

# TGF $\beta$ 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 (TGF $\beta$ ) were dysregulated in SDS hematopoietic stem cells and multipotent progenitors, but not in lineage-committed progenitors. TGF $\beta$  inhibitors (AVID200 and SD208) increased hematopoietic colony formation of SDS patient BM. Finally, TGF $\beta$ 3 and other TGF $\beta$  pathway members were elevated in SDS patient blood plasma. These data establish the TGF $\beta$  pathway as a novel candidate biomarker and therapeutic target in SDS and translate insights from single cell biology into a potential therapy.

Find the latest version:

<http://jci.me/125375/pdf>



1 **TGF $\beta$  signaling underlies hematopoietic dysfunction and bone marrow failure in**  
2 **Shwachman-Diamond Syndrome**

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

7  
8 <sup>1</sup>Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, <sup>2</sup>Department of  
9 Medicine, Harvard Medical School, Boston, MA 02115, USA. <sup>3</sup>Department of Biostatistics and  
10 Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. <sup>4</sup>Department of  
11 Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. <sup>5</sup>Division of  
12 Hematology/Oncology, Boston Children's Hospital and Dana Farber Cancer Institute, Boston, MA  
13 02115, USA. <sup>6</sup>Broad Institute of Harvard and MIT, Cambridge, MA 02141, USA. <sup>7</sup>Division of Bone  
14 Marrow Transplantation and Immune Deficiency, Cincinnati Children's Hospital Medical Center,  
15 Cincinnati, Ohio, 45229, USA. <sup>8</sup>Beth Israel Deaconess Medical Center Genomics, Proteomics,  
16 Bioinformatics and Systems Biology Center, Division of Interdisciplinary Medicine and  
17 Biotechnology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA, 02215.

18  
19 \*These authors contributed equally to this work

20  
21 †Corresponding author:

22 **Carl D. Novina**  
23 Dana-Farber Cancer Institute, Dana 1420B  
24 450 Brookline Avenue, Boston, MA 02215  
25 [carl\\_novina@dfci.harvard.edu](mailto:carl_novina@dfci.harvard.edu)  
26 Tel: 617-582-7961  
27 Fax: 617-582-7962

28  
29 Conflict of interest statement: The authors have declared that no conflict of interest exists.

32 **ABSTRACT**

33 Shwachman-Diamond Syndrome (SDS) is a rare and clinically-heterogeneous bone  
34 marrow (BM) failure syndrome caused by mutations in the Shwachman-Bodian-Diamond  
35 Syndrome (*SBDS*) gene. Although SDS was described over 50 years ago, the molecular  
36 pathogenesis is poorly understood due, in part, to the rarity and heterogeneity of the affected  
37 hematopoietic progenitors. To address this, we used single cell RNA sequencing to profile scant  
38 hematopoietic stem and progenitor cells from SDS patients. We generated a single cell map of  
39 early lineage commitment and found that SDS hematopoiesis was left-shifted with selective loss  
40 of granulocyte-monocyte progenitors. Transcriptional targets of transforming growth factor-beta  
41 ( $TGF\beta$ ) were dysregulated in SDS hematopoietic stem cells and multipotent progenitors, but not  
42 in lineage-committed progenitors.  $TGF\beta$  inhibitors (AVID200 and SD208) increased  
43 hematopoietic colony formation of SDS patient BM. Finally,  $TGF\beta 3$  and other  $TGF\beta$  pathway  
44 members were elevated in SDS patient blood plasma. These data establish the  $TGF\beta$  pathway  
45 as a candidate biomarker and therapeutic target in SDS and translate insights from single cell  
46 biology into a potential therapy.

47

48 **INTRODUCTION**

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

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

67

## 68 **COMBINED RESULTS AND DISCUSSION**

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

82 transcriptional signatures of lineage commitment. This approach is well suited to capture cells  
83 along the CD34<sup>+</sup> differentiation spectrum, which is a subject of evolving understanding in human  
84 BM(10, 11).

85 A major challenge for studying a rare patient population is that biological variables and  
86 batch effects can obscure disease signatures. To classify single cells with respect to  
87 hematopoietic lineage commitment (and not other unrelated variables), we designed a supervised  
88 dimensionality reduction analysis. Specifically, we performed bulk RNA-seq on FACS-purified  
89 HSPC subpopulations(12) from normal BM to derive an mRNA expression signature that  
90 distinguished HSCs, multipotent progenitors (MPPs), common myeloid progenitors (CMPs),  
91 multilymphoid progenitors (MLPs), granulocyte-monocyte progenitors (GMPs), and  
92 megakaryocyte-erythroid progenitors (MEPs) (Supplementary Figure 1). We then analyzed this  
93 signature in single cell RNA-seq datasets from both normal and SDS BM to predict the identity of  
94 each cell. Data were visualized using t-distributed stochastic neighbor embedding (tSNE; Figure  
95 1, Supplementary Table 2)(13). For simplicity, SDS cells are masked in Figure 1.

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

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

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

126 We next compared gene expression between normal and SDS cells within each cluster  
127 except for GMP, which was excluded due to the low number of GMP in untreated SDS patients.  
128 Overall, 1680 genes were differentially expressed in at least one cluster (FDR < 0.05,  $|\log_2(\text{fold}$   
129  $\text{change})| > 1$ , Supplementary Table 3). Strikingly, 81.5% of all differentially expressed genes were  
130 unique to either HSC/MPP or CMP (Figure 3a). An additional 9.8% were commonly affected in  
131 HSC/MPP and CMP, but not in MLP/CLP or MEP. Overall, these data demonstrate that despite  
132 the general biochemical functions of the SBDS protein, *SBDS* mutations differentially affect the

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

135 The Inflammatory Response was enriched among differentially-expressed genes in both  
136 the HSC/MPP and CMP clusters (maximum  $p$ -value  $4.98 \times 10^{-5}$  and  $1.18 \times 10^{-3}$ , respectively).  
137 However, the genes contributing to the enrichment differed between the clusters (Figure 3b).  
138 Transforming growth factor-beta (TGF $\beta$ ) was the top regulator predicted for the HSC/MPP  
139 inflammatory response ( $p=4.03 \times 10^{-15}$ ,  $z$ -score=0.891). It was also a significant upstream regulator  
140 among all differentially-expressed genes in HSC/MPP ( $p=1.27 \times 10^{-2}$ ,  $z$ -score=0.417).  
141 Dysregulation of these TGF $\beta$  targets was most significant in HSC/MPP, with lesser or no effect in  
142 other HSPC populations (Figure 3c). TGF $\beta$  induces context-dependent effects on cell growth,  
143 survival, inflammation, and extracellular matrix. TGF $\beta$ 1 and TGF $\beta$ 3 have potent growth inhibitory  
144 effects on HSC(15-17). Thus, we hypothesized that activation of TGF $\beta$  in SDS HSC/MPP may  
145 contribute to BM failure in SDS.

146 To confirm activation of TGF $\beta$  signaling in SDS BM, we assessed TGF $\beta$  dependent  
147 phosphorylation and nuclear translocation of the transcriptional co-activator protein Mothers  
148 Against Decapentaplegic Homolog 2 (p-SMAD2). A subset of CD34+ cells from SDS BM had  
149 elevated levels of nuclear p-SMAD2 that were outside the normal range (Figure 4a, b). Treating  
150 SDS cells with AVID200, a decoy receptor trap designed to specifically neutralize TGF $\beta$ 1 and  
151 TGF $\beta$ 3, reduced the p-SMAD2 signal. The same trend was observed to varying degrees in two  
152 additional sample pairs (Figure 4c). These data are consistent with our single cell RNA-seq  
153 analysis demonstrating selective activation of the TGF $\beta$  pathway in the HSC/MPP subset of SDS  
154 CD34+ cells.

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

158 (Supplementary Table 1) in methylcellulose supplemented with AVID200 and SD208, which  
159 inhibits TGF $\beta$ R1 kinase activity(19). Both compounds improved hematopoietic colony formation  
160 in SDS patient samples, but not in normal donor controls (Figure 4d, Supplementary Figure 3b,  
161 Supplementary Table 4). Taken together, our data support a model in which activation of TGF $\beta$ R1  
162 kinase activity by TGF $\beta$ 1 and/or TGF $\beta$ 3 lead to increased concentration of p-SMAD2 in the  
163 nucleus and transcription of inflammatory response genes in SDS HSC/MPP (Figure 4e).

164 To determine whether SDS patients express elevated levels of TGF $\beta$  ligands, we  
165 screened blood plasma proteins from six SDS patients and six normal controls (Supplementary  
166 Table 1) using SOMAscan; a highly-sensitive, aptamer-based proteomic platform(20). TGF $\beta$ 3 was  
167 significantly upregulated in SDS patient plasma, along with several other factors that were  
168 annotated to a network of TGF $\beta$ -associated factors (Figure 4f, Supplementary Figure 4). These  
169 and other dysregulated plasma proteins that were common across clinically-heterogeneous  
170 patients could serve as diagnostic biomarkers for SDS (Supplementary Table 5). Further studies  
171 are required to determine the levels of TGF $\beta$ 3 in the BM compartment and identify the cell types  
172 that produce it.

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



183 syndrome, cancer, and pulmonary fibrosis, among others(30). Our work suggests that TGF $\beta$ 1/3  
184 inhibition by an agent such as AVID200 could be an effective therapy across clinically-  
185 heterogeneous SDS patients and different marrow failure disorders.

186

## 187 **METHODS**

188 Detailed methods are provided as Supplementary Material.

189

## 190 **AUTHOR CONTRIBUTIONS**

191 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  
192 K.C.M. collected patient samples and clinical information; C.E.J, M.R-G., O.V-B., and D.D.T.  
193 performed experiments; A.Sa., L.J., and S.Y. performed computational analyses; C.E.J., A.Sa.,  
194 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  
195 manuscript; all authors provided critical reviews of the manuscript.

196

## 197 **ACKNOWLEDGMENTS**

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

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

211

## 212 REFERENCES

- 213 1. Menne TF, et al. The Shwachman-Bodian-Diamond syndrome protein mediates  
214 translational activation of ribosomes in yeast. *Nature genetics*. 2007;39:486-95.
- 215 2. Ganapathi KA, et al. The human Shwachman-Diamond syndrome protein, SBDS,  
216 associates with ribosomal RNA. *Blood*. 2007;110:1458-65.
- 217 3. Finch AJ, et al. Uncoupling of GTP hydrolysis from eIF6 release on the ribosome causes  
218 Shwachman-Diamond syndrome. *Genes & development*. 2011;25:917-29.
- 219 4. Burwick N, Coats SA, Nakamura T, and Shimamura A. Impaired ribosomal subunit  
220 association in Shwachman-Diamond syndrome. *Blood*. 2012;120:5143-52.
- 221 5. Austin KM, et al. Mitotic spindle destabilization and genomic instability in Shwachman-  
222 Diamond syndrome. *J Clin Invest*. 2008;118(4):1511-8.
- 223 6. Huang JN, and Shimamura A. Clinical spectrum and molecular pathophysiology of  
224 Shwachman-Diamond syndrome. *Current opinion in hematology*. 2010;18:30-5.
- 225 7. Myers KC, et al. Variable clinical presentation of Shwachman-Diamond syndrome: update  
226 from the North American Shwachman-Diamond Syndrome Registry. *The Journal of*  
227 *pediatrics*. 2014;164:866-70.
- 228 8. Ramskold D, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual  
229 circulating tumor cells. *Nat Biotechnol*. 2012;30(8):777-82.
- 230 9. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, and Sandberg R. Smart-  
231 seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods*.  
232 2013;10(11):1096-8.

- 233 10. Notta F, Doulatov S, Laurenti E, Poepl A, Jurisica I, and Dick JE. Isolation of single  
234 human hematopoietic stem cells capable of long-term multilineage engraftment. *Science*  
235 *(New York, NY)*. 2011;333:218-21.
- 236 11. Velten L, et al. Human haematopoietic stem cell lineage commitment is a continuous  
237 process. *Nat Cell Biol*. 2017;19(4):271-81.
- 238 12. Laurenti E, et al. The transcriptional architecture of early human hematopoiesis identifies  
239 multilevel control of lymphoid commitment. *Nature immunology*. 2013;14:756-63.
- 240 13. van der Maaten L, Hinton, G. Visualizing High-Dimensional Data Using t-SNE. *Journal of*  
241 *Machine Learning Research*. 2008(9):2579-605.
- 242 14. Thomas J, Liu F, and Link DC. Mechanisms of mobilization of hematopoietic progenitors  
243 with granulocyte colony-stimulating factor. *Curr Opin Hematol*. 2002;9(3):183-9.
- 244 15. Hatzfeld J, et al. Release of early human hematopoietic progenitors from quiescence by  
245 antisense transforming growth factor beta 1 or Rb oligonucleotides. *J Exp Med*.  
246 1991;174(4):925-9.
- 247 16. Scandura JM, Boccuni P, Massague J, and Nimer SD. Transforming growth factor beta-  
248 induced cell cycle arrest of human hematopoietic cells requires p57KIP2 up-regulation.  
249 *Proc Natl Acad Sci U S A*. 2004;101(42):15231-6.
- 250 17. Challen GA, Boles NC, Chambers SM, and Goodell MA. Distinct hematopoietic stem cell  
251 subtypes are differentially regulated by TGF-beta1. *Cell Stem Cell*. 2010;6(3):265-78.
- 252 18. Dror Y, and Freedman MH. Shwachman-Diamond syndrome: An inherited preleukemic  
253 bone marrow failure disorder with aberrant hematopoietic progenitors and faulty marrow  
254 microenvironment. *Blood*. 1999;94(9):3048-54.
- 255 19. Uhl M, Aulwurm S, Wischhusen J, Weiler M, Ma JY, Almirez R, et al. SD-208, a novel  
256 transforming growth factor beta receptor I kinase inhibitor, inhibits growth and  
257 invasiveness and enhances immunogenicity of murine and human glioma cells in vitro and  
258 in vivo. *Cancer Res*. 2004;64(21):7954-61.

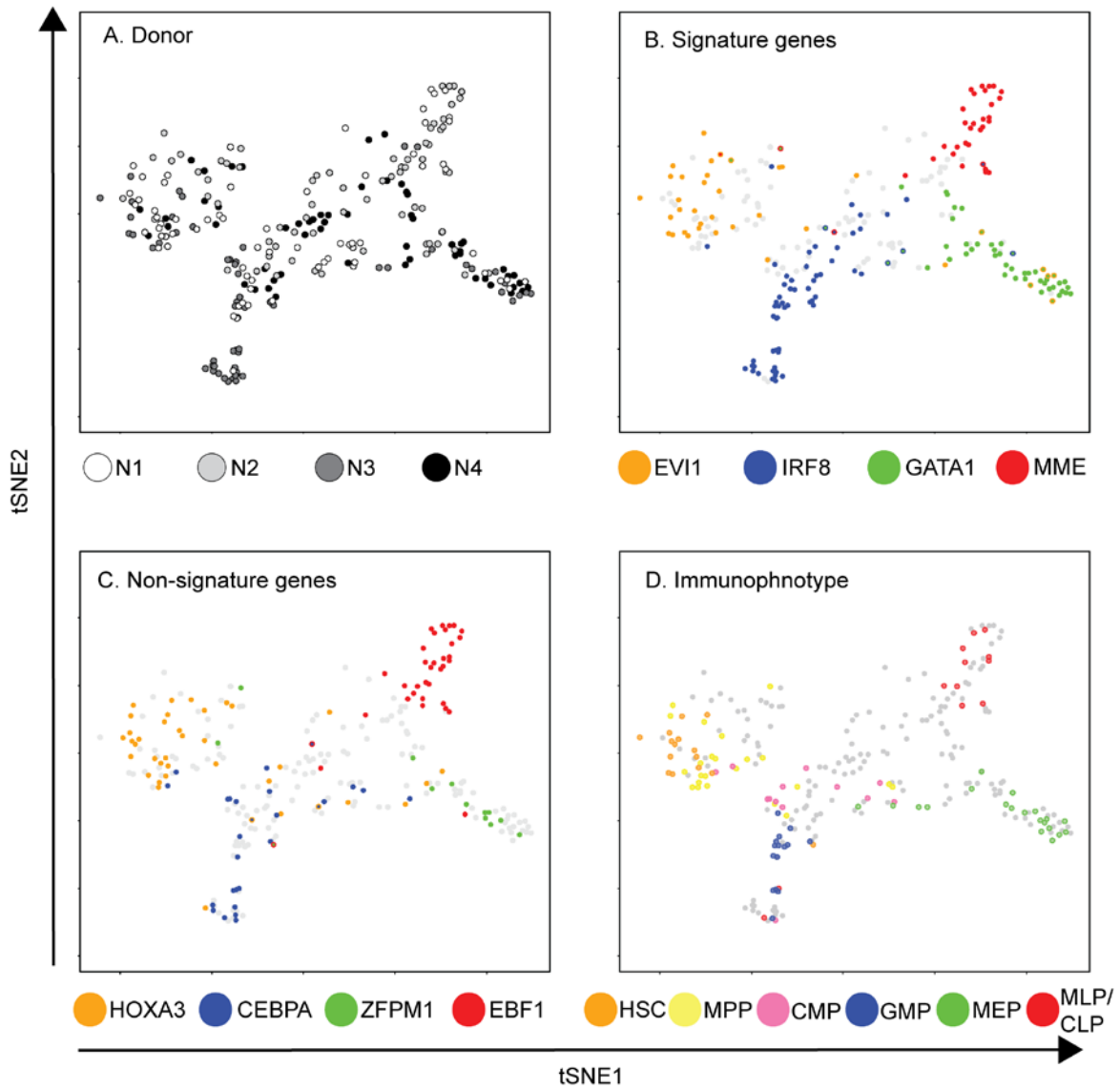
- 259 20. Gold L, Walker JJ, Wilcox SK, and Williams S. Advances in human proteomics at high  
260 scale with the SOMAscan proteomics platform. *N Biotechnol.* 2012;29(5):543-9.
- 261 21. Turlakis ME, Zhong J, Gandhi R, Zhang S, Chen L, Durie PR, et al. Deficiency Of Sbds  
262 In The Mouse Pancreas Leads To Features Of Shwachman-Diamond Syndrome, With  
263 Loss Of Zymogen Granules. *Gastroenterology.* 2012.
- 264 22. Zambetti NA, Bindels EMJ, Van Strien PMH, Valkhof MG, Adisty MN, Hoogenboezem RM,  
265 et al. Deficiency of the ribosome biogenesis gene Sbds in hematopoietic stem and  
266 progenitor cells causes neutropenia in mice by attenuating lineage progression in  
267 myelocytes. *Haematologica.* 2015;100:1285-93.
- 268 23. Tulpule A, Kelley JM, Lensch MW, McPherson J, Park IH, Hartung O, et al. Pluripotent  
269 Stem Cell Models of Shwachman-Diamond Syndrome Reveal a Common Mechanism for  
270 Pancreatic and Hematopoietic Dysfunction. *Cell stem cell.* 2013;12:727-36.
- 271 24. Villani AC SR, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A,  
272 Zheng S, Lazo S, Jardine L, Dixon D, Stephenson E, Nilsson E, Grundberg I, McDonald  
273 D, Filby A, Li W, De Jager PL, Rozenblatt-Rosen O, Lane AA, Haniffa M, Regev A,  
274 Hacohen N. Single-cell RNA-seq reveals new types of human blood dendritic cells,  
275 monocytes, and progenitors. *Science.* 2017;356(6335).
- 276 25. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting  
277 the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*  
278 *(New York, NY).* 2016;352:189-96.
- 279 26. Kumar RM, Cahan P, Shalek AK, Satija R, DaleyKeyser AJ, Li H, et al. Deconstructing  
280 transcriptional heterogeneity in pluripotent stem cells. *Nature.* 2014;516:56-61.
- 281 27. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of human  
282 brain transcriptome diversity at the single cell level. *Proceedings of the National Academy*  
283 *of Sciences.* 2015;112:201507125.

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

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

290 30. Herberitz S, Sawyer JS, Stauber AJ, Gueorguieva I, Driscoll KE, Estrem ST, et al. Clinical  
291 development of galunisertib (LY2157299 monohydrate), a small molecule inhibitor of  
292 transforming growth factor-beta signaling pathway. *Drug Des Devel Ther*. 2015;9:4479-  
293 99.

294  
295  
296  
297  
298  
299  
300



302

303 **Figure 1. Supervised dimensionality reduction maps lineage commitment of CD34+ cells**

304 **from healthy donors.** tSNE plot of hematopoietic lineage commitment was derived from an

305 empirically-defined gene expression signature. Shown here are cells from four normal donors

306 ( $n^{N1}=70$ ,  $n^{N2}=58$ ,  $n^{N3}=69$ ,  $n^{N4}=59$ ,  $n^{total}=256$ ). Cells are colored based on (a) donor identity, (b)

307 mRNA expression of selected signature genes, (c) mRNA expression of lineage-restricted genes

308 reported elsewhere<sup>12</sup>, and (d) immunophenotypes. For (b,c), color indicates TPM>1 for the

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

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

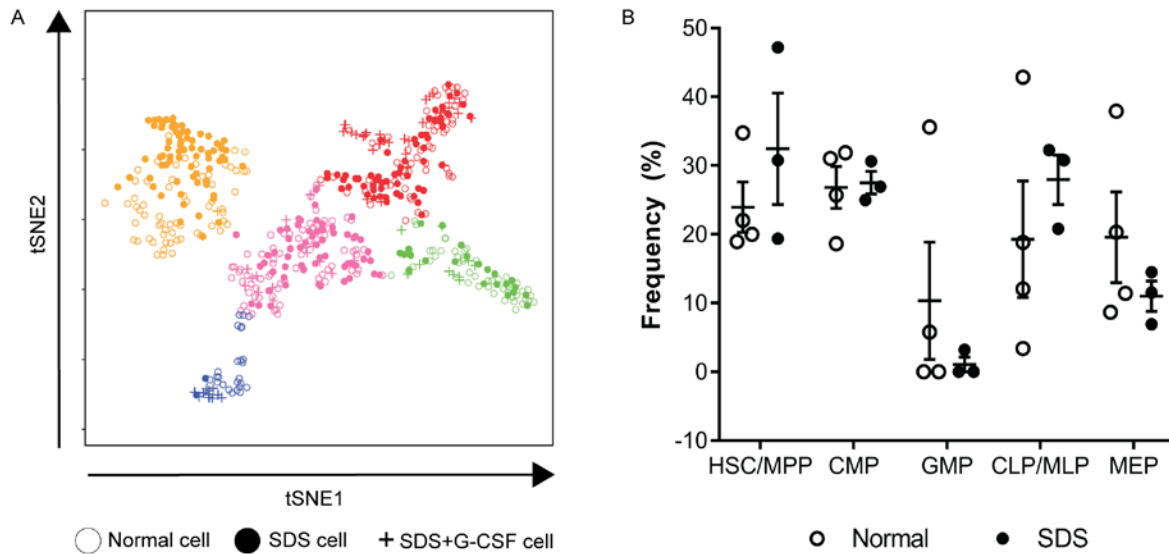
331

332

333

334

335 **FIGURE 2**



336

337

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

347

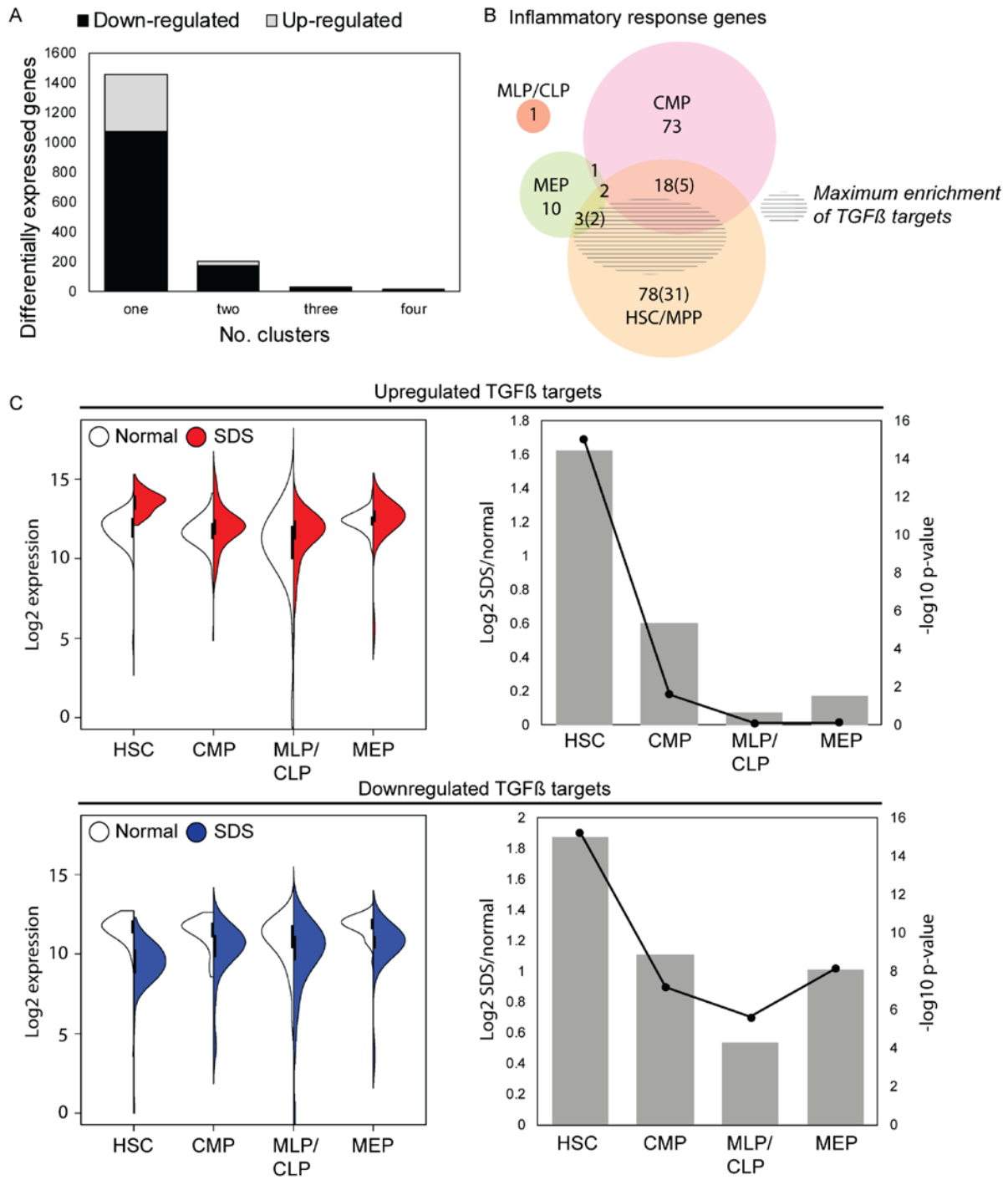
348

349

350

351





353

354

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

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

357 each cluster – HSC/MPP, CMP, MLP/CLP, or MEP. To aid biological interpretation, this gene set  
358 was filtered to focus on genes with FDR adjusted p-value < .05 and log<sub>2</sub>(fold change) >|1| in at  
359 least one cluster. Plotted are the number of genes that were either up- or down-regulated in one,  
360 two, three or four clusters. GMP was excluded due to the paucity of SDS GMP. Inset pie chart  
361 shows the proportion of differentially expressed genes in each cluster. (b) Venn diagram of  
362 differentially expressed genes in each cluster that were annotated to the “Inflammatory Response”  
363 function in Ingenuity Pathway Analysis. The shaded region shows the area of maximal enrichment  
364 of TGFβ targets (p=4.03x10<sup>-15</sup>). (c) Left: split violin for the summed expression of 25 upregulated  
365 TGFβ targets and 52 down-regulated TGFβ targets in SDS HSC/MPP. Right: Log<sub>2</sub> fold changes  
366 (primary axis, bars) and p-values (secondary axis, lines) for the gene sets plotted in ‘b’.  
367 Significance was determined by two-way ANOVA, with Holm-Sidak’s multiple comparisons test.

368

369

370

371

372

373

374

375

376

377

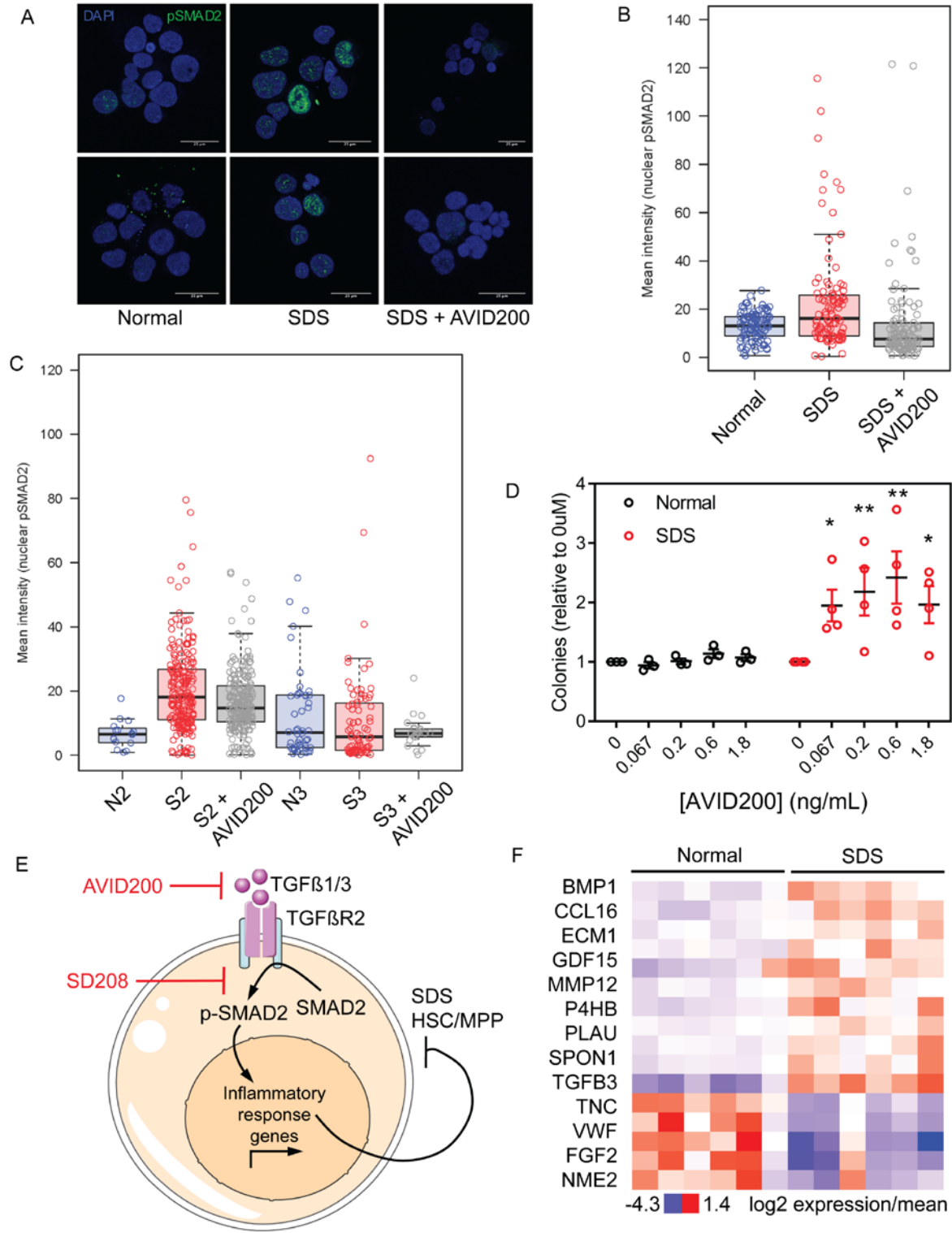
378

379

380

381

382 **FIGURE 4**



383

384

385 **Figure 4. TGF $\beta$  pathway activation through TGF $\beta$ R1 suppresses hematopoiesis in SDS BM**  
386 **progenitors.** a) Representative images showing DAPI and phospho-SMAD2 staining of primary  
387 BM CD34<sup>+</sup> cells from adult normal donor BM and pediatric SDS BM, either untreated or treated  
388 with AVID200. b) Mean intensity of phospho-SMAD2 staining in individual CD34<sup>+</sup> nuclei from  
389 samples depicted in panel (a). Significance was determined by two-way ANOVA, with Holm-  
390 Sidak's multiple comparisons test. Error bars= minimum and maximum values, excluding outliers  
391 that exceed median+1.5\*IQR. \*\*p<0.01, \*\*\*p<0.001. c) Mean intensity of phospho-SMAD2  
392 staining in individual CD34<sup>+</sup> nuclei in two additional pairs of SDS and normal donor BM samples.  
393 Error bars= minimum and maximum values, excluding outliers that exceed median+1.5\*IQR.  
394 \*\*p<0.01, \*\*\*\*p<0.0001. d) Number of colonies formed by adult normal donor and pediatric SDS  
395 patient BM-derived mononuclear cells with increasing concentrations of AVID200, normalized to  
396 the 0uM treatment. Significance was determined relative to the 0uM treatment by two-way  
397 ANOVA, with Holm-Sidak's multiple comparisons test. Error bars=SEM. \*p<0.05, \*\*p<0.01. e)  
398 Model for the role of TGF $\beta$  signaling in SDS BM failure. TGF $\beta$ 1 and/or TGF $\beta$ 3 ligands (targets of  
399 AVID200 inhibitor) activate signaling through the TGF $\beta$ R1 receptor (target of SD208 inhibitor) on  
400 SDS HSC/MPP. Our data suggest that TGF $\beta$  ligands are primarily derived from a CD34<sup>-</sup> cell type  
401 in BM because TGF $\beta$  ligand mRNAs were not detected in CD34<sup>+</sup> HSPC. Increased TGF $\beta$ R1  
402 signaling leads to increased concentrations of nuclear phospho-SMAD2 and transcription of  
403 inflammatory response genes, which impairs HSC/MPP function. This model predicts that  
404 therapeutic inhibition of TGF $\beta$  signaling in HSC/MPP will improve hematopoietic function in SDS  
405 patients. f) Expression of extracellular proteins annotated to a TGF $\beta$  network that was enriched  
406 among dysregulated proteins in SDS patient plasma. Asterisks indicate TGF $\beta$  family ligands.  
407  
408  
409

## 410 SUPPLEMENTAL METHODS

411 **Sample processing.** For scRNA-seq of all SDS samples, and normal donors N1 and N2: 7-20  
412 ml of fresh BM were diluted to 35ml in MACS buffer (PBS/2mM EDTA/0.5% BSA), layered onto  
413 15ml Ficoll-paque (GE Healthcare, Uppsala, Sweden), and spun for 30 min at 1400 rpm and 20°C  
414 with no brakes. Mononuclear cells were collected from the interface, washed once, pelleted for 5  
415 min at 1200 rpm and 20°C, and resuspended at 40 ul per  $10^7$  cells in MACS buffer + 1 ul/ml  
416 RNaseOUT (Thermo Fisher Scientific, Waltham, MA, USA). CD34+ cells were positively selected  
417 on an AutoMACS instrument using the Indirect CD34 MicroBead Kit (Miltenyi, Bergisch Gladbach,  
418 Germany), and singulated on the C1 Instrument (Fluidigm, San Francisco, CA, USA). cDNA  
419 libraries were prepared using the SMARTer Ultra Low RNA Kit (Clontech, Mountain View, CA,  
420 USA). For samples N3 and N4, protocol conditions were modified to ascertain immunophenotypes  
421 from single cells, and in accordance with the newest available methods. For these samples: red  
422 blood cells were lysed with ammonium chloride (Stem Cell Technologies, Vancouver, CA).  
423 Mononuclear cells were pelleted for 5 min at 1200 rpm and 20°C, washed twice, and resuspended  
424 in PBS + 1 ul/ml RNaseOUT. Cells were stained as described below. Single CD34+ cells were  
425 sorted into 5ul TCL buffer (Qiagen, Hilden, Germany) in 96 well plates using a FACS Aria II  
426 instrument (BD, Franklin Lakes, NJ, USA) on index mode. Two technical replicates of 100 cells  
427 from each gated CD34+ subset – HSC, MPP, MLP, CMP, GMP, MEP – were sorted into 5 ul TCL  
428 buffer in separate 96 well plates. cDNA libraries were prepared using the SMART-Seq v4 Ultra  
429 Low RNA Kit (Clontech). Libraries from all samples were sequenced on a HiSeq 2500 Instrument  
430 (Illumina, San Diego, CA) to a read depth of ~3 M paired-end, 25 bp reads per single cell, or ~12  
431 M paired-end, 25 bp reads per 100 cells.

432 **Antibodies and staining.** Cells were stained at a density of  $1 \times 10^6$  per 100 ul in PBS + 1 ul/ml  
433 RNaseOUT because staining buffers contain proteins that can inhibit SMARTer-seq (Clontech)  
434 cDNA synthesis reactions. The staining panel was adapted from an analysis of human cord blood

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

442

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

450

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

458

459 **tSNE analysis.** We divided TPM values by 10 to better reflect the complexity of single cell libraries  
460 which is estimated to be ~100,000 transcripts(5). The data were log2 transformed ( $\log_2(TPM/10$

461 +1)). The expression of the 79 genes identified by bulk data across the 583 single cells was used  
462 for Principal Component Analysis (PCA) in the Seurat Package in R(6). Using a jackstraw  
463 approach implemented in the Seurat package with num.replicate = 200 and each time randomly  
464 permuting three genes, the top four principal components (PCs) were identified as significant ( $p$ -  
465 value  $< 1 \times 10^{-4}$ ). To aid visualization, these top four PCs, were subject to t-distributed Stochastic  
466 Neighbor Embedding (t-SNE)(7) analysis in Seurat with 2000 iterations.

467

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

473

474 **Differential gene expression and pathway analysis.** Differential gene expression analysis was  
475 performed on SDS versus normal cells in each cluster (and in all clusters combined) using the  
476 MAST package in R(9)  $p$ -values were adjusted for multiple testing using the “p.adjust” function in  
477 R with “fdr” method(10) We focused on genes with an FDR adjusted  $p$ -value  $< 0.05$  and  $|\log_2(\text{fold}$   
478  $\text{change})| > 1$  in at least one cluster. Enriched pathways and functions were determined in Ingenuity  
479 Pathway Analysis (Qiagen) using the 11094 detected genes as the reference gene set. Split violin  
480 plots were generated using the “vioplot” package and “vioplot2” function in R.

481

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

487 relevant samples. 25,000-50,000 cells were spun onto coverslips (ES0117580, Azer Scientific,  
488 Morgantown, PA) using a cytospin instrument (Thermo Shandon) at 380rpm for 5min; fixed with  
489 4% PFA in 1X PBS for 10min at room temperature (RT); washed 2X with 1X PBS; permeabilized  
490 with 0.3% TritonX in 1X PBS solution for 10min at RT; washed 2X with 1X PBS; blocked in 10%  
491 FBS, 0.1% NP40 in 1X PBS for 1h at RT; incubated with 1:250 anti-p-smad2 (Invitrogen, 44-  
492 244G) in blocking solution for 14-16h at 4°C; washed 3X with 0.1% NP40 in 1X PBS at RT for  
493 10min; incubated with 1:1,000 diluted anti-rabbit IgG-Alexa488 antibody (Invitrogen, A21206) in  
494 blocking solution for 1h at RT; and washed 3X with 0.1% NP40 in 1X PBS at RT for 10min. Stained  
495 coverslips were mounted on glass slides with VectaShield with DAPI (H-1200, Vector  
496 Laboratories, Burlingame, CA) diluted 1:1 in VectaShield without DAPI (H-1000). Slides were  
497 imaged on a LeicaSP5 confocal microscope with constant laser power (30% for DAPI, 70% for  
498 Alexa488) and identical resolution, offset, and gain settings for all slides. Z stack images were  
499 captured with 40-80µm step range, and the plane with the best nuclear representation was  
500 analyzed using Fiji software. Background was calculated using four randomly selected empty  
501 regions for each image. Mean signal intensity for p-SMAD2 (Alexa Fluor-488) was calculated  
502 within each nucleus, and background signal was subtracted.

503

504 **Colony formation assays.** Primary BM-derived mononuclear cultured for 24h in StemSpan  
505 SFEM II (Stem Cell Technologies) supplemented with 100 ng/mL of SCF, TPO, Flt3L and 20  
506 ng/mL of IL-3 (PreproTech, Rocky Hill, NJ). Cells were resuspended at 10,000 cells/mL for control  
507 and 20,000 cells/mL for SDS in the presence or absence of 0, 0.25, 0.5, 1, or 5 µM SD208 (Tocris,  
508 Bristol, UK), and incubated for 1hr at 37°C/5% CO<sub>2</sub>. 200 µL of cell suspension was mixed with 3  
509 mL of Methocult H4434 (Stem Cell Technologies), and 1 mL was plated in triplicate in a SmartDish  
510 6-well plate (Stem Cell Technologies). After 14 days of growth at 37°C/5% CO<sub>2</sub>, colonies were  
511 manually scored by two independent, blinded investigators using standard criteria(11).



512

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

522

523 **Statistics.** In figure 2a, statistical significance was determined by the chi-squared test; the  
524 frequency of cells in each cluster was compared between SDS and normal. In Figure 2b, 3c, 4d,  
525 and Extended Data Figure 3b, statistical significance was determined by two-way ANOVA with  
526 Holm-Sidak's multiple correction test in GraphPad Prism 7. In Figure 2b, the frequency of cells  
527 was compared between SDS and normal cells within each cluster. In Figure 3c, log<sub>2</sub> expression  
528 was compared between SDS and normal cells within each cluster. In Figure 4d and  
529 Supplementary Figure 3b, relative colony number was compared between each drug dose and  
530 the 0 $\mu$ M treatment. In Figure 4b and 4c, statistical significance was determined by one-way  
531 ANOVA with Holm-Sidak's multiple correction test in GraphPad Prism 7; SDS samples were  
532 compared to normal samples that were stained and imaged concurrently.

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

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

538

#### 539 **SUPPLEMENTAL METHODS REFERENCES**

540 1. Laurenti E, et al. The transcriptional architecture of early human hematopoiesis identifies  
541 multilevel control of lymphoid commitment. *Nat Immunol.* 2013;14(7):756-63.

542 2. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-  
543 21.

544 3. Li B, and Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or  
545 without a reference genome. *BMC Bioinformatics.* 2011;12:323.

546 4. Jiang L, Chen H, Pinello L, and Yuan GC. GiniClust: detecting rare cell types from single-  
547 cell gene expression data with Gini index. *Genome Biol.* 2016;17(1):144.

548 5. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, et al.  
549 Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq.  
550 *Science.* 2016;352(6282):189-96.

551 6. Satija R, Farrell JA, Gennert D, Schier AF, and Regev A. Spatial reconstruction of single-  
552 cell gene expression data. *Nat Biotechnol.* 2015;33(5):495-502.

553 7. Maaten LJPVD, and Hinton GE. *Visualizing High-Dimensional Data using t-SNE.* 2008.

554 8. Rousseeuw P. *Rousseeuw, P.J.: Silhouettes: A Graphical Aid to the Interpretation and*  
555 *Validation of Cluster Analysis. Comput. Appl. Math. 20, 53-65.* 1987.

556 9. Finak G, et al. MAST: a flexible statistical framework for assessing transcriptional changes  
557 and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.*  
558 2015;16:278.

559 10. Benjamini Y, and Hochberg Y. Controlling the False Discovery Rate: A Practical and  
560 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B*  
561 *(Methodological).* 1995;57(1):289-300.

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

565

566