Supplementary Figure 1

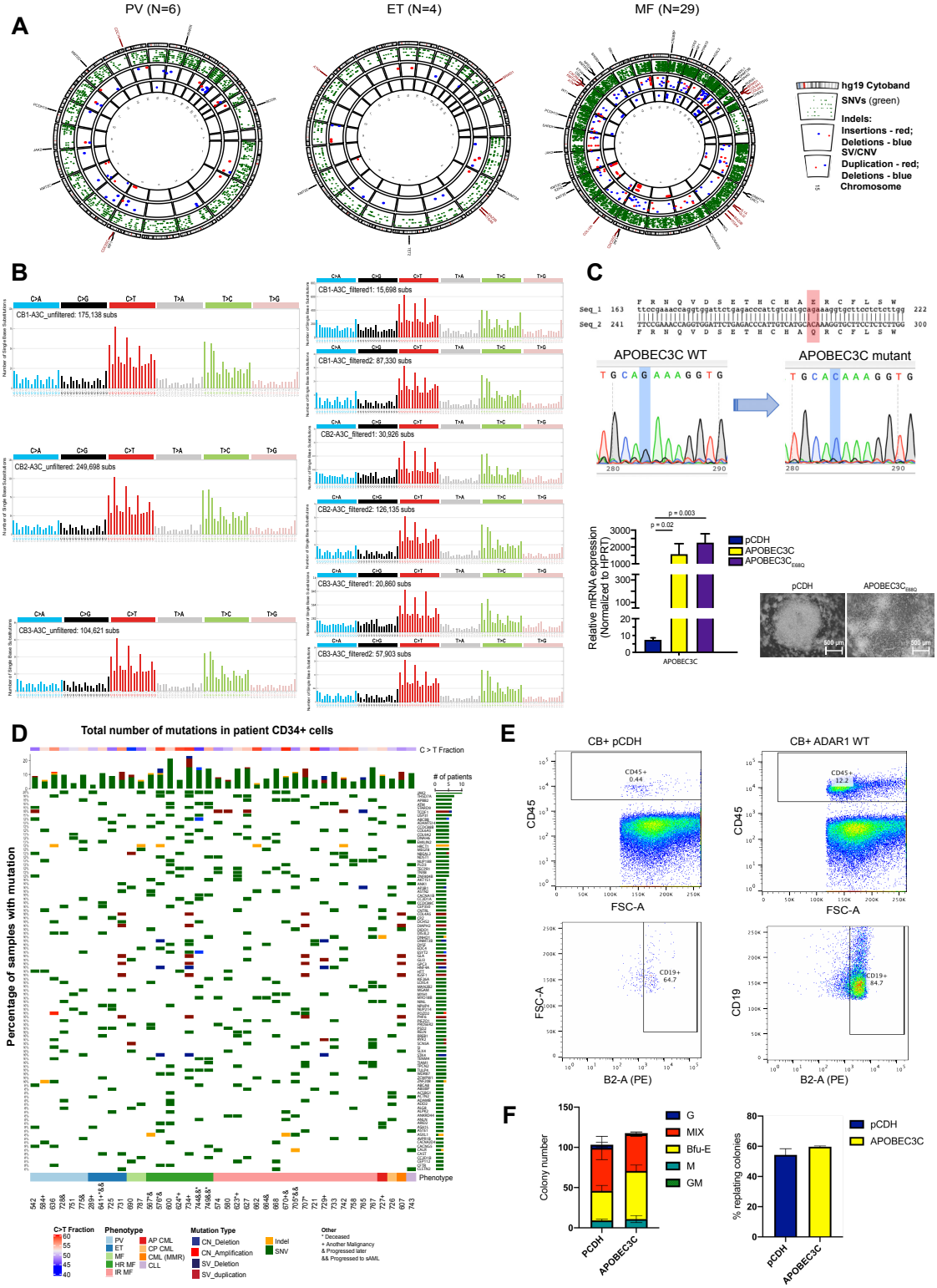


Figure S1. Top DNA mutations in MPN peripheral blood samples. Related to Figure 1. A. Circos plots depicting single-nucleotide variants (SNVs) and structural variants (SVs) in MPN-associated samples. Labels indicate the top 30 mutated genes (red font) 69 MPN-associated genes (black font) [38] B. Mutational profiles observed after APOBEC3C expression in CD34+ cord blood stem cells. Mutations were detected using ensemble variant calling followed by extensive filtering of germline variants. Three types of germline filtering were used: (i) no filtering (annotated as unfiltered); (ii) filtering against all known germline variants in gnomAD and dbSNP (filtered1); (iii) filtering against all germline variants detected in CD34+ normal blood cells generated in the current study (filtered2). In all cases, the mutational profiles were consistent. Mutational profiles are shown using single base substitutions with six subtypes: C>A, C>G, C>T, T>A, T>C, T>G. Underneath each subtype are 16 bars reflecting the sequence contexts determined by the four possible bases immediately 5' and 3' to each mutated base. C. Top: Sanger sequencing analysis of APOBEC3CE68Qmutant. A G-to-C point mutation at position 68 results in an amino acid change from Glutamic Acid to Glutamine in the catalytic domain of APOBEC3C and catalytic inactivity. Bottom left: relative mRNA expression of APOBEC3C in cord blood CD34+ cells transduced with a lentiviral backbone (pCDH), APOBEC3C, or APOBEC3C mutant ($APOBEC3C_{E68Q}$) n = 3. Bottom right: brightfield images of cord blood CD34+ cells transduced with pCDH or $APOBEC3C_{E68Q}$. D. Top mutations in MPN patients from peripheral blood including single nucleotide variants (SNVs), copy number (CN) variants and structural variants (SVs). MPN disease stage depicted in colored bar at the bottom of the figure. *, patient deceased since sample collection; +, patient had another malignancy; &, patient progressed after sample collection, and &&, patient progressed to AML after sample collection. Color of alterations signifies the type of alteration. Fraction of C-to-T transitions is colored according to percent in the legend. E. Flow gating for cord blood transduced with ADAR1 overexpressing lentiviral vectors that were engrafted. The CD19 was gated on CD45+ human cells. F. Colony assay with cord blood CD34+ cells overexpressing pCDH backbone or APOBEC3C. The self-renewal capacity as measured by relapting assay is shown on the right.

Supplementary Figure 2

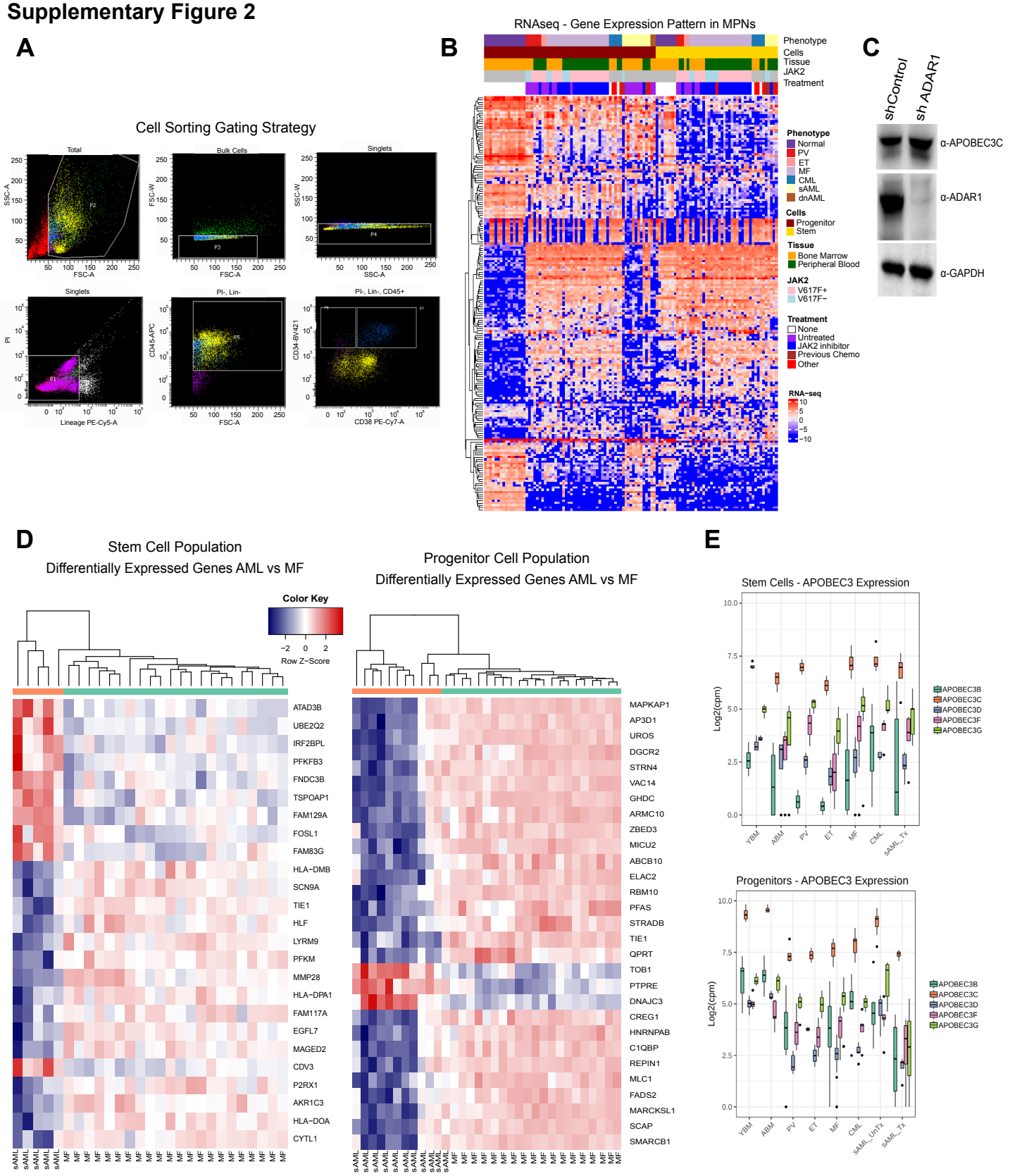
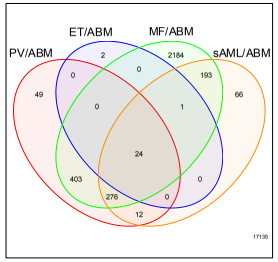


Figure S2. Differential Gene Expression in FACS-purified MPN stem cells and progenitors. Related to Figure 2. A. A representative gating strategy for FACS-purified stem cell (CD34+CD38-Lin-) and progenitor (CD34+CD38+Lin-) populations from 54 unique patients and 24 young and aged healthy controls. B. A heatmap of RNA-Seq expression of the top 1% of genes ranked by variance. Annotation is shown as a stack of colored bars representing phenotype, cell type, source tissue, mutation status, and the treatment type. C. TF1a shADAR1 cells show ablation of ADAR1 protein by western blot. Western blot analysis of whole cell extracts (WCE) prepared from TF1a shADAR1 transduced cells with indicated antibodies confirms knockdown of ADAR1 compared with TF1a shControl cells. GAPDH is a protein loading control. D. Heatmap showing the top 25 differentially expressed genes in AML stem cells compared with MF stem cells (987 total DE genes). Heatmap showing the top 25 differentially expressed genes in AML progenitors compared with MF progenitors (678 total DE genes). E. Expression of APOBEC3 family genes in FACS-purified stem cells and progenitors from normal aged (ABM), normal young (YBM), MPN and AML samples.

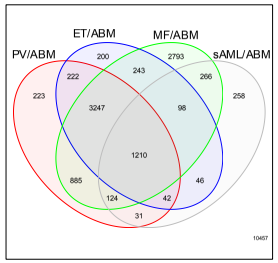
Supplementary Figure 3

A

Stem Cell RNAseq - Differentially Expressed Genes



Progenitor Cell RNAseq - Differentially Expressed Genes



B

Top 10 Unique Genes per Condition Stem Cell Population

PV vs. ABM			ET vs. ABM			MF vs. ABM			CML vs. ABM			AML vs. ABM		
SYM	logFC	adj.P.Val	SYM	logFC	adj.P.Val	SYM	logFC	adj.P.Val	SYM	logFC	adj.P.Val	SYM	logFC	adj.P.Val
IFITM1	1.8	0.01	TM4SF1	3.5	0.03	ZNF587	-2.1	2.4E-07	APLP2	-1.5	8.0E-03	IER2	2.5	6.0E-03
IRF9	1.5	0.01	GP1BB	2.4	0.05	GOLGA8N	-3.8	3.4E-07	GUSB	1.4	0.01	PLIN2	3.4	0.01
KIAA2013	1.6	0.01				RPS17	-4.0	4.2E-07	GLIPR2	2.4	0.01	ADGRG6	-3.1	0.02
CEP70	1.9	0.01				SECISBP2L	-1.8	4.2E-07	RASGRP4	1.9	0.01	FSCN1	1.7	0.02
SLC35A4	1.6	0.01				STRN	-1.6	9.8E-07	CAPG	1.8	0.01	OTUD1	2.2	0.02
BMP2K	-1.3	0.02				TAF4	-2.0	1.1E-06	CTSW	1.8	0.02	PTPN14	4.0	0.02
POR	1.4	0.02				KTN1	-1.2	1.1E-06	TMEM165	1.4	0.02	IFI44L	-3.0	0.03
HLA-DRB1	3.5	0.02				SACS	-2.0	1.2E-06	CYFIP2	1.4	0.02	SPARC	2.0	0.03
ARID5B	1.7	0.02				SKI	-3.3	2.2E-06	TPST2	1.3	0.02	MED28	-1.2	0.03
LDOC1	1.7	0.02				TFAP4	-2.8	2.6E-06	TRABD	1.2	0.02	GNAI3	1.4	0.03

C

Top Signaling Pathways per Stem Cell Population by Gene Set Enrichment Analysis

PATHWAY	AML vs. MF logFC	AML vs. MF adj.P.Val
BIOCARTA TALL1 PATHWAY	-0.59	0.02
REACTOME PHOSPHORYLATION OD CD3 AND TCR ZETA CHAINS	0.88	0.02
KEGG ASTHMA	0.89	0.02
REACTOME TRANSLOCATION OF ZAP70 TO IMMUNOLOGICAL SYNPASE	0.93	0.02
BIOCARTA NOS1 PATHWAY	0.46	0.02
KRGG TYPE 1 DIABETES MELLITUS	0.70	0.02
REACTOME GENERATION OF SECOND MESSENGER MOLECULES	0.65	0.02
REACTOME PD1 SIGNALING	0.73	0.02
KEGG RETINOL METABOLISM	0.60	0.02
KEGG GRAFT VERSUS HOST DISEASE	0.75	0.03
KEGG STEROID BIOSYNTHESIS	0.50	0.03
KEGG RIBOFLAVIN METABOLISM	0.48	0.03
KEGG AUTOIMMUNE THYROID DISEASE	0.73	0.03
KEGG BUTANOATE METABOLISM	0.48	0.03
KEGG ALLOGRAFT REJECTION	0.68	0.04
KEGG SULFUR METABOLISM	0.56	0.04
BIOCARTA EIF PATHWAY	-0.45	0.04
BIOCARTA RNA PATHWAY	-0.48	0.04
BIOCARTA IGF1MTOR PATHWAY	-0.40	0.04
BIOCARTA TEL PATHWAY	-0.41	0.04
KEGG GNRH SIGNALING PATHWAY	-0.35	0.05

D

Top 10 Unique Genes per Condition Progenitor Cell Population

PV vs. ABM			ET vs. ABM			MF vs. ABM			CML vs. ABM			AML vs. ABM		
SYM	logFC	adj.P.Val	SYM	logFC	adj.P.Val	SYM	logFC	adj.P.Val	SYM	logFC	adj.P.Val	SYM	logFC	adj.P.Val
ALDH3A2	1.5	2.7E-05	PROS1	2.0	7.5E-05	CSNK1G2	7.1	9.7E-17	ITGAM1	2.8	3.5E-07	AHNAK	2.4	4.3E-05
GPR27	2.5	3.7E-05	MN1	2.0	2.7E-04	NELFB	7.1	1.0E-15	GTSE1	2.2	1.2E-06	CSF1R	2.4	1.3E-04
ACOX2	1.9	9.3E-05	VWF	2.6	1.1E-03	PSD	7.8	3.2E-15	NLN	1.5	5.1E-06	PARP1	-1.6	2.1E-04
ITGAL	1.7	1.5E-04	ELK3	1.4	1.5E-03	ZC3H3	6.8	4.2E-14	SELL	-3.0	9.0E-06	TLE4	1.7	2.4E-04
ARAP2	1.0	2.7E-04	TAGAP	1.9	1.6E-03	AP5Z1	6.4	4.5E-14	KIF18B	1.9	1.1E-05	KIF1B	1.8	2.6E-04
CCDC149	1.8	3.6E-04	DLC1	1.4	2.0E-03	TMEM259	7.9	4.5E-14	PCDH9	-4.4	1.6E-05	LGALS1	1.9	4.5E-04
DNAH1	1.6	6.4E-04	TUBB1	1.8	2.3E-03	ABHD17A	8.4	4.6E-14	ATP7B	2.1	1.9E-05	PPA2	-1.2	5.2E-04
ZNF862	1.9	6.8E-04	B3GLCT	1.3	2.4E-03	FAAP100	6.0	5.9E-14	PARKCQ	1.2	1.9E-05	CPPED1	-1.6	5.4E-04
MOV10	0.9	1.0E-03	ZBTB8A	1.2	2.7E-03	ASPSCR1	5.2	6.8E-14	TBC1D4	1.5	1.9E-05	ACY3	3.7	5.4E-04
SLC25A4	1.6	1.3E-03	MDFIC	1.3	3.0E-03	CNOT3	5.1	8.1E-14	LDLRAD3	1.7	2.4E-05	MPO	-1.5	6.6E-04

E

Top Signaling Pathways per Progenitor Cell Population by Gene Set Enrichment Analysis

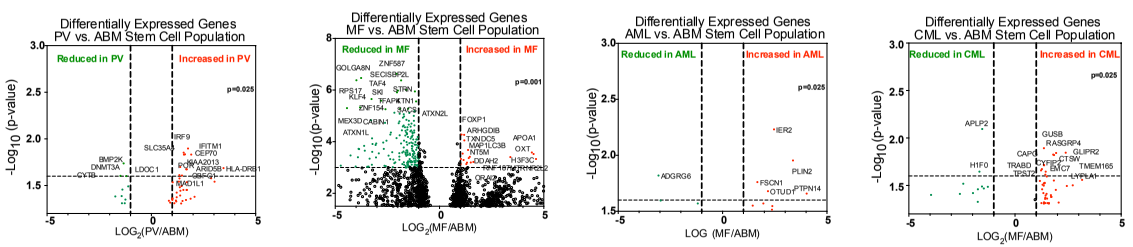
PATHWAY	AML vs. MF logFC	AML vs. MF adj.P.Val
REACTOME HORMONE SENSITIVE LIPASE HSL MEDIATED TRIACYLGLYCEROL_HYDROLYSIS	0.72	0.01
BIOCARTA NOS1 PATHWAY	0.56	0.01
REACTOME EICOSANOID LIGAND BINDING RECEPTORS	-0.87	0.05
REACTOME DARPP32 EVENTS	0.49	0.05
BIOCARTA STATHMIN PATHWAY	0.57	0.05
BIOCARTA NO1 PATHWAY	0.53	0.06
BIOCARTA CHEMICAL PATHWAY	0.42	0.07

F

Top Enriched Pathways per Progenitor Cell Population (PANTHER)

PATHWAY	sAML vs. MF FDR
WNT SIGNALING PATHWAY	1.05E-04
ANGIOGENESIS	3.35E-04
CCKR SIGNALING MAP	3.35E-04
HETEROTRIMERIC G-PROTEIN SIGNALING PATHWAY -GQ ALPHA AND GO ALPHA MEDIATED PATHWAY	4.72E-04
PI3 KINASE PATHWAY	0.002
ALZHEIMER DISEASE PRESENILLIN PATHWAY	0.006
HUNTINGTON DISEASE	0.006
INTEGRIN SIGNALING PATHWAY	0.010
INFLAMMATION MEDIATED BY CHEMOKINE AND CYTOKINE SIGNALING	0.012
DOPAMINE RECEPTOR MEDIATED PATHWAY	0.013
ALZHEIMER DISEASE-AMYLOID SECRETASE PATHWAY	0.013
MUSCARINIC ACETYLCHOLINE RECEPTOR 1 AND 3 SIGNALING PATHWAY	0.017
ENDOTHELIN SIGNALING PATHWAY	0.017
PARKINSON DISEASE	0.017
UBIQUITIN PROTEASOME PATHWAY	0.017
HETEROTRIMERIC G-PROTEIN SIGNALING PATHWAY -GI ALPHA AND GS ALPHA MEDIATED PATHWAY	0.017
ANGIOTENSIN II-STIMULATED SIGNALING THROUGH G PROTEINS AND BETA-ARRESTIN	0.021
P53 PATHWAY FEEDBACK LOOPS 2	0.029
RAS PATHWAY	0.037
TOLL RECEPTOR SIGNALING PATHWAY	0.041

G



H

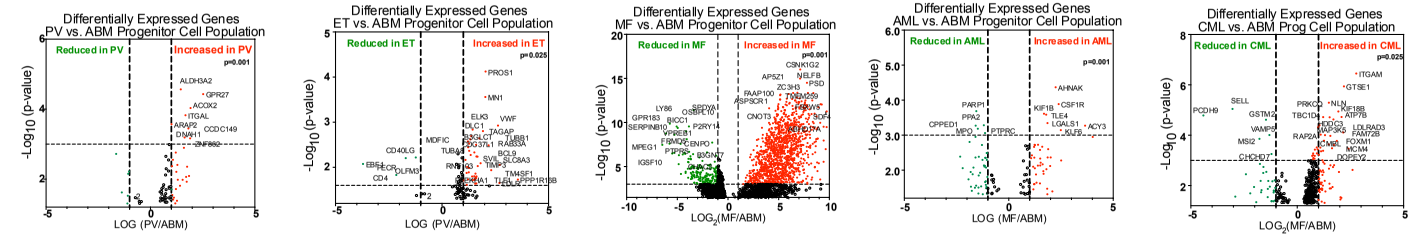
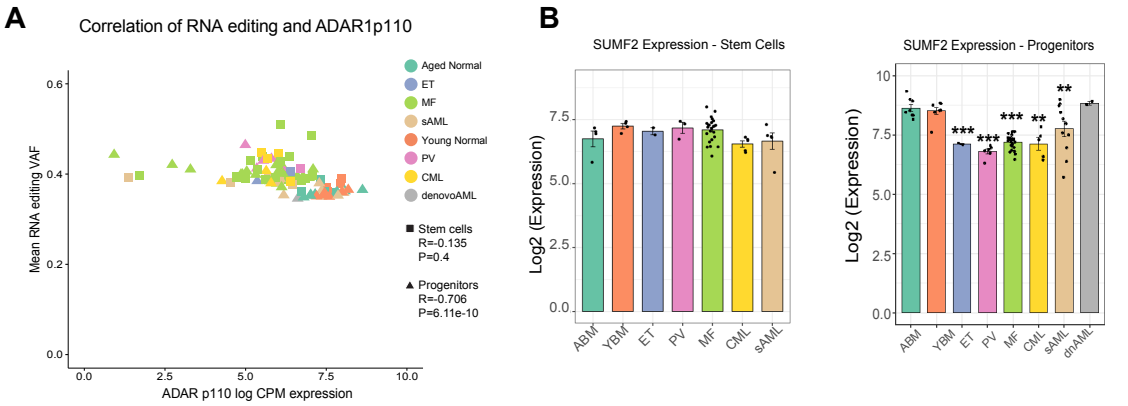
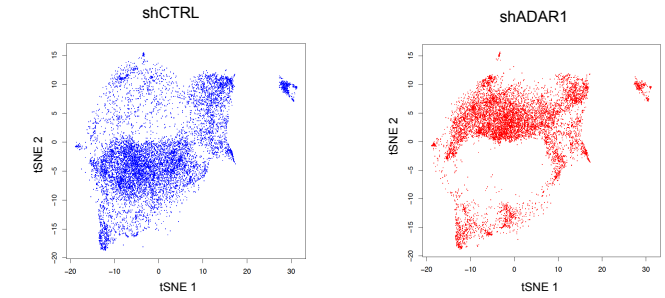


Figure S3. Differentially Expressed Splice Isoforms and Signaling Pathways in MPN and AML compared with aged bone marrow stem and progenitor cells. Related to Figure 3. A. Stem and progenitor cell differential transcript expression analysis between patients with various MPN phenotypes and AML compared with aged normal bone marrow. The Venn diagram shows the overlap of significantly different genes (adjusted p-val < 0.05) between the comparisons. *Adjusted Statistical significance values have been used. B. The top 10 statistically significantly expressed genes in the stem cell population are listed in the table. C. Gene Set Enrichment Analysis (GSEA) results for the comparison of AML vs MF in Stem Cells. Top pathways from KEGG, REACTOME, and BIOCARTA at adj.P.Val < 0.05 are shown along with the log fold change (logFC) of the pathway and the significance. D. The top 10 statistically significant genes in the progenitor cell population are listed in the table. E. GSEA results for the comparison of AML vs MF in Progenitors. Top pathways from KEGG, REACTOME, and BIOCARTA at adj.P.Val < 0.1 are shown along with the log fold change (logFC) of the pathway and the significance. F. Top enriched pathways per progenitor population by PANTHER analysis of sequencing data. G. Volcano plot of stem cell population of MPN and AML compared to ABM for genes with an adjusted p-value of < 0.025 (PV, AML and CML) or an adjusted p-value of < 0.001 (MF). H. Volcano plot for the progenitor cell population of MPN compared to ABM for genes with an adjusted p-value of < 0.001 (PV, MF and AML), and an adjusted p-value of < 0.025 (CML and ET).

Supplementary Figure 4

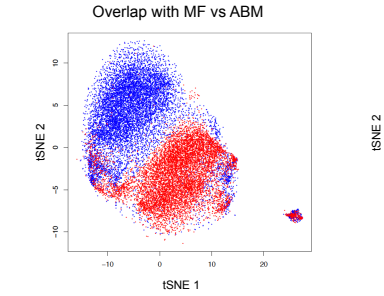


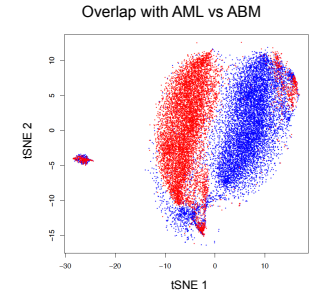


D

Top 10 Differentially Expressed Genes

shCTRL vs shADAR1 SYM	IFDR
RPL37A	1.2E-65
RPL41	4.3E-62
RPS28	2.8E-59
RPL39	5.1E-59
RPL37	2.3E-51
RPL34	1.4E-50
RPS27	4.7E-49
RPL35	2.4E-44
RPL36	5.8E-39
RPS21	3.3E-37





G

Top 10 Ovelapping Genes per Condition in Each Comparison

MF vs. ABM SYM	IFDR	AML vs. ABM SYM	IFDR
RPL37A	1.2E-65	RPL37A	1.2E-65
RPL41	4.3E-62	RPL41	4.3E-62
RPL39	5.1E-59	RPS28	2.8E-59
RPL37	2.3E-51	RPL39	5.1E-59
RPL34	1.4E-50	RPL37	2.3E-51
RPS27	4.7E-49	RPL34	1.4E-50
RPL35	2.4E-44	RPS27	4.7E-49
RPL36	5.8E-39	RPL35	2.4E-44
RPS21	3.3E-37	RPL36	5.8E-39
RPL35A	4.7E-34	RPS21	3.3E-37

H

Genes	# of sites in MPN data	# of sites in Cordblood ADAR Overexpression	Overlap
CDK13	27	12	4
SUMF2	50	21	11
IL6ST	13	0	0
STAT3	42	7	2

I

Gene	Position	Variant Classification
CDK13	7:40105692	Intron
	7:40105735	Intron
	7:39990548	Intron
	7:39990527	Missense
STAT3	17:40473342	Intron
	17:40471041	Intron

J

Gene	Position	Variant Classification
SUMF2	7:56146707	Missense
	7:56137111	Intron
	7:56137115	Intron
	7:56142670	Intron
	7:56142684	Intron
	7:56146659	Missense
	7:56146667	Missense
	7:56138064	Intron
	7:56132455	Intron
	7:56132456	Intron
	7:56138063	Intron

Figure S4. A-to-I RNA Editing Events in MPN and AML Compared with Normal Aged Bone Marrow Stem and Progenitor cells. Related to Figure 4. A. Correlation of normalized and Log2-transformed CPM data for the ADAR1 p110 isoform with mean A-to-I RNA editing in stem (square) or progenitor (triangle) population of each MPN subtype and AML. Each color represents a normal or disease phenotype. B. Normalized and Log2 transformed RNA-Seq expression data for SUMF2 in stem cells and progenitors plotted by phenotype. The results of t-tests (ns = not significant; $p < 0.05$ = *; $p < 0.01$ = **, $p < 0.005$ = ***) between each phenotype and the Aged Bone Marrow (ABM) Normal group are shown. C. tSNE scRNA-seq analysis of cord blood CD34+ cells transduced with lentiviral backbone control or shRNA targeting ADAR1 (shADAR1). D. Top ten differentially expressed genes between control shRNA (shCTRL) and shRNA targeting ADAR1 (shADAR1) in cord blood CD34+ cells. E. tSNE plot of genes expressed in scRNA-seq analysis cord blood CD34+ cells transduced with lentiviral backbone control (shCTRL) or shRNA targeting ADAR1 (shADAR1) that are found in the STRING interactome seeded with genes differentially edited between MF and aged normal bone marrow (ABM). Blue = shCTRL; red = shADAR1. F. tSNE plot of genes expressed in scRNA-seq analysis cord blood CD34+ cells transduced with lentiviral backbone control (shCTRL) or shRNA targeting ADAR1 (shADAR1) that are found in the STRING interactome seeded with genes differentially edited between AML and aged normal bone marrow (ABM). Blue = shCTRL; red = shADAR1. G. The top ten overlapping genes in MF compared with ABM and AML compared with ABM comparison that are affected by lentiviral ADAR1 shRNA knockdown. H. Overlapping A-to-I RNA edited sites between MPN RNA-seq and cord blood overexpressing ADAR1 dataset in highly edited transcripts. I. The positions of RNA editing in CDK13 and STAT3 are shown. J. The positions of RNA editing in SUMF2 are shown.

Supplemental Figure 5

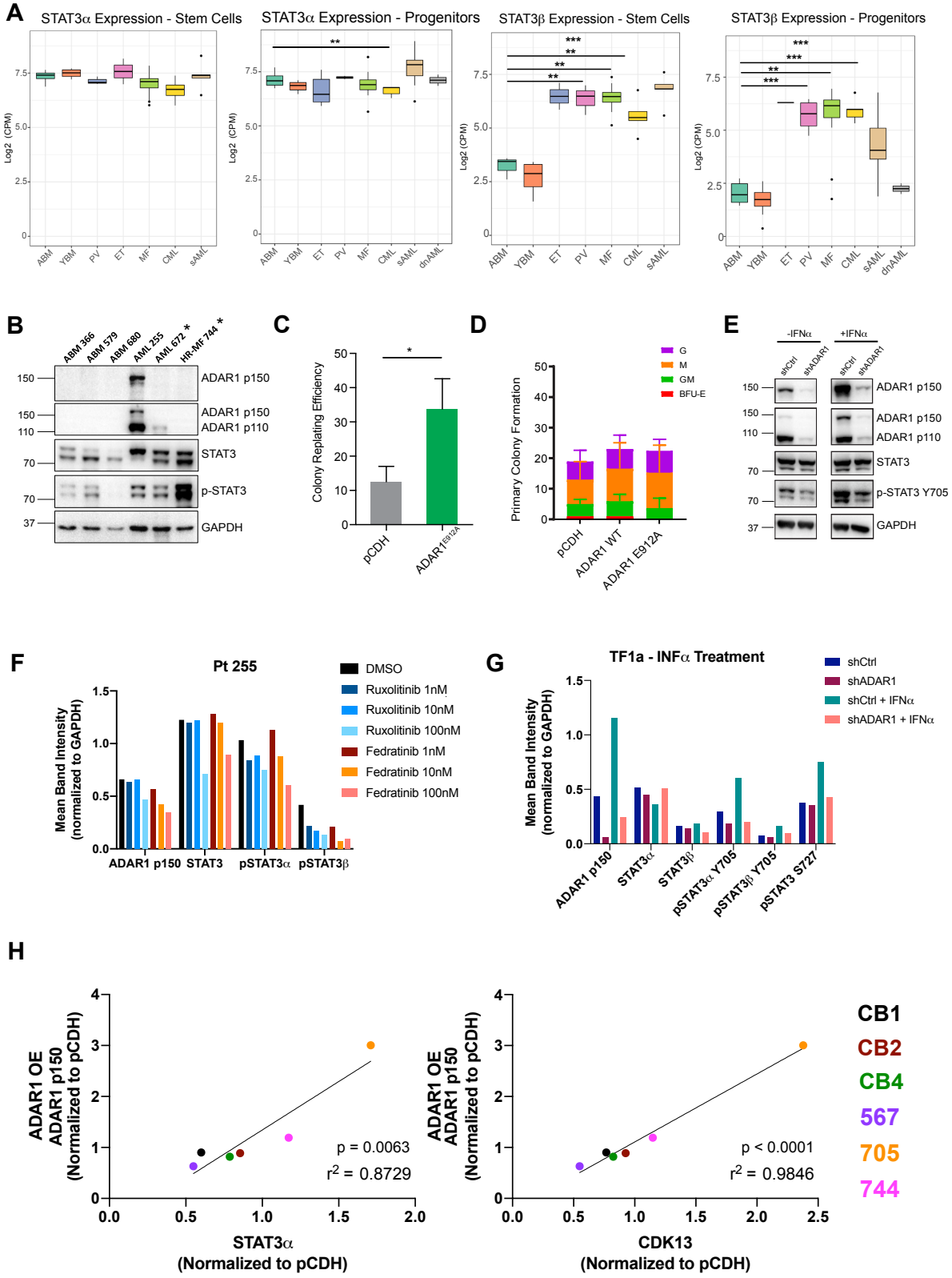


Figure S5. ADAR1-induced STAT3 Intronic Editing and Isoform Switching in LSC. Related to Figure 5. A. Expression of the STAT3α and STAT3β isoform levels in normal young (YBM), normal aged (ABM), and MPN and AML stem cells and progenitors using normalized and Log2-transformed RNA-Seq analyses. (Student's t-test, $p < 0.05$= *). B. ADAR1, STAT3, and pSTAT3 protein levels in CD34+ primary patient samples. SDS-PAGE Western blot analysis of whole cell extract isolated from aged bone marrow (ABM, lanes 1 - 3), acute myeloid leukemia (AML, lanes 4 and 5), and high-risk myelofibrosis (HR-MF, lane 6) CD34+ purified cells. *, clinically treated with ruxolitinib. C. Self-renewal capacity, as measured by colony replating assays, in MF CD34+ cells transduced with pCDH backbone or an ADAR1E912A deaminase deficient mutant. D. Total number of colonies formed in primary MF CD34+ cells overexpressing ADAR1, ADAR1E912A mutant, or pCDH backbone (n=8). E. Western blot of whole cell extracts from TF1a parental, shCtrl and shRNA-mediated ADAR1 knockdown cells treated with or without IFNα. F. Densitometry analysis of western blot of sAML (patient 255) CD34+ cells treated with FDA approved JAK2 inhibitors (ruxolitinib and fedratinib) compared with a JAK3 inhibitor (FM-381) at concentrations of 1nM, 10nM, and 100 nM. G. Densitometry analysis of western blot of whole cell extracts from TF1a parental, shCtrl and shRNA-mediated ADAR1 knockdown cells treated with IFNα. H. Correlation of ADAR1 p150 expression with the expression of STAT3α and CDK13. The CD34+ cells from cord blood (n = 3; CB1, CB2, CB4), sAML and high-risk MF samples (n = 3; 567, 705,744) were transduced with pCDH or ADAR1 overexpressing vectors. The relative gene expression levels were measured by qRT-PCR and normalized to HPRT values.