

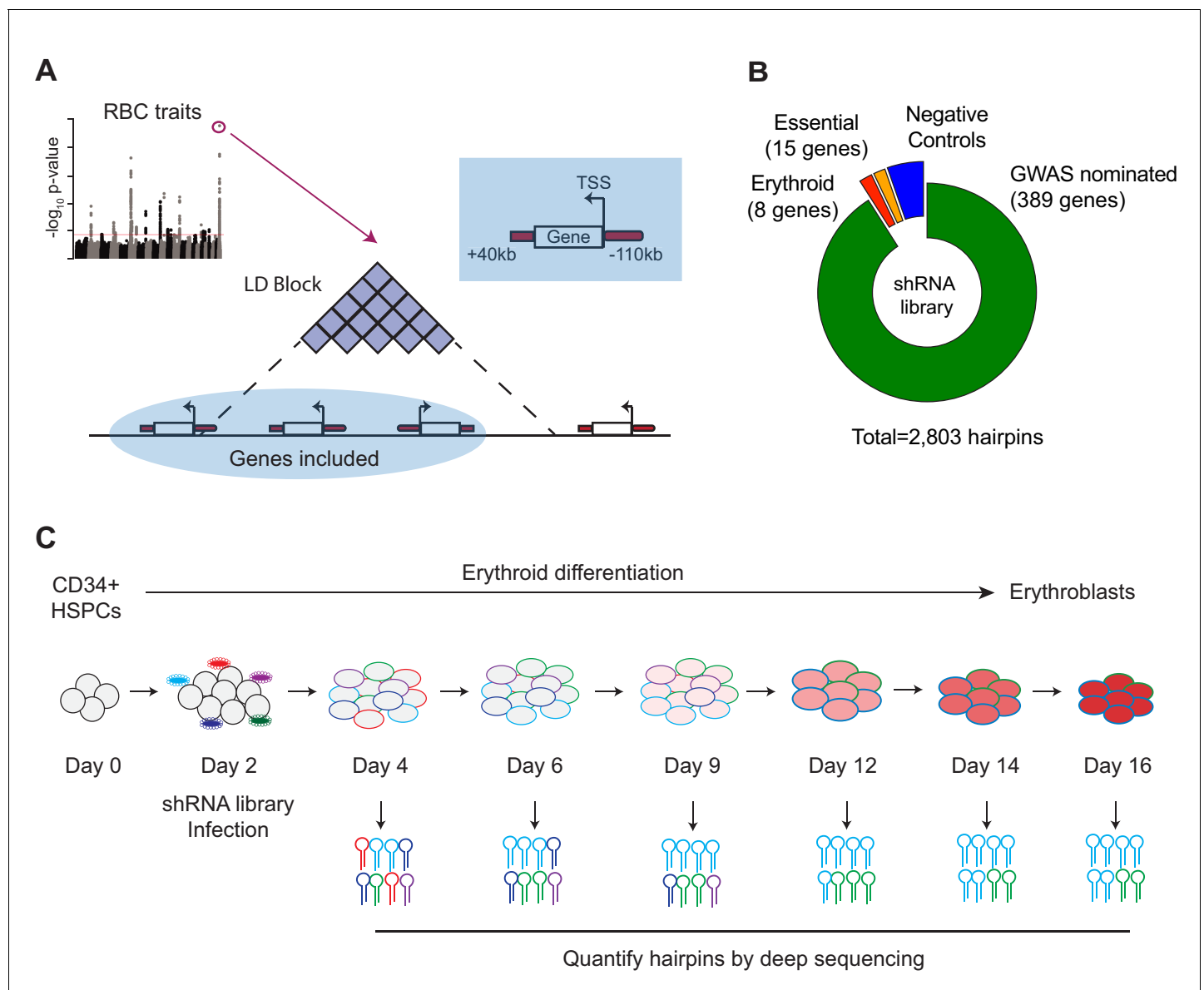


---

## Figures and figure supplements

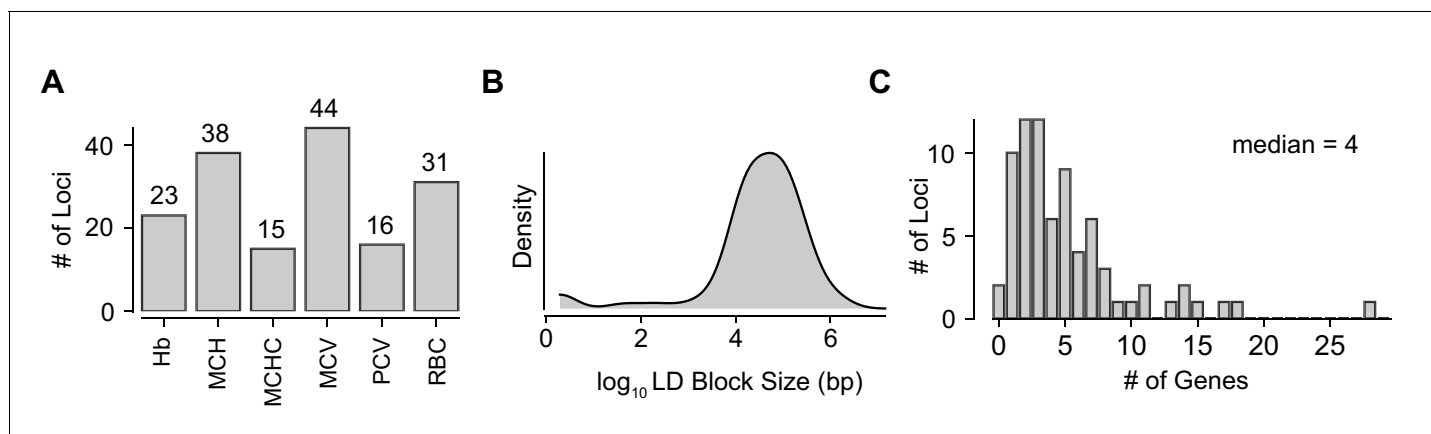
Gene-centric functional dissection of human genetic variation uncovers regulators of hematopoiesis

**Satish K Nandakumar *et al***



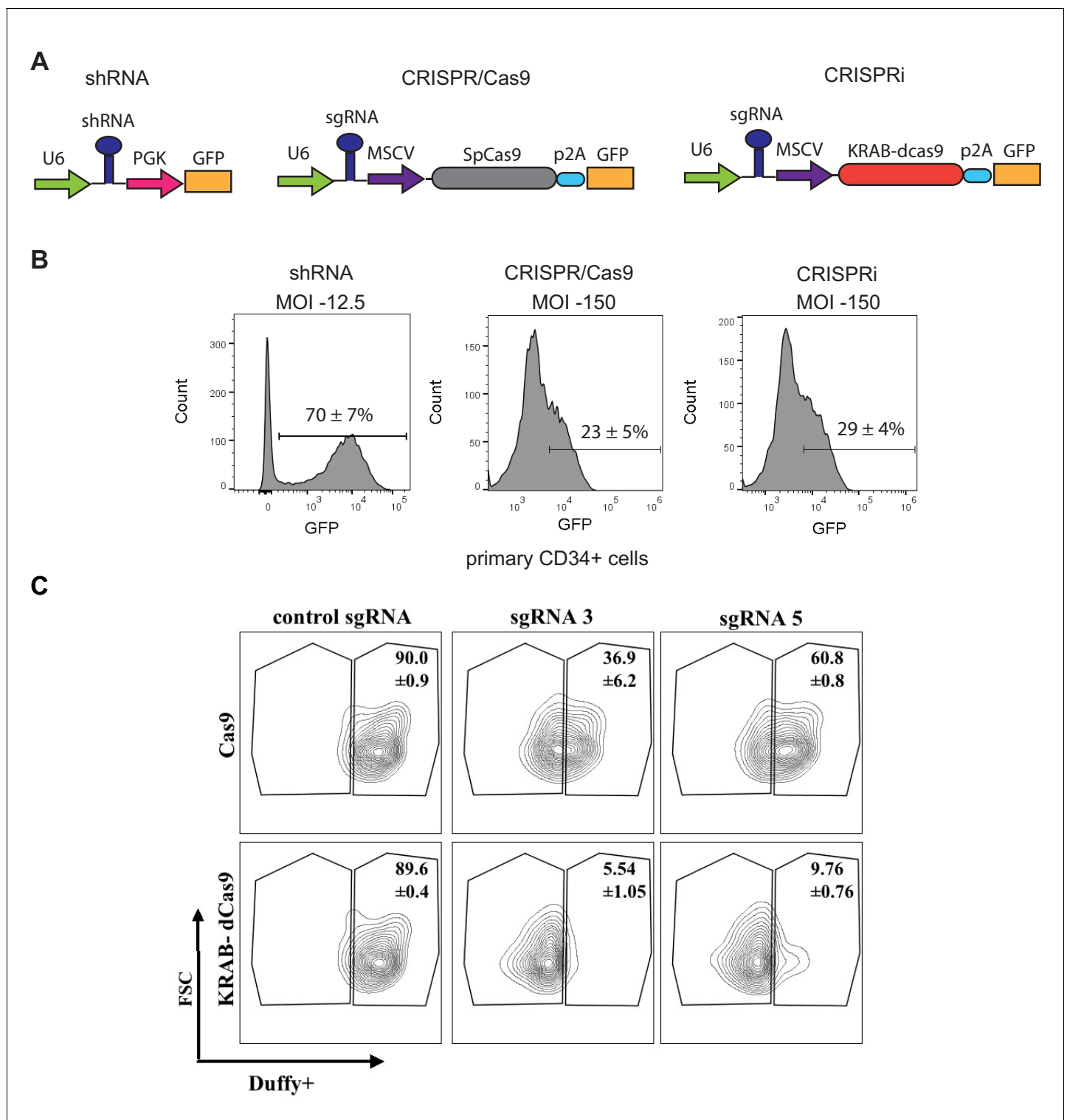
**Figure 1.** Design and Execution of an shRNA Screen Using Blood Cell Trait GWAS Hits to Identify Genetic Actors in Erythropoiesis. **(A)** Overview of shRNA library design. 75 loci associated with red blood cell traits (*van der Harst et al., 2012*) were used as the basis to calculate 75 genomic windows of LD 0.8 or greater from the sentinel SNP. Genes with a start site within 110 kb or end site within 40 kb of the LD-defined genomic windows were chosen as candidates to target in the screen. **(B)** Compositional makeup of the library, depicted as number of genes and number of hairpins for each of the four included subcategories; GWAS-nominated genes, erythroid genes, essential genes, and negative control genes (**Figure 1—source data 2**). **(C)** Primary CD34<sup>+</sup> hematopoietic stem and progenitor cells (HSPCs) isolated from three independent donors were cultured for a period of 16 days in erythroid differentiation conditions. At day 2, cells were infected with the shRNA library, and the abundances of each shRNA were measured at days 4, 6, 9, 12, 14, and 16 using deep sequencing.

DOI: <https://doi.org/10.7554/eLife.44080.002>



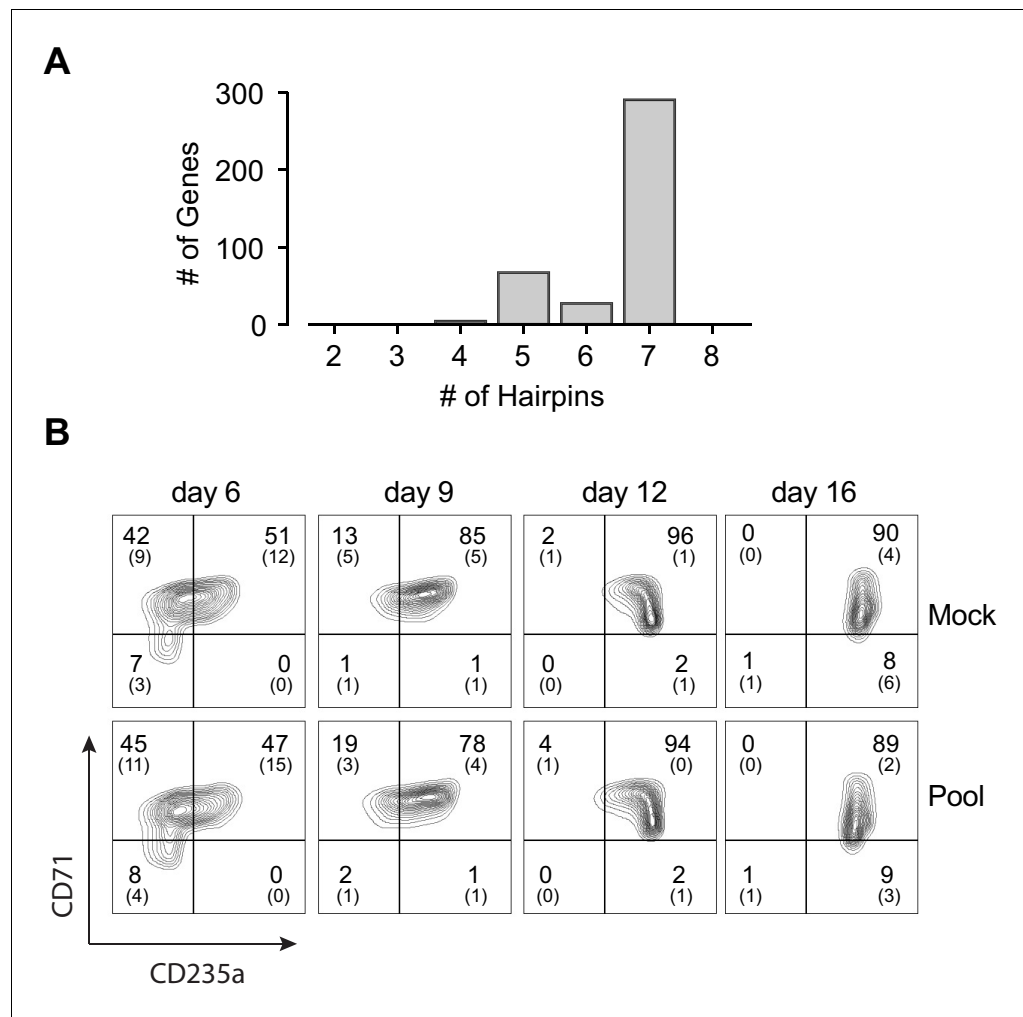
**Figure 1—figure supplement 1.** Characteristics of GWAS Loci and Gene Selection for Pooled Screen. (A) Counts of loci from among the original 75 annotated with linkage to each of the six RBC traits, hemoglobin (Hb), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), packed cell volume (PCV), and red blood cell count (RBC). Some loci were associated with multiple traits. Detailed information on each loci available in **Figure 1—source data 1**. (B) Kernel density plot showing the  $\log_{10}$  sizes in bp of the LD-defined genomic windows used to find overlapping genes. (C) Histogram showing distribution of number of genes selected using the LD window method at each locus. A median of 4 genes were present at each.

DOI: <https://doi.org/10.7554/eLife.44080.003>



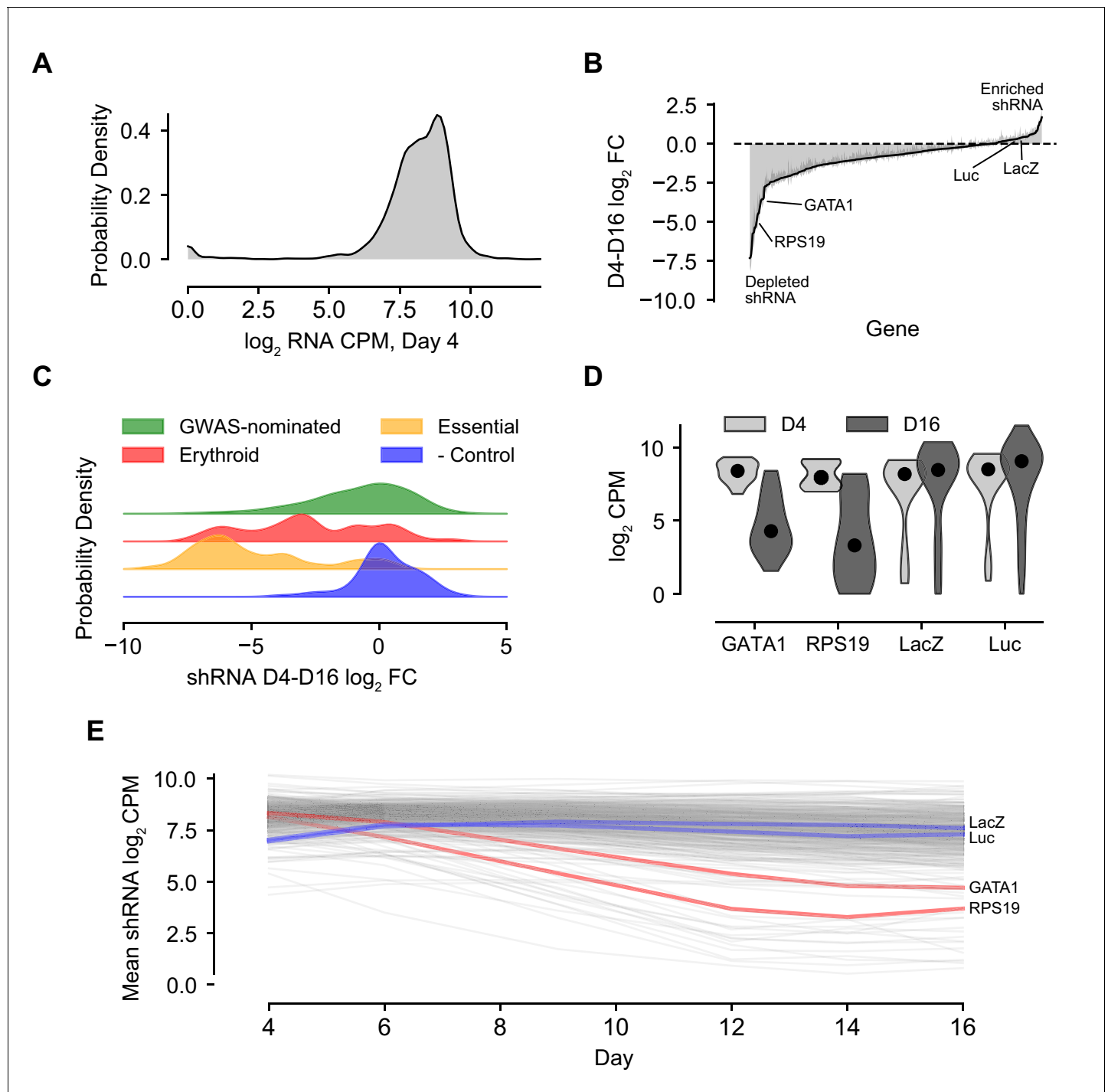
**Figure 1—figure supplement 2.** Feasibility of Loss of Function Approaches to Perform Pooled Screens in Primary Hematopoietic Stem and Progenitor Cells (HSPCs). (A) Schematic of the loss of function lentiviral constructs tested for pooled screens in primary CD34<sup>+</sup> cells. (B) FACS plots showing the proportion of infected GFP<sup>+</sup> cells 4 days after transduction with the respective constructs (MOI -multiplicity of infection). FACS analysis was performed in independent analyzers. (C) Efficient silencing of Duffy surface antigen in primary CD34<sup>+</sup> derived erythroid cells by targeting the promoter region using CRISPRi compared to CRISPR constructs.

DOI: <https://doi.org/10.7554/eLife.44080.004>



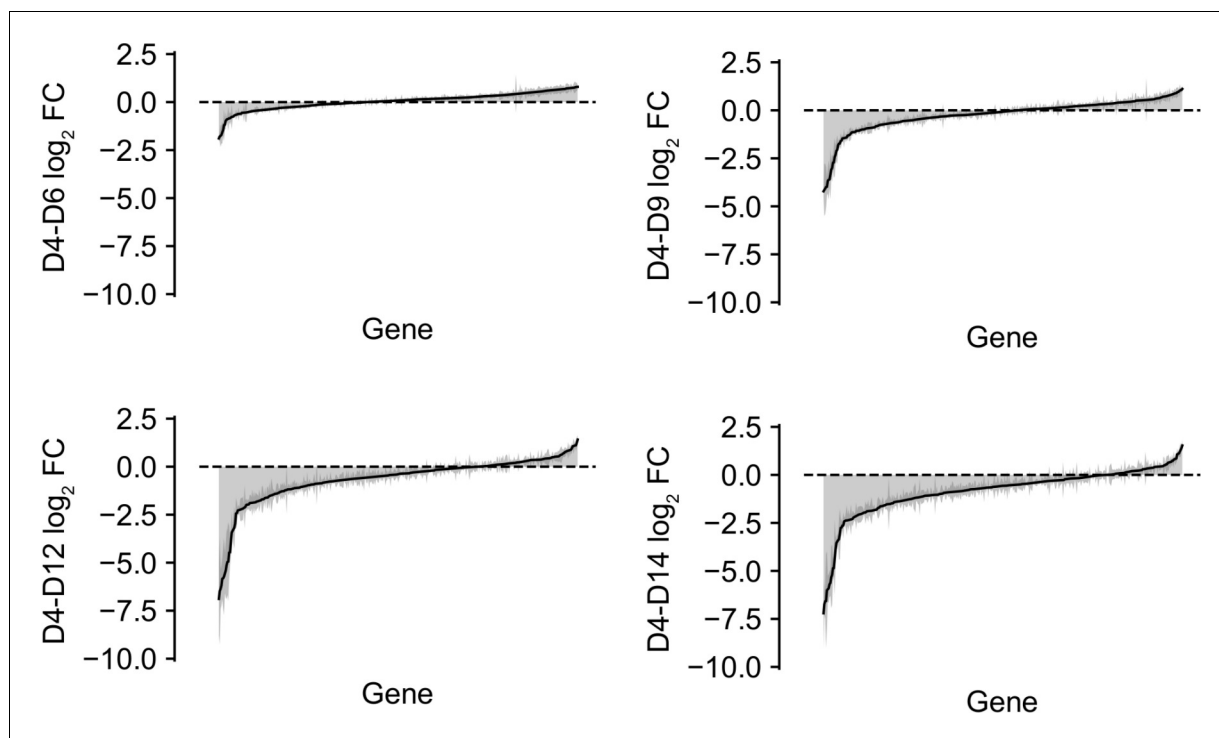
**Figure 1—figure supplement 3.** Pooled shRNA screen in primary HSPCs undergoing erythroid differentiation. (A) Histogram showing distribution of number of independent hairpins included in the library to target each of the candidate's genes. (B) Representative FACS plots of erythroid cell surface markers CD71 (transferrin receptor) and CD235a (Glycophorin A) expression at various time points during erythroid differentiation at which deep sequencing of shRNAs was performed. Percentages in each quadrant is represented as mean and standard deviation of 3 experiments from independent donors that were uninfected (Mock) or infected with the shRNA library (Pool).

DOI: <https://doi.org/10.7554/eLife.44080.005>



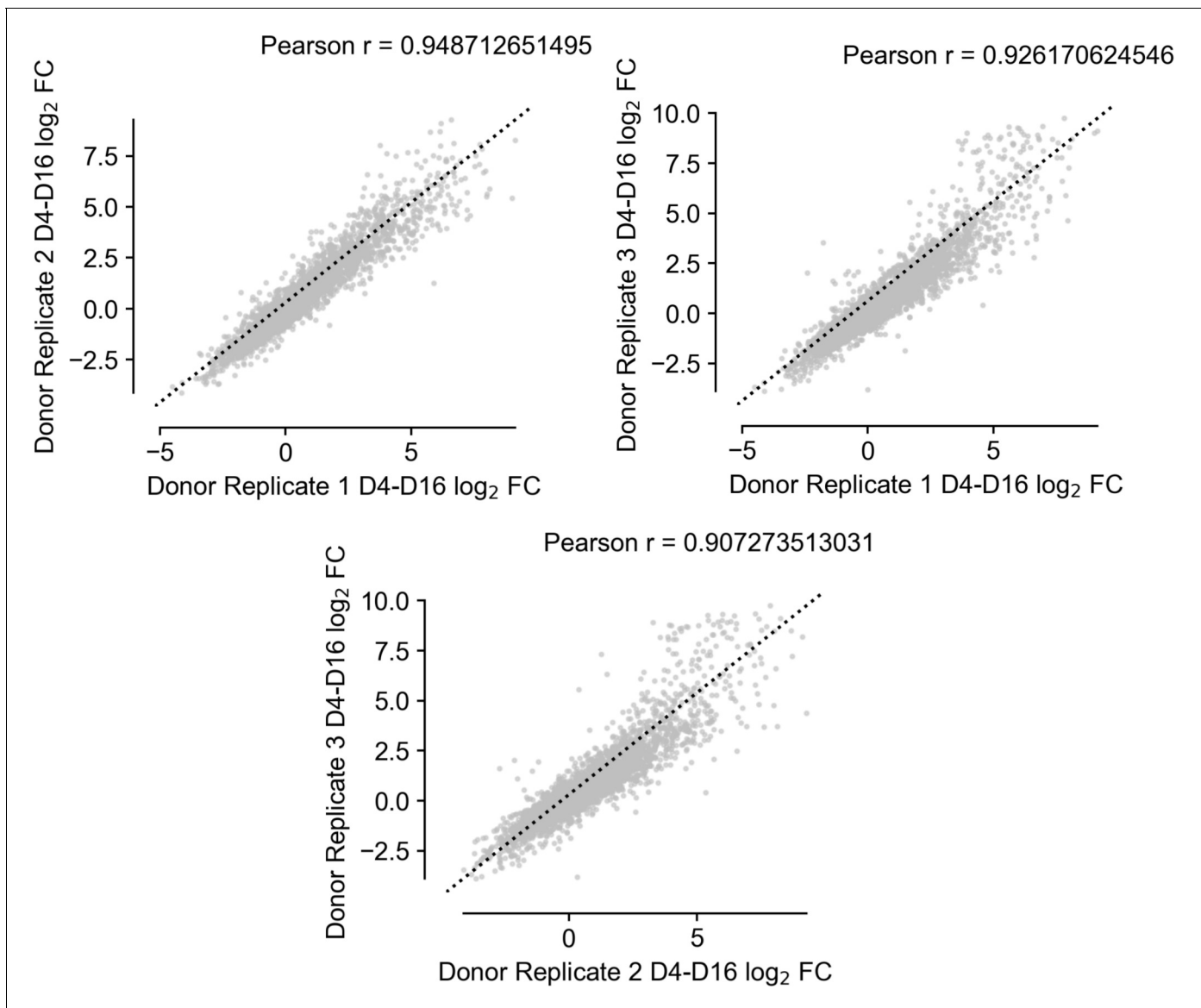
**Figure 2.** Summary Characterization of shRNA Screen Outcomes. (A) Kernel density plot showing library representation as log<sub>2</sub> shRNA CPM across all hairpins. (B) shRNA abundance log<sub>2</sub> fold changes from day 4 to day 16. Represented values are the mean of hairpin abundance log<sub>2</sub> fold changes across hairpins for each gene and two standard deviations. (C) Kernel density plots representing the day 4 to day 16 log<sub>2</sub> fold changes of hairpin abundances for each of the subcategories of the library, including GWAS-nominated genes, known erythroid essential genes, essential genes to cell viability, and orthogonal genes serving as negative controls. (D) Violin plot of day 4 and day 16 log<sub>2</sub> CPM for known actors GATA1 and RPS19 and negative controls LacZ and luciferase. (E) Log<sub>2</sub> hairpin counts averaged for known actors GATA1 and RPS19 as well as negative controls LacZ and luciferase across the course of the experiment. Gray lines depict the universe of all other gene traces in the library for context.

DOI: <https://doi.org/10.7554/eLife.44080.008>



**Figure 2—figure supplement 1.** shRNA abundance log<sub>2</sub> fold changes from day four to each of the other time points. Represented values are the mean of hairpin abundance log<sub>2</sub> fold changes across hairpins for each gene and two standard deviations.

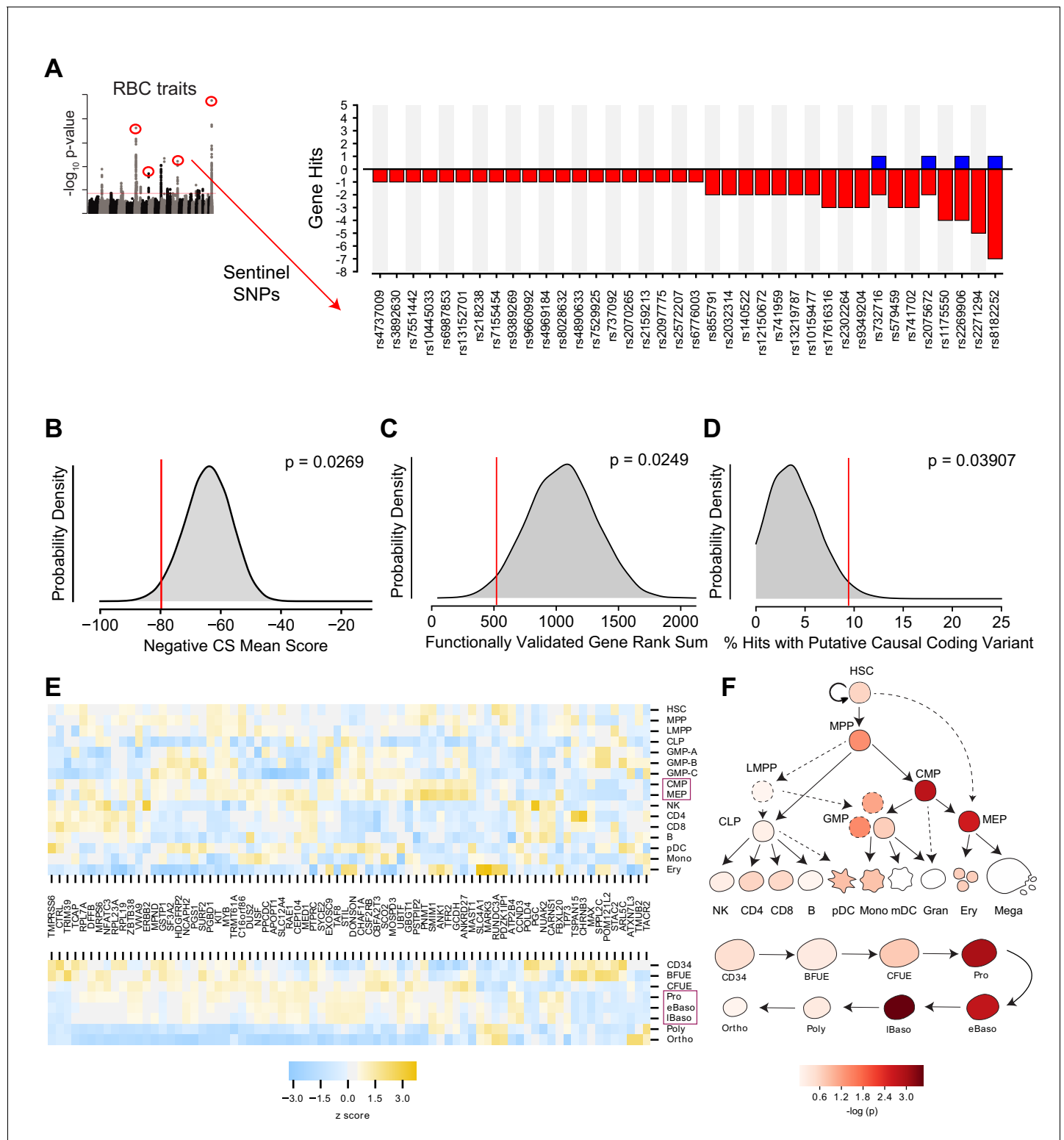
DOI: <https://doi.org/10.7554/eLife.44080.009>



**Figure 2—figure supplement 2.** Scatter plots showing agreement of replicate observations across independent CD34<sup>+</sup> donor populations.

DOI: <https://doi.org/10.7554/eLife.44080.010>



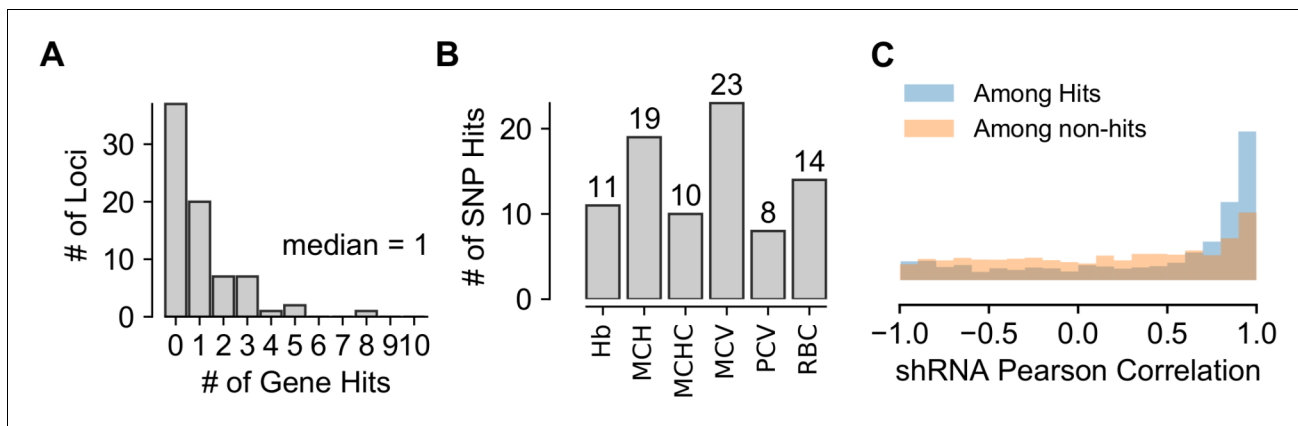


**Figure 3.** Statistical Modeling of Gene Effect Accounting for Off-target shRNA Confounders. (A) Bar graph showing the 38 of 75 loci in the screen with at least one corresponding statistically significant ( $FDR < 0.1$ ,  $\beta > 0.1$ ) gene effect causing either a positive or negative  $\log_2$  fold change in shRNA abundance. Statistical model output for each gene in screen available in **Figure 3—source data 1**. (B) Kernel density plot showing the expected distributions of K562 essentiality scores using permuted gene hit sets from the library. (C) Hairpin rank sums for permuted sets of 5 genes. The red line indicates the enriched rank sums for 5 ‘gold standard’ genes included in the library, *CCND3*, *SH2B3*, *MYB*, *KIT*, and *RBM38*, for each which a genetic basis of action has already been established. (D) Permuted distribution of % inclusion of predicted coding variants among the set of identified hits. (E) *Figure 3 continued on next page*

*Figure 3 continued*

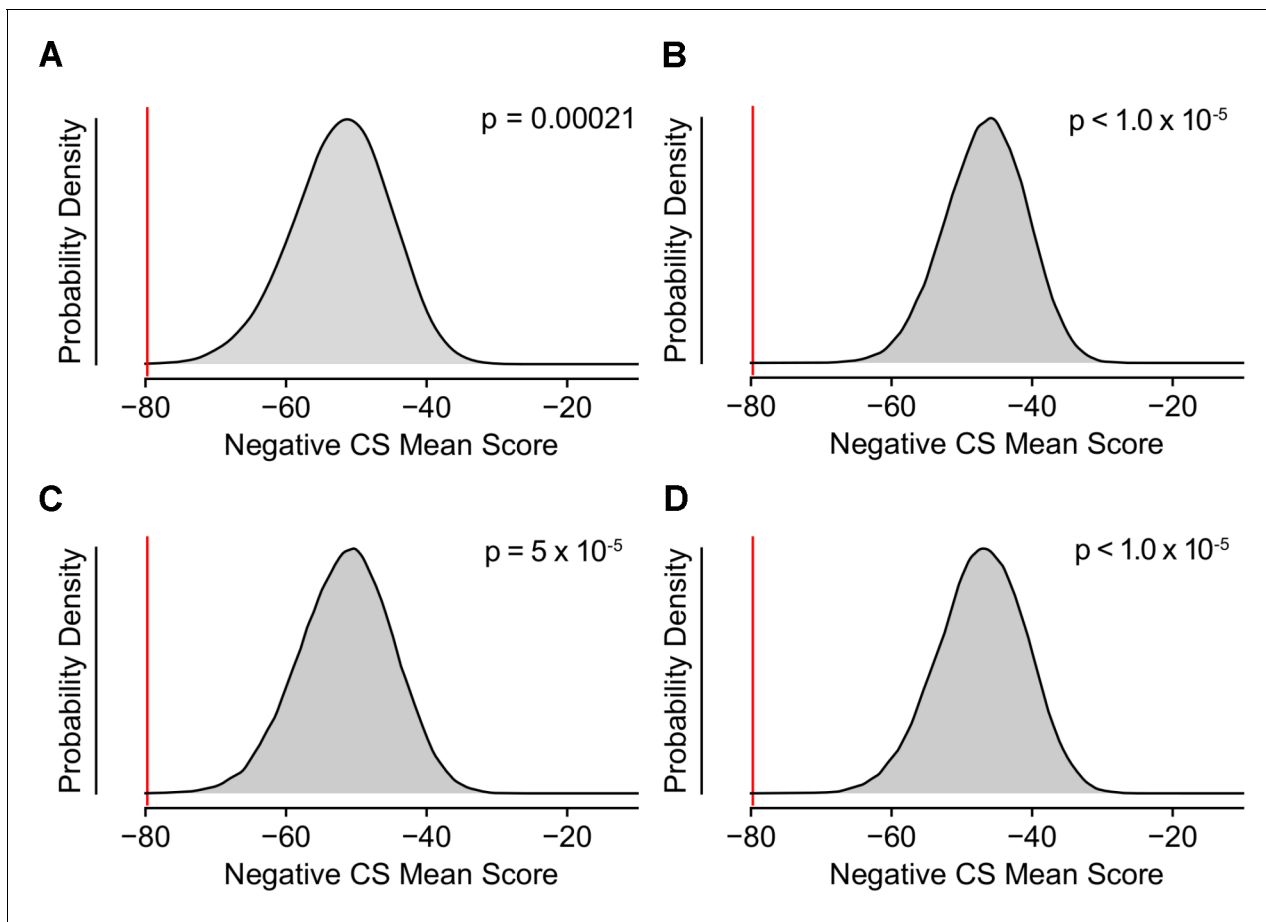
Heat map depicting strength of expression (as z scores within each gene) for each of the 77 identified hit genes across hematopoietic lineages (top) and throughout the specific stages of adult erythropoiesis (bottom). Purple boxes highlight the cell types that were enriched for expression of hit genes. (F) Calculated enrichment of the identified hit genes for expression across hematopoietic lineages (top) and throughout the specific stages of adult erythropoiesis (bottom). In both cases, cellular states corresponding to those along the erythropoietic lineage had elevated probability of expressing genes from the hit set as compared to other genes from the library.

DOI: <https://doi.org/10.7554/eLife.44080.011>



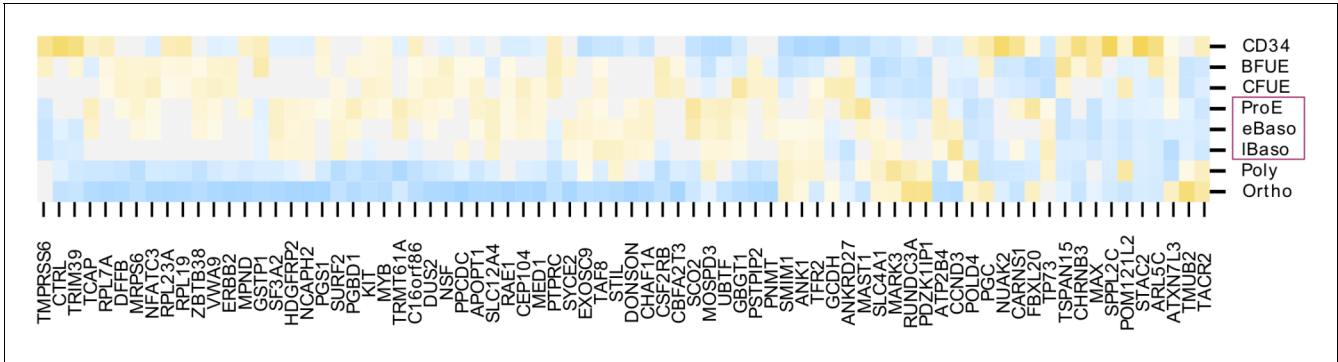
**Figure 3—figure supplement 1.** Additional Characterization of Modeling Outcomes. (A) Histogram showing the number of gene hits identified at each of the 40 loci with at least one significant gene effect detected. Statistical model output for each gene in screen available in **Figure 3—source data 1**. (B) Bar graph showing the number of gene hits identified for each of the six red blood cell traits used in the original GWAS to identify the studied loci. (C) Density-normalized histogram showing the Pearson correlation of hairpin measurements for both genes nominated as hits and genes not nominated as hits.

DOI: <https://doi.org/10.7554/eLife.44080.012>

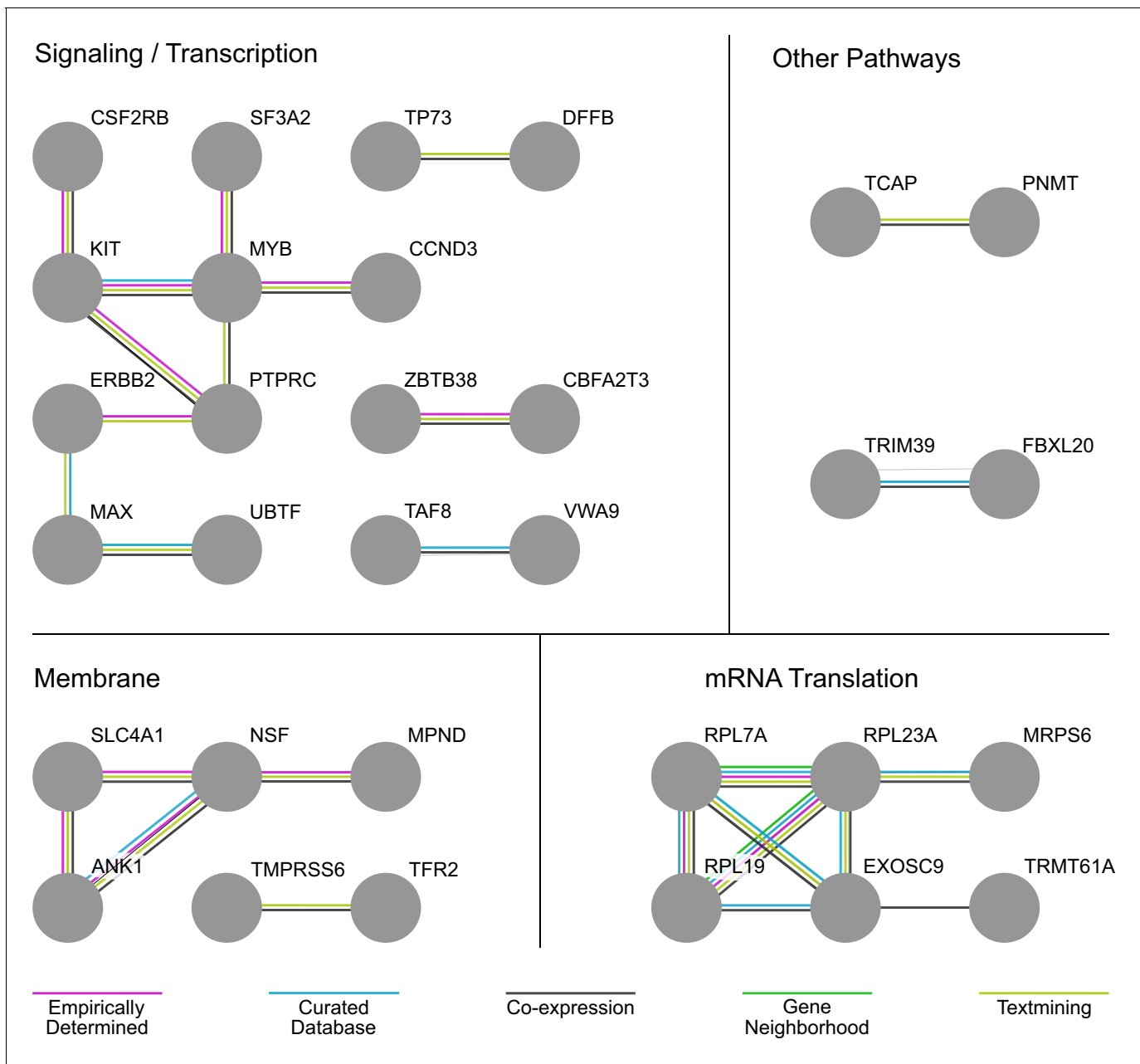


**Figure 3—figure supplement 2.** K562 Essentiality Scores Comparing Hit Genes vs. Genes Implicated by Other Traits. (A) Permuted enrichment of essentiality among the set of hit genes vs. randomly chosen sets of genes from the human genome. (B) Permuted enrichment of essentiality among the set of hit genes vs. genes implicated by a separate GWAS for LDL cholesterol levels. (C) Permuted enrichment of essentiality among the set of hit genes vs. genes implicated by a separate GWAS for HDL cholesterol levels. (D) Permuted enrichment of essentiality among the set of hit genes vs. genes implicated by a separate GWAS for blood triglyceride levels.

DOI: <https://doi.org/10.7554/eLife.44080.013>

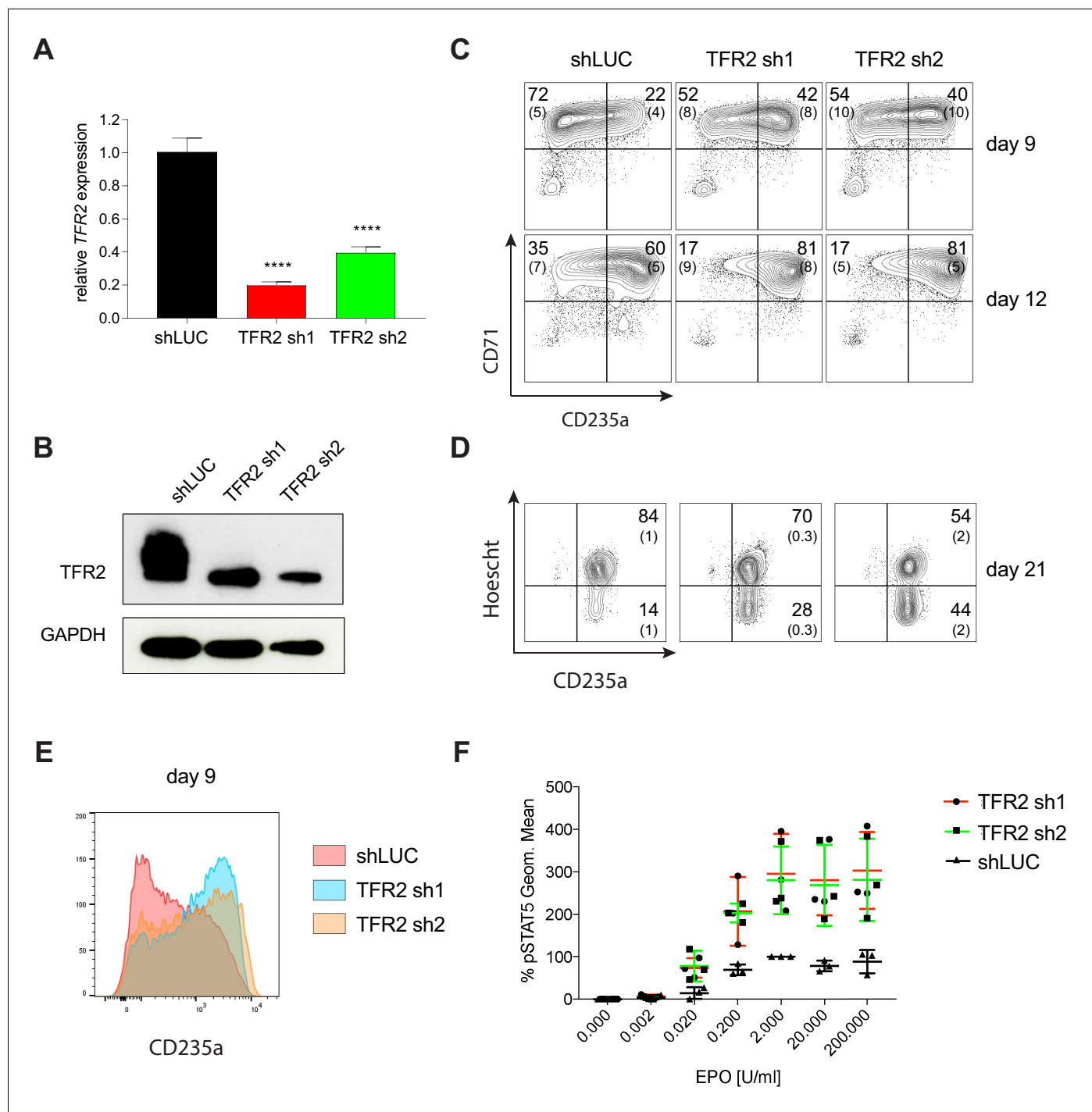


**Figure 3—figure supplement 3.** Heat map depicting strength of expression (as z scores within each gene) for each of the 77 identified hit genes throughout the specific stages of fetal erythropoiesis. Purple boxes highlight the cell types that were enriched for expression of hit genes.  
DOI: <https://doi.org/10.7554/eLife.44080.014>



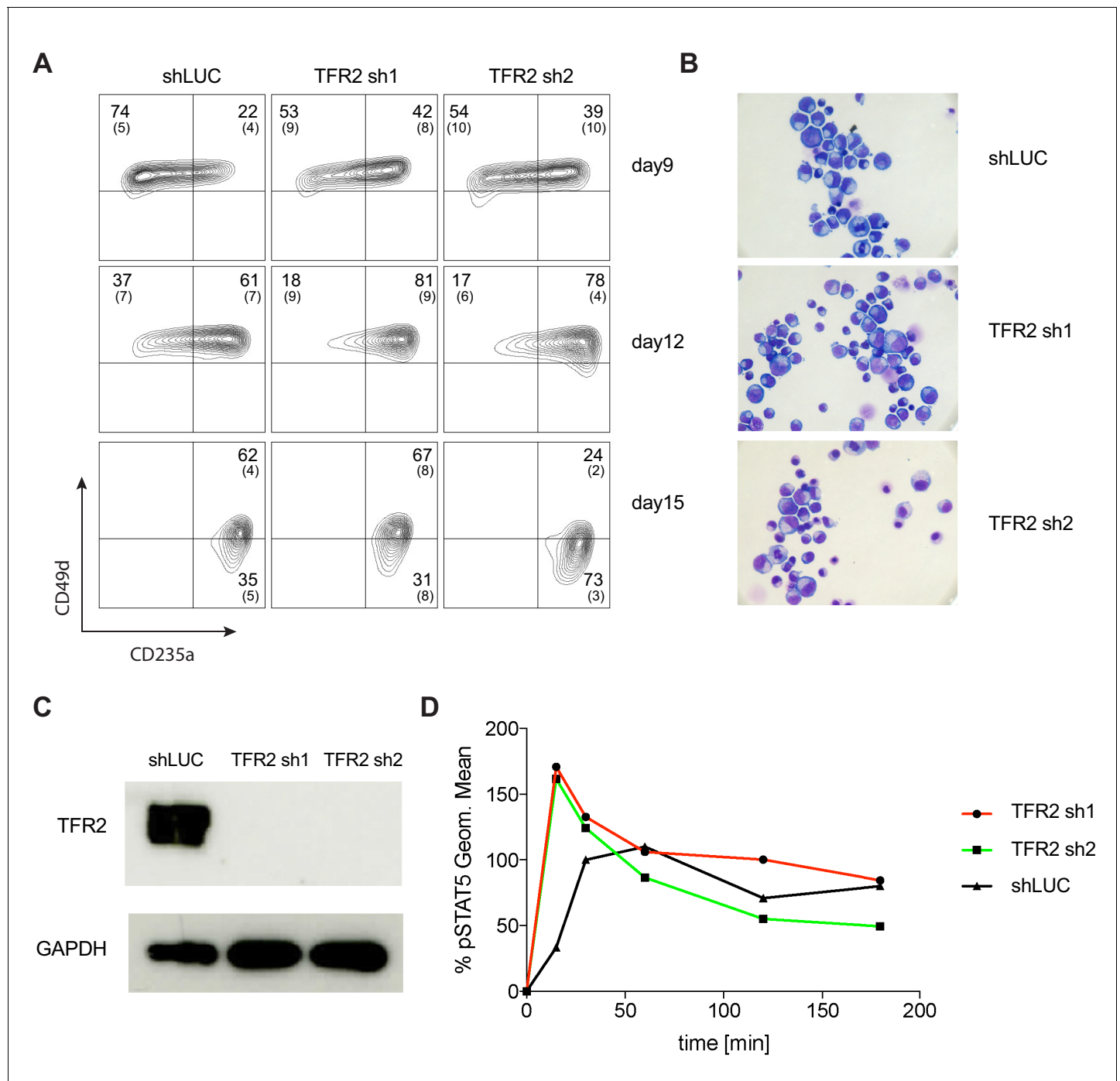
**Figure 4.** Analysis of Interactions Among Members of the Hit Set Identifies Signaling/Transcription, Membrane, and mRNA Translation-Related Subnetworks Important to Erythropoiesis. STRING interaction network analysis identifies signaling/transcription, membrane, and mRNA translation-related subnetworks important to erythropoiesis embedded in the genes identified in the screen hit set. Edges connecting the network are color-coded according to the evidence supporting the interaction. In STRING, this evidence can derive from empirical determination, curation in a database, co-expression of the respective gene nodes, genomic proximity, and text-mining of published literature.

DOI: <https://doi.org/10.7554/eLife.44080.016>



**Figure 5.** Transferrin receptor two is a Negative Regulator of Human Erythropoiesis. (A) Quantitative RT-PCR and (B) Western blot showing the expression of TFR2 in human CD34<sup>+</sup> cells five days post-infection with the respective lentiviral shRNAs targeting TFR2 (TFR2 sh1 and sh2) and a control luciferase gene (shLUC). (C) Representative FACS plots of erythroid cell surface markers CD71 (transferrin receptor) and CD235a (Glycophorin A) expression at various time points during erythroid differentiation. Percentages in each quadrant are represented as mean and standard deviation of 3 independent experiments (D) Hoechst staining showing more enucleated cells after TFR2 knockdown at day 21 of erythroid culture. (E) Representative histogram plots showing increased expression of CD235a (Glycophorin A) after TFR2 knockdown (F) Enhanced pSTAT5 response after TFR2 knockdown in UT7/EPO cells.

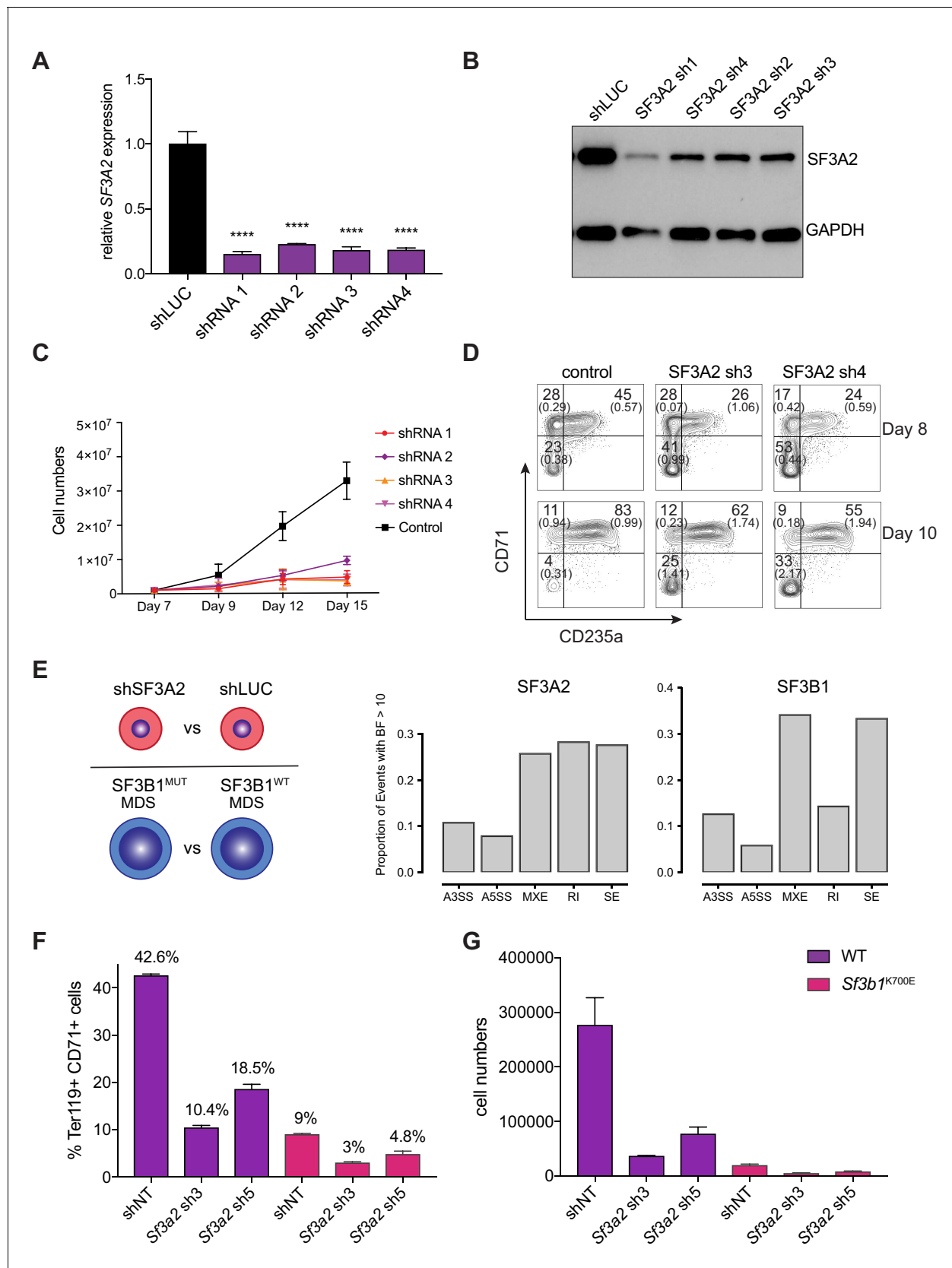
DOI: <https://doi.org/10.7554/eLife.44080.017>



**Figure 5—figure supplement 1.** Additional Analysis Showing Transferrin Receptor two is a Negative Regulator of Human Erythropoiesis. (A) Representative FACS plots of alternate erythroid cell surface markers CD49d ( $\alpha 4$  integrin) and CD235a (Glycophorin A) expression at various time points during erythroid differentiation. (B) May-Grunwald Giemsa staining showing more differentiated erythroid cells after TFR2 knockdown at day 18 of erythroid culture. (C) Western blot showing downregulation of TFR2 in UT7/EPO cells. (D) Time-dependent absolute value of Mean Fluorescence Intensity (MFI) of STAT5 in UT7/Epo cells after TFR2 knockdown.

DOI: <https://doi.org/10.7554/eLife.44080.018>



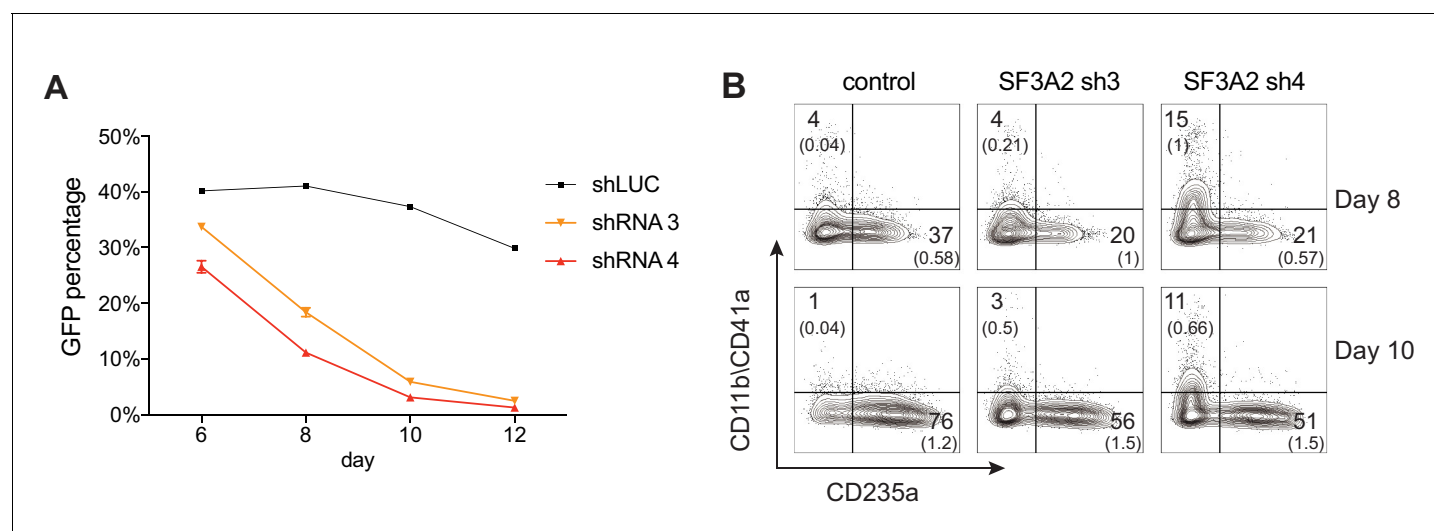


**Figure 6.** SF3A2 is a Key regulator of Human Erythropoiesis and Modulates Erythropoiesis Defects in a Murine Model of MDS. (A) Quantitative RT-PCR and (B) Western blot showing the expression of SF3A2 in human CD34<sup>+</sup> cells five days post-infection with the respective lentiviral shRNAs targeting Figure 6 continued on next page

## Figure 6 continued

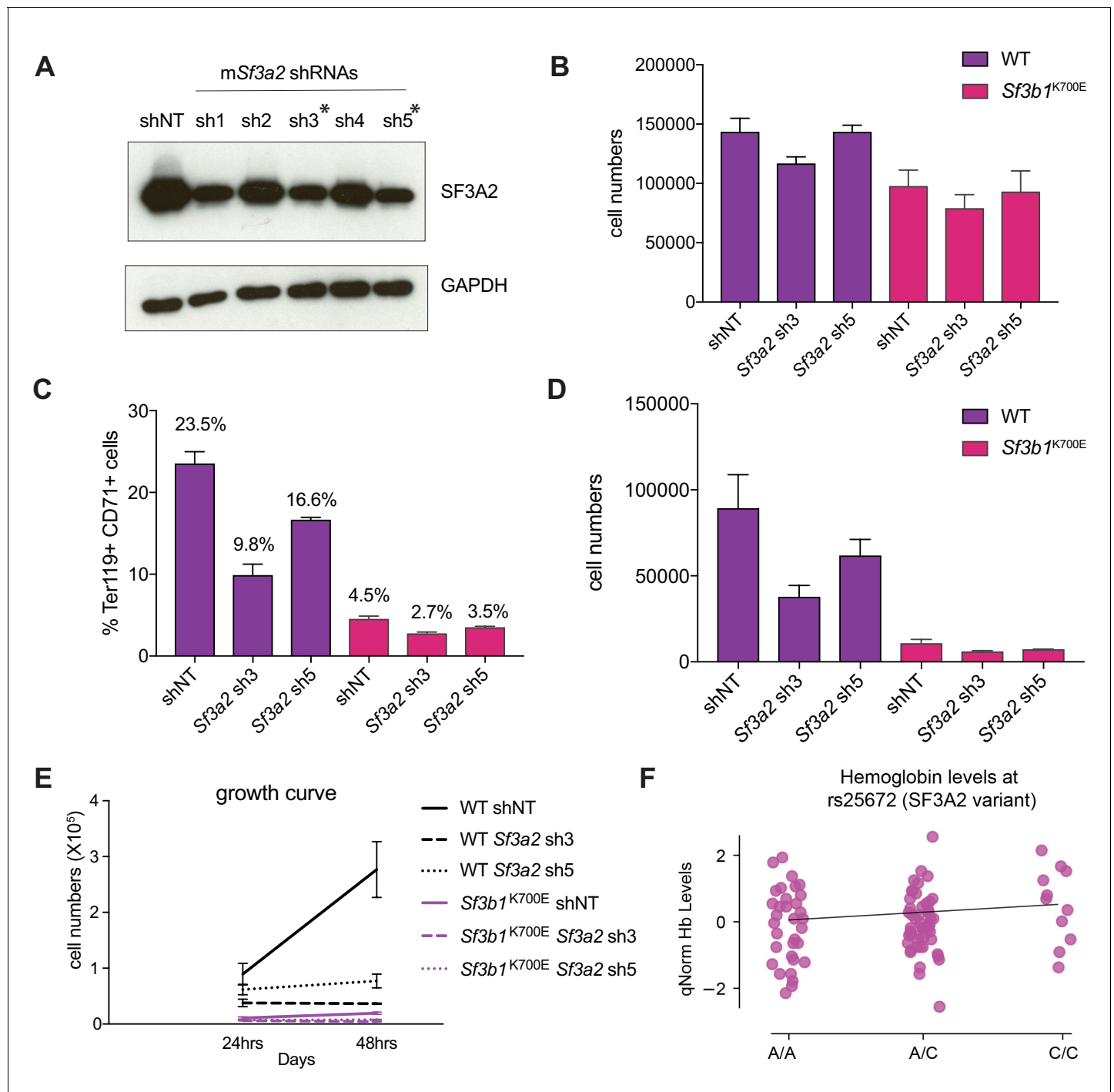
SF3A2 (sh1-4) and a control luciferase gene (shLUC). (C) Growth curves showing that downregulation of SF3A2 results in reduced total cell numbers during erythroid differentiation from three independent experiments. (D) Representative FACS plots of erythroid cell surface markers CD71 (transferrin receptor) and CD235a (Glycophorin A) expression at various time points during erythroid differentiation. Percentages in each quadrant are represented as mean and standard deviation of three independent experiments (E) Altered splicing events identified by RNA-Seq analysis of stage matched erythroid cells (shSF3A2 vs. shLUC). Overlapping changes observed in SF3B1 mutant BM cells from MDS patients (Obeng et al) (**Figure 6—source data 5 and 6**). Differentially expressed genes and pathway analysis available in **Figure 6—source data 1–4**. (F) Lineage negative bone marrow cells from wildtype (WT) and *Sf3b1*<sup>K700E</sup> mice were infected with shRNAs targeting murine *Sf3a2* gene co-expressing a reporter GFP gene. Percentage of Ter119<sup>+</sup> CD71<sup>+</sup> erythroid cells within the GFP compartment after 48 hr in erythroid differentiation. (G) Total cell numbers of GFP<sup>+</sup> erythroid cells after 48 hr in erythroid differentiation.

DOI: <https://doi.org/10.7554/eLife.44080.019>



**Figure 6—figure supplement 1.** Additional Analysis Showing SF3A2 is Required for Human Erythropoiesis. **(A)** shRNAs targeting *SF3A2* co-expressing a reporter GFP gene was infected into human CD34<sup>+</sup> cells and cultured in erythroid conditions. GFP expression at various time points from three independent experiments show that downregulation of *SF3A2* results in reduced cell numbers. **(B)** Representative FACS plots of erythroid (CD235a) and non-erythroid cell surface markers (CD11b/CD41 a) and at various time points showing an increase in non-erythroid lineages upon *SF3A2* downregulation. Cells were gated on the GFP positive population.

DOI: <https://doi.org/10.7554/eLife.44080.020>



**Figure 6—figure supplement 2.** Additional Analysis of Erythropoiesis Defects Observed in *Sf3b1*<sup>K700E</sup> Murine Erythroid Cells upon SF3A2 knockdown. (A) Knockdown efficiency of shRNAs targeting *SF3A2* in murine erythroleukemia (MEL) cells by western blot. (B) Total cell numbers of GFP +shRNA expressing bone marrow cells from wildtype (WT) and *Sf3b1*<sup>K700E</sup> mice at the start of murine erythroid differentiation. (C) Percentage of Ter119<sup>+</sup> CD71<sup>+</sup> erythroid cells within GFP compartment and (D) Total cell numbers of GFP<sup>+</sup> erythroid cells after 24 hr in erythroid differentiation. (E) Growth curves of GFP<sup>+</sup> erythroid cells during erythroid culture. (F) Putative but insignificant interaction between *SF3A2* variant alleles (rs25672) and hemoglobin levels in MDS patients with *SF3B1* mutations.

DOI: <https://doi.org/10.7554/eLife.44080.021>