JCI The Journal of Clinical Investigation

# TGFβ signaling underlies hematopoietic dysfunction and bone marrow failure in Shwachman-Diamond Syndrome

Cailin E. Joyce, … , Akiko Shimamura, Carl D. Novina

*J Clin Invest.* 2019. https://doi.org/10.1172/JCI125375.

**Concise Communication** In-Press Preview **Hematology**

Shwachman-Diamond Syndrome (SDS) is a rare and clinically-heterogeneous bone marrow (BM) failure syndrome caused by mutations in the Shwachman-Bodian-Diamond Syndrome (*SBDS*) gene. Although SDS was described over 50 years ago, the molecular pathogenesis is poorly understood due, in part, to the rarity and heterogeneity of the affected hematopoietic progenitors. To address this, we used single cell RNA sequencing to profile scant hematopoietic stem and progenitor cells from SDS patients. We generated a single cell map of early lineage commitment and found that SDS hematopoiesis was left-shifted with selective loss of granulocyte-monocyte progenitors. Transcriptional targets of transforming growth factor-beta (TGFβ) were dysregulated in SDS hematopoietic stem cells and multipotent progenitors, but not in lineage-committed progenitors. TGFβ inhibitors (AVID200 and SD208) increased hematopoietic colony formation of SDS patient BM. Finally, TGFβ3 and other TGFβ pathway members were elevated in SDS patient blood plasma. These data establish the TGFβ pathway as a novel candidate biomarker and therapeutic target in SDS and translate insights from single cell biology into a potential therapy.



# TGFβ signaling underlies hematopoietic dysfunction and bone marrow failure in Shwachman-Diamond Syndrome

Cailin E. Joyce[1,2]*, Assieh Saadatpour[3,4]*, Melisa Ruiz-Gutierrez[5], Ozge Vargel Bolukbasi[5], Lan Jiang[3,4], Dolly D. Thomas[1,2], Sarah Young[6], Inga Hofmann[5], Colin A. Sieff[5], Kasiani C. Myers[7], Jennifer Whangbo[5], Towia A. Libermann[2,8], Chad Nusbaum[6], Guo-Cheng Yuan[3,4], Akiko Shimamura[5], Carl D. Novina[1,2,6†]

[1]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, [2]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA. [3]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. [4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [5]Division of Hematology/Oncology, Boston Children's Hospital and Dana Farber Cancer Institute, Boston, MA 02115, USA. [6]Broad Institute of Harvard and MIT, Cambridge, MA 02141, USA. [7]Division of Bone Marrow Transplantation and Immune Deficiency, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, 45229, USA. [8]Beth Israel Deaconess Medical Center Genomics, Proteomics, Bioinformatics and Systems Biology Center, Division of Interdisciplinary Medicine and Biotechnology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA, 02215.

*These authors contributed equally to this work

[†]Corresponding author:

**Carl D. Novina**
Dana-Farber Cancer Institute, Dana 1420B
450 Brookline Avenue, Boston, MA 02215
carl_novina@dfci.harvard.edu
Tel: 617-582-7961
Fax: 617-582-7962

Conflict of interest statement: The authors have declared that no conflict of interest exists.

## ABSTRACT

Shwachman-Diamond Syndrome (SDS) is a rare and clinically-heterogeneous bone marrow (BM) failure syndrome caused by mutations in the Shwachman-Bodian-Diamond Syndrome (*SBDS*) gene. Although SDS was described over 50 years ago, the molecular pathogenesis is poorly understood due, in part, to the rarity and heterogeneity of the affected hematopoietic progenitors. To address this, we used single cell RNA sequencing to profile scant hematopoietic stem and progenitor cells from SDS patients. We generated a single cell map of early lineage commitment and found that SDS hematopoiesis was left-shifted with selective loss of granulocyte-monocyte progenitors. Transcriptional targets of transforming growth factor-beta (TGFβ) were dysregulated in SDS hematopoietic stem cells and multipotent progenitors, but not in lineage-committed progenitors. TGFβ inhibitors (AVID200 and SD208) increased hematopoietic colony formation of SDS patient BM. Finally, TGFβ3 and other TGFβ pathway members were elevated in SDS patient blood plasma. These data establish the TGFβ pathway as a candidate biomarker and therapeutic target in SDS and translate insights from single cell biology into a potential therapy.

## INTRODUCTION

Shwachman-Diamond Syndrome (SDS) is an inherited bone marrow (BM) failure syndrome associated with biallelic, hypomorphic mutations in the Shwachman-Bodian-Diamond Syndrome (*SBDS*) gene. SBDS is a pleiotropic protein that facilitates basic cellular processes such as ribosomal subunit joining and mitotic spindle assembly(1-5). Despite the simple genetic underpinnings of SDS, clinical heterogeneity driven by differences in the primarily affected blood cell lineages complicates diagnosis and treatment. BM failure typically manifests first in the myeloid lineage, but erythroid and megakaryocyte dysfunction may co-occur to varying degrees.

The only curative treatment for BM failure in SDS patients is hematopoietic stem cell (HSC) transplant. Unfortunately, outcomes are limited by the inability to predict which patients will develop complications, such as progression to clonal disease, that outweigh significant transplant risks. The development of rational therapies that could supplant or delay transplant requires a deeper understanding of the pathways that underlie cell type-specific responses to *SBDS* mutations. These pathways have been difficult to assess due to limitations of animal models and the paucity of human primary cells that can be obtained from BM failure patients. Here, we leverage recent technological advances in single cell profiling to directly examine the molecular pathogenesis of SDS in primary patient BM. Our findings implicate the TGFβ pathway as a potential therapeutic target in SDS and demonstrate the power of single cell transcriptomics to shed new light on rare and intractable diseases.

## COMBINED RESULTS AND DISCUSSION

Despite the basic cellular functions of SBDS, only certain cell types manifest dysfunction in SDS. BM hypocellularity and peripheral cytopenias involving multiple lineages(6, 7) are hallmarks of SDS, suggesting defects in the CD34+ hematopoietic stem and progenitor cell (HSPC) pool. We hypothesized that the dynamic subpopulations that comprise the HSPC pool may exhibit selective responses to *SBDS* mutations that influence clinical presentation. To simultaneously examine the consequences of *SBDS* mutations across HSPC subpopulations, we performed single cell RNA sequencing (RNA-seq) on $CD34^+$ cells freshly isolated from normal donor (n=4, ranging from 25-29 years old) and SDS patient (n=4, ranging from 11-26 years old) BM. The SDS patients all exhibited BM hypocellularity or cytopenias at the time of sampling; one patient was being treated with G-CSF for severe neutropenia (Supplementary Table 1) and is discussed separately below. We selected $CD34^+$ cells from the mononuclear fraction without gating on additional markers, sequenced single cells using the SMART-seq approach for full length cDNA amplification (Clontech)(8, 9) and classified HSPC a posteriori based on

transcriptional signatures of lineage commitment. This approach is well suited to capture cells along the $CD34^+$ differentiation spectrum, which is a subject of evolving understanding in human BM(10, 11).

A major challenge for studying a rare patient population is that biological variables and batch effects can obscure disease signatures. To classify single cells with respect to hematopoietic lineage commitment (and not other unrelated variables), we designed a supervised dimensionality reduction analysis. Specifically, we performed bulk RNA-seq on FACS-purified HSPC subpopulations(12) from normal BM to derive an mRNA expression signature that distinguished HSCs, multipotent progenitors (MPPs), common myeloid progenitors (CMPs), multilymphoid progenitors (MLPs), granulocyte-monocyte progenitors (GMPs), and megakaryocyte-erythroid progenitors (MEPs) (Supplementary Figure 1). We then analyzed this signature in single cell RNA-seq datasets from both normal and SDS BM to predict the identity of each cell. Data were visualized using t-distributed stochastic neighbor embedding (tSNE; Figure 1, Supplementary Table 2)(13). For simplicity, SDS cells are masked in Figure 1.

Cells from four normal donors were interspersed in a configuration that suggested population structure related to hematopoietic lineage commitment (Figure 1a). To associate regions of the map with specific lineages, we examined the expression of select mRNAs that are associated with stem, myeloid, erythroid, and lymphoid fate(11). We examined a set of mRNAs that were present in our 79-signature (Figure 1b), and a set that was absent from our signature as independent validation (Figure 1c). Most cells primarily expressed mRNAs associated with one fate, and expression of the different lineage-predictive mRNAs was concentrated in distinct regions of the tSNE map (Figure 1b,c). To confirm patterns of lineage commitment as determined by mRNA expression, we examined indexed surface marker intensities on a subset of normal cells. Gated HSCs, MPPs, MLPs, CMPs, GMPs or MEPs accounted for 68% of indexed cells; an additional 9% were $CD34^+CD90^-CD38^+CD10^+CD45RA^+$ common lymphoid progenitors (CLPs); the remaining 23% fell outside of defined gates and possibly represent transitional or

unconventional HSPC states. Cells that did fall within defined gates clustered in distinct regions of the map that were consistent with mRNA expression patterns (Figure 1d). Thus, supervised transcriptional mapping distinguished the major branches of hematopoiesis among randomly sampled $CD34^+$ cells.

We used this single cell map of normal hematopoietic lineage commitment as a baseline from which to examine alterations in the cellular architecture of SDS hematopoiesis. Figure 2a shows the same map as in Figure 1, with cells from SDS patients unmasked. SDS and normal cells were intermixed, but their distribution and relative frequencies differed ($\chi^2$ $p$<0.0001). We quantified these changes using *k*-means clustering. Five clusters were defined based on maximum silhouette value and named for the most enriched immunophenotypic subpopulation within the cluster (Figure 2a). CMP, MLP/CLP, GMP and MEP each designated a distinct cluster whereas HSC and MPP were enriched in the same cluster. Untreated SDS patients had a stark reduction in GMPs and a modest increase in HSC/MPP (Figure 2b). The reduction in GMP was evident even in the absence of symptomatic neutropenia (Supplementary Figure 2), suggesting that it contributes to the neutropenia predisposition in SDS patients. G-CSF treatment in one patient rescued loss of GMP and depleted HSC/MPP from the BM (Figure 2b), consistent with the drug's known mechanism(14). We therefore excluded cells from this treated patient from comparative gene expression analyses.

We next compared gene expression between normal and SDS cells within each cluster except for GMP, which was excluded due to the low number of GMP in untreated SDS patients. Overall, 1680 genes were differentially expressed in at least one cluster (FDR<0.05, |log2(fold change)| >1, Supplementary Table 3). Strikingly, 81.5% of all differentially expressed genes were unique to either HSC/MPP or CMP (Figure 3a). An additional 9.8% were commonly affected in HSC/MPP and CMP, but not in MLP/CLP or MEP. Overall, these data demonstrate that despite the general biochemical functions of the SBDS protein, *SBDS* mutations differentially affect the

frequency (as for GMP) or gene expression characteristics (as for HSC/MPP or CMP) of HSPC subpopulations. In contrast, the MLP/CLP and MEP populations are relatively unaffected.

The Inflammatory Response was enriched among differentially-expressed genes in both the HSC/MPP and CMP clusters (maximum *p*-value $4.98 \times 10^{-5}$ and $1.18 \times 10^{-3}$, respectively). However, the genes contributing to the enrichment differed between the clusters (Figure 3b). Transforming growth factor-beta (TGFβ) was the top regulator predicted for the HSC/MPP inflammatory response ($p=4.03 \times 10^{-15}$, z-score=0.891). It was also a significant upstream regulator among all differentially-expressed genes in HSC/MPP ($p=1.27 \times 10^{-2}$, z-score=0.417). Dysregulation of these TGFβ targets was most significant in HSC/MPP, with lesser or no effect in other HSPC populations (Figure 3c). TGFβ induces context-dependent effects on cell growth, survival, inflammation, and extracellular matrix. TGFβ1 and TGFβ3 have potent growth inhibitory effects on HSC(15-17). Thus, we hypothesized that activation of TGFβ in SDS HSC/MPP may contribute to BM failure in SDS.

To confirm activation of TGFβ signaling in SDS BM, we assessed TGFβ dependent phosphorylation and nuclear translocation of the transcriptional co-activator protein Mothers Against Decapentaplegic Homolog 2 (p-SMAD2). A subset of CD34+ cells from SDS BM had elevated levels of nuclear p-SMAD2 that were outside the normal range (Figure 4a, b). Treating SDS cells with AVID200, a decoy receptor trap designed to specifically neutralize TGFβ1 and TGFβ3, reduced the p-SMAD2 signal. The same trend was observed to varying degrees in two additional sample pairs (Figure 4c). These data are consistent with our single cell RNA-seq analysis demonstrating selective activation of the TGFβ pathway in the HSC/MPP subset of SDS CD34+ cells.

BM cells from SDS patients exhibit impaired hematopoietic colony formation in vitro(18) (Supplementary Figure 3a). To determine whether attenuation of TGFβ signaling improves SDS hematopoiesis, we cultured primary BM mononuclear cells from SDS patients and normal donors

(Supplementary Table 1) in methylcellulose supplemented with AVID200 and SD208, which inhibits TGFβR1 kinase activity(19). Both compounds improved hematopoietic colony formation in SDS patient samples, but not in normal donor controls (Figure 4d, Supplementary Figure 3b, Supplementary Table 4). Taken together, our data support a model in which activation of TGFβR1 kinase activity by TGFβ1 and/or TGFβ3 lead to increased concentration of p-SMAD2 in the nucleus and transcription of inflammatory response genes in SDS HSC/MPP (Figure 4e).

To determine whether SDS patients express elevated levels of TGFβ ligands, we screened blood plasma proteins from six SDS patients and six normal controls (Supplementary Table 1) using SOMAscan; a highly-sensitive, aptamer-based proteomic platform(20). TGFβ3 was significantly upregulated in SDS patient plasma, along with several other factors that were annotated to a network of TGFβ-associated factors (Figure 4f, Supplementary Figure 4). These and other dysregulated plasma proteins that were common across clinically-heterogeneous patients could serve as diagnostic biomarkers for SDS (Supplementary Table 5). Further studies are required to determine the levels of TGFβ3 in the BM compartment and identify the cell types that produce it.

Although SDS was reported over 50 years ago and progress has been made using animal and cellular models(3, 21-23), the molecular mechanisms leading to BM failure remain unclear. Here we leveraged advanced single cell technologies to perform the first direct analysis of primary human SDS hematopoietic progenitors. Whereas most single cell transcriptomic studies have focused on dissecting and characterizing cell types(24-27), this study demonstrates the power of single cell transcriptomics to uncover a key disease mechanism in rare cells. Our data add to an emerging body of evidence linking inflammation to BM dysfunction, including Fanconi Anemia (FA) where the pathogenic mechanism of TGFβ is thought to be suppression of homologous recombination repair(28, 29). We demonstrate a broader role for TGFβ in a mechanistically distinct BM failure syndrome. TGFβ inhibitors are already in clinical trials to treat myelodysplastic

syndrome, cancer, and pulmonary fibrosis, among others(30). Our work suggests that TGFβ1/3 inhibition by an agent such as AVID200 could be an effective therapy across clinically-heterogeneous SDS patients and different marrow failure disorders.

## METHODS

Detailed methods are provided as Supplementary Material.

## AUTHOR CONTRIBUTIONS

C.E.J., C.D.N., A.Sh., and C.D.N. designed experiments; I.H., C.A.S., A.Sh., M.R-G., and K.C.M. collected patient samples and clinical information; C.E.J, M.R-G., O.V-B., and D.D.T. performed experiments; A.Sa., L.J., and S.Y. performed computational analyses; C.E.J., A.Sa., M.R-G., O.V-B., C.D.N., A.Sh., G-C.Y., and T.A.L. analyzed data; C.E.J and C.D.N. wrote the manuscript; all authors provided critical reviews of the manuscript.

## ACKNOWLEDGMENTS

This work was supported by a Department of Defense Idea Award W81XWH-14-1-0124 and National Institutes of Health grant 1 R01 DK102165 to C.D.N; National Institutes of Health grant R24 DK099808 to A.Sh., and a Claudia Adams Barr Award from Dana-Farber Cancer Institute to G-C.Y. C.E.J. was supported by National Institutes of Health training grant T32 CA070083 and postdoctoral fellowship F32 HL124941. We thank Formation Biologics (Austin, TX, USA) for supplying AVID200. We thank Dr. Donna Neuberg for critical scientific advice. We thank the Dana-Farber Cancer Institute Flow Cytometry Core (Boston, MA, USA), especially John Daley, Michael Buonopane, and Alexander Heubeck, for providing technical expertise and equipment; Broad Technology Labs (Cambridge, MA, USA), especially Jim Bochicchio and Caroline Cusick for providing project management support; the Beth Israel Deaconess Medical

Center Genomics, Proteomics, Bioinformatics and Systems Biology Center, especially Simon T. Dillon, Xuesong Gu, Hasan Out for providing technical and analytical support (Boston, MA, USA); and the Broad Genomics Platform (Cambridge, MA, USA).

## REFERENCES

1. Menne TF, et al. The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast. *Nature genetics.* 2007;39:486-95.
2. Ganapathi KA, et al. The human Shwachman-Diamond syndrome protein, SBDS, associates with ribosomal RNA. *Blood.* 2007;110:1458-65.
3. Finch AJ, et al. Uncoupling of GTP hydrolysis from eIF6 release on the ribosome causes Shwachman-Diamond syndrome. *Genes & development.* 2011;25:917-29.
4. Burwick N, Coats SA, Nakamura T, and Shimamura A. Impaired ribosomal subunit association in Shwachman-Diamond syndrome. *Blood.* 2012;120:5143-52.
5. Austin KM, et al. Mitotic spindle destabilization and genomic instability in Shwachman-Diamond syndrome. *J Clin Invest.* 2008;118(4):1511-8.
6. Huang JN, and Shimamura A. Clinical spectrum and molecular pathophysiology of Shwachman-Diamond syndrome. *Current opinion in hematology.* 2010;18:30-5.
7. Myers KC, et al. Variable clinical presentation of Shwachman-Diamond syndrome: update from the North American Shwachman-Diamond Syndrome Registry. *The Journal of pediatrics.* 2014;164:866-70.
8. Ramskold D, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol.* 2012;30(8):777-82.
9. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, and Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods.* 2013;10(11):1096-8.

10. Notta F, Doulatov S, Laurenti E, Poeppl A, Jurisica I, and Dick JE. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science (New York, NY).* 2011;333:218-21.

11. Velten L, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol.* 2017;19(4):271-81.

12. Laurenti E, et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nature immunology.* 2013;14:756-63.

13. van der Maaten L, Hinton, G. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research.* 2008(9):2579-605.

14. Thomas J, Liu F, and Link DC. Mechanisms of mobilization of hematopoietic progenitors with granulocyte colony-stimulating factor. *Curr Opin Hematol.* 2002;9(3):183-9.

15. Hatzfeld J, et al. Release of early human hematopoietic progenitors from quiescence by antisense transforming growth factor beta 1 or Rb oligonucleotides. *J Exp Med.* 1991;174(4):925-9.

16. Scandura JM, Boccuni P, Massague J, and Nimer SD. Transforming growth factor beta-induced cell cycle arrest of human hematopoietic cells requires p57KIP2 up-regulation. *Proc Natl Acad Sci U S A.* 2004;101(42):15231-6.

17. Challen GA, Boles NC, Chambers SM, and Goodell MA. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. *Cell Stem Cell.* 2010;6(3):265-78.

18. Dror Y, and Freedman MH. Shwachman-Diamond syndrome: An inherited preleukemic bone marrow failure disorder with aberrant hematopoietic progenitors and faulty marrow microenvironment. *Blood.* 1999;94(9):3048-54.

19. Uhl M, Aulwurm S, Wischhusen J, Weiler M, Ma JY, Almirez R, et al. SD-208, a novel transforming growth factor beta receptor I kinase inhibitor, inhibits growth and invasiveness and enhances immunogenicity of murine and human glioma cells in vitro and in vivo. *Cancer Res.* 2004;64(21):7954-61.

20. Gold L, Walker JJ, Wilcox SK, and Williams S. Advances in human proteomics at high scale with the SOMAscan proteomics platform. *N Biotechnol.* 2012;29(5):543-9.

21. Tourlakis ME, Zhong J, Gandhi R, Zhang S, Chen L, Durie PR, et al. Deficiency Of Sbds In The Mouse Pancreas Leads To Features Of Shwachman-Diamond Syndrome, With Loss Of Zymogen Granules. *Gastroenterology.* 2012.

22. Zambetti NA, Bindels EMJ, Van Strien PMH, Valkhof MG, Adisty MN, Hoogenboezem RM, et al. Deficiency of the ribosome biogenesis gene Sbds in hematopoietic stem and progenitor cells causes neutropenia in mice by attenuating lineage progression in myelocytes. *Haematologica.* 2015;100:1285-93.

23. Tulpule A, Kelley JM, Lensch MW, McPherson J, Park IH, Hartung O, et al. Pluripotent Stem Cell Models of Shwachman-Diamond Syndrome Reveal a Common Mechanism for Pancreatic and Hematopoietic Dysfunction. *Cell stem cell.* 2013;12:727-36.

24. Villani AC SR, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A, Zheng S, Lazo S, Jardine L, Dixon D, Stephenson E, Nilsson E, Grundberg I, McDonald D, Filby A, Li W, De Jager PL, Rozenblatt-Rosen O, Lane AA, Haniffa M, Regev A, Hacohen N. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science.* 2017;356(6335).

25. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (New York, NY).* 2016;352:189-96.

26. Kumar RM, Cahan P, Shalek AK, Satija R, DaleyKeyser AJ, Li H, et al. Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature.* 2014;516:56-61.

27. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences.* 2015;112:201507125.

28. Zhang H, Kozono DE, O'Connor KW, Vidal-Cardenas S, Rousseau A, Hamilton A, et al. TGF-beta Inhibition Rescues Hematopoietic Stem Cell Defects and Bone Marrow Failure in Fanconi Anemia. *Cell Stem Cell.* 2016;18(5):668-81.

29. Zhou L, McMahon C, Bhagat T, Alencar C, Yu Y, Fazzari M, et al. Reduced SMAD7 leads to overactivation of TGF-beta signaling in MDS that can be reversed by a specific inhibitor of TGF-beta receptor I kinase. *Cancer Res.* 2011;71(3):955-63.

30. Herbertz S, Sawyer JS, Stauber AJ, Gueorguieva I, Driscoll KE, Estrem ST, et al. Clinical development of galunisertib (LY2157299 monohydrate), a small molecule inhibitor of transforming growth factor-beta signaling pathway. *Drug Des Devel Ther.* 2015;9:4479-99.

**FIGURE 1**

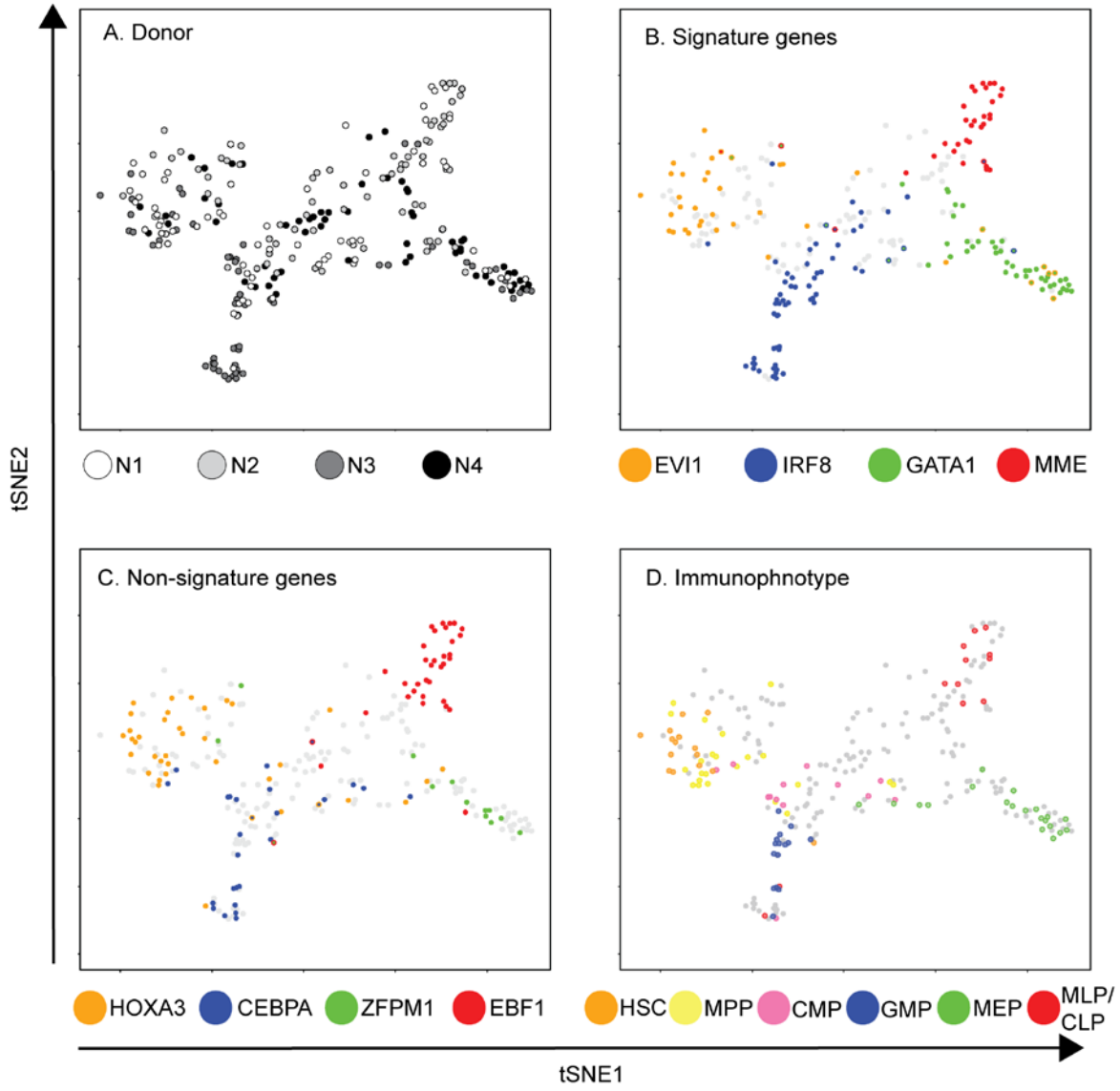


**Figure 1. Supervised dimensionality reduction maps lineage commitment of CD34+ cells from healthy donors.** tSNE plot of hematopoietic lineage commitment was derived from an empirically-defined gene expression signature. Shown here are cells from four normal donors ($n^{N1}$=70, $n^{N2}$=58, $n^{N1}$=69, $n^{N1}$=59, $n^{total}$=256). Cells are colored based on (a) donor identity, (b) mRNA expression of selected signature genes, (c) mRNA expression of lineage-restricted genes reported elsewhere[12], and (d) immunophenotypes. For (b,c), color indicates TPM>1 for the

indicated stem- (orange), myeloid- (blue), erythroid- (green), or lymphoid- (red) enriched mRNA. The presence of two colors indicates co-expression. Grey indicates TPM<1 for all four factors. For (d), color indicates membership in a gated immunophenotypic subset as shown in Extended Data Figure 1a, b. Grey indicates cells that were ungated or sorted without indexing. Numerical axes derived from tSNE are arbitrary, and therefore not shown.

**FIGURE 2**

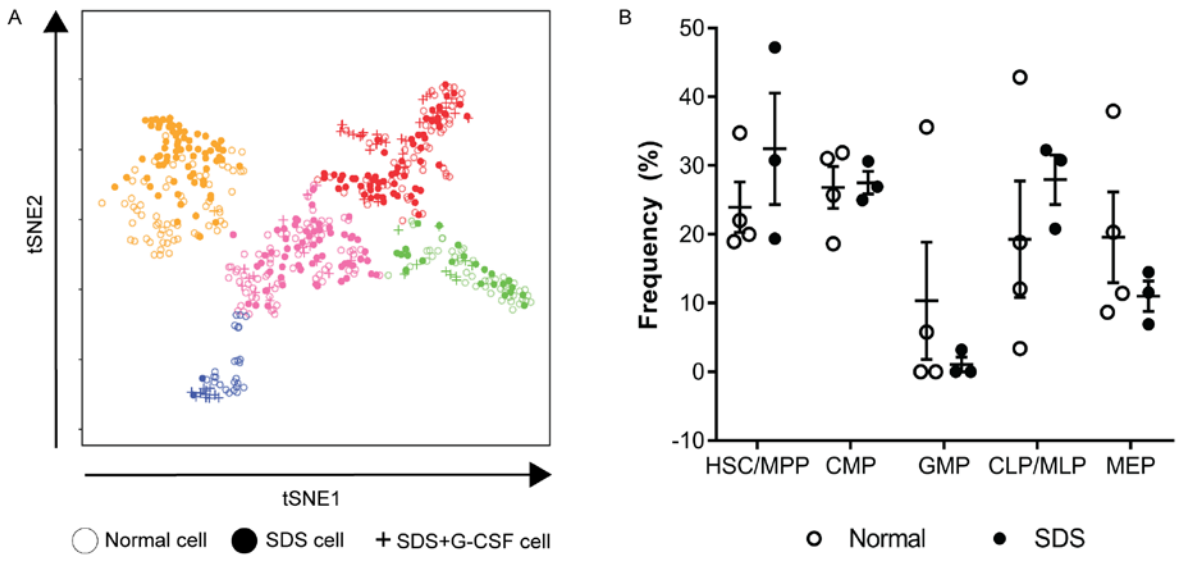


**Figure 2. The cellular architecture of early hematopoiesis is altered in SDS.** (a) tSNE plot of hematopoietic lineage commitment showing cells from normal donors as in Figure 1, untreated SDS patients ($n^{SDS1.1}$=72, $n^{SDS1.2}$=62, $n^{SDS2.1}$=78, $n^{total}$=212), and an SDS patient who was being treated with 4.2ug/kg/day G-CSF ($n^{SDS2.2}$=71). Clusters were determined using 'partitioning around medoids' version of *k*-means clustering (*k*=5), and labeled based on the enrichment of index sorted HSC, MPP, MLP, CMP, GMP and MEP as shown in Figure 1d. The sum of normal cells and SDS cells in each cluster is significantly changed using the $\chi^2$ test. (b) Relative frequencies of HSPC subpopulations for normal donors and untreated SDS patients. Error bars=SEM.

**FIGURE 3**

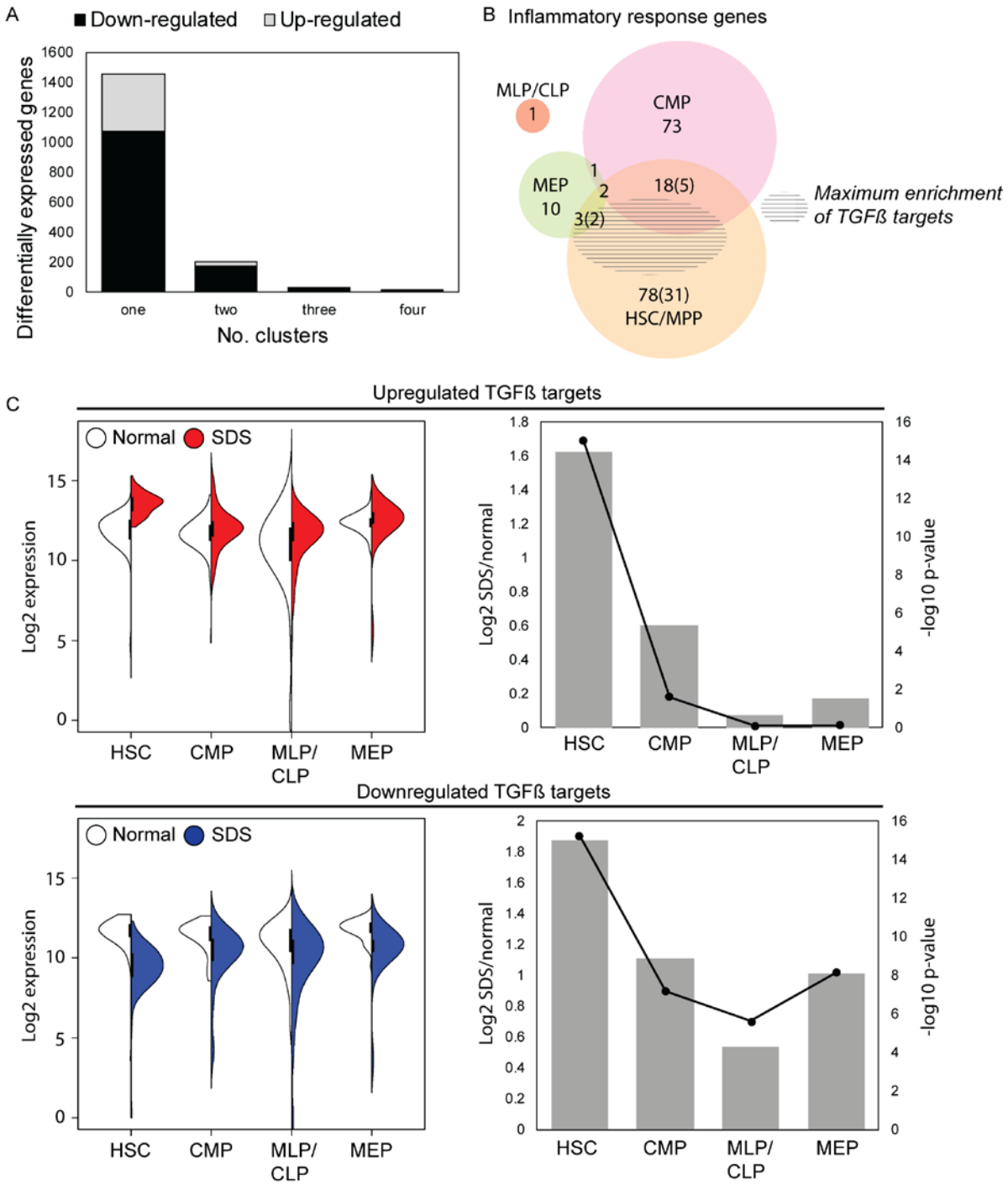


**Figure 3. TGFβ signaling is selectively activated in SDS stem and multipotent progenitors.**

**(**a) Differentially expressed genes were identified among all SDS versus normal cells and within

each cluster – HSC/MPP, CMP, MLP/CLP, or MEP. To aid biological interpretation, this gene set was filtered to focus on genes with FDR adjusted p-value < .05 and log2(fold change) >|1| in at least one cluster. Plotted are the number of genes that were either up- or down-regulated in one, two, three or four clusters. GMP was excluded due to the paucity of SDS GMP. Inset pie chart shows the proportion of differentially expressed genes in each cluster. (b) Venn diagram of differentially expressed genes in each cluster that were annotated to the "Inflammatory Response" function in Ingenuity Pathway Analysis. The shaded region shows the area of maximal enrichment of TGFβ targets ($p=4.03x10^{-15}$). (c) Left: split violin for the summed expression of 25 upregulated TGFβ targets and 52 down-regulated TGFβ targets in SDS HSC/MPP. Right: Log2 fold changes (primary axis, bars) and *p*-values (secondary axis, lines) for the gene sets plotted in 'b'. Significance was determined by two-way ANOVA, with Holm-Sidak's multiple comparisons test.

**FIGURE 4**

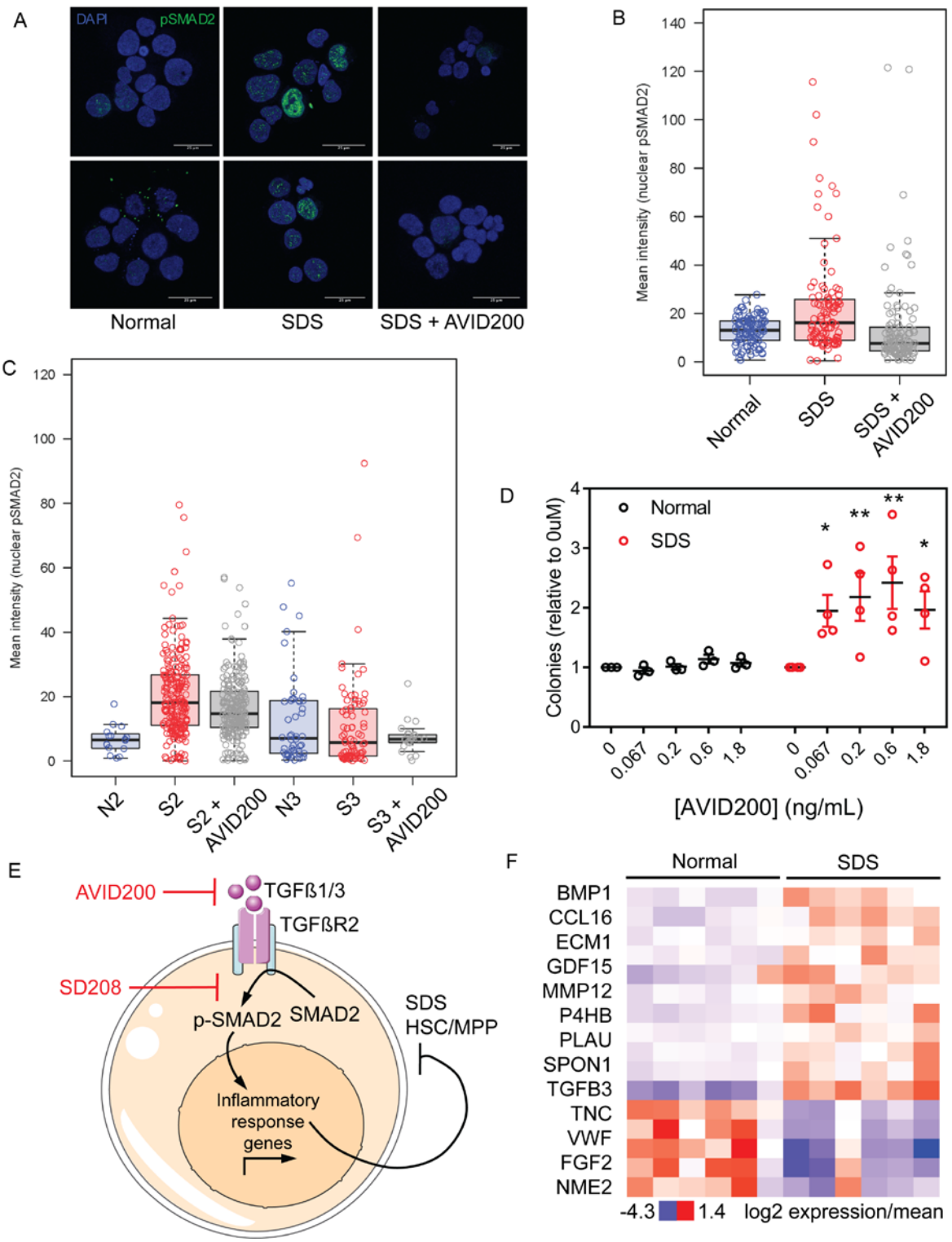
A
DAPI
pSMAD2
Normal
SDS
SDS + AVID200
B
Mean intensity (nuclear pSMAD2)
Normal
SDS
SDS + AVID200
C
Mean intensity (nuclear pSMAD2)
N2
S2
S2 + AVID200
N3
S3
S3 + AVID200
D
Normal
SDS
Colonies (relative to 0uM)
[AVID200] (ng/mL)
E
AVID200
TGFß1/3
TGFßR2
SD208
p-SMAD2
SMAD2
SDS
HSC/MPP
Inflammatory response genes
F
Normal
SDS
BMP1
CCL16
ECM1
GDF15
MMP12
P4HB
PLAU
SPON1
TGFB3
TNC
VWF
FGF2
NME2
-4.3
1.4
log2 expression/mean

**Figure 4. TGFβ pathway activation through TGFβR1 suppresses hematopoiesis in SDS BM progenitors.** a) Representative images showing DAPI and phospho-SMAD2 staining of primary BM CD34+ cells from adult normal donor BM and pediatric SDS BM, either untreated or treated with AVID200. b) Mean intensity of phospho-SMAD2 staining in individual CD34+ nuclei from samples depicted in panel (a). Significance was determined by two-way ANOVA, with Holm-Sidak's multiple comparisons test. Error bars= minimum and maximum values, excluding outliers that exceed median+1.5*IQR. **p<0.01, ***p<0.001. c) Mean intensity of phospho-SMAD2 staining in individual CD34+ nuclei in two additional pairs of SDS and normal donor BM samples. Error bars= minimum and maximum values, excluding outliers that exceed median+1.5*IQR. **p<0.01, ****p<0.0001. d) Number of colonies formed by adult normal donor and pediatric SDS patient BM-derived mononuclear cells with increasing concentrations of AVID200, normalized to the 0uM treatment. Significance was determined relative to the 0uM treatment by two-way ANOVA, with Holm-Sidak's multiple comparisons test. Error bars=SEM. *p<0.05, **p<0.01. e) Model for the role of TGFβ signaling in SDS BM failure. TGFβ1 and/or TGFβ3 ligands (targets of AVID200 inhibitor) activate signaling through the TGFβR1 receptor (target of SD208 inhibitor) on SDS HSC/MPP. Our data suggest that TGFβ ligands are primarily derived from a $CD34^-$ cell type in BM because TGFβ ligand mRNAs were not detected in CD34+ HSPC. Increased TGFβR1 signaling leads to increased concentrations of nuclear phospho-SMAD2 and transcription of inflammatory response genes, which impairs HSC/MPP function. This model predicts that therapeutic inhibition of TGFβ signaling in HSC/MPP will improve hematopoietic function in SDS patients. f) Expression of extracellular proteins annotated to a TGFβ network that was enriched among dysregulated proteins in SDS patient plasma. Asterisks indicate TGFβ family ligands.

## SUPPLEMENTAL METHODS

**Sample processing.** For scRNA-seq of all SDS samples, and normal donors N1 and N2: 7-20 ml of fresh BM were diluted to 35ml in MACS buffer (PBS/2mM EDTA/0.5% BSA), layered onto 15ml Ficoll-paque (GE Healthcare, Uppsala, Sweden), and spun for 30 min at 1400 rpm and 20°C with no brakes. Mononuclear cells were collected from the interface, washed once, pelleted for 5 min at 1200 rpm and 20°C, and resuspended at 40 ul per $10^7$ cells in MACS buffer + 1 ul/ml RNaseOUT (Thermo Fisher Scientific, Waltham, MA, USA). CD34+ cells were positively selected on an AutoMACS instrument using the Indirect CD34 MicroBead Kit (Miltenyi, Bergisch Gladbach, Germany), and singulated on the C1 Instrument (Fluidigm, San Francisco, CA, USA). cDNA libraries were prepared using the SMARTer Ultra Low RNA Kit (Clontech, Mountain View, CA, USA). For samples N3 and N4, protocol conditions were modified to ascertain immunophenotypes from single cells, and in accordance with the newest available methods. For these samples: red blood cells were lysed with ammonium chloride (Stem Cell Technologies, Vancouver, CA). Mononuclear cells were pelleted for 5 min at 1200 rpm and 20°C, washed twice, and resuspended in PBS + 1 ul/ml RNaseOUT. Cells were stained as described below. Single CD34+ cells were sorted into 5ul TCL buffer (Qiagen, Hilden, Germany) in 96 well plates using a FACS Aria II instrument (BD, Franklin Lakes, NJ, USA) on index mode. Two technical replicates of 100 cells from each gated CD34+ subset – HSC, MPP, MLP, CMP, GMP, MEP – were sorted into 5 ul TCL buffer in separate 96 well plates. cDNA libraries were prepared using the SMART-Seq v4 Ultra Low RNA Kit (Clontech). Libraries from all samples were sequenced on a HiSeq 2500 Instrument (Illumina, San Diego, CA) to a read depth of ~3 M paired-end, 25 bp reads per single cell, or ~12 M paired-end, 25 bp reads per 100 cells.

**Antibodies and staining.** Cells were stained at a density of $1x10^6$ per 100 ul in PBS + 1 ul/ml RNaseOUT because staining buffers contain proteins that can inhibit SMARTer-seq (Clontech) cDNA synthesis reactions. The staining panel was adapted from an analysis of human cord blood

progenitors(1). in accordance with the parameters of our flow cytometer. Antibodies used were: brilliant violet 421-anti-CD90 (BD 562556, 1:20), alexa fluor 488-anti-CD34 (Biolegend, San Diego, CA 343518, 1:20), brilliant violet 711-anti-CD38 (BD 563965, 1:20), allophycocyanin-anti-CD45RA (BD 550855, 1:5), phycoerythrin-anti-CD135 (BD 558996, 1:5), and allophycocyanin-cyanine 7-anti-CD10 (Biolegend 312212, 1:20). Live/dead staining was performed immediately prior to sorting using Zombie Aqua Fixable Viability Dye (Biolegend). Cells were sorted on a FACSAria II instrument (BD), and data analysis was performed in FlowJo v10.0.8.

**Data processing and availability.** Paired-end reads were mapped to the hg38 human transcriptome (Gencode v24) using STAR v2.4.2a(2). Aligned reads are available through dbGaP (phs001845.v1.p1). Gene expression levels were quantified as transcript-per-million (TPM) in RSEM(3). Cells with at least 1000 expressed genes (defined by TPM>1) and genes expressed in at least 50 single cells were kept. This resulted in 11094 genes in 583 single cells. The same set of 11094 genes was analyzed to derive lineage signature genes from 100 cell libraries made from FACS-purified CD34+ subsets.

**Gene selection based on bulk expression data.** We used the Gini index(4) to identify cell type-specific genes from HSC, MPP, CLP, CMP, MEP, and GMP 100 cell libraries. We first calculated maximum TPM value of each gene, and genes with maximum value lower than the 20-quantile of all maximum values were filtered out because those genes could have high Gini index due to their low expression. We then identified the top 500 high Gini index genes for each of the biological ($n$=2) and technical ($n$=2) replicates for each cell type. The cell type specific gene signatures were chosen as the intersection of high Gini genes across all replicates for each cell type.

**tSNE analysis.** We divided TPM values by 10 to better reflect the complexity of single cell libraries which is estimated to be ~100,000 transcripts(5). The data were log2 transformed (log2(TPM/10

+1)). The expression of the 79 genes identified by bulk data across the 583 single cells was used for Principal Component Analysis (PCA) in the Seurat Package in R(6). Using a jackstraw approach implemented in the Seurat package with num.replicate = 200 and each time randomly permuting three genes, the top four principal components (PCs) were identified as significant (*p*-value < $1x10^{-4}$). To aid visualization, these top four PCs, were subject to t-distributed Stochastic Neighbor Embedding (t-SNE)(7) analysis in Seurat with 2000 iterations.

**Clustering analysis.** The tSNE coordinates were used for partitioning around medoids (PAM), a more robust version of *k*-means clustering implemented in the "cluster" package in R with default parameters (https://stat.ethz.ch/R-manual/R-devel/library/cluster/html/pam.html). To determine the optimal *k*, we assessed the average Silhouette value(8) for each clustering result (from *k*=2 to *k*=10) and selected *k* =5, which gave the largest mean Silhouette value.

**Differential gene expression and pathway analysis.** Differential gene expression analysis was performed on SDS versus normal cells in each cluster (and in all clusters combined) using the MAST package in R(9) *p*-values were adjusted for multiple testing using the "p.adjust" function in R with "fdr" method(10) We focused on genes with an FDR adjusted *p*-value < 0.05 and |log2(fold change)| >1 in at least one cluster. Enriched pathways and functions were determined in Ingenuity Pathway Analysis (Qiagen) using the 11094 detected genes as the reference gene set. Split violin plots were generated using the ''vioplot'' package and "vioplot2" function in R.

**Immunofluorescent staining and imaging.** Primary BM-derived mononuclear cells were cultured for 30-32h in StemSpan SFEM II (Stem Cell Technologies) supplemented with 100 ng/mL of SCF, TPO, Flt3L and 20 ng/mL of IL-3 (PreproTech, Rocky Hill, NJ). CD34+ cells were sorted using CD34 Microbeads (Millitenyi) according to manufacturer's protocol, and allowed to recover in culture medium for 14-16h, plus an additional 2h in the presence of 0.6μg/ml AVID200 for

relevant samples. 25,000-50,000 cells were spun onto coverslips (ES0117580, Azer Scientific, Morgantown, PA) using a cytospin instrument (Thermo Shandon) at 380rpm for 5min; fixed with 4% PFA in 1X PBS for 10min at room temperature (RT); washed 2X with 1X PBS; permeabilized with 0.3% TritonX in 1X PBS solution for 10min at RT; washed 2X with 1X PBS; blocked in 10% FBS, 0.1% NP40 in 1X PBS for 1h at RT; incubated with 1:250 anti-p-smad2 (Invitrogen, 44-244G) in blocking solution for 14-16h at 4°C; washed 3X with 0.1% NP40 in 1X PBS at RT for 10min; incubated with 1:1,000 diluted anti-rabbit IgG-Alexa488 antibody (Invitrogen, A21206) in blocking solution for 1h at RT; and washed 3X with 0.1% NP40 in 1X PBS at RT for 10min. Stained coverslips were mounted on glass slides with VectaShield with DAPI (H-1200, Vector Laboratories, Burlingame, CA) diluted 1:1 in VectaShield without DAPI (H-1000). Slides were imaged on a LeicaSP5 confocal microscope with constant laser power (30% for DAPI, 70% for Alexa488) and identical resolution, offset, and gain settings for all slides. Z stack images were captured with 40-80μm step range, and the plane with the best nuclear representation was analyzed using Fiji software. Background was calculated using four randomly selected empty regions for each image. Mean signal intensity for p-SMAD2 (Alexa Fluor-488) was calculated within each nucleus, and background signal was subtracted.

**Colony formation assays.** Primary BM-derived mononuclear cultured for 24h in StemSpan SFEM II (Stem Cell Technologies) supplemented with 100 ng/mL of SCF, TPO, Flt3L and 20 ng/mL of IL-3 (PreproTech, Rocky Hill, NJ). Cells were resuspended at 10,000 cells/mL for control and 20,000 cells/mL for SDS in the presence or absence of 0, 0.25, 0.5, 1, or 5 μM SD208 (Tocris, Bristol, UK), and incubated for 1hr at 37˚C/5% $CO_2$. 200 μL of cell suspension was mixed with 3 mL of Methocult H4434 (Stem Cell Technologies), and 1 mL was plated in triplicate in a SmartDish 6-well plate (Stem Cell Technologies). After 14 days of growth at 37˚C/5% $CO_2$, colonies were manually scored by two independent, blinded investigators using standard criteria(11).

**SOMAscan proteomic analysis.** SOMAscan (SomaLogic, Boulder, CO) was performed on 50 ul of EDTA-plasma from six patients and six normal controls at the BIDMC Genomics, Proteomics, Bioinformatics and Systems Biology Center. Samples were prepared and run using the SOMAscan Assay Kit for Human Plasma, 1.3k (cat. # 900-00011), according to the manufacturer's protocol. Five pooled controls and one no-protein buffer control provided in the kit were run in parallel with the samples. Median normalization and calibration of the data was performed according to the standard quality control protocols at SomaLogic. All samples passed the established quality control criteria. Proteins with $p$-values<0.01 were analyzed. Benjamini-Hochberg adjusted $p$-values are reported in Extended Data Table 4.

**Statistics.** In figure 2a, statistical significance was determined by the chi-squared test; the frequency of cells in each cluster was compared between SDS and normal. In Figure 2b, 3c, 4d, and Extended Data Figure 3b, statistical significance was determined by two-way ANOVA with Holm-Sidak's multiple correction test in GraphPad Prism 7. In Figure 2b, the frequency of cells was compared between SDS and normal cells within each cluster. In Figure 3c, log2 expression was compared between SDS and normal cells within each cluster. In Figure 4d and Supplementary Figure 3b, relative colony number was compared between each drug dose and the 0uM treatment. In Figure 4b and 4c, statistical significance was determined by one-way ANOVA with Holm-Sidak's multiple correction test in GraphPad Prism 7; SDS samples were compared to normal samples that were stained and imaged concurrently.

**Study approval.** Subjects provided written, informed consent for protocols approved by the institutional review board of Boston Children's Hospital (Boston, MA) and Dana-Farber Cancer Institute (Boston, MA), in accordance with the Declaration of Helsinki's Ethical Principles of

Medical Research Involving Human Subjects. All subjects provided informed consent prior to their participation in the study.

## SUPPLEMENTAL METHODS REFERENCES


1. Laurenti E, et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat Immunol.* 2013;14(7):756-63.
2. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.
3. Li B, and Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
4. Jiang L, Chen H, Pinello L, and Yuan GC. GiniClust: detecting rare cell types from single-cell gene expression data with Gini index. *Genome Biol.* 2016;17(1):144.
5. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science.* 2016;352(6282):189-96.
6. Satija R, Farrell JA, Gennert D, Schier AF, and Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol.* 2015;33(5):495-502.
7. Maaten LJPVD, and Hinton GE. *Visualizing High-Dimensional Data using t-SNE.* 2008.
8. Rousseeuw P. *Rousseeuw, P.J.: Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Comput. Appl. Math. 20, 53-65.* 1987.
9. Finak G, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* 2015;16:278.
10. Benjamini Y, and Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological).* 1995;57(1):289-300.

11. Eaves C., Lambie K. Atlas of Human Hematopoietic Colonies: An Introduction to the Recognition of Colonies Produced by Human Hematopoietic Progenitor Cells Cultured in Methylcellulose Media. *StemCell Technologies.* 1995.