Molecular analysis of the erythroid phenotype of a patient with *BCL11A* haploinsufficiency

Marja W. Wessels, Marjon H. Cnossen, Thamar B. van Dijk, Nynke Gillemans, K.L. Juliëtte Schmidt, Kirsten van Lom, Divya S. Vinjamur, Steven Coyne, Ryo Kurita, Yukio Nakamura, Stella A. de Man, Rolph Pfundt, Zakia Azmani, Rutger W.W. Brouwer, Daniel E. Bauer, Mirjam C.G.N. van den Hout, Wilfred F.J. van IJcken and Sjaak Philipsen

Supplemental Materials and Methods
Supplemental Tables 1 and 2 (separate excel file)
Supplemental Figures 1-4

Supplemental Materials and Methods

Hematological analyses

Complete blood counts were performed with an automated hemocytometer used for routine measurements of patient blood samples. High performance liquid chromatography (HPLC) for quantification of HbF, HbA and HbA2 ($\alpha_2\delta_2$) was performed on a Variant II hemoglobin testing system (Bio-Rad Laboratories, Hercules, CA). Age-adjusted reference values were obtained from the Erasmus MC Clinical Chemistry and Hematology Departments.

Flow cytometry

Single-cell suspensions of peripheral blood erythrocytes were washed twice with PBS and then resuspended in PBS containing 1% (w/v) bovine serum albumin and 1 mM EDTA. Approximately 10^6 cells were incubated for 30 minutes at room temperature with CD117-BV421 (562435, dilution 1:100); CD71-FITC (555536, dilution 1:400) CD44-PE (550989, dilution 1:200) and CD235a-APC (561775, dilution 1:2000) antibodies in a final volume of 100 µl. Antibodies were from BD Biosciences, Franklin Lake, NJ. The cells were washed and ~50,0000 cells were measured on a Fortessa instrument (BD Biosciences). Data were analyzed with FlowJo software v10 (FlowJo, Ashland, OR).

Cell culture

~7 ml of blood was collected and the buffy coat was isolated by centrifugation. White cells were isolated from the interphase after Ficoll gradient purification, and washed with PBS. For initial expansion, cells were cultured at a density of 1-2x10^6 cells/ml in serum-free medium (StemSpan; Stem Cell Technologies, Vancouver, BC) enriched with lipids (40 ng/ml cholesterol-rich lipid mix; Sigma-Aldrich, St Louis, MO) and supplemented with erythropoietin (EPO; 2 u/ml, Orthobiotech, Tilburg, NL), dexamethasone (Dex; 1 µM, Sigma-Aldrich) and stem cell factor (SCF; 50 ng/ml, home-made from a COS producer cell line). The erythroblast cultures were expanded until Day 17 by daily partial medium changes and addition of fresh factors, keeping cell densities between 1.5–2x10^6 cells/ml. Differentiation was induced by withdrawal of Dex and SCF, increasing the EPO concentration to 10 units/ml, and adding iron-saturated ferritin (1 mg/ml) and human plasma (5% v/v). Cells were harvested at Day 20. Proliferation kinetics and cell size distributions were monitored using an electronic cell counter (CASY; Roche, Bielefeld, DE). HUDEP1 and HUDEP2 cells were grown under similar conditions, with the addition of Doxycycline (1 µg/ml) [1]. The cells were maintained at a density between 0.5-2x10^6 cells/ml; fresh medium was added every 2-3 days. Differentiation of HUDEP cells was induced by withdrawal of Doxycycline, Dex and SCF, increasing the EPO concentration to 10 units/ml, and adding iron-saturated ferritin (1 mg/ml) and human plasma to 5% (v/v). Cells were harvested at 48 hours of differentiation.

BCL11A null HUDEP2 cells were generated by RNP electroporation of wildtype HUDEP2 cells with 2xNLS-Cas9 purified protein (QB3 MacroLab, University of California, Berkeley) and an

sgRNA targeting exon 2 of *BCL11A* (ACAGAUGAUGAACCAGACCA, Synthego) according to previously described RNP electroporation method [2]. Following recovery from electroporation, bulk edited cells were plated by limiting dilution in 96 well plates (30 cells per plate) to obtain single cell clones. Three *BCL11A* null clones were identified by PCR amplification of the targeted region using primers BCL11A_exon2_FP (ACTGCTTGGCTACAGCACCT) and BCL11A_exon2_RP (TACATGATGTGGTGGGATGG), followed by Sanger sequencing of the PCR product and identification of clones with frameshift alleles using TIDE analysis [3].

RNA analysis

RNA was isolated from ~1x10^6 cells using TRI reagent (Sigma-Aldrich). For RT-qPCR analysis, cDNA was synthesized using 2 µg of total RNA together with oligo dT, RNase OUT, and SuperScript reverse transcriptase II (all ThermoFisher Scientific, Waltham, MA) in a total volume of 20 µl for 1 hour at 42°C. 0.2 µl of cDNA was used for amplification by RT-qPCR. Primer design for amplification of the human β-like globins has been described [4].

Gene	globin encoded	Primer pair	PCR product (bp)
HBE	ε	CCCTGGCCCATAAGTACCAC	105
		TTTCTCTCAAGGCCAAGCCC	
HBG1	Aγ	GGTGACCGTTTTGGCAATCC	106
		GTATCTGGAGGACAGGGCAC	
HBG2	Gγ	Identical to *HBG1*	
HBBP1	pseudogene	GTGCAATAATGGGCAACCCC	96
		TGCCTTTGAGGTCATCCGTG	
HBB	β	GCCCTGGCCCACAAGTATC	109
		GCCCTTCATAATATCCCCCAGTT	
PSMD1		CCCCTACCTGTGTCATTGGC	106
		ACTTCCAGAGGGGCAGGATA	

PCR amplification was performed on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) using Platinum Taq DNA polymerase (ThermoFisher Scientific). PCR conditions were 3 min 95°C, 40 cycles [30 sec 95°C, 25 sec 60°C, 15 sec 72°C], 5 sec 60°C, 5 min 95°C. Specific PCR product accumulation was monitored by SYBR Green dye fluorescence (Sigma-Aldrich). Dissociation curves were used to assess the homogeneity of PCR products. To obtain expression values, Ct values were transformed by computing 2^{-Ct}. Human β-like globin (ε+γ+β) expression was used to calculate the contribution of individual β-like globins to the total. Expression of *PSMD1*, encoding a proteasomal subunit, was used as a control to normalize the data. *PSMD1* was selected from the 'L1000 invariant genes' list [5].

Total RNA samples were checked for quality on a 2100 Bioanalyzer using the RNA nano assay (Agilent Technologies, Santa Clara, CA). All samples had RNA integrity values greater than

8. RNA-seq libraries were prepared according to the Illumina TruSeq stranded mRNA protocol using 200 ng total RNA (Illumina, San Diego, CA). Cluster generation was performed according to the TruSeq SR Rapid Cluster kit v2 (cBot) Reagents Preparation Guide (Illumina). Briefly, RNA-seq libraries were pooled and sequencing-by-synthesis was performed using a HiSeq 2500 with a single-read 50-cycle protocol followed by dual index sequencing. Per library >20x10^6 sequencing reads were obtained.

Bioinformatics

Sequencing reads were trimmed to remove adapter sequences and subsequently mapped against the human reference genome (GRCh38) using HiSat2 (version 2.1.0) [6]. Gene expression values were called using htseq-count version 0.9.1 [7] and Ensembl release 91 gene and transcript annotation (https://www.ensembl.org/biomart/).

Sample quality control and differential expression analysis were performed in the R environment for statistical computing version 3.6.3 (http://www/R-project.org), using the packages DESeq2 version 1.26.0 [8] and tidyverse version 1.3.0 (https://github.com/tidyverse). Vulcano plots for the two differential tests patient *versus* HUDEP1 and patient *versus* HUDEP2 have been created with EnhancedVolcano version 1.4.0.

Chromatin accessibility assays

To assess chromatin accessibility, 50,000 cells were subjected to the ATAC-seq procedure as described [9] with minor modifications. Briefly, nuclei were isolated by washing the cells in detergent-containing buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% octylphenoxy poly(ethyleneoxy)ethanol (IGEPAL CA-630, Sigma-Aldrich). Nuclei were treated immediately for 30 minutes at 37 °C with the Nextera Tn5 transposase mixture, which inserts sequencing adapters into accessible regions of chromatin. Next, the samples were purified using the MinElute PCR Purification Kit (Qiagen, Hilden, DE) according to the manufacturer's protocol and eluted with 10 µl elution buffer. Samples were amplified using the NEBNext High Fidelity PCR master mix (New England Biolabs, Ipswich, USA) and afterwards purified using a MinElute PCR purification kit (Qiagen) according to the manufacturer's instructions. One microliter was loaded on a 2100 Bioanalyzer using a DNA 1000 assay to determine the library concentration and for quality check. ATAC-seq libraries were pooled and sequencing-by-synthesis was performed using the HiSeq 2500 with a paired end 50-cycle protocol followed by dual index sequencing. Per library >14x10^6 sequencing reads were obtained.

Bioinformatics

Sequencing reads were trimmed to remove adapter sequences and subsequently mapped against the human reference genome (GRCh38) using HiSat2 (version 2.1.0), and converted to bed format using in-house developed code. Coverage views of *HBB* locus have been created with the Integrated Genome Browser [10].

All Next Generation Sequencing data have been uploaded to the European Nucleotide Archive accession number PRJEB31728.

DNA methylation analysis of the HBG1/2 promoters

Genomic DNA from cultured cells (1-2x10^6) was extracted by overnight treatment incubation with lysis buffer and proteinase K, followed by phenol-chloroform extraction, ethanol precipitation and RNase digestion. Bisulfite conversion was carried out using the Epitect kit (Qiagen, Hilden, DE) following the manufacturer's protocol. Samples were amplified using primers specific for the *HBG1/2* promoters.

U HBG1/2 FW GTTTTGGTATTTTTTATGATGGGAG

U HBG1/2 RV AACCTTATCCTCCTCTATAAAATAACC

Amplified fragments were loaded on 2% agarose gels and extracted from the gel using a NucleoSpin Extract II kit (Macherey-Nagel, Düren, DE). Fragments were then ligated to pGEM-t easy vector (Promega, Madison, WI) and transfected into *E. coli* DH12 competent cells. Colony PCR was performed on >20 colonies and the PCR fragments were sequenced by Sanger sequencing [11]. Analysis of methylated and unmethylated CpGs was executed using two online programs: BISMA (http://services.ibc.uni-stuttgart.de/BDPC/BISMA/) and QUMA (http://quma.cdb.riken.jp/).

Protein analysis

Whole cell protein lysates were prepared from 2-5x10^6 cells. Cells were lysed in lysis buffer (20mM Tris-HCl pH 7.4, 150mM NaCl, 1% NP40, 0.1% SDS, 0.5% sodium deoxycholate, 5mM EDTA), supplemented with complete protease inhibitor mix (Roche), on ice for 15 minutes and centrifuged at 4°C for 10 minutes at 15 000 rpm. Ten to 50 µg of protein was subjected to SDS-polyacrylamide gel electrophoresis, and electrotransferred to nitrocellulose membranes (Schleicher & Schuell; Dassel, DE). Nitrocellulose membranes were blocked in PBS containing 1% BSA and 0.1% Tween (Sigma-Aldrich), probed with the appropriate combinations of primary and secondary antibodies and analyzed using the Odyssey Infrared Imaging System (Li-Cor, Lincoln, NE). For LRF, chemiluminescence detection was applied using the Pierce ECL Western Blotting Substrate (32106, Thermo Fisher Scientific) and a Typhoon FLA9500 imaging system (General Electric, Boston, MA). Primary antibodies used are: BCL11A (ab 19487 mouse monoclonal raised against aa 172-434, Abcam, Cambridge, UK); MYB (sc-516, rabbit polyclonal, Santa Cruz Biotechnology, Dallas, TX); LIN28B (4196 rabbit polyclonal, Cell Signaling Technology, Danvers, MA), KLF1 (home-made rabbit polyclonal); LRF (sc-33683 Armenian hamster monoclonal, Santa Cruz); β-globin (sc-130321 mouse monoclonal, Santa Cruz); γ-globin (sc-21756 mouse monoclonal, Santa Cruz); β-tubulin (T8328 mouse monoclonal, Sigma-Aldrich).

Supplementary References

1. Kurita R, Suda N, Sudo K, et al. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. PLoS One. 2013;8:e59890.
2. Wu Y, Zeng J, Roscoe BP, et al. Highly efficient therapeutic gene editing of human hematopoietic stem cells. Nat Med. 2019;25:776-783.
3. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. 2014;42:e168.
4. Korporaal A, Gillemans N, Heshusius S, et al. Hemoglobin switching in mice carrying the Klf1(Nan) variant. Haematologica. 2021;106:464-473.
5. Subramanian A, Narayan R, Corsello SM, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017;171:1437-1452 e1417.
6. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357-360.
7. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166-169.
8. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
9. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10:1213-1218.
10. Freese NH, Norris DC, Loraine AE. Integrated genome browser: visual analytics platform for genomics. Bioinformatics. 2016;32:2089-2095.
11. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74:5463-5467.

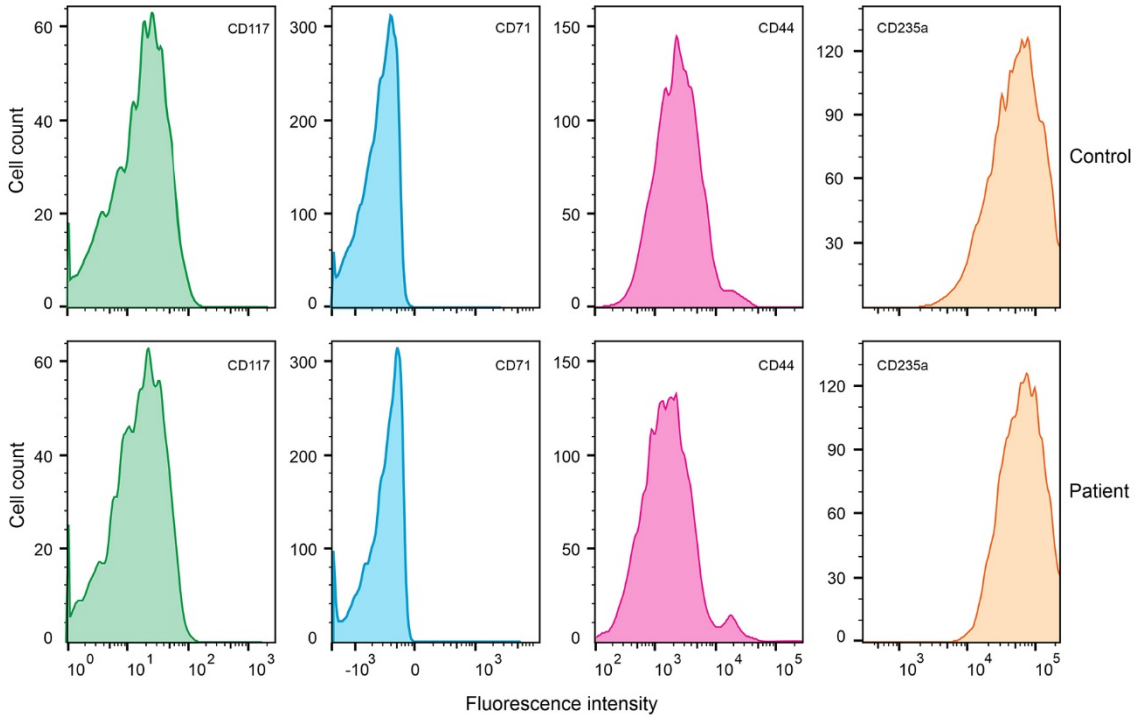
Supplemental Table 1. Complete blood counts

CBC	Mother	Father	Child	Unit
WBC	7.31	5.64	9.07	[10^9/L]
RBC	4.96	5.65	4.7	[1 0^12/L]
HGB	9.3	9.5	7.8	[mmol/L]
HCT	0.446	0.456	0.364	[L/L]
MCV	89.9	80.7	77.4	[fl]
MCH	1875	1681	1660	[amol]
MCHC	20.9	20.8	21.4	[mmol/L]
PLT	265	264	218	[10^9/L]
RDW-SO	44.9	37.2	33.8	[fl]
RDW-CV	14	12.8	12.4	[%]
PDW	11.4	13.1	10.3	[fl]
MPV	9.6	10.9	9.2	[fl]
P-LCR	23.4	32	19.5	[%]
PCT	0.26	0.29	0.2	[%]
NEUT	4.17	2.64	5.58	[10^9/L]
LYMPH	2.53	1.84	2.48	[10^9/L]
MONO	0.51	0.5	0.92	[10^9/L]
EO	0.09	0.6	0.08	[10^9/L]
BASO	0.01	0.06	0.01	[10^9/L]
RET	1.31	0.98	1.15	[%]
IRF	8.9	6.4	5.1	[%]
LFR	91.1	93.6	94.9	[%]
MFR	8	5.3	4.8	[%]
HFR	0.9	1.1	0.3	[%]
RET-He	2265	2137	1976	[amol]
IG	0.01	0.01	0.01	[10^9/L]
NEUT	57.1	46.8	61.6	[%]
LYMPH	34.6	32.6	27.3	[%]
MONO	7	8.9	10.1	[%]
EO	1.2	10.6	0.9	[%]
BASO	0.1	1.1	0.1	[%]
RET	0.065	0.0554	0.0541	[10^12/L]
IG	0.1	0.2	0.1	[%]

Supplemental Table 2. RNA-seq data analysis

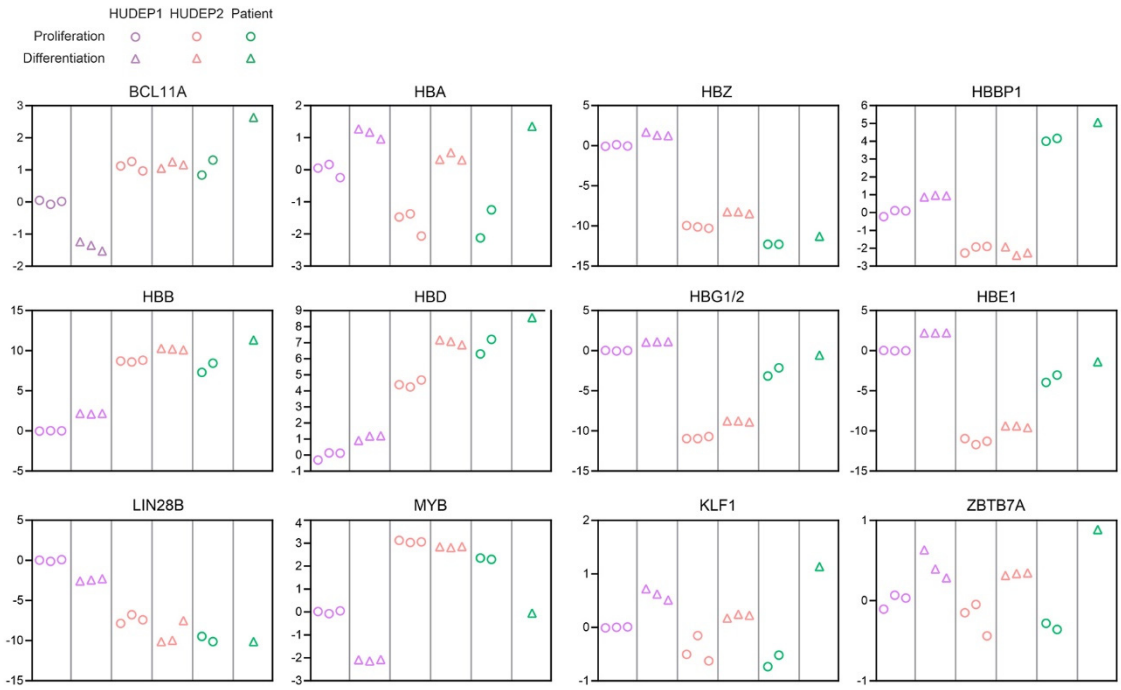
Provided as separate excel file.

Supplemental Figure 1



Supplemental Figure 1. *Flow cytometry analysis of peripheral blood erythrocytes*.
Flow cytometry data for CD117 (KIT), CD71 (transferrin receptor), CD44 (Indian blood type) and CD235a (glycophorin A) are shown for control and patient's erythrocytes.

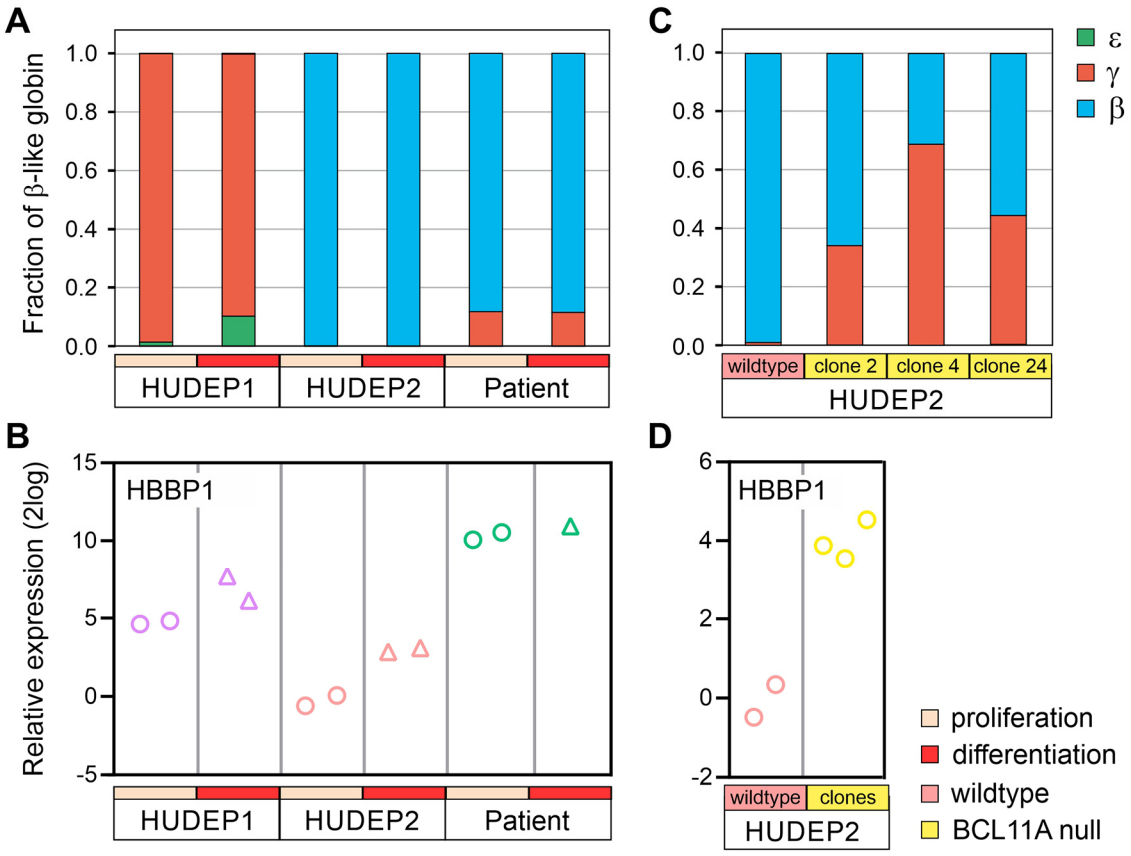
Supplemental Figure 2



Supplemental Figure 2. *RNA-seq analysis of selected genes*.

Expression levels of selected genes, as measured by RNA-seq analysis, are displayed. To allow comparison between samples and individual genes, expression levels in undifferentiated HUDEP1 cells were set to 1. Data were 2log-transformed; note logarithmic scales.

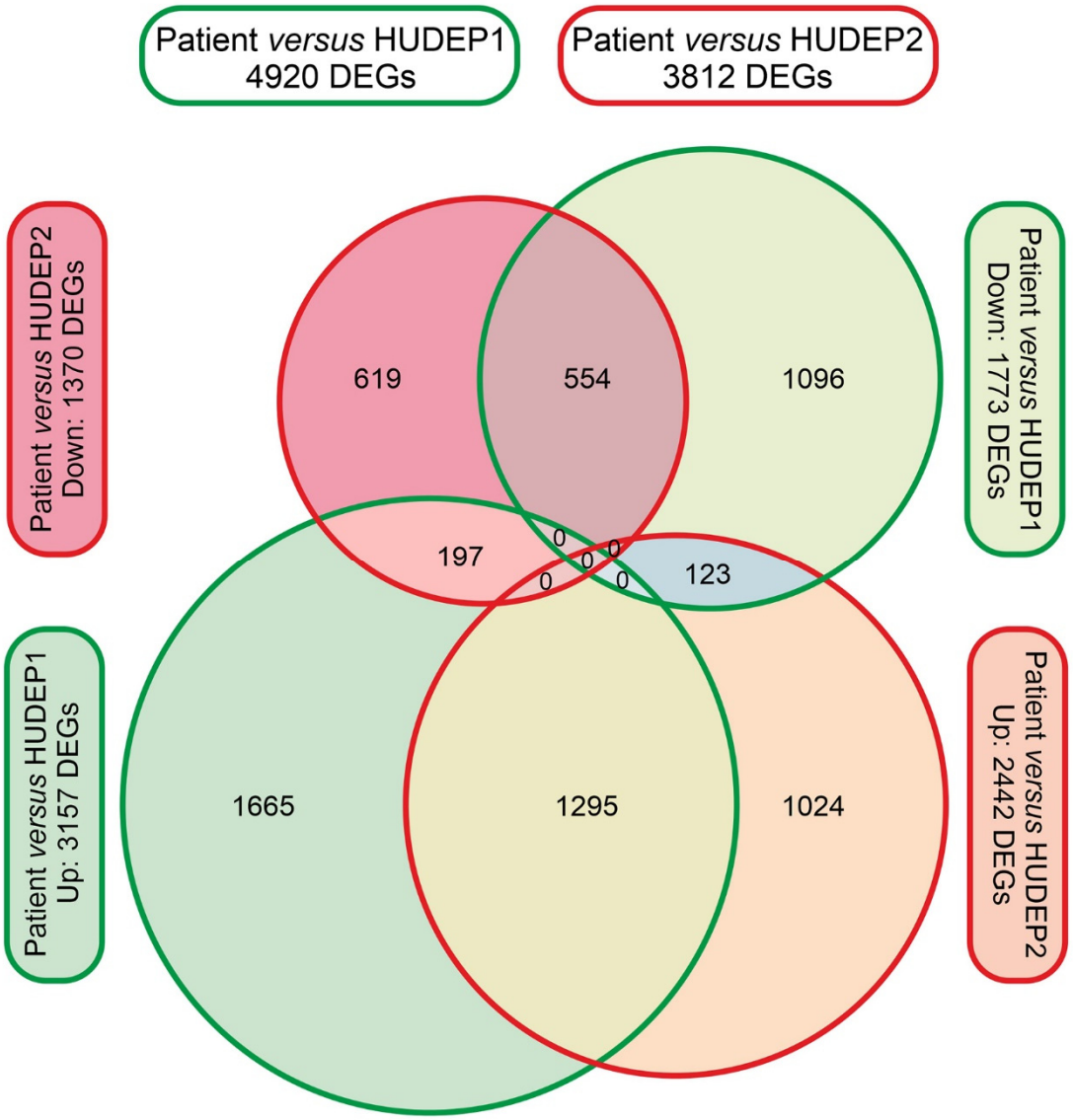
Supplemental Figure 3



Supplemental Figure 3. *RT-qPCR analysis of β-like globin expression*.

A) Expression of ε-, γ-, and β-globin. **B**) Expression of *HBBP1*. **C**) Expression of ε-, γ-, and β-globin in wildtype and BCL11A null HUDEP2 cells. **D**) Expression of *HBBP1* in wildtype and *BCL11A* null HUDEP2 cells. Expression was measured by RT-qPCR analysis of RNA isolated from HUDEP1 and HUDEP2 cells, and patient HEPs as indicated. Cells in **A**) and **B**) were grown under proliferation or differentiation conditions (48 hours of differentiation); cells in **C**) and **D**) were grown under proliferation conditions. Data in **B**) and **D**) were 2log-transformed; note logarithmic scales. Expression of *PSMD1* was used to normalize the data.

Supplemental Figure 4



Supplemental Figure 4. *Venn diagram of differentially expressed genes*.
Differentially expressed genes (DEGs) were selected based on 2log fold-change > |1| and adjusted p-value < 0.01 (Suppl Table 2).