Stem Cell Reports, Volume *10*

Supplemental Information

Dissecting the Contributions of Cooperating Gene Mutations to Cancer Phenotypes and Drug Responses with Patient-Derived iPSCs

Chan-Jung Chang, Andriana G. Kotini, Malgorzata Olszewska, Maria Georgomanoli, Julie Teruya-Feldstein, Henrik Sperber, Roberto Sanchez, Robert DeVita, Timothy J. Martins, Omar Abdel-Wahab, Robert K. Bradley, and Eirini P. Papapetrou

Figure S1

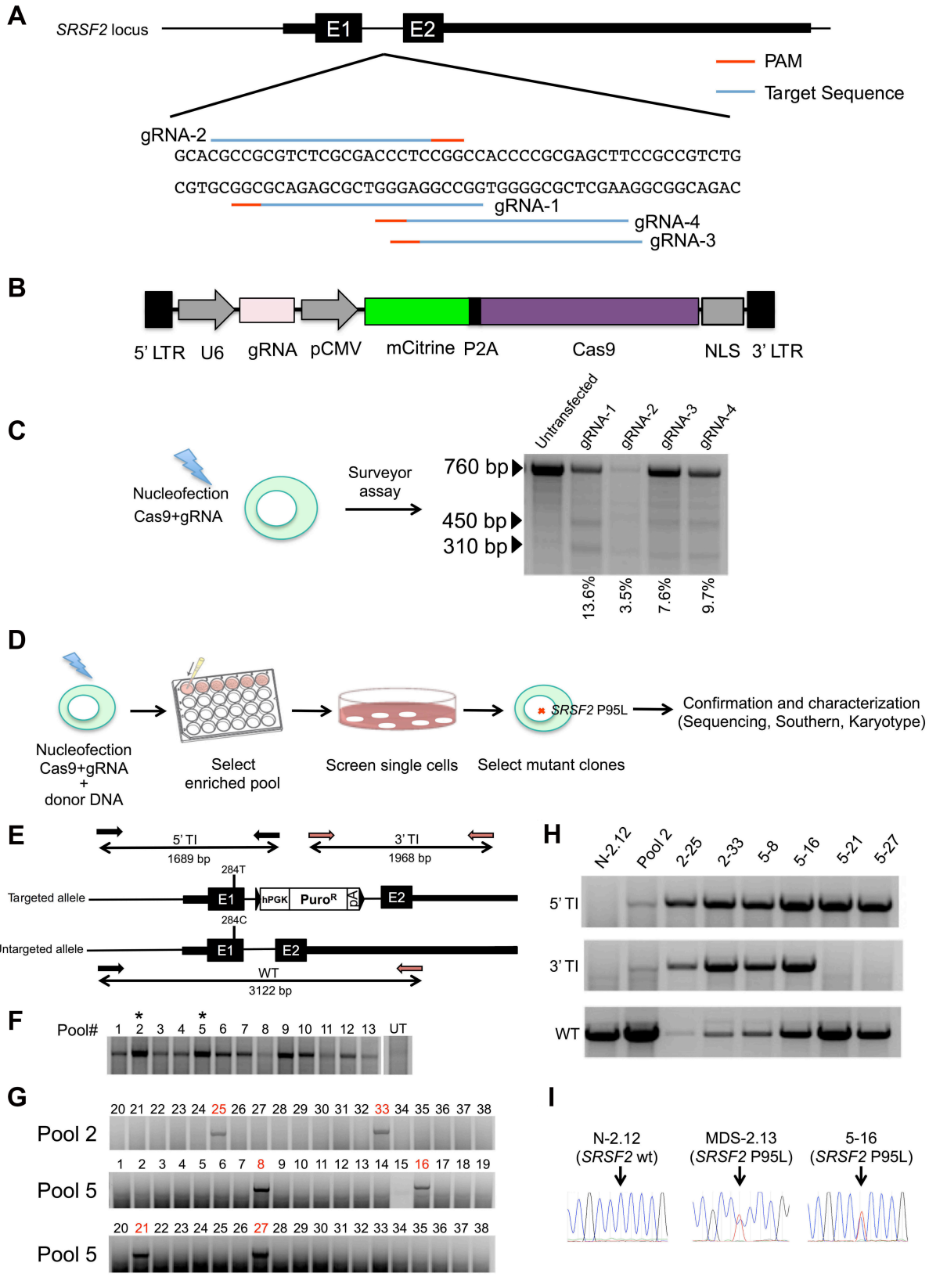
A
SRSF2 locus
E1
E2
PAM
Target Sequence
gRNA-2
GCACGCCGCGTCTCGCGACCCTCCGGCCACCCCGCGAGCTTCCGCCGTCTG
CGTGCGGCGCAGAGCGCTGGGAGGCCGGTGGGGCGCTCGAAGGCGGCAGAC
gRNA-1
gRNA-4
gRNA-3
B
5' LTR
U6
gRNA
pCMV
mCitrine
P2A
Cas9
NLS
3' LTR
C
Nucleofection
Cas9+gRNA
Surveyor
assay
Untransfected
gRNA-1
gRNA-2
gRNA-3
gRNA-4
760 bp
450 bp
310 bp
13.6%
3.5%
7.6%
9.7%
D
Nucleofection
Cas9+gRNA
+
donor DNA
Select
enriched pool
Screen single cells
Select mutant clones
SRSF2 P95L
Confirmation and characterization
(Sequencing, Southern, Karyotype)
E
5' TI
1689 bp
3' TI
1968 bp
284T
Targeted allele
E1
hPGK
Puro^R
E2
284C
Untargeted allele
E1
E2
WT
3122 bp
F
Pool#
1 2 3 4 5 6 7 8 9 10 11 12 13 UT
G
Pool 2
20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Pool 5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Pool 5
20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
H
N-2.12
Pool 2
2-25
2-33
5-8
5-16
5-21
5-27
5' TI
3' TI
WT
I
N-2.12
(SRSF2 wt)
MDS-2.13
(SRSF2 P95L)
5-16
(SRSF2 P95L)

Figure S1. Introduction of the *SRSF2* P95L mutation in normal iPSCs. Related to Figure 1

(A) The SRSF2 locus with the DNA sequences within intron 1 targeted by each of four gRNAs. Blue lines indicate target sequences and red lines the PAM sequence.
(B) gRNA/Cas9 lentiviral plasmid expressing Cas9 linked to mCitrine by a P2A peptide driven by the CMV immediate early promoter (pCMV) and a gRNA driven by the U6 promoter. NLS: nuclear localization signal; LTR: long terminal repeat.
(C) Surveyor assay after nucleofection of the plasmid shown in B encoding each of the gRNAs shown in a in N-2.12 iPSCs. The 450 bp and 310 bp bands correspond to surveyor nuclease cleavage products. Percentages denote the intensity of the 450 bp and 310 bp bands relative to the intensity of all 3 bands per lane. gRNA-1 and gRNA-4 were selected for subsequent experiments.
(D) Scheme of the experimental strategy for isolating gene edited clones. The N-2.12 cells were transfected with the gRNA/Cas9 and donor plasmids and divided in 48 pools. Selected pools were single-cell subcloned and individual clones were screened initially by PCR and further characterized by DNA sequencing, Southern blot and karyotyping.
(E) Scheme of the *SRSF2* locus with the position and amplicon length of the primers used for PCR screening indicated.
(F) Representative gel image of screening pools of transfected cells using the 5' TI (targeted integration) set of PCR primers shown in e. Pools 2 and 5 with the highest band intensity were selected.
(G) Representative gel image of screening of single-cell clones subcloned from pools 2 and 5, as indicated, using the 5' TI set of PCR primers shown in e. Positive clones are shown in red.
(H) Further testing of 6 selected single-cell clones (from D), with the 5'TI and 3' TI targeted integration-specific primer sets and the WT primer set to detect the untargeted allele (primer positions shown in E). All 6 clones were mono-allelically targeted (an untargeted allele could be detected in all). Clones 2-25, 2-33, 5-8 and 5-16 could be confirmed to have targeted integration of the intact donor sequence.
(I) Sanger sequencing of the selected gene edited clone 5.16, confirming the introduced 284C>T mutation. The parental *SRSF2* WT N-2.12 line and the patient-derived *SRSF2* mutant MDS-2.13 line are shown as controls.

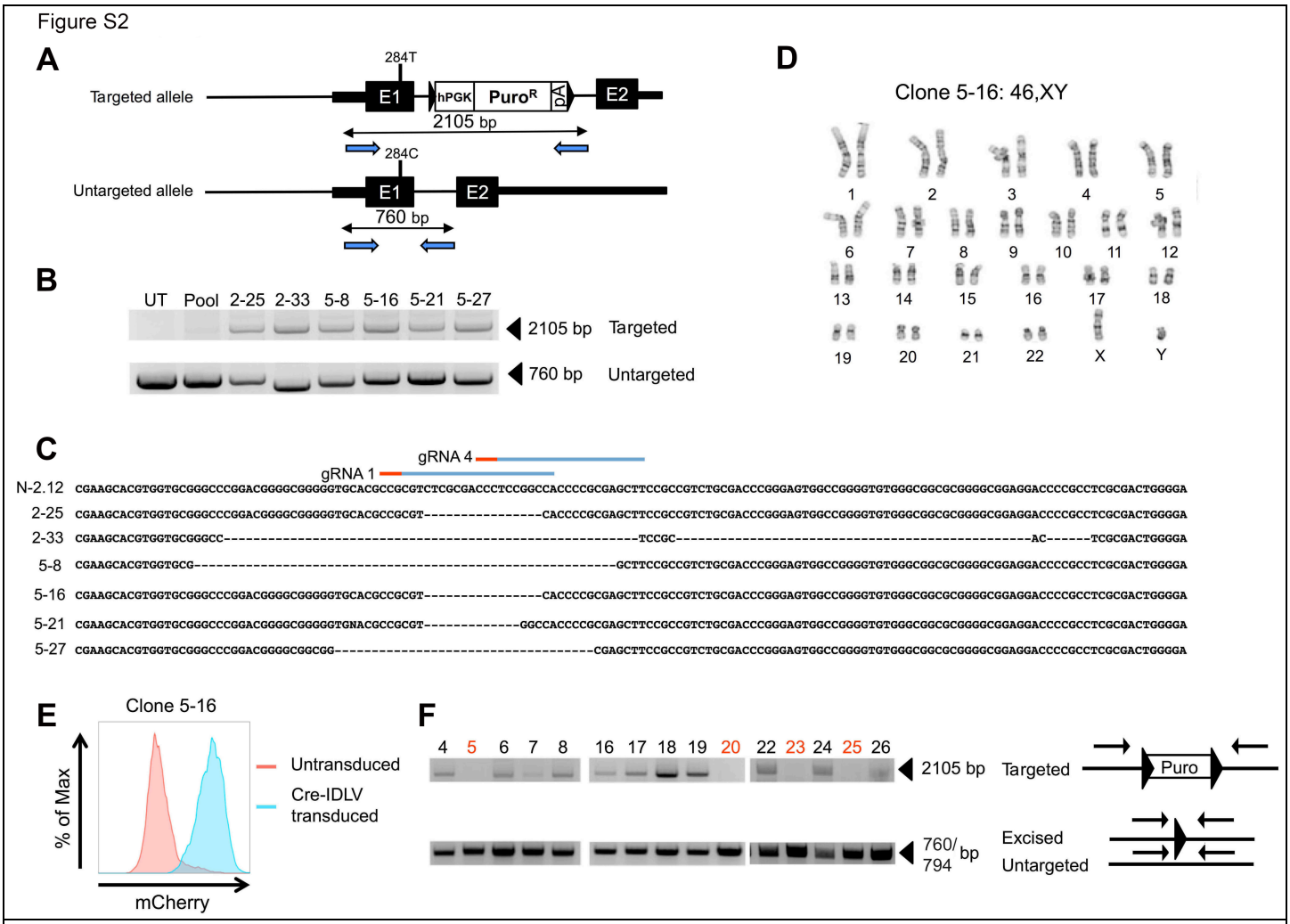

Figure S2. Characterization and cassette excision of *SRSF2* mutant iPSCs. Related to Figure 1

(A) Scheme of the *SRSF2* locus with the position and amplicon length of PCR primers used for additional characterization and sequencing of the untargeted allele indicated.
(B) PCR of the 6 targeted clones with the primers shown in A. The 2105 bp band corresponds to the targeted allele and the 760 bp band to the untargeted allele. Note that the 760 bp band is migrating faster in some clones (2-33, 5-8), found by sequencing to harbor large deletions encompassing the gRNA cleavage site mediated through NHEJ.
(C) Sequencing of the untargeted allele (purified 760 bp band shown in B) in the 6 targeted clones. All 6 clones contain deletions mediated through NHEJ.
(D) Karyotyping of the targeted 5-16 clone showing a normal male karyotype.
(E) mCherry expression in clone 5-16 48 hours after transduction with an IDLV expressing Cre recombinase linked through a P2A peptide with mCherry.
(F) Screening of single-cell clones after transduction of clone 5-16 with the mCherry-Cre IDLV for excision of the selection cassette with primers shown in A. Cassette-excised clones (shown in red) lose the 2105 bp band.

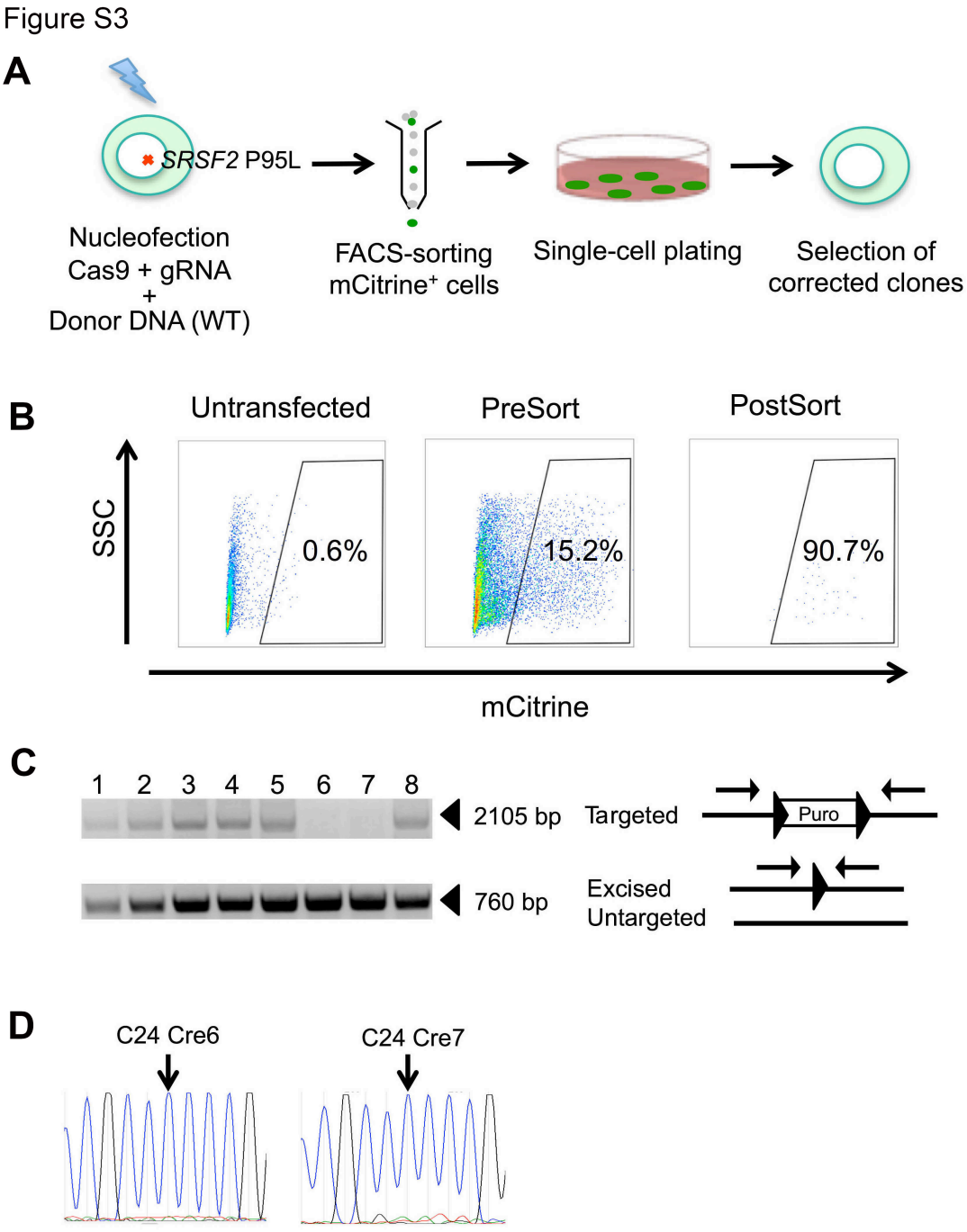

Figure S3. Correction of the *SRSF2* P95L mutation in MDS patient-derived iPSCs. Related to Figure 1

(A) Scheme of experimental strategy for isolating corrected clones. The MDS-2.13 cells were transfected with the gRNA/Cas9 and donor plasmids, mCitrine^{+} cells were FACS-sorted 72 hours after transfection and plated as single cells at clonal density.
(B) FACS-sorting of mCitrine^{+} cells after transfection with the gRNA/Cas9 and donor plasmids. (C) Screening of single cell clones for excision of the selection cassette after transduction of clone C24 with the mCherry-Cre IDLV. Cassette-excised clones (6 and 7) lose the 2105 bp band.
(D) Sequencing of the two targeted and cassette-excised clones C24 Cre6 and C24 Cre7 (clones 6 and 7 shown in C), confirming correction of the 284C>T mutation.

Figure S4

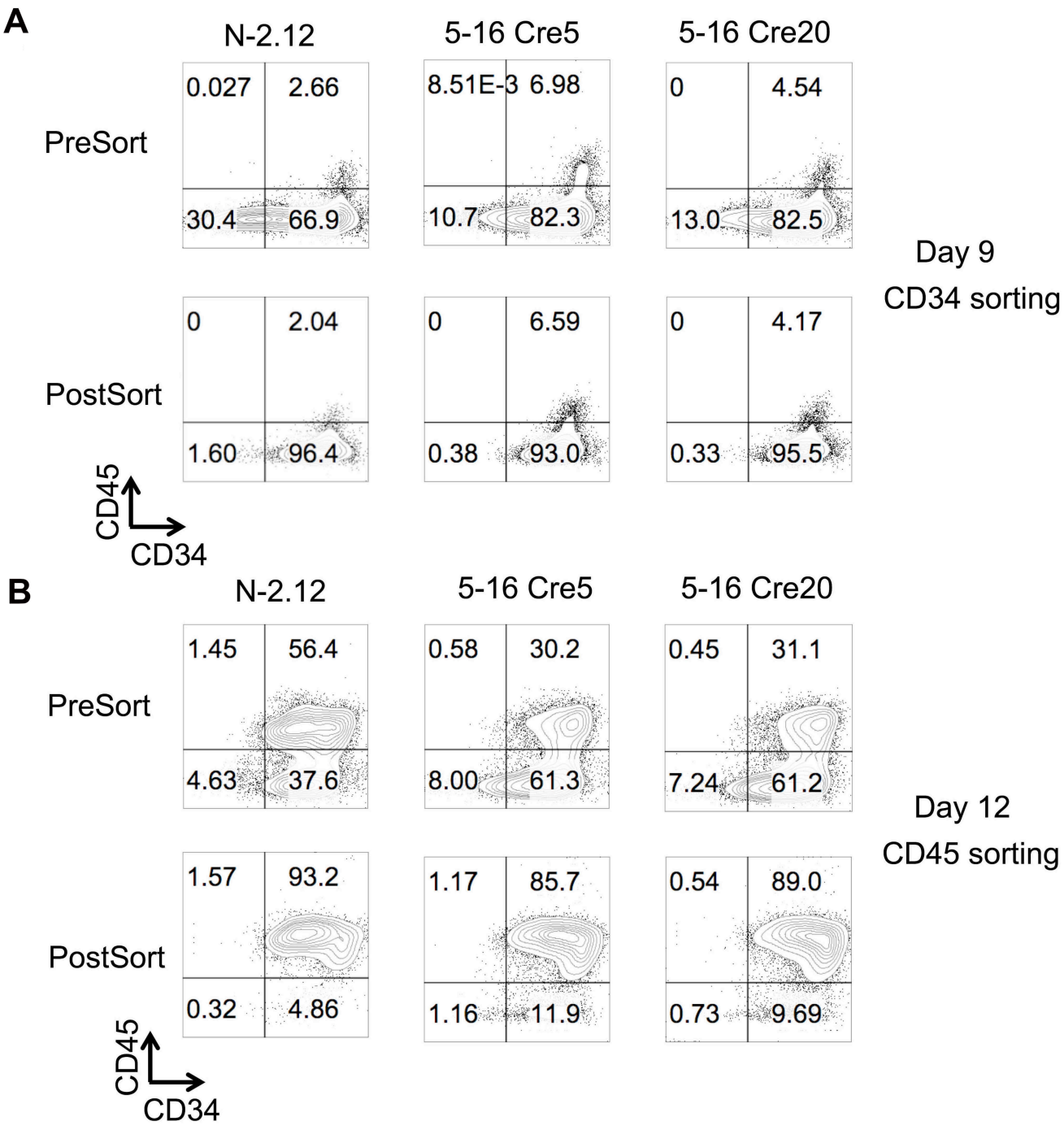

Figure S4. Sorting of isogenic SRSF2 WT and mutant iPSC-derived CD34$^+$ and CD45$^+$ hematopoietic progenitors for RNA sequencing analysis. Related to Figure 2

(A and B) Flow cytometric assessment of cell purity of N-2.12, 5-16 Cre5 and 5-16 Cre20 cells, sorted by magnetic sorting (MACS) for CD34 on day 9 (A) and CD45 on day 12 (B) of hematopoietic differentiation. Cell purity of CD34 sorting ranged from 98%-99.5%. Purity of CD45 sorting ranged from 87%-95%.

Figure S5

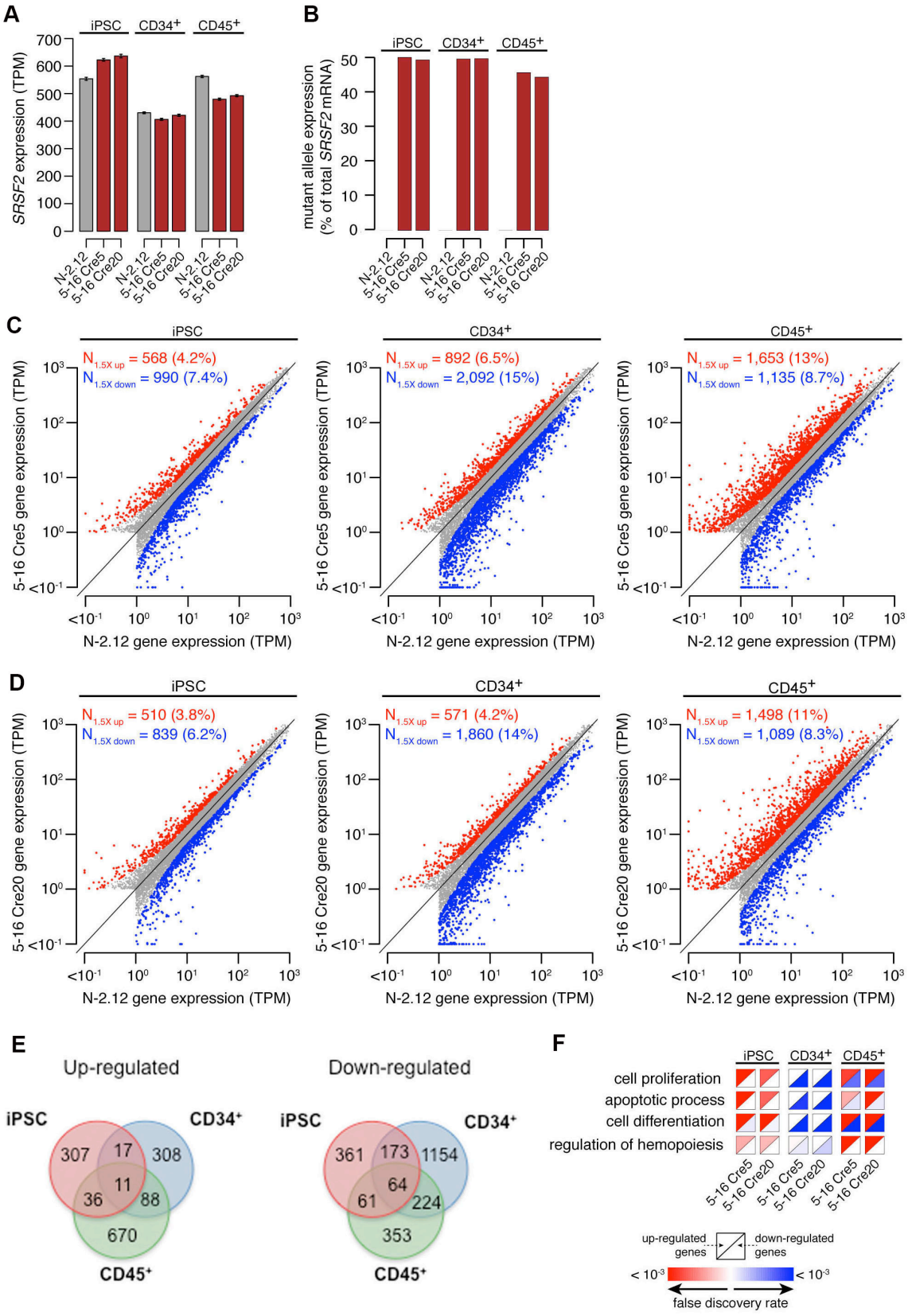
A
iPSC
CD34+
CD45+
SRSF2 expression (TPM)
N-2.12
5-16 Cre5
5-16 Cre20
B
mutant allele expression (% of total SRSF2 mRNA)
C
N1.5X up = 568 (4.2%)
N1.5X down = 990 (7.4%)
N1.5X up = 892 (6.5%)
N1.5X down = 2,092 (15%)
N1.5X up = 1,653 (13%)
N1.5X down = 1,135 (8.7%)
5-16 Cre5 gene expression (TPM)
N-2.12 gene expression (TPM)
D
N1.5X up = 510 (3.8%)
N1.5X down = 839 (6.2%)
N1.5X up = 571 (4.2%)
N1.5X down = 1,860 (14%)
N1.5X up = 1,498 (11%)
N1.5X down = 1,089 (8.3%)
5-16 Cre20 gene expression (TPM)
E
Up-regulated
Down-regulated
307
17
308
11
36
88
670
361
173
1154
64
61
224
353
F
cell proliferation
apoptotic process
cell differentiation
regulation of hemopoiesis
up-regulated genes
down-regulated genes
false discovery rate

Figure S5. Gene expression alterations in *SRSF2* mutant iPSC-derived hematopoietic cells. Related to Figure 2

(A) Total *SRSF2* mRNA levels as measured by RNA-seq. TPM, transcripts per million.
(B) Mutant *SRSF2* mRNA levels as a percentage of total *SRSF2* mRNA as measured by RNA-seq.
(C and D) Scatter plots of coding gene expression in *SRSF2* mutant 5-16 Cre5 and 5-16 Cre20 cells relative to the parental *SRSF2* WT N-2.12 cells. Red and blue dots represent up- and down-regulated coding genes in 5-16 Cre5 or 5-16 Cre20 vs. N-2.12 cells, respectively, where up/down-regulation are defined as increases/decreases in gene expression of magnitude ≥1.5- fold with a Bayes factor ≥100 as estimated by Wagenmakers's framework. Percentages indicate the fraction of expressed coding genes that are up or down-regulated. Coding genes with expression ≥1 TPM in at least one sample are included. TPM, transcripts per million as estimated by RSEM.
(E) Venn diagrams depicting the differentially expressed genes (upregulated, left and downregulated, right) in *SRSF2* mutant (either 5-16 Cre5 or 5-16 Cre20) vs *SRSF2* WT vs *SRSF2* WT (N-2.12) cells at the undifferentiated (iPSC) state, $CD34^+$ or $CD45^+$ state.
(F) Gene Ontology (GO) enrichment for differentially expressed coding genes in the indicated samples compared to the corresponding N-2.12 cells (undifferentiated iPSCs, $CD34^+$ or $CD45^+$).

Figure S6

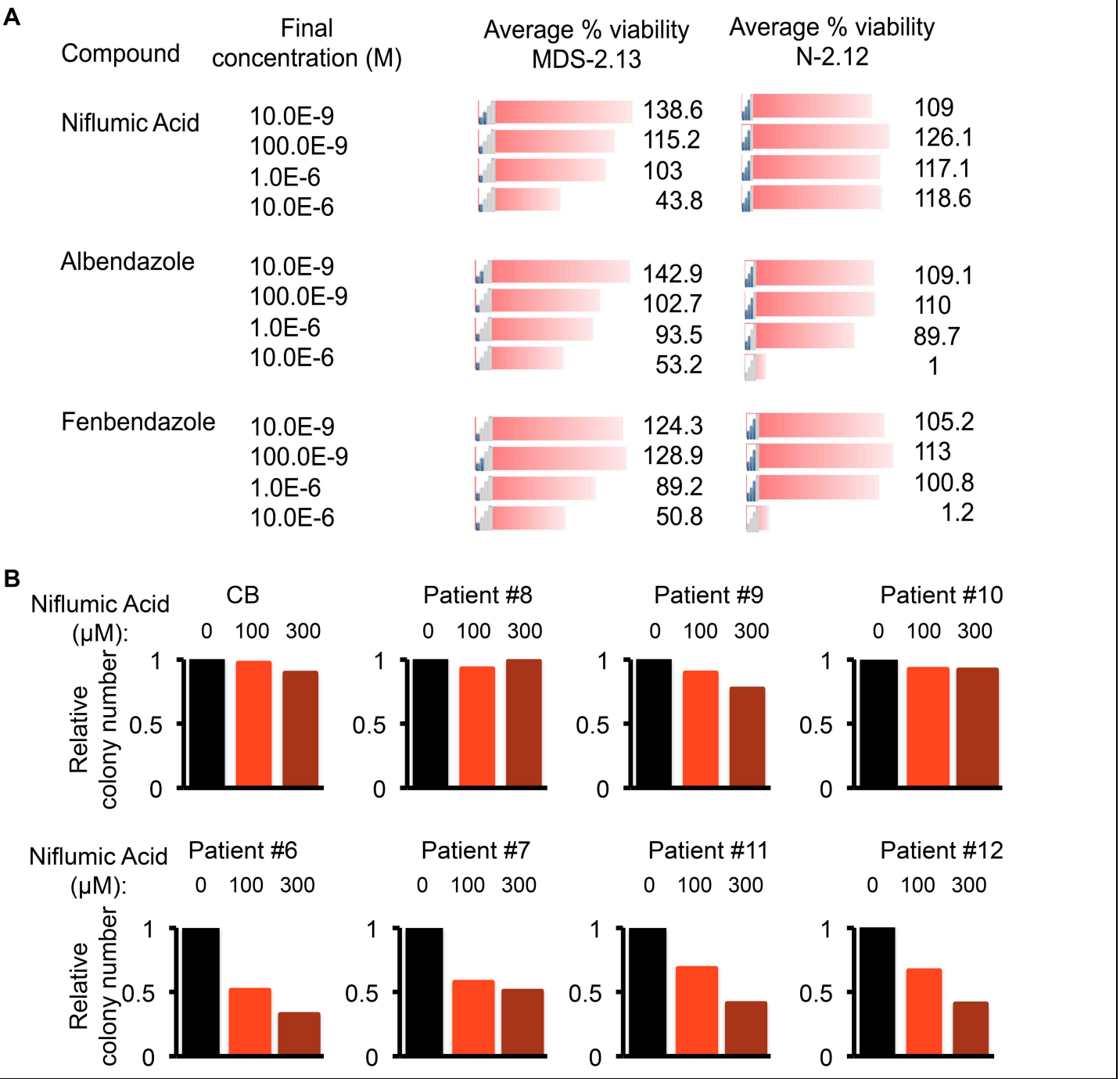

Figure S6. Selected hit compounds with preferential activity on MDS compared to isogenic normal iPSCs. Related to Figure 6

(A) Screen results of the three selected hit compounds niflumic acid, albendazole and fenbendazole at the indicated concentrations in the MDS-2.13 line and the isogenic normal N-2.12 line. Average percent viability at each compound concentration was calculated as: (Signal-Blank)/(DMSO Control-Blank) x 100.

(B) Treatment of primary cells from patients with MDS and sAML with (lower panels) or without (upper panels) chromosome 7q loss (see Table S6 for details on patient characteristics) and of cord blood $CD34^+$ cells (CB) plated in methylcellulose cultures with niflumic acid. The relative colony number was calculated from the total number of colonies on day 14 of methylcellulose culture relative to the DMSO-treated (untreated) cells.

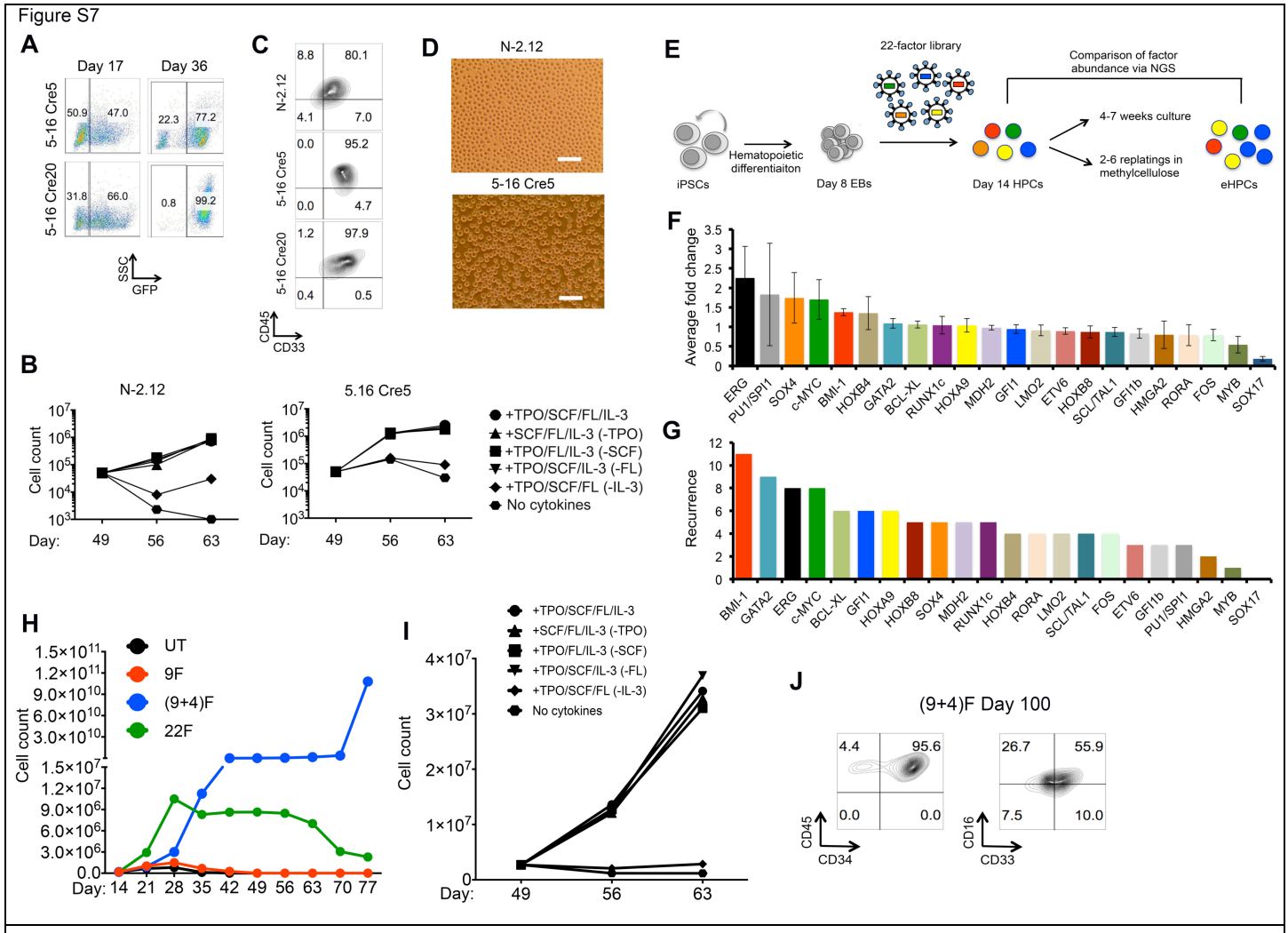

Figure S7. Characterization of expandable iPSC-HPCs. Related to Figure 7

(A) Selective expansion of the GFP$^+$ cells indicated by the increase in the fraction of GFP$^+$ cells between days 17 and 36 of culture in hematopoietic differentiation conditions.
(B) Expansion of eHPCs grown in liquid culture supplemented with different cytokine combinations, as indicated. Removal of TPO, SCF or FL from the medium has no effect on cell growth. In contrast, IL-3 is necessary for cell expansion.
(C and D) Immunophenotype (C) and representative bright-field images (D) of eHPCs after freezing and thawing after 50 days of culture.
(E) Scheme of experimental setup to identify factors that become enriched over time in eHPC culture.
(F) Average fold change and SEM in the representation of the 22 factors over time in liquid or methylcellulose culture across 14 comparisons from 11 independent experiments using different iPSC lines calculated by barcode next-generation sequencing.
(G) A recurrence score for each gene was calculated as the number of times the gene was found to be enriched (>1.1 fold change) across 14 comparisons from 11 independent experiments.
(H) Expansion of eHPCs in liquid culture estimated by cell counts. One of two independent experiments is shown. UT: untransduced.
(I) Expansion of 9+4F eHPCs grown in liquid culture supplemented with different cytokine combinations, as indicated. 9+4F eHPCs are dependent on IL-3.
(J) Immunophenotype of 9+4F eHPCs on day 100 of culture.

Table S1

Cell line name	7q status	SRSF2 status	Parental line	Comments
N-2.12	Normal	WT/WT	NA	Patient-derived (Kotini et al. 2015)
5-16 Cre5	Normal	WT/P95L	N-2.12	CRISPR-edited (this study)
5-16 Cre20	Normal	WT/P95L	N-2.12	CRISPR-edited (this study)
MDS-2.13	Del(7q)	WT/P95L	NA	Patient-derived (Kotini et al. 2015)
C24 Cre6	Del(7q)	WT/WT	MDS-2.13	CRISPR-edited (this study)
MDS-2.A3C	Normal	WT/P95L	MDS-2.13	Spontaneously corrected (Kotini et al. 2015)
Cre10	Del(7q)	WT/WT	N-2.12	Genetically engineered (Kotini et al. 2015)
8Cre21	Del(7q)	WT/WT	N-2.12	Genetically engineered (Kotini et al. 2015)

Table S1. Cell lines used in this study. Related to Figures 3, 4

Table S2

Cell line name	Dysplastic changes
N-2.12	No dysplastic morphology noted
5-16 Cre5	<1% Micromegakaryocytes, 30% hypolobular myeloid cells, 30%-40% hypogranular myeloid cells
5-16 Cre20	<1% Micromegakaryocytes, 30% hypolobular myeloid cells, 30%-40% hypogranular myeloid cells
MDS-2.13	100% hyperlobulated/hypolobulated/hypogranular myeloid cells
C24 Cre6	ND
MDS-2.A3C	<1% Micromegakaryocytes, 30% hypolobular myeloid cells, 30%-40% hypogranular myeloid cells
Cre10	No dysplastic morphology noted
8Cre21	No dysplastic morphology noted

Table S2. Quantification of dysplastic changes. Related to Figure 3

Table S6						
Patient	**Sample Source**	**Diagnosis**	**Stage**	**Sample Drawn**	**BM Blasts %**	**Cytogenetic analysis**
Patient 6	PB	MDS-RCMD	Refractory	Pre-Treatment	BM: 2.5% M, 3.5% F; PB: 0% F	43-48,XX,t(3;16)(q?21;24),del(5)(q13q23),**-7**,-12,-20,+1~5mar(cp7)[7]/46,XX[13]
Patient 7	PB	MDS-RAEB2	Diagnosis	Pre-Treatment	BM: 13% M, 4.4% F; PB: 9%	45~47,X,add(Y)(q11.23),add(4)(q12),der(5;17)(p10;q10),-6,**del(7)(q22q34)**,+8,-18,add(18)(q11.2),-20,-22,+1~3r,+mar1,+2~4mar[cp20]
Patient 8	BM	MDS-RCMD	Diagnosis	Pre-Treatment	1% M, 1% F	47,XY,+8[12]/46,XY[8]
Patient 9	PB	MDS-RAEB2	Diagnosis	Pre-Treatment	BM: 17.5% M, 21.9% F; PB: 16%	46, XX[20]
Patient 10	BM	MDS/MPN	Residual	Pre-Treatment	1% M, 1.7% F	46,XY[20]
Patient 11	BM	sAML	Refractory	Pre-Treatment	15% M, 21% F	43,XY,del(1)(q32),der(1)del(1)(p36.1)t(1;7)(q21;q11.2)add(7)(q36),der(2)inv(2)(p23q35)add(2)(p11.2),del(3)(q21q2?1),del(4)(q22q25),add(5)(q31),**-7**, der(7)t(1;7),add(8)(p11.2),-11,-12,-13,add(14)(p11.2),add(15)(p11.2), del(15)(q?11.2q24),-16,del(17)(p11.2),add(20)(p13),add(21)(q22),+2~4mar[cp6]/ 41~43,sl,18[cp3]/41~43,sl,del(18)(p11.1p11.2)[cp3]//46,XX[8]
Patient 12	BM	sAML	Refractory	Pre-Treatment	49% M, 34% F	45,XY, **-7**, [4]/46, sl, +mar[12]/47,sdl1, +12 [4]
	PB: peripheral blood	RCMD: Refractory anemia with multilineage displasia			M: morphology	
	BM: bone marrow	RAEB: Refractory anemia with excess blasts			F: Flow cytometry	
		MDS/MPN: Myelodysplastic/myeloproliferative overlap syndrome				
		sAML: Secondary acute myeloid leukemia (from prior MDS)				
Table S6. Patient characteristics. Related to Figure 6						

Table S7

Gene Name	Entrez	Cloned from:	cDNA size (bp)	References	Barcode
ERG	NM_182918	pINDUCER21-ERG	1440	Doulatov S et al. Cell Stem Cell, 2013; Batta K et al. Cell Rep, 2014; Sugimura R et al. Nature, 2017	TGAA
HOXA9	NM_152739	pINDUCER21-HOXA9	819	Doulatov S et al. Cell Stem Cell, 2013; Sugimura R et al. Nature, 2017	TAAT
RORA	NM_134261	pINDUCER21-RORA	1572	Doulatov S et al. Cell Stem Cell, 2013	CAAA
MYB	NM_001130173	pINDUCER21-MYB	2286	Doulatov S et al. Cell Stem Cell, 2013	CTAA
SOX4	NM_003107	pINDUCER21-SOX4	1425	Doulatov S et al. Cell Stem Cell, 2013	CCAA
ETV6	NM_001987	pENTR-ETV6	1359	Pereira CF et al. Cell Stem Cell, 2013	GTAA
GFI1b	NM_004188	pENTR-GFI1B	993	Pereira CF et al. Cell Stem Cell, 2013	GGAA
FOS	NM_005252	PENTR-FOS	1143	Pereira CF et al. Cell Stem Cell, 2013; Sandler V et al. Nature, 2014; Lis R et al. Nature, 2017	CCAT
GATA2	NM_001145661	pENTR-hGATA2	1413	Pereira CF et al. Cell Stem Cell, 2013; Batta K et al. Cell Rep, 2014	CGAT
GFI1	NM_001127215	pENTR-GFI1	1269	Sandler V et al. Nature, 2014; Lis R et al. Nature, 2017	GCAA
PU1/SPI1	NM_001080547	PENTR-PU1	795	Sandler V et al. Nature, 2014; Sugimura R et al. Nature, 2017; Lis R et al. Nature, 2017	GAAT
RUNX1c	NM_001754	pENTR-RUNX1c	1443	Sandler V et al. Nature, 2014; Batta K et al. Cell Rep, 2014; Riddell et al. Cell, 2014; Sugimura R et al. Nature, 2017; Lis R et al. Nature, 2017	TCAT
LMO2	NM_005574	pCMV6-LMO2	684	Batta K et al. Cell Rep, 2014; Riddell et al. Cell, 2014	ATAT
SCL/TAL1	NM_001290404	pENTR-TAL1	996	Batta K et al. Cell Rep, 2014	ACAT
HOXB4	NM_024015	pENTR-HOXB4	756	Kyba M et al. Cell, 2002	TTAA
c-MYC	NM_002467	pLM-mCerulean-cMYC	1320	Hirose S et al. Stem Cell Reports, 2013	AGAT
BCL-XL	NM_138578	pCDH-puro-BCL-XL	710	Hirose S et al. Stem Cell Reports, 2013	TCAA
HMGA2	NM_003483	pMXS-hs-HMGA2	330	Cavazzana-Calvo M et al. Nature, 2010	GTAT
SOX17	NM_022454	pENTR-SOX17	1245	Nakajima-Takagi Y et al. Blood, 2013	ACAC
HOXB8	NM_024016	pENTR-HOXB8	732	Redecke V et al. Nature Methods, 2013	TAAA
BMI-1	NM_005180	pT3-EF1a-BMI1	981	Rizo A et al. Blood, 2008	CGAA
MDH2	NM_001282404	pNIC28-MDH2	696	Shojaei F et al. Blood, 2004	GAAA

Table S7. Library genes for generation of eHPCs. Related to Figure 7

Supplemental Experimental Procedures

CRISPR/Cas9 genome editing

A lentiviral plasmid (gRNA/Cas9) co-expressing a human codon optimized Cas9 with a nuclear localization signal (from George Church, Addgene plasmid # 41815) linked to mCitrine by a P2A peptide driven by the CMV immediate early promoter and a gRNA driven by the U6 promoter was constructed (shown in Figure S1B). gRNAs targeting the *SRSF2* locus within the first intron (cutting site between 201 bp and 215 bp from the 284 C>T mutation site, sequences shown in Figure S1A) were amplified in a two-step overlapping PCR reaction downstream of the U6 promoter sequence in a single fragment and cloned in the gRNA/Cas9 plasmid. A donor plasmid containing a floxed cassette consisting of a human PGK promoter driven puromycin resistance gene and 5' (1442 bp) and 3' (1244 bp) homology arms consisting of nucleotides 74732743 – 74734184 and 74731499 – 74732742 (hg19 human genome assembly), respectively, was constructed. The entire 5'+3' homology sequence was amplified from N-2.12 and MDS-2.13 gDNA (alleles harboring 284C and 284T, respectively) in two independent PCR reactions, subcloned and sequenced, and one allele was chosen for subsequent cloning in the donor plasmid.

iPSCs were cultured in 10 μM Y-27632 for at least one hour before nucleofection and dissociated into single cells with accutase. 1 million cells were used for nucleofection with 5 μg of CRISPR/Cas9 plasmid(s) and 10 μg of donor plasmid using Nucleofector II (Lonza) and program B-16 and replated on MEFs. In the experiments aimed at evaluating the different gRNAs, only the gRNA/Cas9 plasmid was transfected and cells were harvested 3 days after nucleofection for surveyor assay. Following PCR with primers F: GTG GAC AAC CTG ACC TAC CG and R: GGT CGA CCG AGA TCG AGA AC, the PCR product was treated with Surveyor Nuclease using the Surveyor Mutation Detection kit (IDT) and analyzed in an agarose gel stained with EtBr. Band intensity was quantified by Image Lab (Bio-rad). In the experiments aimed at selecting single edited clones, iPSCs were harvested with accutase 3 days after nucleofection and plated on MEFs either at clonal density (500 cells per well of a 6-well plate) or divided in pools in 24-well plates or sorted for expression of mCitrine and plated at clonal density. After 7-10 days of plating single cells, single colonies were picked in separate wells of a 6-well plate, allowed to grow for approximately 3-6 days and screened by PCR. Approximately 100 cells were picked directly into a 0.2 ml tube, pelleted and lysed. PCR was performed with primers: 5'TI-F: CCC AGG TTT AGG GCG AAG TT, 5'TI-R: AAG AAT GTG CGA GAC CCA GG, 3'TI-F: AGC AAC AGA TGG AAG GCC TC, 3'TI-R: GCC AGT TGC TTG TTC CAA GG, WT-F: CCC AGG TTT AGG GCG AAG TT, WT-R: GCC AGT TGC TTG TTC CAA GG. For detection and sequencing of the untargeted allele and of the excised allele, PCR primers F and R (sequences listed above) were used. Southern Blot analyses were performed as previously described.(Papapetrou and Sadelain, 2011) Transduction with a Cre-expressing IDLV was performed as previously described.(Papapetrou and Sadelain, 2011)

Human iPSC culture and hematopoietic differentiation

Culture of human iPSCs on mouse embryonic fibroblasts (MEFs) or in feeder-free conditions was performed as previously described.(Papapetrou and Sadelain, 2011) For hematopoietic differentiation, spin embryoid bodies (EBs) were prepared and cultured in APEL medium, as described.(Ng et al., 2008) Briefly, cells were dissociated into single cells with accutase and plated at 3,500 cells per well in round-bottom low-attachment 96-well plates in APEL medium containing 30 ng/ml bone morphogenetic protein 4 (BMP4) and 10nM Y-27632. The plates were centrifuged at 800 rpm for 5 min to induce EB aggregation. After 24 hours, the medium was replaced by APEL medium containing 30 ng/mL BMP4 and 50 ng/mL FGF2. After 2 days, the cytokine cocktail was changed to: 20 ng/ml vascular endothelial growth factor (VEGF), 10 ng/ml FGF2, 100 ng/ml stem cell factor (SCF), 20 ng/ml Flt3 ligand (Flt3L), 20 ng/ml thrombopoietin (TPO), 40 ng/ml IL-3. At day 8, EBs were collected and resuspended in Stem Pro34 SFM medium with 1% nonessential amino acids (NEAA), 1 mM L-glutamine and 0.1 mM β-mercaptoethanol (2ME), supplemented with 100 ng/ml SCF, 20 ng/ml Flt3L, 20 ng/ml TPO, 40 ng/ml IL-3. The medium was thereafter replaced every two days. At the end of the EB differentiation culture (day 8, 10, 12 or 14, depending on the assay) the cells were collected and dissociated with accutase into single cells for downstream readouts.

Gene expression analysis by qRT-PCR
RNA was isolated with the RNeasy mini kit (Qiagen). Reverse transcription was performed with Superscript III (Life Technologies) and qPCR was performed with the SsoFast EvaGreen Supermix (Bio-Rad) using primers SRSF2-F: CCC GGA CTC ACA CCA CAG, SRSF2-R: ACC GAG ATC GAG AAC GAG TG, ACTIN-F: TGA AGT GTG ACG TGG ACA TC, ACTIN-R: GGA GGA GCA ATG ATC TTG AT. Reactions were carried out in triplicate in a 7500 Fast Real-Time PCR System (Applied Biosystems).

RNA sequencing
Magnetic cell sorting of $CD34^+$ and $CD45^+$ cells was performed on day 9 and 12 of hematopoietic differentiation, respectively, using the MACS cell separation microbeads and reagents (Miltenyi Biotec). Total RNA from undifferentiated iPSCs, sorted $CD34^+$ and sorted $CD45^+$ cells was extracted with the RNeasy mini kit (Qiagen). PolyA-tailed mRNA was selected with beads from 1µg total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs). cDNAs were generated using random hexamers and ligated to barcoded Illumina adaptors with the NEXTflex Rapid Directional RNA-Seq Library Prep Kit (Bioo Scientific). Sequencing of 75 nucleotide-long single-end reads was performed in a NextSeq-500 (Illumina).

Genome annotations and read mapping
Illumina reads were mapped to the UCSC hg19 (NCBI GRCh37) genome assembly as previously described.(Dvinge et al., 2014) Briefly, genome annotations from MISO v2.0,(Katz et al., 2010) UCSC knownGene,(Meyer et al., 2013) and Ensembl 71(Flicek et al., 2013) were merged to create a single genome annotation. Reads were then sequentially mapped, wherein RSEM and Bowtie(Langmead et al., 2009) were used to map reads to the transcriptome, and Tophat(Trapnell et al., 2009) was used to map remaining unaligned reads.

Differential splicing and gene expression analysis
Isoforms that were differentially spliced in mutant versus wild-type cells were identified as previously described.(Ilagan et al., 2015) Briefly, MISO(Katz et al., 2010) was used to estimate isoform abundance in each sample individually using v2.0 of its annotations. Subsequent analyses were confined to splicing events for which there were at least 20 identifiable reads, meaning reads that uniquely supported a single isoform, between each of the two samples used in each comparison. Differentially spliced isoforms were then defined as those that exhibited an absolute difference in isoform abundance of at least 10% and a Bayes factor of at least 5. The Bayes factor was computed using Wagenmaker's alternative to the binomial proportion test.(Wagenmakers et al., 2010) Differentially expressed genes were identified using a similar method, where the Bayes factor was computed using counts of reads supporting a particular gene relative to the total number of reads in expressed genes. Differentially expressed genes were defined as those with change in expression of magnitude ≥1.5- fold with a Bayes factor ≥100. Gene expression values were computed by RSEM and normalized using the TMM method relative to all coding genes.(Robinson and Oshlack, 2010)

Data analysis and visualization
Statistical analysis of gene and isoform expression measurements was performed in R with Bioconductor.(Huber et al., 2015) Gene Ontology analysis was performed using GOseq with a bias correction based on numbers of mapped reads per gene.(Young et al., 2010) Plots were created with ggplot2 [H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.].

Flow cytometry and cell sorting
For flow cytometry, the following antibodies were used: CD34-PE (clone 563, BD Pharmingen), CD45-APC (clone HI30, BD Pharmingen), CD33-PECF594 (clone WM53, BD Biosciences) and CD16-BV510 (clone 3G8, BD Biosciences). Cell viability was assessed with DAPI (Life Technologies). Cells were then assayed on a BD Fortessa and data were analyzed with FlowJo software (Tree Star). Sorting of $mCitrine^+$ cells was performed on a BD FACS Aria II.

Cytological analyses
Approximately 200,000 cells from liquid hematopoietic differentiation cultures or methylcellulose cultures were washed twice with PBS containing 2% FBS and resuspended in PBS. Cytospins were prepared on

slides using a Shandon CytoSpin III cytocentrifuge (Thermo Electron Corporation). Slides were then air-dried for 30 mins and stained with the Hema 3 staining kit (Fisher Scientific Company LLC). The slides were read on a Nikon Eclipse Ci microscope in a manner blinded to the disease status and genotype and digital images were taken with a Nikon DS-Ri2 camera and NIS-Elements D4.40.00 software.

Cell growth assays

For growth competition assays, N-2.12-GFP, a clonal line previously generated from the N-2.12 line after transduction with a lentiviral vector expressing eGFP,(Kotini et al., 2015) was mixed with each test iPSC line 1:1 and 3.5 x 10^3 cells/well were plated in duplicate round bottom ultra low attachment 96-well plates with APEL medium and cytokines to initiate hematopoietic differentiation. 36 wells were harvested the next day (day 2) and on days 4, 8 and 12 and the GFP^- cells were measured by flow cytometry. The relative population size of GFP^- cells at each time point was calculated relative to the population size on day 2.

Clonogenic assays

For methylcellulose assays, the cells were resuspended in StemPro-34 SFM medium at a concentration of 3 x 10^4/ml. 500µl of cell suspension were mixed with 2.5 ml MethoCult GF+ (H4435, Stem cell technologies) and 1 ml was plated in duplicate 35-mm dishes. Colonies were scored after 14 days and averaged between the duplicate dishes.

Treatment with splicing inhibitor drugs

E7107 was kindly provided by H3 Biomedicine Inc. Cpd-1, Cpd-2, and Cpd-3 were synthesized by WuXi AppTec (Shanghai, China) based on published structures from Araki et al(Araki et al., 2015). For drug treatment assays, the 5-16 Cre20 line was mixed 1:1 with the N-2.12-GFP clone. The drugs were added on day 8 of hematopoietic differentiation at concentrations: E7107 0.1 nM, Cpd-1 5 uM, Cpd-2 0.5 uM, Cpd-3 0.5 uM. The GFP^- fraction was measured by flow cytometry and the relative population size of GFP^- cells at each time point was calculated relative to the population size of untreated cells. In a separate experiment, the 5-16 Cre20 line was differentiated and treated on day 8 of hematopoietic differentiation with the drugs at concentrations as above and total RNA was isolated on days 10 and 12, reverse transcribed and expression levels of *DNAJB1* and *EIF4A1* pre-mRNA and mature mRNA were measured by real-time PCR.

Small molecule screen

The small molecule screen was performed at the University of Washington Quellos High Throughput Screening Core. The MicroSource Discovery Systems "Spectrum Collection" library was used which contains 2,000 compounds covering a wide range of biological activities and structural diversity, including known drugs, experimental bioactives and pure natural products. The library was supplied on 96well formatted storage plates (10mM in DMSO) and was reformatted into 384-well screening plates. These plates were sequentially subjected to 10-fold dilutions thereby generating 384-well plates containing the 2,000 chemical entities at 10, 1.0, 0.1 and 0.01 mM in DMSO.

The experimental conditions were optimized in pilot experiments. Using robotic equipment, 384-well plates were first coated with 20ul of Matrigel (BD Biosciences) diluted 1:20 in DMEM/F12 media for 30 minutes at RT. The iPSCs were subsequently plated as a single cell suspension (dissociated with accutase) at a density of 1,500 per well in 50 ul TeSR medium (Stem Cell Technologies) with 10uM Rock inhibitor Y-27632. The next day, the cells were washed and the compounds (in 50 nl) were added to wells containing cells with media (50µL) with slotted pins effectively making a 1 x 1,000 fold dilution of compound stock. The final concentration of compounds ranged from 10, 1.0, 0.1 and 0.01 µM at a solvent (DMSO) concentration of 0.1%. Pins were successively washed with 100% DMSO, 100% Methanol and air dried prior to another round of compound addition. Controls (blanks, solvent treated cells and Mitomycin C in 16 concentrations ranging from 10µM to 0.7pM) were included in each plate (each 384-well plate contained 320 wells of compound entities, 32 blanks, 16 wells of DMSO solvent controls and 16 wells of Mitomycin C control). On day 5, luminescence was measured using Promega's CellTiter-Glo assay per the manufacturer's suggested conditions with the exception that 5µL of reagent was used per 50µL culture mix per well and recorded on a Perkin Elmer Envision Multilabel Detector Plate Reader. Percent viability at each compound concentration was calculated as: (Signal-Blank)/(DMSO Control-Blank) x 100.

Treatment of primary patient cells with niflumic acid

Patient samples were obtained at the Fred Hutchinson Cancer Research Center with informed consent under a protocol approved by a local Institutional Review Board. After thawing the cells were cultured in X-VIVO 15 media with 20% BIT 9500 serum substitute (Stem CellTechnologies), 1% nonessential amino acids (NEAA), 1 mM L-glutamine and 0.1 mM b-mercaptoethanol (2ME), supplemented with 100 ng/ml stem cell factor (SCF), 100 ng/ml Flt3 ligand (Flt3L), 100 ng/ml thrombopoietin (TPO) and 20 ng/ml IL-3 for 2 days to induce cell proliferation. 50,000-200,000 cells, depending on the total cell number, and equal for each patient sample, were treated with different concentrations of niflumic acid (50-300 μM) or DMSO (untreated), added each day for 3 days. Live cells were quantified by cell counting in a hemocytometer 1, 3 and 5 days after the beginning of treatment. In parallel, $1x10^4$-$5x10^4$ cells were plated in methylcellulose media (H4435, Stem cell technologies) in duplicates, as detailed above, and DMSO or niflumic acid was added to the medium in the beginning of methylcellulose cultures, at 100 μM and 300 μM. The colonies were scored after 14 days.

Generation of expandable iPSC-derived HPCs

A lentiviral library of 22 human cDNAs encoding the factors shown in Table S7, tagged to unique 4-nt barcodes, was constructed in the mP2A lentiviral backbone, as previously described(Kotini et al., 2015). Packaging was performed as described previously.(Kotini et al., 2015; Papapetrou and Sadelain, 2011) Each vector plasmid was independently co-transfected with the two packaging plasmids and the supernatants were pooled and concentrated 400-1000-fold. iPSCs were differentiated for 8 days along the hematopoietic lineage, as described above, and day 8 embryoid bodies were transduced with the library at a multiplicity of infection (MOI) that gives approximately 30% transduced cells (which is expected to yield a low vector copy number per cell). The cells were thereafter cultured in Stem Pro34 SFM medium with 1% nonessential amino acids (NEAA), 1 mM L-glutamine and 0.1 mM β-mercaptoethanol (2ME), supplemented with 100 ng/ml SCF, 20 ng/ml Flt3L, 20 ng/ml TPO and 40 ng/ml IL-3. For calculation of barcode enrichment, cell samples were collected 6-8 days after transduction (starting point) and after liquid culture for a total of 42 to 62 days or after 2-6 serial replatings in methylcellulose (end point). Genomic DNA was isolated and a 300-bp vector sequence containing the barcode was amplified in 12 PCR cycles with primers barcode-F: ATCTTGTTCAATGGCCGATC and WPRE-R-300: GAGCTGACAGGTGGTGGC. The universal Illumina TruSeq adaptors were tagged in a second round of 15 cycles of PCR amplification and the PCR product was purified and sequenced in the Illumina platform as previously described(Kotini et al., 2015). Sequencing reads (75 nt) were trimmed of vector sequences for barcode extraction with a Python programming code and the representation of each barcode sequence was calculated as a percentage of total number of reads of all barcodes. Fold enrichment was calculated between paired samples and averaged across all samples for each gene (Figure S7F). Enrichment was not calculated for pairs of samples with fewer than 500 reads at any time point. The recurrence score (Figure S7G) was calculated as the number of independent experiments a gene was found to be enriched (fold change >1.1).

Statistical analysis

Statistical analysis was performed with GraphPad Prism software. Data are shown as the mean with standard error of the mean (SEM). Pairwise comparisons between different groups were performed using a two-sided unpaired unequal variance t-test. For all analyses, $P < 0.05$ was considered statistically significant. Investigators were not blinded to the different groups.

Supplemental References

Dvinge, H., Ries, R.E., Ilagan, J.O., Stirewalt, D.L., Meshinchi, S., and Bradley, R.K. (2014). Sample processing obscures cancer-specific alterations in leukemic transcriptomes. Proceedings of the National Academy of Sciences of the United States of America *111*, 16802-16807.

Flicek, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S.*, et al.* (2013). Ensembl 2013. Nucleic acids research *41*, D48-55.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T.*, et al.* (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nature methods *12*, 115-121.

Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P., and Bradley, R.K. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res *25*, 14-26.

Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature methods *7*, 1009-1015.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B.*, et al.* (2013). The UCSC Genome Browser database: extensions and updates 2013. Nucleic acids research *41*, D64-69.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol *11*, R25.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Wagenmakers, E.J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). Bayesian hypothesis testing for psychologists: a tutorial on the Savage-Dickey method. Cognitive psychology *60*, 158-189.

Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome biology *11*, R14.