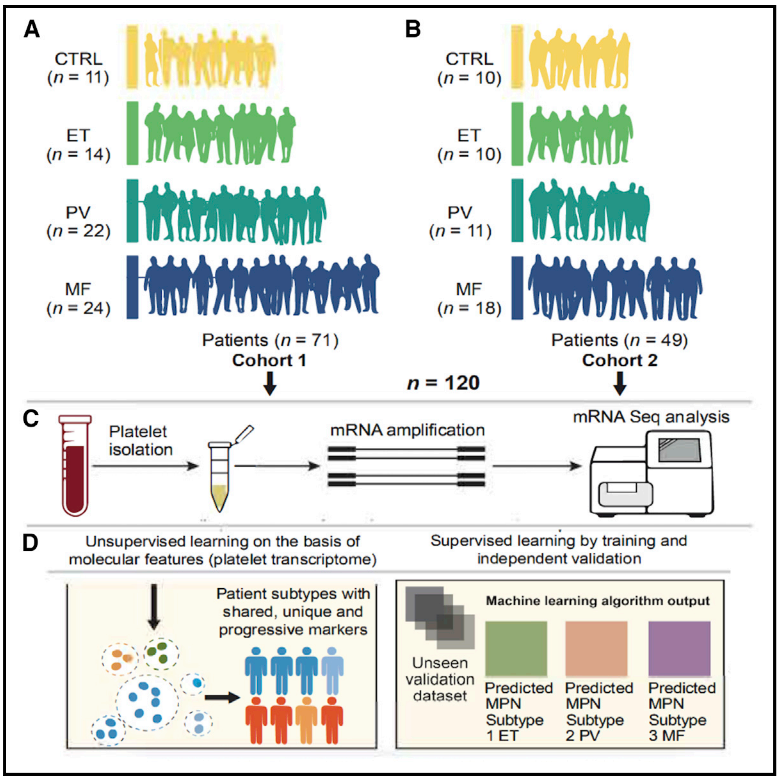
Cell Reports
Medicine

Article

Platelet transcriptome identifies progressive markers and potential therapeutic targets in chronic myeloproliferative neoplasms

Graphical abstract



Highlights

- Platelet transcriptome yields progressive markers across MPN subtypes
- Lasso-penalized multinomial regression model predicts advanced MPNs
- Impaired protein homeostasis and an integrated stress response feature in MPN progression

Authors

Zhu Shen, Wenfei Du, Cecelia Perkins, ..., Jason Gotlib, James Zehnder, Anandi Krishnan

Correspondence

anandi.krishnan@stanford.edu

In brief

Shen et al. leverage two independent MPN patient cohorts to identify progressive platelet transcriptomic markers, which also enable an externally validated prediction for advanced MPNs. The platelet RNA-seq data identify impaired protein homeostasis in MPN progression and offer potential targets of therapy.

Shen et al., 2021, Cell Reports Medicine 2, 100425
October 19, 2021
https://doi.org/10.1016/j.xcrm.2021.100425

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Platelet transcriptome identifies progressive markers and potential therapeutic targets in chronic myeloproliferative neoplasms

Zhu Shen,[1] Wenfei Du,[1] Cecelia Perkins,[2] Lenn Fechter,[2] Vanita Natu,[3] Holden Maecker,[4] Jesse Rowley,[5] Jason Gotlib,[2,6,8] James Zehnder,[2,6,7,8] and Anandi Krishnan[2,7,9,10,*]

[1]Department of Statistics, Stanford University, Stanford, CA, USA
[2]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA
[3]Stanford Functional Genomics Facility, Stanford University School of Medicine, Stanford, CA, USA
[4]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA
[5]Department of Internal Medicine, University of Utah, Salt Lake City, UT, USA
[6]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA
[7]Department of Pathology, Stanford University, Stanford, CA, USA
[8]These authors contributed equally
[9]Lead contact
[10]Twitter: @anandi_krishnan
*Correspondence: anandi.krishnan@stanford.edu
https://doi.org/10.1016/j.xcrm.2021.100425

SUMMARY

Predicting disease progression remains a particularly challenging endeavor in chronic degenerative disorders and cancer, thus limiting early detection, risk stratification, and preventive interventions. Here, profiling the three chronic subtypes of myeloproliferative neoplasms (MPNs), we identify the blood platelet transcriptome as a proxy strategy for highly sensitive progression biomarkers that also enables prediction of advanced disease via machine-learning algorithms. The MPN platelet transcriptome reveals an incremental molecular reprogramming that is independent of patient driver mutation status or therapy. Subtype-specific markers offer mechanistic and therapeutic insights, and highlight impaired proteostasis and a persistent integrated stress response. Using a LASSO model with validation in two independent cohorts, we identify the advanced subtype MF at high accuracy and offer a robust progression signature toward clinical translation. Our platelet transcriptome snapshot of chronic MPNs demonstrates a proof-of-principle for disease risk stratification and progression beyond genetic data alone, with potential utility in other progressive disorders.

INTRODUCTION

The classic Philadelphia chromosome-negative (Ph^-) MPNs,[5–8] are clonal disorders of the bone marrow that comprise three clinical phenotypes– essential thrombocythemia (ET), polycythemia vera (PV), and myelofibrosis (MF). These myeloid neoplasms are defined by a combination of morphologic, clinical, laboratory, and cytogenetic/molecular genetic features. The existing genetic landscape[9–11] of MPNs primarily involves mutations in three driver genes that lead to constitutive JAK-STAT signaling (*JAK2, CALR, MPL*). Several additional non-driver mutations (see references[9–15] for details) as well as cytogenetic[7] and epigenetic[16,17] abnormalities also contribute to disease initiation and progression, and impact both overall survival and potential for progression to acute myeloid leukemia (AML)[18]. Depending on the MPN and stage of disease, these patients may exhibit debilitating constitutional symptoms such as fatigue, pruritus, night sweats, and weight loss; thrombohemorrhagic diathesis and extramedullary hematopoiesis; and an increased propensity for transformation to AML. Although an increase in one or more blood cell lineages contributes to these morbid sequelae, the qualitative abnormalities of myeloid cells that increase vascular risk or disease progression are not well understood. Taken together, a limited understanding exists regarding how genotypic variability contributes to diverse phenotypic presentations and disease natural histories. We were motivated by the current clinical need[7,9,10] and the potential for deeper integration of clinical features and genetics with gene expression profiling to improve stratification and management of chronic blood disorders, such as MPNs.

Blood platelets play critical roles in multiple processes and diseases[19–21], from their traditional function in hemostasis and wound healing to inflammation, immunity, cancer metastasis, and angiogenesis. Platelets originate from bone marrow precursor megakaryocytes as anucleate fragments with a distinctive discoid shape. Platelets contain a complex transcriptional landscape of messenger RNAs (mRNAs), unspliced pre-mRNAs, rRNAs, tRNAs and microRNAs[21–24]. Most platelet RNA

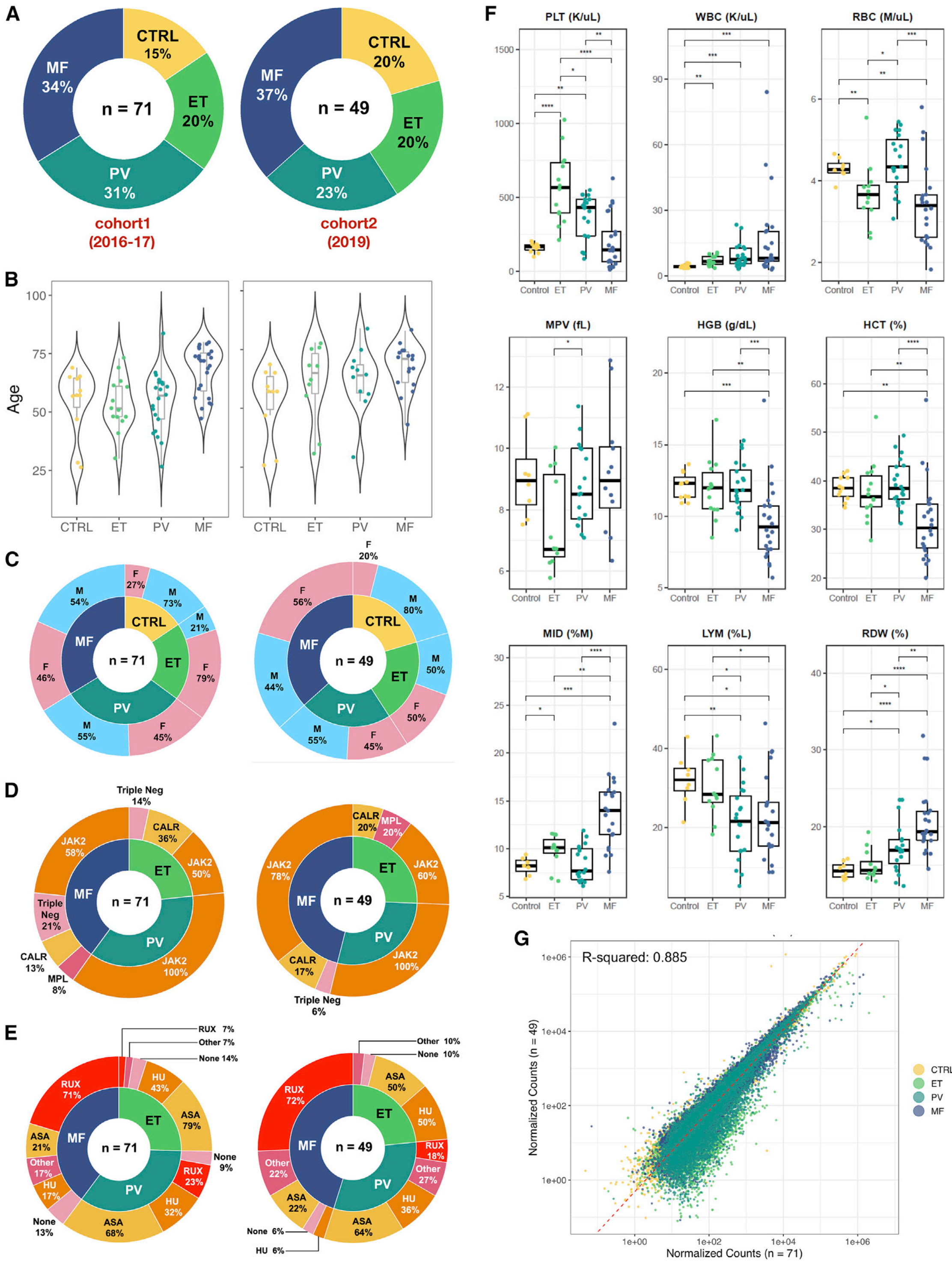


Figure 1. Two independent MPN clinical cohorts and closely replicated patient variables

(A) Similarity in distribution of MPN subtypes between two cohorts of MPN patients (Stanford single-center; approximately 2 years apart: cohort 1: 2016-17, n = 71; and cohort 2: 2019, n = 49); the majority subtype is MF in both cohorts (34% of n = 71 and 37% of n = 49).

(B) Comparable distribution of age across MPN subtypes in the two cohorts. Violin plots of patient age from each MPN subtype reflect clinical expectation, with an almost identical match between the two cohorts. Slightly higher median age noted in the second cohort for ET and PV patients alone.

(C) Comparable and balanced distribution of gender across MPN subtypes in the two cohorts. Larger percentage of male healthy controls in both cohorts and smaller percentage of males in ET cohort 1 noted.

(legend continued on next page)

CellPress
OPEN ACCESS

expression results from the transcription of nuclear DNA in the megakaryocyte[22,25], and thus reflects the status of the megakaryocyte at the time of platelet release into the circulation, as well as subsequent splicing events, selective packaging, and intercellular RNA transfer. There is emerging evidence[26–30] that the molecular signature of platelets may be changed in disease conditions where these processes are altered, including via intercellular transfer[27,29] of cytosolic RNA. In the context of MPNs, the platelet transcriptome therefore not only represents a critical biomarker of megakaryocytic activity[31–35], but also provides a snapshot of the underlying hemostatic, thrombotic, and inflammatory derangements associated with these hematologic neoplasms and the potential impact of treatment[34].

To date, no one composite study[4,15,36,37–40] has evaluated the disease-relevant platelet transcriptome in a sizeable clinical cohort of all three subtypes of Ph^- MPNs. Here, we extend our prior preliminary work[36] toward a comprehensive analysis of disease-relevant[41] platelet RNA-sequencing (RNA-seq) in two temporally independent and mutually validating cohorts of all three MPN subtypes, ET, PV, and MF (primary or post ET/PV secondary). We demonstrate marked differences in platelet gene expression across the MPN spectrum, which also permits robust validated (temporal and geographical) prediction of MF. In addition to identifying novel gene expression signatures impacted by the JAK1/JAK2 inhibitor ruxolitinib (RUX), platelet profiling reveals MPN-altered pathways that may be targets for future drug development.

RESULTS

Two independent MPN clinical cohorts and closely replicated platelet transcriptome

We prepared highly purified leukocyte-depleted[42,43] platelets from peripheral blood samples of two cohorts (approximately 2 years apart; Stanford single-center) of patients with a diagnosis of MPN (including provisional) at the time of sample acquisition, and included healthy controls in each cohort as reference (cohort 1, n = 71, and cohort 2, n = 49; Figure 1, Tables S1A and S1B). Only 2 patients (2%) received a change in diagnosis from MPN (MF) to a non-MPN phenotype; and were therefore excluded from all downstream analyses (Figure 2A principal component analysis plot panel-3 identifies these 2 outliers). Our two-cohort study was specifically designed (before knowledge of other subsequent studies) for the explicit purpose of validation, not only of intercohort RNA-seq results[44] but also to evaluate temporal validation[45] of our prediction model (see STAR Methods, Figures 5C, 5E, 5F). Figure S1 demonstrates our established[42,46] high-quality and highly efficient experimental framework toward a rigorous platelet RNA-seq approach. Clinical features of the MPN patients are shown in Figure 1 and listed in Tables S1A and S1B. The distribution of key variables was closely matched between the two cohorts by MPN subtype (Figure 1A), age (Figure 1B), gender (Figure 1C), *JAK2*/*CALR* mutation status (Figure 1D), and treatment (Figure 1E). Any interpatient variability in patient age, gender, and treatment were adjusted as confounding factors in all downstream gene expression analyses (see STAR Methods). Clinical laboratory measures (Figure 1F) at the time of sampling reflected the phenotype of the MPN subtypes (please see figure legend for detailed statistical comparisons). The two cohorts of platelet transcriptome data (Figure 1G, normalized transcript counts) adjusted for patient age, gender, and all treatment as confounding factors were also highly correlated (R^2 = 0.89), thus demonstrating high intercohort validation of gene expression that then enabled us to combine our two RNA-seq datasets into a final integrated dataset of enhanced statistical power for downstream analyses, especially prediction modeling (Figure 5). Together, this platelet transcriptome compendium comprises 118 human peripheral blood samples from healthy controls (n = 21) and World Health Organization–defined MPN patients (24 ET, 33 PV and 40 MF) that include 7 untreated, and 92 either on cytoreductives/biologics (e.g., ruxolitinib, hydroxyurea, interferon-alpha), anti-thrombotic agents (e.g., aspirin, warfarin), or a combination of the two; and captures the real-life diversity among MPN patients. Our cross-sectional design here capturing patients from all three MPN subtypes also serves as a practical alternative to the longitudinal approach; though ideal, it is likely difficult to implement in these chronic disorders which often occur over decades.

MPN platelet transcriptome distinguishes disease phenotype and reveals phenotype- and JAK-inhibitor specific signatures

Given the phenotypic overlap, yet also differences in disease behavior and prognosis between ET, PV, and MF, we

(D) Matched distribution of primarily *JAK2* and *CALR* mutational status across MPN subtypes in the two cohorts. JAK2 is the most common mutation across all three subtypes; 100% of PV and over 50% of ET and MF patients have JAK2 mutation in both cohorts. Mismatch between cohorts on the MPL/triple negative patients is noted as a natural consequence of the rarer prevalence of these mutations; and therefore, not the primary focus of this work.

(E) Diversity of MPN patient therapies across the two cohorts reflecting current clinical practice. The majority being aspirin (ASA) in ET/PV patients and the *JAK*-inhibitor, ruxolitinib, in MF. Note that a given patient may be on more than one treatment and therefore, the total treatment percentage in this graphic may not equal 100. To control for any inter-patient variability, all treatment, in addition to patient age and gender are adjusted as confounding factors in downstream gene expression analyses.

(F) Representative clinical laboratory variables, as boxplots, measured at the same date and time as platelet sampling. Compared to controls, MPN patients show larger variance (inter-quartile range [IQR]), and reflect current clinical knowledge. Groups differ primarily only with respect to blood cell counts (platelet/RBC/WBC); and show the greatest differential in MF. Note higher platelet count in ET with a concomitant lower mean platelet volume, higher red blood cell count in PV and lower red blood cell count in MF with concomitantly lower hemoglobin, hematocrit, and higher red cell width. Wilcoxson signed rank tests marked by asterisks indicate a statistically significant difference between any two groups (*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; ****$p \leq 0.0001$).

(G) High correlation (R^2 = 0.89) of platelet gene expression as assessed by normalized counts of matched transcripts in each cohort between each of controls, essential thrombocythemia (ET), polycythemia vera (PV), and myelofibrosis (MF). The two-cohort collective sample size by subtype constitutes each of ET (n = 24), PV (n = 33), primary or post ET/PV secondary MF (n = 42), healthy donors (n = 21), and totals (n = 120), affording increased statistical power for subsequent analyses.

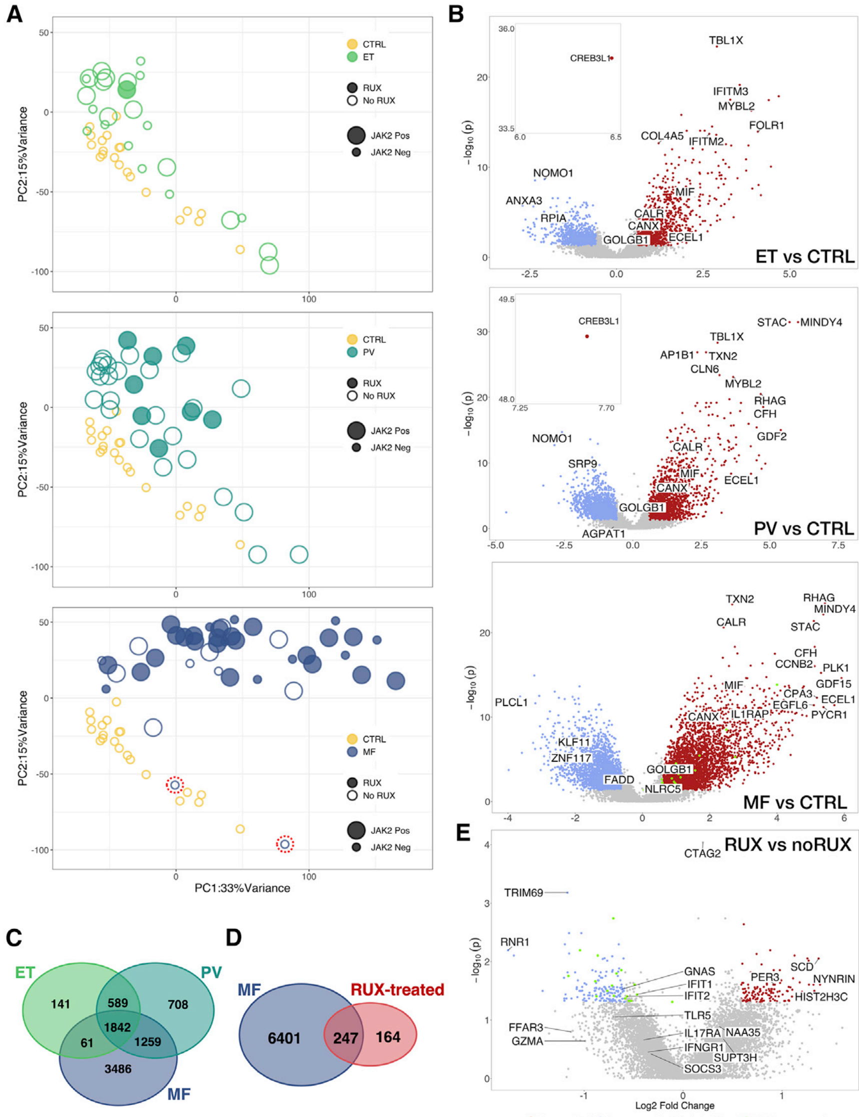


Figure 2. MPN platelet transcriptome distinguishes disease phenotype and reveals phenotype- and JAK-inhibitor specific signatures

(A) Unsupervised principal component analysis (PCA) of normalized platelet gene expression counts adjusted for age, gender, treatment, and experimental batch. Three panels of PC1 and PC2 colored by MPN subtype; and each contrasted with controls (n = 21; yellow): ET (n = 24; top, light green), PV (n = 33; middle, dark green), and MF (n = 42; bottom, dark blue). Circles filled or open mark presence or absence of ruxolitinib treatment and size of circles, smaller or larger, indicate presence or absence of *JAK2* mutation. The first two principal components account for 48% of total variance in the data.

(B) Volcano plots (three panels as A above of ET, PV, and MF) of differential gene expression showing statistical significance (negative log10 of p values) versus log2 fold change of each gene. Significant upregulated and downregulated genes are those with p values (FDR) smaller or equal to 0.05 and absolute value of fold changes larger or equal to 1.5.

(C and D) Venn Diagram comparisons of MPN differential gene expression lists. In (C), each of ET, PV and MF is contrasted with controls; identifying gene sets that are shared (n = 1504, FDR < 0.05) as well as unique to each subtype. In (D), differential in the RUX-treated cohort is contrasted with MF versus controls. Differential in gene expression in RUX-treated cohort is a fraction of the total differential noted in the MF transcriptome.

(E) Volcano plot of differential gene expression between MPN patients treated with ruxolitinib and not. A small subset of overlapping differential genes that are upregulated in MF (B) (bottom panel) and suppressed in the RUX-treated cohort (E) are colored in green.

hypothesized that there may be MPN subtype-specific differences in gene expression that are independent of *JAK2/CALR/MPL* mutation status. In addition, we hypothesized that MPN platelets are likely to be enriched for subtype-specific biomarkers that may otherwise be missed in blood/plasma/serum sources[37–39]. Therefore, we compared platelet transcriptomic expression in each of the three MPN subtypes with that of controls and discovered a shared gene set that is also progressively differentiated across the MPN spectrum (ET/PV to MF as shown in Figures 2A–2C). First, unsupervised principal component analysis (PCA) of MPN patients and controls data (Figure 2A) confirmed that the collective variability from the first two principal components (accounting for 48% of total variance), after adjusting for age, gender, treatment, and experimental batch, was MPN disease status, with increasing differentiation by subtype. Next, differential gene expression analysis (DGEA), as shown in a volcano plot (Figures 2B and 2C), efficiently distinguished each of the MPN subtypes and resulted in highly significant expression signatures (adjusted p value/FDR < 0.05) with 2634 genes differentially regulated in ET (1364 up and 1269 down), 4398 in PV (2098 up and 2300 down), and 6648 in MF (3965 up and 2683 down). A subset of 100+ long non-coding RNAs and pseudogenes also constituted the significant (FDR < 0.05) differential expression across MPNs (Table S2A-C).

Specifically, DGEA also uncovered shared and unique genes between all three MPN phenotypes, thus offering a potential core set of genes involved in MPN pathogenesis (Figure 2C, with associated heatmaps in Figure 3 and pathways in Figure 4). The shared gene set at FDR < 0.01 constituted 654 upregulated genes, with a predominance of molecular pathways involving myeloid cell activation in immune response and membrane protein proteolysis, and 360 downregulated genes, reflecting negative regulation of hematopoiesis and negative regulation of transmembrane receptor protein serine/threonine kinase signaling as a consistent pathogenetic mechanism. The upregulated genes belonged to the endoplasmic reticulum/ER-Golgi intermediate

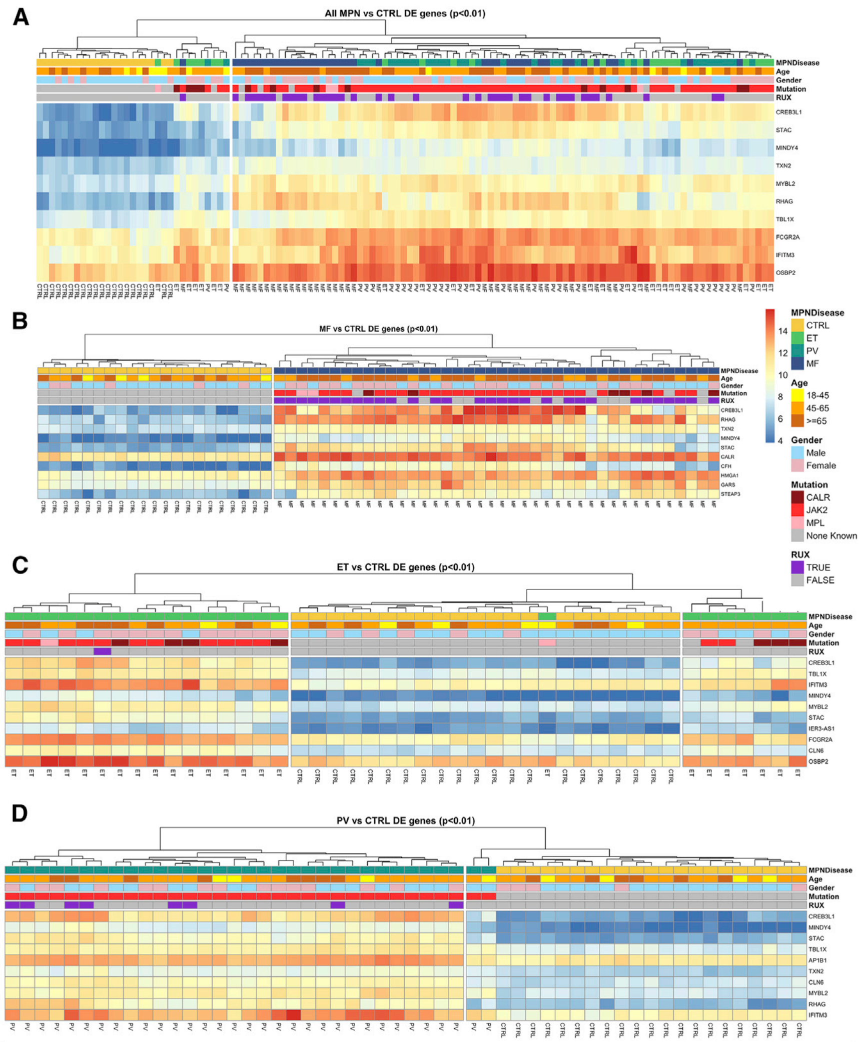


Figure 3. Graded differential expression by MPN phenotype and driver mutation status

(A–D) Hierarchically clustered heatmaps of the top 10 differentially expressed genes (DEGs) from controls (FDR < 0.01) of all MPN (A) and MF, ET, and PV each separately (B–D). Colored annotation is provided to indicate MPN subtype, age, gender, mutation, and ruxolitinib treatment. Rows indicate gradation in gene expression on a blue (low) to red (high) scale. Columns indicate sample type from controls (CTRL) to ET, PV, and MF.

compartment; they also included a particularly high expression of the cAMP-response element binding transcription factor, *CREB3L1*[47] implicated in cell differentiation and inhibition of cell proliferation, with concomitant high expression of ER chaperones[48] calreticulin (*CALR*), calnexin (*CANX*)*;* transport *factors:* golgin *(GOLGB1)* and folate receptor *FOLR1.* Platelet alpha granule proteins (*F5, VWF, MMP14*), several collagens (*COL10A1, COL18A1, COL6A3*), immune/inflammatory (*IFITM2/3/10, FCGR2A, TMEM179B*), *Cathepsins (C/Z/D, MIF, PTGES2*) and proliferation mediator genes (*CDK1, CCNG1, BMP9/GDF2, LAPTM4B, PSENEN*) also constituted the MPN shared set. Downregulated genes, on the other hand, were predominantly within transcription factor complexes, and included opposing expression of *CREB1* (vis-à-vis *CREB3L1* above), calcium-calmodulin protein kinases, *CAMK4*, *SMAD1* and β-catenin *CTNNB1* together pointing to dysregulated calcium (Ca^{2+}) homeostasis in the ER lumen.

Differential markers in each of ET, PV and MF also highlight candidate genes as potential mediators of the pro-thrombotic and pro-fibrotic phenotypes in MPNs. In ET and PV, a strong thromboinflammatory profile[6,40] is revealed by the upregulation of several interferon inducible transmembrane genes (*IFITM2, IFITM3, IFITM10, IFIT3, IFI6, IFI27L1, IFI27L2*); interleukin receptor accessory kinases/proteins (*IRAK1, IL15, IL1RAP, IL17RC)*; and several solute carrier family genes (*SLC16A1, SLC25A1, SLC26A8, SLC2A9*) as glucose and other metabolic transport proteins; and coagulation factor V (*F5)*. In MF, fibrosis-specific markers were identified by an additional focused comparison of MF patients versus ET and PV (Figure S2), showing increased expression of several pro-fibrotic growth factors (*FGFR1, FGFR3, FGFRL1*); matrix metalloproteinases (*MMP8, MMP14*); vascular endothelial growth factor A (*VEGFA*); insulin growth factor binding protein (*IGFBP7*); and cell cycle regulators (*CCND1, CCNA2, CCNB2, CCNF*). GSEA of this MF-focused comparison again highlighted potential underlying molecular dysregulation in MPNs that likely contribute to the fibrotic phenotype (e.g., unfolded protein response, mTORc1 signaling, MYC/E2F targets, oxidative phosphorylation; Figure 4).

Having defined differential gene expression signatures by MPN subtype, we then explored how platelet gene expression profiles differed in patients by treatment, focusing for this work on the JAK1/JAK2 inhibitor, ruxolitinib (RUX, Figure 2E). DGEA on the platelet transcriptome in RUX-treated patients identified over 400 core significant genes changed in response to treatment (Figure 2E; Tables S4 and S5). At-least two-fold reduction in expression was noted in genes associated with interferon-stimulation (*IFI6, TRIM69, LY6G5C*), platelet-mediated apoptosis, *FASLG*, G-protein coupled receptor, *GPR88,* calcium-calmodulin protein kinase, *CAMK2A* and fibroblast growth factor binding protein, FGFBP2, followed by over 150 genes with at least one-fold reduction in expression in the RUX-treated cohort (Table S5). These included genes downregulated within classical pathways of adaptive immune response (*TNFSF13B, B2M, HMGB1*); response to oxidative stress (*COX1, COX2, COX15, TP53*); and myeloid activation (*TNIP2, FTH1, TOLLIP, RAB6A*).

A

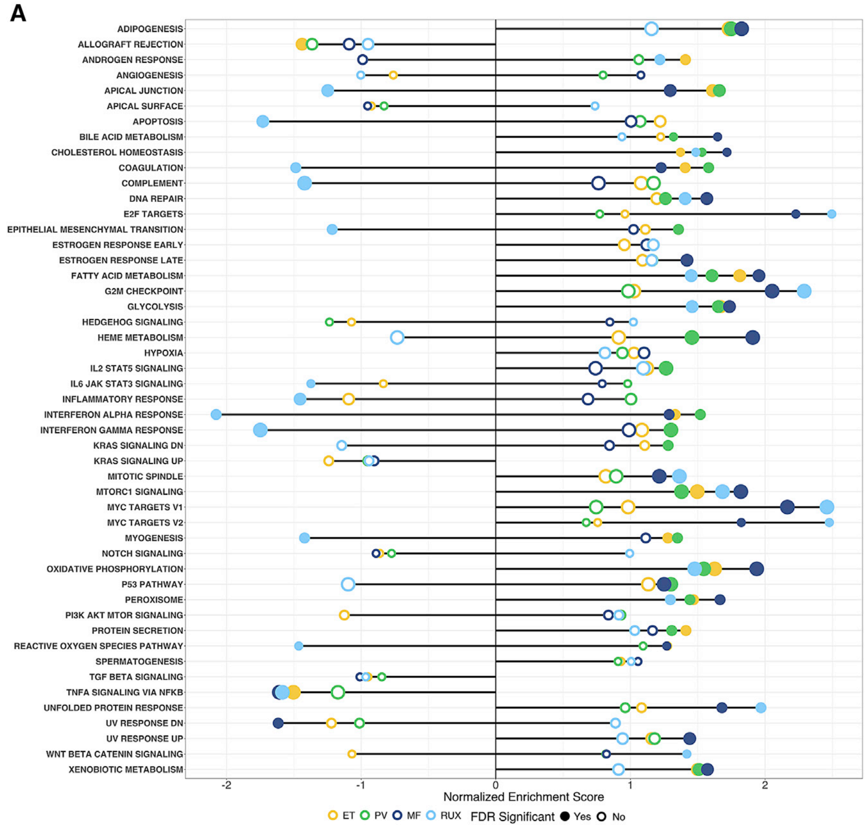


Figure 4. Altered immune, metabolic, and proteostatic pathways underlie each MPN phenotype

Pathway-enrichment analysis of genes with MPN subtype-specific expression (color indicated; light green ET, dark green PV, and dark blue MF) overlayed with ruxolitinib-specific expression (light blue). Each point represents a pathway; the x axis gives the normalized enrichment score, which reflects the degree to which each pathway is over- or under-represented at the top or bottom of the ranked list of differentially expressed genes, normalized to account for differences in gene set size and in correlations between gene sets and the expression dataset. The y axis lists the detail-level node of the most enriched pathways; solid lines mark GSEA-recommended[49] Bonferroni-corrected statistical significance criterion of FDR < 0.25 for exploratory analyses. Dotted lines mark FDR > 0.25 and therefore, not sufficiently significant, yet visualized alongside solid lines to retain overall context (upper-level parent nodes of the detail-level pathways are provided in Table S3A–S3C). Multiple immune and inflammatory pathways are consistently significantly enriched across ET, PV, and MF (and suppressed in the ruxolitinib-treated cohort). MF is differentiated from ET and PV through dysregulation of additional molecular processes for cellular development, proliferation, metabolism, and DNA damage.

In addition to confirming previous observations[7,50–52] on the anti-inflammatory and immunosuppressive effects of *JAK* inhibition by RUX (e.g., downregulation in our RUX-treated cohort of *IL1RAP, CXCR5, CPNE3, ILF3*), we identified new gene clusters responsive to RUX in the inhibition of type I interferon (e.g., *IFIT1, IFIT2, IFI6*); chromatin regulation (*HIST2H3A/C, HIST1H2BK, H2AFY, SMARCA4, SMARCC2*); and epigenetic methylation in mitochondrial genes (*ATP6, ATP8, ND1-6 and NDUFA5*). Recent literature probing the mechanisms of action of ruxolitinib in other disease settings, including SARS-CoV-2,[53,54] confirm our observations in MPNs.

A direct comparison restricted to only differentially expressed genes (FDR < 0.05) in *RUX-treated versus not* with *MF versus healthy controls* revealed less than 5% overlap (Figure 2D), reflecting potentially the extent of the impact of treatment by ruxolitinib relative to the substantive disease burden in MF. Focusing further on the directionality of the changes observed, we found just 18 genes that were increased in MF and suppressed in the RUX-treated cohort (Figure 2E, colored green), and 9 vice versa (Tables S3A S3B). Despite the small overlap, we capitalized on the converging genes to better define molecular and physiological pathways underlying the effect of RUX in MPNs. The 18 RUX-downregulated genes followed expected mapping to immunosuppression through interferon- and cytokine-mediated signaling pathways. The nine genes that were upregulated with RUX in MF mapped to previously undescribed effect of the drug in select G-protein-coupled receptor and chemokine activity (e.g. *CXCR5, GPR128/ADGRG7*), semaphorin signaling (*SEMA3C*) and circadian regulation (*PER3*). A subcohort analysis of RUX-treated and RUX-naive MF patients alone also identified downregulation of interferon-stimulated genes (e.g., *SLFN12L,* G-protein-coupled receptors (*S1PR5*), fibroblast growth factor binding protein (*FGFBP2*), and tyrosine kinase (*LCK*).

Graded differential expression by MPN phenotype and driver mutation status

Unsupervised hierarchical clustering was used to define the nature of MPN platelet transcriptome variability more precisely from controls, as well as between and within MPN subtypes. Figure 3 reveals a spectrum of expression in the MPN platelet transcriptomic profile using just the top 10 highly significant differentially expressed genes by disease status: (Figure 3A) all MPN versus controls, (Figure 3B) MF versus controls, and (Figures 3C and 3D) ET and PV versus controls. As shown in Figure 3A, all MPN patients clustered into two distinct groups: a larger group of 87 ET, PV, and MF patients clustered independently from the 21 controls; whereas a smaller group of 10 ET, PV, and MF patients formed a homogeneous cluster of their own closer to controls reflecting a varying gradation with respect to the top-10 gene expression by MPN subtype (patient variables annotated above the heatmaps offer additional context, particularly on mutation status and RUX therapy). In the larger cluster, while we observed a graded overlap in platelet RNA signatures

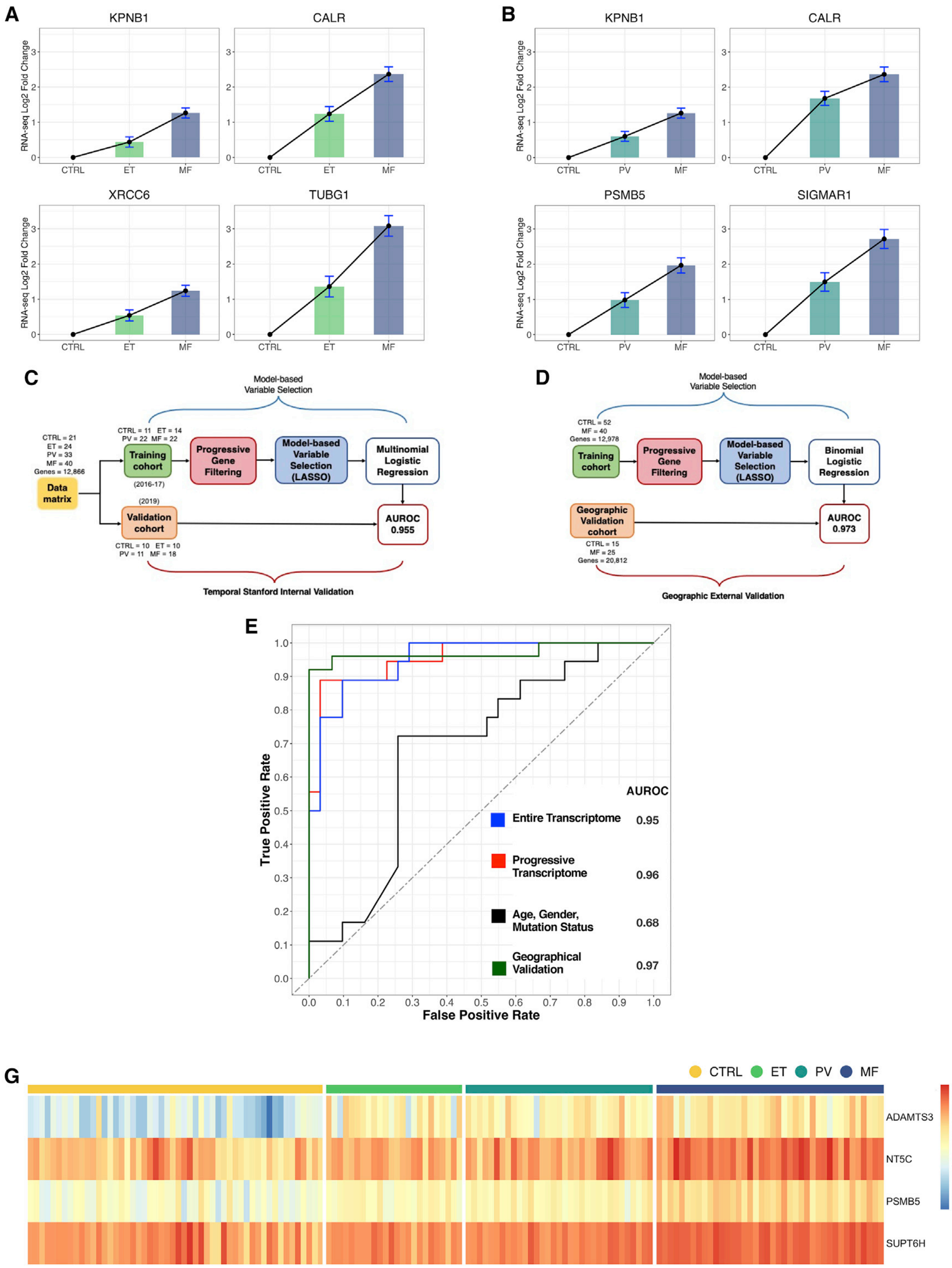


Figure 5. Prediction of MF based on unique and progressive MPN platelet transcriptome

(A and B) Top few genes (out of 3000+) demonstrating monotonic progressive gene expression (log 2 fold change in expression y axis, FDR < 0.01, *Mann-Kendall* test with Bonferroni correction) across *x*-axes (A), CTRL-to-ET-to-MF and (B, CTRL-to-PV-to-MF.

(C) Lasso-penalized multinomial logistic regression model under temporal validation i.e., trained on Stanford cohort 1 (n = 71, 2016–2017, Figure 1A) and validated on Stanford cohort 2 (n = 49, 2019, Figure 1A) as test set.

(D) Lasso-penalized multinomial logistic regression model under geographical validation using two independently published datasets for training (cohort 3, n = 31 healthy controls in addition to Stanford cohorts 1 and 2) and validation (cohort 4, n = 25 MF and n = 15 healthy controls).

(legend continued on next page)

between ET and PV, a more distinct expression pattern characterized the more advanced population of MF patients (PC1 correlates with MF disease risk by the Dynamic International Prognostic Scoring System[55] [DIPSS]; Figure S3). These data collectively highlight the importance of phenotype-modifying genes that are independent of *JAK2*/*CALR*/*MPL* mutation status.

Untangling other mechanisms beyond the few genes that are recurrently mutated is critical for defining subtype-specific risk and for identifying molecular pathways for targeted therapy. Accordingly, we sought to refine the molecular classification of MPN by associating platelet gene expression profiles with the corresponding subtype, and yielded a core set of 10 highly significant preferentially expressed genes for each: (1) MF (Figure 3B), defined by high mRNA expression of proteostasis-associated *CREB3L1* and *CALR,* and megakaryocyte-erythroid differentiation stage-associated *RHAG*;[56,57] (2) ET (Figure 3C) marked by comparatively high expression of interferon-related genes *IFITM2/3,* immune response *FCGR2A*, and proliferation-associated *STAT5* target *OSBP2*; and (3) PV (Figure 3D) marked by overlapping signatures with ET in inflammation-associated *IFITM3* and *TBL1X*, and the B-myb promoter *MYBL2*, with MF in the maturation-associated *RHAG* at variable expression, and with both ET and MF in the high expression of *CREB3L1* and cell survival-associated *MINDY4* and *STAC*.

Altered immune, metabolic, and proteostatic pathways underlie each MPN phenotype

Our analysis of MPN platelet RNA-seq enabled identification of altered MPN pathways that might be amenable to drug therapy. To understand the biological significance of transcriptional changes, we performed pathway-enrichment analysis and identified signaling pathways that are differentially activated between MPN subtypes (Figure 4). Gene set enrichment analysis (GSEA; see STAR Methods) of Hallmark gene sets found that MPN (stratified by subtypes, ET, PV, and MF) induces genes related to pathways with known immune modulatory functions (Figure 4, notably interferon alpha response in ET, PV, and MF, and IL2 STAT5 signaling, and interferon gamma response specifically enriched in PV). Moreover, among the most enriched gene sets, MPN pathology induces robust activation of oxidative phosphorylation (OXPHOS) and mTORC1 signaling pathways, with increasing enrichment and significance by MPN subtype (FDR < 0.0001 in MF). Pathways of reactive oxygen species (ROS) production paralleled activation of mTORC1 in MF. Other complementary metabolic pathways paralleled OXPHOS activation, with significant enrichment of bile and fatty acid metabolism, cholesterol homeostasis, and adipogenesis, most pronounced in MF and variably expressed in ET and PV. Coagulation- and complement-associated gene sets were expectedly enriched across ET, PV, and MF. What is particularly noteworthy is that in MF, cell cycle progression and proliferation pathways reveal significant enrichment (FDR < 0.001) around c-MYC and E2F targets, and G2M checkpoint pathways; and unfolded protein response emerges as a key factor, likely attributable to ER stress (see *CREB3L1* and *CALR* overexpression in Figures 2 and ,3). Representative GSEA profiles are shown in Figure 4 and the full list of enriched pathways and gene sets detailed in Tables S4A–S4C. The MPN pathways exhibiting significant transcriptional regulation by GSEA are consistent with our observations at the individual gene level for upregulated and downregulated transcripts, specifically those upregulated in MF. Taken together, these data demonstrate that in addition to immune factors such as type I/II interferons and dysregulation of interleukin-dependent inflammatory responses, which have been linked to MPNs, platelet transcriptional signatures of proliferation, metabolic and proteostasis signaling are a feature of MPN pathogenesis (Figure S4 captures the relative enrichment by subtype of MPN molecular pathway categories as a concept model of MPN progression).

Prediction of MF based on shared, unique, and progressive MPN platelet transcriptome

Current knowledge of MPN genetic, cytogenetic, or epigenetic abnormalities are limited[7,8] in their ability to enable prediction of disease progression or evolution of a given patient, from ET/PV phenotype to MF. In order to investigate the potential of platelet transcriptomic parameters to enable MF prediction, we constructed LASSO penalized[58] regression classifiers (machine-learning R package *glmnet)* to discriminate MPN subtypes from each other, and from healthy controls (Figures 5A–5E). We apply two rigorous[45] external validation conditions (Figures 5C and 5D): (1) training and independent temporal validation (Figure 5C) leveraging the Stanford two-cohort design, and (2) geographical validation (Figure 5D). We used two independently published platelet transcriptome datasets: first, from Rondina et al.[3] on an additional n = 31 healthy donors integrated with the Stanford datasets as training; and second, from Guo et al.[4] n = 25 MF and n = 15 healthy donors as geographical validation of the Lasso algorithm (Figure 5D). Our temporal validation constituted three types of models: (1) baseline (no transcriptome, but age, gender, and driver mutation status as reference variable information available not only for patients but also healthy donors); (2) full platelet transcriptome ; and (3) subset platelet transcriptome that exhibits progressive differentiation from controls to ET to MF or controls to PV to MF (> 3000 genes, top few of each comparison visualized to demonstrate progressive gene expression, (Figures 5A and 5B) (progressive subset is selected unbiased as part of the Lasso learning procedure). Comparison of the classification potential among the three

(E) Receiver operating curves (ROC) toward MF prediction under conditions of temporal (C) and geographical (D) validation. Temporal validation uses three models: (i) baseline, with no gene expression data but patient age, gender, and mutation status alone; (ii) entire MPN platelet transcriptome; and (iii) MPN progressive genes alone. Outperformance of the progressive transcriptome model (red curve, AUROC = 0.96) vis-a-vis the entire transcriptome dataset (blue curve, AUROC = 0.95) and lastly, the baseline model without gene expression (black curve, AUROC = 0.68). Geographical validation using the progressive transcriptome model also demonstrates independent, high MF predictive accuracy (green curve, AUROC = 0.97).

(F) Heatmap of top recurring Lasso-selected progressive genes for each of controls (left column, CTRL, yellow bar), ET (light green), PV (dark green), and MF (dark blue). Rows indicate gradation in gene expression on a blue (low) to red (high) scale. Columns indicate sample type (CTRL, ET, PV, and MF).

models demonstrated that the progressive platelet transcriptome model (Figure 5E red curve) substantially outperformed the baseline model (Figure 5E black curve) and was slightly better than the full transcriptome model (Figure 5E blue versus red curves) in the classification of ET, PV, and MF. Predicted probabilities for all three models are shown in Tables S5A–S5C. Lasso logistic regression classifiers to predict MF with each of the models under the first temporal two-cohort training-validation split of baseline, full, and progressive transcriptome each achieved area under the receiver-operating characteristic curve (AUROC) of 0.68, 0.95, and 0.96, respectively. Outperformance of the progressive transcriptome model was validated in our subsequent independent external geographic validation (Figure 5E green curve) at an AUROC of 0.97.

Recurrent top 4 genes from our progressive transcriptome Lasso are visualized as a heatmap in Figure 5F, clearly capturing the incremental gradient in gene expression between controls, ET, PV, and MF. These include *ADAMTS3* (ADAM metallopeptidase protease[59] with likely roles in *VEGF* signaling, tissue remodeling, and expression of related collagens through profibrotic *PAR1* and *TGF-β* signaling), *PSMB5* (implicated in proteasomal degradation/UPR activation[60] and identified previously in MPNs[61]), *NT5C* related to *PI3K* signaling[62,63] and *SUPT6H*/SPT6,[64,65] a tumor-initiating histone chaperone associated with chromatin remodeling. Lasso-selected candidate markers reflect the MPN clinical spectrum and capture the underlying pathology.

A key aspect of the layered Lasso modeling demonstrated here is our use of an approach that can be developed in future work to incorporate additional predictors to MF (e.g., *JAK2* V617F allele burden[66] or other genetic variants[67] beyond the driver mutations). Figure S5 demonstrates this through a second base model for prediction to MF from ET or PV alone (no transcriptome, but platelet count and hemoglobin levels as two additional clinical parameters that were available for all patients in addition to age, gender, and driver mutation status).

Essentially, our analyses with both the Lasso regression as well as the progressive transcriptome (Figure 5A) enable a small gene signature that could be implemented in a PCR-based prognostic assay and performed in any standard clinical laboratory. Figure S6 demonstrates a robust model (AUROC of 0.86) with just the top progressive genes (Figure 5A) or the Lasso-selected features (Figure 5G) incorporated with patient age, gender, driver mutation status, and the two clinical laboratory variables, platelet count and hemoglobin levels. Identification of genes with significant prognostic impact in this rich dataset in a therapy-agnostic manner may prove useful for stratifying patients at high risk for progression and may be translated toward a widely applicable clinical laboratory assay.

DISCUSSION

Here, we present a comprehensive catalog of the platelet transcriptome in chronic progressive MPN with immediate relevance to defining subtype-specific molecular differences and predicting the advanced phenotype, MF. Recent data[68] identify the timing of MPN driver mutation acquisition to be very early in life, even before birth, with life-long clonal expansion and evolution. These new findings highlight the importance of early progression biomarkers and the substantial opportunity for early detection and intervention strategies in these disorders.

Our analyses confirm and extend many important observations made previously either *in* vitro,[69–71] in other transcriptome or microarray analyses[15,38,40,4,72,73] including our own early work,[36] or using animal models.[31,74–76] By highlighting intersecting mechanisms in transcription across MPN, and by annotating MPN subtype-specific gene signatures, this dataset facilitates predictive machine-learning algorithms, that aid in MPN classification and potential prognostication.

The platelet transcriptome is significantly reprogrammed in the MPN setting, with a wealth of transcript associations that may be missed in using conventional tissue sources such as serum, plasma, whole blood, or bulk bone marrow. While previous bulk RNA-seq studies on MPNs by us and others analyzed far fewer samples,[36–38,4] select MPN subtypes,[40,4] or non-specific source tissue[37–39] that may be underpowered[41] for candidate genes, we here analyzed, by next-generation RNA sequencing, 120 purified platelet samples from healthy controls and all three subtypes; and identified clinically interpretable transcriptomic signatures for each of the three subtypes. Each subtype showed both overlapping and progressively divergent transcriptional pathways, suggesting both a shared signature across all MPN, and unique biological trajectories. Pathway-enrichment analyses confirmed the existence of a shared inflammatory milieu[37,77–82] among MPN. We also confirmed that the JAK1/JAK2 inhibitor ruxolitinib was associated with inhibition of inflammatory as well as interferon-mediated signaling pathways. Additional previously undescribed insights into the mechanisms of action of RUX in MF included genes implicated in protein maturation, chaperone-mediated protein complex assembly, and circadian rhythm. These and other gene signatures and pathways identified may help guide candidate drugs to be used alone or in combination with RUX for the treatment of MPNs. Whether MPN oncogenic driver mutations increase inflammation or mutations are acquired in response to inflammatory stimuli is unclear from this work and remains an active area of investigation.[79,80,83,84]

The 10 genes most significant (FDR < 0.001) of the commonly expressed genes across MPN indicated a gradation in platelet gene expression, with overlapping signatures in ET and PV (*e.g. IFITM2, MYBL2)* and a substantial difference with MF (*e.g. CREB3L1, CALR*) that was independent of driver mutation status or treatment. Hence, while over 1500 genes were commonly differentially expressed across MPN, their abundance and function could differ between subtypes. The nature of the separation of transcriptomic clusters between ET, PV, and MF suggest also that they represent diverse cell states along a continuous spectrum of MPN, in line with the clinical overlap of these neoplasms.

Another observation relates to the association of the differential genes with signaling pathways: As indicated above, all three MPN subtypes showed a positive enrichment in immune modulation pathways, independent of mutational status. Whether this response reflects a causal effect of inflammation on bone marrow biology remains to be elucidated. Indeed, the platelet transcriptomic signatures could also reflect intercell interactions of platelets with other immune cells, including as transient aggregates with neutrophils, granulocytes, and dendritic cells. Nevertheless, observations that MPN transcriptomic biomarkers correlated robustly with immune factors such as type I/II interferons and dysregulation of interleukin-dependent inflammatory responses across ET, PV, and MF suggest opportunities for use of these and other subtype-specific genes as biomarkers for prognosis as well as design of therapies and prediction of response.

Our data closely overlaps with recent MPN platelet studies: thrombo-inflammatory signatures in PV from Gangaraju and Prchal et al.[40] (*BCL2, CXCL1, MMP7, PGLYRP1, CKB, BSG, CFL1*, and more) or fibrosis-associated signatures in MF from Guo et al.[4] (*CCND1, H2AFX*, and *CEP55*). Most notably in our data on MF, high expression of ER stress and unfolded protein response (UPR) biomarkers (e.g.*, CREB3L1, CALR)* associated with impaired proteostasis signaling and emerged as a key feature of MPN pathobiology. Indeed, recently published works from distinct research groups[69,70,85] highlight protein quality control in ER-associated degradation and proteostasis[69,70] deregulation as a primary effector of myeloid transformation, highlighting the importance of protein homeostasis for normal hematopoiesis. These findings too are in line with reports[86] implicating chronic ER stress, malfunctioning protein quality control, and loss of proteostasis as aggravating factors in age-related disorders.

Most importantly, platelet gene expression profiling in MPN offers directions for prediction of MF. Applying machine-learning algorithms of LASSO penalized regression under two conditions of external validation[45]—temporal (using our two-cohort design) and geographical (independently published datasets on healthy donors[3,4] and MF[4])—we uniquely discriminate MPN subtypes from each other and healthy controls, using three model types, and predict MF at high accuracy. The highest performing model used a set of progressively differentiated MPN genes at an area under the (ROC) curve of 0.96 (temporal) and 0.97 (geographical), and rendered a core signature of < 5 candidate markers as top predictors of disease progression. It will be of interest to determine what machine-learning algorithms based on a defined platelet gene expression classifier on potential new MPN datasets (ideally longitudinal) can be used to more precisely predict the probability and/or timing of an individual's risk of progression from ET/PV to secondary MF.

In conclusion, using platelet transcriptome profiling, we observed dynamic shifts in MPN immune inflammatory profile and preferential expression of interferon-, proliferation-, and proteostasis-associated genes as a progressive gradient across the three MPN subtypes. Our findings highlight that MPN progression may be influenced by defects in protein homeostasis (impaired protein folding and an accumulation of misfolded proteins within the endoplasmic reticulum) and an abnormal integrated stress response—consistent with recent studies[61,69,70,85] indicating dysregulated proteostasis as a primary effector of myeloid transformation. While this work has been focused on the overarching progressive platelet transcriptome across MPNs, these data open an important avenue for utilizing platelet RNA signatures to better understand specific MPN complications such as the risk of thrombosis and bleeding, or fibrosis, and transformation to AML. Altogether, this study in chronic MPNs provides a comprehensive framework for exploiting the platelet transcriptome and may inform future studies toward mechanistic understanding and therapeutic development in MPNs and potentially, other age-related disorders.

Limitations of the study

There are several limitations to our study. First, our data are not longitudinal by design but rather the closest practical alternative of cross-sectional snapshots of all three MPN subtypes with the goal of achieving a well-powered dataset in these chronic disorders. In this regard, the progressive or progression terminology used here refer strictly to trends in gene expression and do not imply study of longitudinal clinical progression. Therefore, subsidiary longitudinal evaluation of the disease as well as treatment markers identified here is warranted. Second, our focus for this study has been on the platelet transcriptome alone. Future investigations focused on ascertaining overlap between our platelet-derived molecular alterations with those of other cell types, specifically, parent megakaryocytes, CD34+ cells, granulocytes/immune cells, and even whole blood will be required to identify additional functional aspects of bone marrow pathology. Follow-on studies comparing MPN platelet transcriptomic data with other published data on non-malignant controls would also add significant value to the current study. Such integrative analyses may necessitate advanced systems genomics methods that compare or combine data without biases and batch effects inherent in each cell data type. Third, we recognize that our choice of ribosomal RNA depletion to home in on platelet mRNA signatures leaves out additional diversity in the platelet RNA repertoire (and will be important future work). Lastly, in our Lasso predictive modeling, we demonstrate two rigorous approaches of external validation (temporal and geographical) and identify a core signature toward MPN risk stratification or early detection of progression. Yet, substantive future biological and computational validations are needed to advance our findings toward clinical decision making or personalized medicine.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects
- METHOD DETAILS
  - Platelet isolation
  - Next-generation RNA sequencing
  - Platelet transcriptome analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Pathway/gene set enrichment for differentially expressed (DE) genes
  - Predictive model generation and external validation

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2021.100425.

ACKNOWLEDGMENTS

This work was funded by US National Institutes of Health grants 1K08HG010061-01A1 and 3UL1TR001085-04S1 (research re-entry award) to A.K., 1S10OD018220 and 1S10OD021763 shared instrumentation grants to the Stanford Functional Genomics Facility and the Stanford Research Computing Center, and the Charles and Ann Johnson Foundation to J.G. A.K. would like to thank mentors, Profs. Richard Becker (University of Cincinnati, OH), Andrew Weyrich (University of Utah, UT), Harry Greenberg (Stanford University, CA), Rob Tibshirani (Stanford University, CA), and Stephen Montgomery (Stanford University, CA) for their support in her unique return to research and NIH funding after a 5-year hiatus that then led to the work outlined in this manuscript. All authors thank the patients at the Stanford Cancer Center for their generous participation in this research, and the Stanford Functional Genomics Facility for genomic data storage. Authors also thank Drs. Belinda Guo and Wendy Erber (University of Western Australia, Australia) for the gene counts file of their published data (PMID 31426129) used for our independent validation. A.K. extends special thanks to colleagues Dr. Kellie Machlus (Harvard University, MA) and Dr. Eric Pietras (University of Colorado, CO) for their critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

A.K., J.G., and J.Z. conceived of the overall study. A.K. designed the study and secured funding that initiated this research. L.F., C.P., and J.G. provided samples and clinical annotation and reviewed the clinical data. A.K. designed the experimental plan with input from J.R., H.M., J.G., and J.Z. A.K. coordinated, performed, and oversaw the sample acquisition and processing. V.N. performed RNA isolation and library preparation. A.K. coordinated and oversaw sample sequencing. Z.S., W.D., and A.K. performed and interpreted the computational analyses. A.K., J.R., H.M., J.G., and J.Z. interpreted the data. A.K. wrote and edited the manuscript; C.P., W.D., J.R., H.M., J.G., and J.Z. critically reviewed and edited the manuscript. All authors approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. One or more of the authors of this paper received support from a program designed to increase minority representation in science. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: May 8, 2021
Revised: August 8, 2021
Accepted: September 23, 2021
Published: October 19, 2021

REFERENCES

1. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
2. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw *33*, 1–22.
3. Rondina, M.T., Voora, D., Simon, L.M., Schwertz, H., Harper, J.F., Lee, O., Bhatlekar, S.C., Li, Q., Eustes, A.S., Montenont, E., et al. (2020). Longitudinal RNA-seq analysis of the repeatability of gene expression and splicing in human platelets identifies a platelet *SELP* splice QTL. Circ. Res. *126*, 501–516.
4. Guo, B.B., Linden, M.D., Fuller, K.A., Phillips, M., Mirzai, B., Wilson, L., et al. (2020). Platelets in myeloproliferative neoplasms have a distinct transcript signature in the presence of marrow fibrosis. Br. J. Haematol. *188*, 272–282. https://doi.org/10.1111/bjh.16152.
5. Finazzi, G., De Stefano, V., and Barbui, T. (2013). Are MPNs vascular diseases? Curr. Hematol. Malig. Rep. *8*, 307–316.
6. Barbui, T., and Falanga, A. (2016). Molecular biomarkers of thrombosis in myeloproliferative neoplasms. Thromb. Res. *140* (*Suppl 1*), S71–S75.
7. Spivak, J.L. (2017). Myeloproliferative Neoplasms. N. Engl. J. Med. *376*, 2168–2181.
8. Spivak, J.L., Considine, M., Williams, D.M., Talbot, C.C., Jr., Rogers, O., Moliterno, A.R., Jie, C., and Ochs, M.F. (2014). Two clinical phenotypes in polycythemia vera. N. Engl. J. Med. *371*, 808–817.
9. Rumi, E., and Cazzola, M. (2017). Diagnosis, risk stratification, and response evaluation in classical myeloproliferative neoplasms. Blood *129*, 680–692.
10. Vainchenker, W., and Kralovics, R. (2017). Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. Blood *129*, 667–679.
11. Zoi, K., and Cross, N.C. (2017). Genomics of Myeloproliferative Neoplasms. J. Clin. Oncol. *35*, 947–954.
12. Nguyen, H.M., and Gotlib, J. (2012). Insights into the molecular genetics of myeloproliferative neoplasms. Am Soc Clin Oncol Educ Book *2021*, 411–418.
13. Oh, S.T., and Gotlib, J. (2010). JAK2 V617F and beyond: Role of genetics and aberrant signaling in the pathogenesis of myeloproliferative neoplasms. Expert Rev. Hematol. *3*, 323–337.
14. Hinds, D.A., Barnholt, K.E., Mesa, R.A., Kiefer, A.K., Do, C.B., Eriksson, N., Mountain, J.L., Francke, U., Tung, J.Y., Nguyen, H.M., et al. (2016). Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. Blood *128*, 1121–1128.
15. Rampal, R., Al-Shahrour, F., Abdel-Wahab, O., Patel, J.P., Brunel, J.-P., Mermel, C.H., Bass, A.J., Pretz, J., Ahn, J., Hricik, T., et al. (2014). Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. Blood *123*, e123–e133.
16. Mascarenhas, J., Roper, N., Chaurasia, P., and Hoffman, R. (2011). Epigenetic abnormalities in myeloproliferative neoplasms: a target for novel therapeutic strategies. Clin. Epigenetics *2*, 197–212.
17. Tefferi, A., Abdel-Wahab, O., Cervantes, F., Crispino, J.D., Finazzi, G., Girodon, F., et al. (2011). Mutations with epigenetic effects in myeloproliferative neoplasms and recent progress in treatment: Proceedings from the 5th International Post-ASH Symposium. Blood Cancer J. *1*, e7. https://doi.org/10.1038/bcj.2011.4.
18. Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N., et al. (2016). Genomic Classification and Prognosis in Acute Myeloid Leukemia. N. Engl. J. Med. *374*, 2209–2221.
19. Rondina, M.T., Weyrich, A.S., and Zimmerman, G.A. (2013). Platelets as cellular effectors of inflammation in vascular diseases. Circ. Res. *112*, 1506–1519.
20. Weyrich, A.S., and Zimmerman, G.A. (2013). Platelets in lung biology. Annu. Rev. Physiol. *75*, 569–591.
21. Weyrich, A.S. (2014). Platelets: More than a sack of glue. Hematology (Am. Soc. Hematol. Educ. Program) *2014*, 400–403.
22. Rowley, J.W., Schwertz, H., and Weyrich, A.S. (2012). Platelet mRNA: The meaning behind the message. Curr. Opin. Hematol. *19*, 385–391.

23. Schubert, S., Weyrich, A.S., and Rowley, J.W. (2014). A tour through the transcriptional landscape of platelets. Blood *124*, 493–502.

24. Simon, L.M., Edelstein, L.C., Nagalla, S., Woodley, A.B., Chen, E.S., Kong, X., Ma, L., Fortina, P., Kunapuli, S., Holinstat, M., et al. (2014). Human platelet microRNA-mRNA networks associated with age and gender revealed by integrated plateletomics. Blood *123*, e37–e45.

25. Davizon-Castillo, P., Rowley, J.W., and Rondina, M.T. (2020). Megakaryocyte and platelet tanscriptomics for discoveries in human health and disease. Arterioscler. Thromb. Vasc. Biol. *40*, 1432–1440.

26. Best, M.G., Sol, N., Kooi, I., Tannous, J., Westerman, B.A., Rustenburg, F., Schellen, P., Verschueren, H., Post, E., Koster, J., et al. (2015). RNA-Seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. Cancer Cell *28*, 666–676.

27. Clancy, L., and Freedman, J.E. (2016). Blood-derived extracellular RNA and platelet pathobiology: Adding pieces to a complex circulating puzzle. Circ. Res. *118*, 374–376.

28. Campbell, R.A., Franks, Z., Bhatnagar, A., Rowley, J.W., Manne, B.K., Supiano, M.A., Schwertz, H., Weyrich, A.S., and Rondina, M.T. (2018). Granzyme A in human platelets regulates the synthesis of proinflammatory cytokines by monocytes in aging. J. Immunol. *200*, 295–304.

29. Cunin, P., Bouslama, R., Machlus, K.R., Martinez-Bonet, M., Lee, P.Y., Wactor, A., Nelson-Maney, N., Morris, A., Guo, L., Weyrich, A., et al. (2019). Megakaryocyte emperipolesis mediates membrane transfer from intracytoplasmic neutrophils to platelets. elife *8*, e44031.

30. Middleton, E.A., Rowley, J.W., Campbell, R.A., Grissom, C.K., Brown, S.M., Beesley, S.J., Schwertz, H., Kosaka, Y., Manne, B.K., Krauel, K., et al. (2019). Sepsis alters the transcriptional and translational landscape of human and murine platelets. Blood *134*, 911–923.

31. Wen, Q.J., Yang, Q., Goldenson, B., Malinge, S., Lasho, T., Schneider, R.K., Breyfogle, L.J., Schultz, R., Gilles, L., Koppikar, P., et al. (2015). Targeting megakaryocytic-induced fibrosis in myeloproliferative neoplasms by AURKA inhibition. Nat. Med. *21*, 1473–1480.

32. Gilles, L., Finke, C., Lasho, T.L., Pardanani, A., Tefferi, A., and Crispino, J. (2012). Aberrant megakaryocyte gene expression contributes to primary myelofibrosis. Blood *120*, 2867.

33. Wen, Q.J., Woods, B., Yang, Q., Sophia, C., Lillu, G., Rapaport, F., et al. (2016). Activation of JAK/STAT signaling in megakaryocytes is necessary and sufficient for myeloproliferation in vivo. Blood *128*, 949.

34. Woods, B., Chen, W., Chiu, S., Marinaccio, C., Fu, C., Gu, L., Bulic, M., Yang, Q., Zouak, A., Jia, S., et al. (2019). Activation of JAK/STAT signaling in megakaryocytes sustains myeloproliferation *in vivo*. Clin. Cancer Res. *25*, 5901–5912.

35. Krause, D.S., and Crispino, J.D. (2013). Molecular pathways: induction of polyploidy as a novel differentiation therapy for leukemia. Clin. Cancer Res. *19*, 6084–6088.

36. Krishnan, A., Zhang, Y., Perkins, C., Gotlib, J., and Zehnder, J.L. (2017). Platelet transcriptomic signatures in myeloproliferative neoplasms. Blood *130*, 5288, 5288.

37. Skov, V., Larsen, T.S., Thomassen, M., Riley, C.H., Jensen, M.K., Bjerrum, O.W., Kruse, T.A., and Hasselbalch, H.C. (2011). Whole-blood transcriptional profiling of interferon-inducible genes identifies highly upregulated IFI27 in primary myelofibrosis. Eur. J. Haematol. *87*, 54–60.

38. Skov, V., Larsen, T.S., Thomassen, M., Riley, C.H., Jensen, M.K., Bjerrum, O.W., Kruse, T.A., and Hasselbalch, H.C. (2012). Molecular profiling of peripheral blood cells from patients with polycythemia vera and related neoplasms: Identification of deregulated genes of significance for inflammation and immune surveillance. Leuk. Res. *36*, 1387–1392.

39. Schischlik, F., Jäger, R., Rosebrock, F., Hug, E., Schuster, M., Holly, R., Fuchs, E., Milosevic Feenstra, J.D., Bogner, E., Gisslinger, B., et al. (2019). Mutational landscape of the transcriptome offers putative targets for immunotherapy of myeloproliferative neoplasms. Blood *134*, 199–210.

40. Gangaraju, R., Song, J., Kim, S.J., Tashi, T., Reeves, B.N., Sundar, K.M., Thiagarajan, P., and Prchal, J.T. (2020). Thrombotic, inflammatory, and HIF-regulated genes and thrombosis risk in polycythemia vera and essential thrombocythemia. Blood Adv. *4*, 1115–1130.

41. Cummings, B.B., Marshall, J.L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A.R., et al.; Genotype-Tissue Expression Consortium (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci. Transl. Med. *9*, eaal5209. https://doi.org/10.1126/scitranslmed.aal5209.

42. Rowley, J.W., Oler, A.J., Tolley, N.D., Hunter, B.N., Low, E.N., Nix, D.A., Yost, C.C., Zimmerman, G.A., and Weyrich, A.S. (2011). Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes. Blood *118*, e101–e111.

43. Amisten, S. (2012). A rapid and efficient platelet purification protocol for platelet gene expression studies. Methods Mol. Biol. *788*, 155–172.

44. Kukurba, K.R., and Montgomery, S.B. (2015). RNA sequencing and analysis. Cold Spring Harb. Protoc. *2015*, 951–969.

45. Moons, K.G., Kengne, A.P., Grobbee, D.E., Royston, P., Vergouwe, Y., Altman, D.G., and Woodward, M. (2012). Risk prediction models: II. External validation, model updating, and impact assessment. Heart *98*, 691–698.

46. Rowley, J.W., Weyrich, A.S., and Bray, P.F. (2019). The platelet transcriptome in health and disease. In Platelets (Elsevier), pp. 139–153.

47. Sampieri, L., Di Giusto, P., and Alvarez, C. (2019). CREB3 Transcription Factors: ER-Golgi Stress Transducers as Hubs for Cellular Homeostasis. Front. Cell Dev. Biol. *7*, 123.

48. Clark, R.A., Li, S.-L., Pearson, D.W., Leidal, K.G., Clark, J.R., Denning, G.M., Reddick, R., Krause, K.-H., and Valente, A.J. (2002). Regulation of calreticulin expression during induction of differentiation in human myeloid cells. Evidence for remodeling of the endoplasmic reticulum. J. Biol. Chem. *277*, 32369–32378.

49. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

50. Arcaini, L., and Cazzola, M. (2018). Benefits and risks of JAK inhibition. Blood *132*, 675–676.

51. Vainchenker, W., Leroy, E., Gilles, L., Marty, C., Plo, I., and Constantinescu, S.N. (2018). JAK inhibitors for the treatment of myeloproliferative neoplasms and other disorders. F1000Res. *7*, 82.

52. Kleppe, M., Koche, R., Zou, L., van Galen, P., Hill, C.E., Dong, L., De Groote, S., Papalexi, E., Hanasoge Somasundara, A.V., Cordner, K., et al. (2018). Dual Targeting of Oncogenic Activation and Inflammatory Signaling Increases Therapeutic Efficacy in Myeloproliferative Neoplasms. Cancer Cell *33*, 785–787.

53. Zhou, T., Georgeon, S., Moser, R., Moore, D.J., Caflisch, A., and Hantschel, O. (2014). Specificity and mechanism-of-action of the JAK2 tyrosine kinase inhibitors ruxolitinib and SAR302503 (TG101348). Leukemia *28*, 404–407.

54. Yan, B., Freiwald, T., Chauss, D., Wang, L., West, E., Mirabelli, C., Zhang, C.J., Nichols, E.-M., Malik, N., Gregory, R., et al. (2021). SARS-CoV-2 drives JAK1/2-dependent local complement hyperactivation. Sci. Immunol. *6*, eabg0833.

55. Passamonti, F., Cervantes, F., Vannucchi, A.M., Morra, E., Rumi, E., Pereira, A., Guglielmelli, P., Pungolino, E., Caramella, M., Maffioli, M., 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*, 1703–1708.

56. Zeddies, S., Jansen, S.B., di Summa, F., Geerts, D., Zwaginga, J.J., van der Schoot, C.E., von Lindern, M., and Thijssen-Timmer, D.C. (2014).

MEIS1 regulates early erythroid and megakaryocytic cell fate. Haematologica *99*, 1555–1564.

57. Caparrós-Pérez, E., Teruel-Montoya, R., López-Andreo, M.J., Llanos, M.C., Rivera, J., Palma-Barqueros, V., Blanco, J.E., Vicente, V., Martínez, C., and Ferrer-Marín, F. (2017). Comprehensive comparison of neonate and adult human platelet transcriptomes. PLoS ONE *12*, e0183042.

58. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. J. R. Stat. Soc. B *58*, 267–288.

59. Mead, T.J., and Apte, S.S. (2018). ADAMTS proteins in human disorders. Matrix Biol. *71-72*, 225–239.

60. Wang, C.Y., Li, C.Y., Hsu, H.P., Cho, C.Y., Yen, M.C., Weng, T.Y., Chen, W.C., Hung, Y.H., Lee, K.T., Hung, J.H., et al. (2017). PSMB5 plays a dual role in cancer development and immunosuppression. Am. J. Cancer Res. *7*, 2103–2120.

61. Skov, V., Larsen, T.S., Thomassen, M., Riley, C., Jensen, M.K., Bjerrum, O.W., Kruse, T.A., and Hasselbalch, H.C. (2010). Increased expression of proteasome-related genes in patients with pimary myelofibrosis. Blood *116*, 4117.

62. Moniz, L.S., Surinova, S., Ghazaly, E., Velasco, L.G., Haider, S., Rodríguez-Prados, J.C., Berenjeno, I.M., Chelala, C., and Vanhaesebroeck, B. (2017). Phosphoproteomic comparison of Pik3ca and Pten signalling identifies the nucleotidase NT5C as a novel AKT substrate. Sci. Rep. *7*, 39985.

63. Fruman, D.A., Chiu, H., Hopkins, B.D., Bagrodia, S., Cantley, L.C., and Abraham, R.T. (2017). The PI3K Pathway in Human Disease. Cell *170*, 605–635.

64. Brès, V., Yoh, S.M., and Jones, K.A. (2008). The multi-tasking P-TEFb complex. Curr. Opin. Cell Biol. *20*, 334–340.

65. Frydman, G.H., Tessier, S.N., Wong, K.H.K., Vanderburg, C.R., Fox, J.G., Toner, M., Tompkins, R.G., and Irimia, D. (2020). Megakaryocytes contain extranuclear histones and may be a source of platelet-associated histones during sepsis. Sci. Rep. *10*, 4621.

66. Vannucchi, A.M., Antonioli, E., Guglielmelli, P., Pardanani, A., and Tefferi, A. (2008). Clinical correlates of JAK2V617F presence or allele burden in myeloproliferative neoplasms: a critical reappraisal. Leukemia *22*, 1299–1307.

67. Tefferi, A., Lasho, T.L., Guglielmelli, P., Finke, C.M., Rotunno, G., Elala, Y., Pacilli, A., Hanson, C.A., Pancrazzi, A., Ketterling, R.P., et al. (2016). Targeted deep sequencing in polycythemia vera and essential thrombocythemia. Blood Adv. *1*, 21–30.

68. Williams, N., Lee, J., Moore, L., Baxter, E.J., Hewinson, J., Dawson, K.J., Menzies, A., Godfrey, A.L., Green, A.R., Campbell, P.J., and Nangalia, J. (2020). Phylogenetic reconstruction of myeloproliferative neoplasm reveals very early origins and lifelong evolution. BioRxiv, 2020.2011.2009. 374710. https://doi.org/10.1101/2020.11.09.374710.

69. LaFave, L.M., and Levine, R.L. (2016). Targeting a regulator of protein homeostasis in myeloproliferative neoplasms. Nat. Med. *22*, 20–21.

70. Osorio, F.G., Soria-Valles, C., Santiago-Fernández, O., Bernal, T., Mittelbrunn, M., Colado, E., Rodríguez, F., Bonzon-Kulichenko, E., Vázquez, J., Porta-de-la-Riva, M., et al. (2016). Loss of the proteostasis factor AIRAPL causes myeloid transformation by deregulating IGF-1 signaling. Nat. Med. *22*, 91–96.

71. Merlinsky, T.R., Levine, R.L., and Pronier, E. (2019). Unfolding the Role of Calreticulin in Myeloproliferative Neoplasm Pathogenesis. Clin. Cancer Res. *25*, 2956–2962.

72. Wong, W.J., Baltay, M., Getz, A., Fuhrman, K., Aster, J.C., Hasserjian, R.P., and Pozdnyakova, O. (2019). Gene expression profiling distinguishes prefibrotic from overtly fibrotic myeloproliferative neoplasms and identifies disease subsets with distinct inflammatory signatures. PLoS ONE *14*, e0216810.

73. Rontauroli, S., Castellano, S., Guglielmelli, P., Zini, R., Bianchi, E., Genovese, E., Carretta, C., Parenti, S., Fantini, S., Mallia, S., et al. (2021). Gene expression profile correlates with molecular and clinical features in patients with myelofibrosis. Blood Adv. *5*, 1452–1462.

74. Mullally, A., Bruedigam, C., Poveromo, L., Heidel, F.H., Purdon, A., Vu, T., Austin, R., Heckl, D., Breyfogle, L.J., Kuhn, C.P., et al. (2013). Depletion of Jak2V617F myeloproliferative neoplasm-propagating stem cells by interferon-α in a murine model of polycythemia vera. Blood *121*, 3692–3702.

75. Dunbar, A., Nazir, A., and Levine, R. (2017). Overview of transgenic mouse models of myeloproliferative neoplasms (MPNs). Curr. Protoc. Pharmacol. *77*, 14 40 11–14 40 19.

76. Matsuura, S., Thompson, C.R., Belghasem, M.E., Bekendam, R.H., Piasecki, A., Leiva, O., Ray, A., Italiano, J., Yang, M., Merill-Skoloff, G., et al. (2020). Platelet dysfunction and thrombosis in JAK2^{V617F}-mutated primary myelofibrotic mice. Arterioscler. Thromb. Vasc. Biol. *40*, e262–e272.

77. Barbui, T., Carobbio, A., Finazzi, G., Vannucchi, A.M., Barosi, G., Antonioli, E., Guglielmelli, P., Pancrazzi, A., Salmoiraghi, S., Zilio, P., et al.; AGIMM and IIC Investigators (2011). Inflammation and thrombosis in essential thrombocythemia and polycythemia vera: different role of C-reactive protein and pentraxin 3. Haematologica *96*, 315–318.

78. Geyer, H.L., Dueck, A.C., Scherber, R.M., and Mesa, R.A. (2015). Impact of inflammation on myeloproliferative neoplasm symptom development. Mediators Inflamm. *2015*, 284706.

79. Hasselbalch, H.C., and Bjørn, M.E. (2015). MPNs as inflammatory diseases: The evidence, consequences, and perspectives. Mediators Inflamm. *2015*, 102476.

80. Koschmieder, S., and Chatain, N. (2020). Role of inflammation in the biology of myeloproliferative neoplasms. Blood Rev. *42*, 100711.

81. Marin Oyarzún, C.P., and Heller, P.G. (2019). Platelets as mediators of thromboinflammation in chronic myeloproliferative neoplasms. Front. Immunol. *10*, 1373.

82. Mughal, T.I., Pemmaraju, N., Radich, J.P., Deininger, M.W., Kucine, N., Kiladjian, J.J., Bose, P., Gotlib, J., Valent, P., Chen, C.C., et al. (2019). Emerging translational science discoveries, clonal approaches, and treatment trends in chronic myeloproliferative neoplasms. Hematol. Oncol. *37*, 240–252.

83. Lussana, F., Carobbio, A., Salmoiraghi, S., Guglielmelli, P., Vannucchi, A.M., Bottazzi, B., Leone, R., Mantovani, A., Barbui, T., and Rambaldi, A. (2017). Driver mutations (JAK2V617F, MPLW515L/K or CALR), pentraxin-3 and C-reactive protein in essential thrombocythemia and polycythemia vera. J. Hematol. Oncol. *10*, 54.

84. Curto-Garcia, N., Harrison, C., and McLornan, D.P. (2020). Bone marrow niche dysregulation in myeloproliferative neoplasms. Haematologica *105*, 1189–1200.

85. Liu, L., Inoki, A., Fan, K., Mao, F., Shi, G., Jin, X., Zhao, M., Ney, G., Jones, M., Sun, S., et al. (2020). ER-associated degradation preserves hematopoietic stem cell quiescence and self-renewal by restricting mTOR activity. Blood *136*, 2975–2986.

86. Kaushik, S., and Cuervo, A.M. (2015). Proteostasis and aging. Nat. Med. *21*, 1406–1415.

87. Manne, B.K., Denorme, F., Middleton, E.A., Portier, I., Rowley, J.W., Stubben, C., Petrey, A.C., Tolley, N.D., Guo, L., Cody, M., et al. (2020). Platelet gene expression and function in patients with COVID-19. Blood *136*, 1317–1329.

88. Gnatenko, D.V., Dunn, J.J., McCorkle, S.R., Weissmann, D., Perrotta, P.L., and Bahou, W.F. (2003). Transcript profiling of human platelets using microarray and serial analysis of gene expression. Blood *101*, 2285–2293.

89. Raghavachari, N., Xu, X., Harris, A., Villagra, J., Logun, C., Barb, J., Solomon, M.A., Suffredini, A.F., Danner, R.L., Kato, G., et al. (2007). Amplified expression profiling of platelet transcriptome reveals changes in arginine metabolic pathways in patients with sickle cell disease. Circulation *115*, 1551–1562.

90. Weyrich, A.S., and Zimmerman, G.A. (2003). Evaluating the relevance of the platelet transcriptome. Blood *102*, 1550–1551.

91. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

92. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

93. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Peripheral blood samples from healthy adults and patients with one of three subtypes of chronic myeloproliferative neoplasms	This study	Deidentified
Platelets from healthy adults and patients with one of three subtypes of chronic myeloproliferative neoplasms	This study	N/A
Deposited data		
Deidentified human/patient platelet RNA-sequencing data	This study	dbGaP accession # Database: PHS-0021-21. v1.P1
Deidentified human/patient platelet RNA-sequencing data	Guo et al[4]	Data secured via corr author, W. Erber
Deidentified human platelet RNA-sequencing data	Rondina et al[3]	NCBI Bioproject ID Database: 531691
Software and algorithms		
DeSeq2	Bioconductor Open Source	N/A
LASSO/glmnet	CRAN R Package Open Source	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Anandi Krishnan (anandi.krishnan@stanford.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

De-identified human/patient platelet transcriptomic data have been deposited at NIH dbGaP repository and are available upon request from dbGaP. To request access, please contact dbGaP following the instructions at https://dbgap.ncbi.nlm.nih.gov/. Accession numbers are listed in the Key resources table.

This paper does not report original code. Any analyses applied are based on previously available software and established R packages, primarily i) DESeq2 (Michael Love[1]) with the package: http://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html and ii) Lasso/glmnet (Robert Tibshirani[2]; https://www.jstatsoft.org/v33/i01/) with https://cran.r-project.org/web/packages/glmnet/glmnet.pdf.

RNA-sequencing data from this work (original FASQ files from paired-end sequencing of all 120 samples) will be deposited to the NIH genomic data repository dbGAP under public accession # Database: PHS-0021-21.v1.P1. Previously published RNA-sequencing data used in this work as geographically independent validation cohorts are from Rondina et al.[3] (PMID 31852401, healthy donors) and Guo et al.[4] (PMID 31426129, MF patients and healthy donors). Source data from the work of Rondina et al.[3] is publicly available at NIH NCBI Bioproject ID 531691; and that of Guo et al.[4] is secured through reaching the corresponding author, Dr. Wendy Erber.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

All MPN peripheral blood samples were obtained under written informed patient consent and were fully anonymized. Study approval was provided by the Stanford University Institutional Review Board.

All relevant ethical regulations were followed. Subject demographic data and other relevant clinical variables are reported in Table S1A and S1B and Figure 1.

We collected blood from ninety-five MPN patients enrolled in the Stanford University and Stanford Cancer Institute Hematology Tissue Bank from December 2016- December 2019 after written informed consent from patients or their legally authorized representatives (Stanford IRB approval #18329). Eligibility criteria included age $\geq$ 18 years and Stanford MPN clinic diagnosis of essential thrombocythemia, polycythemia vera or myelofibrosis (defined using the consensus criteria at the time of this study). We use the term 'myelofibrosis' to encompass both primary myelofibrosis and myelofibrosis that evolved from essential thrombocythemia or polycythemia vera. Electronic medical records review of all subjects was performed by the clinical consultants (J.G. and L.F.), study data manager (C.P.), and the study principal investigator (A.K.). For controls, blood was collected from twenty-one asymptomatic adult donors selected at random from the Stanford Blood Center. All donors were asked for consent for genetic research. For both MPN patients and healthy controls, blood was collected into acid citrate-dextrose (ACD, 3.2%) sterile yellow-top tubes (Becton, Dickinson and Co.) and platelets were isolated by established[28,30,42,43] purification protocols. Blood was processed within 4 h of collection for all samples. The time from whole blood collection to platelet isolation was similar between healthy donors and MPN patients.

Power calculations based on preliminary data indicated 20 PMF, 25 PV and 25 ET patients to achieve 80% power and identify statistically significant differential expression. This was based on the total number of genes as 12000 and the top 3326 genes as prognostic in PMF, 1190 in PV and 385 in ET with a desired minimum fold change of 2 and a False Discovery Rate FDR < 0.05. Our final sample set exceeded our preliminary calculations.

METHOD DETAILS

Platelet isolation

Human platelets were isolated and leuko-depleted using established methods ([28,42,43] with excellent reproducibility[19,25,28,30,87] resulting in a highly purified population of fewer than 3 leukocytes/10^7 platelets (> 99.9% purity) as counted by hemocytometer. Briefly, the ACD-tube whole blood was first centrifuged at 200xg for 20min at room temperature (RT). The platelet-rich plasma (PRP) was removed and Prostaglandin E1 was added to the PRP to prevent exogenous platelet activation. The PRP was then centrifuged at 1000xg for 20min at RT. The platelet pellet was re-suspended in warmed (37°C) PIPES saline glucose (PSG). Leukocytes were depleted using CD45+ magnetic beads (Miltenyi Biotec). Isolated platelets were lysed in Trizol for RNA extraction. Post RNA-seq analysis of an index leukocyte transcript (*PTPRC; CD45*) confirmed that the samples were highly purified platelet preparations (subsequent bioinformatic analyses also adjusted for PTPRC expression for absolute removal of any CD45 expression in our analyses). Two reference markers of platelet activation, *P-selectin (SELP)* and *Glycoprotein IIbIIIA (CD41/ITGA2B)* were expectedly higher in all MPN than healthy controls but were not statistically significantly different between MPN subtypes; indicating that any expression difference was not due to experimental artifacts. In addition, we know from prior work (J.R.) that regardless of activation status, RNA-seq reliably estimates mRNA expression patterns in platelets. We also know from rigorous prior work[88–90] that several abundant platelet mRNAs are well-known leukocyte or red cell transcripts; and do not immediately imply contamination by these classes but rather that platelets express gene products that are also present in other cell lineages. Sixty-five of our top 100 abundant platelet transcripts matched exactly with those of the top 100 abundant genes from the three previous studies cited[42,88]; and a composite pathway analysis with the top 100 abundant genes from this as well as the previous studies matched identically.

Next-generation RNA sequencing

For next-generation RNA-sequencing (RNA-seq), 1x10^9 isolated platelets were lysed in Trizol and then DNase treated. Total RNA was isolated, and an Agilent bio-analyzer was used to quantify the amount and quality. The RNA yield was estimated by measuring absorbance at 260 nm on the Nanodrop 2000 (Thermo Fisher), and RNA purity was determined by calculating 260/280 nm and 260/230 nm absorbance ratios. RNA integrity was assessed on the Agilent Bioanalyzer using the RNA 6000 Nano Chip kit (Agilent Technologies). An RNA integrity number (RIN) was assigned to each sample by the accompanying Bioanalyzer Expert 2100 software. To control for variable RNA quality, RNA sequencing was only performed for samples with a RIN score of 7 or higher. RNA-seq libraries were constructed with removal of ribosomal RNA using the KAPA Stranded RNA-Seq kit with RiboErase (Roche). The RNA extraction and library preparation were performed by the same technician to minimize confounding effects. cDNA libraries were constructed following the Illumina TrueSeq Stranded mRNA Sample Prep Kit protocol and dual indexed. The average size and quality of each cDNA library were determined by the Agilent Bioanalyzer 2100, and concentrations were determined by Qubit for proper dilutions and balancing across samples. Twelve pooled samples with individual indices were run on an Illumina HiSeq 4000 (Patterned flow cell with Hiseq4000 SBS v3 chemistry) as 2 X 75bp paired end sequencing with a coverage goal of 40M reads/sample. Output BCL files were FASTQ-converted and demultiplexed.

Platelet transcriptome analysis

Picard, Samtools, and other metrics were used to evaluate data quality. Processed reads were aligned against the reference human transcriptome GRCh37/hg19 using RSEM[91] and bowtie2[92], and expression at gene level determined by calculating raw gene count. Only genes that passed expression threshold were used; genes were considered expressed if, in all samples, they had at least 10 counts (genes with low counts are automatically filtered by built-in functions in DeSeq2, see below). A total of 12,924 genes were considered expressed. Gene expression data was library-size-corrected, variance-stabilized, and log2-transformed using the R package DESeq2[1]. We refer to this version of the data as "raw data" as it is not corrected for any confounders of gene expression variability. DESeq2 was used to call differential expression while adjusting for patient age, gender and treatment as confounding variables and controlling for multiple comparisons using the Benjamini-Hochberg defined false discovery rate (FDR). Significant variance in expressed transcripts were pre-specified as those transcripts with an FDR < 0.05 and a log2 fold change $\geq$ 0.5 in MPN, as compared to healthy controls (the entire differential transcriptome was applied in the instances of downstream Gene Set Enrichment Analysis and the Lasso prediction modeling).

QUANTIFICATION AND STATISTICAL ANALYSIS

Continuous data were summarized as medians and IQRs and categorical data are presented as frequencies and percentages. To compare differences in clinical variables between healthy controls and MPN subtypes (ET, PV, and MF), we use box and whisker plots and conduct pairwise Wilcoxon signed ranked tests. For unsupervised clustering and visualization, we performed principal component analyses (identifying MPN subtypes by color, treatment by filled or open circles, and *JAK2* mutation status by shape) using built-in functions of the DeSeq2 R package. We generated a heatmap of all of the top highly significant genes (FDR < 0.01) using the pheatmap R package and its built-in functions for hierarchical cluster analysis on the sample-to-sample Euclidean distance matrix of the expression data. All analyses were performed using the R studio interface.

Pathway/gene set enrichment for differentially expressed (DE) genes

Gene set enrichment analysis (GSEA)[49] was performed on the entire DE gene set for each MPN subtype, using the Cancer Hallmarks gene sets from MSigDB[93]. The 'GSEA Pre-ranked' function was used with a metric score that combines fold change and adjusted p value together for improved gene ranking. We used default settings with 10,000 gene set permutations to generate *p* and *q* values, and compared MPN subtypes in the overall cohort, and the ruxolitinib-treated subgroup and the ruxolitinib-naive subgroup separately. In these analyses, to allow for a broad comparison, we assessed all transcripts that were differentially expressed according to FDR/adjusted $p < 0.25$ as recommended by the authors of GSEA[49].

Predictive model generation and external validation

At the conception of this study late 2016, and our early work[36], we did not identify any publicly available RNA sequencing data on MPN platelets. This prompted our specific two-cohort design for the express purpose of temporal external validation as an essential step in rigorous prediction modeling[45]. A subsequent independent publication[4] facilitated an additional geographically independent external validation of our model.

We used Lasso penalized regression[58] for our model to predict MF from the either healthy controls, ET or PV. Among a variety of statistical machine-learning algorithms that have been used in prediction modeling, Lasso is favored for its flexibility and simplicity; and its ability to identify the least set of significant factors from high dimensional data. We evaluated platelet transcriptomic features with clinical features (age, gender and mutation status for the entire dataset including healthy donors, and in MPN patients alone, platelet and hemoglobin values). Normalized gene counts data were split into training (used for constructing multinomial logistic models) and validation (used for model evaluation and generalization) cohorts. Separately, we assessed the progressive and monotonic upward or downward trend in gene expression using Mann-Kendall trend test (multiple comparison adjusted with the Benjamini-Hochberg method) to normalized gene counts and identified statistically significant progressive genes across all three MPN subtypes.

Three multinomial logistic models were constructed: first, with Lasso selected predictors from all genes, second, with Lasso selected predictors from progressive genes and third, a baseline model using age, gender, and mutation status (*JAK2 and CALR*) as predictors. Model outputs correspond to probabilities of having a CTRL, ET, PV, or MF phenotype (sum of these four probability values totaling 1). Potential interpretation of these probabilities includes MPN risk assessment, e.g., a patient with higher probabilities of PV and MF would indicate higher risk than one with higher probabilities of CTRL or ET. The[4] dataset on MF platelet RNA seq served as an independent test set (Figure 5D schematic) while data from our cohorts at Stanford and additional external data on healthy donors from[3] constituted an integrated training cohort (R package Limma was applied for bioinformatic correction of any batch effects).

ROC curves were used to evaluate the different prediction models and discriminate outcomes. ROC curves demonstrate the trade-off between true positive and false positive rates, ideal being high true positive rate (sensitivity) and low false positive rate (specificity) the area under the curve (AUROC) as close to 1 as possible. True positive rate (TPR) is defined as correctly predicting an MF patient as MF; and false positive rate (FPR) as falsely predicting a non-MF patient as MF.