Platelet transcriptome yields progressive markers in chronic myeloproliferative neoplasms and identifies putative targets of therapy

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*]

Affiliations

[1]Department of Statistics, Stanford University, Stanford, CA
[2]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA
[3]Stanford Functional Genomics Facility, Stanford University School of Medicine, Stanford, CA
[4]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA
[5]Department of Internal Medicine, University of Utah, Salt Lake City, UT
[6]Department of Medicine, Stanford University School of Medicine, Stanford, CA
[7]Department of Pathology, Stanford University, Stanford, CA
[8]These authors contributed equally
[#]Current Affiliation: Google LLC, Mountain View, CA

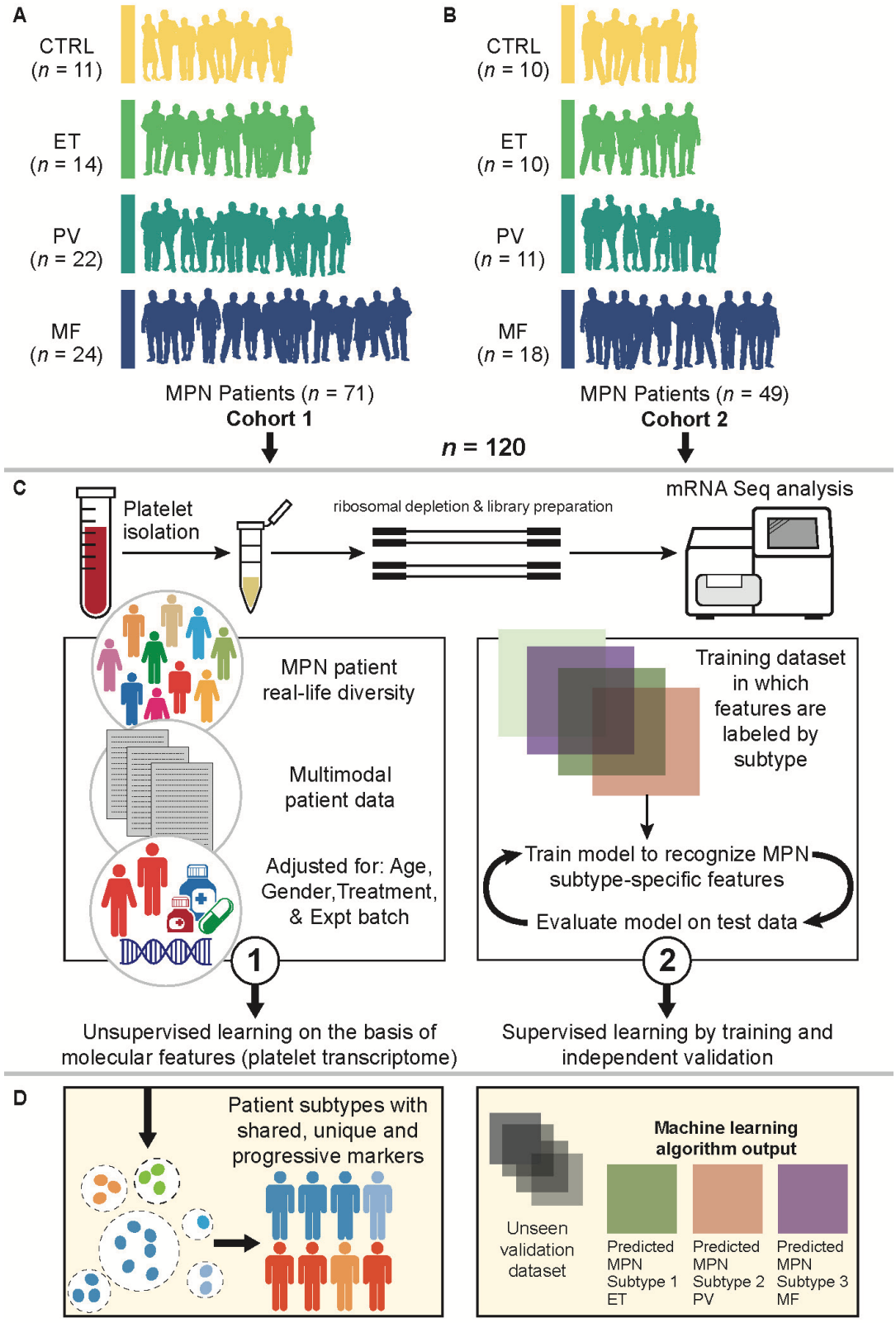
[*]Corresponding Author:
Anandi Krishnan, PhD
Stanford University, Stanford, CA
anandi.krishnan@stanford.edu

Highlights

Leveraging two independent and mutually validating MPN patient cohorts, we identify progressive transcriptomic markers that also enable validated prediction in MPNs.

Our platelet RNA-Seq data identifies impaired protein homeostasis as prominent in MPN progression and offers putative targets of therapy.

VISUAL ABSTRACT



Abstract

Predicting disease natural history remains a particularly challenging endeavor in chronic degenerative disorders and cancer, thus limiting early detection, risk stratification, and preventive interventions. Here, profiling the spectrum of chronic myeloproliferative neoplasms (MPNs), as a model, we identify the blood platelet transcriptome as a proxy strategy for highly sensitive progression biomarkers that also enable prediction via machine learning algorithms. Using RNA sequencing (RNA-seq), we derive disease-relevant gene expression in purified platelets from 120 peripheral blood samples constituting two independently collected and mutually validating patient cohorts of the three MPN subtypes – essential thrombocythemia, ET (n=24), polycythemia vera, PV (n=33), and primary or post ET/PV secondary myelofibrosis, MF (n=42), as well as healthy donors (n=21). The MPN platelet transcriptome discriminates each clinical phenotype and reveals an incremental molecular reprogramming that is independent of patient driver mutation status or therapy. Leveraging this dataset, in particular the progressive expression gradient noted across MPN, we develop a machine learning model (Lasso-penalized regression) predictive of the advanced subtype MF at high accuracy under two conditions of validation: i) temporal internal (AUC-ROC of 0.96) and ii) geographic external cohorts (AUC-ROC of 0.86). Lasso-derived signatures offer a robust core set of < 10 MPN progression markers. Mechanistic insights from our data highlight impaired protein homeostasis as a prominent driver of MPN evolution, with persistent integrated stress response. We also identify JAK inhibitor-specific signatures and other interferon, proliferation, and proteostasis-associated markers as putative targets for MPN-directed therapy. Our platelet transcriptome snapshot of chronic MPNs demonstrates a methodological avenue for disease risk stratification and progression beyond genetic data alone, with potential utility in a wide range of age-related disorders.

82 **Introduction**

83 The classic Philadelphia chromosome-negative (Ph^-) MPNs,(Barbui and Falanga, 2016; Finazzi et al., 2013;
84 Spivak, 2017; Spivak et al., 2014) are clonal disorders of the bone marrow that comprise three clinical
85 phenotypes– essential thrombocythemia (ET), polycythemia vera (PV), and myelofibrosis (MF). These
86 myeloid neoplasms are defined by a combination of morphologic, clinical, laboratory, and
87 cytogenetic/molecular genetic features. The existing genetic landscape (Rumi and Cazzola, 2017;
88 Vainchenker and Kralovics, 2017; Zoi and Cross, 2017) of MPNs primarily involves mutations in three driver
89 genes that lead to constitutive JAK-STAT signaling (*JAK2, CALR, MPL*). Several additional non-driver
90 mutations (see references (Hinds et al., 2016; Nguyen and Gotlib, 2012; Oh and Gotlib, 2010; Rampal et
91 al., 2014; Rumi and Cazzola, 2017; Vainchenker and Kralovics, 2017; Zoi and Cross, 2017) for details) as
92 well as cytogenetic (Spivak, 2017) and epigenetic (Mascarenhas et al., 2011; Tefferi et al., 2011)
93 abnormalities also contribute to disease initiation and progression, and impact both overall survival and
94 potential for progression to acute myeloid leukemia (AML) (Papaemmanuil et al., 2016). Depending on the
95 MPN and stage of disease, these patients may exhibit debilitating constitutional symptoms such as
96 fatigue, pruritis, night sweats, and weight loss; thrombohemorrhagic diathesis and extramedullary
97 hematopoiesis; and an increased propensity for transformation to AML. Although an increase in one or
98 more blood cell lineages contributes to these morbid sequelae, the qualitative abnormalities of myeloid
99 cells that increase vascular risk or disease progression are not well understood. Taken together, a limited
100 understanding exists regarding how genotypic variability contributes to diverse phenotypic presentations
101 and disease natural histories. We were motivated by the current clinical need (Rumi and Cazzola, 2017;
102 Spivak, 2017; Vainchenker and Kralovics, 2017) and the potential for deeper integration of clinical features
103 and genetics with gene expression profiling to improve stratification and management of chronic blood
104 disorders, such as MPNs.

105

106 Blood platelets play critical roles in multiple processes and diseases (Rondina et al., 2013; Weyrich, 2014;
107 Weyrich and Zimmerman, 2013), from their traditional function in hemostasis and wound healing to
108 inflammation, immunity, cancer metastasis, and angiogenesis. Platelets originate from bone marrow
109 precursor megakaryocytes as anucleate fragments with a distinctive discoid shape. Platelets contain a

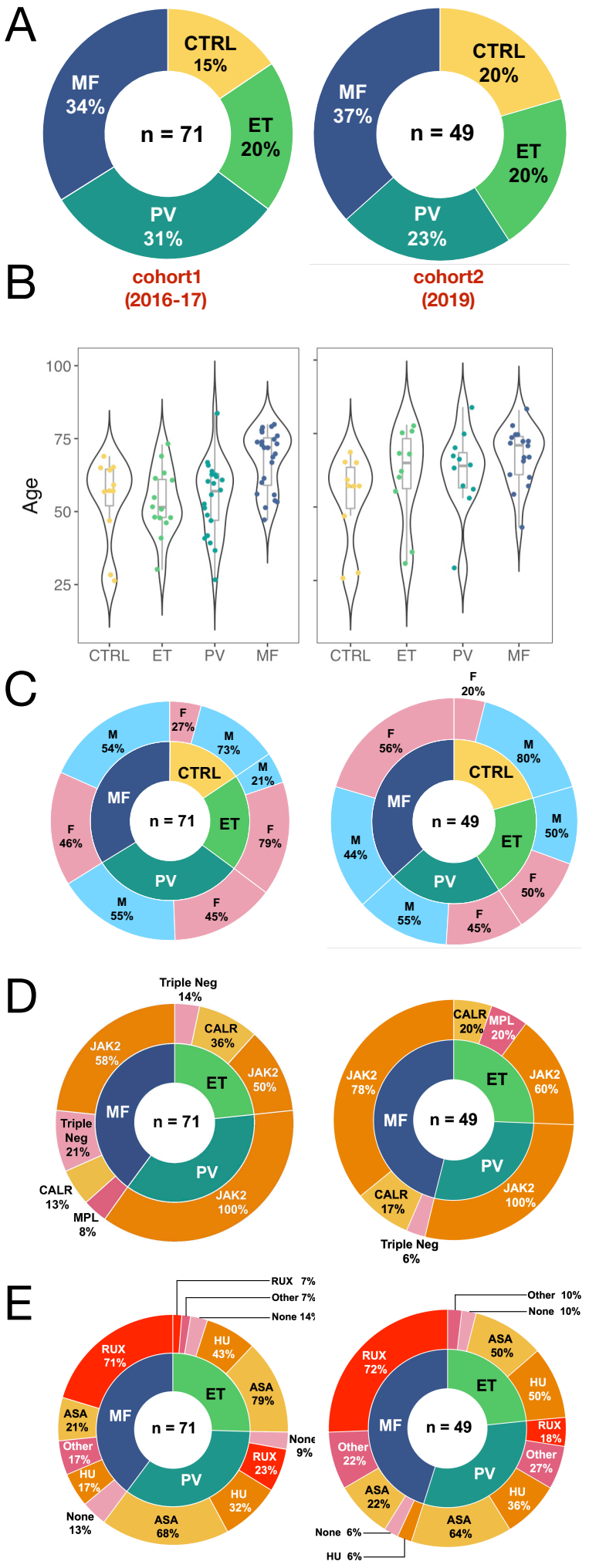
110 complex transcriptional landscape of messenger RNAs (mRNAs), unspliced pre-mRNAs, rRNAs, tRNAs
111 and microRNAs (Rowley et al., 2012; Schubert et al., 2014; Simon et al., 2014; Weyrich, 2014). Most
112 platelet RNA expression results from the transcription of nuclear DNA in the megakaryocyte (Davizon-
113 Castillo et al., 2020; Rowley et al., 2012), and thus reflects the status of the megakaryocyte at the time of
114 platelet release into the circulation, as well as subsequent splicing events, and selective packaging and
115 inter-cellular RNA transfer. There is emerging evidence (Best et al., 2015; Campbell et al., 2018; Clancy
116 and Freedman, 2016; Cunin et al., 2019; Middleton et al., 2019) that the molecular signature of platelets
117 may be changed in disease conditions where these processes are altered, including via inter-cellular
118 transfer (Clancy and Freedman, 2016; Cunin et al., 2019) of cytosolic RNA. In the context of MPNs, the
119 platelet transcriptome therefore not only represents a critical biomarker of megakaryocytic activity (Gilles et
120 al., 2012; Krause and Crispino, 2013; Wen et al., 2016; Wen et al., 2015; Woods et al., 2019), but also
121 provides a snapshot of the underlying hemostatic, thrombotic, and inflammatory derangements associated
122 with these hematologic neoplasms and the potential impact of treatment (Woods et al., 2019).
123
124 To date, no one composite study including our own (Krishnan et al., 2017) (Gangaraju et al., 2020; Guo et
125 al., 2019; Rampal et al., 2014; Schischlik et al., 2019; Skov et al., 2011; Skov et al., 2012) has evaluated
126 the disease-relevant platelet transcriptome in a sizeable clinical cohort of all three subtypes of Ph^- MPNs.
127 Here, we extend our prior preliminary work(Krishnan et al., 2017) toward a comprehensive analysis of
128 disease-relevant (Cummings et al., 2017) platelet RNA-sequencing (RNA-seq) in two temporally
129 independent and mutually validating cohorts of all three MPN subtypes, ET, PV, and MF (primary or post
130 ET/PV secondary). We demonstrate marked differences in platelet gene expression across the MPN
131 spectrum, which also permits robust validated (temporal internal and geographic external) prediction of
132 MF. In addition to identifying novel gene expression signatures impacted by the JAK1/JAK2 inhibitor
133 ruxolitinib (RUX), platelet profiling reveals MPN-altered pathways that may be targets for future drug
134 development.
135
136
137

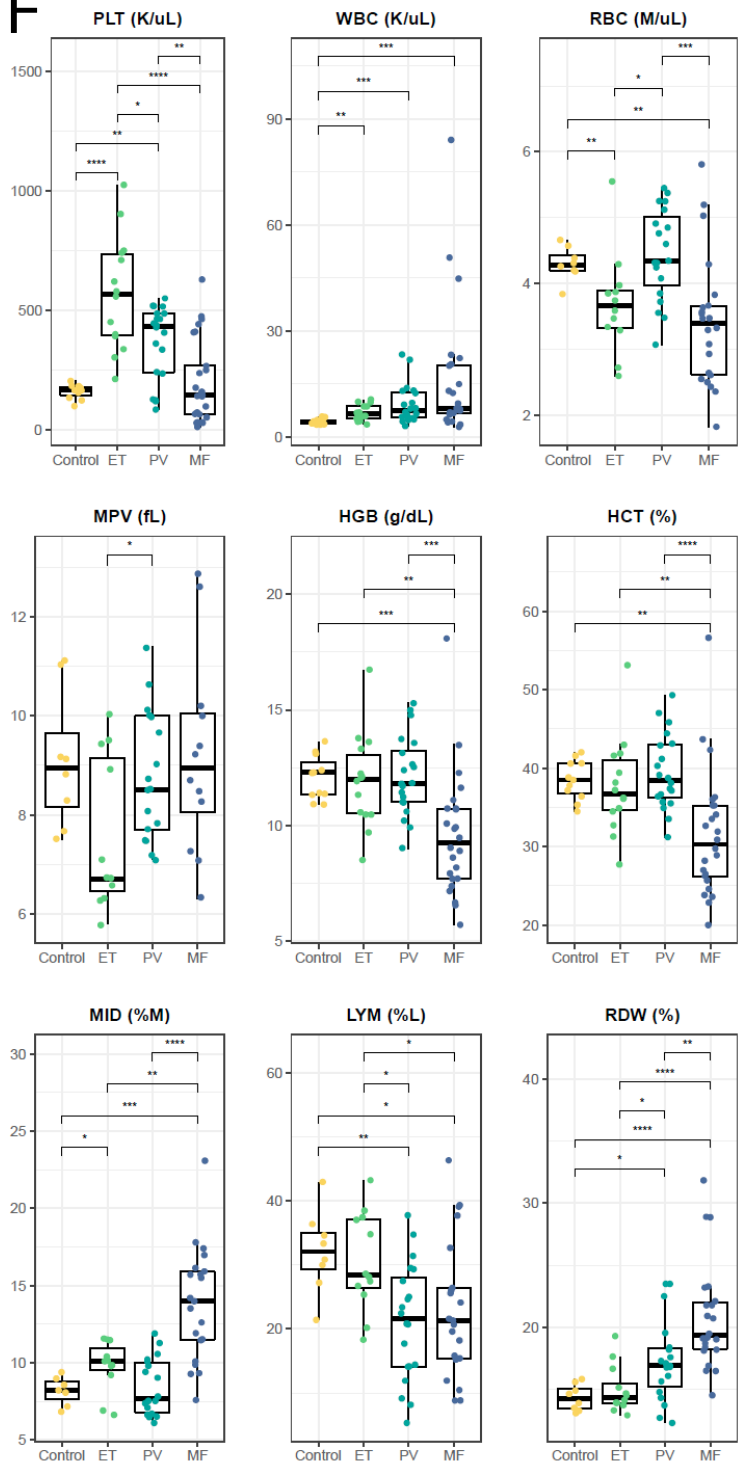
138 **Results**

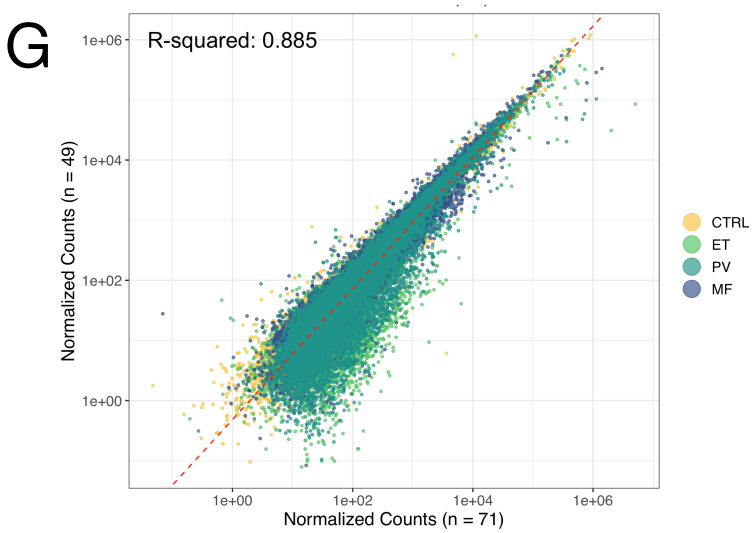
139 **Two independent MPN clinical cohorts and closely replicated platelet transcriptome**

140 We prepared highly purified leukocyte-depleted (Amisten, 2012; Rowley et al., 2011) platelets from
141 peripheral blood samples of two cohorts of World Health Organization-defined MPN patients
142 (approximately 2 years apart; Stanford single-center); and included healthy controls in each cohort as
143 reference (cohort 1, n = 71, and cohort 2, n=49; **Figure 1A**). Our two-cohort study was specifically
144 designed (before knowledge of other subsequent studies) for the purpose of validation, not only of inter-
145 cohort RNA-seq results (Kukurba and Montgomery, 2015) but also for temporal validation (Moons et al.,
146 2012) of our prediction model (see Methods). **Figure S1** demonstrates our established(Rowley et al., 2011;
147 Rowley et al., 2019) high-quality and highly efficient experimental framework toward a rigorous platelet
148 RNA-seq approach. Clinical features of the MPN patients are shown in **Figure 1**; and listed in **Table S1A-**
149 **B**. The distribution of key variables was closely matched between the two cohorts by MPN subtype
150 (**Figure 1A**), age (**B**), gender (**C**), *JAK2/CALR* mutation status (**D**) and treatment (**E**). Any inter-patient
151 variability in patient age, gender and treatment are adjusted as confounding factors in all downstream gene
152 expression analyses (see Methods). Clinical laboratory measures (**Figure 1F**) at the time of sampling
153 reflected the phenotype of the MPN subtypes (please see figure legend for detailed statistical
154 comparisons).The two cohorts of platelet transcriptome data (**Figure 1G**, normalized transcript counts)
155 adjusted for patient age, gender and all treatment as confounding factors were also highly correlated (R^2 =
156 0.89), thus demonstrating high inter-cohort validation of gene expression that then enabled us to combine
157 our two RNA-seq datasets into a final integrated data set of enhanced statistical power for downstream
158 analyses. Together, this platelet transcriptome compendium spans 120 human peripheral blood samples
159 from healthy controls (n=21) and patients diagnosed with an MPN at the time of sample acquisition (24 ET,
160 33 PV and 42 MF) that include seven untreated, and 92 either on cytoreductives/biologics (*e.g.* ruxolitinib,
161 hydroxyurea, interferon-alpha), anti-thrombotic agents (e.g. aspirin, warfarin), or a combination and
162 captures the real-life diversity among MPN patients. Our cross-sectional design here capturing patients
163 from all three MPN subtypes also serves as a practical alternative to the longitudinal approach, though
164 ideal, is likely difficult to implement in these chronic disorders with often decades of disease.

FIGURE 1


A
CTRL 15%
ET 20%
PV 31%
MF 34%
n = 71
CTRL 20%
ET 20%
PV 23%
MF 37%
n = 49
cohort1 (2016-17)
cohort2 (2019)
B
Age
CTRL
ET
PV
MF
C
n = 71
n = 49
D
Triple Neg 14%
CALR 36%
JAK2 50%
JAK2 100%
MPL 8%
CALR 13%
Triple Neg 21%
JAK2 58%
CALR 20%
MPL 20%
JAK2 60%
JAK2 100%
CALR 17%
Triple Neg 6%
JAK2 78%
E
RUX 7%
Other 7%
None 14%
HU 43%
ASA 79%
RUX 23%
HU 32%
ASA 68%
None 13%
HU 17%
Other 17%
ASA 21%
RUX 71%
None 9%
Other 10%
None 10%
ASA 50%
HU 50%
RUX 18%
Other 27%
HU 36%
ASA 64%
HU 6%
None 6%
ASA 22%
Other 22%
RUX 72%


F
PLT (K/uL)
WBC (K/uL)
RBC (M/uL)
MPV (fL)
HGB (g/dL)
HCT (%)
MID (%M)
LYM (%L)
RDW (%)
Control
ET
PV
MF

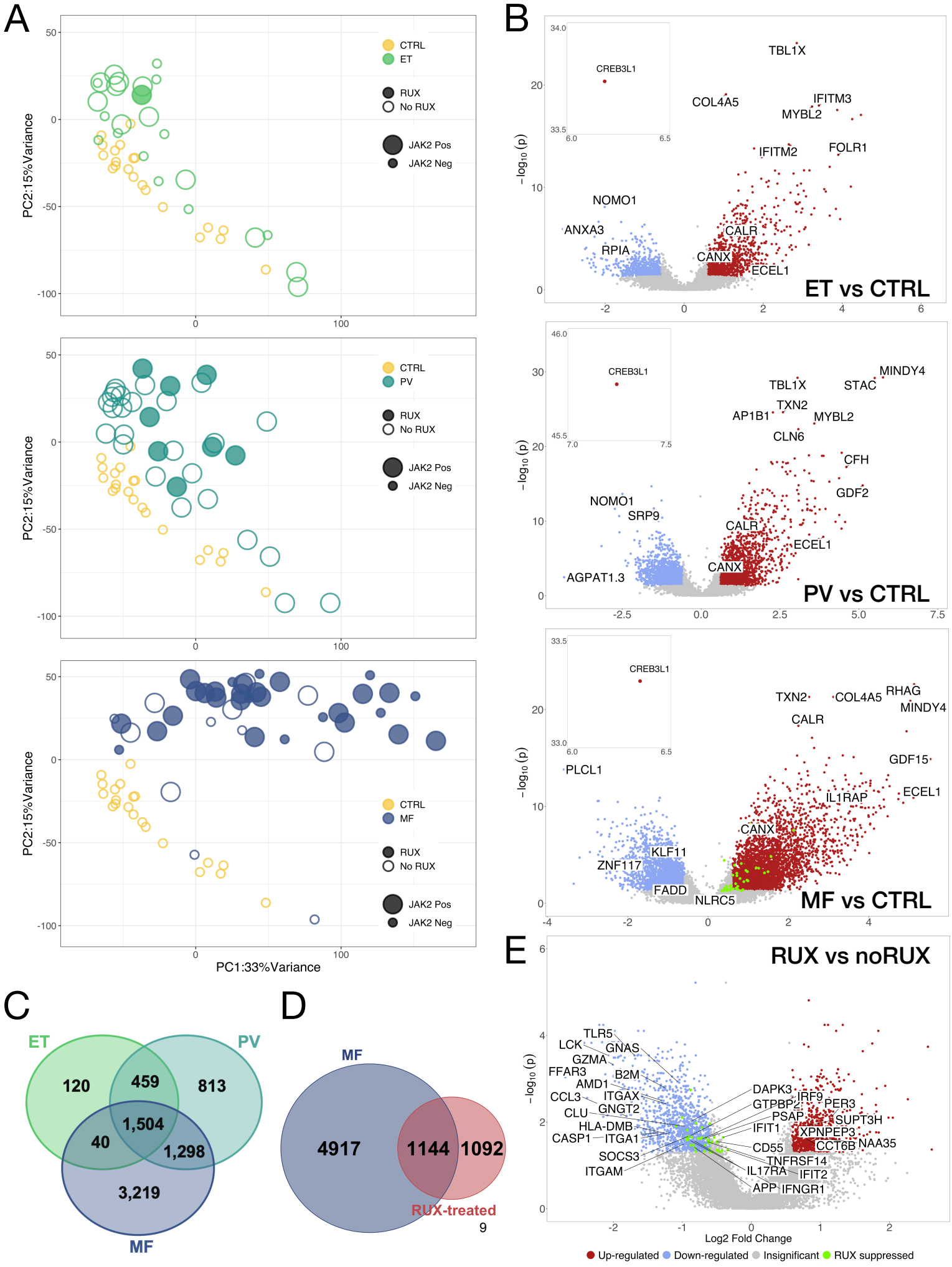

G
R-squared: 0.885
Normalized Counts (n = 49)
Normalized Counts (n = 71)
CTRL
ET
PV
MF

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

167 Given the phenotypic overlap, yet also differences in disease behavior and prognosis between ET, PV and
168 MF, we hypothesized that there may be MPN subtype-specific differences in gene expression that are
169 independent of *JAK2/CALR/MPL* mutation status. In addition, we hypothesized that MPN platelets are
170 likely to be enriched for subtype-specific biomarkers that may otherwise be missed in blood/plasma/serum
171 sources (Schischlik et al., 2019; Skov et al., 2011; Skov et al., 2012). Therefore, we compared platelet
172 transcriptomic expression in each of the three MPN subtypes with that of controls and discovered a
173 shared gene set that is also progressively differentiated across the MPN spectrum (ET/PV to MF **Figure 2**
174 **A-C**). First, unsupervised principal component analysis (PCA) of MPN patients and controls data (**Figure**
175 **2A**) confirmed that the main source of variability, after adjusting for age, gender, treatment and
176 experimental batch, was MPN disease status, with increasing differentiation by subtype (the first two
177 principal components account for 48% of total variance in original data). Next, differential gene expression
178 analysis (DGEA, volcano plot**, Figure 2B, C**) efficiently distinguished each of the MPN subtypes and
179 resulted in highly significant expression signatures (adjusted p-value/FDR <0.05) with 2090 genes
180 differentially regulated in ET (1202 up and 888 down), 3999 in PV (1954 up and 2045 down), and 5930 in
181 MF (3591 up and 2339 down).

182 Specifically, DGEA also uncovered shared and unique genes between all three MPN phenotypes, thus
183 offering a potential core set of genes involved in MPN pathogenesis (**Figure 2C**). The shared gene set at
184 FDR < 0.01 constituted 583 up-regulated genes, with a predominance of molecular pathways involving
185 *myeloid cell activation in immune response*; and 188 down-regulated genes, with *myeloid cell*
186 *differentiation* as the predominant and consistent pathogenetic mechanism. The upregulated genes
187 belonged to the endoplasmic reticulum/ER-Golgi intermediate compartment, and included a particularly
188 high expression of the cAMP-response element binding transcription factor, *CREB3L1*(Sampieri et al.,
189 2019) implicated in cell differentiation and inhibition of cell proliferation, with concomitant high expression
190 of ER chaperones(Clark et al., 2002) calreticulin (*CALR*), calnexin (*CANX*)*;* transport *factors:* golgin
191 *(GOLGB1)* and folate receptor *FOLR1.* Platelet alpha granule proteins *(F5, VWF, TGFβ, MMP14),* several

FIGURE 2


A
CTRL
ET
RUX
No RUX
JAK2 Pos
JAK2 Neg
PV
MF
PC2:15%Variance
PC1:33%Variance
B
CREB3L1
TBL1X
COL4A5
IFITM3
MYBL2
IFITM2
FOLR1
NOMO1
ANXA3
RPIA
CALR
CANX
ECEL1
ET vs CTRL
MINDY4
STAC
TXN2
AP1B1
CLN6
CFH
GDF2
SRP9
AGPAT1.3
PV vs CTRL
RHAG
PLCL1
GDF15
IL1RAP
KLF11
ZNF117
FADD
NLRC5
MF vs CTRL
C
ET
PV
MF
120
459
813
1,504
40
1,298
3,219
D
MF
4917
1144
1092
RUX-treated
E
RUX vs noRUX
TLR5
LCK
GNAS
GZMA
FFAR3
B2M
AMD1
CCL3
ITGAX
CLU
GNGT2
HLA-DMB
CASP1
ITGA1
SOCS3
ITGAM
DAPK3
IRF9
GTPBP2
PER3
PSAP
SUPT3H
IFIT1
XPNPEP3
CD55
CCT6B
NAA35
TNFRSF14
IL17RA
IFIT2
APP
IFNGR1
Log2 Fold Change
Up-regulated
Down-regulated
Insignificant
RUX suppressed

192 collagens (*COL10A1, COL18A1, COL6A3*), immune/inflammatory (*IFITM2/3/10, FCGR2A, TMEM179B*),
193 *Cathepsins (C/Z/D, MIF, PTGES2*) and proliferation mediator genes (*CDK1, CCNG1, BMP9/GDF2,*
194 *LAPTM4B, PSENEN*) also constituted the MPN shared set. Downregulated genes, on the other hand, were
195 predominantly within transcription factor complexes, and included opposing expression of *CREB1* (vis-à-
196 vis *CREB3L1* above), calcium-calmodulin protein kinases, *CAMK4* and β-catenin *CTNNB1*, together
197 pointing to dysregulated calcium (Ca^{2+}) homeostasis in the ER lumen.

198 Having defined differential gene expression signatures by MPN subtype, we then explored how platelet
199 gene expression profiles differed in patients by treatment, focusing for this work on the JAK1/JAK2
200 inhibitor, ruxolitinib (RUX, **Figure 2E**). DGEA on the platelet transcriptome in RUX-treated patients revealed
201 a significant contrast from that of the MF or MPN platelet transcriptome with a notable shift toward down-
202 regulation (**Figures 2E vs 2B** bottom panel; ratio of downregulated:upregulated genes at 2053:458 in RUX-
203 treated patients versus 2379:3682 in MF at FDR <0.05). Over a five-fold decrease in gene expression was
204 noted in a G-protein coupled receptor, *GPR41* (also called free fatty acid receptor, *FFAR3*), followed by
205 over 500 genes with *at-least* two-fold reduction in expression in the RUX-treated cohort. These included
206 genes downregulated within classically defined membrane compartments, such as the immunological
207 synapse (*e.g. GZMA/B, LCK* proto-oncogene, and *ZAP70* with previously unidentified or emerging
208 roles(Campbell et al., 2018) of expression in platelets), and protein complexes within GTPase (*e.g. GNAS,*
209 *GNGT2, GNG8/11*), MHC *(B2M*, *HLA-DMB, HLA-DOA/DOB)* or cell adhesion contexts (*ITGA1, ITGAX/L,*
210 *ITGB7*).

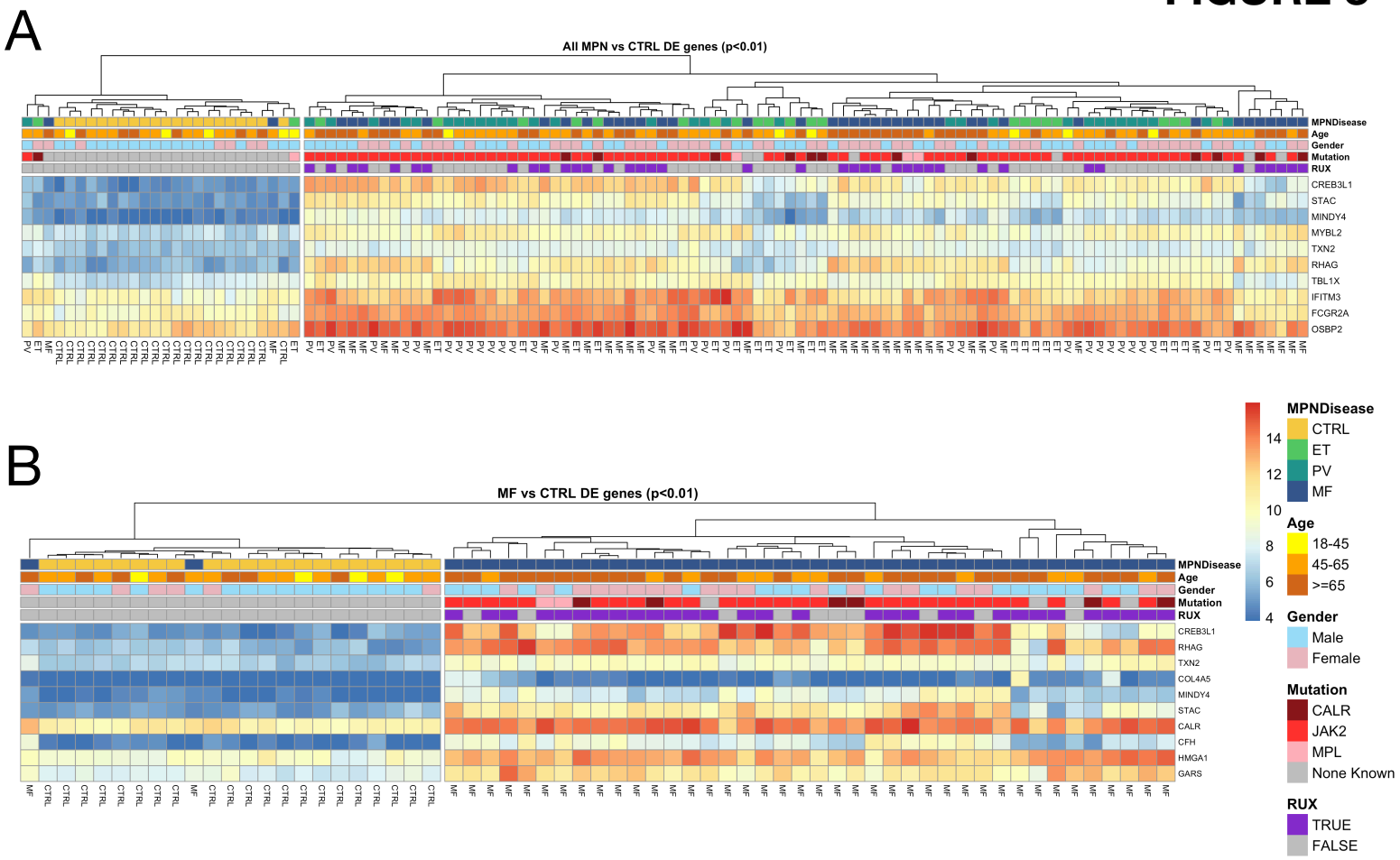
211 In addition to confirming previous observations (Arcaini and Cazzola, 2018; Kleppe et al., 2018; Spivak,
212 2017; Vainchenker et al., 2018) on the anti-inflammatory and immunosuppressive effects of *JAK* inhibition
213 by RUX (e.g. downregulation in our RUX-treated cohort of *CCL3, IL17RA, TLR5, TLR7, TNFRSF13/14*, and
214 *CASP1*), we identified new gene clusters responsive to RUX in the inhibition of type I interferon (e.g. *IFIT1,*
215 *IFIT2, IRF9*), interferon-gamma (e.g. *IFNGR1, SOCS3, DAPK3, CXCL16*), platelet degranulation (e.g. *APP,*
216 *CLU, GTPBP2, PF4, PDGFB, TUB4A, TTN*) and amine metabolic processes (e.g. *AMD1, ITGAM, PAOX,*
217 *HNMT*).

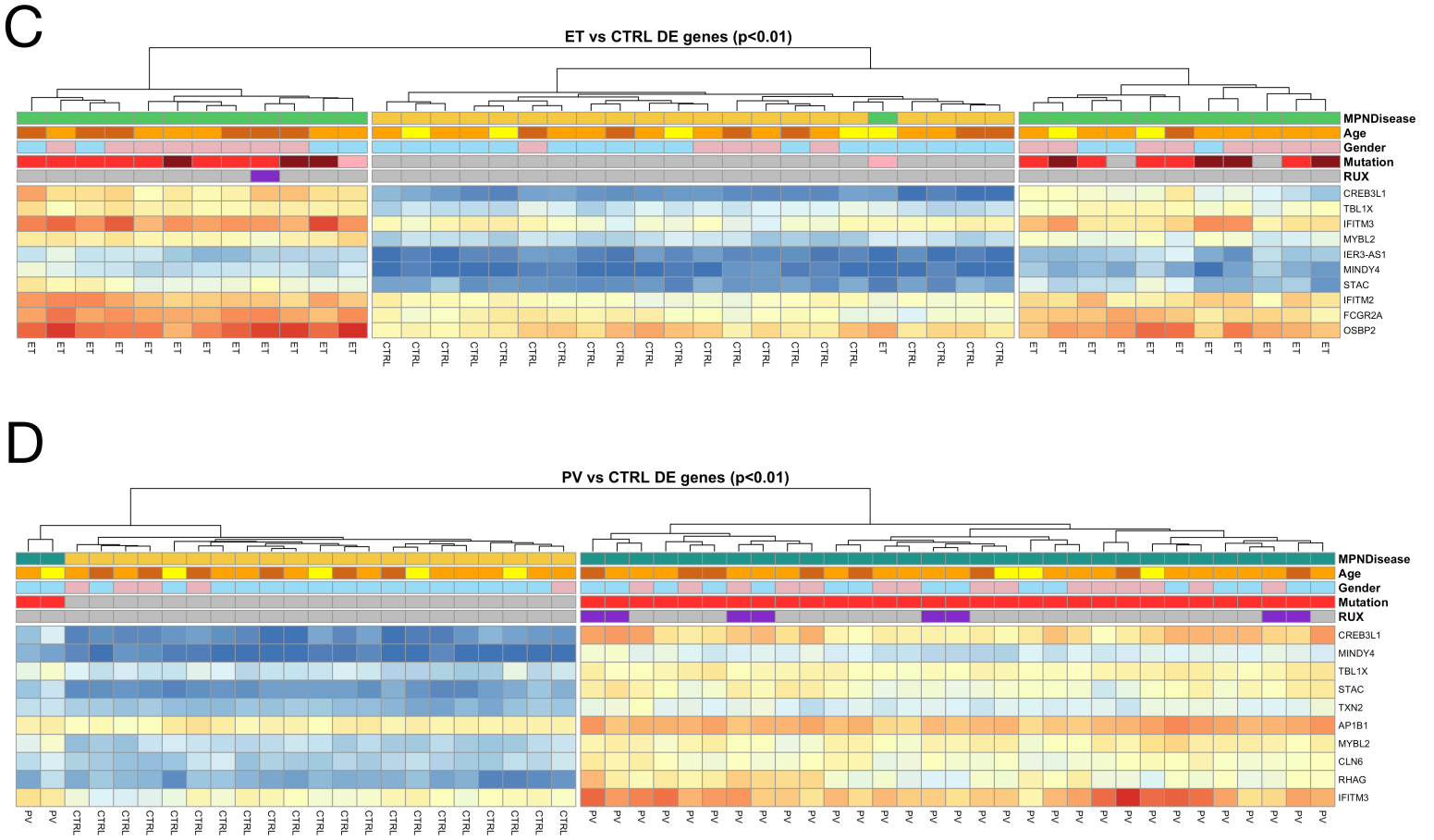
218 A direct comparison restricted to only differentially expressed genes (FDR<0.05) in *RUX-treated vs not* with
219 *MF vs healthy controls* revealed just 18% overlap (**Figure 2D**). Also, contrasting the total differentially
220 expressed genes (FDR<0.05) between the two groups (**Figure 2D**) reveal that the differential in the RUX-
221 treated cohort is roughly one-third of that in the MF disease cohort, reflecting the extent of the impact of
222 treatment by ruxolitinib relative to the substantive disease burden in MF that is beyond RUX treatment.
223 Focusing further on the directionality of the changes observed, we found just 39 genes that were increased
224 in MF and suppressed in the RUX-treated cohort (**Figure 2E**, colored green); and 15 vice versa (**Table**
225 **S2A-B**). Despite the small overlap, we capitalized on the converging genes to better define molecular and
226 physiological pathways underlying the effect of RUX in MPNs. The 39 RUX-down-regulated genes followed
227 expected mapping to immunosuppression through integrin and interferon-mediated signaling pathways
228 and indicated platelet crosstalk with other classical immune cells (granulocytes, neutrophils and T-cells
229 through genes such as *CD55, PSAP and TYROBP*). Interestingly, the 15 genes that were upregulated with
230 RUX in MF mapped to previously undescribed effect of the drug in improving protein maturation (*e.g.*
231 *NAA35, XPNPEP3*), toxin transport (*e.g. CCT6B*), chaperone-mediated protein complex assembly (*e.g.*
232 *CCT6B, SUPT3H*), and circadian regulation (*PER3*).

233 **Graded differential expression by MPN phenotype and driver mutation status**

234 Unsupervised hierarchical clustering was used to more precisely define the nature of MPN platelet
235 transcriptome variability from controls, as well as between and within MPN subtypes. **Figure 3** reveals
236 spectrum of expression in the MPN platelet transcriptomic profile using just the top 10 highly significant
237 differentially expressed genes by disease status: (a) all MPN vs controls, (b) MF vs controls, and (c, and d)
238 ET- and PV- vs controls. As shown in **Figure 3A**, all MPN patients clustered into two distinct groups: a
239 larger group of 94 ET, PV and MF patients clustered independently from the 21 controls and showed a
240 varying gradation with respect to the top-10 gene expression by MPN subtype, whereas the other, a
241 smaller group of 5 ET, PV and MF patients was more homogeneous and overlapped with the controls.
242 Importantly, post-RNA-seq re-review of the MPN diagnosis of these 5 distinctly clustered outliers identified
243 them as 2 MF patients later annotated as autoimmune myelofibrosis (as opposed to classical MF) with
244 *JAK2* negative or triple negative mutational status, and the other 3 as *CALR*$^{+}$ ET, *MPL*$^{+}$ ET, and *JAK2*$^{+}$ PV.

FIGURE 3

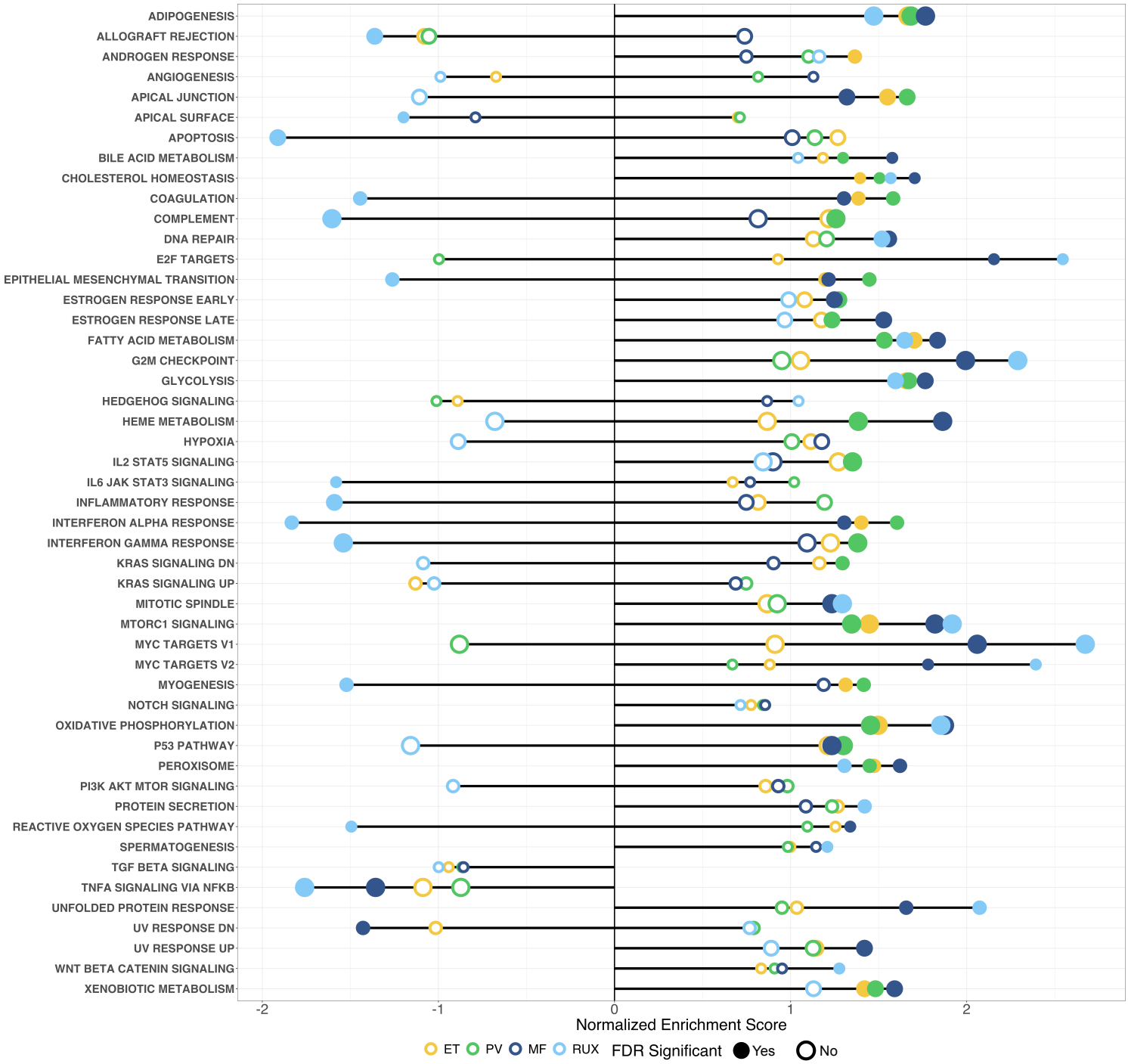

A
All MPN vs CTRL DE genes (p<0.01)
MPNDisease
Age
Gender
Mutation
RUX
CREB3L1
STAC
MINDY4
MYBL2
TXN2
RHAG
TBL1X
IFITM3
FCGR2A
OSBP2
B
MF vs CTRL DE genes (p<0.01)
MPNDisease
Age
Gender
Mutation
RUX
CREB3L1
RHAG
TXN2
COL4A5
MINDY4
STAC
CALR
CFH
HMGA1
GARS
14
12
10
8
6
4
MPNDisease
CTRL
ET
PV
MF
Age
18-45
45-65
>=65
Gender
Male
Female
Mutation
CALR
JAK2
MPL
None Known
RUX
TRUE
FALSE


C
ET vs CTRL DE genes (p<0.01)
MPNDisease
Age
Gender
Mutation
RUX
CREB3L1
TBL1X
IFITM3
MYBL2
IER3-AS1
MINDY4
STAC
IFITM2
FCGR2A
OSBP2
D
PV vs CTRL DE genes (p<0.01)
MPNDisease
Age
Gender
Mutation
RUX
CREB3L1
MINDY4
TBL1X
STAC
TXN2
AP1B1
MYBL2
CLN6
RHAG
IFITM3

245 In the larger cluster, while we observed a graded overlap in platelet RNA signatures between ET and PV, a
246 more distinct expression pattern characterized the more advanced population of MF patients (*PC1*
247 *correlates with MF disease risk by the Dynamic International Prognostic Scoring System (Passamonti et al.,*
248 *2010), DIPSS*, **Figure S2**). These, and our additional granular differential analyses by driver mutations
249 (sub-heatmaps not shown), highlight the importance of phenotype-modifying genes that are independent
250 of JAK2/*CALR*/*MPL* mutation status.
251
252 Untangling other mechanisms beyond the few genes that are recurrently mutated is critical for defining
253 subtype-specific risk and for identifying molecular pathways for targeted therapy. Accordingly, we sought
254 to refine the molecular classification of MPN by associating platelet gene expression profiles with the
255 corresponding subtype, and yielded a core set of 10 highly-significant preferentially expressed genes for
256 each: (i) MF (**Figure 3B**), defined by high mRNA expression of proteostasis-associated *CREB3L1* and
257 *CALR,* and megakaryocyte-erythroid differentiation stage-associated *RHAG* (Caparros-Perez et al., 2017;
258 Zeddies et al., 2014) ii) ET (**Figure 3C**), marked by comparatively high expression of interferon-related
259 genes *IFITM2* and *IFITM3*, and proliferation-associated *STAT5* target *OSBP2*; and iii) PV (**Figure 3D**),
260 marked by overlapping signatures with ET in inflammation-associated *IFITM3* or *TBL1X*, and the B-myb
261 promoter, *MYBL2*; with MF in the maturation-associated *RHAG* at variable expression; and with both ET
262 and MF in the high expression of *CREB3L1*, and cell survival-associated *MINDY4* and *STAC*.
263
264 **Altered immune, metabolic, and proteostatic pathways underlie each MPN phenotype**
265 Our analysis of MPN platelet RNA-seq enabled identification of altered MPN pathways that might be
266 amenable to drug therapy. To understand the biological significance of transcriptional changes, we
267 performed pathway-enrichment analysis and identified signaling pathways that are differentially activated
268 between MPN subtypes (**Figure 4**). Gene set enrichment analysis (GSEA, see Methods) of Hallmark gene
269 sets found that MPN (stratified by subtypes, ET, PV and MF) induces genes related to pathways with
270 known immune modulatory functions (**Figure 4**, notably interferon alpha response in ET, PV and MF, and
271 IL2 STAT5 signaling, and interferon gamma response specifically enriched in PV). Moreover, among the
272 most enriched gene sets, MPN pathology induces robust activation of oxidative phosphorylation

FIGURE 4

A


ADIPOGENESIS
ALLOGRAFT REJECTION
ANDROGEN RESPONSE
ANGIOGENESIS
APICAL JUNCTION
APICAL SURFACE
APOPTOSIS
BILE ACID METABOLISM
CHOLESTEROL HOMEOSTASIS
COAGULATION
COMPLEMENT
DNA REPAIR
E2F TARGETS
EPITHELIAL MESENCHYMAL TRANSITION
ESTROGEN RESPONSE EARLY
ESTROGEN RESPONSE LATE
FATTY ACID METABOLISM
G2M CHECKPOINT
GLYCOLYSIS
HEDGEHOG SIGNALING
HEME METABOLISM
HYPOXIA
IL2 STAT5 SIGNALING
IL6 JAK STAT3 SIGNALING
INFLAMMATORY RESPONSE
INTERFERON ALPHA RESPONSE
INTERFERON GAMMA RESPONSE
KRAS SIGNALING DN
KRAS SIGNALING UP
MITOTIC SPINDLE
MTORC1 SIGNALING
MYC TARGETS V1
MYC TARGETS V2
MYOGENESIS
NOTCH SIGNALING
OXIDATIVE PHOSPHORYLATION
P53 PATHWAY
PEROXISOME
PI3K AKT MTOR SIGNALING
PROTEIN SECRETION
REACTIVE OXYGEN SPECIES PATHWAY
SPERMATOGENESIS
TGF BETA SIGNALING
TNFA SIGNALING VIA NFKB
UNFOLDED PROTEIN RESPONSE
UV RESPONSE DN
UV RESPONSE UP
WNT BETA CATENIN SIGNALING
XENOBIOTIC METABOLISM
-2
-1
0
1
2
Normalized Enrichment Score
ET
PV
MF
RUX
FDR Significant
Yes
No

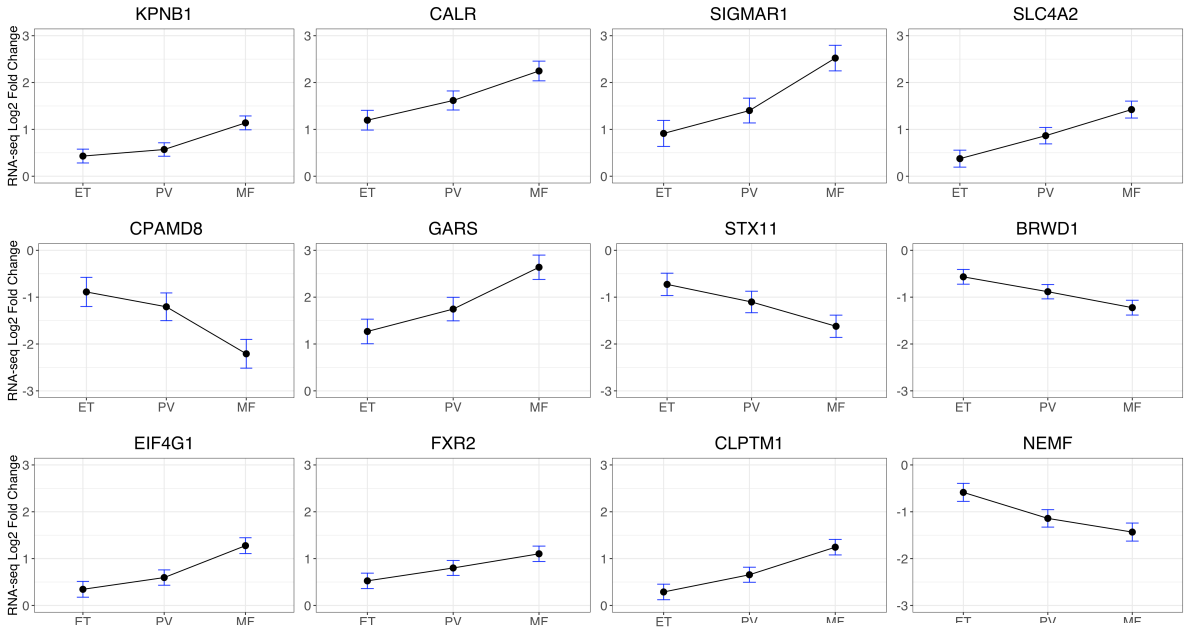
273 (OXPHOS) and mTORC1 signaling pathways, with increasing enrichment and significance by MPN subtype
274 (FDR <0.0001 in MF). Pathways of reactive oxygen species (ROS) production paralleled activation of
275 mTORC1 in MF. Other complementary metabolic pathways paralleled OXPHOS activation, with significant
276 enrichment of bile and fatty acid metabolism, cholesterol homeostasis and adipogenesis, most
277 pronounced in MF and variably expressed in ET and PV. Coagulation- and complement-associated gene
278 sets were expectedly enriched across ET, PV and MF. What is particularly noteworthy is that in MF, cell
279 cycle progression and proliferation pathways reveal significant enrichment (FDR <0.001) around c-MYC
280 and E2F targets, and G2M checkpoint pathways; and unfolded protein response emerges as a key factor,
281 likely attributable to ER stress (see *CREB3L1, CALR* overexpression in **Figures 2,3**). Representative GSEA
282 profiles are shown in **Figure 4** and the full list of enriched pathways and gene sets detailed in **Table S3A-**
283 **C**. The MPN pathways exhibiting significant transcriptional regulation by GSEA are consistent with our
284 observations at the individual gene level for upregulated and downregulated transcripts, specifically those
285 upregulated in MF. Taken together, these data demonstrate that in addition to immune factors such as
286 type I/II interferons and dysregulation of interleukin-dependent inflammatory responses, which have been
287 linked to MPNs, platelet transcriptional signatures of proliferation, metabolic, and proteostasis signaling
288 are a feature of MPN pathogenesis (**Figure S3** captures the relative enrichment by subtype of MPN
289 molecular pathway categories as a concept model of MPN progression).
290

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

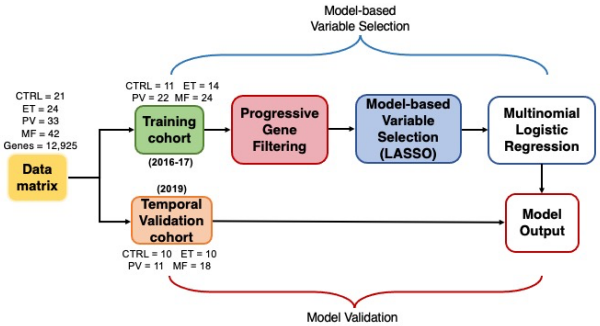
292 Current knowledge of MPN genetic, cytogenetic or epigenetic abnormalities are limited(Spivak, 2017;
293 Spivak et al., 2014) in their ability to enable prediction of disease progression or evolution of a given
294 patient, from ET/PV phenotype to MF. In order to investigate the potential of platelet transcriptomic
295 parameters to enable MF prediction, we constructed LASSO penalized (Tibshirani, 1996) regression
296 classifiers (machine learning R package *glmnet)* to discriminate MPN subtypes from each other, and from
297 healthy controls (**Figure 5A-E**). We apply two rigorous (Moons et al., 2012) validation conditions **(Figure**
298 **5B-C):** i) training and independent temporal validation (**Figure 5B**) leveraging our two cohort design
299 (**Figure 1** cohort 1 n=71 and cohort 2 n=49 noting the two different time periods and populations reflecting

FIGURE 5

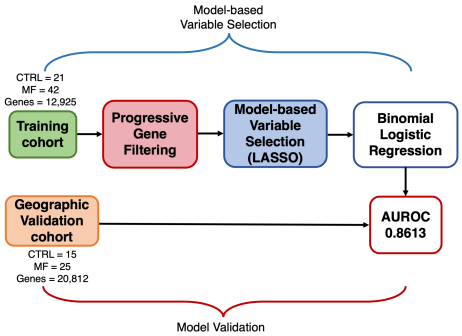
A



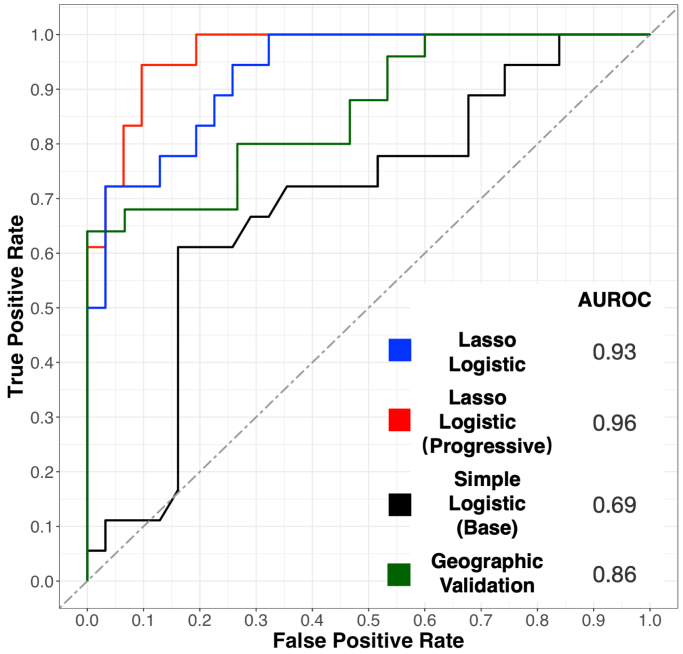
B



C



D



E

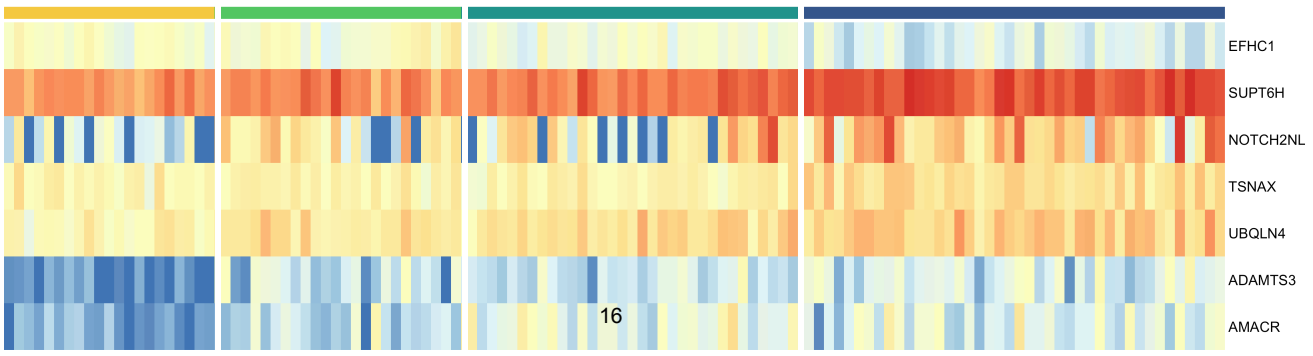
Entire Platelet Transcriptome Lasso		Actual			
		CTRL	ET	PV	MF
Input of 12,925 genes					
Predicted	CTRL	20	2	1	0
	ET	0	12	5	5
	PV	0	10	20	8
	MF	1	0	7	29
	Total	21	24	33	42
	Correct	20	12	20	29
Sensitivity					69.1
Specificity		89.7			

12 Most Progressive Genes		Actual			
		CTRL	ET	PV	MF
Input of 12 genes					
Predicted	CTRL	21	0	0	0
	ET	0	16	6	1
	PV	0	7	23	7
	MF	0	1	4	34
	Total	21	24	33	42
	Correct	21	16	23	34
Sensitivity					81.0
Specificity		93.6			

Progressive Transcriptome Lasso		Actual			
		CTRL	ET	PV	MF
Input of 4,458 genes					
Predicted	CTRL	19	5	1	1
	ET	2	5	4	3
	PV	0	14	20	6
	MF	0	0	8	32
	Total	21	24	33	42
	Correct	19	5	20	32
Sensitivity					76.2
Specificity		89.7			

Base Model w/o Platelet Transcriptome		Actual			
		CTRL	ET	PV	MF
No gene expression					
Predicted	CTRL	13	4	0	6
	ET	1	4	1	3
	PV	0	7	16	15
	MF	7	9	16	18
	Total	21	24	33	42
	Correct	13	4	16	18
Sensitivity					42.9
Specificity		59.0			

F



300 real-life heterogeneity) and ii) geographic external (**Figure 5C**) on an independently published (Guo et al.,
301 2020) dataset (n=25 MF and n=15 healthy donors). First, in our temporal internal validation, three types of
302 models were constructed based on i) baseline (no transcriptome, but driver mutation, age and gender, as
303 reference variables available for all patients); ii) full platelet transcriptome integrated with the above
304 baseline; and iii) subset platelet transcriptome that exhibits progressive differentiation from controls/ET/PV
305 to MF (>4000 genes, top 12 shown in **Figure 5A**) integrated with the above baseline (progressive subset is
306 selected unbiased as part of the Lasso learning process). Comparison of the classification potential among
307 the three models demonstrated that the progressive platelet transcriptome model substantially
308 outperformed the baseline model (**Figure 5D** black vs red curves) and was slightly better than the full
309 transcriptome model (**Figure 5D** blue vs red curves) in the classification of ET, PV and MF. Predicted
310 probabilities for all three models are shown in **Table S4A-C** and identical conclusions were drawn with
311 confusion matrices as well (**Figure 5E**). Lasso logistic regression classifiers to predict MF with each of the
312 models under the first temporal two-cohort training-validation split - baseline, full, and progressive
313 transcriptome achieved area under the receiver-operating characteristic curve (AUROC) of 0.69, 0.93, and
314 0.96 respectively. Outperformance of the progressive transcriptome model was validated in our
315 subsequent independent external geographic cohort too (**Figure 5D** brown curve) at an AUROC of 0.86.
316 Recurrent top 7 genes from our progressive transcriptome Lasso are visualized as a heatmap in **Figure 5F**
317 clearly capturing the incremental gradient in gene expression between controls, ET, PV and MF. A key
318 aspect of the layered Lasso modeling demonstrated here is our use of an approach that can be developed
319 in future work to incorporate additional predictors to MF (e.g. *JAK2* V617F allele burden (Vannucchi et al.,
320 2008) or other genetic variants (Tefferi et al., 2016) beyond the driver mutations).
321

322 **Discussion**

323 Here, we present a comprehensive catalog of the platelet transcriptome in chronic progressive MPN with
324 immediate relevance to defining subtype-specific molecular differences and predicting the advanced
325 phenotype, myelofibrosis. Recent data(Williams et al., 2020) identify the timing of MPN driver mutation
326 acquisition to be very early in life, even before birth, with life-long clonal expansion and evolution. These

327 new findings highlight the importance of early progression biomarkers and the substantial opportunity for
328 early detection and intervention strategies in these disorders.

329 Our analyses confirm and extend many important observations made previously either in vitro (LaFave and
330 Levine, 2016; Merlinsky et al., 2019; Osorio et al., 2016), in other transcriptome or microarray analyses
331 (Gangaraju et al., 2020; Guo et al., 2019; Rampal et al., 2014; Rontauroli et al., 2021; Skov et al., 2012;
332 Wong et al., 2019) including our own early work(Krishnan et al., 2017), or using animal models (Dunbar et
333 al., 2017; Matsuura et al., 2020; Mullally et al., 2013; Wen et al., 2015). By highlighting intersecting
334 mechanisms in transcription across MPN, and by annotating MPN subtype-specific gene signatures, this
335 dataset facilitates predictive machine learning algorithms, that aid in MPN classification and potential
336 prognostication.

337 The platelet transcriptome is significantly reprogrammed in the MPN setting, with a wealth of transcript
338 associations that may be missed in using conventional tissue sources such as serum, plasma, whole blood
339 or bulk bone marrow. While previous bulk RNA-seq studies on MPNs by us and others analyzed far fewer
340 samples (Guo et al., 2019; Krishnan et al., 2017; Skov et al., 2011; Skov et al., 2012), select MPN subtypes
341 (Gangaraju et al., 2020; Guo et al., 2019), or non-specific source tissue (Schischlik et al., 2019; Skov et al.,
342 2011; Skov et al., 2012) that may be underpowered (Cummings et al., 2017) for candidate genes, we here
343 analyzed, by next generation RNA sequencing, 120 purified platelet samples from healthy controls and all
344 three subtypes; and identified clinically interpretable transcriptomic signatures for each of the three
345 subtypes. Each subtype showed both overlapping and progressively divergent transcriptional pathways,
346 suggesting both a shared signature across all MPN, and unique biological trajectories. Pathway-
347 enrichment analyses confirmed the existence of a shared inflammatory milieu (Barbui et al., 2011; Geyer et
348 al., 2015; Hasselbalch and Bjorn, 2015; Koschmieder and Chatain, 2020; Marin Oyarzún and Heller, 2019;
349 Mughal et al., 2019; Skov et al., 2011) among MPN. We also confirmed that the JAK1/JAK2 inhibitor
350 ruxolitinib was associated with inhibition of inflammatory as well as interferon-mediated signaling
351 pathways. Additional previously undescribed insights into the mechanisms of action of RUX in MF included
352 genes implicated in protein maturation, chaperone-mediated protein complex assembly, and circadian

353 rhythm. These and other gene signatures and pathways identified may help guide candidate drugs to be
354 used alone or in combination with RUX for the treatment of MPNs. Whether MPN oncogenic driver
355 mutations increase inflammation or mutations are acquired in response to inflammatory stimuli is unclear
356 from this work and remains an active area of investigation.(Curto-Garcia et al., 2020; Hasselbalch and
357 Bjorn, 2015; Koschmieder and Chatain, 2020; Lussana et al., 2017)

358 The 10 genes most significant (FDR < 0.001) of the commonly expressed genes across MPN indicated a
359 gradation in platelet gene expression, with overlapping signatures in ET and PV (*e.g. IFITM2, MYBL2)* and
360 a substantial difference with MF (*e.g. CREB3L1, CALR*) that was independent of driver mutation status or
361 treatment. Hence, while over 1500 genes were commonly differentially expressed across MPN, their
362 abundance and function could differ between subtypes. The nature of the separation of transcriptomic
363 clusters between ET, PV and MF suggest also that they represent diverse cell states along a continuous
364 spectrum of MPN, in line with the clinical overlap of these disorders.
365

366 Another observation relates to the association of the differential genes with signaling pathways: as
367 indicated above, all three MPN subtypes showed a positive enrichment in immune modulation pathways,
368 independent of mutational status. Whether this response reflects a causal effect of inflammation on bone
369 marrow biology remains to be elucidated. Indeed, the platelet transcriptomic signatures could also reflect
370 inter-cell interactions of platelets with other immune cells, including as transient aggregates with
371 neutrophils, granulocytes and dendritic cells. Nevertheless, observations that MPN transcriptomic
372 biomarkers correlated robustly with immune factors such as type I/II interferons and dysregulation of
373 interleukin-dependent inflammatory responses across ET, PV and MF suggest opportunities for use of
374 these and other subtype-specific genes as biomarkers for prognosis as well as design of therapies and
375 prediction of response.
376

377 Our data closely overlaps with recent MPN platelet studies: thrombo-inflammatory signatures in PV from
378 Gangaraju and Prchal et al (Gangaraju et al., 2020) (*BCL2, CXCL1, MMP7, PGLYRP1, CKB, BSG, CFL1*
379 and more) or fibrosis-associated signatures in MF from Guo and Erber et al. (*CCND1, H2AFX* and *CEP55*).

380 Most notably in our data on MF, high expression of ER stress and unfolded protein response (UPR)
381 biomarkers *(e.g. CREB3L1, CALR)* associated with impaired proteostasis signaling; and emerged as a key
382 feature of MPN pathobiology. Indeed, recently published work from distinct research groups (Liu et al.,
383 2020; Osorio et al., 2016)(LaFave and Levine, 2016) highlight protein quality control in ER-associated
384 degradation[57] and proteostasis (Osorio et al., 2016)(LaFave and Levine, 2016) deregulation as a primary
385 effector of myeloid transformation highlighting the importance of protein homeostasis for normal
386 hematopoiesis. These findings too are in line with reports (Kaushik and Cuervo, 2015) implicating chronic
387 ER stress, malfunctioning protein quality control, and loss of proteostasis as aggravating factors in age-
388 related disorders.
389
390 Most importantly, platelet gene expression profiling in MPN offers directions for prediction of
391 myelofibrosis. Applying machine-learning algorithms of LASSO penalized regression under two
392 stringent(Moons et al., 2012) conditions of temporal (using our two cohort design) and geographic
393 (independently published dataset(Guo et al., 2020)) validation, we uniquely discriminate MPN subtypes
394 from each other, and healthy controls using three model types and predict MF at high accuracy. The
395 highest performing model used a set of progressively differentiated MPN genes at an area under the (ROC)
396 curve of 0.96 (temporal internal) and 0.86 (geographic external); and rendered a core signature of <10
397 candidate markers as top predictors of disease progression. It will be of interest to determine what
398 machine-learning algorithms based on a defined platelet gene expression classifier on potential new MPN
399 datasets can be used to more precisely predict the probability and/or timing of an individual's risk of
400 progression from ET/PV to secondary MF.
401
402 In conclusion, using platelet transcriptome profiling, we observed dynamic shifts in MPN immune
403 inflammatory profile and preferential expression of interferon-, proliferation-, and proteostasis-associated
404 genes as a progressive gradient across the three MPN subtypes. Our findings highlight that MPN
405 progression may be influenced by defects in protein homeostasis (impaired protein folding and an
406 accumulation of misfolded proteins within the endoplasmic reticulum) and an abnormal integrated stress
407 response – consistent with recent studies[57](Liu et al., 2020; Osorio et al., 2016)(LaFave and Levine, 2016)

408 indicating dysregulated proteostasis as a primary effector of myeloid transformation. While this particular
409 work has been focused on the overarching progressive platelet transcriptome across MPNs, these data
410 open an important avenue for utilizing platelet RNA signatures to better understand specific MPN
411 complications such as the risk of thrombosis, bleeding, and transformation to AML. Altogether, this study
412 demonstrated in chronic MPNs provides a comprehensive framework for exploiting the platelet
413 transcriptome and may inform future studies toward mechanistic understanding and therapeutic
414 development not only in MPNs, but also more broadly as potential progression markers in other age-
415 related disorders.
416
417 **Limitations of the study**
418 There are several limitations to our study. First, our data are not longitudinal by design but rather the
419 closest practical alternative of cross-sectional snapshots of all three MPN subtypes with the goal of
420 achieving a
421 well-powered dataset in these chronic disorders. In this regard, the progressive or progression terminology
422 used here refer strictly to trends in gene expression and do not imply study of longitudinal clinical
423 progression. Therefore, subsidiary longitudinal evaluation of the disease as well as treatment markers
424 identified here is warranted. Second, our focus for this study has been on the platelet transcriptome alone.
425 Future investigations focused on ascertaining overlap between our platelet-derived molecular alterations
426 with those of other cell types, specifically, parent megakaryocytes, CD34+ cells, other immune cells and
427 even whole blood may identify additional functional aspects of bone marrow pathology. Third, we
428 recognize that our choice of ribosomal RNA depletion to home in on platelet mRNA signatures leaves out
429 additional diversity in the platelet RNA repertoire (and will be important future work). Lastly, in our Lasso
430 predictive modeling, we demonstrate two rigorous approaches of validation (two-cohort temporal internal
431 and geographically independent external) and identify a core signature toward MPN risk stratification or
432 early detection of progression. Yet, substantive future biological and computational validations are needed
433 in order to advance our findings toward clinical decision making or personalized medicine.
434
435

436 **Methods**

437 Ethical Approval

438 All MPN peripheral blood samples were obtained under written informed patient consent and were fully
439 anonymized. Study approval was provided by the Stanford University Institutional Review Board.
440 All relevant ethical regulations were followed.

441 Subjects and Specimen Collection

442 We collected blood from ninety-five MPN patients enrolled in the Stanford University and Stanford Cancer
443 Institute Hematology Tissue Bank from December 2016- December 2019 after written informed consent
444 from patients or their legally authorized representatives (Stanford IRB approval #18329). Eligibility criteria
445 included age ≥18 years and Stanford MPN clinic diagnosis of essential thrombocythemia, polycythemia
446 vera or myelofibrosis (defined using the consensus criteria at the time of this study). We use the term
447 'myelofibrosis' to encompass both primary myelofibrosis and myelofibrosis that evolved from essential
448 thrombocythemia or polycythemia vera. Electronic medical records review of all subjects was performed
449 by the clinical consultants (J.G. and L.F.), study data manager (C.P.), and the study principal investigator
450 (A.K.). For controls, blood was collected from twenty-one asymptomatic adult donors selected at random
451 from the Stanford Blood Center. All donors were asked for consent for genetic research. For both MPN
452 patients and healthy controls, blood was collected into acid citrate-dextrose (ACD, 3.2%) sterile yellow-top
453 tubes (Becton, Dickinson and Co.) and platelets were isolated by established(Amisten, 2012; Campbell et
454 al., 2018; Middleton et al., 2019; Rowley et al., 2011) purification protocols. Blood was processed within 4h
455 of collection for all samples. The time from whole blood collection to platelet isolation was similar between
456 healthy donors and MPN patients.

457

458 Platelet Isolation

459 Human platelets were isolated and leuko-depleted using established methods ((Amisten, 2012; Campbell
460 et al., 2018; Rowley et al., 2011) with excellent reproducibility(Campbell et al., 2018; Davizon-Castillo et al.,
461 2020; Manne et al., 2020; Middleton et al., 2019; Rondina et al., 2013) resulting in a highly purified
462 population of fewer than 3 leukocytes/10^7 platelets (>99.9% purity) as counted by hemocytometer. Briefly,
463 the ACD-tube whole blood was first centrifuged at 200xg for 20min at room temperature (RT). The platelet-
464 rich plasma (PRP) was removed and Prostaglandin E1 was added to the PRP to prevent exogenous
465 platelet activation. The PRP was then centrifuged at 1000xg for 20min at RT. The platelet pellet was re-
466 suspended in warmed (37 deg C) PIPES saline glucose (PSG). Leukocytes were depleted using CD45+
467 magnetic beads (Miltenyi Biotec). Isolated platelets were lysed in Trizol for RNA extraction. Post RNA-seq
468 analysis of an index leukocyte transcript (*PTPRC; CD45*) confirmed that the samples were highly purified
469 platelet preparations (subsequent bioinformatic analyses also adjusted for PTPRC expression for absolute
470 removal of any CD45 expression in our analyses). Two reference markers of platelet activation, *P-selectin*

471 *(SELP)* and *Glycoprotein IIbIIIA (CD41/ITGA2B)* were expectedly higher in all MPN than healthy controls
472 but were not statistically significantly different between MPN subtypes; indicating that any expression
473 difference was not due to experimental artefacts. In addition, we know from prior work (J.R.) that
474 regardless of activation status, RNA-seq reliably estimates mRNA expression patterns in platelets. We also
475 know from rigorous prior work(Gnatenko et al., 2003; Raghavachari et al., 2007; Weyrich and Zimmerman,
476 2003) that several abundant platelet mRNAs are well-known leukocyte or red cell transcripts; and do not
477 immediately imply contamination by these classes but rather that platelets express gene products that are
478 also present in other cell lineages. Sixty-five of our top 100 abundant platelet transcripts matched exactly
479 with those of the top 100 abundant genes from the three previous studies cited(Gnatenko et al., 2003;
480 Rowley et al., 2011); and a composite pathway analysis with the top 100 abundant genes from this as well
481 as the previous studies matched identically.
482
483 Next Generation RNA Sequencing
484 For next generation RNA-sequencing (RNA-seq), $1x10^9$ isolated platelets were lysed in Trizol and then
485 DNAse treated. Total RNA was isolated, and an Agilent bio-analyzer was used to quantify the amount and
486 quality.
487 The RNA yield was estimated by measuring absorbance at 260 nm on the Nanodrop 2000 (Thermo Fisher),
488 and RNA purity was determined by calculating 260/280 nm and 260/230 nm absorbance ratios. RNA
489 integrity was assessed on the Agilent Bioanalyzer using the RNA 6000 Nano Chip kit (Agilent
490 Technologies). An RNA integrity number (RIN) was assigned to each sample by the accompanying
491 Bioanalyzer Expert 2100 software.
492 To control for variable RNA quality, RNA sequencing was only performed for samples with a RIN score of 7
493 or higher. RNA-seq libraries were constructed with removal of ribosomal RNA using the KAPA Stranded
494 RNA-Seq kit with RiboErase (Roche). The RNA extraction and library preparation were performed by the
495 same technician to minimize confounding effects. cDNA libraries were constructed following the Illumina
496 TrueSeq Stranded mRNA Sample Prep Kit protocol and dual indexed. The average size and quality of each
497 cDNA library were determined by the Agilent Bioanalyzer 2100, and concentrations were determined by
498 Qubit for proper dilutions and balancing across samples. Twelve pooled samples with individual indices
499 were run on an Illumina HiSeq 4000 (Patterned flow cell with Hiseq4000 SBS v3 chemistry) as 2 X 75bp
500 paired end sequencing with a coverage goal of 40M reads/sample. Output BCL files were FASTQ-
501 converted and demultiplexed.
502
503 Platelet Transcriptome Analysis
504 Picard, Samtools, and other metrics were used to evaluate data quality. Processed reads were aligned
505 against the reference human transcriptome GRCh37/hg19 using RSEM(Li and Dewey, 2011) and
506 bowtie2(Langmead and Salzberg, 2012), and expression at gene level determined by calculating raw gene
507 count. Only genes that passed expression threshold were used; genes were considered expressed if, in all

508 samples, they had at least 10 counts (genes with low counts are automatically filtered by built-in functions
509 in DeSeq2, see below). A total of 12,924 genes were considered expressed. Gene expression data was
510 library-size-corrected, variance-stabilized, and log2-transformed using the R package DESeq2(Love et al.,
511 2014).
512 We refer to this version of the data as "raw data" as it is not corrected for any confounders of gene
513 expression variability. DESeq2 was used to call differential expression while adjusting for patient age,
514 gender and treatment as confounding variables and controlling for multiple comparisons using the
515 Benjamini-Hochberg defined false discovery rate (FDR). Significant variance in expressed transcripts were
516 pre-specified as those transcripts with an FDR <0.05 and a log2 fold change ≥ 0.5 in MPN, as compared
517 to healthy controls (the entire differential transcriptome was applied in the instances of downstream Gene
518 Set Enrichment Analysis and the Lasso prediction modeling)
519
520 Statistical analysis
521 Continuous data were summarized as medians and IQRs and categorical data are presented as
522 frequencies and percentages. To compare differences in clinical variables between healthy controls and
523 MPN subtypes (ET, PV and MF), we use box and whisker plots and conduct pairwise Wilcoxon signed
524 ranked tests. For unsupervised clustering and visualization, we performed principal component
525 analyses (identifying MPN subtypes by color, treatment by filled or open circles, and *JAK2* mutation status
526 by shape) using built-in functions of the DeSeq2 R package. We generated a heatmap of all of the top
527 highly significant genes (FDR < 0.01) using the pheatmap R package and its built-in functions for
528 hierarchical cluster analysis on the sample-to-sample Euclidean distance matrix of the expression data. All
529 analyses were performed using the R studio interface.
530
531 Pathway/Gene set enrichment analysis for differentially expressed (DE) genes.
532 Gene set enrichment analysis (GSEA)(Subramanian et al., 2005) was performed on the entire DE gene set
533 for each MPN subtype, using the Cancer Hallmarks gene sets from MSigDB(Liberzon et al., 2015). The
534 'GSEA Pre-ranked' function was used with a metric score that combines fold change and adjusted p-value
535 together for improved gene ranking. We used default settings with 10,000 gene set permutations to
536 generate *p* and *q* values, and compared MPN subtypes in the overall cohort, and the ruxolitinib-treated
537 subgroup and the ruxolitinib-naive subgroup separately. In these analyses, to allow for a broad
538 comparison, we assessed all transcripts that were differentially expressed according to FDR/adjusted p <
539 0.25 as recommended by the authors of GSEA (Subramanian et al., 2005).
540
541 Predictive model generation and external validation
542 At the conception of this study late 2016, and our early work(Krishnan et al., 2017), we did not identify any
543 publicly available RNA sequencing data on MPN platelets. This prompted our specific two-cohort design

544 for the express purpose of temporal external validation as an essential step in rigorous prediction
545 modeling(Moons et al., 2012). A subsequent independent publication(Guo et al., 2020) facilitated a
546 geographically independent and unbiased external validation of our model.

547 We used Lasso penalized regression(Tibshirani, 1996) for our model to predict MF from the either healthy
548 controls, ET or PV. Among a variety of statistical machine learning algorithms that have been used in
549 prediction modeling, Lasso is favored for its flexibility and simplicity; and its ability to identify the least set
550 of significant factors from high dimensional data. We evaluated patient features (age, gender and mutation
551 status) and platelet transcriptomic features. Normalized gene counts data were split into training (used for
552 constructing multinomial logistic models) and validation (used for model evaluation and generalization)
553 cohorts. Separately, we assessed the progressive and monotonic upward or downward trend in gene
554 expression, we applied the Mann-Kendall trend test (multiple comparison adjusted with the Benjamini-
555 Hochberg method) to normalized gene counts and identified statistically significant progressive genes
556 across all three MPN subtypes.
557
558 Three multinomial logistic models were constructed: first, with Lasso selected predictors from all genes,
559 second, with Lasso selected predictors from progressive genes and third, a baseline model using age,
560 gender and mutation status (*JAK2 and CALR*) as predictors. Model outputs correspond to probabilities of
561 having a CTRL, ET, PV or MF phenotype (sum of these four probability values totaling 1). Potential
562 interpretation of these probabilities includes MPN risk assessment, *e.g.* a patient with higher probabilities
563 of PV and MF would indicate higher risk than one with higher probabilities of CTRL or ET.
564
565 ROC curves and confusion matrices were used to evaluate the different prediction models and
566 discriminate outcomes. ROC curves demonstrate the trade-off between true positive and false positive
567 rates, ideal being high true positive rate (sensitivity) and low false positive rate (specificity) the area under
568 the curve (AUROC) as close to 1 as possible. True positive rate (TPR) is defined as correctly predicting an
569 MF patient as MF; and false positive rate (FPR) as falsely predicting a non-MF patient as MF. The
570 confusion matrices were constructed by predicting each patient as the MPN subtype with highest
571 predicted probability.
572
573
574
575
576
577

578 **Authorship Contributions** A. Krishnan, J. Gotlib and J. Zehnder conceived of the overall study. A.
579 Krishnan designed the study and secured funding. L.F., C.P., and J.G. provided samples and clinical
580 annotation and reviewed the clinical data. A.K. designed the experimental plan with input from J.R., H.M.,
581 J.G. and J.Z.
582 A.K. coordinated, performed and oversaw the sample acquisition and processing. V.N. performed RNA
583 isolation and library preparation. A.K. coordinated and oversaw sample sequencing. Z.S., W.D. and A.K.
584 performed and interpreted the computational analyses. A.K. J.R., H.M., J.G. and J.Z. interpreted the data.
585 A.K. wrote and edited the manuscript; C.P., W.D., J.R., H.M., J.G. and J.Z. critically reviewed and edited
586 the manuscript. [+]J.G. and J.Z. contributed equally. All authors approved the final manuscript.
587

588 **Acknowledgements** This work was funded by US National Institutes of Health grants 1K08HG010061-
589 01A1 and 3UL1TR001085-04S1 to A.K and the Charles and Ann Johnson Foundation to J.G.
590 A.K. would like to thank mentors, Profs. Richard Becker (University of Cincinnati, OH), Andrew Weyrich
591 (University of Utah, UT), Harry Greenberg (Stanford University, CA), Rob Tibshirani (Stanford University,
592 CA) and Stephen Montgomery (Stanford University, CA) for their support in her unique return to research
593 and NIH funding after a 5-year hiatus that then led to the work outlined in this manuscript. All authors
594 thank the patients at the Stanford Cancer Center for their generous participation in this research, and the
595 Stanford Functional Genomics Facility for genomic data storage. Authors also thank Drs. Belinda Guo and
596 Wendy Erber (University of Western Australia, Australia) for the gene counts file of their published data
597 (PMID 31426129) used for our independent validation. A.K. extends special thanks to colleagues Dr. Kellie
598 Machlus (Harvard University, MA), and Dr. Eric Pietras (University of Colorado, CO) for their critical reading
599 of the manuscript.
600

601 **Ethics Declarations/Competing Interests** The authors declare no competing interests.
602
603
604

605 Bibliography
606

607 Amisten, S. (2012). A rapid and efficient platelet purification protocol for platelet gene expression studies.
608 Methods in molecular biology (Clifton, NJ) *788*, 155-172.
609 Arcaini, L., and Cazzola, M. (2018). Benefits and risks of JAK inhibition. Blood *132*, 675-676.
610 Barbui, T., Carobbio, A., Finazzi, G., Vannucchi, A. M., Barosi, G., Antonioli, E., Guglielmelli, P., Pancrazzi,
611 A., Salmoiraghi, S., Zilio, P., Ottomano, C.*, et al.* (2011). Inflammation and thrombosis in essential
612 thrombocythemia and polycythemia vera: different role of C-reactive protein and pentraxin 3.
613 Haematologica *96*, 315-318.
614 Barbui, T., and Falanga, A. (2016). Molecular biomarkers of thrombosis in myeloproliferative neoplasms.
615 Thromb Res *140 Suppl 1*, S71-75.
616 Best, M. G., Sol, N., Kooi, I., Tannous, J., Westerman, B. A., Rustenburg, F., Schellen, P., Verschueren, H.,
617 Post, E., Koster, J., Ylstra, B.*, et al.* (2015). RNA-Seq of Tumor-Educated Platelets Enables Blood-Based
618 Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. Cancer Cell *28*, 666-676.
619 Campbell, R. A., Franks, Z., Bhatnagar, A., Rowley, J. W., Manne, B. K., Supiano, M. A., Schwertz, H.,
620 Weyrich, A. S., and Rondina, M. T. (2018). Granzyme A in human platelets regulates the synthesis of
621 proinflammatory cytokines by monocytes in aging. J Immunol *200*, 295-304.
622 Caparros-Perez, E., Teruel-Montoya, R., Lopez-Andreo, M. J., Llanos, M. C., Rivera, J., Palma-Barqueros,
623 V., Blanco, J. E., Vicente, V., Martinez, C., and Ferrer-Marin, F. (2017). Comprehensive comparison of
624 neonate and adult human platelet transcriptomes. PLoS One *12*, e0183042.
625 Clancy, L., and Freedman, J. E. (2016). Blood-Derived Extracellular RNA and Platelet Pathobiology: Adding
626 Pieces to a Complex Circulating Puzzle. Circulation research *118*, 374-376.
627 Clark, R. A., Li, S.-L., Pearson, D. W., Leidal, K. G., Clark, J. R., Denning, G. M., Reddick, R., Krause, K.-H.,
628 and Valente, A. J. (2002). Regulation of calreticulin expression during induction of differentiation in human
629 myeloid cells. Evidence for remodeling of the endoplasmic reticulum. J Biol Chem *277*, 32369-32378.
630 Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R., Bolduc, V., Waddell,
631 L. B., Sandaradura, S. A., O'Grady, G. L., Estrella, E.*, et al.* (2017). Improving genetic diagnosis in
632 Mendelian disease with transcriptome sequencing. Sci Transl Med *9*.
633 Cunin, P., Bouslama, R., Machlus, K. R., Martinez-Bonet, M., Lee, P. Y., Wactor, A., Nelson-Maney, N.,
634 Morris, A., Guo, L., Weyrich, A., Sola-Visner, M.*, et al.* (2019). Megakaryocyte emperipolesis mediates
635 membrane transfer from intracytoplasmic neutrophils to platelets. elife *8*.
636 Curto-Garcia, N., Harrison, C., and McLornan, D. P. (2020). Bone marrow niche dysregulation in
637 myeloproliferative neoplasms. Haematologica *105*, 1189-1200.
638 Davizon-Castillo, P., Rowley, J. W., and Rondina, M. T. (2020). Megakaryocyte and Platelet
639 Transcriptomics for Discoveries in Human Health and Disease. Arterioscler Thromb Vasc Biol *40*, 1432-
640 1440.

641 Dunbar, A., Nazir, A., and Levine, R. (2017). Overview of Transgenic Mouse Models of Myeloproliferative
642 Neoplasms (MPNs). Curr Protoc Pharmacol *77*, 14 40 11-14 40 19.
643 Finazzi, G., De Stefano, V., and Barbui, T. (2013). Are MPNs vascular diseases? Current hematologic
644 malignancy reports *8*, 307-316.
645 Gangaraju, R., Song, J., Kim, S. J., Tashi, T., Reeves, B. N., Sundar, K. M., Thiagarajan, P., and Prchal, J.
646 T. (2020). Thrombotic, inflammatory, and HIF-regulated genes and thrombosis risk in polycythemia vera
647 and essential thrombocythemia. Blood Adv *4*, 1115-1130.
648 Geyer, H. L., Dueck, A. C., Scherber, R. M., and Mesa, R. A. (2015). Impact of inflammation on
649 myeloproliferative neoplasm symptom development. Mediators Inflamm *2015*, 284706.
650 Gilles, L., Finke, C., Lasho, T. L., Pardanani, A., Tefferi, A., and Crispino, J. (2012). Aberrant megakaryocyte
651 gene expression contributes to primary myelofibrosis. Blood *120*, 2867-2867.
652 Gnatenko, D. V., Dunn, J. J., McCorkle, S. R., Weissmann, D., Perrotta, P. L., and Bahou, W. F. (2003).
653 Transcript profiling of human platelets using microarray and serial analysis of gene expression. Blood *101*,
654 2285-2293.
655 Guo, B. B., Linden, M. D., Fuller, K. A., Phillips, M., Mirzai, B., Wilson, L., Chuah, H., Liang, J., Howman, R.,
656 Grove, C. S., Malherbe, J. A.*, et al.* (2019). Platelets in myeloproliferative neoplasms have a distinct
657 transcript signature in the presence of marrow fibrosis. Br J Haematol.
658 Guo, B. B., Linden, M. D., Fuller, K. A., Phillips, M., Mirzai, B., Wilson, L., Chuah, H., Liang, J., Howman, R.,
659 Grove, C. S., Malherbe, J. A.*, et al.* (2020). Platelets in myeloproliferative neoplasms have a distinct
660 transcript signature in the presence of marrow fibrosis. Br J Haematol *188*, 272-282.
661 Hasselbalch, H. C., and Bjorn, M. E. (2015). MPNs as Inflammatory Diseases: The Evidence,
662 Consequences, and Perspectives. Mediators Inflamm *2015*, 102476.
663 Hinds, D. A., Barnholt, K. E., Mesa, R. A., Kiefer, A. K., Do, C. B., Eriksson, N., Mountain, J. L., Francke, U.,
664 Tung, J. Y., Nguyen, H. M., Zhang, H.*, et al.* (2016). Germ line variants predispose to both JAK2 V617F
665 clonal hematopoiesis and myeloproliferative neoplasms. Blood *128*, 1121-1128.
666 Kaushik, S., and Cuervo, A. M. (2015). Proteostasis and aging. Nature Medicine *21*, 1406-1415.
667 Kleppe, M., Koche, R., Zou, L., van Galen, P., Hill, C. E., Dong, L., De Groote, S., Papalexi, E., Hanasoge
668 Somasundara, A. V., Cordner, K., Keller, M.*, et al.* (2018). Dual Targeting of Oncogenic Activation and
669 Inflammatory Signaling Increases Therapeutic Efficacy in Myeloproliferative Neoplasms. Cancer Cell *33*,
670 785-787.
671 Koschmieder, S., and Chatain, N. (2020). Role of inflammation in the biology of myeloproliferative
672 neoplasms. Blood Rev *42*, 100711.
673 Krause, D. S., and Crispino, J. D. (2013). Molecular pathways: induction of polyploidy as a novel
674 differentiation therapy for leukemia. Clin Cancer Res *19*, 6084-6088.
675 Krishnan, A., Zhang, Y., Perkins, C., Gotlib, J., and Zehnder, J. L. (2017). Platelet Transcriptomic
676 Signatures in Myeloproliferative Neoplasms. Blood *130*, 5288.

677 Kukurba, K. R., and Montgomery, S. B. (2015). RNA Sequencing and Analysis. Cold Spring Harb Protoc
678 *2015*, 951-969.
679 LaFave, L. M., and Levine, R. L. (2016). Targeting a regulator of protein homeostasis in myeloproliferative
680 neoplasms. Nat Med *22*, 20-21.
681 Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-
682 359.
683 Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or
684 without a reference genome. BMC Bioinformatics *12*, 323.
685 Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The
686 Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst *1*, 417-425.
687 Liu, L., Inoki, A., Fan, K., Mao, F., Shi, G., Jin, X., Zhao, M., Ney, G., Jones, M. A., Sun, S., Dou, Y.*, et al.*
688 (2020). ER associated degradation preserves hematopoietic stem cell quiescence and self-renewal by
689 restricting mTOR activity. Blood.
690 Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for
691 RNA-seq data with DESeq2. Genome Biol *15*, 550.
692 Lussana, F., Carobbio, A., Salmoiraghi, S., Guglielmelli, P., Vannucchi, A. M., Bottazzi, B., Leone, R.,
693 Mantovani, A., Barbui, T., and Rambaldi, A. (2017). Driver mutations (JAK2V617F, MPLW515L/K or CALR),
694 pentraxin-3 and C-reactive protein in essential thrombocythemia and polycythemia vera. J Hematol Oncol
695 *10*, 54.
696 Manne, B. K., Denorme, F., Middleton, E. A., Portier, I., Rowley, J. W., Stubben, C., Petrey, A. C., Tolley, N.
697 D., Guo, L., Cody, M., Weyrich, A. S.*, et al.* (2020). Platelet gene expression and function in patients with
698 COVID-19. Blood *136*, 1317-1329.
699 Marin Oyarzún, C. P., and Heller, P. G. (2019). Platelets as mediators of thromboinflammation in chronic
700 myeloproliferative neoplasms. Front Immunol *10*, 1373.
701 Mascarenhas, J., Roper, N., Chaurasia, P., and Hoffman, R. (2011). Epigenetic abnormalities in
702 myeloproliferative neoplasms: a target for novel therapeutic strategies. Clin Epigenetics *2*, 197-212.
703 Matsuura, S., Thompson, C. R., Belghasem, M. E., Bekendam, R. H., Piasecki, A., Leiva, O., Ray, A.,
704 Italiano, J., Yang, M., Merill-Skoloff, G., Chitalia, V. C.*, et al.* (2020). Platelet Dysfunction and Thrombosis in
705 JAK2(V617F)-Mutated Primary Myelofibrotic Mice. Arterioscler Thromb Vasc Biol *40*, e262-e272.
706 Merlinsky, T. R., Levine, R. L., and Pronier, E. (2019). Unfolding the Role of Calreticulin in Myeloproliferative
707 Neoplasm Pathogenesis. Clin Cancer Res *25*, 2956-2962.
708 Middleton, E. A., Rowley, J. W., Campbell, R. A., Grissom, C. K., Brown, S. M., Beesley, S. J., Schwertz,
709 H., Kosaka, Y., Manne, B. K., Krauel, K., Tolley, N. D.*, et al.* (2019). Sepsis alters the transcriptional and
710 translational landscape of human and murine platelets. Blood *134*, 911-923.

711 Moons, K. G., Kengne, A. P., Grobbee, D. E., Royston, P., Vergouwe, Y., Altman, D. G., and Woodward, M.
712 (2012). Risk prediction models: II. External validation, model updating, and impact assessment. Heart *98*,
713 691-698.
714 Mughal, T. I., Pemmaraju, N., Radich, J. P., Deininger, M. W., Kucine, N., Kiladjian, J. J., Bose, P., Gotlib,
715 J., Valent, P., Chen, C. C., Barbui, T.*, et al.* (2019). Emerging translational science discoveries, clonal
716 approaches, and treatment trends in chronic myeloproliferative neoplasms. Hematol Oncol *37*, 240-252.
717 Mullally, A., Bruedigam, C., Poveromo, L., Heidel, F. H., Purdon, A., Vu, T., Austin, R., Heckl, D., Breyfogle,
718 L. J., Kuhn, C. P., Kalaitzidis, D.*, et al.* (2013). Depletion of Jak2V617F myeloproliferative neoplasm-
719 propagating stem cells by interferon-alpha in a murine model of polycythemia vera. Blood *121*, 3692-3702.
720 Nguyen, H. M., and Gotlib, J. (2012). Insights into the molecular genetics of myeloproliferative neoplasms.
721 Am Soc Clin Oncol Educ Book, 411-418.
722 Oh, S. T., and Gotlib, J. (2010). JAK2 V617F and beyond: role of genetics and aberrant signaling in the
723 pathogenesis of myeloproliferative neoplasms. Expert Rev Hematol *3*, 323-337.
724 Osorio, F. G., Soria-Valles, C., Santiago-Fernandez, O., Bernal, T., Mittelbrunn, M., Colado, E., Rodriguez,
725 F., Bonzon-Kulichenko, E., Vazquez, J., Porta-de-la-Riva, M., Ceron, J.*, et al.* (2016). Loss of the
726 proteostasis factor AIRAPL causes myeloid transformation by deregulating IGF-1 signaling. Nat Med *22*,
727 91-96.
728 Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V. I., Paschka, P., Roberts, N. D., Potter, N. E.,
729 Heuser, M., Thol, F., Bolli, N., Gundem, G.*, et al.* (2016). Genomic Classification and Prognosis in Acute
730 Myeloid Leukemia. N Engl J Med *374*, 2209-2221.
731 Passamonti, F., Cervantes, F., Vannucchi, A. M., Morra, E., Rumi, E., Pereira, A., Guglielmelli, P.,
732 Pungolino, E., Caramella, M., Maffioli, M., Pascutto, C.*, et al.* (2010). A dynamic prognostic model to
733 predict survival in primary myelofibrosis: a study by the IWG-MRT (International Working Group for
734 Myeloproliferative Neoplasms Research and Treatment). Blood *115*, 1703-1708.
735 Raghavachari, N., Xu, X., Harris, A., Villagra, J., Logun, C., Barb, J., Solomon, M. A., Suffredini, A. F.,
736 Danner, R. L., Kato, G., Munson, P. J.*, et al.* (2007). Amplified expression profiling of platelet transcriptome
737 reveals changes in arginine metabolic pathways in patients with sickle cell disease. Circulation *115*, 1551-
738 1562.
739 Rampal, R., Al-Shahrour, F., Abdel-Wahab, O., Patel, J. P., Brunel, J.-P., Mermel, C. H., Bass, A. J., Pretz,
740 J., Ahn, J., Hricik, T., Kilpivaara, O.*, et al.* (2014). Integrated genomic analysis illustrates the central role of
741 JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. Blood *123*, e123-133.
742 Rondina, M. T., Weyrich, A. S., and Zimmerman, G. A. (2013). Platelets as cellular effectors of inflammation
743 in vascular diseases. Circulation research *112*, 1506-1519.
744 Rontauroli, S., Castellano, S., Guglielmelli, P., Zini, R., Bianchi, E., Genovese, E., Carretta, C., Parenti, S.,
745 Fantini, S., Mallia, S., Tavernari, L.*, et al.* (2021). Gene expression profile correlates with molecular and
746 clinical features in patients with myelofibrosis. Blood Adv *5*, 1452-1462.

747 Rowley, J. W., Oler, A. J., Tolley, N. D., Hunter, B. N., Low, E. N., Nix, D. A., Yost, C. C., Zimmerman, G.
748 A., and Weyrich, A. S. (2011). Genome-wide RNA-seq analysis of human and mouse platelet
749 transcriptomes. Blood *118*, e101-111.
750 Rowley, J. W., Schwertz, H., and Weyrich, A. S. (2012). Platelet mRNA: the meaning behind the message.
751 Curr Opin Hematol *19*, 385-391.
752 Rowley, J. W., Weyrich, A. S., and Bray, P. F. (2019). The platelet transcriptome in health and disease. In
753 Platelets, (Elsevier), pp. 139-153.
754 Rumi, E., and Cazzola, M. (2017). Diagnosis, risk stratification, and response evaluation in classical
755 myeloproliferative neoplasms. Blood *129*, 680-692.
756 Sampieri, L., Di Giusto, P., and Alvarez, C. (2019). CREB3 Transcription Factors: ER-Golgi Stress
757 Transducers as Hubs for Cellular Homeostasis. Front Cell Dev Biol *7*, 123.
758 Schischlik, F., Jager, R., Rosebrock, F., Hug, E., Schuster, M., Holly, R., Fuchs, E., Milosevic Feenstra, J.
759 D., Bogner, E., Gisslinger, B., Schalling, M.*, et al.* (2019). Mutational landscape of the transcriptome offers
760 putative targets for immunotherapy of myeloproliferative neoplasms. Blood *134*, 199-210.
761 Schubert, S., Weyrich, A. S., and Rowley, J. W. (2014). A tour through the transcriptional landscape of
762 platelets. Blood *124*, 493-502.
763 Simon, L. M., Edelstein, L. C., Nagalla, S., Woodley, A. B., Chen, E. S., Kong, X., Ma, L., Fortina, P.,
764 Kunapuli, S., Holinstat, M., McKenzie, S. E.*, et al.* (2014). Human platelet microRNA-mRNA networks
765 associated with age and gender revealed by integrated plateletomics. Blood *123*, e37-45.
766 Skov, V., Larsen, T. S., Thomassen, M., Riley, C. H., Jensen, M. K., Bjerrum, O. W., Kruse, T. A., and
767 Hasselbalch, H. C. (2011). Whole-blood transcriptional profiling of interferon-inducible genes identifies
768 highly upregulated IFI27 in primary myelofibrosis. European Journal of Haematology *87*, 54-60.
769 Skov, V., Larsen, T. S., Thomassen, M., Riley, C. H., Jensen, M. K., Bjerrum, O. W., Kruse, T. A., and
770 Hasselbalch, H. C. (2012). Molecular profiling of peripheral blood cells from patients with polycythemia
771 vera and related neoplasms: identification of deregulated genes of significance for inflammation and
772 immune surveillance. Leuk Res *36*, 1387-1392.
773 Spivak, J. L. (2017). Myeloproliferative Neoplasms. N Engl J Med *376*, 2168-2181.
774 Spivak, J. L., Considine, M., Williams, D. M., Talbot, C. C., Jr., Rogers, O., Moliterno, A. R., Jie, C., and
775 Ochs, M. F. (2014). Two clinical phenotypes in polycythemia vera. N Engl J Med *371*, 808-817.
776 Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A.,
777 Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a
778 knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A
779 *102*, 15545-15550.
780 Tefferi, A., Abdel-Wahab, O., Cervantes, F., Crispino, J. D., Finazzi, G., Girodon, F., Gisslinger, H., Gotlib,
781 J., Kiladjian, J. J., Levine, R. L., Licht, J. D.*, et al.* (2011). Mutations with epigenetic effects in

782 myeloproliferative neoplasms and recent progress in treatment: Proceedings from the 5th International
783 Post-ASH Symposium. Blood Cancer J *1*, e7.
784 Tefferi, A., Lasho, T. L., Guglielmelli, P., Finke, C. M., Rotunno, G., Elala, Y., Pacilli, A., Hanson, C. A.,
785 Pancrazzi, A., Ketterling, R. P., Mannarelli, C.*, et al.* (2016). Targeted deep sequencing in polycythemia
786 vera and essential thrombocythemia. Blood Adv *1*, 21-30.
787 Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical
788 Society Series B (Methodological), 267-288.
789 Vainchenker, W., and Kralovics, R. (2017). Genetic basis and molecular pathophysiology of classical
790 myeloproliferative neoplasms. Blood *129*, 667-679.
791 Vainchenker, W., Leroy, E., Gilles, L., Marty, C., Plo, I., and Constantinescu, S. N. (2018). JAK inhibitors for
792 the treatment of myeloproliferative neoplasms and other disorders. F1000Res *7*, 82.
793 Vannucchi, A. M., Antonioli, E., Guglielmelli, P., Pardanani, A., and Tefferi, A. (2008). Clinical correlates of
794 JAK2V617F presence or allele burden in myeloproliferative neoplasms: a critical reappraisal. Leukemia *22*,
795 1299-1307.
796 Wen, Q. J., Woods, B., Yang, Q., Sophia, C., Lillu, G., Rapaport, F., Levine, R. L., and Crispino, J. (2016).
797 Activation of JAK/STAT signaling in megakaryocytes is necessary and sufficient for myeloproliferation in
798 vivo. Blood *128*, 949-949.
799 Wen, Q. J., Yang, Q., Goldenson, B., Malinge, S., Lasho, T., Schneider, R. K., Breyfogle, L. J., Schultz, R.,
800 Gilles, L., Koppikar, P., Abdel-Wahab, O.*, et al.* (2015). Targeting megakaryocytic-induced fibrosis in
801 myeloproliferative neoplasms by AURKA inhibition. Nat Med *21*, 1473-1480.
802 Weyrich, A. S. (2014). Platelets: more than a sack of glue. Hematology Am Soc Hematol Educ Program
803 *2014*, 400-403.
804 Weyrich, A. S., and Zimmerman, G. A. (2003). Evaluating the relevance of the platelet transcriptome. Blood
805 *102*, 1550-1551.
806 Weyrich, A. S., and Zimmerman, G. A. (2013). Platelets in lung biology. Annual review of physiology *75*,
807 569-591.
808 Williams, N., Lee, J., Moore, L., Baxter, E. J., Hewinson, J., Dawson, K. J., Menzies, A., Godfrey, A. L.,
809 Green, A. R., Campbell, P. J., and Nangalia, J. (2020). Phylogenetic reconstruction of myeloproliferative
810 neoplasm reveals very early origins and lifelong evolution. BioRxiv, 2020.2011.2009.374710.
811 Wong, W. J., Baltay, M., Getz, A., Fuhrman, K., Aster, J. C., Hasserjian, R. P., and Pozdnyakova, O. (2019).
812 Gene expression profiling distinguishes prefibrotic from overtly fibrotic myeloproliferative neoplasms and
813 identifies disease subsets with distinct inflammatory signatures. PLoS One *14*, e0216810.
814 Woods, B., Chen, W., Chiu, S., Marinaccio, C., Fu, C., Gu, L., Bulic, M., Yang, Q., Zouak, A., Jia, S.,
815 Suraneni, P. K.*, et al.* (2019). Activation of JAK/STAT Signaling in Megakaryocytes Sustains
816 Myeloproliferation In Vivo. Clin Cancer Res *25*, 5901-5912.

817 Zeddies, S., Jansen, S. B., di Summa, F., Geerts, D., Zwaginga, J. J., van der Schoot, C. E., von Lindern,
818 M., and Thijssen-Timmer, D. C. (2014). MEIS1 regulates early erythroid and megakaryocytic cell fate.
819 Haematologica *99*, 1555-1564.
820 Zoi, K., and Cross, N. C. (2017). Genomics of Myeloproliferative Neoplasms. J Clin Oncol *35*, 947-954.
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

842 **Figure Legends**
843
844 **Figure 1: Two independent MPN clinical cohorts and closely replicated patient variables.**
845 **A,** Similarity in distribution of MPN subtypes between two cohorts of MPN patients (Stanford single-center;
846 approximately 2 years apart: cohort 1: 2016-17, n=71; and cohort 2: 2019, n=49); the majority subtype is
847 MF in both cohorts (34% of n=71 and 37% of n=49). **B**, Comparable distribution of age across MPN
848 subtypes in the two cohorts. Violin plots of patient age from each MPN subtype reflect clinical expectation,
849 with a fairly identical match between the two control MF groups between the two cohorts. Slightly higher
850 median age noted in the second cohort for ET and PV patients alone. **C,** Comparable and balanced
851 distribution of gender across MPN subtypes in the two cohorts. Larger percentage of male healthy controls
852 in both cohorts and smaller percentage of males in ET cohort 1 noted. **D,** Matched distribution of primarily
853 *JAK2* and *CALR* mutational status across MPN subtypes in the two cohorts. JAK2 is the most common
854 mutation across all three subtypes; 100% of PV and over 50% of ET and MF patients have JAK2 mutation
855 in both cohorts. Mismatch between cohorts on the MPL/triple negative patients is noted as a natural
856 consequence of the rarer prevalence of these mutations; and therefore, not the primary focus of this work.
857 **E,** Diversity of MPN patient therapies across the two cohorts reflecting current clinical practice. The
858 majority being aspirin (ASA) in ET/PV patients and the *JAK*-inhibitor, ruxolitinib in MF. Note that a given
859 patient may be on more than one treatment and therefore, the total treatment percentage in this graphic
860 may not equal 100. To control for any inter-patient variability, all treatment, in addition to patient age and
861 gender are adjusted as confounding factors in downstream gene expression analyses. **F,** Representative
862 clinical laboratory variables, as box plots, measured at the same date and time as platelet sampling.
863 Compared to controls, MPN patients show larger variance (inter-quartile range, IQR), and reflect current
864 clinical knowledge. Groups differ primarily only with respect to blood cell counts (platelet/RBC/WBC); and
865 show the greatest differential in MF. Note higher platelet count in ET with a concomitant lower mean
866 platelet volume, higher red blood cell count in PV and lower red blood cell count in MF with concomitantly
867 lower hemoglobin, hematocrit and higher red cell width. Wilcoxson signed rank tests marked by asterisks
868 indicate a statistically significant difference between any two groups (* $p<=0.05$; ** $p<=0.01$; *** $p<=0.001$;
869 **** $p<=0.0001$). **G,** High correlation ($R^2 = 0.89$) of platelet gene expression as assessed by normalized
870 counts of matched transcripts in each cohort between each of controls, ET, PV and MF. The two-cohort
871 collective sample size totals n=120, affording increased statistical power for subsequent analyses.
872
873
874
875
876
877

878 **Figure 2: MPN platelet transcriptome distinguishes disease phenotype and reveals phenotype- and**
879 **JAK-inhibitor specific signatures.**
880 **A,** Unsupervised principal component analysis (PCA) of normalized platelet gene expression counts
881 adjusted for age, gender, treatment and experimental batch. Three panels of PC1 and PC2 colored by
882 MPN subtype; and each contrasted with controls (n=21, yellow): ET (n=24, top, light green), PV (n=33,
883 middle, dark green) and MF (n=42, bottom, dark blue). Circles filled or open mark presence or absence of
884 ruxolitinib treatment; and size of circles, smaller or larger, indicate presence or absence of *JAK2* mutation.
885 The first two principal components account for 48% of total variance in the data. **B,** Volcano plots (three
886 panels as **a** above of ET, PV and MF) of differential gene expression showing statistical significance
887 (negative log10 of p-values) versus log2 fold change of each gene. Significant up-regulated and down-
888 regulated genes are those with p-values (FDR) smaller or equal to 0.01 and absolute value of fold changes
889 larger or equal to 1.5**. C,D,** Venn Diagram comparisons of MPN differential gene expression lists. In **c,** each
890 of ET, PV and MF is contrasted with controls; identifying gene sets that are shared (n=1504, FDR < 0.05)
891 as well as unique to each subtype. In **D,** differential in the RUX-treated cohort is contrasted with MF vs
892 controls. Differential in gene expression from ruxolitinib is a fraction (approximately one-third) of the total
893 differential noted in the MF transcriptome. **E,** Volcano plot of differential gene expression between MPN
894 patients treated with ruxolitinib and not. A small subset of overlapping differential genes that are
895 upregulated in MF (**B,** bottom panel) and suppressed in the RUX-treated cohort (**E**) are colored in green.
896
897
898
899
900
901
902
903
904

905 **Figure 3: Graded differential expression by MPN phenotype and driver mutation status**

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

911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931

932 **Figure 4: Altered immune, metabolic, and proteostatic pathways underlie each MPN phenotype**

933 **A,** Pathway-enrichment analysis of genes with MPN subtype-specific expression (color indicated; light
934 green ET, dark green PV, and dark blue MF) overlayed with ruxolitinib-specific expression (light blue). Each
935 point represents a pathway; the *x*-axis gives the normalized enrichment score, which reflects the degree to
936 which each pathway is over or under-represented at the top or bottom of the ranked list of differentially
937 expressed genes, normalized to account for differences in gene set size and in correlations between gene
938 sets and the expression data set. The y-axis lists the detail-level node of the most enriched pathways;
939 solid lines mark GSEA-recommended(Subramanian et al., 2005) Bonferroni-corrected statistical
940 significance criterion of FDR < 0.25 for exploratory analyses. Dotted lines mark FDR > 0.25 and therefore,
941 not sufficiently significant, yet visualized alongside solid lines to retain overall context (upper-level parent
942 nodes of the detail-level pathways are provided in **Table S3A-C**). Multiple immune and inflammatory
943 pathways are consistently significantly enriched across ET, PV and MF (and suppressed in the ruxolitinib-
944 treated cohort). MF is differentiated from ET and PV through dysregulation of additional molecular
945 processes for cellular development, proliferation, metabolism and DNA damage.

946

947

948

949

950

951

952

953

954

955

956

957

958

959 **Figure 5: Prediction of MF based on unique and progressive MPN platelet transcriptome**

960 **A,** Top 12 (out of 4500+) monotonic progressive gene expression across all three MPN subtypes (FDR <
961 0.0001, *Mann-Kendall* test with Bonferroni correction. From ET to MF (*x*-axis), both monotonic increase
962 (*e.g. KPNB1, CALR, and SIGMAR1*) and decrease (*e.g. CPAMD8, BRWD1, NEMF*) noted in log 2 fold
963 change in expression (*y*-axis). **B,** Lasso-penalized multinomial logistic regression model under temporal
964 validation i.e. trained on cohort 1 (n=71, Figure 1a) and validated on cohort 2 (n=49, Figure 1a) as test set.
965 **C,** Lasso-penalized multinomial logistic regression model under geographic validation on an independently
966 published dataset (n=25 MF, n=15 controls). **D,** Receiver Operating Curves (ROC) toward MF prediction
967 under conditions of temporal (**B**) and geographic (**C**) validation. Temporal validation uses three layered
968 models: i) baseline, with no gene expression data but patient age, gender and mutation alone; ii) adding
969 entire MPN platelet transcriptome and iii) adding MPN progressive genes alone. Outperformance of the
970 progressive transcriptome model (red curve, AUROC=0.96) vis-a-vis the entire transcriptome dataset (blue
971 curve, AUROC=0.93) and lastly, the baseline model without gene expression (black curve, AUROC=0.69).
972 Geographic validation using the progressive transcriptome model also demonstrates independent high MF
973 predictive accuracy (brown curve, AUROC=0.86). **E,** Confusion matrices evaluating the different
974 multinomial Lasso logistic models. **F,** Heatmap of top recurring Lasso-selected progressive genes for each
975 of controls (left, CTRL, yellow), ET (light green), PV (dark green) and MF (dark blue). Rows indicate
976 gradation in gene expression on a blue (low) to red (high) scale. Columns indicate sample type (CTRL, ET,
977 PV and MF).

978

979

980

981

982

983

984

985

986

987 **Supplementary Figure Legends**
988

989 **Figure S1: Platelets isolated for RNA-sequencing. A.** Flow cytometry analysis for purity of platelets
990 (CD61+) isolated from whole blood**.** Population (quadrant 3, outlined in red) showing 96% pure platelets
991 prior to enrichment by depletion of CD45 leukocytes. **B.** CD45 magnetic microbeads were used to deplete
992 leukocytes, further enriching platelet population. Platelets enriched from magnetic sorting were confirmed
993 for ~99% platelet purity using a Cell-Dyn Hematology Analyzer prior to isolation of RNA. Standard trizol
994 RNA extraction protocol was used to extract RNA. **C.** Representative electropherogram results from
995 Agilent Bioanalyzer showing RNA isolated from platelets with an RNA Integrity Number of 9.4**. D***.*
996 Electropherogram showing a clean product following cDNA synthesis and amplification. Similar procedure
997 was carried out for all samples and by the same individual.

998

999 **Figure S2: Correlation of MF platelet transcriptome with disease risk by Dynamic International**
.000 **Prognostic Scoring System (DIPSS)**
.001 **A,** Unsupervised principal component analysis (PCA) of MF (n=42) normalized platelet gene expression
.002 counts adjusted for age, gender, treatment and experimental batch. PC1 and PC2 colored by DIPSS score
.003 (0-6 denoting increasing MF risk). Circles filled or open mark presence or absence of ruxolitinib treatment;
.004 and size of circles, smaller or larger, indicate presence or absence of *JAK2* mutation.

.005

.006 **Figure S3: Relative molecular trajectories of MPN progression**
.007 The relative enrichment of MPN molecular pathways reflecting MPN progression within specific biological
.008 process categories.(Liberzon et al., 2015) Each color-indicated line represents a pathway category. *x*-axis
.009 captures sample type from controls (CTRL) to ET, PV and MF, and the y-axis gives the normalized
.010 enrichment score, which reflects the degree to which each pathway is over or under-represented at the top
.011 or bottom of the ranked list of differentially expressed genes, normalized to account for differences in gene
.012 set size and in correlations between gene sets and the expression data set.

.013

.014

.015 **Table Descriptions**

.016 **Table S1:** MPN patient characteristics from two time-separated cohorts at the Stanford Cancer Institute.

.017

.018 **Table S2:** Overlap of differentially expressed genes between the RUX-treated and the MF patient cohorts.

.019

.020 **Table S3:** Full data for all molecular pathways identified in each of ET, PV and MF patient cohorts.

.021

.022 **Table S4:** Predicted probabilities for each of CTRL, ET, PV and MF using all three Lasso logistic models.

.023

.024

Table S1A. MPN Patient Characteristics Cohort1

	ET	PV	MF	Control	TOTAL
Subject Count	n=14	n=21	n=24	n=11	n=70
Sample Count	(n=14)	(n=22)	(n=24)	(n=11)	(n=71)
Median age, y (range)	52 (30-73)	57 (27-84)	72 (47-80)	57 (26-69)	60 (26-84)
Male, n (%)	3 (21)	11 (52)	13 (54)	8 (73)	35 (50)
Race					
Asian	3 (21)	1 (5)	1 (4)	2 (18)	7 (10)
Black	0	2 (10)	0	0	2 (3)
Native American	0	0	0	0	0
White	8 (57)	11 (52)	18 (75)	7 (64)	44 (63)
Other	2 (14)	6 (29)	3 (13)	2 (18)	13 (19)
Unknown	1 (7)	1 (5)	2 (8)	0	4 (6)
Ethnicity					
Hispanic or Latino	0	1 (5)	2 (8)	2 (18)	5 (7)
Not Hispanic or Latino	13 (93)	18 (86)	21 (88)	9 (82)	61 (87)
Unknown	1 (7)	2 (9)	1 (4)	0	4 (6)
MPN Mutation present, n (%)					
CALR	5 (36)	0	3 (13)		8 (11)
JAK2	7 (50)	21 (100)	14 (58)		42 (60)
MPL	0	0	2 (8)		2 (3)
Triple negative (no *CALR* , *JAK2* or *MPL* mutation)	2 (14)	0	5 (21)		7 (10)
Mutation status unknown	0	0	0 (0)	11	11 (16)
	ET	PV	MF		TOTAL
	n=14	n=21	n=24		n=59
Number of current therapies, n (%					
0	2 (14)	2 (10)	3 (13)		18 (26)
1	5 (36)	11 (52)	13 (54)		29 (41)
2+	7 (50)	8 (38)	8 (33)		23 (33)
Type of current therapy*, n (%)					
Aspirin	11 (79)	15 (71)	5 (21)		31 (44)
JAK Inhibitor	1 (7)	5 (24)	17 (71)		23 (33)
Hydroxyurea	6 (43)	7 (33)	4 (17)		17 (24)
Other**	1 (7)	1 (7)	4 (17)		6 (8)
None	2 (14)	0	3 (13)		5 (7)

*Not mutually exclusive

** 'Other' treatments include Warfarin, PEG interferon, Anagrelide, Luspatercept or Procrit in either cohort

Table S1B. MPN Patient Characteristics Cohort2

	ET	PV	MF	Control	TOTAL
Subject Count	n=10	n=10	n=16	n=10	n=46
Sample Count	(n=10)	(n=11)	(n=18)	(n=10)	(n=49)
Median age, y (range)	65 (31-78)	65 (29-84)	69 (43-83)	57 (26-69)	65 (26-84)
Male, n (%)	5 (50)	6 (60)	8 (50)	8 (80)	27 (59)
Race					
Asian	1 (10)	1 (10)	6 (38)	5 (50)	13 (28)
Black	0	0	1 (6)	0	1 (2)
Native American	0	0	0	2 (20)	2 (4)
White	8 (80)	8 (80)	7 (44)	2 (20)	25 (54)
Other	0	1 (10)	2 (13)	1 (10)	4 (9)
Unknown	1 (10)	0	0	0	1 (2)
Ethnicity					
Hispanic or Latino	0	1 (10)	1 (6)	1 (10)	3 (6)
Not Hispanic or Latino	9 (90)	9 (90)	15 (94)	9 (90)	42 (91)
Unknown	1 (10)	0	0	0	1 (2)
MPN Mutation present, n (%)					
CALR	2 (20)	0	3 (19)		5 (11)
JAK2	6 (60)	10 (100)	11 (68)		27 (59)
MPL	2 (20)	0	0		2 (4)
Triple negative (no *CALR* , *JAK2* or *MPL* mutation)	0	0	2 (13)		2 (4)
Mutation status unknown	0	0	0	10	10 (22)
	ET	PV	MF		TOTAL
	n=10	n=10	n=16		n=36
Number of current therapies, n (%)					
0	1 (10)	0	1 (6)		2 (6)
1	7 (70)	6 (60)	10 (63)		23 (64)
2+	2 (20)	4 (40)	5 (31)		11 (31)
Type of current therapy*, n (%)					
Aspirin	5 (50)	7 (70)	4 (25)		16 (44)
JAK Inhibitor	0	2 (20)	11 (69)		13 (36)
Hydroxyurea	5 (50)	4 (40)	1 (6)		10 (28)
Other**	1 (10)	2 (20)	4 (25)		7 (19)
None	1 (10)	0	1 (6)		2 (6)

TABLE S2A FOR FIGURE 2D-E

Gene	Name	baseMean_n	log2FoldChar	lfcSE_mf	stat_mf	pvalue_mf	padj_mf	baseMean_r	log2FoldChar	lfcSE_rux	stat_rux	pvalue_rux	padj_rux
IFIT2	interferon induced protein with tetratricopeptide repeats 2	129.384803	0.9920401	0.37060592	3.18653678	0.00143987	0.00487399	171.851723	-0.3453671	0.3579696	-2.6922496	0.00709718	0.04449282
IFIT1	interferon induced protein with tetratricopeptide repeats 1	72.4408011	1.40286011	0.38905001	3.79213932	0.00014936	0.0007243	100.076705	-0.4101571	0.36086417	-3.0618261	0.00219991	0.02263801
STK16	serine/threonine kinase 16	244.802755	1.05639738	0.16078705	6.39906564	1.56E-10	6.65E-09	226.844043	-0.4222584	0.16697508	-2.6791402	0.00738115	0.04567426
WDR83OS	WD repeat domain 83 opposite strand	795.037346	0.39346143	0.16711237	2.28049953	0.02257808	0.04835915	743.232856	-0.4583307	0.17990465	-2.6424942	0.00822979	0.04906111
DMTN	dematin actin binding protein	27421.2887	0.34577347	0.16016732	2.2663339	0.02343095	0.04996231	25158.15	-0.4805121	0.18917939	-2.6862843	0.00722516	0.04499458
JOSD1	Josephin domain containing 1	1233.45453	0.61524931	0.16890599	3.62439611	0.00028964	0.00125782	1046.72807	-0.4929592	0.18421387	-2.6949632	0.00703964	0.04427998
CD55	CD55 molecule (Cromer blood group)	1171.57591	0.82601754	0.17929011	4.40374652	1.06E-05	7.91E-05	1134.44707	-0.4978962	0.18592924	-2.944897	0.00323062	0.02830625
STUB1	STIP1 homology and U-box containing protein 1	562.789649	0.45436161	0.16503396	2.70269459	0.00687799	0.01797597	526.212423	-0.4988989	0.18468655	-2.6859747	0.00723186	0.04501325
YIPF3	Yip1 domain family member 3	1419.49555	0.41398667	0.16446889	2.52812523	0.01146734	0.02762937	1269.40529	-0.5036147	0.17984043	-2.9263638	0.0034295	0.02922411
OTUD5	OTU deubiquitinase 5	1096.62533	0.71434274	0.18471809	3.76993238	0.00016329	0.00077953	2095.01698	-0.5389033	0.19035997	-3.056888	0.00223648	0.02279463
SHARPIN	SHANK associated RH domain interactor	783.471597	0.55828677	0.2060955	2.78758199	0.0053103	0.01448509	831.550038	-0.5673856	0.22765763	-2.637477	0.00835253	0.04945374
L2HGDH	L-2-hydroxyglutarate dehydrogenase	63.3000848	0.97871381	0.22488487	3.88258297	0.00010335	0.00053817	485.320677	-0.5747827	0.18385617	-2.931365	0.00337476	0.0289454
TXNDC17	thioredoxin domain containing 17	52.0113013	0.54886962	0.21409354	2.29933888	0.0214857	0.04642081	53.6199211	-0.5975114	0.22919577	-2.6379027	0.00834205	0.04942045
POLD4	DNA polymerase delta 4, accessory subunit	436.101852	0.65763515	0.15115069	4.29606121	1.74E-05	0.00011857	445.186375	-0.6016293	0.18401998	-3.4218647	0.00062193	0.01086728
KLHDC8B	kelch domain containing 8B	1561.67143	1.19613668	0.276078	4.13415839	3.56E-05	0.00021821	1387.63295	-0.6116034	0.28867266	-2.7306207	0.00632152	0.04163646
PPP1R3F	protein phosphatase 1 regulatory subunit 3F	23.7905062	0.4860071	0.20347342	2.43396562	0.01493441	0.03438667	47.3532212	-0.6137633	0.21229873	-2.9030969	0.00369492	0.03065107
FAM89B	family with sequence similarity 89 member B	1515.68987	0.54469774	0.19827833	2.74458083	0.00605882	0.01618292	1435.00378	-0.6188486	0.21650888	-2.9466294	0.00321258	0.02821744
RENBP	renin binding protein	143.177964	0.93993558	0.22659105	3.96559779	7.32E-05	0.00040315	277.574729	-0.6243111	0.22421852	-2.9916588	0.00277466	0.02589422
PSAP	prosaposin	20385.4829	0.47597507	0.21886709	2.40446018	0.01619637	0.03674893	18741.5687	-0.6622052	0.22869444	-3.0409614	0.00235824	0.02358241
YIF1B	Yip1 interacting factor homolog B, membrane trafficking protein	3688.02924	0.57131556	0.22110123	2.54222831	0.01101482	0.02676862	3138.57453	-0.6709532	0.24491986	-2.8369756	0.00455431	0.03421591
SNHG11	small nucleolar RNA host gene 11	222.462972	0.69777747	0.25603154	2.53711279	0.01117709	0.02706628	156.473797	-0.6803125	0.27563341	-2.6581199	0.00785779	0.04740653
CCDC157	coiled-coil domain containing 157	53.0184143	0.71644334	0.20109786	3.58587547	0.00033595	0.00142776	31.1621345	-0.7177901	0.21898232	-3.3185123	0.00090498	0.01343499
EGFL7	EGF like domain multiple 7	1566.22711	0.79442401	0.23828563	3.37474024	0.00073885	0.00277505	1436.6383	-0.721958	0.25220845	-3.0555248	0.00224667	0.02284926
HIST2H2AA4	H2A clustered histone 19	39.7664675	0.40018242	0.33237648	4.60077552	4.21E-06	3.72E-05	1621.4983	-0.7436426	0.27496848	-3.0295511	0.00244918	0.02417893
NECAB3	N-terminal EF-hand calcium binding protein 3	570.05801	2.12442071	0.31908519	6.1164525	9.57E-10	3.00E-08	470.107939	-0.7751935	0.30924274	-3.0338032	0.00241492	0.0239668
ARSA	arylsulfatase A	399.688136	0.68556973	0.22088373	2.94944946	0.00318341	0.00951488	391.592812	-0.8003901	0.22791563	-3.4381874	0.00058562	0.01049579
CRYM	crystallin mu	547.013042	1.49161234	0.34369578	3.87719594	0.00010567	0.00054713	494.909076	-0.8057958	0.34135695	-2.6483936	0.00808753	0.04846338
TYROBP	transmembrane immune signaling adaptor TYROBP	289.219362	0.62584715	0.26172204	2.4685591	0.01356583	0.0317215	321.811877	-0.8129676	0.28139401	-2.7469143	0.00601588	0.04053845
PANX2	pannexin 2	15.8975401	0.83101183	0.4287281	2.27170664	0.02310424	0.04936339	20.2681033	-0.8361811	0.39906675	-2.8349631	0.0045831	0.03435082
E2F1	E2F transcription factor 1	3353.7609	1.25957957	0.28803002	4.11073841	3.94E-05	0.00023807	2993.57865	-0.8660957	0.2877488	-3.069719	0.0021426	0.02229121
CAPN11	calpain 11	249.938075	0.73843799	0.31239843	2.70547592	0.00682066	0.0178514	237.626168	-0.879357	0.3117876	-2.8336613	0.00460181	0.03439577
ISCU	iron-sulfur cluster assembly enzyme	2640.30277	0.46725575	0.21113478	2.28947383	0.02205184	0.04740355	2656.96355	-0.8807938	0.22312194	-4.1635732	3.13E-05	0.00182161
RNASET2	ribonuclease T2	463.913557	0.71297435	0.29309875	2.47033721	0.01349857	0.03161	473.07086	-0.9190513	0.30431874	-3.0077029	0.0026323	0.02504596
SOCS3	suppressor of cytokine signaling 3	164.860539	0.71182374	0.35338286	2.44376466	0.01453491	0.03361652	245.03243	-0.9192828	0.35079287	-3.0851113	0.00203476	0.02153676
MYLIP	myosin regulatory light chain interacting protein	246.76322	0.88916952	0.28483088	3.12270613	0.00179197	0.00587056	323.661539	-0.9285208	0.28953842	-3.2548403	0.00113456	0.01532834
GCH1	GTP cyclohydrolase 1	196.538295	1.56544269	0.29605834	4.81830051	1.45E-06	1.50E-05	261.637501	-0.9340261	0.32161913	-2.9887997	0.00280076	0.02604443
SEC16B	SEC16 homolog B, endoplasmic reticulum export factor	23.3779755	0.86707448	0.34163561	2.36965822	0.01780454	0.03976941	29.0798944	-0.9419619	0.30314071	-3.0509623	0.00228109	0.02312774
KIF17	kinesin family member 17	25.7468948	0.74720877	0.38120544	2.38418655	0.01711692	0.0384662	25.1884183	-1.0058894	0.37141035	-3.5767729	0.00034786	0.00784489
C20orf96	chromosome 20 open reading frame 96	133.551884	0.96736288	0.33940809	3.24554673	0.00117225	0.00409576	117.425883	-1.0953035	0.31253257	-3.3557657	0.00079146	0.01242691

TABLE S2B FOR FIGURE 2D-E

Gene	Name	baseMean_n	log2FoldChai	lfcSE_mf	stat_mf	pvalue_mf	padj_mf	baseMean_r	log2FoldChai	lfcSE_rux	stat_rux	pvalue_rux	padj_rux
PER3	period circad	111.957039	-0.9386283	0.40729465	-2.8430956	0.00446777	0.01257161	112.241192	1.17070798	0.37618026	3.60833205	0.00030817	0.00729464
LMO7	LIM domain	31.7560914	-0.5369593	0.30677707	-2.4118959	0.01586981	0.03616034	37.5290771	0.9946848	0.30176353	3.48718415	0.00048814	0.00947248
ACP6	acid phospha	11.2109527	-0.7942913	0.33690928	-2.4076161	0.01605706	0.0364969	22.3270702	0.81014563	0.26688183	3.0188037	0.00253775	0.02462779
CCT6B	chaperonin c	11.7153836	-1.3906478	0.29147977	-4.2607228	2.04E-05	0.00013575	9.80126886	0.80075571	0.29792345	3.02007961	0.00252708	0.02458468
PTPN13	protein tyros	34.7328799	-0.6428562	0.29130888	-2.4827627	0.01303679	0.03066754	31.3818615	0.78672467	0.30250055	3.14031537	0.00168766	0.01934506
SUPT3H	SPT3 homolc	18.8390024	-0.6587125	0.25841129	-2.5628934	0.01038039	0.02551953	20.4477301	0.77915543	0.23265576	3.35733208	0.00078699	0.01239676
ABCD3	ATP binding	118.286551	-1.0586785	0.29531833	-3.5013588	0.00046289	0.0018736	124.424884	0.72414816	0.2742921	2.83459571	0.00458837	0.03435396
OSBPL1A	oxysterol bin	78.0950641	-0.8199805	0.24479793	-3.2001268	0.00137367	0.0046892	64.8271803	0.70280155	0.25618935	2.86981839	0.00410708	0.03237371
SDHAP2	succinate del	147.905281	-1.1115055	0.25329666	-4.0493863	5.14E-05	0.00029949	41.8543292	0.68034055	0.30559375	2.81468344	0.00488253	0.03577639
XPNPEP3	X-prolyl amir	320.419149	-0.5869994	0.15112661	-3.7032133	0.00021289	0.00097083	153.739278	0.66608027	0.20866706	3.19591774	0.00139387	0.01742058
NAA35	N-alpha-acet	104.224825	-0.5297798	0.18470699	-2.8405352	0.00450379	0.0126427	100.010439	0.65052325	0.21114958	3.11613651	0.00183237	0.0202212
FAM208B	transcription	354.606289	-0.5935055	0.21185634	-2.6477893	0.008102	0.02068555	221.35102	0.62794713	0.22017187	2.91378687	0.00357074	0.02994223
MRPL30	mitochondria	83.7718103	-1.9365889	0.32223414	-5.0617373	4.15E-07	5.13E-06	150.273887	0.61368392	0.21670208	2.80779998	0.00498812	0.0362923
FTO	FTO alpha-ke	190.141504	-0.5404297	0.20025566	-2.675799	0.00745514	0.01927003	211.802206	0.59214993	0.2199967	2.80218296	0.00507581	0.03674717
SMC5	structural ma	173.372081	-0.642283	0.18958046	-3.5283203	0.00041821	0.00171638	213.799901	0.57553258	0.1845234	3.23007355	0.00123758	0.01632014

TABLE S3A FOR FIGURE 4

NAME	SIZE	NES	FDR q-val	Lower Level	Upper Level	Number of pathway genes DE in ET at 5% FD	% of DE genes in pathway
HALLMARK_FATTY_ACID_METABOLISM	130	1.7014078	0.037811887	FATTY_ACID_METABOLISM	metabolic	39	0.3
HALLMARK_ADIPOGENESIS	172	1.6664369	0.037547365	ADIPOGENESIS	development	39	0.226744186
HALLMARK_GLYCOLYSIS	147	1.6560266	0.02866926	GLYCOLYSIS	metabolic	27	0.183673469
HALLMARK_APICAL_JUNCTION	127	1.5503955	0.07132768	APICAL_JUNCTION	cellular component	27	0.212598425
HALLMARK_OXIDATIVE_PHOSPHORYLATION	183	1.4968516	0.10739506	OXIDATIVE_PHOSPHORYLATION	metabolic	38	0.207650273
HALLMARK_PEROXISOME	81	1.4731966	0.11777763	PEROXISOME	cellular component	15	0.185185185
HALLMARK_MTORC1_SIGNALING	190	1.4468399	0.13046366	MTORC1_SIGNALING	signaling	45	0.236842105
HALLMARK_XENOBIOTIC_METABOLISM	128	1.4197437	0.1478639	XENOBIOTIC_METABOLISM	metabolic	27	0.2109375
HALLMARK_INTERFERON_ALPHA_RESPONSE	89	1.4017038	0.15807928	INTERFERON_ALPHA_RESPONSE	immune	17	0.191011236
HALLMARK_CHOLESTEROL_HOMEOSTASIS	65	1.3945726	0.15155487	CHOLESTEROL_HOMEOSTASIS	metabolic	17	0.261538462
HALLMARK_COAGULATION	78	1.3851715	0.15126099	COAGULATION	immune	13	0.166666667
HALLMARK_ANDROGEN_RESPONSE	85	1.3650767	0.16553695	ANDROGEN_RESPONSE	signaling	17	0.2
HALLMARK_MYOGENESIS	116	1.312356	0.23891975	MYOGENESIS	development	30	0.25862069
HALLMARK_IL2_STAT5_SIGNALING	158	1.2709328	0.31069982	IL2_STAT5_SIGNALING	signaling	35	0.221518987
HALLMARK_APOPTOSIS	138	1.2679595	0.29717794	APOPTOSIS	pathway	34	0.246376812
HALLMARK_PROTEIN_SECRETION	91	1.2668278	0.28126517	PROTEIN_SECRETION	pathway	20	0.21978022
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	48	1.2546115	0.2918593	REACTIVE_OXYGEN_SPECIES_PATHWAY	pathway	9	0.1875
HALLMARK_INTERFERON_GAMMA_RESPONSE	177	1.2263	0.33761492	INTERFERON_GAMMA_RESPONSE	immune	33	0.186440678
HALLMARK_COMPLEMENT	156	1.2184482	0.33609262	COMPLEMENT	immune	30	0.192307692
HALLMARK_P53_PATHWAY	163	1.2127665	0.33313882	P53_PATHWAY	proliferation	30	0.18404908
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITIO	115	1.1977124	0.35076702	EPITHELIAL_MESENCHYMAL_TRANSITIO	development	29	0.252173913
HALLMARK_BILE_ACID_METABOLISM	70	1.1829697	0.3671251	BILE_ACID_METABOLISM	metabolic	16	0.228571429
HALLMARK_ESTROGEN_RESPONSE_LATE	137	1.1747813	0.36919993	ESTROGEN_RESPONSE_LATE	signaling	28	0.204379562
HALLMARK_KRAS_SIGNALING_DN	72	1.1631252	0.37978804	KRAS_SIGNALING_DN	signaling	18	0.25
HALLMARK_UV_RESPONSE_UP	125	1.1437556	0.40672106	UV_RESPONSE_UP	DNA damage	29	0.232
HALLMARK_DNA_REPAIR	143	1.1289358	0.42510718	DNA_REPAIR	DNA damage	20	0.13986014
HALLMARK_HYPOXIA	139	1.1145871	0.44307953	HYPOXIA	pathway	26	0.18705036
HALLMARK_ESTROGEN_RESPONSE_EARLY	136	1.0792629	0.50996	ESTROGEN_RESPONSE_EARLY	signaling	31	0.227941176
HALLMARK_G2M_CHECKPOINT	191	1.0562203	0.54596955	G2M_CHECKPOINT	proliferation	23	0.120418848
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	103	1.0348771	0.57947785	UNFOLDED_PROTEIN_RESPONSE	pathway	14	0.13592233
HALLMARK_SPERMATOGENESIS	66	0.9981173	0.6495295	SPERMATOGENESIS	development	12	0.181818182
HALLMARK_E2F_TARGETS	198	0.9283918	0.7953193	E2F_TARGETS	proliferation	26	0.131313131
HALLMARK_MYC_TARGETS_V1	193	0.91181886	0.80799156	MYC_TARGETS_V1	proliferation	26	0.134715026
HALLMARK_MYC_TARGETS_V2	57	0.88163126	0.84851307	MYC_TARGETS_V2	proliferation	5	0.087719298
HALLMARK_HEME_METABOLISM	186	0.86716205	0.8543608	HEME_METABOLISM	metabolic	30	0.161290323
HALLMARK_MITOTIC_SPINDLE	189	0.86634165	0.83247405	MITOTIC_SPINDLE	proliferation	23	0.121693122
HALLMARK_PI3K_AKT_MTOR_SIGNALING	91	0.8582202	0.8246167	PI3K_AKT_MTOR_SIGNALING	signaling	23	0.252747253
HALLMARK_WNT_BETA_CATENIN_SIGNALING	30	0.8327523	0.8497261	WNT_BETA_CATENIN_SIGNALING	signaling	8	0.266666667
HALLMARK_INFLAMMATORY_RESPONSE	143	0.8159283	0.8573132	INFLAMMATORY_RESPONSE	immune	26	0.181818182
HALLMARK_NOTCH_SIGNALING	26	0.7755206	0.90025306	NOTCH_SIGNALING	signaling	5	0.192307692
HALLMARK_APICAL_SURFACE	26	0.7005199	0.96442187	APICAL_SURFACE	cellular component	7	0.269230769
HALLMARK_IL6_JAK_STAT3_SIGNALING	65	0.6706368	0.9649063	IL6_JAK_STAT3_SIGNALING	immune	11	0.169230769
HALLMARK_KRAS_SIGNALING_UP	124	-1.1302108	1	KRAS_SIGNALING_UP	signaling	25	0.201612903
HALLMARK_TNFA_SIGNALING_VIA_NFKB	162	-1.087878	0.8927672	TNFA_SIGNALING_VIA_NFKB	signaling	26	0.160493827
HALLMARK_ALLOGRAFT_REJECTION	153	-1.078077	0.6434493	ALLOGRAFT_REJECTION	immune	26	0.169934641
HALLMARK_UV_RESPONSE_DN	109	-1.0154132	0.7239723	UV_RESPONSE_DN	DNA damage	24	0.220183486
HALLMARK_TGF_BETA_SIGNALING	48	-0.9414553	0.83871853	TGF_BETA_SIGNALING	signaling	7	0.145833333
HALLMARK_HEDGEHOG_SIGNALING	21	-0.8908152	0.85840535	HEDGEHOG_SIGNALING	signaling	4	0.19047619
HALLMARK_ANGIOGENESIS	23	-0.6727009	0.98062736	ANGIOGENESIS	development	4	0.173913043

TABLE S3B FOR FIGURE 4

NAME	SIZE	NES	FDR q-val	Lower Level	Upper Level	Number of pathway genes DE in PV at 5% FDR	% of genes in lower level pathway
HALLMARK_ADIPOGENESIS	172	1.6837702	0.054820366	ADIPOGENESIS	development	66	0.38372093
HALLMARK_GLYCOLYSIS	147	1.6706107	0.03275582	GLYCOLYSIS	metabolic	47	0.319727891
HALLMARK_APICAL_JUNCTION	127	1.6615859	0.023792129	APICAL_JUNCTION	cellular component	47	0.37007874
HALLMARK_INTERFERON_ALPHA_RESPONSE	89	1.604319	0.034332387	INTERFERON_ALPHA_RESPONSE	immune	32	0.359550562
HALLMARK_COAGULATION	78	1.5822302	0.036073864	COAGULATION	immune	22	0.282051282
HALLMARK_FATTY_ACID_METABOLISM	130	1.5319052	0.053782526	FATTY_ACID_METABOLISM	metabolic	55	0.423076923
HALLMARK_CHOLESTEROL_HOMEOSTASIS	65	1.5053315	0.06399146	CHOLESTEROL_HOMEOSTASIS	metabolic	27	0.415384615
HALLMARK_XENOBIOTIC_METABOLISM	128	1.4816432	0.0728615	XENOBIOTIC_METABOLISM	metabolic	44	0.34375
HALLMARK_OXIDATIVE_PHOSPHORYLATION	183	1.4538035	0.08933777	OXIDATIVE_PHOSPHORYLATION	metabolic	60	0.327868852
HALLMARK_PEROXISOME	81	1.4486157	0.08525506	PEROXISOME	cellular component	32	0.395061728
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	115	1.4472731	0.07870331	EPITHELIAL_MESENCHYMAL_TRANSITION	development	52	0.452173913
HALLMARK_MYOGENESIS	116	1.4151324	0.09862933	MYOGENESIS	development	44	0.379310345
HALLMARK_HEME_METABOLISM	186	1.3841854	0.12258336	HEME_METABOLISM	metabolic	43	0.231182796
HALLMARK_INTERFERON_GAMMA_RESPONSE	177	1.381199	0.116602294	INTERFERON_GAMMA_RESPONSE	immune	59	0.333333333
HALLMARK_IL2_STAT5_SIGNALING	158	1.351629	0.14442815	IL2_STAT5_SIGNALING	signaling	60	0.379746835
HALLMARK_MTORC1_SIGNALING	190	1.346513	0.14140692	MTORC1_SIGNALING	signaling	67	0.352631579
HALLMARK_P53_PATHWAY	163	1.2994498	0.19858885	P53_PATHWAY	proliferation	49	0.300613497
HALLMARK_BILE_ACID_METABOLISM	70	1.2975764	0.19079606	BILE_ACID_METABOLISM	metabolic	35	0.5
HALLMARK_KRAS_SIGNALING_DN	72	1.2946832	0.1855612	KRAS_SIGNALING_DN	signaling	21	0.291666667
HALLMARK_ESTROGEN_RESPONSE_EARLY	136	1.2740546	0.20825383	ESTROGEN_RESPONSE_EARLY	signaling	44	0.323529412
HALLMARK_COMPLEMENT	156	1.2571237	0.22866465	COMPLEMENT	immune	56	0.358974359
HALLMARK_PROTEIN_SECRETION	91	1.2355688	0.25597104	PROTEIN_SECRETION	pathway	37	0.406593407
HALLMARK_ESTROGEN_RESPONSE_LATE	137	1.2347167	0.2465166	ESTROGEN_RESPONSE_LATE	signaling	40	0.291970803
HALLMARK_DNA_REPAIR	143	1.2044455	0.2877144	DNA_REPAIR	DNA damage	43	0.300699301
HALLMARK_INFLAMMATORY_RESPONSE	143	1.1924322	0.29885033	INFLAMMATORY_RESPONSE	immune	47	0.328671329
HALLMARK_APOPTOSIS	138	1.1372687	0.4020361	APOPTOSIS	pathway	50	0.362318841
HALLMARK_UV_RESPONSE_UP	125	1.1277058	0.4092707	UV_RESPONSE_UP	DNA damage	46	0.368
HALLMARK_ANDROGEN_RESPONSE	85	1.1016694	0.45312434	ANDROGEN_RESPONSE	signaling	28	0.329411765
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	48	1.09504	0.4526312	REACTIVE_OXYGEN_SPECIES_PATHWAY	pathway	15	0.3125
HALLMARK_IL6_JAK_STAT3_SIGNALING	65	1.020292	0.62188655	IL6_JAK_STAT3_SIGNALING	immune	22	0.338461538
HALLMARK_HYPOXIA	139	1.0069741	0.6352679	HYPOXIA	pathway	40	0.287769784
HALLMARK_SPERMATOGENESIS	66	0.98429364	0.6728391	SPERMATOGENESIS	development	24	0.363636364
HALLMARK_PI3K_AKT_MTOR_SIGNALING	91	0.9816164	0.6587411	PI3K_AKT_MTOR_SIGNALING	signaling	35	0.384615385
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	103	0.9508966	0.7094241	UNFOLDED_PROTEIN_RESPONSE	pathway	33	0.32038835
HALLMARK_G2M_CHECKPOINT	191	0.9497592	0.69136053	G2M_CHECKPOINT	proliferation	52	0.272251309
HALLMARK_MITOTIC_SPINDLE	189	0.9236967	0.7320612	MITOTIC_SPINDLE	proliferation	57	0.301587302
HALLMARK_WNT_BETA_CATENIN_SIGNALING	30	0.9090693	0.7433333	WNT_BETA_CATENIN_SIGNALING	signaling	11	0.366666667
HALLMARK_NOTCH_SIGNALING	26	0.84303105	0.8612048	NOTCH_SIGNALING	signaling	10	0.384615385
HALLMARK_ANGIOGENESIS	23	0.81546026	0.89189714	ANGIOGENESIS	development	10	0.434782609
HALLMARK_UV_RESPONSE_DN	109	0.78969043	0.9145725	UV_RESPONSE_DN	DNA damage	35	0.321100917
HALLMARK_KRAS_SIGNALING_UP	124	0.7474986	0.9528914	KRAS_SIGNALING_UP	signaling	40	0.322580645
HALLMARK_APICAL_SURFACE	26	0.71194917	0.9694058	APICAL_SURFACE	cellular component	10	0.384615385
HALLMARK_MYC_TARGETS_V2	57	0.6694619	0.97536016	MYC_TARGETS_V2	proliferation	13	0.228070175
HALLMARK_ALLOGRAFT_REJECTION	153	-1.0539784	1	ALLOGRAFT_REJECTION	immune	58	0.379084967
HALLMARK_HEDGEHOG_SIGNALING	21	-1.0109023	1	HEDGEHOG_SIGNALING	signaling	10	0.476190476
HALLMARK_E2F_TARGETS	198	-0.99681187	0.9029633	E2F_TARGETS	proliferation	54	0.272727273
HALLMARK_MYC_TARGETS_V1	193	-0.88047856	1	MYC_TARGETS_V1	proliferation	47	0.243523316
HALLMARK_TNFA_SIGNALING_VIA_NFKB	162	-0.87274617	0.9463062	TNFA_SIGNALING_VIA_NFKB	signaling	42	0.259259259
HALLMARK_TGF_BETA_SIGNALING	48	-0.8644262	0.80427504	TGF_BETA_SIGNALING	signaling	13	0.270833333

TABLE S3C FOR FIGURE 4

NAME	SIZE	NES	FDR q-val	Lower Level	Upper Level	Number of pathway genes DE in MF at 5% FDR	% of genes in lower level pathway
HALLMARK_E2F_TARGETS	198	2.1546834	0	E2F_TARGETS	proliferation	137	0.691919192
HALLMARK_MYC_TARGETS_V1	193	2.0593772	0	MYC_TARGETS_V1	proliferation	139	0.720207254
HALLMARK_G2M_CHECKPOINT	191	1.9931763	0	G2M_CHECKPOINT	proliferation	129	0.67539267
HALLMARK_OXIDATIVE_PHOSPHORYLATION	183	1.8738438	0	OXIDATIVE_PHOSPHORYLATION	metabolic	121	0.661202186
HALLMARK_HEME_METABOLISM	186	1.8635341	0	HEME_METABOLISM	metabolic	101	0.543010753
HALLMARK_FATTY_ACID_METABOLISM	130	1.8342843	0	FATTY_ACID_METABOLISM	metabolic	90	0.692307692
HALLMARK_MTORC1_SIGNALING	190	1.820083	0	MTORC1_SIGNALING	signaling	116	0.610526316
HALLMARK_MYC_TARGETS_V2	57	1.7807441	0.000117	MYC_TARGETS_V2	proliferation	42	0.736842105
HALLMARK_ADIPOGENESIS	172	1.7658876	0.000104	ADIPOGENESIS	development	103	0.598837209
HALLMARK_GLYCOLYSIS	147	1.7646972	9.38E-05	GLYCOLYSIS	metabolic	96	0.653061224
HALLMARK_CHOLESTEROL_HOMEOSTASIS	65	1.7045	0.000824	CHOLESTEROL_HOMEOSTASIS	metabolic	36	0.553846154
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	103	1.6564009	0.002390033	UNFOLDED_PROTEIN_RESPONSE	pathway	62	0.601941748
HALLMARK_PEROXISOME	81	1.6210088	0.003139739	PEROXISOME	cellular component	44	0.543209877
HALLMARK_XENOBIOTIC_METABOLISM	128	1.590422	0.003769141	XENOBIOTIC_METABOLISM	metabolic	69	0.5390625
HALLMARK_BILE_ACID_METABOLISM	70	1.5770305	0.004575105	BILE_ACID_METABOLISM	metabolic	42	0.6
HALLMARK_DNA_REPAIR	143	1.5569813	0.006037511	DNA_REPAIR	DNA damage	77	0.538461538
HALLMARK_ESTROGEN_RESPONSE_LATE	137	1.5287629	0.008452902	ESTROGEN_RESPONSE_LATE	signaling	62	0.452554745
HALLMARK_UV_RESPONSE_UP	125	1.4194556	0.038833655	UV_RESPONSE_UP	DNA damage	62	0.496
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	48	1.3385755	0.091422826	REACTIVE_OXYGEN_SPECIES_PATHWAY	pathway	27	0.5625
HALLMARK_APICAL_JUNCTION	127	1.3195742	0.10590828	APICAL_JUNCTION	cellular component	60	0.472440945
HALLMARK_INTERFERON_ALPHA_RESPONSE	89	1.3047196	0.11625483	INTERFERON_ALPHA_RESPONSE	immune	39	0.438202247
HALLMARK_COAGULATION	78	1.3030975	0.11295785	COAGULATION	immune	29	0.371794872
HALLMARK_ESTROGEN_RESPONSE_EARLY	136	1.248276	0.18546781	ESTROGEN_RESPONSE_EARLY	signaling	67	0.492647059
HALLMARK_P53_PATHWAY	163	1.234459	0.20173946	P53_PATHWAY	proliferation	67	0.411042945
HALLMARK_MITOTIC_SPINDLE	189	1.2343034	0.1938696	MITOTIC_SPINDLE	proliferation	89	0.470899471
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	115	1.2155784	0.21863407	EPITHELIAL_MESENCHYMAL_TRANSITION	development	56	0.486956522
HALLMARK_MYOGENESIS	116	1.1875826	0.26661995	MYOGENESIS	development	51	0.439655172
HALLMARK_HYPOXIA	139	1.1767125	0.27842367	HYPOXIA	pathway	63	0.45323741
HALLMARK_SPERMATOGENESIS	66	1.1446724	0.33960944	SPERMATOGENESIS	development	38	0.575757576
HALLMARK_ANGIOGENESIS	23	1.1293957	0.36409512	ANGIOGENESIS	development	14	0.608695652
HALLMARK_INTERFERON_GAMMA_RESPONSE	177	1.0933464	0.4439761	INTERFERON_GAMMA_RESPONSE	immune	76	0.429378531
HALLMARK_PROTEIN_SECRETION	91	1.0866075	0.44756335	PROTEIN_SECRETION	pathway	41	0.450549451
HALLMARK_APOPTOSIS	138	1.0091037	0.646069	APOPTOSIS	pathway	56	0.405797101
HALLMARK_WNT_BETA_CATENIN_SIGNALING	30	0.951642	0.7780404	WNT_BETA_CATENIN_SIGNALING	signaling	16	0.533333333
HALLMARK_PI3K_AKT_MTOR_SIGNALING	91	0.9299124	0.81221837	PI3K_AKT_MTOR_SIGNALING	signaling	48	0.527472527
HALLMARK_KRAS_SIGNALING_DN	72	0.90283114	0.8560608	KRAS_SIGNALING_DN	signaling	28	0.388888889
HALLMARK_IL2_STAT5_SIGNALING	158	0.899449	0.83995366	IL2_STAT5_SIGNALING	signaling	82	0.518987342
HALLMARK_HEDGEHOG_SIGNALING	21	0.8669753	0.88709676	HEDGEHOG_SIGNALING	signaling	8	0.380952381
HALLMARK_NOTCH_SIGNALING	26	0.85548687	0.8872063	NOTCH_SIGNALING	signaling	8	0.307692308
HALLMARK_COMPLEMENT	156	0.81541675	0.93541306	COMPLEMENT	immune	66	0.423076923
HALLMARK_IL6_JAK_STAT3_SIGNALING	65	0.77015305	0.9759494	IL6_JAK_STAT3_SIGNALING	immune	23	0.353846154
HALLMARK_ANDROGEN_RESPONSE	85	0.74873525	0.9777931	ANDROGEN_RESPONSE	signaling	42	0.494117647
HALLMARK_INFLAMMATORY_RESPONSE	143	0.74842304	0.9552385	INFLAMMATORY_RESPONSE	immune	58	0.405594406
HALLMARK_ALLOGRAFT_REJECTION	153	0.73908633	0.94275194	ALLOGRAFT_REJECTION	immune	73	0.477124183
HALLMARK_KRAS_SIGNALING_UP	124	0.68830496	0.96209025	KRAS_SIGNALING_UP	signaling	57	0.459677419
HALLMARK_UV_RESPONSE_DN	109	-1.4277561	0.083994046	UV_RESPONSE_DN	DNA damage	55	0.504587156
HALLMARK_TNFA_SIGNALING_VIA_NFKB	162	-1.3566213	0.06609435	TNFA_SIGNALING_VIA_NFKB	signaling	59	0.364197531
HALLMARK_TGF_BETA_SIGNALING	48	-0.8572026	1	TGF_BETA_SIGNALING	signaling	17	0.354166667
HALLMARK_APICAL_SURFACE	26	-0.78979635	0.91649055	APICAL_SURFACE	cellular component	12	0.461538462

TABLE S4A FOR FIGURE 5

BaselineModel	Prob MF	Prob CTRL	Prob ET	Prob PV	Predicted Label	True Label
SC1	0.49946851	0.21088983	0.28964055	1.11E-06	MF	CTRL
SC2	0.17313416	0.6221824	0.20468126	2.19E-06	CTRL	CTRL
SC3	0.31755445	0.52021055	0.16223334	1.66E-06	CTRL	CTRL
SC4	0.46839901	0.40929621	0.12230357	1.21E-06	MF	CTRL
SC5	0.34174147	0.27051068	0.38774631	1.54E-06	ET	CTRL
SC6	0.35311209	0.49435478	0.15253159	1.55E-06	CTRL	CTRL
SC7	0.04384405	0.70093254	0.25522044	2.97E-06	CTRL	CTRL
SC8	0.54798584	0.34972082	0.10229235	9.93E-07	MF	CTRL
SC9	0.31755445	0.52021055	0.16223334	1.66E-06	CTRL	CTRL
SC10	0.03760596	0.70348856	0.25890244	3.04E-06	CTRL	CTRL
SC11	0.47906753	0.21880723	0.30212408	1.16E-06	MF	CTRL
SC12	0.49946851	0.21088983	0.28964055	1.11E-06	MF	CTRL
SC13	0.17497578	0.62131113	0.20370814	4.95E-06	CTRL	CTRL
SC14	0.31945493	0.52125581	0.15928559	3.68E-06	CTRL	CTRL
SC15	0.46970483	0.4114475	0.11884504	2.63E-06	MF	CTRL
SC16	0.18093936	0.59698477	0.22207403	1.85E-06	CTRL	CTRL
SC17	0.35491859	0.49571503	0.14936297	3.41E-06	CTRL	CTRL
SC18	0.04466564	0.69488913	0.26043825	6.99E-06	CTRL	CTRL
SC19	0.54882815	0.35224905	0.09892066	2.15E-06	MF	CTRL
SC20	0.31945493	0.52125581	0.15928559	3.68E-06	CTRL	CTRL
SC21	0.03833779	0.69679827	0.26485675	7.18E-06	CTRL	CTRL
SE1	0.5133069	9.97E-08	0.07940075	0.40729225	MF	ET
SE2	0.07847406	2.52E-07	0.29490564	0.62662005	PV	ET
SE3	0.39780089	6.40E-06	0.60219245	2.61E-07	ET	ET
SE4	0.64036061	1.36E-08	0.05452121	0.30511817	MF	ET
SE5	0.24503946	5.39E-06	0.75495501	1.34E-07	ET	ET
SE6	0.73483467	8.84E-09	0.03787395	0.22729137	MF	ET
SE7	0.0569035	9.17E-08	0.21978776	0.72330865	PV	ET
SE8	0.20143601	5.25E-08	0.1561111	0.64245285	PV	ET
SE9	0.71734113	8.88E-09	0.01768409	0.26497478	MF	ET
SE10	0.55541867	1.86E-08	0.0706567	0.37392462	MF	ET
SE11	0.71449101	3.68E-06	0.28550516	1.52E-07	MF	ET
SE12	0.40626859	2.78E-08	0.04514816	0.54858322	PV	ET
SE13	0.03443013	9.28E-06	0.96556047	1.26E-07	ET	ET
SE14	0.16179602	5.93E-08	0.16945659	0.66874734	PV	ET
SE15	0.01114379	0.91532818	0.07352639	1.63E-06	CTRL	ET
SE16	0.08296214	0.86011895	0.05691423	4.67E-06	CTRL	ET
SE17	0.49128388	5.14E-07	0.14742804	0.36128756	MF	ET
SE18	0.27042245	3.55E-06	0.72957305	9.54E-07	ET	ET
SE19	0.8007541	1.18E-06	0.1992444	3.23E-07	MF	ET
SE20	0.5581823	3.29E-06	0.44181327	1.15E-06	MF	ET
SE21	0.26716076	0.70384874	0.02898551	5.00E-06	CTRL	ET
SE22	0.37034437	6.30E-07	0.18874882	0.44090618	PV	ET
SE23	0.15582006	8.25E-07	0.27330065	0.57087846	PV	ET
SE24	0.19529814	0.7717507	0.0329457	5.46E-06	CTRL	ET
SP1	0.1876452	7.24E-07	0.16399186	0.64836222	PV	PV
SP2	0.03672485	9.07E-07	0.34467399	0.61860025	PV	PV
SP3	0.29063342	6.36E-07	0.13691815	0.57244779	PV	PV
SP4	0.12493024	3.26E-07	0.24192746	0.63314198	PV	PV
SP5	0.12493024	3.26E-07	0.24192746	0.63314198	PV	PV
SP6	0.10866825	3.33E-07	0.24731745	0.64401396	PV	PV
SP7	0.18649136	3.01E-07	0.22247073	0.59103761	PV	PV
SP8	0.33541885	2.43E-07	0.1784552	0.48612571	PV	PV
SP9	0.43237194	2.21E-07	0.26628041	0.30134743	MF	PV
SP10	0.51136554	1.90E-07	0.22800492	0.26062936	MF	PV
SP11	0.32110412	2.66E-07	0.32101923	0.35787638	PV	PV
SP12	0.46719368	1.93E-07	0.14124335	0.39156277	MF	PV
SP13	0.13381786	3.22E-07	0.23903662	0.6271452	PV	PV
SP14	0.49151576	1.98E-07	0.23758414	0.27089991	MF	PV
SP15	0.39388407	2.36E-07	0.28509173	0.32102397	MF	PV
SP16	0.65517739	5.88E-08	0.13342462	0.21139793	MF	PV
SP17	0.64173796	5.10E-08	0.22477568	0.1334863	MF	PV
SP18	0.59584868	6.42E-08	0.14841432	0.25573694	MF	PV
SP19	0.3539788	7.01E-08	0.33937712	0.30664402	MF	PV
SP20	0.56481869	6.67E-08	0.15558949	0.27959176	MF	PV
SP21	0.26208387	7.71E-08	0.19761741	0.54029864	PV	PV
SP22	0.85350766	3.37E-08	0.07022668	0.07626562	MF	PV
SP23	0.18897532	7.09E-08	0.36663169	0.44439292	PV	PV
SP24	0.94518877	1.64E-08	0.03142179	0.02338942	MF	PV
SP25	0.41065936	6.78E-08	0.32190246	0.26743812	MF	PV
SP26	0.6831077	5.60E-08	0.12578675	0.19110549	MF	PV
SP27	0.89785627	1.99E-08	0.07679965	0.02534406	MF	PV
SP28	0.24371938	2.01E-07	0.16560077	0.59067965	PV	PV
SP29	0.06688324	3.02E-07	0.23764578	0.69547068	PV	PV
SP30	0.03136111	3.52E-07	0.50175486	0.46688369	ET	PV
SP31	0.32354796	1.90E-07	0.29256421	0.38388764	PV	PV
SP32	0.3789129	1.70E-07	0.26386632	0.35722061	MF	PV
SP33	0.18709416	2.45E-07	0.3703878	0.4425178	PV	PV
SM1	0.59003146	8.88E-08	0.07666674	0.33330172	MF	MF
SM2	0.62277003	6.84E-07	0.3772289	3.93E-07	MF	MF
SM3	0.59774643	1.00E-07	0.15974318	0.24251028	MF	MF
SM4	0.13096469	0.75053565	0.11849661	3.05E-06	CTRL	MF
SM5	0.30644062	1.76E-07	0.1469833	0.5465759	PV	MF
SM6	0.55834704	1.12E-07	0.17759669	0.26405615	MF	MF
SM7	0.42624715	1.91E-08	0.10591047	0.46784237	PV	MF
SM8	0.76339209	3.01E-07	0.23660761	8.87E-09	MF	MF
SM9	0.28590108	2.42E-08	0.13001031	0.58408859	PV	MF
SM10	0.24174699	0.60809925	0.15015272	1.05E-06	CTRL	MF
SM11	0.11050506	5.04E-08	0.33109484	0.55840005	PV	MF
SM12	0.37651273	3.37E-08	0.23844572	0.38504152	PV	MF
SM13	0.44006858	3.00E-08	0.21509822	0.34483317	MF	MF
SM14	0.44006858	3.00E-08	0.21509822	0.34483317	MF	MF
SM15	0.22567883	0.62150695	0.15281315	1.07E-06	CTRL	MF
SM16	0.48369858	2.75E-08	0.19892789	0.3173735	MF	MF
SM17	0.70073177	9.68E-09	0.0566706	0.24259762	MF	MF
SM18	0.06551786	0.79623597	0.13824338	2.80E-06	CTRL	MF
SM19	0.04734544	1.47E-07	0.3438354	0.60881902	PV	MF
SM20	0.55377556	1.97E-07	0.08384442	0.36237983	MF	MF
SM21	0.19935267	2.15E-07	0.1402618	0.66038532	PV	MF
SM22	0.3640921	0.58638587	0.04952061	1.42E-06	CTRL	MF
SM23	0.4290058	2.16E-07	0.10497753	0.46601645	PV	MF
SM24	0.42155947	9.94E-06	0.57843035	2.40E-07	ET	MF
SM25	0.1145962	9.41E-06	0.88539398	4.05E-07	ET	MF
SM26	0.61639163	1.69E-07	0.15541019	0.22819801	MF	MF
SM27	0.40541212	0.50276348	0.0918235	8.98E-07	CTRL	MF
SM28	0.49454829	1.91E-07	0.20152285	0.30392868	MF	MF
SM29	0.4852824	0.48101284	0.03370383	9.34E-07	MF	MF
SM30	0.2291585	2.04E-07	0.29555227	0.47528902	PV	MF
SM31	0.16312404	1.66E-07	0.14075965	0.69611614	PV	MF
SM32	0.63011177	1.09E-06	0.36988707	7.04E-08	MF	MF
SM33	0.12972092	1.65E-07	0.14369092	0.72658799	PV	MF
SM34	0.58174048	1.12E-07	0.08041231	0.33784711	MF	MF
SM35	0.46969907	0.46602127	0.06427878	8.77E-07	MF	MF
SM36	0.30347741	1.56E-07	0.12379673	0.5727257	PV	MF
SM37	0.15128236	1.66E-07	0.14187358	0.70684389	PV	MF
SM38	0.27699257	1.95E-07	0.28242403	0.44058321	PV	MF
SM39	0.42680357	1.39E-07	0.10565562	0.46754068	PV	MF
SM40	0.48749718	1.43E-06	0.51250129	1.03E-07	ET	MF
SM41	0.56658718	1.38E-07	0.18011823	0.25329446	MF	MF
SM42	0.51554243	1.24E-07	0.09147803	0.39297942	MF	MF

TABLE S4B FOR FIGURE 5

LassoEntireTranscriptome	Prob MF	Prob CTRL	Prob ET	Prob PV	Predicted Label	True Label
SC1	0.07357186	0.79419263	0.07178174	0.06045377	CTRL	CTRL
SC2	0.04612924	0.69264823	0.17171096	0.08951157	CTRL	CTRL
SC3	0.07811852	0.660229	0.14277087	0.11888161	CTRL	CTRL
SC4	0.10010098	0.82995907	0.06497066	0.00496929	CTRL	CTRL
SC5	0.12307686	0.66549992	0.15445039	0.05697283	CTRL	CTRL
SC6	0.12844862	0.45189218	0.26411421	0.15554499	CTRL	CTRL
SC7	0.01320116	0.93235173	0.0408182	0.01362892	CTRL	CTRL
SC8	0.06403348	0.81034699	0.0633857	0.06223383	CTRL	CTRL
SC9	0.06533504	0.66300753	0.16100064	0.11065679	CTRL	CTRL
SC10	0.23146097	0.43564296	0.16479959	0.16809649	CTRL	CTRL
SC11	0.0209689	0.95313242	0.01636771	0.00953098	CTRL	CTRL
SC12	0.06581626	0.69235801	0.19491178	0.04691395	CTRL	CTRL
SC13	0.04343552	0.71247377	0.10864172	0.13544899	CTRL	CTRL
SC14	0.08320871	0.4242659	0.21313927	0.27938612	CTRL	CTRL
SC15	0.01495888	0.86306414	0.11193103	0.01004596	CTRL	CTRL
SC16	0.07770854	0.63865331	0.16185016	0.12178799	CTRL	CTRL
SC17	0.08611759	0.48766773	0.22562694	0.20058774	CTRL	CTRL
SC18	0.29671582	0.22241633	0.25905852	0.22180933	MF	CTRL
SC19	0.15501736	0.692294	0.13082275	0.0218659	CTRL	CTRL
SC20	0.19226571	0.63151862	0.13519102	0.04102465	CTRL	CTRL
SC21	0.11133302	0.71032903	0.1103095	0.06802846	CTRL	CTRL
SE1	0.19996796	0.21502081	0.30264332	0.28236791	ET	ET
SE2	0.2866448	0.05511718	0.12132675	0.53691127	PV	ET
SE3	0.02204216	0.01051677	0.08255454	0.88488653	PV	ET
SE4	0.00852019	0.00625849	0.42095697	0.56426437	PV	ET
SE5	0.24624808	0.04019885	0.29437155	0.41918153	PV	ET
SE6	0.09520893	0.05752465	0.58459243	0.26267399	ET	ET
SE7	0.05994874	0.23518538	0.42489599	0.27996989	ET	ET
SE8	0.34550082	0.00025328	0.00093314	0.65331277	PV	ET
SE9	0.04682479	0.59867758	0.11205565	0.24244198	CTRL	ET
SE10	0.15254686	0.15712326	0.34050812	0.34982176	PV	ET
SE11	0.14832344	0.04995711	0.10517022	0.69654923	PV	ET
SE12	0.03519869	0.27949403	0.55445872	0.13084856	ET	ET
SE13	0.04807492	0.03254274	0.64829381	0.27108852	ET	ET
SE14	0.18366989	0.08164519	0.49534584	0.23933908	ET	ET
SE15	0.01281414	0.02225936	0.29709025	0.66783626	PV	ET
SE16	0.01887296	0.06544241	0.86919418	0.04649046	ET	ET
SE17	0.13162121	0.10917115	0.3586427	0.40056493	PV	ET
SE18	0.04001409	0.06979393	0.74846167	0.14173031	ET	ET
SE19	0.04184098	0.1506486	0.68038471	0.12712571	ET	ET
SE20	0.02771808	0.06654651	0.37558325	0.53015217	PV	ET
SE21	0.28117238	0.09114888	0.41312479	0.21455395	ET	ET
SE22	0.06602456	0.72035592	0.16724456	0.04637496	CTRL	ET
SE23	0.04469749	0.12206715	0.46125942	0.37197595	ET	ET
SE24	0.07446199	0.04568588	0.62651255	0.25333958	ET	ET
SP1	0.03668847	0.0211452	0.0446094	0.89755693	PV	PV
SP2	0.33715772	0.00010903	0.11379649	0.54893677	PV	PV
SP3	0.35790373	0.19086844	0.36146422	0.08976362	ET	PV
SP4	0.01700819	0.00431399	0.02340773	0.95527009	PV	PV
SP5	0.09821527	0.00757737	0.03946439	0.85474298	PV	PV
SP6	0.03238098	0.03313465	0.40659243	0.52789194	PV	PV
SP7	0.13829992	0.00010538	0.00918822	0.85240649	PV	PV
SP8	0.03396235	0.03313001	0.61579602	0.31711162	ET	PV
SP9	0.3834045	0.11970206	0.16216939	0.33472404	MF	PV
SP10	0.15958077	0.04635514	0.13557419	0.6584899	PV	PV
SP11	0.0858162	0.06459505	0.17752515	0.67206359	PV	PV
SP12	0.04788229	0.03121279	0.1526225	0.76828242	PV	PV
SP13	0.38173428	0.15136627	0.26317436	0.2037251	MF	PV
SP14	0.68577221	0.00013043	0.02834159	0.28575578	MF	PV
SP15	0.0587266	0.05917421	0.19457307	0.68752612	PV	PV
SP16	0.40697138	0.07420896	0.19111005	0.3277096	MF	PV
SP17	0.87690757	8.92E-05	0.01432512	0.10867807	MF	PV
SP18	0.08474341	0.21313249	0.45661725	0.24550686	ET	PV
SP19	0.03115224	0.24101067	0.09449864	0.63333845	PV	PV
SP20	0.10015363	0.05678425	0.22459931	0.61846282	PV	PV
SP21	0.52476863	0.00013275	0.03218335	0.44291527	MF	PV
SP22	0.07900261	0.06139403	0.40547877	0.4541246	PV	PV
SP23	0.29629959	0.05038667	0.26011446	0.39319929	PV	PV
SP24	0.03743081	0.03032527	0.56979059	0.36245333	ET	PV
SP25	0.05028259	0.01496752	0.05177761	0.88297228	PV	PV
SP26	0.50788296	0.00017877	0.0583705	0.43356777	MF	PV
SP27	0.17993075	0.30160504	0.27527181	0.24319241	CTRL	PV
SP28	0.47351048	5.76E-05	0.00978477	0.51664718	PV	PV
SP29	0.39300875	7.42E-05	0.05209467	0.55482243	PV	PV
SP30	0.16847079	0.05469597	0.49450727	0.28232597	ET	PV
SP31	0.05862645	0.08309406	0.3407866	0.5174929	PV	PV
SP32	0.06307594	0.14195859	0.33953868	0.45542679	PV	PV
SP33	0.12067244	0.08724957	0.23759679	0.5544812	PV	PV
SM1	0.34547349	0.05044775	0.14938404	0.45469472	PV	MF
SM2	0.72527848	1.06E-05	0.00220486	0.27250609	MF	MF
SM3	0.96149569	2.02E-05	0.00324471	0.03523942	MF	MF
SM4	0.9740871	1.97E-05	0.00319674	0.02269647	MF	MF
SM5	0.99345637	1.26E-05	2.11E-06	0.00652894	MF	MF
SM6	0.1832057	6.45E-05	0.00021155	0.81651821	PV	MF
SM7	0.88082121	6.32E-05	0.00866603	0.11044953	MF	MF
SM8	0.90848114	7.07E-05	0.0167937	0.07465444	MF	MF
SM9	0.55527788	0.05566364	0.26909537	0.11996312	MF	MF
SM10	0.91005375	2.92E-05	0.0043746	0.08554243	MF	MF
SM11	0.85034614	4.35E-05	0.00630903	0.14330135	MF	MF
SM12	0.94801843	1.92E-05	0.00528342	0.04667898	MF	MF
SM13	0.34094634	0.00038295	0.05391199	0.60475872	PV	MF
SM14	0.98356214	5.27E-05	0.00543424	0.01095097	MF	MF
SM15	0.92308727	6.68E-05	0.01341379	0.06343213	MF	MF
SM16	0.97496182	2.39E-05	0.00529473	0.01971952	MF	MF
SM17	0.10710982	0.05858161	0.56839148	0.26591708	ET	MF
SM18	0.91155368	4.44E-05	0.00778837	0.08061353	MF	MF
SM19	0.96344501	3.75E-05	0.00884464	0.02767286	MF	MF
SM20	0.14725811	0.07227272	0.22282827	0.5576409	PV	MF
SM21	0.12597553	7.46E-05	0.04856829	0.82538158	PV	MF
SM22	0.17210884	0.14773467	0.49851445	0.18164204	ET	MF
SM23	0.09764916	0.31705449	0.19208867	0.39320768	PV	MF
SM24	0.10722018	0.13444123	0.32297367	0.43536493	PV	MF
SM25	0.07184981	0.04924395	0.21672412	0.66218213	PV	MF
SM26	0.64572897	0.0001273	0.01597899	0.33816475	MF	MF
SM27	0.8994235	9.67E-05	0.0096464	0.0908334	MF	MF
SM28	0.40436848	0.11202924	0.22499662	0.25860566	MF	MF
SM29	0.15763516	0.0882717	0.5128225	0.24127065	ET	MF
SM30	0.20169913	0.14182315	0.45857332	0.1979044	ET	MF
SM31	0.96424113	5.65E-05	0.0087693	0.02693302	MF	MF
SM32	0.71300954	4.84E-05	0.00538926	0.28155282	MF	MF
SM33	0.91791998	0.00011718	0.01321452	0.06874832	MF	MF
SM34	0.93484707	4.87E-05	0.00551097	0.05959326	MF	MF
SM35	0.97186554	2.51E-05	0.00285369	0.02525568	MF	MF
SM36	0.2032858	0.12274248	0.36585575	0.30811598	ET	MF
SM37	0.97406502	2.04E-05	0.00464503	0.02126959	MF	MF
SM38	0.95364427	4.47E-05	0.00854102	0.03777004	MF	MF
SM39	0.84730496	3.48E-05	0.00932884	0.14333145	MF	MF
SM40	0.94773462	2.11E-05	0.005377	0.04686732	MF	MF
SM41	0.79599919	7.35E-05	0.01738577	0.1865415	MF	MF
SM42	0.98837833	7.78E-06	0.00291253	0.00870136	MF	MF

TABLE S4C FOR FIGURE 5

LassoProgressiveTranscriptome	Prob.MF	Prob.CTRL	Prob.ET	Prob.PV	Predicted.Label	True.Label
SC1	0.11953611	0.71843336	0.12143054	0.04059999	CTRL	CTRL
SC2	0.08829207	0.65287485	0.15421191	0.10462117	CTRL	CTRL
SC3	0.08433741	0.73842337	0.10393016	0.07330907	CTRL	CTRL
SC4	0.09262521	0.76346086	0.1036935	0.04022043	CTRL	CTRL
SC5	0.01932603	0.92544463	0.04281476	0.01241459	CTRL	CTRL
SC6	0.07149869	0.63405687	0.19031315	0.1041313	CTRL	CTRL
SC7	0.06848145	0.64477237	0.16156917	0.12517702	CTRL	CTRL
SC8	0.05516125	0.49276264	0.42541846	0.02665766	CTRL	CTRL
SC9	0.04431217	0.72719871	0.20868197	0.01980714	CTRL	CTRL
SC10	0.05660253	0.92176886	0.0168756	0.00475301	CTRL	CTRL
SC11	0.03192008	0.94463998	0.01238369	0.01105626	CTRL	CTRL
SC12	0.09145398	0.73079533	0.10511631	0.07263438	CTRL	CTRL
SC13	0.05637921	0.66260818	0.19410558	0.08690704	CTRL	CTRL
SC14	0.09888137	0.57919413	0.20244362	0.11948088	CTRL	CTRL
SC15	0.02540633	0.92322874	0.03063168	0.02073325	CTRL	CTRL
SC16	0.0213214	0.81374254	0.14348819	0.02144787	CTRL	CTRL
SC17	0.11345581	0.42883025	0.24227892	0.21543503	CTRL	CTRL
SC18	0.11566515	0.45400666	0.22440045	0.20592775	CTRL	CTRL
SC19	0.11988058	0.57194328	0.21554779	0.09262835	CTRL	CTRL
SC20	0.21795155	0.21085206	0.32141498	0.24978142	ET	CTRL
SC21	0.08885575	0.44802453	0.31105044	0.15206928	CTRL	CTRL
SE1	0.19185468	0.09191865	0.17409445	0.54213222	PV	ET
SE2	0.01399148	0.01972524	0.4478244	0.51845888	PV	ET
SE3	0.05722022	0.08559313	0.38029826	0.47688838	PV	ET
SE4	0.08698831	0.53854449	0.24628877	0.12817843	CTRL	ET
SE5	0.09186676	0.25435415	0.57695489	0.07682421	ET	ET
SE6	0.09663369	0.10049923	0.4591548	0.34371229	ET	ET
SE7	0.40474693	0.08526226	0.19384274	0.31614807	MF	ET
SE8	0.06922732	0.02250272	0.75734628	0.15092368	ET	ET
SE9	0.01572648	0.16252891	0.80037845	0.02136616	ET	ET
SE10	0.17562682	0.04173692	0.12670612	0.65593014	PV	ET
SE11	0.14660458	0.06789208	0.52808015	0.2574232	ET	ET
SE12	0.0463228	0.10029021	0.63177597	0.22161103	ET	ET
SE13	0.04255582	0.11662293	0.41440928	0.42641197	PV	ET
SE14	0.2185422	0.25760322	0.21117579	0.3126788	PV	ET
SE15	0.01057582	0.08374178	0.84393473	0.06174767	ET	ET
SE16	0.00449068	0.00192856	0.02590285	0.96767791	PV	ET
SE17	0.01903182	0.01728913	0.64316002	0.32051903	ET	ET
SE18	0.00572101	0.00331862	0.35138092	0.63957946	PV	ET
SE19	0.00325131	0.95466921	0.02587854	0.01620094	CTRL	ET
SE20	0.03759891	0.18315572	0.643838	0.13540738	ET	ET
SE21	0.18645176	0.00017396	0.00090842	0.81246586	PV	ET
SE22	0.15360613	0.10199412	0.30342456	0.4409752	PV	ET
SE23	0.16073993	0.10342331	0.50532766	0.2305091	ET	ET
SE24	0.22586081	0.10030674	0.34199576	0.33183669	ET	ET
SP1	0.19221145	0.07147305	0.28434365	0.45197185	PV	PV
SP2	0.02349815	0.03222311	0.59471882	0.34955993	ET	PV
SP3	0.47920737	0.00013532	0.07246738	0.44818993	MF	PV
SP4	0.09621192	0.44765593	0.22459326	0.23153889	CTRL	PV
SP5	0.03798588	0.02294882	0.10783982	0.83122548	PV	PV
SP6	0.40109412	0.11779933	0.20819719	0.27290936	MF	PV
SP7	0.69185283	0.0001112	0.02638358	0.28165239	MF	PV
SP8	0.03316713	0.03143724	0.18277648	0.75261915	PV	PV
SP9	0.82129962	7.56E-05	0.01070769	0.16791711	MF	PV
SP10	0.14681444	0.10440946	0.4797281	0.269048	ET	PV
SP11	0.17276924	0.04793645	0.17414574	0.60514857	PV	PV
SP12	0.01975231	0.02220021	0.2360549	0.72199259	PV	PV
SP13	0.41755415	0.11270952	0.37697671	0.09275962	MF	PV
SP14	0.20120179	0.00010916	0.08443416	0.71425489	PV	PV
SP15	0.00300561	0.00268999	0.01324519	0.98105921	PV	PV
SP16	0.41340682	0.06240253	0.28130433	0.24288633	MF	PV
SP17	0.01836935	0.00275444	0.00949489	0.96938132	PV	PV
SP18	0.00322868	0.00316688	0.0379184	0.95568604	PV	PV
SP19	0.01308491	0.0674486	0.25035323	0.66911326	PV	PV
SP20	0.1672399	0.00018564	0.01899022	0.81358424	PV	PV
SP21	0.04038372	0.03620113	0.37041285	0.55300231	PV	PV
SP22	0.16717994	0.09799465	0.28079194	0.45403348	PV	PV
SP23	0.29694351	0.11103733	0.22991483	0.36210433	PV	PV
SP24	0.14308173	0.06505827	0.39927802	0.39258198	ET	PV
SP25	0.03646578	0.05231417	0.24461009	0.66660996	PV	PV
SP26	0.06712364	0.16173878	0.47108922	0.30004837	ET	PV
SP27	0.78295956	5.23E-05	0.0079824	0.20900575	MF	PV
SP28	0.433091	0.00015938	0.0892332	0.47751643	PV	PV
SP29	0.03153207	0.17701039	0.12512352	0.66633403	PV	PV
SP30	0.29065578	0.02885029	0.18209062	0.49840332	PV	PV
SP31	0.48891873	0.00020191	0.04341581	0.46746355	MF	PV
SP32	0.16717628	0.05266671	0.19057934	0.58957766	PV	PV
SP33	0.07874545	0.07579153	0.20123648	0.64422654	PV	PV
SM1	0.96081643	6.22E-05	0.00685308	0.03226834	MF	MF
SM2	0.23812451	0.00022218	0.04288461	0.7187687	PV	MF
SM3	0.97125321	3.20E-05	0.00599066	0.02272414	MF	MF
SM4	0.66640731	0.05494684	0.16370925	0.1149366	MF	MF
SM5	0.97348601	0.00013582	0.00814972	0.01822845	MF	MF
SM6	0.19344042	0.05171374	0.25055417	0.50429166	PV	MF
SM7	0.88022239	6.56E-05	0.01355858	0.10615341	MF	MF
SM8	0.50017467	6.15E-06	0.00109647	0.49872271	MF	MF
SM9	0.7219964	0.05344073	0.11843809	0.10612478	MF	MF
SM10	0.99138657	1.03E-05	0.0023857	0.00621745	MF	MF
SM11	0.86817871	9.61E-05	0.01348832	0.11823685	MF	MF
SM12	0.38025776	9.61E-05	0.05128223	0.56836389	PV	MF
SM13	0.82106053	9.70E-05	0.01215689	0.16668555	MF	MF
SM14	0.25972913	0.03295531	0.09528282	0.61203273	PV	MF
SM15	0.38344599	0.00010386	0.03489953	0.58155062	PV	MF
SM16	0.9278259	2.44E-05	0.00447889	0.06767077	MF	MF
SM17	0.96537233	2.58E-05	0.00349486	0.03110696	MF	MF
SM18	0.91336306	2.33E-05	0.0043131	0.08230051	MF	MF
SM19	0.17109463	0.07291532	0.70670552	0.04928453	ET	MF
SM20	0.8758595	0.00012594	0.01424829	0.10976627	MF	MF
SM21	0.94931514	5.63E-05	0.00588428	0.04474428	MF	MF
SM22	0.91070855	9.39E-05	0.00982785	0.0793697	MF	MF
SM23	0.66332026	5.42E-05	0.00776878	0.32885678	MF	MF
SM24	0.99306956	1.33E-05	0.00191682	0.00500031	MF	MF
SM25	0.98202752	3.00E-05	0.00148722	0.01645529	MF	MF
SM26	0.09562794	0.08234163	0.6721281	0.14990233	ET	MF
SM27	0.9651225	6.69E-05	0.00466513	0.03014546	MF	MF
SM28	0.17432784	0.03004939	0.23250196	0.56312082	PV	MF
SM29	0.15556784	0.22649225	0.56815492	0.04978499	ET	MF
SM30	0.97557981	8.54E-06	4.76E-06	0.02440689	MF	MF
SM31	0.13202501	0.12187325	0.21700943	0.52909231	PV	MF
SM32	0.94511522	5.33E-05	0.01034426	0.0444872	MF	MF
SM33	0.97357977	2.76E-05	0.00656565	0.01982698	MF	MF
SM34	0.89163289	5.47E-05	0.01093877	0.09737362	MF	MF
SM35	0.88896737	4.24E-05	0.00825228	0.10273797	MF	MF
SM36	0.18829398	0.09058049	0.54093635	0.18018919	ET	MF
SM37	0.90303977	7.67E-05	0.01798562	0.07889789	MF	MF
SM38	0.93539392	4.24E-05	0.00680592	0.05775779	MF	MF
SM39	0.1915096	0.3249661	0.27248031	0.211044	CTRL	MF
SM40	0.9738553	3.88E-05	0.00416779	0.02193814	MF	MF
SM41	0.20355817	0.08323456	0.42714689	0.28606039	ET	MF
SM42	0.86830344	8.70E-05	0.01480884	0.11680076	MF	MF