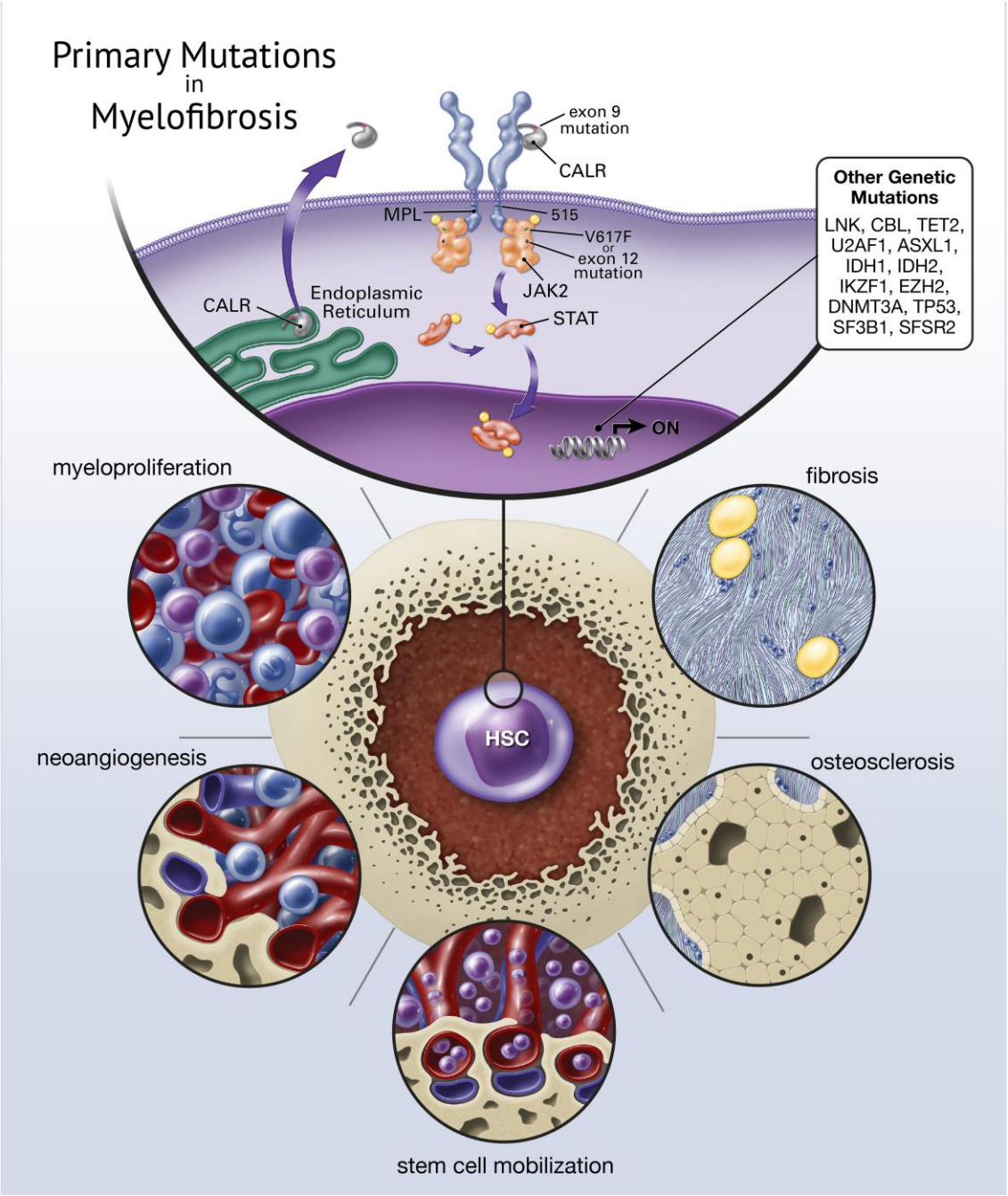

Figure 6. Primary mutations in myelofibrosis. In this figure important mutations and their relationship with the crucial JAK-STAT signaling pathway in MF are highlighted. Reproduced with permission, Garmezy *et al.*, 2021

1.1.9 Therapeutic options

Currently available drug therapies for MPNs are not curative or disease-modifying, indeed they only are of palliative benefit (Tefferi and Pardanani, 2015). The main goals of such treatments are to prevent thrombohemorrhagic

complications and alleviate symptoms. The main treatments for PV are phlebotomy and aspirin to keep hematocrit below 45%; furthermore, cytoreductive therapy with hydroxyurea (or interferon-α and busulfan as second-line drugs) is required in high-risk PV patients. Ruxolitinib (which is covered in more detail further down in this section) or other JAK2 inhibitors are recommended for use only in case of marked splenomegaly that is not responding to the previously cited drugs or severe and protracted pruritus (Spivak, 2017; Tefferi and Barbui, 2020).

Since ET is a more indolent disease, observation alone is recommended for very-low risk patients, who are often asymptomatic, while low-risk patients may require at least once-daily aspirin therapy, and only for high-risk ET patients, cytoreductive therapy is recommended (Tefferi and Barbui, 2020).

PMF is associated with the worst prognosis and with the greatest symptom burden. Treatment of these patients might include observation, participation in investigational drug trials, or conventional drug therapy. The latter is exploited to treat symptoms such as anemia, splenomegaly, non-hepatosplenic EMH, EMH-associated pulmonary hypertension, bone pain, or constitutional symptoms. The main indications for treatment in PMF are anemia and symptomatic splenomegaly. Prednisone, androgens (e.g. danazol), thalidomide, or lenalidomide are used to treat anemia (Cervantes, Mesa and Barosi, 2007); lenalidomide acts also on thrombocytopenia and splenomegaly, even if it is associated with cytopenias (Tefferi, Cortes, *et al.*, 2006). Symptomatic splenomegaly in PMF is treated preferentially with hydroxyurea, while patients who received frequent red blood cell transfusions or patients with splenomegaly refractory to hydroxyurea are often managed by splenectomy (Mishchenko and Tefferi, 2010). Additionally, in the presence of extreme leukocytosis or thrombocytosis, cytoreductive therapy might be needed.

The first target therapy for the treatment of PMF approved by the FDA (Food and Drug Administration) is ruxolitinib (Mascarenhas and Hoffman, 2012). This belongs to the type I kinase inhibitor class which binds and stabilizes the active conformation of the kinase, specifically it is an adenosine triphosphate-competitive JAK1/JAK2 inhibitor. The efficacy of ruxolitinib has been compared to either placebo or best available therapy in two randomized phase III clinical trials: COMFORT-1 and COMFORT-2 trial, respectively (Harrison *et al.*, 2012;

Verstovsek *et al.*, 2012). Ruxolitinib treatment showed to reduce constitutional symptoms, splenomegaly, and serum levels of inflammatory cytokines. Still, important hematological toxicity is related to this therapy, leading to anemia, thrombocytopenia, and immunosuppression (Harrison *et al.*, 2012; Verstovsek *et al.*, 2012; Deininger *et al.*, 2015). Furthermore, no substantial drug effect on JAK2V617F allele burden or bone marrow fibrosis has been observed, suggesting that Ruxolitinib does not perform a selective inhibition of the malignant clone. However, a slight but significant improvement in survival has been reported (Cervantes *et al.*, 2013). To improve the treatment of MPNs, increase the therapeutic efficacy, and reduce side effects several clinical trials focused on testing combinatorial protocols of ruxolitinib and other drugs (Mascarenhas, 2014). These experimental drugs include:

(i) other JAK inhibitors (e.g. fedratinib, pacratinib, momelotinib);
(ii) drugs acting on the PI3K/Akt/mTOR pathway, which is indirectly activated by JAK/STAT (e.g. everolimus);
(iii) telomerase inhibitors (e.g. imetelstat).

Nevertheless, many JAK inhibitors were halted in the development due to severe adverse effects reported during clinical trials. Another promising drug, currently under investigation, is Pelabresib (CPI-0610), a BET (bromodomain and extraterminal domain) proteins inhibitor. Preliminary data showed that this therapy demonstrated signals of clinical activity in MF patients by modifying the expression of genes involved in NFkB signaling (Kremyanskaya *et al.*, 2021).

To date, the only curative approach for PMF is allogenic hematopoietic stem cell transplantation (HSCT) (Kröger *et al.*, 2015). However, it carries a substantial mortality risk and it is associated with a considerable risk of life-threatening infections and graft *versus* host disease (McLornan *et al.*, 2019). Other limiting factors for HSCT include advanced age, medical comorbidities, and donor availability. For these reasons, HSCT is generally recommended for younger patients with few comorbidities and who are classified as high risk by current risk stratification strategies. Given the median age at diagnosis of 65 years (Szuber *et al.*, 2019), it is evident that HSCT is feasible only in a small subset of patients. Accordingly, a risk-adapted therapeutic approach is the best strategy to follow (Figure 7). Investigational drug therapy is recommended in high-risk cases where HSCT is not feasible. While in low-risk diseases with absence of symptoms

observation alone is suggested (Tefferi, 2016; Gangat and Tefferi, 2020). In this context, the extreme importance of risk stratification methods comes to light. The description of currently available prognostic scores for MPN is detailed in the following paragraph.

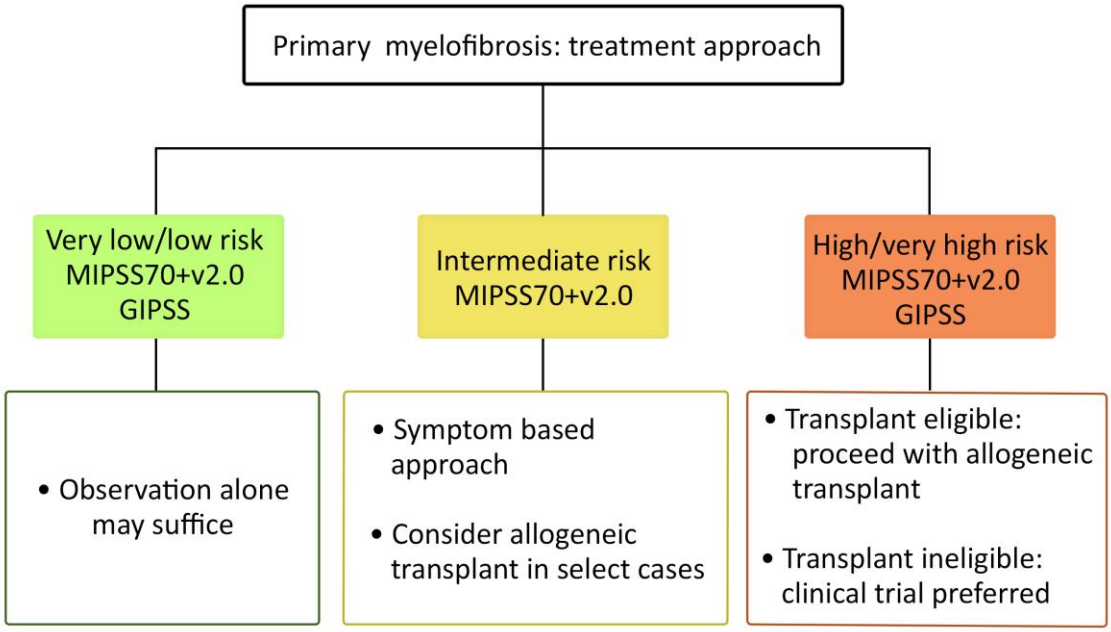

Figure 7. Primary myelofibrosis: risk-adapted treatment approach. MIPSS70+v2.0: Mutation Enhanced Prognostic Scoring System; GIPSS: Genetically Inspired Prognostic Scoring System. Adapted from Gangat and Tefferi, 2020

1.1.10 Current risk stratification strategies

Advances in next-generation sequencing allowed the identification and characterization of several genomic aberrations in MPN patients, thus leading to a considerable evolution of prognostic assessment tools for MPNs. The first prognostic score to be designed in 2009 when only clinical and laboratory features were available, is the International Prognostic Scoring System (IPSS); the factors included in this score are obtained at the time of diagnosis and they are age, hemoglobin, leukocyte count, peripheral blast percentage, and constitutional symptoms (Cervantes *et al.*, 2009). The IPSS was subsequently improved, in 2010, in order to assess prognosis also through the course of the disease, by developing the Dynamic International Prognostic Scoring System

(DIPSS). It is based on the same factors, but these are obtained at any time point, and anemia is given greater weight (Passamonti *et al.*, 2010). Further refinements led, in 2011, to the DIPSS-plus score, which includes also red cell transfusion need, degree of thrombocytopenia less than 100 x 10^9/L, and unfavorable karyotype. Risk categories generated by this score are low, intermediate 1, intermediate 2, and high, with respective median survivals of 15.4, 6.5, 2.9, and 1.3 years (Gangat *et al.*, 2011). The main limitation of these prognostic models lies in the fluctuations to which laboratory parameters are subject in transfused patients, in particular hemoglobin values. This was overcome by developing the contemporary prognostic scores, which include genetics in addition to laboratory parameters. They are defined as Mutation Enhanced Prognostic Scoring Systems and comprise MIPSS70, MIPSS70-plus, and MIPSS70-plus version 2.0 (Ayalew Tefferi, Guglielmelli, Lasho, *et al.*, 2018; Guglielmelli, Terra L. Lasho, *et al.*, 2018). The main novelty brought by these models concerns high molecular risk (HMR) mutations, which encompass variants in ASXL1, EZH2, SRSF2, and IDH1, while U2AF1Q157 was added only in MIPSS70-plus version 2.0. Furthermore, in this version of the score improvements have been made regarding anemia and karyotype:

(i) anemia was adjusted for sex, thus developing different categories for severe anemia and moderate anemia (Nicolosi *et al.*, 2018);

(ii) in light of a study on cytogenetic abnormalities in over 1,000 PMF patients, revised cytogenetic risk categories were defined as very high-risk karyotype (VHR), unfavorable karyotype, and favorable karyotype (Ayalew Tefferi, Nicolosi, Mudireddy, *et al.*, 2018).

In 2018, a simpler scoring system was designed using only genetic markers: the genetically-inspired prognostic scoring system (GIPSS). This is based on VHR, unfavorable karyotype, ASXL1, SRSF2 U2AF1Q157, and the absence of Type 1 CALR mutations.

Prognostic score	Prognostic variables	Median survival					
		Very low	Low	Int-1	Int-2	High	Very high
IPSS	- Age > 65 years - constitutional symptoms - hemoglobin < 10g/dl - leukocyte count > 25 x 10(9)/l - circulating blasts > 1% at diagnosis		135 months	95 months	48 months	27 months	
DIPSS	- Age > 65 years - constitutional symptoms - hemoglobin < 10g/dl - leukocyte count > 25 x 10(9)/l - circulating blasts > 1% at anytime		14.6 years	7.4 years	4 years	2.3 years	
DIPSS-plus	- Age > 65 years - constitutional symptoms - hemoglobin < 10g/dl - leukocyte count > 25 x 10(9)/l - circulating blasts > 1% - unfavorable karyotype - platelet count < 100 x 10(9)/l - transfusion needs at anytime		185 months	78 months	35 months	16 months	
MIPSS70	- Hemoglobin < 10g/dl - leukocyte count > 25 x 10(9)/l - platelet count < 100 x 10(9)/l - circulating blasts > 2% - bone marrow fibrosis grade ≥ 2 - constitutional symptoms - absence of CALR type-1 mutation - presence of HMR mutation (ie, ASXL1, EZH2, SRSF2, IDH1/2) - presence of two or more HMR mutations at anytime		27.7 years	7.1 years		2.3 years	
MIPSS70-plus	Same as MIPSS70 plus - unfavorable karyotype at anytime		20 years	6.3 years		3.9 years	1.7 years

MIPSS70-plus Version 2.0	- Severe anemia - moderate anemia - circulating blasts ≥ 2% - constitutional symptoms - very high risk karyotype / unfavorable karyotype - ≥2 HMR mutations / One HMR mutation / Type 1-like CALR absent - additional HMR: U2AF1Q157 at anytime	Not reached	16.4 years	7.7 years		4.1 years	1.8 years
GIPSS	- Very high risk karyotype - unfavorable karyotype - ASXL1 - SRSF2 - U2AF1Q157 - absence of type 1-like CALR at anytime	26.4 years	8 years			4.2 years	2 years

Table 1. Contemporary prognostic scoring systems for primary myelofibrosis. Int-1: intermediate-1; Int-2: intermediate-2; IPSS: International Prognostic Scoring System; DIPSS: Dynamic International Prognostic Scoring System; DIPSS-Plus: Dynamic International Prognostic Scoring System plus thrombocytopenia, karyotype and transfusion needs; MIPSS70: Mutation-enhanced International Prognostic Scoring System for transplant-age patients (age ≤ 70 years); GIPSS: Genetically Inspired Prognostic Scoring System for all age groups; HMR: high molecular risk mutations. Adapted from Gangat and Tefferi, 2020.

In addition to these prognostic models, whose main purpose is to select PMF patients for allogeneic transplantation, a clinical-molecular system was designed in 2019 to assess prognosis after transplantation: the myelofibrosis transplant scoring system (MTSS) (Gagelmann *et al.*, 2019). Overall, primary and secondary myelofibrosis patients are managed in the same way, as they share common histopathologic features and clinical manifestations, and the same prognostic models are used to predict survival. However, it has been demonstrated that these prognostic models are less accurate in distinguishing different risk categories in SMF (Hernández-Boluda *et al.*, 2014; Masarova *et al.*, 2017). For this reason, the Myelofibrosis Secondary to polycythemia vera and essential thrombocythemia-Prognostic Model (MYSEC-PM) was developed in a collaborative project conducted in post-ET/PV MF (Passamonti *et al.*, 2017). The observation that only ASCT is potentially curative, but its use is justified only in high-risk patients due

to the incidence of mortality and severe adverse events, highlights the need for more specific and more accurate prognostic models in PMF and SMF. In this context, further improvements in risk stratification for MF may come from gene expression profiling (GEP). Indeed several studies showed that GEP can provide valuable prognostic information, improving risk classification in hematologic malignancies such as AML (Bullinger *et al.*, 2004; Ng *et al.*, 2016) and MDS (Mills *et al.*, 2009; Pellagatti *et al.*, 2013; Shiozawa *et al.*, 2017).

1.2 Machine learning applied to cancer

Machine learning (ML) is a subdomain of artificial intelligence (AI), which provides systems the ability to automatically learn and improve from experience, by finding patterns in large datasets. In this context, computer programs are not explicitly programmed, but they are developed to access data and to use it to learn for themselves. Examples, direct experience, or instruction are used as prior training to identify new patterns, classify new data, or predict novel trends (Mitchell, 1997). ML is more powerful than statistics and probability, even if it still draws strongly from them because it allows making decisions or inferences that conventional statistical methods could not allow (Mitchell, 1997).

ML methods are commonly divided into two main types: (i) supervised learning and (ii) unsupervised learning. In the first method, the input data is estimated or mapped to the desired output using a labeled set of training data. Whereas there is no notion of the output during the unsupervised learning process, and no labeled examples are provided. Consequently, finding patterns or discovering the groups of the input data is up to the learning model. This procedure can be seen as a classification problem in supervised learning, in other words, it refers to the process of categorization of data into a set of finite classes (Kourou *et al.*, 2015).

In many domains of clinical practice, AI-based technologies have been successfully applied, such as decision-support systems, prediction and diagnosis, natural language processing, and image recognition (Harwich and Laycock, 2018).

Particularly, machine learning has found great applications in cancer research. In this context, the main types of predictive tasks that can be addressed with ML are three:

(i) the prediction of cancer susceptibility (risk assessment): in this case, the likelihood of developing a type of cancer is sought;

(ii) the prediction of cancer recurrence: in this case, the likelihood of redeveloping a type of cancer after complete or partial remission is sought;

(iii) the prediction of cancer survival: in this case, the main objective is the prediction of disease-specific or overall survival after cancer diagnosis or treatment.

At first, ML has been primarily used for cancer diagnosis and detection, indeed artificial neural networks (ANNs) and decision trees have been used for nearly 40 years for this purpose (Simes, 1985; Maclin *et al.*, 1991; Cicchetti, 1992). For instance, they are successfully applied in tumors detection and classification via X-ray images and in the classification of malignancies from proteomic and genomic (microarray) assays. Afterward, ML began to be used also for cancer prediction and prognosis, as part of a growing trend towards personalized medicine (Weston and Hood, 2004; Di Sanzo *et al.*, 2017; Goetz and Schork, 2018). In a review that analyzed studies regarding the cancer prediction and prognosis based on ML methods, Kourou et al. (Kourou *et al.*, 2015) noticed that the lack of external validation or testing was a common problem shared by several works, affecting the evaluation of the predictive performance of the models. However, different evaluation techniques that split the initial datasets into subsets are increasingly employed to, at least partially, overcome this issue. Among these techniques, the most widely used are:

(i) Holdout Method;

(ii) Random Sampling;

(iii) Bootstrap:

(iv) Cross-Validation.

In the first method, the whole dataset is split into two separate sets: the training and the test sets. The training set is then exploited to generate the classification model, while the model's performance is estimated using the test set. A similar approach to the Holdout method is random sampling. This approach consists in repeating several times the Holdout method, choosing the training and test instances randomly, to obtain a better estimation of the accuracy. In the bootstrap approach the samples are divided into training and test sets with replacement, in other words, after they have been chosen for training they are placed again into the entire data set. Lastly, in the cross-validation approach, different validation cycles are performed, in which each sample is used only once for testing and the same number of times for training. Averaging the different validation cycles allows obtaining final accuracy results. As this method is applied for this thesis work, it will be described in more detail in section 3.5. As a matter of fact, the authors stated that the application of ML methods could improve the accuracy of cancer susceptibility, recurrence, and survival prediction (Kourou *et al.*, 2015). Indeed, as reported by Cruz et al., the application of ML techniques allowed a 15%-20% improvement of the accuracy of cancer prediction outcomes in the last years (Cruz and Wishart, 2007). In many studies regarding disease prognosis and prediction based on ML techniques, gene expression profiles, clinical variables, and histological parameters are fed as input to the prognostic procedure in a complementary manner (Kourou *et al.*, 2015).

Over the last 10 years, various tests based on gene expression quantification have been introduced in clinical practice for different diseases. For instance, such tests are used in breast cancer to provide molecular stratification and to assess the risk of relapse (Yersal and Barutca, 2014; Dai *et al.*, 2015; Vieira and Schmitt, 2018). In particular, an example of a successful application of such tests in clinical practice for the identification of breast cancer molecular subtypes is the PAM50 test, which is based on nearest centroids, a linear classification machine learning algorithm. PAM50 was developed using microarray and quantitative reverse transcriptase polymerase chain reaction (RT-qPCR) data, and it was subsequently converted into a test based on the Nanostring nCounter platform, which was approved by Food and Drug Administration (FDA) as a predictive test called Prosigna (Wallden *et al.*, 2015). Independent studies confirmed the

prognostic value of this method and its derivatives (Bastien *et al.*, 2014; Nielsen *et al.*, 2014; Ohnstad *et al.*, 2017). Moreover, starting from the algorithm developed by Parker et al. (Parker *et al.*, 2009), also RNA-seq profiles have been used for PAM50 classification. PAM50 is based on a set of 50 genes, whose expression is measured to assign an intrinsic subtype (Luminal A, Luminal B, Normal-like, Her2-Enriched, and Basal) to a given breast cancer sample. Such classification is obtained based on the correlation of the gene expression of the sample with the five subtype centroids (Parker *et al.*, 2009).

The nearest centroid method for cancer class prediction from gene expression was further improved by Tibshirani et al. leading to the "nearest shrunken centroids" approach (Tibshirani *et al.*, 2002). By using a "de-noised" version of the centroids, this method allows identifying subsets of genes that best characterize each class. In other terms, this algorithm's extension involves shifting class-based centroids towards the overall centroid, thus performing an automatic feature selection by removing those input variables that are shrunk down to the value of the data centroid, as they are less useful at discriminating between the class labels. The amount of shrinkage applied to the centroids represents a hyperparameter that can be tuned for the specific dataset, making this method particularly appropriate for a dataset with a large number of input variables, considering that some of them may be irrelevant or noisy. This is exactly the case of a microarray experiment, where there are a huge number of genes (variables), but a much smaller number of samples, and relevant genes that distinguish samples have to be extracted among them. The effectiveness of this approach was demonstrated by the authors by building models to classify small round blue cell tumors and leukemias (Tibshirani *et al.*, 2002). Since this was the approach used in this thesis work to build the classifier for myelofibrosis samples, more details are described in the materials and methods chapter, section 3.4.

2 Aim of the thesis

Myeloproliferative neoplasms (MPNs) are a group of related hematologic cancers characterized by an excessive proliferation of terminally differentiated myeloid cells. The three disorders included in the so-called “classic” MPNs are polycythemia vera (PV), which involves elevated red-cell counts; essential thrombocythemia (ET), which is defined by elevated platelets in the blood; and myelofibrosis (MF), which is characterized by bone marrow fibrosis. PV and ET are chronic conditions that can progress to myelofibrosis (secondary MF), giving rise to post-PV (PPV) and post-ET (PET) myelofibrosis. However, MF can also occur without pre-existing conditions (primary MF).

During the past decade, several aberrations have been identified and characterized in MPN patients, such as the dysregulation of JAK2 signaling in 2005 and the CALR mutations in 2013. Still, the molecular pathogenesis of these malignancies remains incompletely understood and, for most MPNs, allogenic stem cell transplantation is the only curative treatment option. Diagnostic features and outcomes in patients show strong heterogeneity, probably as a consequence of a complex genetic landscape in addition to changes in gene expression. Currently, risk stratification of MPN patients is based mainly on clinical features and the presence of driver mutations. Gene expression data have already been used to provide prognostic indications in hematological diseases; nevertheless, none of the existing models for MPNs integrates transcriptomic data. Therefore, there is a need to better characterize the transcriptomic profile of these disorders in order to elaborate a signature able to add more robustness to clinically widely accepted scoring systems.

The main aim of this project is to exploit gene expression profiles (GEPs) of granulocytes from patients affected by primary and secondary myelofibrosis to predict survival and distinguish “high risk” myelofibrosis patients, with inferior overall survival, from “low risk” ones.

The first step consists of identifying survival-related transcripts exploiting a Cox regression analysis.

The second step is focused on the optimization of the number of transcripts composing the molecular signature, through the development of a robust classification model, based on a supervised learning technique, with potential for translation into the clinic.

As a third step, we studied the correlation between gene expression profiles of myelofibrosis patients and their clinical characteristics. In particular, we compared the model with current predictive scores in terms of prognostic power and investigated how, if combined with them as an independent prognostic factor, it can improve the ability to direct patients toward the best available therapeutic strategy.

3 Materials and methods

3.1 Patients and samples

Samples of granulocytes were obtained from 114 patients with a diagnosis of PMF or SMF recruited from 5 Italian centers.

The dataset included:

- 35 pre-PMF;
- 37 overt PMF;
- 26 PET-MF;
- 16 PPV-MF.

The 2016 World Health Organization criteria (Arber *et al.*, 2016) were used for the diagnosis of PMF, whereas PET-MF and PPV-MF were diagnosed according to the International Working Group for Myeloproliferative neoplasms Research and Treatment criteria (Barosi *et al.*, 2008). The study was conducted in accordance with the Declaration of Helsinki and was approved by local ethics committees. All subjects provided informed written consent.

Granulocytes were isolated from peripheral blood using a density gradient-based method. Following centrifugation, granulocytes and erythrocytes formed a cell pellet at the bottom of the tube, and granulocytes were purified by means of red blood cell lysis reagent. The 5 different units provided frozen granulocyte pellets and lysed cells.

The different units provided also information on the mutational status of the sample. Specifically, gene mutations were detected in DNA from peripheral blood cells (Rotunno *et al.*, 2016; Guglielmelli, Terra L Lasho, *et al.*, 2018):

- real-time quantitative polymerase chain reaction was used to identify JAK2V617F and MPLW515x mutations;
- CALR mutations were detected by capillary electrophoresis, followed by bidirectional sequencing, and classified as type 1/type 1-like or type 2/type 2-like (P Guglielmelli *et al.*, 2015).
- Next-generation sequencing (NGS) was used to detect high molecular risk (HMR) mutations (i.e., ASXL1, EZH2, IDH1, IDH2, and SRSF2).

3.2 RNA extraction and gene expression profiling

Gene expression profiling was performed, after total cellular RNA was isolated from stored frozen granulocyte pellets, using HG-U219 Array Strips (Affymetrix), as previously described (Norfo *et al.*, 2014). Partek Genomics Suite (GS) software, version 7.0 (Partek Inc., St. Louis, MO) was used to normalize the probe level data and convert them into expression values by means of robust multiarray average (rma). To assess whether major sources of variation can be explained by batch effects, an exploratory principal component analysis (PCA) was performed. In particular, the effects of the unit of origin, disease diagnosis, and driver mutation were evaluated. PCA plot showed that samples clustered together according to the clinical unit of origin (Figure 8), while diagnosis and driver mutation displayed smaller effects.

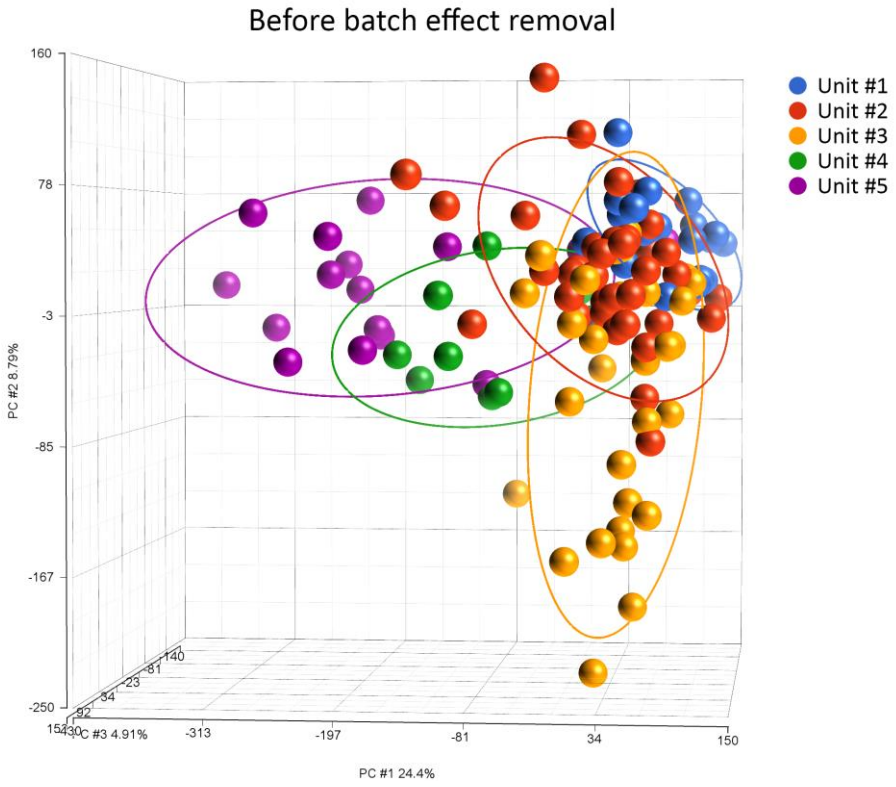

Figure 8. Principal component analysis (PCA) of samples before removing batch effect. Samples are colored according to unit of origin. Ellipses enclose samples according to unit of origin.

The existing batch effect due to the clinical unit of origin was removed using the "remove batch effect" function in Partek GS. This is based on a mixed-model analysis of variance and performs adjustment of the gene expression matrix, removing differences between batches. To verify the batch effect adjustment, PCA was performed on the adjusted data. The resulting PCA plot (Figure 9) showed that batch effect correction was effective in removing unit effect, indeed samples clustering according to the unit was no longer observed, but only the effect of disease diagnosis and driver mutation was still visible.

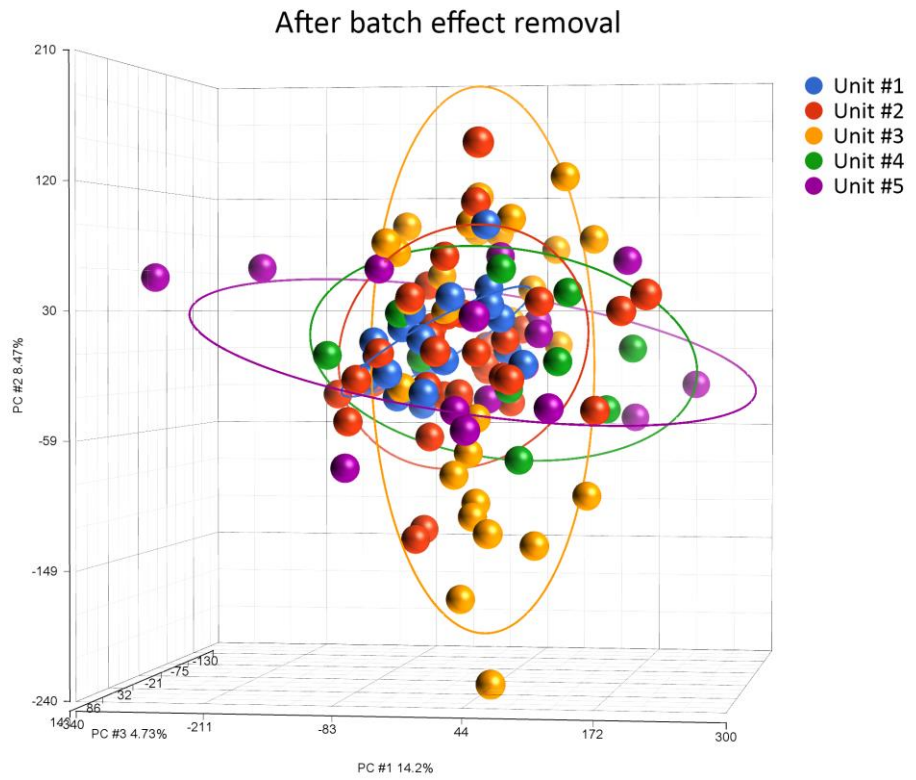

Figure 9. Principal component analysis (PCA) of samples after removing batch effect. Samples are colored according to unit of origin. Ellipses enclose samples according to unit of origin.

3.3 Identification of survival-related transcripts

Probe sets, in Affymetrix arrays, are groups of probes of the microarray targeting the same transcript with multiple measurements. We decided to perform a probe set-level Cox regression analysis, rather than a gene-level analysis, in order not to lose information coming from different probe sets recognizing the same gene. This approach has been successfully applied in other studies (Bonome *et al.*, 2008). Official gene symbols were subsequently associated with probe sets by using the current NetAffx annotation file (HG-U219, release 36).

3.3.1 Assumptions checking

To apply the Cox model correctly, we tested if the proportional hazard assumption was satisfied for the model, using the *cox.zph* function from the R

package *survival* (v. 3.1.8). The analysis confirmed that for all the 832 probe sets correlated with patients' survival the proportional hazard assumption was satisfied.

3.3.2 Cox regression

In order to identify probe sets whose expression correlates with patient survival, Cox regression analysis was performed using the *coxph* function of the R package *survival* (v. 3.1.8). Probe sets with a p-value < 0.005 were selected for further analyses.

3.3.3 Supervised clustering

Survival-related probe sets were used to perform a supervised hierarchical clustering exploiting the function *Heatmap* of the R package *ComplexHeatmap* (v.1.20.0). Clinical features of samples group that share similar gene expression patterns for these probe sets, resulting from this analysis, were evaluated. Risk classes were defined based on the results of the analysis.

3.4 Classification model construction

A classification model was built starting from the list of probe sets resulting from Cox regression analysis and risk classes defined with hierarchical clustering. The "nearest shrunken centroids" supervised learning technique, implemented in the *pamr.train* function of the R package *pamr* (v.1.56.1) was applied for the classification model generation. Specifically, the following steps were performed:

- for each risk class defined with hierarchical clustering a standardized centroid was calculated. The centroid is defined as the median expression of each probe set in each class divided by the standard deviation in each class;
- centroids were shrunk towards the overall centroid by a fixed quantity called threshold. Thirty different thresholds were tested to find the one

that gives the lowest error rate (Figure 10). If a probe set was shrunk to zero for all classes, then it did not survive the threshold and was eliminated from the prediction rule.

The latter step made it possible to optimize the number of probe sets in the model by excluding from the prediction rule those whose expression was not sufficiently different between the different risk classes (and thus not contributing to the classification).

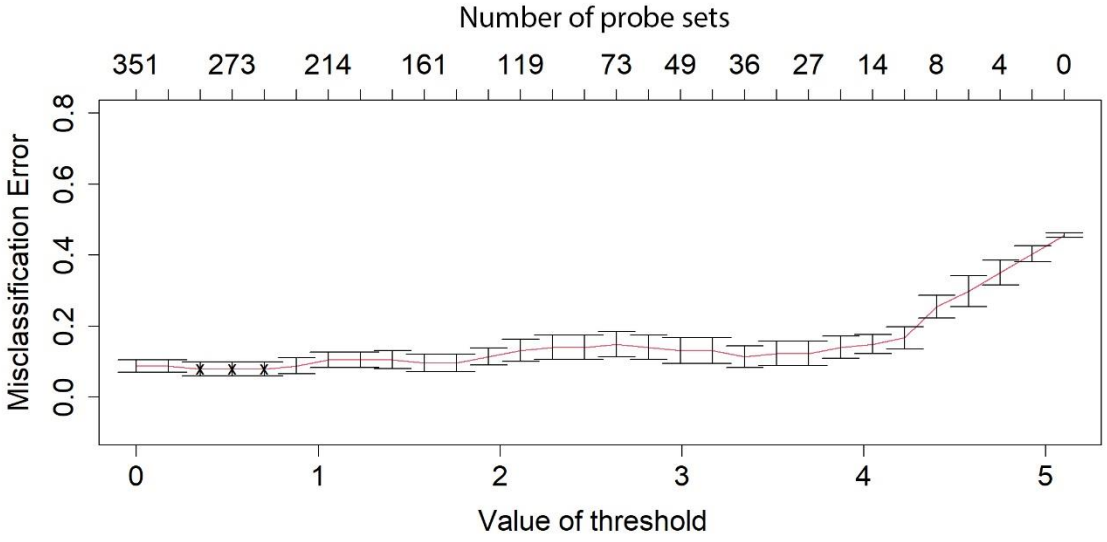

Figure 10. The cross-validated misclassification error curve, from nearest shrunken centroid classifier. Y axis is the misclassification error, bottom X axis is the value of threshold, top X axis is the number of surviving probe sets for the corresponding threshold.

3.5 Classification model validation

A major problem encountered for the classification model validation was the absence of an external independent validation set. Moreover, the limited number of samples included in our dataset makes classical methods based on splitting the dataset into training and test sets unsuitable. To overcome these issues, the resulting model was cross-validated with a k-fold cross-validation method, with k = 20, using the *pamr.cv* function in the R package *pamr*. In particular, the following steps were performed:

- for each threshold (described in the previous section), the whole dataset was split into k smaller sets (k=20) and a model was defined using k -1 of the folds;
- the remaining part of the data was used for testing the model, predicting the class of the test samples on the basis of the nearest centroid, in terms of squared distance;
- the process was repeated until each fold was used as a test set;
- a cross-validated misclassification error was obtained for each threshold.

3.6 Classification model optimization

To further optimize the number of probe sets, model generation and cross-validation were repeated $m - 1$ times (where m is the total number of probe sets). Starting from 2 probe sets, each time 1 probe set was sequentially added from the top of the rank-ordered probe set list based on the hazard ratio (HR), until all of the probe sets were used.

A similar approach has been successfully applied by van't Veer et al. (Van't Veer *et al.*, 2002) who optimized the number of genes in a "prognosis classifier" in breast cancer by sequentially adding subsets of 5 genes from the top of a rank-ordered list based on the correlation coefficient of the gene expression with disease outcome.

The model's performance was assessed with the use of the cross-validated misclassification error calculated with the *pamr.cv* function from the *pamr* R package. At the end of the process, the model with the optimized set of probe sets that provided the lowest error was selected.

3.7 Statistical analyses

The prognostic value of the optimized model was tested on the same dataset analyzing available clinical data. Differences in the distribution of numerical variables were evaluated using the Mann-Withney U test or the Kruskal-Wallis test for the comparison of 2 groups or >2 groups, respectively, exploiting the

functions *wilcox.test* and *kruskal.test* in the R package *stats* (v.4.0.0). Categorical variables were compared using the Fisher's exact test for 2x2 contingency tables or χ^2 test for all other cases, applying the functions *fisher.test* and *chisq.test* from the R package *stats*. Overall survival (OS) was calculated from the date of sample collection to the date of death or last follow-up. When calculating leukemia-free survival (LFS), the date of leukemic transformation was used in place of the date of death. OS and LFS analyses were performed with the Kaplan-Meier method, using the *survplot* function of the R package *rms* (v.6.0-1); and a log-rank test was used to compare curves, exploiting the *survdiff* function of the *survival* R package (v.3.1-12). Univariate and multivariate analyses were carried out using Cox proportional-hazards regression for OS, exploiting the *coxph* function of the R package *survival* and the function *analyse_multivariate* from the *survivalAnalysis* R package (v.0.2.0). A p-value < 0.05 was considered statistically significant.

3.8 Pathway enrichment analysis

To better elucidate the biological meaning of genes that compose the model, a pathway enrichment analysis was performed using the genes of the model that, having a hazard ratio greater than 1, are associated with an increased risk of death. The analysis was performed using EnrichR (R package *enrichR*, v.3.0) and ToppFun from the ToppGene Suite (Chen *et al.*, 2009).

3.9 Integration with contemporary prognostic models

To strengthen our findings and assess if our model was able to improve the prognostic power of existing prognostic scores for MF, we designed two new combined models by integrating information from our gene expression-based classification within DIPSS and MIPSS70 (Figure 11; Figure 12). Since we observed that our model was particularly efficient in identifying high-risk (HR) patients within DIPSS and MIPSS70 intermediate classes, we decided to design the new combined models in the following way:

- MODIFIED DIPSS: intermediate-1 (INT-1) and intermediate-2 (INT-2) patients identified as HR according to our expression signature were assigned to the higher class (INT-2 and HIGH), respectively;
- MODIFIED MIPSS70: intermediate (INT) patients identified as HR according to our expression signature were assigned to the higher class (HIGH).

To compare models for prediction of survival, the Akaike information criterion (AIC) score was determined with the *stepAIC* function of the MASS R package (v.7.3-51.5).

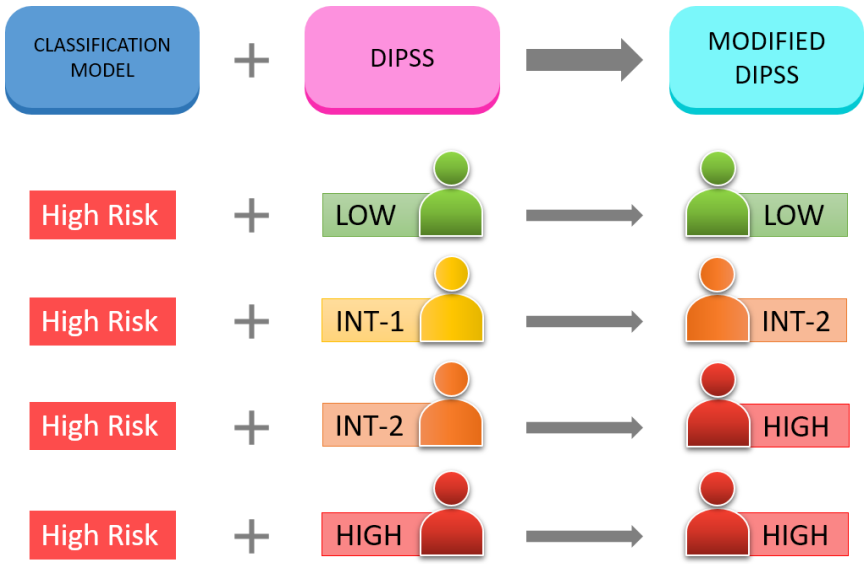

Figure 11. Design of integrated DIPSS prognostic model, rules for the construction of a modified DIPSS classification. The figure represents how gene expression-based classification was incorporated in DIPSS prognostic model. Intermediate-1 (INT-1) and Intermediate-2 (INT-2) patients identified as HR according to our expression signature were assigned to the higher class (Intermediate-2 (INT-2) and HIGH, respectively).

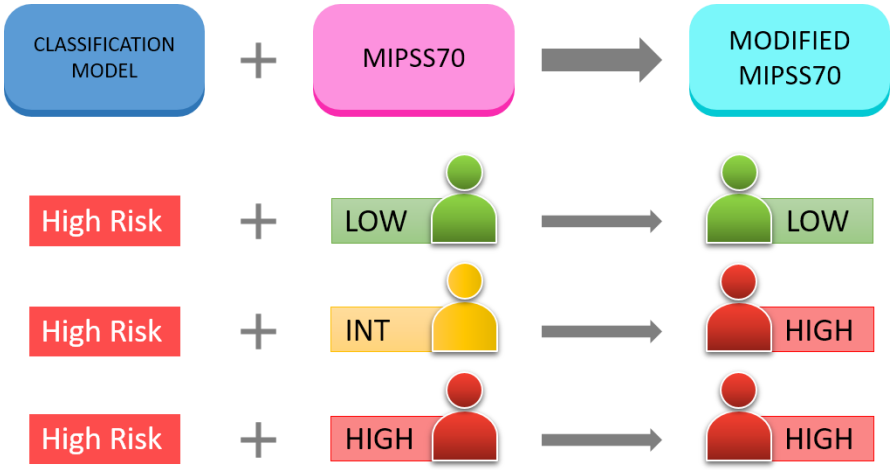

Figure 12. Design of integrated MIPSS70 prognostic model, rules for the construction of a modified MIPSS70 classification. The figure represents how gene expression-based classification was incorporated in MIPSS70 prognostic model. Intermediate (INT) patients identified as HR according to our expression signature were assigned to the higher class (HIGH).

4 Results

4.1 Patients

A total of 114 patients (35 Pre-PMF, 37 Overt-PMF, 26 PET-MF, and 16 PPV-MF) were analyzed. Table 2 shows the distribution of patients in MF clinical subtypes. Some dissimilarities were evident among these groups concerning clinical characteristics, as expected.

Indeed, pre-PMF patients displayed increased hemoglobin (Hb) levels and were less frequently anemic; moreover, platelet count was increased in this group and splenomegaly was more frequent. Hb was increased also in PPV-MF as well as white blood cell counts. Anemia was less frequent in these patients while leukocytosis was more common. PET-MF showed increased platelet counts. All PPV-MF patients harbored a JAK2 V617F mutation except for one patient displaying a mutation on exon 12. In contrast, pre-PMF displayed an increased frequency of SRSF2 mutation and more than 2 high molecular risk (HMR) mutations (namely ASXL1, EZH2, SRSF2, and IDH1/2). Leukemic transformation occurred in 13 patients while 49 died from causes related to the disease.

Variable	Pre-PMF (n=35)	Overt-PMF (n=37)	PET-MF (n=26)	PPV-MF (n=16)	p-value
Follow up, median (95% CI), y	6.88 (3.39-NA)	5.54 (3.26-6.36)	4.55 (2.91-NA)	4.18 (1.72-6.79)	6.15E-01
Males	19 (54.3)	21 (56.8)	13 (50.0)	7 (43.8)	8.33E-01
Age					

Median (range), y	62.90 (34.6-80.9)	63.80 (31.1-84.3)	65.80 (32.1-81.0)	71.10 (42.6-85.6)	9.62E-02
> 65 y,	13 (37.1)	15 (40.5)	14 (53.8)	11 (68.8)	1.38E-01
Hemoglobin					
Median (range), g/dL	12.40 (8.0-16.6)	11.20 (5.2-15.3)	10.75 (6.5-14.6)	12.55 (9.2-15.9)	**2.52E-04**
< 10 g/dL	2 (5.7)	9 (24.3)	9 (34.6)	2 (12.5)	**2.75E-02**
Leukocytes					
Median (range), x 10^9/L	8.70 (3.6-41.0)	10.00 (2.8-89.0)	9.58 (2.3-104.0)	14.90 (5.9-88.7)	**1.35E-02**
> 25 x 10^9/L	3 (8.8)	6 (16.2)	3 (11.5)	7 (46.7)	**9.09E-03**
Platelets					
Median (range), x 10^9/L	410.0 (72-1299)	179.0 (22-1252)	377.5 (61-1568)	224.5 (20-1271)	**6.54E-03**
< 100 x 10^9/L	2 (5.9)	9 (25.0)	2 (8.3)	1 (6.2)	6.19E-02
Circulating blast, ≥ 1%	5 (15.2)	8 (25.8)	8 (34.8)	3 (20.0)	3.77E-01
Circulating blast, ≥ 2%	5 (15.2)	4 (12.9)	6 (26.1)	2 (13.3)	5.81E-01
BM fibrosis grade ≥ 2	-	33 (97.1)	23 (100)	13 (92.9)	4.44E-01
Constitutional symptoms	5 (14.7)	11 (30.6)	7 (26.9)	5 (31.2)	4.08E-01
Splenomegaly	16 (45.7)	31 (86.1)	22 (84.6)	10 (71.4)	**6.24E-04**
Driver mutation					
JAK2 V617F	12 (34.3)	16 (43.2)	11 (42.3)	15 (93.8)	**8.04E-04**
JAK2 ex12	0	0	0	1 (6.2)	-
CALR unspecified	0	0	1 (3.8)	0	-
CALR Type 1	4 (11.4)	10 (27.0)	7 (26.9)	0	1.99E-01
CALR Type 2	5 (14.3)	3 (8.1)	0	0	1.31E-01
MPL	5 (14.3)	3 (8.1)	3 (11.5)	0	7.07E-01
Triple Negative	9 (25.7)	5 (13.5)	4 (15.4)	0	3.69E-01
ASXL1 mutation (n evaluable, total = 85)	25	32	15	13	
n (%)	11 (44.0)	10 (31.2)	5 (33.3)	6 (46.2)	6.81E-01
EZH2 mutation (n evaluable, total = 82)	24	27	16	15	
n (%)	3 (12.5)	2 (7.4)	0	1 (6.7)	5.27E-01
SRSF2 mutation (n evaluable, total = 81)	23	27	16	15	
n (%)	7 (30.4)	1 (3.7)	1 (6.2)	0	**5.98E-03**
IDH1/2 mutation (n evaluable, total = 81)	23	27	16	15	
n (%)	2 (8.7)	1 (6.7)	0	1 (6.7)	6.36E-01
HMR (n evaluable, total = 86)	25	29	17	15	
n (%)	14 (56.0)	12 (41.4)	6 (35.3)	7 (46.7)	5.66E-01
≥ 2	8 (32.0)	2 (6.9)	0	1 (6.7)	**6.82E-03**
DIPSS (n evaluable, total = 107)	31	36	25	15	
Low	14 (45.2)	9 (25.0)	4 (16.0)	3 (20.0)	
Intermediate-1	10 (32.3)	13 (36.1)	13 (52.0)	5 (33.3)	
Intermediate-2	3 (9.7)	12 (33.3)	4 (16.0)	5 (33.3)	
High	4 (12.9)	2 (5.6)	4 (16.0)	2 (13.3)	3.89E-01

MIPSS70 (n evaluable, total = 73)	21	25	15	12	
Low	10 (47.6)	2 (8.0)	0	1 (8.3)	
Intermediate	4 (19.0)	14 (56.0)	10 (66.7)	5 (41.7)	
High	7 (33.3)	9 (36.0)	5 (33.3)	6 (50.0)	**1.71E-03**
MYSEC-PM (n evaluable, total = 41)			23	15	
Low	-	-	7 (30.4)	2 (13.3)	
Intermediate-1	-	-	9 (39.1)	7 (46.7)	
Intermediate-2	-	-	4 (17.4)	3 (20.0)	
High	-	-	3 (13.0)	3 (20.0)	6.70E-01
Progression to leukemia	8 (22.9)	3 (8.1)	2 (7.7)	0	6.14E-02
Death	13 (37.1)	17 (45.9)	9 (34.6)	10 (62.5)	2.78E-01

Table 2. Clinical and laboratory characteristics of patients included in our dataset, divided according to diagnosis. Unless otherwise noted, data are n (%). Significant p-values (< 0.05) are highlighted in bold. NA, not available. -, missing value.

4.2 Identification of a gene signature that correlates with overall survival

Cox regression analysis identified 832 probe sets (corresponding to 596 genes) correlated with patients' survival, among them 433 genes were associated with inferior survival (Table 15). According to supervised hierarchical clustering (Figure 13), this list split our dataset into 2 main branches composed of 62 (left) and 52 (right) samples.

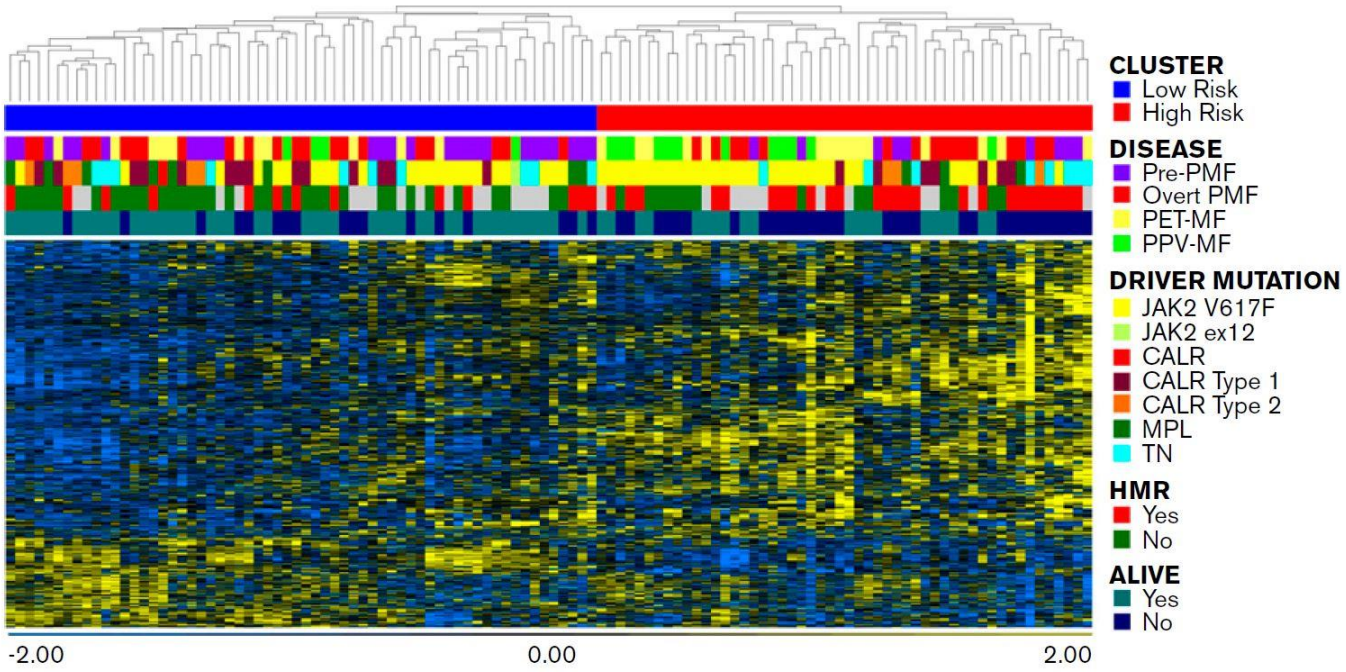

Figure 13. Hierarchical clustering of samples according to the expression of 832 probe sets that correlated with OS. The dataset was divided into 2 clusters.

As shown in Figure 14 and Table 3, the cluster on the right was characterized by significant inferior OS (p-value=4.38e-6, log-rank test) compared with the one on the left. Moreover, the frequency of dead patients was significantly higher in the right branch (p-value=3.08e-4, Fisher's exact test) (Table 3), therefore we named this cluster High Risk (HR) while the other one was termed Low Risk (LR).

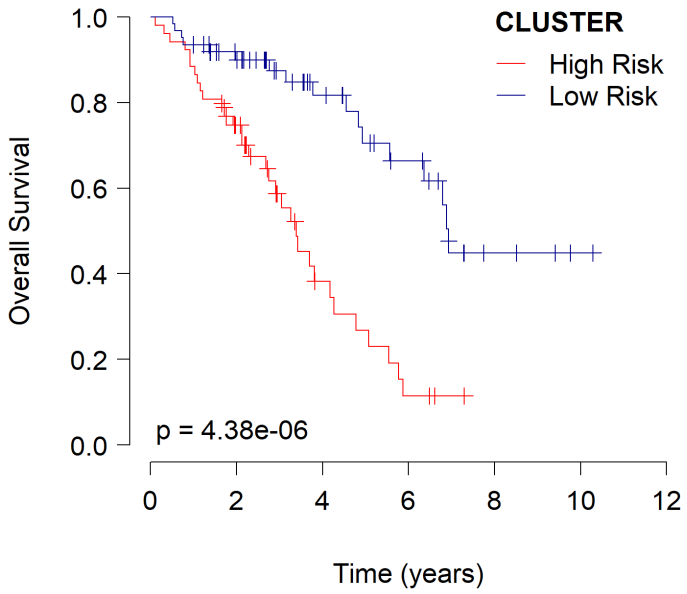

Figure 14. Kaplan-Meier curve comparing OS of samples belonging to the "High Risk" cluster and samples belonging to the "Low Risk" cluster.

The majority of pre-PMF samples were included in the LR cluster while SMF samples' frequency was higher in the HR (Figure 13; Table 3). The HR group was also enriched in patients harboring JAK2V617F mutation; moreover, the frequency of homozygous mutation was increased in this subgroup while JAK2 heterozygosity was more frequent in the LR group. Furthermore, 26 out of 40 evaluable samples (65 %) in the HR group harbored at least 1 HMR mutation (Figure 13; Table 3). Despite this, we identified a significant difference only in the distribution of patients with ASXL1 mutations (p-value=1.05e-4, Fisher's exact test). Considering clinical characteristics, the HR group displayed features of patients with predictable inferior survival, as it was characterized by the presence of clinical and molecular features representing well-known risk factors in myelofibrosis. Indeed, the median age at the time of sample collection was higher in this group compared to LR. Furthermore, white blood cell count was increased in HR patients, while hemoglobin levels and platelet count were decreased compared to LR. HR group exhibited increased frequencies of patients with more than 1% or 2% circulating blasts, splenomegaly, and BM fibrosis grade ≥ 2 (Table

3). Taken together these results demonstrated that Cox regression analysis led to the identification of genes whose expression correlates with OS in MF patients. These genes identified two patient subgroups characterized by high-level differences in terms of clinical and molecular features, and with divergent OS.

Variable	Low Risk (n=62)	High Risk (n=52)	p-value
OS, median (95% CI), y	6.93 (5.56-NA)	3.39 (2.68-4.27)	**4.38E-06**
Males	32 (51.6)	28 (53.8)	8.52E-01
Disease			
Pre-PMF	26 (41.9)	9 (17.3)	
Overt PMF	21 (33.9)	16 (30.8)	
PET-MF	11 (17.7)	15 (28.8)	
PPV-MF	4 (6.5)	12 (23.1)	**5.16E-03**
Age			
Median (range), y	62.1 (31.1-81)	68.5 (32.0-85.6)	**5.71E-04**
> 65 y	21 (33.9)	32 (61.5)	**4.56E-03**
Hemoglobin			
Median (range), g/dL	12.1 (6.5-16.6)	11.2 (5.2-14.6)	**7.74E-03**
< 10 g/dL	7 (11.3)	15 (28.8)	**3.05E-02**
Leukocytes			
Median (range), x 10^9/L	8.0 (2.3-38.7)	14.4 (3.15-104)	**3.28E-07**
> 25 x 10^9/L	1 (1.6)	18 (35.3)	**1.27E-06**
Platelets			
Median (range), x 10^9/L	417 (22-1568)	223 (20-1271)	**2.54E-03**
< 100 x 10^9/L	4 (6.8)	10 (19.6)	5.07E-02
Circulating blast, ≥ 1%	7 (12.1)	17 (38.6)	**2.31E-03**
Circulating blast, ≥ 2%	5 (8.6)	12 (27.3)	**1.61E-02**
BM fibrosis grade ≥ 2	31 (52.5)	38 (80.9)	**3.79E-03**
Constitutional symptoms	11 (17.7)	17 (34)	7.80E-02
Splenomegaly	34 (54.8)	45 (91.8)	**1.57E-05**
Driver mutation			
JAK2 V617F	23 (37.1)	31 (59.6)	**2.36E-02**
JAK2 ex12	1 (1.6)	0	1.00E+00
CALR unspecified	1 (1.6)	0	1.00E+00
CALR Type 1	14 (22.6)	7 (13.5)	2.35E-01
CALR Type 2	5 (8.1)	3 (5.8)	7.26E-01
MPL	8 (12.9)	3 (5.8)	3.40E-01

Triple Negative	10 (16.1)	8 (15.4)	1.00E+00
JAK2 V617F			
Heterozygous	16 (72.7%)	5 (16.7 %)	
Homozygous	6 (27.3%)	25 (83.3 %)	**6.28E-05**
CALR Type 1 absence	48 (77.4)	45 (86.5)	2.35E-01
ASXL1 mutation (n evaluable, total = 85)	45	40	
n (%)	8 (17.8)	24 (60)	**1.05E-04**
EZH2 mutation (n evaluable, total =82)	46	36	
n (%)	3 (8.3)	3 (6.5)	1.00E+00
SRSF2 mutation (n evaluable, total =81)	45	36	
n (%)	5 (11.1)	4 (11.1)	1.00E+00
IDH1/2 mutation (n evaluable, total =81)	45	36	
n (%)	3 (6.7)	1 (2.8)	6.25E-01
HMR (n evaluable, total =86)	46	40	
n (%)	13 (28.3)	26 (65)	**1.03E-03**
≥ 2	5 (10.9)	6 (15)	7.48E-01
Progression to leukemia	4 (6.5)	9 (17.3)	8.25E-02
Death	17 (27.4)	32 (61.5)	**3.08E-04**

Table 3. Clinical characteristics of patients included in our dataset, divided according to hierarchical clustering obtained using the list of 832 probe sets from cox regression analysis. A High Risk and a Low Risk group were considered for this analysis. Unless otherwise noted, data are n (%). Significant p-values (< 0.05) are highlighted in bold.

4.3 Classification model performance

The list of 832 probe sets derived from cox regression analysis was used to build a gene expression-based classification model for the two risk categories defined according to hierarchical clustering. As detailed in the materials and methods chapter, sections 3.4, 3.5, and 3.6, a supervised learning technique, the "nearest shrunken centroids" method, was applied to construct the model. It was then cross-validated using a 20-fold cross-validation strategy, exploiting the pamr.cv R function to estimate the misclassification error and build a robust classifier. Specifically, thirty different thresholds were evaluated during the 20-fold cross-validation to identify the model with the lowest misclassification error. To further optimize the number of probe sets in the model, the whole process was repeated 831 times, starting from a model with only 2 probe sets, and then sequentially

adding 1 probe set from the top of the rank-ordered list of 832 probe sets. The model with the lowest cross-validated misclassification error was selected among the 831 models.

The process of probe set number optimization allowed lowering the misclassification error from 9.65% (model with 832 probe sets) to 6.14% (model with 351 probe sets). In particular, we observed a dramatic change in the number of wrong predictions, in terms of cross-validated misclassification error, as the number of probe sets employed varied. The accuracy improved when increasing the number of probe sets until the optimal number of probe sets was reached. Nevertheless, beyond the optimal number of probe sets the accuracy became slightly worse (Figure 15).

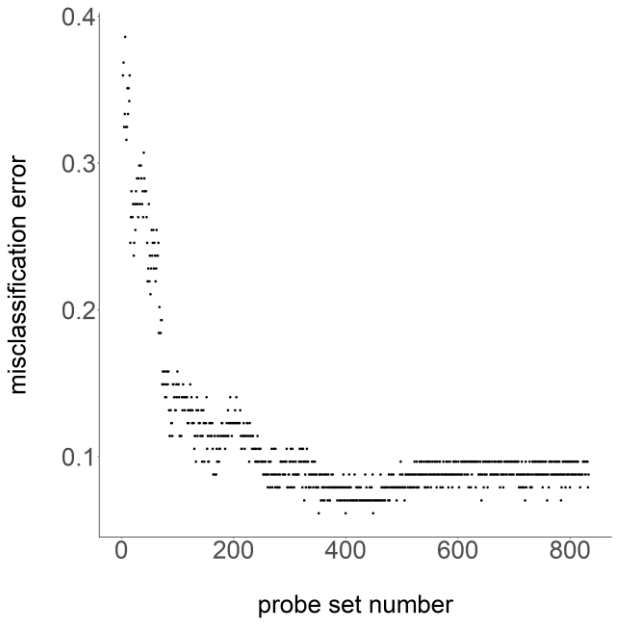

Figure 15. The cross-validated misclassification error as a function of the number of discriminating probe sets used in the model. Y axis is the misclassification error, X axis is the number of probe sets.

We obtained an optimized model using the first 351 probe sets of the list, of which 273 (corresponding to 201 genes) survived the cross-validation threshold (Table 16).

We used this optimized model to assign a classification (high-risk or low-risk) to the 114 MF samples of our dataset. According to this classification, LR and HR groups identified were composed of 60 and 54 patients respectively (Table 4). We observed a high level of concordance (92.98%) with the classification defined with supervised clustering with 832 probe sets. Patients in the HR group displayed several detrimental characteristics. The HR class was enriched in samples derived from dead patients (34 vs 15, p-value=6.01e-5, Fisher's exact test) and was characterized by a significant inferior survival compared to the LR group (p-value=1.78e-7, log-rank test) (Figure 16A; Table 4). Moreover, the LR group was enriched in pre-PMF samples while overt PMF samples were equally distributed between the two classes. On the contrary, higher percentages of PET-MF and PPV-MF were present in the HR group.

Considering molecular characteristics, the HR group was enriched in patients harboring JAK2V617F mutation (Table 4), and particularly, homozygosity was more frequent in this group rather than LR one. Similarly, the low-risk group was enriched in patients harboring JAK2V617F heterozygous mutation and this is consistent with the most recent molecular findings. Indeed, it has been reported that MPNs with heterozygous JAK2V617F mutation have a generally favorable prognosis, while the presence of JAK2V617F homozygous mutation is associated with PV diagnosis and an increased frequency of myelofibrosis transformation. The frequency of patients with at least one HMR mutation was increased in the HR group; again, this group was enriched in ASXL1 mutated patients compared to LR one (Table 4). This finding is consistent with results reported by Grinfeld et al. (Grinfeld *et al.*, 2018) where they observed an increased risk of myelofibrosis transformation and reduced event-free survival, considering both AML transformation and death, within a subgroup of patients characterized by one or more mutations in 16 myeloid cancer genes. Myelofibrosis patients within this subgroup displayed inferior overall survival compared with patients with heterozygous JAK2V617F mutation, considered as reference. Unfortunately, only for 36 samples, we have additional information regarding other important mutations in MPNs, such as CBL, TET2, DNMT3A, and TP53 mutations. Due to the small number of samples harboring mutations in the latter genes, we were not able to draw any conclusions on this topic.

Next, we studied the distribution of clinical variables included in contemporary prognostic models and observed that HR classification correlated with the presence of clinical markers of inferior survival. Indeed, HR group was characterized by: (i) higher patients' age at sample collection; (ii) inferior hemoglobin levels; (iii) inferior platelet counts; (iv) superior white blood cell counts; (v) increased incidence of splenomegaly; (vi) circulating blasts greater than 1% or 2%; (vii) BM fibrosis grade ≥ 2; (viii) constitutional symptoms (Table 4). Even if 9 out of 13 patients who developed secondary AML clustered within the HR group, this difference failed to reach the statistical significance (Table 4). Nevertheless, survival analysis revealed that LFS was significantly reduced in the HR group (p-value=1.9e-2, Log-rank test) (Figure 16B). Collectively, these results demonstrated that belonging to the HR group represented a risk factor for survival (hazard ratio=4.736; 95% CI=2.5-8.9; p-value=1.48e-6) and for leukemic transformation (hazard ratio=3.976; 95% CI=1.2-13.6; p-value=2.75e-2). Of particular interest, this is true also considering samples stratified according to diagnosis. Indeed, HR patients displayed inferior OS when PMF and SMF samples were separately considered (Figure 17).

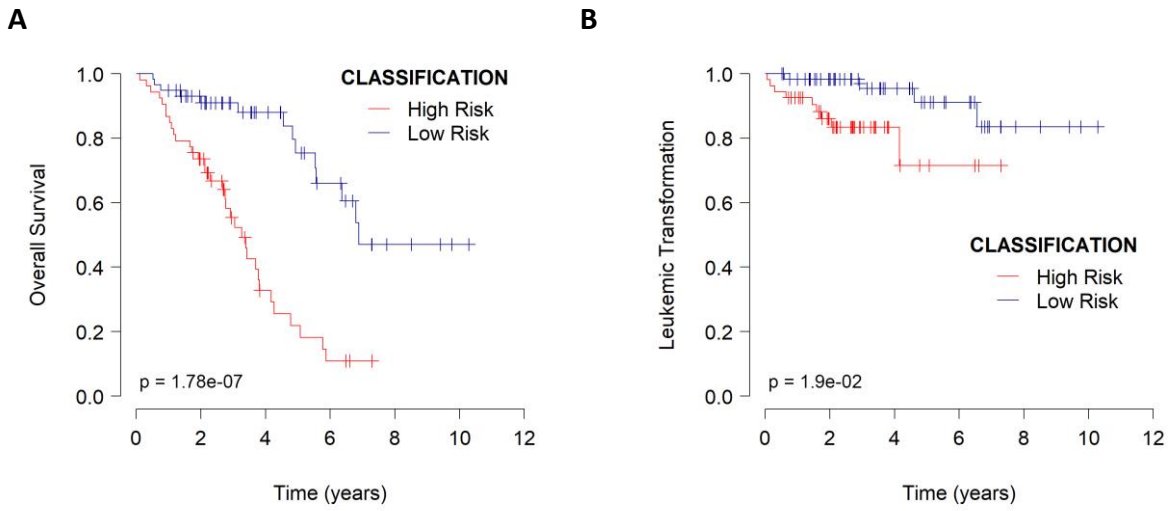

Figure 16. OS and LFS. (A) Kaplan-Meier analysis demonstrated a statistically significant difference in terms of OS between the high-risk and low-risk groups, as identified by gene expression–based classifier. (B) The high-risk group showed an inferior LFS compared with the low-risk group.

Variable	Low Risk (n=60)	High Risk (n=54)	p-value
OS, median (95% CI), y	6.93 (5.56-NA)	3.26 (2.68-3.81)	**1.78E-07**
Males	32 (53.3)	28 (51.9)	1.00E+00
Disease			
Pre-PMF	25 (41.7)	10 (18.5)	
Overt PMF	20 (33.3)	17 (31.5)	
PET-MF	11 (18.3)	15 (27.8)	
PPV-MF	4 (6.7)	12 (22.2)	**1.17E-02**
Age			
Median (range), y	61.4 (32.1-81)	69.95 (31.1-85.6)	**5.73E-05**
> 65	18 (30)	35 (64.8)	**3.15E-04**
Hemoglobin			
Median (range), g/dL	12.1 (6.5-16.6)	11.15 (5.2-14.4)	**1.89E-03**
< 10 g/dL	6 (10)	16 (29.6)	**9.44E-03**
Leukocytes			
Median (range), x 10^9/L	8.00 (2.3-38.7)	14.2 (3.15-104)	**3.98E-06**
> 25 x 10^9/L	1 (1.7)	18 (34)	**3.80E-06**
Platelets			
Median (range), x 10^9/L	386 (22-1568)	226 (20-1271)	**6.85E-03**
< 100 x 10^9/L	4 (7)	10 (18.9)	8.62E-02
Circulating blast, ≥ 1%	8 (14.3)	16 (34.8)	**1.94E-02**
Circulating blast, ≥ 2%	5 (8.9)	12 (26.1)	**3.13E-02**
BM fibrosis grade ≥ 2	33 (55.9)	36 (76.6)	**3.97E-02**
Constitutional symptoms	10 (16.7)	18 (34.6)	**4.78E-02**
Splenomegaly	34 (56.7)	45 (88.2)	**2.93E-04**
Driver mutation			
JAK2 V617F	22 (36.7)	32 (59.3)	**2.38E-02**
JAK2 ex12	1 (1.7)	0	1.00E+00
CALR unspecified	1 (1.7)	0	1.00E+00
CALR Type 1	14 (23.3)	7 (13)	2.26E-01
CALR Type 2	6 (10)	2 (3.7)	2.77E-01
MPL	8 (13.3)	3 (5.6)	2.11E-01
Triple Negative	8 (13.3)	10 (18.5)	6.08E-01
JAK2 V617F			
Heterozygous	14 (66.7)	7 (22.6)	
Homozygous	7 (33.3)	24 (77.4)	**3.41E-03**

CALR Type 1 absence	46 (76.7)	47 (87)	2.26E-01
ASXL1 mutation (n evaluable, total =85)	46	39	
n (%)	9 (19.6)	23 (59)	**2.79E-04**
EZH2 mutation (n evaluable, total =82)	48	34	
n (%)	3 (6.2)	3 (8.8)	6.88E-01
SRSF2 mutation (n evaluable, total =81)	47	34	
n (%)	4 (8.5)	5 (14.7)	4.81E-01
IDH1/2 mutation (n evaluable, total =81)	47	34	
n (%)	3 (6.4)	1 (2.9)	6.36E-01
HMR (n evaluable, total =86)	48	38	
n (%)	14 (29.2)	25 (65.8)	**1.02E-03**
≥ 2	4 (8.3)	7 (18.4)	2.03E-01
DIPSS (n evaluable, total =107)	57	50	
Low	23 (40.4)	7 (14)	
Intermediate-1	24 (42.1)	17 (34)	
Intermediate-2	8 (14)	16 (32)	
High	2 (3.5)	10 (20)	**6.00E-04**
MIPSS70 (n evaluable, total =73)	41	32	
Low	12 (29.3)	1 (3.1)	
Intermediate	22 (53.7)	11 (34.4)	
High	7 (17.1)	20 (62.5)	**1.01E-04**
MYSEC-PM (n evaluable, total =38)	14	24	
Low	5 (35.7)	4 (16.7)	
Intermediate-1	5 (35.7)	11 (45.8)	
Intermediate-2	3 (21.4)	4 (16.7)	
High	1 (7.1)	5 (20.8)	4.36E-01
Progression to leukemia	4 (6.7)	9 (16.7)	1.40E-01
Death	15 (25)	34 (63)	**6.01E-05**

Table 4. Clinical and molecular characteristics of patients included in our dataset, classified according to our gene expression–based model. Unless otherwise noted, data are n (%). Significant p-values (< 0.05) are highlighted in bold.

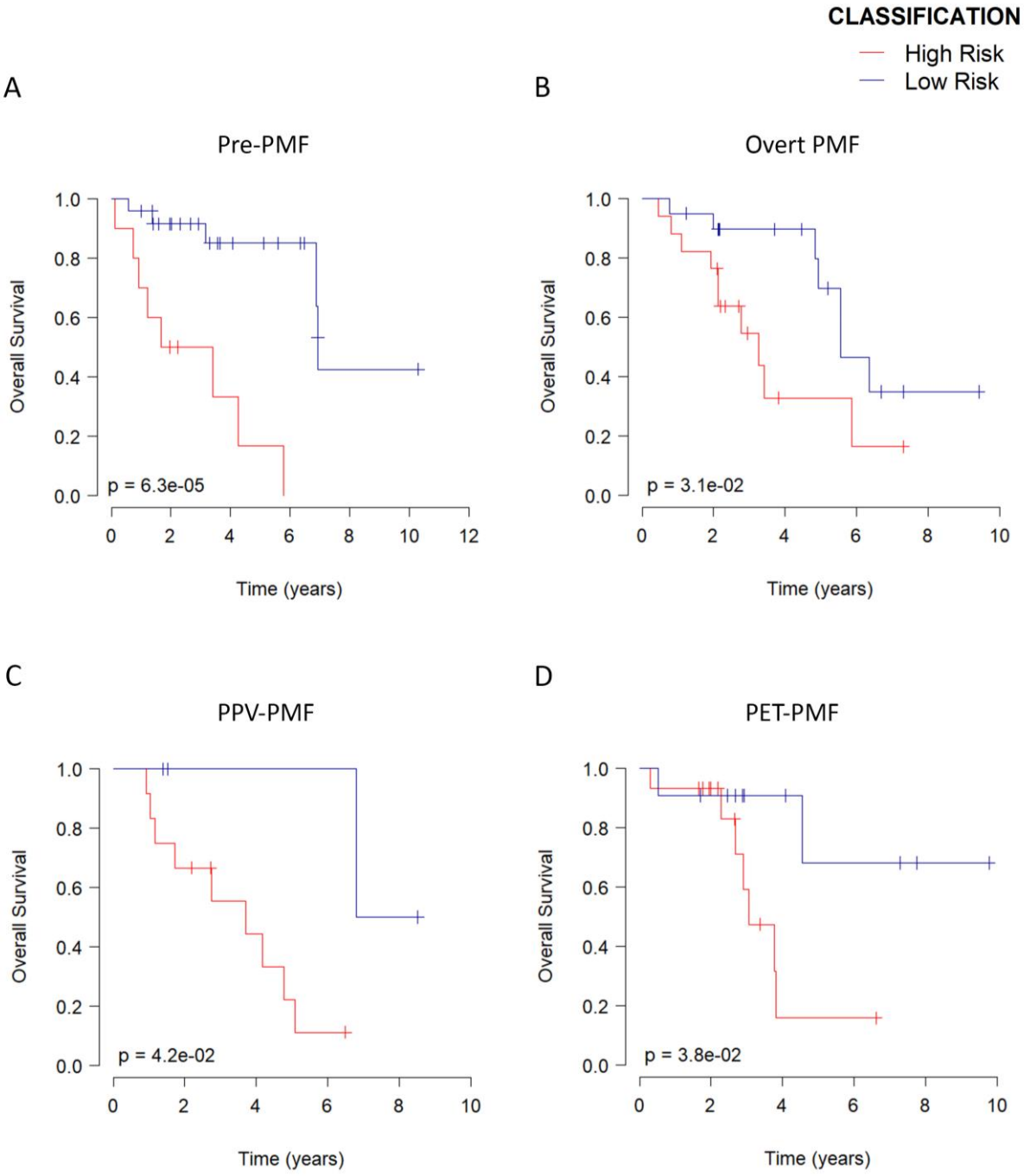

Figure 17. Overall survival of patients with different diagnosis. Each of the four Kaplan-Meier represents OS of samples classified according to our gene expression-based classifier and divided on the basis of disease diagnosis. The result of Log-Rank p-value is shown. Samples belonging to HR group display always a significant inferior OS.

4.4 Comparison with contemporary prognostic models

To analyze the correlation between GEP-based and prognostic models' classifications, we decided to consider DIPSS, MIPSS70, and MYSEC-PM scores because they were available for most of our samples. In contrast, due to the lack of cytogenetic information for most of the samples, it was not possible to determine the classification based on the most recent MIPSS70 version 2.0 prognostic score. The analysis showed that the LR group was significantly enriched in patients belonging to DIPSS Low and Intermediate-1 categories (p-value=6.00e-4, χ^2 test), while the majority of patients classified as DIPSS Intermediate-2 (16/24 [66.7%]) or High (10/12 [83.3%]) risk clustered within the gene expression-defined high-risk group (Table 4). Interestingly, DIPSS Intermediate-1 and Intermediate-2 patients belonging to the HR group according to our classification showed a significantly inferior survival compared to LR ones (p-value=5.07e-4 and p-value=2.53e-2, respectively, log-rank test) (Figure 19) and Cox-regression analysis confirmed that high-risk classification represented a risk factor for inferior survival in both DIPSS Intermediate-1 and Intermediate-2 groups (Table 5). High-risk classification retained its significance in multivariate analysis when evaluated in the context of DIPSS classification (p-value=5.69e-5) (Table 6), but also when considering risk factors included in DIPSS prognostic model (p-value=4.96e-3) (Table 7).

DIPSS	Survival	
	HR (95% CI)	p-value
LOW	<0.01	9.99e-01
INT-1	8.326 (2.1-32.7)	**2.40e-03**
INT-2	3.92 (1.1-14.1)	**3.69e-02**
HIGH	2.415 (0.3-20.1)	4.15e-01

Table 5. HR for death comparing the high-risk and low-risk groups in the context of DIPSS classification.

Survival	HR (95% CI)	p-value
DIPSS		
INT-1	7.606 (1.7-34.1)	**8.00e-03**
INT-2	10.980 (2.5-48.6)	**1.59e-03**
HIGH	22.050 (4.7-103.4)	**8.77e-05**
High Risk category	4.801 (2.2-10.3)	**5.69e-05**

Table 6. Results of multivariate analyses.

Variable	Univariable		Multivariable	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Age sample > 65	5.134 (2.5-10.5)	**6.91-e06**	4.166 (1.8-9.6)	**8.46E-04**
Hemoglobin < 10 g/dL	1.679 (0.9-3.3)	1.28E-01	0.807 (0.3-1.9)	6.33E-01
Leukocytes > 25 x 10^9/L	3.663 (2.0-6.8)	**4.76E-05**	0.990 (0.4-2.4)	9.82E-01
Circulating blasts >= 1%	4.096 (2.2-7.8)	**1.47E-05**	3.709 (1.8-7.7)	**4.16E-04**
Constitutional symptoms	2.179 (1.2-4.1)	**1.40E-02**	1.149 (0.6-2.3)	6.92E-01
High risk category	5.61 (2.8-11.1)	**6.18E-07**	4.293 (1.6-11.9)	**4.96E-03**

Table 7. Univariable and multivariable regression analysis of prognostic factors for overall survival in patients classified according to DIPSS model.

In addition, as regards SMF, we found that the distinction between HR and LR patients in terms of survival reached the statistical significance for MYSEC-PM lower-risk categories (Low plus Intermediate-1) while approached the significance for higher risk classes (Intermediate-2 plus High) (p-value=1.39e-2 and p-value=7.26e-2 respectively, log-rank test) (Figure 18).

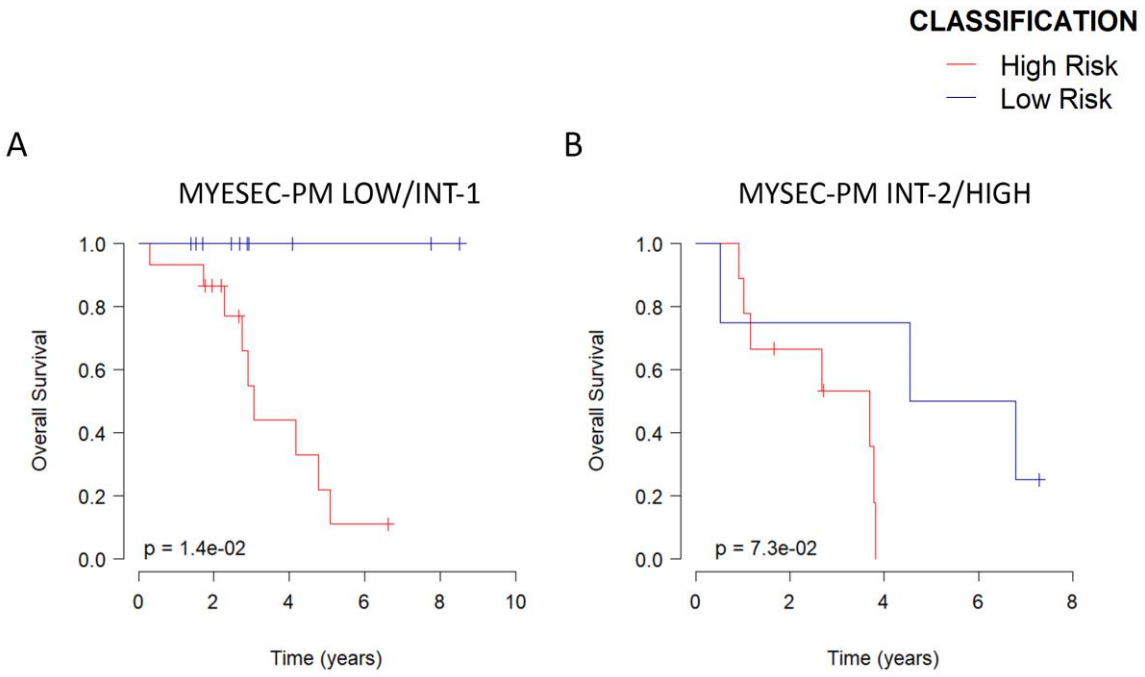

Figure 18. Overall survival of SMF patients stratified according to MYSEC-PM. Each of the two Kaplan-Meier represents OS of SMF samples classified according to our gene expression-based classifier and divided according to MYSEC-PM model. In panel A Low and Intermediate-1 (INT-1) risk categories are grouped together while in panel B Intermediate-2 (INT-2) and High risk samples are considered. The results of Log-Rank p-value are shown.

Most MIPSS70 High risk patients clustered within the HR group, while the LR group displayed higher frequencies of Low and Intermediate MIPSS70 samples (p-value=1.01e-4, χ^2 test) (Table 4). Nevertheless, gene expression-based classification distinguished high-risk patients from low-risk ones with different survival within MIPSS70 Intermediate and High categories (p-value=1.28e-2 and p-value=8.59e-3 respectively, log-rank test) (Figure 19). The high-risk classification represented a risk factor for inferior survival for patients belonging to these prognostic categories (Table 8), also retaining its significance in multivariate analysis when considering MIPSS70 classification (p-value=5.12e-4, Wald test) (Table 9), or factors included in MIPSS70 model (p-value=1.12e-4, Wald test) (Table 10).

MIPSS70	Survival	
	HR (95% CI)	p-value
LOW	1	NaN
INT	5.682 (1.3-25.7)	**2.41e-02**
HIGH	5.984 (1.3-26.6)	**1.88e-02**

Table 8. HR for death comparing the high-risk and low-risk groups in the context of MIPSS70 classification.

Survival	HR (95% CI)	p-value
MIPSS70		
INT	9.385 (1.2-75.5)	**3.53e-02**
HIGH	13.750 (1.7-110.9)	**1.39e-02**
High Risk category	5.417 (2.1-14.1)	**5.12e-04**

Table 9. Results of multivariate analyses.

Variable	Univariable		Multivariable	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Hemoglobin < 10 g/dL	1.671 (0.80-3.49)	1.72E-01	0.989 (0.3-3.2)	9.86E-01
Leukocytes > 25 x 10^9/L	3.977 (1.91-8.26)	**2.16E-04**	0.545 (0.2-1.7)	3.00E-01
Platelets < 100 10^9/L	2.7 (1.25-5.85)	**1.18E-02**	4.747 (1.6-14.0)	**4.71E-03**
Circulating blasts >= 2%	5.127 (2.27-11.57)	**8.27E-05**	2.937 (1.1-8.1)	**3.67E-02**
BM fibrosis grade >= 2	1.322 (0.63-2.76)	4.57E-01	0.992 (0.4-2.6)	9.86E-01
Constitutional symptoms	2.28 (1.16-4.48)	**1.67E-02**	1.811 (0.7-4.6)	2.12E-01
HMR	4.453 (2.21-8.97)	**2.94E-05**	1.804 (0.7-4.9)	2.47E-01
CALR Type 1 absence	1.65 (0.58-4.68)	3.46E-01	0.450 (0.1-1.5)	1.95E-01
HMR >= 2	4.503 (2.07-9.81)	**1.55E-04**	4.340 (1.3-14.0)	**1.43E-02**
High risk category	9.703 (3.95-23.83)	**7.18E-07**	12.850 (3.5-46.9)	**1.12E-04**

Table 10. Univariable and multivariable regression analysis of prognostic factors for overall survival in patients classified according to MIPSS70 model.

Taken together, these results demonstrated that a gene expression-based classifier might also identify groups of patients characterized by different

outcomes in the context of contemporary prognostic models, thus representing a useful tool to complement existing prognostic models.

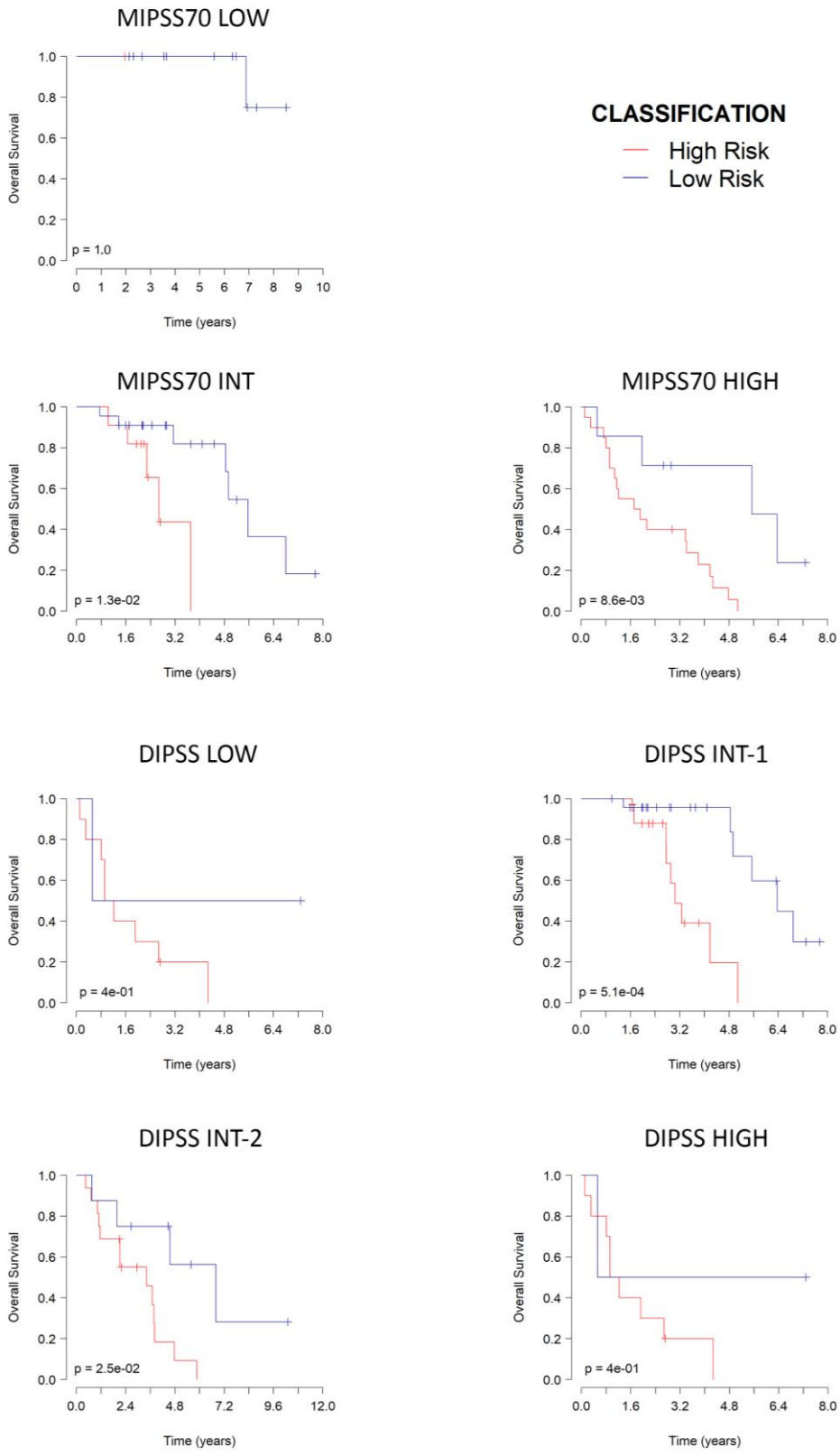

Figure 19. Survival curves for samples stratified according to MIPSS70 and DIPSS and classified in the high-risk and low-risk groups using our gene expression–based model. P-values were calculated using the log-rank test.

4.5 Patients potentially misclassified by DIPSS and MIPSS70

Of particular interest, we identified within our dataset two patients belonging to the DIPSS and MIPSS70 intermediate-2 or high-risk classes, that our model classified as low-risk. These patients are still alive after a longer period than the reported median survival for the reference prognostic class. Specifically, one patient with pre-PMF, triple-negative for driver mutations and bearing an ASXL1 mutation, was classified as intermediate-2 by DIPSS and intermediate by MIPSS70. The reported median survival for intermediate-2 DIPSS class is 4 years, and for intermediate MIPSS70 class is 7.1 years, but this patient is still alive after 10 years from diagnosis and sample collection. The second patient has a diagnosis of PET-MF and it bears an MPL mutation and an ASXL1 mutation; it was classified as high-risk by both DIPSS and MIPSS70. The reported median survival for the high-risk DIPSS class is 1.5 years, and for the high-risk MIPSS70 class is 2.3 years, but this patient is still alive after 7.3 years from sample collection and 10 years from diagnosis. Details regarding these particular cases from our dataset are summarized in Table 11.

Patient	Diagnosis	Driver Mutation	HMR	AML-t	DIPSS	DIPSS MS (y)	MIPSS70	MIPSS70 MS (y)	OS – TS (y)	OS – TD (y)
patient 1	pre-PMF	Triple Negative	ASXL1	no	INT-2	4	INT	7.1	10	10
patient 2	PET-PMF	MPL	ASXL1	no	HIGH	1.5	HIGH	2.3	7.3	10

Table 11. Low-risk patients, classified as intermediate/high risk by DIPSS and MIPSS70. HMR: high molecular risk; AML-t: acute myeloid leukemia transformation; DIPSS MS: reference DIPSS prognostic class median survival; MIPSS70 MS: reference MIPSS70 prognostic class median survival; OS – TS: overall survival – time from sample collection; OS – TD: overall survival – time from diagnosis.

Even more interesting are six samples that our model identified as high risk, but they are classified as intermediate risk by existing scores. These samples are deceased or leukemia transformed in a shorter time than the reported median survival for the reference prognostic class. Conceivably, these patients could have benefited from the treatments reserved for high-risk groups. Particularly, two

patients had a diagnosis of PPV-MF and were JAK2V67-positive (homozygous), they were classified as intermediate by DIPSS and MIPSS70, but as high risk by our model. The reported median survival for intermediate-1 and intermediate-2 DIPSS risk classes is 14.2 years and 4 years, respectively; while for intermediate MIPSS70 risk class is 7.1 years. These two samples were deceased after 1 and 2.8 years from sample collection, respectively; and after 1 and 3.7 years from diagnosis, respectively. Additionally, two patients with a diagnosis of PET-MF were classified as intermediate-1 by DIPSS and as high-risk by our model (MIPSS70 classification was not available for these samples). One of them was JAK2V617F-positive (homozygous), while the other was triple negative. They were both deceased after 3 years from sample collection and after 3.7 and 4.8 years, respectively, from diagnosis. The remaining two patients had a diagnosis of pre-PMF and overt-PMF, they were JAK2V617-positive (heterozygous) and triple-negative, respectively, and they were classified as intermediate-1 by DIPSS and intermediate by MIPSS70, but as high risk by our model. These patients were deceased after 1.7 and 3.3 years from sample collection, respectively; and after 1.7 and 4 years from diagnosis. Details regarding these particular cases from our dataset are summarized in Table 12.

Patient	Diagnosis	Driver Mutation	HMR	AML-t	DIPSS	DIPSS MS (y)	MIPSS70	MIPSS70 MS (y)	OS – TS (y)	OS – TD (y)
patient 3	PPV-MF	JAK2V617F homozygous	Wild Type	no	INT-2	4	INT	7.1	1	1
patient 4	PPV-MF	JAK2V617F homozygous	ASXL1	no	INT-1	14.2	/	/	3.7	2.8
patient 5	PET-MF	JAK2V617F homozygous	/	no	INT-1	14.2	/	/	3.7	3
patient 6	PET-MF	Triple Negative	/	no	INT-1	14.2	/	/	4.8	3
patient 7	pre-PMF	JAK2V617F heterozygous	ASXL1, EZH2	yes	INT-1	14.2	INT	7.1	1.7	1.7
patient 8	overt-PMF	Triple Negative	/	no	INT-1	14.2	/	/	4	3.3

Table 12. High-risk patients, classified as intermediate risk by DIPSS and MIPSS70. HMR: high molecular risk; AML-t: acute myeloid leukemia transformation; DIPSS MS: reference DIPSS prognostic class median survival; MIPSS70 MS: reference MIPSS70 prognostic class median survival; OS – TS: overall survival – time from sample collection; OS – TD: overall survival – time from diagnosis.

4.6 Pathway enrichment analysis

Pathway enrichment analysis was performed using the 98 genes of the model having a hazard ratio greater than 1, which represent genes that are associated with an increased risk. The same analysis was performed with two different tools: EnrichR (results are shown in Table 13) and ToppFun from the ToppGene Suite (results are shown in Table 14).

The analysis revealed an enrichment in several pathways related to general cancer development, such as DNA damage response and angiogenesis, but also some other different interesting pathways involved in cytokines regulation, inflammation, and pathways related to myeloproliferative neoplasms.

A significant enrichment is observed in pathways related to folate biosynthesis, which is essential for DNA synthesis. It has been shown that the stem cell-derived clonal proliferation and the increased cell turnover, that characterizes chronic myeloproliferative disorders, may lead to a progressive depletion of the biochemical factors involved in cell proliferation, in particular those involved in DNA synthesis, such as folate (Vener *et al.*, 2010).

Our analysis shows that also the BMP signaling is enriched, indeed bone morphogenetic proteins have been found to be overexpressed in the bone marrow of PMF and it has been suggested that BMP family members may play an important role in the pathogenesis of myelofibrosis in PMF (Bock *et al.*, 2008).

Another element highlighted by the enrichment analysis is the Hippo pathway. It is well known that this pathway regulates cell proliferation and apoptosis and its deregulation is involved in cancer development. Interestingly, the Hippo pathway has been shown to be involved in the development of lethal myelofibrosis in murine models. Indeed, Stoner and colleagues showed that inactivation of heterozygous Hippo kinase in a JAK2-V617F murine model led to accelerated development of lethal myelofibrosis, mimicking what happens when MPN progress towards myelofibrotic transformation (Stoner *et al.*, 2019).

The analysis revealed also enrichment in genes involved in signaling events mediated by VEGF receptors, which are possibly involved in the pathophysiology

of myeloproliferative neoplasms. Indeed, VEGF/VEGFR pathways are involved in the proliferation, migration, and survival of cells (Medinger and Passweg, 2014). Moreover, a VEGF/VEGFR-1 autocrine loop has been hypothesized to occur in the neoplastic cells of Ph- MPNs. This hypothesis was supported by the observation of co-localization and increased levels of VEGFR-1 and VEGF in megakaryocytes, macrophages ad myeloid precursors of Ph- MPNs (Boiocchi *et al.*, 2011). VEGF, along with other growth factors, pro-inflammatory cytokines, and extracellular matrix components, has been shown to be secreted by aberrant immature megakaryocytes, which are a quintessential feature of myelofibrosis (Gangat and Tefferi, 2020).

Different pathways related to interleukin signaling are found to be enriched; of particular interest is IL-12 whose levels have been shown to have a prognostic value in primary myelofibrosis (Tefferi *et al.*, 2011). IL-1 should also be noted, as it is one of the master regulators of the inflammatory state and it has been implicated in various pathological diseases including MPN (Rai *et al.*, 2019). Moreover, it has been observed that IL-1β is secreted by JAK2 mutant hematopoietic stem cells (HSC), causing apoptosis of mesenchymal cells, thus affecting the survival of normal HSC and favoring mutant HSC expansion over normal HSC (Gangat and Tefferi, 2020).

The platelet-derived growth factor receptor beta (PDGFRβ) signaling pathway is another noteworthy element of the list of enriched pathways. Constitutive activation of PDGFR has been acknowledged as one of the features of myeloproliferative diseases (Cross and Reiter, 2002; Tokita *et al.*, 2007) and the upregulation of the PDGF pathway has been indicated as a hallmark of myelofibrosis. Bone marrow stromal cells of myelofibrosis patients show increased expression of PDGFRβ, which correlated with the grade of myelofibrosis (Bedekovics *et al.*, 2013). Kramer et al. further investigated the precise role of PDGFRβ signaling in myelofibrosis exploiting a Gata-1^{low} mouse model for myelofibrosis. They observed increased PDGFRβ and PDGF-B protein expression in the overt fibrotic bone marrow and analyzed the modifications that occur during the initiation and progression of myelofibrosis. Findings reported by Kramer et al. suggest that increased PDGFR expression could be exploited as

a useful early biomarker for myelofibrosis (Kramer *et al.*, 2020; Marneth and Mullally, 2020).

Another interesting pathway that is highlighted by this enrichment analysis is the MAP kinase pathway, which, along with NF-kB, plays an important role in cytokine overproduction in myelofibrosis and for this reason, optimal control of inflammatory pathophysiology in myelofibrosis may be achieved with the inhibition of these pathways (Fisher *et al.*, 2019). In addition, mutations in this pathway showed association with survival outcomes in myelofibrosis (Coltro *et al.*, 2020).

Lastly, the mTOR pathway is found to be enriched. The PI3K/AKT/mTOR pathway is frequently activated in human cancers and it contributes to cell growth, proliferation, survival, apoptosis, autophagy, and angiogenesis (Guertin and Sabatini, 2007; Engelman, 2009). It is also involved in inflammation and pathogenesis of MPNs (Bartalucci, Guglielmelli and Vannucchi, 2013) and it has been suggested that the mTOR pathway may be a clinically relevant target in myelofibrosis (Guglielmelli, Barosi, *et al.*, 2011).

Enriched Terms	p-value	odds ratio	genes
Hippo-Merlin Signaling Dysregulation WP4541	3.11E-04	9.25	ITGAM;INSR;CTNNA1;ITGAV;KRAS
MAPK Cascade WP422	3.80E-04	24.14	JUN;KRAS;MAP2K6
Signal transduction through IL1R WP4496	5.59E-04	20.92	JUN;IRAK3;MAP2K6
VEGFA-VEGFR2 Signaling Pathway WP3888	1.08E-03	5.58	KL;JUN;GRB10;CTNNA1;ITGAV;MAP2K6
Bone Morphogenic Protein (BMP) Signaling and Regulation WP1425	1.52E-03	41.44	SMAD1;BMP2
BMP2-WNT4-FOXO1 Pathway in Human Primary Endometrial Stromal Cell Differentiation WP3876	1.79E-03	37.67	SMAD1;BMP2
MFAP5-mediated ovarian cancer cell motility and invasiveness WP3301	1.79E-03	37.67	JUN;ITGAV
DNA Damage Response (only ATM dependent) WP710	2.10E-03	7.95	JUN;INSR;KRAS;LDLR
IL-1 signaling pathway WP195	2.49E-03	12.05	JUN;IRAK3;MAP2K6
ESC Pluripotency Pathways WP3931	2.55E-03	7.52	SMAD1;JUN;FGF13;MAP2K6
Oncostatin M Signaling Pathway WP2374	4.00E-03	10.11	TIMP3;KRAS;LDLR
Folate Metabolism WP176	4.18E-03	9.94	DHFR;INSR;LDLR

BMP Signaling Pathway in Eyelid Development WP3927	4.26E-03	23.01	SMAD1;JUN
Angiogenesis WP1539	6.11E-03	18.83	SMAD1;TIMP3
Differentiation of white and brown adipocyte WP2895	6.62E-03	18.01	SMAD1;BMP2
Insulin Signaling WP481	7.96E-03	5.39	JUN;INSR;GRB10;MAP2K6
Selenium Micronutrient Network WP15	8.70E-03	7.54	KYNU;INSR;LDLR
PI3K-AKT-mTOR signaling pathway and therapeutic opportunities WP3844	9.45E-03	14.79	GRB10;KRAS
TNF alpha Signaling Pathway WP231	1.05E-02	7.03	JUN;KRAS;MAP2K6
White fat cell differentiation WP4149	1.07E-02	13.80	KLF5;CTNNA1

Table 13. Pathways associated with model's genes having HR>1. Analysis performed with EnrichR.

Enriched terms	p-value	genes
Folate and Pterine Biosynthesis	1.13E-05	GCH1,GGH,DHFR
Folate biosynthesis	2.21E-05	GCH1,GGH,DHFR
MAP00790 Folate biosynthesis	2.94E-05	GCH1,GGH,DHFR
Bone morphogenetic proteins signaling	2.70E-04	BMP2,SMAD1
Tetrahydrofolate biosynthesis	2.70E-04	GCH1,DHFR
Hippo-Merlin Signaling Dysregulation	4.53E-04	KRAS,INSR,ITGAM,ITGAV,CTNNA1
Signaling events mediated by VEGFR1 and VEGFR2	4.61E-04	GRB10,ITGAV,CTNNA1,MAP2K6
Signal transduction through IL1R	5.59E-04	JUN,IRAK3,MAP2K6
PDGFR-beta signaling pathway	5.63E-04	KRAS,PAG1,GRB10,ITGAV,JUN
IL12 signaling mediated by STAT4	6.15E-04	PIAS2,IL18RAP,JUN
MAPK Cascade	6.74E-04	KRAS,JUN,MAP2K6
IL1-mediated signaling events	7.36E-04	JUN,IRAK3,MAP2K6
Signal transduction through IL1R	8.02E-04	JUN,IRAK3,MAP2K6

Table 14. Pathways associated with model's genes having HR>1. Analysis performed with ToppFun from the ToppGene Suite.

4.7 Evaluation of integrated models

Two new prognostic models were designed, as detailed in the materials and methods section, by integrating information coming from our gene expression-based classification within DIPSS and MIPSS70. These new models were named MODIFIED DIPSS and MODIFIED MIPSS70, respectively. The Akaike information criteria (AIC) was 313.2 for DIPSS and 303.6 for MODIFIED DIPSS (Figure 20). Similarly, AIC values were 229.5 for MIPSS70 and 226.6 for

MODIFIED MIPSS70 (Figure 21). As a smaller AIC value means a better fit for the model (Portet, 2020), these AIC values suggested that both combined models outperformed DIPSS and MIPSS70 classification.

It must be acknowledged that lower AIC values shown by integrated models may be due to some extent to the effect of overfitting. Nevertheless, our observations suggest that gene expression can potentially represent a new level of information with prognostic value and that it can be integrated with existing scores. Obviously, validation with an external independent dataset has to be performed in order to confirm the results and to think for a real clinical application.

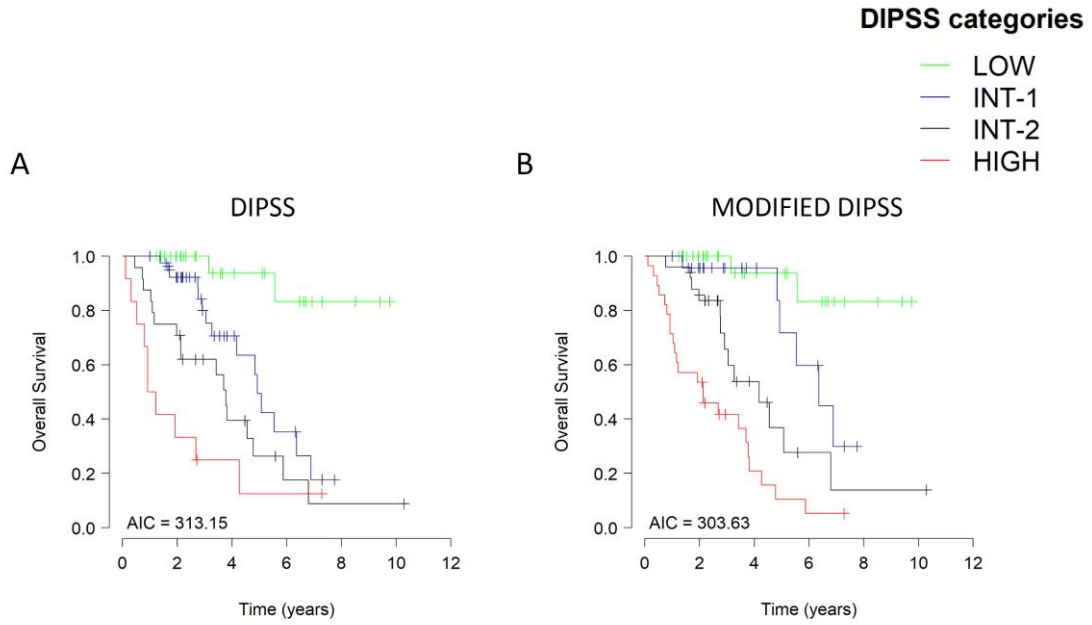

Figure 20. Accuracy of integrated DIPSS prognostic model. Kaplan Meier curves represents the overall survival of patients stratified according to (A) DIPSS and (B) integrated classification (Modified DIPSS). Akaike information criteria (AIC) was determined in order to compare prediction accuracy of the different models represented.

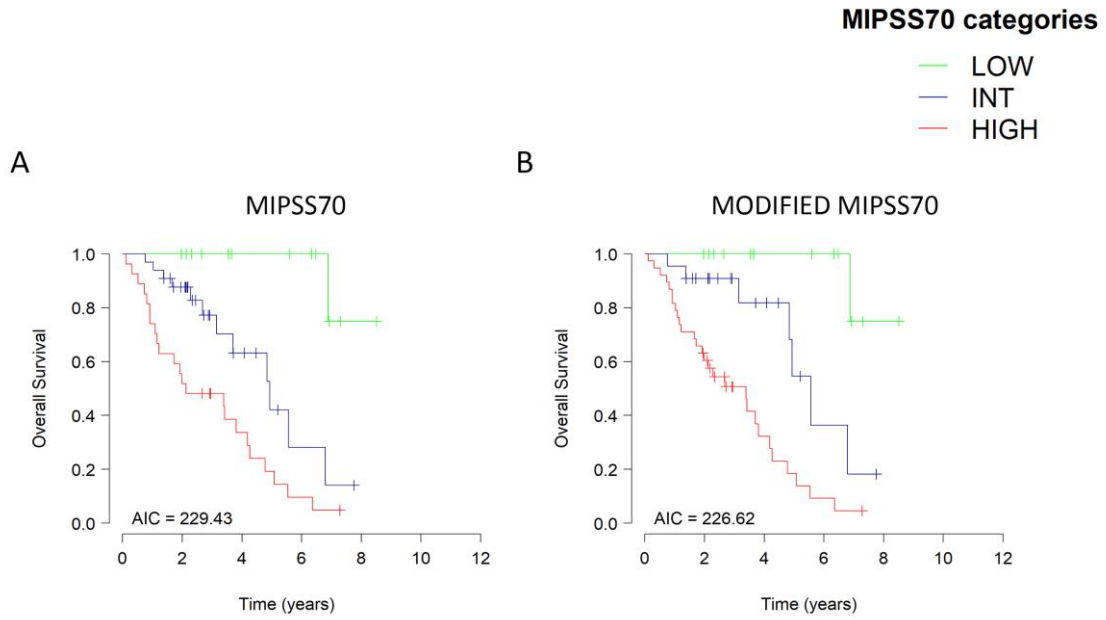

Figure 21. Accuracy of integrated MIPSS70 prognostic model. Kaplan Meier curves represents the overall survival of patients stratified according to (A) MIPSS70 and (B) integrated classification (Modified MIPSS70). Akaike information criteria (AIC) was determined in order to compare prediction accuracy of the different models represented.

5 Discussion

Myelofibrosis is a complex hematologic disorder arising from the mutation of hematopoietic stem cells. Excessive proliferation of cells from the neoplastic clones gives rise to granulocyte and megakaryocyte hyperplasia, while the altered interaction between hematopoietic and stromal cells in the bone marrow microenvironment leads to the development of fibrosis that is the hallmark of the disease (Gangat and Tefferi, 2020). Like PV and ET, PMF originates from the acquisition by hematopoietic stem/progenitor cells of somatic driver mutations in JAK2, MPL, or CALR genes, even if almost 5% of patients do not harbor any of these and, therefore, are considered triple-negative (Rumi and Cazzola, 2017). Several other mutations might be present in PMF patients, among them those occurring in EZH2, ASXL1, SRSF2, and IDH1/2 genes are termed "High Molecular Risk" due to their negative impact on overall survival and leukemia-free survival (Mills *et al.*, 2009).

Myelofibrosis can be either primary or secondary to PV and ET, moreover, a pre-fibrotic stage and an overt fibrotic stage must be distinguished within PMF depending on BM fibrosis degree and represent distinct entities. In agreement with the clinical features of patients included in our dataset, PET-MF was characterized by increased platelet counts (Rotunno *et al.*, 2016; Palandri *et al.*, 2018) that was accompanied by higher hemoglobin levels and inferior frequency of splenomegaly in pre-PMF (Guglielmelli *et al.*, 2017; Mudireddy *et al.*, 2018). Increased hemoglobin was evident also in PPV-MF which was characterized also by increased leukocyte count and by the presence of JAK2V617F mutation in the totality of patients (Rotunno *et al.*, 2016; Palandri *et al.*, 2018). Despite these

differences, primary and secondary myelofibrosis cases are currently managed in the same way (Ayalew Tefferi, Guglielmelli, Pardanani, *et al.*, 2018).

To study whether gene expression might provide clinical and prognostic information at any time during the disease course we included in our dataset both samples at diagnosis and during follow-up and correlated granulocytes gene expression profile with patient's features at that time point. Many prognostic models were developed in the last decade allowing risk stratification in MF patients based on clinical features (i.e. IPSS, DIPSS), but also on molecular and cytogenetic characteristics (i.e. DIPSS-plus, MIPSS70, MIPSS70+v2.0, GIPSS) (Cervantes *et al.*, 2009; Passamonti *et al.*, 2010; Gangat *et al.*, 2011; Ayalew Tefferi, Guglielmelli, Lasho, *et al.*, 2018; Guglielmelli, Terra L. Lasho, *et al.*, 2018). These models were developed for prognostication in PMF but are applied also in SMF cases even if a specific model was recently developed (MYSEC-PM) (Passamonti *et al.*, 2017). More recently, Grinfeld et al. underlined the impact of driver mutations on patients' prognosis by defining a classification scheme for MPNs and related disorders based on the type of genomic alterations harbored by patients (Grinfeld *et al.*, 2018).

Despite recent advances, allogenic stem cell transplantation remains the only curative option for MF, but due to the high rate of transplant-related death or severe morbidities, guidelines suggest considering it only for patients with an expected survival lower than 5 years. Conversely, current drug therapies do not modify the natural history of the disease and are intended to treat MF-related symptoms. According to the different prognostic models, allogenic stem cell transplantation is suggested only for the higher risk classes while, for other patients, observation only is the preferred choice with the introduction of drug therapies in the presence of symptoms (Gangat and Tefferi, 2020).

Since clinical decision-making in MF is mainly influenced by survival prediction, it is fundamental to develop easy-to-use models that can identify patients at the highest risk of death. Gene expression analysis was recently adopted to predict the risk of recurrence in breast cancer (Güler, 2017), but gene expression signatures were developed also to predict survival in hematopoietic malignancies such as AML (i.e. LSC-17) (Ng *et al.*, 2016). Therefore, we decided to study the

gene expression profile of MF granulocytes to evaluate its impact on disease phenotype and patients' outcomes. We focused on granulocytes since they represent an easily accessible myeloid cell population that belongs to the neoplastic clone, unlike peripheral blood mononuclear cells that comprise also lymphoid elements (Mead and Mullally, 2017).

As described in section 4.1, our dataset included 114 patients with a diagnosis of PMF or SMF coming from 5 centers in Italy. In section 4.2 we identified survival-related transcripts by means of cox regression analysis, which were subsequently used to construct a classifier based on the expression of 201 genes that could identify two groups of patients. The classifier was built using the "nearest shrunken centroids" supervised learning technique, and the number of genes was optimized by subsequent iterations. To overcome the issue of the absence of an external independent validation set, the model was validated exploiting 20-fold cross-validation.

High-risk patients identified by our model displayed an inferior overall survival and leukemia-free survival compared with low-risk cases. Our results demonstrated that gene expression-based classification showed a good agreement with contemporary prognostic models, indeed high-risk classification correlated with the presence of several detrimental features, such as advanced age, decreased hemoglobin levels (<10 g/dL), and platelet count (<100 x 109/L), increased WBC count (>25 x 109/L), circulating blasts ≥1% and 2%, presence of constitutional symptoms and splenomegaly. Interestingly, in a recent report by Penna et al. (Penna *et al.*, 2019), the absence of the same detrimental features was correlated to long survival (>20 years) in patients with primary myelofibrosis; this condition was similar to that observed in the low-risk group identified by our gene expression signature. The frequency of patients with at least one HMR mutation was increased in the high-risk group, the presence of these variants correlates with MF diagnosis and more aggressive disease with inferior overall survival (Grinfeld *et al.*, 2018). Moreover, our findings underlined the impact of JAK2V617F allele burden on overall survival, since the high-risk group showed an increased frequency of homozygous patients while heterozygous ones clustered within the low-risk group. These data are in

agreement with Grinfeld et al. which demonstrated that patients with a JAK2V617F heterozygous disease have favorable outcomes (Grinfeld *et al.*, 2018).

To further evaluate the gene expression-based classification model performance, we contextualized results in section 4.4, integrating available information relating to contemporary prognostic models. While 82.5% of patients in the low-risk category were classified as Low or Intermediate-1 risk according to DIPSS, only 52% of high-risk patients came from DIPSS Intermediate-2 and High classes, thus indicating that in our model 48% of high-risk patients had been upgraded from lowest DIPSS categories. The same is true for MIPSS70 classification, again 83% of low-risk patients belonged to MIPSS70 Low and Intermediate risk classes while only 62.5% of high-risk cases belonged to MIPSS70 High-risk category.

Our gene expression-based classifier was able to distinguish high-risk patients from low-risk ones within intermediate risk classes. Indeed, in patients stratified according to DIPSS, belonging to the high-risk group represented a risk factor for inferior survival in both Intermediate-1 and Intermediate-2 categories. The high-risk group was characterized by significantly inferior OS compared with the low-risk group: 3.05 years (high-risk) vs 6.36 years (low-risk), and 3.42 years (high-risk) vs 6.79 years (low-risk) in the Intermediate-1 and Intermediate-2 classes, respectively. Likewise, within the MIPSS70 Intermediate group, high-risk patients displayed a median overall survival equal to 2.68 years, significantly lower compared to the low-risk group (5.56 years). It is worth noting that intermediate-risk classes represent the most challenging patients' categories, for whom determining the optimal therapeutic strategy is more difficult. Furthermore, the multivariable analysis demonstrated that gene expression-based classification might represent a risk factor for inferior survival independent from both DIPSS and MIPSS70 classification and factors included in these models. Therefore, this analysis suggests that gene expression analysis might provide additional information other than those included in contemporary prognostic models, and it might improve risk stratification in myelofibrosis.

This was further supported by the observation, described in section 4.7, that incorporating our gene expression-based classification in the context of DIPSS

and MIPSS70 allowed a better distinction between risk categories. Gene expression-based prognostic evaluation in breast cancer patients has recently come into clinical use thanks to the development of tests based on real-time qRT-PCR (OncotypeDX), microarray (MammaPrint), and NanoString (Prosigna) technologies (Güler, 2017). The latter was developed starting from a microarray-based test, PAM50, to identify breast cancer molecular subtypes (Parker *et al.*, 2009). It should be emphasized that PAM50 is based on a precursory version of the nearest shrunken centroids, the linear classification machine learning algorithm used for this thesis work. Specifically, as detailed in section 1.2, PAM50 exploits the nearest centroids algorithm, which involves first summarizing the dataset into a set of centroids (without shrinking them), then using the centroids to make predictions for new samples. The extension to this method was developed by Tibshirani et al., giving rise to the nearest shrunken centroids, which involves shifting class-based centroids toward the overall centroid of the entire dataset (Tibshirani *et al.*, 2002). By means of the NanoString platform, Ng and colleagues were able to validate the prognostic value of the LSC17 gene signature in AML (Ng *et al.*, 2016). This technique might be of particular interest for the validation of our signature since it allows the evaluation of a high number of targets and has a rapid turnaround time (24h-48h).

Additionally, the pathway enrichment analysis described in section 4.6, helped to better elucidate the biological meaning of the genes included in our classification model. The analysis revealed an enrichment in several pathways related to general cancer development, such as DNA damage response and angiogenesis, but also some other different interesting pathways involved in cytokines regulation, inflammation, and pathways related to myeloproliferative neoplasms.

As a whole, this study demonstrates the correlation between GEP and MF clinical features and provides the proof of concept that gene expression analysis should be considered to complement risk stratification in MF, thus allowing a more effective clinical management of patients. To date, treatment algorithms suggest for DIPSS Intermediate-1 patients observation-only accompanied by palliative drug therapy against symptoms, while ASCT and enrollment in clinical trials must be considered for patients belonging to Intermediate-2 and High risk categories (Tefferi, 2018). In section 4.5 we highlighted that HR patients with an

expected median survival inferior to 5 years might also be identified within DIPSS Intermediate-1 and 2 classes; since they are characterized by an expected median survival inferior to 5 years they might benefit from treatment options reserved to DIPSS High risk category. The same is true when we consider Intermediate risk patients according to MIPSS70 classification.

One of the main limitations of our study is given by the limited number of samples included in our dataset. For this reason, we have to consider all the myelofibrosis subtypes together and we cannot build a single model for each subtype, or at least two separate models for primary and secondary myelofibrosis. Furthermore, we do not have a proper validation set. Although the k-fold cross-validation that we performed helps to overcome this issue providing a certain level of confidence, it is clear that a large independent validation set would be the best option to validate our model. As a matter of fact, as reported by Kourou et al. (Kourou *et al.*, 2015), the lack of external validation is a rather common problem in studies regarding cancer prediction and prognosis based on machine-learning methods. In addition, it has to be considered that myelofibrosis is a rare condition and RNA is a perishable material, making it difficult to collect large numbers of samples. Notwithstanding, it has been reported that the application of machine-learning methods can improve the accuracy of cancer susceptibility, recurrence, and survival prediction (Cruz and Wishart, 2007; Kourou *et al.*, 2015).

Overall, we believe that, given the robustness of these analyses, these data may still be helpful to the hematological community to uncover the possibility that gene expression profiles might integrate contemporary prognostic models. Further studies might be able to prospectively validate this evidence in a larger cohort of myelofibrosis cases, thus improving the identification of patients with expected inferior survival that can benefit from clinical trials participation or ASCT.

6 Appendix

Table 15. The list of 832 probe sets whose expression is related to survival resulting from Cox regression analysis. Genes with p-value < 0.005 were selected.

Probe set ID	Gene Symbol	Gene Description	Hazard Ratio	p-value
11760334_at	EMG1	EMG1 N1-specific pseudouridine methyltransferase	8.31E-04	2.16E-03
11731897_a_at	PTHLH	parathyroid hormone-like hormone	1.98E-03	1.04E-03
11735432_a_at	OBP2B	odorant binding protein 2B	2.31E-03	1.40E-03
11759789_x_at	NLGN4Y	neuroligin 4, Y-linked	4.13E-03	2.41E-03
11724453_a_at	KCNJ1	potassium channel, inwardly rectifying subfamily J, member 1	9.71E-03	2.41E-03
11749111_a_at	MTTP	microsomal triglyceride transfer protein	1.03E-02	1.75E-03
11730240_a_at	SLC2A12	solute carrier family 2 (facilitated glucose transporter), member 12	1.18E-02	2.41E-03
11738213_a_at	SLC17A3	solute carrier family 17 (organic anion transporter), member 3	1.25E-02	2.43E-03
11750388_a_at	HSF5	heat shock transcription factor family member 5	1.98E-02	2.06E-03
11741742_at	LEUTX	leucine twenty homeobox	2.12E-02	4.54E-03
11764026_at	PPARA	peroxisome proliferator-activated receptor alpha	2.32E-02	1.58E-03
11728353_at	MSMP	microseminoprotein, prostate associated	2.37E-02	1.72E-03
11738100_a_at	C20orf195	chromosome 20 open reading frame 195	2.39E-02	2.02E-04
11761028_at	GTF3C5	general transcription factor IIIC subunit 5	2.42E-02	4.26E-03
11748669_x_at	PDCD2	programmed cell death 2	2.42E-02	6.27E-04
11736943_at	WFIKKN1	WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain containing 1	2.50E-02	3.40E-03
11738011_a_at	MYLK2	myosin light chain kinase 2	2.51E-02	4.56E-03
11762515_at	ODF2L	outer dense fiber of sperm tails 2-like	2.65E-02	4.29E-03
11721124_s_at	MMP11	matrix metallopeptidase 11	2.75E-02	4.46E-03
11722968_a_at	TRPV4	transient receptor potential cation channel, subfamily V, member 4	2.75E-02	3.01E-04
11726223_a_at	FBLN7	fibulin 7	2.98E-02	2.49E-03

11726673_s_at	ASCL1	achaete-scute family bHLH transcription factor 1	2.99E-02	2.74E-03
11727836_a_at	CPZ /// GPR78	carboxypeptidase Z /// G protein-coupled receptor 78	3.02E-02	3.01E-03
11737930_a_at	FSHB	follicle stimulating hormone, beta polypeptide	3.16E-02	4.24E-03
11747904_a_at	PIP5K1C	phosphatidylinositol-4-phosphate 5-kinase, type I, gamma	3.20E-02	3.20E-03
11725802_a_at	USP46	ubiquitin specific peptidase 46	3.32E-02	9.86E-04
11725578_at	BEND3	BEN domain containing 3	3.48E-02	2.77E-03
11734561_a_at	WBSCR17	Williams-Beuren syndrome chromosome region 17	3.67E-02	1.48E-03
11730124_s_at	ADAMTSL2	ADAMTS like 2	3.68E-02	3.56E-03
11734424_at	GPR26	G protein-coupled receptor 26	4.23E-02	9.26E-05
11751169_a_at	PGAP3	post-GPI attachment to proteins 3	4.26E-02	1.62E-03
11719076_a_at	REEP1	receptor accessory protein 1	4.27E-02	1.96E-03
11723872_a_at	ADGRG1	adhesion G protein-coupled receptor G1	4.43E-02	4.65E-03
11761131_at	NOL12	nucleolar protein 12	4.56E-02	3.38E-04
11733741_x_at	SYMPK	symplekin	4.60E-02	9.28E-04
11739766_at	PNPLA3	patatin-like phospholipase domain containing 3	4.78E-02	2.86E-03
11729557_s_at	EYA2	EYA transcriptional coactivator and phosphatase 2	4.87E-02	3.52E-03
11746153_a_at	VWA3B	von Willebrand factor A domain containing 3B	4.88E-02	2.62E-03
11744652_a_at	HAPLN2	hyaluronan and proteoglycan link protein 2	4.91E-02	1.08E-03
11746108_a_at	MPP2	membrane protein, palmitoylated 2	4.92E-02	2.72E-03
11729919_a_at	ADAMTS18	ADAM metallopeptidase with thrombospondin type 1 motif 18	5.15E-02	1.96E-03
11739914_a_at	CRELD1	cysteine rich with EGF-like domains 1	5.31E-02	3.07E-03
11755591_x_at	MSTO1	misato 1, mitochondrial distribution and morphology regulator	5.35E-02	2.93E-03
11722534_at	RPRM	reprimo, TP53 dependent G2 arrest mediator candidate	5.44E-02	3.97E-03
11741162_a_at	SLC12A4	solute carrier family 12 (potassium/chloride transporter), member 4	5.68E-02	1.29E-03
11725634_at	MRPS25	mitochondrial ribosomal protein S25	5.75E-02	4.39E-03
11740922_a_at	C17orf80	chromosome 17 open reading frame 80	5.79E-02	3.89E-03
11758321_s_at	GATSL2	GATS protein-like 2	5.93E-02	4.16E-03
11745606_a_at	LRRC18	leucine rich repeat containing 18	6.04E-02	4.07E-03
11747480_a_at	CENPT	centromere protein T	6.59E-02	1.03E-03
11740185_at	KCNJ4	potassium channel, inwardly rectifying subfamily J, member 4	6.68E-02	4.03E-03
11734991_a_at	WNT3A	wingless-type MMTV integration site family, member 3A	6.78E-02	2.67E-03
11732007_at	GLIS1	GLIS family zinc finger 1	6.83E-02	4.96E-03
11751058_a_at	OTX1	orthodenticle homeobox 1	6.85E-02	3.17E-03
11721459_a_at	TAPBPL	TAP binding protein-like	6.86E-02	2.02E-04
11748115_a_at	PKNOX2	PBX/knotted 1 homeobox 2	6.90E-02	4.08E-03
11716965_a_at	ATN1	atrophin 1	7.41E-02	3.99E-03

11762432_a_at	WDR31	WD repeat domain 31	7.57E-02	2.15E-03
11738483_a_at	TNFRSF13B	tumor necrosis factor receptor superfamily, member 13B	7.71E-02	1.57E-03
11757004_a_at	KCTD10	potassium channel tetramerization domain containing 10	7.78E-02	2.93E-03
11737917_at	PDLIM2	PDZ and LIM domain 2 (mystique)	8.07E-02	6.41E-04
11721983_a_at	PID1	phosphotyrosine interaction domain containing 1	8.21E-02	3.50E-03
11746304_a_at	NOP14	NOP14 nucleolar protein	8.41E-02	3.25E-03
11726571_a_at	SEMA7A	semaphorin 7A, GPI membrane anchor (John Milton Hagen blood group)	8.48E-02	3.88E-03
11717112_a_at	REPIN1	replication initiator 1	8.58E-02	1.96E-03
11721770_x_at	CCDC64	coiled-coil domain containing 64	8.80E-02	3.06E-03
11717402_s_at	ARF6	ADP-ribosylation factor 6	8.81E-02	2.78E-04
11715346_at	EBI3	Epstein-Barr virus induced 3	8.81E-02	2.62E-03
11733630_at	LRRC10	leucine rich repeat containing 10	8.84E-02	4.63E-03
11723612_a_at	TOX2	TOX high mobility group box family member 2	9.00E-02	5.12E-04
11720673_a_at	PPM1A	protein phosphatase, Mg2+/Mn2+ dependent, 1A	9.18E-02	1.93E-03
11722494_s_at	HIP1R	huntingtin interacting protein 1 related	9.59E-02	1.04E-03
11751711_s_at	SLC48A1	solute carrier family 48 (heme transporter), member 1	9.86E-02	1.46E-03
11738032_x_at	FAM153C	family with sequence similarity 153, member C, pseudogene	9.88E-02	1.68E-03
11762096_at	KCNQ2	potassium channel, voltage gated KQT-like subfamily Q, member 2	1.02E-01	2.81E-03
11735821_s_at	AQP12A /// AQP12B	aquaporin 12A /// aquaporin 12B	1.05E-01	4.54E-03
11739406_x_at	FPGS	folylpolyglutamate synthase	1.07E-01	2.16E-03
11726510_a_at	SMPD3	sphingomyelin phosphodiesterase 3, neutral membrane (neutral sphingomyelinase II)	1.11E-01	2.07E-03
11726458_s_at	GSPT2	G1 to S phase transition 2	1.12E-01	3.17E-03
11727004_s_at	KXD1	KxDL motif containing 1	1.12E-01	1.81E-03
11758905_x_at	TMED4	transmembrane p24 trafficking protein 4	1.12E-01	4.72E-04
11756999_a_at	CYB561D2	cytochrome b561 family, member D2	1.14E-01	3.44E-03
11739667_at	GALNT6	polypeptide N-acetylgalactosaminyltransferase 6	1.15E-01	1.50E-03
11746529_x_at	TNFRSF14	tumor necrosis factor receptor superfamily, member 14	1.21E-01	5.17E-04
11763429_at	MOK	MOK protein kinase	1.22E-01	3.44E-03
11742634_a_at	PABPN1L	poly(A) binding protein, nuclear 1-like (cytoplasmic)	1.28E-01	3.75E-03
11727282_x_at	ADAMTS7	ADAM metallopeptidase with thrombospondin type 1 motif 7	1.29E-01	3.83E-03
11734812_at	TMEM86A	transmembrane protein 86A	1.32E-01	3.42E-03
11722374_a_at	GGA2	golgi-associated, gamma adaptin ear containing, ARF binding protein 2	1.32E-01	2.08E-03
11726353_at	CD180	CD180 molecule	1.32E-01	2.74E-03

11754747_x_at	QTRT1	queuine tRNA-ribosyltransferase 1	1.41E-01	3.38E-03
11717401_at	ARF6	ADP-ribosylation factor 6	1.46E-01	2.07E-03
11757321_a_at	TAPBP	TAP binding protein (tapasin)	1.53E-01	1.51E-03
11727656_s_at	EPHB1	EPH receptor B1	1.53E-01	1.50E-04
11715524_a_at	COTL1	coactosin-like F-actin binding protein 1	1.58E-01	1.51E-03
11725960_s_at	CALM2 /// CALM3	calmodulin 2 (phosphorylase kinase, delta) /// calmodulin 3 (phosphorylase kinase, delt	1.62E-01	3.66E-03
11721669_a_at	KLHDC4 /// LOC105371397	kelch domain containing 4 /// uncharacterized LOC105371397	1.65E-01	1.72E-03
11736090_a_at	OLIG2	oligodendrocyte lineage transcription factor 2	1.76E-01	2.90E-04
11759223_a_at	EDF1	endothelial differentiation-related factor 1	1.77E-01	1.79E-03
11735370_a_at	IQSEC2	IQ motif and Sec7 domain 2	1.85E-01	3.70E-03
11744315_at	SGK223	homolog of rat pragma of Rnd2	1.85E-01	7.12E-04
11717508_at	IRF4	interferon regulatory factor 4	1.88E-01	2.48E-03
11745772_x_at	TNFRSF14	tumor necrosis factor receptor superfamily, member 14	1.90E-01	2.13E-03
11743194_x_at	TNFRSF14	tumor necrosis factor receptor superfamily, member 14	1.95E-01	1.22E-03
11728523_a_at	MYCL	v-myc avian myelocytomatosis viral oncogene lung carcinoma derived homolog	2.02E-01	3.34E-03
11728576_a_at	PYROXD2	pyridine nucleotide-disulphide oxidoreductase domain 2	2.05E-01	2.31E-03
11763978_a_at	SRRT	serrate, RNA effector molecule	2.08E-01	1.06E-03
11731578_s_at	ACSS1	acyl-CoA synthetase short-chain family member 1	2.12E-01	1.57E-03
11736260_a_at	SSPN	sarcospan	2.17E-01	4.76E-03
11716273_x_at	POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	2.24E-01	6.40E-05
11716272_a_at	POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	2.26E-01	7.14E-04
11750008_a_at	POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	2.28E-01	4.96E-04
11724933_a_at	FAM173A	family with sequence similarity 173, member A	2.34E-01	4.09E-03
11720327_a_at	POLR2C	polymerase (RNA) II (DNA directed) polypeptide C, 33kDa	2.35E-01	5.03E-04
11721460_s_at	TAPBPL	TAP binding protein-like	2.37E-01	7.64E-06
11745894_x_at	TNFRSF14	tumor necrosis factor receptor superfamily, member 14	2.41E-01	8.06E-04
11756172_x_at	TBC1D22A	TBC1 domain family, member 22A	2.50E-01	4.67E-03
11735695_a_at	ARF5	ADP-ribosylation factor 5	2.57E-01	1.13E-03
11741419_x_at	KCNQ1	potassium channel, voltage gated KQT-like subfamily Q, member 1	2.63E-01	3.11E-03
11726165_at	ZNF518B	zinc finger protein 518B	2.65E-01	7.14E-04
11718305_a_at	PDLIM2	PDZ and LIM domain 2 (mystique)	2.66E-01	6.30E-04
11754156_x_at	ARF5	ADP-ribosylation factor 5	2.75E-01	1.57E-03
11744872_x_at	ARF5	ADP-ribosylation factor 5	2.75E-01	1.25E-03

11755617_s_at	EPHB1	EPH receptor B1	2.81E-01	4.50E-03
11746363_x_at	SYNGR2	synaptogyrin 2	2.82E-01	1.75E-03
11718889_x_at	PIP5K1A	phosphatidylinositol-4-phosphate 5-kinase, type I, alpha	2.85E-01	1.38E-03
11742938_at	ASCL2	achaete-scute family bHLH transcription factor 2	2.90E-01	1.87E-03
11750009_x_at	POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	2.92E-01	2.63E-04
11754609_x_at	STMN3	stathmin-like 3	2.92E-01	3.53E-03
11743434_a_at	CHST11	carbohydrate (chondroitin 4) sulfotransferase 11	2.99E-01	6.75E-05
11744190_a_at	SPNS3	spinster homolog 3 (Drosophila)	3.05E-01	3.26E-03
11734503_x_at	ALOX15	arachidonate 15-lipoxygenase	3.08E-01	3.83E-03
11750605_a_at	ATF6B /// TNXB	activating transcription factor 6 beta /// tenascin XB	3.13E-01	3.06E-03
11744067_s_at	SYNGR2	synaptogyrin 2	3.14E-01	1.49E-03
11744871_a_at	ARF5	ADP-ribosylation factor 5	3.25E-01	2.49E-03
11730422_at	FLVCR1	feline leukemia virus subgroup C cellular receptor 1	3.25E-01	2.00E-04
11720226_s_at	TBC1D1	TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 1	3.26E-01	2.23E-03
11746076_a_at	PRAP1 /// ZNF511 /// ZNF511-PRAP1	proline-rich acidic protein 1 /// zinc finger protein 511 /// ZNF511-PRAP1 readthrough	3.26E-01	3.48E-03
11744191_x_at	SPNS3	spinster homolog 3 (Drosophila)	3.31E-01	4.59E-03
11742985_at	LY9	lymphocyte antigen 9	3.33E-01	2.85E-03
11736188_a_at	ORMDL3	ORMDL sphingolipid biosynthesis regulator 3	3.37E-01	3.70E-03
11744176_at	PPP1R9B	protein phosphatase 1, regulatory subunit 9B	3.37E-01	9.25E-04
11715439_at	DAP	death-associated protein	3.39E-01	4.87E-03
11727657_at	EPHB1	EPH receptor B1	3.49E-01	1.66E-05
11755873_a_at	ACSS1	acyl-CoA synthetase short-chain family member 1	3.53E-01	3.45E-03
11718456_at	SLC43A2	solute carrier family 43 (amino acid system L transporter), member 2	3.53E-01	6.85E-04
11739658_a_at	LTB	lymphotoxin beta (TNF superfamily, member 3)	3.84E-01	1.22E-03
11716511_x_at	SRRM2	serine/arginine repetitive matrix 2	3.84E-01	4.91E-03
11739657_a_at	LTB	lymphotoxin beta (TNF superfamily, member 3)	3.89E-01	2.67E-03
11717588_at	CLPP	caseinolytic mitochondrial matrix peptidase proteolytic subunit	3.99E-01	4.95E-03
11715807_a_at	STARD10	StAR-related lipid transfer domain containing 10	4.03E-01	3.43E-04
11725746_a_at	PITPNC1	phosphatidylinositol transfer protein, cytoplasmic 1	4.10E-01	1.40E-03
11723325_a_at	KCNQ1	potassium channel, voltage gated KQT-like subfamily Q, member 1	4.16E-01	3.71E-03
11763447_x_at	TRDC	T cell receptor delta constant	4.28E-01	1.98E-03

11715727_s_at	OAZ2	ornithine decarboxylase antizyme 2	4.50E-01	4.31E-04
11717162_a_at	SLC29A1	solute carrier family 29 (equilibrative nucleoside transporter), member 1	4.53E-01	4.26E-03
11758214_s_at	SRRM2	serine/arginine repetitive matrix 2	4.60E-01	2.64E-03
11740244_a_at	PHOSPHO1	phosphatase, orphan 1	4.62E-01	5.22E-04
11725186_x_at	KLF12	Kruppel-like factor 12	4.64E-01	4.53E-03
11719228_a_at	CAMK1	calcium/calmodulin-dependent protein kinase I	4.67E-01	1.56E-03
11738435_x_at	CD74	CD74 molecule, major histocompatibility complex, class II invariant chain	4.84E-01	2.83E-03
11717661_a_at	PPP1R16B	protein phosphatase 1, regulatory subunit 16B	4.86E-01	3.28E-03
11737845_x_at	FAM102A	family with sequence similarity 102, member A	5.05E-01	4.19E-03
11726254_s_at	CD74	CD74 molecule, major histocompatibility complex, class II invariant chain	5.12E-01	4.19E-03
11740554_a_at	PHOSPHO1	phosphatase, orphan 1	5.18E-01	9.90E-04
11715726_a_at	OAZ2	ornithine decarboxylase antizyme 2	5.22E-01	4.62E-03
11720225_a_at	TBC1D1	TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 1	5.22E-01	4.59E-03
11731422_s_at	FCGR3A	Fc fragment of IgG, low affinity IIIa, receptor (CD16a)	5.37E-01	4.03E-04
11722635_at	IL2RB	interleukin 2 receptor, beta	5.38E-01	3.30E-03
11726255_x_at	CD74	CD74 molecule, major histocompatibility complex, class II invariant chain	5.41E-01	4.40E-03
11715394_s_at	CD81	CD81 molecule	5.41E-01	3.76E-04
11722403_a_at	PLEKHO1	pleckstrin homology domain containing, family O member 1	5.62E-01	5.83E-05
11732275_at	CCL5	chemokine (C-C motif) ligand 5	5.68E-01	2.76E-04
11731941_at	PRSS33	protease, serine, 33	5.84E-01	2.54E-03
11754313_s_at	ARL4C	ADP-ribosylation factor like GTPase 4C	5.93E-01	6.36E-04
11732276_x_at	CCL5	chemokine (C-C motif) ligand 5	6.00E-01	1.61E-04
11722681_at	RBP7	retinol binding protein 7, cellular	6.02E-01	5.00E-03
11732331_s_at	ARL4C	ADP-ribosylation factor like GTPase 4C	6.05E-01	1.86E-03
11753810_a_at	CCL5	chemokine (C-C motif) ligand 5	6.29E-01	1.71E-04
11726046_s_at	FCGR2B	Fc fragment of IgG, low affinity IIb, receptor (CD32)	6.39E-01	4.29E-03
11733736_a_at	CD2	CD2 molecule	6.44E-01	4.28E-03
11741517_s_at	FCGR2B	Fc fragment of IgG, low affinity IIb, receptor (CD32)	6.56E-01	4.93E-03
11727609_at	KLRB1	killer cell lectin-like receptor subfamily B, member 1	6.58E-01	2.66E-04
11741899_s_at	FCGR2B	Fc fragment of IgG, low affinity IIb, receptor (CD32)	6.60E-01	4.61E-03
11762318_x_at	IL23A	interleukin 23, alpha subunit p19	6.64E-01	3.01E-03
11758555_s_at	GPR183	G protein-coupled receptor 183	6.68E-01	4.39E-03
11728560_at	GZMK	granzyme K	6.80E-01	1.80E-03
11755180_x_at	TCF7	transcription factor 7 (T-cell specific, HMG-box)	6.84E-01	1.61E-03

11738124_x_at	SIGLEC10	sialic acid binding Ig-like lectin 10	6.87E-01	3.37E-03
11741190_a_at	P2RY10	purinergic receptor P2Y, G-protein coupled, 10	6.92E-01	4.90E-03
11761918_x_at	TRBC1 /// TRBV19	T cell receptor beta constant 1 /// T cell receptor beta variable 19	6.96E-01	4.57E-03
11722379_at	GNG11	guanine nucleotide binding protein (G protein), gamma 11	7.03E-01	2.28E-03
11763233_x_at	TRAC /// TRAJ17 /// TRAV20 /// TRDV2	T-cell receptor alpha constant /// T cell receptor alpha joining 17 /// T cell receptor	7.23E-01	3.17E-03
11719120_a_at	KYNU	kynureninase	1.26E+00	4.20E-03
11718394_at	JUN	jun proto-oncogene	1.36E+00	4.66E-03
11741830_s_at	RFPL4A /// RFPL4AL1	ret finger protein-like 4A /// ret finger protein-like 4A-like 1	1.38E+00	3.95E-03
11735945_x_at	UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	1.38E+00	3.71E-03
11718397_s_at	JUN	jun proto-oncogene	1.40E+00	1.32E-03
11718395_s_at	JUN	jun proto-oncogene	1.40E+00	8.18E-04
11739844_at	UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	1.40E+00	3.93E-03
11738959_s_at	CARD17	caspase recruitment domain family, member 17	1.40E+00	5.90E-04
11718396_x_at	JUN	jun proto-oncogene	1.44E+00	6.24E-04
11726814_x_at	KDM5D	lysine (K)-specific demethylase 5D	1.44E+00	3.20E-03
11737166_at	ANKRD34B	ankyrin repeat domain 34B	1.46E+00	5.20E-04
11720029_a_at	LDLR	low density lipoprotein receptor	1.46E+00	2.77E-03
11737167_at	ANKRD34B	ankyrin repeat domain 34B	1.49E+00	8.26E-05
11733646_x_at	KYNU	kynureninase	1.49E+00	4.66E-03
11723209_s_at	KBTBD6	kelch repeat and BTB (POZ) domain containing 6	1.54E+00	2.80E-03
11718998_x_at	DHFR	dihydrofolate reductase	1.56E+00	4.54E-03
11724178_a_at	DAAM2	dishevelled associated activator of morphogenesis 2	1.58E+00	1.34E-04
11749598_a_at	IL18RAP	interleukin 18 receptor accessory protein	1.58E+00	4.07E-03
11729424_s_at	CCRL2	chemokine (C-C motif) receptor-like 2	1.59E+00	1.77E-03
11732084_a_at	TFEC	transcription factor EC	1.59E+00	6.09E-04
11727171_at	TRPS1	trichorhinophalangeal syndrome I	1.59E+00	2.97E-03
11721733_a_at	GCH1	GTP cyclohydrolase 1	1.60E+00	1.44E-03
11743416_s_at	C5orf30	chromosome 5 open reading frame 30	1.60E+00	1.22E-04
11743497_at	BMP2	bone morphogenetic protein 2	1.61E+00	4.97E-04
11718152_a_at	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)	1.63E+00	2.37E-03
11755369_a_at	KDM5D	lysine (K)-specific demethylase 5D	1.64E+00	1.33E-03
11743498_at	BMP2	bone morphogenetic protein 2	1.65E+00	4.31E-04
11752282_a_at	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)	1.66E+00	1.02E-03
11738958_at	CARD17	caspase recruitment domain family, member 17	1.68E+00	1.79E-03

11721216_s_at	TMEM106B	transmembrane protein 106B	1.68E+00	4.94E-03
11721654_at	AGPAT4	1-acylglycerol-3-phosphate O-acyltransferase 4	1.69E+00	4.90E-03
11741990_s_at	CCRL2	chemokine (C-C motif) receptor-like 2	1.69E+00	3.24E-03
11747134_a_at	TRPS1	trichorhinophalangeal syndrome I	1.72E+00	1.38E-03
11763837_s_at	UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	1.75E+00	2.37E-03
11756285_s_at	IGF2BP3	insulin-like growth factor 2 mRNA binding protein 3	1.75E+00	5.73E-04
11721734_s_at	GCH1	GTP cyclohydrolase 1	1.76E+00	6.25E-04
11728571_a_at	PLCL1	phospholipase C-like 1	1.77E+00	1.38E-03
11738960_x_at	CARD17	caspase recruitment domain family, member 17	1.78E+00	3.93E-04
11717861_a_at	EGR1	early growth response 1	1.79E+00	2.49E-03
11719000_x_at	DHFR	dihydrofolate reductase	1.79E+00	4.05E-03
11725294_at	USP9Y	ubiquitin specific peptidase 9, Y-linked	1.79E+00	1.91E-04
11731430_a_at	CYP27A1	cytochrome P450, family 27, subfamily A, polypeptide 1	1.80E+00	2.10E-03
11723044_at	SMAD1	SMAD family member 1	1.80E+00	1.12E-05
11733413_a_at	SLC26A6	solute carrier family 26 (anion exchanger), member 6	1.80E+00	2.07E-03
11759902_at	HACD4	3-hydroxyacyl-CoA dehydratase 4	1.81E+00	1.56E-03
11726895_a_at	IRAK3	interleukin 1 receptor associated kinase 3	1.81E+00	6.17E-04
11718056_a_at	BTBD3	BTB (POZ) domain containing 3	1.81E+00	4.15E-03
11729722_a_at	PIAS2	protein inhibitor of activated STAT 2	1.81E+00	9.96E-05
11749782_x_at	BCAT1	branched chain amino-acid transaminase 1, cytosolic	1.82E+00	1.21E-03
11724606_a_at	MAP2K6	mitogen-activated protein kinase kinase 6	1.82E+00	3.49E-03
11745012_a_at	KDM5D	lysine (K)-specific demethylase 5D	1.82E+00	2.20E-03
11726896_a_at	IRAK3	interleukin 1 receptor associated kinase 3	1.82E+00	3.07E-04
11740702_a_at	HMGB2	high mobility group box 2	1.83E+00	1.64E-03
11739711_a_at	PHACTR2	phosphatase and actin regulator 2	1.84E+00	1.53E-03
11747333_a_at	HSD17B4	hydroxysteroid (17-beta) dehydrogenase 4	1.84E+00	4.33E-03
11747501_a_at	SMAD1	SMAD family member 1	1.85E+00	3.66E-05
11722818_a_at	GGH	gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)	1.86E+00	1.68E-05
11725691_a_at	VWA5A	von Willebrand factor A domain containing 5A	1.86E+00	1.30E-03
11724376_at	PAG1	phosphoprotein membrane anchor with glycosphingolipid microdomains 1	1.86E+00	2.06E-03
11750599_a_at	OSBPL6	oxysterol binding protein-like 6	1.87E+00	2.39E-04
11725603_a_at	HOXA10-HOXA9 /// HOXA9	HOXA10-HOXA9 readthrough /// homeobox A9	1.88E+00	1.30E-03
11755017_a_at	CHCHD7	coiled-coil-helix-coiled-coil-helix domain containing 7	1.89E+00	4.12E-03
11744572_a_at	KLF5	Kruppel-like factor 5 (intestinal)	1.89E+00	3.79E-04
11751292_a_at	PIAS2	protein inhibitor of activated STAT 2	1.89E+00	1.91E-05

11748095_a_at	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)	1.91E+00	2.01E-04
11720716_a_at	FGF13	fibroblast growth factor 13	1.92E+00	1.62E-03
11721215_a_at	TMEM106B	transmembrane protein 106B	1.92E+00	2.12E-03
11752193_a_at	RGL4	ral guanine nucleotide dissociation stimulator-like 4	1.93E+00	4.63E-03
11732299_at	KIAA1715	KIAA1715	1.93E+00	1.13E-03
11744985_a_at	FNDC3A	fibronectin type III domain containing 3A	1.93E+00	4.45E-03
11757461_s_at	BTBD3	BTB (POZ) domain containing 3	1.94E+00	1.03E-03
11741924_a_at	GRB10	growth factor receptor bound protein 10	1.94E+00	3.11E-03
11738049_a_at	ATP2C2	ATPase, Ca++ transporting, type 2C, member 2	1.95E+00	8.11E-04
11759085_s_at	KIAA1715	KIAA1715	1.95E+00	3.65E-04
11731541_at	SESTD1	SEC14 and spectrin domains 1	1.96E+00	1.19E-04
11731542_at	SESTD1	SEC14 and spectrin domains 1	1.96E+00	2.67E-05
11726894_a_at	IRAK3	interleukin 1 receptor associated kinase 3	1.96E+00	2.08E-04
11744581_a_at	KHDRBS3	KH domain containing, RNA binding, signal transduction associated 3	1.97E+00	5.53E-04
11738392_at	LIPN	lipase, family member N	1.97E+00	6.59E-05
11730313_a_at	ERMAP	erythroblast membrane-associated protein (Scianna blood group)	1.98E+00	1.87E-03
11723092_at	FNIP2	folliculin interacting protein 2	1.98E+00	8.64E-07
11718026_a_at	NCOA7	nuclear receptor coactivator 7	1.98E+00	2.77E-03
11731291_a_at	PIAS2	protein inhibitor of activated STAT 2	1.99E+00	5.23E-05
11742017_a_at	IRAK3	interleukin 1 receptor associated kinase 3	2.00E+00	8.20E-05
11750007_a_at	S100Z	S100 calcium binding protein Z	2.00E+00	2.51E-04
11748859_a_at	TFEC	transcription factor EC	2.01E+00	1.38E-03
11717631_s_at	KLHL9	kelch-like family member 9	2.01E+00	1.40E-03
11733954_at	SLC22A15	solute carrier family 22, member 15	2.02E+00	4.10E-03
11758315_s_at	PER2	period circadian clock 2	2.04E+00	1.79E-03
11749728_a_at	KIAA1715	KIAA1715	2.04E+00	7.32E-04
11729941_at	TMEM56	transmembrane protein 56	2.04E+00	1.21E-03
11728361_a_at	CHCHD7	coiled-coil-helix-coiled-coil-helix domain containing 7	2.04E+00	4.58E-03
11758115_s_at	PTBP2	polypyrimidine tract binding protein 2	2.05E+00	4.31E-03
11740045_a_at	PHACTR2	phosphatase and actin regulator 2	2.05E+00	1.16E-03
11722977_at	HOXB5	homeobox B5	2.05E+00	6.57E-05
11759585_at	RAB18	RAB18, member RAS oncogene family	2.06E+00	4.13E-03
11743640_a_at	FNDC3A	fibronectin type III domain containing 3A	2.08E+00	3.25E-03
11729243_s_at	FBN2	fibrillin 2	2.08E+00	1.10E-05
11731539_at	SESTD1	SEC14 and spectrin domains 1	2.09E+00	9.85E-05
11723091_s_at	FNIP2	folliculin interacting protein 2	2.10E+00	1.95E-06
11733477_at	SUCNR1	succinate receptor 1	2.11E+00	1.21E-04
11742035_a_at	PLCL1	phospholipase C-like 1	2.12E+00	5.76E-04
11746816_a_at	PIAS2	protein inhibitor of activated STAT 2	2.13E+00	1.41E-04
11729827_at	FAM110B	family with sequence similarity 110, member B	2.13E+00	4.16E-04
11741441_a_at	FGF13	fibroblast growth factor 13	2.13E+00	1.33E-03

11727403_at	GPC4	glypican 4	2.13E+00	3.05E-04
11745030_at	LINC00597	long intergenic non-protein coding RNA 597	2.14E+00	6.19E-04
11743881_s_at	MREG	melanoregulin	2.14E+00	2.10E-03
11755694_a_at	AGFG1	ArfGAP with FG repeats 1	2.16E+00	2.47E-03
11731605_s_at	HSD17B12	hydroxysteroid (17-beta) dehydrogenase 12	2.16E+00	9.48E-04
11727397_s_at	CTNNA1	catenin (cadherin-associated protein), alpha 1	2.16E+00	3.11E-03
11755851_a_at	UBR4	ubiquitin protein ligase E3 component n-recognin 4	2.16E+00	4.57E-03
11748385_a_at	KIAA0430	KIAA0430	2.17E+00	3.92E-03
11740018_a_at	INSR	insulin receptor	2.18E+00	1.05E-03
11758391_s_at	ZADH2	zinc binding alcohol dehydrogenase domain containing 2	2.18E+00	1.12E-03
11722062_at	REEP3	receptor accessory protein 3	2.18E+00	2.18E-03
11764018_s_at	CLCN3	chloride channel, voltage-sensitive 3	2.20E+00	4.47E-03
11719179_a_at	PVR	poliovirus receptor	2.22E+00	3.52E-03
11759429_a_at	KLF7	Kruppel-like factor 7 (ubiquitous)	2.22E+00	3.86E-03
11733955_at	SLC22A15	solute carrier family 22, member 15	2.22E+00	4.37E-03
11732481_a_at	ITGAM	integrin, alpha M (complement component 3 receptor 3 subunit)	2.22E+00	3.09E-03
11726661_s_at	GPN3	GPN-loop GTPase 3	2.24E+00	3.06E-03
11721532_x_at	EIF4G3	eukaryotic translation initiation factor 4 gamma, 3	2.25E+00	3.71E-03
11717963_a_at	KRAS	Kirsten rat sarcoma viral oncogene homolog	2.25E+00	3.37E-03
11748797_a_at	AGPAT4	1-acylglycerol-3-phosphate O-acyltransferase 4	2.26E+00	4.39E-03
11732297_at	KIAA1715	KIAA1715	2.26E+00	1.13E-03
11749780_a_at	BCAT1	branched chain amino-acid transaminase 1, cytosolic	2.26E+00	5.18E-04
11719334_a_at	RCBTB1	regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 1	2.27E+00	5.68E-04
11730513_a_at	LSMEM1	leucine-rich single-pass membrane protein 1	2.30E+00	2.73E-03
11748211_a_at	OSBPL9	oxysterol binding protein-like 9	2.30E+00	4.77E-03
11740173_a_at	SOCS5	suppressor of cytokine signaling 5	2.30E+00	2.94E-03
11749895_a_at	BCAT1	branched chain amino-acid transaminase 1, cytosolic	2.32E+00	1.77E-03
11732318_a_at	KHDRBS3	KH domain containing, RNA binding, signal transduction associated 3	2.33E+00	5.77E-04
11731543_at	SESTD1	SEC14 and spectrin domains 1	2.33E+00	4.43E-03
11729210_at	ORC2	origin recognition complex subunit 2	2.35E+00	1.81E-03
11729320_a_at	SOCS5	suppressor of cytokine signaling 5	2.36E+00	9.26E-04
11716847_a_at	SLC43A3	solute carrier family 43, member 3	2.37E+00	4.27E-04
11754805_a_at	SPIDR	scaffolding protein involved in DNA repair	2.37E+00	2.65E-03
11753440_x_at	GBA	glucosidase, beta, acid	2.38E+00	7.57E-04
11750386_s_at	CTNNA1	catenin (cadherin-associated protein), alpha 1	2.38E+00	9.72E-04
11760189_at	ZSCAN30	zinc finger and SCAN domain containing 30	2.38E+00	2.07E-03
11758251_s_at	HOXB3	homeobox B3	2.38E+00	8.73E-04
11757515_s_at	ITGAV	integrin alpha V	2.38E+00	4.73E-03

11725295_s_at	USP9Y	ubiquitin specific peptidase 9, Y-linked	2.38E+00	1.19E-05
11734230_a_at	LOC145783 /// ZNF280D	uncharacterized LOC145783 /// zinc finger protein 280D	2.39E+00	3.86E-03
11741582_a_at	SLC3A2	solute carrier family 3 (amino acid transporter heavy chain), member 2	2.39E+00	1.03E-03
11755585_a_at	AZI2	5-azacytidine induced 2	2.39E+00	1.44E-03
11754951_a_at	NCOA7	nuclear receptor coactivator 7	2.40E+00	2.24E-03
11736111_a_at	ARHGAP18	Rho GTPase activating protein 18	2.42E+00	2.10E-03
11757777_s_at	KL	klotho	2.42E+00	9.92E-06
11729857_at	H6PD	hexose-6-phosphate dehydrogenase (glucose 1-dehydrogenase)	2.43E+00	4.43E-03
11715435_s_at	TIMP3	TIMP metallopeptidase inhibitor 3	2.43E+00	3.68E-03
11720580_a_at	TCF12	transcription factor 12	2.43E+00	4.94E-03
11719963_a_at	HOMER3	homer scaffolding protein 3	2.44E+00	7.04E-05
11736112_a_at	ARHGAP18	Rho GTPase activating protein 18	2.44E+00	1.02E-03
11716187_a_at	ABCB6 /// ATG9A	ATP binding cassette subfamily B member 6 (Langereis blood group) /// autophagy related	2.44E+00	2.17E-03
11743090_a_at	SLC36A4	solute carrier family 36 (proton/amino acid symporter), member 4	2.45E+00	7.78E-05
11718818_s_at	SLC26A6	solute carrier family 26 (anion exchanger), member 6	2.45E+00	1.10E-03
11731449_s_at	SEC62	SEC62 homolog, preprotein translocation factor	2.45E+00	4.91E-03
11753435_x_at	SLC3A2	solute carrier family 3 (amino acid transporter heavy chain), member 2	2.45E+00	1.21E-04
11753439_a_at	GBA /// GBAP1	glucosidase, beta, acid /// glucosidase, beta, acid pseudogene 1	2.45E+00	6.59E-04
11749112_a_at	RB1	retinoblastoma 1	2.46E+00	4.90E-03
11745231_a_at	TCAIM	T cell activation inhibitor, mitochondrial	2.47E+00	4.82E-03
11753615_a_at	ARL8B	ADP-ribosylation factor like GTPase 8B	2.47E+00	1.00E-03
11739771_a_at	OXR1	oxidation resistance 1	2.48E+00	4.92E-03
11756435_a_at	GLMP	glycosylated lysosomal membrane protein	2.48E+00	3.70E-03
11725293_at	USP9Y	ubiquitin specific peptidase 9, Y-linked	2.49E+00	7.20E-05
11722588_a_at	BCAP29	B-cell receptor-associated protein 29	2.49E+00	1.67E-04
11729242_at	FBN2	fibrillin 2	2.50E+00	2.21E-04
11753434_a_at	SLC3A2	solute carrier family 3 (amino acid transporter heavy chain), member 2	2.50E+00	4.72E-04
11744076_a_at	OLAH	oleoyl-ACP hydrolase	2.52E+00	2.48E-04
11722482_x_at	MFF	mitochondrial fission factor	2.52E+00	2.34E-03
11754529_x_at	SLC3A2	solute carrier family 3 (amino acid transporter heavy chain), member 2	2.52E+00	3.83E-03
11741149_x_at	SLC3A2	solute carrier family 3 (amino acid transporter heavy chain), member 2	2.52E+00	2.21E-03
11749673_x_at	CTNNA1	catenin (cadherin-associated protein), alpha 1	2.52E+00	1.53E-03
11733071_a_at	SLC3A2	solute carrier family 3 (amino acid transporter heavy chain), member 2	2.53E+00	1.41E-04
11737747_a_at	GMFB	glia maturation factor, beta	2.54E+00	1.47E-03
11730054_s_at	SNX16	sorting nexin 16	2.54E+00	1.35E-03

11718999_x_at	DHFR	dihydrofolate reductase	2.55E+00	3.46E-03
11738455_a_at	C20orf197	chromosome 20 open reading frame 197	2.55E+00	2.48E-03
11739772_s_at	OXR1	oxidation resistance 1	2.56E+00	2.95E-03
11744588_a_at	CCDC17	coiled-coil domain containing 17	2.57E+00	1.52E-03
11742774_a_at	CCDC112	coiled-coil domain containing 112	2.57E+00	2.84E-03
11746697_a_at	SLC3A2	solute carrier family 3 (amino acid transporter heavy chain), member 2	2.58E+00	5.46E-04
11763856_a_at	HOMER3	homer scaffolding protein 3	2.58E+00	1.69E-05
11763707_at	CCDC93	coiled-coil domain containing 93	2.59E+00	4.10E-03
11759932_at	CLK4	CDC like kinase 4	2.59E+00	8.99E-04
11723756_at	ZFR	zinc finger RNA binding protein	2.60E+00	4.18E-03
11750239_a_at	VPS13B	vacuolar protein sorting 13 homolog B (yeast)	2.60E+00	3.00E-03
11736668_at	CLEC4E	C-type lectin domain family 4, member E	2.60E+00	2.09E-03
11718491_x_at	FABP5	fatty acid binding protein 5 (psoriasis-associated)	2.60E+00	2.38E-03
11741276_s_at	TMEM144	transmembrane protein 144	2.60E+00	1.21E-03
11732350_a_at	CCNA1	cyclin A1	2.60E+00	3.02E-03
11754811_x_at	CTNNA1	catenin (cadherin-associated protein), alpha 1	2.60E+00	1.48E-03
11753763_x_at	CDKN3	cyclin-dependent kinase inhibitor 3	2.61E+00	1.95E-03
11723315_a_at	CAPN3	calpain 3	2.61E+00	1.94E-03
11753616_s_at	ARL8B	ADP-ribosylation factor like GTPase 8B	2.61E+00	1.58E-04
11739710_a_at	PHACTR2	phosphatase and actin regulator 2	2.61E+00	1.04E-04
11747308_a_at	VTA1	vesicle (multivesicular body) trafficking 1	2.61E+00	3.10E-03
11759340_at	DENND1B	DENN/MADD domain containing 1B	2.62E+00	3.48E-04
11727396_a_at	CTNNA1	catenin (cadherin-associated protein), alpha 1	2.62E+00	5.48E-04
11730052_a_at	SNX16	sorting nexin 16	2.62E+00	1.88E-03
11739368_a_at	DMXL2	Dmx-like 2	2.63E+00	2.57E-05
11758934_x_at	RAB27A	RAB27A, member RAS oncogene family	2.63E+00	4.32E-03
11754771_s_at	LMBR1	limb development membrane protein 1	2.63E+00	1.15E-03
11724378_s_at	PAG1	phosphoprotein membrane anchor with glycosphingolipid microdomains 1	2.63E+00	4.69E-04
11761460_s_at	CBWD1 /// CBWD2 /// CBWD3 /// CBWD5 /// CBWD6 /// CBWD7	COBW domain containing 1 /// COBW domain containing 2 /// COBW domain containing 3 ///	2.64E+00	8.15E-04
11737782_a_at	PROSC	proline synthetase co-transcribed homolog (bacterial)	2.65E+00	4.99E-03
11726539_a_at	OSBPL9	oxysterol binding protein-like 9	2.65E+00	4.07E-03
11720067_a_at	APH1B	APH1B gamma secretase subunit	2.65E+00	3.89E-03
11739351_a_at	AMOT	angiomotin	2.66E+00	7.53E-05
11718490_s_at	FABP5 /// FABP5P2	fatty acid binding protein 5 (psoriasis-associated) /// fatty acid binding protein 5 ps	2.67E+00	1.06E-03
11735156_a_at	ZNF585A	zinc finger protein 585A	2.67E+00	7.08E-04
11736801_a_at	PLAG1	pleiomorphic adenoma gene 1	2.67E+00	6.21E-07
11763359_at	DOC2B	double C2-like domains, beta	2.69E+00	3.38E-03

11732697_s_at	COL24A1	collagen, type XXIV, alpha 1	2.70E+00	3.53E-03
11738932_x_at	RPS4Y2	ribosomal protein S4, Y-linked 2	2.70E+00	3.71E-03
11750881_a_at	LMBR1	limb development membrane protein 1	2.70E+00	3.84E-03
11733422_a_at	LRRFIP2	leucine rich repeat (in FLII) interacting protein 2	2.71E+00	4.30E-03
11725485_at	DIRC2	disrupted in renal carcinoma 2	2.71E+00	1.91E-03
11736793_a_at	NXT2	nuclear transport factor 2-like export factor 2	2.72E+00	3.27E-04
11750207_a_at	HSD17B12	hydroxysteroid (17-beta) dehydrogenase 12	2.72E+00	9.12E-05
11718974_at	LPL	lipoprotein lipase	2.73E+00	1.29E-04
11719460_s_at	WDR41	WD repeat domain 41	2.73E+00	4.40E-04
11743288_at	PEAK1	pseudopodium-enriched atypical kinase 1	2.74E+00	1.68E-03
11729813_a_at	ZADH2	zinc binding alcohol dehydrogenase domain containing 2	2.74E+00	3.94E-03
11727719_a_at	BCAP29	B-cell receptor-associated protein 29	2.76E+00	1.23E-03
11748561_a_at	DAAM2	dishevelled associated activator of morphogenesis 2	2.77E+00	2.28E-03
11762112_a_at	HLA-F	major histocompatibility complex, class I, F	2.77E+00	2.80E-04
11758284_s_at	KCTD9	potassium channel tetramerization domain containing 9	2.77E+00	4.94E-03
11752453_a_at	GPD2	glycerol-3-phosphate dehydrogenase 2	2.77E+00	1.05E-03
11716844_s_at	DNAJC13	DnaJ (Hsp40) homolog, subfamily C, member 13	2.77E+00	1.87E-03
11744221_a_at	ASUN	asunder spermatogenesis regulator	2.78E+00	4.54E-03
11749005_a_at	CTNNA1	catenin (cadherin-associated protein), alpha 1	2.79E+00	2.51E-04
11753788_x_at	CDKN3	cyclin-dependent kinase inhibitor 3	2.81E+00	1.14E-03
11732351_at	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)	2.86E+00	8.83E-04
11722591_s_at	BCAP29	B-cell receptor-associated protein 29	2.86E+00	5.17E-04
11757749_s_at	ARNT	aryl hydrocarbon receptor nuclear translocator	2.88E+00	2.09E-03
11735058_a_at	ARNT2	aryl-hydrocarbon receptor nuclear translocator 2	2.88E+00	4.51E-03
11740524_x_at	CCDC112	coiled-coil domain containing 112	2.89E+00	2.72E-03
11743265_a_at	ENSA	endosulfine alpha	2.90E+00	2.12E-03
11722010_a_at	AZI2	5-azacytidine induced 2	2.91E+00	4.09E-04
11728413_at	SERPINB8	serpin peptidase inhibitor, clade B (ovalbumin), member 8	2.91E+00	1.19E-04
11730885_a_at	FAM63B	family with sequence similarity 63, member B	2.92E+00	5.05E-04
11730828_a_at	CPSF2	cleavage and polyadenylation specific factor 2	2.92E+00	2.89E-03
11736060_a_at	KIF5B	kinesin family member 5B	2.93E+00	3.81E-03
11745780_x_at	OXR1	oxidation resistance 1	2.94E+00	8.76E-04
11732694_at	PTH2R	parathyroid hormone 2 receptor	2.94E+00	1.45E-03
11745144_a_at	CLEC4E	C-type lectin domain family 4, member E	2.94E+00	3.63E-03
11722041_s_at	TEX2	testis expressed 2	2.95E+00	2.96E-04
11753719_a_at	GMFB	glia maturation factor, beta	2.97E+00	3.42E-03
11722589_x_at	BCAP29	B-cell receptor-associated protein 29	2.97E+00	2.42E-04
11717352_a_at	RPRD1A	regulation of nuclear pre-mRNA domain containing 1A	2.98E+00	8.39E-04

11732701_a_at	CSGALNACT2	chondroitin sulfate N-acetylgalactosaminyltransferase 2	2.98E+00	3.12E-04
11745794_s_at	OXR1	oxidation resistance 1	3.00E+00	1.26E-03
11717680_at	UNC13B	unc-13 homolog B (C. elegans)	3.00E+00	7.83E-05
11755003_a_at	ADCY4	adenylate cyclase 4	3.00E+00	2.02E-04
11720726_at	UBR1	ubiquitin protein ligase E3 component n-recognin 1	3.00E+00	5.18E-04
11747052_s_at	GALNT4 /// POC1B-GALNT4	polypeptide N-acetylgalactosaminyltransferase 4 /// POC1B-GALNT4 readthrough	3.01E+00	6.44E-04
11746739_a_at	ORC2	origin recognition complex subunit 2	3.01E+00	1.63E-03
11740689_a_at	PHACTR2	phosphatase and actin regulator 2	3.01E+00	5.83E-05
11728289_a_at	TBC1D2	TBC1 domain family, member 2	3.02E+00	6.56E-04
11726031_a_at	CSNK1G3	casein kinase 1, gamma 3	3.03E+00	9.51E-05
11747166_a_at	WDR41	WD repeat domain 41	3.03E+00	6.76E-05
11717555_at	SETD7	SET domain containing (lysine methyltransferase) 7	3.03E+00	7.17E-04
11759678_at	RAB12	RAB12, member RAS oncogene family	3.05E+00	2.50E-03
11749440_a_at	WDR41	WD repeat domain 41	3.05E+00	1.88E-05
11751176_a_at	PRUNE2	prune homolog 2 (Drosophila)	3.05E+00	1.45E-03
11754263_s_at	CHD9	chromodomain helicase DNA binding protein 9	3.06E+00	2.71E-04
11728047_a_at	RAB3IP	RAB3A interacting protein	3.06E+00	1.95E-04
11743091_x_at	SLC36A4	solute carrier family 36 (proton/amino acid symporter), member 4	3.07E+00	4.65E-05
11748711_a_at	EIF4G3	eukaryotic translation initiation factor 4 gamma, 3	3.08E+00	2.07E-03
11739349_a_at	AMOT	angiomotin	3.08E+00	3.49E-03
11757902_a_at	DYNC1LI2	dynein, cytoplasmic 1, light intermediate chain 2	3.08E+00	4.42E-03
11748341_a_at	TPP1	tripeptidyl peptidase I	3.09E+00	1.89E-04
11717004_s_at	FKBP9 /// FKBP9P1	FK506 binding protein 9 /// FK506 binding protein 9 pseudogene 1	3.10E+00	4.42E-03
11719401_a_at	C1orf122	chromosome 1 open reading frame 122	3.10E+00	6.11E-04
11740601_a_at	APH1B	APH1B gamma secretase subunit	3.10E+00	8.27E-04
11722579_a_at	CHMP4A	charged multivesicular body protein 4A	3.10E+00	3.04E-03
11721801_at	ARL8B	ADP-ribosylation factor like GTPase 8B	3.10E+00	7.67E-04
11732746_a_at	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)	3.11E+00	2.33E-03
11719461_a_at	WDR41	WD repeat domain 41	3.11E+00	5.76E-05
11752817_s_at	TPP1	tripeptidyl peptidase I	3.11E+00	1.83E-03
11717679_a_at	UNC13B	unc-13 homolog B (C. elegans)	3.12E+00	1.52E-04
11764066_s_at	EPB41L5	erythrocyte membrane protein band 4.1 like 5	3.12E+00	2.03E-06
11745789_a_at	TMEM135	transmembrane protein 135	3.13E+00	4.35E-03
11754548_a_at	RMDN1	regulator of microtubule dynamics 1	3.14E+00	4.45E-03
11758255_s_at	CSNK1G3	casein kinase 1, gamma 3	3.14E+00	8.64E-05
11749441_x_at	WDR41	WD repeat domain 41	3.14E+00	7.73E-05
11715739_s_at	PLPP3	phospholipid phosphatase 3	3.14E+00	3.69E-03

11753248_a_at	TPP1	tripeptidyl peptidase I	3.15E+00	8.45E-05
11735352_x_at	SLC37A2	solute carrier family 37 (glucose-6-phosphate transporter), member 2	3.15E+00	4.51E-03
11748342_x_at	TPP1	tripeptidyl peptidase I	3.16E+00	1.38E-04
11759287_at	DNAJB4	DnaJ (Hsp40) homolog, subfamily B, member 4	3.16E+00	7.61E-04
11747893_a_at	AGPAT4	1-acylglycerol-3-phosphate O-acyltransferase 4	3.17E+00	1.18E-03
11746388_x_at	SPIDR	scaffolding protein involved in DNA repair	3.18E+00	2.24E-04
11746092_a_at	ROGDI	rogdi homolog	3.19E+00	1.13E-03
11731792_x_at	SLC27A6	solute carrier family 27 (fatty acid transporter), member 6	3.19E+00	1.40E-03
11746537_x_at	EIF4G3	eukaryotic translation initiation factor 4 gamma, 3	3.19E+00	5.91E-04
11737831_a_at	FNBP1L	formin binding protein 1-like	3.21E+00	3.03E-03
11748990_a_at	KIF1B	kinesin family member 1B	3.21E+00	1.02E-03
11722826_a_at	NCAPG	non-SMC condensin I complex subunit G	3.22E+00	4.96E-03
11740688_a_at	PHACTR2	phosphatase and actin regulator 2	3.23E+00	4.71E-05
11721979_at	SLC36A1	solute carrier family 36 (proton/amino acid symporter), member 1	3.23E+00	2.60E-03
11748216_a_at	STAU2	staufen double-stranded RNA binding protein 2	3.24E+00	5.61E-04
11717526_a_at	RDX	radixin	3.25E+00	2.33E-03
11730053_a_at	SNX16	sorting nexin 16	3.25E+00	1.25E-04
11745424_a_at	SMC2	structural maintenance of chromosomes 2	3.26E+00	4.69E-03
11745386_a_at	ZNF638	zinc finger protein 638	3.26E+00	2.21E-03
11725032_a_at	TPP1	tripeptidyl peptidase I	3.26E+00	3.85E-05
11753451_a_at	WDR41	WD repeat domain 41	3.29E+00	2.99E-04
11739546_a_at	NEDD1	neural precursor cell expressed, developmentally down-regulated 1	3.29E+00	1.71E-03
11755603_a_at	ARL8B	ADP-ribosylation factor like GTPase 8B	3.30E+00	3.41E-03
11718972_at	LPL	lipoprotein lipase	3.31E+00	3.64E-04
11735155_s_at	ZNF585A	zinc finger protein 585A	3.32E+00	2.84E-03
11718571_at	NT5DC3	5'-nucleotidase domain containing 3	3.33E+00	6.17E-05
11755243_a_at	SMG7	SMG7 nonsense mediated mRNA decay factor	3.34E+00	1.63E-03
11719502_a_at	DHX36	DEAH (Asp-Glu-Ala-His) box polypeptide 36	3.34E+00	1.94E-03
11751069_a_at	GLMP	glycosylated lysosomal membrane protein	3.34E+00	8.47E-04
11740266_at	JRKL	JRK-like	3.35E+00	1.09E-03
11725211_a_at	PLCXD1	phosphatidylinositol-specific phospholipase C, X domain containing 1	3.35E+00	1.83E-03
11751719_a_at	AGPAT4	1-acylglycerol-3-phosphate O-acyltransferase 4	3.35E+00	4.02E-04
11756005_x_at	COMMD4	COMM domain containing 4	3.36E+00	1.05E-03
11748641_a_at	ZADH2	zinc binding alcohol dehydrogenase domain containing 2	3.36E+00	2.22E-03
11722976_x_at	HOXB5	homeobox B5	3.37E+00	3.99E-03
11731358_at	ADAMTS3	ADAM metallopeptidase with thrombospondin type 1 motif 3	3.37E+00	2.14E-03

11724302_a_at	EEA1	early endosome antigen 1	3.37E+00	5.67E-04
11718246_a_at	SPIDR	scaffolding protein involved in DNA repair	3.38E+00	9.96E-04
11722935_x_at	CHML	choroideremia-like (Rab escort protein 2)	3.38E+00	2.42E-03
11737819_x_at	LOXHD1	lipoxygenase homology domains 1	3.39E+00	2.89E-04
11737418_a_at	ERCC6L2	excision repair cross-complementation group 6-like 2	3.39E+00	1.92E-03
11749408_a_at	ZFR	zinc finger RNA binding protein	3.39E+00	3.08E-03
11758603_s_at	PANK3	pantothenate kinase 3	3.39E+00	3.20E-03
11752669_a_at	SREK1IP1	SREK1-interacting protein 1	3.39E+00	7.44E-04
11763384_a_at	FBN2	fibrillin 2	3.40E+00	2.84E-03
11716445_s_at	COMMD4 /// LOC440292 /// LOC646670 /// LOC732265	COMM domain containing 4 /// COMM domain-containing protein 4-like /// uncharacterized	3.41E+00	6.59E-05
11719505_x_at	DHX36	DEAH (Asp-Glu-Ala-His) box polypeptide 36	3.43E+00	3.03E-03
11741638_a_at	PLAG1	pleiomorphic adenoma gene 1	3.45E+00	1.36E-06
11717664_s_at	KLHL12	kelch-like family member 12	3.45E+00	3.95E-03
11746542_x_at	SPIDR	scaffolding protein involved in DNA repair	3.46E+00	1.25E-03
11755167_s_at	LINC00847	long intergenic non-protein coding RNA 847	3.47E+00	1.13E-03
11762234_a_at	MAX	MYC associated factor X	3.47E+00	4.39E-03
11756029_a_at	ASCC1	activating signal cointegrator 1 complex subunit 1	3.47E+00	2.13E-04
11732370_a_at	CUX1	cut-like homeobox 1	3.49E+00	2.87E-03
11728056_a_at	CASP10	caspase 10	3.50E+00	1.68E-05
11750434_x_at	SPIDR	scaffolding protein involved in DNA repair	3.51E+00	1.40E-04
11759428_a_at	KLF7	Kruppel-like factor 7 (ubiquitous)	3.51E+00	3.31E-03
11719863_a_at	TTLL12	tubulin tyrosine ligase-like family member 12	3.52E+00	1.39E-04
11748069_a_at	PTH2R	parathyroid hormone 2 receptor	3.53E+00	2.29E-03
11725722_at	ATAD2	ATPase family, AAA domain containing 2	3.55E+00	6.29E-04
11749683_a_at	COL4A3BP	collagen, type IV, alpha 3 (Goodpasture antigen) binding protein	3.56E+00	1.45E-03
11749354_x_at	WDR41	WD repeat domain 41	3.56E+00	6.02E-05
11753249_x_at	TPP1	tripeptidyl peptidase I	3.58E+00	1.44E-04
11728964_s_at	ZNF585A /// ZNF585B	zinc finger protein 585A /// zinc finger protein 585B	3.58E+00	6.43E-04
11729814_a_at	ZADH2	zinc binding alcohol dehydrogenase domain containing 2	3.59E+00	6.55E-04
11717525_s_at	RDX	radixin	3.59E+00	1.97E-04
11762368_at	KLHL8	kelch-like family member 8	3.60E+00	8.90E-04
11752885_x_at	TPP1	tripeptidyl peptidase I	3.61E+00	1.49E-04
11746299_a_at	CNTLN	centlein, centrosomal protein	3.61E+00	2.90E-05
11762431_at	RSPH10B /// RSPH10B2	radial spoke head 10 homolog B (Chlamydomonas) /// radial spoke head 10 homolog B2 (Chl	3.62E+00	4.43E-03

11720360_a_at	PHIP	pleckstrin homology domain interacting protein	3.64E+00	3.98E-03
11749803_s_at	ARL8B	ADP-ribosylation factor like GTPase 8B	3.64E+00	2.05E-03
11729211_a_at	ORC2	origin recognition complex subunit 2	3.65E+00	1.73E-03
11721051_at	C5orf51	chromosome 5 open reading frame 51	3.65E+00	1.95E-03
11723721_a_at	SLAIN2	SLAIN motif family member 2	3.65E+00	1.76E-03
11715119_s_at	C2CD4B	C2 calcium-dependent domain containing 4B	3.68E+00	1.87E-03
11726267_a_at	KDM4C	lysine (K)-specific demethylase 4C	3.68E+00	3.01E-03
11726508_a_at	CDKN2B	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	3.69E+00	7.77E-05
11731090_s_at	CACHD1	cache domain containing 1	3.71E+00	1.47E-03
11761374_x_at	AGPAT4	1-acylglycerol-3-phosphate O-acyltransferase 4	3.71E+00	2.70E-04
11756539_a_at	TCTEX1D1	Tctex1 domain containing 1	3.73E+00	2.22E-03
11759623_at	C11orf71	chromosome 11 open reading frame 71	3.74E+00	2.94E-04
11720964_s_at	DCK	deoxycytidine kinase	3.78E+00	2.76E-03
11728700_a_at	BMPR2	bone morphogenetic protein receptor type II	3.79E+00	2.95E-04
11722219_at	WBP4	WW domain binding protein 4	3.79E+00	3.63E-03
11721802_s_at	ARL8B	ADP-ribosylation factor like GTPase 8B	3.80E+00	1.05E-03
11743925_s_at	SBNO1	strawberry notch homolog 1 (Drosophila)	3.81E+00	1.94E-03
11749908_x_at	SYT11	synaptotagmin XI	3.81E+00	7.30E-04
11747296_a_at	ARMC8	armadillo repeat containing 8	3.81E+00	6.85E-04
11727392_s_at	COX15	cytochrome c oxidase assembly homolog 15 (yeast)	3.82E+00	1.91E-03
11751136_a_at	CCDC30	coiled-coil domain containing 30	3.83E+00	1.87E-05
11758081_s_at	ENSA	endosulfine alpha	3.86E+00	4.89E-03
11730068_a_at	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)	3.88E+00	2.25E-03
11733899_a_at	TROVE2	TROVE domain family, member 2	3.90E+00	1.54E-03
11737818_a_at	LOXHD1	lipoxygenase homology domains 1	3.92E+00	1.67E-03
11739709_a_at	PHACTR2	phosphatase and actin regulator 2	3.93E+00	1.73E-05
11732366_a_at	SCAPER	S-phase cyclin A-associated protein in the ER	3.94E+00	3.26E-03
11748278_x_at	ASCC1	activating signal cointegrator 1 complex subunit 1	3.94E+00	9.36E-05
11728320_a_at	CSTB	cystatin B (stefin B)	3.95E+00	3.72E-03
11722481_a_at	MFF	mitochondrial fission factor	3.95E+00	1.63E-03
11751928_x_at	EIF4G3	eukaryotic translation initiation factor 4 gamma, 3	3.95E+00	5.22E-04
11723814_s_at	ARHGAP29	Rho GTPase activating protein 29	3.96E+00	1.11E-03
11720565_a_at	NAV2	neuron navigator 2	3.97E+00	3.11E-04
11739712_a_at	PHACTR2	phosphatase and actin regulator 2	3.99E+00	1.13E-04
11743400_s_at	CDC27	cell division cycle 27	3.99E+00	3.16E-04
11732261_a_at	RRAGB	Ras-related GTP binding B	4.00E+00	3.85E-03
11725342_a_at	VPS37A	vacuolar protein sorting 37 homolog A (S. cerevisiae)	4.01E+00	1.16E-03
11716144_a_at	PDCD6IP	programmed cell death 6 interacting protein	4.05E+00	5.50E-04
11718514_s_at	TSPAN14	tetraspanin 14	4.07E+00	4.10E-04

11724923_at	PPP1R9A	protein phosphatase 1, regulatory subunit 9A	4.07E+00	4.36E-06
11729410_a_at	ZNF415	zinc finger protein 415	4.07E+00	8.49E-04
11738185_s_at	PDE8B	phosphodiesterase 8B	4.09E+00	5.46E-06
11719287_a_at	PEX11B	peroxisomal biogenesis factor 11 beta	4.13E+00	3.04E-03
11745652_s_at	EPCAM	epithelial cell adhesion molecule	4.13E+00	4.41E-03
11745669_a_at	KANSL1L	KAT8 regulatory NSL complex subunit 1 like	4.15E+00	2.74E-03
11758681_s_at	ASCC1	activating signal cointegrator 1 complex subunit 1	4.18E+00	1.41E-03
11756310_a_at	NXT2	nuclear transport factor 2-like export factor 2	4.19E+00	2.62E-03
11758383_s_at	USP6NL	USP6 N-terminal like	4.20E+00	3.78E-03
11729100_a_at	CFAP70	cilia and flagella associated protein 70	4.23E+00	8.44E-04
11727547_s_at	RBBP8	retinoblastoma binding protein 8	4.27E+00	1.33E-03
11722616_at	UBLCP1	ubiquitin-like domain containing CTD phosphatase 1	4.27E+00	7.83E-04
11727625_x_at	ASCC1	activating signal cointegrator 1 complex subunit 1	4.28E+00	1.96E-03
11733077_a_at	PPP1R12A	protein phosphatase 1, regulatory subunit 12A	4.28E+00	3.28E-03
11737126_x_at	CTAGE15 /// CTAGE4 /// CTAGE6 /// CTAGE8 /// CTAGE9 /// LOC101060696	CTAGE family, member 15 /// CTAGE family, member 4 /// CTAGE family, member 6 /// CTAGE	4.31E+00	8.23E-04
11739978_s_at	AGO3	argonaute RISC catalytic component 3	4.31E+00	1.00E-03
11758223_s_at	GNPDA1	glucosamine-6-phosphate deaminase 1	4.32E+00	1.28E-03
11727624_a_at	ASCC1	activating signal cointegrator 1 complex subunit 1	4.32E+00	2.43E-04
11741967_a_at	CDYL	chromodomain protein, Y-like	4.33E+00	2.03E-03
11730834_a_at	KATNA1	katanin p60 (ATPase containing) subunit A 1	4.33E+00	3.32E-04
11736479_s_at	FAM72A /// FAM72B /// FAM72C /// FAM72D	family with sequence similarity 72, member A /// family with sequence similarity 72, me	4.35E+00	3.63E-03
11763859_x_at	UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	4.35E+00	1.37E-03
11716052_s_at	LAMC1	laminin, gamma 1 (formerly LAMB2)	4.37E+00	1.27E-03
11750018_at	HIST2H2BF	histone cluster 2, H2bf	4.37E+00	3.14E-03
11739117_a_at	EFR3A	EFR3 homolog A	4.39E+00	9.94E-04
11723067_s_at	FAM21A /// FAM21C	family with sequence similarity 21, member A /// family with sequence similarity 21, me	4.40E+00	1.11E-03
11727869_s_at	GALK2	galactokinase 2	4.44E+00	5.20E-04
11761365_at	MIR181A1HG	MIR181A1 host gene	4.44E+00	6.40E-04
11720971_at	TOP2A	topoisomerase (DNA) II alpha	4.45E+00	4.99E-03
11721948_a_at	RAB3GAP2	RAB3 GTPase activating protein subunit 2 (non-catalytic)	4.47E+00	4.16E-03
11740119_x_at	SMC2	structural maintenance of chromosomes 2	4.51E+00	2.19E-03

11742910_a_at	ANKRD50	ankyrin repeat domain 50	4.53E+00	4.34E-04
11736036_a_at	TADA3	transcriptional adaptor 3	4.56E+00	3.23E-03
11750036_a_at	DDIAS	DNA damage-induced apoptosis suppressor	4.57E+00	1.29E-03
11753095_a_at	SLC22A15	solute carrier family 22, member 15	4.58E+00	1.39E-03
11749411_a_at	SPIDR	scaffolding protein involved in DNA repair	4.59E+00	1.56E-04
11741315_a_at	PLSCR4	phospholipid scramblase 4	4.63E+00	1.32E-03
11721553_a_at	GABBR1	gamma-aminobutyric acid (GABA) B receptor, 1	4.66E+00	1.42E-04
11748277_a_at	ASCC1	activating signal cointegrator 1 complex subunit 1	4.74E+00	6.97E-04
11743088_s_at	PDE8B	phosphodiesterase 8B	4.74E+00	3.86E-04
11735751_a_at	SLC43A3	solute carrier family 43, member 3	4.76E+00	4.97E-03
11749353_a_at	WDR41	WD repeat domain 41	4.76E+00	9.94E-06
11737863_s_at	PIGF	phosphatidylinositol glycan anchor biosynthesis class F	4.77E+00	4.88E-03
11736792_a_at	NXT2	nuclear transport factor 2-like export factor 2	4.77E+00	1.31E-04
11721423_a_at	AMZ2	archaelysin family metallopeptidase 2	4.79E+00	2.92E-03
11736163_a_at	CDKN2B	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	4.80E+00	3.15E-04
11728057_a_at	CASP10	caspase 10	4.87E+00	1.01E-04
11729812_at	ZADH2	zinc binding alcohol dehydrogenase domain containing 2	4.92E+00	1.17E-04
11715305_s_at	HOXA10	homeobox A10	4.92E+00	2.96E-03
11741388_a_at	TSPAN14	tetraspanin 14	4.94E+00	1.11E-04
11727720_x_at	BCAP29	B-cell receptor-associated protein 29	4.95E+00	1.46E-04
11727886_a_at	ZNF25	zinc finger protein 25	4.99E+00	1.47E-04
11744324_at	SLX4IP	SLX4 interacting protein	5.01E+00	1.39E-03
11745223_a_at	ATP2C1	ATPase, Ca++ transporting, type 2C, member 1	5.01E+00	5.11E-04
11754676_a_at	CENPE	centromere protein E	5.04E+00	3.09E-03
11741948_a_at	NF1	neurofibromin 1	5.04E+00	4.07E-03
11727277_a_at	TMEM169	transmembrane protein 169	5.07E+00	4.30E-03
11715722_a_at	GOLM1	golgi membrane protein 1	5.09E+00	6.26E-04
11723545_a_at	PLD1	phospholipase D1, phosphatidylcholine-specific	5.10E+00	1.87E-03
11732850_x_at	CTAGE4 /// CTAGE8 /// CTAGE9 /// LOC101060696	CTAGE family, member 4 /// CTAGE family, member 8 /// CTAGE family, member 9 /// cTAGE	5.12E+00	1.04E-03
11716095_s_at	KLF6	Kruppel-like factor 6	5.14E+00	3.12E-03
11719123_a_at	TIMELESS	timeless circadian clock	5.15E+00	3.25E-03
11760536_a_at	HEXA	hexosaminidase A (alpha polypeptide)	5.19E+00	3.45E-03
11740404_a_at	GALNT10	polypeptide N-acetylgalactosaminyltransferase 10	5.22E+00	3.86E-04
11732747_at	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)	5.23E+00	1.07E-05
11736396_a_at	RFT1	RFT1 homolog	5.24E+00	8.36E-04
11727008_at	SORT1	sortilin 1	5.25E+00	4.82E-03

11747720_a_at	BIRC5	baculoviral IAP repeat containing 5	5.27E+00	3.51E-03
11752596_s_at	TPP1	tripeptidyl peptidase I	5.29E+00	2.32E-04
11722975_at	HOXB5	homeobox B5	5.31E+00	4.85E-04
11719485_x_at	EPS15	epidermal growth factor receptor pathway substrate 15	5.35E+00	1.23E-03
11718952_at	EPB41L5	erythrocyte membrane protein band 4.1 like 5	5.41E+00	4.03E-06
11735676_a_at	OSGIN1	oxidative stress induced growth inhibitor 1	5.43E+00	2.20E-03
11757936_s_at	GCSH /// GCSHP3 /// LOC641746 /// NDUFS1	glycine cleavage system protein H (aminomethyl carrier) /// glycine cleavage system pro	5.50E+00	2.64E-03
11737107_x_at	ARL17A /// ARL17B /// LOC100294341 /// LOC100996709	ADP-ribosylation factor like GTPase 17A /// ADP-ribosylation factor like GTPase 17B ///	5.52E+00	2.75E-03
11724249_a_at	SLC39A8	solute carrier family 39 (zinc transporter), member 8	5.54E+00	3.82E-04
11757469_s_at	TPP1	tripeptidyl peptidase I	5.55E+00	2.60E-04
11743684_a_at	GPR21 /// RABGAP1	G protein-coupled receptor 21 /// RAB GTPase activating protein 1	5.55E+00	2.55E-03
11746197_a_at	HUS1	HUS1 checkpoint clamp component	5.64E+00	6.33E-04
11736059_a_at	KIF5B	kinesin family member 5B	5.65E+00	3.48E-03
11715637_a_at	UGP2	UDP-glucose pyrophosphorylase 2	5.73E+00	3.88E-03
11722436_a_at	GOSR2	golgi SNAP receptor complex member 2	5.77E+00	8.51E-04
11726000_x_at	CPNE4	copine IV	5.77E+00	1.05E-05
11723813_at	ARHGAP29	Rho GTPase activating protein 29	5.78E+00	3.75E-03
11729445_at	FAM150B	family with sequence similarity 150, member B	5.95E+00	7.03E-04
11725343_a_at	VPS37A	vacuolar protein sorting 37 homolog A (S. cerevisiae)	6.06E+00	1.09E-03
11722917_s_at	IGF2BP3	insulin-like growth factor 2 mRNA binding protein 3	6.06E+00	1.68E-03
11719449_at	RAB14	RAB14, member RAS oncogene family	6.07E+00	4.62E-03
11755605_s_at	MALAT1	metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)	6.14E+00	4.59E-03
11738080_x_at	ANKRD20A1 /// ANKRD20A2 /// ANKRD20A3 /// ANKRD20A4	ankyrin repeat domain 20 family, member A1 /// ankyrin repeat domain 20 family, member	6.14E+00	2.81E-05
11751884_s_at	PSME4	proteasome activator subunit 4	6.16E+00	1.47E-03
11750768_a_at	OSBPL6	oxysterol binding protein-like 6	6.29E+00	3.84E-03
11743976_at	MYL12A	myosin light chain 12A	6.34E+00	4.19E-03
11722593_s_at	BCAP29	B-cell receptor-associated protein 29	6.36E+00	4.55E-04
11739028_s_at	CLTC	clathrin, heavy chain (Hc)	6.36E+00	1.61E-03

11739563_a_at	ITPR1	inositol 1,4,5-trisphosphate receptor, type 1	6.39E+00	1.93E-03
11731819_a_at	TBC1D8B	TBC1 domain family, member 8B (with GRAM domain)	6.40E+00	3.81E-05
11726722_at	WDFY2	WD repeat and FYVE domain containing 2	6.42E+00	1.67E-03
11743685_a_at	GPR21 /// RABGAP1	G protein-coupled receptor 21 /// RAB GTPase activating protein 1	6.48E+00	2.81E-03
11715977_a_at	VGLL4	vestigial-like family member 4	6.52E+00	2.52E-03
11718329_a_at	NAPB	N-ethylmaleimide-sensitive factor attachment protein, beta	6.53E+00	3.85E-03
11745269_s_at	CHD9	chromodomain helicase DNA binding protein 9	6.54E+00	4.55E-03
11736397_at	RFT1	RFT1 homolog	6.56E+00	4.41E-03
11717742_a_at	GPRC5C	G protein-coupled receptor, class C, group 5, member C	6.65E+00	1.65E-03
11728300_at	CCNE2	cyclin E2	6.70E+00	1.87E-05
11740972_a_at	MAPRE2	microtubule-associated protein, RP/EB family, member 2	6.71E+00	1.54E-03
11739384_a_at	ROCK2	Rho-associated, coiled-coil containing protein kinase 2	6.72E+00	2.93E-03
11721869_at	HIF1AN	hypoxia inducible factor 1, alpha subunit inhibitor	6.77E+00	3.76E-03
11737922_a_at	SERPINB8	serpin peptidase inhibitor, clade B (ovalbumin), member 8	6.79E+00	2.95E-04
11759608_at	EPB41L5	erythrocyte membrane protein band 4.1 like 5	7.03E+00	5.98E-06
11756273_a_at	RBM12B	RNA binding motif protein 12B	7.05E+00	1.86E-03
11738111_a_at	LOXHD1	lipoxygenase homology domains 1	7.08E+00	8.14E-06
11735041_a_at	C1orf52	chromosome 1 open reading frame 52	7.10E+00	4.99E-03
11722801_s_at	SS18	synovial sarcoma translocation, chromosome 18	7.15E+00	2.16E-03
11729960_x_at	DHFRL1	dihydrofolate reductase like 1	7.21E+00	2.82E-03
11745653_a_at	CASP10	caspase 10	7.29E+00	2.57E-03
11747648_a_at	TCAIM	T cell activation inhibitor, mitochondrial	7.38E+00	1.05E-03
11722252_x_at	NEK2	NIMA-related kinase 2	7.43E+00	4.64E-03
11756942_s_at	ING5	inhibitor of growth family member 5	7.49E+00	1.74E-03
11746840_a_at	PPP1R9A	protein phosphatase 1, regulatory subunit 9A	7.51E+00	4.50E-04
11755516_a_at	KIF16B	kinesin family member 16B	7.52E+00	1.59E-05
11730725_a_at	GABBR1	gamma-aminobutyric acid (GABA) B receptor, 1	7.53E+00	4.28E-06
11728156_s_at	LACC1	laccase (multicopper oxidoreductase) domain containing 1	7.70E+00	4.60E-05
11725344_a_at	VPS37A	vacuolar protein sorting 37 homolog A (S. cerevisiae)	7.70E+00	1.32E-05
11731989_at	HESX1	HESX homeobox 1	7.81E+00	1.32E-03
11756873_a_at	ALDH8A1	aldehyde dehydrogenase 8 family, member A1	7.90E+00	1.64E-04
11722690_at	SPTBN1	spectrin, beta, non-erythrocytic 1	7.94E+00	4.77E-03
11741370_a_at	ATP6V1E1	ATPase, H+ transporting, lysosomal 31kDa, V1 subunit E1	7.96E+00	4.56E-03

11746594_a_at	MCM2	minichromosome maintenance complex component 2	8.22E+00	3.63E-03
11726835_a_at	DNAJC21	DnaJ (Hsp40) homolog, subfamily C, member 21	8.27E+00	4.12E-03
11748180_a_at	SERPINB8	serpin peptidase inhibitor, clade B (ovalbumin), member 8	8.30E+00	9.03E-08
11755837_a_at	ARHGEF5 /// LOC102725117	Rho guanine nucleotide exchange factor 5 /// rho guanine nucleotide exchange factor 5-l	8.35E+00	3.20E-05
11737592_at	EYS	eyes shut homolog (Drosophila)	8.37E+00	2.21E-04
11735379_a_at	CEP162	centrosomal protein 162kDa	8.54E+00	5.58E-04
11740267_at	CCDC18	coiled-coil domain containing 18	8.85E+00	3.19E-03
11747235_a_at	KANSL1L	KAT8 regulatory NSL complex subunit 1 like	8.85E+00	3.77E-03
11753254_a_at	SGOL2	shugoshin-like 2 (S. pombe)	8.85E+00	8.62E-04
11731736_at	PGAP1	post-GPI attachment to proteins 1	8.86E+00	1.54E-03
11716818_a_at	VSIG4	V-set and immunoglobulin domain containing 4	8.89E+00	1.82E-03
11730796_x_at	PSPH	phosphoserine phosphatase	9.13E+00	3.35E-04
11754428_a_at	ZFYVE26	zinc finger, FYVE domain containing 26	9.15E+00	1.01E-04
11728496_a_at	GNRH1	gonadotropin releasing hormone 1	9.32E+00	4.97E-03
11761061_at	UQCRC1	ubiquinol-cytochrome c reductase core protein I	9.33E+00	4.71E-03
11755302_a_at	COL24A1	collagen, type XXIV, alpha 1	9.34E+00	3.34E-03
11717369_a_at	ARSD	arylsulfatase D	9.37E+00	2.96E-03
11722204_a_at	FBXO22	F-box protein 22	9.39E+00	4.48E-03
11752733_a_at	CSNK1G3	casein kinase 1, gamma 3	9.41E+00	7.93E-05
11727545_at	PANK3	pantothenate kinase 3	9.45E+00	4.42E-04
11731185_at	PPM1E	protein phosphatase, Mg2+/Mn2+ dependent, 1E	9.45E+00	1.50E-03
11724153_at	CELF6	CUGBP, Elav-like family member 6	9.46E+00	2.15E-06
11762388_at	LOC145783 /// ZNF280D	uncharacterized LOC145783 /// zinc finger protein 280D	9.66E+00	2.68E-03
11725448_at	PLD6	phospholipase D family, member 6	9.98E+00	3.24E-04
11762063_a_at	EPB41L5	erythrocyte membrane protein band 4.1 like 5	1.01E+01	1.59E-04
11729983_s_at	PARP11	poly(ADP-ribose) polymerase family member 11	1.02E+01	2.66E-03
11756957_a_at	DCTN2	dynactin 2 (p50)	1.02E+01	5.23E-06
11740857_x_at	AGO3	argonaute RISC catalytic component 3	1.03E+01	3.12E-03
11755942_x_at	C19orf54	chromosome 19 open reading frame 54	1.03E+01	3.09E-04
11758682_s_at	TMEM261	transmembrane protein 261	1.05E+01	1.71E-03
11746974_a_at	LDLR	low density lipoprotein receptor	1.08E+01	3.09E-04
11754566_a_at	RNF217	ring finger protein 217	1.08E+01	1.87E-04
11727868_a_at	GALK2	galactokinase 2	1.15E+01	8.92E-05
11749733_a_at	JRKL	JRK-like	1.19E+01	1.87E-03
11728493_a_at	ZNF197	zinc finger protein 197	1.19E+01	3.86E-03
11746499_a_at	GSTO2	glutathione S-transferase omega 2	1.21E+01	3.92E-03
11721426_at	FZD5	frizzled class receptor 5	1.22E+01	4.03E-03
11717645_a_at	GPR107	G protein-coupled receptor 107	1.22E+01	2.49E-04

11751667_a_at	COQ5	coenzyme Q5, methyltransferase	1.24E+01	1.77E-06
11747208_a_at	MYSM1	Myb-like, SWIRM and MPN domains 1	1.25E+01	4.76E-03
11729834_at	EIF5A2	eukaryotic translation initiation factor 5A2	1.25E+01	5.31E-04
11731540_at	SESTD1	SEC14 and spectrin domains 1	1.27E+01	4.98E-04
11732381_a_at	USP6NL	USP6 N-terminal like	1.28E+01	1.30E-04
11752333_a_at	ITGAV	integrin alpha V	1.29E+01	7.10E-05
11731887_at	KIF14	kinesin family member 14	1.29E+01	3.72E-03
11727293_x_at	ALG10B	ALG10B, alpha-1,2-glucosyltransferase	1.33E+01	6.41E-04
11763326_s_at	FER	fer (fps/fes related) tyrosine kinase	1.33E+01	1.51E-03
11755905_a_at	ST8SIA5	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 5	1.43E+01	3.60E-03
11763766_x_at	STYXL1	serine/threonine/tyrosine interacting-like 1	1.46E+01	4.08E-03
11730362_a_at	KIAA1107	KIAA1107	1.51E+01	9.22E-06
11760436_a_at	DST	dystonin	1.54E+01	1.69E-03
11757502_x_at	SCRG1	stimulator of chondrogenesis 1	1.56E+01	6.06E-04
11748825_x_at	TTLL12	tubulin tyrosine ligase-like family member 12	1.56E+01	1.69E-03
11756992_a_at	MARC2	mitochondrial amidoxime reducing component 2	1.59E+01	5.77E-04
11732259_at	SH2D4A	SH2 domain containing 4A	1.59E+01	4.83E-03
11723068_at	CRHBP	corticotropin releasing hormone binding protein	1.61E+01	1.46E-03
11736249_x_at	KIAA0101	KIAA0101	1.68E+01	4.26E-03
11723845_a_at	DEFB1	defensin, beta 1	1.68E+01	2.47E-03
11739812_a_at	HDGFRP3	hepatoma-derived growth factor, related protein 3	1.72E+01	1.08E-04
11755557_a_at	HACE1	HECT domain and ankyrin repeat containing E3 ubiquitin protein ligase 1	1.74E+01	4.24E-03
11758572_s_at	HELLS	helicase, lymphoid-specific	1.75E+01	4.12E-03
11756950_a_at	CELF6	CUGBP, Elav-like family member 6	1.80E+01	7.59E-07
11764258_at	LOC102723906	uncharacterized LOC102723906	1.86E+01	9.21E-04
11721593_at	CHST1	carbohydrate (keratan sulfate Gal-6) sulfotransferase 1	1.87E+01	4.67E-03
11734291_at	ADAMTS17	ADAM metallopeptidase with thrombospondin type 1 motif 17	1.89E+01	1.03E-03
11762002_at	NCOA7	nuclear receptor coactivator 7	1.91E+01	3.91E-03
11742890_at	MND1	meiotic nuclear divisions 1	1.92E+01	1.28E-03
11730219_s_at	SPG20	spastic paraplegia 20 (Troyer syndrome)	2.09E+01	7.98E-05
11735419_at	HTR1D	5-hydroxytryptamine (serotonin) receptor 1D, G protein-coupled	2.14E+01	4.66E-03
11729886_at	C18orf54	chromosome 18 open reading frame 54	2.15E+01	3.27E-04
11719404_a_at	HIPK2	homeodomain interacting protein kinase 2	2.16E+01	1.24E-03
11734295_x_at	ZNF772	zinc finger protein 772	2.16E+01	1.99E-03
11724055_x_at	SGOL2	shugoshin-like 2 (S. pombe)	2.16E+01	1.83E-03
11759900_at	TMEM144	transmembrane protein 144	2.30E+01	2.42E-03
11756989_x_at	KRT8	keratin 8, type II	2.39E+01	4.89E-03
11742358_at	OR5AN1	olfactory receptor, family 5, subfamily AN, member 1	2.57E+01	4.55E-04

11731744_a_at	TRDMT1	tRNA aspartic acid methyltransferase 1	2.58E+01	2.82E-04
11725074_a_at	FAM126A	family with sequence similarity 126, member A	2.63E+01	7.29E-04
11724831_at	STS	steroid sulfatase (microsomal), isozyme S	2.63E+01	1.88E-03
11745398_a_at	NF1	neurofibromin 1	2.64E+01	6.90E-04
11753792_a_at	NEK3	NIMA-related kinase 3	2.74E+01	4.11E-03
11759123_at	TMEM150C	transmembrane protein 150C	2.79E+01	3.86E-03
11721524_s_at	ZNF706	zinc finger protein 706	2.88E+01	1.92E-03
11752681_s_at	TMEM135	transmembrane protein 135	3.01E+01	6.85E-04
11716300_x_at	ITGAV	integrin alpha V	3.07E+01	4.59E-05
11737127_a_at	ITGA1	integrin alpha 1	3.31E+01	1.69E-03
11731698_s_at	ALG10 /// ALG10B	ALG10, alpha-1,2-glucosyltransferase /// ALG10B, alpha-1,2-glucosyltransferase	3.57E+01	8.07E-04
11752403_a_at	STEAP3	STEAP family member 3, metalloreductase	3.77E+01	3.90E-03
11733378_at	DCAF4L1	DDB1 and CUL4 associated factor 4-like 1	3.83E+01	4.08E-03
11733887_at	FBXO42	F-box protein 42	3.96E+01	1.87E-03
11736850_a_at	FSD1L	fibronectin type III and SPRY domain containing 1-like	4.12E+01	2.99E-03
11721688_at	ID4	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein	4.29E+01	2.00E-03
11725191_s_at	ADGRG6	adhesion G protein-coupled receptor G6	4.59E+01	1.22E-04
11749804_a_at	CELF6	CUGBP, Elav-like family member 6	4.66E+01	9.15E-07
11753973_x_at	MUC1	mucin 1, cell surface associated	4.72E+01	9.91E-04
11737997_x_at	LRRC3	leucine rich repeat containing 3	4.75E+01	5.13E-04
11730112_a_at	DEPDC1	DEP domain containing 1	5.31E+01	2.71E-03
11759595_x_at	ERVK-6	endogenous retrovirus group K, member 6	5.35E+01	1.07E-03
11736958_at	TNFSF15	tumor necrosis factor (ligand) superfamily, member 15	5.96E+01	2.75E-05
11763568_a_at	NXPE2	neurexophilin and PC-esterase domain family, member 2	6.13E+01	4.52E-04
11731511_s_at	DBT	dihydrolipoamide branched chain transacylase E2	6.16E+01	9.54E-04
11750398_a_at	HTR2A	5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled	6.23E+01	1.07E-03
11758961_at	FUT9	fucosyltransferase 9 (alpha (1,3) fucosyltransferase)	6.74E+01	2.97E-03
11758943_s_at	AK4	adenylate kinase 4	7.07E+01	1.43E-03
11757573_s_at	FZD5	frizzled class receptor 5	7.57E+01	3.74E-03
11755806_at	ZNF407	zinc finger protein 407	7.79E+01	2.43E-03
11741058_a_at	SATB1	SATB homeobox 1	8.20E+01	4.18E-03
11724397_s_at	PRRG1	proline rich Gla (G-carboxyglutamic acid) 1	8.46E+01	7.59E-05
11741005_a_at	PARG	poly (ADP-ribose) glycohydrolase	1.05E+02	2.43E-03
11761621_at	TIPIN	TIMELESS interacting protein	1.35E+02	4.83E-03
11759160_at	IFT81	intraflagellar transport 81	1.62E+02	3.96E-03
11761425_at	GALNT11	polypeptide N-acetylgalactosaminyltransferase 11	3.77E+02	6.58E-05

Table 16. The list of 273 probsets surviving the cross validation threshold used for model construction.

Probe set ID	Gene Symbol	Gene Description	Hazard Ratio	p-value
11764026_at	PPARA	peroxisome proliferator-activated receptor alpha	2.32E-02	1.58E-03
11738100_a_at	C20orf195	chromosome 20 open reading frame 195	2.39E-02	2.02E-04
11748669_x_at	PDCD2	programmed cell death 2	2.42E-02	6.27E-04
11726223_a_at	FBLN7	fibulin 7	2.98E-02	2.49E-03
11725802_a_at	USP46	ubiquitin specific peptidase 46	3.32E-02	9.86E-04
11751169_a_at	PGAP3	post-GPI attachment to proteins 3	4.26E-02	1.62E-03
11761131_at	NOL12	nucleolar protein 12	4.56E-02	3.38E-04
11733741_x_at	SYMPK	symplekin	4.60E-02	9.28E-04
11755591_x_at	MSTO1	misato 1, mitochondrial distribution and morphology regulator	5.35E-02	2.93E-03
11741162_a_at	SLC12A4	solute carrier family 12 (potassium/chloride transporter), member 4	5.68E-02	1.29E-03
11725634_at	MRPS25	mitochondrial ribosomal protein S25	5.75E-02	4.39E-03
11758321_s_at	GATSL2	GATS protein-like 2	5.93E-02	4.16E-03
11747480_a_at	CENPT	centromere protein T	6.59E-02	1.03E-03
11751058_a_at	OTX1	orthodenticle homeobox 1	6.85E-02	3.17E-03
11721459_a_at	TAPBPL	TAP binding protein-like	6.86E-02	2.02E-04
11716965_a_at	ATN1	atrophin 1	7.41E-02	3.99E-03
11721983_a_at	PID1	phosphotyrosine interaction domain containing 1	8.21E-02	3.50E-03
11746304_a_at	NOP14	NOP14 nucleolar protein	8.41E-02	3.25E-03
11717112_a_at	REPIN1	replication initiator 1	8.58E-02	1.96E-03
11721770_x_at	CCDC64	coiled-coil domain containing 64	8.80E-02	3.06E-03
11717402_s_at	ARF6	ADP-ribosylation factor 6	8.81E-02	2.78E-04
11723612_a_at	TOX2	TOX high mobility group box family member 2	9.00E-02	5.12E-04
11722494_s_at	HIP1R	huntingtin interacting protein 1 related	9.59E-02	1.04E-03
11751711_s_at	SLC48A1	solute carrier family 48 (heme transporter), member 1	9.86E-02	1.46E-03
11739406_x_at	FPGS	folylpolyglutamate synthase	1.07E-01	2.16E-03
11726510_a_at	SMPD3	sphingomyelin phosphodiesterase 3, neutral membrane (neutral sphingomyelinase II)	1.11E-01	2.07E-03
11726458_s_at	GSPT2	G1 to S phase transition 2	1.12E-01	3.17E-03
11727004_s_at	KXD1	KxDL motif containing 1	1.12E-01	1.81E-03
11758905_x_at	TMED4	transmembrane p24 trafficking protein 4	1.12E-01	4.72E-04
11756999_a_at	CYB561D2	cytochrome b561 family, member D2	1.14E-01	3.44E-03
11739667_at	GALNT6	polypeptide N-acetylgalactosaminyltransferase 6	1.15E-01	1.50E-03

11746529_x_at	TNFRSF14	tumor necrosis factor receptor superfamily, member 14	1.21E-01	5.17E-04
11722374_a_at	GGA2	golgi-associated, gamma adaptin ear containing, ARF binding protein 2	1.32E-01	2.08E-03
11726353_at	CD180	CD180 molecule	1.32E-01	2.74E-03
11754747_x_at	QTRT1	queuine tRNA-ribosyltransferase 1	1.41E-01	3.38E-03
11717401_at	ARF6	ADP-ribosylation factor 6	1.46E-01	2.07E-03
11757321_a_at	TAPBP	TAP binding protein (tapasin)	1.53E-01	1.51E-03
11727656_s_at	EPHB1	EPH receptor B1	1.53E-01	1.50E-04
11715524_a_at	COTL1	coactosin-like F-actin binding protein 1	1.58E-01	1.51E-03
11725960_s_at	CALM2 /// CALM3	calmodulin 2 (phosphorylase kinase, delta) /// calmodulin 3 (phosphorylase kinase, delt	1.62E-01	3.66E-03
11721669_a_at	KLHDC4 /// LOC105371397	kelch domain containing 4 /// uncharacterized LOC105371397	1.65E-01	1.72E-03
11736090_a_at	OLIG2	oligodendrocyte lineage transcription factor 2	1.76E-01	2.90E-04
11759223_a_at	EDF1	endothelial differentiation-related factor 1	1.77E-01	1.79E-03
11744315_at	SGK223	homolog of rat pragma of Rnd2	1.85E-01	7.12E-04
11717508_at	IRF4	interferon regulatory factor 4	1.88E-01	2.48E-03
11745772_x_at	TNFRSF14	tumor necrosis factor receptor superfamily, member 14	1.90E-01	2.13E-03
11743194_x_at	TNFRSF14	tumor necrosis factor receptor superfamily, member 14	1.95E-01	1.22E-03
11728523_a_at	MYCL	v-myc avian myelocytomatosis viral oncogene lung carcinoma derived homolog	2.02E-01	3.34E-03
11728576_a_at	PYROXD2	pyridine nucleotide-disulphide oxidoreductase domain 2	2.05E-01	2.31E-03
11763978_a_at	SRRT	serrate, RNA effector molecule	2.08E-01	1.06E-03
11731578_s_at	ACSS1	acyl-CoA synthetase short-chain family member 1	2.12E-01	1.57E-03
11716273_x_at	POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	2.24E-01	6.40E-05
11716272_a_at	POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	2.26E-01	7.14E-04
11750008_a_at	POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	2.28E-01	4.96E-04
11724933_a_at	FAM173A	family with sequence similarity 173, member A	2.34E-01	4.09E-03
11720327_a_at	POLR2C	polymerase (RNA) II (DNA directed) polypeptide C, 33kDa	2.35E-01	5.03E-04
11721460_s_at	TAPBPL	TAP binding protein-like	2.37E-01	7.64E-06
11745894_x_at	TNFRSF14	tumor necrosis factor receptor superfamily, member 14	2.41E-01	8.06E-04
11756172_x_at	TBC1D22A	TBC1 domain family, member 22A	2.50E-01	4.67E-03
11735695_a_at	ARF5	ADP-ribosylation factor 5	2.57E-01	1.13E-03

11741419_x_at	KCNQ1	potassium channel, voltage gated KQT-like subfamily Q, member 1	2.63E-01	3.11E-03
11726165_at	ZNF518B	zinc finger protein 518B	2.65E-01	7.14E-04
11718305_a_at	PDLIM2	PDZ and LIM domain 2 (mystique)	2.66E-01	6.30E-04
11754156_x_at	ARF5	ADP-ribosylation factor 5	2.75E-01	1.57E-03
11744872_x_at	ARF5	ADP-ribosylation factor 5	2.75E-01	1.25E-03
11755617_s_at	EPHB1	EPH receptor B1	2.81E-01	4.50E-03
11746363_x_at	SYNGR2	synaptogyrin 2	2.82E-01	1.75E-03
11742938_at	ASCL2	achaete-scute family bHLH transcription factor 2	2.90E-01	1.87E-03
11750009_x_at	POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	2.92E-01	2.63E-04
11754609_x_at	STMN3	stathmin-like 3	2.92E-01	3.53E-03
11743434_a_at	CHST11	carbohydrate (chondroitin 4) sulfotransferase 11	2.99E-01	6.75E-05
11744190_a_at	SPNS3	spinster homolog 3 (Drosophila)	3.05E-01	3.26E-03
11734503_x_at	ALOX15	arachidonate 15-lipoxygenase	3.08E-01	3.83E-03
11750605_a_at	ATF6B /// TNXB	activating transcription factor 6 beta /// tenascin XB	3.13E-01	3.06E-03
11744067_s_at	SYNGR2	synaptogyrin 2	3.14E-01	1.49E-03
11744871_a_at	ARF5	ADP-ribosylation factor 5	3.25E-01	2.49E-03
11730422_at	FLVCR1	feline leukemia virus subgroup C cellular receptor 1	3.25E-01	2.00E-04
11720226_s_at	TBC1D1	TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 1	3.26E-01	2.23E-03
11746076_a_at	PRAP1 /// ZNF511 /// ZNF511-PRAP1	proline-rich acidic protein 1 /// zinc finger protein 511 /// ZNF511-PRAP1 readthrough	3.26E-01	3.48E-03
11744191_x_at	SPNS3	spinster homolog 3 (Drosophila)	3.31E-01	4.59E-03
11742985_at	LY9	lymphocyte antigen 9	3.33E-01	2.85E-03
11736188_a_at	ORMDL3	ORMDL sphingolipid biosynthesis regulator 3	3.37E-01	3.70E-03
11744176_at	PPP1R9B	protein phosphatase 1, regulatory subunit 9B	3.37E-01	9.25E-04
11715439_at	DAP	death-associated protein	3.39E-01	4.87E-03
11727657_at	EPHB1	EPH receptor B1	3.49E-01	1.66E-05
11755873_a_at	ACSS1	acyl-CoA synthetase short-chain family member 1	3.53E-01	3.45E-03
11718456_at	SLC43A2	solute carrier family 43 (amino acid system L transporter), member 2	3.53E-01	6.85E-04
11739658_a_at	LTB	lymphotoxin beta (TNF superfamily, member 3)	3.84E-01	1.22E-03
11716511_x_at	SRRM2	serine/arginine repetitive matrix 2	3.84E-01	4.91E-03
11739657_a_at	LTB	lymphotoxin beta (TNF superfamily, member 3)	3.89E-01	2.67E-03
11717588_at	CLPP	caseinolytic mitochondrial matrix peptidase proteolytic subunit	3.99E-01	4.95E-03
11715807_a_at	STARD10	StAR-related lipid transfer domain containing 10	4.03E-01	3.43E-04

11725746_a_at	PITPNC1	phosphatidylinositol transfer protein, cytoplasmic 1	4.10E-01	1.40E-03
11723325_a_at	KCNQ1	potassium channel, voltage gated KQT-like subfamily Q, member 1	4.16E-01	3.71E-03
11763447_x_at	TRDC	T cell receptor delta constant	4.28E-01	1.98E-03
11715727_s_at	OAZ2	ornithine decarboxylase antizyme 2	4.50E-01	4.31E-04
11717162_a_at	SLC29A1	solute carrier family 29 (equilibrative nucleoside transporter), member 1	4.53E-01	4.26E-03
11758214_s_at	SRRM2	serine/arginine repetitive matrix 2	4.60E-01	2.64E-03
11740244_a_at	PHOSPHO1	phosphatase, orphan 1	4.62E-01	5.22E-04
11725186_x_at	KLF12	Kruppel-like factor 12	4.64E-01	4.53E-03
11719228_a_at	CAMK1	calcium/calmodulin-dependent protein kinase I	4.67E-01	1.56E-03
11738435_x_at	CD74	CD74 molecule, major histocompatibility complex, class II invariant chain	4.84E-01	2.83E-03
11717661_a_at	PPP1R16B	protein phosphatase 1, regulatory subunit 16B	4.86E-01	3.28E-03
11737845_x_at	FAM102A	family with sequence similarity 102, member A	5.05E-01	4.19E-03
11726254_s_at	CD74	CD74 molecule, major histocompatibility complex, class II invariant chain	5.12E-01	4.19E-03
11740554_a_at	PHOSPHO1	phosphatase, orphan 1	5.18E-01	9.90E-04
11720225_a_at	TBC1D1	TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 1	5.22E-01	4.59E-03
11731422_s_at	FCGR3A	Fc fragment of IgG, low affinity IIIa, receptor (CD16a)	5.37E-01	4.03E-04
11722635_at	IL2RB	interleukin 2 receptor, beta	5.38E-01	3.30E-03
11726255_x_at	CD74	CD74 molecule, major histocompatibility complex, class II invariant chain	5.41E-01	4.40E-03
11715394_s_at	CD81	CD81 molecule	5.41E-01	3.76E-04
11722403_a_at	PLEKHO1	pleckstrin homology domain containing, family O member 1	5.62E-01	5.83E-05
11732275_at	CCL5	chemokine (C-C motif) ligand 5	5.68E-01	2.76E-04
11731941_at	PRSS33	protease, serine, 33	5.84E-01	2.54E-03
11754313_s_at	ARL4C	ADP-ribosylation factor like GTPase 4C	5.93E-01	6.36E-04
11732276_x_at	CCL5	chemokine (C-C motif) ligand 5	6.00E-01	1.61E-04
11732331_s_at	ARL4C	ADP-ribosylation factor like GTPase 4C	6.05E-01	1.86E-03
11753810_a_at	CCL5	chemokine (C-C motif) ligand 5	6.29E-01	1.71E-04
11726046_s_at	FCGR2B	Fc fragment of IgG, low affinity IIb, receptor (CD32)	6.39E-01	4.29E-03
11733736_a_at	CD2	CD2 molecule	6.44E-01	4.28E-03
11741517_s_at	FCGR2B	Fc fragment of IgG, low affinity IIb, receptor (CD32)	6.56E-01	4.93E-03
11727609_at	KLRB1	killer cell lectin-like receptor subfamily B, member 1	6.58E-01	2.66E-04

11741899_s_at	FCGR2B	Fc fragment of IgG, low affinity IIb, receptor (CD32)	6.60E-01	4.61E-03
11762318_x_at	IL23A	interleukin 23, alpha subunit p19	6.64E-01	3.01E-03
11758555_s_at	GPR183	G protein-coupled receptor 183	6.68E-01	4.39E-03
11728560_at	GZMK	granzyme K	6.80E-01	1.80E-03
11755180_x_at	TCF7	transcription factor 7 (T-cell specific, HMG-box)	6.84E-01	1.61E-03
11741190_a_at	P2RY10	purinergic receptor P2Y, G-protein coupled, 10	6.92E-01	4.90E-03
11761918_x_at	TRBC1 /// TRBV19	T cell receptor beta constant 1 /// T cell receptor beta variable 19	6.96E-01	4.57E-03
11722379_at	GNG11	guanine nucleotide binding protein (G protein), gamma 11	7.03E-01	2.28E-03
11763233_x_at	TRAC /// TRAJ17 /// TRAV20 /// TRDV2	T-cell receptor alpha constant /// T cell receptor alpha joining 17 /// T cell receptor	7.23E-01	3.17E-03
11719120_a_at	KYNU	kynureninase	1.26E+00	4.20E-03
11741830_s_at	RFPL4A /// RFPL4AL1	ret finger protein-like 4A /// ret finger protein-like 4A-like 1	1.38E+00	3.95E-03
11718397_s_at	JUN	jun proto-oncogene	1.40E+00	1.32E-03
11718395_s_at	JUN	jun proto-oncogene	1.40E+00	8.18E-04
11739844_at	UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	1.40E+00	3.93E-03
11738959_s_at	CARD17	caspase recruitment domain family, member 17	1.40E+00	5.90E-04
11718396_x_at	JUN	jun proto-oncogene	1.44E+00	6.24E-04
11737166_at	ANKRD34B	ankyrin repeat domain 34B	1.46E+00	5.20E-04
11720029_a_at	LDLR	low density lipoprotein receptor	1.46E+00	2.77E-03
11737167_at	ANKRD34B	ankyrin repeat domain 34B	1.49E+00	8.26E-05
11733646_x_at	KYNU	kynureninase	1.49E+00	4.66E-03
11723209_s_at	KBTBD6	kelch repeat and BTB (POZ) domain containing 6	1.54E+00	2.80E-03
11718998_x_at	DHFR	dihydrofolate reductase	1.56E+00	4.54E-03
11724178_a_at	DAAM2	dishevelled associated activator of morphogenesis 2	1.58E+00	1.34E-04
11749598_a_at	IL18RAP	interleukin 18 receptor accessory protein	1.58E+00	4.07E-03
11729424_s_at	CCRL2	chemokine (C-C motif) receptor-like 2	1.59E+00	1.77E-03
11732084_a_at	TFEC	transcription factor EC	1.59E+00	6.09E-04
11727171_at	TRPS1	trichorhinophalangeal syndrome I	1.59E+00	2.97E-03
11721733_a_at	GCH1	GTP cyclohydrolase 1	1.60E+00	1.44E-03
11743416_s_at	C5orf30	chromosome 5 open reading frame 30	1.60E+00	1.22E-04
11743497_at	BMP2	bone morphogenetic protein 2	1.61E+00	4.97E-04
11718152_a_at	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)	1.63E+00	2.37E-03
11743498_at	BMP2	bone morphogenetic protein 2	1.65E+00	4.31E-04
11752282_a_at	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)	1.66E+00	1.02E-03

11738958_at	CARD17	caspase recruitment domain family, member 17	1.68E+00	1.79E-03
11721654_at	AGPAT4	1-acylglycerol-3-phosphate O-acyltransferase 4	1.69E+00	4.90E-03
11741990_s_at	CCRL2	chemokine (C-C motif) receptor-like 2	1.69E+00	3.24E-03
11747134_a_at	TRPS1	trichorhinophalangeal syndrome I	1.72E+00	1.38E-03
11763837_s_at	UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	1.75E+00	2.37E-03
11756285_s_at	IGF2BP3	insulin-like growth factor 2 mRNA binding protein 3	1.75E+00	5.73E-04
11721734_s_at	GCH1	GTP cyclohydrolase 1	1.76E+00	6.25E-04
11738960_x_at	CARD17	caspase recruitment domain family, member 17	1.78E+00	3.93E-04
11719000_x_at	DHFR	dihydrofolate reductase	1.79E+00	4.05E-03
11731430_a_at	CYP27A1	cytochrome P450, family 27, subfamily A, polypeptide 1	1.80E+00	2.10E-03
11723044_at	SMAD1	SMAD family member 1	1.80E+00	1.12E-05
11733413_a_at	SLC26A6	solute carrier family 26 (anion exchanger), member 6	1.80E+00	2.07E-03
11759902_at	HACD4	3-hydroxyacyl-CoA dehydratase 4	1.81E+00	1.56E-03
11726895_a_at	IRAK3	interleukin 1 receptor associated kinase 3	1.81E+00	6.17E-04
11718056_a_at	BTBD3	BTB (POZ) domain containing 3	1.81E+00	4.15E-03
11729722_a_at	PIAS2	protein inhibitor of activated STAT 2	1.81E+00	9.96E-05
11749782_x_at	BCAT1	branched chain amino-acid transaminase 1, cytosolic	1.82E+00	1.21E-03
11724606_a_at	MAP2K6	mitogen-activated protein kinase kinase 6	1.82E+00	3.49E-03
11726896_a_at	IRAK3	interleukin 1 receptor associated kinase 3	1.82E+00	3.07E-04
11740702_a_at	HMGB2	high mobility group box 2	1.83E+00	1.64E-03
11739711_a_at	PHACTR2	phosphatase and actin regulator 2	1.84E+00	1.53E-03
11747333_a_at	HSD17B4	hydroxysteroid (17-beta) dehydrogenase 4	1.84E+00	4.33E-03
11747501_a_at	SMAD1	SMAD family member 1	1.85E+00	3.66E-05
11722818_a_at	GGH	gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)	1.86E+00	1.68E-05
11725691_a_at	VWA5A	von Willebrand factor A domain containing 5A	1.86E+00	1.30E-03
11724376_at	PAG1	phosphoprotein membrane anchor with glycosphingolipid microdomains 1	1.86E+00	2.06E-03
11750599_a_at	OSBPL6	oxysterol binding protein-like 6	1.87E+00	2.39E-04
11725603_a_at	HOXA10-HOXA9 /// HOXA9	HOXA10-HOXA9 readthrough /// homeobox A9	1.88E+00	1.30E-03
11755017_a_at	CHCHD7	coiled-coil-helix-coiled-coil-helix domain containing 7	1.89E+00	4.12E-03
11744572_a_at	KLF5	Kruppel-like factor 5 (intestinal)	1.89E+00	3.79E-04
11751292_a_at	PIAS2	protein inhibitor of activated STAT 2	1.89E+00	1.91E-05

11748095_a_at	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)	1.91E+00	2.01E-04
11720716_a_at	FGF13	fibroblast growth factor 13	1.92E+00	1.62E-03
11721215_a_at	TMEM106B	transmembrane protein 106B	1.92E+00	2.12E-03
11752193_a_at	RGL4	ral guanine nucleotide dissociation stimulator-like 4	1.93E+00	4.63E-03
11732299_at	KIAA1715	KIAA1715	1.93E+00	1.13E-03
11744985_a_at	FNDC3A	fibronectin type III domain containing 3A	1.93E+00	4.45E-03
11757461_s_at	BTBD3	BTB (POZ) domain containing 3	1.94E+00	1.03E-03
11741924_a_at	GRB10	growth factor receptor bound protein 10	1.94E+00	3.11E-03
11738049_a_at	ATP2C2	ATPase, Ca++ transporting, type 2C, member 2	1.95E+00	8.11E-04
11759085_s_at	KIAA1715	KIAA1715	1.95E+00	3.65E-04
11731541_at	SESTD1	SEC14 and spectrin domains 1	1.96E+00	1.19E-04
11731542_at	SESTD1	SEC14 and spectrin domains 1	1.96E+00	2.67E-05
11726894_a_at	IRAK3	interleukin 1 receptor associated kinase 3	1.96E+00	2.08E-04
11744581_a_at	KHDRBS3	KH domain containing, RNA binding, signal transduction associated 3	1.97E+00	5.53E-04
11738392_at	LIPN	lipase, family member N	1.97E+00	6.59E-05
11730313_a_at	ERMAP	erythroblast membrane-associated protein (Scianna blood group)	1.98E+00	1.87E-03
11723092_at	FNIP2	folliculin interacting protein 2	1.98E+00	8.64E-07
11718026_a_at	NCOA7	nuclear receptor coactivator 7	1.98E+00	2.77E-03
11731291_a_at	PIAS2	protein inhibitor of activated STAT 2	1.99E+00	5.23E-05
11742017_a_at	IRAK3	interleukin 1 receptor associated kinase 3	2.00E+00	8.20E-05
11750007_a_at	S100Z	S100 calcium binding protein Z	2.00E+00	2.51E-04
11748859_a_at	TFEC	transcription factor EC	2.01E+00	1.38E-03
11717631_s_at	KLHL9	kelch-like family member 9	2.01E+00	1.40E-03
11733954_at	SLC22A15	solute carrier family 22, member 15	2.02E+00	4.10E-03
11758315_s_at	PER2	period circadian clock 2	2.04E+00	1.79E-03
11749728_a_at	KIAA1715	KIAA1715	2.04E+00	7.32E-04
11729941_at	TMEM56	transmembrane protein 56	2.04E+00	1.21E-03
11728361_a_at	CHCHD7	coiled-coil-helix-coiled-coil-helix domain containing 7	2.04E+00	4.58E-03
11758115_s_at	PTBP2	polypyrimidine tract binding protein 2	2.05E+00	4.31E-03
11740045_a_at	PHACTR2	phosphatase and actin regulator 2	2.05E+00	1.16E-03
11722977_at	HOXB5	homeobox B5	2.05E+00	6.57E-05
11759585_at	RAB18	RAB18, member RAS oncogene family	2.06E+00	4.13E-03
11743640_a_at	FNDC3A	fibronectin type III domain containing 3A	2.08E+00	3.25E-03
11729243_s_at	FBN2	fibrillin 2	2.08E+00	1.10E-05
11731539_at	SESTD1	SEC14 and spectrin domains 1	2.09E+00	9.85E-05
11723091_s_at	FNIP2	folliculin interacting protein 2	2.10E+00	1.95E-06
11733477_at	SUCNR1	succinate receptor 1	2.11E+00	1.21E-04

11746816_a_at	PIAS2	protein inhibitor of activated STAT 2	2.13E+00	1.41E-04
11729827_at	FAM110B	family with sequence similarity 110, member B	2.13E+00	4.16E-04
11741441_a_at	FGF13	fibroblast growth factor 13	2.13E+00	1.33E-03
11727403_at	GPC4	glypican 4	2.13E+00	3.05E-04
11745030_at	LINC00597	long intergenic non-protein coding RNA 597	2.14E+00	6.19E-04
11743881_s_at	MREG	melanoregulin	2.14E+00	2.10E-03
11755694_a_at	AGFG1	ArfGAP with FG repeats 1	2.16E+00	2.47E-03
11731605_s_at	HSD17B12	hydroxysteroid (17-beta) dehydrogenase 12	2.16E+00	9.48E-04
11727397_s_at	CTNNA1	catenin (cadherin-associated protein), alpha 1	2.16E+00	3.11E-03
11755851_a_at	UBR4	ubiquitin protein ligase E3 component n-recognin 4	2.16E+00	4.57E-03
11748385_a_at	KIAA0430	KIAA0430	2.17E+00	3.92E-03
11740018_a_at	INSR	insulin receptor	2.18E+00	1.05E-03
11758391_s_at	ZADH2	zinc binding alcohol dehydrogenase domain containing 2	2.18E+00	1.12E-03
11722062_at	REEP3	receptor accessory protein 3	2.18E+00	2.18E-03
11764018_s_at	CLCN3	chloride channel, voltage-sensitive 3	2.20E+00	4.47E-03
11759429_a_at	KLF7	Kruppel-like factor 7 (ubiquitous)	2.22E+00	3.86E-03
11733955_at	SLC22A15	solute carrier family 22, member 15	2.22E+00	4.37E-03
11732481_a_at	ITGAM	integrin, alpha M (complement component 3 receptor 3 subunit)	2.22E+00	3.09E-03
11726661_s_at	GPN3	GPN-loop GTPase 3	2.24E+00	3.06E-03
11721532_x_at	EIF4G3	eukaryotic translation initiation factor 4 gamma, 3	2.25E+00	3.71E-03
11717963_a_at	KRAS	Kirsten rat sarcoma viral oncogene homolog	2.25E+00	3.37E-03
11748797_a_at	AGPAT4	1-acylglycerol-3-phosphate O-acyltransferase 4	2.26E+00	4.39E-03
11732297_at	KIAA1715	KIAA1715	2.26E+00	1.13E-03
11749780_a_at	BCAT1	branched chain amino-acid transaminase 1, cytosolic	2.26E+00	5.18E-04
11719334_a_at	RCBTB1	regulator of chromosome condensation (RCC1) and BTB (POZ) domain containing protein 1	2.27E+00	5.68E-04
11730513_a_at	LSMEM1	leucine-rich single-pass membrane protein 1	2.30E+00	2.73E-03
11748211_a_at	OSBPL9	oxysterol binding protein-like 9	2.30E+00	4.77E-03
11740173_a_at	SOCS5	suppressor of cytokine signaling 5	2.30E+00	2.94E-03
11749895_a_at	BCAT1	branched chain amino-acid transaminase 1, cytosolic	2.32E+00	1.77E-03
11732318_a_at	KHDRBS3	KH domain containing, RNA binding, signal transduction associated 3	2.33E+00	5.77E-04
11731543_at	SESTD1	SEC14 and spectrin domains 1	2.33E+00	4.43E-03
11729210_at	ORC2	origin recognition complex subunit 2	2.35E+00	1.81E-03
11729320_a_at	SOCS5	suppressor of cytokine signaling 5	2.36E+00	9.26E-04

11716847_a_at	SLC43A3	solute carrier family 43, member 3	2.37E+00	4.27E-04
11753440_x_at	GBA	glucosidase, beta, acid	2.38E+00	7.57E-04
11750386_s_at	CTNNA1	catenin (cadherin-associated protein), alpha 1	2.38E+00	9.72E-04
11760189_at	ZSCAN30	zinc finger and SCAN domain containing 30	2.38E+00	2.07E-03
11757515_s_at	ITGAV	integrin alpha V	2.38E+00	4.73E-03
11734230_a_at	LOC145783 /// ZNF280D	uncharacterized LOC145783 /// zinc finger protein 280D	2.39E+00	3.86E-03
11755585_a_at	AZI2	5-azacytidine induced 2	2.39E+00	1.44E-03
11754951_a_at	NCOA7	nuclear receptor coactivator 7	2.40E+00	2.24E-03
11736111_a_at	ARHGAP18	Rho GTPase activating protein 18	2.42E+00	2.10E-03
11757777_s_at	KL	klotho	2.42E+00	9.92E-06
11729857_at	H6PD	hexose-6-phosphate dehydrogenase (glucose 1-dehydrogenase)	2.43E+00	4.43E-03
11715435_s_at	TIMP3	TIMP metallopeptidase inhibitor 3	2.43E+00	3.68E-03
11719963_a_at	HOMER3	homer scaffolding protein 3	2.44E+00	7.04E-05
11736112_a_at	ARHGAP18	Rho GTPase activating protein 18	2.44E+00	1.02E-03
11716187_a_at	ABCB6 /// ATG9A	ATP binding cassette subfamily B member 6 (Langereis blood group) /// autophagy related	2.44E+00	2.17E-03
11743090_a_at	SLC36A4	solute carrier family 36 (proton/amino acid symporter), member 4	2.45E+00	7.78E-05
11718818_s_at	SLC26A6	solute carrier family 26 (anion exchanger), member 6	2.45E+00	1.10E-03

7 References

Abdel-Wahab, O. *et al.* (2011) 'DNMT3A mutational analysis in primary myelofibrosis, chronic myelomonocytic leukemia and advanced phases of myeloproliferative neoplasms.', *Leukemia*, pp. 1219–1220. doi: 10.1038/leu.2011.82.

Abdel-Wahab, O. *et al.* (2012) 'ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression.', *Cancer cell*, 22(2), pp. 180–193. doi: 10.1016/j.ccr.2012.06.032.

Abdel-Wahab, O. I. and Levine, R. L. (2009) 'Primary myelofibrosis: update on definition, pathogenesis, and treatment.', *Annual review of medicine*. United States, 60, pp. 233–245. doi: 10.1146/annurev.med.60.041707.160528.

Adamson, J. W. *et al.* (1976) 'Polycythemia Vera: Stem-Cell and Probable Clonal Origin of the Disease', *New England Journal of Medicine*. Massachusetts Medical Society, 295(17), pp. 913–916. doi: 10.1056/NEJM197610212951702.

Antonioli, E. *et al.* (2005) 'Clinical implications of the JAK2 V617F mutation in essential thrombocythemia', *Leukemia*, 19(10), pp. 1847–1849. doi: 10.1038/sj.leu.2403902.

Araki, M. *et al.* (2016) 'Activation of the thrombopoietin receptor by mutant calreticulin in CALR-mutant myeloproliferative neoplasms.', *Blood*. United States, 127(10), pp. 1307–1316. doi: 10.1182/blood-2015-09-671172.

Arber, D. A. *et al.* (2016) 'The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia', *Blood*, 127(20), pp. 2391–2405. doi: 10.1182/blood-2016-03-643544.

Balduini, C. L. *et al.* (1991) 'Platelet aggregation in platelet-rich plasma and whole blood in 120 patients with myeloproliferative disorders.', *American journal of clinical pathology*. England, 95(1), pp. 82–86. doi: 10.1093/ajcp/95.1.82.

Barbui, T. *et al.* (2011) 'Survival and disease progression in essential thrombocythemia are significantly influenced by accurate morphologic diagnosis: an international study.', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. United States, 29(23), pp. 3179–3184. doi: 10.1200/JCO.2010.34.5298.

Barbui, T. *et al.* (2012) 'Development and validation of an International Prognostic Score of thrombosis in World Health Organization-essential thrombocythemia (IPSET-thrombosis).', *Blood*. United States, 120(26), pp. 5128–33; quiz 5252. doi: 10.1182/blood-2012-07-444067.

Barbui, T. *et al.* (2013) 'Problems and pitfalls regarding WHO-defined diagnosis of early/prefibrotic primary myelofibrosis versus essential thrombocythemia.', *Leukemia*. England, 27(10), pp. 1953–1958. doi: 10.1038/leu.2013.74.

Barosi, G. *et al.* (2008) 'Proposed criteria for the diagnosis of post-polycythemia vera and post-essential thrombocythemia myelofibrosis: a consensus statement from the International Working Group for Myelofibrosis Research and Treatment.', *Leukemia*. England, pp. 437–438. doi: 10.1038/sj.leu.2404914.

Barosi, G. *et al.* (2012) 'Evidence that prefibrotic myelofibrosis is aligned along a clinical and biological continuum featuring primary myelofibrosis.', *PloS one*, 7(4), p. e35631. doi: 10.1371/journal.pone.0035631.

Bartalucci, N., Guglielmelli, P. and Vannucchi, A. M. (2013) 'Rationale for targeting the PI3K/Akt/mTOR pathway in myeloproliferative neoplasms.', *Clinical lymphoma, myeloma & leukemia*. United States, 13 Suppl 2, pp. S307-9. doi: 10.1016/j.clml.2013.07.011.

Bastien, R. R. L. *et al.* (2014) 'Clinical validation of the Prosigna breast cancer prognostic gene signature assay on formalin-fixed paraffin embedded breast cancer tumors with comparison to standard molecular markers.', *Journal of Clinical Oncology*. Wolters Kluwer, 32(15_suppl), pp. e11518–e11518. doi:

10.1200/jco.2014.32.15_suppl.e11518.

Baxter, E. J. *et al.* (2005) 'Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders.', *Lancet (London, England)*. England, 365(9464), pp. 1054–1061. doi: 10.1016/S0140-6736(05)71142-9.

Bedekovics, J. *et al.* (2013) 'Platelet derived growth factor receptor-beta (PDGFRβ) expression is limited to activated stromal cells in the bone marrow and shows a strong correlation with the grade of myelofibrosis.', *Virchows Archiv : an international journal of pathology*. Germany, 463(1), pp. 57–65. doi: 10.1007/s00428-013-1434-0.

Ben-Neriah, Y. *et al.* (1986) 'The Chronic Myelogenous Leukemia--Specific P210 Protein is the Product of the bcr/abl Hybrid Gene', *Science*. American Association for the Advancement of Science, 233(4760), pp. 212–214. Available at: http://www.jstor.org/stable/1697189.

Bennett, J. H. (1845) 'Case of hypertrophy of the spleen and liver, in which death took place from suppuration of the blood', *Edinburgh Med Sug J*, 64, pp. 413–423.

Bock, O. *et al.* (2008) 'Bone morphogenetic proteins are overexpressed in the bone marrow of primary myelofibrosis and are apparently induced by fibrogenic cytokines', *The American journal of pathology*. 2008/03/18. American Society for Investigative Pathology, 172(4), pp. 951–960. doi: 10.2353/ajpath.2008.071030.

Boiocchi, L. *et al.* (2011) 'Increased expression of vascular endothelial growth factor receptor 1 correlates with VEGF and microvessel density in Philadelphia chromosome-negative myeloproliferative neoplasms.', *Journal of clinical pathology*. England, 64(3), pp. 226–231. doi: 10.1136/jcp.2010.083386.

Bonome, T. *et al.* (2008) 'A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer.', *Cancer research*, 68(13), pp. 5478–5486. doi: 10.1158/0008-5472.CAN-07-6595.

Bruns, I. *et al.* (2014) 'Megakaryocytes regulate hematopoietic stem cell quiescence through CXCL4 secretion.', *Nature medicine*, 20(11), pp. 1315–1320. doi: 10.1038/nm.3707.

Buhr, T. *et al.* (2012) 'European Bone Marrow Working Group trial on reproducibility of World Health Organization criteria to discriminate essential thrombocythemia from prefibrotic primary myelofibrosis', *Haematologica*. 2011/11/04. Ferrata Storti Foundation, 97(3), pp. 360–365. doi: 10.3324/haematol.2011.047811.

Bullinger, L. *et al.* (2004) 'Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia.', *The New England journal of medicine*. United States, 350(16), pp. 1605–1616. doi: 10.1056/NEJMoa031046.

Calvi, L. M. *et al.* (2003) 'Osteoblastic cells regulate the haematopoietic stem cell niche.', *Nature*. England, 425(6960), pp. 841–846. doi: 10.1038/nature02040.

Campbell, P. J. *et al.* (2005) 'Definition of subtypes of essential thrombocythaemia and relation to polycythaemia vera based on JAK2 V617F mutation status: a prospective study.', *Lancet (London, England)*. England, 366(9501), pp. 1945–1953. doi: 10.1016/S0140-6736(05)67785-9.

Cervantes, F. *et al.* (2009) 'New prognostic scoring system for primary myelofibrosis based on a study of the International working group for myelofibrosis research and treatment', *Blood*, 113(13), pp. 2895–2901. doi: 10.1182/blood-2008-07-170449.

Cervantes, F. *et al.* (2013) 'Three-year efficacy, safety, and survival findings from COMFORT-II, a phase 3 study comparing ruxolitinib with best available therapy for myelofibrosis.', *Blood*. United States, 122(25), pp. 4047–4053. doi: 10.1182/blood-2013-02-485888.

Cervantes, F., Mesa, R. and Barosi, G. (2007) 'New and old treatment modalities in primary myelofibrosis.', *Cancer journal (Sudbury, Mass.)*. United States, 13(6), pp. 377–383. doi: 10.1097/PPO.0b013e31815a7c0a.

Chachoua, I. *et al.* (2016) 'Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants.', *Blood*. United States, 127(10), pp. 1325–1335. doi: 10.1182/blood-2015-11-681932.

Chen, E. *et al.* (2015) 'Distinct effects of concomitant Jak2V617F expression and Tet2 loss in mice promote disease progression in myeloproliferative neoplasms.', *Blood*, 125(2), pp. 327–335. doi: 10.1182/blood-2014-04-567024.

Chen, J. *et al.* (2009) 'ToppGene Suite for gene list enrichment analysis and candidate gene prioritization', *Nucleic Acids Research*, 37(suppl_2), pp. W305–W311. doi: 10.1093/nar/gkp427.

Cicchetti, D. V (1992) 'Neural networks and diagnosis in the clinical laboratory: state of the art.', *Clinical chemistry*. England, pp. 9–10.

Ciurea, S. O. *et al.* (2007) 'Pivotal contributions of megakaryocytes to the biology of idiopathic myelofibrosis.', *Blood*, 110(3), pp. 986–993. doi: 10.1182/blood-2006-12-064626.

Coltro, G. *et al.* (2020) 'RAS/CBL mutations predict resistance to JAK inhibitors in myelofibrosis and are associated with poor prognostic features', *Blood advances*. American Society of Hematology, 4(15), pp. 3677–3687. doi: 10.1182/bloodadvances.2020002175.

Cortelazzo, S. *et al.* (1995) 'Hydroxyurea for patients with essential thrombocythemia and a high risk of thrombosis.', *The New England journal of medicine*. United States, 332(17), pp. 1132–1136. doi: 10.1056/NEJM199504273321704.

Cross, N. C. P. and Reiter, A. (2002) 'Tyrosine kinase fusion genes in chronic myeloproliferative diseases', *Leukemia*, 16(7), pp. 1207–1212. doi: 10.1038/sj.leu.2402556.

Cruz, J. A. and Wishart, D. S. (2007) 'Applications of machine learning in cancer prediction and prognosis', *Cancer informatics*. Libertas Academica, 2, pp. 59–77. Available at: https://pubmed.ncbi.nlm.nih.gov/19458758.

Dai, X. *et al.* (2015) 'Breast cancer intrinsic subtype classification, clinical use and future trends', *American Journal of Cancer Research*, 5(10), pp. 2929–2943.

Daley, G. Q., Van Etten, R. A. and Baltimore, D. (1990) 'Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome.', *Science (New York, N.Y.)*. United States, 247(4944), pp. 824–830. doi: 10.1126/science.2406902.

Dameshek, W. (1951) 'Some Speculations on the Myeloproliferative Syndromes', *Blood*, 6(4), pp. 372–375. doi: https://doi.org/10.1182/blood.V6.4.372.372.

Deininger, M. *et al.* (2015) 'The effect of long-term ruxolitinib treatment on JAK2p.V617F allele burden in patients with myelofibrosis.', *Blood*, 126(13), pp. 1551–1554. doi: 10.1182/blood-2015-03-635235.

Delhommeau, F. *et al.* (2006) 'Oncogenic mechanisms in myeloproliferative disorders.', *Cellular and molecular life sciences : CMLS*. Switzerland, 63(24), pp. 2939–2953. doi: 10.1007/s00018-006-6272-7.

Elf, S. *et al.* (2016) 'Mutant Calreticulin Requires Both Its Mutant C-terminus and the Thrombopoietin Receptor for Oncogenic Transformation.', *Cancer discovery*, 6(4), pp. 368–381. doi: 10.1158/2159-8290.CD-15-1434.

Elliott, M. A. and Tefferi, A. (2005) 'Thrombosis and haemorrhage in polycythaemia vera and essential thrombocythaemia.', *British journal of haematology*. England, 128(3), pp. 275–290. doi: 10.1111/j.1365-2141.2004.05277.x.

Engelman, J. A. (2009) 'Targeting PI3K signalling in cancer: opportunities, challenges and limitations.', *Nature reviews. Cancer*. England, 9(8), pp. 550–562. doi: 10.1038/nrc2664.

Epstein, E. and Goedel, A. (1934) 'Hemorrhagic thrombocythemia with a cascular, sclerotic spleen', *Virchows Arch*, 293, pp. 233–248.

Fialkow, P. J. *et al.* (1981) 'Evidence that essential thrombocythemia is a clonal disorder with origin in a multipotent stem cell', *Blood*, 58(5), pp. 916–919. doi: 10.1182/blood.V58.5.916.916.

Fialkow, P. J., Gartler, S. M. and Yoshida, A. (1967) 'Clonal origin of chronic myelocytic leukemia in man', *Proceedings of the National Academy of Sciences of the United States of America*, 58(4), pp. 1468–1471. doi: 10.1073/pnas.58.4.1468.

Finazzi, G. and Barbui, T. (2007) 'How I treat patients with polycythemia vera.', *Blood*. United States, 109(12), pp. 5104–5111. doi: 10.1182/blood-2006-12-038968.

Finazzi, G. and Harrison, C. (2005) 'Essential thrombocythemia.', *Seminars in hematology*. United States, 42(4), pp. 230–238. doi: 10.1053/j.seminhematol.2005.05.022.

Fisher, D. A. C. *et al.* (2017) 'Mass cytometry analysis reveals hyperactive NF Kappa B

signaling in myelofibrosis and secondary acute myeloid leukemia.', *Leukemia*, 31(9), pp. 1962–1974. doi: 10.1038/leu.2016.377.

Fisher, D. A. C. *et al.* (2019) 'Cytokine production in myelofibrosis exhibits differential responsiveness to JAK-STAT, MAP kinase, and NFκB signaling', *Leukemia*, 33(8), pp. 1978–1995. doi: 10.1038/s41375-019-0379-y.

Gagelmann, N. *et al.* (2019) 'Comprehensive clinical-molecular transplant scoring system for myelofibrosis undergoing stem cell transplantation.', *Blood*. United States, 133(20), pp. 2233–2242. doi: 10.1182/blood-2018-12-890889.

Gangat, N. *et al.* (2011) 'DIPSS plus: a refined Dynamic International Prognostic Scoring System for primary myelofibrosis that incorporates prognostic information from karyotype, platelet count, and transfusion status.', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. United States, 29(4), pp. 392–397. doi: 10.1200/JCO.2010.32.2446.

Gangat, N. and Tefferi, A. (2020) 'Myelofibrosis biology and contemporary management', *British Journal of Haematology*, 191(2), pp. 152–170. doi: 10.1111/bjh.16576.

Garmezy, B. *et al.* (2021) 'A provider's guide to primary myelofibrosis: pathophysiology, diagnosis, and management', *Blood Reviews*. Elsevier Ltd, 45(April 2020), p. 100691. doi: 10.1016/j.blre.2020.100691.

Geyer, H. L. and Mesa, R. A. (2014) 'Therapy for myeloproliferative neoplasms: when, which agent, and how?', *Blood*. United States, 124(24), pp. 3529–3537. doi: 10.1182/blood-2014-05-577635.

Goetz, L. H. and Schork, N. J. (2018) 'Personalized medicine: motivation, challenges, and progress.', *Fertility and sterility*, 109(6), pp. 952–963. doi: 10.1016/j.fertnstert.2018.05.006.

Green, A. R. (1996) 'Pathogenesis of polycythaemia vera.', *Lancet (London, England)*. England, 347(9005), pp. 844–845. doi: 10.1016/s0140-6736(96)91338-0.

Grinfeld, J. *et al.* (2018) 'Classification and Personalized Prognosis in Myeloproliferative Neoplasms', *New England Journal of Medicine*, 379(15), pp. 1416–1430. doi:

10.1056/nejmoa1716614.

Guertin, D. A. and Sabatini, D. M. (2007) 'Defining the role of mTOR in cancer.', *Cancer cell*. United States, 12(1), pp. 9–22. doi: 10.1016/j.ccr.2007.05.008.

Guglielmelli, P. *et al.* (2007) 'MicroRNA expression profile in granulocytes from primary myelofibrosis patients.', *Experimental hematology*. Netherlands, 35(11), pp. 1708–1718. doi: 10.1016/j.exphem.2007.08.020.

Guglielmelli, P., Biamonte, F., *et al.* (2011) 'EZH2 mutational status predicts poor survival in myelofibrosis.', *Blood*. United States, 118(19), pp. 5227–5234. doi: 10.1182/blood-2011-06-363424.

Guglielmelli, P., Barosi, G., *et al.* (2011) 'Safety and efficacy of everolimus, a mTOR inhibitor, as single agent in a phase 1/2 study in patients with myelofibrosis', *Blood*, 118(8), pp. 2069–2076. doi: 10.1182/blood-2011-01-330563.

Guglielmelli, Paola *et al.* (2015) 'Small RNA Sequencing Uncovers New miRNAs and moRNAs Differentially Expressed in Normal and Primary Myelofibrosis CD34+ Cells.', *PloS one*, 10(10), p. e0140445. doi: 10.1371/journal.pone.0140445.

Guglielmelli, P *et al.* (2015) 'Validation of the differential prognostic impact of type 1/type 1-like versus type 2/type 2-like CALR mutations in myelofibrosis', *Blood cancer journal*. Nature Publishing Group, 5(10), pp. e360–e360. doi: 10.1038/bcj.2015.90.

Guglielmelli, P. *et al.* (2017) 'Presentation and outcome of patients with 2016 WHO diagnosis of prefibrotic and overt primary myelofibrosis.', *Blood*. United States, 129(24), pp. 3227–3236. doi: 10.1182/blood-2017-01-761999.

Guglielmelli, P., Lasho, Terra L, *et al.* (2018) 'MIPSS70: Mutation-Enhanced International Prognostic Score System for Transplantation-Age Patients With Primary Myelofibrosis.', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. United States, 36(4), pp. 310–318. doi: 10.1200/JCO.2017.76.4886.

Guglielmelli, P., Lasho, Terra L., *et al.* (2018) 'MIPSS70: Mutation-enhanced international prognostic score system for transplantation-age patients with

primary myelofibrosis', *Journal of Clinical Oncology*, 36(4), pp. 310–318. doi: 10.1200/JCO.2017.76.4886.

Guglielmelli, P. and Vannucchi, A. M. (2016) 'The prognostic impact of bone marrow fibrosis in primary myelofibrosis.', *American journal of hematology*. United States, pp. E454-5. doi: 10.1002/ajh.24482.

Güler, E. N. (2017) 'Gene Expression Profiling in Breast Cancer and Its Effect on Therapy Selection in Early-Stage Breast Cancer.', *European journal of breast health*, 13(4), pp. 168–174. doi: 10.5152/ejbh.2017.3636.

Harrison, C. *et al.* (2012) 'JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis.', *The New England journal of medicine*. United States, 366(9), pp. 787–798. doi: 10.1056/NEJMoa1110556.

Harrison, C. N. *et al.* (1999) 'A large proportion of patients with a diagnosis of essential thrombocythemia do not have a clonal disorder and may be at lower risk of thrombotic complications.', *Blood*. United States, 93(2), pp. 417–424.

Harwich, E. and Laycock, K. (2018) 'Thinking on its own: AI in the NHS', *Reform*, Jan(January), pp. 1–60. Available at: https://reform.uk/research/thinking-its-own-ai-nhs.

Hernández-Boluda, J.-C. *et al.* (2014) 'The International Prognostic Scoring System does not accurately discriminate different risk categories in patients with post-essential thrombocythemia and post-polycythemia vera myelofibrosis.', *Haematologica*, pp. e55-7. doi: 10.3324/haematol.2013.101733.

Hess, G. *et al.* (1994) 'Molecular analysis of the erythropoietin receptor system in patients with polycythaemia vera.', *British journal of haematology*. England, 88(4), pp. 794–802. doi: 10.1111/j.1365-2141.1994.tb05119.x.

Heuck, G. (1879) 'Zwei Fälle von Leukämie mit eigenthümlichem Blut- resp. Knochenmarksbefund', *Archiv für pathologische Anatomie und Physiologie und für klinische Medicin*, 78(3), pp. 475–496. doi: 10.1007/BF01878089.

How, J., Hobbs, G. S. and Mullally, A. (2019) 'Mutant calreticulin in myeloproliferative neoplasms.', *Blood*, 134(25), pp. 2242–2248. doi: 10.1182/blood.2019000622.

Hungerford, D. A. and Nowell, P. C. (1960) 'A minute chromosome in human chronic granulocytic leukemia', *Science*, 132, pp. 1497–1499.

Hussein, K. *et al.* (2009) 'MicroRNA expression profiling of megakaryocytes in primary myelofibrosis and essential thrombocythemia', *Platelets*. Institute of Pathology, 30625, Hannover, Germany. Hussein.Kais@MH-Hannover.de, 20(6), pp. 391–400. doi: 10.1080/09537100903114537.

Iland, H. J. *et al.* (1983) 'Essential thrombocythemia: clinical and laboratory characteristics at presentation.', *Transactions of the Association of American Physicians*. United States, 96, pp. 165–174.

Jacobson, R. J., Salo, A. and Fialkow, P. J. (1978) 'Agnogenic Myeloid Metaplasia: A Clonal Proliferation of Hematopoietic Stem Cells With Secondary Myelofibrosis', *Blood*, 51(2), pp. 189–194. doi: https://doi.org/10.1182/blood.V51.2.189.189.

Jacquelin, S. *et al.* (2018) 'Jak2V617F and Dnmt3a loss cooperate to induce myelofibrosis through activated enhancer-driven inflammation.', *Blood*. United States, 132(26), pp. 2707–2721. doi: 10.1182/blood-2018-04-846220.

James, C. *et al.* (2005) 'A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera.', *Nature*. England, 434(7037), pp. 1144–1148. doi: 10.1038/nature03546.

James, C. *et al.* (2006) 'Detection of JAK2 V617F as a first intention diagnostic test for erythrocytosis', *Leukemia*, 20(2), pp. 350–353. doi: 10.1038/sj.leu.2404069.

Kaushansky, K. *et al.* (1994) 'Promotion of megakaryocyte progenitor expansion and differentiation by the c-Mpl ligand thrombopoietin.', *Nature*. England, 369(6481), pp. 568–571. doi: 10.1038/369568a0.

Khan, I. *et al.* (2013) 'AKT is a therapeutic target in myeloproliferative neoplasms.', *Leukemia*, 27(9), pp. 1882–1890. doi: 10.1038/leu.2013.167.

Kittur, J. *et al.* (2007) 'Clinical correlates of JAK2V617F allele burden in essential thrombocythemia.', *Cancer*. United States, 109(11), pp. 2279–2284. doi: 10.1002/cncr.22663.

Klampfl, T. *et al.* (2013) 'Somatic Mutations of Calreticulin in Myeloproliferative

Neoplasms', *New England Journal of Medicine*. Massachusetts Medical Society, 369(25), pp. 2379–2390. doi: 10.1056/NEJMoa1311347.

Kourou, K. *et al.* (2015) 'Machine learning applications in cancer prognosis and prediction', *Computational and Structural Biotechnology Journal*. Elsevier B.V., 13, pp. 8–17. doi: 10.1016/j.csbj.2014.11.005.

Kralovics, R. *et al.* (2005) 'A gain-of-function mutation of JAK2 in myeloproliferative disorders.', *The New England journal of medicine*. United States, 352(17), pp. 1779–1790. doi: 10.1056/NEJMoa051113.

Kramer, F. *et al.* (2020) 'Platelet-derived growth factor receptor β activation and regulation in murine myelofibrosis', *Haematologica*. 2019/10/31. Ferrata Storti Foundation, 105(8), pp. 2083–2094. doi: 10.3324/haematol.2019.226332.

Kremyanskaya, M. *et al.* (2021) 'Pelabresib (CPI-0610) Monotherapy in Patients with Myelofibrosis - Update of Clinical and Translational Data from the Ongoing Manifest Trial', *Blood*, 138, p. 141. doi: https://doi.org/10.1182/blood-2021-150172.

Kröger, N. M. *et al.* (2015) 'Indication and management of allogeneic stem cell transplantation in primary myelofibrosis: a consensus process by an EBMT/ELN international working group.', *Leukemia*. England, 29(11), pp. 2126–2133. doi: 10.1038/leu.2015.233.

Lasho, T. L. *et al.* (2012) 'SRSF2 mutations in primary myelofibrosis: significant clustering with IDH mutations and independent association with inferior overall and leukemia-free survival.', *Blood*. United States, 120(20), pp. 4168–4171. doi: 10.1182/blood-2012-05-429696.

Leiva, O. *et al.* (2017) 'The role of the extracellular matrix in primary myelofibrosis.', *Blood cancer journal*, 7(2), p. e525. doi: 10.1038/bcj.2017.6.

Levine, R. L. *et al.* (2005) 'Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis.', *Cancer cell*. United States, 7(4), pp. 387–397. doi: 10.1016/j.ccr.2005.03.023.

Lin, C. H. S., Kaushansky, K. and Zhan, H. (2016) 'JAK2(V617F)-mutant vascular niche contributes to JAK2(V617F) clonal expansion in myeloproliferative neoplasms', *Blood cells, molecules & diseases*. 2016/11/04, 62, pp. 42–48. doi: 10.1016/j.bcmd.2016.09.004.

Maclin, P. S. *et al.* (1991) 'Using neural networks to diagnose cancer.', *Journal of medical systems*. United States, 15(1), pp. 11–19. doi: 10.1007/BF00993877.

Marchioli, R. *et al.* (2013) 'Cardiovascular events and intensity of treatment in polycythemia vera.', *The New England journal of medicine*. United States, 368(1), pp. 22–33. doi: 10.1056/NEJMoa1208500.

Marneth, A. E. and Mullally, A. (2020) 'Busy signal: platelet-derived growth factor activation in myelofibrosis', *Haematologica*. Ferrata Storti Foundation, 105(8), pp. 1988–1990. doi: 10.3324/haematol.2020.253708.

Marty, C. *et al.* (2014) 'Calr Mutants Retroviral Mouse Models Lead to a Myeloproliferative Neoplasm Mimicking an Essential Thrombocythemia Progressing to a Myelofibrosis', *Blood*, 124(21), p. 157. doi: 10.1182/blood.V124.21.157.157.

Marty, C. *et al.* (2016) 'Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis.', *Blood*. United States, 127(10), pp. 1317–1324. doi: 10.1182/blood-2015-11-679571.

Masarova, L. *et al.* (2017) 'Patients with post-essential thrombocythemia and post-polycythemia vera differ from patients with primary myelofibrosis.', *Leukemia research*, 59, pp. 110–116. doi: 10.1016/j.leukres.2017.06.001.

Mascarenhas, J. (2014) 'Rationale for combination therapy in myelofibrosis.', *Best practice & research. Clinical haematology*. Netherlands, 27(2), pp. 197–208. doi: 10.1016/j.beha.2014.07.009.

Mascarenhas, J. and Hoffman, R. (2012) 'Ruxolitinib: the first FDA approved therapy for the treatment of myelofibrosis.', *Clinical cancer research : an official journal of the American Association for Cancer Research*. United States, 18(11), pp. 3008–3014. doi: 10.1158/1078-0432.CCR-11-3145.

McLornan, D. P. *et al.* (2019) 'State-of-the-art review: allogeneic stem cell transplantation for myelofibrosis in 2019', *Haematologica*. Guy's and St. Thomas' NHS Foundation Trust, Department of Haematology, Guy's Tower, Great Maze Pond, London, UK donal.mclornan@nhs.net., 104(4), pp. 659–668. doi: 10.3324/haematol.2018.206151.

Mead, A. J. and Mullally, A. (2017) 'Myeloproliferative neoplasm stem cells.', *Blood*, 129(12), pp. 1607–1616. doi: 10.1182/blood-2016-10-696005.

Medinger, M. and Passweg, J. (2014) 'Angiogenesis in myeloproliferative neoplasms, new markers and future directions', *Memo*. 2014/05/22. Springer Vienna, 7, pp. 206–210. doi: 10.1007/s12254-014-0142-z.

Michiels, J. J. *et al.* (2006) 'Clinical and laboratory features, pathobiology of platelet-mediated thrombosis and bleeding complications, and the molecular etiology of essential thrombocythemia and polycythemia vera: therapeutic implications.', *Seminars in thrombosis and hemostasis*. United States, 32(3), pp. 174–207. doi: 10.1055/s-2006-939431.

Mills, K. I. *et al.* (2009) 'Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome.', *Blood*. United States, 114(5), pp. 1063–1072. doi: 10.1182/blood-2008-10-187203.

Mishchenko, E. and Tefferi, A. (2010) 'Treatment options for hydroxyurea-refractory disease complications in myeloproliferative neoplasms: JAK2 inhibitors, radiotherapy, splenectomy and transjugular intrahepatic portosystemic shunt.', *European journal of haematology*. England, 85(3), pp. 192–199. doi: 10.1111/j.1600-0609.2010.01480.x.

Mitchell, T. (1997) *Machine Learining*. Edited by McGraw Hill.

Moulard, O. *et al.* (2014) 'Epidemiology of myelofibrosis, essential thrombocythemia, and polycythemia vera in the European Union.', *European journal of haematology*. England, 92(4), pp. 289–297. doi: 10.1111/ejh.12256.

Mudireddy, M. *et al.* (2018) 'Prefibrotic versus overtly fibrotic primary myelofibrosis: clinical, cytogenetic, molecular and prognostic comparisons.', *British journal of*

haematology. England, pp. 594–597. doi: 10.1111/bjh.14838.

Nangalia, J. *et al.* (2013) 'Somatic CALR Mutations in Myeloproliferative Neoplasms with Nonmutated JAK2', *New England Journal of Medicine*. Massachusetts Medical Society, 369(25), pp. 2391–2405. doi: 10.1056/NEJMoa1312542.

Nangalia, J. and Green, T. R. (2014) 'The evolving genomic landscape of myeloproliferative neoplasms', *Hematology*, 2014(1), pp. 287–296. doi: 10.1182/asheducation-2014.1.287.

Ng, S. W. K. *et al.* (2016) 'A 17-gene stemness score for rapid determination of risk in acute leukaemia', *Nature*. Nature Publishing Group, 540(7633), pp. 433–437. doi: 10.1038/nature20598.

Nicolosi, M. *et al.* (2018) 'Sex and degree of severity influence the prognostic impact of anemia in primary myelofibrosis: analysis based on 1109 consecutive patients.', *Leukemia*, 32(5), pp. 1254–1258. doi: 10.1038/s41375-018-0028-x.

Nielsen, T. *et al.* (2014) 'Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens.', *BMC cancer*, 14, p. 177. doi: 10.1186/1471-2407-14-177.

Norfo, R. *et al.* (2014) 'miRNA-mRNA integrative analysis in primary myelofibrosis CD34+ cells: role of miR-155/JARID2 axis in abnormal megakaryopoiesis.', *Blood*, 124(13), pp. e21-32. doi: 10.1182/blood-2013-12-544197.

NOWELL, P. C. and HUNGERFORD, D. A. (1960) 'Chromosome studies on normal and leukemic human leukocytes.', *Journal of the National Cancer Institute*. United States, 25, pp. 85–109.

Ohnstad, H. O. *et al.* (2017) 'Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up.', *Breast cancer research : BCR*, 19(1), p. 120. doi: 10.1186/s13058-017-0911-9.

Ortmann, C. A. *et al.* (2015) 'Effect of Mutation Order on Myeloproliferative Neoplasms', *New England Journal of Medicine*. Massachusetts Medical Society, 372(7), pp. 601–612. doi: 10.1056/NEJMoa1412098.

Palandri, F. *et al.* (2018) 'Differences in presenting features, outcome and prognostic models in patients with primary myelofibrosis and post-polycythemia vera and/or post-essential thrombocythemia myelofibrosis treated with ruxolitinib. New perspective of the MYSEC-PM in a large m', *Seminars in hematology*. United States, 55(4), pp. 248–255. doi: 10.1053/j.seminhematol.2018.05.013.

Parker, J. S. *et al.* (2009) 'Supervised risk predictor of breast cancer based on intrinsic subtypes.', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27(8), pp. 1160–1167. doi: 10.1200/JCO.2008.18.1370.

Passamonti, F. *et al.* (2010) 'A dynamic prognostic model to predict survival in primary myelofibrosis: A study by the IWG-MRT (International Working Group for Myeloproliferative Neoplasms Research and Treatment)', *Blood*, 115(9), pp. 1703–1708. doi: 10.1182/blood-2009-09-245837.

Passamonti, F., Elena, C., *et al.* (2011) 'Molecular and clinical features of the myeloproliferative neoplasm associated with JAK2 exon 12 mutations.', *Blood*. United States, 117(10), pp. 2813–2816. doi: 10.1182/blood-2010-11-316810.

Passamonti, F., Maffioli, M., *et al.* (2011) 'Myeloproliferative neoplasms: from JAK2 mutations discovery to JAK2 inhibitor therapies', *Oncotarget*. Impact Journals LLC, 2(6), pp. 485–490. doi: 10.18632/oncotarget.281.

Passamonti, F. *et al.* (2017) 'A clinical-molecular prognostic model to predict survival in patients with post polycythemia vera and post essential thrombocythemia myelofibrosis.', *Leukemia*. England, 31(12), pp. 2726–2731. doi: 10.1038/leu.2017.169.

Pellagatti, A. *et al.* (2013) 'Identification of gene expression-based prognostic markers in the hematopoietic stem cells of patients with myelodysplastic syndromes', *Journal of Clinical Oncology*, 31(28), pp. 3557–3564. doi: 10.1200/JCO.2012.45.5626.

Penna, D. *et al.* (2019) '20+ Years and alive with primary myelofibrosis: Phenotypic signature of very long-lived patients.', *American journal of hematology*. United States, 94(3), pp. 286–290. doi: 10.1002/ajh.25351.

Pikman, Y. *et al.* (2006) 'MPLW515L is a novel somatic activating mutation in

myelofibrosis with myeloid metaplasia.', *PLoS medicine*, 3(7), p. e270. doi: 10.1371/journal.pmed.0030270.

Portet, S. (2020) 'A primer on model selection using the Akaike Information Criterion', *Infectious Disease Modelling*, 5, pp. 111–128. doi: 10.1016/j.idm.2019.12.010.

Prchal, J. F. and Axelrad, A. A. (1974) 'Letter: Bone-marrow responses in polycythemia vera.', *The New England journal of medicine*. United States, 290(24), p. 1382. doi: 10.1056/nejm197406132902419.

Radaelli, F. *et al.* (2008) 'Second malignancies in essential thrombocythemia (ET): a retrospective analysis of 331 patients with long-term follow-up from a single institution.', *Hematology (Amsterdam, Netherlands)*. England, 13(4), pp. 195–202. doi: 10.1179/102453308X316022.

Rai, S. *et al.* (2019) 'IL-1β Secreted from Mutant Cells Carrying JAK2-V617Ffavors Early Clonal Expansion and Promotes MPN Disease Initiation and Progression', *Blood*, 134(Supplement_1), p. 307. doi: 10.1182/blood-2019-129800.

Rampal, R., Ahn, J., *et al.* (2014) 'Genomic and functional analysis of leukemic transformation of myeloproliferative neoplasms.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(50), pp. E5401-10. doi: 10.1073/pnas.1407792111.

Rampal, R., Al-Shahrour, F., *et al.* (2014) 'Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis.', *Blood*, 123(22), pp. e123-33. doi: 10.1182/blood-2014-02-554634.

Rollison, D. E. *et al.* (2008) 'Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001-2004, using data from the NAACCR and SEER programs.', *Blood*. United States, 112(1), pp. 45–52. doi: 10.1182/blood-2008-01-134858.

Rotunno, G. *et al.* (2016) 'Epidemiology and clinical relevance of mutations in postpolycythemia vera and postessential thrombocythemia myelofibrosis: A study on 359 patients of the AGIMM group.', *American journal of hematology*. United States, 91(7), pp. 681–686. doi: 10.1002/ajh.24377.

ROWLEY, J. D. (1973) 'A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining', *Nature*, 243(5405), pp. 290–293. doi: 10.1038/243290a0.

Rumi, E. and Cazzola, M. (2017) 'Diagnosis, risk stratification, and response evaluation in classical myeloproliferative neoplasms.', *Blood*, 129(6), pp. 680–692. doi: 10.1182/blood-2016-10-695957.

Di Sanzo, M. *et al.* (2017) 'Clinical Applications of Personalized Medicine: A New Paradigm and Challenge.', *Current pharmaceutical biotechnology*. Netherlands, 18(3), pp. 194–203. doi: 10.2174/1389201018666170224105600.

Scott, L. M. *et al.* (2007) 'JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis.', *The New England journal of medicine*, 356(5), pp. 459–468. doi: 10.1056/NEJMoa065202.

Shimizu, T. *et al.* (2016) 'Loss of Ezh2 synergizes with JAK2-V617F in initiating myeloproliferative neoplasms and promoting myelofibrosis.', *The Journal of experimental medicine*, 213(8), pp. 1479–1496. doi: 10.1084/jem.20151136.

Shiozawa, Y. *et al.* (2017) 'Gene expression and risk of leukemic transformation in myelodysplasia', *Blood*, 130(24), pp. 2642–2653. doi: 10.1182/blood-2017-05-783050.

Shtivelman, E. *et al.* (1985) 'Fused transcript of abl and bcr genes in chronic myelogenous leukaemia', *Nature*, 315(6020), pp. 550–554. doi: 10.1038/315550a0.

Simes, R. J. (1985) 'Treatment selection for cancer patients: application of statistical decision theory to the treatment of advanced ovarian cancer.', *Journal of chronic diseases*. England, 38(2), pp. 171–186. doi: 10.1016/0021-9681(85)90090-6.

Spivak, J. L. (2003) 'Diagnosis of the myeloproliferative disorders: resolving phenotypic mimicry.', *Seminars in hematology*. United States, 40(1 Suppl 1), pp. 1–5. doi: 10.1053/shem.2003.50026.

Spivak, J. L. (2017) 'Myeloproliferative Neoplasms.', *The New England journal of medicine*. United States, 376(22), pp. 2168–2181. doi: 10.1056/NEJMra1406186.

Stam, K. *et al.* (1985) 'Evidence of a New Chimeric bcr/c-abl mRNA in Patients with Chronic Myelocytic Leukemia and the Philadelphia Chromosome', *New England Journal of Medicine*. Massachusetts Medical Society, 313(23), pp. 1429–1433. doi: 10.1056/NEJM198512053132301.

Stoner, S. A. *et al.* (2019) 'Hippo kinase loss contributes to del(20q) hematologic malignancies through chronic innate immune activation', *Blood*. American Society of Hematology, 134(20), pp. 1730–1744. doi: 10.1182/blood.2019000170.

Szuber, N. *et al.* (2019) '3023 Mayo Clinic Patients With Myeloproliferative Neoplasms: Risk-Stratified Comparison of Survival and Outcomes Data Among Disease Subgroups.', *Mayo Clinic proceedings*. England, 94(4), pp. 599–610. doi: 10.1016/j.mayocp.2018.08.022.

Tefferi, A. (2005) 'Pathogenesis of myelofibrosis with myeloid metaplasia.', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. United States, 23(33), pp. 8520–8530. doi: 10.1200/JCO.2004.00.9316.

Tefferi, A., Cortes, J., *et al.* (2006) 'Lenalidomide therapy in myelofibrosis with myeloid metaplasia.', *Blood*. United States, 108(4), pp. 1158–1164. doi: 10.1182/blood-2006-02-004572.

Tefferi, A., Lasho, T. L., *et al.* (2006) 'The clinical phenotype of wild-type, heterozygous, and homozygous JAK2V617F in polycythemia vera.', *Cancer*. United States, 106(3), pp. 631–635. doi: 10.1002/cncr.21645.

Tefferi, A. *et al.* (2011) 'Circulating interleukin (IL)-8, IL-2R, IL-12, and IL-15 levels are independently prognostic in primary myelofibrosis: a comprehensive cytokine profiling study.', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. United States, 29(10), pp. 1356–1363. doi: 10.1200/JCO.2010.32.9490.

Tefferi, A. *et al.* (2012) 'IDH mutations in primary myelofibrosis predict leukemic transformation and shortened survival: clinical evidence for leukemogenic collaboration with JAK2V617F.', *Leukemia*, 26(3), pp. 475–480. doi: 10.1038/leu.2011.253.

Tefferi, A. *et al.* (2013) 'Survival and prognosis among 1545 patients with contemporary

polycythemia vera: an international study.', *Leukemia*, 27(9), pp. 1874–1881. doi: 10.1038/leu.2013.163.

Tefferi, A, Thiele, J., *et al.* (2014) 'An overview on CALR and CSF3R mutations and a proposal for revision of WHO diagnostic criteria for myeloproliferative neoplasms.', *Leukemia*. England, 28(7), pp. 1407–1413. doi: 10.1038/leu.2014.35.

Tefferi, A, Guglielmelli, P., *et al.* (2014) 'CALR and ASXL1 mutations-based molecular prognostication in primary myelofibrosis: an international study of 570 patients.', *Leukemia*. England, 28(7), pp. 1494–1500. doi: 10.1038/leu.2014.57.

Tefferi, Ayalew, Guglielmelli, P., Larson, D. R., *et al.* (2014) 'Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis.', *Blood*, 124(16), pp. 2507–13; quiz 2615. doi: 10.1182/blood-2014-05-579136.

Tefferi, Ayalew, Lasho, T. L., Tischer, A., *et al.* (2014) 'The prognostic advantage of calreticulin mutations in myelofibrosis might be confined to type 1 or type 1-like CALR variants', *Blood*. American Society of Hematology, 124(15), pp. 2465–2466. doi: 10.1182/blood-2014-07-588426.

Tefferi, Ayalew, Wassie, E. A., Guglielmelli, P., *et al.* (2014) 'Type 1 versus Type 2 calreticulin mutations in essential thrombocythemia: a collaborative study of 1027 patients.', *American journal of hematology*. United States, 89(8), pp. E121-4. doi: 10.1002/ajh.23743.

Tefferi, A, Lasho, T. L., *et al.* (2014) 'Type 1 vs type 2 calreticulin mutations in primary myelofibrosis: differences in phenotype and prognostic impact.', *Leukemia*. England, pp. 1568–1570. doi: 10.1038/leu.2014.83.

Tefferi, A, Finke, C. M., *et al.* (2014) 'U2AF1 mutations in primary myelofibrosis are strongly associated with anemia and thrombocytopenia despite clustering with JAK2V617F and normal karyotype.', *Leukemia*. England, pp. 431–433. doi: 10.1038/leu.2013.286.

Tefferi, A. (2016) 'Myeloproliferative neoplasms: A decade of discoveries and treatment advances', *American Journal of Hematology*, 91(1), pp. 50–58. doi:

10.1002/ajh.24221.

Tefferi, A. *et al.* (2016) 'Targeted deep sequencing in primary myelofibrosis', *Blood advances*. American Society of Hematology, 1(2), pp. 105–111. doi: 10.1182/bloodadvances.2016000208.

Tefferi, Ayalew, Mudireddy, M., *et al.* (2018) 'Blast phase myeloproliferative neoplasm: Mayo-AGIMM study of 410 patients from two separate cohorts.', *Leukemia*, 32(5), pp. 1200–1210. doi: 10.1038/s41375-018-0019-y.

Tefferi, Ayalew, Lavu, S., *et al.* (2018) 'JAK2 exon 12 mutated polycythemia vera: Mayo-Careggi MPN Alliance study of 33 consecutive cases and comparison with JAK2V617F mutated disease.', *American journal of hematology*. United States, pp. E93–E96. doi: 10.1002/ajh.25017.

Tefferi, Ayalew, Guglielmelli, P., Lasho, T. L., *et al.* (2018) 'MIPSS70+ Version 2.0: Mutation and Karyotype-Enhanced International Prognostic Scoring System for Primary Myelofibrosis.', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. United States, pp. 1769–1770. doi: 10.1200/JCO.2018.78.9867.

Tefferi, Ayalew, Guglielmelli, P., Pardanani, A., *et al.* (2018) 'Myelofibrosis Treatment Algorithm 2018', *Blood Cancer Journal*, 8(8). doi: 10.1038/s41408-018-0109-0.

Tefferi, A. (2018) 'Primary myelofibrosis: 2019 update on diagnosis, risk-stratification and management', *American Journal of Hematology*, 93(12), pp. 1551–1560. doi: 10.1002/ajh.25230.

Tefferi, A *et al.* (2018) 'Prognostic significance of ASXL1 mutation types and allele burden in myelofibrosis.', *Leukemia*. England, pp. 837–839. doi: 10.1038/leu.2017.318.

Tefferi, Ayalew, Nicolosi, M., Mudireddy, M., *et al.* (2018) 'Revised cytogenetic risk stratification in primary myelofibrosis: analysis based on 1002 informative patients.', *Leukemia*, 32(5), pp. 1189–1199. doi: 10.1038/s41375-018-0018-z.

Tefferi, Ayalew, Finke, C. M., Lasho, T. L., *et al.* (2018) 'U2AF1 mutation types in primary myelofibrosis: phenotypic and prognostic distinctions', *Leukemia*.

2018/02/27. Nature Publishing Group UK, 32(10), pp. 2274–2278. doi: 10.1038/s41375-018-0078-0.

Tefferi, A. (2021) 'Primary myelofibrosis: 2021 update on diagnosis, risk-stratification and management', *American Journal of Hematology*, 96(1), pp. 145–162. doi: 10.1002/ajh.26050.

Tefferi, A. and Barbui, T. (2015) 'Polycythemia vera and essential thrombocythemia: 2015 update on diagnosis, risk-stratification and management.', *American journal of hematology*. United States, 90(2), pp. 162–173. doi: 10.1002/ajh.23895.

Tefferi, A. and Barbui, T. (2020) 'Polycythemia vera and essential thrombocythemia: 2021 update on diagnosis, risk-stratification and management.', *American journal of hematology*. United States, 95(12), pp. 1599–1613. doi: 10.1002/ajh.26008.

Tefferi, A. and Pardanani, A. (2015) 'Myeloproliferative neoplasms: A contemporary review', *JAMA Oncology*, 1(1), pp. 97–105. doi: 10.1001/jamaoncol.2015.89.

Thiele, J. *et al.* (1991) 'Ultrastructure of bone marrow tissue in so-called primary (idiopathic) myelofibrosis-osteomyelosclerosis (agnogenic myeloid metaplasia). I. Abnormalities of megakaryopoiesis and thrombocytes.', *Journal of submicroscopic cytology and pathology*. Italy, 23(1), pp. 93–107.

Thiele, J. *et al.* (2009) 'Bone marrow fibrosis and diagnosis of essential thrombocythemia.', *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. United States, pp. e220-1; author reply e222-3. doi: 10.1200/JCO.2009.24.3485.

Tibshirani, R. *et al.* (2002) 'Diagnosis of multiple cancer types by shrunken centroids of gene expression.', *Proceedings of the National Academy of Sciences of the United States of America*, 99(10), pp. 6567–6572. doi: 10.1073/pnas.082099299.

Tokita, K. *et al.* (2007) 'Chronic idiopathic myelofibrosis expressing a novel type of TEL-PDGFRB chimaera responded to imatinib mesylate therapy', *Leukemia*, 21(1), pp. 190–192. doi: 10.1038/sj.leu.2404397.

Tonkin, J. *et al.* (2012) 'Myeloproliferative neoplasms: diagnosis, management and treatment.', *Nursing standard (Royal College of Nursing (Great Britain) : 1987)*. England, 26(51), pp. 44–51. doi: 10.7748/ns2012.08.26.51.44.c9243.

Ugo, V. *et al.* (2004) 'Multiple signaling pathways are involved in erythropoietin-independent differentiation of erythroid progenitors in polycythemia vera.', *Experimental hematology*. Netherlands, 32(2), pp. 179–187. doi: 10.1016/j.exphem.2003.11.003.

Van't Veer, L. J. *et al.* (2002) 'Gene expression profiling predicts clinical outcome of breast cancer', *Nature*, 415(6871), pp. 530–536. doi: 10.1038/415530a.

Vannucchi, A. M. *et al.* (2002) 'Development of myelofibrosis in mice genetically impaired for GATA-1 expression (GATA-1(low) mice).', *Blood*. United States, 100(4), pp. 1123–1132. doi: 10.1182/blood-2002-06-1913.

Vannucchi, A. M. *et al.* (2007) 'Clinical profile of homozygous JAK2 617V>F mutation in patients with polycythemia vera or essential thrombocythemia.', *Blood*. United States, 110(3), pp. 840–846. doi: 10.1182/blood-2006-12-064287.

Vannucchi, A. M. *et al.* (2013) 'Mutations and prognosis in primary myelofibrosis', *Leukemia*. Nature Publishing Group, 27(9), pp. 1861–1869. doi: 10.1038/leu.2013.119.

Vannucchi, A. M., Guglielmelli, P. and Tefferi, A. (2009) 'Advances in understanding and management of myeloproliferative neoplasms.', *CA: a cancer journal for clinicians*. United States, 59(3), pp. 171–191. doi: 10.3322/caac.20009.

Vaquez, H. (1892) 'Sur une forme spéciale de cyanose s' accompagnant d'hyperglobulie excessive et persistante', *CR Soc Biol (Paris)*, 44, pp. 384–388.

Vardiman, J. W. *et al.* (2009) 'The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes.', *Blood*. United States, 114(5), pp. 937–951. doi: 10.1182/blood-2009-03-209262.

Varricchio, L., Mancini, A. and Migliaccio, A. R. (2009) 'Pathological interactions between hematopoietic stem cells and their niche revealed by mouse models of

primary myelofibrosis', *Expert review of hematology*, 2(3), pp. 315–334. doi: 10.1586/ehm.09.17.

Vener, C. *et al.* (2010) 'Oxidative stress is increased in primary and post-polycythemia vera myelofibrosis', *Experimental Hematology*. ISEH - Society for Hematology and Stem Cells, 38(11), pp. 1058–1065. doi: 10.1016/j.exphem.2010.07.005.

Verstovsek, S. *et al.* (2012) 'A double-blind, placebo-controlled trial of ruxolitinib for myelofibrosis.', *The New England journal of medicine*, 366(9), pp. 799–807. doi: 10.1056/NEJMoa1110557.

Vieira, A. F. and Schmitt, F. (2018) 'An update on breast cancer multigene prognostic tests-emergent clinical biomarkers', *Frontiers in Medicine*, 5(SEP), pp. 1–12. doi: 10.3389/fmed.2018.00248.

Wallden, B. *et al.* (2015) 'Development and verification of the PAM50-based Prosigna breast cancer gene signature assay.', *BMC medical genomics*, 8, p. 54. doi: 10.1186/s12920-015-0129-6.

Wang, J. C. *et al.* (2002) 'Plasma matrix metalloproteinase and tissue inhibitor of metalloproteinase in patients with agnogenic myeloid metaplasia or idiopathic primary myelofibrosis', *British Journal of Haematology*. John Wiley & Sons, Ltd, 119(3), pp. 709–712. doi: https://doi.org/10.1046/j.1365-2141.2002.03874.x.

Weston, A. D. and Hood, L. (2004) 'Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine.', *Journal of proteome research*. United States, 3(2), pp. 179–196. doi: 10.1021/pr0499693.

Wolanskyj, A. P. *et al.* (2005) 'JAK2 mutation in essential thrombocythaemia: clinical associations and long-term prognostic relevance.', *British journal of haematology*. England, 131(2), pp. 208–213. doi: 10.1111/j.1365-2141.2005.05764.x.

Wolanskyj, A. P. *et al.* (2006) 'Essential thrombocythemia beyond the first decade: life expectancy, long-term complication rates, and prognostic factors.', *Mayo Clinic proceedings*. England, 81(2), pp. 159–166. doi: 10.4065/81.2.159.

Yang, Y. *et al.* (2016) 'Loss of Ezh2 cooperates with Jak2V617F in the development of myelofibrosis in a mouse model of myeloproliferative neoplasm', *Blood*. 2016/04/14. American Society of Hematology, 127(26), pp. 3410–3423. doi: 10.1182/blood-2015-11-679431.

Yersal, O. and Barutca, S. (2014) 'Biological subtypes of breast cancer: Prognostic and therapeutic implications', *World Journal of Clinical Oncology*, 5(3), pp. 412–424. doi: 10.5306/wjco.v5.i3.412.

Yin, T. and Li, L. (2006) 'The stem cell niches in bone.', *The Journal of clinical investigation*, 116(5), pp. 1195–1201. doi: 10.1172/JCI28568.