bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

***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 externally validated prediction in MPNs.

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Predicting disease progression remains a particularly challenging endeavor in chronic degenerative disorders and cancer, thus limiting early detection, risk stratification, and preventive interventions. Here, profiling the spectrum of chronic myeloproliferative neoplasms (MPNs) as a model, we identify the blood platelet transcriptome as a proxy for highly sensitive progression biomarkers that also enables prediction of advanced disease 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 time-separated cohorts of patients diagnosed with one of three MPN subtypes at sample acquisition – essential thrombocythemia, ET (n=24), polycythemia vera, PV (n=33), and primary or post ET/PV secondary myelofibrosis, MF (n=42), and healthy donors (n=21). The MPN platelet transcriptome reveals an incremental molecular reprogramming that is independent of patient driver mutation status or therapy and discriminates each clinical phenotype. Leveraging this dataset that shows a characteristic progressive expression gradient across MPN, we develop a machine learning model (Lasso-penalized regression) and predict advanced subtype MF at high accuracy and under two conditions of external validation: i) temporal: our two Stanford cohorts, AUC-ROC of 0.96; and ii) geographical: independently published data of an additional n=25 MF and n=46 healthy donors, AUC-ROC of 0.97). Lasso-derived signatures offer a robust core set of < 5 MPN transcriptome markers that are progressive in expression. 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 proof of principle for disease risk stratification and progression beyond genetic data alone, with potential utility in other progressive disorders.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

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

Blood platelets play critical roles in multiple processes and diseases (Rondina et al., 2013; Weyrich, 2014; Weyrich and Zimmerman, 2013), from their traditional function in hemostasis and wound healing to inflammation, immunity, cancer metastasis, and angiogenesis. Platelets originate from bone marrow

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

precursor megakaryocytes as anucleate fragments with a distinctive discoid shape. Platelets contain a complex transcriptional landscape of messenger RNAs (mRNAs), unspliced pre-mRNAs, rRNAs, tRNAs and microRNAs (Rowley et al., 2012; Schubert et al., 2014; Simon et al., 2014; Weyrich, 2014). Most platelet RNA expression results from the transcription of nuclear DNA in the megakaryocyte (Davizon-Castillo et al., 2020; Rowley et al., 2012), and thus reflects the status of the megakaryocyte at the time of platelet release into the circulation, as well as subsequent splicing events, and selective packaging and inter-cellular RNA transfer. There is emerging evidence (Best et al., 2015; Campbell et al., 2018; Clancy and Freedman, 2016; Cunin et al., 2019; Middleton et al., 2019) that the molecular signature of platelets may be changed in disease conditions where these processes are altered, including via inter-cellular transfer (Clancy and Freedman, 2016; Cunin et al., 2019) of cytosolic RNA. In the context of MPNs, the platelet transcriptome therefore not only represents a critical biomarker of megakaryocytic activity (Gilles et al., 2012; Krause and Crispino, 2013; Wen et al., 2016; Wen et al., 2015; Woods et al., 2019), but also provides a snapshot of the underlying hemostatic, thrombotic, and inflammatory derangements associated with these hematologic neoplasms and the potential impact of treatment (Woods et al., 2019).

To date, no one composite study (Krishnan et al., 2017) (Gangaraju et al., 2020; Guo et al., 2019; Rampal et al., 2014; Schischlik et al., 2019; Skov et al., 2011; Skov et al., 2012) has evaluated the disease-relevant platelet transcriptome in a sizeable clinical cohort of all three subtypes of Ph^- MPNs. Here, we extend our prior preliminary work(Krishnan et al., 2017) toward a comprehensive analysis of disease-relevant (Cummings et al., 2017) platelet RNA-sequencing (RNA-seq) in two temporally independent and mutually validating cohorts of all three MPN subtypes, ET, PV, and MF (primary or post ET/PV secondary). We demonstrate marked differences in platelet gene expression across the MPN spectrum, which also permits robust validated (temporal and geographical) prediction of MF. In addition to identifying novel gene expression signatures impacted by the JAK1/JAK2 inhibitor ruxolitinib (RUX), platelet profiling reveals MPN-altered pathways that may be targets for future drug development.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Two independent MPN clinical cohorts and closely replicated platelet transcriptome

We prepared highly purified leukocyte-depleted (Amisten, 2012; Rowley et al., 2011) platelets from peripheral blood samples of two cohorts (approximately 2 years apart; Stanford single-center) of patients with a diagnosis of MPN (including provisional) at the time of sample acquisition, and included healthy controls in each cohort as reference (cohort 1, n = 71, and cohort 2, n=49; **Figure 1**, **Table S1A-B**). Only 2 patients (2%) received a change in diagnosis from MPN (MF) to a non-MPN phenotype; and were therefore excluded from all downstream analyses (**Figure 2A** principal component analysis plot panel-3 identifies these 2 outliers). Our two-cohort study was specifically designed (before knowledge of other subsequent studies) for the explicit purpose of validation, not only of inter-cohort RNA-seq results (Kukurba and Montgomery, 2015) but also to evaluate temporal validation (Moons et al., 2012) of our prediction model (see Methods, **Figure 5C,E,F**). **Figure S1** demonstrates our established(Rowley et al., 2011; Rowley et al., 2019) high-quality and highly efficient experimental framework toward a rigorous platelet RNA-seq approach. Clinical features of the MPN patients are shown in **Figure 1**; and listed in **Table S1A-B**. The distribution of key variables was closely matched between the two cohorts by MPN subtype (**Figure 1A**), age (**B**), gender (**C**), *JAK2*/*CALR* mutation status (**D**) and treatment (**E**). Any inter-patient variability in patient age, gender and treatment were adjusted as confounding factors in all downstream gene expression analyses (see Methods). Clinical laboratory measures (**Figure 1F**) at the time of sampling reflected the phenotype of the MPN subtypes (please see figure legend for detailed statistical comparisons). The two cohorts of platelet transcriptome data (**Figure 1G**, normalized transcript counts) adjusted for patient age, gender and all treatment as confounding factors were also highly correlated (R^2 = 0.89), thus demonstrating high inter-cohort validation of gene expression that then enabled us to combine our two RNA-seq datasets into a final integrated data set of enhanced statistical power for downstream analyses, especially prediction modeling (**Figure 5**). Together, this platelet transcriptome compendium comprises 118 human peripheral blood samples from healthy controls (n=21) and World Health Organization-defined MPN patients (24 ET, 33 PV and 40 MF) that include seven untreated, and 92 either on cytoreductives/biologics (*e.g.* ruxolitinib, hydroxyurea, interferon-alpha), anti-thrombotic agents (e.g.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

aspirin, warfarin), or a combination and captures the real-life diversity among MPN patients. Our cross-sectional design here capturing patients from all three MPN subtypes also serves as a practical alternative to the longitudinal approach, though ideal, is likely difficult to implement in these chronic disorders with often decades of disease.

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

Given the phenotypic overlap, yet also differences in disease behavior and prognosis between ET, PV and MF, we hypothesized that there may be MPN subtype-specific differences in gene expression that are independent of *JAK2*/*CALR*/*MPL* mutation status. In addition, we hypothesized that MPN platelets are likely to be enriched for subtype-specific biomarkers that may otherwise be missed in blood/plasma/serum sources (Schischlik et al., 2019; Skov et al., 2011; Skov et al., 2012). Therefore, we compared platelet transcriptomic expression in each of the three MPN subtypes with that of controls and discovered a shared gene set that is also progressively differentiated across the MPN spectrum (ET/PV to MF **Figure 2 A-C**). First, unsupervised principal component analysis (PCA) of MPN patients and controls data (**Figure 2A**) confirmed that the collective variability from the first two principal components (accounting for 48% of total variance), after adjusting for age, gender, treatment and experimental batch, was MPN disease status, with increasing differentiation by subtype. Next, differential gene expression analysis (DGEA, volcano plot**, Figure 2B, C**) efficiently distinguished each of the MPN subtypes and resulted in highly significant expression signatures (adjusted p-value/FDR <0.05) with 2634 genes differentially regulated in ET (1364 up and 1269 down), 4398 in PV (2098 up and 2300 down), and 6648 in MF (3965 up and 2683 down). A subset of 100+ long non-coding RNAs and pseudogenes also constituted the significant (FDR<0.05) differential expression across MPNs (**Table S2A-C**).

Specifically, DGEA also uncovered shared and unique genes between all three MPN phenotypes, thus offering a potential core set of genes involved in MPN pathogenesis (**Figure 2C,** with associated heatmaps in Figure 3 and pathways in Figure 4). The shared gene set at FDR < 0.05 constituted 654 up-regulated genes, with a predominance of molecular pathways involving *myeloid cell activation in immune response*

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and membrane protein proteolysis; and 361 down-regulated genes, reflecting negative regulation of hematopoeisis and negative regulation of transmembrane receptor protein serine/threonine kinase signaling as a consistent pathogenetic mechanism. The upregulated genes belonged to the endoplasmic reticulum/ER-Golgi intermediate compartment, and included a particularly high expression of the cAMP-response element binding transcription factor, *CREB3L1*(Sampieri et al., 2019) implicated in cell differentiation and inhibition of cell proliferation, with concomitant high expression of ER chaperones(Clark et al., 2002) calreticulin (*CALR*), calnexin (*CANX);* transport *factors:* golgin *(GOLGB1)* and folate receptor *FOLR1.* Platelet alpha granule proteins *(F5, VWF, MMP14),* several collagens (*COL10A1, COL18A1, COL6A3*), immune/inflammatory (*IFITM2/3/10, FCGR2A, TMEM179B*), *Cathepsins (C/Z/D, MIF, PTGES2*) and proliferation mediator genes (*CDK1, CCNG1, BMP9/GDF2, LAPTM4B, PSENEN*) also constituted the MPN shared set. Downregulated genes, on the other hand, were predominantly within transcription factor complexes, and included opposing expression of *CREB1* (vis-à-vis *CREB3L1* above), calcium-calmodulin protein kinases, *CAMK4*, *SMAD1* and β-catenin *CTNNB1* together pointing to dysregulated calcium (Ca^{2+}) homeostasis in the ER lumen.

Differential markers in each of ET, PV and MF also highlight candidate genes as potential mediators of the pro-thrombotic and pro-fibrotic phenotypes in MPNs. In ET and PV, a strong thromboinflammatory profile (Barbui and Falanga, 2016; Gangaraju et al., 2020) is revealed by the upregulation of several interferon inducible transmembrane genes (*IFITM2, IFITM3, IFITM10, IFIT3, IFI6, IFI27L1, IFI27L2*) , interleukin receptor accessory kinases/proteins (*IRAK1, IL15, IL1RAP, IL17RC) and* several solute carrier family genes (*SLC16A1, SLC25A1, SLC26A8, SLC2A9)* as glucose and other metabolic transport proteins, and coagulation factor V (*F5)*. In MF, fibrosis-specific markers were identified by an additional focused comparison of MF patients versus ET and PV (**Figure S2**), showing increased expression of several pro-fibrotic growth factors (*FGFR1, FGFR3, FGFRL1*), matrix metalloproteinases (*MMP8, MMP14*), vascular endothelial growth factor A (*VEGFA*), insulin growth factor binding protein (*IGFBP7*), and cell cycle regulators (*CCND1, CCNA2, CCNB2, CCNF*). GSEA of this MF-focused comparison again highlighted potential underlying molecular dysregulation in MPNs that likely contribute to the fibrotic phenotype (*e.g.* unfolded protein response, mTORc1 signaling, MYC/E2F targets, oxidative phosphorylation, **Figure 4**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

In addition to confirming previous observations (Arcaini and Cazzola, 2018; Kleppe et al., 2018; Spivak, 2017; Vainchenker et al., 2018) on the anti-inflammatory and immunosuppressive effects of *JAK* inhibition by RUX (e.g. downregulation in our RUX-treated cohort of *IL1RAP, CXCR5, CPNE3, ILF3*), we identified new gene clusters responsive to RUX in the inhibition of type I interferon (e.g. *IFIT1, IFIT2, IFI6*), chromatin regulation (*HIST2H3A/C, HIST1H2BK, H2AFY, SMARCA4, SMARCC2*), and epigenetic methylation in mitochondrial genes (*ATP6, ATP8, ND1-6 and NDUFA5*). Recent literature probing the mechanisms of action of ruxolitinib in other disease settings, including SARS-CoV-2 (Yan et al., 2021; Zhou et al., 2014) confirm our observations in MPNs.

A direct comparison restricted to only differentially expressed genes (FDR<0.05) in *RUX-treated vs not* with *MF vs healthy controls* revealed less than 5% overlap (**Figure 2D**), reflecting potentially the extent of the impact of treatment by ruxolitinib relative to the substantive disease burden in MF. Focusing further on the directionality of the changes observed, we found just 18 genes that were increased in MF and suppressed in the RUX-treated cohort (**Figure 2E**, colored green); and 9 vice versa (**Table S3A-B**). Despite the small overlap, we capitalized on the converging genes to better define molecular and physiological pathways underlying the effect of RUX in MPNs. The 18 RUX-down-regulated genes followed expected mapping to

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

immunosuppression through interferon- and cytokine-mediated signaling pathways. The 9 genes that were upregulated with RUX in MF mapped to previously undescribed effect of the drug in select G-protein-coupled receptor and chemokine activity (*e.g. CXCR5, GPR128/ADGRG7*), semaphorin signaling (*SEMA3C*) and circadian regulation (*PER3*). A sub-cohort analysis of RUX-treated and RUX-naïve MF patients alone also identified downregulation of interferon-stimulated genes (*e.g. SLFN12L),* G-protein-coupled receptors *(S1PR5), f*ibroblast growth factor binding protein *(FGFBP2) and* tyrosine kinase *(LCK).*

Graded differential expression by MPN phenotype and driver mutation status

Unsupervised hierarchical clustering was used to more precisely define the nature of MPN platelet transcriptome variability from controls, as well as between and within MPN subtypes. **Figure 3** reveals a spectrum of expression in the MPN platelet transcriptomic profile using just the top 10 highly significant differentially expressed genes by disease status: (a) all MPN vs controls, (b) MF vs controls, and (c, and d) ET- and PV- vs controls. As shown in **Figure 3A**, all MPN patients clustered into two distinct groups: a larger group of 87 ET, PV and MF patients clustered independently from the 21 controls, whereas a smaller group of 10 ET, PV and MF patients formed a homogeneous cluster of their own closer to controls reflecting a varying gradation with respect to the top-10 gene expression by MPN subtype (patient variables annotated above the heatmaps offer additional context, particularly on mutation status and RUX therapy). In the larger cluster, while we observed a graded overlap in platelet RNA signatures between ET and PV, a more distinct expression pattern characterized the more advanced population of MF patients (*PC1 correlates with MF disease risk by the Dynamic International Prognostic Scoring System (Passamonti et al., 2010), DIPSS*, **Figure S3**). These data collectively highlight the importance of phenotype-modifying genes that are independent of JAK2/*CALR*/*MPL* mutation status.

Untangling other mechanisms beyond the few genes that are recurrently mutated is critical for defining subtype-specific risk and for identifying molecular pathways for targeted therapy. Accordingly, we sought to refine the molecular classification of MPN by associating platelet gene expression profiles with the corresponding subtype, and yielded a core set of 10 highly-significant preferentially expressed genes for each: (i) MF (**Figure 3B**), defined by high mRNA expression of proteostasis-associated *CREB3L1* and

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CALR, and megakaryocyte-erythroid differentiation stage-associated *RHAG* (Caparros-Perez et al., 2017; Zeddies et al., 2014) ii) ET (**Figure 3C**), marked by comparatively high expression of interferon-related genes *IFITM2/3,* immune response *FCGR2A*, and proliferation-associated *STAT5* target *OSBP2*; and iii) PV (**Figure 3D**), marked by overlapping signatures with ET in inflammation-associated *IFITM3* and *TBL1X*, and the B-myb promoter, *MYBL2*; with MF in the maturation-associated *RHAG* at variable expression; and with both ET and MF in the high expression of *CREB3L1*, and cell survival-associated *MINDY4* and *STAC*.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

upregulated in MF. Taken together, these data demonstrate that in addition to immune factors such as type I/II interferons and dysregulation of interleukin-dependent inflammatory responses, which have been linked to MPNs, platelet transcriptional signatures of proliferation, metabolic, and proteostasis signaling are a feature of MPN pathogenesis (**Figure S4** captures the relative enrichment by subtype of MPN molecular pathway categories as a concept model of MPN progression).

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

Current knowledge of MPN genetic, cytogenetic or epigenetic abnormalities are limited(Spivak, 2017; Spivak et al., 2014) in their ability to enable prediction of disease progression or evolution of a given patient, from ET/PV phenotype to MF. In order to investigate the potential of platelet transcriptomic parameters to enable MF prediction, we constructed LASSO penalized (Tibshirani, 1996) regression classifiers (machine learning R package *glmnet)* to discriminate MPN subtypes from each other, and from healthy controls (**Figure 5A-E**). We apply two rigorous (Moons et al., 2012) external validation conditions **(Figure 5C,D):** i) training and independent temporal validation (**Figure 5C**) leveraging the Stanford two-cohort design and ii) geographical validation (**Figure 5D**) using two independently published platelet transcriptome datasets: first, from Rondina et al., 2020 on an additional n=31 healthy donors integrated with the Stanford datasets as training; and second, from Guo et al., 2020 n=25 MF and n=15 healthy donors as geographical validation of the Lasso algorithm (**Figure 5D**). Our temporal validation constituted three types of models i) baseline (no transcriptome, but age, gender and driver mutation status as reference variable information available not only for patients but also healthy donors); ii) full platelet transcriptome integrated with the above baseline; and iii) subset platelet transcriptome that exhibits progressive differentiation from controls to ET to MF or controls to PV to MF (>3000 genes, top few of each comparison visualized to demonstrate progressive gene expression, **Figure 5A-B**) integrated with the above baseline (progressive subset is selected unbiased as part of the Lasso learning procedure). Comparison of the classification potential among the three models demonstrated that the progressive platelet transcriptome model (**Figure 5E** red curve**)** substantially outperformed the baseline model (**Figure 5E** black curve) and was slightly better than the full transcriptome model (**Figure 5E** blue vs red curves) in

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the classification of ET, PV and MF. Predicted probabilities for all three models are shown in **Table S5A-C**. Lasso logistic regression classifiers to predict MF with each of the models under the first temporal two-cohort training-validation split of baseline, full, and progressive transcriptome each achieved area under the receiver-operating characteristic curve (AUROC) of 0.68, 0.95, and 0.96 respectively. Outperformance of the progressive transcriptome model was validated in our subsequent independent external geographic validation (**Figure 5E** green curve) at an AUROC of 0.97. Recurrent top 4 genes from our progressive transcriptome Lasso are visualized as a heatmap in **Figure 5F** clearly capturing the incremental gradient in gene expression between controls, ET, PV and MF. These include *ADAMTS3* (ADAM metallopeptidase protease(Mead and Apte, 2018) with likely roles in *VEGF* signaling, tissue remodeling and expression of related collagens through profibrotic *PAR1* and *TGF-β* signaling), *PSMB5* (implicated in proteasomal degradation/UPR activation(Wang et al., 2017) and identified previously in MPNs (Skov et al., 2010)), *NT5C* related to *PI3K* signaling (Fruman et al., 2017; Moniz et al., 2017) and *SUPT6H*/*SPT6(Bres et al., 2008; Frydman et al., 2020)*, a tumor-initiating histone chaperone associated with chromatin remodeling. Lasso-selected candidate markers capture the underlying MPN pathology and offer potential therapeutic targets.

A key aspect of the layered Lasso modeling demonstrated here is our use of an approach that can be developed in future work to incorporate additional predictors to MF (e.g. *JAK2* V617F allele burden (Vannucchi et al., 2008) or other genetic variants (Tefferi et al., 2016) beyond the driver mutations). **Figure S5** demonstrates this through a second base model for prediction to MF from ET or PV alone (no transcriptome, but platelet count and hemoglobin levels as two additional clinical parameters that were available for all patients in addition to age, gender and driver mutation status).

Discussion

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

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

Our data closely overlaps with recent MPN platelet studies: thrombo-inflammatory signatures in PV from Gangaraju and Prchal et al (Gangaraju et al., 2020) (*BCL2, CXCL1, MMP7, PGLYRP1, CKB, BSG, CFL1* and more) or fibrosis-associated signatures in MF from Guo et al. (*CCND1, H2AFX, CEP55* and several others collectively reflected in the external validation of our Lasso algorithm).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Most notably in our data on MF, high expression of ER stress and unfolded protein response (UPR) biomarkers *(e.g. CREB3L1, CALR)* associated with impaired proteostasis signaling; and emerged as a key feature of MPN pathobiology. Indeed, recently published work from distinct research groups (Liu et al., 2020; Osorio et al., 2016),(LaFave and Levine, 2016) highlight protein quality control in ER-associated degradation and proteostasis (Osorio et al., 2016),(LaFave and Levine, 2016) deregulation as a primary effector of myeloid transformation highlighting the importance of protein homeostasis for normal hematopoiesis. These findings too are in line with reports (Kaushik and Cuervo, 2015) implicating chronic ER stress, malfunctioning protein quality control, and loss of proteostasis as aggravating factors in age-related disorders.

Most importantly, platelet gene expression profiling in MPN offers directions for prediction of myelofibrosis. Applying machine-learning algorithms of LASSO penalized regression under two conditions of external validation (Moons et al., 2012): temporal (using our two cohort design) and geographical (independently published datasets on healthy donors(Guo et al., 2020; Rondina et al., 2020) and MF (Guo et al., 2020), we uniquely discriminate MPN subtypes from each other, and healthy controls using three model types and predict MF at high accuracy. The highest performing model used a set of progressively differentiated MPN genes at an area under the (ROC) curve of 0.96 (temporal) and 0.97 (geographical); and rendered a core signature of <5 candidate markers as top predictors of disease progression. It will be of interest to determine what machine-learning algorithms based on a defined platelet gene expression classifier on potential new MPN datasets (ideally longitudinal) can be used to more precisely predict the probability and/or timing of an individual's risk of progression from ET/PV to secondary MF.

In conclusion, using platelet transcriptome profiling, we observed dynamic shifts in MPN immune inflammatory profile and preferential expression of interferon-, proliferation-, and proteostasis-associated genes as a progressive gradient across the three MPN subtypes. Our findings highlight that MPN progression may be influenced by defects in protein homeostasis (impaired protein folding and an accumulation of misfolded proteins within the endoplasmic reticulum) and an abnormal integrated stress response – consistent with recent studies[57],(Liu et al., 2020; Osorio et al., 2016),(LaFave and Levine, 2016)

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

indicating dysregulated proteostasis as a primary effector of myeloid transformation. While this particular work has been focused on the overarching progressive platelet transcriptome across MPNs, these data open an important avenue for utilizing platelet RNA signatures to better understand specific MPN complications such as the risk of thrombosis and bleeding, or fibrosis and transformation to AML. Altogether, this study demonstrated in chronic MPNs provides a comprehensive framework for exploiting the platelet transcriptome and may inform future studies toward mechanistic understanding and therapeutic development in MPNs, and potentially other age-related disorders.

Limitations of the study

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Ethical Approval

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

Subjects and Specimen Collection

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

Platelet Isolation

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(SELP) and *Glycoprotein IIbIIIA (CD41/ITGA2B)* were expectedly higher in all MPN than healthy controls but were not statistically significantly different between MPN subtypes; indicating that any expression difference was not due to experimental artefacts. In addition, we know from prior work (J.R.) that regardless of activation status, RNA-seq reliably estimates mRNA expression patterns in platelets. We also know from rigorous prior work(Gnatenko et al., 2003; Raghavachari et al., 2007; Weyrich and Zimmerman, 2003) that several abundant platelet mRNAs are well-known leukocyte or red cell transcripts; and do not immediately imply contamination by these classes but rather that platelets express gene products that are also present in other cell lineages. Sixty-five of our top 100 abundant platelet transcripts matched exactly with those of the top 100 abundant genes from the three previous studies cited(Gnatenko et al., 2003; Rowley et al., 2011); and a composite pathway analysis with the top 100 abundant genes from this as well as the previous studies matched identically.

Next Generation RNA Sequencing

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

Platelet Transcriptome Analysis

Picard, Samtools, and other metrics were used to evaluate data quality. Processed reads were aligned against the reference human transcriptome GRCh37/hg19 using RSEM(Li and Dewey, 2011) and bowtie2(Langmead and Salzberg, 2012), and expression at gene level determined by calculating raw gene count. Only genes that passed expression threshold were used; genes were considered expressed if, in all samples, they had at least 10 counts (genes with low counts are automatically filtered by built-in functions in DeSeq2, see below). A total of 12,924 genes were considered expressed. Gene expression data was

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

library-size-corrected, variance-stabilized, and log2-transformed using the R package DESeq2(Love et al., 2014). We refer to this version of the data as "raw data" as it is not corrected for any confounders of gene expression variability. DESeq2 was used to call differential expression while adjusting for patient age, gender and treatment as confounding variables and controlling for multiple comparisons using the Benjamini-Hochberg defined false discovery rate (FDR). Significant variance in expressed transcripts were pre-specified as those transcripts with an FDR <0.05 and a log2 fold change ≥ 0.5 in MPN, as compared to healthy controls (the entire differential transcriptome was applied in the instances of downstream Gene Set Enrichment Analysis and the Lasso prediction modeling).

Statistical analysis

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

Pathway/Gene set enrichment analysis for differentially expressed (DE) genes.

Gene set enrichment analysis (GSEA)(Subramanian et al., 2005) was performed on the entire DE gene set for each MPN subtype, using the Cancer Hallmarks gene sets from MSigDB(Liberzon et al., 2015). The 'GSEA Pre-ranked' function was used with a metric score that combines fold change and adjusted p-value together for improved gene ranking. We used default settings with 10,000 gene set permutations to generate *p* and *q* values, and compared MPN subtypes in the overall cohort, and the ruxolitinib-treated subgroup and the ruxolitinib-naive subgroup separately. In these analyses, to allow for a broad comparison, we assessed all transcripts that were differentially expressed according to FDR/adjusted p < 0.25 as recommended by the authors of GSEA (Subramanian et al., 2005).

Predictive model generation and external validation

At the conception of this study late 2016, and our early work(Krishnan et al., 2017), we did not identify any publicly available RNA sequencing data on MPN platelets. This prompted our specific two-cohort design for the express purpose of temporal external validation as an essential step in rigorous prediction modeling(Moons et al., 2012). A subsequent independent publication(Guo et al., 2020) facilitated an additional geographically independent external validation of our model.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We used Lasso penalized regression(Tibshirani, 1996) for our model to predict MF from the either healthy controls, ET or PV. Among a variety of statistical machine learning algorithms that have been used in prediction modeling, Lasso is favored for its flexibility and simplicity; and its ability to identify the least set of significant factors from high dimensional data. We evaluated platelet transcriptomic features with clinical features (age, gender and mutation status for the entire dataset including healthy donors, and in MPN patients alone, platelet and hemoglobin values). Normalized gene counts data were split into training (used for constructing multinomial logistic models) and validation (used for model evaluation and generalization) cohorts. Separately, we assessed the progressive and monotonic upward or downward trend in gene expression, we applied the Mann-Kendall trend test (multiple comparison adjusted with the Benjamini-Hochberg method) to normalized gene counts and identified statistically significant progressive genes across all three MPN subtypes.

Three multinomial logistic models were constructed: first, with Lasso selected predictors from all genes, second, with Lasso selected predictors from progressive genes and third, a baseline model using age, gender and mutation status (*JAK2 and CALR*) as predictors. Model outputs correspond to probabilities of having a CTRL, ET, PV or MF phenotype (sum of these four probability values totaling 1). Potential interpretation of these probabilities includes MPN risk assessment, *e.g.* a patient with higher probabilities of PV and MF would indicate higher risk than one with higher probabilities of CTRL or ET. The (Guo et al., 2020) dataset on MF platelet RNA seq served as an independent test set (Fig. 5D schematic) while data from our cohorts at Stanford and additional external data on healthy donors from (Rondina et al., 2020) constituted an integrated training cohort (R package Limma was applied for bioinformatic correction of any batch effects).

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data Sharing Statement

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

A private link for editors and reviewers will be made available at a Stanford web-based data repository.

Ethics Declarations/Competing Interests The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bibliography

Amisten, S. (2012). A rapid and efficient platelet purification protocol for platelet gene expression studies. Methods in molecular biology (Clifton, NJ) *788*, 155-172.

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

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

Barbui, T., and Falanga, A. (2016). Molecular biomarkers of thrombosis in myeloproliferative neoplasms. Thromb Res *140 Suppl 1*, S71-75.

Best, M. G., Sol, N., Kooi, I., Tannous, J., Westerman, B. A., Rustenburg, F., Schellen, P., Verschueren, H., Post, E., Koster, J., Ylstra, B.*, et al.* (2015). RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. Cancer Cell *28*, 666-676.

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

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

Caparros-Perez, E., Teruel-Montoya, R., Lopez-Andreo, M. J., Llanos, M. C., Rivera, J., Palma-Barqueros, V., Blanco, J. E., Vicente, V., Martinez, C., and Ferrer-Marin, F. (2017). Comprehensive comparison of neonate and adult human platelet transcriptomes. PLoS One *12*, e0183042.

Clancy, L., and Freedman, J. E. (2016). Blood-Derived Extracellular RNA and Platelet Pathobiology: Adding Pieces to a Complex Circulating Puzzle. Circulation research *118*, 374-376.

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

Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R., Bolduc, V., Waddell, L. B., Sandaradura, S. A., O'Grady, G. L., Estrella, E.*, et al.* (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med *9*.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

694 Davizon-Castillo, P., Rowley, J. W., and Rondina, M. T. (2020). Megakaryocyte and Platelet
695 Transcriptomics for Discoveries in Human Health and Disease. Arterioscler Thromb Vasc Biol *40*, 1432-
696 1440.
697 Dunbar, A., Nazir, A., and Levine, R. (2017). Overview of Transgenic Mouse Models of Myeloproliferative
698 Neoplasms (MPNs). Curr Protoc Pharmacol *77*, 14 40 11-14 40 19.
699 Finazzi, G., De Stefano, V., and Barbui, T. (2013). Are MPNs vascular diseases? Current hematologic
700 malignancy reports *8*, 307-316.
701 Fruman, D. A., Chiu, H., Hopkins, B. D., Bagrodia, S., Cantley, L. C., and Abraham, R. T. (2017). The PI3K
702 Pathway in Human Disease. Cell *170*, 605-635.
703 Frydman, G. H., Tessier, S. N., Wong, K. H. K., Vanderburg, C. R., Fox, J. G., Toner, M., Tompkins, R. G.,
704 and Irimia, D. (2020). Megakaryocytes contain extranuclear histones and may be a source of platelet-
705 associated histones during sepsis. Sci Rep *10*, 4621.
706 Gangaraju, R., Song, J., Kim, S. J., Tashi, T., Reeves, B. N., Sundar, K. M., Thiagarajan, P., and Prchal, J.
707 T. (2020). Thrombotic, inflammatory, and HIF-regulated genes and thrombosis risk in polycythemia vera
708 and essential thrombocythemia. Blood Adv *4*, 1115-1130.
709 Geyer, H. L., Dueck, A. C., Scherber, R. M., and Mesa, R. A. (2015). Impact of inflammation on
710 myeloproliferative neoplasm symptom development. Mediators Inflamm *2015*, 284706.
711 Gilles, L., Finke, C., Lasho, T. L., Pardanani, A., Tefferi, A., and Crispino, J. (2012). Aberrant megakaryocyte
712 gene expression contributes to primary myelofibrosis. Blood *120*, 2867-2867.
713 Gnatenko, D. V., Dunn, J. J., McCorkle, S. R., Weissmann, D., Perrotta, P. L., and Bahou, W. F. (2003).
714 Transcript profiling of human platelets using microarray and serial analysis of gene expression. Blood *101*,
715 2285-2293.
716 Guo, B. B., Linden, M. D., Fuller, K. A., Phillips, M., Mirzai, B., Wilson, L., Chuah, H., Liang, J., Howman, R.,
717 Grove, C. S., Malherbe, J. A*., et al.* (2019). Platelets in myeloproliferative neoplasms have a distinct
718 transcript signature in the presence of marrow fibrosis. Br J Haematol.
719 Guo, B. B., Linden, M. D., Fuller, K. A., Phillips, M., Mirzai, B., Wilson, L., Chuah, H., Liang, J., Howman, R.,
720 Grove, C. S., Malherbe, J. A*., et al.* (2020). Platelets in myeloproliferative neoplasms have a distinct
721 transcript signature in the presence of marrow fibrosis. Br J Haematol *188*, 272-282.
722 Hasselbalch, H. C., and Bjorn, M. E. (2015). MPNs as Inflammatory Diseases: The Evidence,
723 Consequences, and Perspectives. Mediators Inflamm *2015*, 102476.
724 Hinds, D. A., Barnholt, K. E., Mesa, R. A., Kiefer, A. K., Do, C. B., Eriksson, N., Mountain, J. L., Francke, U.,
725 Tung, J. Y., Nguyen, H. M., Zhang, H*., et al.* (2016). Germ line variants predispose to both JAK2 V617F
726 clonal hematopoiesis and myeloproliferative neoplasms. Blood *128*, 1121-1128.
727 Kaushik, S., and Cuervo, A. M. (2015). Proteostasis and aging. Nature Medicine *21*, 1406-1415.
728 Kleppe, M., Koche, R., Zou, L., van Galen, P., Hill, C. E., Dong, L., De Groote, S., Papalexi, E., Hanasoge
729 Somasundara, A. V., Cordner, K., Keller, M*., et al.* (2018). Dual Targeting of Oncogenic Activation and

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Inflammatory Signaling Increases Therapeutic Efficacy in Myeloproliferative Neoplasms. Cancer Cell *33*, 785-787.

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

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

Krishnan, A., Zhang, Y., Perkins, C., Gotlib, J., and Zehnder, J. L. (2017). Platelet Transcriptomic Signatures in Myeloproliferative Neoplasms. Blood *130*, 5288.

Kukurba, K. R., and Montgomery, S. B. (2015). RNA Sequencing and Analysis. Cold Spring Harb Protoc *2015*, 951-969.

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

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

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

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

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.* (2020). ER associated degradation preserves hematopoietic stem cell quiescence and self-renewal by restricting mTOR activity. Blood.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

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

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

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

Mascarenhas, J., Roper, N., Chaurasia, P., and Hoffman, R. (2011). Epigenetic abnormalities in myeloproliferative neoplasms: a target for novel therapeutic strategies. Clin Epigenetics *2*, 197-212.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

764 Matsuura, S., Thompson, C. R., Belghasem, M. E., Bekendam, R. H., Piasecki, A., Leiva, O., Ray, A.,
765 Italiano, J., Yang, M., Merill-Skoloff, G., Chitalia, V. C*., et al.* (2020). Platelet Dysfunction and Thrombosis in
766 JAK2(V617F)-Mutated Primary Myelofibrotic Mice. Arterioscler Thromb Vasc Biol *40*, e262-e272.
767 Mead, T. J., and Apte, S. S. (2018). ADAMTS proteins in human disorders. Matrix Biol *71-72*, 225-239.
768 Merlinsky, T. R., Levine, R. L., and Pronier, E. (2019). Unfolding the Role of Calreticulin in Myeloproliferative
769 Neoplasm Pathogenesis. Clin Cancer Res *25*, 2956-2962.
770 Middleton, E. A., Rowley, J. W., Campbell, R. A., Grissom, C. K., Brown, S. M., Beesley, S. J., Schwertz,
771 H., Kosaka, Y., Manne, B. K., Krauel, K., Tolley, N. D*., et al.* (2019). Sepsis alters the transcriptional and
772 translational landscape of human and murine platelets. Blood *134*, 911-923.
773 Moniz, L. S., Surinova, S., Ghazaly, E., Velasco, L. G., Haider, S., Rodriguez-Prados, J. C., Berenjeno, I.
774 M., Chelala, C., and Vanhaesebroeck, B. (2017). Phosphoproteomic comparison of Pik3ca and Pten
775 signalling identifies the nucleotidase NT5C as a novel AKT substrate. Sci Rep *7*, 39985.
776 Moons, K. G., Kengne, A. P., Grobbee, D. E., Royston, P., Vergouwe, Y., Altman, D. G., and Woodward, M.
777 (2012). Risk prediction models: II. External validation, model updating, and impact assessment. Heart *98*,
778 691-698.
779 Mughal, T. I., Pemmaraju, N., Radich, J. P., Deininger, M. W., Kucine, N., Kiladjian, J. J., Bose, P., Gotlib,
780 J., Valent, P., Chen, C. C., Barbui, T*., et al.* (2019). Emerging translational science discoveries, clonal
781 approaches, and treatment trends in chronic myeloproliferative neoplasms. Hematol Oncol *37*, 240-252.
782 Mullally, A., Bruedigam, C., Poveromo, L., Heidel, F. H., Purdon, A., Vu, T., Austin, R., Heckl, D., Breyfogle,
783 L. J., Kuhn, C. P., Kalaitzidis, D*., et al.* (2013). Depletion of Jak2V617F myeloproliferative neoplasm-
784 propagating stem cells by interferon-alpha in a murine model of polycythemia vera. Blood *121*, 3692-3702.
785 Nguyen, H. M., and Gotlib, J. (2012). Insights into the molecular genetics of myeloproliferative neoplasms.
786 Am Soc Clin Oncol Educ Book, 411-418.
787 Oh, S. T., and Gotlib, J. (2010). JAK2 V617F and beyond: role of genetics and aberrant signaling in the
788 pathogenesis of myeloproliferative neoplasms. Expert Rev Hematol *3*, 323-337.
789 Osorio, F. G., Soria-Valles, C., Santiago-Fernandez, O., Bernal, T., Mittelbrunn, M., Colado, E., Rodriguez,
790 F., Bonzon-Kulichenko, E., Vazquez, J., Porta-de-la-Riva, M., Ceron, J*., et al.* (2016). Loss of the
791 proteostasis factor AIRAPL causes myeloid transformation by deregulating IGF-1 signaling. Nat Med *22*,
792 91-96.
793 Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V. I., Paschka, P., Roberts, N. D., Potter, N. E.,
794 Heuser, M., Thol, F., Bolli, N., Gundem, G*., et al.* (2016). Genomic Classification and Prognosis in Acute
795 Myeloid Leukemia. N Engl J Med *374*, 2209-2221.
796 Passamonti, F., Cervantes, F., Vannucchi, A. M., Morra, E., Rumi, E., Pereira, A., Guglielmelli, P.,
797 Pungolino, E., Caramella, M., Maffioli, M., Pascutto, C*., et al.* (2010). A dynamic prognostic model to
798 predict survival in primary myelofibrosis: a study by the IWG-MRT (International Working Group for
799 Myeloproliferative Neoplasms Research and Treatment). Blood *115*, 1703-1708.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

Rondina, M. T., Voora, D., Simon, L. M., Schwertz, H., Harper, J. F., Lee, O., Bhatlekar, S. C., Li, Q., Eustes, A. S., Montenont, E., Campbell, R. A., *et al.* (2020). Longitudinal RNA-Seq Analysis of the Repeatability of Gene Expression and Splicing in Human Platelets Identifies a Platelet SELP Splice QTL. Circulation research *126*, 501-516.

Rondina, M. T., Weyrich, A. S., and Zimmerman, G. A. (2013). Platelets as cellular effectors of inflammation in vascular diseases. Circulation research *112*, 1506-1519.

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

Rowley, J. W., Oler, A. J., Tolley, N. D., Hunter, B. N., Low, E. N., Nix, D. A., Yost, C. C., Zimmerman, G. A., and Weyrich, A. S. (2011). Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes. Blood *118*, e101-111.

Rowley, J. W., Schwertz, H., and Weyrich, A. S. (2012). Platelet mRNA: the meaning behind the message. Curr Opin Hematol *19*, 385-391.

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

Rumi, E., and Cazzola, M. (2017). Diagnosis, risk stratification, and response evaluation in classical myeloproliferative neoplasms. Blood *129*, 680-692.

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

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

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

Simon, L. M., Edelstein, L. C., Nagalla, S., Woodley, A. B., Chen, E. S., Kong, X., Ma, L., Fortina, P., Kunapuli, S., Holinstat, M., McKenzie, S. E., *et al.* (2014). Human platelet microRNA-mRNA networks associated with age and gender revealed by integrated plateletomics. Blood *123*, e37-45.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Skov, V., Larsen, T. S., Thomassen, M., Riley, C., Jensen, M. K., Bjerrum, O. W., Kruse, T. A., and Hasselbalch, H. C. (2010). Increased Expression of Proteasome-Related Genes In Patients with Primary Myelofibrosis. Blood *116*, 4117.

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

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

Spivak, J. L. (2017). Myeloproliferative Neoplasms. N Engl J Med *376*, 2168-2181.

Spivak, J. L., Considine, M., Williams, D. M., Talbot, C. C., Jr., Rogers, O., Moliterno, A. R., Jie, C., and Ochs, M. F. (2014). Two clinical phenotypes in polycythemia vera. N Engl J Med *371*, 808-817.

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

Tefferi, A., Abdel-Wahab, O., Cervantes, F., Crispino, J. D., Finazzi, G., Girodon, F., Gisslinger, H., Gotlib, J., Kiladjian, J. J., Levine, R. L., Licht, J. D.*, et al.* (2011). Mutations with epigenetic effects in myeloproliferative neoplasms and recent progress in treatment: Proceedings from the 5th International Post-ASH Symposium. Blood Cancer J *1*, e7.

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological), 267-288.

Vainchenker, W., and Kralovics, R. (2017). Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. Blood *129*, 667-679.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

Weyrich, A. S. (2014). Platelets: more than a sack of glue. Hematology Am Soc Hematol Educ Program *2014*, 400-403.

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

Weyrich, A. S., and Zimmerman, G. A. (2013). Platelets in lung biology. Annual review of physiology *75*, 569-591.

Williams, N., Lee, J., Moore, L., Baxter, E. J., Hewinson, J., Dawson, K. J., Menzies, A., Godfrey, A. L., Green, A. R., Campbell, P. J., and Nangalia, J. (2020). Phylogenetic reconstruction of myeloproliferative neoplasm reveals very early origins and lifelong evolution. BioRxiv, 2020.2011.2009.374710.

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

Woods, B., Chen, W., Chiu, S., Marinaccio, C., Fu, C., Gu, L., Bulic, M., Yang, Q., Zouak, A., Jia, S., Suraneni, P. K.*, et al.* (2019). Activation of JAK/STAT Signaling in Megakaryocytes Sustains Myeloproliferation In Vivo. Clin Cancer Res *25*, 5901-5912.

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

Zeddies, S., Jansen, S. B., di Summa, F., Geerts, D., Zwaginga, J. J., van der Schoot, C. E., von Lindern, M., and Thijssen-Timmer, D. C. (2014). MEIS1 regulates early erythroid and megakaryocytic cell fate. Haematologica *99*, 1555-1564.

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

Zoi, K., and Cross, N. C. (2017). Genomics of Myeloproliferative Neoplasms. J Clin Oncol *35*, 947-954.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

A, Unsupervised principal component analysis (PCA) of normalized platelet gene expression counts adjusted for age, gender, treatment and experimental batch. Three panels of PC1 and PC2 colored by MPN subtype; and each contrasted with controls (n=21, yellow): ET (n=24, top, light green), PV (n=33, middle, dark green) and MF (n=42, bottom, dark blue). Circles filled or open mark presence or absence of ruxolitinib treatment; and size of circles, smaller or larger, indicate presence or absence of *JAK2* mutation. The first two principal components account for 48% of total variance in the data. **B,** Volcano plots (three panels as **a** above of ET, PV and MF) of differential gene expression showing statistical significance (negative log10 of p-values) versus log2 fold change of each gene. Significant up-regulated and down-regulated genes are those with p-values (FDR) smaller or equal to 0.05 and absolute value of fold changes larger or equal to 1.5**. C,D,** Venn Diagram comparisons of MPN differential gene expression lists. In **c,** each of ET, PV and MF is contrasted with controls; identifying gene sets that are shared (n=1504, FDR < 0.05) as well as unique to each subtype. In **D,** differential in the RUX-treated cohort is contrasted with MF vs controls. Differential in gene expression in RUX-treated cohort is a fraction of the total differential noted in the MF transcriptome. **E,** Volcano plot of differential gene expression between MPN patients treated with ruxolitinib and not. A small subset of overlapping differential genes that are upregulated in MF (**B,** bottom panel) and suppressed in the RUX-treated cohort (**E**) are colored in green.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

A, B Top few genes (out of 3000+) demonstrating monotonic progressive gene expression (log 2 fold change in expression *y*-axis, FDR < 0.01, *Mann-Kendall* test with Bonferroni correction) across *x*-axes **A**, CTRL-to-ET-to-MF and **B**, CTRL-to-PV-to-MF. **C,** Lasso-penalized multinomial logistic regression model under temporal validation *i.e.* trained on Stanford cohort 1 (n=71, 2016-17, Figure 1a) and validated on Stanford cohort 2 (n=49, 2019, Figure 1a) as test set. **D,** Lasso-penalized multinomial logistic regression model under geographical validation using two independently published datasets for training (cohort 3, n=31 healthy controls in addition to Stanford 1 & 2 cohorts) and validation (cohort 4, n=25 MF and n=15 healthy controls). **E,** Receiver Operating Curves (ROC) toward MF prediction under conditions of temporal (**C**) and geographical (**D**) validation. Temporal validation uses three layered models: i) baseline, with no gene expression data but patient age, gender and mutation alone; ii) adding entire MPN platelet transcriptome and iii) adding MPN progressive genes alone. Outperformance of the progressive transcriptome model (red curve, AUROC=0.96) vis-a-vis the entire transcriptome dataset (blue curve, AUROC=0.95) and lastly, the baseline model without gene expression (black curve, AUROC=0.68). Geographical validation using the progressive transcriptome model also demonstrates independent high MF predictive accuracy (green curve, AUROC=0.97). **F,** Heatmap of top recurring Lasso-selected progressive genes for each of controls (left column, CTRL, yellow bar), ET (light green), PV (dark green) and MF (dark blue). Rows indicate gradation in gene expression on a blue (low) to red (high) scale. Columns indicate sample type (CTRL, ET, PV and MF).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure Legends

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

Figure S2: Volcano plots (showing statistical significance, negative log10 of p-values, versus log2 fold change of each gene) of differential gene expression analysis of a sub-cohort of MF versus ET and PV patients combined. Significant up- and down-regulated genes are those with p-values (FDR) smaller or equal to 0.05 and absolute value of fold changes larger or equal to 1.5**.** Possible mediators of fibrosis-associated genes are highlighted and include pro-fibrotic growth factors and other matrix proteins.

Figure S3: Correlation of MF platelet transcriptome with disease risk by Dynamic International Prognostic Scoring System (DIPSS)

A, Unsupervised principal component analysis (PCA) of MF (n=42) normalized platelet gene expression counts adjusted for age, gender, treatment and experimental batch. PC1 and PC2 colored by DIPSS score (0-6 denoting increasing MF risk). Circles filled or open mark presence or absence of ruxolitinib treatment; and size of circles, smaller or larger, indicate presence or absence of *JAK2* mutation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S4: Relative molecular trajectories of MPN progression

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

Figure S5: Receiver Operating Curves (ROC) toward MF prediction from ET or PV alone with temporal validation and three layered models: i) baseline, with no gene expression data but patient age, gender and mutation status (as Figure 5E) and an additional two clinical variables: platelet count and hemoglobin; ii) adding entire MPN platelet transcriptome to the above baseline and iii) adding MPN progressive genes alone. Outperformance of the progressive transcriptome model (red curve, AUROC=0.97) vis-a-vis the entire transcriptome dataset (blue curve, AUROC=0.93) and the baseline model without gene expression (black curve, AUROC=0.83).

Figure S6: Unsupervised principal component analysis of platelet transcriptomic data from Stanford MPN & healthy donor cohorts (n=120) integrated with that of independently published healthy donors (n=31) (Rondina et al., 2020). Colors indicate controls (n=52, yellow): ET (n=24, top, light green), PV (n=33, middle, dark green) and MF (n=42, bottom, dark blue). Circles filled or open mark presence or absence of ruxolitinib treatment; and size of circles, smaller or larger, indicate presence or absence of *JAK2* mutation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table Descriptions

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

Table S2: List of non-coding genes at significant (FDR<0.05) differential expression in each comparison of ET, PV and MF versus healthy controls.

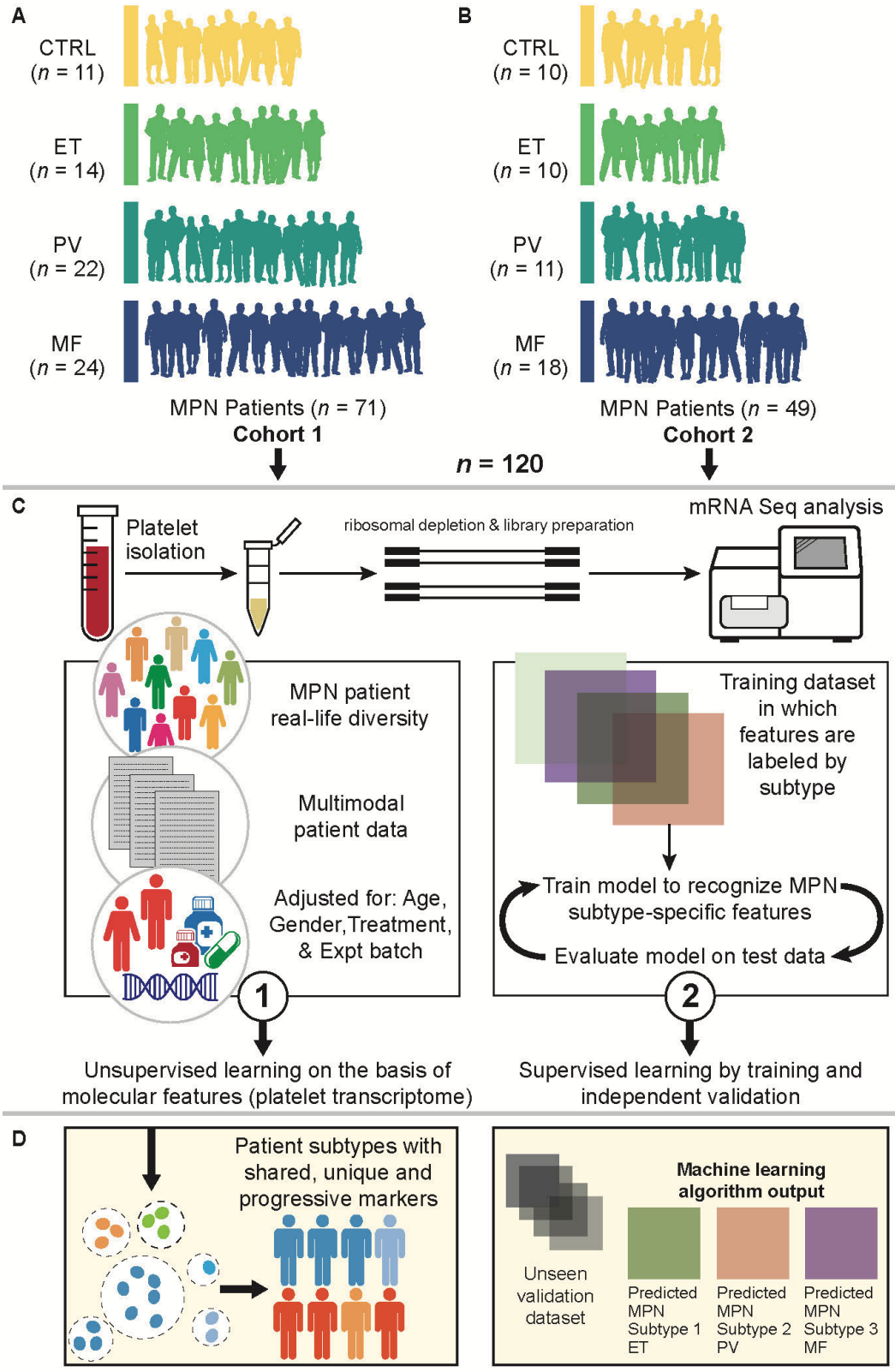
Table S3: Overlap of differentially expressed genes between the RUX-treated and the MF patient cohorts.

Table S4: Full data for all molecular pathways identified in each of ET, PV and MF patient cohorts.

Table S5: Predicted probabilities for each of CTRL, ET, PV and MF using all three Lasso logistic models.

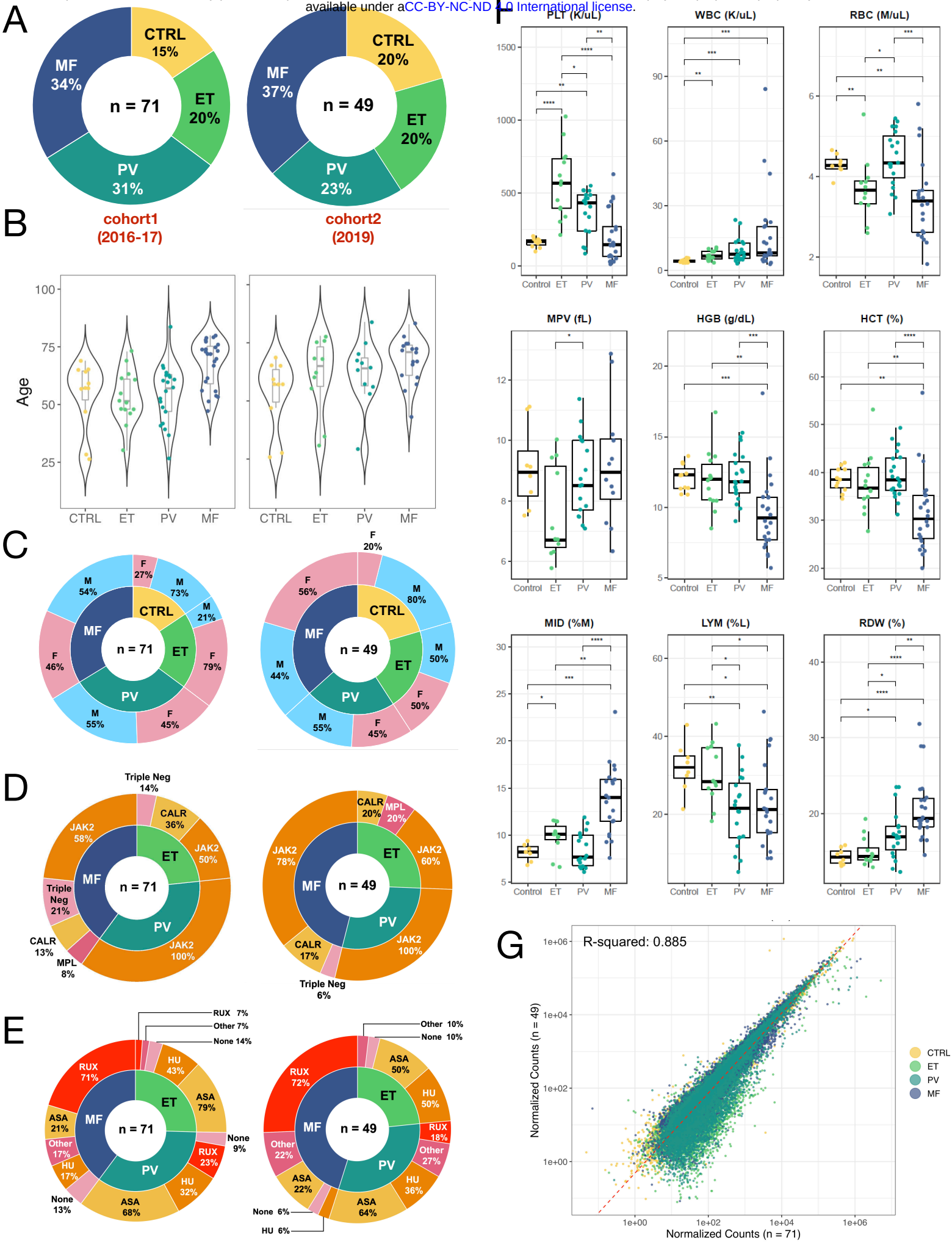
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

VISUAL ABSTRACT



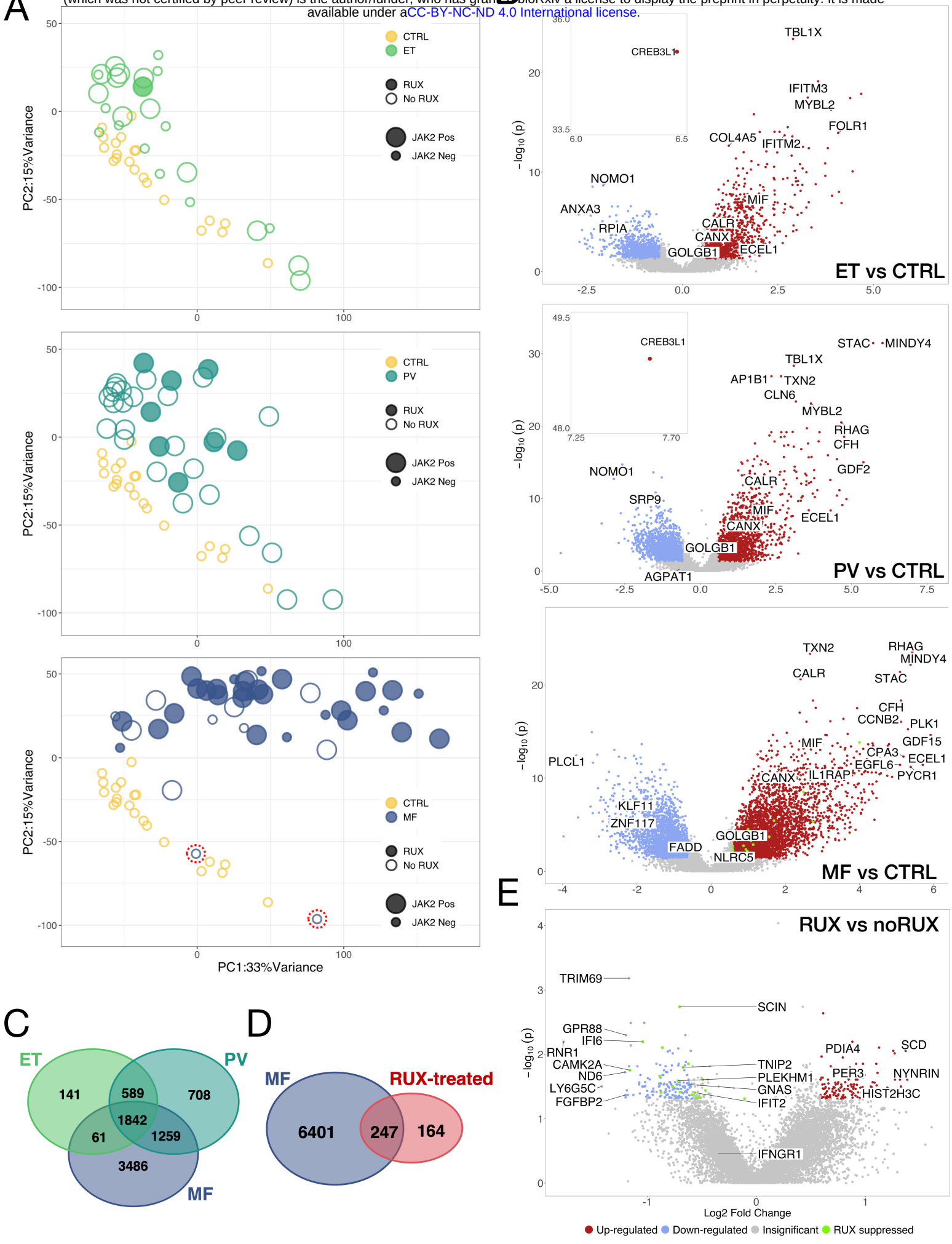
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE 1



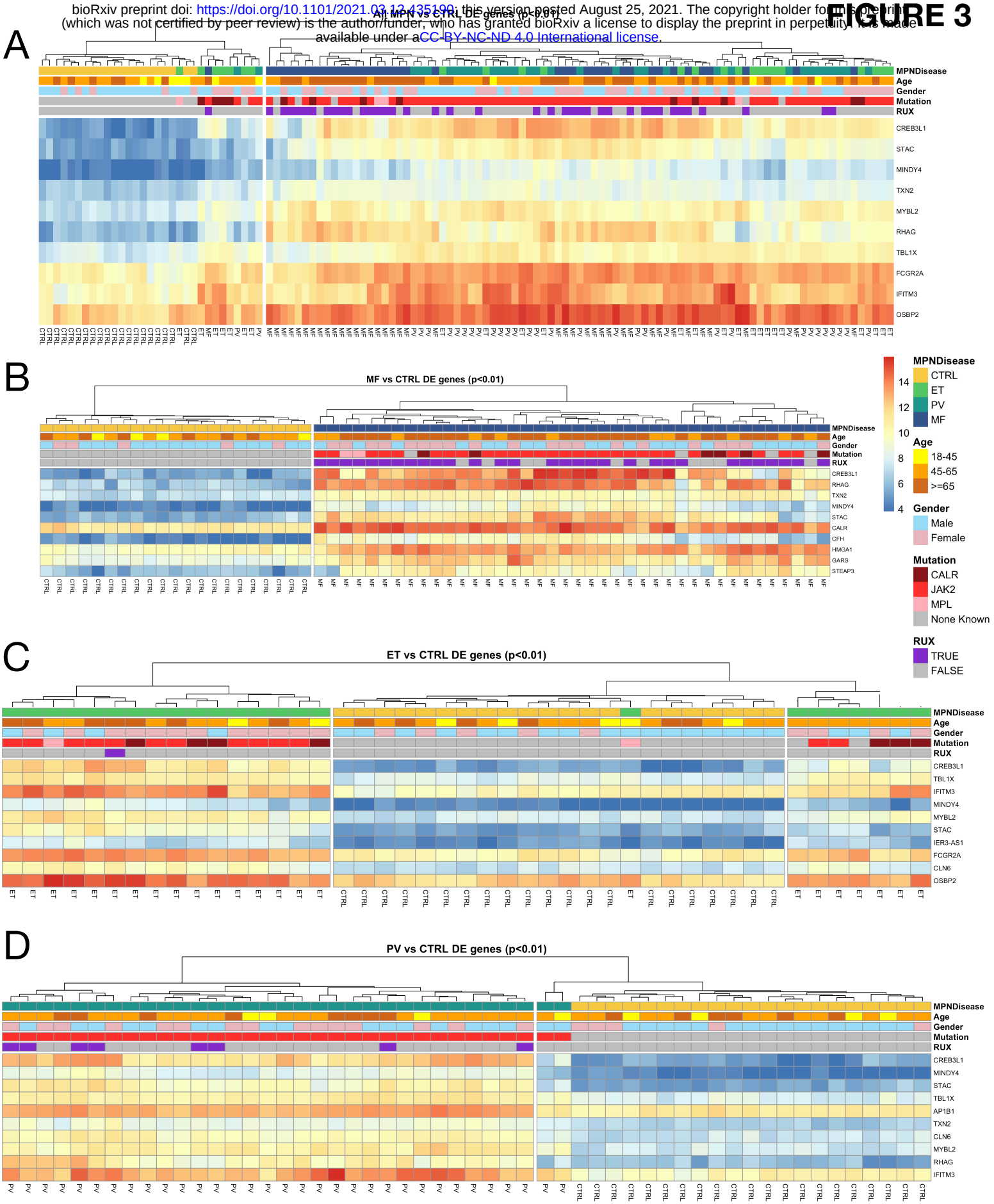
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE 2



bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

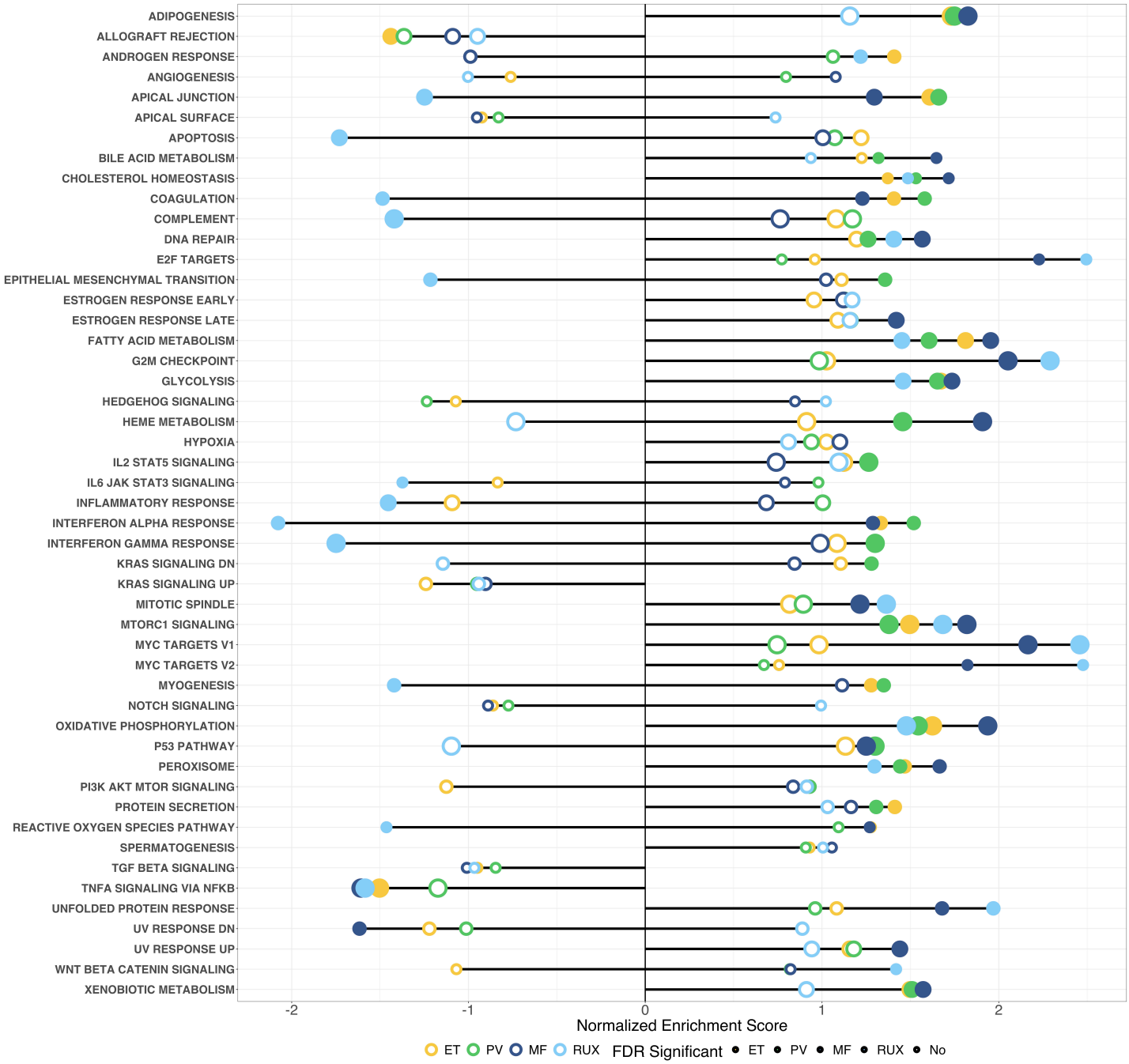
FIGURE 3



bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

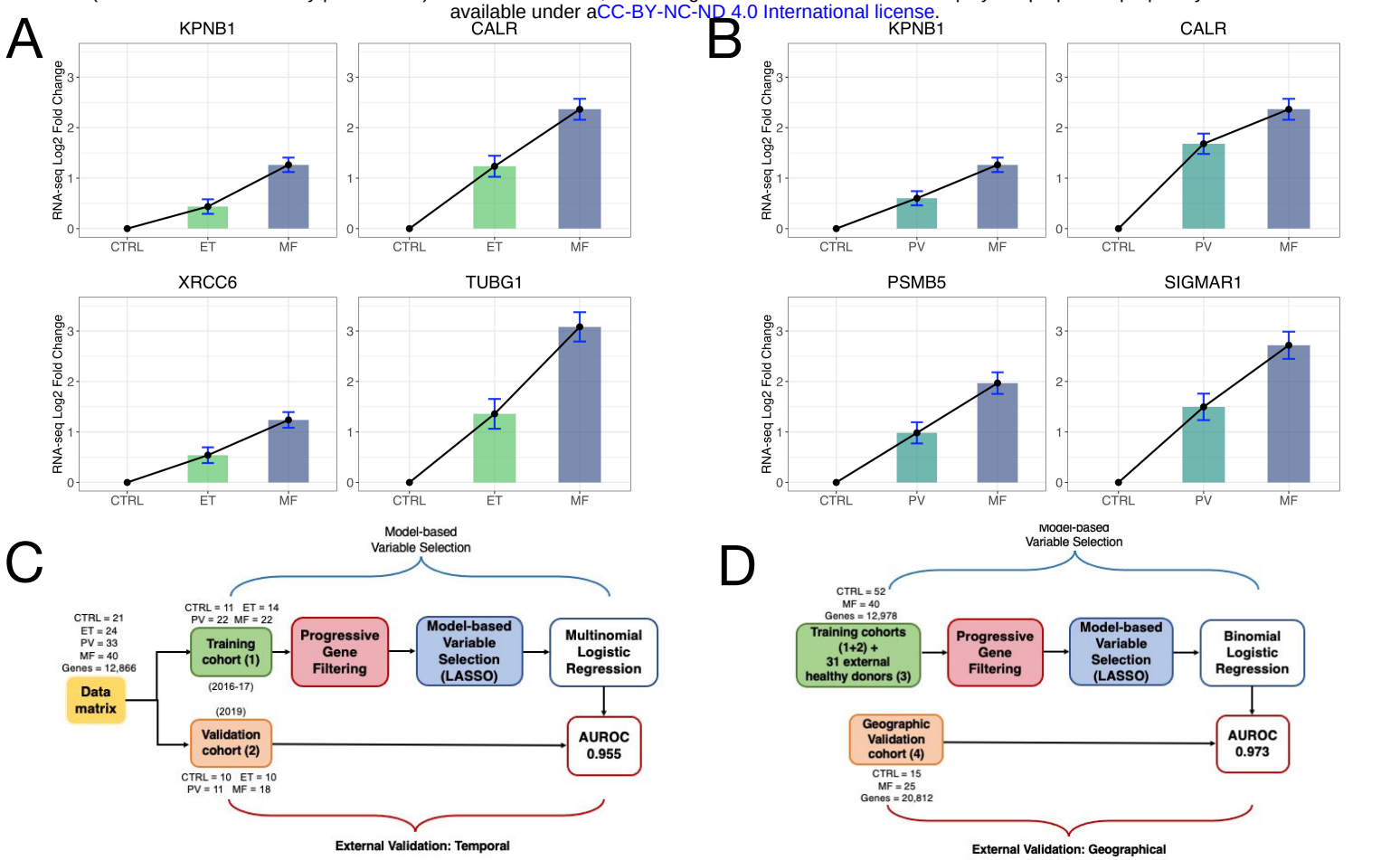
FIGURE 4

A



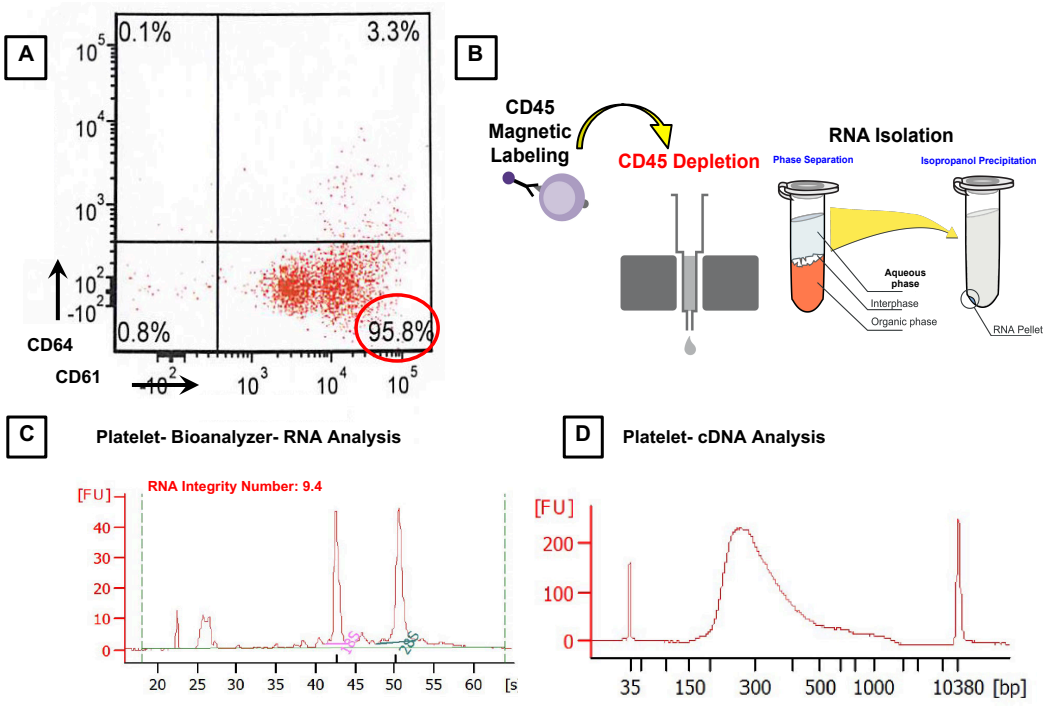
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE 5



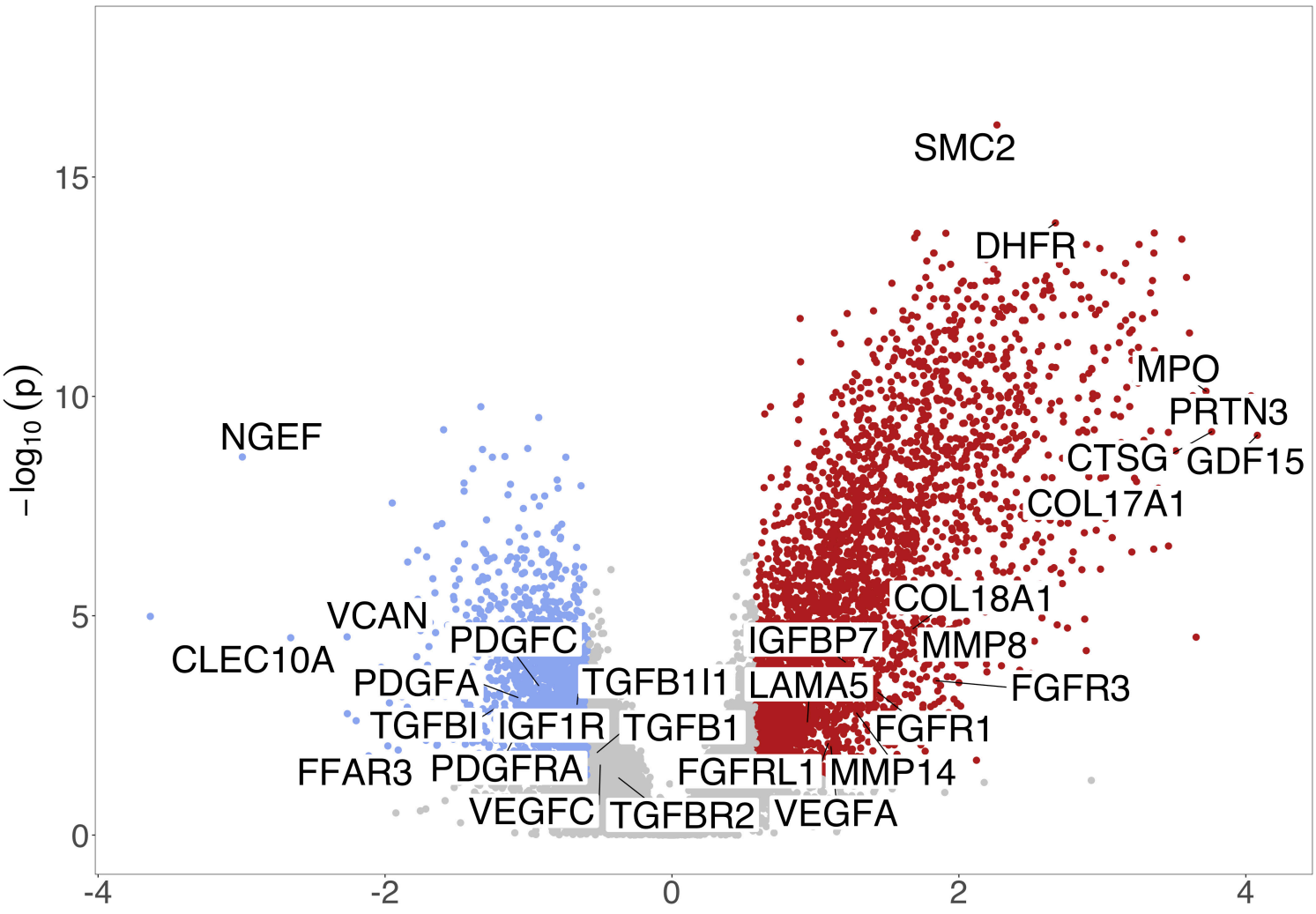
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE S1



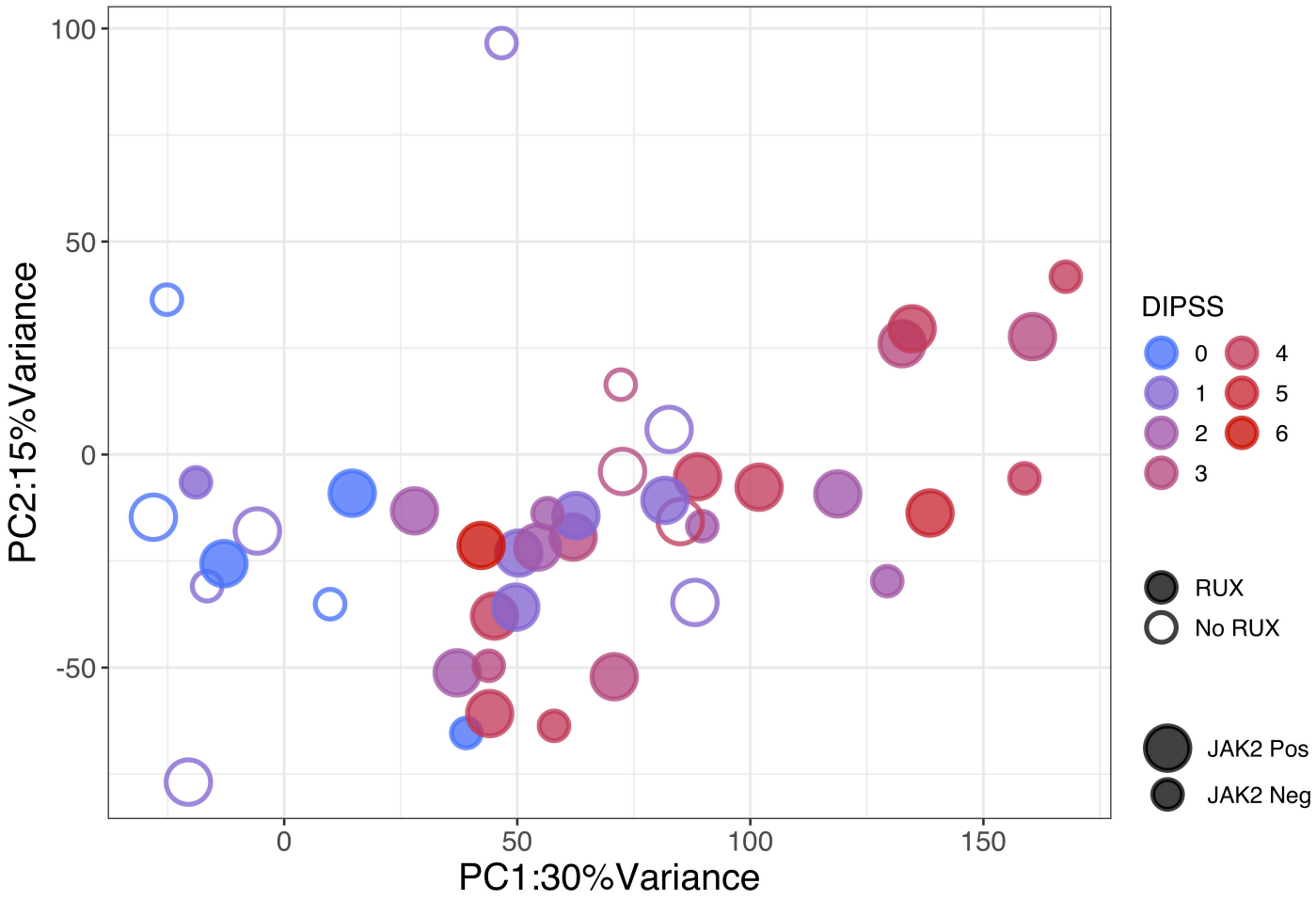
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE S2



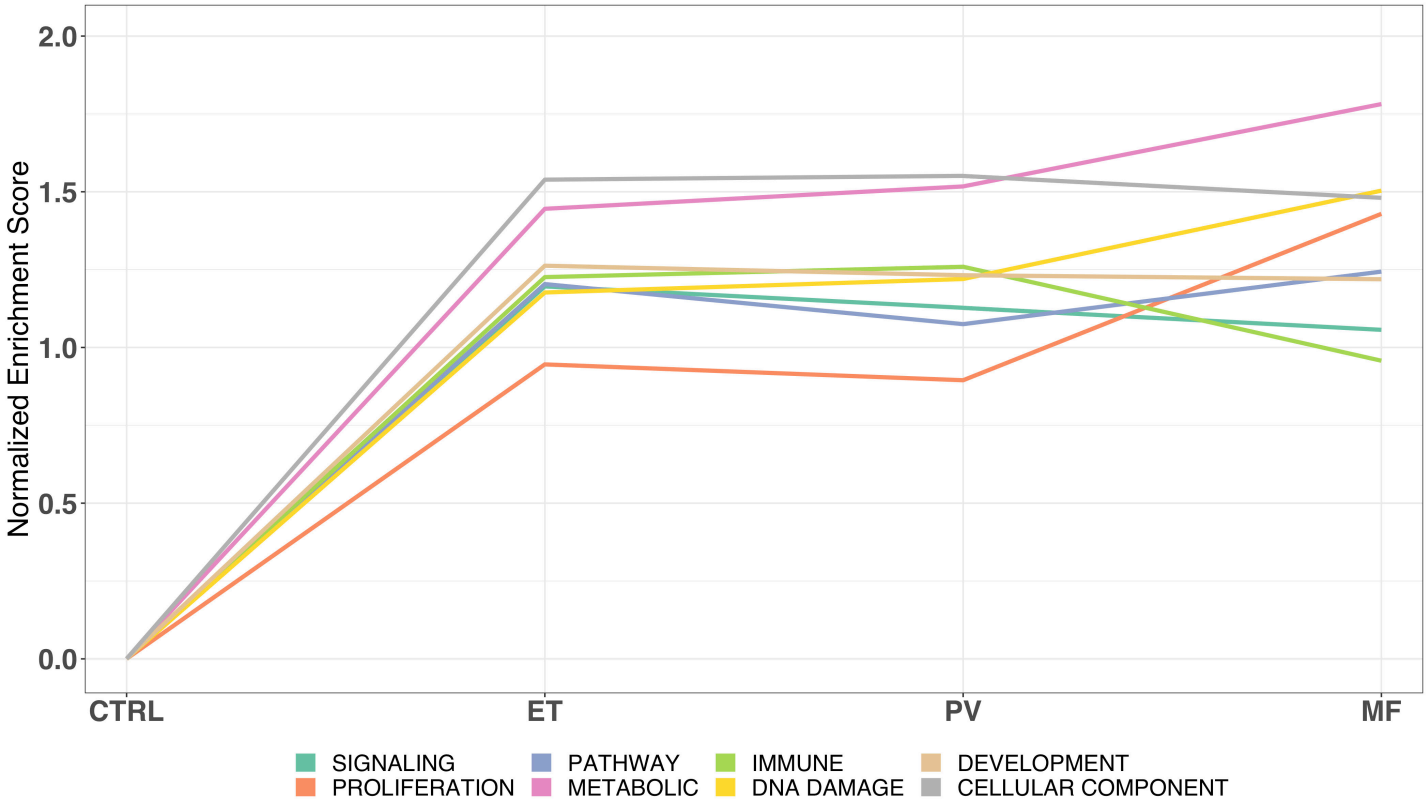
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE S3



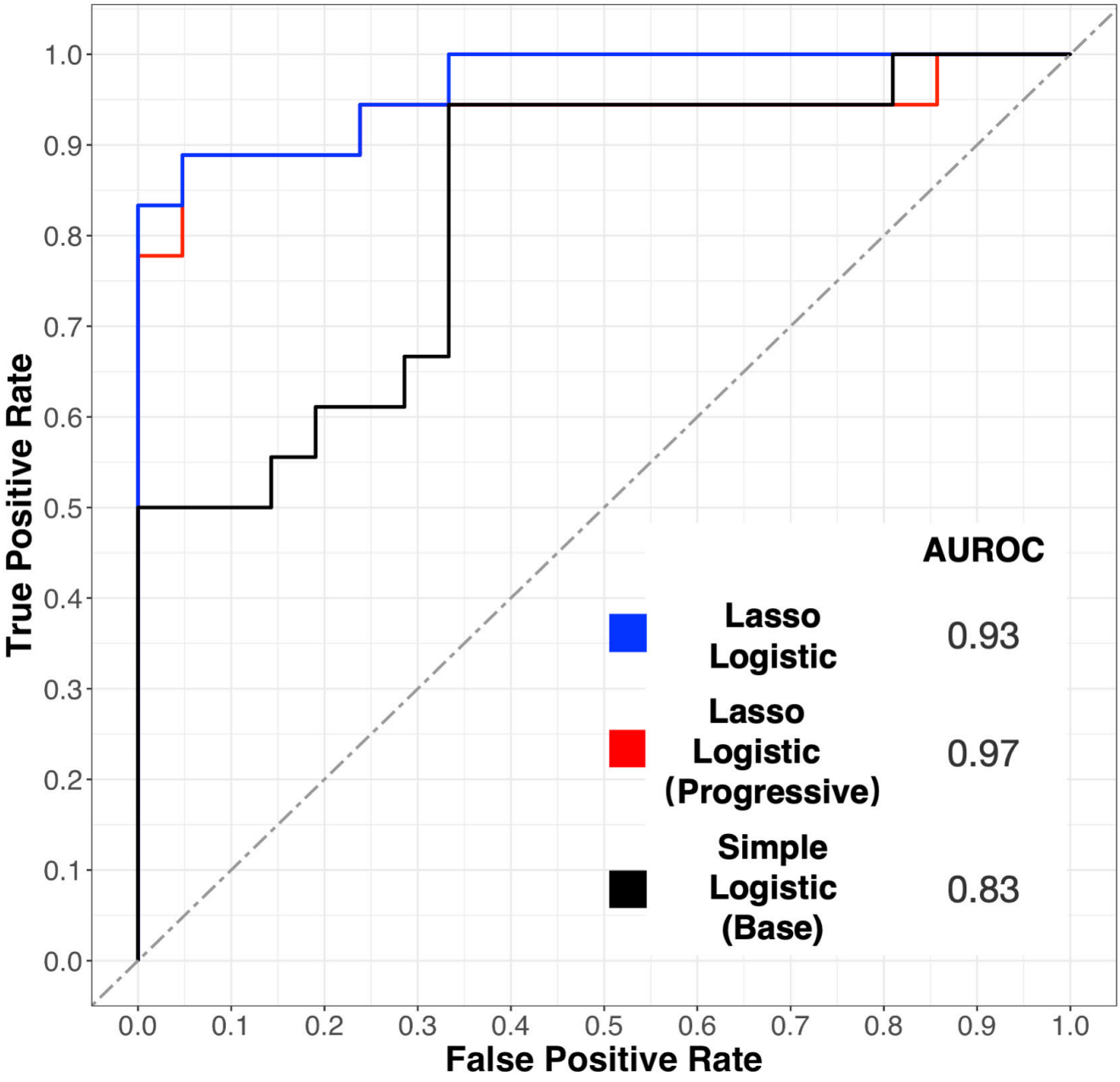
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE S4



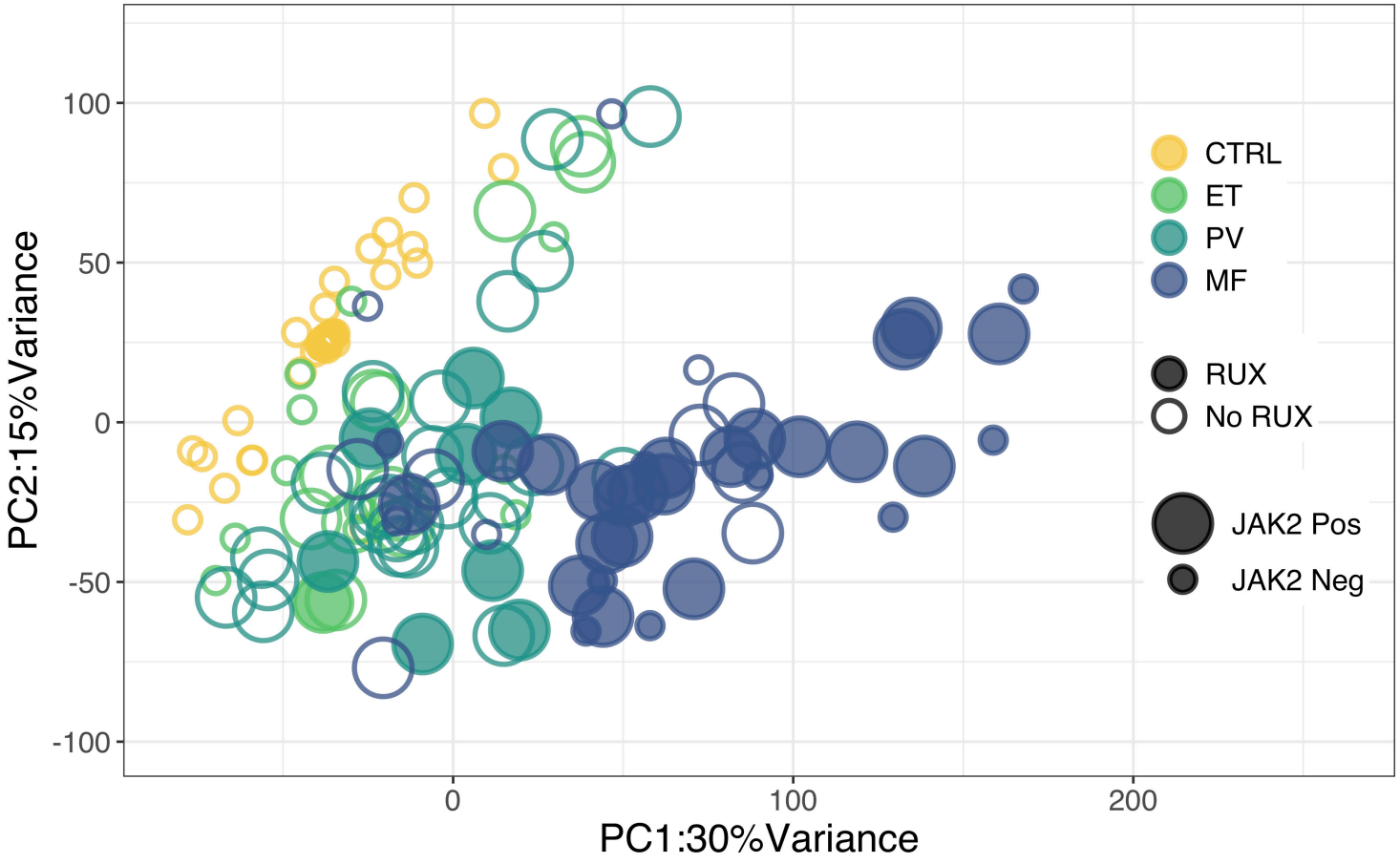
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE S5



bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE S6



bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S1A. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S1B. 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)

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TableS2A: Non-coding RNA ET vs CTRL FDR <0.05

Gene	baseMean	log2FoldChar	lfcSE	stat	pvalue	padj
C6orf223	75.0957846	3.01740668	0.39682251	6.82495279	8.80E-12	1.89E-09
TREML3P	465.150144	2.09076139	0.32879191	5.89156432	3.83E-09	3.56E-07
PTENP1	90.153903	-1.3414763	0.24299574	-5.2308836	1.69E-07	9.03E-06
TXLNGY	70.2890279	-1.150555	0.58601854	-5.1656862	2.40E-07	1.21E-05
WASIR2	43.3167126	2.40131956	0.41838737	4.93956149	7.83E-07	3.24E-05
LINC00926	49.189988	-0.8460768	0.51299087	-4.9383514	7.88E-07	3.25E-05
GSTT2	12.2152526	2.55343922	0.56455669	4.69096864	2.72E-06	8.92E-05
C15orf54	1110.16259	-2.0410416	0.3592438	-4.6843096	2.81E-06	9.20E-05
HHLA3	37.3416835	1.20795875	0.25156831	4.50879144	6.52E-06	0.00018558
LINC00888	45.0286505	1.49897176	0.32604995	4.35281255	1.34E-05	0.0003319
PRKY	36.6660387	-1.3818357	0.56903696	-4.3422151	1.41E-05	0.00034309
STAG3L4	72.7622771	0.92848143	0.20317944	4.33567135	1.45E-05	0.00035144
APTR	16.6293968	1.44901666	0.28977528	4.30802291	1.65E-05	0.00039174
SNHG11	223.496342	1.2259314	0.26232414	4.27792252	1.89E-05	0.00043732
LINC02324	150.716456	2.06322552	0.40319397	4.21767817	2.47E-05	0.00055134
ZNF542P	1859.97136	-1.6557883	0.33591721	-4.156471	3.23E-05	0.00067835
LINC01816	21.3715116	1.3807136	0.29931732	4.09022598	4.31E-05	0.0008491
CATSPER2P	57.1797883	0.9437991	0.22974592	3.99439539	6.49E-05	0.00116548
MT1L	22.0458816	1.35090139	0.42718061	3.93530705	8.31E-05	0.00142349
LINC00667	124.738641	-0.8112539	0.19697244	-3.9224974	8.76E-05	0.00147775
CRYZL2P	258.679365	0.9859327	0.24756806	3.87823669	0.00010522	0.00170493
SNHG20	1252.2434	1.2819875	0.30780752	3.86610294	0.00011059	0.00177189
MIR663AHG	47.2413146	-1.3000552	0.49967854	-3.7753047	0.00015981	0.00231027
LINC01686	30.4591717	-1.272036	0.2827493	-3.712935	0.00020487	0.00280708
DANCR	129.858339	1.26433721	0.28617305	3.70619362	0.0002104	0.00287308
XIST	477.026751	0.402829	0.60233484	-3.6774546	0.00023557	0.0031214
TRPC2	270.438119	1.07771801	0.30611814	3.65984563	0.00025237	0.00328639
MALAT1	20950.7202	-1.0068962	0.29864363	-3.6554008	0.00025678	0.00332703
SCARNA7	64.2222264	-1.295771	0.44242288	-3.6494433	0.00026281	0.00338469
LINC00642	12.3262931	-1.5927968	0.40141176	-3.6265377	0.00028725	0.00362325
CA5BP1	74.5936639	0.80044947	0.20835618	3.59252265	0.00032749	0.0039638
CLUHP3	13.3693335	-1.7876459	0.53150826	-3.5820634	0.00034089	0.00407612
LINC01137	100.470031	-1.1064498	0.27397923	-3.5463582	0.0003906	0.00456854
LINC02397	8.39863836	-0.4896785	0.52319022	-3.5155152	0.0004389	0.00491624
MAP3K14	32.5368126	-1.0545683	0.35999171	-3.4896245	0.0004837	0.00527618
LINC00665	11.3426597	-1.1114831	0.34125107	-3.4647271	0.00053077	0.00563905
DANT2	19.2328865	1.30273456	0.40333808	3.46491806	0.00053039	0.00563905
BCYRN1	65.7484022	0.98090182	0.2570155	3.44071329	0.00058018	0.00602874
GVINP1	109.956496	-1.0567755	0.33540489	-3.3952595	0.00068564	0.0069025
LINC00847	59.7405718	-0.6915249	0.18909998	-3.3837347	0.00071507	0.00713186
ANKRD18DP	27.8528501	-1.242929	0.3179281	-3.2793234	0.00104056	0.00938842
MRPS31P5	21.5300564	-0.8834363	0.28268472	-3.262172	0.00110562	0.00984423
STAG3L5P	23.7187195	-1.1991666	0.38147882	-3.2496866	0.00115532	0.01015576
SNHG17	97.5070658	0.71664689	0.20435316	3.22397972	0.00126422	0.01076833
MIR22HG	220.562225	0.85378561	0.25582948	3.21620071	0.001299	0.01100373
RRP7BP	18.7179955	-0.8557298	0.41881081	-3.1437826	0.00166779	0.01319669
SCARNA10	386.403772	-1.1667638	0.48238085	-3.1418756	0.00167869	0.01326663
LINC01128	87.4916407	-1.2902006	0.35116866	-3.1249391	0.00177842	0.01387577
H19	313.096933	-1.4106261	0.40469355	-3.0841625	0.00204126	0.01543059
ZNF767P	78.9102273	-0.6771986	0.21151393	-3.0634223	0.00218821	0.01627244
HERC2P9	578.492242	0.6527631	0.20593559	3.05601307	0.00224302	0.01657589
SCARNA5	74.3397954	-0.7742817	0.51523561	-3.0345135	0.00240924	0.01748753
LINC02035	11.7211718	-1.2695788	0.49286829	-3.0326511	0.00242416	0.01755161
SNHG7	251.166361	0.69348644	0.21685007	3.0289978	0.00245367	0.01766584
DLEU2	235.773583	-0.7535201	0.2241744	-3.0166561	0.0025558	0.01823786
GOLGA6L17	55.5492253	1.06601878	0.33402009	2.95993916	0.003077	0.02085809
FAM86C2P	18.425204	-1.0645685	0.29991799	-2.9411981	0.00326945	0.02172768
NEAT1	9909.64484	0.7001828	0.25633566	2.90839679	0.00363287	0.02330036
SCARNA6	22.7104976	-0.7153145	0.53288452	-2.9014381	0.00371454	0.0236986
ZNF702P	21.8809927	-0.8420264	0.27417687	-2.8990126	0.0037434	0.02381927
LINC00861	46.3183483	-1.0949497	0.5138199	-2.8733688	0.0040612	0.02523712
FIRRE	37.6610274	1.02952133	0.31430573	2.85820879	0.0042604	0.02604005
ADCY10P1	11.6649385	0.92403279	0.31285727	2.81794766	0.00483317	0.02840688
MIR3142HG	70.9205333	-0.997961	0.32252358	-2.7559206	0.00585272	0.03275385
TMEM198B	72.6128627	0.6382592	0.22147004	2.75136171	0.00593481	0.03308372
SNORD17	131.067946	-0.667343	0.46677813	-2.7442571	0.0060648	0.03367705
MHENCR	16.7707593	-0.6273489	0.2707653	-2.7211937	0.00650466	0.03561234
LINC00649	34.8121712	-0.8196205	0.31894703	-2.7171589	0.0065845	0.03591189
LINC00507	12.8647012	-0.937725	0.3151483	-2.6796713	0.00736945	0.03913138
DUXAP10	88.9531782	1.50094331	0.43338865	2.65319513	0.00797338	0.04158308
LINC00939	10.5088034	-1.3030862	0.45072083	-2.6300681	0.00853678	0.04367164
LINC01534	25.7796284	-0.8038282	0.27676242	-2.6119351	0.00900314	0.04540276
HRAT92	1485.80332	-1.0832085	0.34066179	-2.6074679	0.00912146	0.04577096
MYOSLID	87.8957944	-1.1170454	0.31880485	-2.5758862	0.00999835	0.04870838
MINCR	13.2401831	0.86586544	0.36167785	2.57431683	0.01004383	0.04887438

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TableS2B: Non-coding RNA PV vs CTRL FDR <0.05

NonCoding	baseMean	log2FoldChar	lfcSE	stat	pvalue	padj
C6orf223	75.0957846	3.50457905	0.37516835	7.74424677	9.62E-15	1.32E-12
TREML3P	465.150144	2.22703835	0.31177084	6.33483255	2.38E-10	1.24E-08
LINC00888	45.0286505	2.13237434	0.30857425	6.23805196	4.43E-10	2.18E-08
SNHG11	223.496342	1.6816479	0.25099007	5.96409366	2.46E-09	9.86E-08
PTENP1	90.153903	-1.4831523	0.23303634	-5.8924654	3.80E-09	1.43E-07
LINC02324	150.716456	2.81469222	0.38051391	5.65878374	1.52E-08	4.85E-07
DLEU2	235.773583	-1.3397294	0.21600145	-5.4394386	5.34E-08	1.45E-06
LINC00926	49.189988	-1.3962078	0.48804234	-5.1800064	2.22E-07	5.06E-06
XIST	477.026751	-0.3947063	0.58284501	-5.1332665	2.85E-07	6.27E-06
GSTT2	12.2152526	2.76495054	0.53909957	5.05602905	4.28E-07	8.81E-06
C15orf54	1110.16259	-2.1935255	0.33967559	-5.0521926	4.37E-07	8.96E-06
MT1L	22.0458816	1.95650187	0.39983386	5.04663161	4.50E-07	9.14E-06
LINC00667	124.738641	-0.9853913	0.19020589	-4.8295123	1.37E-06	2.37E-05
ZFAS1	770.677886	-1.2114361	0.22524652	-4.8229585	1.41E-06	2.43E-05
LINC00989	7443.56582	-1.7140028	0.2883521	-4.7368801	2.17E-06	3.46E-05
HHLA3	37.3416835	1.21112993	0.24179048	4.66917186	3.02E-06	4.57E-05
LINC01686	30.4591717	-1.5719219	0.26994642	-4.6114996	4.00E-06	5.73E-05
WASIR2	43.3167126	2.21529958	0.39558587	4.59329224	4.36E-06	6.18E-05
ANKRD18DP	27.8528501	-1.6770181	0.30143038	-4.5711024	4.85E-06	6.74E-05
LINC01816	21.3715116	1.52578952	0.28522809	4.55807406	5.16E-06	7.11E-05
LINC00649	34.8121712	-1.5136472	0.30369137	-4.5565699	5.20E-06	7.15E-05
GOLGA6L17I	55.5492253	1.63824758	0.31605436	4.51703481	6.27E-06	8.32E-05
ZNF542P	1859.97136	-1.798552	0.31826653	-4.5072002	6.57E-06	8.62E-05
LINC00642	12.3262931	-2.042877	0.37968977	-4.503952	6.67E-06	8.75E-05
CRYZL2P	258.679365	1.08465028	0.23747578	4.33426424	1.46E-05	0.0001652
STAG3L4	72.7622771	0.90594418	0.1961482	4.33093021	1.48E-05	0.00016714
TXLNGY	70.2890279	-0.9067054	0.56425165	-4.2988946	1.72E-05	0.00018892
SNHG20	1252.2434	1.40246076	0.29260331	4.27605526	1.90E-05	0.00020689
CA5BP1	74.5936639	0.91103632	0.20094491	4.18177511	2.89E-05	0.00029187
ZNF702P	21.8809927	-1.2225697	0.26151007	-4.1777818	2.94E-05	0.00029588
MALAT1	20950.7202	-1.171433	0.28424084	-4.1329002	3.58E-05	0.00034627
BMS1P1	79.8644082	-1.7747868	0.37809336	-4.0680245	4.74E-05	0.00043264
LINC01137	100.470031	-1.2547907	0.26157908	-4.0597675	4.91E-05	0.00044601
DANCR	129.858339	1.31387428	0.27296095	4.05506872	5.01E-05	0.0004522
ZNF767P	78.9102273	-0.8299822	0.20414764	-3.7913733	0.00014982	0.00111033
HRAT92	1485.80332	-1.5534809	0.32263046	-3.7672017	0.00016509	0.00119394
MYOSLID	87.8957944	-1.5261351	0.30266783	-3.7608812	0.00016932	0.00122109
TMEM198B	72.6128627	0.84693583	0.21305753	3.69494549	0.00021993	0.00150595
APTR	16.6293968	1.21911742	0.27828674	3.63909417	0.0002736	0.00179506
CLUHP3	13.3693335	-1.7654836	0.50442982	-3.6277408	0.00028591	0.00186231
GVINP1	109.956496	-1.18737	0.31763794	-3.6259605	0.00028789	0.00187156
TRPC2	270.438119	1.02325572	0.29103878	3.61105738	0.00030495	0.00196372
LINC02397	8.39863836	-0.9272924	0.49896179	-3.5157165	0.00043857	0.00265444
PRKY	36.6660387	-0.6522394	0.5446327	-3.5019697	0.00046183	0.002756
SCARNA10	386.403772	-1.1561412	0.45680221	-3.4994599	0.0004662	0.00277564
LINC00847	59.7405718	-0.7064881	0.18273056	-3.497419	0.00046978	0.00279308
SNHG17	97.5070658	0.7648344	0.19732728	3.48693675	0.00048859	0.00287842
LINC01684	72.9422645	-0.8374995	0.22681547	-3.4154821	0.00063669	0.00357247
SCARNA7	64.2222264	-1.0862699	0.41597786	-3.389683	0.00069974	0.00384851
SCARNA5	74.3397954	-0.8163805	0.48934128	-3.3778654	0.00073051	0.00398082
DUBR	187.751247	-1.0820991	0.25786418	-3.3186905	0.00090441	0.00476108
FIRRE	37.6610274	1.16533918	0.298336	3.31347572	0.00092144	0.0048251
LINC00665	11.3426597	-1.0038774	0.31759928	-3.3081006	0.00093931	0.00488685
LINC02444	13.0463175	-1.0599281	0.32132346	-3.3021517	0.00095946	0.00497559
LINC01534	25.7796284	-1.0153562	0.26444037	-3.2969935	0.00097726	0.00503857
NBR2	58.6527179	-0.9722597	0.25788029	-3.2919838	0.00099483	0.00511367
HERC2P9	578.492242	0.69898328	0.19878371	3.26856103	0.00108096	0.00546254
CATSPER2P	57.1797883	0.73922277	0.22145466	3.26239388	0.00110476	0.00554576
LINC01814	233.117887	-0.6690812	0.18260855	-3.2504475	0.00115224	0.0057216
MIR663AHG	47.2413146	-0.9156299	0.47258609	-3.2443505	0.00117719	0.00581285
DUXAP8	32.55787	2.14072931	0.45810437	3.24227464	0.0011858	0.00584315
MINCR	13.2401831	1.11772442	0.34085298	3.18798989	0.00143266	0.00681445
MAP3K14	32.5368126	-0.9413573	0.33839139	-3.1511518	0.00162628	0.00754008
SNORD17	131.067946	-0.6403155	0.44086192	-3.1340518	0.0017241	0.00793643
LINC01750	177.053867	-1.4234717	0.33241096	-3.127026	0.00176584	0.00807081
LINC00339	26.2887957	-0.6963345	0.22024313	-3.0888501	0.00200933	0.00896182
DUXAP10	88.9531782	1.6724524	0.40923458	3.08881751	0.00200955	0.00896182
KIR3DX1	14.8612819	-1.2451781	0.33497592	-3.0635191	0.0021875	0.0095925
LINC02506	38.9670994	-1.2046384	0.61525799	-3.0347445	0.0024074	0.01035905
CYTOR	2654.42929	-0.8516447	0.22981261	-3.0289564	0.002454	0.01051483
LINC00861	46.3183483	-1.6546468	0.48926921	-3.0058821	0.00264812	0.01115243
LINC00939	10.5088034	-1.3816894	0.42167137	-2.9611514	0.00306491	0.01252244
CD99P1	55.4656821	0.65326192	0.222176	2.9205659	0.00349396	0.01385738
MIR924HG	11.5870364	-1.0780476	0.33546596	-2.9006194	0.00372426	0.01459973
PVT1	15.5919309	0.94085793	0.28946548	2.87051385	0.00409805	0.01579555
SNHG7	251.166361	0.6466723	0.209039	2.84118188	0.00449467	0.01705852
MIR22HG	220.562225	0.74126187	0.24501429	2.81135092	0.0049334	0.01839266
LINC02280	97.8623144	-1.0964884	0.29677203	-2.809708	0.00495865	0.01845471
SCARNA6	22.7104976	-0.7162347	0.50641926	-2.794333	0.00520069	0.01918351
DPY19L2P2	143.786668	-0.5757799	0.18913839	-2.7860995	0.00533465	0.01954317
LINC02470	32.4988864	-0.9286029	0.4958935	-2.7704619	0.00559769	0.02032162
MRPS31P5	21.5300564	-0.6547881	0.26799378	-2.6799403	0.00736353	0.02521529
MHENCR	16.7707593	-0.6072653	0.25783567	-2.6691333	0.00760473	0.02587055
LINC00987	328.482169	-0.7574809	0.24131329	-2.6348109	0.00841842	0.02813282
ADCY10P1	11.6649385	0.82871557	0.29791244	2.63415321	0.00843474	0.02816543
MIR600HG	16.3163891	-1.061752	0.40965187	-2.601873	0.00927162	0.03033791
MFSD14C	26.8218097	0.63014847	0.22700835	2.5819244	0.00982511	0.03177724
MIR3142HG	70.9205333	-0.8877215	0.30563757	-2.5790587	0.009907	0.03199383
LINC00507	12.8647012	-0.877875	0.29571184	-2.5736074	0.01006444	0.03243715
RRP7BP	18.7179955	-0.7307527	0.3925651	-2.5630313	0.01037627	0.03324231
CASC15	109.834567	-1.1867731	0.29877595	-2.5444431	0.01094522	0.03477924
LINC02384	244.239028	-0.7122509	0.26305089	-2.5379227	0.01115126	0.03530318
FAM86C2P	18.425204	-0.9222917	0.28345133	-2.5240664	0.0116006	0.03641214
FLJ37453	56.4820867	-0.555238	0.20008201	-2.5023982	0.01233551	0.0383026
CCT6P3	1203.53388	0.75605772	0.2797982	2.4911242	0.01273396	0.03930785
H19	313.096933	-1.1234946	0.38181068	-2.4894083	0.01279559	0.03943188
STAG3L5P	23.7187195	-0.9215094	0.35668759	-2.4602428	0.0138843	0.04206156
JPX	66.9551008	-0.4771308	0.18121018	-2.4525777	0.01418367	0.04276708
LINC00641	26.8406871	-0.6002245	0.37330883	-2.4526097	0.01418241	0.04276708
LINC02043	120.520369	-0.6538486	0.23137881	-2.4482651	0.0143546	0.04316108
C22orf46	49.6938689	0.54728016	0.20942785	2.4267886	0.01523313	0.04528732
MIAT	66.4307001	-1.1591788	0.46142626	-2.4248697	0.01531388	0.0455031
CABIN1	638.674222	0.56993214	0.21951572	2.42385113	0.0153569	0.04560985
NEAT1	9909.64484	0.53094747	0.24554853	2.41930854	0.01555004	0.04600296
CROCCP3	13.3700881	-0.8506598	0.36721557	-2.4181241	0.01560075	0.04609998

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TableS2C: Non-coding RNA MF vs CTRL FDR <0.05

NonCoding	baseMean	log2FoldChar	lfcSE	stat	pvalue	padj
C6orf223	75.0957846	3.84925472	0.38666035	7.84777206	4.23E-15	6.26E-13
LINC00888	45.0286505	2.42391258	0.32199856	6.56687072	5.14E-11	1.73E-09
PTENP1	90.153903	-1.8011229	0.24670387	-6.5497114	5.76E-11	1.92E-09
C15orf54	1110.16259	-2.9640006	0.35355393	-6.2452986	4.23E-10	1.08E-08
MT1L	22.0458816	2.78634493	0.40966418	6.22977483	4.67E-10	1.19E-08
C19orf48	48.0423819	2.9065684	0.36954989	6.02052087	1.74E-09	3.61E-08
LINC00989	7443.56582	-2.2948674	0.30263259	-6.0085973	1.87E-09	3.81E-08
ZNF542P	1859.97136	-2.5229016	0.33258228	-5.8678926	4.41E-09	7.98E-08
LINC01534	25.7796284	-1.9739679	0.28211136	-5.8661839	4.46E-09	8.05E-08
GSTT2	12.2152526	3.47898362	0.53431154	5.75026388	8.91E-09	1.45E-07
ANKRD18DP	27.8528501	-2.2382841	0.31762958	-5.6985565	1.21E-08	1.89E-07
CYTOR	2654.42929	-1.6378818	0.24246011	-5.6714256	1.42E-08	2.17E-07
LINC01684	72.9422645	-1.4909858	0.24087484	-5.6042808	2.09E-08	3.01E-07
ZFAS1	770.677886	-1.4117909	0.23772439	-5.2548054	1.48E-07	1.61E-06
HRAT92	1485.80332	-2.3223765	0.33688713	-5.1532328	2.56E-07	2.54E-06
DUBR	187.751247	-1.7233738	0.27190998	-5.0761456	3.85E-07	3.59E-06
ANKRD19P	19.471925	1.74989585	0.29287354	5.07476435	3.88E-07	3.61E-06
LINC21686	30.4591717	-1.8406634	0.28538987	-4.9921145	5.97E-07	5.22E-06
SNHG16	191.220158	2.14395362	0.32520774	4.98536418	6.18E-07	5.38E-06
LINC00649	34.8121712	-1.856628	0.31974426	-4.9732403	6.58E-07	5.67E-06
LINC01750	177.053867	-2.1716641	0.34657725	-4.8166023	1.46E-06	1.14E-05
LINC01814	233.117887	-1.0378028	0.19343576	-4.7641016	1.90E-06	1.42E-05
LINC00987	328.482169	-1.4109395	0.25466342	-4.6931536	2.69E-06	1.91E-05
LINC01137	100.470031	-1.5641997	0.27574691	-4.690281	2.73E-06	1.94E-05
CASC15	109.834567	-2.0423557	0.31382911	-4.6640021	3.10E-06	2.17E-05
DLEU2	235.773583	-1.2102223	0.22806966	-4.5884905	4.46E-06	2.96E-05
LINC02324	150.716456	2.37210095	0.39242154	4.57539	4.75E-06	3.11E-05
ZNF767P	78.9102273	-1.0880231	0.21682766	-4.567845	4.93E-06	3.21E-05
PLEKHA8P1	621.977908	-1.2849808	0.24211292	-4.5492048	5.38E-06	3.46E-05
MYOSLID	87.8957944	-2.0213922	0.31770192	-4.5468155	5.45E-06	3.49E-05
LINC02280	97.8623144	-1.8474682	0.31193838	-4.5421083	5.57E-06	3.56E-05
LINC00504	1120.33438	-1.0354366	0.2059126	-4.5242059	6.06E-06	3.83E-05
NBR2	58.6527179	-1.4207186	0.27250788	-4.477846	7.54E-06	4.63E-05
LINC01089	894.053589	-1.3348437	0.24987274	-4.4707395	7.79E-06	4.77E-05
LINC02043	120.520369	-1.2657337	0.24474069	-4.437156	9.12E-06	5.47E-05
DANCR	129.858339	1.52853179	0.28653605	4.38466641	1.16E-05	6.72E-05
PSMB1	170.125138	1.09773111	0.22453505	4.35147346	1.35E-05	7.65E-05
SDHAP2	145.012207	-1.2303083	0.26217482	-4.3387713	1.43E-05	8.05E-05
LINC00892	167.397507	-1.7056888	0.30722265	-4.3298409	1.49E-05	8.34E-05
BMS1P1	79.8644082	-2.012152	0.39003967	-4.3162698	1.59E-05	8.78E-05
SENCR	946.214517	-1.3136766	0.26392184	-4.3116623	1.62E-05	8.93E-05
RRN3P2	84.2678934	-1.1188777	0.23253151	-4.2440865	2.19E-05	0.00011582
MIR3142HG	70.9205333	-1.6739083	0.32088788	-4.2413144	2.22E-05	0.00011708
FLJ37453	56.4820867	-1.012941	0.21270854	-4.2371584	2.26E-05	0.00011878
CENPBD1	14.2066354	1.53308695	0.30915007	4.2338025	2.30E-05	0.00012042
LINC00667	124.738641	-0.9048188	0.20115045	-4.1864051	2.83E-05	0.00014401
WASIR2	43.3167126	2.06636628	0.40641727	4.13441385	3.56E-05	0.00017323
CRDCCP2	98.6861521	2.19923408	0.38260752	4.08341571	4.44E-05	0.00020808
FAM86DP	14.5169738	1.52483018	0.31964464	4.07371578	4.63E-05	0.00021569
AATBC	11.395935	-1.8680216	0.53971244	-4.0432788	5.27E-05	0.00024142
CABIN1	638.674222	0.99587112	0.23165777	4.03590745	5.44E-05	0.00024781
LINC00926	49.189988	-1.1995556	0.48788134	-4.0348291	5.46E-05	0.00024877
ANKRD36BP1	325.779604	-0.9239995	0.40056465	-4.0064062	6.16E-05	0.00027647
LINC20847	59.7405718	-0.8609489	0.19395813	-3.9885236	6.65E-05	0.00029517
SNHG8	62.2331747	1.77596926	0.33745158	3.96351214	7.39E-05	0.00032378
PATL2	85.1502874	-1.1535473	0.26285674	-3.9452764	7.97E-05	0.00034576
NORAD	43931.6513	-1.2349456	0.26471024	-3.9051679	9.42E-05	0.00039956
SDAD1P1	50.8608729	-1.0286205	0.23667886	-3.8901865	0.00010017	0.00042047
XIST	477.026751	0.05334253	0.56999316	3.8870875	0.00010145	0.00042491
OR2L1P	26.9232927	-1.7038327	0.35570319	-3.8855423	0.0001021	0.00042748
LINC01011	1857.31016	-1.3047424	0.27794532	-3.8132315	0.00013716	0.00055165
NAPSB	26.7641143	-1.9246707	0.46870081	-3.7515264	0.00017576	0.0006823
MAP3K14	32.5368126	-1.3673378	0.35411677	-3.7362889	0.00018676	0.00071876
KIR3DX1	14.8612819	-1.7346787	0.35274896	-3.7314849	0.00019035	0.00073151
RNY3	66.6363011	-2.245866	0.41199114	-3.715451	0.00020284	0.00077189
LINC00642	12.3262931	-1.9131791	0.39143144	-3.7090913	0.000208	0.00078988
KCNQ1OT1	1447.97863	-0.8490558	0.20376262	-3.7030339	0.00021304	0.00080639
LINC00664	11.8542399	1.92932098	0.47598487	3.69481315	0.00022005	0.00082879
TREML3P	465.150144	1.31624434	0.32613807	3.68898416	0.00022515	0.00084504
LINC02384	244.239028	-1.1297206	0.27690621	-3.6680042	0.00024445	0.00090768
DSTNP2	38.619312	-1.0716359	0.24258436	-3.6545063	0.0002576	0.000947
LINC00458	12.1236705	-1.5995151	0.33898291	-3.6530438	0.00025915	0.00095193
CENPBD1P1	838.764428	-1.7548008	0.35406296	-3.636671	0.00027618	0.00100777
TXLNGY	70.2890279	-1.1367665	0.55391361	-3.63007	0.00028334	0.00102977
MHENCR	16.7707593	-0.9736327	0.27490611	-3.5390558	0.00040156	0.00139824
LINC00861	46.3183483	-2.2999795	0.49114025	-3.5380516	0.00040309	0.00140319
PMS2P3	33.3887481	-0.8416465	0.20858349	-3.5288822	0.00041732	0.00144723
LINC00339	26.2887957	-0.8622708	0.23366017	-3.5206165	0.00043055	0.00148629
NATB8	484.35185	-1.2873628	0.29952305	-3.5030674	0.00045993	0.00157255
CA5BP1	74.5936639	0.83085313	0.21244458	3.50239685	0.00046109	0.00157542
TSSC2	37.4347069	-0.9203052	0.22763462	-3.4978156	0.00046909	0.00160001
LINC02444	13.0463175	-1.268217	0.33813191	-3.4949496	0.00047415	0.0016143
LINC02470	32.4988864	-1.6620767	0.49562	-3.4642924	0.00053163	0.00178262
LINC01151	213.576384	-1.537713	0.32492946	-3.4635869	0.00053302	0.00178684
LINC00853	238.87796	-1.2414011	0.27894586	-3.4501682	0.00056024	0.00186688
ZNF876P	54.5082085	-1.4674708	0.34348969	-3.4390529	0.00058375	0.00193472
LINC00654	19.721148	1.16059595	0.29309285	3.43792025	0.0005862	0.00194033
LINC01252	165.339	-1.2212189	0.29084044	-3.4009394	0.00067155	0.00217412
DDX11L2	1595.81206	-1.8195402	0.37295676	-3.332444	0.00086087	0.00268573
PVT1	15.5919309	1.18516535	0.30260021	3.30211314	0.00095959	0.00295221
LINC00507	12.8647012	-1.2613179	0.31369938	-3.2873518	0.00101134	0.00309514
TPTEP1	11751.0583	-1.2727427	0.30596783	-3.2703809	0.00107403	0.00325446
LINC00211	132.800892	-1.0251375	0.25600732	-3.2697174	0.00107655	0.00326133
LINC01128	87.4916407	1.33945109	0.34191928	3.24643533	0.0011686	0.00349819
DPY19L2P2	143.786668	-0.7117207	0.20040391	-3.1993383	0.00137743	0.00402067
MEG3	80.2978477	1.73620284	0.38197238	3.19780244	0.00138479	0.00403949
LINC01023	21.3273686	-1.254761	0.28902067	-3.1963476	0.00139179	0.00405407
LINC01554	62.3176692	-1.1036027	0.28563993	-3.1802952	0.00147125	0.0042442
RP9P	8.68990266	1.54067348	0.35805712	3.13658874	0.00170926	0.00482371
LINC00887	104.333292	-0.822617	0.22478043	-3.1289935	0.00175406	0.00492854
CLUHP3	13.3693335	-2.0200173	0.50300891	-3.1246873	0.00177994	0.0049958
LINC00939	10.5088034	-1.7572943	0.43158763	-3.0936198	0.00197731	0.00548158
ZNF271P	3339.00226	-0.8802831	0.23609113	-3.0917169	0.00199003	0.00551091
LINC01795	28.9384037	-1.2958961	0.30171105	-3.0898006	0.00200291	0.00554539
RPL13P5	75.2700979	-0.9073819	0.23262205	-3.0680376	0.0021547	0.00590968
GAS5	176.304687	0.95379675	0.25178637	3.06538052	0.00217393	0.00595609
C22orf46	49.6938689	0.7239239	0.22108964	3.05586455	0.00224413	0.00612883
SMG1P6	57.2689943	-0.8151813	0.24537983	-3.046962	0.00231167	0.00628794
LINC01063	26.1884359	-1.3784185	0.34543662	-3.0228055	0.00250443	0.00674101
STAG3L4	72.7622771	0.66863518	0.20761899	3.00092844	0.00269158	0.00715788
LINC01184	38.5808434	0.80377467	0.23940408	2.99345573	0.00275838	0.00730532
FAM86C2P	18.425204	-1.1628587	0.29859267	-2.965532	0.0030216	0.00788398
SMG1P3	644.071646	0.81256718	0.24436196	2.94867348	0.00319141	0.00826503
PMS2CL	25.8032747	0.83094234	0.24927838	2.92143307	0.00348425	0.00890158
TMEM198B	72.6128627	0.69179423	0.22548098	2.91508497	0.00355592	0.00905232
PRKY	36.6660387	-0.9333695	0.53703017	-2.9096474	0.00361837	0.00918223
METTL21EP	278.204212	-1.5113507	0.3636686	-2.8742766	0.00404954	0.01009717
MIR663AHG	47.2413146	-1.3251973	0.47518648	-2.8660008	0.00415693	0.01031298
ADCY10P1	11.6649385	0.96994059	0.3120233	2.8274912	0.00469143	0.01145567
RNF216P1	64.6137973	-0.5634066	0.19736585	-2.82227	0.0047685	0.01161081
LINC01278	58.7808631	-0.5292111	0.17725508	-2.8042596	0.00504323	0.01218748
GOLGA6L17I	55.5492253	1.03013772	0.33046056	2.78566006	0.00534189	0.0128297
LINC21609	208.697433	-0.8548099	0.25354358	-2.7854671	0.00534507	0.01283494
LUCAT1	48.8223033	-1.5891878	0.42466551	-2.7801663	0.00543311	0.01300509
ZNF37BP	75.4392195	0.89984018	0.26714029	2.77181393	0.00557449	0.01329158
MRPS31P5	21.5300564	-0.7478336	0.28337671	-2.7572935	0.0058282	0.01378666
LINC01419	38.8357312	-1.1366257	0.31345657	-2.7403589	0.00613721	0.01439587
GVINP1	109.956496	-1.0251732	0.33201393	-2.7308965	0.00631623	0.01475661
HHLA3	37.3416835	0.72769188	0.25488441	2.72533404	0.00642365	0.01499123
LINC00839	12.3545856	1.14473198	0.38260357	2.72484513	0.00643317	0.01500901
LINC01215	201.470396	-1.0340353	0.28407824	-2.7220529	0.00648778	0.01511895
FLJ31356	11.4187566	-1.0117298	0.3214889	-2.6762292	0.00744557	0.01704532
ZNF262P	64.8195105	0.67652434	0.21900717	2.61517745	0.0089181	0.01988671
ZNF702P	21.8809927	-0.7820876	0.2740289	-2.6150489	0.00892147	0.01988973
LINC02035	11.7211718	-1.3663674	0.4661166	-2.6137882	0.00895445	0.01993998
HERC2P9	578.492242	0.59817535	0.20991832	2.58594184	0.00971133	0.02136194
LINC02397	8.39863836	-0.8624786	0.49877722	-2.5666771	0.01026782	0.0223908
LINC00570	53.1436234	1.40997289	0.4416368	2.5463411	0.01088588	0.02348779
LINC00923	23.6829463	-0.8895214	0.28993549	-2.5461532	0.01089174	0.02349649
SCARNA5	74.3397954	-0.5831727	0.48933085	-2.5339695	0.01127786	0.02421981
SNORD17	131.067946	-0.5911693	0.4469745	-2.5113488	0.01202708	0.02560232
FTX	119.322397	-0.5502938	0.19548787	-2.4806369	0.01311479	0.02758008
PHKAP2	77.0350077	1.17583723	0.33398038	2.46398158	0.01374032	0.02872651
DUXAP8	32.557787	1.73132653	0.46302854	2.45978164	0.01390216	0.02899419
LINC01873	346.254692	-0.6039854	0.21317898	-2.4388367	0.01473463	0.03049802
GUSBP2	16.0448974	-0.6714681	0.2968501	-2.3976668	0.01649987	0.03363766
APTR	16.6293968	0.85641601	0.2931767	2.38717079	0.0169786	0.03444988
BCYRN1	65.7484022	0.69823226	0.25959698	2.38062361	0.01728336	0.03490099
DLEU1	258.981612	-0.6272033	0.22571218	-2.3766527	0.01747053	0.0353088
RNY1	16.7816357	-1.2523941	0.41652201	-2.3741376	0.01759	0.03551289
GNRHR2	25.8548559	-0.8182842	0.26204263	-2.3678358	0.01789247	0.03604267
SNHG11	223.496342	0.66829802	0.26456663	2.36201322	0.01817599	0.03650526
LINC21134	40.267996	-0.7470089	0.25443656	-2.3359298	0.0194949	0.03885692
RPL21P28	22.89044	0.83407889	0.31125076	2.30486828	0.02117395	0.04173163
MIR924HG	11.5870364	-0.817217	0.34696709	-2.2886215	0.02210136	0.04318895
CECR7	14.0445451	0.74367615	0.3060054	2.24545098	0.0247392	0.04755869

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S3A: Genes Upregulated in MF and Downregulated in the RUX-treated cohort

Gene	Name
APLP1	amyloid beta precursor like protein 1
IFI6	interferon alpha inducible protein 6
GCH1	GTP cyclohydrolase 1
CAMK2A	calcium/calmodulin dependent protein kinase II alpha
TNIP2	TNFAIP3 interacting protein 2
IFIT1	interferon induced protein with tetratricopeptide repeats 1
NDRG3	NDRG family member 3
IFIT2	interferon induced protein with tetratricopeptide repeats 2
PLEKHM1	pleckstrin homology and RUN domain containing M1
JOSD1	Josephin domain containing 1
MYLIP	myosin regulatory light chain interacting protein
FAM89B	family with sequence similarity 89 member B
SCIN	scinderin
NPL	N-acetylneuraminate pyruvate lyase
ODF3B	outer dense fiber of sperm tails 3B
EIF1AY	eukaryotic translation initiation factor 1A Y-linked
IFI27L2	interferon alpha inducible protein 27 like 2
ISCU	iron-sulfur cluster assembly enzyme

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S3B: Genes Downreglated in MF and Upregulated n the RUX-treated cohort

Gene	Name
LINC00926	long intergenic non-protein coding RNA 926
SEMA3C	semaphorin 3C
ADGRG7	adhesion G protein-coupled receptor G7
CXCR5	C-X-C motif chemokine receptor 5
LOC101927151	uncharacterized LOC101927151
SMC5	structural maintenance of chromosomes 5
PER3	period circadian regulator 3
SNORD17	small nucleolar RNA, C/D box 17
LMO7	LIM domain 7

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S4A: Gene Set Enrichment Pathway Analysis ET vs CTRL

NAME	SIZE	NES	FDR q-val	Lower Level	Upper Level
HALLMARK_FATTY_ACID_METABOLISM	129	1.8118953	0.007511497	FATTY_ACID_METABOLISM	metabolic
HALLMARK_ADIPOGENESIS	172	1.7318747	0.0114795	ADIPOGENESIS	development
HALLMARK_GLYCOLYSIS	146	1.6732337	0.016614022	GLYCOLYSIS	metabolic
HALLMARK_APICAL_JUNCTION	127	1.6097152	0.022050116	APICAL_JUNCTION	cellular component
HALLMARK_OXIDATIVE_PHOSPHORYLATION	183	1.6249464	0.024344018	OXIDATIVE_PHOSPHORYLATION	metabolic
HALLMARK_MTORC1_SIGNALING	190	1.4962126	0.06486081	MTORC1_SIGNALING	signaling
HALLMARK_XENOBIOTIC_METABOLISM	128	1.497276	0.07428097	XENOBIOTIC_METABOLISM	metabolic
HALLMARK_PEROXISOME	81	1.4680862	0.076201215	PEROXISOME	cellular component
HALLMARK_COAGULATION	78	1.406872	0.10213155	COAGULATION	immune
HALLMARK_ANDROGEN_RESPONSE	85	1.4086183	0.11024309	ANDROGEN_RESPONSE	signaling
HALLMARK_PROTEIN_SECRETION	91	1.4120365	0.11874847	PROTEIN_SECRETION	pathway
HALLMARK_ALLOGRAFT_REJECTION	151	-1.4392167	0.12846054	ALLOGRAFT_REJECTION	immune
HALLMARK_CHOLESTEROL_HOMEOSTASIS	64	1.372009	0.13062237	CHOLESTEROL_HOMEOSTASIS	metabolic
HALLMARK_TNFA_SIGNALING_VIA_NFKB	162	-1.5032148	0.14304653	TNFA_SIGNALING_VIA_NFKB	signaling
HALLMARK_INTERFERON_ALPHA_RESPONSE	89	1.3321728	0.16736604	INTERFERON_ALPHA_RESPONSE	immune
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWA	48	1.2765079	0.22986871	REACTIVE_OXYGEN_SPECIES_PATHWA	pathway
HALLMARK_MYOGENESIS	115	1.2783359	0.24290702	MYOGENESIS	development
HALLMARK_APOPTOSIS	138	1.2210994	0.3065594	APOPTOSIS	pathway
HALLMARK_BILE_ACID_METABOLISM	70	1.2247	0.31768104	BILE_ACID_METABOLISM	metabolic
HALLMARK_UV_RESPONSE_DN	109	-1.2210813	0.3375144	UV_RESPONSE_DN	DNA damage
HALLMARK_DNA_REPAIR	143	1.1955484	0.34101444	DNA_REPAIR	DNA damage
HALLMARK_KRAS_SIGNALING_UP	123	-1.2409247	0.38903093	KRAS_SIGNALING_UP	signaling
HALLMARK_UV_RESPONSE_UP	125	1.1575215	0.4082876	UV_RESPONSE_UP	DNA damage
HALLMARK_COMPLEMENT	156	1.079815	0.43777266	COMPLEMENT	immune
HALLMARK_P53_PATHWAY	162	1.1330146	0.44727734	P53_PATHWAY	proliferation
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	103	1.082889	0.44814357	UNFOLDED_PROTEIN_RESPONSE	pathway
HALLMARK_KRAS_SIGNALING_DN	72	1.1056865	0.44971326	KRAS_SIGNALING_DN	signaling
HALLMARK_IL2_STAT5_SIGNALING	158	1.1219251	0.45266697	IL2_STAT5_SIGNALING	signaling
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITIC	115	1.1106548	0.45790106	EPITHELIAL_MESENCHYMAL_TRANSITIO	development
HALLMARK_INTERFERON_GAMMA_RESPONSE	177	1.0849764	0.4614395	INTERFERON_GAMMA_RESPONSE	immune
HALLMARK_ESTROGEN_RESPONSE_LATE	135	1.0897365	0.46908423	ESTROGEN_RESPONSE_LATE	signaling
HALLMARK_WNT_BETA_CATENIN_SIGNALING	30	-1.0684173	0.4762125	WNT_BETA_CATENIN_SIGNALING	signaling
HALLMARK_G2M_CHECKPOINT	191	1.0252559	0.52272725	G2M_CHECKPOINT	proliferation
HALLMARK_PI3K_AKT_MTOR_SIGNALING	92	-1.1251472	0.5303781	PI3K_AKT_MTOR_SIGNALING	signaling
HALLMARK_HEDGEHOG_SIGNALING	21	-1.0717624	0.53431135	HEDGEHOG_SIGNALING	signaling
HALLMARK_HYPOXIA	138	1.0261731	0.5390404	HYPOXIA	pathway
HALLMARK_INFLAMMATORY_RESPONSE	143	-1.0926526	0.5442097	INFLAMMATORY_RESPONSE	immune
HALLMARK_MYC_TARGETS_V1	193	0.9829576	0.598892	MYC_TARGETS_V1	proliferation
HALLMARK_ESTROGEN_RESPONSE_EARLY	136	0.9546907	0.62060666	ESTROGEN_RESPONSE_EARLY	signaling
HALLMARK_E2F_TARGETS	198	0.9593918	0.6301633	E2F_TARGETS	proliferation
HALLMARK_SPERMATOGENESIS	66	0.93009776	0.65188074	SPERMATOGENESIS	development
HALLMARK_HEME_METABOLISM	186	0.91314536	0.6677364	HEME_METABOLISM	metabolic
HALLMARK_MITOTIC_SPINDLE	189	0.817258	0.828359	MITOTIC_SPINDLE	proliferation
HALLMARK_TGF_BETA_SIGNALING	48	-0.9497621	0.8393127	TGF_BETA_SIGNALING	signaling
HALLMARK_APICAL_SURFACE	26	-0.92444056	0.8456369	APICAL_SURFACE	cellular component
HALLMARK_MYC_TARGETS_V2	57	0.7569921	0.8991134	MYC_TARGETS_V2	proliferation
HALLMARK_IL6_JAK_STAT3_SIGNALING	65	-0.8349971	0.9259255	IL6_JAK_STAT3_SIGNALING	immune
HALLMARK_ANGIOGENESIS	23	-0.76122874	0.9430061	ANGIOGENESIS	development
HALLMARK_NOTCH_SIGNALING	26	-0.861826	0.95274657	NOTCH_SIGNALING	signaling

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S4B: Gene Set Enrichment Pathway Analysis PV vs CTRL

NAME	SIZE	NES	FDR q-val	Lower Level	Upper Level
HALLMARK_ADIPOGENESIS	172	1.7491294	0.012118747	ADIPOGENESIS	development
HALLMARK_GLYCOLYSIS	146	1.6530863	0.01930149	GLYCOLYSIS	metabolic
HALLMARK_APICAL_JUNCTION	127	1.660242	0.025757005	APICAL_JUNCTION	cellular component
HALLMARK_FATTY_ACID_METABOLISM	129	1.6055912	0.026302228	FATTY_ACID_METABOLISM	metabolic
HALLMARK_COAGULATION	78	1.5808175	0.02970178	COAGULATION	immune
HALLMARK_OXIDATIVE_PHOSPHORYLATION	183	1.5442894	0.039326318	OXIDATIVE_PHOSPHORYLATION	metabolic
HALLMARK_CHOLESTEROL_HOMEOSTASIS	64	1.5314032	0.04042827	CHOLESTEROL_HOMEOSTASIS	metabolic
HALLMARK_XENOBIOTIC_METABOLISM	128	1.5096257	0.042610623	XENOBIOTIC_METABOLISM	metabolic
HALLMARK_INTERFERON_ALPHA_RESPONSE	89	1.518906	0.04321455	INTERFERON_ALPHA_RESPONSE	immune
HALLMARK_HEME_METABOLISM	186	1.4567754	0.0673379	HEME_METABOLISM	metabolic
HALLMARK_PEROXISOME	81	1.4418689	0.07086366	PEROXISOME	cellular component
HALLMARK_MTORC1_SIGNALING	190	1.3795087	0.12471676	MTORC1_SIGNALING	signaling
HALLMARK_MYOGENESIS	115	1.3494041	0.14121492	MYOGENESIS	development
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITIC	115	1.356965	0.1420532	EPITHELIAL_MESENCHYMAL_TRANSITIO	development
HALLMARK_BILE_ACID_METABOLISM	70	1.3197786	0.17186424	BILE_ACID_METABOLISM	metabolic
HALLMARK_P53_PATHWAY	162	1.2998974	0.17200567	P53_PATHWAY	proliferation
HALLMARK_INTERFERON_GAMMA_RESPONSE	177	1.3008407	0.18022385	INTERFERON_GAMMA_RESPONSE	immune
HALLMARK_PROTEIN_SECRETION	91	1.3067324	0.18193667	PROTEIN_SECRETION	pathway
HALLMARK_KRAS_SIGNALING_DN	72	1.2802348	0.18998314	KRAS_SIGNALING_DN	signaling
HALLMARK_DNA_REPAIR	143	1.2603142	0.20102613	DNA_REPAIR	DNA damage
HALLMARK_IL2_STAT5_SIGNALING	158	1.2645485	0.20397907	IL2_STAT5_SIGNALING	signaling
HALLMARK_ALLOGRAFT_REJECTION	151	-1.3648527	0.2617964	ALLOGRAFT_REJECTION	immune
HALLMARK_UV_RESPONSE_UP	125	1.1793895	0.34141245	UV_RESPONSE_UP	DNA damage
HALLMARK_COMPLEMENT	156	1.171932	0.34395382	COMPLEMENT	immune
HALLMARK_ESTROGEN_RESPONSE_LATE	135	1.1602746	0.35328636	ESTROGEN_RESPONSE_LATE	signaling
HALLMARK_HEDGEHOG_SIGNALING	21	-1.2357339	0.35349664	HEDGEHOG_SIGNALING	signaling
HALLMARK_TNFA_SIGNALING_VIA_NFKB	162	-1.1725798	0.3754302	TNFA_SIGNALING_VIA_NFKB	signaling
HALLMARK_ESTROGEN_RESPONSE_EARLY	136	1.1259818	0.4149885	ESTROGEN_RESPONSE_EARLY	signaling
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWA	48	1.0937103	0.475638	REACTIVE_OXYGEN_SPECIES_PATHWA	pathway
HALLMARK_APOPTOSIS	138	1.0725611	0.5131383	APOPTOSIS	pathway
HALLMARK_ANDROGEN_RESPONSE	85	1.061999	0.51994646	ANDROGEN_RESPONSE	signaling
HALLMARK_INFLAMMATORY_RESPONSE	143	1.0041803	0.65263337	INFLAMMATORY_RESPONSE	immune
HALLMARK_IL6_JAK_STAT3_SIGNALING	65	0.97972196	0.6693822	IL6_JAK_STAT3_SIGNALING	immune
HALLMARK_G2M_CHECKPOINT	191	0.98495644	0.67889416	G2M_CHECKPOINT	proliferation
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	103	0.9614258	0.6946754	UNFOLDED_PROTEIN_RESPONSE	pathway
HALLMARK_HYPOXIA	138	0.94022757	0.7245265	HYPOXIA	pathway
HALLMARK_PI3K_AKT_MTOR_SIGNALING	92	0.9294137	0.7290277	PI3K_AKT_MTOR_SIGNALING	signaling
HALLMARK_SPERMATOGENESIS	66	0.9083337	0.7548022	SPERMATOGENESIS	development
HALLMARK_MITOTIC_SPINDLE	189	0.8941381	0.765224	MITOTIC_SPINDLE	proliferation
HALLMARK_UV_RESPONSE_DN	109	-1.0125421	0.8153306	UV_RESPONSE_DN	DNA damage
HALLMARK_WNT_BETA_CATENIN_SIGNALING	30	0.8177549	0.90225345	WNT_BETA_CATENIN_SIGNALING	signaling
HALLMARK_ANGIOGENESIS	23	0.7963376	0.91691846	ANGIOGENESIS	development
HALLMARK_KRAS_SIGNALING_UP	123	-0.950727	0.9232399	KRAS_SIGNALING_UP	signaling
HALLMARK_E2F_TARGETS	198	0.77223194	0.9336657	E2F_TARGETS	proliferation
HALLMARK_NOTCH_SIGNALING	26	-0.7739246	0.9437037	NOTCH_SIGNALING	signaling
HALLMARK_MYC_TARGETS_V1	193	0.74576426	0.94463915	MYC_TARGETS_V1	proliferation
HALLMARK_MYC_TARGETS_V2	57	0.6715539	0.97912604	MYC_TARGETS_V2	proliferation
HALLMARK_APICAL_SURFACE	26	-0.8297998	0.9997721	APICAL_SURFACE	cellular component
HALLMARK_TGF_BETA_SIGNALING	48	-0.8463163	1	TGF_BETA_SIGNALING	signaling

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S4C: Gene Set Enrichment Pathway Analysis MF vs CTRL

NAME	SIZE	NES	FDR q-val	Lower Level	Upper Level
HALLMARK_E2F_TARGETS	198	2.2282386	0	E2F_TARGETS	proliferation
HALLMARK_MYC_TARGETS_V1	193	2.1653433	0	MYC_TARGETS_V1	proliferation
HALLMARK_G2M_CHECKPOINT	191	2.0521796	0	G2M_CHECKPOINT	proliferation
HALLMARK_FATTY_ACID_METABOLISM	129	1.9544679	0	FATTY_ACID_METABOLISM	metabolic
HALLMARK_OXIDATIVE_PHOSPHORYLATION	183	1.9378474	0	OXIDATIVE_PHOSPHORYLATION	metabolic
HALLMARK_HEME_METABOLISM	186	1.9081693	0	HEME_METABOLISM	metabolic
HALLMARK_ADIPOGENESIS	172	1.8256093	0	ADIPOGENESIS	development
HALLMARK_MTORC1_SIGNALING	190	1.8196555	9.69E-05	MTORC1_SIGNALING	signaling
HALLMARK_MYC_TARGETS_V2	57	1.8229932	0.000109	MYC_TARGETS_V2	proliferation
HALLMARK_CHOLESTEROL_HOMEOSTASIS	64	1.7167392	0.000656	CHOLESTEROL_HOMEOSTASIS	metabolic
HALLMARK_GLYCOLYSIS	146	1.7352828	0.000722	GLYCOLYSIS	metabolic
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	103	1.6793369	0.001606297	UNFOLDED_PROTEIN_RESPONSE	pathway
HALLMARK_PEROXISOME	81	1.6652691	0.001979208	PEROXISOME	cellular component
HALLMARK_BILE_ACID_METABOLISM	70	1.6472547	0.002317972	BILE_ACID_METABOLISM	metabolic
HALLMARK_DNA_REPAIR	143	1.567265	0.006026251	DNA_REPAIR	DNA damage
HALLMARK_XENOBIOTIC_METABOLISM	128	1.5719225	0.00610644	XENOBIOTIC_METABOLISM	metabolic
HALLMARK_TNFA_SIGNALING_VIA_NFKB	162	-1.6079856	0.014967728	TNFA_SIGNALING_VIA_NFKB	signaling
HALLMARK_UV_RESPONSE_DN	109	-1.6162648	0.026313726	UV_RESPONSE_DN	DNA damage
HALLMARK_UV_RESPONSE_UP	125	1.4404038	0.030453881	UV_RESPONSE_UP	DNA damage
HALLMARK_ESTROGEN_RESPONSE_LATE	135	1.4200132	0.036488954	ESTROGEN_RESPONSE_LATE	signaling
HALLMARK_APICAL_JUNCTION	127	1.2958041	0.13213581	APICAL_JUNCTION	cellular component
HALLMARK_INTERFERON_ALPHA_RESPONSE	89	1.2880113	0.13591722	INTERFERON_ALPHA_RESPONSE	immune
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWA	48	1.2698575	0.15439267	REACTIVE_OXYGEN_SPECIES_PATHWA	pathway
HALLMARK_P53_PATHWAY	162	1.2502935	0.1753588	P53_PATHWAY	proliferation
HALLMARK_COAGULATION	78	1.2284198	0.20240758	COAGULATION	immune
HALLMARK_MITOTIC_SPINDLE	189	1.2146338	0.21652985	MITOTIC_SPINDLE	proliferation
HALLMARK_PROTEIN_SECRETION	91	1.1643198	0.3081479	PROTEIN_SECRETION	pathway
HALLMARK_ESTROGEN_RESPONSE_EARLY	136	1.121265	0.40073216	ESTROGEN_RESPONSE_EARLY	signaling
HALLMARK_MYOGENESIS	115	1.1136562	0.40471503	MYOGENESIS	development
HALLMARK_HYPOXIA	138	1.1011097	0.42094424	HYPOXIA	pathway
HALLMARK_ANGIOGENESIS	23	1.0774357	0.46749738	ANGIOGENESIS	development
HALLMARK_SPERMATOGENESIS	66	1.0559838	0.50928974	SPERMATOGENESIS	development
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITIC	115	1.0231813	0.58039236	EPITHELIAL_MESENCHYMAL_TRANSITIO	development
HALLMARK_APOPTOSIS	138	1.0045238	0.61114854	APOPTOSIS	pathway
HALLMARK_INTERFERON_GAMMA_RESPONSE	177	0.98938555	0.63126415	INTERFERON_GAMMA_RESPONSE	immune
HALLMARK_ALLOGRAFT_REJECTION	151	-1.0892317	0.6535871	ALLOGRAFT_REJECTION	immune
HALLMARK_NOTCH_SIGNALING	26	-0.8892041	0.7645577	NOTCH_SIGNALING	signaling
HALLMARK_ANDROGEN_RESPONSE	85	-0.9891044	0.76488686	ANDROGEN_RESPONSE	signaling
HALLMARK_APICAL_SURFACE	26	-0.9520981	0.7780008	APICAL_SURFACE	cellular component
HALLMARK_KRAS_SIGNALING_UP	123	-0.9037519	0.8283934	KRAS_SIGNALING_UP	signaling
HALLMARK_TGF_BETA_SIGNALING	48	-1.0094684	0.84966916	TGF_BETA_SIGNALING	signaling
HALLMARK_PI3K_AKT_MTOR_SIGNALING	92	0.83679307	0.9083634	PI3K_AKT_MTOR_SIGNALING	signaling
HALLMARK_WNT_BETA_CATENIN_SIGNALING	30	0.8222267	0.9099532	WNT_BETA_CATENIN_SIGNALING	signaling
HALLMARK_KRAS_SIGNALING_DN	72	0.844743	0.91847396	KRAS_SIGNALING_DN	signaling
HALLMARK_IL6_JAK_STAT3_SIGNALING	65	0.79050213	0.93626595	IL6_JAK_STAT3_SIGNALING	immune
HALLMARK_HEDGEHOG_SIGNALING	21	0.84735936	0.9408993	HEDGEHOG_SIGNALING	signaling
HALLMARK_COMPLEMENT	156	0.76320857	0.9481598	COMPLEMENT	immune
HALLMARK_IL2_STAT5_SIGNALING	158	0.74106455	0.9485353	IL2_STAT5_SIGNALING	signaling
HALLMARK_INFLAMMATORY_RESPONSE	143	0.6850725	0.96872264	INFLAMMATORY_RESPONSE	immune

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S5A: Predicted Probabilities of Lasso Base Model (Age, Gender, Mutation Status)

Sample# DeID	Prob.MF	Prob.CTRL	Prob.ET	Prob.PV	Predicted.Label	True.Label
S1428	0.72463124	2.41E-08	0.070519432	0.204849304	MF	MF
S990146	0.215659422	0.708605109	0.075733984	1.48E-06	CTRL	CTRL
S7542	0.533503343	7.10E-09	0.466496595	5.52E-08	MF	ET
S0105	0.85252	2.51E-09	0.14747999	6.66E-09	MF	MF
S3039	0.056567154	3.76E-08	0.451098397	0.492334411	PV	MF
S6784	0.207247245	2.47E-08	0.081579043	0.711173687	PV	MF
S6544	0.270275612	1.44E-08	0.048769711	0.680954662	PV	PV
S8075	0.174742547	1.31E-08	0.167745461	0.657511979	PV	ET
S6398	0.574564107	1.90E-08	0.212942309	0.212493565	MF	ET
S7033	0.76804637	1.00E-09	0.231953623	5.76E-09	MF	ET
S990145	0.133824878	0.616157626	0.25001612	1.38E-06	CTRL	CTRL
S5385	0.735574218	7.70E-09	0.08945948	0.174966294	MF	MF
S8676	0.803311737	4.22E-09	0.050064493	0.146623766	MF	PV
S1765.y	0.646282024	1.35E-08	0.153382603	0.200335359	MF	MF
S7782	0.807425128	1.07E-08	0.032222196	0.160352665	MF	PV
S3147	0.779750326	1.46E-08	0.043469588	0.176780072	MF	MF
S990147	0.006784112	0.893179214	0.100035242	1.43E-06	CTRL	CTRL
S5702	0.009539467	0.889763967	0.100694801	1.76E-06	CTRL	ET
S990150	0.005402597	0.895044774	0.099551383	1.25E-06	CTRL	CTRL
S990149	0.179422458	0.741838283	0.07873791	1.35E-06	CTRL	CTRL
S2541	0.008094881	1.53E-08	0.208944436	0.782960667	PV	ET
S1562	0.233223038	1.97E-08	0.065717981	0.701058961	PV	MF
S8016	0.23566708	0.690297758	0.07403361	1.55E-06	CTRL	ET
S0574	0.005636387	2.13E-08	0.283906456	0.710457136	PV	PV
S5311	0.679527782	1.41E-09	0.320472211	6.08E-09	MF	MF
S0359	0.646282024	1.35E-08	0.153382603	0.200335359	MF	MF
S8413	0.166509902	1.40E-08	0.178683538	0.654806546	PV	MF
S990144	0.386802572	0.560754705	0.052442461	2.62E-07	CTRL	CTRL
S990142	0.065013348	0.848043174	0.086942718	7.60E-07	CTRL	CTRL
S8702	0.198949623	2.66E-08	0.087543785	0.713506565	PV	PV
S3335	0.135728957	1.80E-08	0.227450256	0.636820769	PV	PV
S7297	0.712129102	2.66E-08	0.077471657	0.210399214	MF	MF
S2657	0.678559055	1.12E-08	0.128871522	0.192569411	MF	MF
S1522	0.889803236	4.64E-10	0.110196759	4.69E-09	MF	MF
S1765.z	0.646282024	1.35E-08	0.153382603	0.200335359	MF	MF
S990143	0.179422458	0.741838283	0.07873791	1.35E-06	CTRL	CTRL
S8571	0.72463124	2.41E-08	0.070519432	0.204849304	MF	PV
S990140	0.313325334	0.505826547	0.180847829	2.90E-07	CTRL	CTRL
S7428	0.175168261	3.31E-08	0.107674955	0.717156751	PV	PV
S9059	0.693754412	1.03E-08	0.117864416	0.188381162	MF	MF
S0103	0.736623778	2.18E-08	0.064127621	0.199248579	MF	ET
S5881	0.712129102	2.66E-08	0.077471657	0.210399214	MF	PV
S990148	0.499403466	0.457237593	0.043358658	2.83E-07	MF	CTRL
S6482	0.611558786	1.61E-08	0.181374804	0.207066394	MF	PV
S6952	0.664890449	0.30552776	0.029581501	2.89E-07	MF	MF
S9313.y	0.183182968	1.22E-08	0.157305841	0.659511178	PV	PV
S2308	0.735574218	7.70E-09	0.08945948	0.174966294	MF	ET
S6874	0.182908513	3.08E-08	0.100577501	0.716513955	PV	MF
S2946	0.815845484	9.65E-09	0.029125167	0.155029339	MF	ET

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S5B: Predicted Probabilities of Lasso Model with the Entire Transcriptome

Sample#	Del	Prob.MF	Prob.CTRL	Prob.ET	Prob.PV	Predicted.La	True.Label
S3147		0.12395032	0.06361095	0.09790658	0.71453214	PV	MF
S7033		0.19659506	0.08203161	0.22459043	0.4967829	PV	ET
S7428		0.06388445	0.05482461	0.13857067	0.74272027	PV	PV
S6544		0.01264897	0.02531482	0.04853313	0.91350308	PV	PV
S0359		0.05075509	0.03008909	0.04769623	0.8714596	PV	MF
S6482		0.03318443	0.03831813	0.08172735	0.8467701	PV	PV
S6874		0.6565228	8.44E-05	0.02096776	0.32242508	MF	MF
S8702		0.00560781	0.00202412	0.00347268	0.98889539	PV	PV
S990148		0.02070988	0.89551519	0.0412124	0.04256253	CTRL	CTRL
S8676		0.00688092	0.00571897	0.0101142	0.97728591	PV	PV
S7297		0.70371895	0.00010441	0.01761245	0.27856419	MF	MF
S2946		0.02586105	0.02701811	0.59058045	0.35654039	ET	ET
S3335		0.02952875	0.05703482	0.19155367	0.72188276	PV	PV
S3039		0.23361967	0.00017637	0.03511815	0.73108582	PV	MF
S1562		0.0589415	2.72E-05	0.07769311	0.86333823	PV	MF
S8413		0.69923559	0.00018365	0.03854927	0.2620315	MF	MF
S990144		0.00931965	0.94026804	0.03396616	0.01644615	CTRL	CTRL
S1765.y		0.90660249	8.97E-05	0.02020768	0.07310016	MF	MF
S6952		0.28272304	0.00010059	0.06797471	0.64920165	PV	MF
S990150		0.04668521	0.66484371	0.14354843	0.14492265	CTRL	CTRL
S990142		0.03484686	0.70948013	0.12376817	0.13190485	CTRL	CTRL
S990147		0.02387115	0.82905306	0.10868893	0.03838686	CTRL	CTRL
S2657		0.21022434	0.14793153	0.35334701	0.28849713	ET	MF
S990145		0.04457762	0.55890729	0.31135881	0.08515628	CTRL	CTRL
S0103		0.09417498	0.15376317	0.36503756	0.38702429	PV	ET
S1428		0.62708746	0.00014542	0.02790394	0.34486318	MF	MF
S5385		0.77873205	0.00010495	0.02608396	0.19507904	MF	MF
S9059		0.19711924	2.16E-05	0.00628247	0.79657665	PV	MF
S5311		0.93289883	6.17E-05	0.01392972	0.0531097	MF	MF
S0574		0.01840554	0.02799563	0.10456313	0.8490357	PV	PV
S2308		0.00510199	0.00525284	0.01752037	0.97212481	PV	ET
S990146		0.08880267	0.35034659	0.43806959	0.12278115	ET	CTRL
S8016		0.01332511	0.02108113	0.36751933	0.59807443	PV	ET
S8075		0.03637302	0.11283222	0.25223217	0.59856259	PV	ET
S9313.y		0.01265993	0.01520904	0.05035531	0.92177573	PV	PV
S2541		0.06394786	0.13346632	0.32592943	0.4766564	PV	ET
S1522		0.25669537	0.1920827	0.29396999	0.25725194	ET	MF
S5702		0.03994486	0.76294024	0.11564505	0.08146986	CTRL	ET
S1765.z		0.88281977	0.00010015	0.020947	0.09613309	MF	MF
S5881		0.17256702	6.51E-05	0.05169277	0.77567515	PV	PV
S990149		0.04579218	0.6788359	0.13618693	0.13918499	CTRL	CTRL
S990143		0.00996377	0.88831275	0.07604564	0.02567785	CTRL	CTRL
S7782		0.0496517	0.0449496	0.11325015	0.79214856	PV	PV
S990140		0.05219702	0.47797688	0.32468732	0.14513878	CTRL	CTRL
S0105		0.15503438	0.08436133	0.53905446	0.22154983	ET	MF
S7542		0.11561118	0.12378457	0.54402197	0.21658227	ET	ET
S8571		0.34181523	4.62E-05	0.00959242	0.64854619	PV	PV
S6784		0.59728542	0.00010083	0.02811422	0.37449953	MF	MF
S6398		0.05058235	0.09084023	0.47936468	0.37921274	ET	ET

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.12.435190; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S5C: Predicted Probabilities of Lasso Model with the Progressive Transcriptome

Sample# DeID	Prob.MF	Prob.CTRL	Prob.ET	Prob.PV	Predicted.Label	True.Label
S3147	0.092073998	0.00493465	0.031344267	0.871647085	PV	MF
S7033	0.159484242	0.013514786	0.11672636	0.710274612	PV	ET
S7428	0.018488282	0.008499175	0.125530701	0.847481841	PV	PV
S6544	0.001328213	0.002355917	0.028773588	0.967542283	PV	PV
S0359	0.010627975	0.001140442	0.001253889	0.986977693	PV	MF
S6482	0.004419731	0.009786142	0.179595073	0.806199054	PV	PV
S6874	0.217134582	1.28E-05	0.003925044	0.778927546	PV	MF
S8702	0.03779986	0.00271328	0.026403773	0.933083087	PV	PV
S990148	0.001265836	0.977512385	0.015728321	0.005493459	CTRL	CTRL
S8676	0.002146306	0.000258697	0.001015756	0.996579241	PV	PV
S7297	0.386461413	6.85E-05	0.013583582	0.599886493	PV	MF
S2946	0.002646809	0.001465367	0.049279281	0.946608543	PV	ET
S3335	0.000551306	0.001708487	0.904558334	0.093181873	ET	PV
S3039	0.083757526	0.000348078	0.268427663	0.647466732	PV	MF
S1562	0.110977219	4.04E-05	0.006545455	0.882436965	PV	MF
S8413	0.219666538	6.68E-05	0.019332678	0.760933953	PV	MF
S990144	9.25E-05	0.991629353	0.007701818	0.000576321	CTRL	CTRL
S1765.y	0.919405903	5.09E-05	0.013330491	0.067212723	MF	MF
S6952	0.023268789	4.25E-05	0.098810469	0.877878287	PV	MF
S990150	0.008180225	0.891813454	0.048001318	0.052005003	CTRL	CTRL
S990142	0.010046736	0.840480139	0.109791528	0.039681596	CTRL	CTRL
S990147	0.00273875	0.918542798	0.066699239	0.012019214	CTRL	CTRL
S2657	0.055533721	0.016873468	0.251680583	0.675912228	PV	MF
S990145	0.02068156	0.549791974	0.42137892	0.008147546	CTRL	CTRL
S0103	0.041596134	0.034167556	0.583429166	0.340807143	ET	ET
S1428	0.224592498	6.36E-05	0.011940365	0.763403542	PV	MF
S5385	0.926392637	4.18E-05	0.011415157	0.062150377	MF	MF
S9059	0.115405357	1.27E-05	0.00101406	0.883567923	PV	MF
S5311	0.962316012	2.35E-05	0.015219642	0.022440855	MF	MF
S0574	0.000797008	0.00142649	0.219763071	0.778013431	PV	PV
S2308	0.002745334	0.001480741	0.28371321	0.712060714	PV	ET
S990146	0.037220556	0.507160101	0.395266592	0.060352752	CTRL	CTRL
S8016	0.001302611	0.002971883	0.132241562	0.863483943	PV	ET
S8075	0.015486707	0.124415352	0.442418325	0.417679616	ET	ET
S9313.y	0.00193569	0.002765518	0.028682678	0.966616115	PV	PV
S2541	0.037133276	0.176393665	0.251033344	0.535439715	PV	ET
S1522	0.430429563	0.078512377	0.363995484	0.127062577	MF	MF
S5702	0.010989962	0.774793968	0.15882085	0.05539522	CTRL	ET
S1765.z	0.928962973	7.08E-05	0.007488947	0.063477328	MF	MF
S5881	0.021403102	3.44E-05	0.034411008	0.944151475	PV	PV
S990149	0.004708259	0.691336231	0.210213379	0.09374213	CTRL	CTRL
S990143	0.00038944	0.78654355	0.211859405	0.001207605	CTRL	CTRL
S7782	0.009267968	0.009626422	0.144706684	0.836398926	PV	PV
S990140	0.007793924	0.464113827	0.469603441	0.058488808	ET	CTRL
S0105	0.150151999	0.042619713	0.485353304	0.321874984	ET	MF
S7542	0.005179512	0.007342405	0.871411071	0.116067012	ET	ET
S8571	0.034490309	5.56E-06	0.000692632	0.9648115	PV	PV
S6784	0.23455979	9.07E-05	0.008421458	0.7569281	PV	MF
S6398	0.023367024	0.025093319	0.169208568	0.782331089	PV	ET