Figure S1

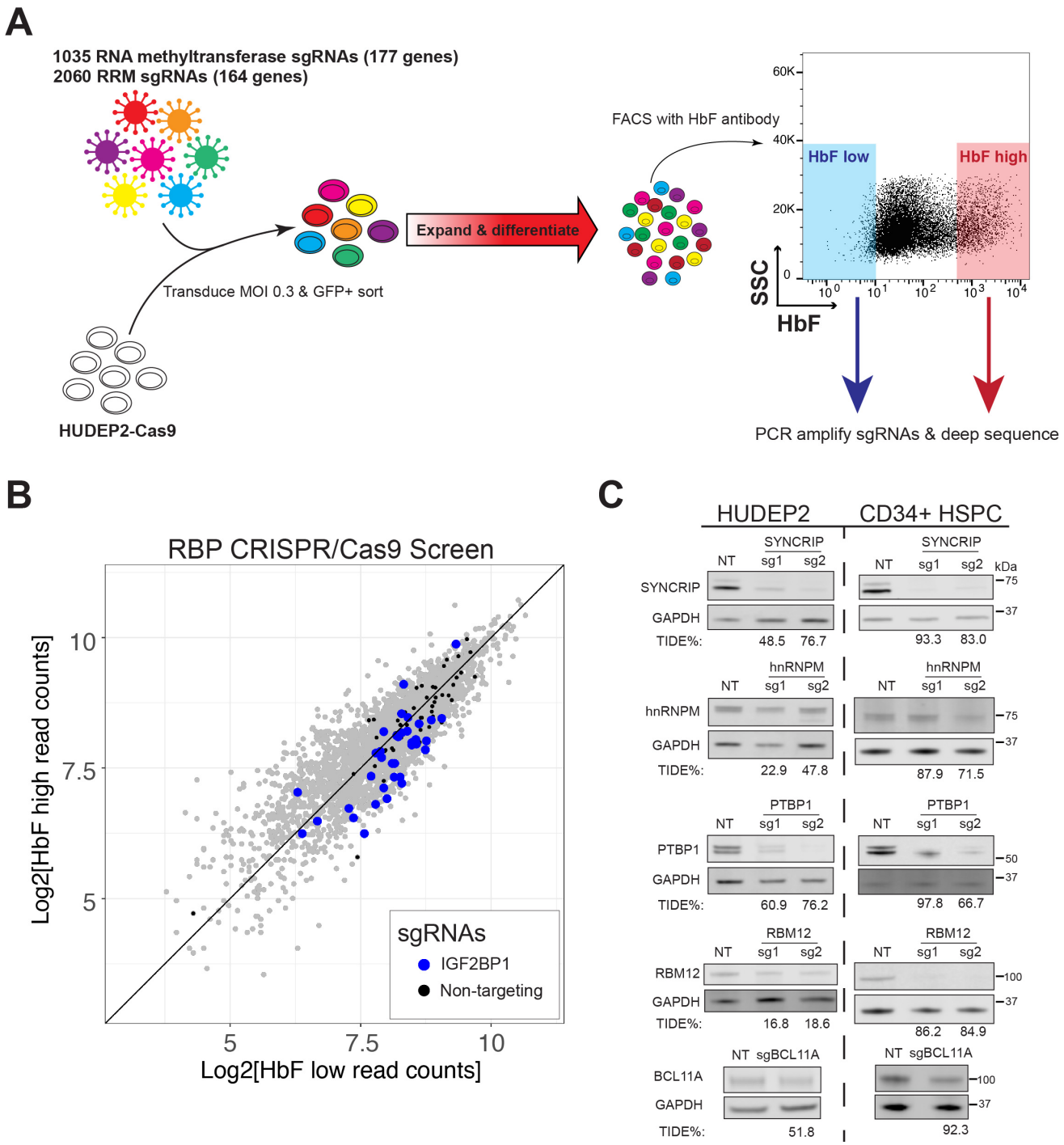

Figure S1. RBP focused CRISPR/Cas9 screen.
A) Experimental outline of CRISPR/Cas9 screening strategy. Cas9 expressing HUDEP2 cells were transduced at multiplicity of infection (MOI) of 0.3 with a lentiviral sgRNA library harboring sgRNAs targeting RNA methyltransferase and RRM domains. GFP+ cells were sorted, expanded, and differentiated for 7 days at which HbF high and low cells were sorted via fluorescence activated cell sorting (FACS) with an HbF antibody. sgRNAs enriched in each population were identified via next generation sequencing.
B) Scatterplot displaying RBP-focused CRISPR/Cas9 screening results. Each dot represents an sgRNA. Highlighted are non-targeting (NT) control sgRNAs and sgRNAs targeting *IGF2BP1*.
C) Western blot analysis of HUDEP2 and CD34+ HSPCs with indicated antibodies and sgRNAs. The sgRNA against BCL11A (positive control) in HUDEP2 cells targeted exon 2, while in CD34+ cells the sgRNA targeted the +58 enhancer. Editing efficiency as measured by TIDE analysis is displayed.

Figure S2

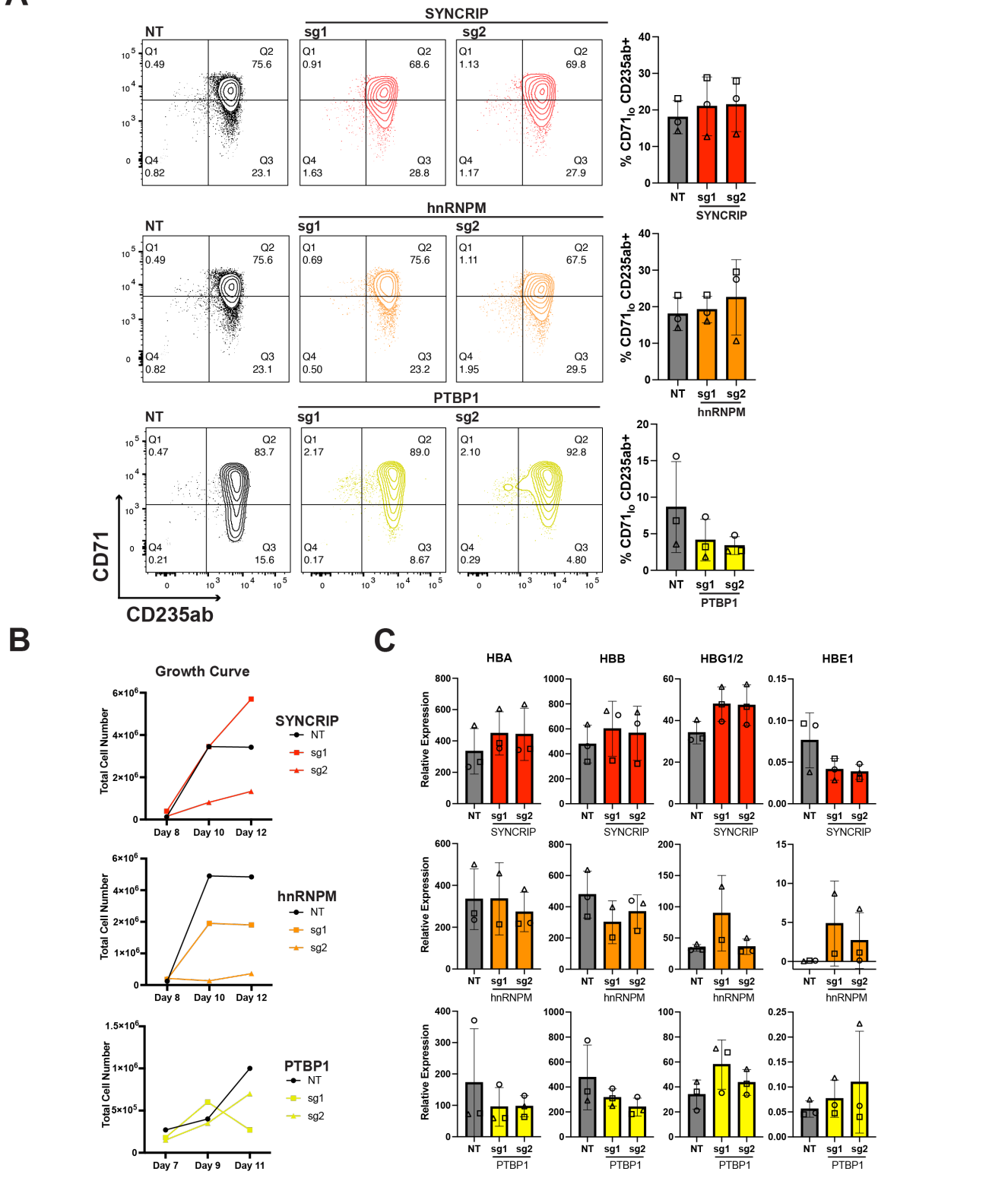

Figure S2. SYNCRIP, hnRNPM, and PTBP1 targeted HSPC maturation and growth.
CRISPR/Cas9 depletion of SYNCRIP, hnRNPM, and PTBP1 in human CD34+ HSPCs via 2 independent sgRNAs. n=3 biological replicates.
A) *Left:* Representative contour plots of CD71 vs CD235ab at day 15 of differentiation. *Right:* Quantification of $CD71_{lo}CD235ab+$ cells. Plotted are mean ± standard deviation.
B) Growth curve displaying total cell number.
C) RT-qPCR analysis of globin genes at day 12 of differentiation. Results are normalized to *HPRT1*. Plotted are mean ± standard deviation.

Figure S3

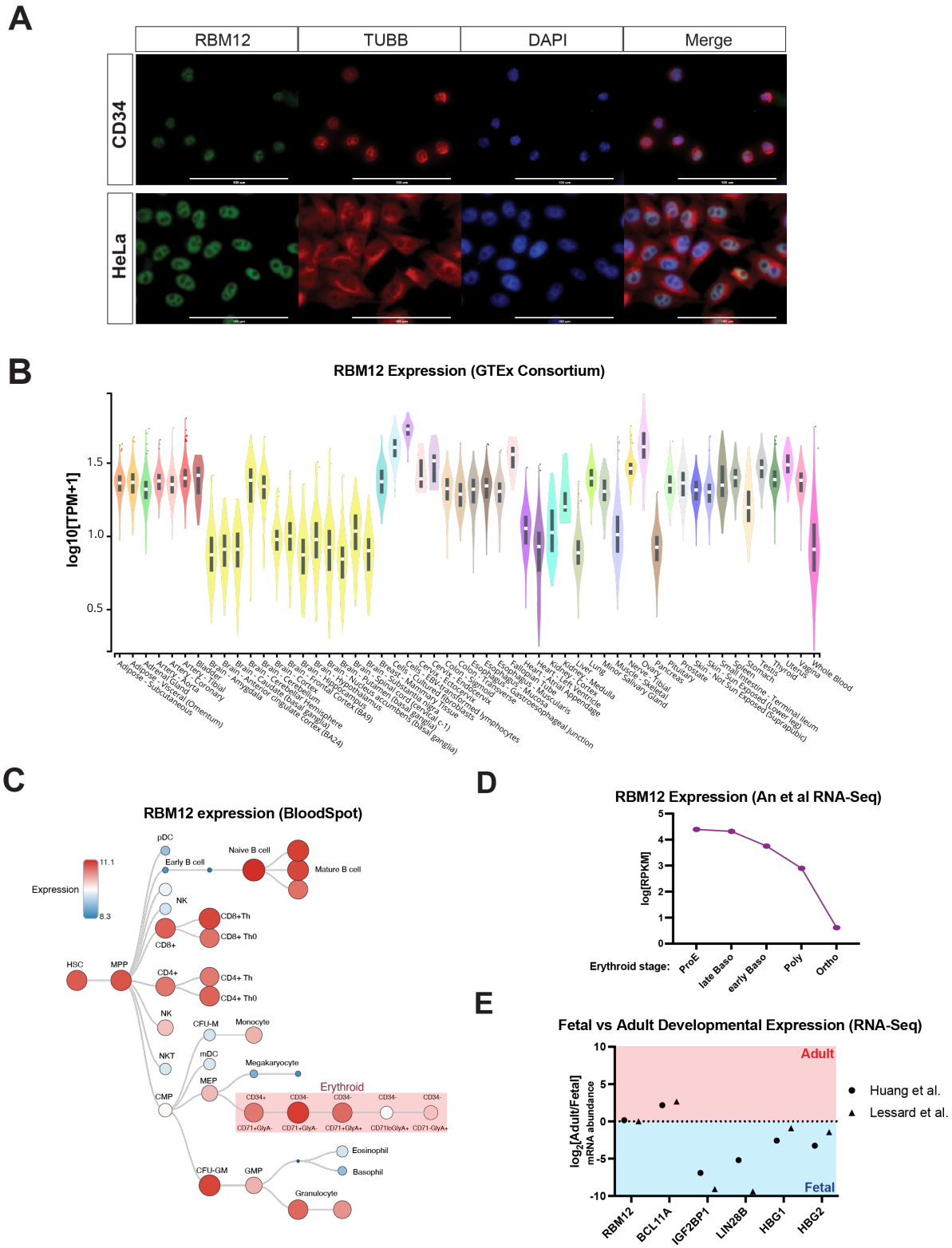

Figure S3. RBM12 expression.
A) Immunofluorescence staining of primary human erythroblasts (top) and HeLa cells (bottom) stained with antibodies for RBM12, TUBB, and DAPI.
B) *RBM12* tissue expression data from GTEx database.[1] TPM = transcript per million.
C) *RBM12* expression in various hematopoietic cell types via Bloodspot microarray analysis.[2]
D) *RBM12* expression across erythroid differentiation.[3] RPKM = reads per kilobase per million.
E) Expression of *RBM12* and developmentally specific genes in RNA-Seq data of erythroblasts isolated from human fetal liver and adult bone marrow from two independent datasets.[4,5]

Figure S4

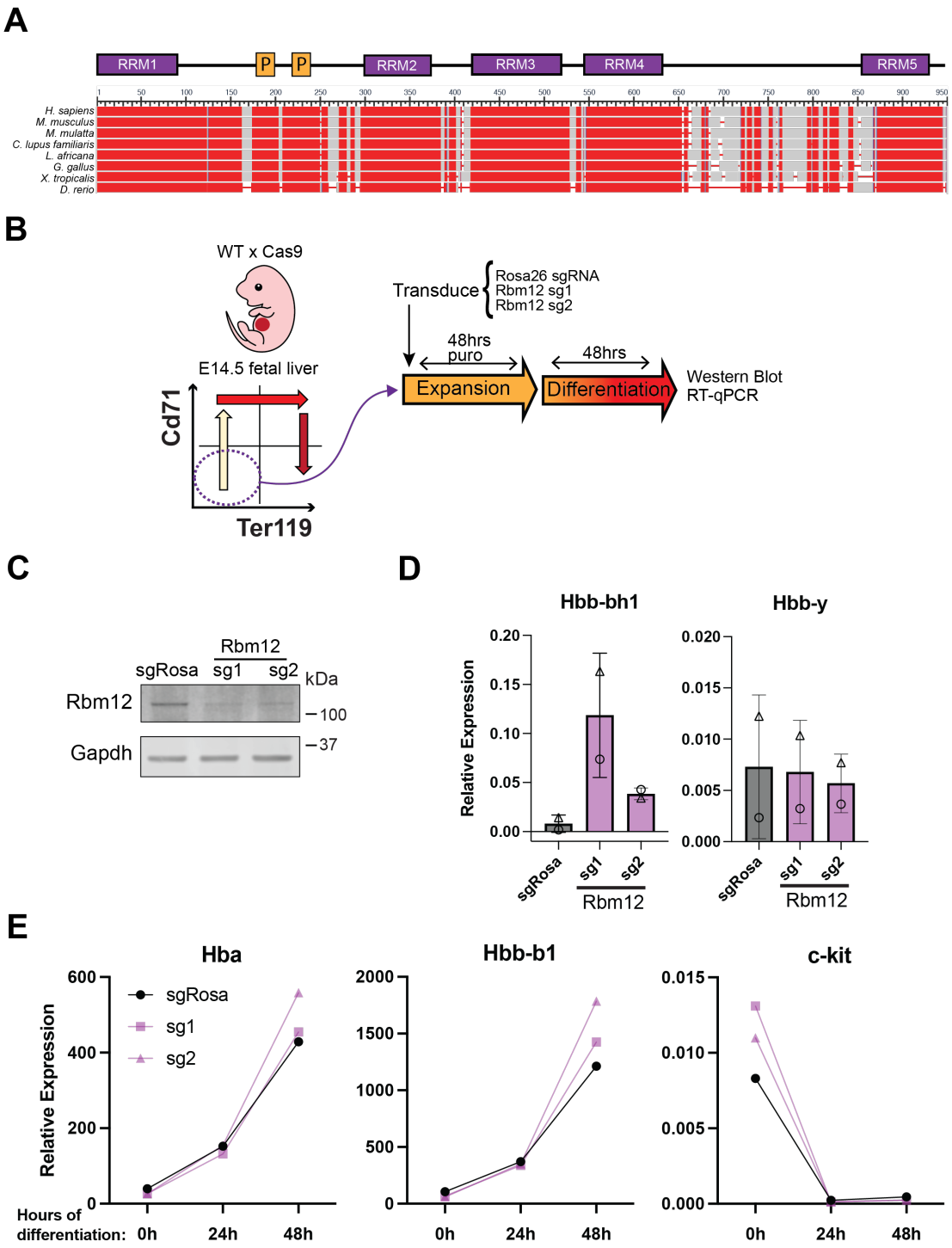

Figure S4. RBM12 mediated β-type globin regulation is conserved across species.

A) Conservation of RBM12 protein across various species via NCBI COBALT alignment tool.[6] Red regions denote high conservation.

B) Schematic of murine erythroblast experiments. Erythroid progenitors were isolated from E14.5 fetal livers via Cd71 depletion. Progenitors were then transduced with retrovirus harboring sgRNAs targeting *Rosa26* (control) and *Rbm12* (2 sgRNAs). Infected cells were enriched via 48 hours of puromycin selection during expansion. Cells were then differentiated for 48 hours at which samples were harvested for western blotting and RT-qPCR analyses.

C) Western blot analysis of murine erythroblasts with indicated antibodies.

D) RT-qPCR of *Hbb-bh1* and *Hbb-y* of murine erythroblasts at 48 hours of differentiation. Results are normalized to *Actb*. Plotted are mean ± standard deviation (n=2).

E) RT-qPCR of adult murine globin transcripts *Hba* and *Hbb-b1* along with early differentiation marker *Kit* of murine erythroblasts at 0, 24, and 48 hours of differentiation. Results are normalized to *Actb.* Plotted are mean ± standard deviation (n=2).

Figure S5

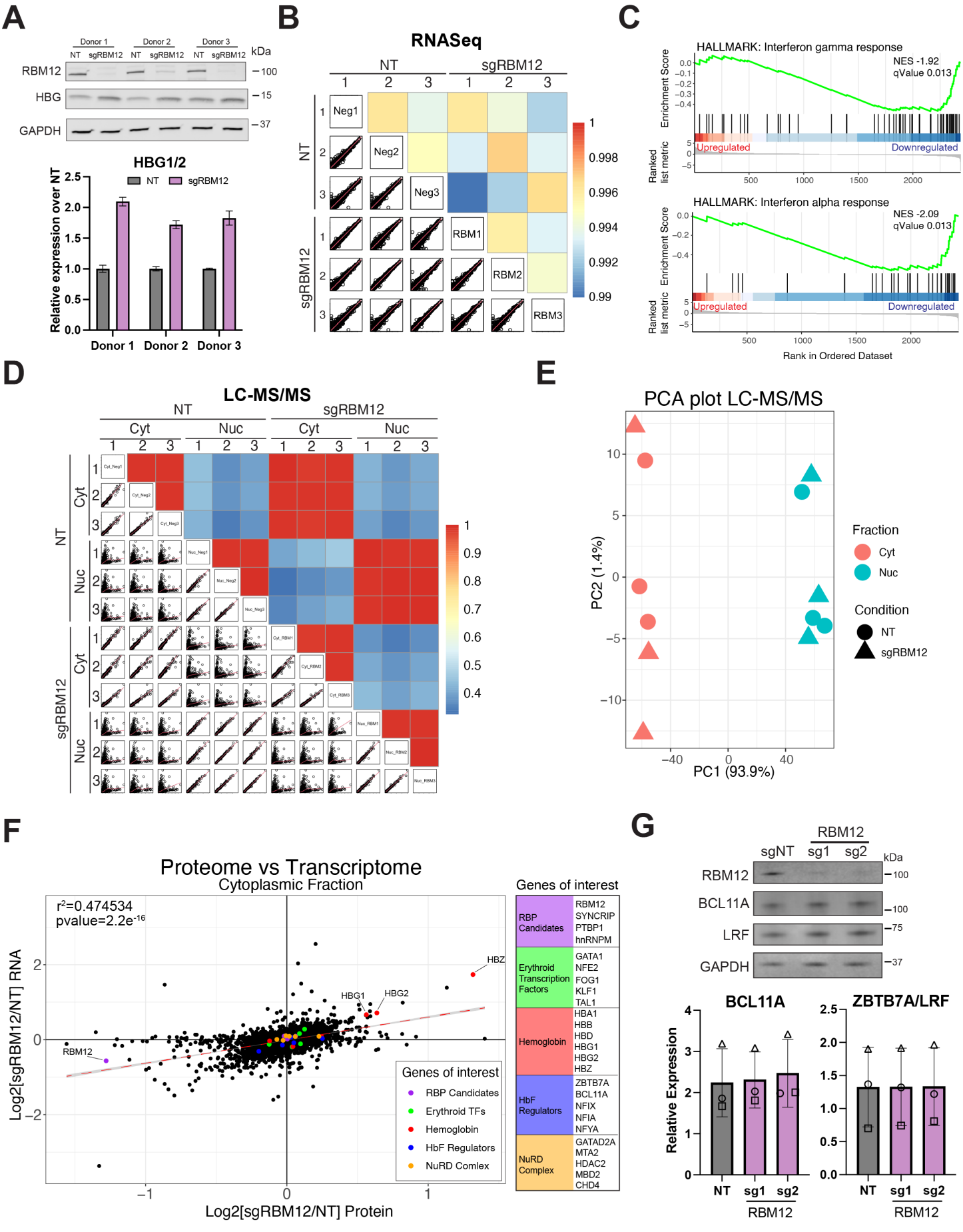
A
Donor 1
Donor 2
Donor 3
NT
sgRBM12
kDa
RBM12
HBG
GAPDH
100
15
37
HBG1/2
Relative expression over NT
B
RNASeq
Neg1
Neg2
Neg3
RBM1
RBM2
RBM3
C
HALLMARK: Interferon gamma response
NES -1.92
qValue 0.013
HALLMARK: Interferon alpha response
NES -2.09
qValue 0.013
Enrichment Score
Ranked list metric
Upregulated
Downregulated
Rank in Ordered Dataset
D
LC-MS/MS
Cyt
Nuc
E
PCA plot LC-MS/MS
PC1 (93.9%)
PC2 (1.4%)
Fraction
Condition
F
Proteome vs Transcriptome
Cytoplasmic Fraction
r²=0.474534
pvalue=2.2e-16
Log2[sgRBM12/NT] RNA
Log2[sgRBM12/NT] Protein
HBZ
HBG1
HBG2
RBM12
Genes of interest
RBP Candidates
Erythroid TFs
Hemoglobin
HbF Regulators
NuRD Comlex
RBP Candidates
RBM12
SYNCRIP
PTBP1
hnRNPM
Erythroid Transcription Factors
GATA1
NFE2
FOG1
KLF1
TAL1
Hemoglobin
HBA1
HBB
HBD
HBG1
HBG2
HBZ
HbF Regulators
ZBTB7A
BCL11A
NFIX
NFIA
NFYA
NuRD Complex
GATAD2A
MTA2
HDAC2
MBD2
CHD4
G
RBM12
sgNT
sg1
sg2
kDa
RBM12
BCL11A
LRF
GAPDH
100
100
75
37
BCL11A
ZBTB7A/LRF
Relative Expression
NT
sg1
sg2
RBM12

Figure S5. RNA-Seq and LC-MS/MS in RBM12 KO primary human HSPCs.
CRISPR/Cas9-mediated depletion of RBM12 in human CD34+ HSPCs. n=3 biological replicates.
A) *Top:* Western blot analysis with indicated antibodies. *Bottom: HBG1/2* RT-qPCR analysis. Results are normalized to *HPRT1*. Plotted are mean ± standard deviations of technical replicates.
B) Pairwise Pearson correlation of RNA-Seq read counts. Color scale display absolute quantification of r values. Red line on scatterplots indicates LOESS smoothed regression line.
C) Gene set enrichment analysis of differentially expressed transcripts. Shown are gene set signatures from the MSigDB hallmark collection. NES = normalized enrichment score.
D) Pairwise Pearson correlation of LC-MS/MS protein abundance.
E) PCA of protein abundance quantifications from LC-MS/MS. Color scale display absolute quantification of r values. Red line on scatterplots indicates LOESS smoothed regression line.
F) Scatter plot of protein vs mRNA abundance from LC-MS/MS and RNA-Seq experiments, respectively. Detected proteins/transcripts were down sampled to those detected in the LC-MS/MS data. r^2 = Pearson correlation coefficient. Red dotted line represents linear model smoothing.
G) *Top*: Western blot analysis with indicated antibodies. *Bottom*: *BCL11A* and *ZBTB7A/LRF* RT-qPCR analysis. Results are normalized to *HPRT1*. Plotted are mean ± standard deviations. n=3 biological replicates.

Figure S6

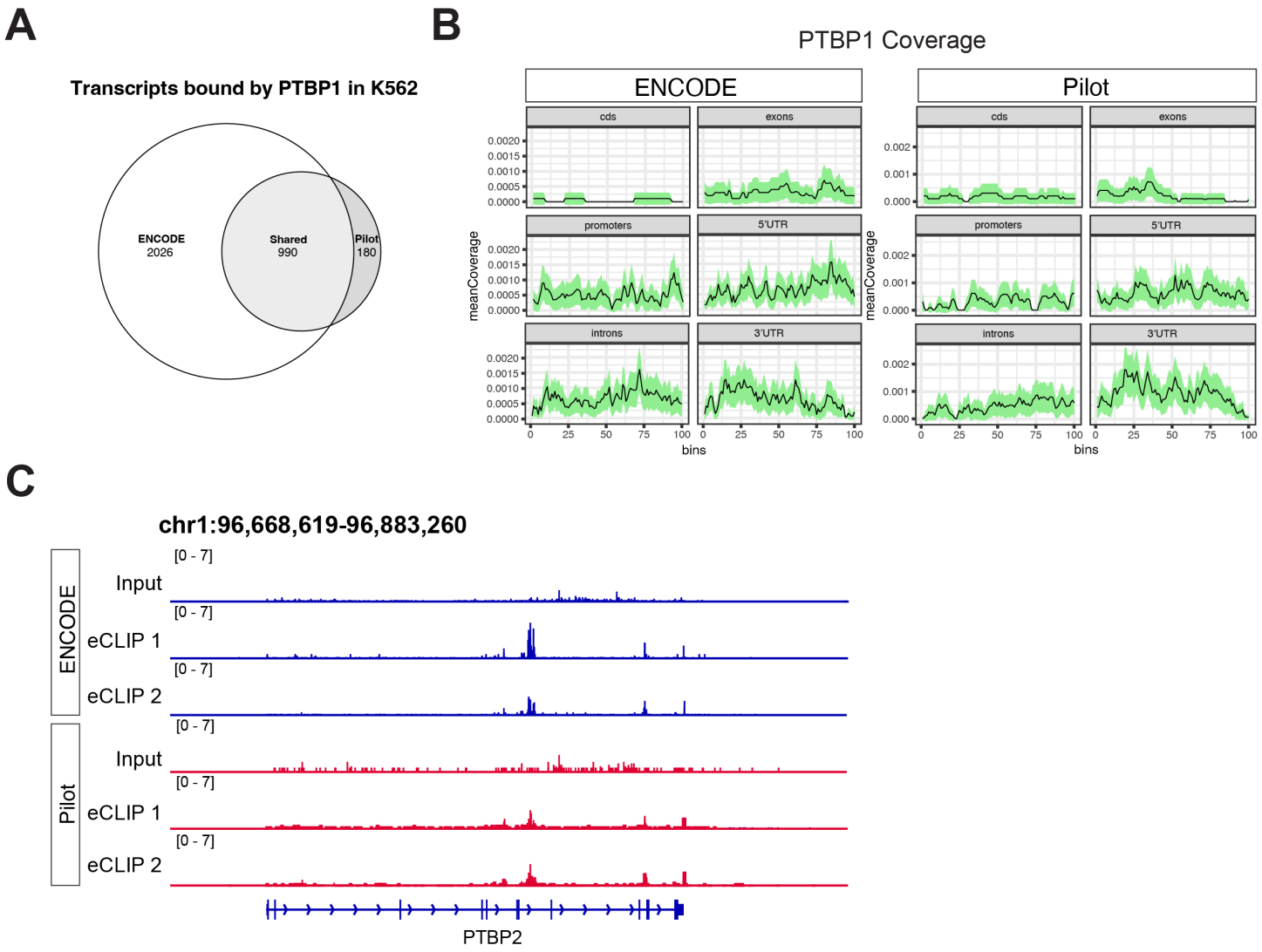

Figure S6. Comparison of PTBP1 eCLIP-Seq in K562 pilot experiment to ENCODE experiment.
A) Venn diagram displaying transcripts commonly found between the experimental pilot and published ENCODE eCLIP-Seq dataset.
B) Coverage profile of PTBP1 binding. Each feature was divided into 100 bins of equal length.
C) Representative Integrative Genomics Viewer (IGV) snapshot of a known PTBP1 target, *PTBP2*.[7] Displayed are the size-matched input control and 2 replicate eCLIP tracks of the ENCODE and pilot PTBP1 eCLIP-Seq in K562 experiments.

Figure S7

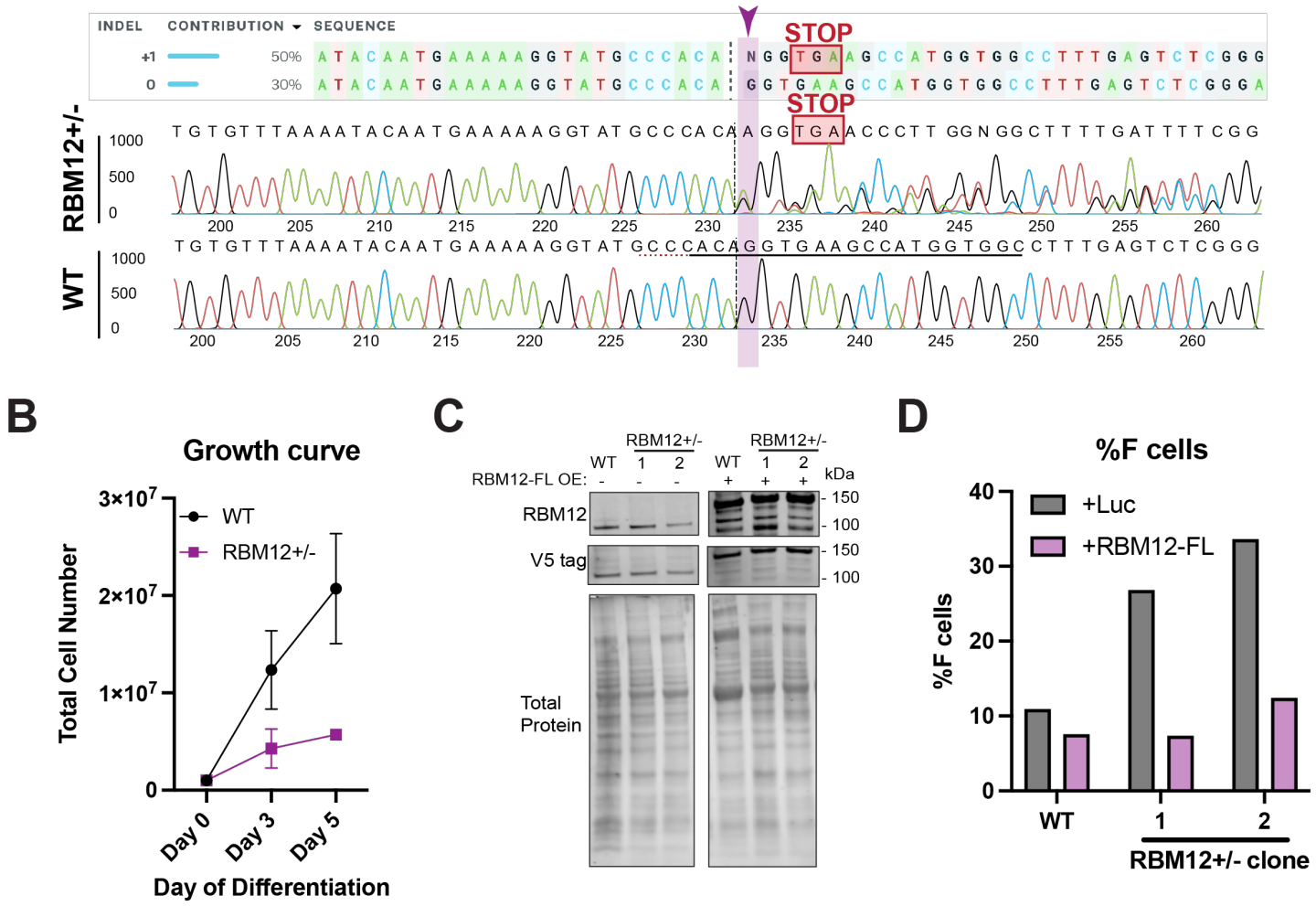

Figure S7. RBM12+/- HUDEP2 clones.
A) Sanger sequencing and mapping of edits found in RBM12+/- clones via SYNTHEGO ICE webtool. Highlighted in purple is the 1bp insertion that creates a premature stop codon.
B) Growth curve of RBM12+/- clones during differentiation. n = 2 independent clones.
C) Western blot analysis with indicated antibodies. Total Protein stain serves as the loading control.
D) HbF flow cytometry at day 5 of differentiation of *RBM12*+/- clones infected with luciferase or *RBM12*-FL OE constructs.

Figure S8

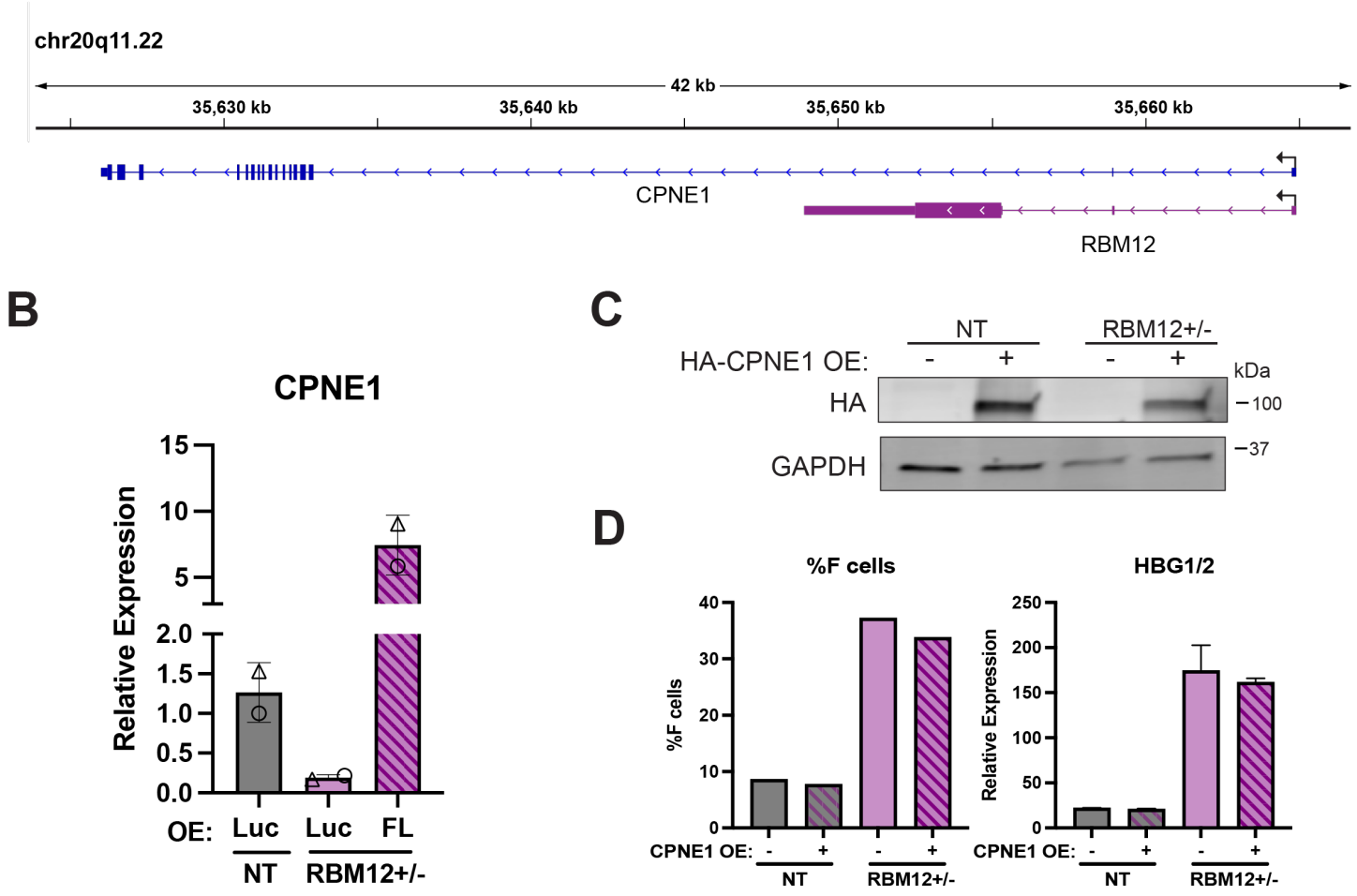

Figure S8. CPNE1 rescue.
A) Genome track view of the *CPNE1*/*RBM12* locus.
B) *CPNE1* RT-qPCR analysis. Results are normalized to *HPRT1*. Plotted are mean ± standard deviations (n=2).
C) Western blot analysis with indicated antibodies.
D) HbF flow cytometry (*left*) and *HBG1/2* RT-qPCR (*right*) of RBM12+/- clones infected with *CPNE1* OE constructs.

Figure S9

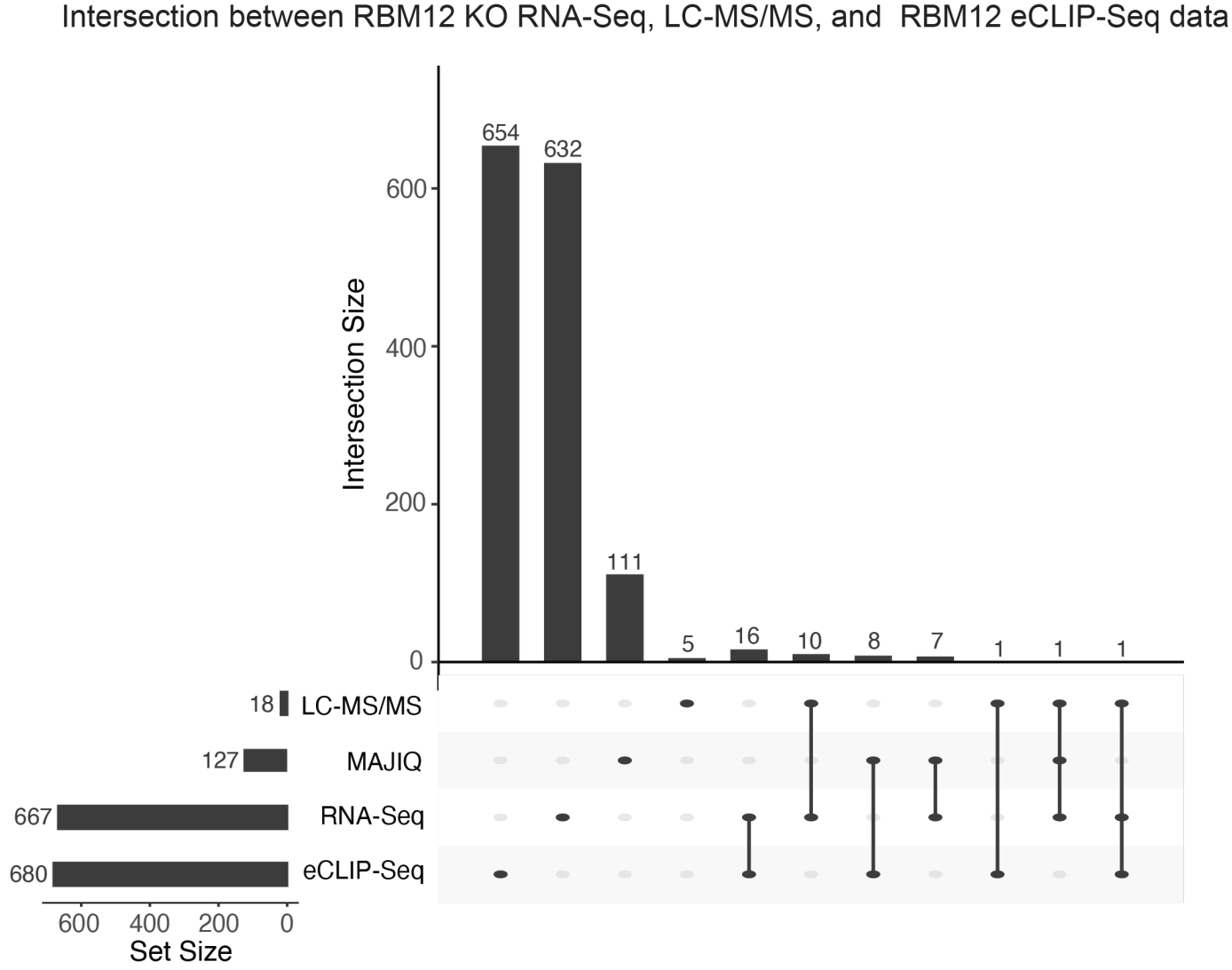

Figure S9. Intersection between RBM12 bound genes and DEGs/DSGs.
Upset plot intersecting RBM12 bound genes with differentially expressed (both transcript and protein) and spliced genes identified in RBM12 KO erythroblasts. MS = differentially expressed proteins identified from LC-MS/MS (q-value < 0.1). MAJIQ = DSGs identified from MAJIQ analysis (dPSI > 0.15). RNA-Seq = differentially expressed genes identified from RNA-Seq dataset (fold-change > 1.5, adjusted p-value < 0.05). eCLIP-Seq = RBM12 bound genes identified by Irreproducible Discovery Rate (IDR) analysis.

Table S1. GSEA of DEGs in RBM12 KO erythroblasts at day 7 of differentiation.
GSEA was done with RNA-Seq DEGs detected in RBM12 KO erythroblasts (FDR < 0.05) using the clusterProfiler package in R against the C2 MSigDB collection. Gene sets in this collection include online pathway databases and biomedical literature. NES = normalized enrichment score.

Description	NES	p.adjust
SANA_RESPONSE_TO_IFNG_UP	-2.46634	0.00034
BLANCO_MELO_COVID19_SARS_COV_2_INFECTION_CALU3_CELLS_UP	-2.38244	0.00034
KEGG_ASTHMA	-2.37860	0.00012
XIE_ST_HSC_S1PR3_OE_UP	-2.33133	0.00157
BLANCO_MELO_BRONCHIAL_EPITHELIAL_CELLS_INFLUENZA_A_DEL_NS1_INFECTION_UP	-2.28491	0.00004
TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_3D_UP	-2.27890	0.00626
REACTOME_INTERFERON_SIGNALING	-2.27708	0.00157
DAUER_STAT3_TARGETS_DN	-2.24768	0.00507
SEITZ_NEOPLASTIC_TRANSFORMATION_BY_8P_DELETION_UP	-2.22697	0.01246
KIM_GLIS2_TARGETS_UP	-2.20537	0.00533
ICHIBA_GRAFT_VERSUS_HOST_DISEASE_D7_UP	-2.19778	0.01444
WIELAND_UP_BY_HBV_INFECTION	-2.17661	0.01246
BOSCO_INTERFERON_INDUCED_ANTIVIRAL_MODULE	-2.16993	0.02448
BLANCO_MELO_HUMAN_PARAINFLUENZA_VIRUS_3_INFECTION_A594_CELLS_UP	-2.15886	0.00483
KEGG_GRAFT_VERSUS_HOST_DISEASE	-2.13462	0.01571
KEGG_TYPE_I_DIABETES_MELLITUS	-2.12901	0.01263
WALLACE_PROSTATE_CANCER_RACE_UP	-2.10440	0.00281
BLANCO_MELO_RESPIRATORY_SYNCYTIAL_VIRUS_INFECTION_A594_CELLS_UP	-2.09637	0.00653
STAMBOLSKY_TARGETS_OF_MUTATED_TP53_DN	-2.08801	0.03802
HECKER_IFNB1_TARGETS	-2.08294	0.03251
WUNDER_INFLAMMATORY_RESPONSE_AND_CHOLESTEROL_UP	-2.07779	0.03351
REACTOME_PTEN_REGULATION	-2.07523	0.01914
BLANCO_MELO_BETA_INTERFERON_TREATED_BRONCHIAL_EPITHELIAL_CELLS_UP	-2.07249	0.00395
LEE_DIFFERENTIATING_T_LYMPHOCYTE	-2.04863	0.02177
ZHONG_RESPONSE_TO_AZACITIDINE_AND_TSA_DN	-2.04308	0.03185
JINESH_BLEBBISHIELD_VS_LIVE_CONTROL_DN	-2.04264	0.01121
JINESH_BLEBBISHIELD_TO_IMMUNE_CELL_FUSION_PBSHMS_DN	-2.04142	0.00750
NUYTTEN_EZH2_TARGETS_UP	-2.02687	0.00008
DOUGLAS_BMI1_TARGETS_DN	-1.99944	0.00918
SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_UP	-1.96192	0.02072
CHYLA_CBFA2T3_TARGETS_UP	-1.89142	0.02704
REACTOME_ADAPTIVE_IMMUNE_SYSTEM	-1.88911	0.00533
NUYTTEN_NIPP1_TARGETS_UP	-1.87173	0.00626
WANG_RESPONSE_TO_GSK3_INHIBITOR_SB216763_UP	-1.80902	0.04759
REACTOME_TRANSLATION	-1.78564	0.03590
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION	1.85251	0.04068
NABA_CORE_MATRISOME	1.91527	0.02448
REACTOME_ASSEMBLY_OF_COLLAGEN_FIBRILS_AND_OTHER_MULTIMERIC_STRUCTURES	1.94258	0.01981
REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX	1.95772	0.02839
REACTOME_ION_CHANNEL_TRANSPORT	1.97103	0.01248
MIKKELSEN_NPC_HCP_WITH_H3K27ME3	2.01146	0.01248
MIKKELSEN_MCV6_HCP_WITH_H3K27ME3	2.02142	0.01194
NABA_MATRISOME	2.09061	0.00066
REACTOME_STIMULI_SENSING_CHANNELS	2.11414	0.00483
MIKKELSEN_MEF_HCP_WITH_H3K27ME3	2.18529	0.00128
BLANCO_MELO_MERS_COV_INFECTION_MCR5_CELLS_UP	2.38516	0.00002

Table S2. LC-MS/MS of sgRBM12 versus NT day 7 differentiated primary human erythroblasts.
Values shown are normalized log2 fold change of proteins that reached the q value cutoff of 0.1. n=3 biological replicates. P values were calculated by student's t test. Q values were adjusted p values to account for multiple testing.

GENE SYMBOL	LOG2 FOLD CHANGE	P VALUE	Q VALUE
CPNE1	-1.329	0.009	0.076
RBM12	-1.279	0.001	0
RPL22L1	-0.976	0.0003	0.046
GLCE	-0.935	0.001	0.033
CLYBL	-0.865	0.006	0.080
DDX60	-0.797	0.001	0.041
SSU72	-0.694	0.001	0.040
IFI35	-0.688	0.0001	0.035
OSBPL1A	-0.670	0.0001	0.037
HADH	-0.640	0.0003	0.035
DNASE2	-0.606	0.002	0.072
SMAP1	-0.507	0.0005	0.077
TNKS2	0.566	6.47e-06	0.037
PRKACA	0.570	7.66e-06	0.031
BCCIP	0.816	0.0008	0.031
TNKS	0.914	0.0004	0.046
ESPN	1.132	0.001	0.026
ANKDD1A	1.453	0.001	0.031

Table S3. sgRNA sequences

Human	
Target	**Sequence**
NT	GCAGCTCGACCTCAAGCCGT
BCL11A exon2 (HUDEP2)	TGAACCAGACCACGGCCCGT
BCL11A +58 (CD34+)	CTAACAGTTGCTTTTATCAC
SYNCRIP sg1	TGAGAAAGCTGGACCTATAT
SYNCRIP sg2	TATTCCTAAGAGTAAAACCA
hnRNPM sg1	CATCATCCGGCAGGCTCTCT
hnRNPM sg2	TATAGCTTGCACAGCTTCAA
PTBP1 sg1	AGTCTCCGAGGAGGATCTCA
PTBP1 sg2	TTGGCGTGCTGGGCGCTCAC
RBM12 sg1	GTGTCGTCATAAACAGTACA
RBM12 sg2	TATGGCTATCAAGTAATCCC
Mouse	
Target	**Sequence**
sgRosa	AAGATGGGCGGGAGTCTTC
Rbm12 sg1	GCTGTGGTCATCCGTTTGCA
Rbm12 sg2	GCAAGGTCTCCCAATTGTGG

Table S4. RT-qPCR primers

Human	
Gene	**Sequence**
HBG1/2 FW	TGGCAAGAAGGTGCTGACTTC
HBG1/2 RV	GCAAAGGTGCCCTTGAGATC
HBB FW	TGGGCAACCCTAAGGTGAAG
HBB RV	GTGAGCCAGGCCATCACTAAA
HBE1 FW	TGCACTGTGACAAGCTGCAT
HBE1 RV	CCTTGCCAAAGTGAGTAGCC
HBA FW	AAGACCTACTTCCCGCACTTC
HBA RV	GTTGGGCATGTCGTCCAC
HPRT1 FW	GAAAAGGACCCCACGAAGTGT
HPRT1 RV	AGTCAAGGGCATATCCTACAACA
BCL11A FW	ACAAACGGAAACAATGCAATGG
BCL11A RV	TTTCATCTCGATTGGTGAAGGG
ZBTB7A/LRF FW	CGAGTGCAACATCTGCAAGG
ZBTB7A/LRF RV	TTCAGGTCGTAGTTGTGGGC
GATA1 FW	TGCGGCCTCTATCACAAGATG
GATA1 RV	CTGCCCGTTTACTGACAATCA
BAND3 FW	CCTATACGCTTCCTCTTTGTGTT
BAND3 RV	CCATGTAGGCATCTATGCGGA
ALAS2 FW	CAGCGCAATGTCAAGCAC
ALAS2 RV	TAGATGCCATGCTTGGAGAG
CPNE1 FW	CACTGCGTGACCTTGGTTCA
CPNE1 RV	CTCCCACATCCTGTAAAAGGAC
Mouse	
Gene	**Sequence**
Actb FW	ACACCCGCCACCAGTTC
Actb RV	TACAGCCCGGGGAGCAT
Hba FW	GTGGATCCCGTCAACTTCAAG
Hba RV	CAAGGTCACCAGCAGGCAGT
Hbb-b1 FW	AGCTCCACTGTGACAAGCTG
Hbb-b1 RV	CCAGCACAATCACGATCATA
Hbb-bh1 FW	AGGCAGCTATCACAAGCATCTG
Hbb-bh1 RV	AACTTGTCAAAGAATCTCTGAGTCCA
Hbb-y FW	ACAGCTTTGGGAACTTGTCCTC
Hbb-y RV	TCTCCAAAAGCAGTCAGCACC
c-kit FW	AGCAGATCTCGGACAGCACC
c-kit RV	TGCAGTTTGCCAAGTTGGAG

Supplemental Materials and Methods

HUDEP2 cell culture

HUDEP2 cells were maintained in StemSpan SFEM (StemCell Technologies, #09650) supplemented with 50ng/ml recombinant human stem cell factor (hSCF, Peprotech, #300-07), 1uM dexamethasone (Sigma, #D4902), 1ug/ml doxycycline (Sigma, #D9891), 3IU/ml erythropoietin (EPO, Amgen, #55513-144-10), and 1% penicillin/streptomycin (ThermoFisher Scientific, #15140122) at a density of 100,000 – 1,000,000 cells/ml. Cells were differentiated for 5-7 days in Iscove's Modified Dulbecco's Medium (IMDM, Mediatech, #MT10016CV) supplemented with 50ng/ml hSCF, 3IU/ml EPO, 1ug/ml doxycycline, 5% fetal bovine serum, 320 ug/ml holo-transferrin (Sigma, #T4132), 10ug/ml insulin (Sigma, #I9278), and 10ug/ml heparin (Sigma, #H3149) at a density of 500,000-2,000,000 cells/ml. Cells were cultured at 37C in 5% CO2.

Primary human CD34+ cell culture

Primary human CD34+ hematopoietic stem and progenitor cells (HSPCs) were obtained from the Fred Hutch Cooperative Center of Excellence in Hematology and cultured in a 3-phase erythroid differentiation system.[8] The base culture medium consists of IMDM, 5% human AB serum (Sigma, #H4522), 10ug/ml heparin, 3IU/ml EPO, and 2% penicillin/streptomycin. During phase I (day 0 to day 6), base medium was supplemented with 100ng/ml hSCF, 1ng/ml IL-3 (Peprotech, #200-03), and 200ug/ml holo-transferrin. During phase II (day 7 to day 11), base medium was supplemented with 100ng/ml hSCF and 200ug/ml holo-transferrin. Cells were maintained at a density of 100,000-1,000,000 cell/ml during phases I and II. During phase III (day 12 to day 21), base medium was supplemented with 1mg/ml holo-transferrin and cell density was adjusted to 1,000,000-5,000,000 cells/ml. Cells were cultured at 37C in 5% CO2.

CRISPR/Cas9-based targeting in HUDEP2 cells

sgRNAs were subcloned into a LRG2.1T lentiviral vector which contains a GFP reporter.[9] HUDEP2 cells with stably integrated Cas9 were transduced with sgRNA lentivirus via spinfection: 400,000 cells were spun at 2200rpm at room temperature for 90 minutes with 8ug/ml polybrene (Sigma, #H9268), 10mM HEPES (ThermoFisher Scientific, #J63578.AP), and viral supernatant. 48 hours post transduction, GFP positive cells were sorted via fluorescence activated cell sorting (FACS).

RNP electroporation

RNP complexes were assembled by incubating 100pmol of sgRNA (Synthego) with 50pmol of spCas9 (IDT) for 10minutes at room temperature. RNP complexes were electroporated into 50,000-100,000 CD34+ HSPCs at least 6 hours post thaw using the P3 Primary Cell 4D-Nucleofector X Kit (Lonza, #V4XP-3032) and the Amaxa Nucleofector (Lonza) with program DZ-100. Post-electroporation, cells were transferred to 1ml of phase I CD34+ media. The next morning, cells underwent full media change.

Plasmids

spCas9 was subcloned into a lentiviral EFS-Cas9-P2A-puro vector and sgRNAs were subcloned into a lentiviral U6-sgRNA-EFS-GFP vector. RBM12 and Luciferase cDNAs were subcloned into a lentiviral EF1a-cDNA-ERLBD-V5-IRES-BSD vector. CPNE1 cDNA was subcloned into a lentiviral EFS-CPNE1-3HA-P2A-mCherry vector.

RT-qPCR

RNA was extracted from cells using RNEasy Mini Kit (Qiagen, #74106) including an on-column DNase digestion using RNase-free DNase set (Qiagen, #79254). 100-200ng RNA was reverse-transcribed to cDNA via iScript Reverse Transcription Supermix (BioRad, #1708841). Quantitative PCR was performed with 2ng equivalent of RNA to cDNA, .45uM each FW and RV primers and 2x Power SYBR Green Master Mix (Life Technologies, #4367660).

Western Blotting

2 million cells were lysed in 50ul RIPA buffer (150mM NaCl, 1% IGEPPAL, 0.5% Na-Deoxycholate, 0.1% SDS, 25mM Tris pH 7.4) for 30 minutes on ice and then subsequently sonicated for 5 cycles at 30sec on/off “easy” mode cycles on the BioRuptor® Pico (Diagenode, Denville, NJ). Samples were spun down at 20,000xg for 10min and supernatant (lysate) were collected. 10ug of lysates were boiled at 95C for 10min in protein sample loading buffer (LI-COR, #928-40004) and resolved on NuPAGE Bis-Tris 4-12% polyacrylamide gels (ThermoFisher Scientific, #NP0322) using the MES running buffer (ThermoFisher Scientific, #NP0002). Proteins were then transferred onto Immobilon-FL PVDF membranes (Millipore Sigma, #IFPL00010) at 30V overnight in 4C. Total proteins were visualized with the Revert 700 total protein stain (LI-COR, #927-11011). After stain removal, membranes were blocked in Odyssey blocking buffer (LI-COR, #927-50000) for 1 hour at room temperature. Membranes were then incubated with primary antibody diluted in blocking buffer overnight at 4C. After 3 rounds of washes in Tris-buffered saline with Tween-20 (TBS-T), membranes were incubated in fluorescent secondary antibody for 1 hour at room temperature in the dark. Membranes underwent 3 rounds of washes in TBS-T and were then imaged on the Odyssey imaging system (LI-COR). Primary antibodies used were: RBM12 (1:2000, Abcam, #ab72319), GAPDH (1:1000, Santa Cruz Biotech, #sc-32233), HBG (1:5000, Novus Biologicals, #NB110-41084), HBB (1:4000, Santa Cruz Biotech, #sc-21757), BCL11A (1:1000, Cell Signaling Tech, #75432S), ZBTB7A/LRF (1:1000, ThermoFisher Scientific, #14-3309-82), V5 (1:1000, ThermoFisher Scientific, #MA5-15253). Fluorescent secondary antibodies used were: Mouse IRDye 800CW (1:10000, LI-COR, #926-32212), Mouse IRDye 680CW (1:10000, LI-COR, #926-68072), Rabbit IRDye 800CW (1:10000, LI-COR, #926-32213), Rabbit IRDye 680CW (1:10000, LI-COR, #926-68074), Sheep/Goat IRDye 800CW (1:15000, LI-COR, #926-32214), Hamster IRDye 680CW (1:10000, Rockland, #620-144-440).

HPLC

Hemolysates were prepped via osmotic lysis. 1 million cells were lysed in 100ul MilliQ water, spun down (20,000xg, 10min), and the supernatant were collected. Hemolysates were analyzed by cation-exchange high performance liquid chromatography (HPLC) with a Hitachi D-7000 Series (Hitachi Instruments Inc., San Jose, CA) and a weak cation-exchange column (Poly CAT A: 35mmx4.6mm, Poly LC, Inc., Columbia, MD). Hemoglobin isotype peaks were eluted with a linear gradient of phase B (20mM Bis-Tris, 2mM KCN, 0.2M NaCl, pH 6.55) from 0% to 90% at phase A_{410nm} (20mM Bis-Tris, 2mM KCN, pH 6.95). Reference isotypes were hemolysates from normal human cord blood samples and a commercial standard containing similar levels of HbF, A, S, and C (Helena Laboratories, Beaumont, TX).

Flow cytometry

250,000 cells were fixed in 0.05% glutaraldehyde (Sigma, #G6257) for 10 minutes at room temperature and then subsequently washed 3 times in 0.1% BSA/PBS (Sigma, #A7906). Cells were blocked with BSA/PBS for 10 minutes before staining with primary anti-CD71 PE (1:100, BioLegend, #334106) and anti-CD235ab PECy7 (1:200, BioLegend, #306620) antibodies for 20 minutes in the dark at room temperature. Samples were rinsed once and permeabilized with 0.1% triton for 5 minutes, rinsed, and stained with anti-HbF antibody (1:500, Novus Biologicals, #NB110-41084)

conjugated to AF647 for 30 minutes at room temperature in the dark. Samples were then rinsed twice and resuspended in 300ul of 0.1%BSA/PBS in FACS tubes for analysis on the FACSCanto cytometer (BD Biosciences).

May-Grünwald Giemsa stain

200,000 cells were spun onto poly-L-lysine coated microscope slides (Sigma, #P0425) at 1000rpm for 5 minutes. After drying at room temperature for 5 minutes, slides were stained with May-Grunwald stain (Sigma, #MG1L) for 5 minutes, rinsed in distilled (DI) water, then stained with 1:20 diluted Giemsa stain (Sigma, #GS-500) for 15 minutes, followed by DI water rinsing. After drying, samples were sealed with a coverslip and Cytoseal 60 (ThermoFisher Scientific, #8310-4). Images were captured by bright field microscopy using an Olympus BX60 microscope fitted with an Infinity Lite B camera (Olympus), and the coupled Image Capture software.

Immunofluorescence Staining

500,000 cells were fixed with 2% formaldehyde in PBS at room temperature for 10 minutes. Cells were then washed with PBS for a total of 3 washes and subsequently permeabilized with 0.1% triton in 3%BSA/PBS (w/v). After washing, samples were blocked in 3%BSA/PBS for 1 hour at room temperature and then stained with anti-RBM12 (1:100, Santa Cruz, #sc-514258) and anti-TUBB (1:1000, Abcam, #ab6046) overnight at 4C, washed, and then stained with anti-mouse IgG AF488 (1:1000, ThermoFisher Scientific, #A-21202) and anti-rabbit IgG PE (1:1000, Biolegend, #406421) for 1 hour and room temperature. Cells were then spun onto poly-L-lysine coated microscope slides and mounted with ProLong Diamond Antifade Mountant with DAPI (ThermoFisher Scientific, #P36966) and a coverslip. Images were captured on the EVOS FL auto imaging system (ThermoFisher Scientific).

RNA-Seq

RNA samples were extracted as described above. 100ng of RNA was used for library preparation using the TruSeq Stranded mRNA protocol (Illumina). Libraries were sequenced on the Illumina NextSeq2000 instrument as paired-end 150bp reads. Reads were processed using the ENCODE3 long RNA-Seq pipeline (https://www.encodeproject.org/pipelines/ENCPL002LPE/). Differentially expressed genes were identified using the DESeq2 R package. Differentially spliced genes were identified using the MAJIQ software.[10]

LC-MS/MS

Subcellular fractionation was performed as previously described.[11] 4 million cells were lysed in 100ul cellular extraction buffer (20mM HEPES pH 7.9, 150mM NaCl, 0.5mM $MgCl_2$, 0.5% IGEPAL, 10% glycerol, protease inhibitor, 1mM PMSF) and incubated on ice for 5 minutes. Samples were spun down on a microcentrifuge for 20 minutes at max speed at 4C. Supernatant was harvested as the cytoplasmic fraction; the nuclei pellet was rinsed in cellular extraction buffer for a total of 3 times and lysed with 50ul of 50mM NH_4HCO_3 (plus protease inhibitors and PMSF) via sonication at 30s on/off for 45 cycles at 100% amplitude. Cytoplasmic and nuclear lysates were then prepared for LC-MS/MS analysis by running 25ug of protein on NuPAGE Bis Tris 10% polyacrylamide gel for 0.5cm from the well using the MES running buffer. The gel was then fixed in 50% methanol/10% acetic acid in water (v/v) for 10 minutes followed by staining and de-staining with the Novex Colloidal Blue Staining Kit (ThermoFisher Scientific, #LC6025) according to the manufacturer's protocol. The stained gel regions were excised, reduced with TCEP, alkylated with iodoacetamide, and digested with trypsin. Tryptic peptides of each sample were labeled with a specific TMTpro 16-plex reagent. Samples were mixed 1:1 and subjected to high pH fractionation into 14 fractions (5, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 50, and 80% acetonitrile). Fractions

were combined to yield a final total of 6 fractions. A total of 6 LC-MS/MS sample runs were performed on the Thermo Q Exactive HF mass spectrometer using a 175-minute LC-MS/MS method.

eCLIP-Seq

eCLIP-Seq was performed following the eCLIP standard operating procedure from ENCODE and as described in Van Nostrand et. al.[12] Samples were UV-crosslinked (254nm, 400mJ/cm^3) at a density of 20 million cells/ml in PBS followed by lysis in 1ml lysis iCLIP lysis buffer (50mM Tris-HCl pH 7.4, 100mM NaCl, 1% IGEPAL, 0.1% SDS, 0.5% sodium deoxycholate, protease inhibitor cocktail III [EMD Millipore, #539134]). RNA protected by the RBP-RNA complexes were fragmented via controlled RNase I digestion (Invitrogen, #AM2294) and then subject to immunoprecipitation with anti-RBM12 antibody (10ug, Santa Cruz, #sc-514259) that has been pre-coupled with anti-mouse IgG magnetic beads (ThermoFisher Scientific, #11201D) overnight at 4C. Inputs were collected after immunoprecipitation before subjecting samples to a series of stringent washes with high salt (50mM Tris-HCl pH 7.4, 1M NaCl, 1mM EDTA, 1% IGEPAL, 0.1% SDS, 0.5% sodium deoxycholate) and wash buffers (20mM Tris-HCl pH 7.4, 10mM $MgCl_2$ 0.2% Tween-20). Samples were equilibrated in FastAP buffer (10mM Tris-HCl pH 7.5, 5mM $MgCl_2$, 100mM KCl, 0.2% Triton) before FastAP (ThermoFisher Scientific, #EF0651) and subsequent T4 PNK treatments (NEB, #M0201S). Samples were then washed with wash and high salt buffers before equilibration in RNA ligase buffer (50mM Tris-HCl pH 7.5, 10mM $MgCl_2$). RNA barcode adapters were ligated onto samples using high concentration RNA Ligase (NEB, #M0437M). After washing with wash and high salt buffers, samples (plus the reserved inputs) were run on a standard protein gel and transferred to a nitrocellulose membrane, where the RNA was isolated from a region encompassing 75kDa above the expected size of RBP. RNA was purified via urea/proteinase K digestion followed by pheno/chloroform extraction and column cleanup (Zymo, #R1016). Inputs were then FastAP and PNK treated along with RNA barcode adaptor ligated. All samples were then reverse transcribed to cDNA (Agilent, #600107) with the addition of a DNA adaptor (containing a randomer of 10 random bases) ligation. cDNA was cleaned via ExoSAP-IT (Applied Biosystems, #78201) and Silane treatments (ThermoFisher Scientific, #37002D) and an aliquot of it was subject to qPCR for PCR cycle determination. Libraries were then PCR amplified (Q5, NEB, #M0494S) and size selected via DNA gel electrophoresis extraction. Libraries were sequenced on the Illumina NextSeq2000 instrument as paired-end 50bp reads. Reads were processed using the ENCODE3 eCLIP-Seq standard analysis pipeline (https://www.encodeproject.org/documents/3b1b2762-269a-4978-902e-0e1f91615782/@@download/attachment/eCLIP_analysisSOP_v2.0.pdf). eCLIP-Seq peaks were analyzed using the RNA Centric Annotation System (RCAS) R package.[13]

Supplemental References

1. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580–5.
2. Bagger FO, Kinalis S, Rapin N. BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. *Nucleic Acids Res*. 2019;47(D1):D881–D885.
3. An X, Schulz VP, Li J, et al. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood*. 2014;123(22):3466–3477.
4. Huang P, Keller CA, Giardine B, et al. Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes and Development*. 2017;31(16):1704–1713.
5. Lessard S, Beaudoin M, Orkin SH, Bauer DE, Lettre G. 14q32 and let-7 microRNAs regulate transcriptional networks in fetal and adult human erythroblasts. *Human Molecular Genetics*. 2018;27(8):1411–1420.
6. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*. 2007;23(9):1073–9.
7. Hu J, Qian H, Xue Y, Fu X-D. PTB/nPTB: master regulators of neuronal fate in mammals. *Biophys Rep*. 2018;4(4):204–214.
8. Hu J, Liu J, Xue F, et al. Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood*. 2013;121(16):3246–3253.
9. Grevet JD, Lan X, Hamagami N, et al. Domain-focused CRISPR screen identifies HRI as a fetal hemoglobin regulator in human erythroid cells. *Science (1979)*. 2018;361(6399):285–290.
10. Vaquero-Garcia J, Barrera A, Gazzara MR, et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. 2016;
11. Agosto LM, Gazzara MR, Radens CM, et al. Deep profiling and custom databases improve detection of proteoforms generated by alternative splicing. *Genome Research*. 2019;29(12):2046–2055.
12. van Nostrand EL, Pratt GA, Shishkin AA, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature Methods*. 2016;13(6):508–514.
13. Uyar B, Yusuf D, Wurmus R, et al. RCAS: An RNA centric annotation system for transcriptome-wide regions of interest. *Nucleic Acids Research*. 2017;45(10):.