Stem Cell Reports
Article

ISSCR

OPEN ACCESS

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

Chan-Jung Chang,[1,2,3,4] Andriana G. Kotini,[1,2,3,4] Malgorzata Olszewska,[1,2,3,4] Maria Georgomanoli,[1,2,3,4] Julie Teruya-Feldstein,[5] Henrik Sperber,[1,2,3,4] Roberto Sanchez,[6] Robert DeVita,[6] Timothy J. Martins,[7] Omar Abdel-Wahab,[8,9] Robert K. Bradley,[10,11] and Eirini P. Papapetrou[1,2,3,4,*]

[1]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1044A, New York, NY 10029, USA
[2]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[3]Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[4]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[5]Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[6]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[7]Quellos High Throughput Screening Core, Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA 98109, USA
[8]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[9]Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[10]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[11]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
*Correspondence: eirini.papapetrou@mssm.edu
https://doi.org/10.1016/j.stemcr.2018.03.020

SUMMARY

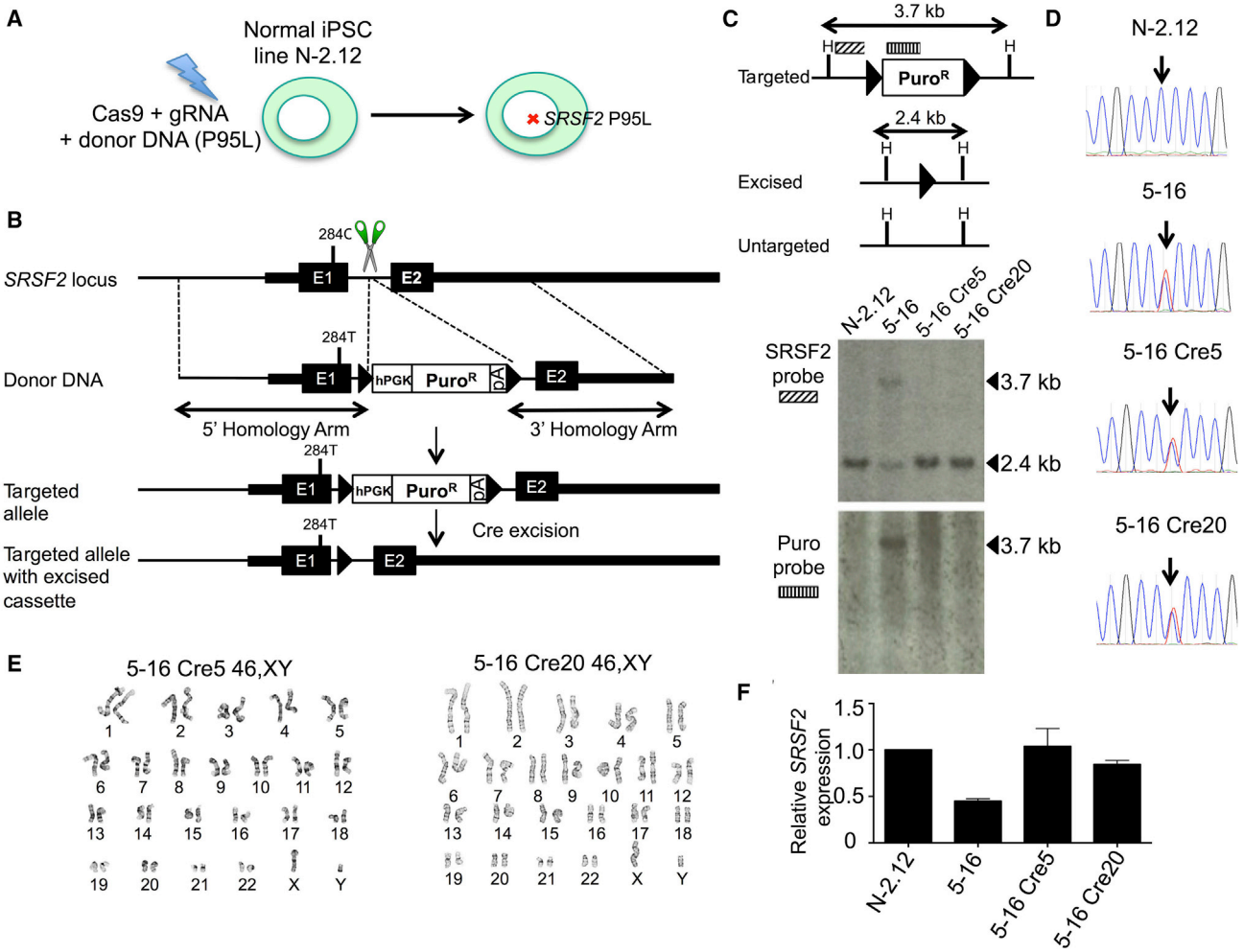
Connecting specific cancer genotypes with phenotypes and drug responses constitutes the central premise of precision oncology but is hindered by the genetic complexity and heterogeneity of primary cancer cells. Here, we use patient-derived induced pluripotent stem cells (iPSCs) and CRISPR/Cas9 genome editing to dissect the individual contributions of two recurrent genetic lesions, the splicing factor *SRSF2* P95L mutation and the chromosome 7q deletion, to the development of myeloid malignancy. Using a comprehensive panel of isogenic iPSCs—with none, one, or both genetic lesions—we characterize their relative phenotypic contributions and identify drug sensitivities specific to each one through a candidate drug approach and an unbiased large-scale small-molecule screen. To facilitate drug testing and discovery, we also derive *SRSF2*-mutant and isogenic normal expandable hematopoietic progenitor cells. We thus describe here an approach to dissect the individual effects of two cooperating mutations to clinically relevant features of malignant diseases.

INTRODUCTION

Precision oncology aims to match patients with drugs that target the specific genetic makeup of their malignant cells to enhance their chances of response. Pairing specific mutations with drug responses is central to this goal, but hindered by the immense genetic diversity and complexity of both hematologic malignancies and solid tumors. The vast majority of human malignancies contain more than one recurrent genetic lesion that cooperatively confer their phenotypic characteristics. In addition to associations of specific mutations to drug responses, associations of mutations to cellular phenotypes would also be valuable to help determine the most useful functional assays to guide drug testing. Furthermore, understanding the relative contribution of each genetic lesion to the malignant phenotype can aid the understanding of mechanisms of oncogenesis and the importance of the order of mutation acquisition.

Current experimental methods afford limited opportunities for drawing genotype-to-phenotype and genotype-to-drug response associations. Very few associations of gene mutations to drug responses can be made directly in the clinic due to the complexity and heterogeneity of cancer genomes, the relative rarity of some cancers and genotypes and the constrains imposed by currently administered therapies. Patient-derived xenograft models have been used to predict therapeutic responses but variable clonal dynamics and genetic drifts hamper systematic connections of specific mutations with drug responses (Gao et al., 2015; Papapetrou, 2016). Genetically engineered mouse models offer precise genetic models amenable to studies of mutational cooperation, but can be limited by species-determined differences in disease phenotypes, downstream mechanisms, and drug specificities (Day et al., 2015; Gould et al., 2015; Kronke et al., 2015).

Myelodysplastic syndrome (MDS) genomes typically harbor two or more genetic lesions, including gene mutations and larger-scale genetic abnormalities, mainly deletions of the long arm of chromosomes 5, 7, and 20 (Papaemmanuil et al., 2013). Mutations in genes encoding splicing factors have recently been discovered as the most frequent class of mutations in MDS (Graubert et al., 2012; Papaemmanuil et al., 2011; Yoshida et al., 2011). Serine/arginine-rich splicing factor 2 (*SRSF2*) is a member of the serine/arginine-rich (SR) protein family that contributes to both constitutive and alternative splicing by binding to RNA sequence motifs that are enriched in exons, called exonic splicing enhancer (ESE) sequences. Mutations of *SRSF2*



(legend on next page)

are found in 20%–30% of MDS patients and, less frequently, in other hematologic malignancies and solid tumors and are almost always heterozygous missense substitutions at codon P95 (P95 L/R/H) (Dvinge et al., 2016; Papaemmanuil et al., 2013; Yoshida et al., 2011). Somatic loss of one copy of the long arm of chromosome 7 (del(7q)) is a characteristic cytogenetic abnormality in MDS and other myeloid malignancies, associated with unfavorable prognosis and can co-occur with the *SRSF2* P95 mutation in patients with MDS and acute myeloid leukemia (AML) (Papaemmanuil et al., 2013, 2016).

Here we combined patient-derived induced pluripotent stem cells (iPSCs) with the CRISPR/Cas9 system to interrogate the contributions of the *SRSF2* P95 mutation and of the del(7q) to cellular phenotype and drug responses. We find that the *SRSF2* P95 mutation confers dysplastic morphology and other phenotypic characteristics to iPSC-derived hematopoietic progenitor cells (iPSC-HPCs) in support of a role early in the transformation process, while del(7q)-iPSC-HPCs exhibit a more severe differentiation block, concomitant with disease progression—findings consistent with clinical observations and population genetics analyses. We show that SRSF2 mutant iPSC-HPCs are preferentially sensitive to splicing modulator drugs and identify candidate compounds preferentially targeting del(7q) cells through an unbiased large-scale small-molecule screen. To facilitate drug testing and screening, we report the derivation of iPSC-derived expandable HPCs (eHPCs) that can be grown like conventional cell lines while maintaining specific drug sensitivities. These results demonstrate the power of patient-derived iPSCs and genome editing in dissecting the individual contributions of cooperating genetic lesions to clinically relevant cancer features.

RESULTS

Introduction of the *SRSF2* P95L Mutation in Normal Patient-Derived iPSCs

We previously derived normal and MDS iPSC lines from a patient with MDS harboring *SRSF2 P95L* mutation and del(7q) (Kotini et al., 2015, 2017). The MDS-2.13 line was derived from the MDS clone of this patient and harbors the *SRSF2 P95L* mutation and a deletion of chr(7q), and contains no additional mutations recurrently found in myeloid malignancies, as determined by whole-exome sequencing of the iPSC line and of the starting patient cells (Kotini et al., 2015). The N-2.12 line originated from normal bone marrow (BM) hematopoietic cells of the same patient, as it was not found to share any common somatic variants with the patient's MDS clone by whole-exome sequencing (Kotini et al., 2015). To study the effects of the *SRSF2* P95L mutation in isolation, we first introduced

Figure 1. Introduction of the *SRSF2* P95L Mutation in Normal iPSCs and Correction of the *SRSF2* P95L Mutation in MDS Patient-Derived iPSCs

(A and B) Schematic representation of the gene-editing strategy for the generation of *SRSF2* P95L iPSCs. Depicted from top to bottom in (B): the *SRSF2* locus with the position of the 284C > T mutation and the position targeted by the gRNAs (scissors) shown; the donor DNA with 5′ and 3′ homology arms; scheme of the targeted allele with the 284T mutation introduced and the selection cassette integrated; scheme of the targeted allele following Cre recombinase-mediated excision of the floxed selection cassette. Triangles depict *loxP* sites. E1, exon 1; E2, exon 2; hPGK, human phosphoglycerate kinase promoter; $Puro^R$, puromycin resistance gene; pA, poly-adenylation signal.

(C) Southern blot analysis of CRISPR/Cas9-edited clones before and after targeting and excision of the selection cassette with probes shown in the upper panel. H depicts HindIII restriction sites.

(D) Sanger sequencing of the parental line, the targeted clone 5-16 and two excised derivative clones 5-16 Cre5 and 5-16 Cre20, harboring heterozygous 284T *SRSF2* mutation. The arrow shows the position of the 284C (WT) or 284T (mutant) nucleotide.

(E) Karyotypic analysis of the two targeted and excised clones showing normal male karyotype.

(F) Total *SRSF2* expression by qRT-PCR, showing reduced expression by approximately half in the intermediate targeted clone 5-16, relative to the parental N-2.12 line, presumably because the pA in the selection cassette, shown in (B), results in premature transcription termination and inactivation of the targeted allele. Expression of total *SRSF2* is restored to normal levels in the two clones (5-16 Cre5 and 5-16 Cre20) after excision of the cassette. Error bars represent standard deviation.

(G) Schematic representation of the gene-editing strategy for the correction of the *SRSF2* P95L iPSC line MDS-2.13, harboring a chr7q deletion in addition to the *SRSF2* P95L mutation.

(H) Identification of monoallelically targeted clones with primers shown in the right panel. The 760 bp band (lower) corresponds to the untargeted allele. The 2,105 bp band (upper) corresponds to the targeted allele.

(I) Sanger sequencing of the parental *SRSF2* mutant MDS-2.13 line and five targeted clones found positive by PCR screening for targeted integration of the donor, as shown in (H). The arrow shows the position of the 284C (WT) or 284T (mutant) nucleotide. Clones 8, 18, and 24 are targeted in the mutant allele and therefore corrected (homozygous for 284C), whereas clones 7 and 21 are targeted in the WT allele and remain mutant (heterozygous 284C/T).

(J) Sanger sequencing of the PCR band corresponding to the untargeted allele, shown in (H), to test for indels generated through non-homologous end joining. Clone 8 harbors a 19-nt deletion, clone 18 a 1-nt (G) insertion, whereas clone 24 has an intact untargeted allele and was therefore selected for further experiments.

the mutation into the iPSC line N-2.12 (Figure 1A) (Kotini et al., 2015). We designed four guide RNAs (gRNAs) targeting the first intron of the *SRSF2* gene and a donor plasmid containing a selection cassette (Figure 1B). We selected two gRNAs, which we co-transfected with the donor DNA (Figures S1A–S1C). Cells with targeted integration (TI) of the donor DNA were detected by PCR, but no puromycin-resistant colonies could be retrieved, presumably because expression of the puromycin resistance gene from the *SRSF2* locus was not sufficient for successful selection. We therefore attempted to obtain targeted clones by first selecting pools of transfected cells enriched for targeting events and subsequent screening of single-cell clones (Figure S1D). TI of the donor could be detected in all 48 pools of approximately 20,000 transfected cells. Two pools (no. 2 and no. 5) with the strongest signal were selected. Two out of 48 and 4 out of 48 targeted clones were found after single-cell subcloning of the two pools, respectively (Figures S1E–S1G). These six clones were tested with a second set of TI-specific primers, DNA sequencing of the introduced 284C > T mutation, as well as detection and sequencing of the untargeted allele (Figures S1H, S1I, and S2A–S2C). All six clones contained indels in the untargeted allele, which were restricted to intronic sequences (Figure S2C). Out of 4 clones with confirmed TI of the intact donor (Figure S1H), clone 5-16, harboring the smallest indel, a deletion of 16 nt at a distance of 125 and 193 bp from the splice donor and splice acceptor sites, respectively, was selected and confirmed to maintain a normal karyotype (Figures S2C and S2D). Two subclones, 5-16Cre5 and 5-16Cre20, were selected after Cre-mediated excision (Figures S2E and S2F)(Papapetrou et al., 2011; Papapetrou and Sadelain, 2011). Off-target insertions of the donor cassette were excluded and excision of the selection cassette was confirmed by Southern blotting (Figure 1C). The two clones were confirmed to harbor the 284C > T mutation, to maintain a normal karyotype and to express SRSF2 at levels similar to the parental line (Figures 1D–1F).

Correction of the *SRSF2* P95L Mutation in MDS Patient-Derived iPSCs

To further probe the contribution of the *SRSF2* P95L mutation in MDS with cooperating genetic events, we corrected this mutation in the iPSC line MDS-2.13, derived from the MDS clone of the same patient from whom the N-2.12 line was derived and harboring a del(7q) in addition to the *SRSF2* P95L mutation (Figure 1G) (Kotini et al., 2015, 2017). We used a strategy similar to the one used for introducing the mutation, this time using only gRNA-1, in order to minimize indels in the untargeted allele, a donor plasmid carrying a normal *SRSF2* 284C allele and a modified selection strategy (Figures S3A and S3B). Ten out of 46 clones were found to have TI of the donor cassette in one allele (Figure 1H). In 5 out of 10 the donor was integrated in the mutant allele reverting it to wild-type (WT) (Figure 1I). Clone 24 (C24), found to have an intact untargeted allele, was selected (Figure 1J). Two excised clones (C24 Cre6 and C24 Cre7) were confirmed to be homozygous for the WT *SRSF2* 284C allele (Figures S3C and S3D).

The *SRSF2* P95L Mutation Imparts Splicing Alterations

To assess gene expression and splicing in *SRSF2* mutant iPSCs, we compared the two CRISPR/Cas9-engineered *SRSF2* P95L mutant clones 5-16Cre 5 and 5-16Cre20 with the parental normal N-2.12 line by RNA sequencing of undifferentiated iPSCs, $CD34^+$ cells isolated on day 9 of differentiation and $CD45^+$ cells isolated on day 12 of differentiation (when almost all $CD45^+$ cells are also $CD34^+$) (Figure S4). The SRSF2 mutant iPSC expressed SRSF2 at levels similar to the normal N-2.12 line and expressed the mutant allele at approximately 50% of total levels (Figures S5A and S5B). Global gene expression changes were found that varied with cell type (i.e., undifferentiated pluripotent stem cells, $CD34^+$ hemogenic endothelium, and $CD45^+$ hematopoietic progenitors) (Figures S5C and S5D). Differentially expressed genes included predominantly cell-type-specific genes (Figure S5E). Cell proliferation, differentiation, apoptosis, and regulation of hematopoiesis were among the predominant gene ontology categories (Figure S5F).

SRSF2 mutant cells exhibited genome-wide alterations in ESE motif recognition in all three cell states. C-rich variants of the SSNG (S = C or G) motif recognized by WT SRSF2 were enriched within exons promoted by *SRSF2* P95L, while G-rich variants were enriched within exons repressed by *SRSF2* P95L (Figures 2A and 2B). These findings are consistent with findings previously reported by us and others in a knockin mouse model and in K562 cells expressing mutant SRSF2 ectopically or through gene editing (Kim et al., 2015; Zhang et al., 2015). Cell-type-specific differences in differentially spliced genes were less prominent than those found in differentially expressed genes (Figure 2C). Mis-spliced genes included genes of potential disease relevance with isoforms previously reported to be promoted by mutant *SRSF2*, including *EZH2* and *FYN* (Figures 2D and 2E) (Kim et al., 2015; Zhang et al., 2015). Importantly, iPSC-derived $CD34^+$ and $CD45^+$ hematopoietic progenitors recapitulated a higher fraction of the mis-spliced events that were observed in patient cells than either the knockin mouse or the K562 *SRSF2* mutant models (Figures 2F and 2G), suggesting that iPSCs capture disease-relevant splicing alterations more faithfully than other existing models.

Hematopoietic Phenotypes Exclusively Produced by either *SRSF2* P95L or del(7q)

To connect each of the two genetic lesions—*SRSF2* P95L mutation and chr7q deletion—with specific hematopoietic

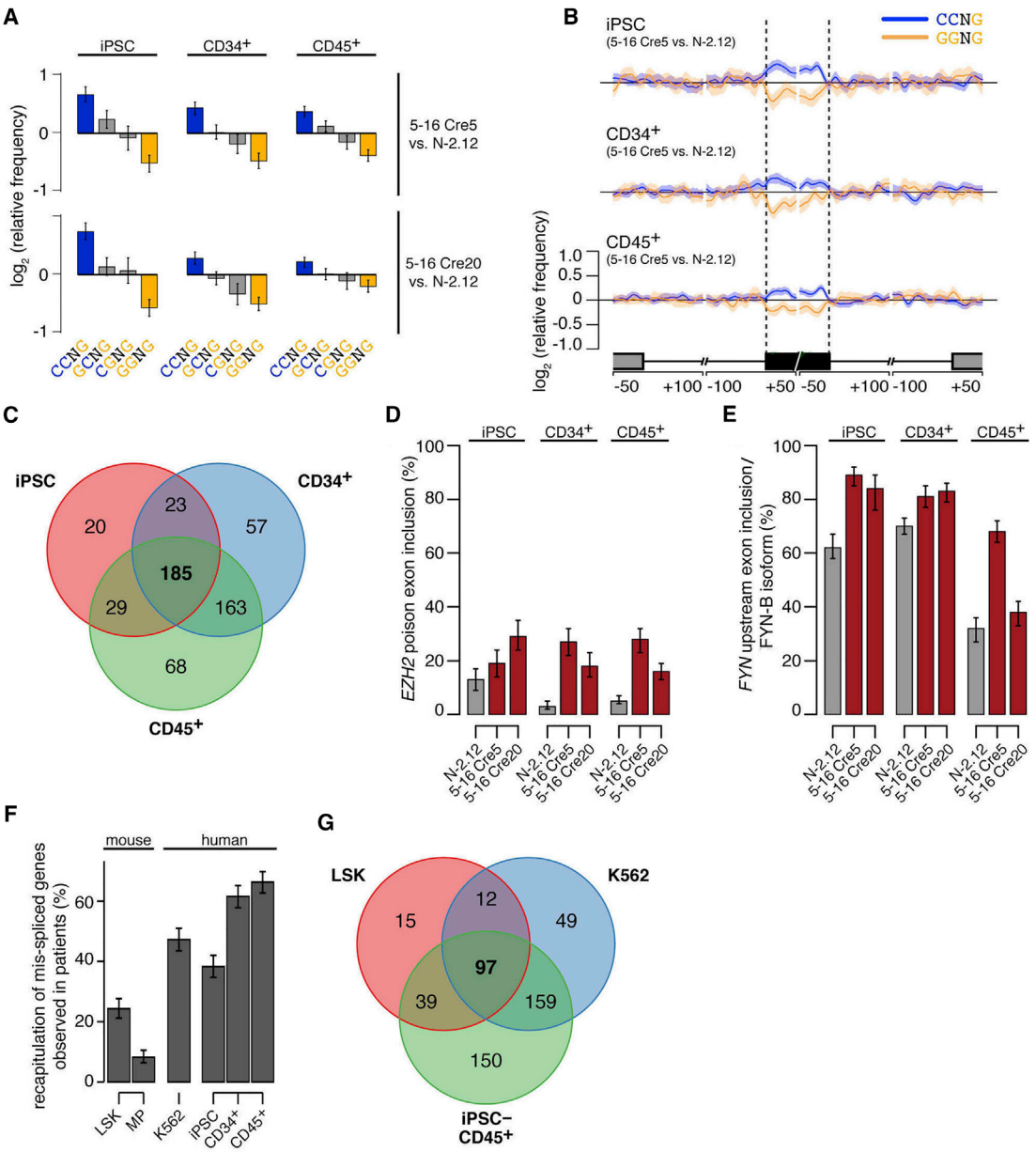


Figure 2. Splicing Alterations in *SRSF2* Mutant iPSCs and iPSC-HPCs

(A) Mean enrichment of all variants of the SSNG (S = C or G) motif in cassette exons that are promoted versus repressed in *SRSF2* mutant 5-16 Cre5 cells (upper panels) or 5-16 Cre20 cells (lower panels) compared with the isogenic *SRSF2* WT N-2.12 cells. Error bars indicate 95% confidence intervals estimated by bootstrapping.

(B) Relative occurrence of CCNG and GGNG motifs within and adjacent to cassette exons that are differentially spliced in association with the *SRSF2* mutation in undifferentiated iPSCs (top), $CD34^+$ cells (middle), and $CD45^+$ cells (lower). Shading indicates 95% confidence interval by bootstrapping. The schematic illustrates a portion of a metagene centered on the differentially spliced cassette exon. From left to right, the features are the upstream exon (gray box) and intron (black line), the cassette exon (black box, vertical dashed lines), and the downstream intron (black line) and exon (gray box). Horizontal axis, genomic coordinates defined with respect to the 5′ and 3′ splice sites where 0 is the splice site itself. Vertical axis, relative frequency of the indicated motifs over genomic loci containing cassette exons promoted versus repressed by *SRSF2* mutations (log scale).

(C) Venn diagram depicting the overlap between genes that are differentially spliced in *SRSF2* mutant (in either 5-16 Cre5 or 5-16 Cre20) versus *SRSF2* WT (N-2.12) cells at the undifferentiated (iPSC) state, $CD34^+$ or $CD45^+$ state.

(legend continued on next page)

phenotypes, we assembled a comprehensive panel of isogenic iPSC lines with both, one or none of the two genetic lesions (Figure 3A; Table S1). The patient-derived lines N-2.12 and MDS-2.13, as well as the lines derived from them through CRISPR/Cas9 (5-16Cre5, 5-16Cre20, and C24Cre6), were described above. We also analyzed a line, MDS-2.A3C, derived from MDS-2.13 after spontaneous correction of the del(7q), as well as two lines, Cre10 and 8Cre21, derived from N-2.12 by engineered del(7q) through a previously described inverted Cre-loxP genetic engineering strategy (Kotini et al., 2015). This panel was subjected to hematopoietic differentiation and detailed phenotypic characterization (Figure 3B). Both normal and *SRSF2* P95L mutant lines without del(7q) gave rise to comparable percentages of $CD45^+$ HPCs (80%–90% on day 14 of differentiation) (Figures 3C and 3D). In contrast, all del(7q) lines generated very few $CD45^+$ cells (<10%), and these levels remained unchanged even after correction of the *SRSF2* P95L mutation (Figures 3C and 3D, compare MDS-2.13 with C24 Cre6). These results suggest that the *SRSF2* P95L mutation does not affect the numbers of $CD45^+$ myeloid progenitors. Hematopoietic progenitors and mature cells from *SRSF2* P95L mutant iPSCs displayed dysplastic morphological features, whereas cells with isolated del(7q) showed no apparent morphological changes of dysplasia (Figure 3E; Table S2).

Hematopoietic Phenotypes Cooperatively Produced by *SRSF2 P95L* and del(7q)

We previously identified reduced growth and viability as a disease-relevant phenotype of MDS-iPSC-derived HPCs (Kotini et al., 2015, 2017). HPCs derived from both *SRSF2* P95L mutant and del(7q)- iPSCs grew at a slower rate and exhibited increased cell death, compared with the normal isogenic cells (Figures 4A and 4B). Viability was even lower in HPCs derived from iPSCs harboring both the *SRSF2* P95L mutation and the chr7q deletion. This defect was partially restored by correcting either the *SRSF2* P95L mutation (C24 Cre6 line) or the chr7q deletion (MDS-2.A3C line), albeit without reaching normal levels (Figures 4A and 4B). *SRSF2*-mutant HPCs consistently gave rise to lower numbers of methylcellulose colonies than normal cells, whereas del(7q) cells gave rise to very few if any colonies (Figure 4C). The patient-derived line MDS-2.13 did not give rise to any colonies before or after correction of the *SRSF2* mutation, whereas correction of the chr7q deletion (MDS-2.A3C line) partially restored colony formation, albeit not to normal levels (Figure 4C). These results suggest that the impaired viability and clonogenicity phenotypes are cooperatively generated by both the *SRSF2* P95L mutation and the chr7q deletion.

Taken together, our phenotypic analysis of an expansive panel of isogenic iPSCs allowed us to determine phenotypes that are specific to each isolated genetic lesion—morphologic dysplasia to the *SRSF2* P95L mutation and loss of myeloid progenitors to the del(7q)—as well as phenotypes that are cooperatively produced by both lesions—impaired clonogenicity, viability, and growth (Figures 4D and 4E).

The *SRSF2* P95L Mutation Confers Selective Susceptibility to Splicing Inhibitors

We recently reported that *SRSF2*-mutant primary human leukemias preferentially respond to treatment with the splicing inhibitor E7107 (Lee et al., 2016). To precisely assess the selectivity of this growth inhibitory effect for the *SRSF2* P95L mutation, we tested the growth of iPSC-derived HPCs with isolated *SRSF2* mutation (5-16Cre20) and isolated chr7q deletion (Cre10) in the presence of the E7107 drug (Figure 5A). These experiments showed selective growth inhibition of *SRSF2* mutant, but not of isogenic normal or of cells with isolated del(7q) (Figure 5A). We then sought to test other classes of splicing-modulating drugs for selectivity against *SRSF2*-mutant cells. Cpd-1, Cpd-2, and Cpd-3 are small-molecule inhibitors of CDC-like kinases and serine-arginine protein kinases, which modulate splicing (Araki et al., 2015). All three compounds could inhibit the *SRSF2*-mutant line, whereas they had no effect in the growth of the isogenic normal or del(7q)

(D) Percentage of *EZH2* transcripts harboring the poison cassette exon in the indicated cell lines and cell types, as measured by RNA sequencing (RNA-seq) data. Error bars indicate 95% confidence intervals. *SRSF2* mutant cells exhibit preferential inclusion of a "poison" cassette exon of *EZH2*, giving rise to an isoform of *EZH2* that contains a premature termination codon, resulting in degradation by nonsense-mediated RNA decay.

(E) Percentage of *FYN* transcripts corresponding to the FYN-B isoform in the indicated cell lines and cell types, as measured by RNA-seq data. Error bars indicate 95% confidence intervals. *SRSF2* mutant cells exhibit preferential inclusion of the upstream exon 7 (exon 7A) of *FYN*, a gene encoding a tyrosine kinase, giving rise to an isoform (FYN-B) with altered kinase activity.

(F) Fraction of mis-spliced genes observed in patients that are also mis-spliced in each indicated model system. The overlap between genes exhibiting differential splicing in association with *SRSF2* mutational status in either an AML or CMML cohort (Kim et al., 2015) and genes exhibiting differential splicing in each system was computed. All analyses were restricted to human genes with orthologous mouse genes to avoid potential biases resulting from incomplete orthology assignments.

(G) Venn diagram depicting the overlap between genes that are differentially spliced in *SRSF2* mutant versus *SRSF2* WT cells, comparing data from mice ($Lineage^-$ $Sca1^+$ $c\text{-}Kit^+$ [LSK] cells), K562 cells (Kim et al., 2015) and iPSC-derived $CD45^+$ cells.

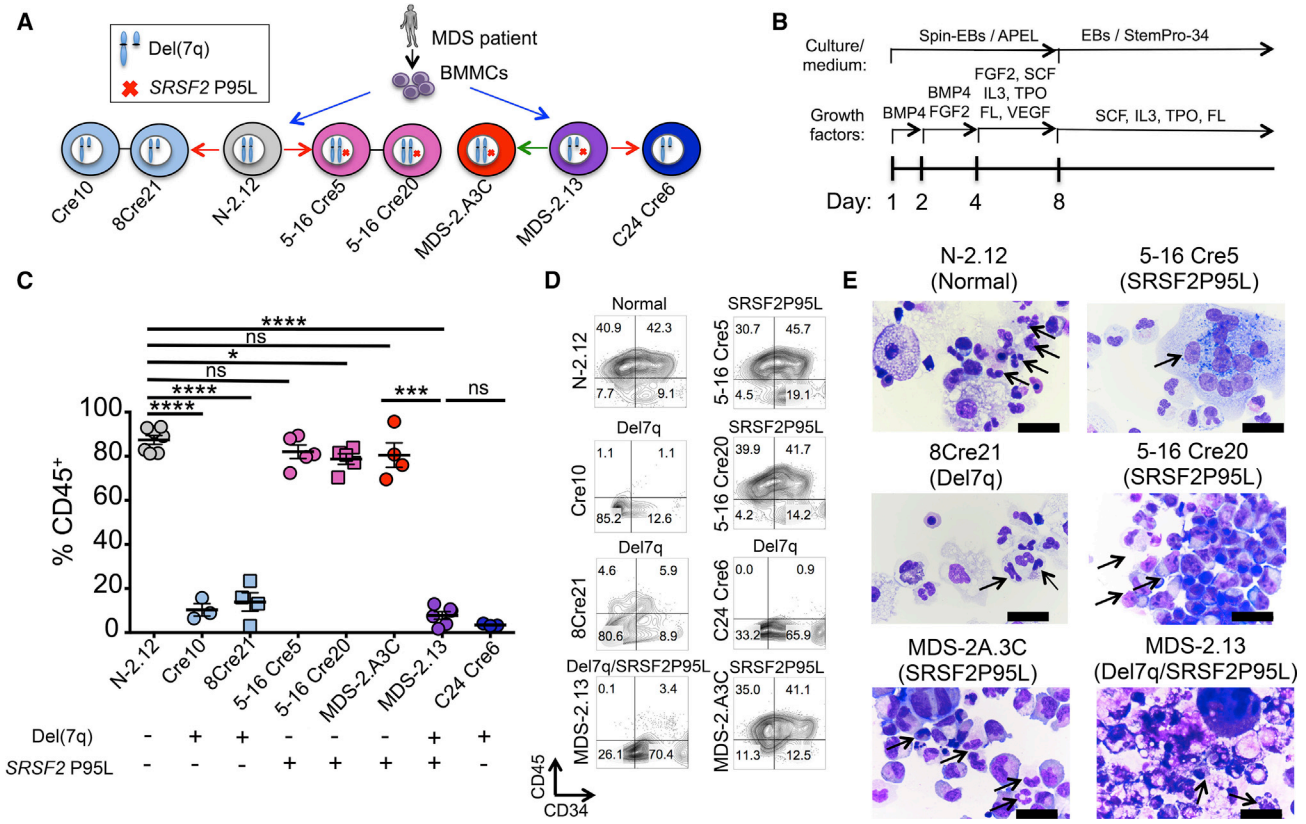


Figure 3. Hematopoietic Phenotypes Exclusively Contributed by *SRSF2* P95L or del(7q)

(A) All isogenic iPSC lines used in this study. Lines N-2.12 (WT *SRSF2*) and MDS-2.13 (mutant *SRSF2*) were derived from BM mononuclear cells (BMMCs) of an MDS patient in parallel in the same reprogramming experiment (Kotini et al., 2015, 2017). The latter also contains a chr7q deletion. Lines 5-16 Cre5 and 5-16 Cre20 were derived from N-2.12 through CRISPR/Cas9-mediated editing to introduce the *SRSF2* P95L mutation (Figure 1). C24 Cre6 was derived from MDS-2.13 through CRISPR/Cas9-mediated correction of the *SRSF2* P95L mutation (Figure 1). MDS-2.A3C was derived from MDS-2.13 through spontaneous correction of the chr7q deletion and maintains the *SRSF2* P95L mutation (Kotini et al., 2015). Cre10 and 8Cre21 were derived from N-2.12 through engineering of del(7q) with a modified Cre/*loxP* strategy (Kotini et al., 2015). The red arrows depict CRISPR/Cas9-mediated gene editing. The green arrow depicts spontaneous conversion.

(B) Hematopoietic differentiation scheme.

(C) CD45 expression on day 14 of hematopoietic differentiation. Mean and SEM from independent differentiation experiments are shown for each line. *$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$; ns, not significant.

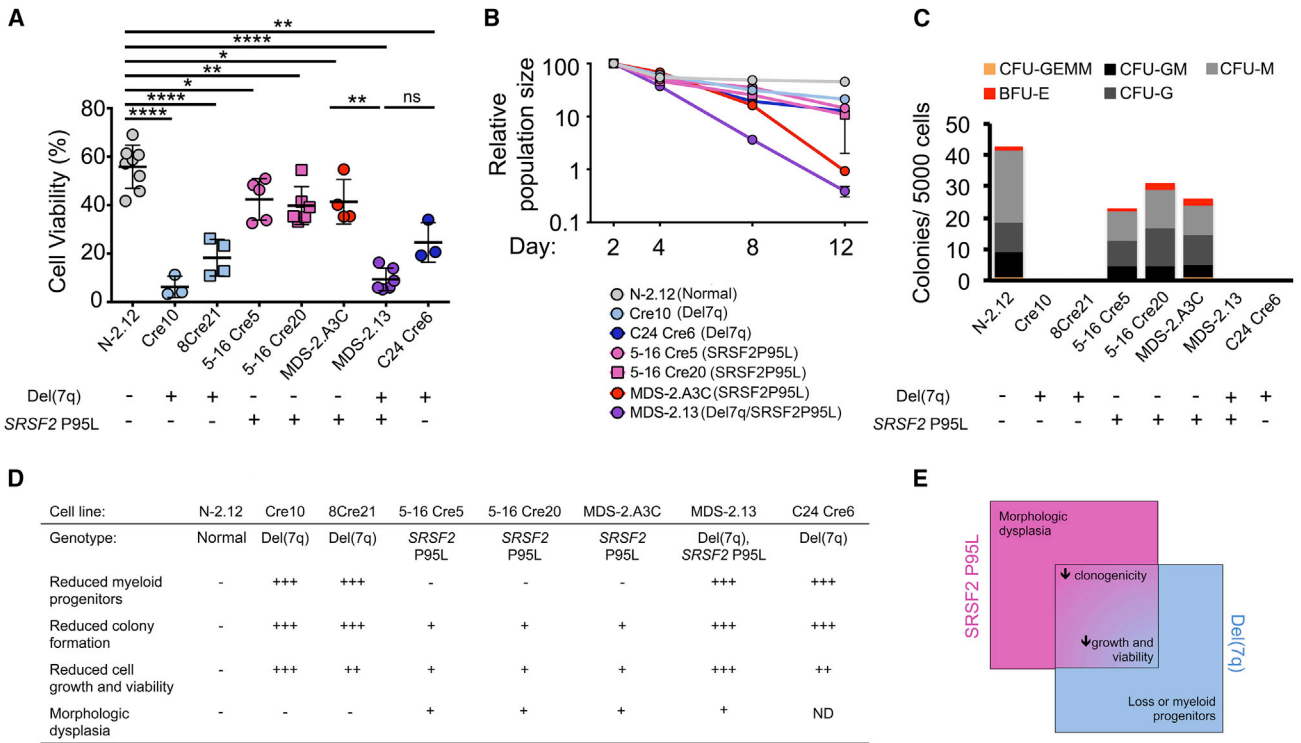
(D) CD34 and CD45 expression on day 14 of hematopoietic differentiation in representative experiments.

(E) May-Giemsa staining of cytospin preparations of hematopoietic cells derived from the indicated iPSC lines after 14 days of hematopoietic differentiation culture. Upper and middle left panels: arrows indicate normal myeloid cells. Upper right panel: arrow indicates giant multinucleated myeloid progenitor cells. Remaining panels: arrows indicate hypogranular and/or hypolobular myeloid cells. Scale bars, 10 μm; ns, not significant.

HPCs (Figure 5A). Quantification of un-spliced and spliced mRNA levels of two genes very sensitive to inhibition of splicing, *DNAJB1* and *EIF4A1* (Eskens et al., 2013), confirmed that the effect of these compounds was mediated through splicing inhibition (Figure 5B).

A Small-Molecule Screen Identifies a Hit Compound against del(7q) Cells

To identify putative new compounds with selective growth-inhibitory activity against *SRSF2*-mutant and/or del(7q) cells, we screened a library of 2,000 compounds containing Food and Drink Administration-approved drugs, natural products, and other bioactive compounds at four different concentrations using a luminescence-based viability assay in the MDS-2.13 and isogenic normal N-2.12 cells. This experimental design allowed us to take advantage of parallel screening in normal isogenic control cells to filter hits with unspecific effects in cell growth inhibition. Primary hits were defined as compounds that inhibited the growth of the MDS-iPSC line, but not of the



Cell line:	N-2.12	Cre10	8Cre21	5-16 Cre5	5-16 Cre20	MDS-2.A3C	MDS-2.13	C24 Cre6
Genotype:	Normal	Del(7q)	Del(7q)	*SRSF2* P95L	*SRSF2* P95L	*SRSF2* P95L	Del(7q), *SRSF2* P95L	Del(7q)
Reduced myeloid progenitors	-	+++	+++	-	-	-	+++	+++
Reduced colony formation	-	+++	+++	+	+	+	+++	+++
Reduced cell growth and viability	-	+++	++	+	+	+	+++	++
Morphologic dysplasia	-	-	-	+	+	+	+	ND

Figure 4. Cooperative Hematopoietic Phenotypes of *SRSF2* P95L and del(7q)

(A) Cell viability measured by DAPI staining on day 14 of hematopoietic differentiation. Mean and SEM from independent differentiation experiments are shown for each line. *p < 0.05, **p < 0.01, ****p < 0.0001; ns, not significant.

(B) Growth competition assay. The cells were mixed 1:1 with the N-2.12 line stably expressing GFP at the beginning of hematopoietic differentiation and followed for 12 days by flow cytometry. The relative population size was calculated as the percentage of GFP^- cells at each time point relative to the population size at day 2. Error bars represent SEM from one to four independent differentiation experiments.

(C) Methylcellulose assays on day 14 of hematopoietic differentiation. The number of colonies from 5,000 seeded cells is shown. CFU-GEMM, colony-forming unit-granulocyte, erythrocyte, monocyte, megakaryocyte; CFU-GM, CFU-granulocyte, monocyte; CFU-G, colony-forming unit-granulocyte; CFU-M, CFU-monocyte; BFU-E, burst-forming unit-erythrocyte.

(D) Summary of phenotypic analyses.

(E) Scheme of phenotypes contributed exclusively by each genetic lesion or cooperatively by both.

control normal iPSC line in a compound dose-dependent manner. A total of 66 compounds had a dose-dependent inhibitory effect on MDS cells. Thirty-four of these also inhibited the normal cells, whereas the remaining 32 appeared selective for the MDS cells (Figure 6A; Tables S3, S4, and S5). These were evaluated based on physical-chemical properties, propensity for promiscuous reactivity, overall medicinal chemistry viability, and drug-like properties (Arrowsmith et al., 2015; Dahlin et al., 2015; Dahlin and Walters, 2014). Three compounds—niflumic acid, a cyclooxygenase-2 inhibitor; albendazole, and fenbendazole, two broad-spectrum benzimidazole-derivative antihelminthic agents (Figure S6A)—were further tested with a competitive growth assay (Figure 6B). Only niflumic acid showed detectable selectivity against the MDS line (Figure 6C). We further validated this result in a larger set of isogenic iPSC lines following hematopoietic differentiation (Figure 6D). Niflumic acid selectively inhibited the growth of HPCs from two del(7q) iPSC clones Cre10 and 8Cre21, but not that of HPCs derived from the two SRSF2-edited clones. Moreover, it inhibited an additional del(7q) iPSC line derived from a different MDS patient (MDS-3.1) (Kotini et al., 2015, 2017), but had no effect on the growth of the del(7q)-corrected MDS-2A.3C clone.

We next treated primary cells from BM or peripheral blood of patients with MDS and secondary AML with or without chr7/7q abnormalities (Figures 6E and S6B; Table S6). Treatment with niflumic acid suppressed the growth of cells with monosomy 7 or del(7q), both in liquid and methylcellulose-based culture, while it had little effect in cells from patients without del7/7q or in cord blood $CD34^+$ cells. These results further support a selective effect of niflumic acid in suppressing the growth of del(7q)-MDS cells.

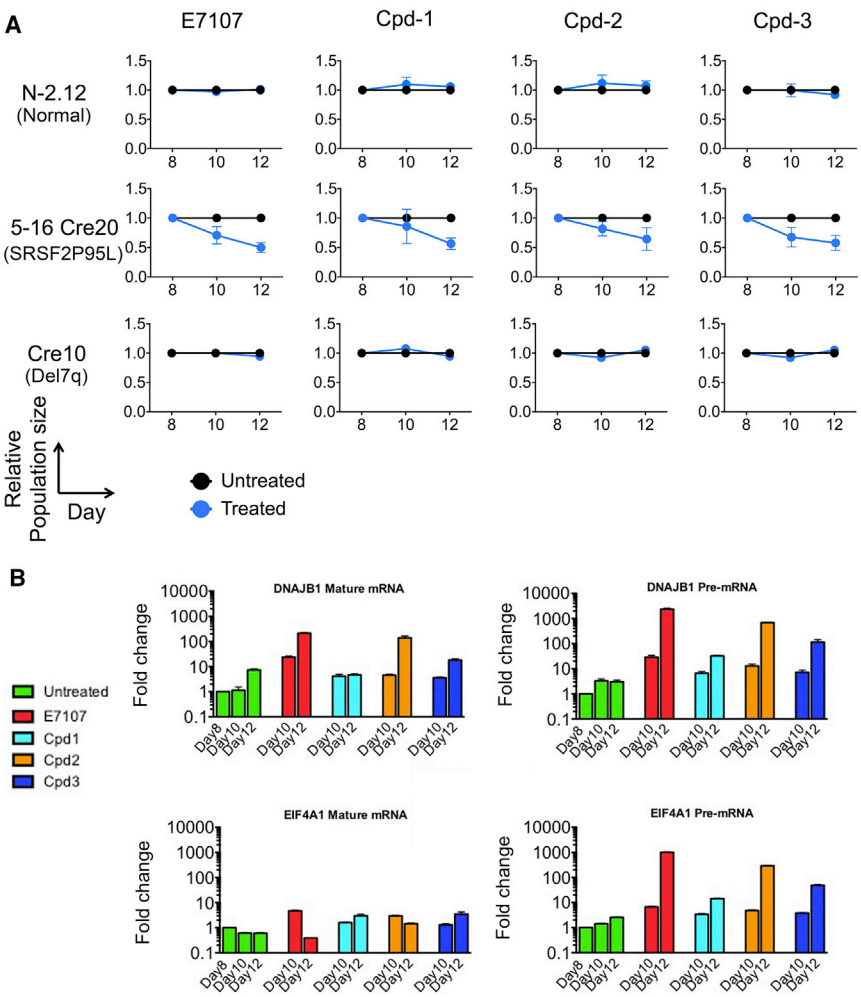


Figure 5. The *SRSF2* P95L Mutation Confers Selective Sensitivity to Splicing Inhibitor Drugs

(A) The normal N-2.12, SRSF2-mutant 5-16 Cre20 and del(7q) Cre10 lines were mixed 1:1 with a GFP-marked clone derived from the parental normal N-2.12 line to allow easy tracking of the relative rations of the two cell populations by flow cytometry. The cells were treated with E7107 0.1 nM, Cpd-1 5 μM, Cpd-2 0.5 μM, or Cpd-3 0.5 μM, as indicated, starting on day 8 of hematopoietic differentiation. The relative population size was calculated as the percentage of GFP^- cells in the treated cells relative to the percentage of GFP^- cells in the untreated cells at each time point. Error bars represent SEM from three independent experiments.

(B) Modulation of un-spliced pre-mRNA and mature mRNA of genes *DNAJB1* and *EIF4A1*, (chosen on the basis of their documented sensitivity to splicing inhibition), in hematopoietic cells differentiated from the SRSF2 mutant 5-16 Cre20 line and treated with the indicated drugs starting at day 8 of differentiation. Fold-change overexpression levels of day 8 untreated cells is shown. The un-spliced precursor pre-mRNA levels of both genes is markedly increased, whereas the spliced (mature) mRNA levels are largely unchanged, consistent with splicing inhibition by the compounds.

Generation of Expandable iPSC-Derived HPCs

A major limiting step in the use of iPSC models in drug testing is the requirement to generate the appropriate cell types in sufficient cell numbers through *in vitro* differentiation procedures, which generally require specialized expertise and are often difficult to scale up. We thus sought to derive from our iPSCs HPCs that could be expanded and maintained as conventional hematopoietic cell lines. We assembled a lentiviral library of 22 factors (22F) that may confer enhanced self-renewal of HPCs in various contexts (Table S7) (Doulatov et al., 2013; Lis et al., 2017; Pereira et al., 2013; Riddell et al., 2014; Sandler et al., 2014; Sugimura et al., 2017). Day 8 embryoid bodies (EBs) from normal and SRSF2-mutant iPSCs transduced with the library gave rise to hematopoietic cells that could be expanded in culture for at least several weeks and serially replated in methylcellulose-based media (Figures 7A–7D and S7A). In contrast, control untransduced HPCs from both normal and SRSF2-mutant iPSCs did not serially replate or expand past day 28 of hematopoietic differentiation culture (Figure 7D). The eHPCs exhibited morphology and immunophenotype of myeloid progenitors and required interleukin-3 (IL-3) for growth (Figures 7E, 7F, and S7B). They could be frozen and thawed without loss of phenotypic properties (Figures S7C and S7D).

To determine the factors essential for the maintenance of eHPCs, we performed barcode enrichment analysis in cells maintained in culture for 42–130 days or serially replated in methocult compared with the initial cells 6–8 days after transduction in 14 comparisons from 11 independent experiments (Figure S7E). Of the 22F, *SOX17*, *MYB*, *HMGA2*, *GFI1b*, *ETV6*, *FOS*, *SCL/TAL1*, *LMO2*, and *RORA* were dispensable (Figures S7F and S7G). *BMI1*, *GATA2*, *ERG*, *c-MYC*, *RUNX1c*, *BCL-XL*, *HOXA9*, *SOX4*, and *MDH2* were among the most highly enriched in the majority of independent experiments, whereas *GFI1*, *HOXB8*, *HOXB4*, and *PU1/SPI1* were also found enriched, albeit at a lower magnitude and frequency (Figures S7F and S7G). We therefore transduced these 9 factors (*BMI1*, *GATA2*, *ERG*, *c-MYC*, *RUNX1c*, *BCL-XL*, *HOXA9*, *SOX4*, and *MDH2*) with and without the

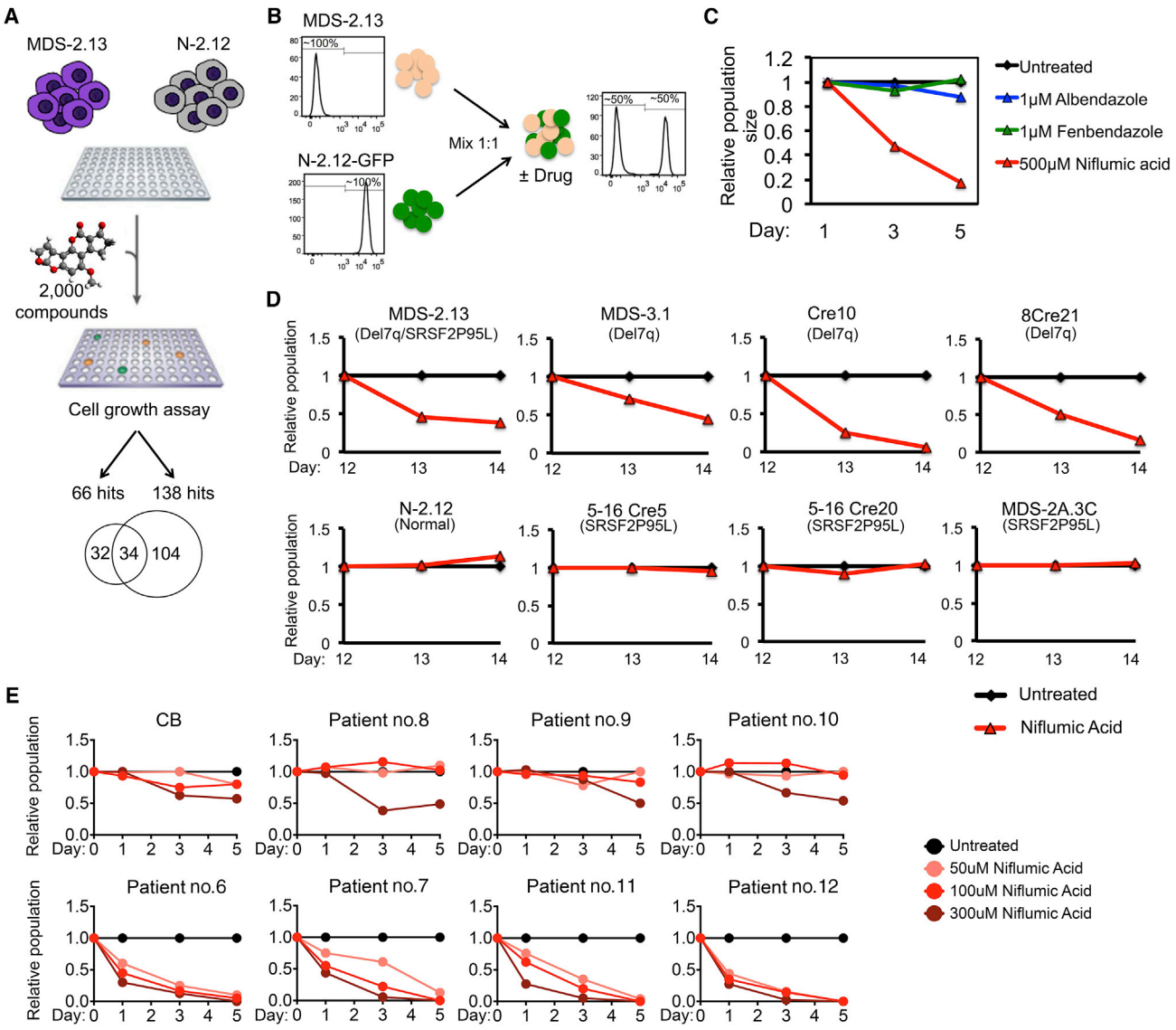


Figure 6. Identification and Validation of Hit Compounds Selective against del(7q) Cells

(A) General scheme of the small-molecule screening strategy, yielding 138 and 66 hits for the N-2.12 and the MDS-2.13 lines, respectively. Thirty-four of these were shared by the two lines, whereas 32 were specific for the MDS-2.13 cells.

(B and C) Competitive growth assay for hit validation. MDS-2.13 iPSCs were mixed at an equal ratio with the normal GFP-marked N-2.12 line to allow easy tracking of the ratio of the MDS to the normal cells. The cells were treated with the indicated drugs at the indicated concentrations (C). The relative population size was calculated as the percentage of GFP^- cells in the treated cells relative to the percentage of GFP^- cells in the untreated cells at each time point.

(D) Competitive growth assay in HPCs derived from the indicated iPSC lines. The cells were differentiated to hematopoietic cells for 12 days, mixed at an equal ratio with the normal GFP-marked N-2.12 line, differentiated in parallel, and treated with niflumic acid at 100–500 μM for 2 days. The relative population size was calculated as the percentage of GFP^- cells in the treated cells relative to the percentage of GFP^- cells in the untreated cells at each time point.

(E) Treatment of primary cells from patients with MDS and secondary AML 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) with niflumic acid. The cells were thawed on day −2 and treated with niflumic acid on days 0, 1, and 2. The relative population was calculated from cell numbers counted on a hemocytometer on days 1, 3, and 5, relative to the DMSO-treated (untreated) cells.

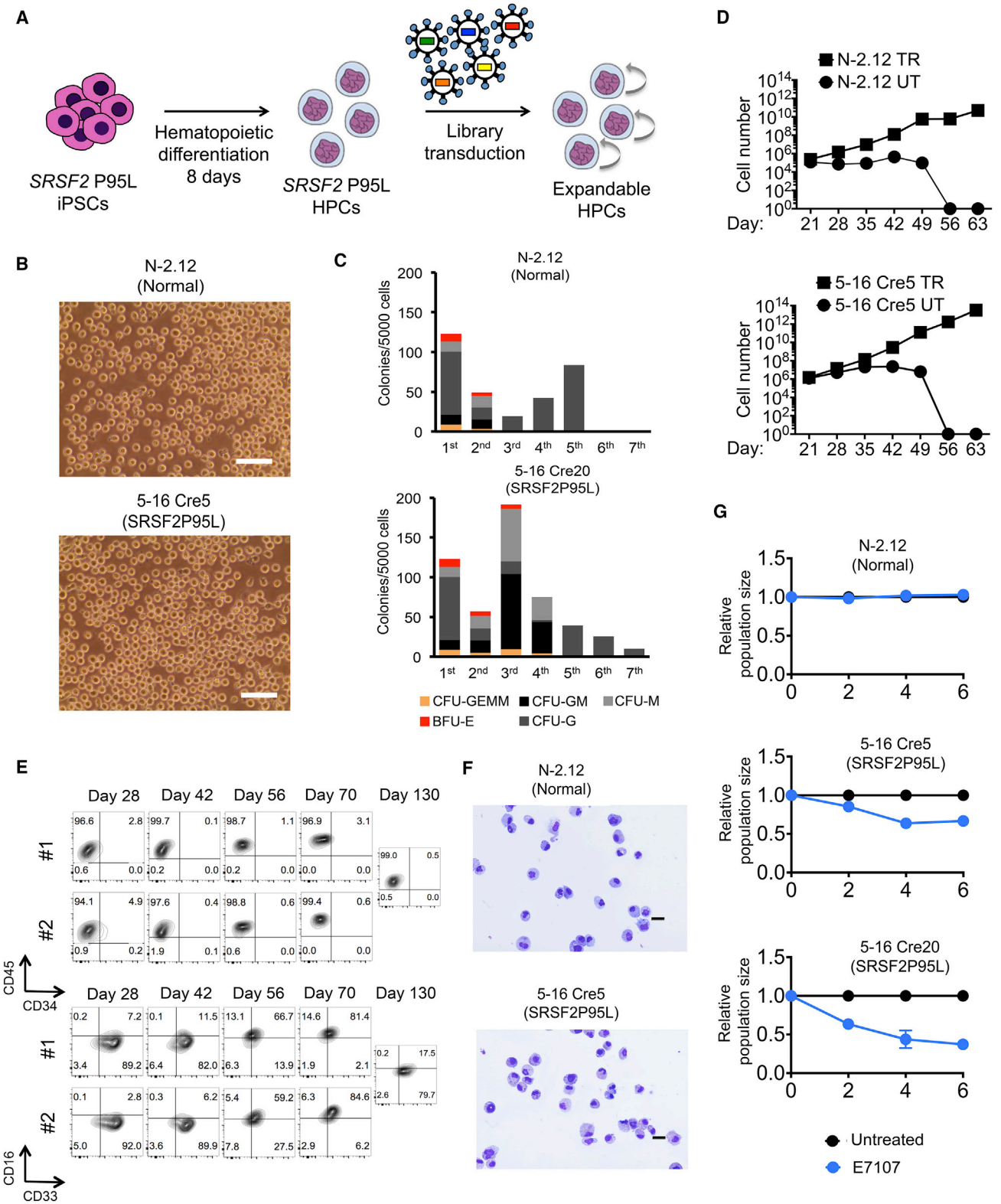


Figure 7. Establishment of Expandable *SRSF2* P95L and Isogenic Normal HPCs

(A) Experimental scheme. Normal and SRSF2-mutant iPSCs were transduced with the lentiviral library (Table S7) at day 8 of hematopoietic differentiation to generate expandable HPCs (eHPCs).

(legend continued on next page)

additional 4 factors ([9+4F], *BMI1*, *GATA2*, *ERG*, *c-MYC*, *RUNX1c*, *BCL-XL*, *HOXA9*, *SOX4*, *MDH2*, *GFI1*, *HOXB8*, *HOXB4*, and *PU1/SPI1*) in SRSF2-mutant HPCs. The 9+4F combination, but not the 9 factors alone, generated IL-3-dependent eHPCs that could be expanded for more than 100 days (Figures S7H and S7I). Notably, the 9+4F eHPCs were capable of higher expansion than the 22F cells in two independent experiments and upregulated CD34 expression (Figure S7J). Importantly, the selective inhibition of SRSF2-mutant, but not normal cells, by the splicing inhibitor E7107 was maintained in eHPCs (Figure 7G). These results demonstrate that expandable tissue-specific progenitor cell lines can be derived from isogenic mutant and normal iPSCs and used to model therapeutic responses.

DISCUSSION

Here, we used somatic cell reprogramming and genome editing to study the individual contributions of two cooperating genetic lesions to cellular phenotypes and therapeutic vulnerabilities. *SRSF2*-mutant iPSCs recapitulated several hallmarks of *SRSF2*-mutant MDS, including impaired hematopoiesis, splicing defects, and sensitivity to splicing inhibition. The *SRSF2* mutation in isolation conferred relatively mild perturbation of hematopoiesis, compared with the profound phenotypic impairment mediated by the chr7q deletion (Kotini et al., 2015, 2017). This is consistent with the *SRSF2* mutation being an early, potentially initiating, event, supported by its frequent presence in the dominant clone and in individuals with clonal hematopoiesis of indeterminate potential (CHIP) (Genovese et al., 2014; Jaiswal et al., 2014; Papaemmanuil et al., 2013; Xie et al., 2014). Furthermore, the morphologic findings in our model support a role for the *SRSF2* mutation in driving overt dysplasia, which might explain its lower relative frequency in CHIP compared with MDS (Malcovati et al., 2017; Sperling et al., 2017). In contrast, del(7q) resulted in a profound differentiation block in our model, consistent with a role in advanced disease and its typical association with progression to high-risk disease.

By comparing *SRSF2* WT and mutant cells at different differentiation stages, we could find cell-type-specific gene expression and splicing changes. Although these may be simply due to diverse sets of genes being expressed at different differentiation stages, cell-type-specific expression of other splicing factors or global changes in splicing efficiency may also contribute to these stage-specific effects (Pimentel et al., 2016; Wong et al., 2013), and may conceivably modify disease phenotypes depending on patient BM cell composition. Our model captured a greater percentage of the genes characterized as mis-spliced in patients than previous models (Figure 2F). This suggests that the iPSC lines may be a more faithful model of mutant SRSF2, particularly for molecular studies into disease mechanisms. These results are consistent with the observation that, while splicing regulation itself is highly conserved between human and mouse, most specific isoforms are not (Barbosa-Morais et al., 2012; Merkin et al., 2012; Yeo et al., 2005). Therefore, disease-relevant targets of mutant SRSF2 may be missed in the mouse. In contrast to K562 or other immortalized cell lines, iPSCs enable the study of the *SRSF2* P95L mutation in the context of a normal diploid genome both in isolation and together with cooperating mutations. All isogenic iPSC lines generated here harbor two *SRSF2* alleles, maintaining correct stoichiometry of mutant to WT protein. In contrast, K562 cells typically contain three copies of the *SRSF2* gene (Zhang et al., 2015). This is likely crucial in view of recent evidence on the importance of the relative expression levels of mutant and WT splicing factor genes (Fei et al., 2016; Lee et al., 2016). Furthermore, while SRSF2 mutations are early, potentially initiating events in MDS and are also found in patients with CHIP, they have so far mostly been investigated in the context of full blown leukemia. The iPSC model provides the opportunity to study their effects in isolation as they would present in pre-leukemic hematopoietic stem/progenitor cells before the onset of overt disease. Thus, the mutant iPSC lines we describe here can greatly facilitate future investigations into the mechanisms by which splicing gene mutations drive malignancy and the identification of relevant targets for therapeutic intervention.

(B) Representative bright-field images of eHPC cultures 49 days after the beginning of differentiation. Scale bars, 25 μm.
(C) Serial replating of eHPCs derived from normal and SRSF2-mutant iPSCs in methylcellulose culture. The first plating was performed on day 14 from the beginning of differentiation and cells were replated every 14 days thereafter. CFU-GEMM, colony-forming unit-granulocyte, erythrocyte, monocyte, megakaryocyte; CFU-GM, CFU-granulocyte, monocyte; CFU-G, colony-forming unit-granulocyte; CFU-M, CFU-monocyte; BFU-E, burst-forming unit-erythrocyte.
(D) Expansion of eHPCs in liquid culture estimated by cell counts. TR, transduced with library; UT, untransduced.
(E) eHPCs are immunophenotypically stable for several weeks.
(F) May-Giemsa staining of cytospin preparations of day 49 eHPCs derived from the indicated iPSC lines. Scale bars, 100 μm.
(G) eHPCs derived from the N-2.12 and the 5-16 Cre20 iPSC lines were treated every 2 days with 0.1 nM E7107. The relative population size was calculated as the number of the treated cells relative to the number of untreated cells at each time point. Error bars represent SEM from three independent experiments.

We provide evidence for a potential therapeutic value of inhibiting kinases regulating splicing, supporting splicing modulation as a broader strategy for therapeutic intervention in MDS with splicing gene mutations. Our screening approach highlights new opportunities for drug discovery and drug repurposing that iPSC models present, such as phenotype-based screening in the absence of predetermined targets (Eggert, 2013; Engle and Puppala, 2013) and the ability to use broad phenotypes such as growth and viability as readouts with isogenic controls. Although del(7q) has been known as a recurrent chromosomal abnormality in MDS for decades, the genetic and molecular mechanisms by which it drives MDS remain incompletely understood prohibiting rational development of targeted therapies. While follow-up and translation of our findings with niflumic acid is not straightforward, as the drug exerted its effect at a very high concentration in our study, this study provides proof of principle that our approach can lead to the identification of lead compounds targeting del(7q) and other intractable genetic lesions for future drug development. Parallel testing in isogenic controls, as we show here, can enhance sensitivity and specificity and obviate the need for multiple normal controls to account for genetic background differences.

A major limiting step in the use of iPSC-derived HPCs in disease modeling and drug testing is the derivation of the desired cell types through directed differentiation. This step introduces variability, increases experimentation time and cost, and limits scalability. Here we report the generation of expandable tissue-specific progenitor cells that can be grown similarly to conventional cell lines from iPSCs. These "secondary" cell lines can overcome some of the limitations of iPSC disease models, while maintaining disease-relevant features. Such cell lines can be widely accessible and particularly valuable for precise genetic studies of the effects of driver mutations in the context of a normal diploid genome, as most established immortalized cell lines harbor complex karyotypes with frequent gene copy-number gains or losses. Indeed, no human hematopoietic cell lines harboring splicing factor mutations in the context of a normal diploid genome currently exist. Our results suggest that different factor combinations can potentially sustain expandable cultures. Future experiments could define a minimal factor combination required to generate such expandable cells, which can, further, be expressed in an inducible system.

EXPERIMENTAL PROCEDURES

CRISPR/Cas9 Genome Editing

A plasmid co-expressing Cas9 linked to mCitrine by a P2A peptide driven by the cytomegalovirus promoter and a gRNA driven by the U6 promoter was constructed and gRNAs targeting the *SRSF2* locus within the first intron were designed and inserted. A donor plasmid containing a floxed cassette consisting of a human phosphoglycerate kinase promoter-driven puromycin resistance gene and 5′ and 3′ homology arms was constructed. One million iPSCs were nucleofected with 5 μg of CRISPR/Cas9 plasmid(s) and 10 μg of donor plasmid. After 7–10 days, single colonies were picked, allowed to grow for approximately 3–6 days and screened by PCR.

Human iPSC Culture and Hematopoietic Differentiation

Culture of human iPSCs on mouse embryonic fibroblasts or in feeder-free conditions was performed as described previously (Papapetrou and Sadelain, 2011). For hematopoietic differentiation, spin EBs were prepared and cultured in APEL medium, as described by Ng et al. (2008).

RNA Sequencing

Poly(A)-tailed mRNA was selected with beads using the NEBNext Poly(A) mRNA Magnetic Isolation Module. cDNAs were generated using random hexamers and ligated to barcoded Illumina adaptors with the NEXTflex Rapid Directional RNA-Seq Library Prep Kit (BioScientific). Sequencing of 75-nt-long single-end reads was performed in a NextSeq 500 (Illumina).

Small-Molecule Screen

The MicroSource Discovery Systems "Spectrum Collection" library was used. Using robotic equipment, 384-well plates were first coated with 20 μL of Matrigel (BD Biosciences) diluted 1:20 in DMEM/F12 media for 30 min at room temperature. The iPSCs were subsequently plated as a single-cell suspension at a density of 1,500 per well in 50 μL TeSR medium (STEMCELL Technologies) with 10 μM Rock inhibitor Y-27632. The next day, the cells were washed and the compounds were added at a final concentration range from 10, 1.0, 0.1 to 0.01 μM. On day 5, luminescence was measured using Promega's CellTiter-Glo assay.

ACCESSION NUMBERS

The accession number for the data reported in this paper is GEO: SRP108379.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and seven tables and can be found with this article online at https://doi.org/10.1016/j.stemcr.2018.03.020.

AUTHOR CONTRIBUTIONS

C.-J.C. performed the experiments, analyzed the data, and assisted with manuscript preparation. A.G.K., M.O., and M.G. performed the experiments. J.T.-F. performed cytological analyses. R.K.B. analyzed and interpreted RNA sequencing data and contributed to discussion of the results and manuscript preparation. O.A.-W. provided the splicing inhibitor drugs and contributed to discussion of the results and critical reading of the manuscript. H.S. analyzed barcode enrichment data. T.J.M. performed and analyzed

the small-molecule screen. R.S. and R.D. evaluated the small-molecule screen results. E.P.P. conceived, designed, and supervised the study, analyzed data, and prepared the manuscript with assistance from C.J.C. and R.K.B.

ACKNOWLEDGMENTS

We thank Chrystel Husser for excellent technical assistance. We thank H3 Biomedicine for kindly providing E7107 and Elli Papaemmanuil for helpful comments. This work was supported by NIH grants R01HL121570 and R01HL137219, by a Damon Runyon-Rachleff Innovation Award from the Damon Runyon Cancer Research Foundation, by a Pershing Square Sohn Prize from the Pershing Square Sohn Cancer Research Alliance, and by grants from the Edward P. Evans Foundation, the Ellison Medical Foundation, and the Henry and Marilyn Taub Foundation, all to E.P.P. R.K.B. was supported by NIH grants R01HL128239 and R01DK103854, the Edward P. Evans Foundation, the Ellison Medical Foundation (AG-NS-1030-13), and the Department of Defense Bone Marrow Failure Research Program (BM150092). The small-molecule screen was funded by a Quellos Research Acceleration Award by the Institute for Stem Cell and Regenerative Medicine, University of Washington, Seattle, WA.

Received: January 9, 2018
Revised: March 23, 2018
Accepted: March 23, 2018
Published: April 19, 2018

REFERENCES

Araki, S., Dairiki, R., Nakayama, Y., Murai, A., Miyashita, R., Iwatani, M., Nomura, T., and Nakanishi, O. (2015). Inhibitors of CLK protein kinases suppress cell growth and induce apoptosis by modulating pre-mRNA splicing. PLoS One *10*, e0116929.

Arrowsmith, C.H., Audia, J.E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P.E., Brown, P.J., Bunnage, M.E., et al. (2015). The promise and peril of chemical probes. Nat. Chem. Biol. *11*, 536–541.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science *338*, 1587–1593.

Dahlin, J.L., Inglese, J., and Walters, M.A. (2015). Mitigating risk in academic preclinical drug discovery. Nat. Rev. Drug Discov. *14*, 279–294.

Dahlin, J.L., and Walters, M.A. (2014). The essential roles of chemistry in high-throughput screening triage. Future Med. Chem. *6*, 1265–1290.

Day, C.P., Merlino, G., and Van Dyke, T. (2015). Preclinical mouse cancer models: a maze of opportunities and challenges. Cell *163*, 39–53.

Doulatov, S., Vo, L.T., Chou, S.S., Kim, P.G., Arora, N., Li, H., Hadland, B.K., Bernstein, I.D., Collins, J.J., Zon, L.I., et al. (2013). Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. Cell Stem Cell *13*, 459–470.

Dvinge, H., Kim, E., Abdel-Wahab, O., and Bradley, R.K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. Nat. Rev. Cancer *16*, 413–430.

Eggert, U.S. (2013). The why and how of phenotypic small-molecule screens. Nat. Chem. Biol. *9*, 206–209.

Engle, S.J., and Puppala, D. (2013). Integrating human pluripotent stem cells into drug development. Cell Stem Cell *12*, 669–677.

Eskens, F.A., Ramos, F.J., Burger, H., O'Brien, J.P., Piera, A., de Jonge, M.J., Mizui, Y., Wiemer, E.A., Carreras, M.J., Baselga, J., et al. (2013). Phase I pharmacokinetic and pharmacodynamic study of the first-in-class spliceosome inhibitor E7107 in patients with advanced solid tumors. Clin. Cancer Res. *19*, 6296–6304.

Fei, D.L., Motowski, H., Chatrikhi, R., Prasad, S., Yu, J., Gao, S., Kielkopf, C.L., Bradley, R.K., and Varmus, H. (2016). Wild-type U2AF1 antagonizes the splicing program characteristic of U2AF1-mutant tumors and is required for cell survival. PLoS Genet. *12*, e1006384.

Gao, H., Korn, J.M., Ferretti, S., Monahan, J.E., Wang, Y., Singh, M., Zhang, C., Schnell, C., Yang, G., Zhang, Y., et al. (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat. Med. *21*, 1318–1325.

Genovese, G., Kahler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N. Engl. J. Med. *371*, 2477–2487.

Gould, S.E., Junttila, M.R., and de Sauvage, F.J. (2015). Translational value of mouse models in oncology drug development. Nat. Med. *21*, 431–439.

Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E., et al. (2012). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat. Genet. *44*, 53–57.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med. *371*, 2488–2498.

Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.B., Murphy, M.E., et al. (2015). SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. Cancer Cell *27*, 617–630.

Kotini, A.G., Chang, C.J., Boussaad, I., Delrow, J.J., Dolezal, E.K., Nagulapally, A.B., Perna, F., Fishbein, G.A., Klimek, V.M., Hawkins, R.D., et al. (2015). Functional analysis of a chromosomal deletion associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells. Nat. Biotechnol. *33*, 646–655.

Kotini, A.G., Chang, C.J., Chow, A., Yuan, H., Ho, T.C., Wang, T., Vora, S., Solovyov, A., Husser, C., Olszewska, M., et al. (2017). Stage-specific human induced pluripotent stem cells map the progression of myeloid transformation to transplantable leukemia. Cell Stem Cell *20*, 315–328 e317.

Kronke, J., Fink, E.C., Hollenbach, P.W., MacBeth, K.J., Hurst, S.N., Udeshi, N.D., Chamberlain, P.P., Mani, D.R., Man, H.W., Gandhi, A.K., et al. (2015). Lenalidomide induces ubiquitination and degradation of CK1alpha in del(5q) MDS. Nature *523*, 183–188.

Lee, S.C., Dvinge, H., Kim, E., Cho, H., Micol, J.B., Chung, Y.R., Durham, B.H., Yoshimi, A., Kim, Y.J., Thomas, M., et al. (2016). Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. Nat. Med. *22*, 672–678.

Lis, R., Karrasch, C.C., Poulos, M.G., Kunar, B., Redmond, D., Duran, J.G.B., Badwe, C.R., Schachterle, W., Ginsberg, M., Xiang, J., et al. (2017). Conversion of adult endothelium to immunocompetent haematopoietic stem cells. Nature *545*, 439–445.

Malcovati, L., Galli, A., Travaglino, E., Ambaglio, I., Rizzo, E., Molteni, E., Elena, C., Ferretti, V.V., Catricala, S., Bono, E., et al. (2017). Clinical significance of somatic mutation in unexplained blood cytopenia. Blood *129*, 3371–3378.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in mammalian tissues. Science *338*, 1593–1599.

Ng, E.S., Davis, R., Stanley, E.G., and Elefanty, A.G. (2008). A protocol describing the use of a recombinant protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as spin embryoid bodies. Nat. Protoc. *3*, 768–776.

Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J.S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., et al. (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N. Engl. J. Med. *365*, 1384–1395.

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N., et al. (2016). Genomic classification and prognosis in acute myeloid leukemia. N. Engl. J. Med. *374*, 2209–2221.

Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A., et al. (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood *122*, 3616–3627, quiz 3699.

Papapetrou, E.P. (2016). Patient-derived induced pluripotent stem cells in cancer research and precision oncology. Nat. Med. *22*, 1392–1401.

Papapetrou, E.P., Lee, G., Malani, N., Setty, M., Riviere, I., Tirunagari, L.M., Kadota, K., Roth, S.L., Giardina, P., Viale, A., et al. (2011). Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells. Nat. Biotechnol. *29*, 73–78.

Papapetrou, E.P., and Sadelain, M. (2011). Generation of transgene-free human induced pluripotent stem cells with an excisable single polycistronic vector. Nat. Protoc. *6*, 1251–1273.

Pereira, C.F., Chang, B., Qiu, J., Niu, X., Papatsenko, D., Hendry, C.E., Clark, N.R., Nomura-Kitabayashi, A., Kovacic, J.C., Ma'ayan, A., et al. (2013). Induction of a hemogenic program in mouse fibroblasts. Cell stem cell *13*, 205–218.

Pimentel, H., Parra, M., Gee, S.L., Mohandas, N., Pachter, L., and Conboy, J.G. (2016). A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. Nucleic Acids Res. *44*, 838–851.

Riddell, J., Gazit, R., Garrison, B.S., Guo, G., Saadatpour, A., Mandal, P.K., Ebina, W., Volchkov, P., Yuan, G.C., Orkin, S.H., et al. (2014). Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. Cell *157*, 549–564.

Sandler, V.M., Lis, R., Liu, Y., Kedem, A., James, D., Elemento, O., Butler, J.M., Scandura, J.M., and Rafii, S. (2014). Reprogramming human endothelial cells to haematopoietic cells requires vascular induction. Nature *511*, 312–318.

Sperling, A.S., Gibson, C.J., and Ebert, B.L. (2017). The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. Nat. Rev. Cancer *17*, 5–19.

Sugimura, R., Jha, D.K., Han, A., Soria-Valles, C., da Rocha, E.L., Lu, Y.F., Goettel, J.A., Serrao, E., Rowe, R.G., Malleshaiah, M., et al. (2017). Haematopoietic stem and progenitor cells from human pluripotent stem cells. Nature *545*, 432–438.

Wong, J.J., Ritchie, W., Ebner, O.A., Selbach, M., Wong, J.W., Huang, Y., Gao, D., Pinello, N., Gonzalez, M., Baidya, K., et al. (2013). Orchestrated intron retention regulates normal granulocyte differentiation. Cell *154*, 583–595.

Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A., et al. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat. Med. *20*, 1472–1478.

Yeo, G.W., Van Nostrand, E., Holste, D., Poggio, T., and Burge, C.B. (2005). Identification and analysis of alternative splicing events conserved in human and mouse. Proc. Natl. Acad. Sci. USA *102*, 2850–2855.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64–69.

Zhang, J., Lieu, Y.K., Ali, A.M., Penson, A., Reggio, K.S., Rabadan, R., Raza, A., Mukherjee, S., and Manley, J.L. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proc. Natl. Acad. Sci. USA *112*, E4726–E4734.

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

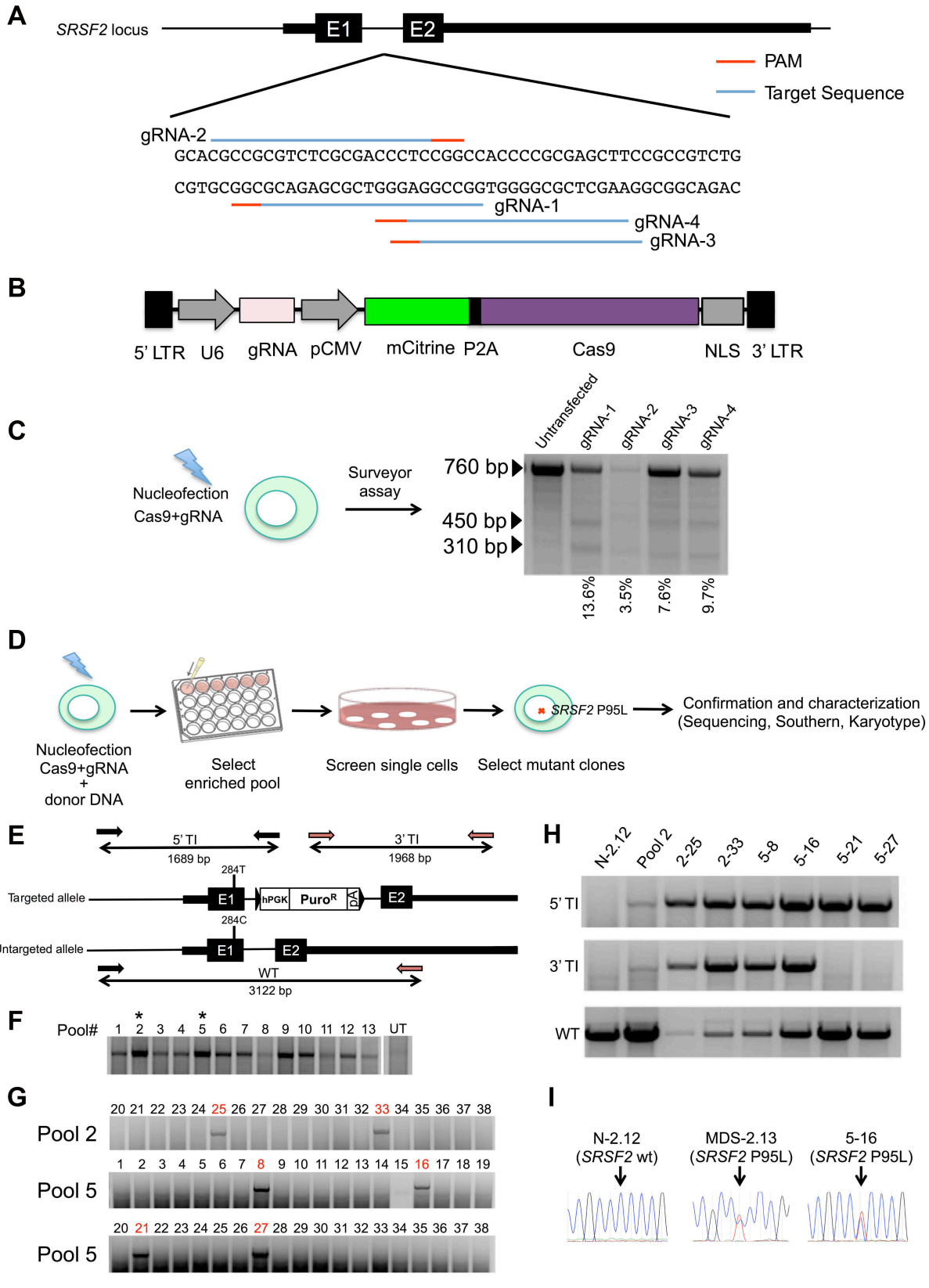

A
SRSF2 locus
E1
E2
PAM
Target Sequence
gRNA-2
GCACGCCGCGTCTCGCGACCCTCCGGCCACCCCGCGAGCTTCCGCCGTCTG
CGTGCGGCGCAGAGCGCTGGGAGGCCGGTGGGGCGCTCGAAGGCGGCAGAC
gRNA-1
gRNA-4
gRNA-3
B
5' LTR
U6
gRNA
pCMV
mCitrine
P2A
Cas9
NLS
3' LTR
C
Nucleofection
Cas9+gRNA
Surveyor
assay
Untransfected
gRNA-1
gRNA-2
gRNA-3
gRNA-4
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.

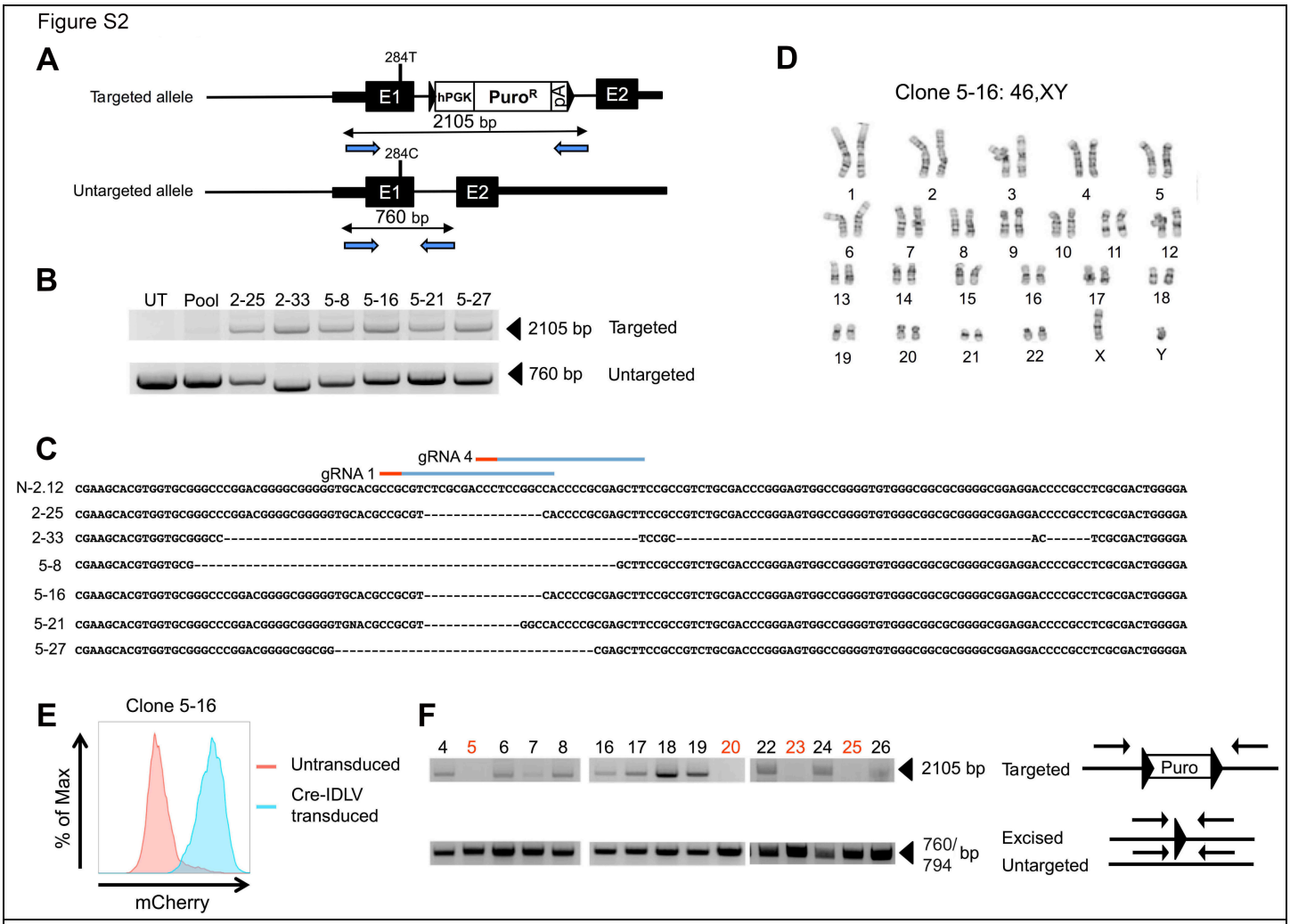


Figure S2. Characterization and cassette excision of *SRSF2* mutant iPSCs. Related to Figure 1

(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.

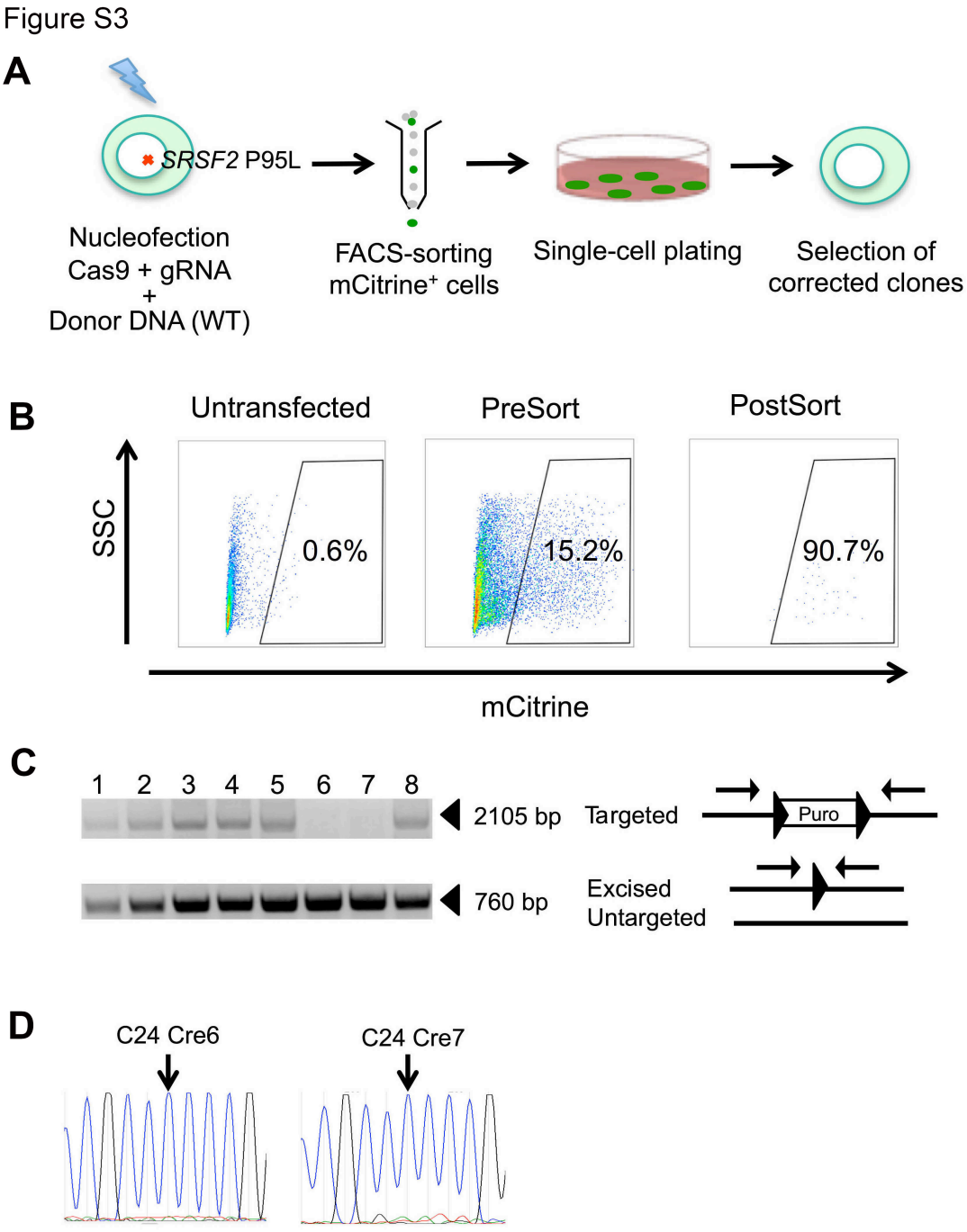


Figure S3. Correction of the *SRSF2* P95L mutation in MDS patient-derived iPSCs. Related to Figure 1

(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

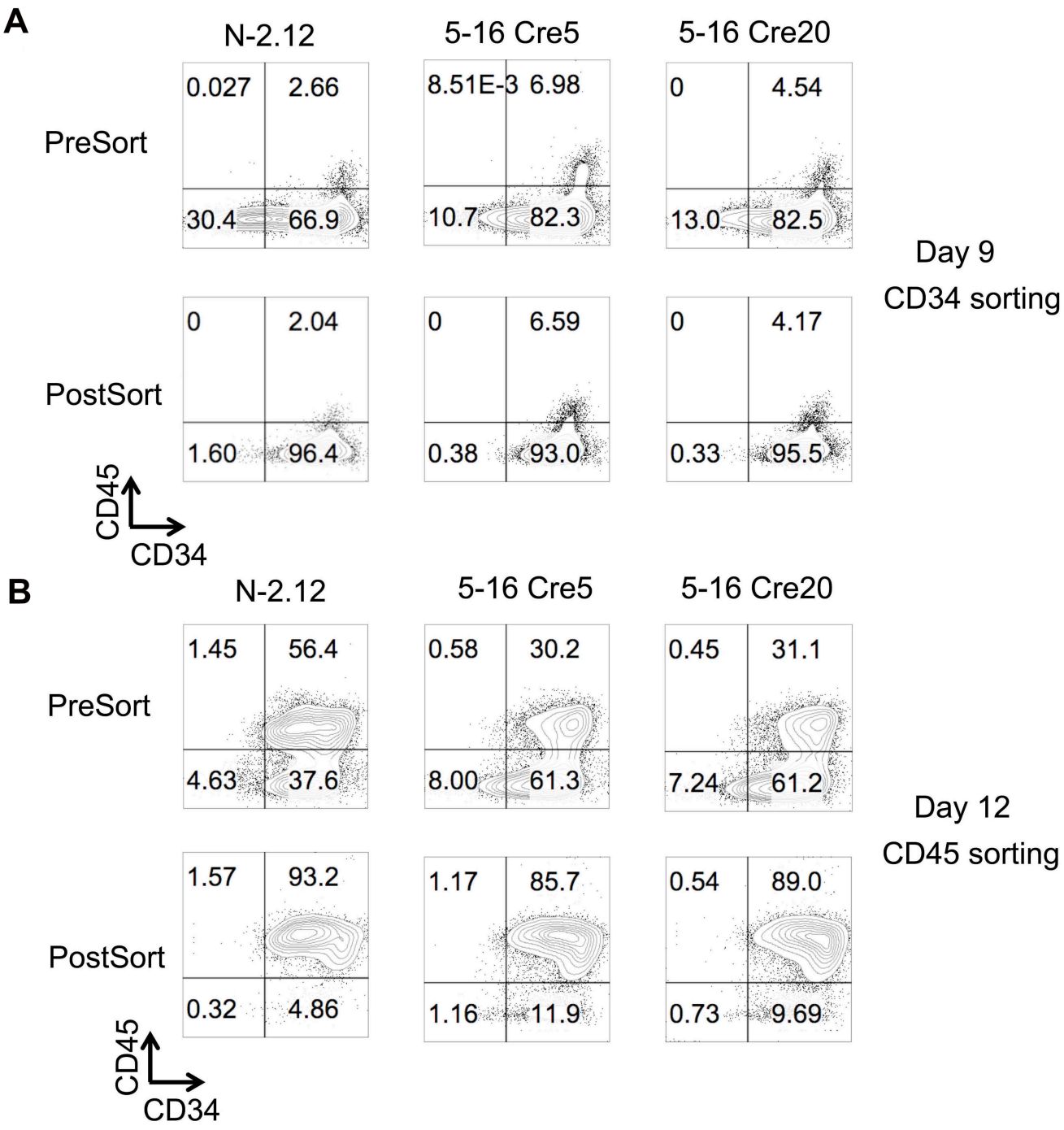


Figure S4. Sorting of isogenic SRSF2 WT and mutant iPSC-derived CD34$^+$ and CD45$^+$ hematopoietic progenitors for RNA sequencing analysis. Related to Figure 2

(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

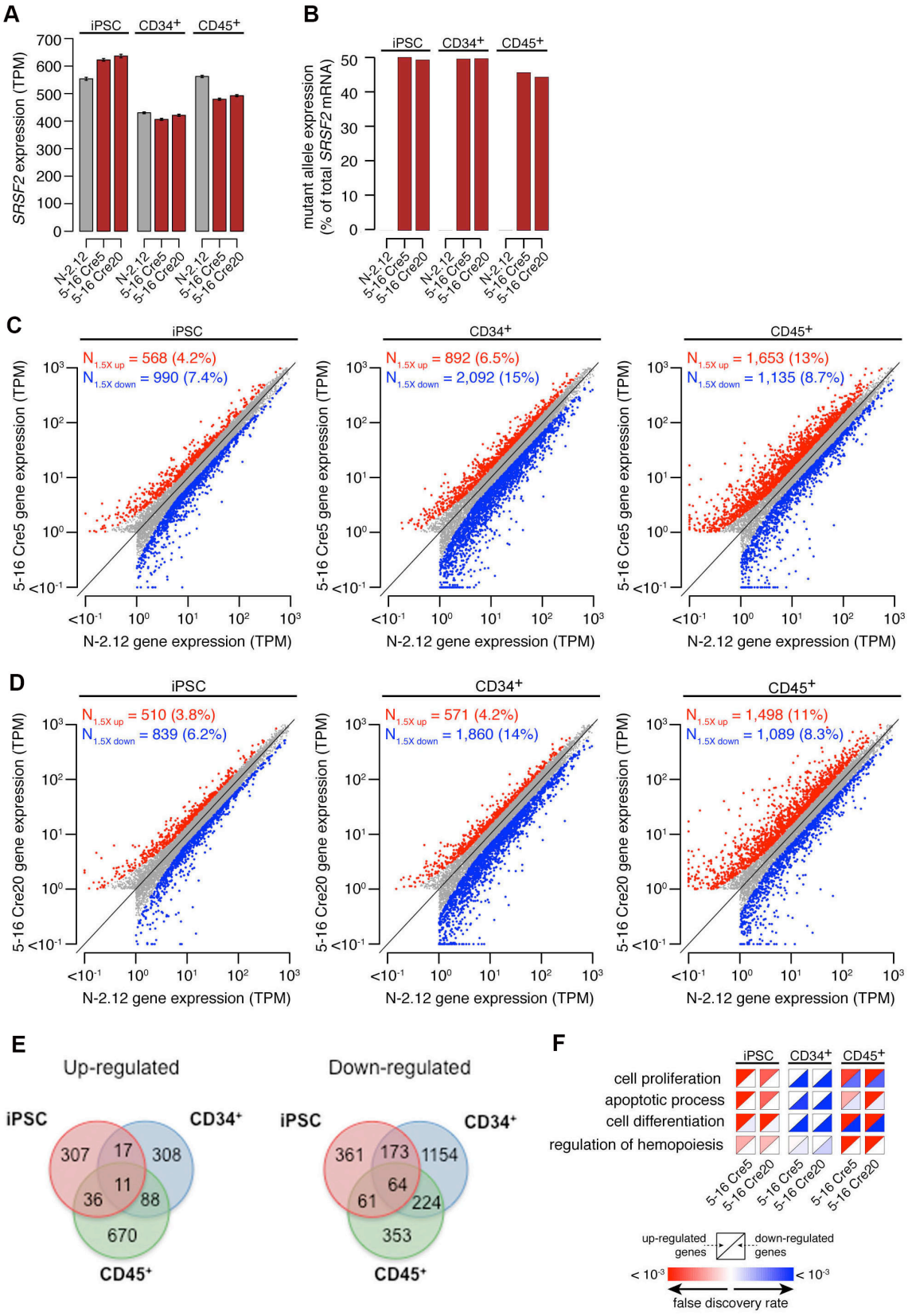

A
iPSC
CD34+
CD45+
SRSF2 expression (TPM)
N-2.12
5-16 Cre5
5-16 Cre20
B
mutant allele expression (% of total SRSF2 mRNA)
C
N1.5X up = 568 (4.2%)
N1.5X down = 990 (7.4%)
N1.5X up = 892 (6.5%)
N1.5X down = 2,092 (15%)
N1.5X up = 1,653 (13%)
N1.5X down = 1,135 (8.7%)
5-16 Cre5 gene expression (TPM)
N-2.12 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

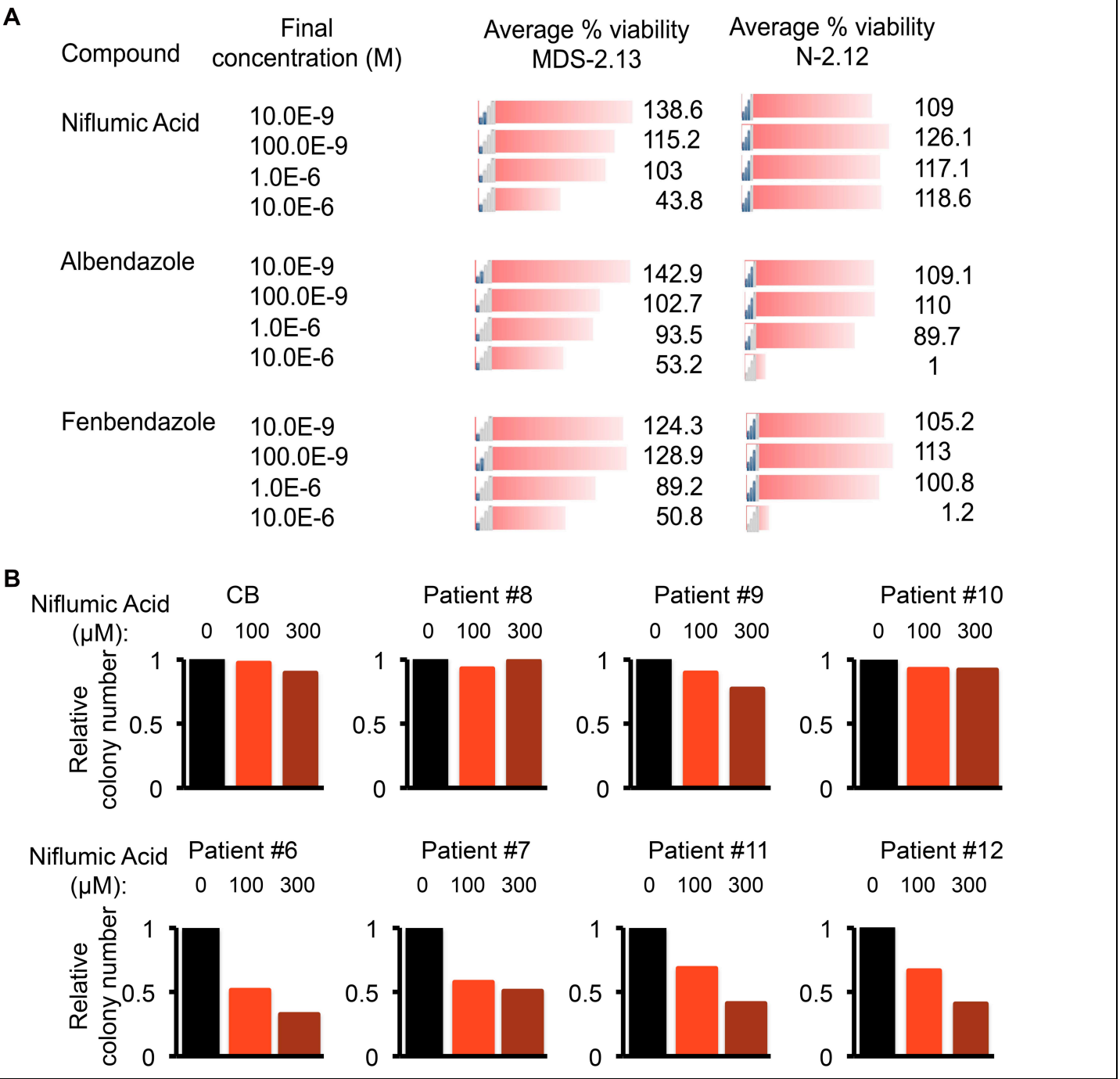


Figure S6. Selected hit compounds with preferential activity on MDS compared to isogenic normal iPSCs. Related to Figure 6

(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.

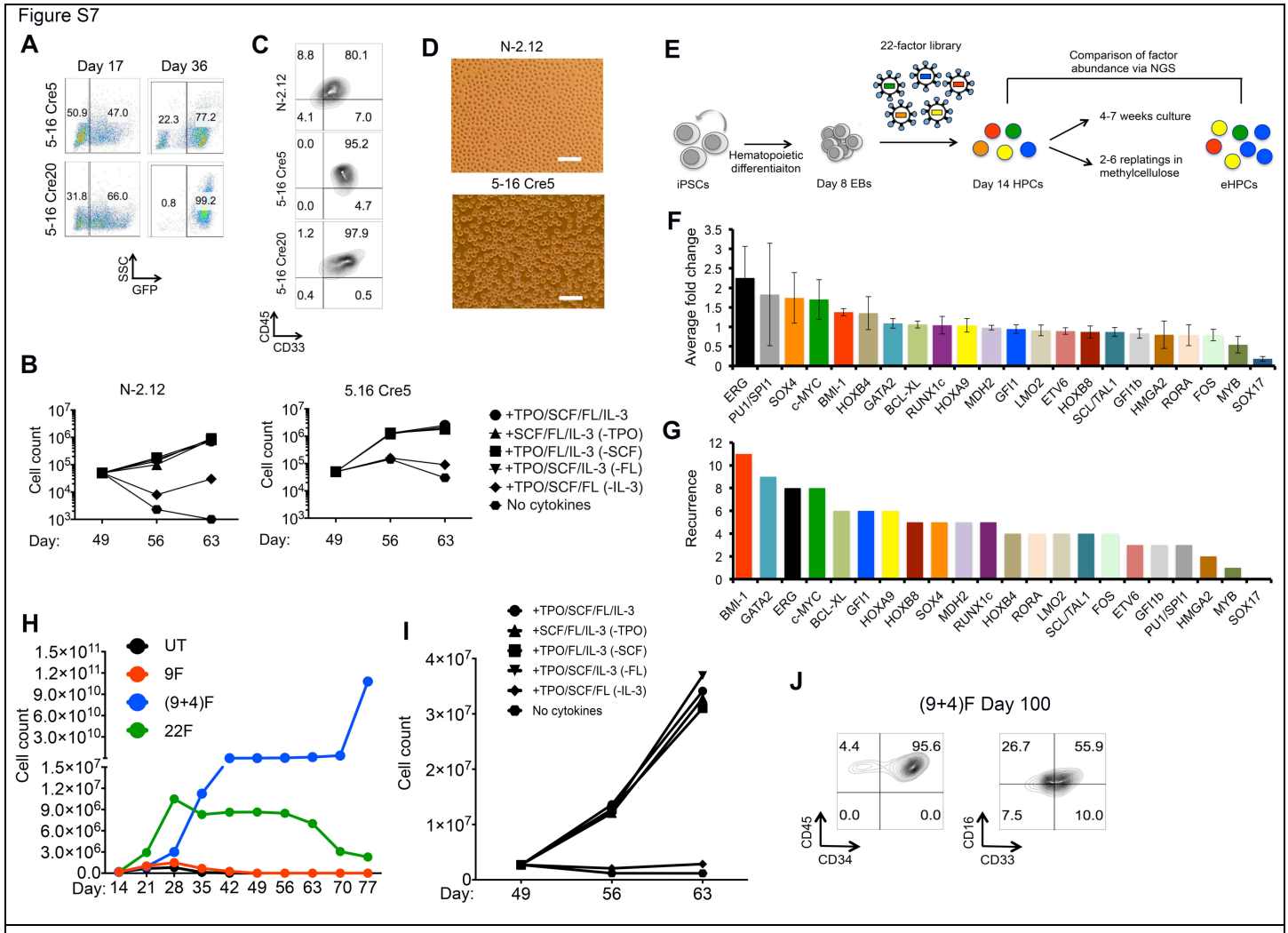


Figure S7. Characterization of expandable iPSC-HPCs. Related to Figure 7

(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.