Abstract

Phenotypic and Transcriptional Consequences of Shwachman-Bodian-Diamond Syndrome (SBDS) Deficiency during Hematopoietic Development

Wing Man Stephanie Ng
Masters of Science
Institute of Medical Science
University of Toronto
2017

Shwachman-Diamond syndrome (SDS) is a genetic disorder characterized by exocrine pancreatic and hematological dysfunction. As hematological complications are major concerns for mortality with limited treatment options, the mechanisms underlying the SDS hematopoietic defects need to be elucidated to advance the development of effective treatments. We hypothesized that the onset of SDS hematopoietic defects occurred in a certain hematopoietic developmental stage. In this study, definitive hematopoietic differentiation of SDS compared to normal hiPSCs showed reduced CFU-GM and BFU-E formation, increased venous and endothelial potential, and reduced viability in embryoid bodies containing hemogenic and vascular endothelia. Over-representation analysis of dysregulated genes in SDS compared to normal hiPSCs showed enrichment for pathways including histones, glycosaminoglycan metabolism, and WNT signaling. This study offers insight into the onset of SDS hematopoietic defects and potentially disrupted pathways in SDS hiPSCs, and provides patient hematopoietic cell resources for drug screening and the potential to improve patient outcome.
Acknowledgments

I would like to take this opportunity to thank everyone who has helped me through my degree. Most importantly, I would like to express my heartfelt gratitude to my supervisor Dr. Yigal Dror for giving me the opportunity to study under his supervision, guiding me through experimental design, and supporting me through my degree. I am also appreciative of my committee members, Dr. Freda Miller, Dr. Johanna Rommens, and Dr. Ian Scott for valuable discussion and constructive criticism.

I would like to express my appreciation and give credit to members of the Dror lab, including our lab manager Hongbing Li for maintaining lab operations, providing constructive criticism, and continuing with this study; our clinical research project manager Bozana Zlateska for acquiring patient samples with informed consent; our MSc alumni Alice Luca for pioneering the study of definitive hematopoiesis of SDS hiPSCs, assisting with the establishment of the protocol of definitive hematopoietic differentiation, and continuing with this study; our former research associate Chetankumar Tailor for producing lentivirus, optimizing lentiviral transduction of hiPSCs, and assisting with the establishment of transduced hiPSCs; our MSc candidate Anna Matveev for optimizing lentiviral transduction of hiPSCs; our lab technician Jake Ablaser for performing Western blot and assisting with hiPSC culture and hematopoietic differentiation; our former research fellow Stephanie Heidemann for assisting in flow cytometry; our former summer undergraduate Katrina Sajewycz for assisting in the prioritization of differentially expressed genes; and our PhD candidate Santhosh Dhanraj for his assistance and constructive criticism.

I would also like to express my appreciation to our collaborators. I would like to thank Andrea Ditadi, Marion Kennedy, and Dr. Gordon Keller for providing the protocol of definitive hematopoietic differentiation, assisting with troubleshooting, and providing constructive criticism. I would like to thank the lab technician Weixian Min for his assistance in hiPSC culture and the Grunebaum lab for providing OP9-DL1. I would like to thank the former lab technician Rikesh Gandhi in the Rommens lab for his assistance in real time-quantitative PCR (RT-qPCR). I would like to thank the PhD candidate Jonathon Torchia in the Huang lab for his assistance in RNA-sequencing (RNA-seq) quality assessment. I would like to thank the PhD candidate Michael Liang in the Wilson lab for his assistance in the TopHat-Cufflinks and over-representation analysis.
I would also like to express my appreciation to the facilities that helped us with our project. I would like to thank the Centre for Commercialization of Regenerative Medicine (CCRM) for hiPSC reprogramming and characterization. I would like to thank the Sickkids Embryonic Stem Cell Facility for providing mouse embryonic fibroblasts (MEFs). I would like to thank the Sickkids Laboratory Animal Services for their assistance in irradiation. I would like to thank the Sickkids Molecular Biology Laboratory for Mycoplasma testing. I would like to thank The Centre for Applied Genomics (TCAG) for G-band karyotyping and Sanger sequencing. I would like to thank the Sickkids-University Health Network Flow Cytometry Facility for fluorescence-activated cell sorting (FACS) and purity assessment. I would like to thank the Donnelly Sequencing Centre for RNA quality assessment, complementary DNA (cDNA) library preparation, and RNA-seq.

On a personal note, I would like to thank the Chau (Grandma, Grandpa, Keith, Wendy, Ho Ho, Matthew, Kelly, Justin, Max, and Ruby) and Xu (Theresa and Linus Jr.) families for their hospitality and support. I would also like to thank Kenneth, Florence, and Angus for their support and encouragement.
Statement of Contributions

My supervisor Dr. Yigal Dror and my committee members Dr. Freda Miller, Dr. Johanna Rommens, and Dr. Ian Scott contributed to the experimental design, data interpretation, and thesis revision and approval. Hongbing Li, Alice Luca, Dr. Gordon Keller, Marion Kennedy, and Andrea Ditadi contributed to the experimental design and data interpretation. Santhosh Dhanraj, Anna Matveev, Stephanie Heidemann, and Katrina Sajewycz provided critical data interpretation.

Bozana Zlateska and the Canadian Inherited Marrow Failure Registry acquired patient samples with informed consent. CCRM reprogrammed and characterized hiPSCs. TCAG performed the G-band karyotyping and Sanger sequencing. The Sickkids Embryonic Stem Cell Facility provided mouse embryonic fibroblasts. The Sickkids Molecular Biology Laboratory performed the Mycoplasma testing. Dr. Eyal Grunebaum provided the OP9-DL1 cell line. The Sickkids Laboratory Animal Services provided the irradiator. Jake Ableser performed Western blotting, hiPSC culture, hematopoietic differentiation, and flow cytometry preparation. The Sickkids-University Health Network Flow Cytometry Facility performed FACS and provided flow cytometers. The Donnelly Sequencing Centre assessed RNA quality, prepared cDNA, and sequenced RNA. Chetankumar Tailor produced the lentivirus and transduced hiPSCs.

I contributed to the experimental design; performed cell culture, hematopoietic differentiation, colony assays, and RT-qPCR; prepared cells for karyotyping, Mycoplasma testing, lentiviral transduction, and FACS; acquired flow cytometry data; prepared DNA for Sanger sequencing; performed statistical analysis; interpreted data; and wrote the thesis.

This work was funded by Canadian Institute of Health Research, Research Training Award from Sickkids, and Nicola’s Triathlon for Kids.
# Table of Contents

Abstract ........................................................................................................................................ ii

Table of Contents ........................................................................................................................ iii

Acknowledgments ........................................................................................................................... iii

Statement of Contributions ........................................................................................................... v

List of Abbreviations .................................................................................................................... ix

List of Tables ................................................................................................................................... xiv

List of Figures .................................................................................................................................. xv

Chapter 1 Introduction .................................................................................................................... 1

1  Shwachman-Diamond Syndrome ......................................................................................... 1

   1.1  Clinical Presentation ...................................................................................................... 1

   1.2  *SBDS* Gene ................................................................................................................ 11

   1.3  SBDS Protein Function ................................................................................................ 13

2  Induced Pluripotent Stem Cells ............................................................................................. 23

3  Hematopoiesis ....................................................................................................................... 27

Chapter 2 Research Aims and Hypothesis ................................................................................... 36

Chapter 3 Methods ......................................................................................................................... 39

1  Cell Culture ............................................................................................................................ 39

2  Characterization of hiPSCs ................................................................................................... 41

3  Hematopoietic Differentiation ............................................................................................... 42

4  Flow Cytometry and Analyses .............................................................................................. 46

5  RNA-Seq and Analyses ....................................................................................................... 49

6  RT-qPCR ................................................................................................................................ 49

7  Statistical Analyses ............................................................................................................... 50

Chapter 4 Results ........................................................................................................................... 51

1  Characterization of hiPSCs ................................................................................................... 51
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Colony Forming Unit Assay .......................................................... 53</td>
</tr>
<tr>
<td>3</td>
<td>Flow Cytometry Analyses ................................................................. 55</td>
</tr>
<tr>
<td>3.1</td>
<td>Human Induced Pluripotent Stem Cells ............................................. 56</td>
</tr>
<tr>
<td>3.2</td>
<td>Mesodermal Cells .................................................................................. 59</td>
</tr>
<tr>
<td>3.3</td>
<td>Hemogenic Endothelial Cells ................................................................. 62</td>
</tr>
<tr>
<td>3.4</td>
<td>CD34&lt;sup&gt;+&lt;/sup&gt;/CD43&lt;sup&gt;-&lt;/sup&gt; Population ........................................ 68</td>
</tr>
<tr>
<td>4</td>
<td>RNA-Seq Analyses of Sorted hiPSCs ....................................................... 70</td>
</tr>
<tr>
<td>5</td>
<td>Lentiviral Transduction of SDS hiPSCs ................................................ 80</td>
</tr>
<tr>
<td>6</td>
<td>Optimization of RT-qPCR ..................................................................... 82</td>
</tr>
<tr>
<td></td>
<td>Chapter 5 Discussion ............................................................................. 87</td>
</tr>
<tr>
<td>1</td>
<td>Flow Cytometry Analyses ..................................................................... 88</td>
</tr>
<tr>
<td>2</td>
<td>RNA-seq Analysis of SDS hiPSC ................................................................ 91</td>
</tr>
<tr>
<td></td>
<td>Chapter 6 Conclusion ........................................................................... 104</td>
</tr>
<tr>
<td></td>
<td>Chapter 7 Future Directions .................................................................. 106</td>
</tr>
<tr>
<td></td>
<td>Bibliography .......................................................................................... 112</td>
</tr>
<tr>
<td></td>
<td>Appendices ............................................................................................. 136</td>
</tr>
<tr>
<td>1</td>
<td>Pluripotency of hiPSCs ........................................................................ 136</td>
</tr>
<tr>
<td>2</td>
<td>Genetic Identity of hiPSCs ................................................................... 140</td>
</tr>
<tr>
<td>3</td>
<td>Flow Cytometry Analyses ...................................................................... 143</td>
</tr>
<tr>
<td>3.1</td>
<td>Human Induced Pluripotent Stem Cells ................................................. 143</td>
</tr>
<tr>
<td>3.2</td>
<td>Mesodermal Cells ................................................................................. 145</td>
</tr>
<tr>
<td>3.3</td>
<td>Optimizations in Isolating Hemogenic Endothelium ................................ 157</td>
</tr>
<tr>
<td>3.4</td>
<td>Hemogenic Endothelial Cells ................................................................ 167</td>
</tr>
<tr>
<td>3.5</td>
<td>CD34&lt;sup&gt;+&lt;/sup&gt;/CD43&lt;sup&gt;-&lt;/sup&gt; Population ............................................. 169</td>
</tr>
<tr>
<td>3.6</td>
<td>Early Hematopoietic Progenitors .......................................................... 170</td>
</tr>
<tr>
<td>4</td>
<td>RNA Sequencing ...................................................................................... 174</td>
</tr>
</tbody>
</table>
4.1 Quality ............................................................................................................ 174
4.2 Alignment ....................................................................................................... 180
4.3 Reproducibility ............................................................................................. 183
4.4 Correlation .................................................................................................... 183
4.5 SBDS Splice Variants .................................................................................... 186
4.6 SBDSP1 Splice Variants ................................................................................ 189
# List of Abbreviations

<table>
<thead>
<tr>
<th>A</th>
<th>Adenine</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFP</td>
<td>Alpha Fetoprotein</td>
</tr>
<tr>
<td>AGM</td>
<td>Aorta-Gonad-Mesonephros</td>
</tr>
<tr>
<td>AKT</td>
<td>Serine-Threonine Protein Kinase</td>
</tr>
<tr>
<td>AMP</td>
<td>Adenosine Monophosphate</td>
</tr>
<tr>
<td>AMPK</td>
<td>AMP-activated Protein Kinase</td>
</tr>
<tr>
<td>APC</td>
<td>Allophycocyanin</td>
</tr>
<tr>
<td>APC-CD235a</td>
<td>APC Mouse Anti-Human CD235a</td>
</tr>
<tr>
<td>APC-CD41a</td>
<td>APC Mouse Anti-Human CD41a</td>
</tr>
<tr>
<td>APC-CD56</td>
<td>APC Mouse Anti-Human CD56</td>
</tr>
<tr>
<td>APC-CD73</td>
<td>APC Mouse Anti-Human CD73</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine Triphosphate</td>
</tr>
<tr>
<td>ATP2A2</td>
<td>ATPase Sarcoplasmic/Endoplasmic reticulum Ca2+ Transporting 2</td>
</tr>
<tr>
<td>ATP2B1</td>
<td>ATPase Plasma Membrane Ca2+ Transporting 1</td>
</tr>
<tr>
<td>B</td>
<td>Bone Marrow</td>
</tr>
<tr>
<td>BD</td>
<td>Becton Dickinson</td>
</tr>
<tr>
<td>bFGF</td>
<td>Recombinant Human Fibroblast Growth Factor Basic Protein</td>
</tr>
<tr>
<td>BFU-E</td>
<td>Blast Forming Unit-Erythroid</td>
</tr>
<tr>
<td>BLCFC</td>
<td>Blast Colony Forming Cells</td>
</tr>
<tr>
<td>BM</td>
<td>Bone Marrow</td>
</tr>
<tr>
<td>BMP-4</td>
<td>Recombinant Human Bone Morphogenetic Protein 4</td>
</tr>
<tr>
<td>BV</td>
<td>Brilliant Violet</td>
</tr>
<tr>
<td>BV421-CD184</td>
<td>BV421 Mouse Anti-Human CD184</td>
</tr>
<tr>
<td>BV421-C-KIT</td>
<td>BV421 Mouse Anti-Human CD117</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>C/EBP</td>
<td>CCAAT Enhancer Binding Protein</td>
</tr>
<tr>
<td>CCRM</td>
<td>Centre for Commercialization of Regenerative Medicine</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of Differentiation</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CFU</td>
<td>Colony Forming Unit</td>
</tr>
<tr>
<td>CFU-C</td>
<td>Colony Forming Unit, In Culture</td>
</tr>
<tr>
<td>CFU-E</td>
<td>Colony Forming Unit, Erythroid</td>
</tr>
<tr>
<td>CFU-G</td>
<td>Colony Forming Unit-Granulocyte</td>
</tr>
<tr>
<td>CFU-GEMM</td>
<td>Colony Forming Unit-Granulocyte, Erythroid, Monocyte, and Megakaryocyte</td>
</tr>
<tr>
<td>CFU-GM</td>
<td>Colony Forming Unit-Granulocyte and Monocyte</td>
</tr>
<tr>
<td>CFU-S</td>
<td>Colony Forming Unit, In Spleen</td>
</tr>
<tr>
<td>CHST11</td>
<td>Carbohydrate (Chondroitin 4) Sulfotransferase 11</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CLP</td>
<td>Common Lymphoid Progenitor</td>
</tr>
<tr>
<td>CMP</td>
<td>Common Myeloid Progenitor</td>
</tr>
<tr>
<td>C-MYC</td>
<td>V-Myc Avian Myelocytomatosis Viral Oncogene Homolog</td>
</tr>
<tr>
<td>$C_t$</td>
<td>Threshold Cycle</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>HAND1</td>
<td>Heart And Neural Crest Derivatives Expressed 1</td>
</tr>
<tr>
<td>HE</td>
<td>Hemogenic Endothelium</td>
</tr>
<tr>
<td>HIST1H1A</td>
<td>Histone Cluster 1 H1 Family Member A</td>
</tr>
<tr>
<td>HIST1H2BK</td>
<td>Histone Cluster 1 H2B Family Member K</td>
</tr>
<tr>
<td>HSC</td>
<td>Hematopoietic Stem Cell</td>
</tr>
<tr>
<td>HSCT</td>
<td>HSC Transplantation</td>
</tr>
<tr>
<td>i(7)(q10)</td>
<td>Isochromosome of the Long Arm of Chromosome 7</td>
</tr>
<tr>
<td>ID</td>
<td>Identification</td>
</tr>
<tr>
<td>IIF30</td>
<td>GILT</td>
</tr>
<tr>
<td>IGF-1</td>
<td>Recombinant Human Insulin-like Growth Factor I</td>
</tr>
<tr>
<td>IGV</td>
<td>Integrative Genome Viewer</td>
</tr>
<tr>
<td>IL-11</td>
<td>Recombinant Human Interleukin 11</td>
</tr>
<tr>
<td>IL-3</td>
<td>Recombinant Human Interleukin 3</td>
</tr>
<tr>
<td>IL-6</td>
<td>Recombinant Human Interleukin 6</td>
</tr>
<tr>
<td>IMDM</td>
<td>Iscove's Modified Dulbecco's Medium</td>
</tr>
<tr>
<td>iPSC</td>
<td>Induced Pluripotent Stem Cell</td>
</tr>
<tr>
<td>IQ</td>
<td>Intelligence Quotient</td>
</tr>
<tr>
<td>IRES</td>
<td>Internal ribosome entry site</td>
</tr>
<tr>
<td>iSV.BM</td>
<td>hiPSCs Reprogrammed Using Sendai Virus and Derived from Bone Marrow</td>
</tr>
<tr>
<td>KDR</td>
<td>Kinase Insert Domain Receptor</td>
</tr>
<tr>
<td>KLF4</td>
<td>Kruppel-like Factor 4</td>
</tr>
<tr>
<td>KOSR</td>
<td>KnockOut™ Serum Replacement</td>
</tr>
<tr>
<td>LEF</td>
<td>Lymphoid Enhancer Factor</td>
</tr>
<tr>
<td>LIP</td>
<td>C/EBP-Liver Inhibitory Protein</td>
</tr>
<tr>
<td>LMPP</td>
<td>Lymphoid-primed MPP</td>
</tr>
<tr>
<td>lncRNA</td>
<td>Long Non-coding RNA</td>
</tr>
<tr>
<td>LTC-IC</td>
<td>Long-Term Culture-Initiating Cells</td>
</tr>
<tr>
<td>m</td>
<td>mouse</td>
</tr>
<tr>
<td>MEF</td>
<td>Mouse Embryonic Fibroblast</td>
</tr>
<tr>
<td>MEP</td>
<td>Megakaryocytic-Erythrocytic Progenitor</td>
</tr>
<tr>
<td>MLP</td>
<td>Multi-Lymphoid Progenitor</td>
</tr>
<tr>
<td>MOI</td>
<td>Multiplicity of Infection</td>
</tr>
<tr>
<td>MP</td>
<td>Myeloid Progenitors</td>
</tr>
<tr>
<td>MPP</td>
<td>Multipotent Progenitor</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>mTOR</td>
<td>Mechanistic Target of Rapamycin</td>
</tr>
<tr>
<td>N</td>
<td>Normal</td>
</tr>
<tr>
<td>NAC</td>
<td>No Amplification Control</td>
</tr>
<tr>
<td>NADH</td>
<td>Nicotinamide Adenine Dinucleotide</td>
</tr>
<tr>
<td>NDUFB8</td>
<td>NADH:Ubiquinone Oxidoreductase Subunit B8</td>
</tr>
<tr>
<td>NEUROD1</td>
<td>Neuronal Differentiation 1</td>
</tr>
<tr>
<td>NK</td>
<td>Natural Killer</td>
</tr>
<tr>
<td>NTC</td>
<td>No Template Control</td>
</tr>
<tr>
<td>Oct4</td>
<td>Octamer-binding Transcription Factor 4</td>
</tr>
<tr>
<td>p</td>
<td>p-value</td>
</tr>
</tbody>
</table>

xi
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCF</td>
<td>T Cell Factor</td>
</tr>
<tr>
<td>Tif6</td>
<td>Translation Initiation Factor 6, yeast (human gene EIF6)</td>
</tr>
<tr>
<td>Tm</td>
<td>Melting Temperature</td>
</tr>
<tr>
<td>TPO</td>
<td>Recombinant Human Thrombopoietin</td>
</tr>
<tr>
<td>uORF</td>
<td>Upstream Open Reading Frame</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated Region</td>
</tr>
<tr>
<td>V450</td>
<td>Violet 450</td>
</tr>
<tr>
<td>V450-CD45</td>
<td>V450 Mouse Anti-Human CD45</td>
</tr>
<tr>
<td>VDAC1</td>
<td>Voltage Dependent Anion Channel 1</td>
</tr>
<tr>
<td>VE-cad</td>
<td>Vascular Endothelial Cadherin (CDH5)</td>
</tr>
<tr>
<td>VEGF</td>
<td>Recombinant Human Vascular Endothelial Growth Factor</td>
</tr>
<tr>
<td>VEGFR-2</td>
<td>Vascular Endothelial Growth Factor Receptor-2</td>
</tr>
<tr>
<td>WNT</td>
<td>Wingless-Type MMTV Integration Site</td>
</tr>
<tr>
<td>WNT10B</td>
<td>WNT Family, Member 10B</td>
</tr>
<tr>
<td>αMEM</td>
<td>Minimum Essential Medium α</td>
</tr>
</tbody>
</table>
List of Tables

Table 1. Reagents used in the maintenance of hiPSCs cultured on MEFs. .............................................. 39

Table 2. Reagents used in the maintenance of OP9-DL1 culture. ................................................................. 40

Table 3. Reagents used in the hematopoietic differentiation of hiPSCs cultured on MEFs. ........ 42

Table 4. Fluorophores and cell density of PI only control, FMO controls, and sorting sample of T9 staining reactions. .................................................................................................................................................. 45

Table 5. Components of the stage-specific staining reactions................................................................. 47

Table 6. Fluorophores and cell density of staining reactions................................................................. 48

Table 7. Clinical features of normal subjects and SDS patients........................................................... 51

Table 8. PI viable populations in SDS compared to normal hiPSC-derived developmental stages. .................................................................................................................................................. 68

Table 9. Sequencing depth and coverage of hiPSC RNA-seq .............................................................. 70

Table 10. Transcriptional features of SBDS variants.................................................................................. 72

Table 11. Transcriptional features of SBDSP1 variants.............................................................................. 75

Table 12. Significantly enriched pathways in SDS compared to normal hiPSCs. ......................... 79

Table 13. Concentration of optimized and designed RT-qPCR primers............................................. 83

Supplementary Table 1. Detailed transcriptional features of SBDS variants........................................ 187

Supplementary Table 2. Detailed transcriptional features of SBDSP1 variants.................................... 190
List of Figures

Figure 1. SDS-associated mutations in hiPSCs. ................................................................. 52
Figure 2. SBDS protein expression in hiPSCs. ................................................................. 52
Figure 3. Analyses of CFU Counts at T9+7+14. ............................................................... 54
Figure 4. Definitive hematopoietic differentiation timeline. ............................................. 55
Figure 5. Flow cytometry analyses of T0 hiPSC populations. .......................................... 57
Figure 6. Percentages of T0 hiPSC populations. ............................................................... 58
Figure 7. Flow cytometry analyses of T4 mesodermal populations. ............................... 60
Figure 8. Percentages of T4 mesodermal populations. ..................................................... 61
Figure 9. Flow cytometry analyses of T8 hemogenic endothelial populations. .............. 64
Figure 10. Percentages of T8 hemogenic and vascular endothelial populations. ............ 67
Figure 11. Flow cytometry analyses and percentages of T9 hemogenic and vascular endothelial populations. ......................................................................................... 69
Figure 12. Schematic of SBDS splice variants of hiPSCs .................................................. 72
Figure 13. Expression of SBDS variants in hiPSCs. .......................................................... 73
Figure 14. Schematic of SBDSP1 splice variants of hiPSCs. ............................................. 75
Figure 15. Expression of SBDSP1 variants in hiPSCs. ....................................................... 76
Figure 16. Scatter plot of significantly differentially expressed, non-filtered genes in SDS compared to normal hiPSCs ............................................................... 77
Figure 17. Significantly differentially expressed filtered genes in SDS compared to normal hiPSCs ................................................................................................................. 78
Figure 18. Lentiviral transduction of SDS hiPSCs. ............................................................ 80
Figure 19. Lentiviral plasmid maps. ................................................................. 81
Figure 20. RT-qPCR amplification plot of primers targeting RPLP0. ....................... 84
Figure 21. RT-qPCR dissociation curve of primers targeting RPLP0. ............................ 85
Figure 22. RT-qPCR standard curve of primers targeting RPLP0. ............................... 86

Supplementary Figure 1. Gene expression of pluripotency markers in hiPSCs. ............. 136
Supplementary Figure 2. Protein expression of pluripotency markers in hiPSCs .......... 137
Supplementary Figure 3. Protein expression of pluripotency markers in P55 hiPSCs ....... 138
Supplementary Figure 4. Protein expression of pluripotency markers in P55 hiPSCs ....... 139
Supplementary Figure 5. Gene expression of germ layer markers in hiPSC-derived EBs. ... 140
Supplementary Figure 6. Genetic identity of hiPSCs compared to parental cells. .......... 141
Supplementary Figure 7. Karyotyping of hiPSCs.................................................. 142
Supplementary Figure 8. Compensation and unstained controls of hiPSC markers. ....... 143
Supplementary Figure 9. Fluorescence-minus-one negative controls of hiPSC markers .... 144
Supplementary Figure 10. Compensation and PI only controls of initial mesodermal markers. 145
Supplementary Figure 11. Fluorescence-minus-one negative controls of initial mesodermal markers................................................................. 146
Supplementary Figure 12. Flow cytometry analyses of T4 mesodermal populations with SSEA-3
instead of TRA-1-60. .......................................................................................... 148
Supplementary Figure 13. Compensation and PI only controls of initial mesodermal markers
with SSEA-3 instead of TRA-1-60. ....................................................................... 149
Supplementary Figure 14. Fluorescence-minus-one negative controls of initial mesodermal markers with SSEA-3 instead of TRA-1-60. ................................................................. 150

Supplementary Figure 15. Flow cytometry analyses of T3 mesodermal populations with thorough trypsin removal. ................................................................. 151

Supplementary Figure 16. Compensation and PI only controls of initial mesodermal markers at T3 with thorough trypsin removal. ................................................................. 152

Supplementary Figure 17. Fluorescence-minus-one negative controls of revised mesodermal markers at T3 with thorough trypsin removal. ................................................................. 153

Supplementary Figure 18. Flow cytometry analyses of T4 mesodermal populations with revised mesodermal markers and thorough trypsin removal. ................................................................. 154

Supplementary Figure 19. Compensation and PI only controls of revised mesodermal markers with thorough trypsin removal. ................................................................. 155

Supplementary Figure 20. Fluorescence-minus-one negative controls of revised mesodermal markers with thorough trypsin removal. ................................................................. 156

Supplementary Figure 21. Compensation and PI only controls of hemogenic endothelial markers without collagenase I digestion. ................................................................. 158

Supplementary Figure 22. Fluorescence-minus-one negative controls of hemogenic endothelial markers without collagenase I digestion. ................................................................. 159

Supplementary Figure 23. Flow cytometry analyses of T8 hemogenic endothelial populations without collagenase I digestion. ................................................................. 160

Supplementary Figure 24. Flow cytometry analyses of hemogenic endothelial populations without collagenase I digestion from T5-8. ................................................................. 162

Supplementary Figure 25. Analyses of hemogenic endothelial populations without collagenase I digestion from T5-8. ................................................................. 163

Supplementary Figure 26. Compensation, PI only, and fluorescence-minus-one negative controls of hemogenic endothelial markers at T9+3. ................................................................. 164
Supplementary Figure 27. Flow cytometry analyses of hemogenic endothelial populations at T9+3. ................................................................. 165

Supplementary Figure 28. Analyses of CD56 expression T8 hemogenic endothelial populations with collagenase I digestion................................................................. 166

Supplementary Figure 29. Compensation and PI only controls of hemogenic endothelial markers with collagenase I digestion................................................................. 167

Supplementary Figure 30. Fluorescence-minus-one negative controls of hemogenic endothelial markers with collagenase I digestion................................................................. 168

Supplementary Figure 31. Compensation, PI only, and fluorescence-minus-one negative controls of CD34+/CD43- hemogenic and vascular endothelial population in T9 EBs................................................................. 169

Supplementary Figure 32. Flow cytometry analyses of early hematopoietic progenitors at T9+7. .................................................................................................................... 171

Supplementary Figure 33. Compensation and PI only controls of early hematopoietic progenitor markers........................................................................................................... 172

Supplementary Figure 34. Fluorescence-minus-one negative controls of early hematopoietic progenitor markers........................................................................................................... 173

Supplementary Figure 35. RNA-seq per base sequence quality of hiPSC. ................................................................. 175

Supplementary Figure 36. RNA-seq per sequence quality score of hiPSC................................................................. 176

Supplementary Figure 37. RNA-seq per base sequence content of hiPSC............................................................. 177

Supplementary Figure 38. RNA-seq per sequence GC content of hiPSC................................................................. 178

Supplementary Figure 39. RNA-seq sequence duplication levels of hiPSC. ................................................................. 179

Supplementary Figure 40. RNA-seq alignment information of hiPSCs................................................................. 181

Supplementary Figure 41. RNA-seq transcript coverage of hiPSCs................................................................. 182
Supplementary Figure 42. RNA-seq differential expression heat map of biological replicates of SDS and normal hiPSCs. ................................................................. 184

Supplementary Figure 43. RNA-seq differential expression heat map of hiPSCs. ................. 185

Supplementary Figure 44. SBDS splice variants of hiPSCs visualized by IGV. ......................... 186

Supplementary Figure 45. Sashimi plot of SBDS splice junctions of hiPSCs. ......................... 188

Supplementary Figure 46. SBDSP1 variants of hiPSCs visualized by IGV. ......................... 189

Supplementary Figure 47. Sashimi plot of SBDSP1 splice junctions of hiPSCs. ..................... 191
Chapter 1
Introduction

1 Shwachman-Diamond Syndrome

Shwachman-Diamond syndrome (SDS) is characterized by bone marrow failure, exocrine pancreatic dysfunction, skeletal abnormalities, and neurodevelopmental issues (Bodian et al., 1964; Shwachman et al., 1964). Progression to myelodysplastic syndrome and acute myeloid leukemia has been suggested to be a major concern for mortality in SDS (Huijgens et al., 1977). Inherited through an autosomal recessive mode, SDS is classified as a rare disorder with an estimated incidence of 1 in 76,563 (Ginzberg et al., 2000; Goobie et al., 2001). Biallelic mutations in the Shwachman-Bodian-Diamond Syndrome (SBDS) gene are found in 60-89% of SDS cases (Boocock et al., 2003). SDS is also classified as a ribosomopathy as SBDS plays an essential role in ribosomobiogenesis. SBDS couples the GTP hydrolysis activity of EFL1 (elongation factor like GTPase 1) to the release of the anti-association factor eIF6 (eukaryotic translation initiation factor 6) during pre-60S maturation (Menne et al., 2007; Finch et al., 2011).

**Discovery.** SDS was reported in 1964 as a rare congenital disorder with combined pancreatic and hematological dysfunction. In May 1964, Bodian et al. reported two infant cases 2 months of age and described a new condition with pancreatic lipomatosis and exocrine pancreatic hypoplasia associated with hematological dysfunction, including anemia, thrombocytopenia, neutropenia, leukopenia, and pancytopenia. In November of the same year, Shwachman et al. reported five children and siblings under 3 years of age with pancreatic insufficiency combined with predominantly neutropenia, leading to the recognition of Shwachman-Diamond syndrome. Cases of SDS are often initially misdiagnosed as cystic fibrosis as they share the common presentations of growth failure, frequent early infections, and pancreatic insufficiency (Bodian et al., 1964; Shwachman et al., 1964). However, SDS is distinguished from cystic fibrosis with normal sweat electrolyte concentrations, absence of chronic pulmonary disease, and presentation of hematologic anomalies.

1.1 Clinical Presentation

SDS is a rare disorder characterized by bone marrow failure, exocrine pancreatic dysfunction, skeletal abnormalities, and abnormalities in other organ systems. Diagnosis requires the
combined presence of exocrine pancreatic dysfunction and hematological cytopenia of any lineage (Dror et al., 2011). An analysis of SDS cases showed that the median age of presentation was 1.5 months, the median age of diagnosis was 17 months, and the crude mortality rate (number of deaths per year per 1000 people) was 6% (Hashmi et al., 2011). The mortality rates of SDS were highest before 5 years and after 30 years of age (Donadieu et al., 2012). Slight male predominance in SDS was observed, with 70% of isolated cases being male, although only 44% of affected siblings were male (Ginzberg et al., 1999).

**Hematological Dysfunction**

SDS is the third most commonly recognized inherited bone marrow failure after Diamond-Blackfan anemia and Fanconi anemia (Dror et al., 2005). Hematological complications are the major concerns of mortality in SDS. The hematological features that manifest in SDS include cytopenia (neutropenia, anemia, and thrombocytopenia), hematological abnormalities (macrocytic red blood cells, elevated fetal hemoglobin, and hypocellular bone marrow), and hematological complications (aplastic anemia, clonal marrow cytogenetic abnormalities, myelodysplastic syndrome, and acute myeloid leukemia).

**Neutropenia.** Neutropenia (neutrophil count below 1.5×10^9/L) is the most commonly observed cytopenia in SDS, manifesting in 74-98% of SDS patients (Ginzberg et al., 1999; Hashmi et al., 2011; Myers et al., 2014). The severity of neutropenia in SDS can be mild (neutrophil count between 1.0-1.5×10^9/L) or moderate (neutrophil count between 0.5-1.0×10^9/L), but is often severe (neutrophil count below 0.5×10^9/L) (Hashmi et al., 2011). Neutropenia in SDS can be persistent, but is often intermittent (Ginzberg et al., 1999). Neutropenia in SDS can develop with age, but has also been diagnosed in newborns (Dror et al., 2011).

**Anemia.** Often asymptomatic, anemia (hemoglobin levels below 70g/L) is the second most commonly observed cytopenia in SDS, manifesting in 42-58% of SDS patients (Ginzberg et al., 1999; Hashmi et al., 2011). The severity of anemia in SDS is often mild-moderate (hemoglobin levels above 70g/L but below the age- and gender-adjusted lower limit) rather than severe (hemoglobin levels below 70g/L).

**Thrombocytopenia.** Thrombocytopenia (platelet count below 200×10^9/L) is the least commonly observed cytopenia in SDS, manifesting in 34-38% of SDS patients (Ginzberg et al., 1999;
Hashmi et al., 2011). The severity of thrombocytopenia in SDS can be mild (platelet count between 100-150x10^9/L) and often moderate (platelet count between 10-100x10^9/L), but is rarely severe (platelet count below 20x10^9/L).

**Lymphopenia.** Lymphopenia (lymphocyte count below 1.5x10^9/L) is not commonly observed in SDS (Dror et al., 2001). Although most SDS patients do not present with overt lymphopenia, a study in the immune function of SDS patients discovered varying lymphoid defects, including reduced numbers of B-cells, T-cells, and natural killer (NK) cells. Defects in humoral immunity included reduced serum IgG and its subclasses, reduced circulating B-lymphocytes, reduced B-lymphocyte proliferation *in vitro*, and reduced antibody production. Defects in cellular immunity included reduced circulating T-lymphocytes and reduced T-lymphocyte proliferation *in vitro*.

**Recurrent infections.** Due to neutropenia, viral and bacterial infections are frequent in SDS, especially in early childhood (Ginzberg et al., 1999; Dror et al., 2001). Recurrent viral infections can manifest as respiratory tract infection, gastroenteritis, and unexplained fever. Recurrent bacterial infections can manifest as otitis media, sinusitis, and pneumonia. Severe systemic or deep tissue infections can manifest as septicemia, oral cellulitis and dental abscess, osteomyelitis, abscess, and staphylococcal skin infection.

**Macrocytic red blood cells.** Red blood cells can be normochromic or normocytic, but are predominantly macrocytic in SDS, occurring in 60% of SDS patients after 1 year of age (Dror et al., 2011; Hashmi et al., 2011).

**Elevated fetal hemoglobin levels.** Fetal hemoglobin levels are often elevated in SDS, occurring in 72% of SDS patients after 1 year of age (Dror et al., 2011; Hashmi et al., 2011).

**Hypocellular bone marrow.** Bone marrow cellularity is predominantly hypocellular (cellularity below 25% or cellularity below 50% combined with hematopoietic cells below 30%) with fat replacement in SDS, and rarely normal or hypercellular (Dror et al., 2011; Ginzberg et al., 1999; Hashmi et al., 2011). Patient bone marrows most commonly show maturation delay or arrest in the myeloid lineage, including reduced erythropoiesis or megakaryopoiesis, and often show reduced granulopoiesis as well (Dror et al., 2011; Hashmi et al., 2011).
Aplastic anemia. SDS patients can present with more than one cytopenia. 24% of SDS patients presented with bilineage cytopenia and 19-24% presented with pancytopenia (Ginzberg et al., 1999; Hashmi et al., 2011). SDS can progress to aplastic anemia (bilineage or trilineage cytopenia combined with hypocellular bone marrow in the absence of fibrosis, dysplastic morphology, or clonal evolution), manifesting in 15% of SDS patients at the median age of 3.9 years (Dror et al., 2011; Hashmi et al., 2011). The severity of aplastic anemia in SDS can be moderate (at least two of reticulocyte count below 40×10⁹/L, platelet count below 40×10⁹/L, or neutrophil count below 1.5×10⁹/L) or severe (at least two of reticulocyte count below 40×10⁹/L, platelet count below 20×10⁹/L, or neutrophil count below 0.5×10⁹/L). Aplastic anemia in SDS can be transient or persist over 3 months, and has also been diagnosed in newborns (Kuijpers et al., 2004; Dror et al., 2011). SDS patients with poor prognoses, indicated by age of diagnosis before 3 months and low hematological parameters in the first three complete blood counts (either neutrophil count below 0.5×10⁹/L, hemoglobin levels below 9 g/dL, or platelet count below 100×10⁹/L), have 59% risk of developing aplastic anemia in the next 10 and 20 years (Donadieu et al., 2012).

Clonal marrow cytogenetic abnormalities. SDS patients may acquire chromosomal defects in bone marrow cells known as clonal marrow cytogenetic abnormalities (at least 2 out of 20 bone marrow cells gaining the same cytogenetic abnormality or at least 3 out of 20 cells losing the same chromosome in G-banding, or increased frequency of cytogenetic abnormality in fluorescence in situ hybridization) (Dror et al., 2011). Acquisition of cytogenetic abnormalities and genomic instability in SDS has been suggested to be a consequence of microtubule instability (Austin et al., 2008; Lombardo et al., 2010). Clonal marrow cytogenetic abnormalities in SDS may remain stable, fluctuate in numbers, occur intermittently, regress, or progress to myelodysplastic syndrome and acute myeloid leukemia (Smith et al., 2002; Dror et al., 2002). 44% of cytogenetic abnormalities in SDS were i(7)(q10) (isochromosome of the long arm of chromosome 7), 33% were other varying abnormalities in chromosome 7 (monosomy, deletion, and translocations), and 16% were del(20)(q) (interstitial deletions of the long arm of chromosome 20) (Dror et al., 2005).

i(7)(q10) is the most commonly observed cytogenetic abnormality in SDS (Dror et al., 2005), while also located in the long arm of chromosome 7 is SBDS (7q11). If i(7)(q10) appears concurrently with the splice site mutation (c.258+2T>C) of SBDS, the duplication of
c.258+2T>C can result in increased expression of the SBDS splice variant (Minelli et al., 2009). Thus, i(7)(q10) has been proposed to be a compensatory mechanism to restore SBDS expression. del(20)(q) or int del(20)(q11.21q13.32) is the second most commonly observed cytogenetic abnormality in SDS and results in the deletion of EIF6 on 20q11.22 (Pressato et al., 2012). Thus, del(20)(q) has been proposed to be a compensatory mechanism to restore ribosome biogenesis by circumventing the requirement for eIF6 release. Both i(7)(q10) and del(20)(q) are associated with benign prognoses, although other cytogenetic abnormalities with unknown prognoses often appear concurrently (Pressato et al., 2015). Contrary to the aforementioned benign cytogenetic abnormalities, monosomy 7 is the most commonly observed cytogenetic abnormality associated with malignant prognosis, observed in 50% of SDS patients with myelodysplastic syndrome or acute myeloid leukemia (Donadieu et al., 2012).

**Myelodysplastic syndrome.** Clonal marrow cytogenetic abnormalities in SDS can result in malignant transformation and progression to myelodysplastic syndrome (chronic bilineage or trilineage cytopenia and presence of clonal marrow cytogenetic abnormality with marrow myeloblast count between 5–29%) (Dror et al., 2011). 18% of SDS patients developed clonal and malignant myeloid transformation at the median age of 20 years (Hashmi et al., 2011). Although myelodysplastic syndrome shares similarities with aplastic anemia, myelodysplastic syndrome is distinguished from aplastic anemia by dysplastic morphology with clonal evolution. Dysplastic changes in SDS can occur in erythroid, myeloid, and megakaryocytic lineages, can be mild and transient, and can also occur without cytogenetic abnormalities.

**Acute myeloid leukemia.** Myelodysplastic syndrome in SDS can progress to acute myeloid leukemia (chronic bilineage or trilineage cytopenia and presence of clonal marrow cytogenetic abnormality with marrow myeloblast count of at least 20%) (Dror et al., 2011). 50% of SDS patients diagnosed with myelodysplastic syndrome progressed to leukemia with high male predominance (Dror et al., 2011). Although multiple subtypes of acute myeloid leukemia have been reported (M0 undifferentiated acute myeloblastic leukemia, M1 acute myeloblastic leukemia with minimal maturation, M2 acute myeloblastic leukemia with maturation, M4 acute myelomonocytic leukemia, and M5 acute monocytic leukemia), M6 acute erythroid leukemia is the most commonly observed (Lesesve et al., 2003; Mitsui et al., 2004; Smith et al., 1996; Woods et al., 1981; Dokal et al., 1997). Acute lymphoblastic leukemia (Strevens et al., 1978) and juvenile myelomonocytic leukemia (Caselitz et al., 1979) have also been reported.
**Pancreatic Dysfunction**

SDS is the second most commonly observed pancreatic dysfunction after cystic fibrosis and malabsorption is one of the major concerns for mortality early in life in SDS (Dror *et al.*, 2005). 74-85% of SDS patients presented with pancreatic dysfunction, often diagnosed before 6 months of age (Ginzberg *et al.*, 1999; Hashmi *et al.*, 2011; Mack *et al.*, 1996).

**Exocrine pancreatic enzymes.** As acinar cells are reduced in number and size and the exocrine pancreas is replaced by fat, the secretion of exocrine pancreatic enzymes is insufficient in SDS. Exocrine pancreatic insufficiency in SDS is predominantly diagnosed by reduced levels of exocrine pancreatic enzymes (serum trypsinogen levels below 6μg/L, serum isoamylase levels below 16U/L at age under 3 years, fecal elastase levels below 200μg/g, and reduced fecal chymotrypsin and serum lipase levels) (Ip *et al.*, 2002, Dror *et al.*, 2011). 74% of SDS patients were diagnosed as pancreatic insufficient by reduced trypsinogen levels (Ginzberg *et al.*, 1999).

**Malabsorption.** As the secretion of exocrine pancreatic enzymes is insufficient, the digestion and absorption of fats and fat-soluble vitamins (A, D, E, and K) are reduced, resulting in malnutrition, diarrhea, and steatorrhea. 58% of SDS patients presented with diarrhea, 86% presented with steatorrhea, and 67% presented with steatorrhea requiring treatment (Myers *et al.*, 2014; Ginzberg *et al.*, 1999; Hashmi *et al.*, 2011). Increased fecal fat loss by 72-hr fat balance study is another method to diagnose exocrine pancreatic insufficiency (Dror *et al.*, 2011).

**Pancreatic sufficiency.** The severity of exocrine pancreatic insufficiency is varied and SDS patients can eventually become pancreatic sufficient from the spontaneous improvement of nutrient absorption with age (Dror *et al.*, 2011). Often after 4 years of age, 50% of SDS patients become pancreatic sufficient with normal trypsinogen levels but not amylase (Mack *et al.*, 1996; Ip *et al.*, 2002). Moreover, more males (41%) than females (16%) have become pancreatic sufficient (Ginzberg *et al.*, 1999).

**Fatty Pancreas.** Pancreatic lipomatosis is also varied and only 83% of SDS patients presented with pancreatic lipomatosis (Ginzberg *et al.*, 1999; Myers *et al.*, 2014). The pancreas can be fatty with intact islets of Langerhans and ductal structure, small or enlarged, and often not assessed. Lipomatosis can develop or disappear with age.
**Growth retardation.** Growth retardation is commonly observed before 1 year of age in SDS (Dror *et al*., 2011). 59% of SDS patients presented with developmental delay, 73% presented with failure to thrive, 61% presented with growth retardation, and 26% presented with intrauterine growth retardation (Hashmi *et al*., 2011; Myers *et al*., 2014; Donadieu *et al*., 2012). The mean birth weight is often in the 25th percentile, with height and weight in the 3rd percentile at 1 year of age despite enzyme treatment (Mack *et al*., 1996; Ginzberg *et al*., 1999).

**Endocrine pancreas.** Although the islets of Langerhans appear normal in pancreatic biopsies, rare cases of endocrine pancreatic dysfunction impairing insulin production have been reported in SDS (Hashmi *et al*., 2011; Myers *et al*., 2014). The prevalence of type 1 diabetes in SDS was estimated to be 30-fold higher than the general population (Gana *et al*., 2011). Thus, immunity abnormalities in SDS have been suggested to be associated with autoimmune disease and type 1 diabetes.

**Skeletal Abnormalities**

Other than hematological or exocrine pancreatic dysfunction, skeletal abnormality is the third most commonly observed feature in SDS. Metaphyseal dysplasia in SDS was reported by Burke *et al*. in 1967 and Pringle *et al*. in 1968. Metaphyseal dysostosis has been detected before birth in SDS (Beşer *et al*., 2014). 76% of SDS patients presented with short stature and 49-73% presented with skeletal abnormalities, including metaphyseal dysplasia or dysostosis (46%), thorax narrowing and dystrophy (32%), osteoporosis or osteopenia, scoliosis, and severe bone complications that required surgery (9%) (Ginzberg *et al*., 1999; Hashmi *et al*., 2011; Donadieu *et al*., 2012).

**Skeletal development.** Skeletal abnormalities in SDS are varied in severity and location with age (Mäkitie *et al*., 2004). Longitudinal radiographic analyses of SDS patients showed a delayed appearance of secondary ossification centers that eventually normalized with age. The metaphyses showed widening and irregularity in late childhood. The growth plates showed progressive thickening and irregularity associated with asymmetrical growth. Although metaphyseal and growth plate abnormalities worsened with age, epiphyseal maturation eventually normalized with age. Due to defects in skeletal development, children with SDS often present with short ribs, narrowed thorax, short limbs, short stature, coxa valga, and bowing legs.
Cases of polydactyly are rare. Thinning of cortical bone, osteopenia, and early signs of osteoporosis have been observed before 5 years of age.

**Osteoporosis.** SDS patients are at risk of early onset osteopenia and osteoporosis, characterized by reduced bone mineral density and content, fragility fractures, and vertebral deformities (Toiviainen-Salo et al., 2007). Osteoporosis in SDS is caused by low bone turnover due to reduced number of osteoclasts and osteoblasts, resulting in reduced amount of osteoid and reduced trabecular bone volume.

**Cognitive and Neurological Abnormalities**

**Cognitive abnormalities.** Due to developmental delay, the cognitive function in SDS patients is impaired as indicated by lower intelligence quotient (IQ), learning difficulties, behavioral issues, and social challenges (Kent et al., 1990; Kerr et al., 2010). Cognitive impairment affects both children (65%) and adults (76%) with SDS (Perobelli et al., 2012). Children with SDS can develop behavioral disorders including attention deficit hyperactivity disorder, oppositional defiant disorder, and behavioral issues (withdrawn, somatic complaints, anxiety, depression, acting out, and impulsive, uncooperative, or aggressive behaviors) (Kerr et al., 2010). Moreover, children with SDS have an increased prevalence (6%) of pervasive developmental disorder spectrum. Neuropsychological tests showed lower intellectual reasoning in children with SDS. 20% of children with SDS presented with intellectual disability, 20% presented with severely impaired perceptual intelligence, and 12% presented with severely impaired verbal intelligence. Higher-order language skills, perceptual skills, memory, attention, and academic achievement were also weaker in children with SDS. Due to combined cognitive, behavioral, and medical difficulties, functional independence was also lower in children with SDS.

**Neurological abnormalities.** Cognitive impairment in SDS stems from alterations in the brain’s structure, biochemistry, and connectivity. In 2008, Toiviainen-Salo et al. showed that the volumes of the brain, grey matter, and white matter were reduced in SDS patients (Toiviainen-Salo et al., 2008). Regions of the brain associated with learning, cognition, verbal memory, and IQ were also reduced in SDS patients (Toiviainen-Salo et al., 2008; Booij et al., 2013). In 2013, Booij et al. showed that increased central striatal dopamine transporters in SDS patients were associated with impaired attention and memory. In 2015, Perobelli et al. showed that increased
cortical thickness and altered connectivity in the grey and white matter were also associated with cognitive impairment. Clusters of altered fibers were shown to interfere with inter- and intra-hemispheric connections essential for verbal skills, perceptual skills, visual-motor integration, memory, and executive functions (Perobelli et al., 2015). Other than the reduced size of the brain, increased dopamine transporters, and altered connectivity, other neurological abnormalities, including chiari malformation type I, cerebellar tonsillar ectopia, and myopathy and hypotonia, have been reported in 5% of SDS patients (Myers et al., 2014).

**Cardiac Abnormalities** Cardiomyopathy and congenital heart disease (patent foramen ovale and atrial septal defect, patent ductus arteriosus, ventricular septal defect, atrioventricular septal defect, and dilated cardiomyopathy) have been reported in 11-19% of SDS patients (Hauet et al., 2013; Myers et al., 2014; Le Gloan et al., 2014; Kopel et al., 2011). Heart failure in children with SDS has been suggested to be associated with myocardial fibrosis and necrosis (Nezelof and Le Sec, 1979; Guerrero et al., 1979). Myocardial fibrosis and progression to heart failure have been proposed to be a consequence of increased susceptibility to carditis during concomitant viral infections and myocardial damage in SDS (Savilahti and Rapola, 1984; Hauet et al., 2013). Studies in cardiac function of SDS patients showed alterations in the left ventricular contractile reserve, right ventricular diastolic function, circumferential strain, and left ventricular systolic function, which were suggested to indicate progression to heart failure (Toiviainen-Salo 2008; Ryan et al., 2015).

**Liver abnormalities**

Although often asymptomatic, liver abnormalities in SDS can manifest in early childhood but often improve with age (Brueton et al., 1977; Ginzberg et al., 1999). 15% of SDS patients presented with hepatomegaly and 60% presented with elevated serum aminotransferase. Liver biopsies have shown microvesicular and macrovesicular steatosis, periportal and portal inflammation, periportal, portal, and bridging fibrosis, and glycogenosis in SDS.

**Endocrine abnormalities**

Endocrine abnormalities are variable but have high prevalence in SDS patients (65%) (Myers et al., 2014). SDS patients may present with low stimulated growth hormone levels, abnormal thyrotropin levels, hypopituitarism, hypothyroism, abnormal glucose levels, postprandial
hyperglycemia, elevated follicle-stimulating hormone level, and adrenal insufficiency (Myers et al., 2013).

**Gastrointestinal abnormalities**

Gastrointestinal abnormalities are also variable but less commonly observed in SDS. 8% of SDS patients presented with gastrointestinal abnormalities such as malrotation, bilateral inguinal hemia, and imperforate anus (Myers et al., 2014). 19% of SDS patients presented with transient but severe gastrointestinal symptoms requiring enteral or parenteral feeding or gastrostomy at a median age of 7 months (Donadieu et al., 2012).

**Dental diseases**

44% of SDS patients presented with oral diseases, including carries in primary and permanent dentitions, delayed dental development, enamel hypoplasia, and periodontal disease (Hashmi et al., 2011; Ho et al., 2007; Ginzberg et al., 1999).

**Other Abnormalities**

Other abnormalities include eczema (65%), urological abnormalities (5%) (testicular atrophy and hypospadias), pulmonary hypertension, subglottic stenosis, eye anomaly, ear anomalies (hearing loss), and labial cleft (Myers et al., 2014; Hashmi et al., 2011; Donadieu et al., 2012). In patient leukocytes, age-adjusted telomere length was found to be reduced by 1.4kb/year compared to the normal maximum 60bp/year (Thornley et al., 2002).

**Treatment**

**Primary and secondary complications of bone marrow failure.** Bacterial or fungal infections due to neutropenia are treated with oral antibiotics, while severe recurrent infections or infections concomitant with severe neutropenia are treated with intravenous broad-spectrum antibiotics and granulocyte colony stimulating factor (Dror et al., 2011). Coagulation defects and bleeding episodes due to thrombocytopenia or vitamin K deficiency are treated with xylometazoline nose spray, tranexamic acid, aminocaproic acid, or vitamin K supplement, while severe bleeding and thrombocytopenia are treated with platelet transfusion. Anemia is treated with chronic transfusion and iron-chelation program.
**Hematopoietic stem cell transplantation.** Severe cytopenia, myelodysplastic syndrome, and acute myeloid leukemia require hematopoietic stem cell transplantation (HSCT). HSCT with matched related donor, matched unrelated donor (Okcu et al., 1998), and unrelated umbilical cord blood (Vibhakar et al., 2005) have been reported in SDS. The post-HSCT 5-year event-free survival rate was 60% and 1-year survival rate was 65% (Donadieu et al., 2005; Cesaro et al., 2005). The survival rate of HSCT for non-malignant severe cytopenia was higher than those for myelodysplastic syndrome or acute myeloid leukemia (Donadieu et al., 2012). The survival rate of HSCT using identically matched related donor was also higher than that using matched unrelated donor (Dror et al., 2011). Many complications can arise, including graft failure, graft-versus-host disease, prolonged severe aplasia, infections, veno-occlusive disease, and neurological and renal complications (Dror et al., 2011). As the incidence of regimen-related toxicity in the heart and lungs has been higher in HSCT of SDS patients (Tsai et al., 1990), the intensity of the conditioning regimen can be reduced by avoiding the use of busulfan, cyclophosphamide, and total body radiation to achieve myeloid engraftment and full donor chimerism without grade III-IV graft-versus-host disease (Sauer et al., 2007; Bhatla et al., 2008).

**Pancreatic enzyme supplementation.** Pancreatic insufficiency is treated with pancreatic enzyme replacement with all meals, consisting of lipase (2,000-10000 units/kg body weight/day), protease, amylase or additional H2-receptor antagonist if enzyme supplement fails to improve fat absorption (Dror et al., 2011).

### 1.2 SBDS Gene

**SBDS.** In 2000, Ginzberg et al. analyzed the segregation ratio of 84 patients in 70 families and confirmed that SDS followed an autosomal recessive mode of inheritance. In 2001, Goobie et al. initiated a genome wide scan of 15 families and DNA sequencing mapped the SDS locus to the centromeric region of chromosome 7 and also estimated the incidence of SDS to be 1 in 76,563 according to the incidence of cystic fibrosis and the ratio of SDS-to-cystic fibrosis patients in the Hospital for Sick Children, Toronto over 21 years. In 2002, Popovic et al. refined the mapping of the SDS locus to 7q11. In 2003, Boocock et al. identified the new, uncharacterized causal gene of SDS, Shwachman-Bodian-Diamond syndrome (*SBDS*). *SBDS* is located at chromosome 7q11 with 5 exons spanning 7.9kb, encoding for a 1.6kb mRNA and a 28.8kDa protein with pI 8.9
(Boocock et al., 2003). SBDS mRNA was ubiquitously expressed in many fetal and adult tissues, while the SBDS variant with exon 2 skipping was detected with low mRNA expression.

**SBDSP1.** Located 5.8Mb distally from SBDS is its pseudogene with 97% nucleotide homology, SBDSP1, in a duplicated 305kb genomic segment (Boocock et al., 2003). Recombination between SBDS and SBDSP1 introduces hypomorphic or complete loss of function mutations into SBDS as 74% of SDS-associated alleles were a consequence of gene conversion. 89% of SDS patients carried at least 1 converted allele and 60% carried 2 converted alleles. Although SBDSP1 was identified as an oncogenic long non-coding RNA overexpressed in human colorectal cancer cells as its knockdown resulted in reduced tumor growth and invasion (Shi et al., 2017), the RT-PCR primers and siRNA used were not specific to SBDSP1.

**Common SBDS mutations.** SBDS exon 2 is a mutational hotspot and the most common SDS-associated mutations are c.183_184TA>CT and c.258+2T>C (Boocock et al., 2003). The mutation c.183_184TA>CT introduces a premature stop codon, resulting in a truncated protein (p.Lys62*). The mutation c.258+2T>C disrupts the donor splice site of intron 2 and an upstream cryptic donor splice site at c.251_252 is used, resulting in either an mRNA with 8bp deletion (r.251_258del) encoding a truncated protein due to frameshift (p.Cys84Tyrfs*4), or an mRNA with exon 2 skipping (r.129_258del). To note, the protein expression of both p.Lys62* and p.Cys84Tyrfs*4 were experimentally validated (Orelio et al., 2009). The most common heterozygous mutations in Canada have been c.[183_184TA>CT];[258+2T>C] and c.[183_184TA>CT;258+2T>C];[258+2T>C] (Hashmi et al., 2011). No genotype-phenotype relationship has been found between SBDS mutations and SDS manifestations (Kuijpers et al., 2005; Mäkitie et al., 2004; Donadieu et al., 2012; Hashmi et al., 2011).

**SBDS protein structure.** SBDS is highly conserved across 159 species of archaea and eukaryotes (Boocock et al., 2006). The first reported SBDS protein structures were the SBDS orthologues in Archaeoglobus fulgidus and Methanothermobacterthermautotrophicus (Shammas et al., 2005; Savchenko et al., 2005; Ng et al., 2009). Human SBDS consists of 3 domains (de Oliveira et al., 2010; Finch et al., 2011). The highly conserved N-terminal domain I also known as the FYSH domain (fungal, Yhr087wp, Shwachman) spanning residues S2-S96 contains a mixture of α-helices and β-sheets in the sequence ββaaββaa and is considered to be highly conserved and an SDS-associated mutational hotspot. The central domain II spanning
residues D97-A170 contains three α-helices. The C-terminal domain III spanning residues H171-E250 contains a ferredoxin fold in the sequence βαββαβ which is a putative nucleic acid binding motif. Both N- and C- termini are unstructured. The SBDS protein is localized to both the cytoplasm and nucleus, with greater concentration in the nucleolus, although the function of SBDS in the nucleus is unclear (Austin et al., 2005).

1.3 SBDS Protein Function

Ribosome Biogenesis

Predictions and early evidence. In 2001, Koonin et al. analyzed the archaeal genome and found that the archaeal orthologue of SBDS clustered with RNA processing genes. In 2002 and 2003, Wu et al. and Peng et al. respectively profiled the transcriptome of Saccharomyces cerevisiae and confirmed that the yeast orthologue of SBDS (YRL022C or SDO1) also clustered with RNA processing functions. In 2006, Krogan et al. identified protein interactions through high-throughput affinity-capture mass spectrometry and found that Sdo1p associated with Efl1 and other proteins involved in ribosome biogenesis.

eIF6 release. In April 2007, Menne et al. elucidated the role of SBDS in ribosome biogenesis in Saccharomyces cerevisiae. sdo1Δ mutant yeast strains were either inviable or slow in growth (Winzeler et al., 1999; Menne et al., 2007). Two proteins involved in ribosome biogenesis were integral in determining the function of Sdo1p; Tif6 (translation initiation factor 6, yeast orthologue of human eIF6) and Efl1 (elongation factor like 1). Tif6 is a shuttling factor required for the cytoplasmic export of pre-60S and TIF6 mutations rescued the defects in sdo1Δ. Efl1 is a GTPase required for the release of Tif6 from the pre-60S and efl1Δ is a phenocopy of sdo1Δ as both strains showed reduced 60S subunits. In sdo1Δ, Tif6 nuclear recycling, pre-60S cytoplasmic export, and pre-rRNA processing were impaired. Therefore, Sdo1p was identified as an essential factor in Tif6 release during ribosome biogenesis. In September of the same year, Ganapathi et al. also confirmed the role of SBDS in 60S maturation in HeLa cells. SBDS was localized to the nucleolus during active rRNA transcription and was co-sedimented with 60S and co-precipitated with 28S rRNA (component of 60S), nucleophosmin (ribosomal chaperon), and other ribosomal proteins (Ganapathi et al. 2007).
**GTP stabilizing factor.** In 2011, Finch *et al.* showed that the release of the anti-association factor eIF6 from the pre-60S, catalyzed by SBDS and the GTPase EFL1, required GTP hydrolysis. In 2013, Gijsbers *et al.* analyzed SBDS enzyme kinetics and classified SBDS as a GTP stabilizing factor in the class of GDP/GTP exchange factors, which stabilized the binding of GTP to EFL1 on pre-60S. In 2014, Asano *et al.* showed through isothermal titration calorimetry that domains II-III of SBDS interacted with the intrinsically disordered insertion domain of EFL1.

**Active site protection and proofreading.** In 2015, Weis *et al.* resolved the cryo-EM structures of human SBDS and EFL1 bound to *Dictyostelium discoideum* 60S with or without eIF6 to illustrate the SBDS conformation changes during eIF6 release. SBDS domains were bound to active sites on pre-60S, suggesting that SBDS may function to proofread and protect active sites on the pre-60S and play a role in regulating pre-60S maturation (Weis *et al.*, 2015). Anchored on the P site of the pre-60S, domain I protected and potentially proofread the peptidyl transferase center and polypeptide exit tunnel to prevent premature translation initiation. Domain I also acted as a mimic tRNA to check the function of the P-site through pseudo-translocation as the Sdo1p-bound P-site can activate tRNA binding to the A-site (Ng *et al.*, 2009; Sulima *et al.*, 2014). Binding to the sarcin-ricin loop at the P-stalk base on the pre-60S, domain III protected and potentially proofread active sites of translational GTPases to prevent premature pre-60S maturation (Weis *et al.*, 2015).

**Ribosomopathy**

SDS is classified as a ribosomopathy, a disorder caused by the dysfunction of ribosomal components or factors involved in ribosome biogenesis and maturation (Armistead and Triggs-Raine, 2014). Although their clinical manifestations are heterogeneous, ribosomopathies share the common characteristics of poor growth, bone marrow failure, developmental impairment, predisposition to malignancy, and extra-ribosomal consequences such as cell cycle arrest and mitotic spindle anomalies. The phenotypic heterogeneity of ribosomopathies has been proposed to be a consequence of the differential expression of ribosome-associated genes (Kondrashov *et al.*, 2011), resulting in cell type-specific ribosomal composition and translational control of specific mRNAs. For instance, the translation of *Hox* genes involved in embryonic patterning requires RPL38 to recruit ribosomes to RNA regulons in the 5’ UTR, similar to the internal
ribosome entry site (IRES) (Xue et al., 2015). Moreover, IRES-dependent translation is also affected by DKCI associated with X-linked dyskeratosis congenita and ribosomal proteins associated with Diamond-Blackfan anemia (Yoon et al., 2006; Horos et al., 2012). Similar translational control through an RNA element may also explain the various defective cell types in SDS.

**Extra-ribosomal Consequences of SBDS Deficiency**

Affinity capture and mass spectrometry has identified SBDS interacting proteins with functions other than ribosome biogenesis, suggesting that SBDS may have multiple functions (Ball et al., 2009). SBDS deficiency has been shown to affect multiple biological processes, including rRNA processing (Ganapathi et al., 2007; Rujkijyanont et al., 2009; Luz et al., 2009; Moore et al., 2010), translation of upstream open reading frame (uORF)-dependent protein isoforms (In et al., 2016), actin remodeling and signaling (Wessels et al., 2006; Orelio and Kuijpers, 2009; Leung et al., 2011), microtubule stability (Austin et al., 2008; Orelio et al., 2009; Lombardo et al., 2010), p53-mediated apoptosis and senescence (Elghetany and Tarek, 2002; Kuijpers et al., 2005; Austin et al., 2008; Tourlakis et al., 2012; Tourlakis et al., 2015), Fas-mediated apoptosis (Dror et al., 2001; Rujkijyanont et al., 2008; Watanabe et al., 2009; Ambekar et al., 2010), production of reactive oxygen species (Ambekar et al., 2010; Sen et al., 2011), endoplasmic reticulum stress (Ball et al., 2009; Chatterjee et al., 2015), mitochondrial respiration (Henson et al., 2013; Ravera et al., 2016), and hyperactivation of mTOR and STAT3 (Ravera et al., 2016; Bezzzerri et al., 2016). Whether some of these processes directly involve SBDS or are consequences of reduced ribosome biogenesis remains to be elucidated.

**Actin remodeling.** In 1978, Thong first reported impaired neutrophil chemotaxis in an SDS patient (Aggett et al., 1979; Ruutu et al., 1984; Repo et al., 1987; Szuts et al., 1984). In 2004, Stepanovic et al. demonstrated through spatial gradients of chemotactant fMLP that chemotactic orientation of human SDS polymorphonuclear leukocytes was impaired. In 2006, Wessels et al. showed that SBDS was localized at the anterior pseudopod and newly formed lateral pseudopods of *Dictyostelium discoideum* during chemotaxis. In 2009, Orelio and Kuijpers showed co-localization of SBDS with F-actin and Rac2, a signaling component in actin remodeling, in normal human neutrophils, and delayed directional F-actin polarization in human SDS neutrophils during chemotaxis. In 2011, Leung et al. showed reduced RANKL-TRAF6-
NFkB signaling and reduced Rac2 expression led to impaired monocyte migration and fusion required for osteoclast differentiation in myeloid cell-targeted Sbds conditional knockout mice. As the mechanism between SBDS and Rac2 signaling remains to be elucidated, it has been proposed that SBDS may mediate actin remodeling or SBDS may be localized to regions of increased translation such as actin reorganization.

**Microtubule stabilization.** Other than actin remodeling, SBDS also affects the stability of microtubules. In human bone marrow stromal cells, HeLa cells, and cord blood-derived neutrophils, full-length and truncated SBDS were found to co-localized with the mitotic spindle and centrosomes (Austin et al., 2008; Orello et al., 2009). In human SDS bone marrow stromal cells and lymphoblasts and SBDS siRNA knockdown skin fibroblasts, mitotic abnormalities (multiple centrosomes, multipolar mitotic spindles, and aneuploidy accumulation), cell cycle arrest at G1, and apoptosis were found (Austin et al., 2008). In particular, SBDS increased microtubule stability upon dilution-induced microtubule depolymerization *in vitro*. In a follow-up study, shortened mitotic spindles and reduced spindle acetylation were found in SBDS deficiency, indicating reduced microtubule stability (Lombardo et al., 2010). Wild-type SBDS increased polymerization rate and promoted microtubule bundling of polymerized microtubule *in vitro*, while mutant SBDS reduced microtubule bundling. Treating *SBDS* knockdown human CD34+ cells with taxol, a microtubule stabilizer, rescued proliferation, differentiation, and hematopoietic progenitor colony formation. Both mitotic abnormalities and microtubule instability have been proposed to be mechanisms driving the genomic instability, cytogenetic abnormalities, and malignant transformation associated with SBDS deficiency. It has also been proposed that SBDS may mediate the interconnecting organization of both actin and microtubule.

**P53-mediated apoptosis in hematopoietic cells.** In 2002, Elghetany and Tarek showed that p53 protein was overexpressed in 78% of the bone marrow biopsies from SDS patients, especially in immature cells but was absent in mature cells (band and segmented neutrophils, megakaryocytes, and nucleated red cells). In 2005, Kuijpers et al. confirmed the absence of apoptotic human SDS neutrophils. In 2008, Austin et al. through subjecting human SDS bone marrow stromal cells to ionizing radiation showed that the cell cycle arrested at G1, as indicated by increased p53 protein expression and p21 (p53 downstream mediator) mRNA expression in the apoptotic G1 population. Austin et al. also proposed that cell cycle arrest may be a protective mechanism that
eliminated abnormal cells with DNA damage (2008). In 2015, Zambetti et al. showed through SDS neutropenia murine model, created by transplanting the fetal liver of hematopoietic progenitor-targeted Sbds conditional knockout mice (Sbds^{f/f}) into recipient mice, that terminal differentiation was impaired at the myelocyte-metamyelocyte stage due to p53-mediated apoptosis as indicated by increased mRNA expression of p53 and p53 targets, increased p53 protein expression, and increased apoptotic cells in myelocytes and metamyelocytes.

**P53-mediated senescence in pancreatic cells.** In 2012, Tourlakis et al. showed through pancreas-targeted Sbds conditional knockout mice (Sbds^{P/-R126T}) that exocrine pancreatic defects were caused by senescent cell cycle arrest, as indicated by increased senescence-associated β-galactosidase activity, increased expression of senescence-associated mRNA, and increased p53 (senescence activator) protein expression in acinar cells. Knocking out p53 expression in pancreas-targeted conditional knockout mice (Sbds^{P/-R126T};Trp53^{+/-} double knockout mice) only temporarily rescued SDS pancreatic defects, as apoptotic cells and dedifferentiated acinar cells increased as double knockout mice matured. However, knocking out p53 expression in Sbds knockout mice (Sbds^{R126T/R126T};Trp53^{+/-} double knockout mice) rescued hematopoietic progenitors in the fetal liver and reduced neural apoptosis in the fetal brain.

**Fas-mediated apoptosis and reactive oxygen species.** Increased Fas-mediated apoptosis in SDS was first reported in human SDS bone marrow mononuclear cells (Dror et al., 2001). In human SDS bone marrow mononuclear cells and SBDS shRNA knockdown HeLa cells, Fas stimulation resulted in increased Fas expression and Fas-mediated apoptosis and reduced cell growth and CFU-GM count (Dror et al., 2001; Rujkijyanont et al., 2008). Increased Fas-mediated apoptosis has been shown to be due to increased Fas receptors in the plasma membrane and increased production of reactive oxygen species (Watanabe et al., 2009; Ambekar et al., 2010). In SBDS shRNA knockdown HeLa cells and human myeloid cells, Fas stimulation increased the production of reactive oxygen species and number of apoptotic and necrotic cells, while antioxidant treatment rescued Fas-mediated cell death (Ambekar et al., 2010). Similarly in SBDS knockdown K562 cells, oxidative stress impaired erythroid differentiation, while antioxidant treatment improved proliferation (Sen et al., 2011). Oxidative stress can be contributed by the endoplasmic reticulum stress reported in SBDS siRNA knockdown in human embryonic kidney cell line (Ball et al., 2009) and human SDS lymphoblastoid cells (Chatterjee et al., 2015), as the endoplasmic reticulum is an oxidizing environment that promotes protein
folding and disulfide bond formation (Malhotra and Kaufman, 2007). Oxidative stress is also contributed by impaired mitochondrial respiration (Henson et al., 2013; Ravera et al., 2016).

**Mitochondrial Respiration.** In sdo1Δ yeast, although global protein production was reduced in SDS, the protein expression of Por1 (anion channel of the mitochondrial outer membrane, yeast orthologue of human VDAC1) was increased. Increased Por1 protein expression resulted in reduced growth of sdo1Δ yeast cultured on glycerol and absence of red pigmentation in sdo1Δ yeast cultured on glucose. As the red pigment was an intermediate in the adenine synthesis pathway that required respiration for production, this indicated that mitochondrial respiration was impaired. Increased protein expression of VDAC1 was also confirmed in SBDS shRNA knockdown of a human erythroleukemia cell line (Henson et al., 2013). SBDS knockdown resulted in reduced mitochondrial membrane potential, reduced oxygen consumption, and increased reactive oxygen species. Impaired respiration in the mitochondria has been shown to be due to reduced activity of complex IV in the electron transport chain (Ravera et al., 2016). In human SDS lymphocytes and lymphoblasts, reduced complex IV activity resulted in reduced ATP production and increased AMP accumulation. This energy stress led to the activation of response pathways AMP-activated protein kinase and PI3K/AKT/mTOR, which activated the glycolysis as an alternative method of energy production. Hyperactivation of mTOR and STAT-3 pathways were also found in B cells, granulocytes, monocytes, T cells, NK cells (Bezzerri et al., 2016).

**SDS Patients without Biallelic SBDS Mutations**

Cases of clinically diagnosed SDS lacking biallelic SBDS mutations have been reported, with some cases of confirmed normal SBDS protein expression (Nakashima et al., 2004; Woloszynek et al., 2004; Hashmi et al., 2011; Myers et al., 2014). The Canadian registry suggested that these patients presented with more severe hematological dysfunction but milder pancreatic dysfunction compared to SDS patients with biallelic SBDS mutations (Hashmi et al., 2011). These SDS patients had a higher incidence of severe aplastic anemia, severe bone marrow failure, treatment for cytopenia, reduced hemoglobin levels, and elevated fetal hemoglobin after 1 year of age with increased mean fetal hemoglobin levels, although there was no occurrence of pancreatic enzyme treatment. In contrast, the North American registry suggested that these SDS patients presented with milder hematological dysfunction but more severe pancreatic dysfunction (Myers et al.,
2014). These patients had a lower incidence of hypocellular bone marrow and clonal marrow cytogenetic abnormalities and no occurrence of bone marrow dysplasia or malignant transformation, although there was a higher incidence of enzyme treatment and failure to thrive. The discrepancy between the two registries may be due to varying genetic mutations and underlying molecular mechanisms affected. Other findings of these SDS patients include increased incidence of gait irregularities and no occurrence of elevated liver transaminases (Hashmi et al., 2011; Myers et al., 2014). Recently, DNAJC21 was identified as another causal gene of SDS and biallelic DNAJC21 mutations were found in SDS patients lacking SBDS mutations (Tummala et al., 2016; Dhanraj et al, 2017). Similar to SBDS, DNAJC21 plays a role in ribosome biogenesis through interacting with ribosomal cofactors and nuclear export factor.

**SDS Disease Models**

**Bone marrow samples.** Various disease models have been used to study SDS. Primary cells derived from patients are ideal disease models used to investigate the human hematopoietic defects in SDS. Immunophenotyping showed reduced CD34+ hematopoietic progenitors (Dror and Freedman, 1999), myeloid progenitors, and granulocytes, although monocytes were increased (Mercuri et al., 2015). May-Grünwald Giemsa staining of patients bone marrow samples showed no difference in early myeloid progenitors, reduced myeloblast and promyelocytes, increased intermediate myelocytes, and reduced terminally differentiated cells (metamyelocyte, band neutrophils, and segmented neutrophils) (Orelio et al., 2009). Together, these results indicated that hematopoietic defects in SDS could be found in the more immature hematopoietic progenitors and in the more mature neutrophils.

**CFU assays.** Using patient hematopoietic progenitors in CFU (colony forming unit) assays, defects in the hematopoietic potential of hematopoietic progenitors were demonstrated through reduced myeloid CFUs, including CFU-G (granulocyte; Saunders et al., 1979), CFU-GM (granulocyte, monocyte; Suda et al., 1982; Dror and Freedman, 1999; Kuijpers et al., 2005), CFU-GEMM (granulocyte, erythroid, monocyte, megakaryocyte; Dror and Freedman, 1999), CFU-E (erythroid; Suda et al., 1982), and BFU-E (blast forming unit-erythroid; Kuijpers et al., 2005; Sen et al., 2011). CFU assays of SBDS knockdown human cell lines also showed reduced myeloid and erythroid colony count and size (Sezgin et al., 2013).
**Primary cell lines.** Culture of normal hematopoietic progenitors on SDS stromal cells showed reduced non-adherent cells and reduced CFU-GM, indicating impairment in the bone marrow stromal environment of SDS patients (Dror and Freedman, 1999). In normal human myeloid leukemia cell line and cord blood CD34+ hematopoietic progenitors, the mRNA and protein expression of SBDS was progressively reduced as myeloid progenitors differentiate into neutrophils, suggesting that in SDS, mutant SBDS availability became progressively insufficient for neutrophil development (Orelio et al., 2009).

**Mice.** As human primary cells are short-lived, scarce in rare disease, and unavailable in certain cell types, animal models such as mice and zebrafish are useful in vivo disease models. Mice offer a phylogenetically close in vivo model that recapitulates the defects in SDS. Full Sbds knockout mice (Sbds<sup>−/−</sup>) were embryonically lethal and embryonic developmental defects were evident by E6.5 as shown by reduced Oct4 expression (Zhang et al., 2006). Mutant mice (Sbds<sup>R126T/R126T</sup> and Sbds<sup>R126T/−</sup>) with less severe mutations (c.377G>C, p.R126T) recapitulated SDS defects, such as a small and fatty fetal pancreas, reduced myeloid progenitors and granulocytes in the fetal liver, reduced ossification in the metacarpals, asphyxiating thoracic dystrophy, increased apoptosis and reduced neuronal progenitors in the fetal brain, and reduced saccule expansion in the fetal lung (Tourlakis et al., 2015).

**Hematopoietic defects in mice.** Hematopoietic defects after birth were recapitulated by transplanting Sbds shRNA-silenced murine hematopoietic progenitors into recipient mice (Rawls et al., 2007). These SDS mice showed impaired homing, reconstitution, and proliferation of myeloid progenitors and B-lymphocytes, as indicated by reduced short-term and long-term circulating transduced cells, reduced secondary hematopoietic engraftment, reduced CFU and CFU-GM count, and reduced circulating B-lymphocytes. Another mouse model recapitulated neutropenia in SDS by transplanting the fetal liver of hematopoietic progenitor-targeted Sbds conditional knockout mice (Sbds<sup>ff</sup>) into recipient mice (Zambetti et al., 2015). These mice were presented with neutropenia marked by reduced neutrophils in peripheral blood and bone marrow. Neutropenia was a consequence of impaired terminal differentiation of myelocytes and metamyelocytes, as shown by increased numbers of myelocytes and metamyelocytes but reduced more mature segmented neutrophils, upregulation of primary granule genes but downregulation of more mature secondary and tertiary granule genes, downregulation of activators of terminal myeloid differentiation, and failure to exit the cell cycle to enter terminal differentiation.
Impaired terminal differentiation was a consequence of p53-mediated apoptosis, as indicated by increased gene expression of p53 and p53 targets, increased p53 protein expression, and increased apoptotic cells in myelocytes and metamyelocytes.

**Pancreatic defects in mice.** Pancreatic defects after birth were recapitulated using pancreas-targeted Sbds conditional knockout mice (Sbds<sup>P−/−</sup>, Sbds<sup>P−/R126T</sup>). These SDS mice showed a small and fatty exocrine pancreas with reduced acini count, reduced pancreatic enzyme gene expression, reduced pancreatic zymogens, and reduced granule proteins (Tourlakis et al., 2012). Exocrine pancreatic defects were a consequence of senescent cell cycle arrest, as indicated by increased senescence-associated β-galactosidase activity, increased expression of senescence-associated genes, and increased p53 protein expression in acinar cells. Knocking out p53 expression in pancreas-targeted Sbds conditional knockout mice (Sbds<sup>P−/R126T;Trp53−/−</sup>) only temporarily rescued SDS pancreatic defects, as apoptotic cells and dedifferentiated acinar cells increased as double knockout mice matured. However, knocking out p53 expression in Sbds knockout mice (Sbds<sup>R126T/R126T;Trp53−/−</sup>) rescued hematopoietic progenitors in the fetal liver and reduced neural apoptosis in the fetal brain. p53 could be arresting the cell cycle in response to DNA damage and initiating apoptosis if DNA damage is irreparable. Other than the exocrine pancreas, the endocrine pancreas was also impaired, as indicated by reduced islet size and reduced glucose tolerance. The growth of mice was also impaired, as shown by reduced body mass and absence of body fat.

**Zebrafish.** Zebrafish also offer an in vivo investigation of the pancreatic and skeletal defects, but mainly neutrophilic defects of SDS. Created using morpholino knockdown of sbds, SDS zebrafish recapitulated the early exocrine pancreatic and neutrophilic defects as shown by the absence of the medial exocrine pancreas in the larva and the absence of granulocytes in the embryo (Venkatasubramani and Mayer, 2008). Another SDS zebrafish created using morpholino targeting the start codon of sbds recapitulated pancreatic, neutrophilic, and skeletal defects, as shown by reduced proliferation of pancreatic progenitors and neutrophils, and malformed ear bone and cartilages of the lower jaw and gill arches (Provost et al., 2012). Unlike SDS mice, co-morpholino knockdown of both sbds and p53 in zebrafish and morpholino knockdown of sbds in p53<sup>−/−</sup> mutant zebrafish did not rescue SDS defects. sbds knockout zebrafish (sbds<sup>nu132/nu132</sup>) was also created, recapitulating the growth and neutrophilic defects, as shown by reduced zebrafish size and neutrophil count, although the pancreas appeared normal (Oyarbide et al., 2012).
Atrophy of the pancreas, liver, digestive tract, and eye was also found (Oyarbide et al., 2016).

**hiPSCs.** As species-specific differences in animal models may render phenotypes different from the human disease and drugs effective in mice may be ineffective in humans, human pluripotent stem cells (hPSCs), human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs), are useful *in vitro* disease models that complement animal models. In 2008, Park *et al.* generated SDS hiPSCs derived from patient bone marrow mesenchymal cells using retroviral transduction of OSKM, reprogramming factors that are discussed in later sections. Genotype was characterized by karyotype analysis, DNA fingerprint analysis, and DNA sequencing of SDS-associated mutations (Park *et al.*, 2008). Pluripotency was confirmed by protein markers and gene expression. Differentiation potential was evaluated by gene expression of the three germ layers and teratoma formation in immunodeficient mice. In 2013, Tulpule *et al.* confirmed reduced SBDS protein expression and aberrant polysome profiles (reduced ratios of 80S:40S and 60S:40S) in SDS hPSCs.

**Protease-mediated auto-digestion in hPSCs.** SDS hPSCs recapitulated some aspects of the pancreatic and hematopoietic defects (Tulpule *et al.*, 2013). SDS hPSC-derived acinar cells showed reduced cell numbers, reduced pancreatic enzyme gene expression, increased apoptotic cells, and abnormal morphology. SDS hPSC-derived ductal cells also showed reduced cell count and reduced ductal gene expression. SDS hPSC-derived myeloid cells showed reduced CD45+ cell count, reduced total CFUs, and downregulation of hematopoietic transcription factors. The defects in both SDS hPSC-derived acinar and myeloid cells were a consequence of protease-mediated auto-digestion and resulting apoptosis. In both hPSC-derived acinar and myeloid cells, protease-containing granules were abnormal as shown by increased size and number of pancreatic zymogen granules, which was opposite of the mouse findings, and increased number of myeloid primary azurophilic granules. Treatment with protease inhibitors rescued both hPSC-derived acinar and myeloid cells, indicating that protease released from impaired granules caused protease-mediated auto-digestion and consequent apoptosis.

**Impaired definitive hematopoiesis in hiPSCs.** Comparing Sendai virus-generated hiPSCs derived from an SDS patient (P357) with those derived from a healthy patient (N551), hematopoiesis showed impairment in the embryonic definitive program, although the
extraembryonic primitive program showed no differences (Luca, 2015). Primitive hematopoietic differentiation of SDS hiPSCs showed no differences in generating primitive mesoderm, primitive hemangioblast, primitive early hematopoietic progenitors, and primitive mature blood cells. However, definitive hematopoietic differentiation of SDS hiPSCs showed reduced abundance in successive hematopoietic populations following the definitive mesoderm. SDS hiPSC-derived definitive hemogenic and vascular endothelial population was reduced, as shown by reduced CD34+/CD43−/CD45− population in day 6, 9, and 12 embryoid bodies (EBs) and reduced CD34+/CD31+ population in day 9 EBs. SDS hiPSC-derived definitive early hematopoietic progenitors were also reduced, as shown by reduced CD34+/CD45+ population in day 12 and 15 EBs and reduced CD34+/CD43+ population in day 12 and 15 EBs. SDS hiPSC-derived definitive myeloid progenitors were also reduced, as shown by reduced CD34+/CD45+ population in day 12 and 15 EBs. SDS hiPSC-derived definitive granulocytic and monocytic progenitors were also reduced, as shown by reduced CD34−/CD45+ population in day 12 and 15 EBs. hiPSC-derived definitive CFUs were also reduced, as shown by reduced large and small colonies of CFU-GEMM, CFU-GM, and BFU-E.

2 Induced Pluripotent Stem Cells

Reprogramming factors. Induced pluripotent stem cells (iPSCs) were first reprogrammed from mouse and human somatic cells to pluripotency by overexpressing a combination of pluripotency transcription factors OCT4 (O), SOX2 (S), KLIF4 (K), and C-MYC (M) (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). Other combinations of reprogramming factors are also sufficient, such as OCT4, SOX2, NANOG, and LIN28 (Yu et al., 2007), although activation of OCT4 or SOX2 expression is considered indispensable in reprogramming (Huangfu et al., 2008; Marson et al., 2008; Chen et al., 2011; Buganim et al., 2012). OSK form part of the core transcriptional circuitry of pluripotency, which auto-regulate and co-target regulatory regions of pluripotency genes (Nichols et al., 1998; Boyer et al., 2005; Loh et al., 2006; Masui et al., 2007; Kim et al., 2008). Unlike OSK, C-MYC is not part of the core transcriptional circuitry but rather enhances reprogramming efficiency potentially through promoting cell proliferation and chromatin remodeling (Kim et al., 2008; Wernig et al., 2008). During reprogramming, OKM also repress the transcription of somatic genes, while KLF4 is essential in initiating mesenchymal-to-epithelial transition (Tiemann et al., 2014; Polo et al., 2012; Sridharan et al., 2009; Chen et al., 2011).
Delivery methods. The first generation of iPSCs was reprogrammed using retroviral delivery of reprogramming factors (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). Integrative delivery methods, such as retrovirus, rely on the genomic integration of reprogramming factors with the risk of insertional mutagenesis and incomplete silencing of reprogramming factors, which leads to the reactivation of reprogramming factors and inhibition of differentiation potential (Toivonen et al., 2013; Brouwer et al., 2016). Moreover, the transcriptional profile of hESCs is more similar to non-integrated hiPSCs than integrated hiPSCs (Soldner et al., 2009). Therefore, non-integrative delivery methods are preferred, such as Sendai virus, an RNA virus that introduces reprogramming factors as RNA without integrating into the host genome (Fusaki et al., 2009).

Reprogramming mechanism. Introducing reprogramming factors into the somatic cell induces a series of epigenetic and transcriptional events that revert the epigenome and transcriptome to a pluripotent state (Maherali et al., 2007; Mikkelsen et al., 2008; Koche et al., 2011; Lister et al., 2011; Smith et al., 2010; Nishino et al., 2011; Buganim et al., 2012; Polo et al., 2012; Zunder et al., 2015). During early reprogramming, rapid histone methylation modifies the regulatory regions of pluripotency and developmental genes to a euchromatic state and somatic genes to a heterochromatic state, resulting in transcriptional changes, increased proliferation, reduced cell size, and loss of somatic identity. During the intermediate stage, the partially reprogrammed population either reverts to the original somatic cell type or dedifferentiates to iPSCs, progressing through mesenchymal-to-epithelial transition. During late reprogramming, DNA demethylation of regulatory regions of pluripotency genes establishes stable iPSCs. The result of complete reprogramming are iPSCs with characteristics similar to embryonic stem cells (ESCs).

hiPSC characteristics. Before the discovery of iPSCs, ESCs isolated from the inner cell mass of the blastocyst were the primary source of pluripotent stem cells (Evans and Kaufman, 1981; Martin, 1981; Thomson et al., 1998). Both ESCs and iPSCs are defined by their ability to self-renew and differentiate to almost any cell type (Zhao et al., 2009; Tiscornia et al., 2011; Brouwer et al., 2016). The morphology of iPSCs is a compact colony with distinct borders, containing cells of high nucleus-to-cytoplasm ratio. iPSCs are characterized by their expression of pluripotency markers, high alkaline phosphatase activity, and ability to differentiate into cells of all three germ layers. iPSCs are commonly assessed for normal karyotype and downregulation of transgene expression. Epigenetic profiles of iPSCs assessing DNA demethylation of
pluripotency genes and methylation of somatic genes are used to screen for complete reprogramming.

**Comparison to ESCs.** Transcriptional and epigenetic profiles of iPSCs are similar to ESCs, including mRNA and miRNA profiles, DNA and histone methylation patterns, and chromosome organization. (Marchetto et al., 2009; Chin et al., 2009; Doi et al., 2009; Chin et al., 2010; Guenther et al., 2010; Stadtfeld et al., 2010; Bock et al., 2011; Lister et al., 2011; Ohi et al., 2011; Mills et al., 2013; Shao et al., 2013; Choi et al., 2015; Krijger et al., 2016). Despite similarities with ESCs, iPSC-specific transcriptional signatures and differentially methylated regions capable of transmission through differentiation were reported. However, transcriptional and epigenetic differences between iPSCs and ESCs were diminished upon continuous culture. Moreover, comparing genetically matched iPSCs with ESCs, transcriptional differences were mainly due to genetic background and some reports considered iPSCs transcriptionally and epigenetically indistinguishable from ESCs.

**Differentiation variability.** Similar to ESCs, iPSC lines derived from multiple donors have transcriptional, epigenetic, and differentiation variability mainly due to genetic background, although inherent variability can also be found among iPSC lines derived from the same donor due to varying epigenetic reprogramming events (Osafune et al., 2008; Chen et al., 2009; Hu et al., 2010; Bock et al., 2011; Lister et al., 2011; Mills et al., 2013; Wu et al., 2014). Differentiation variability can be minimized with the use of standardized differentiation protocols and stringent quality screening for complete reprogramming, using criteria such as differentiation potential, histone deposition, and CD30 expression. (Bock et al., 2011; Boulting et al., 2011; Abujarour et al., 2013; Wu et al., 2014). In particular, CD30 was found to be a pluripotency marker that identified hiPSCs with high expression of traditional pluripotency markers and distinguished undifferentiated hiPSCs from differentiated and partially reprogrammed hiPSCs, hence CD30 has been used in sorting for high quality hiPSCs and as a target in lentiviral transduction of hiPSCs (Abujarour et al., 2013; Friedel et al., 2016).

**Somatic memory.** Due to incomplete DNA methylation and incomplete silencing of somatic genes distant from reprogrammed genes, iPSCs were shown to retain residual somatic cell epigenetic and transcriptional memory, which skewed the differentiation potential in favor of the original somatic cell type (Ghosh et al., 2010; Ohi et al., 2011; Lister et al., 2011; Bar-Nur et al., ...
Upon continuous passage, somatic epigenetic and transcriptional memory was shown to either persist or become lost and contribute minimally to transcriptional variation (Kim et al., 2011; Polo et al., 2010; Rouhani et al., 2014). However, comparing iPSCs reprogrammed from different genetically-matched somatic cell types, transcriptional and differentiation variation in iPSCs was found to be mainly contributed by genetic and epigenetic differences (Kajiwara et al., 2012; Shao et al., 2013; Rouhani et al., 2014). Although the effect of somatic memory is debatable, reprogramming from the somatic cell type closest to the differentiated cell of interest has been preferred.

**Genomic instability.** Other than differentiation variability and somatic memory, another consideration in using pluripotent stem cells (PSCs), or any cultured cell line, is genomic instability (Martins-Taylor and Xu, 2012). The incidence of recurrent and non-recurrent karyotypic aberrations and copy number variations was shown to be comparable between iPSCs and ESCs, while the incidence of copy number variations was higher in both PSCs than cultured cell lines (Martins-Taylor et al., 2011; Taapken et al., 2011; Hussein et al., 2011; Laurent et al., 2011). Genomic aberrations other than those originated from donor cells can be acquired during culture adaptation, integrative reprogramming, prolonged culture, and differentiation, resulting in altered gene expression, proliferation rate, and differentiation potential (Mayshar et al., 2010; Martins-Taylor et al., 2011; Hussein et al., 2011; Laurent et al., 2011; Cheng et al., 2012; Mills et al., 2013). Therefore, reprogramming using non-integrative delivery and regular karyotypic monitoring are necessary in minimizing the effects of genomic instability.

**hiPSC Models of Inherited Bone Marrow Failure.** hiPSCs have been used to model various bone marrow failure syndromes (Jung et al., 2015). Fanconi anemia hiPSCs, which required functional DNA repair and absence of oxidative stress during reprogramming, were able to recapitulate the erythroid and myeloid defects, while identifying earlier defects in the hemangioblasts and hematopoietic progenitors. Inhibition of INF-γ and TNFα signaling pathways also rescued the hematopoietic defects in patient hiPSCs. Dyskeratosis congenita hiPSCs were able to recapitulate the impairment in hematopoietic differentiation and telomere maintenance in dyskeratosis congenita. Familial platelet disorder hiPSCs were also able to recapitulate the megakaryocytic defect and genomic instability. Diamond-Blackfan anemia hiPSCs were also able to recapitulate the impairment in erythroid differentiation and 40S biogenesis, while identifying earlier defects in multipotent hematopoietic progenitors. Congenital
amegakaryocytic thrombocytopenia hiPSCs were also able to recapitulate the impairment in megakaryocytic differentiation and TPO/MPL signaling. To summarize, hematopoietic differentiation of hiPSCs is useful in studying the formation of hematopoietic cells, especially multipotent progenitors that are rare in inherited bone marrow failures.

3 Hematopoiesis

Myeloid and lymphoid lineages. Hematopoiesis is the generation of blood cells (Doulatov et al., 2012). Hematopoietic cells are divided into the myeloid and lymphoid lineages. The myeloid lineage consists of erythrocytes, megakaryocytes, monocytes, and granulocytes. Erythrocytes are enucleated and contain hemoglobin that transports oxygen and carbon dioxide (Palis, 2014). Megakaryocytes give rise to platelets that function in blood clotting through endomitosis, in which megakaryocytes become enlarged and polyploid without cell division in preparation for platelet release (Machlus et al., 2014). Monocytes mediate the immune response through chemotactic migration and give rise to phagocytic macrophages and antigen-presenting dendritic cells (Geissmann et al., 2010). Characterized by cytoplasmic granules and segmented nucleus, granulocytes (neutrophils, eosinophils, basophils, and mast cells) mediate the immune response through releasing cytotoxic proteins (Doulatov et al., 2012). The lymphoid lineage consists of thymus (T)-lymphocytes, bone marrow (B)-lymphocytes, and natural killer cells which mediate cell-mediated immunity, humoral immunity, and innate immunity respectively.

Colony assay. Hematopoietic cells were first induced using in vivo colony assay by transplanting mouse bone marrow cells to the spleen of irradiated mouse, generating erythroid, granulocytic, and megakaryocytic colonies in the spleen and their progenitor was known as colony-forming unit, spleen (CFU-S) (Becker et al., 1963). in vitro colony assays of human bone marrow cells co-cultured with feeder layers in agar gel generated mostly granulocytic colonies and their progenitor was known as CFU-in culture (CFU-C) (Pike and Robinson, 1970; Moore et al., 1973). Long-term in vitro culture of human bone marrow cells generated erythroid and myeloid CFUs and their progenitor was known as the long-term culture-initiating cell (LTC-IC) (Gartner and Kaplan, 1980; Sutherland et al., 1989). However, LTC-ICs are not considered hematopoietic stem cells due to undetermined engraftment potential.
**Hematopoietic stem cell.** The term hematopoietic stem cell (HSC) was initially coined to describe a progenitor that gave rise to erythroid and megakaryocytic cells in the yolk sac (Maximow, 1909). Successful bone marrow transplant of irradiated mice and guinea pigs led to the current definition of HSC, a cell capable of self-renewal, giving rise to all lineages of blood, and reconstituting recipient bone marrow long-term (Lorenz et al., 1951). Evidence of human HSC was found through transplanting human bone marrow into immunodeficient mice, which resulted in long-term myeloid and lymphoid reconstitution (Lapidot et al., 1992). To reduce DNA damage due to replicative stress, most HSCs are quiescent and some are slowly dividing every 40-45 weeks in human, only becoming active upon stress (Foudi et al., 2009; Wilson et al., 2008; Shepherd et al., 2004; Catlin et al., 2011). Studies in isolating HSCs have defined its marker expression to be CD34+/CD38−/Thy1+/CD45RA− (Civin et al., 1984; Lansdorp et al., 1990; Baum et al., 1992; Bhatia et al., 1997). CD34 is expressed in both HSCs and hematopoietic progenitors, CD90 is expressed in stem cells, and CD45RA and CD38 are expressed in differentiated progenitors.

**Classical hierarchy.** The classical model of the human hematopoietic hierarchy consists of, in order of progression to maturity, HSCs, lineage progenitors, progenitor intermediates, and terminally differentiated cells, with gradual loss of both differentiation potential and proliferative capacity while maturing (Manz et al., 2002; Doulatov et al., 2010). HSCs give rise to multipotent progenitors (MPPs) that are similar to HSCs in multipotent potential but are only capable of short-term engraftment. MPPs give rise to either the common myeloid progenitors (CMPs) or multi-lymphoid progenitors (MLPs). In the myeloid lineage, CMPs give rise to either granulocytic-monocytic progenitors (GMPs) or megakaryocytic-erythrocytic progenitors (MEPs) (Manz et al., 2002). MLPs give rise to both lymphoid (T-, B-, and natural killer cells) and myeloid progenitors (monocytes, macrophages, dendritic cells) (Doulatov et al., 2010).

**Revised hierarchy.** However, the hematopoietic hierarchy is more complex than expected. Single cell assays of sorted blood compartments showed that the hematopoietic hierarchy of the fetal liver was distinct from that of the adult bone marrow (Notta et al., 2016). The prenatal hierarchy progressed through stages of HSCs, MPPs, oligopotent progenitors, and unipotent progenitors, but the adult hierarchy lacked oligopotent progenitors and progressed directly from multipotent to unipotent progenitors. For instance, in the prenatal hierarchy, MEPs arose from multipotent and oligopotent progenitors, whereas in the adult hierarchy, MEPs arose directly
from MPPs. Fetal-to-adult shift in blood maturation is one of the many developmental switches observed in blood, another embryonic-to-fetal shift in blood formation is also discussed below.

**Hematopoiesis.** There are two major programs of hematopoiesis, the primitive and definitive programs (Sabin, 1920; Dieterlen-Lievre, 1975). The primitive program is initiated extra-embryonically in the blood islands of the yolk sac within the first three weeks of human gestation to provide a transient supply of mainly primitive erythrocytes, but also primitive macrophages and megakaryocytes, to the embryo proper before hematopoiesis is established within the embryo proper through the definitive program (Sabin, 1920; Palis et al., 1999; Xu et al., 2001). After the onset of the primitive program, the definitive program is initiated intra-embryonically at various sites, with the most well characterized site being the dorsal aorta of the aorta-gonad-mesonephros (AGM) region, providing a source of definitive hematopoietic stem cells that generate all hematopoietic lineages, with the distinguishing lineage being the lymphoid lineage (Dieterlen-Lievre, 1975; Huyhn et al., 1995; Medvinsky and Dzierzak, 1996; Tavian et al., 1996).

**Primitive hematopoiesis.** The primitive program of hematopoiesis was earlier known as primitive erythropoiesis as the first visible hematopoietic cells were primitive erythroid cells emerging from the blood islands attached to the endothelium of the yolk sac (Sabin, 1920). Hematopoietic characterization in mice has shown that the primitive program generated primitive erythroid, macrophage, and megakaryocytic progenitors in the yolk sac (Palis et al., 1999). In mice, primitive erythroid progenitors emerge transiently during early embryonic development and generate primitive erythroid colony forming cells in colony assays. Primitive erythroblasts are larger and nucleated upon entry into circulation, but after maturing in circulation, they become erythrocytes that are smaller and enucleated (Kingsley et al., 2004). In humans, the primitive erythroid lineage is identified by the expression of embryonic ζ-globin but is mainly distinguished from the definitive erythroid lineage by the expression of embryonic ε-globin, which is part of the primitive-to-definitive hemoglobin switch (Peschle et al., 1985).

**Hemoglobin switching.** Changes in globin gene expression known as hemoglobin switching distinguish the primitive from the definitive program of erythropoiesis (Fritsch et al., 1980; Peschle et al., 1985; Palis, 2014). Hemoglobin is a tetramer composed of two alpha- and two beta-globin subunits. Both alpha and beta-globin loci consist of globin genes that are
developmentally regulated. Alpha-globin switching occurs at 5-6 weeks of human gestation, in which primitive erythroid cells switch from expressing primitive embryonic ζ-globin to α-globin. The first beta-globin switching occurs at 8 weeks of human gestation, in which erythroid cells switch from expressing primitive embryonic ε-globin to expressing predominantly definitive fetal γ-globin. The switch from embryonic to fetal globin marks the switch from primitive to definitive erythroid program. The second beta-globin switching occurs after birth, in which erythroid cells switch from expressing fetal γ-globin to adult β-globin.

**Primitive hematopoietic cells.** In mice, both primitive erythroid and macrophage progenitors emerge concomitantly from the proximal region of the egg cylinder at mid-primitive streak stage before the formation of the yolk sac (Palis et al., 1999). Unlike adult macrophages that arise through stages of promonocyte and monocyte, mouse fetal macrophages directly arise from primitive macrophage progenitors and have higher proliferative potential and no peroxidase or 5'-nucleotidase activity (Takahashi et al., 1989; Naito et al., 1990). In mice, primitive megakaryocytic progenitors also emerge from the yolk sac (Xu et al., 2001). Primitive megakaryocytic progenitors are smaller and mature more rapidly to compensate for reduced ploidy, as compared to definitive megakaryocytic progenitors with increased ploidy capable of generating more platelets.

**Hemangioblast.** The progenitor of the primitive hematopoietic lineage and the vascular lineage is the bi-potential hemangioblast, which was first proposed due to the close physical association and concomitant emergence of both lineages in the yolk sac (Murray, 1932). Evidence of the hemangioblast was found in mESC-derived blast colony forming cells (BLCFCs) with both primitive hematopoietic and endothelial potential, emerging transiently before the onset of hematopoiesis and vasculogenesis (Choi et al., 1998). The hemangioblast expresses the mesodermal markers Brachyury (also known as T) and KDR (kinase insert domain receptor; also known as Flk1, fetal liver kinase 1; vascular endothelial growth factor receptor 2, VEGFR2; and CD309), but not hematopoietic or endothelial markers C-KIT, VE-cad, CD34, or CD45 (Wilkinson et al., 1990; Yamaguchi et al., 1993; Kabrun et al., 1997; Choi et al., 1998; Fehling et al., 2003). In humans, the primitive hemangioblast arises from the mesoderm at the posterior region of the primitive streak and migrates anteriorly to the yolk sac, differentiating into hematopoietic and endothelial progenitors and generating blood islands (Luckett, 1978; Huber et al., 2004).
Other hematopoiesis in the yolk sac. Distinct from the primitive progenitors but also arising from the yolk sac before the emergence of HSCs are the erythro-myeloid progenitors (EMPs) and the lymphoid-primed multipotent progenitors (LMPPs). Found in the mouse yolk sac, EMPs were known as definitive erythroid progenitors (BFU-E) (Palis et al., 1999) due to the predominately expression of definitive fetal γ-globin and the absence of primitive embryonic ε-globin expression (McGrath et al., 2011). In mouse embryos, EMPs emerged from the yolk sac migrate to the fetal liver for maturation (McGrath et al., 2015). Distinct from primitive progenitors, EMPs give rise to erythroid, megakaryocytic, and other myeloid (macrophage, neutrophil, eosinophil and basophil) lineages. Distinct from definitive progenitors, EMPs have short-term engraftment capability and no lymphoid potential. Other than EMPs, LMPPs with lympho-myeloid potential also emerge from the yolk sac, as supported by B- and T-cell progenitors identified in the mouse yolk sac, contrary to earlier understanding that lymphoid cells arise only from HSCs and are restricted to the definitive program (Yoshimoto et al., 2011; Yoshimoto et al., 2012; Böiers et al., 2013).

Definitive hematopoiesis. The definitive program of hematopoiesis originates in the embryo proper, but not the yolk sac, as shown by the presence of quail hematopoietic cells when quail embryos were grafted onto chick yolk sacs (Dieterlen-Lievre, 1975). The definitive program gives rise to the definitive erythroid lineage, which matures through stages of blast forming unit erythroid (BFU-E) and multi-cluster colony forming unit erythroid (CFU-E), distinguished by the expression of fetal γ-globin (Peschle et al., 1985; Palis et al., 1999). The distinguishing feature of the definitive program is the potential to generate HSCs and HSCs capable of engraftment was found on day 32 of human gestation in the dorsal aorta of the AGM region (de Bruijn et al., 2000; Ivanovs et al., 2011).

Hemogenic endothelium. The progenitor of the definitive hematopoietic lineage is known as the hemogenic endothelium (de Bruijn et al., 2000). As aortic clusters and HSCs were found to emerge from the dorsal aortic endothelium, but not the gonads or mesonephros of the AGM, HSCs and the definitive program were proposed to originate from the aortic endothelium (Garcia-Porrero et al. 1995; Tavian et al., 1996; de Bruijn et al., 2000). Lineage tracing of endothelial cells with acetylated low-density lipoproteins labeling in vivo or inducible VE-cad Cre mice supported the notion that aortic clusters and HSCs originated from the aortic endothelium (Jaffredo et al., 1998; Jaffredo et al., 2000; Zovein et al., 2008). Imaging of
zebrafish and mouse embryo confirmed the budding of HSCs from endothelial cells through endothelial-to-hematopoietic transition at the ventral wall of the dorsal aorta (Bertrand et al., 2010; Kissa and Herbomel, 2010; Boisset et al., 2010). Although the hemogenic endothelium was initially considered to be a bi-potential progenitor similar to the hemangioblast, Runx1 enhancer transgenic mice showed that endothelial cells lost endothelial potential and became hematopoietic before HSCs emerged, hence the hemogenic endothelium was recognized as a hematopoietic progenitor originated from the endothelium with endothelial marker expression but without endothelial potential (Swiers et al., 2013). In the human embryo, the definitive program arises from the mesoderm at the posterior region of the primitive streak and migrates anteriorly, but more posteriorly and ventrally than the hemangioblast (Luckett, 1978; Huber et al., 2004).

**Definitive hematopoiesis in the human embryo.** Studies of the human embryo showed that the formation of hematopoietic tissues began in the splanchnopleural mesoderm, where intra-embryonic myeloid potential was detected at day 19 of human gestation, before the onset of blood circulation at day 21, while lympho-myeloid potential was detected at day 24 (Huyhn et al., 1995; Tavian et al., 1996; Tavian et al., 2001). Aortic clusters of CD34⁺ hematopoietic progenitors was detected in the ventral endothelium of the dorsal aorta at day 27 (Tavian et al., 1996; Tavian et al., 2001), while HSCs capable of long-term reconstitution was detected at day 32 (Ivanovs et al., 2011). Hematopoietic stem cells colonized the fetal liver at day 30-40 and hematopoietic potential was lost in the aorta after day 40 (Tavian et al., 1999; Ivanovs et al., 2011).

**Mesoderm induction.** Generation of hematopoietic cells from PSCs in vitro recapitulates the steps in embryonic hematopoietic development in vivo (Ackermann et al., 2015). During gastrulation, the epiblast ingresses at the primitive streak to establish bilateral symmetry and generate the three germ layers, endoderm, ectoderm, and mesoderm (Kinder et al., 1999). Both primitive streak and mesoderm formation in vivo and mesoderm induction in vitro require BMP4 (bone morphogenic protein 4), bFGF (fibroblast growth factor 2), Nodal-activin signaling, and WNT-β-catenin signaling (Ackermann et al., 2015). BMP4 is essential for mesoderm formation as Bmp4⁻/⁻ mouse embryos failed to develop the mesoderm (Winnier et al., 1995). In hESCs, BMP4 was shown to induce the formation of the hematopoietic mesoderm, hemangioblast, and hematopoietic progenitors, but inhibit ectodermal development (Chadwick et al., 2003; Kennedy
et al., 2007; Pick et al., 2007; Wang and Nakayama, 2009). bFGF is essential for hemangioblast formation as bFGF induces the expression of KDR in mesodermal precursors (Flamme et al., 1995). Nodal-activin and WNT-β-catenin signaling are essential for primitive streak formation as mutant mouse embryos of Nodal or Wnt3 failed to develop the primitive streak, while Wnt3a treatment of hESCs induced both hematopoietic and vascular potentials (Conlon et al., 1994; Liu et al., 1999; Wang and Nakayama, 2009).

**Primitive and definitive specification.** In the human embryo, mesoderm spatial patterning of Nodal-activin and WNT-β-catenin signaling specifies the primitive and definitive programs (Luckett, 1978; Huber et al., 2004; Ditadi et al., 2017). The primitive program arises from the mesoderm at the posterior region of the primitive streak, where Nodal-activin signaling is active, and migrates anteriorly to generate primitive blood islands in the yolk sac at the anterior region of the embryo, where WNT-β-catenin signaling is inhibited. The definitive program migrates more posteriorly and ventrally, where WNT-β-catenin signaling is active but Nodal-activin signaling is absent. Recapitulating the mesodermal signaling patterns during embryonic development, specification of the primitive and definitive programs can be manipulated in the first 72hr of mesoderm induction in hPSCs, in which the primitive program can be specified by the activation of Nodal-activin signaling and inhibition of WNT-β-catenin signaling, whereas the definitive program can be specified by the activation of WNT-β-catenin signaling and inhibition of Nodal-activin signaling (Kennedy et al., 2012; Sturgeon et al., 2014). Mesodermal induction results in hPSC-derived mesodermal cells that express KDR and Brachyury and have hematopoietic and endothelial potential (Carmeliet et al., 1996; Huber et al., 2004). hPSC-derived KDR+ mesodermal cells that give rise to the primitive or definitive program can be distinguished by CD235a expression (Sturgeon et al., 2014). The primitive mesoderm is identified by KDR+/CD235a+ expression and the definitive mesoderm is identified by KDR+/CD235a− expression.

**Hemogenic endothelial specification.** In the avian embryo, the mesoderm was shown to give rise to endothelial cells with different potentials depending on interactions with the endoderm or ectoderm (Pardanaud and Dieterlen-Lievre, 1999). The mesoderm that interacted with the ectoderm gave rise to endothelial cells with only endothelial potential, whereas the mesoderm that interacted with the endoderm gave rise to endothelial cells with both hematopoietic and endothelial potential, hence hematopoietic specification of the mesoderm was accomplished by
mimicking endoderm interaction using KDR ligands VEGF and bFGF. The resulting AGM endothelial population was shown to give rise to either endothelial cells or hemogenic endothelium depending on the strength of Notch1 signaling (Gama-Norton et al., 2015). In the mouse embryo, high levels of Notch1 signaling induced by Dll4 resulted in endothelial cells, whereas low levels of Notch1 signaling induced by Jag1 resulted in hemogenic endothelium and hematopoietic stem cells. Hematopoietic specification of hPSC-derived definitive mesoderm (KDR+/CD235a−) was also shown to give rise to different endothelial populations (CD34+/CD43−) distinguished by the expression of the arterial marker CD184 and venous marker CD73 (Ditadi et al., 2015). The CD34+/CD43− endothelial population was divided into CD184+ arterial endothelium, CD73+ venous endothelium, and CD184+/CD73− hemogenic endothelium.

**Early hematopoietic progenitor specification.** Notch1 signaling and downstream Runx1 expression is essential in giving rise to hematopoietic stem cells as Notch−/− mice and mutant zebrafish failed to establish hematopoietic stem cells long-term (North et al., 2002; Hadland et al., 2004; Burns et al., 2005). hPSC-derived hemogenic endothelium (CD34+/CD43−/CD184−/CD73−) co-cultured with OP9-DL1, which ectopically expressed the Notch ligand DL1 (Holmes et al., 2009), was shown to undergo endothelial-to-hematopoietic transition and give rise to early hematopoietic progenitors that are CD45+, non-adherent, definitive with lymphoid potential, and capable of generating mature definitive hematopoietic cells (Kennedy et al., 2012).

Using the aforementioned strategies of mesoderm induction, definitive specification, and step-wise hematopoietic specification, definitive hematopoietic differentiation of hPSCs can generate intermediates that can be identified by specific marker expressions on specific differentiation days for further investigation of hematopoietic stages. In this study, the definitive hematopoietic differentiation protocol used (Kennedy et al., 2012) began with feeder depletion for 2 days to remove MEFs. Embryoid body formation was induced at day 0. At the same time, mesoderm induction occurred from day 0 to 4. Addition of an activin/nodal inhibitor (SB431542) between day 1.75 to 4 induced the specification of the definitive program. Hematopoietic specification from day 4 to 9 induced the formation of the hemogenic endothelium, as well as vascular endothelium. From day 0 to 8, embryoid bodies were cultured under hypoxia to induce the expression of hypoxia inducible factor, which in turn induced VEGF expression, and both hypoxia inducible factor and VEGF were required for proliferation, hematopoiesis, and vasculogenesis (Yoon et al., 2011). At day 9, CD34+/CD43− cells were plated onto OP9-DL1, a
mouse bone marrow stromal cell line that ectopically expresses the Notch ligand, delta-like 1. Co-culture with OP9-DL1 for 7 days induced the formation of early hematopoietic progenitors that were morphologically round and non-adherent. At day 9+7, 9 days of differentiation as embryoid bodies and 7 days of co-culture with OP9-DL1, early hematopoietic progenitors were seeded in CFU assays at a density of 40,000 cells/dish.
Chapter 2
Research Aims and Hypothesis

Clinical significance. SDS is a rare, multi-system, genetic disorder characterized by combined exocrine pancreatic and hematological dysfunction (Bodian et al., 1964; Shwachman et al., 1964). SDS is often diagnosed in early childhood (Hashmi et al., 2011) and can manifest in neonates (Todorović-Guid et al., 2006). 18% of SDS patients have developed clonal and malignant myeloid transformation at the median age of 20 years (Hashmi et al., 2011) and progression to myelodysplastic syndrome and acute myeloid leukemia is the main concern for mortality in SDS. The only treatment option for severe hematopoietic dysfunction in SDS is hematopoietic stem cell transplantation with 60-65% survival rate (Donadieu et al., 2005; Cesaro et al., 2005). Hence, there is a need to determine the mechanisms underlying the hematopoietic defects in SDS in order to identify drug targets and develop effective treatments.

Molecular mechanisms. The cause of SDS in 81% of SDS patients (Hashmi et al., 2011) are biallelic mutations in the SBDS gene, with c.183_184TA>CT and c.258+2T>C being the most common mutations (Boocock et al., 2003). The SBDS protein is an essential protein that is ubiquitously expressed (Boocock et al., 2003) and the loss of Sbds is embryonically lethal in mice (Zhang et al., 2006). SBDS is a trans-acting factor that participates in the maturation of the pre-60S ribosomal subunit (Menne et al., 2007) and SBDS deficiency leads to impaired global protein synthesis (Ball et al., 2009). Classic ribosomopathies share the common characteristics of poor growth and hematological dysfunction. The downstream hematopoietic consequences of SBDS deficiency appears to include impaired translation of uORF-dependent granulocytic transcription factor C/EBP (In et al., 2016) and p53-mediated apoptosis (Zambetti et al., 2015), which probably lead to impaired terminal neutrophil differentiation, while protease-mediated auto-digestion may induce apoptosis in myeloid progenitors (Tulpule et al., 2013).

Rationale. Hematopoietic defects have also been documented in more immature populations. Hematopoietic progenitors have impaired engraftment (Rawls et al., 2007) and impaired potential to give rise to myeloid progenitor CFU-GEMMs (Dror and Freedman, 1999), while dysregulation in ribosomal genes, apoptosis inhibitors, oncogenes, and tumor suppressors has been found in patient bone marrow mononuclear cells (Rujkijyanont et al., 2007; 2009). Moreover, preliminary findings in our lab have identified the onset of hematopoietic defects.
during definitive hematopoiesis. Modeling hematopoietic development in the embryo, SDS hiPSCs have shown impaired potential of definitive hematopoietic colonies, while primitive hematopoiesis has been shown to be preserved (Luca, 2015). The mechanisms underlying these defects are unclear.

**Hypothesis.** Preliminary findings in our lab have shown that SDS defects manifested during definitive hematopoiesis (Luca, 2015), but analysis of the specific stages and the underlying mechanism have not been completed. We hypothesize that the onset of SDS hematopoietic defects occurs in a certain hematopoietic developmental stage and that this would be reflected by gene dysregulation in certain populations during the hematopoietic differentiation of SDS hiPSCs. Specifically, as SDS hiPSCs were shown to have reduced proliferative capacity (Ruiz-Gutierrez et al., 2016), we expect gene dysregulation to occur in the pluripotent stem cell stage.

**Objective 1.** As a continuation of the previous study, the first aim of this study was to determine if the hiPSCs provide appropriate models of SDS. Assessment of pluripotent markers and differentiation to all germ layers was required to verify the pluripotency of hiPSCs. Genetic characterization of hiPSCs was required to verify fidelity to the parental identity, SDS-associated mutations, and the anticipated normal karyotype. SBDS protein expression of hiPSCs was required to verify that the mutations resulted in reduced SBDS protein expression. CFU assays of hiPSCs were required to verify both hematopoietic potential and hematopoietic defects that recapitulated the defects found in the CFU assays of patient bone marrow cells (Dror and Freedman, 1999).

**Objective 2.** The second aim of this study was to isolate populations generated from definitive hematopoietic differentiation of hiPSCs. As SDS hiPSCs have shown reduced proliferative capacity (Luca, 2015; Ruiz-Gutierrez et al., 2016), hiPSCs were isolated for RNA-seq and phenotypically analyzed for pluripotency status and spontaneous differentiation. Mesodermal and hemogenic endothelial populations were phenotypically analyzed with refined markers and additional hiPSC lines were also investigated to verify preliminary findings in our lab.

**Objective 3.** As oligonucleotide microarray has identified gene dysregulation in patient bone marrow mononuclear cells (Rujkijyanont et al., 2007; 2009), the third aim of this study was to identify gene dysregulation in isolated populations. As SDS-associated SBDS splicing variants
have been shown to be downregulated in SDS (Boocock et al., 2003) and SBDS has been shown to be downregulated during normal neutrophil maturation (Orelio et al., 2009), both known and novel transcripts can be identified and temporal SBDS expression can be quantified to determine the stage-specific requirements for SBDS. To identify potentially disrupted pathways, the SDS transcriptome was analyzed for differential gene expression and over-representation. As ribosomal genes have been shown to be dysregulated in SDS (Rujkijyanont et al., 2007; 2009), differential expression of ribosomal genes could identify the developmental stages affected by ribosomal defects.

**Research methods.** As oligonucleotide microarray was useful in identifying dysregulated genes in SDS patient bone marrow mononuclear cells (Rujkijyanont et al., 2007; 2009), similarly, transcriptome analysis could also be useful in identifying potentially disrupted pathways in SDS. RNA-seq has advantages as a technique to quantify the transcriptome as it can provide accurate quantification and identify splice variants and potentially novel transcripts. To identify gene dysregulation during hematopoietic development, hiPSCs can be used to model SDS, as for our previous studies (Luca, 2015). The self-renewal capability of hiPSCs can provide an unlimited source of material without the challenges of expanding primary patient cell lines, and the pluripotent nature of hiPSCs can generate early embryonic and hematopoietic developmental populations without the ethical controversies of using hESCs. Although using hiPSC lines derived from more than two patients are ideal, using hiPSC lines derived from two patients is common in the literature. Using two patient lines could reduce the effects of genetic difference and attribute observed effects to SDS. Moreover, there was substantial difficulty in generating SDS hiPSCs.
Chapter 3
Methods

1 Cell Culture

**hiPSC reprogramming.** Human primary bone marrow cells from normal subjects (N530 and N551) and SDS patients (P55 and P357) were collected with written and informed consent by the Canadian Inherited Marrow Failure Registry. Bone marrow stromal fibroblasts were generated and reprogrammed to hiPSC using Sendai viral expression of *OCT4, SOX2, KLF4*, and *C-MYC* genes by CCRM.

**hiPSC culture.** The following method of maintaining hiPSCs cultured on MEFs was adopted from the laboratory of Dr. Gordon Keller. The materials used in the thawing, maintenance, passaging, and freezing of hiPSC cultured on MEFs are listed in Table 1. hiPSCs were co-cultured with MEFs in growth medium at 37°C in a humidified 5% CO₂ incubator with daily medium change. hiPSCs were passaged in aggregates using collagenase B at 37°C for 5 minutes and 0.25% trypsin at room temperature for up to 3 minutes. The digestion was stopped with stop medium and aggregates of hiPSCs were lifted using a cell scraper. Aggregates were collected with wash medium and pelleted by centrifugation at 150×g for 3 minutes at room temperature. Aggregates were plated onto new 6-well MEF plates in a 1:2 to 1:6 ratio. MEFs extracted at embryonic day 12.5 and treated with mitomycin C were plated by the Sickkids Embryonic Stem Cell Facility. To improve viability, ROCK inhibitor and additional 10ng/mL bFGF were supplemented in the growth medium after passaging as needed. hiPSCs were also cultured on matrigel according to manufacturer’s instructions (Stemcell Technologies 05850).

Table 1. Reagents used in the maintenance of hiPSCs cultured on MEFs.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Components</th>
<th>Company</th>
<th>Catalogue Number</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wash Medium</strong></td>
<td>Dulbecco’s Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12)</td>
<td>Gibco</td>
<td>11330-032</td>
<td>93%</td>
</tr>
<tr>
<td></td>
<td>Penicillin-Streptomycin, 5000U/mL</td>
<td>Gibco</td>
<td>15070-063</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>L-Glutamine, 100×</td>
<td>Cellgro</td>
<td>25-005-C1</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>KnockOut™ Serum Replacement (KOSR)</td>
<td>Gibco</td>
<td>10828-028</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Growth Medium</strong></td>
<td>DMEM/F12</td>
<td>Gibco</td>
<td>11330-032</td>
<td>77%</td>
</tr>
<tr>
<td></td>
<td>Penicillin-Streptomycin, 5000U/mL</td>
<td>Gibco</td>
<td>15070-063</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>L-Glutamine, 100×</td>
<td>Cellgro</td>
<td>25-005-C1</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>KOSR</td>
<td>Gibco</td>
<td>10828-028</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>MEM Non-essential Amino Acids Solution, 100×</td>
<td>Gibco</td>
<td>11140-050</td>
<td>1%</td>
</tr>
</tbody>
</table>

39
OP9-DL1 Culture. OP9-DL1 cells were gifted by the laboratory of Dr. Eyal Grunebaum. The following method of expanding OP9-DL1 cells was adopted from the laboratory of Dr. Gordon Keller. The materials used in the thawing, maintenance, passaging, and freezing of OP9-DL1 are listed in Table 2. OP9-DL1 cells were cultured on gelatin-coated T-75 flasks in OP9-DL1 medium with medium changed every 3 days. Cells were passaged in single cells using 0.25% trypsin at 37°C for 5 minutes. The digestion was stopped using FBS stop medium. Cells were collected with αMEM and pelleted by centrifugation at 500×g for 5 minutes at room temperature. Cells were resuspended in OP9-DL1 medium and irradiated at 3000rads using the Gammacell® 40 Exactor (Best Theratronics) at Sickkids Laboratory Animal Services. Cells were frozen in OP9-DL1 freezing medium at a density of 10⁶ cells/cryovial. In preparation for hematopoietic differentiation at T9, irradiated OP9-DL1 cells were plated at a density of 10⁶ cells per gelatin-coated 24-well plate.

### Table 2. Reagents used in the maintenance of OP9-DL1 culture.

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Components</th>
<th>Company</th>
<th>Catalogue Number</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gelatin</strong></td>
<td>Gelatin from Porcine Skin Type A</td>
<td>Sigma-Aldrich</td>
<td>G1890-100G</td>
<td>0.1g/100mL</td>
</tr>
<tr>
<td></td>
<td>Sterile Water</td>
<td>Wisent</td>
<td>809-115-CL</td>
<td>100%</td>
</tr>
<tr>
<td><strong>OP9-DL1 Medium</strong></td>
<td>FBS</td>
<td>Gibco</td>
<td>12483-020</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Minimum Essential Medium α (αMEM)</td>
<td>Gibco</td>
<td>12561-056</td>
<td>80%</td>
</tr>
</tbody>
</table>

OP9-DL1 cells were gifted by the laboratory of Dr. Eyal Grunebaum. The following method of expanding OP9-DL1 cells was adopted from the laboratory of Dr. Gordon Keller. The materials used in the thawing, maintenance, passaging, and freezing of OP9-DL1 are listed in Table 2. OP9-DL1 cells were cultured on gelatin-coated T-75 flasks in OP9-DL1 medium with medium changed every 3 days. Cells were passaged in single cells using 0.25% trypsin at 37°C for 5 minutes. The digestion was stopped using FBS stop medium. Cells were collected with αMEM and pelleted by centrifugation at 500×g for 5 minutes at room temperature. Cells were resuspended in OP9-DL1 medium and irradiated at 3000rads using the Gammacell® 40 Exactor (Best Theratronics) at Sickkids Laboratory Animal Services. Cells were frozen in OP9-DL1 freezing medium at a density of 10⁶ cells/cryovial. In preparation for hematopoietic differentiation at T9, irradiated OP9-DL1 cells were plated at a density of 10⁶ cells per gelatin-coated 24-well plate.
**hiPSC lentiviral transduction.** To transduce hiPSCs, hiPSCs were digested with collagenase B at 37°C for 20 minutes and 0.25% trypsin at 37°C for 5 minutes. The digestion was stopped with FBS stop medium and hiPSCs were resuspended into single cells. Single cells were collected with wash medium and pelleted by centrifugation at 500×g for 5 minutes at room temperature. Single cells were filtered through 35µm cell strainer and pelleted again. Single cells were plated onto new MEF plates at a density of 10⁴ cells/well and infected with lentiviral particles at a multiplicity of infection of 100 and with 8µg/mL protamine sulfate for 24 hours. Protamine sulfate and lentivirus packaged with the rescue plasmid pLVX.SIN.EF1a.SBDS-HA.IRES.ZsGreen or mock plasmid pLVX.SIN.EF1a.IRES.ZsGreen were gifted by Chetankumar Tailor. hiPSCs were cultured in growth medium supplemented with additional 10ng/mL bFGF and 10µM ROCK inhibitor for 10 days with daily medium change. After 10 days, hiPSCs were cultured in growth medium supplemented with additional 10ng/mL bFGF with daily medium change.

### 2 Characterization of hiPSCs

**hiPSC characterization.** CCRM assessed fidelity to parental identity by polymerase chain reaction (PCR) of 9 regions of short tandem repeats (STR; CSF1PO, D3S1358 or D21S11, D13S317, D5S818, D7S820, D16S539, THO1, TPOX, and vWA) and gel electrophoresis, assessed gender by PCR of Amelogenin and gel electrophoresis, assessed RNA expression of pluripotency markers by quantitative reverse transcription-PCR (qRT-PCR), assessed protein expression of pluripotent surface markers by flow cytometry or immunocytochemistry (ICC), assessed germ layer differentiation by EB formation and qRT-PCR, assessed *Mycoplasma* contamination, and assessed chromosomal abnormality by G-band karyotyping. Further *Mycoplasma* testing of hiPSCs cultured on matrigel was performed by Sickkids Molecular Biology Laboratory, and further G-band karyotyping of hiPSCs cultured on MEFs was performed by TCAG.

**Sanger sequencing.** Genomic DNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega, A1120) according to manufacturer’s instructions. DNA concentration
was measured by Nanodrop. SBDS exon 2 was PCR amplified from DNA extracted from hiPSCs cultured on matrigel using forward primer 5’- AAA TGG TAA GGC AAA TAC GG -3’, reverse primer 5’- ACC AAG TTC TTT ATT AGA AGT GAC -3’, and the Maxima Hot Start Taq DNA Polymerase (Thermo Fisher Scientific EP0601) according to manufacturer’s instructions. The PCR product was resolved by 1% agarose gel electrophoresis and the 733bp PCR product was extracted by QIAquick Gel Extraction Kit (Qiagen 28704) according to manufacturer’s instructions. Sanger sequencing of PCR amplified SBDS exon 2 was performed by TCAG.

**Western blotting.** Cells were harvested and lysed by RIPA lysis buffer (Santa Cruz Biotechnology sc-24948) according to manufacturer’s instructions. Protein concentration was measured using protein assay reagent concentrate (Bio-Rad 5000006) and BioPhotometer Spectrophotometer (Eppendorf) according to manufacturer’s instructions. Proteins were resolved using 7.5% polyacrylamide gel electrophoresis (Mini-PROTEAN® 3 Cell 165-3301) according to manufacturer’s instructions. Western blot transfer (Mini Trans-Blot® Electrophoretic Transfer Cell, Bio-Rad 170-3930) was performed according to manufacturer’s instructions. The membrane was incubated with blocking solution (TBST/5% skim milk) at room temperature for 1 hour on a rocking platform and washed with TBST. Antibodies were stained according to manufacturer’s instructions. Proteins were detected using ECL Western blotting detection reagent (Amersham Biosciences RPN2106) and Kodak X-OMAT film or Hyperfilm ECL (8”x10”, Amersham Biosciences RPN2114K) according to manufacturer’s instructions.

3 Hematopoietic Differentiation

The following method of definitive hematopoietic differentiation of hiPSCs cultured on MEFs was adopted from the laboratory of Dr. Gordon Keller. The materials used in feeder depletion, EB formation, EB digestion, sorting, OP9-DL1 co-culture, and colony forming unit assay are listed in Table 3.

**Table 3. Reagents used in the hematopoietic differentiation of hiPSCs cultured on MEFs.**

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Components</th>
<th>Company</th>
<th>Catalogue Number</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrigel-GFR</td>
<td>Corning® Matrigel® Matrix Growth Factor Reduced (GFR)</td>
<td>VWR</td>
<td>CACB354230</td>
<td>50%</td>
</tr>
<tr>
<td>Component</td>
<td>Supplier</td>
<td>Part Number</td>
<td>Concentration</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>-------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>Iscove’s Modified Dulbecco’s Medium (IMDM)</td>
<td>Gibco</td>
<td>21056-023</td>
<td>49%</td>
<td></td>
</tr>
<tr>
<td>Penicillin-Streptomycin, 5000U/mL</td>
<td>Gibco</td>
<td>15070-063</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>0.0625% Trypsin</td>
<td>Wisent</td>
<td>325-043-EL</td>
<td>25%</td>
<td></td>
</tr>
<tr>
<td>PBS</td>
<td>Wisent</td>
<td>311-010-CL</td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td>Low Attachment Plates</td>
<td>Corning™ Ultra-Low Attachment Plates 6-well</td>
<td>Thermo Fisher Scientific</td>
<td>07-200-601</td>
<td></td>
</tr>
<tr>
<td>StemPro®-34 Serum Free Medium (SFM) (1X) (SP34)</td>
<td>Gibco</td>
<td>10639-011</td>
<td>98%</td>
<td></td>
</tr>
<tr>
<td>Penicillin-Streptomycin, 5000U/mL</td>
<td>Cellgro</td>
<td>25-005-CI</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>L-Ascorbic Acid</td>
<td>Sigma-Aldrich</td>
<td>1-4544</td>
<td>50µg/mL</td>
<td></td>
</tr>
<tr>
<td>Transferrin</td>
<td>Roche</td>
<td>10652202001</td>
<td>150µg/mL</td>
<td></td>
</tr>
<tr>
<td>L-Glutamine, 100×</td>
<td>Cellgro</td>
<td>25-005-CI</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>1% L-Ascorbic Acid</td>
<td>Sigma-Aldrich</td>
<td>M-6145</td>
<td>0.4mM</td>
<td></td>
</tr>
<tr>
<td>Recombinant Human Bone Morphogenetic Protein 4 (BMP-4)</td>
<td>R&amp;D Systems</td>
<td>314-BP-050</td>
<td>10ng/mL</td>
<td></td>
</tr>
<tr>
<td>bFGF</td>
<td>R&amp;D Systems</td>
<td>234-FSE-025</td>
<td>5ng/mL</td>
<td></td>
</tr>
<tr>
<td>SB 431542 Hydrate</td>
<td>Sigma-Aldrich</td>
<td>S4317-5MG</td>
<td>6µM</td>
<td></td>
</tr>
<tr>
<td>Recombinant Human Vascular Endothelial Growth Factor (VEGF) 165</td>
<td>R&amp;D Systems</td>
<td>314-BP-050</td>
<td>10ng/mL</td>
<td></td>
</tr>
<tr>
<td>Recombinant Human Interleukin 6 (IL-6)</td>
<td>R&amp;D Systems</td>
<td>206-IL-050</td>
<td>10ng/mL</td>
<td></td>
</tr>
<tr>
<td>Recombinant Human Interleukin 11 (IL-11)</td>
<td>R&amp;D Systems</td>
<td>218-IL-025</td>
<td>5ng/mL</td>
<td></td>
</tr>
<tr>
<td>Recombinant Human Stem Cell Factor (SCF)</td>
<td>R&amp;D Systems</td>
<td>255-SC-200</td>
<td>50ng/mL</td>
<td></td>
</tr>
<tr>
<td>Recombinant Human Erythropoietin (EPO)</td>
<td>R&amp;D Systems</td>
<td>287-TC-500</td>
<td>2U/mL</td>
<td></td>
</tr>
<tr>
<td>IL-6</td>
<td>R&amp;D Systems</td>
<td>206-IL-050</td>
<td>10ng/mL</td>
<td></td>
</tr>
<tr>
<td>IL-11</td>
<td>R&amp;D Systems</td>
<td>218-IL-025</td>
<td>5ng/mL</td>
<td></td>
</tr>
<tr>
<td>SCF</td>
<td>R&amp;D Systems</td>
<td>255-SC-200</td>
<td>100ng/mL</td>
<td></td>
</tr>
<tr>
<td>EPO</td>
<td>R&amp;D Systems</td>
<td>287-TC-500</td>
<td>2U/mL</td>
<td></td>
</tr>
<tr>
<td>Recombinant Human Thrombopoietin (TPO)</td>
<td>R&amp;D Systems</td>
<td>288-TP-200</td>
<td>50ng/mL</td>
<td></td>
</tr>
<tr>
<td>Recombinant Human Fms-related tyrosine kinase 3 ligand (Flt-3L)</td>
<td>R&amp;D Systems</td>
<td>308-FKN-025</td>
<td>10ng/mL</td>
<td></td>
</tr>
<tr>
<td>Recombinant Human Interleukin 3 (IL-3)</td>
<td>R&amp;D Systems</td>
<td>203-IL-050</td>
<td>50ng/mL</td>
<td></td>
</tr>
<tr>
<td>Recombinant Human Insulin-like Growth Factor I (IGF-I)</td>
<td>R&amp;D Systems</td>
<td>291-G1-200</td>
<td>25ng/mL</td>
<td></td>
</tr>
<tr>
<td>Collagenase from Clostridium histolyticum Type I</td>
<td>Sigma-Aldrich</td>
<td>C0130-1G</td>
<td>2mg/mL</td>
<td></td>
</tr>
<tr>
<td>FBS</td>
<td>Gibco</td>
<td>12483-020</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>PBS</td>
<td>Wisent</td>
<td>311-010-CL</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>DNaseI</td>
<td>EMD Millipore</td>
<td>260913-10MU</td>
<td>10µg/mL</td>
<td></td>
</tr>
</tbody>
</table>
Staining Medium

<table>
<thead>
<tr>
<th>Medium</th>
<th>Supplier</th>
<th>Catalog Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBS</td>
<td>Gibco</td>
<td>12483-020</td>
<td>10%</td>
</tr>
<tr>
<td>IMDM</td>
<td>Gibco</td>
<td>21056-023</td>
<td>90%</td>
</tr>
<tr>
<td>DNaseI</td>
<td>EMD Millipore</td>
<td>260913-10MU</td>
<td>10µg/mL</td>
</tr>
</tbody>
</table>

T9 Staining Reaction

<table>
<thead>
<tr>
<th>Staining Agent</th>
<th>Supplier</th>
<th>Catalog Number</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phycoerythrin (PE) Mouse Anti-Human CD43, Clone 1G10 (PE-CD43)</td>
<td>BD Pharmingen</td>
<td>560199</td>
<td>5%</td>
</tr>
<tr>
<td>PE-Cyanine 7 (PE-Cy7) Anti-Human CD34, Clone 4H11 (PE-Cy7-CD34)</td>
<td>Affymetrix eBioscience</td>
<td>25-0349-42</td>
<td>1%</td>
</tr>
<tr>
<td>Propidium Iodide (PI)</td>
<td>Thermo Fisher Scientific</td>
<td>P1304MP</td>
<td>1µg/mL</td>
</tr>
<tr>
<td>FBS</td>
<td>Gibco</td>
<td>12483-020</td>
<td>10%</td>
</tr>
</tbody>
</table>

Sorting Medium

<table>
<thead>
<tr>
<th>Medium</th>
<th>Supplier</th>
<th>Catalog Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBS</td>
<td>Gibco</td>
<td>12483-020</td>
<td>1.5%</td>
</tr>
<tr>
<td>IMDM</td>
<td>Gibco</td>
<td>21056-023</td>
<td>98.5%</td>
</tr>
<tr>
<td>DNaseI</td>
<td>EMD Millipore</td>
<td>260913-10MU</td>
<td>10µg/mL</td>
</tr>
</tbody>
</table>

T9+0 to 1 Medium

<table>
<thead>
<tr>
<th>Medium</th>
<th>Supplier</th>
<th>Catalog Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>αMEM</td>
<td>Gibco</td>
<td>12561-056</td>
<td>80%</td>
</tr>
<tr>
<td>BMP-4</td>
<td>R&amp;D Systems</td>
<td>314-BP-050</td>
<td>10ng/mL</td>
</tr>
<tr>
<td>IL-11</td>
<td>R&amp;D Systems</td>
<td>218-IL-025</td>
<td>5ng/mL</td>
</tr>
<tr>
<td>SCF</td>
<td>R&amp;D Systems</td>
<td>255-SC-200</td>
<td>100ng/mL</td>
</tr>
<tr>
<td>TPO</td>
<td>R&amp;D Systems</td>
<td>288-TP-200</td>
<td>30ng/mL</td>
</tr>
<tr>
<td>Flt-3L</td>
<td>R&amp;D Systems</td>
<td>308-FKN-025</td>
<td>20ng/mL</td>
</tr>
</tbody>
</table>

T9+3 to 5 Medium

<table>
<thead>
<tr>
<th>Medium</th>
<th>Supplier</th>
<th>Catalog Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>VEGF</td>
<td>R&amp;D Systems</td>
<td>293-Ve-050</td>
<td>5ng/mL</td>
</tr>
<tr>
<td>HyClone FBS</td>
<td>GE Healthcare</td>
<td>SH30396.03</td>
<td>2%</td>
</tr>
<tr>
<td>IMDM</td>
<td>Gibco</td>
<td>21056-023</td>
<td>98%</td>
</tr>
</tbody>
</table>

Methylcellulose Medium

<table>
<thead>
<tr>
<th>Medium</th>
<th>Supplier</th>
<th>Catalog Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methocult™ H4034 Optimum</td>
<td>Stemcell Technologies</td>
<td>04034</td>
<td></td>
</tr>
<tr>
<td>IL-6</td>
<td>R&amp;D Systems</td>
<td>206-IL-050</td>
<td>10ng/mL</td>
</tr>
<tr>
<td>Flt-3L</td>
<td>R&amp;D Systems</td>
<td>308-FKN-025</td>
<td>10ng/mL</td>
</tr>
</tbody>
</table>

Feeder depletion. Each 6-well plate was coated with a thin layer of cold matrigel-GFR for 30 minutes on ice. After excess matrigel-GFR was removed, the matrigel-GFR-coated plates were incubated at 37°C for at least 4 hours. hiPSCs cultured on MEFs were passaged and plated onto matrigel-GFR-coated plates in a 1:2 ratio. hiPSCs were cultured in growth medium supplemented with 0.0625% matrigel-GFR at 37°C in a humidified 5% CO₂ incubator for 2 days with daily medium change.

Hematopoietic differentiation. Feeder depleted hiPSCs were passaged in aggregates using collagenase B at 37°C for 5 minutes and 0.625% trypsin at room temperature for up to 1 minute. Aggregates were plated onto low attachment plates at a density of 10⁶ cells/well in T0 medium. The time when the aggregates were placed in a humidified 37°C 5% O₂/5% CO₂/90% N₂ hypoxic incubator was designated as time-point zero (T0). After 24 hours, T1 medium was added and EBs were returned to hypoxia at T1. After 17 hours, EBs were settled for 30 minutes at 37°C
and cultured in T1.75 medium. EBs were returned to hypoxia at T1.75, 6 hours ahead of T2. After 53 hours, EBs were pelleted by centrifugation at 60×g for 3 minutes at room temperature and settled for 30 minutes at 37°C. EBs were cultured in T4 medium and returned to hypoxia at T4. After 47 hours, EBs were pelleted, settled, cultured in T6 medium, and returned to hypoxia at T6. After 47 hours, EBs were pelleted, settled, cultured in T8 medium, placed in a humidified 37°C 5% CO₂ incubator at T8, and cultured in normoxia.

**T9 digestion.** Following 20 hours in normoxia, EBs were collected with IMDM and pelleted by centrifugation at 150×g for 3 minutes at room temperature. EBs were digested with 0.25% trypsin at 37°C for 5 minutes and the digestion was stopped with FBS stop medium. EBs were erupted by syringing with a 20G needle 6 times and pelleted by centrifugation at 500×g for 5 minutes at room temperature. EBs were digested with collagenase I at 37°C for 1 hour and further erupted by syringing with a 20G needle 6 times into single cells. Staining medium was added and single cells were pelleted by centrifugation at 500×g for 5 minutes at 4°C. Single cells were filtered through a 35µm cell strainer and pelleted again. Single cells were resuspended in cold staining medium in preparation for staining reactions.

**Preparation of cells for T9 flow cytometry and sorting.** T9 single cells were added to T9 staining reactions, including PI only control, fluorescence-minus-one (FMO) controls, and sorting samples (Table 4), and stained on ice away from light for 30 minutes. Staining medium was added and stained cells were pelleted by centrifugation at 500×g for 5 minutes at 4°C. The controls were resuspended in cold staining medium and sorting samples were resuspended in cold sorting medium at densities according to Table 4. The controls were analyzed and the sorting samples were sorted using the Moflo® Astrios™ or Moflo™ XDP cell sorter (Beckman Coulter) at the Sickkids Flow Cytometry Facility. Each CD34<sup>hi</sup>/CD43<sup>-</sup> population was collected in FBS.

**Table 4. Fluorophores and cell density of PI only control, FMO controls, and sorting sample of T9 staining reactions.**

<table>
<thead>
<tr>
<th>FMOs</th>
<th>PI only</th>
<th>-PE-CD43</th>
<th>-PE-Cy7-CD34</th>
<th>Sorting Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE-CD43</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PE-Cy7-CD34</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>PI</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cells (per 100µL)</td>
<td>5×10⁴</td>
<td>5×10⁴</td>
<td>5×10⁴</td>
<td>10⁶</td>
</tr>
</tbody>
</table>
**OP9-DL1 co-culture.** The CD34$^{hi}$/CD43$^{-}$ sorted cells were pelleted by centrifugation at 500×g for 5 minutes at 4°C. Sorted cells were cultured in T9+0 to 1 medium at a density of 2×10$^4$ cells per well of irradiated OP9-DL1-coated 24-well plate in a humidified 37°C 5% CO$_2$ incubator in normoxia at T9+0. After 24 hours, T9+0 to 1 medium was added and the culture was returned to normoxia at T9+1. After 48 hours, the cells were cultured in T9+3 to 5 medium and returned to normoxia at T9+3. After 48 hours, half of the medium was changed with T9+3 to 5 medium and the culture was returned to normoxia at T9+5.

**CFU assay.** After 46 hours in normoxia, the medium of the culture was filtered through a 35µm cell strainer. Cells were digested with 0.25% trypsin at room temperature for 5 minutes and the digestion was stopped with FBS stop medium. Cells were resuspended to single cells and filtered through 35µm cell strainer. Single cells were pelleted by centrifugation at 500×g for 5 minutes at room temperature. Single cells were plated onto Methocult H4034 according to manufacturer’s instructions. Duplicates of 4×10$^4$ cells per 35mm gridded dish were cultured in a humidified 37°C 5% CO$_2$ incubator at T9+7+0. After 14 days, CFU-GEMM, CFU-GM, and BFU-E were scored at T9+7+14. A colony larger than ¼ of a grid (0.5mm × 0.5mm) was considered large, and a colony equal to or smaller than ¼ of a grid was considered small.

4 Flow Cytometry and Analyses

**T0 hiPSC digestion.** hiPSCs were digested with 0.25% trypsin at 37°C for 5 minutes and the digestion was stopped with FBS stop medium.

**T1-7 EB digestion.** EBs were collected with IMDM and pelleted by centrifugation at 150×g for 3 minutes at room temperature. EBs were digested with 0.25% trypsin at 37°C for 5 minutes and the digestion was stopped with FBS stop medium.

**T8-9 EB digestion.** EBs were collected with IMDM and pelleted by centrifugation at 150×g for 3 minutes at room temperature. EBs were digested with 0.25% trypsin at 37°C for 5 minutes. EBs were dissociated by syringing with a 20G needle 6 times and pelleted by centrifugation at 500×g for 5 minutes at room temperature. EBs were digested with collagenase I at 37°C for 1 hour and further dissociated by syringing with a 20G needle 6 times into single cells. Staining medium was added and single cells were pelleted by centrifugation at 500×g for 5 minutes at 4°C.
**T9+ co-culture digestion.** The medium of the culture was collected. Cells were digested with 0.25% trypsin at room temperature for 5 minutes and the digestion was stopped with FBS stop medium. Cells were resuspended into single cells and pelleted by centrifugation at 500×g for 5 minutes at room temperature.

**Staining.** Cells were resuspended into single cells in staining medium and filtered through 35µm cell strainer. Single cells were pelleted by centrifugation at 500×g for 5 minutes at 4°C and resuspended in cold staining medium. Single cells collected at various stages of hematopoietic differentiation were stained according to Table 5. Single cells were added to PI only control, FMO controls, and the samples (Table 6), and stained on ice away from light for 30 minutes. Cells were diluted with cold staining medium and pelleted by centrifugation at 500×g for 5 minutes at 4°C. The controls and cytometer samples were resuspended in cold staining medium, while the sorting samples were resuspended in cold sorting medium.

**Fluorescence spectral overlap compensation controls.** Each fluorophore was compensated using Anti-Mouse Ig, κ/Negative Control Compensation Particles Set (BD™ CompBead 552843) stained with each fluorophore-conjugated antibody. The staining reaction of each compensation control included 1 drop compensation beads, 1 drop negative beads, and 1µL fluorophore-conjugated antibody. After staining for at least 30 minutes on ice, staining medium was added.

**Table 5. Components of the stage-specific staining reactions.**

<table>
<thead>
<tr>
<th>Staining Reactions</th>
<th>Components</th>
<th>Company</th>
<th>Catalogue Number</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Reaction</td>
<td>PI</td>
<td>Thermo Fisher Scientific</td>
<td>P1304MP</td>
<td>1µg/mL</td>
</tr>
<tr>
<td></td>
<td>FBS</td>
<td>Gibco</td>
<td>12483-020</td>
<td>10%</td>
</tr>
<tr>
<td>Staining Medium</td>
<td>Fluorescein Isothiocyanate (FITC) Mouse Anti-Human TRA-1-60, Clone TRA-1-60 (FITC-TRA-1-60)</td>
<td>BD Pharmingen™</td>
<td>560380</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Peridinin-Chlorophyll Protein-Cyanine 5.5 (PerCP-Cy5.5) Mouse Anti-Human SSEA-4, Clone MC813-70 (PerCP-Cy5.5-SSEA-4)</td>
<td>BD Pharmingen™</td>
<td>561565</td>
<td>5%</td>
</tr>
<tr>
<td>hiPSC</td>
<td>PE Mouse Anti-Human CD30, Clone Ber H8 (PE-CD30)</td>
<td>BD Pharmingen™</td>
<td>550041</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Allophycocyanin (APC) Mouse Anti-Human CD56, Clone B159 (APC-CD56)</td>
<td>BD Pharmingen™</td>
<td>555518</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>Brilliant Violet (BV) 421 Mouse Anti-Human CD184, Clone 12G5 (BV421-CD184)</td>
<td>BD Horizon™</td>
<td>562448</td>
<td>5%</td>
</tr>
</tbody>
</table>
### Mesoderm

<table>
<thead>
<tr>
<th>Basic Reaction</th>
<th>Antibody Description</th>
<th>Supplier</th>
<th>Catalog Number</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE Mouse Anti-Human CD309, Clone 89106 (PE-KDR)</td>
<td>BD Pharmingen™</td>
<td>560494</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>FITC-TRA-1-60</td>
<td>BD Pharmingen™</td>
<td>560380</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>PE-Cy7-CD34</td>
<td>Affymetrix eBioscience</td>
<td>25-0349-42</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>BV421 Mouse Anti-Human CD117, Clone 104D2 (BV421-C-KIT)</td>
<td>BD Horizon</td>
<td>563856</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>APC-CD56 or APC Mouse Anti-Human CD235a, Clone GA-R2 (APC-CD235a)</td>
<td>BD Pharmingen™</td>
<td>561775</td>
<td>1%</td>
<td></td>
</tr>
</tbody>
</table>

### HE

<table>
<thead>
<tr>
<th>Basic Reaction</th>
<th>Antibody Description</th>
<th>Supplier</th>
<th>Catalog Number</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE-Cy7-CD34</td>
<td>Affymetrix eBioscience</td>
<td>25-0349-42</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>PE-CD43 or FITC Mouse Anti-Human CD43, Clone 1G10 (FITC-CD43)</td>
<td>BD Pharmingen™</td>
<td>560199</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>FITC Mouse anti-Human CD144, Clone 55-7H1 (FITC-VE-cad) or PE Mouse Anti-Human CD144, Clone 55-7H1 (PE-VE-cad)</td>
<td>BD Pharmingen™</td>
<td>560411</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>BV421-CD184</td>
<td>BD Horizon™</td>
<td>562448</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>APC Mouse Anti-Human CD73, Clone AD2 (APC-CD73)</td>
<td>BD Pharmingen™</td>
<td>560847</td>
<td>5%</td>
<td></td>
</tr>
</tbody>
</table>

### Table 6. Fluorophores and cell density of staining reactions.

<table>
<thead>
<tr>
<th>PI only</th>
<th>FITC</th>
<th>PE</th>
<th>-PE-Cy7/PerCP-Cy5.5</th>
<th>-BV421/V450</th>
<th>APC</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTTC</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PE</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PE-Cy7/PerCP-Cy5.5</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>BV421/V450</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>APC</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>PI</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cells for Sorter (per 100µL)</th>
<th>5×10⁴</th>
<th>5×10⁴</th>
<th>5×10⁴</th>
<th>5×10⁴</th>
<th>5×10⁴</th>
<th>5×10⁴</th>
<th>10⁶</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cells for Cytometer (per 100µL)</td>
<td>10⁶</td>
<td>10⁶</td>
<td>10⁶</td>
<td>10⁶</td>
<td>10⁶</td>
<td>10⁶</td>
<td>10⁶</td>
</tr>
</tbody>
</table>

**Sample analysis by flow cytometry and sorting.** The cytometer samples were analyzed using the LSRII-CFI, LSRII-GC, or LSRII-Fortessa cell analyzer (BD Biosciences) and the sorting samples were sorted using the Moflo® Astrios™ or Moflo™ XDP cell sorter (Beckman Coulter) at the Sickkids Flow Cytometry Facility. Each sorted population was collected in FBS. Cell
analyses and sorting were analyzed using Flowjo Version 10. Gates were determined using FMOs with less than 1% outliers.

**RNA storage.** Sorted cells were pelleted by centrifugation at 500×g for 5 minutes at 4°C and resuspended in cold PBS. Cells were pelleted again by centrifugation at 2000×g for 5 minutes at 4°C and resuspended in 0.5mL TRIzol® (Thermo Fisher Scientific 15596026). Cells were lysed for 5 minutes at room temperature and stored at -80°C.

5 RNA-Seq and Analyses

RNA was extracted by TRIzol® according to manufacturer’s instructions (Thermo Fisher Scientific 15596026). RNA concentration was assessed using the Qubit RNA Assay Kit and Qubit 2.0 (Invitrogen Q32866). RNA was submitted to the Donnelly Sequencing Centre for library preparation and RNA-seq. The library was prepared using Clontech SMARTer Low Input RNA kit and Nextera XT libraries. The cDNA libraries contained mRNA and long non-coding RNA. RNA-seq was performed by Illumina HiSeq2500 with V4 Chemistry and reagents. Four samples were sequenced per paired end lane with read length of 125bp. RNA-seq analyses included adapter and polyA trimming by FASTQ Toolkit v.1.0, read mapping by TopHat 2 and Bowtie 2 in the RNA-Seq Alignment App v1.0.0, analyzing differential expression by Cuffdiff 2 in Cufflinks Assembly & DE App v2.1.0, and viewing reads by Integrative Genomics Viewer v2.1.2 all under Illumina Basespace (Trapnell et al., 2012). Over-representation analysis was analyzed by InnateDB (Breuer et al., 2013).

6 RT-qPCR

**RNA extraction.** RNA extracted from hiPSCs cultured on matrigel was reverse transcribed into cDNA using Omniscript Reverse Transcription (RT) Kit (Qiagen 205110) according to manufacturer’s instructions.

**Primer design.** Using Primer3 Version 4.0.0 (Untergasser et al., 2012), RT-qPCR primers (Table 13) were designed to target the exon-exon junctions or 3’ untranslated region (UTR) of human mRNA. Primer design parameters include primer size of 18-30bp, primer melting temperature (Tm) of 58-60°C, maximum Tm difference of 2, product Tm of 75-90°C, primer guanine and cytosine (GC) content of 40-60, product size of 100-150, maximum self-
complementarity of 5 or lower, maximum pair complementarity of 5 or lower, maximum 3’ complementarity of 2 or lower, maximum 3’ pair complementarity of 2 or lower, maximum number of nucleotide run of 2, CG clamp of 0-2, and maximum number of GC in the 3’ end of the primer of 2. Primers with mis-primed targets, single nucleotide polymorphisms (SNPs), tendency to form hairpin, homo-, and heterodimers were avoided. Please refer to Table 13 for primer sequences.

**Primer concentration.** Ratios of the concentrations of forward primer-to-reverse primer (FP:RP) were assessed, including 50nM:50nM, 50nM:300nM, 300nM:50nM, and 300nM:300nM. RT-qPCR reactions were performed with no template controls (NTCs) and no amplification control (NAC) using Power SYBR Green PCR Master Mix (Thermo Fisher Scientific 4367659) and Mx3005P qPCR System (Agilent Technologies) according to manufacturer’s instructions. Analyzing RT-qPCR results using MxPro (Agilent Technologies), RT-qPCR primers with FP:RP fulfilling the following criteria were selected for standard curve optimization: standard deviation (SD) below 0.16; absence of primer dimer in dissociation curve, NTC, and 2% agarose gel electrophoresis; lowest threshold cycle (Ct), and highest baseline-corrected, reference dye-normalized fluorescence (dRn) and Tm.

**Standard curve.** RT-qPCR primers with optimized FP:RP were further assessed using serial dilutions of cDNA, including 100ng, 20ng, 4ng, 0.8ng, 0.16ng, and 0.032ng. RT-qPCR primers with FP:RP fulfilling the following criteria passed the complete RT-qPCR optimization: SD below 0.16, coefficient of determination ($R^2$) above 0.985, and amplification efficiency (Eff) of 90-100%.

7  **Statistical Analyses**

Unpaired data (flow cytometry data of hiPSCs) was analyzed using unpaired two-tailed t-test. Paired data with equal variance was analyzed using unpaired two-tailed t-test and data with unequal variance was analyzed by unpaired two-tailed t-test with Welch’s correction. Significance was defined as p-value (p) less than 0.05. Means and standard deviation were graphed by GraphPad Prism 6.
Chapter 4
Results

1 Characterization of hiPSCs

Primary bone marrow stromal cells were obtained from two normal subjects and two SDS patients and reprogrammed to hiPSCs using Sendai viral expression of OCT4, SOX2, KLF4, and c-MYC by CCRM. Two independent hiPSC lines from each individual were used in this study (Table 7). Detailed hiPSC characterization by CCRM is found in the Appendices (Sections 1 and 2). Sanger sequencing of SBDS exon 2 confirmed that P55 hiPSCs carried c.[183_184TA>CT; 258+2T>C];[258+2T>C] and P357 hiPSCs carried c.[183_184TA>CT];[258+2T>C] SDS-associated SBDS mutations (Table 7, Figure 1). Western blotting showed reduced SBDS protein expression in SDS compared to normal hiPSCs (Figure 2). Densitometry was not assessed due to overexposure and difficulty distinguishing bands.

Table 7. Clinical features of normal subjects and SDS patients

<table>
<thead>
<tr>
<th>UPN</th>
<th>hiPSC Lines</th>
<th>Age</th>
<th>Sex</th>
<th>Diagnosis</th>
<th>SBDS Mutations</th>
<th>Clinical Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>N530</td>
<td>E, J</td>
<td>12</td>
<td>F</td>
<td>Normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N551</td>
<td>I, K</td>
<td>41</td>
<td>M</td>
<td>Normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P55</td>
<td>C, F</td>
<td>11</td>
<td>F</td>
<td>SDS</td>
<td>c.[183_184TA&gt;CT; 258+2T&gt;C];[258+2T&gt;C]</td>
<td>Pancreatic insufficiency, mild neutropenia, behavioral issues</td>
</tr>
<tr>
<td>P357</td>
<td>A, D</td>
<td>17</td>
<td>M</td>
<td>SDS</td>
<td>c.[183_184TA&gt;CT]; [258+2T&gt;C]</td>
<td>Pancreatic insufficiency with transient need for enzymes, moderate-severe neutropenia, mild thrombocytopenia, severe metaphyseal dysostosis</td>
</tr>
</tbody>
</table>
Figure 1. SDS-associated mutations in hiPSCs.
Sanger sequencing of sequenced PCR-amplicons of SBDS exon 2 of genomic DNA confirmed SDS-associated mutations in SDS hiPSCs.

Figure 2. SBDS protein expression in hiPSCs.
Western blot with anti-SBDS and anti-β-actin antibodies as control confirmed reduced SBDS expression in PAGE-separated proteins of whole cell extracts of SDS compared to non-SDS (normal) hiPSCs.
2 Colony Forming Unit Assay

Although CFU assays of patient bone marrow and SBDS knockdown hematopoietic progenitors have shown reduced number and size of CFU-GEMM, CFU-GM, CFU-G, CFU-E, and BFU-E (Saunders et al., 1979; Suda et al., 1982; Dror and Freedman, 1999; Kuijpers et al., 2005; Sezgin et al., 2013), CFU assays of SDS hPSCs have shown reduced total colony count but no differences in CFU-G, CFU-M, or CFU-GM lineages (Tulpule et al., 2013). To determine if our SDS hiPSCs were able to recapitulate the SDS hematopoietic defects of reduced myeloid colony count, hiPSCs were differentiated towards myeloid hematopoietic lineage by CFU assay. hiPSCs were feeder depleted for 2 days, subjected to definitive hematopoietic differentiation in the form of EBs for 9 days, sorted for CD34+/CD43− populations, co-cultured on OP9-DL1 for 7 days, cultured in methylcellulose for 14 days, and then scored for large and small colonies of CFU-GEMM, CFU-GM, and BFU-E at T9+7+14 (Figure 3). A colony larger than ¼ of a grid (0.5mm × 0.5mm) was considered to be large, while a colony equal to or smaller than ¼ of a grid was considered to be small. The size of the colony reflected the proliferative potential of each plated cell, while the number of colonies reflected the hematopoietic multipotency of plated cells. To note, CFU-GM in this study also included CFU-G and CFU-M, as the differences among the three were difficult to distinguish. Although there was no significant reduction in CFU-GEMM, there were significant reductions in the number of total colonies, CFU-GM, and BFU-E (Figure 3). Moreover, there were significant reductions in the number of small (p = 0.02) and large (p = 0.002) total colonies, small (p = 0.03) and large (p = 0.002) CFU-GM, large CFU-GEMM (p = 0.046), and small BFU-E (p = 0.02), although there were no significant reductions in small CFU-GEMM (p = 0.07) or large BFU-E (p = 0.4), as BFU-E yielded largely small colonies in the in CFU assays. (Individual lines P55F and N530J have not been scored for colony counts because of technical challenges such as contamination, low cell count, and dried methylcellulose).
**Figure 3. Analyses of CFU Counts at T9+7+14.**

Normal (N530E, N551I, and N551K) and SDS (P55C, P357A, and P357D) hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 9 days, sorted for CD34+/CD43- population, co-cultured on OP9-DL1 for 7 days, plated into methylcellulose at a density of 40000 cells/dish, cultured in methylcellulose for 14 days, and scored for large (greater than 0.5mm × 0.5mm) and small (equal to or less than 0.5mm × 0.5mm) colonies of CFU-GEMM, CFU-GM, and BFU-E at T9+7+14. Bars represent mean count of large and small colonies, error bars represent standard deviation, and p-values comparing total numbers of colonies derived from normal and SDS hiPSCs are shown at the top.
3 Flow Cytometry Analyses

As SDS human bone marrow has shown reduced CD34⁺ hematopoietic stem and progenitor cells (Dror and Freedman, 1999) and mouse Sbds knockdown hematopoietic progenitors have shown both impaired engraftment and myeloid proliferation (Rawls et al., 2007), this evidence suggested that SDS hematopoietic defects could occur in early hematopoietic progenitors. Preliminary findings in our lab also suggested that the onset of hematopoietic defects occurred in the hemogenic endothelium. To follow the course of hematopoietic development in SDS, hiPSCs were subjected to definitive hematopoietic differentiation (Kennedy et al., 2012) and populations (hiPSCs, mesoderm, and hemogenic endothelium) were collected or analyzed by specific surface markers at specific time-points (Figure 4). Sorted hiPSCs were RNA sequenced, while other populations (mesoderm and hemogenic endothelium) were analyzed in preparation for RNA-seq.

Figure 4. Definitive hematopoietic differentiation timeline. hiPSCs were subjected to feeder depletion for 2 days, hematopoietic differentiation in the form of EBs for 9 days, sorting of CD34⁺/CD43⁻ population at T9, OP9-DL1 co-culture for 7 days, and CFU assays for 14 days. Populations were sorted using specific surface markers at specific time-points. Timeline in days is shown at the top, changes in culture conditions are shown in red, cytokine additions are shown below the timeline, sorts are shown in blue arrows, surface markers are shown above the timeline, and incubation conditions are shown at the bottom.
3.1 Human Induced Pluripotent Stem Cells

As SDS hiPSCs have shown reduced proliferation (Ruiz-Gutierrez et al., 2016), to determine if there were any significant differences between SDS and normal hiPSCs, hiPSCs were feeder depleted for 2 days and sorted for hiPSC population using APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ by FACS at T0 (Figure 5, Supplementary Figure 8 - Supplementary Figure 9). Fluorescence spectral overlap compensation controls, fluorescence-minus-one negative controls, and unstained controls are found in the Appendices (Section 3). Populations identified by specified marker(s) in flow cytometry were gated from live cells unless otherwise stated.

Between SDS and normal hiPSCs, there were no significant differences in the percentages of TRA-1-60+, SSEA-4+, and CD56/CD184/TRA-1-60+/SSEA-4+ hiPSCs and CD56/CD184-/TRA-1-60+/SSEA-4+/CD30+ undifferentiated hiPSCs. There was also no significant difference in pluripotency status as reflected by the proportion of CD30+ undifferentiated hiPSCs within CD56/CD184-/TRA-1-60+/SSEA-4+ hiPSCs (Figure 6). There was also no significant difference in spontaneous differentiation as reflected by the percentages of CD56+ early mesodermal and ectodermal cells and CD184+ endodermal cells.

As MEFs were largely eliminated from the analysis during feeder depletion, the percentage of PI- (propidium iodide, viability dye) live cells within singlets reflected the density of live cells within the cultured cells, while the percentages of TRA-1-60+, SSEA-4+, and hiPSCs reflected the density of hiPSCs within the cultured cells.
Figure 5. Flow cytometry analyses of T0 hiPSC populations.
Normal N530J and SDS P357D hiPSCs were feeder depleted for 2 days and sorted for hiPSC population using APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ by FACS, which is visualized here by Flowjo V10.
Figure 6. Percentages of T0 hiPSC populations.

Normal (total, n = 10; N530E, n = 2; N530J, n = 3; N551I, n = 2; N551K, n = 2) and SDS (total, n = 10; P55C, n = 2; P55F, n = 2; P357A, n = 3; P357D, n = 3) hiPSCs were feeder depleted for 2 days and sorted for hiPSC population using APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ by FACS. Percentages of PI+ live cells within singlets (A), CD56+ mesodermal and ectodermal cells (B), CD184+ endodermal cells (C), TRA-1-60+ hiPSCs (D), SSEA-4+ hiPSCs (E), CD30+ undifferentiated hiPSCs (F), CD56+/CD184+/TRA-1-60+/SSEA-4+/CD30+ hiPSCs (G), CD56+/CD184+/TRA-1-60+/SSEA-4+/CD30+ undifferentiated hiPSCs (H), and CD30+ undifferentiated hiPSCs within CD56+/CD184+/TRA-1-60+/SSEA-4+ hiPSCs (I) were analyzed by Flowjo V10. Unpaired two-tailed t-tests were analyzed between SDS and normal hiPSCs, with p-values shown above bars. Bars represent means, error bars represent standard deviation, and the y-axis represent the percentage of population.
3.2 Mesodermal Cells

To investigate the next developmental stage, mesoderm, hiPSCs were feeder depleted for 2 days, subjected to definitive hematopoietic differentiation in the form of EBs for 4 days, and sorted for mesodermal population using FACS at T4. Here, mesodermal populations were sorted using FITC-TRA-1-60/PE-Cy7-CD34/BV421-C-KIT+/APC-CD56+/PE-KDR+ by FACS (Figure 7, Supplementary Figure 10 - Supplementary Figure 11).

Between SDS and normal T4 EBs, there were no significant differences in KDR+, KDR+/C-KIT−, KDR+/CD56+, and TRA-1-60/CD34+/C-KIT+/CD56+/KDR+ mesodermal populations. However, the yield of KDR+ population was lower in both SDS and normal T4 EBs than the expected 60% (Figure 8F; Kennedy et al., 2012), which could be due to the trypsin sensitivity of KDR or issues with differentiation potential due to reagent and cell line quality. The percentage of PI- viable cells were also low (around 50%), which could be due to the mechanical eruption of embryoid bodies. As expected, C-KIT expression at this stage was very low in both SDS and normal T4 EBs (about 1-5%). Intriguingly, SDS T4 EBs showed significantly reduced C-KIT+ and KDR+/C-KIT+ populations compared to normal T4 EBs, although populations were low and might not be meaningful (Figure 8C and I).
Figure 7. Flow cytometry analyses of T4 mesodermal populations.
Normal N530E and SDS P55C hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 4 days, and sorted for mesodermal population using FITC-TRA-1-60/PE-Cy7-CD34/BV421-C-KIT/APC-CD56+/PE-KDR+ by FACS, which is visualized here by Flowjo V10.
Figure 8. Percentages of T4 mesodermal populations.

Normal (total, n = 17; N530E, n = 2; N530J, n = 6; N551L, n = 6; N551K, n = 3) and SDS (total, n = 17; P55F, n = 2; P55F, n = 5; P357A, n = 4; P357D, n = 6) hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 4 days, and sorted for mesodermal population using FITC-TRA-1-60/PE-Cy7-CD34/BV421-C-KIT/APC-CD56/PE-KDR by FACS. Percentages of PI live cells within singlets (A), TRA-1-60+ hiPSCs (B), C-KIT+ population (C), CD34+ hematopoietic cells (D), CD56+ early mesodermal cells (E), KDR+ mesodermal cells (F), KDR+/C-KIT+ mesodermal population (G), KDR+/CD56+ mesodermal population (H), KDR+/C-KIT+ uncharacterized population (I), and TRA-1-60+/CD34+/C-KIT+/CD56+/KDR+ mesodermal population (J) were analyzed by Flowjo V10. Paired two-tailed t-tests were analyzed between SDS and normal T4 EBs, with p-values shown above bars. Bars represent means, error bars represent standard deviation, and the y-axis represent the percentage of population.
3.3 Hemogenic Endothelial Cells

To determine which of the hemogenic, venous, or arterial endothelial populations contributed to the reduction in the CD34+/CD43− hemogenic and vascular endothelial populations previously seen in SDS, the hemogenic endothelium was distinguished from the venous and arterial endotheia with refined marker selection. CD34 marks the endothelial and hematopoietic population, CD43 marks the primitive endothelial and hematopoietic population and late myeloid cells, VE-cad marks the endothelial population, CD184 marks the arterial endothelial population, and CD73 marks the venous endothelial population (Ditadi et al., 2015). Together, the combination of CD34+/CD43−/VE-cad+/CD184−/CD73− defined the definitive hemogenic endothelium. Optimizations of the hemogenic endothelial population are found in the Appendices (Section 3.3). To address the issue of differentiation kinetics, the kinetics of CD34+/CD43− hemogenic and vascular endothelia and CD34+/CD43+/CD184−/CD73− hemogenic endothelium was addressed with a replicate of paired differentiation from day 5 to 8 were assessed (Supplementary Figure 24. Flow cytometry analyses of hemogenic endothelial populations without collagenase I digestion from T5-8). As both SDS and normal lines followed similar differentiation kinetics, analyzing both lines at the same differentiation was appropriate.

hiPSCs were feeder depleted for 2 days, subjected to definitive hematopoietic differentiation in the form of EBs for 8 days, and sorted for hemogenic endothelial population with PE-Cy7-CD34+/FITC-CD43+/PE-VE-cad+/BV421-CD184+APC-CD73− or PE-Cy7-CD34+/PE-CD43+/FITC-VE-cad+/BV421-CD184+APC-CD73− using FACS at T8 (Figure 9, Supplementary Figure 29 - Supplementary Figure 30). The use of FITC-CD43/PE-VE-cad or PE-CD43/FITC-VE-cad was interchangeable, but PE-CD43/FITC-VE-cad was preferred since the brighter FITC marked the positive VE-cad population while the dimmer PE marked the negative CD43 population.

**Hemogenic endothelium.** The populations of interest, hemogenic endothelial populations (CD34+/CD43+/CD184−/CD73− and CD34+/CD43+/VE-cad+/CD184−/CD73−), showed no significant differences between SDS and normal T8 EBs (Figure 10O and S).

**Endothelium.** In terms of the total endothelial population, SDS hiPSCs showed increased potential to generate CD34+ endothelial population at day 8 compared to normal hiPSCs (Figure
10B), although other endothelial populations (VE-cad⁺, CD34⁺/CD43⁻, CD34⁺/VE-cad⁺, CD34⁺/CD43⁻/VE-cad⁺) showed no significant differences (Figure 10D, G, H, L).

**Venous endothelium.** We also analyzed differentiation towards venous endothelium at T8. SDS hiPSCs showed increased potential to generate CD34⁺/CD43⁻/CD184⁻/CD73⁺ venous endothelial population at day 8 compared to normal hiPSCs (Figure 10N), although other venous endothelial populations (CD34⁺/CD73⁺ and CD34⁺/CD43⁻/VE-cad⁺/CD184⁻/CD73⁺) showed no significant differences (Figure 10K and R). Moreover, other populations expressing the venous marker CD73 (CD73⁺ and CD184⁻/CD73⁺) (Figure 10F and X) and CD34⁺/CD184⁻ venous and hemogenic endothelial population (Figure 10V) showed significant increases in SDS compared to normal T8 EBs.

**Arterial endothelium.** Next we analyzed the differentiation towards arterial endothelium at T8. The arterial endothelial populations (CD34⁺/CD43⁻/CD184⁺ and CD34⁺/CD43⁻/VE-cad⁺/CD73⁻/CD184⁺) showed no significant differences between SDS and normal T8 EBs (Figure 10M and Q).

**Other populations.** There were also significant reductions in uncharacterized CD34⁻ populations (CD34⁻/VE-cad⁻, CD34⁻/CD184⁻, CD34⁻/CD73⁻) in SDS compared to normal T8 EBs (Figure 10U-W).
Figure 9. Flow cytometry analyses of T8 hemogenic endothelial populations.
Normal N530E and SDS P55C hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 8 days, and analyzed for hemogenic endothelial population using PE-Cy7-CD34+/PE-CD43+/FITC-VE-cad+/BV421-CD184+/APC-CD73 by analytical flow cytometer, which is visualized here by Flowjo V10.
Figure 10. Percentages of T8 hemogenic and vascular endothelial populations.

SDS (total, n = 9 or 10; N530E, n = 2; N530J, n = 2; N551I, n = 2 or 3; N551K, n = 3) and normal (total, n = 9 or 10; P55C, n = 2; P55F, n = 2; P357A, n = 2 or 3; P357D, n = 3) hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 8 days, and analyzed or sorted for hemogenic endothelial population using PE-Cy7-CD34+/FITC or PE-CD43+/PE or FITC-VE-cad+/BV421-CD184⁻/APC-CD73⁻ by analytical flow cytometer or FACS. Percentages of PI⁻ live cells within singlets (A), CD34⁺ hematopoietic cells (B), CD43⁺ primitive hemopoietic population (C), VE-cad⁺ endothelial population (D), CD184⁺ arterial population (E), CD73⁺ venous population (F), CD34⁺/CD43⁻ endothelial population (G), CD34⁺/VE-cad⁺ endothelial population (H), CD34⁺/CD184⁺ arterial population (I), CD34⁺/CD184⁺ hemogenic and venous endothelial population (J), CD34⁺/CD73⁺ venous endothelial population (K), CD34⁺/CD43⁺/VE-cad⁺ endothelial population (L), CD34⁺/CD43⁻/CD184⁺ arterial endothelial population (M), CD34⁺/CD43⁻/CD184⁺/CD73⁺ venous endothelial population (N), CD34⁺/CD43⁻/CD184⁻/CD73⁻ hemogenic endothelial population (O), proportions of CD184⁺/CD73⁻ hemogenic endothelial, CD184⁻/CD73⁺ venous, and CD184⁺ arterial populations in CD34⁺/CD43⁻ endothelial population (P), CD34⁺/CD43⁻/VE-cad⁺/CD184⁺ arterial endothelial population (Q), CD34⁺/CD43⁻/VE-cad⁺/CD184⁻/CD73⁺ venous population (R), CD34⁺/CD43⁻/VE-cad⁺/CD184⁻/CD73⁻ hemogenic endothelial population (S), proportions of CD184⁻/CD73⁻ hemogenic endothelial, CD184⁻/CD73⁺ venous, and CD184⁺ arterial populations in CD34⁺/VE-cad⁺/CD43⁻ endothelial population (T), CD34⁺/VE-cad⁻ uncharacterized non-endothelial population (U), CD34⁺/CD184⁻ uncharacterized population (V), CD34⁺/CD73⁻ uncharacterized population (W), and CD184⁻/CD73⁻ uncharacterized population were analyzed by Flowjo V10. Paired two-tailed t-tests were analyzed between SDS and normal T8 EBs with p-values shown above bars. Bars represent means, error bars represent standard deviation, and the y-axis represent the percentage of population in (A)-(X) except (P) and (T).
3.4 CD34<sup>+</sup>/CD43<sup>-</sup> Population

To determine if CD34<sup>+</sup>/CD43<sup>-</sup> hemogenic and vascular endothelial populations were reduced, populations in T9 EBs were investigated. hiPSCs were feeder depleted for 2 days, subjected to definitive hematopoietic differentiation in the form of EBs for 9 days, and sorted for PE-Cy7-CD34<sup>+</sup>/PE-CD43<sup>-</sup> population using FACS at T9 to be further cultured on OP9-DL1 (Figure 11, Supplementary Figure 31). The population of interest, CD34<sup>+</sup>/CD43<sup>-</sup> hemogenic and vascular endothelial cells, showed no significant difference between SDS and normal T9 EBs. However, SDS hiPSCs showed reduced potential to differentiate toward CD34<sup>+</sup>/CD43<sup>-</sup> primitive endothelial population at day 9 compared to normal hiPSCs, although this population was inhibited and minimal. Among all differentiation days analyzed (T0, T4, T8, and T9), T9 was the earliest differentiation day with significantly reduced viability in SDS, as there was a significant reduction in the percentage of PI<sup>-</sup> live cells in SDS compared to normal T9 EBs (Table 8).

Table 8. PI<sup>-</sup> viable populations in SDS compared to normal hiPSC-derived developmental stages.

<table>
<thead>
<tr>
<th>Developmental Stage</th>
<th>Differentiation Day</th>
<th>Mean and Standard Deviation</th>
<th>Difference</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Normal (n = x)</td>
<td>SDS (n = y)</td>
<td></td>
</tr>
<tr>
<td>hiPSCs</td>
<td>0</td>
<td>84 ± 17% (n = 10)</td>
<td>80 ± 18% (n = 10)</td>
<td>-4%</td>
</tr>
<tr>
<td>Mesoderm</td>
<td>4</td>
<td>62 ± 26% (n = 17)</td>
<td>54 ± 26% (n = 17)</td>
<td>-8%</td>
</tr>
<tr>
<td>Hemogenic</td>
<td>8</td>
<td>82 ± 12% (n = 10)</td>
<td>71 ± 22% (n = 10)</td>
<td>-11%</td>
</tr>
<tr>
<td>Endothelium</td>
<td>9</td>
<td>83 ± 11% (n = 14)</td>
<td>73 ± 18% (n = 14)</td>
<td>-11%</td>
</tr>
</tbody>
</table>
Figure 11. Flow cytometry analyses and percentages of T9 hemogenic and vascular endothelial populations. SDS (total, n = 14; P55C, n = 3; P55F, n = 4; P357A, n = 4; P357D, n = 3) and normal (total, n = 14; N530E, n = 4; N530J, n = 3; N551I, n = 2; N551K, n = 5) hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 9 days, and sorted for hemogenic and vascular endothelial population using PE-Cy7-CD34+/PE-CD43- by FACS, which is visualized here by Flowjo V10. Percentages of PI- live cells within singlets, CD34+/CD43- hemogenic and vascular endothelial population, and CD34+/CD43+ primitive endothelial population are shown at the bottom. Paired two-tailed t-tests were analyzed between SDS and normal T9 EBs with significant p-values shown above bars. Bars represent means, error bars represent standard deviation, and the y-axis represent the percentage of population.
4 RNA-Seq Analyses of Sorted hiPSCs

As reduced proliferation was found in SDS hiPSCs (Luca, 2015; Ruiz-Gutierrez et al., 2016), hiPSCs were isolated and RNA-sequenced to identify gene dysregulation in SDS. hiPSCs were feeder depleted for 2 days and sorted for APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ by FACS at T0. cDNA libraries were prepared by Clontech SMARTer Low Input RNA kit and Nextera XT libraries and sequenced with 4 samples/paired end lane, 125bp read length by Illumina HiSeq2500 by the Donnelly Sequencing Centre (Table 9). To identify the genes with significant differential expression, processing and analyses of RNA-seq reads included adapter and polyA trimming by FASTQ Toolkit, checking quality by FastQC, aligning reads against Homo sapiens hg19 RefSeq by TopHat 2 and Bowtie 2, analyzing differential expression with fragment bias correction and multi-read correction by Cuffdiff 2, and viewing reads by Integrative Genomics Viewer (Trapnell et al., 2012). RNA-seq quality information is found in the Appendices (Section 4). To determine the affected biological processes, over-representation of differentially expressed genes was analyzed by InnateDB (Breuer et al., 2013).

Table 9. Sequencing depth and coverage of hiPSC RNA-seq.

<table>
<thead>
<tr>
<th>hiPSC Line</th>
<th>Number of Reads</th>
<th>Sequencing Depth</th>
<th>Coding Coverage</th>
<th>UTR Coverage</th>
<th>Intron Coverage</th>
<th>Intergenic Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>N530E</td>
<td>38,328,265</td>
<td>1.6x</td>
<td>106x</td>
<td>56x</td>
<td>0.7x</td>
<td>0.1x</td>
</tr>
<tr>
<td>N530J</td>
<td>48,255,124</td>
<td>2.0x</td>
<td>116x</td>
<td>63x</td>
<td>0.9x</td>
<td>0.2x</td>
</tr>
<tr>
<td>N551I</td>
<td>56,575,204</td>
<td>2.3x</td>
<td>156x</td>
<td>75x</td>
<td>1.3x</td>
<td>0.3x</td>
</tr>
<tr>
<td>N551K</td>
<td>54,063,909</td>
<td>2.2x</td>
<td>152x</td>
<td>74x</td>
<td>1.2x</td>
<td>0.2x</td>
</tr>
<tr>
<td>P55C</td>
<td>48,366,044</td>
<td>2.0x</td>
<td>125x</td>
<td>64x</td>
<td>1.3x</td>
<td>0.2x</td>
</tr>
<tr>
<td>P55F</td>
<td>54,819,194</td>
<td>2.2x</td>
<td>141x</td>
<td>68x</td>
<td>1.4x</td>
<td>0.3x</td>
</tr>
<tr>
<td>P357A</td>
<td>43,222,225</td>
<td>1.7x</td>
<td>110x</td>
<td>57x</td>
<td>1.0x</td>
<td>0.2x</td>
</tr>
<tr>
<td>P357D</td>
<td>51,228,625</td>
<td>2.1x</td>
<td>106x</td>
<td>62x</td>
<td>1.0x</td>
<td>0.2x</td>
</tr>
</tbody>
</table>
Aligned reads of hiPSCs were analyzed for novel SBDS splice variants and differential expression between SDS and normal hiPSCs without fragment bias correction and multi-read correction by Cuffdiff 2. Normal hiPSCs transcribed the full length SBDS transcript (Figure 12, Table 10, Supplementary Figure 44 - Supplementary Figure 45, Supplementary Table 1). P357 transcribed the full length transcript with the nonsense mutation c.183_184TA>CT. As P55 carried c.258+2T>C on both alleles, no full length transcript was detected. As a result of the mutation c.258+2T>C that disrupted the donor splice site of intron 2, SDS hiPSCs transcribed the SBDS splice variant with exon 2 skipping (r.129_258del) due to the usage of the upstream donor splice site of intron 1 (c.128+1_2) and another SBDS splice variant with 8bp deletion 3’ of exon 2 (r.251_258del) due to the usage of the predicted alternative donor splice site within exon 2 (c.251_252GT) as observed (Boocock et al., 2003).

Other potentially novel transcripts were found (Figure 12, Table 8). N530E transcribed an SBDS splice variant with exon 1 deletion and 142nt intronic retention 5’ of exon 2 (r.1_128del; 128_129ins129-142_129-1). P357D transcribed an SBDS splice variant with 64nt deletion 5’ of exon 2 (r.129_192del) due to the usage of an alternative acceptor splice site (c.191_192AG). A similar splicing event (r.129_192del) was found in P55F. P55F transcribed an SBDS splice variant with both 64nt deletion 5’ of exon 2 and 8nt deletion 3’ of exon 2 (r.129_192del; 251_258del) due to the usage of both the alternative donor splice site (c.251_252GT) and the alternative acceptor splice site (c.191_192AG) respectively. These combined splicing events were found in the same transcript as P55 carried both c.183_184TA>CT and c.258+2T>C on the same allele. To note, the potentially novel splicing event r.129_192del was found in one hiPSC line of each patient but not in any normal hiPSC lines.

Consistent with the SBDS downregulation found in patient bone marrow mononuclear cells (Rujkijyanont et al., 2007) and SDS hiPSCs (Tulpule et al., 2013), in this study, SDS hiPSCs also showed SBDS downregulation by log2 fold change of -1.4 (Figure 13A). RNA-seq also showed reduced SBDS transcript expression and coverage (absolute depth of read coverage across the whole transcript) in SDS compared to normal hiPSCs (Figure 13B-C).
Figure 12. Schematic of SBDS splice variants of hiPSCs.
RNA-seq confirmed the expression of the full length SBDS transcript in normal and P357 hiPSCs and SBDS splice variants r.129_258del and r.251_258del in SDS hiPSCs due to the splice site mutation c.258+2T>C. Other potentially novel transcripts found included r.1_128del; 128_129ins129-142-129-1 in N530E, r.129_192del; 251_258del in P55F, and r.129_192del in P357D.

Table 10. Transcriptional features of SBDS variants.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Known</th>
<th>Predicted</th>
<th>Predicted</th>
<th>Potentially Novel</th>
<th>Potentially Novel</th>
<th>Potentially Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SBDS</strong></td>
<td>Full length SBDS</td>
<td>8nt deletion 3’ of E2 (D c.251_252GT)</td>
<td>E2 skipping</td>
<td>64nt deletion 5’ of E2 (A c.191_192AG) and 8bp deletion 3’ of E2 (D c.251_252GT)</td>
<td>64nt deletion 5’ of E2 (A c.191_192AG)</td>
<td>E1 deletion and 142nt intronic retention 5’ of E2</td>
</tr>
<tr>
<td><strong>Protein</strong></td>
<td>Full length p.Lys62*</td>
<td>p.Cys84Tyrfs*4</td>
<td>p.(Glu43Phefs*2)</td>
<td>p.(Glu43Leufs*24)</td>
<td>p.(Glu43Leufs*184) or p.(Glu44_Gln86delins20)</td>
<td>p.(Met1_Val43delins18)</td>
</tr>
</tbody>
</table>

E, exon; D, alternative donor splice site; A, alternative acceptor splice site
Figure 13. Expression of SBDS variants in hiPSCs.
RNA-seq analysis showed significant SBDS downregulation in FPKM in SDS (n = 4) compared to normal (n = 4) hiPSCs (A). SBDS transcript expression in FPKM (error bars represent 95% CI and y-axis is in log2 scale) (B) and coverage (C) between normal (n = 4) and SDS (n = 12) hiPSCs are shown.
Aligned reads of hiPSCs were analyzed for novel \textit{SBDSP1} splice variants and differential expression between SDS and normal hiPSCs without fragment bias correction and multi-read correction by Cuffdiff 2. RNA-seq showed that all hiPSCs transcribed \textit{SBDSP1} variants (Figure 14, Table 11, Supplementary Figure 46 - Supplementary Figure 47, Supplementary Table 2). All hiPSCs transcribed the full length \textit{SBDSP1} (\textit{SBDSP1.1}). All hiPSCs except N530J transcribed the known variant without exon 2 and partial intron 4 (\textit{SBDSP1.2}). P357A transcribed the known variant with exon 5 skipping with 78nt intronic retention 3’ of exon 4 (\textit{SBDSP1.3}). N530J transcribed the known variant with exon 2 and 5 skipping with 78nt intronic retention 3’ of exon 4 (\textit{SBDSP1.4}). Other potentially novel transcripts were found. N530E transcribed a splice variant with 211nt deletion in exon 1 (n.100_310del) due to the usage of an alternative donor splice site (c.100_101GT) and alternative acceptor splice site (c.309_310AG) both within exon 1. P55F transcribed a splice variant with 59nt deletion 5’ of exon 2 (n.412_470del) due to the usage of an alternative acceptor splice site (c.469_470AG). P357 transcribed a splice variant with exon 4 skipping (n.731_895del). RNA-seq showed no significant difference in \textit{SBDSP1} gene expression between SDS and normal hiPSCs.
Figure 14. Schematic of SBDSP1 splice variants of hiPSCs.
RNA-seq confirmed the expression of full length SBDSP1.1 transcript in all hiPSCs, SBDSP1.2 splice variant in all hiPSCs except N530E, SBDSP1.3 in N551I and P357A, and SBDSP1.4 in N530J, N551I, and P357A. Other novel transcripts found included n.100_310del in N530E, n.412_470del in P55F, and n.731_895del in P357D.

Table 11. Transcriptional features of SBDSP1 variants.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Known</th>
<th>Known</th>
<th>Known</th>
<th>Known</th>
<th>Potentially Novel</th>
<th>Potentially Novel</th>
<th>Potentially Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJBDSP1</td>
<td>Full length SJBDSP1</td>
<td>E2 skipping</td>
<td>E5 skipping and 78nt intronic retention 3’ of E4</td>
<td>E2 and E5 skipping and 78nt intronic retention 3’ of E4</td>
<td>211nt deletion in E1 (D c.100_101GT, A c.309_310AG)</td>
<td>59nt deletion 5’ of E2 (A c.469_470AG)</td>
<td>E4 skipping</td>
</tr>
<tr>
<td>hiPSC Lines</td>
<td>N530E, N530J</td>
<td>N530E</td>
<td>N530E</td>
<td>N530J</td>
<td>N530J</td>
<td>N530E</td>
<td>P55F</td>
</tr>
<tr>
<td></td>
<td>N551I, N551K</td>
<td>N551I, N551K</td>
<td>N551I</td>
<td>N551I</td>
<td>N530E</td>
<td>P55F</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P55C, P55F</td>
<td>P55C, P55F</td>
<td>P55C, P55F</td>
<td>P55A</td>
<td>P55A</td>
<td>P55A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P357A, P357D</td>
<td>P357A, P357D</td>
<td>P357A, P357D</td>
<td>P357A</td>
<td>P357A</td>
<td>P357A</td>
<td></td>
</tr>
</tbody>
</table>

E, exon; D, alternative donor splice site; A, alternative acceptor splice site
Figure 15. Expression of SBDS P1 variants in hiPSCs.
RNA-seq analysis showed no differences in SBDS P1 expression in FPKM in SDS (n = 4) compared to normal (n = 4) hiPSCs (A). SBDS transcript expression in FPKM (error bars represent 95% CI and y-axis is in log2 scale) (B) and coverage (C) between normal (n = 4) and SDS (n = 12) hiPSCs are shown.
Aligned reads of hiPSCs were analyzed for differential expression between SDS and normal hiPSCs with fragment bias correction and multi-read correction by Cuffdiff 2. A total of 7953 genes were found to have significant differential expression in SDS compared to normal hiPSCs (Figure 16). 98.8% of the 7953 genes were non-annotated and potentially novel transcripts. Annotated genes were filtered using the criteria that the mean FPKM was greater than 0 in both SDS and normal hiPSCs, resulting in 79 significantly differentially expressed filtered genes (Figure 17). HIST1H1A showed the greatest downregulation and was significantly downregulated in all four SDS hiPSCs by \( \log_2 \) fold change of -4.3, in both P55 hiPSCs by -4.6, and in both P357 hiPSCs by -4.0 compared to the four normal hiPSCs.

![Figure 16](image)

**Figure 16. Scatter plot of significantly differentially expressed, non-filtered genes in SDS compared to normal hiPSCs.**
Aligned reads of feeder depleted CD56/CD184+/TRA-1-60+/SSEA-4+/CD30+ hiPSCs were analyzed for differential expression and a scatter plot of 7953 significantly differentially expressed genes were generated by Cuffdiff 2. The scatter plot shows \( \log_2 \) FPKM of SDS hiPSCs (y-axis, \( n = 4 \)) against that of normal hiPSCs (x-axis, \( n = 4 \)). Each dot represents a gene with significant differential expression.
Figure 17. Significantly differentially expressed filtered genes in SDS compared to normal hiPSCs.

FPKM (top, y-axis in log2 scale, legend above x-axis), log2 fold change (center, legend above x-axis), and q-value (bottom) of 79 significantly differentially expressed filtered genes (0.04 < q < 0.05) in SDS (n = 4) compared to normal (n = 4) hiPSCs are shown. Evidence for SBDS downregulation is outlined in green. Genes with high homology and indistinguishable by aligned transcripts are grouped.
To determine if significantly differentially expressed genes in SDS compared to normal hiPSCs were enriched for certain pathways, over-representation analyses were tested by InnateDB (Breuer et al, 2013). Thirty-three pathways were significantly enriched (q < 0.05, Table 12).

### Table 12. Significantly enriched pathways in SDS compared to normal hiPSCs.

<table>
<thead>
<tr>
<th>Pathway Name</th>
<th>Gene Name</th>
<th>q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Congenital Disorders of Glycosylation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective B4GALT1 causes B4GALT1-CDG (CDG-2d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective B4GALT7 causes EDS, progeroid type</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Congenital Disorders of Glycosaminoglycan Glycosylation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective B3GAT3 causes JDSSDHD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective CHSY1 causes TPBS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective EXT1 causes exostoses 1, TRPS2 and CHDS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective EXT2 causes exostoses 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Congenital Disorders of Sulfation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective PAPSS2 causes SEMD-PA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective SLC26A2 causes chondrodysplasias</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Disorders of Glycosaminoglycan Sulfation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective CHST14 causes EDS, musculocontractural type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective CHST3 causes SEDCJD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defective CHST6 causes MCDC1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diseases associated with glycosaminoglycan metabolism</strong></td>
<td>CD44; CHST11; GNS</td>
<td>0.0245</td>
</tr>
<tr>
<td><strong>Diseases of glycosylation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Glycosaminoglycan metabolism</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lysosomal Storage Disorders</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS I - Hurler syndrome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS II - Hunter syndrome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS IIIA - Sanfilippo syndrome A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS IIIB - Sanfilippo syndrome B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS IIIC - Sanfilippo syndrome C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS IIID - Sanfilippo syndrome D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS IV - Morquio syndrome A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS IV - Morquio syndrome B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS IX - Natowicz syndrome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS VI - Maroteaux-Lamy syndrome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS VII - Sly syndrome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucopolysaccharidoses</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reduction of cytosolic Ca++ levels</strong></td>
<td>ATP2A2; ATP2B1</td>
<td>0.0264</td>
</tr>
<tr>
<td><strong>Ion transport by P-type ATPases</strong></td>
<td>ATP2B1; DPEP3</td>
<td>0.0273</td>
</tr>
<tr>
<td><strong>Arachidonic acid metabolism</strong></td>
<td>CYP1B1</td>
<td>0.0354</td>
</tr>
<tr>
<td><strong>Cell surface interactions at the vascular wall</strong></td>
<td>CD44; ESAM; PROCR; SLC7A8</td>
<td>0.0386</td>
</tr>
<tr>
<td><strong>Basal cell carcinoma</strong></td>
<td>FZD2; WNT10B</td>
<td>0.0413</td>
</tr>
<tr>
<td><strong>DNA Damage/Telomere Stress Induced Senescence</strong></td>
<td>HIST1H1A; HIST1H2BK</td>
<td>0.0492</td>
</tr>
<tr>
<td><strong>Interferon gamma signaling</strong></td>
<td>CD44; IFI30</td>
<td>0.0495</td>
</tr>
</tbody>
</table>
5 Lentiviral Transduction of SDS hiPSCs

In preparation for future experiments, stably transduced SDS hiPSC lines were established by lentiviral transduction of P55C and P357D hiPSCs (Figure 18) using plasmids pLVX.SIN.EF1a.SBDS-HA.IRES.ZsGreen for SBDS rescue and pLVX.SIN.EF1a.IRES.ZsGreen for mock control (Figure 19). Transduced hiPSCs could be subjected to Western blot, hematopoietic differentiation, colony assay, FACS, and RNA-seq to validate previous results.

![Figure 18. Lentiviral transduction of SDS hiPSCs.](image)

SDS P55C (top) and P357D (bottom) hiPSCs were transduced with pLVX.SIN.EF1a.IRES.ZsGreen (left) and pLVX.SIN.EF1a.SBDS-HA.IRES.ZsGreen (right) by lentivirus with 8ug/ml protamine sulfate.
Figure 19. Lentiviral plasmid maps.
EF1α promoter drove the expression of SBDS and ZsGreen in pLVX.SIN.EF1a.SBDS-HA.IRES.ZsGreen (left) and pLVX.SIN.EF1a.IRES.ZsGreen (right).
6 Optimization of RT-qPCR

In preparation for future experiments, RT-qPCR primers were optimized using the cDNA of normal hiPSCs cultured on matrigel (Table 13, Kozera and Rapacz, 2013). Primer concentrations were optimized by testing forward primer-to-reverse primer ratios (50nM:50nM, 50nM:300nM, 300nM:50nM, and 300nM:300nM). Primers and concentrations fulfilling the criteria of standard deviation below 0.16 and absence of primer dimer in both the dissociation curve and no template control were further optimized. Efficiency was assessed by generating standard curves with 5× serial dilutions of cDNA. Amplification plots (Figure 20), dissociation curves (Figure 21), and standard curves were assessed (Figure 22). Primers and concentrations fulfilling the previous criteria and the criteria of efficiency between 90 and 110% and linear regression above 0.958 were selected. Optimized primers could be used to validate RNA-seq results.
Table 13. Concentration of optimized and designed RT-qPCR primers.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Purpose</th>
<th>Primer Name</th>
<th>Sequence</th>
<th>Concentration (nM)</th>
<th>Efficiency (90-110%)</th>
<th>R² (&gt;0.985)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Forward Primer</td>
<td>Reverse Primer</td>
<td>Forward Primer</td>
<td>Reverse Primer</td>
</tr>
<tr>
<td>RPLP0</td>
<td>Reference Gene</td>
<td>RPLP0_805-932</td>
<td>ACTCTGCATTCCTCG</td>
<td>GACAAGGCCAGG</td>
<td>300</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CTTTCT</td>
<td>ACTCGTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPRT1</td>
<td>Reference Gene</td>
<td>HPRT1_183-322</td>
<td>CCTGGCCGTCTGAT</td>
<td>TCTCGAGCAAGA</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TAGTGA</td>
<td>CGTTCAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2M</td>
<td>Reference Gene</td>
<td>B2M_309-425</td>
<td>TCTCTGGTACTACA</td>
<td>GCTGCTTACATGT</td>
<td>300</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CTGAAATTCACC</td>
<td>GCTCGATCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARHGIF12</td>
<td>Positive Control</td>
<td>ARHGIF12_1532-1670</td>
<td>TCGTCCGATCTTCC</td>
<td>ATAGTGCGGATG</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TTGAGT</td>
<td>CAGATCCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBP</td>
<td>Reference Gene</td>
<td>TBP_291-427</td>
<td>CAACACGTGCACCA</td>
<td>ACAGACTATTG</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CCTTAC</td>
<td>GATTTCTGAATAGG</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>GUSB</td>
<td>Reference Gene</td>
<td>GUSB_1788-1933</td>
<td>CCACCTCTGATGTT</td>
<td>ACTCTCGTGCGTG</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CACTGAAG</td>
<td>ACTGTTG</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>Reference Gene</td>
<td>GUSB_805-988</td>
<td>CCACCTACATCGG</td>
<td>CTGACACCTGGG</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GACATCAGC</td>
<td>ACCTTAAAG</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>GUSB</td>
<td>Reference Gene</td>
<td>GUSB_888-1054</td>
<td>AGTAACCTGTTCG</td>
<td>GTGCAAGCTG</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GATACTCGG</td>
<td>CCAATG</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>
Figure 20. RT-qPCR amplification plot of primers targeting RPLP0.
Amplification plot shows the normalized fluorescence (y-axis) at each cycle of amplification (x-axis). Standard curve was generated using six serial dilutions of cDNA shown as six amplification curves (different colors) with three replicates each. Default threshold is shown as a green horizontal line at fluorescence of 0.241. The intercept of each amplification curve and the threshold indicates the threshold cycle.
(C_T) of each dilution. The threshold cycle was used to generate the standard curve. No template control is shown as a horizontal, purple, dotted line at fluorescence of 0.

**Figure 21. RT-qPCR dissociation curve of primers targeting RPLP0.**
Dissociation curve shows the normalized fluorescence (y-axis) and temperature (x-axis). The peak of fluorescence indicates the melting temperature (T_m). No template control is shown as a horizontal, purple, dotted line at fluorescence of 0.
Figure 22. RT-qPCR standard curve of primers targeting RPLP0.
Standard curve shows the threshold cycle (y-axis) at each 5x serial dilutions (x-axis). The slope indicates the efficiency of RT-qPCR. Linear regression (RSq) and efficiency (Eff) are listed at the top.
Chapter 5
Discussion

Characterized by bone marrow failure, exocrine pancreatic dysfunction, and skeletal abnormalities, SDS is a rare, autosomal recessive, multi-system disorder that is mainly caused by mutations in the SBDS gene (Bodian et al., 1964; Shwachman et al., 1964; Ginzberg et al., 2000; Goobie et al., 2001; Boocock et al., 2003). The SBDS gene encodes the SBDS protein that is essential in ribosome biogenesis as SBDS interacts with pre-60S and EFL1 to release eIF6 which inhibits the joining of 60S with 40S (Menne et al., 2007; Finch et al., 2011; Weis et al., 2015). Deficient ribosome production not only reduces global protein production (Ball et al., 2009), but also impedes the translation of isoforms that require multiple ribosome bindings, such as isoforms of the granulocytic transcription factor C/EBP (In et al., 2016). Reduced expression of C/EBP isoforms has been shown to result in impaired terminal differentiation of neutrophils. Impaired terminal differentiation of neutrophils has been found in both human and mice with SDS, although varying in the differentiation stage of arrest. Patient bone marrow has shown reduced myeloid progenitors and their progenies (Orelio et al., 2009), while SDS mice have shown reduced neutrophils due to p53-mediated apoptosis (Zambetti et al., 2015). Myeloid cell formation is also hindered by protease-mediated auto-digestion. hPSC-derived myeloid cells have been shown to produce excess primary azurophilic granules that lead to apoptosis (Tulpule et al., 2013).

Other than neutrophil formation, hematopoietic defects have also been found in earlier hematopoietic progenitors. Patient bone marrow has shown reduced CD34+ hematopoietic stem and progenitor cells (Dror and Freedman, 1999), colony assays of patient bone marrow and hPSCs have shown reduced colony forming potential (Saunders et al., 1979; Suda et al., 1982; Kuijpers et al., 2005; Dror and Freedman, 1999; Tulpule et al., 2013), and mouse Sbds knockdown hematopoietic progenitors have shown impaired engraftment and myeloid proliferation (Rawls et al., 2007). By differentiating SDS hiPSCs towards both the primitive and definitive programs of hematopoietic differentiation and identifying developmental stages by immunophenotyping, preliminary findings in our lab have shown impairment in the definitive rather than the primitive program (Luca, 2015).
As a continuation of the previous study, this study used the same method of definitive hematopoietic differentiation of hiPSCs (Kennedy et al., 2012) with more replicates of hiPSC lines and more markers to refine immunophenotyping. We also aimed to identify gene dysregulation potentially linked to disrupted pathways through RNA-seq. We hypothesized that there were developmental defects with gene dysregulation in SDS compared to normal hiPSCs during hematopoietic development. This study provided the characterization of mesodermal and hemogenic endothelial stages at the cellular level and novel findings in the gene expression of SDS patient cells in the undifferentiated pluripotent stem cell stage. Analyses of later stages of hematopoietic differentiation at the cellular and transcriptome levels could be completed in follow-up studies.

1 Flow Cytometry Analyses

hiPSCs. To determine if the potential to give rise to hematopoietic stages was impaired in SDS compared to normal hiPSCs, hiPSCs were differentiated towards the definitive program by recapitulating embryonic developmental stages, where developmental stages were quantified by immunophenotyping. Representing the embryonic pluripotent stage capable of giving rise to all germ layers, the hiPSC stage in SDS has shown reduced proliferation (Ruiz-Gutierrez et al., 2016). In this study, SDS and normal hiPSCs were comparable in terms of spontaneous differentiation, marked by CD56+ early mesodermal and ectodermal cells and CD184+ endodermal cells, and pluripotent status, marked by CD30+ undifferentiated cells.

Mesoderm. During gastrulation of embryonic development, mesodermal cells in the posterior region of the primitive streak migrate anteriorly to give rise to the hematopoietic lineages (Luckett, 1978; Huber et al., 2004; Ditadi et al., 2017), thus the mesoderm is one of our major developmental stages of interest. SDS hESCs and hiPSCs have shown no differences in mesodermal specification as demonstrated by normal endothelial and cardiac gene expression and mesodermal structure in the teratoma (Tulpule et al., 2013). Consistent with the findings of Tulpule et al., in this study, SDS and normal hiPSCs were comparable in mesodermal formation in day 4 EBs. Although the yield of the KDR+ mesodermal population (Kabrun et al., 1997) was low and might have affected the results, analyses of other markers were not affected. SDS and normal hiPSCs were comparable in their potential to generate the CD56+ mesodermal
progenitors capable of giving rise to all mesodermal lineages (Evseenko et al., 2010) in day 4 EBs.

**Hemogenic endothelium.** During intra-embryonic definitive hemato poiesis, endothelial cells of the dorsal aorta in the AGM region give rise to the hemogenic endothelium that further differentiates to the definitive hematopoietic lineage (de Bruijn et al., 2000). Hence the hemogenic endothelium is one of our hematopoietic developmental stages of interest. SDS and normal hiPSCs were comparable in their potential to generate CD34+/CD43−/CD184−/CD73− and CD34+/CD43−/VE-cad+/CD184+/CD73− hemogenic endothelium in day 8 EBs.

**Endothelia.** However, SDS hiPSCs showed significantly increased potential to generate CD34+ endothelial population containing both hemogenic and vascular endothelia at day 8. Among the three types of endothelia (hemogenic, arterial, and venous), SDS hiPSCs showed significantly increased potential to generate CD34+/CD43+/CD184−/CD73+ venous endothelium and other populations containing the venous endothelium (CD73+, CD184+/CD73+, and CD34+/CD184−). Enhanced venous endothelial potential, which might be a result of reduced hematopoietic potential, has not been reported in SDS and its consequences are unknown.

Endothelial cells have been shown to maintain hematopoietic stem cells and regulate leukocyte trafficking. In the fetal liver, vascular identity affects the hematopoietic stem cell niche (Khan et al., 2016). The transition of portal vessels from arterial to venous identity promotes the migration of hematopoietic stem cells from the liver to the bone marrow. In the bone marrow, vascular identity also affects the metabolic state of hematopoietic stem cells (Itkin et al., 2016). Hematopoietic stem cells in the quiescent state reside in association with arteriolar vessels, while those in the activated state reside in the sinusoid. Hematopoietic stem cells are also supported by endothelial cells through the secretion of angiocrine factors such as stem cell factor (Ding et al., 2012) and Notch ligands (Butler et al., 2010). In addition, endothelial cells support hematopoietic stem cell regeneration, as irradiated mouse bone marrow has been shown to be rescued by transplantation or co-culture with endothelial cells (Chute et al., 2004; Chute et al., 2007; Salter et al., 2009). Moreover, endothelial cells also permit leukocyte trafficking in the venous vessels and sinusoid in the bone marrow (Itkin et al., 2016). As the vasculature supports many tissues, given that SDS can manifest in multiple organ systems, whether defects in the
vasculature in SDS affect the hematopoietic stem cell niche, leukocyte trafficking, and other organs requires further investigation.

**Reduced viability.** SDS has been shown to be associated with Fas-mediated apoptosis (Dror *et al.*, 2001; Watanabe *et al.*, 2009; Ambekar *et al.*, 2010) and p53-mediated apoptosis in myelocyte and metamyelocytes (Zambetti *et al.*, 2015; Tourlakis *et al.*, 2015). Consistent with the aforementioned findings, in this study, SDS hiPSCs also showed significantly reduced viability marked by PI- viable population in day 9 EBs. The reduced viability might be a consequence of the p53-mediated G1 cell cycle arrest and apoptosis due to DNA damage and insufficient translation (Austin *et al.*, 2008). Therefore, reduced CD34+/CD43- hemogenic and vascular endothelial potential in SDS found in the preliminary findings in our lab might be due to reduced viability in SDS day 9 EBs found in this study.

**CFU assay.** During definitive hematopoiesis, the hemogenic endothelium gives rise to hematopoietic stem cells capable of long-term multi-lineage reconstitution (de Bruijn *et al.*, 2000). Although long-term reconstitution capability was not achieved in this hiPSC differentiation, a similar progenitor with multi-lineage potential referred to as early hematopoietic progenitors (Kennedy *et al.*, 2012) was assessed by CFU assays. The number of hematopoietic colonies generated in a CFU assay reflected the proliferation and differentiation of the plated hematopoietic progenitors *in vitro*, while the size of the colony reflected the proliferation of the progenitor *in vitro*.

Studies of CFU assays using patient hematopoietic progenitors have shown reduction in CFU-GEMM (granulocyte, erythroid, monocyte, megakaryocyte), CFU-GM (granulocyte, monocyte), CFU-G (granulocyte), CFU-E (erythroid), BFU-E (erythroid), although CFU assays of SDS hPSC-derived hematopoietic progenitors have shown reduced total colony count, but no differences in CFU-G, CFU-M (monocyte), or CFU-GM (Tulpule *et al.*, 2013). Other studies using patient bone marrow samples have also shown discrepancies. Staining has shown no difference in early myeloid progenitors (Orelio *et al.*, 2009), but immunophenotyping has shown reduced myeloid progenitors (Mercuri *et al.*, 2015).

In this study, SDS day 9+7 early hematopoietic progenitors showed significantly reduced myeloid potential in generating CFU-GM and BFU-E, but no significant reduction in CFU-
GEMM (p = 0.0508), indicating that the onset of hematopoietic defect occurred in hematopoietic progenitors, as SDS hematopoietic progenitors were impaired in differentiating to myeloid progenitors. Whether the onset of hematopoietic defects occurred in earlier stages such as hematopoietic stem cells or multipotent progenitors requires further evidence from hiPSC-derived early hematopoietic progenitors.

Although the hiPSCs in this study were derived from patients who did not present overt anemia, their hiPSCs showed reduced erythroid potential. Previous clinical data and CFU assays have also shown reduced CFU-E (Suda et al., 1982) and BFU-E (Kuijpers et al., 2005; Sen et al., 2011) and that SDS can manifest as anemia (Ginzberg et al., 1999; Hashmi et al., 2011). The presence of erythroid defect in the cell lines of SDS patients without anemia might be due to the fact that inherent erythropoietic impairment might not be compensated in vitro. Although SBDS knockdown in human cell lines reduced erythroid potential (Sezgin et al., 2013), Sbds knockdown in mice did not elicit erythroid defects (Rawls et al., 2007).

2 RNA-seq Analysis of SDS hiPSC

**SDS-associated SBDS transcripts.** The main cause of SDS are biallelic mutations in the SBDS gene and exon 2 is considered a mutational hotspot with the most common mutations being c.183_184TA>CT and c.258+2T>C (Boocock et al., 2003). In this study, the splice site mutation c.258+2T>C resulted in two possible splicing events in SDS hiPSCs. Either a cryptic donor splice site within exon 2 (c.251_252GT) was used to generate the SBDS splice variant with 8nt deletion 3’ of exon 2 (r.251_252GT) as predicted (Boocock et al., 2003) or the upstream donor splice site of intron 1 (c.128+1_2GT) was used to generate the SBDS splice variant without exon 2 (r.129_258del). The nonsense mutation c.183_184TA>CT showed no apparent effect on transcription and P357 hiPSCs with the nonsense mutation on one allele (c.[183_184TA>CT];[258+2T>C]) were able to transcribe the full length SBDS transcript that carried the premature stop codon. The

**Potentially novel SDS-associated SBDS transcripts.** Detected in P357D hiPSCs which carried the mutations c.[183_184TA>CT];[258+2T>C], r.129_192del was found to be a potentially novel SBDS splice variant that used an alternative acceptor splice site (c.191_192AG), resulting in the deletion of 64nt. Detected in P55F hiPSCs which carried the splice site mutation on both
alleles c.[183_184TA>CT; 258+2T>C];[258+2T>C], r.129_192del; 251_258del was found to be a potentially novel SBDS splice variant that showed combined splicing events on the same transcript. To note, the potentially novel splicing event r.129_192del was found in one hiPSC line, but not in the other isogenic line, in both SDS patients but not in any normal hiPSC lines.

Although expressed at low levels in SDS hiPSCs, full length SBDS with the nonsense mutation and r.251_258del have been shown to encode for mutant proteins p.Lys62* and p.Cys84Tyrfs*4 respectively, although by forced expression (Orelio et al., 2009). Similarly, other SBDS transcripts r.129_258del, r.129_192del, and r.129_192del;251_258del might also translate their predicted proteins p.(Glu43Phefs*2), p.(Glu43Leufs*24), and p.(Glu43Leufs*184). Whether these splice variants are translated, stable, or functional requires further experimental validation.

**Potentially novel SBDS transcript.** Detected in N530E hiPSCs, r.1_128del;128_129ins129-142_129-1 was found to be a potentially novel SBDS transcript with exon 1 deletion and 142nt intronic retention 3’ of intron 1, possibly due to a cryptic transcription start site in intron 1. An alternative acceptor splice site was found in intron 1 (c.129-143_129-144AG). An alternative donor splice site upstream of exon 1 or an alternative transcription start site in intron 1 might be used. As this transcript was only found in one normal hiPSC line, whether it is an artifact of hiPSC RNA-seq or a splice variant expressed naturally requires further experimental validation.

**SBDS downregulation.** SBDS downregulation has been shown in patient bone marrow mononuclear cells (Rujkijyanont et al., 2007) and SDS hPSCs (Tulpule et al., 2013). Consistent with the aforementioned findings, in this study, SBDS also showed significant downregulation (log2 fold change of -1.4) in SDS compared to normal hiPSCs, which could be due to non-sense mediated decay.

**SBDSP1 transcripts.** The common mutations in SBDS are introduced by gene conversion with its pseudogene (SBDSP1) with 97% homology (Boocock et al., 2003). In this study, SBDSP1 dysregulation was not detected in SDS compared to normal hiPSCs, but potentially novel SBDSP1 splice variants were identified. Detected in N530E hiPSCs, n.100_310del was found to be a potentially novel splice variant that used an alternative donor splice site (c.100_101GT) and an alternative acceptor splice site (c.309_310AG) both within exon 1, resulting in 211nt deletion in exon 1. Detected in P55F hiPSCs, n.412_470del was found to be a potentially novel splice
variant that used an alternative acceptor splice site (c.469_470AG), resulting in 59nt deletion 5’ of exon 2. Detected in P357 hiPSCs, n.731_895del was found to be a potentially novel splice variant with exon 4 skipping. As these transcripts were only found in one hiPSC line but not in the other isogenic line, whether they are an artifact of hiPSC RNA-seq or a splice variant expressed naturally requires further experimental validation. Whether SBDSP1 has a function in regulating SBDS expression through interactions with SBDS DNA or mRNA is unclear.

**Differential expression.** As reduced global synthesis in SDS could have a negative effect on transcription through the reduced synthesis of transcription factors and RNA polymerase, studying the transcriptome of SDS could provide a global perspective of the mechanisms affected by SBDS deficiency. Oligonucleotide microarray of SDS patient bone marrow mononuclear cells has shown dysregulation in oncogenes and tumor suppressor genes (Rujkijyanont et al., 2007), ribosomal genes (Rujkijyanont et al., 2008), and apoptosis genes (Rujkijyanont et al., 2009). RNA-seq of SDS mouse embryonic fibroblasts has also shown increase lysosome trafficking and activity with reduced ATP and lactate levels (Calamita et al., 2017). To investigate the molecular mechanisms underlying SDS defects, RNA-seq analysis of SDS compared to normal hiPSCs revealed 7953 significantly differentially expressed genes and only 1% were annotated genes with FPKM greater than zero. The remaining 99% non-annotated and potentially novel transcripts, mostly non-coding RNA, requires further annotation and characterization to determine their potential functions. In this study, the RNA-seq analysis focused on identifying annotated, differentially expressed genes to reveal potentially disrupted pathways in SDS.

**HIST1H1A.** HIST1H1A (Histone Cluster 1 H1 Family Member A) showed significant downregulation in both P55 hiPSCs (log2 fold change of -4.6), both P357 hiPSCs (log2 fold change of -4.0), and all four SDS hiPSCs (log2 fold change of -4.3) with the greatest magnitude of downregulation. HIST1H1A encodes for the human linker histone variant H1.1, which is a tissue-specific, somatic variant. Bound to the nucleosome and linker DNA, linker histones (H1) are essential for nucleosome stability, chromatin remodeling, and transcriptional regulation (Millán-Ariño et al., 2016). Although H1.1 has been shown to participate in transcriptional repression (Kandolf, 1994), it has been later recognized to participate in transcriptional activation due to its high mobility and euchromatin localization (Alami et al., 2003; Orrego et al., 2007; Th’ng et al., 2005). H1.1 is classified as a replication-dependent H1 variant with the highest
expression during S phase of the cell cycle in dividing cells (Lennox and Cohen, 1983; Wang et al., 1997). H1.1 is also differentially expressed during development, with high levels of expression during early development and in immature cells (thymus progenitors, neuroblasts, hPSCs), and low levels during late development and in terminally differentiated cells (circulating lymphocytes and neurons) (Lennox and Cohen, 1983; Domínguez et al., 1992; Terme et al., 2011). Therefore, HIST1H1A downregulation in SDS compared to normal hiPSCs found in this study might affect chromatin remodeling and transcription, especially in rapidly dividing cells such as hiPSCs.

Interestingly, SBDS has been shown to potentially interact with other linker histone variants H1.4 (HIST1H1E) and H1.2 (HIST1H1C) (Ball et al., 2009), while its function in the nucleus is unclear (Austin et al., 2005). Ribosomal proteins have also been shown to interact with linker histones (Kalashnikova et al., 2013) to regulate transcription (Ni et al., 2006). Therefore, the connection between SBDS and linker histones might be important in understanding the transcriptional consequences of SBDS deficiency.

Although H1 variants are considered to be redundant as H1.1−/− mice have shown normal development due to the compensatory increase of other H1 variants, H1.1 has shown unique genomic binding profiles compared to other H1 variants such as enrichment at highly methylated exons, suggesting that H1.1 may have a unique epigenetic role (Rabini et al., 2000; Lin et al., 2004; Izzo et al., 2013). In this study, the downregulation of HIST1H1A associated with transcriptional activation in specific genomic regions might lead to the deposition of other transcriptionally repressive linker histones, resulting in transcriptional repression in these specific regions. Therefore, H1.1 might have an important role in transcriptional regulation that might be disrupted by SBDS deficiency. Moreover, compensatory upregulation of other H1 variants was not found, which might indicate a potential detrimental effect of HIST1H1A downregulation.

**Over-representation.** Over-representation analysis of significantly differentially expressed genes in SDS compared to normal hiPSCs revealed several enriched pathways, which are discussed in the following.
Histones. *HIST1H1A* (described above) and *HIST1H2BK* (Histone Cluster 1 H2B Family Member K, encodes for the core histone variant H2BK) were found to be significantly downregulated (log₂ fold change of -4.3) and upregulated (log₂ fold change of 1.3) respectively in SDS compared to normal hiPSCs and participate in the pathway of DNA damage or telomere stress induced senescence. During DNA damage, which often occurs in telomeres, both linker histone and core histones have been shown to be targets of post-translational modifications and degradation in order to recruit the DNA repair machinery and initiate cell cycle arrest (Fernandez-Capetillo *et al.*, 2004; Giannattasio *et al.*, 2005; Thorslund *et al.*, 2015; Hauer *et al.*, 2017). Interestingly, our lab has previously demonstrated shortened telomeres in SDS compared to healthy control subjects (Thornley *et al.*, 2002). Since telomeres contain nucleosomes and are organized in heterochromatin-like structures with enrichment for specific histones, it would be important to clarify whether the loss of SBDS affects chromatin at telomeres and consequently leads to telomere shortening. Upregulation of core histone could lead to global genomic repression, although these excess transcripts could also be targeted for degradation. However, the roles of H1.1 and H2BK have been shown to be functionally redundant among their variants and whether *HIST1H1A* downregulation and *HIST1H2BK* upregulation have an effect in SDS requires further investigation.

Glycosaminoglycan metabolism. *CD44, CHST11* (Carbohydrate (Chondroitin 4) Sulfotransferase 11), and *GNS* (Glucosamine (N-acetyl)-6-Sulfatase) were found to be significantly upregulated (log₂ fold change of 1.3, 1.3, and 1.2 respectively) in SDS compared to normal hiPSCs and participate in the pathway of glycosaminoglycan metabolism. Glycosaminoglycan is a long polysaccharide group synthesized by glycosylation (Jaeken, 2010) and modified by sulfation (Mizumoto *et al.*, 2013). For instance, the sulfation of the glycosaminoglycan, chondroitin, is catalyzed by the sulfotransferase CHST11 (Hiraoka *et al.*, 2000). Targeted by receptors such as the hyaluronan and chondroitin receptor CD44 (Guarnier *et al.*, 1993; Kawashima *et al.*, 2000), glycosaminoglycan is delivered to the lysosome and degraded by lysosomal enzymes, such as GNS (Freeman *et al.*, 1987), as part of its natural turnover. As defects in lysosomal enzymes lead to glycosaminoglycan accumulation in the lysosome, lysosomal storage disorders were generated as part of the pathway (Coutinho *et al.*, 2012).
Lysosome degradation has been shown to be regulated by glycosaminoglycan sulfation patterns as reduced sulfated glycosaminoglycans increased the activity of the lysosomal protease cathepsin D (Lehri-Boufala et al., 2015). CHST11 and GNS play opposite roles in regulating glycosaminoglycan sulfation patterns. CHST11 adds sulfates to position 4 of N-acetylgalactosamine of chondroitin during glycosaminoglycan activation and GNS removes sulfates from glucose-N6S of heparan and keratan sulfate during glycosaminoglycan deactivation. Specific arrangements of sulfate groups on glycosaminoglycans bind to specific ligands or receptors to facilitate the formation of ligand-receptor signaling complex or inhibit ligand-receptor interactions (Gama et al., 2006). For instance, heparan sulfate has important functions in the hematopoietic system, such as interacting with cytokines and growth factors in the microenvironment (Coombe, 2008), maintaining LTC-IC (Gupta et al., 1998), increasing myeloid potential in CFU assays (Bramono et al., 2011), supporting hematopoietic differentiation (Holley et al., 2011), mediating leukocyte migration via CD44 (Murai et al., 2004), and inducing mobilization of hematopoietic stem progenitor cells for engraftment (Albanese et al., 2009). Sulfated glycosaminoglycans also have other roles in development, such as facilitating embryonic patterning of Nodal (Oki et al., 2007), promoting neuronal specification (Pickford et al., 2011), regulating axon growth (Shen, 2014), and inhibiting osteoclastogenesis (Baud’huin et al., 2011).

*CHST11* has also been shown to participate in multiple cellular processes, such as inflammatory skin diseases, cartilage development, neuronal inhibition, and cancer. In terms of chronic inflammatory diseases of the skin, chondroitin-4-sulfate has been shown to be accumulated in the lesions of discoid lupus erythematosus and dermatomyositis, while *CHST11* gene expression in dermal fibroblasts has been shown to be induced by IFN-γ (Kim and Werth, 2011), which is discussed later in the section. In terms of cartilage development, *CHST11* has also been shown to be upregulated in osteoarthritis cartilage (Karlsson et al., 2010) and to be essential in cartilage growth plate and chondrocyte development (Klüppel et al., 2005), while chondroitin-4-sulfate has been shown to interact with hedgehog signaling components essential for growth plate and chondrocyte development (Cortes et al., 2009). In terms of neuronal inhibition, chondroitin-4-sulfate has been shown to be upregulated in glial scar and inhibit neurite extension (Gilbert et al., 2005), as well as to act as a negative guidance cue to neurons and inhibit axonal growth (Wang et al., 2008). In terms of cancer progression, although reduced CHST11 protein and chondroitin-
4-sulfate expression have been implicated in malignancies in Costello syndrome (Klüppel et al., 2012). CHST11 has been shown to be upregulated in aggressive breast cancer cells and to participate in metastatic colonization (Cooney et al., 2011), while increased CHST11 protein expression has been shown to be associated with poor prognosis of ovarian cancer (Oliveira-Ferrer et al., 2015). Therefore, CHST11 upregulation might potentially be associated with growth plate abnormalities, neurological dysfunction, and predisposition to leukemia featured in SDS.

GNS has been shown to participate in WNT signaling, inflammatory response, and response during injury or toxicity. Through desulfation of cell surface heparan sulfate, the avian orthologue of GNS, Qsu1f1, has been shown to activate Wnt signaling (Dhoot et al., 2001), which is discussed later in the section. In response to bacterial infection, lipopolysaccharide-Toll-like receptor 4 signaling has been shown to induce GNS expression to degrade extracellular matrix for cellular infiltration and to inhibit proteoglycan signaling on inflammatory cells (Wells et al., 2005), while GNS upregulation has been shown to correlate with antibody production against polysaccharide-encapsulated bacteria (O’Connor et al., 2017). GNS has also been shown to be upregulated after spinal cord injury in mice (Pajoohesh-Ganji et al., 2012), after exposure to cadmium in human renal epithelial cells (Garrett et al., 2013), and after chemotherapeutic treatment in cancer cells as a part of active lysosomal biogenesis (Zhitomirsky and Assaraf, 2015). Increased lysosomal activity was reported in SDS MEFs (Calamita et al., 2017) and lysosomes may participate in the degradation of proteins, such as excess azurophilic granules in neutrophils and zymogen granules in acinar cells in SDS (Resau et al., 1984; Tulpule et al., 2013).

**Calcium pumps.** ATP2A2 (ATPase Sarcoplasmic/Endoplasmic reticulum Ca2+ Transporting 2) and ATP2B1 (ATPase Plasma Membrane Ca2+ Transporting 1) were found to be significantly upregulated (log₂ fold change of 0.9 and 1.1 respectively) in SDS compared to normal hiPSCs and both encode for ATP-dependent calcium pumps that reduce cytosolic calcium levels either by storage in the sarcoplasmic or endoplasmic reticulum or transport out of the cell respectively. Upregulation of calcium pumps might serve to counteract elevated cytosolic calcium levels (Ravera et al., 2016) and might result in endoplasmic reticulum stress in SDS (Ball et al., 2009). Elevated calcium levels in SDS has been shown to be a consequence of mTOR hyperactivation in response to reduced ATP levels, resulting in switching the energy production to glycolysis
(Ravera et al., 2016) due to impaired mitochondrial respiration (Ball et al., 2009; Henson et al., 2013).

Although mitochondrial pathways were not found to be disrupted in SDS compared to normal hiPSCs, an accessory subunit of complex I NADH dehydrogenase of the electron transport chain (Emahazion and Brookes, 1998), NDUF8 (NADH:Ubiquinone Oxidoreductase Subunit B8), was found to be downregulated (log₂ fold change of -0.9), consistent with the downregulation of mitochondrial genes (Calamita et al., 2017) and reduced mitochondrial respiration (Ball et al., 2009; Henson et al., 2013) reported in SDS. Although energy pathways were not found to be disrupted in SDS compared to normal hiPSCs, a regulatory subunit of the energy sensor AMPK (AMP-activated protein kinase), PRKAB1 (Protein Kinase AMP-Activated Non-Catalytic Subunit Beta 1), was found to be upregulated (log₂ fold change of 1.1), consistent with increased AMPK activity reported in SDS (Ravera et al., 2016).

**Arachidonic acid metabolism.** CYP1B1 (Cytochrome P450 Family 1 Subfamily B Member 1) and DPEP3 (Dipeptidase 3) were found to be significantly downregulated (log₂ fold change of -1.5 and -2.9 respectively) in SDS compared to normal hiPSCs and participate in the pathway of arachidonic acid (polyunsaturated omega-6 fatty acid) metabolism. Dipeptidase participates in the conversion of arachidonic acid into leukotrienes and is highly expressed in the testis (Yoshitake et al., 2011). Some members of the cytochrome P450 family are involved in arachidonic acid metabolism, but CYP1B1 participates in the metabolism of drugs and steroids, such as estradiol (Hanna et al., 2000), anticancer drugs (Rochat et al., 2001), and resveratrol (Potter et al., 2002), and has been shown to be expressed in malignant tumors but not corresponding normal tissues (Murray et al., 2001). As arachidonic acid has been shown to induce global protein synthesis through activating PI3K/AKT/mTOR pathway (Neeli et al., 2003), DPEP3 downregulation might reduce arachidonic acid conversion and increase arachidonic acid availability to activate mTOR and increase translation in SDS.

**Vascular cell surface interactions.** CD44 (log₂ fold change of 1.3), ESAM (Endothelial Cell-Selective Adhesion Molecule, log₂ fold change of 1.7), PROCR (Protein C Receptor, log₂ fold change of -1.1), and SLC7A8 (Solute Carrier Family 7 (Amino Acid Transporter Light Chain, L System) Member 8, log₂ fold change of -0.9) were found to be significantly dysregulated in SDS compared to normal hiPSCs and participate in the pathway of vascular cell surface interactions.
CD44 plays many roles, such as a receptor for hyaluronan and components of the extracellular matrix, as a scaffold protein for growth factors and metalloproteinases, as a co-receptor that mediates signaling of receptor tyrosine kinases, and as a transmembrane protein that links the plasma membrane to the actin cytoskeleton via crosslinking proteins (Ponta et al., 2003). ESAM is a tight junction protein expressed on mouse hematopoietic stem cells (Oguro et al., 2013), endothelial cells, megakaryocytes, and platelets (Nasdala et al., 2002). ESAM has a redundant function in physiological angiogenesis, but plays a role in tumor angiogenesis (Ishida et al., 2003) and regulating thrombus growth and stability (Stalker et al., 2009). PROCR is a protein C receptor expressed on endothelial cells and neutrophils and participates in the protective mechanism that limits blood coagulation, inflammatory responses, and endothelial cell apoptosis (Esmon, 2003; Sturn et al., 2003). SLC7A8 is a subunit of the system L amino acid transporter 2 that is an energy-independent exchanger of neutral amino acids and thyroid hormone transporter (Bassi et al., 1999; Braun et al., 2011). Each of these genes plays an independent role but were linked by the fact that they are membrane proteins expressed in endothelial cells.

**WNT signaling.** FZD2 (Frizzled Class Receptor 2) and WNT10B (WNT Family, Member 10B) were found to be significantly upregulated (log2 fold change of 1.5 and 2.0 respectively) in SDS compared to normal hiPSCs and participate in the pathway of basal cell carcinoma. Both genes are components of WNT (Wingless-Type MMTV Integration Site) signaling, which plays many roles in embryonic and tissue development, while dysregulation in components of WNT signaling are associated with cancer (Clevers and Nusse, 2012). During hematopoietic differentiation, activation of WNT signaling during mesoderm induction has been shown to promote the definitive program and inhibit the primitive program (Sturgeon et al., 2014). In this study, as both mesodermal and hemogenic endothelial potentials were comparable between SDS and normal hiPSCs, upregulation of WNT components in hiPSCs might not have an effect during early hematopoiesis.

FZD2 is a WNT receptor involved in neuronal development (Bhat et al., 1998; Rodriguez et al., 2005) and cardiac organization (Toyofuku et al., 2000). WNT10B is a WNT ligand that activates WNT signaling for hematopoietic development (Lento et al., 2013), mammary gland development (Lane and Leder, 1997), adipose inhibition (Longo et al., 2004), bone formation (Bennet et al., 2005), brain development (Lekven et al., 2003), skin epithelial differentiation (Ouji et al., 2006), hair follicle development (Ouji et al., 2006), and limb development (Ugur and
Tolun, 2008), while WNT10B upregulation is found in breast and bone cancer (Bui et al., 1997; Chen et al., 2008). In terms of hematopoietic development, WNT10B is expressed throughout the hematopoietic system (fetal liver (Austin et al., 1997), bone marrow (Van Den Berg et al., 1998), marrow microenvironment (Congdon et al., 2008), myeloid cells, erythrocytes, and immature B cells) and induces the proliferation of hematopoietic progenitors (Austin et al., 1997) especially during hematopoietic regeneration after injury (Congdon et al., 2008). CD44 is also a target gene and positive regulator of WNT signaling (Schmitt et al., 2014) and CD44 was upregulated (log2 fold change of 1.3) along with the two WNT genes.

WNT signaling has been shown to regulate ribosomobiogenesis through inducing the expression of the transcription factor c-MYC, which regulates the transcription of rRNA, ribosomal protein genes, rRNA processing genes, ribosome assembly genes, and nuclear-cytoplasmic shuttling genes (Chaillou et al., 2014), although c-MYC was not upregulated with the two WNT genes in this study. Interestingly, hiPSCs derived from patients with X-linked dyskeratosis congenita, another ribosomopathy and inherited bone marrow failure syndrome with impaired telomerase and rRNA processing, has shown reduced WNT signaling (Gu et al., 2015). Collectively, this evidence indicates that WNT upregulation in SDS compared to normal hiPSCs might have relevant developmental consequences.

**IFN-γ signaling.** CD44 and IFI30 (Gamma-Interferon-Inducible Lysosomal Thiol Reductase, GILT), were found to be upregulated (log2 fold change of 1.3) and downregulated (log2 fold change of -1.0) respectively in SDS compared to normal hiPSCs and participate in the pathway of IFN (interferon)-γ signaling. During inflammatory response, CD44 regulates the production of the cytokine IFN-γ by T cells (Blass et al., 2001) and natural killer cells (Sague et al., 2004). IFN-γ induces the expression of GILT, which reduces disulfide bonds in disulfide-rich antigens during antigen processing in antigen-presenting cells (Singh and Cresswell, 2010) and reduces disulfide bonds in lysosomes (Arunachalam et al., 2000). As IFN-γ signaling is found in antigen presenting cells and IFN-γ transcripts were not detected in either SDS or normal hiPSCs, GILT upregulation might be involved with increased lysosomal activity (Calamita et al., 2017) rather than IFN-γ signaling in SDS hiPSCs.

**Absence of ribosomal gene dysregulation.** Genes involved with rRNA transcription, pre-rRNA processing, and ribosomal proteins have been shown to be dysregulated in SDS patient bone
marrow mononuclear cells (Rujkijyanont et al., 2009). Intriguingly, despite the above findings and the major role of SBDS in ribosome biogenesis, ribosomal genes were not dysregulated in SDS compared to normal hiPSCs in this study. This leads to the question whether ribosomopathy is found in SDS compared to normal hiPSCs. Ribosome biogenesis-associated genes might have milder levels of dysregulation, and consequently statistical significance might only be reached if more patient and control cell lines were used. Also, SDS and normal hiPSCs were comparable in terms of pluripotent status and spontaneous differentiation, as well as their mesodermal and hemogenic endothelial potentials, suggesting that ribosome biogenesis might not be severely impaired at these early stages. Sufficient ribosome biogenesis could be due to compensatory mechanisms or low requirements for protein synthesis at this developmental stage. The onset of impairment might be found later in early hematopoietic progenitors as myeloid colonies were reduced in CFU assays, which enumerated the number of hematopoietic progenitors.

Other than ribosomal gene dysregulation, other features of SDS ribosomopathy include reduced monosomes (80S) (Menne et al., 2007; Sen et al., 2011) as well as polysomes (Menne et al., 2007) in human and yeast model of SBDS deficiency. Contradictory results have been reported regarding 60S:40S ratios as studies have reported reduced (Menne et al., 2007), increased (Luca, 2015; Calamita et al., 2017), or no differences (Finch et al., 2011; Wong et al., 2011) in 60S:40S ratio. Sucrose density gradient of P357 SDS hiPSCs have shown aberrant ribosome profile including reduced 80S and increased 60S:40S ratio (Luca, 2015). The absence of ribosomal gene dysregulation might be due to differences in ribosomal composition between hiPSCs and marrow mononuclear cells, as the composition of ribosomal proteins on the ribosome has been shown to be specific to cell type and developmental stage (Xue and Barna, 2012; Slavov et al., 2015).

**Limitations.** As shown in both RNA-seq of hiPSC lines and flow cytometry analysis of hiPSC-derived populations, one of the limitations of using hiPSC is inherent variability in both transcription and differentiation. Differentiation variability has been reported in both hiPSC and hESC lines (Osafune et al., 2008; Bock et al., 2011) and other modalities of differentiation such as neural differentiation of hiPSCs (Hu et al., 2010).

In the case of patient variability in which hiPSCs are derived from different donors, differentiation variability has been shown to be mainly due to genetic background (Kajiwara et al., 2012). The role of the specific mutated alleles in the cellular or biochemical phenotype could
be elucidated by reducing genetic-related variability by using a larger cohort, isogenic cell lines with corrected mutations or transgene expression, or genetically related individuals with different phenotypes (Sandoe and Eggan, 2013). Hence, hiPSCs of both SDS patients were transduced with either SBDS-ZsGreen to create rescued lines or ZsGreen to create mock controls using lentivirus in order to minimize both differentiation and transcriptional variability due to genetic background.

In the case of inter-clone variability in which hiPSCs are derived from the same donor, varying epigenetic events during reprogramming contribute to variability as individual hiPSC clones have been shown to have unique differentially methylated DNA regions (Lister et al., 2011). This reprogramming variability could be minimized by screening for high quality, completely reprogrammed iPSC through assessing histone deposition pattern, marker expression, and differentiation efficiency. Histone variant H2A.X deposition pattern distinguishes miPSCs with the potential to develop into viable mice (Wu et al., 2014) through tetraploid complementation, the standard for testing miPSC competence (Zhao et al., 2009). Surface marker CD30 distinguishes undifferentiated hiPSCs from other reprogramming derivatives (Abujarour et al., 2013) and the marker combination Nanoghigh/Sox2high/CD54high distinguishes completely reprogrammed miPSCs from partially reprogrammed miPSCs (Zunder et al., 2015). In this study, the standard hiPSC screening, including gene and protein expression of pluripotency markers and gene expression of germ layer markers upon differentiation, and karyotyping, was sufficient as hiPSC lines derived from the same individual have been shown to be sufficiently similar (Mills et al., 2013). Differentiation efficiency could be assessed (Bock et al., 2011) and standardized protocols could be employed to reproduce differentiation efficiency (Boulting et al., 2011). In addition, inter-clone variability is also contributed by genomic instability as copy number variations acquired during reprogramming and prolonged culture have been shown to affect hematopoietic differentiation potential (Mills et al., 2013). In this study, although high resolution copy number variation arrays were not used, high passage hiPSCs were karyotyped to screen for clonal changes. Moreover, interpreting RNA-seq data require a cautious approach as hiPSCs might express hiPSC-specific transcripts and artefacts and further validation in other model systems is necessary.

In the case of technical variability in which replicates are generated from the same hiPSC line, reagent, instrument, and technique variability contribute to technical variability. Reagent and
technique variability was minimized by paired differentiation, differentiating one SDS hiPSC line with one normal hiPSC line, and instrument variability was minimized by fluorescence-minus-one negative controls. Technical variability could be further minimized by increasing replicates.

On the other, the in vitro differentiation used was limited in modeling in vivo hematopoiesis in which there is a complex microenvironment, gradients of factors, and growth factors (discussions with Dr. Herman Yeger). However, hiPSCs could potentially reproduce the microenvironment using a combination of hiPSC-derived mesenchymal stromal cells, osteoprogenitors, endothelial cells, and other cells of the bone marrow microenvironment to create an in vitro system of the bone marrow for disease modeling and drug screening. Furthermore, a potential in vitro human system could be created by interconnecting various hiPSC-derived tissue types representing major organ system and used as an alternative in vitro physiological model used to establish risks before clinical trials.
Chapter 6
Conclusion

To address whether there was a developmental defect in SDS compared to normal hiPSCs during definitive hematopoietic differentiation, developmental stages were quantified by immunophenotyping. As no significant differences in spontaneous differentiation, pluripotent status, mesoderm induction, and hemogenic endothelial specification were found between SDS and normal hiPSCs, there was no impairment in early definitive hematopoiesis in SDS. However, significant reduction in viability was found in SDS compared to normal embryoid bodies by day 9 where hemogenic and vascular endothelial cells were identified, suggesting that apoptosis, which was reported in more mature mouse and human hematopoietic cells in SDS (Dror et al., 2001; Rukijijyanont et al., 2008; Orelio et al., 2009; Watanabe et al., 2009; Ambekar et al., 2010; Sen et al., 2011; Tulpule et al., 2013; Zambetti et al., 2015), might affect the formation of the hematopoietic system. An impairment was apparent in later hematopoietic stages. SDS hiPSCs showed significantly reduced myeloid potential, specifically of the CFU-GM and BFU-E, in CFU assays compared to normal hiPSCs, consistent with the findings in the CFU assays of human bone marrow hematopoietic progenitors (Suda et al., 1982; Dror and Freedman, 1999; Kuijpers et al., 2005; Sen et al., 2011). Reduced number of colonies indicated either reduced formation of hematopoietic progenitors or reduced capability of formed progenitors to further proliferate and develop to blood colonies. Other non-hematopoietic populations were also assessed. SDS hiPSCs showed significantly increased potential to generate the venous endothelium and associated populations by day 8 compared to normal hiPSCs. Increased venous potential might be a result of reduced hematopoietic potential or might affect hematopoietic development. In summary, the onset of SDS hematopoietic defects occurred in hematopoietic progenitors as indicated by CFU assays. Whether the potential to differentiate to early hematopoietic progenitors is impaired requires further immunophenotypic characterization as discussed in Future Directions.

To address whether there was gene dysregulation in SDS compared to normal hiPSCs, RNA from sorted hiPSCs were subjected to RNA-sequencing and analyzed for differential gene expression. Seventy-nine annotated filtered genes were significantly differentially expressed. HIST1H1A encoding for a linker histone variant was significantly downregulated with the greatest magnitude in both P55 hiPSC lines, both P357 hiPSC lines, and all four SDS hiPSC
lines compared to all four normal hiPSCs, suggesting that nucleosome stability, chromatin remodeling, and transcriptional regulation might be affected. *SBDS* was significantly downregulated in SDS compared to normal hiPSCs, consistent with the findings reported in patient bone marrow mononuclear cells (Rujkijyanont *et al*., 2007) and SDS hiPSCs (Tulpule *et al*., 2013). Alignment of *SBDS* revealed novel transcripts including r.1_128del;128_129ins129-142_129-1 and SDS-associated transcripts r.129_192del and r.129_192del; 251_258del, suggesting potential translation to novel mutant proteins. Over-representation analysis showed that the significantly differentially expressed genes were enriched for several pathways, including glycosaminoglycan metabolism, WNT signaling, and linker and core histones. Interestingly, ribosomal pathways were not enriched in SDS compared to normal hiPSCs, suggesting that ribosomal biogenesis might not be severely impaired in SDS hiPSCs. The hypothesis that there was gene dysregulation in SDS compared to normal hiPSCs was supported by differential expression analysis of hiPSCs, while RNA-seq of further stages including mesoderm, hemogenic endothelium, and early hematopoietic progenitors could provide insight into potentially disrupted pathways during early hematopoietic development.
Chapter 7
Future Directions

Immunophenotyping and RNA-seq. In this study, RNA-seq revealed differential gene expression and potentially disrupted pathways in SDS compared to normal hiPSCs. To determine if these gene dysregulation and pathways could also affect further stages of embryonic hematopoietic formation, the hiPSC-derived mesodermal and hemogenic endothelial populations optimized and isolated in this study could be RNA-sequenced to identify temporal and differential gene expression and potentially disrupted pathways in SDS compared to normal developmental populations.

In this study, early developmental stages (pluripotent stem cells, mesoderm, and hemogenic endothelium) were comparable between SDS and normal hiPSCs and myeloid colony formation was significantly reduced in SDS compared to normal hiPSCs, but hematopoietic stages between hemogenic endothelium and myeloid progenitors were not quantified due to low cell counts. To identify the onset of hematopoietic impairment, hiPSC-derived early hematopoietic progenitors could be quantified by immunophenotyping. Immunophenotyping and isolation of early hematopoietic progenitors could be challenging as cell counts of round, non-adherent early hematopoietic progenitors showed a ten-time reduction in SDS compared to normal hiPSCs. To improve the yield of early hematopoietic progenitors, analysis of differentiation kinetics after day 9+7 could identify the differentiation day with maximum abundance of early hematopoietic progenitors, while the expansion of early hematopoietic progenitors could also be optimized through refining medium formulation. To investigate the gene dysregulation underlying the onset of hematopoietic defects, isolated hiPSC-derived early hematopoietic progenitors could be RNA-sequenced to identify differential gene expression and potentially disrupted pathways relevant to the impaired expansion of early hematopoietic progenitors in SDS compared to normal hiPSCs.

Furthermore, as myeloid maturation is impaired in SDS, to investigate the gene dysregulation underlying myeloid defects, isolated hiPSC-derived myeloid progenitors could be RNA-sequenced to identify differential gene expression and potentially disrupted pathways relevant to impaired myeloid differentiation in SDS compared to normal hiPSCs. Disrupted pathways in early hematopoietic progenitors and myeloid progenitors could provide critical insight into drug targets against hematopoietic defects in SDS and improve patient care and outcome.
Collectively, RNA-seq of hematopoietic developmental stages (hiPSCs, mesoderm, hemogenic endothelium, early hematopoietic progenitors, and myeloid progenitors) could establish the temporal and differential gene expression in SDS during hematopoietic development. This could also reveal the temporal requirement of SBDS during hematopoietic development and identify disrupted pathways that link SBDS deficiency and ribosomal defects to hematopoietic defects.

Validation of RNA-seq could include validating gene and protein expression of relevant differentially expressed genes by real time-qPCR and Western blot. The SBDS-ZsGreen and ZsGreen lentiviral-transduced SDS hiPSC lines established in this study could also be used for gene and protein validation to minimize variability due to genetic background, although SBDS overexpression might lead to increased protein synthesis and generate false positives when comparing with SBDS deficiency. The hematopoietic consequences of these genes could be assessed by CFU assays of SDS hiPSCs with shRNA knockdown or transgene overexpression. Lastly, a recommendation for future RNA-seq may involve generating strand-specific cDNA libraries for RNA-seq to distinguish between sense and antisense RNA (Mills et al., 2013).

**Venous potential.** In this study, SDS hiPSCs showed increased potential to generate the venous endothelium (CD34+/CD43+/CD73+/184+) and associated populations (CD73+, CD184+/CD73+, CD34+/CD184+) compared to normal hiPSCs. We could investigate if venous endothelial potential could affect hematopoietic potential to provide insight into the dysregulation of vascular and hematopoietic commitment in SDS. To validate the venous endothelial identity of these populations, these populations could be assessed for gene and protein expression of NRP2 and EPHB4 venous markers, secretion of angiocrine factors, and vasculature formation and integration upon engraftment onto immune-deficient mice (Sriram et al., 2015; Ditadi et al., 2015).

As repression of arterial genes (Sox17 and Notch1) improved hematopoietic potential during endothelial-to-hematopoietic transition in mice due to the derepression of hematopoietic transcription factors Runx1 and Gata2 (known to be repressed by Sox17; Lizama et al., 2015), commitment to the hematopoietic lineage might require the inhibition of the vascular endothelial lineages. As increased venous potential was found in SDS compared normal hiPSCs in this study, we could investigate if increased venous commitment correlated with reduced hematopoietic commitment. To investigate if the presence of the venous endothelium could
affect the hematopoietic potential in SDS, the output of early hematopoietic progenitors could be quantified by the immunophenotyping of CD34+/CD45+ population generated from co-culturing sorted hiPSC-derived venous and hemogenic endothelium and culturing sorted hiPSC-derived hemogenic endothelium alone. As non-cell autonomous effects of SDS stromal cells on normal hematopoietic progenitors have previously been described (Dror and Freedman, 1999), it would be important to assess the defects of the SDS microenvironment using hiPSC-derived cells.

To determine the genes that could be associated with the increased venous endothelial commitment in SDS, RNA-seq of hiPSC-derived venous endothelium optimized and isolated in this study could identify differentially expressed genes and potentially disrupted pathways in vascular differentiation in SDS compared to normal hiPSCs. To validate if these genes could affect hematopoietic potential in SDS, the output of early hematopoietic progenitors could be quantified by the immunophenotyping of CD34+/CD45+ population generated from rescuing gene expression by transgene overexpression or shRNA silencing in SDS embryoid bodies after mesoderm induction and definitive specification (at day 4 or after).

To determine if increased venous endothelial cells could be found in SDS in vivo, the venous endothelium could be visualized by immunostaining of the aorta-gonad-mesonephros region and relevant hematopoietic systems of endothelial cell-targeted Sbds conditional knockout mice using the venous marker EphB4. Immunostaining using hematopoietic stem cell markers could also permit the visualization of hematopoietic stem cells and determine if increased venous endothelium could affect the hematopoietic stem cell niche.

An alternative to using mice is to recreate the aforementioned bone marrow microenvironment using hiPSCs to study the effects of venous endothelium on the hematopoietic potential.

**Reduced viability.** In this study, reduced viability was found in SDS compared to normal embryoid bodies at day 9, a population containing hemogenic and vascular endothelia. It would be important to investigate whether and how apoptosis could affect the formation of the hematopoietic system or vasculature and identify the onset of hematopoietic defects in SDS. To determine if increased apoptosis could be found in hemogenic, arterial, or venous endothelial population, gene and protein expression of p53 and p53 targets could be assessed (Zambetti et al., 2015) and a caspase activity assay could be used to detect apoptosis in isolated populations.
To determine if increased apoptosis could be found \textit{in vivo}, TUNEL assay of the aorta-gonad-mesonephros region and relevant hematopoietic systems of endothelial cell-targeted \textit{Sbds} conditional knockout mice could be used to visualize apoptotic cells in hematopoietic tissues.

\textbf{Ribosomal defect.} Although P357 hiPSCs showed aberrant polysome profile (Luca, 2015), the polysome profile of P55 hiPSCs was not assessed in this study and details of defective ribosome profiles need to be understood. To determine if a ribosomal defect could also be found in P55 hiPSCs, the polysome profile could be assessed to confirm reduced 80S and aberrant polysome profiles characteristic of SDS. In this study, dysregulation in ribosome-associated genes was absent in SDS compared to normal hiPSCs at the RNA level. To determine if ribosome-associated proteins could show aberrant expression in SDS hiPSCs, monosomes, polysomes, and their associated proteins could be purified to determine the ribosome composition in SDS compared to non-SDS hiPSCs, providing insight into how SBDS deficiency and reduced global protein production could affect ribosome composition and translational control.

\textbf{Histones.} In this study, the linker histone variant \textit{HIST1H1A} (encodes for H1.1) was downregulated and the core histone variant \textit{HIST1H2BK} (encodes for H2BK) was upregulated in SDS compared to normal hiPSCs. If their gene and protein expression could be validated and their overexpression or knockdown could rescue the SDS hematopoietic defect in CFU assays, it would be important to investigate how H1.1 depletion and H2BK deposition could affect transcriptional regulation in SDS hiPSCs and understand the epigenetic mechanisms that could result in global changes in gene expression in SDS. Although histone variants are considered to be redundant, the unique deposition pattern of H1.1 compared to other H1 variants might result in variant-specific gene regulation (Izzo et al., 2013). To ask whether changes in variant expression could be associated with changes in variant deposition pattern in SDS, ChIP-seq using variant-specific antibodies could identify the genomic location of variant-specific deposition in SDS compared to non-SDS hiPSCs. In addition, cell cycle-specific deposition of these variants could also be identified by arresting cells at certain phases of the cell cycle.

As raising variant-specific antibodies could be challenging, alternatively, C-terminal tag sequences could be inserted into the endogenous variant genes through CRISPR, although high sequence homology among variants could also render gene editing challenging. Alternatively, expressing exogenous fusion proteins at extremely low levels could be used to study H2BK
upregulation, but introducing exogenous histones could disrupt histone levels due to compensatory effects. Regardless of the method of variant isolation, ChIP-seq could identify SDS-associated changes in histone variant deposition.

Differential variant deposition could be interpreted as either transcriptional repression through inhibiting promoter accessibility or transcriptional activation through recruiting transcriptional regulators, histone modifiers, and chromatin remodeling complexes. To determine if differential variant deposition was associated with differential gene expression in SDS, both ChIP-seq data and RNA-seq data found in this study could be correlated to identify variant-specific differential gene expression in SDS compared to non-SDS hiPSCs.

**Glycosaminoglycan metabolism.** In this study, genes involved in glycosaminoglycan metabolism, *CHST11, GNS,* and *CD44*, were upregulated. If their gene and protein expression could be validated and their knockdown could rescue the SDS hematopoietic defect in CFU assays, it would be important to investigate how glycosaminoglycan metabolism could affect hematopoietic development in SDS and to understand the sulfation patterns of glycosaminoglycans that could result in developmental regulation. Glycosaminoglycans (heparin, keratin, chondroitin, hyaluronan) act as signaling molecules that facilitate ligand receptor interactions and regulate embryonic (Oki et al., 2007), hematopoietic (Holley et al., 2011), bone (Baud’huin et al., 2011), and neuronal (Pickford et al., 2011) development.

The activity and specificity of glycosaminoglycans are regulated by specific sulfation patterns (Gama et al., 2006). The three upregulated genes of interest participate in glycosaminoglycan metabolism through either regulating the sulfation patterns of glycosaminoglycans or targeting glycosaminoglycans to the lysosome for degradation. CHST11 adds sulfates to chondroitin, GNS removes sulfates from heparan and keratan sulfate, and CD44 targets hyaluronan and chondroitin to the lysosome. Therefore, upregulation of these three genes could lead to increased chondroitin sulfate and reduced hyaluronan, chondroitin, heparan and keratan sulfate.

To determine if glycosaminoglycan imbalance could be found in SDS compared to non-SDS hiPSCs, non-sulfated glycosaminoglycans could be quantified by toluidine staining and sulfated glycosaminoglycans could be quantified by immuno-detection following electrophoresis and capillary blotting (Volpi and Maccari, 2015). To investigate the molecular interactions affected
by the dysregulation of sulfated glycosaminoglycans, targeted candidates of sulfated glycosaminoglycan binding protein could be identified by literature search or sulfated glycosaminoglycan microarray to identify candidates. One known interaction occurs between heparan sulfate and TGF-β cytokines (Coome, 2008), thus reduced heparan sulfate could result in reduced TGF-β signaling. To investigate if TGF-β signaling could be reduced in SDS compared to non-SDS hiPSCs, SMAD2/3 luciferase assay could be used to quantify TGF-β signaling (Lohcharoenkal et al., 2014).

**WNT signaling.** With regards to explaining the enrichment in WNT signaling in this study, a WNT10B ligand and FZD2 receptor were upregulated in SDS compared to normal hiPSCs. FZD2 has been shown to be essential in neuronal development (Bhat et al., 1998; Rodriguez et al., 2005) and cardiac organization (Toyofuku et al., 2000), while WNT10B has been shown to be essential in hematopoietic development (Lento et al., 2013), bone formation (Bennet et al., 2005), and brain development (Lekven et al., 2003). Therefore, WNT upregulation could result in dramatic consequences in development in SDS. If the gene and protein expression of the WNT genes could be validated and their knockdown could rescue the SDS hematopoietic defect in CFU assays, we could investigate how increased WNT signaling could result in hematopoietic changes in SDS. To determine if WNT signaling is increased in SDS compared to non-SDS hiPSCs, WNT signaling could be quantified by TCF/LEF luciferase reporter assay. As SDS hiPSCs has shown reduced proliferation (Ruiz-Gutierrez et al., 2016), to determine if increased WNT signaling could affect SDS hiPSC proliferation, proliferation could be assessed by carboxyfluorescein succinimidyl ester proliferation assay with a WNT inhibitor. To determine the transcriptional consequences of increased WNT signaling, literature search could identify experimentally confirmed WNT targets within the significantly upregulated genes found in this study, such as CD44. Other upregulated genes with TCF and LEF binding sequences in their promoters could also be potential WNT targets. If the gene and protein expression of WNT targets could be validated, we could investigate if and how these WNT targets affect hematopoietic development by shRNA silencing and CFU assays. Alternatively, CRISPR approaches could be used to tackle several other questions (discussions with Dr. Herman Yeger).

**Potentially novel transcripts.** In this study, potentially novel SBDS transcripts were only found in one of the two isogenic lines. In particular, SBDS r.129_192del was found in one of the two isogenic hiPSC lines of each patient but not in normal hiPSC lines. Aberrant transcripts were
also found in non-SDS and from pseudogene. We could investigate how SDS-associated SBDS splice variants could impair pre-60S maturation. To determine if these transcripts could be artifacts of RNA-seq or hiPSC cell line generation, additional transcript validation such as real time-qPCR, droplet digital PCR, or Sanger sequencing of cloned cDNAs could be used to confirm transcript expression and sequenced RNA findings. To determine if potentially novel SBDS transcripts could translate predicted proteins, overexpression of cDNA with fluorophore tag could be used to visualize cellular localization and co-sedimentation with ribosomal fractions. To determine if these predicted proteins, as well as p.Lys62* and p.Cys84Tryfs*4 mutant SBDS, could be functional, eIF6 release assay could be used to quantify residual activity (Finch et al., 2011). To determine the defects of these mutant proteins in pre-60S maturation, interactions of the mutant proteins with pre-60S could be identified by X-ray crystallography to provide insight into the binding sites and drug targets that could promote eIF6 release.

**Therapeutic strategies.** The current treatment for hematopoietic issues in SDS is limited. Neutropenia can be controlled by granulocyte colony stimulating factor with variable effectiveness (Dror et al., 2011) and severe hematological dysfunction requires hematopoietic stem cell transplantation with survival rates of 60-65% (Donadieu et al., 2005; Cesaro et al., 2005). In culture models, various molecules have been shown to improve SDS defects. Antioxidant treatment has been shown to reduce reactive oxygen species and improve cell growth (Ambekar et al., 2010; Sen et al., 2011). Pan-caspase inhibitor has been shown to inhibit apoptosis and improve cell survival of hematopoietic cells (Tulpule et al., 2013). Broad-spectrum protease inhibitor has been shown to reduce protease auto-digestion and hematopoietic cell death (Tulpule et al., 2013). Leucine treatment has been shown to promote protein synthesis, rescue mitochondrial defect, and increase erythroid potential (Ravera et al., 2016). The efficacy in improving myeloid function of these molecules and other cytoplasmic molecules that could target eIF6 release could be assessed by CFU assays of SDS hiPSC-derived early hematopoietic progenitors. In addition to being a resource for drug screening, SDS hiPSC-derived early hematopoietic progenitors could also provide a foundation to study gene therapy, particularly towards transplantation of the patient-derived hematopoietic stem cells with gene correction of SBDS mutations or SBDS insertion into genomic safe harbors.
Bibliography


Ambekar, Chhaya, et al. "SBDS-deficiency results in deregulation of reactive oxygen species leading to increased cell death and decreased cell growth." *Pediatric Blood & Cancer* 55.6 (2010): 1138-44.


Bhat, Krishna Moorthi. "Frizzled and Frizzled 2 play a partially redundant role in Wingless signaling and have similar requirements to Wingless in neurogenesis." Cell 95.7 (1998): 1027-36.


Bhatla, D, et al. "Reduced-intensity conditioning is effective and safe for transplantation of patients with Shwachman-Diamond syndrome." Bone Marrow Transplantation 42.3 (2008): 159-65.


—. "Transplantation of vascular endothelial cells mediates the hematopoietic recovery and survival of lethally irradiated mice." Blood 109.6 (2007): 2365-72.


Coombe, Deirdre R. "Biological implications of glycosaminoglycan interactions with haemopoietic cytokines." Immunology and Cell Biology 86.7 (2008): 598-607.

—. "Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development." Nature Immunology 11.7 (2010): 585-93.


Elghetany, M Tarek and Blanche P Alter. "p53 protein overexpression in bone marrow biopsies of patients with Shwachman-Diamond syndrome has a prevalence similar to that of patients with refractory anemia." *Archives of Pathology & Laboratory Medicine* 126.4 (2002): 452-5.


Lane, Timothy F and Philip Leder. "Wnt-10b directs hypermorphic development and transformation in mammary glands of male and female mice." *Oncogene* 15.18 (1997): 2133-44.


Maximow, A. "The lymphocyte as a stem cell common to different blood elements in embryonic development and during the post-fetal life of mammals (1909)." Originally in German: Folia Haematologica 8.1909, 125-34. English translation: *Cellular Therapy and Transplantation* 1.3 (2009). e.000032.01


Medvinsky, A and E Dzierzak. "Definitive hematopoiesis is autonomously initiated by the AGM region." *Cell* 86.6 (1996): 897-906.


—. "Leukaemia-related gene expression in bone marrow cells from patients with the preleukaemic disorder Shwachman-Diamond syndrome." *British Journal of Haematology* 137.6 (2007): 537-44.


Sabin FR. "Studies on the origin of blood vessels and of red corpuscles as seen in the living blastoderm of the chick during the second day of incubation." *Contributions to Embryology* 9 (1920): 213-62.


Appendices

1 Pluripotency of hiPSCs

Pluripotency of hiPSCs was confirmed by gene and protein expression of pluripotency markers and gene expression of germ layer markers following differentiation. hiPSCs expressed OCT4 and NANOG transcripts in levels above 80% as assessed by qRT-PCR with no significant differences between SDS and normal hiPSCs (Supplementary Figure 1). However, OCT4 expression was significantly higher in P55 (n = 2) compared to normal hiPSCs. hiPSCs expressed pluripotency nuclear proteins OCT4 and NANOG and surface markers SSEA-4, TRA-1-81, and TRA-1-60 as assessed by flow cytometry or ICC Supplementary Figure 4 - Supplementary Figure 3). Upon differentiation, hiPSC-derived EBs showed increased expression of endodermal AFP, mesodermal HAND1, and ectodermal NEUROD1 transcripts with no significant differences between SDS and normal hiPSCs (Supplementary Figure 5). However, HAND1 expression was significantly higher in P55 compared to normal hiPSCs.

Supplementary Figure 1. Gene expression of pluripotency markers in hiPSCs.

RNA expression levels of OCT4 and NANOG was above 80% in all hiPSCs (n = 1 per line) as assessed by qRT-PCR by CCRM. Unpaired two-tailed t-tests were analyzed between normal hiPSCs and SDS, P55, and P357 hiPSCs with p-values shown at the top.
Supplementary Figure 2. Protein expression of pluripotency markers in hiPSCs.
Protein expression of nuclear proteins OCT4 (green) and NANOG (green), as well as surface markers SSEA-4 (red), TRA-1-81 (red), and TRA-1-60 (red) in hiPSCs was assessed by ICC with DAPI nuclear counter stain (blue) by CCRM. hiPSCs were reprogrammed using Sendai virus derived from bone marrow stromal cells (iSV.BM). The scale bars represent 200µm.
Supplementary Figure 3. Protein expression of pluripotency markers in P55 hiPSCs.
Protein expression of nuclear proteins OCT4 and NANOG (green), as well as surface markers TRA-1-60 and SSEA-4 (red) of SDS hiPSCs P55C and P55F was assessed by ICC with DAPI nuclear counter stain (blue) by CCRM.
Supplementary Figure 4. Protein expression of pluripotency markers in P55 hiPSCs. Protein expression of surface markers SSEA-4 and TRA-1-60 and nuclear protein OCT4 of SDS P55C and P55F hiPSCs was assessed by flow cytometry with unstained control by CCRM.
Supplementary Figure 5. Gene expression of germ layer markers in hiPSC-derived EBs.
Upon differentiation, hiPSC-derived EBs (n = 1 per line) showed increased expression of endodermal AFP, mesodermal HAND1, and ectodermal NEUROD1 transcripts compared to undifferentiated hiPSCs as assessed by qRT-PCR by CCRM. Unpaired two-tailed t-tests were analyzed between normal hiPSCs and SDS, P55, and P357 hiPSCs with p-values shown at the top. y-axis is in log_{10} scale.

2 Genetic Identity of hiPSCs

The fidelity of hiPSCs to their parental identity was confirmed by profiling the Amelogenin gene and 9 short tandem repeats by PCR and gel electrophoresis and karyotyping. hiPSCs showed band sizes consistent with their parental bone marrow stromal cells (Supplementary Figure 6, P55 not shown). hiPSCs showed normal karyotype consistent with parental cells (Supplementary Figure 7). However, P55C displayed 1 nonclonal finding in 40 cells counted, 44, XX, del(2)(p15),-4,-19, in its first karyotyping. Further karyotyping showed 2 findings with missing chromosome(s) in 20 cells counted, 44, XX, -4, -7 and 45, XX, -20. Further karyotyping showed 2 findings with missing chromosome, 45, XX, -8 and 45, XX, -16. As each abnormal finding was not found in subsequent karyotyping, the P55C karyotype was considered to be normal.
Supplementary Figure 6. Genetic identity of hiPSCs compared to parental cells.
Genetic fidelity of hiPSCs (bottom 2 rows) to their respective parental bone marrow stromal cells (top row) was confirmed by PCR amplification of Amelogenin and 9 short tandem repeats and resolved by gel electrophoresis (band size in base pairs) by CCRM.
Supplementary Figure 7. Karyotyping of hiPSCs.
The chromosome modal number, gender, and rearrangements of hiPSCs were identified by G-band karyotyping by CCRM.
3 Flow Cytometry Analyses

In this section, compensation beads and unstained controls were used to establish fluorescence spectral overlap compensation while the fluorescence-minus-one controls were used to establish gates by identifying the negative population (Methods Section 4).

3.1 Human Induced Pluripotent Stem Cells

Supplementary Figure 8. Compensation and unstained controls of hiPSC markers.
Compensation beads stained with each of APC-CD56, BV421-CD184, FITC-TRA-1-60, PerCP-Cy5.5-SSEA-4, and PE-CD30 and unstained hiPSCs were analyzed to establish compensation visualized by Flowjo V10.
Supplementary Figure 9. Fluorescence-minus-one negative controls of hiPSC markers.
hiPSCs stained with the combination of markers identifying the hiPSC population but excluding each marker, APC-CD56, BV421-CD184, FITC-TRA-1-60, PerCP-Cy5.5-SSEA-4, and PE-CD30, were analyzed to establish gates by identifying the negative population visualized by Flowjo V10.
3.2 Mesodermal Cells

Supplementary Figure 10. Compensation and PI only controls of initial mesodermal markers.
Compensation beads stained with each of FITC-TRA-1-60, PE-Cy7-CD34, BV421-C-KIT, APC-CD56, and PE-KDR and T4 cells stained with PI only were analyzed to establish compensation visualized by Flowjo V10.
Supplementary Figure 11. Fluorescence-minus-one negative controls of initial mesodermal markers.
T4 cells stained with the initial combination of markers identifying the mesodermal population but excluding each marker, FITC-TRA-1-60, PE-Cy7-CD34, BV421-C-KIT, APC-CD56, and PE-KDR, were analyzed to establish gates by identifying the negative population visualized by Flowjo V10.
**Pluripotency markers.** Since the aforementioned strategy in isolating mesodermal population resulted in low yield, new strategies were employed to increase the yield of KDR+ population. The pluripotent marker used in the negative selection of mesodermal population was assessed. Since both TRA-1-60 and SSEA-3 are pluripotent markers with the most rapid loss of expression upon mesodermal induction (Ramirez et al., 2011), SSEA-3 was analyzed in place of TRA-1-60 to determine if the mesodermal population could be increased. The percentages of mesodermal population negatively selected using SSEA-3 (FITC-SSEA-3/PE-Cy7-CD34/BV421-C-KIT+/APC-CD56+/PE-KDR+) were less than those negatively selected using TRA-1-60 (FITC-TRA-1-60/PE-Cy7-CD34/BV421-C-KIT+/APC-CD56+/PE-KDR+) in both SDS and normal T3 and T4 EBs (Supplementary Figure 12 - Supplementary Figure 14). Therefore, TRA-1-60 had been a suitable marker for mesodermal isolation.

**Differentiation Day.** The kinetics of mesodermal induction was assessed. As individual hiPSC lines might progress through differentiation at different rates, the time point that yielded maximum KDR+ population could vary among hiPSC lines. To determine if the low KDR+ yield was due to the variation in kinetics, mesodermal population was analyzed at T3 instead of T4. At this differentiation day, KDR+ yield (96-98%) and SSEA-3/CD34/C-KIT+/CD235a/KDR+ mesodermal yield (14-20%) was comparable to that at T4 (Supplementary Figure 15 - Supplementary Figure 17). Therefore, T4 had been a suitable time point for mesodermal isolation.

**Mesodermal marker.** The mesodermal marker used in the positive selection of mesodermal population was reconsidered. CD56 was replaced with CD235a to eliminate the mesodermal population involved in primitive hematopoiesis (Sturgeon et al., 2014). Hence, mesodermal populations were later sorted using FITC-TRA-1-60/PE-Cy7-CD34/BV421-C-KIT+/APC-CD235a/PE-KDR+.

**Trypsin removal.** The method of digesting EBs into single cells was assessed. As KDR is trypsin-sensitive, small amounts of trypsin activity readily cleaves KDR, leading to low KDR detection by flow cytometry. To determine if the low KDR+ yield was due to insufficient trypsin elimination, five total washes were performed after trypsin digestion. Using this method, KDR+ yield (92-94%) and TRA-1-60/CD34/C-KIT+/CD235a/KDR+ mesodermal yield (28-30%) were improved (Supplementary Figure 18- Supplementary Figure 20).
Supplementary Figure 12. Flow cytometry analyses of T4 mesodermal populations with SSEA-3 instead of TRA-1-60. Normal N551K and SDS P357A hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 4 days, and sorted for mesodermal population using FITC-SSEA-3/PE-Cy7-CD34/BV421-C-KIT/APC-CD56+/PE-KDR+ by FACS visualized by Flowjo V10.
Supplementary Figure 13. Compensation and PI only controls of initial mesodermal markers with SSEA-3 instead of TRA-1-60. Compensation beads stained with each of PE-Cy7-CD34 and BV421-C-KIT, T4 cells stained with each of FITC-SSEA-3, APC-CD56, and PE-KDR, and T4 cells stained with PI only were analyzed to establish compensation visualized by Flowjo V10.
Supplementary Figure 14. Fluorescence-minus-one negative controls of initial mesodermal markers with SSEA-3 instead of TRA-1-60.

T4 cells stained with the initial combination of markers identifying the mesodermal population but excluding each marker, FITC-SSEA-3, PE-Cy7-CD34, BV421-C-KIT, APC-CD56, and PE-KDR, were analyzed to establish gates by identifying the negative population visualized by Flowjo V10.
Supplementary Figure 15. Flow cytometry analyses of T3 mesodermal populations with thorough trypsin removal.
Normal N530J and SDS P357A hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 3 days, and analyzed for mesodermal population using FITC-SSEA-3/PE-Cy7-CD34/BV421-C-KIT/APC-CD56/PE-KDR by analytical flow cytometer visualized by Flowjo V10.
Supplementary Figure 16. Compensation and PI only controls of initial mesodermal markers at T3 with thorough trypsin removal.

Compensation beads stained with each of PE-Cy7-CD34 and BV421-C-KIT, T3 cells stained with each of FITC-SSEA-3, APC-CD56, and PE-KDR, and T3 cells stained with PI only along with prior thorough trypsin removal were analyzed to establish compensation visualized by Flowjo V10.
Supplementary Figure 17. Fluorescence-minus-one negative controls of revised mesodermal markers at T3 with thorough trypsin removal.
T3 cells stained with the initial combination of markers identifying the mesodermal population but excluding each marker, FITC-SSEA-3, PE-Cy7-CD34, BV421-C-KIT, APC-CD56, and PE-KDR, were analyzed to establish gates by identifying the negative population visualized by Flowjo V10.
Supplementary Figure 18. Flow cytometry analyses of T4 mesodermal populations with revised mesodermal markers and thorough trypsin removal.

Normal N551I and SDS P357D hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 4 days, and analyzed for mesodermal population using FITC-TRA-1-60/PE-Cy7-CD34/BV421-C-KIT*/APC-CD235a*/PE-KDR* by analytical flow cytometer visualized by Flowjo V10.
Supplementary Figure 19. Compensation and PI only controls of revised mesodermal markers with thorough trypsin removal. Compensation beads stained with each of PE-Cy7-CD34, APC-CD235a, and BV421-C-KIT, T4 cells stained with each of FITC-TRA-1-60 and PE-KDR, and T4 cells stained with PI only along with prior thorough trypsin removal were analyzed to establish compensation visualized by Flowjo V10.
Supplementary Figure 20. Fluorescence-minus-one negative controls of revised mesodermal markers with thorough trypsin removal.

T4 cells stained with the revised combination of markers identifying the mesodermal population but excluding each marker, FITC-TRA-1-60, PE-Cy7-CD34, BV421-C-KIT, APC-CD235a, and PE-KDR, were analyzed to establish gates by identifying the negative population visualized by Flowjo V10.
3.3 Optimizations in Isolating Hemogenic Endothelium

**Differentiation day.** To determine the kinetics of the hemogenic endothelial population, EBs at T5 to T8 were analyzed using PE-Cy7-CD34+/PE-CD43+/BV421-CD184+/APC-CD73- (Supplementary Figure 24 – Supplementary Figure 25) and cells on OP9-DL1 co-culture at T9+3 were analyzed using PE-Cy7-CD34+/PE-CD43+/BV421-CD73- (Supplementary Figure 26 – Supplementary Figure 27) by analytical flow cytometry. As T9+3 cells of N530E were 100% positive for CD73, T9+3 was not a suitable day for hemogenic endothelium collection. The hemogenic endothelial population first appeared in T7.5 EBs and peaked in T7.75 EBs in both N551K and P357D. Throughout T7 to T8 time points, the yield of HE was low, 5% or less. Moreover, there was no distinct CD34+/CD43-/CD184-/CD73- hemogenic endothelial population clearly separated from CD34+/CD43-/CD184-/CD73+ venous endothelial population. As the yields of hemogenic endothelium were similar in both T7.75 and T8 EBs, T8 was chosen as the differentiation day to collect hemogenic endothelium.

**CD56 expression.** CD56 has been shown to be expressed in mesodermal progenitors (Evseenko et al., 2010) and was expressed at T4 in this study. CD56 was initially considered as a marker in the negative selection for hemogenic endothelium if CD56 expression were decreased from T4 to T8. Analyzing APC-CD56 expression by analytical flow cytometry, 94.8% live cells expressed CD56 in N551K and 85.3% in P357D (Supplementary Figure 28). Hence CD56 was excluded in the marker combination for hemogenic endothelium.
Supplementary Figure 21. Compensation and PI only controls of hemogenic endothelial markers without collagenase I digestion. Compensation beads stained with each of PE-Cy7-CD34, FITC-CD43, PE-VE-cad, BV421-CD184, and APC-CD73 and T8 cells stained with PI only without prior collagenase I digestion were analyzed to establish compensation shown here by Flowjo V10.
Supplementary Figure 22. Fluorescence-minus-one negative controls of hemogenic endothelial markers without collagenase I digestion.

T8 cells stained with the combination of markers identifying the hemogenic endothelial population but excluding each marker, PE-Cy7-CD34, FITC-CD43, PE-VE-cad, BV421-CD184, and APC-CD73, were analyzed to establish gates by identifying the negative population visualized by Flowjo V10.
Supplementary Figure 23. Flow cytometry analyses of T8 hemogenic endothelial populations without collagenase I digestion.
Normal N551K and SDS P357D hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 8 days, and analyzed for hemogenic endothelial population using PE-Cy7-CD34/FITC-CD43/PE-VE-cad/BV421-CD184/APC-CD73 by FACS visualized by Flowjo V10.
Supplementary Figure 24. Flow cytometry analyses of hemogenic endothelial populations without collagenase I digestion from T5-8.

Normal N551K and SDS P357D hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 5-8 days, and analyzed for hemogenic endothelial population using PE-Cy7-CD34+/PE-CD43+/BV421-CD184+/APC-CD73 by analytical flow cytometer. Compensation beads, samples, and gating strategies were analyzed by Flowjo V10.
Supplementary Figure 25. Analyses of hemogenic endothelial populations without collagenase I digestion from T5-8.
Normal N551K and SDS P357D hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 5-8 days, and analyzed for hemogenic endothelial population using PE-Cy7-CD34+/PE-CD43/BV421-CD184+/APC-CD73 by analytical flow cytometer. Percentages of CD184+/CD73- population within CD34+/CD43- and that of CD34+/CD43- population in N551K and P357D from T5 to T8 are displayed (legend at the bottom).
Supplementary Figure 26. Compensation, PI only, and fluorescence-minus-one negative controls of hemogenic endothelial markers at T9+3.

Compensation beads stained with each of PE-Cy7-CD34, FITC-CD43, PE-VE-cad, and BV421-CD73 were analyzed to establish compensation, while T9+3 cells stained with the combination of markers identifying the hemogenic endothelial population but excluding each marker, PE-Cy7-CD34, FITC-CD43, PE-VE-cad, and BV421-CD73, were analyzed to establish gates by identifying the negative population visualized by Flowjo V10.
Supplementary Figure 27. Flow cytometry analyses of hemogenic endothelial populations at T9+3.
Normal N530E hiPSC were feeder depleted for 2 days, subjected to hematopoietic differentiation for 9 days, sorted for CD34+/CD43−, cultured on OP9-DL1 for 3 days, and analyzed for hemogenic endothelial population using PE-Cy7-CD34+/PE-CD43−/BV421-CD73 by analytical flow cytometer as visualized by Flowjo V10.
Supplementary Figure 28. Analyses of CD56 expression T8 hemogenic endothelial populations with collagenase I digestion.

Normal N551K and SDS P357D hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 8 days, and analyzed for early mesodermal population using APC-CD56 by FACS. Compensation beads, unstained control, FMO controls, SDS and normal samples were analyzed by Flowjo V10.
3.4 Hemogenic Endothelial Cells

Supplementary Figure 29. Compensation and PI only controls of hemogenic endothelial markers with collagenase I digestion.
Compensation beads stained with each of PE-Cy7-CD34, PE-CD43, FITC-VE-cad, BV421-CD184, and APC-CD73 and T8 cells stained with PI only along with prior collagenase I digestion were analyzed to establish compensation visualized by Flowjo V10.
Supplementary Figure 30. Fluorescence-minus-one negative controls of hemogenic endothelial markers with collagenase I digestion.

T8 cells stained with the combination of markers identifying the hemogenic endothelial population but excluding each marker, PE-Cy7-CD34, PE-CD43, FITC-VE-cad, BV421-CD184, and APC-CD73, were analyzed to establish gates by identifying the negative population visualized by Flowjo V10.
3.5 CD34⁺/CD43⁻ Population

Supplementary Figure 31. Compensation, PI only, and fluorescence-minus-one negative controls of CD34⁺/CD43⁻ hemogenic and vascular endothelial population in T9 EBs. T9 cells stained with each of PE-Cy7-CD34, PE-CD43, and PI only were used to establish compensation, while T9 cells stained with the combination of markers identifying the hemogenic and vascular endothelial population but excluding each marker, PE-Cy7-CD34 and PE-CD43, were used to establish gates by identifying the negative population visualized by Flowjo V10.
3.6 Early Hematopoietic Progenitors

In preparation for future investigations of later hematopoietic stages, the population of early hematopoietic progenitors were identified by flow cytometry. hiPSCs were feeder depleted for 2 days, subjected to definitive hematopoietic differentiation in the form of EBs for 9 days, sorted for CD34⁺/CD43⁻ population, co-cultured on OP9-DL1 for 7 days, and sorted for early hematopoietic progenitors using FITC-CD11b/APC-CD41a/CD235a'/PE-CD33'/V450-CD45⁻/PE-Cy7-CD34⁺ by FACS at T9+7 (Supplementary Figure 32 - Supplementary Figure 34). These early hematopoietic progenitors at T9+7 do not express CD45 that marks the endothelial to hematopoietic transition and T cell potential (Sturgeon et al., 2014), but emergence of non-adherent round cells was observed in this analysis and other T9+7 cultures used in seeding CFU assays. Cells were found to be auto-fluorescent in FITC channel (Supplementary Figure 32, PI only) and OP9-DL1 was positive in both CD11b and CD41a/CD235a (Supplementary Figure 32, OP9-DL1). As harvested cells comprised of both early hematopoietic progenitors and OP9-DL1, the initial gate (Supplementary Figure 32, SSC-H and FSC-H) showed 12.1% viable cells in P55F and 14.3% in N530J, with substantial OP9-DL1 population. However, the percentages of viable cells that were CD11b'/CD41a'/CD235a'/CD33'/CD45'/CD34⁺ early hematopoietic progenitors were high, 78.4% in P55F and 87.8% in N530J. Despite high early hematopoietic progenitors yield in terms of viable cells, the limiting factor was low early hematopoietic progenitor density in OP9-DL1 co-culture. Unfortunately, further investigation in early hematopoietic progenitors is beyond the scope of my MSc timeline.
Supplementary Figure 32. Flow cytometry analyses of early hematopoietic progenitors at T9+7.

Normal N530J and SDS P55F hiPSCs were feeder depleted for 2 days, subjected to hematopoietic differentiation for 9 days, sorted for CD34+/CD43- population, co-cultured on OP9-DL1 for 7 days, and sorted for early hematopoietic progenitors using FITC-CD11b/APC-CD41a/CD235a/PE-CD33/V450-CD45/PE-Cy7-CD34+ by FACS visualized by Flowjo V10.
Supplementary Figure 33. Compensation and PI only controls of early hematopoietic progenitor markers.
Compensation beads stained with each of FITC-CD11b, APC-CD41a, CD235a, PE-CD33, V450-CD45, and PE-Cy7-CD34 and T9+7 cells stained with PI only were analyzed to establish compensation visualized by Flowjo V10.
Supplementary Figure 34. Fluorescence-minus-one negative controls of early hematopoietic progenitor markers.

T9+7 cells stained with the combination of markers identifying the hematopoietic progenitor population but excluding each marker, FITC-CD11b, APC-CD41a, CD235a, PE-CD33, V450-CD45, and PE-Cy7-CD34, were analyzed to establish gates by identifying the negative population visualized by Flowjo V10.
4 RNA Sequencing

4.1 Quality

Prior to read alignment, the quality of trimmed reads were assessed by FastQC v1.0.0.

**Per base sequence quality and per sequence quality score.** Phred score (Q) is a quality score that measures the reliability of assigning nucleobases to chromatogram peaks. Q greater than 30 represents very good quality, Q between 20 and 30 represents reasonably good quality, and Q lower than 20 represents poor quality. Per base sequence quality shows the Phred score for reads at each position (Supplementary Figure 35). Poorer quality at positions closer to the 3’ end of reads (bases 110-125) is common. Per sequence quality score shows the mean Phred score per read (Supplementary Figure 36). The expected distribution peaks at a Q of 36-38. hiPSC samples passed both the per base sequence quality and per sequence quality score.

**Per base sequence content and per sequence GC content.** Per base sequence content shows the nucleobase content across all bases (Supplementary Figure 37). The expected content of each nucleobase is 25%. Large deviations at positions closer to the 5’ end of reads (bases 1-15) is common due to random hexamer priming. hiPSC samples did not pass the per base sequence content due to high A content of polyA tails despite prior trimming. Additional trimming was not necessary as downstream alignment by TopHat 2 and Bowtie 2 further processed polyA tails. Similar to per base sequence content, per sequence GC content measures the randomness of the cDNA library. Per sequence GC content shows the distribution of mean GC content over all sequences (Supplementary Figure 38). The expected distribution of GC content is to peak at 50%. hiPSC samples passed the per sequence GC content.

**Sequence duplication levels.** Sequence duplication levels show the relative percentage of duplicated sequences (Supplementary Figure 39). Although the expected duplication level is under 50%, high duplication levels are common in RNA-seq due to low diversity in the cDNA library or over-sequencing. hiPSC samples failed with sequence duplication levels over 50% due to polyA tails in P357A and polyN (unassigned nucleobase) sequences in other hiPSC lines, but no bias in specific sequences was detected.
Supplementary Figure 35. RNA-seq per base sequence quality of hiPSC.
RNA-seq reads of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ hiPSCs were trimmed by FASTQ Toolkit v.1.0 and assessed for per base sequence quality by FastQC v1.0.0. Per base sequence quality shows Phred scores (y-axis) for all reads at each position (x-axis). Q greater than 30 represents very good quality (green area), between 20 and 30 represents reasonably good quality (yellow area), and lower than 20 represents poor quality (red area).
**Supplementary Figure 36. RNA-seq per sequence quality score of hiPSC.**

RNA-seq reads of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60*/PerCP-Cy5.5-SSEA-4*/PE-CD30+ hiPSCs were trimmed by FASTQ Toolkit v.1.0 and assessed for per sequence quality by FastQC v1.0.0. Per sequence quality score shows the mean Phred score (x-axis) per read (y-axis). Q greater than 30 represents very good quality, between 20 and 30 represents reasonably good quality, and lower than 20 represents poor quality.
Supplementary Figure 37. RNA-seq per base sequence content of hiPSC.
RNA-seq reads of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ hiPSCs were trimmed by FASTQ Toolkit v.1.0 and assessed for per bp sequence content by FastQC v1.0.0. Per base sequence content shows the percentage of each nucleobase (y-axis), T, C, A, and G (legend at top right corner), at each position along the 125bp read (x-axis).
Supplementary Figure 38. RNA-seq per sequence GC content of hiPSC.
RNA-seq reads of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ hiPSCs were trimmed by FASTQ Toolkit v.1.0 and assessed for per sequence GC content by FastQC v1.0.0. Per sequence GC content shows the number of reads (y-axis) with a particular mean percentage of GC (x-axis). The actual distribution is shown with the theoretical distribution (legend at top right corner).
Supplementary Figure 39. RNA-seq sequence duplication levels of hiPSC.
RNA-seq reads of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ hiPSCs were trimmed by FASTQ Toolkit v.1.0 and assessed for sequence duplication level by FastQC v1.0.0. Sequence duplication levels show the relative percentage (y-axis) of particular duplication sequences (x-axis) with the sequence duplication level shown at the top.
4.2 Alignment

After read quality assessment, trimmed reads were aligned against *Homo sapiens* hg19 RefSeq by TopHat 2 and Bowtie 2. RNA-seq generated a mean 49.36±2.19 million reads per sample (n = 8). 75.99±0.86% of reads were aligned to the reference genome, 74.10±0.89% were uniquely aligned to a single loci, 8.18±0.67% were aligned to abundant transcripts (mitochondrial and ribosomal sequences), and 28.94±0.66% were aligned across splice junctions. 47.54±0.60% of aligned reads were mapped to coding regions. 32.94±0.67% of aligned reads were mapped to untranslated regions, 15.14±1.91% were mapped to intronic regions, and 4.39±0.24% were mapped to intergenic regions (Supplementary Figure 40). Transcript coverage (normalized read density of each position along the transcript normalized to 100bp) hiPSC samples showed high coverage uniformity (Supplementary Figure 41).
Supplementary Figure 40. RNA-seq alignment information of hiPSCs.
Trimmed RNA-seq reads of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ hiPSCs were aligned against Homo sapiens hg19 RefSeq by TopHat 2 and Bowtie 2. Alignment information for each hiPSC line includes (A) number of reads, number of million reads generated in RNA-seq; (B) aligned reads, percentage of reads passing filter that aligned to the reference genome for both paired end reads; (C) uniquely aligned reads, percentage of reads uniquely aligned to a single loci; (D) mitochondrial and ribosomal reads, percentage of reads aligned to abundant transcripts such as mitochondrial and ribosomal sequences; (E) spliced aligned reads, percentage of reads aligned across splice junctions; and (F) alignment regions, percentages of aligned reads mapped to coding regions, untranslated regions, intronic regions, and intergenic regions (legend on right).
**Supplementary Figure 41.** RNA-seq transcript coverage of hiPSCs.
Trimmed RNA-seq reads of feeder depleted APC-CD56+/BV421-CD184+/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ hiPSCs were aligned against *Homo sapiens* hg19 RefSeq by TopHat 2 and Bowtie 2. Transcript coverage shows normalized coverage or read density of each position along the transcript normalized to 100bp (legend at top right corner).
4.3 Reproducibility

Aligned reads of hiPSC samples were assessed for reproducibility between normal biological replicates N530 and N551, as well as between SDS biological replicates P55 and P357 with fragment bias correction and multi-read correction by Cuffdiff 2. Differential expression heat map, which shows the expression levels of significantly differentially expressed genes across RNA-seq samples, showed highest to lowest positive correlation between N551I and N551K, between N530E and N530J, between N530 and N551, between P55C and P55F, between P357A and P357D, and between P55 and P357 (Supplementary Figure 42).

4.4 Correlation

Correlation between hiPSC samples were assessed using differential expression heat map with fragment bias correction and multi-read correction by Cuffdiff 2. In contrast to aforementioned findings in correlation, normal hiPSCs did not cluster together and SDS hiPSCs also did not cluster together (Supplementary Figure 43).
Supplementary Figure 42. RNA-seq differential expression heat map of biological replicates of SDS and normal hiPSCs. Aligned RNA-seq reads of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60*/PerCP-Cy5.5-SSEA-4*/PE-CD30* hiPSCs were analyzed using differential expression heat map by Cuffdiff 2. Differential expression heat map shows the expression level (legend below) of genes with significant differential expression (y-axis) across biological replicates (x-axis) of normal (left heat map) and SDS (right heat map). Gene dendrogram (left of y-axis) and sample dendrogram (top of x-axis) show clustering and correlation.
Supplementary Figure 43. RNA-seq differential expression heat map of hiPSCs. Aligned RNA-seq reads of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60*/PerCP-Cy5.5-SSEA-4*/PE-CD30* hiPSCs were analyzed using differential expression heat map by Cuffdiff 2. Differential expression heat map shows the expression level (legend below) of genes with significant differential expression (y-axis) across hiPSCs (x-axis). Gene dendrogram (left of y-axis) and sample dendrogram (top of x-axis) show clustering and correlation.
4.5 SBDS Splice Variants

Supplementary Figure 44. SBDS splice variants of hiPSCs visualized by IGV. 
SBDS transcripts of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ hiPSCs were visualized by IGV. RNA-seq confirmed the expression of SBDS splice variants r.[129_258del, 251_258del] in SDS hiPSCs. The expression of the full length (FL) SBDS transcript was confirmed in normal and P357 hiPSCs. Other novel transcripts were found including r.1_128del; 128_129ins129-142_129-1 in N530E, r.129_192del; 251_258del in P55F, and r.129_192del in P357D. Each splice variant is shown with its Cufflinks (CUFF) tracking ID.
**Supplementary Table 1. Detailed transcriptional features of SBDS variants.**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Known</th>
<th>Predicted</th>
<th>Predicted</th>
<th>Novel</th>
<th>Novel</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SBDS Variants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Full length SBDS</td>
<td>8nt deletion 3’ of exon 2 (alternative splice donor site c.251_252GT)</td>
<td>Exon 2 skipping</td>
<td>64nt deletion 5’ of exon 2 (alternative splice acceptor site c.191_192AG)</td>
<td>64nt deletion 5’ of exon 2 (alternative splice acceptor site c.191_192AG) and 8nt deletion 3’ of exon 2 (alternative splice donor site c.251_252GT)</td>
<td>Exon 1 skipping and 142nt intronic retention 5’ of exon 2</td>
</tr>
<tr>
<td><strong>Protein</strong></td>
<td>Full length, p.Lys62*</td>
<td>p.Cys84Tyrfs*4</td>
<td>p.(Glu43Phefs*2)</td>
<td>p.(Glu43Leufs*24)</td>
<td>p.(Glu43Leufs*184)</td>
<td></td>
</tr>
<tr>
<td><strong>Chromosome Position</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 1</td>
<td>chr7.66460277-66460588</td>
<td>chr7.66460277-66460606</td>
<td>chr7.66460277-66460606</td>
<td>chr7.66460277-66460606</td>
<td>chr7.66460277-66460601</td>
<td></td>
</tr>
<tr>
<td>Exon 2</td>
<td>chr7.66459199-66459328</td>
<td>chr7.66459207-66459328</td>
<td>chr7.66459199-66459264</td>
<td>chr7.66459207-66459264</td>
<td>chr7.66459199-66459470</td>
<td></td>
</tr>
<tr>
<td>Exon 3</td>
<td>chr7.66458204-66458404</td>
<td>chr7.66458204-66458404</td>
<td>chr7.66458204-66458404</td>
<td>chr7.66458204-66458404</td>
<td>chr7.66458204-66458404</td>
<td></td>
</tr>
<tr>
<td>Exon 4</td>
<td>chr7.66456124-66456288</td>
<td>chr7.66456124-66456288</td>
<td>chr7.66456124-66456288</td>
<td>chr7.66456124-66456288</td>
<td>chr7.66456124-66456288</td>
<td></td>
</tr>
<tr>
<td>Exon 5</td>
<td>chr7.66452690-66453486</td>
<td>chr7.66452688-66453486</td>
<td>chr7.66452688-66453486</td>
<td>chr7.66452688-66453486</td>
<td>chr7.66452691-66453486</td>
<td></td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td>Exon 1</td>
<td>312</td>
<td>330</td>
<td>330</td>
<td>325</td>
<td>0</td>
</tr>
<tr>
<td>Exon 2</td>
<td>130</td>
<td>122</td>
<td>0</td>
<td>66</td>
<td>58</td>
<td>272</td>
</tr>
<tr>
<td>Exon 3</td>
<td>201</td>
<td>201</td>
<td>201</td>
<td>201</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>Exon 4</td>
<td>165</td>
<td>165</td>
<td>165</td>
<td>165</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>Exon 5</td>
<td>797</td>
<td>799</td>
<td>799</td>
<td>799</td>
<td>796</td>
<td></td>
</tr>
<tr>
<td><strong>Coding RNA Position</strong></td>
<td>Exon 1</td>
<td>r.1-184_128</td>
<td>r.1-202_128</td>
<td>r.1-202_128</td>
<td>r.1-197_128</td>
<td></td>
</tr>
<tr>
<td>Exon 2</td>
<td>r.129_258</td>
<td>r.129_250</td>
<td>r.193_258</td>
<td>r.193_250</td>
<td>r.129-142_258</td>
<td></td>
</tr>
<tr>
<td>Exon 3</td>
<td>r.259_459</td>
<td>r.259_459</td>
<td>r.259_459</td>
<td>r.259_459</td>
<td>r.259_459</td>
<td></td>
</tr>
<tr>
<td>Exon 4</td>
<td>r.460_624</td>
<td>r.460_624</td>
<td>r.460_624</td>
<td>r.460_624</td>
<td>r.460_624</td>
<td></td>
</tr>
<tr>
<td>Exon 5</td>
<td>r.625_1421</td>
<td>r.625_1423</td>
<td>r.625_1423</td>
<td>r.625_1423</td>
<td>r.625_1420</td>
<td></td>
</tr>
<tr>
<td><strong>Tracking ID</strong></td>
<td>UPN CUFF</td>
<td>N530E CUFF:24411.2</td>
<td>N530I CUFF:34310.1</td>
<td>P55C CUFF:32077.2</td>
<td>P55C CUFF:32077.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N551I CUFF:34345.1</td>
<td>P55F CUFF:45002.3</td>
<td>P55F CUFF:45002.1</td>
<td>P55F CUFF:45002.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N551K CUFF:34363.1</td>
<td>P357A CUFF:31700.3</td>
<td>P357A CUFF:31700.1</td>
<td>P357D CUFF:36256.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P357A CUFF:31700.2</td>
<td>P357D CUFF:36256.3</td>
<td>P357D CUFF:36256.2</td>
<td>P357D CUFF:36256.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>P357D CUFF:36256.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The CUFF tracking ID was generated during initial alignment and was assigned to each transcript; the TCONS tracking ID was generated during transcript assembly and differential expression and was assigned to each isoform or splice variant.*
Supplementary Figure 45. Sashimi plot of *SBDS* splice junctions of hiPSCs.

*SBDS* junction reads of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ hiPSCs were visualized using sashimi plot by IGV. Coverage was reduced in *SBDS* exon 2 in SDS. Junction reads spanning exon 1 and 3 summarizes the splicing events resulting in deletions in exon 2 in SDS transcripts. Genomic coordinates and *SBDS* reference sequence are shown at the bottom. Histograms represent coverage of exon or alignment. Arcs represent junction reads, reads spanning the splice junction connecting exons. Both the number and thickness of the arc represent junction depth, the number of junction reads.
4.6 *SBDSP1* Splice Variants

Supplementary Figure 46. *SBDSP1* variants of hiPSCs visualized by IGV. *SBDSP1* transcripts of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60*/PerCP-Cy5.5-SSEA-4*/PE-CD30* hiPSCs were visualized by IGV. Each splice variant is shown with its Cufflinks (CUFF) tracking ID.
### Supplementary Table 2. Detailed transcriptional features of SBDSPI variants.

<table>
<thead>
<tr>
<th>SBDSPI.1</th>
<th>SBDSPI.2</th>
<th>SBDSPI.3</th>
<th>SBDSPI.4</th>
<th>SBDSPI.1</th>
<th>SBDSPI.1</th>
<th>SBDSPI.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known</td>
<td>Known</td>
<td>Exon 5 skipping and 78nt intronic retention 3' of exon 4</td>
<td>Exon 2 and 5 skipping and 78nt intronic retention 3' of exon 4</td>
<td>211nt deletion in exon 1</td>
<td>59nt deletion 5' of exon 2</td>
<td>Exon 4 skipping</td>
</tr>
</tbody>
</table>

#### Chromosome Position

<table>
<thead>
<tr>
<th>Feature</th>
<th>Size (nt)</th>
<th>Non-coding RNA Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 1</td>
<td>411</td>
<td>chr7.72299952-72300362</td>
</tr>
<tr>
<td>Exon 2</td>
<td>122</td>
<td>chr7.72301272-72301393</td>
</tr>
<tr>
<td>Exon 3</td>
<td>197</td>
<td>chr7.72302182-72302378</td>
</tr>
<tr>
<td>Exon 4</td>
<td>165</td>
<td>chr7.72304344-72304508</td>
</tr>
<tr>
<td>Exon 5</td>
<td>823</td>
<td>chr7.72307156-72307978</td>
</tr>
</tbody>
</table>

#### Non-coding RNA Position

<table>
<thead>
<tr>
<th>Tracking ID</th>
<th>CUFF</th>
<th>N530E CUFF:24479.3</th>
<th>N530I CUFF:34437.2</th>
<th>N551I CUFF:34501.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCONS_00140386</td>
<td>TCONS_00140391</td>
<td>N530E CUFF:24479.2</td>
<td>N530I CUFF:34437.1</td>
<td>N551I CUFF:34501.1</td>
</tr>
<tr>
<td>TCONS_00140387</td>
<td>TCONS_00140390</td>
<td>N530E CUFF:24479.1</td>
<td>P55F CUFF:45159.1</td>
<td>P357D CUFF:36412.2</td>
</tr>
</tbody>
</table>
Supplementary Figure 47. Sashimi plot of $SBDSPI$ splice junctions of hiPSCs.
$SBDSPI$ junction reads of feeder depleted APC-CD56/BV421-CD184/FITC-TRA-1-60+/PerCP-Cy5.5-SSEA-4+/PE-CD30+ hiPSCs were visualized using sashimi plot by IGV. Genomic coordinates and $SBDSPI$ reference sequence are shown at the bottom. Histograms represent coverage of exon or alignment. Arcs represent junction reads, reads spanning the splice junction connecting exons. The number and thickness of the arc represent junction depth, the number of junction reads.