Supporting Information

Kanjee et al. 10.1073/pnas.1711310114

SI Materials and Methods

Parasite Cell Line Generation. The tdTomato sequence was amplified from a plasmid containing the tdTomato coding sequence (96) using the following primers: pCG110-F 5′–AGTACCTAGGATGGTGAGCAAGGGCGAG and pCG111-RC 5′–AGTACTCGAGTTACTTGTACAGCTCGTCCATGC. The PCR product was digested with AvrII and XhoI (New England Biolabs) and cloned into pEcDamHI (97). For transfection, 100 μg of tdTomato-containing plasmid were cotransfected with the plasmid pINT into 3D7*attB* and Dd2*attB* parasites as described previously (98). Cultures were selected with G418 (125 μg/mL for Dd2 and 250 μg/mL for 3D7), 2.5 μg/mL blasticidin, and 2.5 nM WR99210 for 7 d after transfection followed by continuous selection with blasticidin and WR99210 alone.

Generation of JK-1 shRNA Knockdowns. Lentivirus of the pLKO plasmid containing the shRNA against *GYPA* (TRCN0000116455) was obtained from the Broad Institute, Cambridge, MA. Lentiviral transduction into JK-1 cells was performed based on existing protocols used for $CD34^+$ hematopoietic stem cells (14, 26).

Quantitative Surface Proteomics. The quantitative surface proteomics is based on previously described methods (56, 87). Briefly, 2×10^7 cells of each cell type were washed with PBS. Surface sialic acid residues were oxidized with sodium metaperiodate (Thermo Fisher Scientific) and then were biotinylated with aminooxy-biotin (Biotium). The reaction was quenched, and the biotinylated cells were incubated in a 1% Triton X-100 lysis buffer. Biotinylated glycoproteins were enriched with high-affinity streptavidin agarose beads (Pierce) and washed extensively. Captured protein was denatured with dithiotreitol (Sigma-Aldrich), alkylated with iodoacetamide (IAA; Sigma), and digested on-bead with trypsin (Promega) in 200 mM Hepes (pH 8.5) for 3 h. Tryptic peptides were collected and labeled using TMT reagents (56). The reaction was quenched with hydroxylamine, and TMT-labeled samples were combined in a 1:1:1:1:1:1:1:1 ratio. Labeled peptides were enriched and desalted; then 75% of the total sample was separated into six fractions using tip-based strong cation exchange as previously described (56), and 10% of the total sample was subjected to mass spectrometry unfractionated.

Mass spectrometry data were acquired using an Orbitrap Fusion coupled with an UltiMate 3000 Nano LC (Thermo Fisher Scientific). Peptides were separated on a 75-cm PepMap C18 column (Thermo Fisher Scientific). Peptides were separated using a 90-min gradient of 3–33% acetonitrile in 0.1% formic acid at a flow rate of 200 nL/min (fractionated samples) or a 180-min gradient with otherwise identical parameters (unfractionated sample). Each analysis used a MultiNotch MS3-based TMT method. The scan sequence began with an MS1 spectrum [Orbitrap analysis, resolution 120,000, 400–1,400 Thompsons (Th), automatic gain control (AGC) target 2×10^5, maximum injection time 50 ms]. MS2 analysis consisted of collision-induced dissociation [quadrupole ion trap analysis, AGC 15,000, normalized collision energy (NCE) 35, maximum injection time 120 ms]. The top 10 precursors were selected for MS3 analysis, in which precursors were fragmented by high-energy collision dissociation before Orbitrap analysis (NCE 55, maximum AGC 2×10^5, maximum injection time 150 ms, isolation specificity 0.5 Th, resolution 60,000). Mass spectra were processed using a SEQUEST-based in-house software pipeline as previously described (56). Data were searched using the human UniProt database (April 2014) concatenated with common contaminants (56) and were filtered to a final protein-level false-discovery rate of 1%. Proteins were quantified by summing TMT reporter ion counts across all peptide-spectral matches using in-house software as previously described (56), excluding peptide-spectral matches with poor-quality MS3 spectra [a combined signal:noise (S/N) ratio of less than 250 across all TMT reporter ions]. For protein quantitation, reverse and contaminant proteins were removed, and then each reporter ion channel was summed across all quantified proteins and normalized assuming equal protein loading across all samples. Fold change for each protein was calculated according to (average S/N ratio of BSG knockouts/average S/N ratio of JK-1 controls) or (S/N ratio of cRBC sample/average S/N ratio of JK-1 controls). Protein quantitation values were exported for further analysis in Excel (Microsoft). Gene Ontology Cellular Compartment terms were downloaded from www.uniprot.org, and *P* values (significance A) were calculated and adjusted with the Benjamini–Hochberg method using Perseus version 1.2.0.16 (95).

Comparison of RBC Proteomes. The complete list of 667 proteins identified in the quantitative surface proteomics was compared with published proteomes from the following publications: PubMed ID (PMID) 16861337 (99), PMID 18346024 (100), PMID 24555563 (101), PMID 22954596 (102), PMID 19778645 (103), PMID 18494517 (104), PMID 18614565 (105), PMID 27006477 (106), and PMID 27452463 (57). Datasets were ranked by hierarchical clustering using Gene Cluster 3.0 (82) with a Euclidian distance similarity metric and visualized using TreeView version 1.1.6*r4* (83).

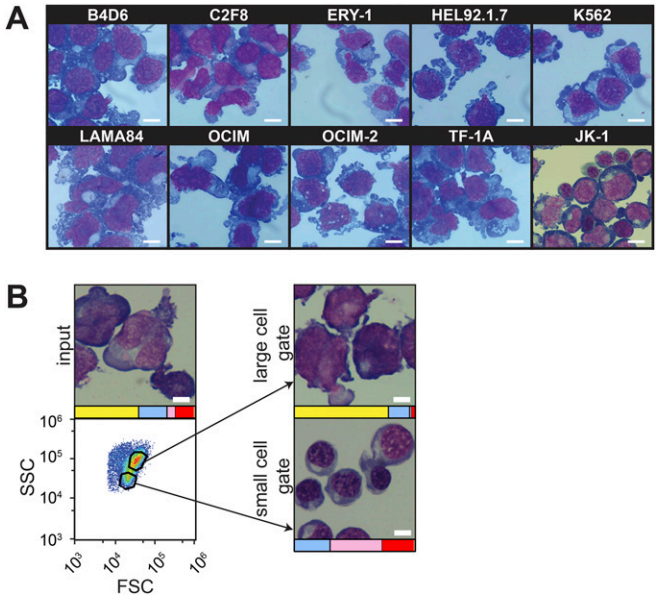

Fig. S1. (*A*) Images of 10 different erythroleukemia cell lines during typical in vitro culture: B4D6 (33), C2F8 (33), Ery-1 (31), HEL92.1.7 (36), K562 (32), LAMA-84 (34), OCIM (37), OCIM-2 (37), TF-1A (35), and JK-1 (22). (Scale bars, 20 μm.) (*B*) Sorting of cells based on cell-size parameters leads to the enrichment of differentiated cells in the small cell gate. The relative proportions of the different cell populations are shown in the bars beneath the microscopy images, color coded according to the key in Fig. 1*A*. (Scale bars, 10 μm.) All cells were stained with May–Grünwald Giemsa.

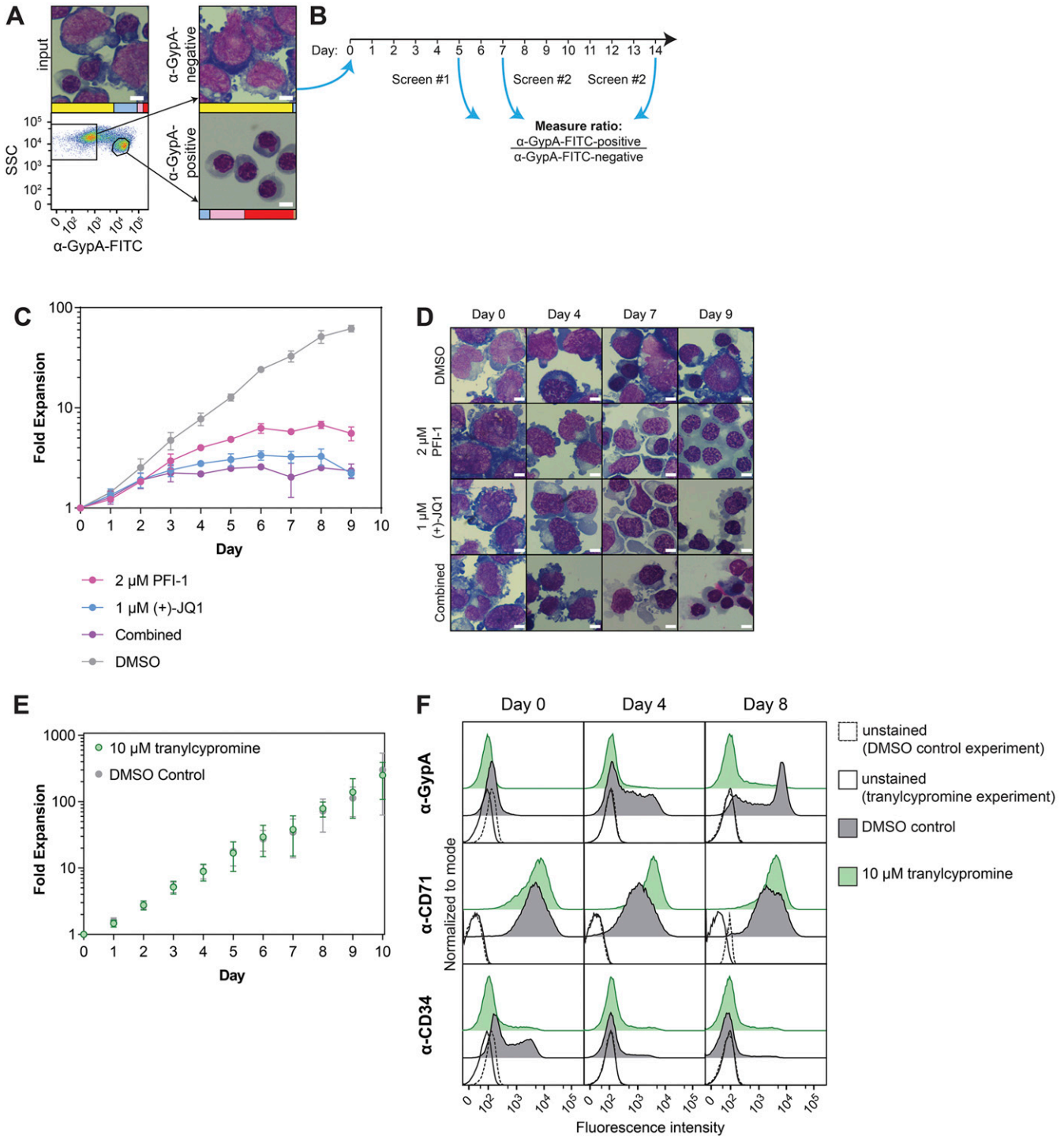

Fig. S2. (*A*) GypA levels increase as JK-1 cells differentiate. An undifferentiated population of JK-1 cells was stained with α-GypA-FITC antibody and sorted into GypA$^+$ and GypA$^-$ fractions. GypA$^-$ cells correspond to undifferentiated proerythroblasts, while the GypA fraction corresponds to differentiated polychromatic and orthochromatic cells. Relative proportions of the different cell populations are shown in the bars beneath the microscopy images, color coded according to the key in Fig. 1*A*. (Scale bars: 10 μm.) (*B*) Schematic of the epigenetic library screen. Undifferentiated JK-1 cells were obtained by sorting for the α-GypA$^-$ population at day 0. Cells were screened for 5 d (screen 1) or 7 and 14 d (screen 2), and differentiation of JK-1 cells was assessed by measuring the ratio of α-GypA$^+$ to α-GypA$^-$ cells. (*C* and *D*) Plots of fold-expansion (*C*) and microscopy images (*D*) for induction with DMSO control, 2 μM PFI-1, 1 μM (+)-JQ1, or a combination of 2 μM PFI-1 plus 1 μM (+)-JQ1. The fold-expansion of the cells is higher in the 2-μM PFI-1 conditions. The addition of (+)-JQ1 alone or in combination with PFI-1 leads to reduced cell expansion and a more distended cell morphology. Data in *C* are the average and SD from two replicates. (*E*) Tranylcypromine maintains JK-1 cells in an undifferentiated state. Growth curve of JK-1 cells treated with DMSO or with 10 μM tranylcypromine. Data are the average and SD of three or four biological replicates. (*F*) Treatment of JK-1 cells with 10 μM tranylcypromine maintains the cells in an undifferentiated state (compared to a DMSO-control) as shown by representative flow cytometry plots of cells stained with GypA, CD71, and CD34 over a course of 8 d.

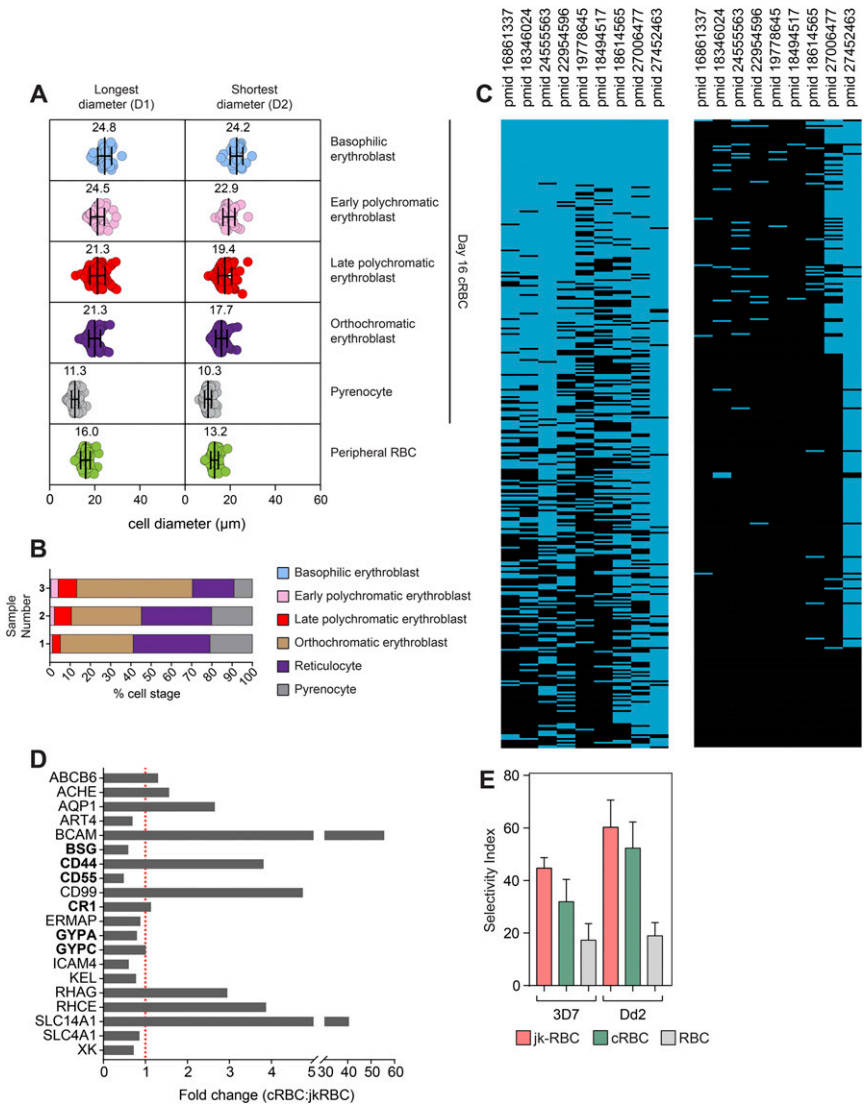

Fig. S3. Comparison of jkRBC cells to cRBCs and RBCs. (*A*) The longest (D1) and shortest (D2) diameters of day 16 cRBC cells at different stages of erythropoiesis and peripheral RBCs were measured in 50–100 cells using May–Grünwald Giemsa–stained images. The average (the numerically indicated value and the tall vertical line) and SDs (indicated by whiskers) for each range are shown. (*B*) The relative proportion of erythroid cells from three independent cRBC cultures at 16/17 d post thaw were measured in 500–800 cells. (*C*) The 677 jkRBC proteins identified by quantitative proteomics were compared with published RBC whole-proteome datasets as shown in the heatmap. The presence of a jkRBC protein in a published dataset is indicated by a blue box. The heatmap is split into two; the right heatmap is a continuation of the bottom of the left heatmap. The following datasets were compared: PMID 16861337 (99), PMID 18346024 (100), PMID 24555563 (101), PMID 22954596 (102), PMID 19778645 (103), PMID 18494517 (104), PMID 18614565 (105), PMID 27006477 (106), and PMID 27452463 (57). (*D*) Comparison of fold-change between cRBCs and jkRBCs for blood group proteins identified by surface proteomics. The majority of cRBC proteins have a relative abundance close to that found in jkRBCs, with a few exceptions (e.g., BCAM, SLC14A1, and CD99). Proteins that are known to be associated with *P. falciparum* invasion are indicated in bold. (*E*) The propensity for multiple invasion events was determined via a selectivity index (61): A higher selectivity index indicates a greater number of multiple invasions into a single host cell. The selectivity index of jkRBCs was comparable to that of cRBCs but was higher than that of RBCs. Data shown are the average and SD are from four biological replicates.

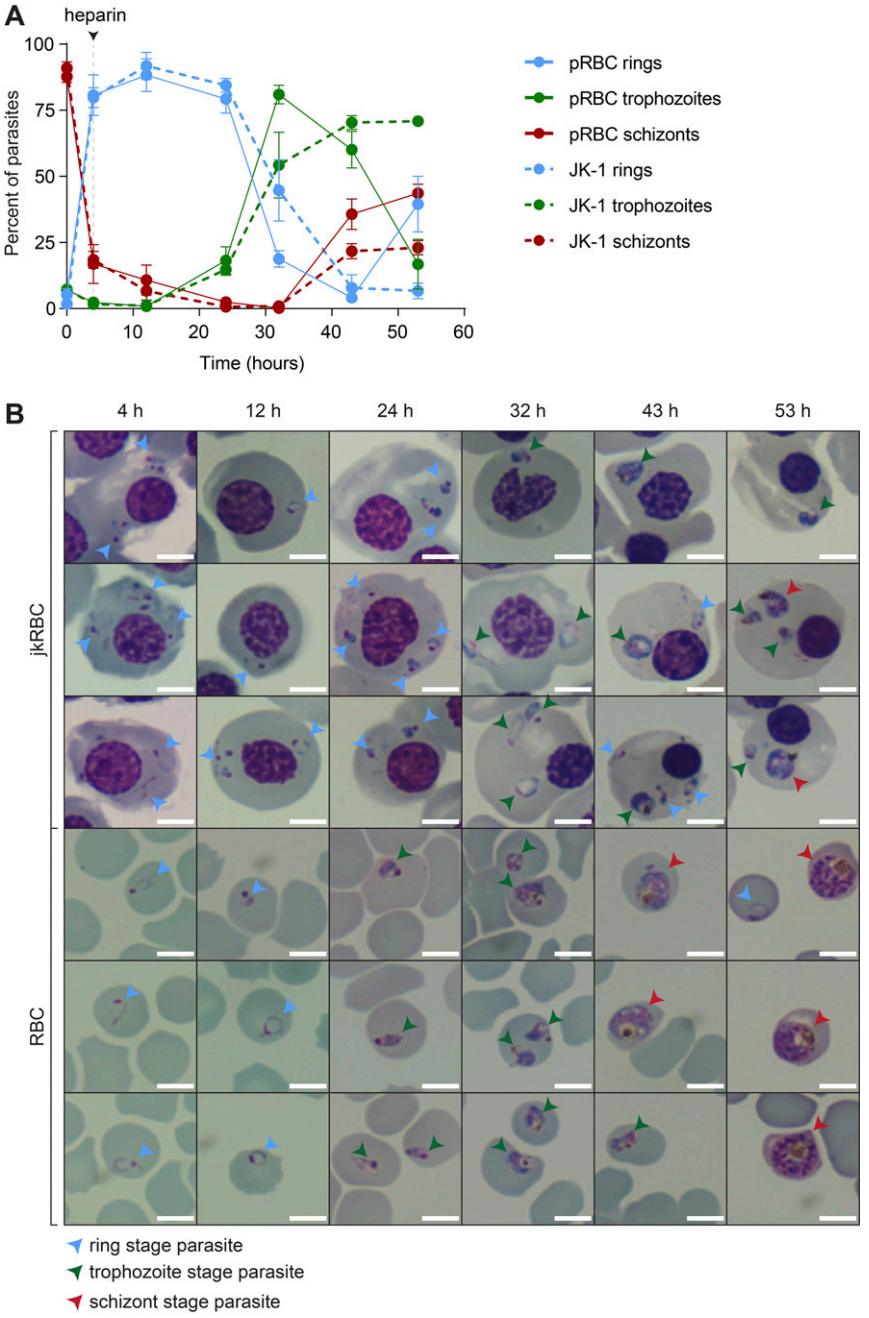

Fig. S4. Growth of parasites in jkRBCs is impeded. (*A*) *P. falciparum* strain 3D7 parasites were purified using a Percoll gradient to enrich for late-stage schizonts, and parasites were allowed to invade jkRBCs and RBCs for 4 h; then heparin was added to block further invasion (76, 107). Parasites were followed over one complete cycle, and the proportion of rings, trophozoites, and schizonts was evaluated by slide microscopy. Formation of trophozoites and schizonts was delayed in jkRBCs compared with RBCs. Data shown represent the average and SD of two biological replicates. (*B*) Representative images of parasites during the experiment with the ring-, trophozoite-, and schizont-stage parasites. (Scale bars: 10 μm.)

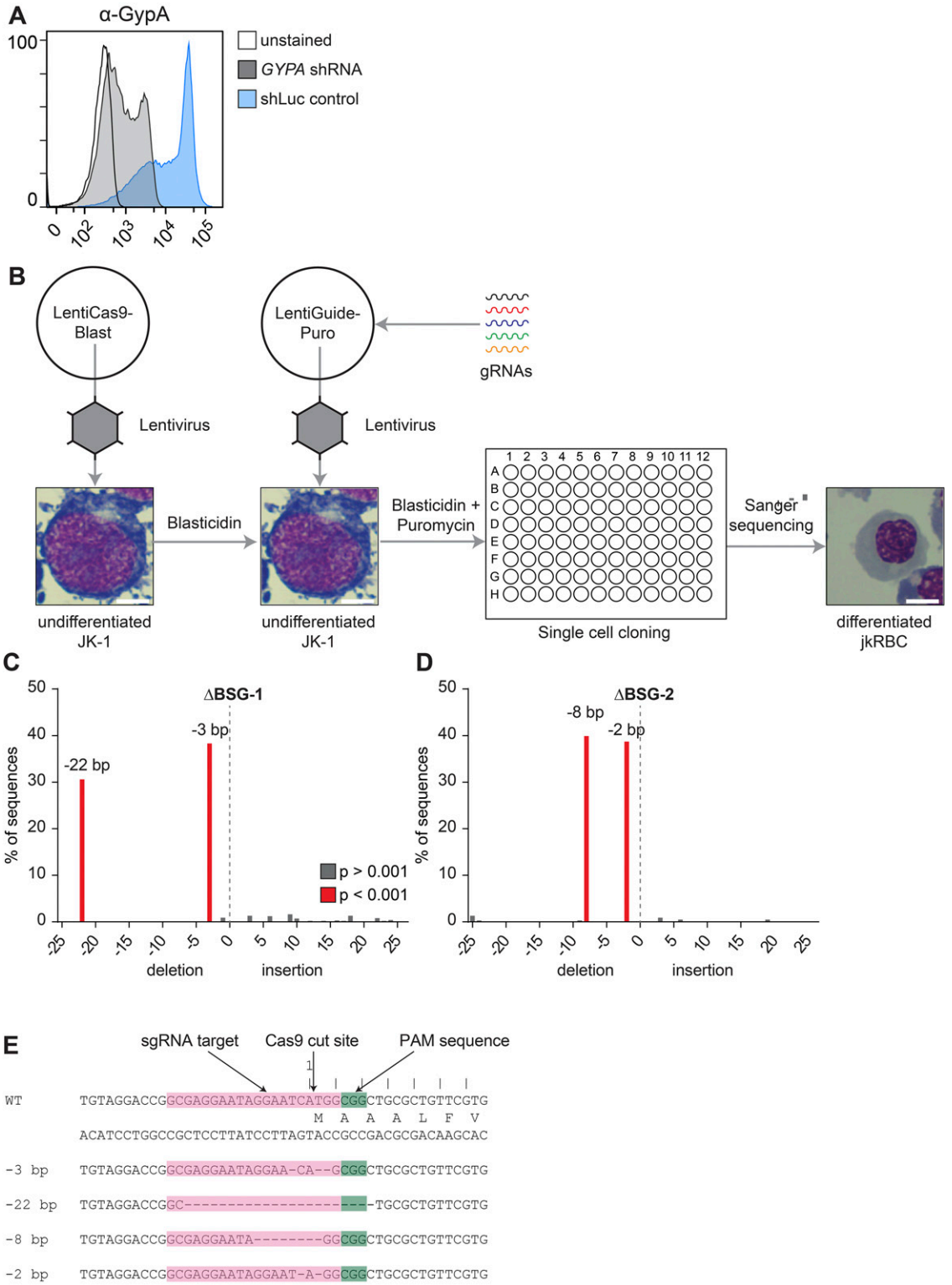

Fig. S5. Genetic manipulation of JK-1 cells. (*A*) An undifferentiated population of JK-1 cells was transduced with lentivirus targeting either *GYPA* or luciferase (shLuc) as a control (14, 26). Flow cytometry analysis with α-GypA–FITC staining revealed an ∼20-fold knockdown (based on comparison of mean fluorescence intensity). (*B*) Generation of CRISPR/Cas9 knockouts and validation of *BSG* knockout by TIDE (86) analysis. The schematic shows the steps in generating gene knockouts using JK-1 cells. Cells were transduced with a lentivirus containing the LentiCas9-Blast vector and then were selected with blasticidin. The single-guide RNA was cloned into LentiGuide-Puro, this vector was transduced into the Cas9-expressing cells, and cells were coselected with blasticidin and puromycin. After 2 wk of selection, cells were plated to obtain clones by limiting dilution. Deletions were validated by Sanger sequencing and TIDE analysis. (*C*) The Δ*BSG*-1 clone has two prominent deletions of −3 bp and −22 bp. (*D*) The Δ*BSG*-2 clone has two prominent deletions of −2 bp and −8 bp. (*E*) The location of the deletions was mapped based on the Sanger sequencing data and TIDE analysis.

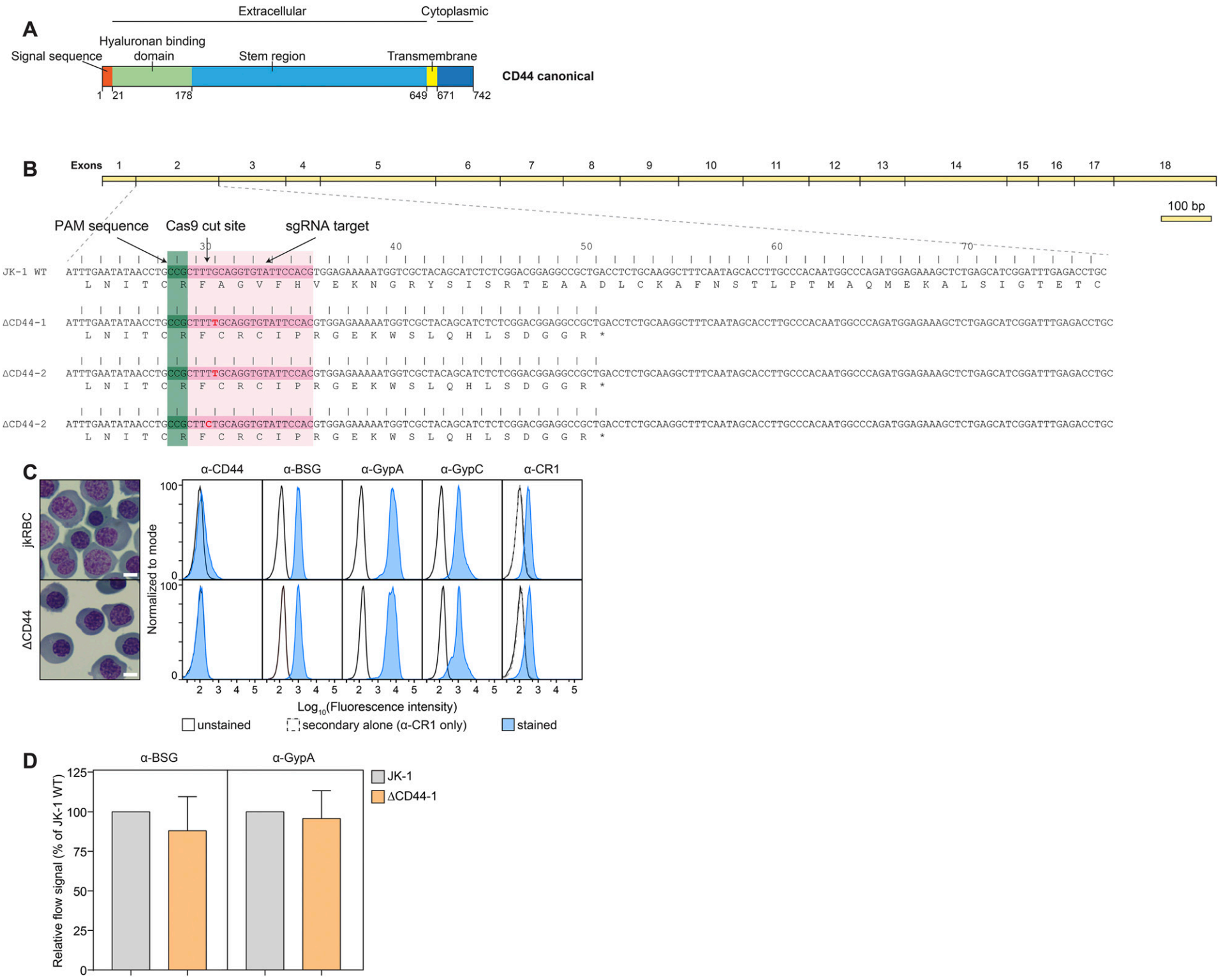

Fig. S6. Generation of the Δ*CD44*-knockout line. (*A*) Domain structure of the full-length CD44 protein indicating the N-terminal hyaluronan-binding domain in the extracellular region, the single-pass transmembrane helix, and the C-terminal cytoplasmic domain. Numbers below the figure represent amino acid positions. (*B*) Exon structure of *CD44* and location of the CD44-1 sgRNA-binding site. In the Δ*CD44*-1 and Δ*CD44*-2 clones we observe single-base insertions (indicated in red bold type) that result in a premature stop codon (*) and a truncated protein. Importantly, the truncated protein does not have a transmembrane domain, which is located in the C terminus of the gene. (*C*) Representative microscopy images of wild-type jkRBCs and Δ*CD44*-knockout cells are shown. Flow cytometry comparison of levels of CD44, BSG, GypA, GypC, and CR1 in WT and Δ*CD44*-knockout jkRBC cells demonstrating specific loss of CD44 signal while levels of other surface markers remain unaffected. (Scale bars: 10 μm.) (*D*) Measurement of the relative flow cytometry signals for BSG and GypA in JK-1 WT and Δ*CD44*-1knockout lines. Data shown are the average and SD from four independent experiments.

Dataset S1. Epigenetic modifiers tested for the ability to induce differentiation of JK-1 cells

Dataset S1

This dataset shows the epigenetic modifiers ranked by cluster analysis. The Cayman Chemicals (https://www.caymanchem.com/Home) catalog number for each compound is listed. Chemical Abstract Service (CAS) numbers were obtained from Cayman Chemicals or from the SciFinder software (https://scifinder.cas.org/). Functional annotation of the targets of each compound was compiled from available literature, and the compounds were grouped into the following target categories: histone acetyltransferases, histone deacetylases, histone methyltransferases, histone demethylases, DNA methyltransferases, sirtuins, bromodomains, and others.

Dataset S2. Cell-surface proteomic analysis of JK-1 WT, Δ*BSG*, Δ*CD44*, and cRBC lines

Dataset S2

The full list of the 677 proteins identified from the surface proteomics analysis is shown in the No_Filter worksheet. All proteins were identified from the UniProt database with the exception of PNP, which was identified in the Trembl database. Classification of the identified proteins is given: UniProt ID, gene symbol, description, and Gene Ontology Cellular Compartment (GOCC) term classification: CS, cell surface; IPM, integral to plasma membrane; M, membrane; Nuc, nuclear; PM, plasma membrane; ShG, short GO; XC, extracellular. "Short GO'' refers to a subset of proteins annotated by GO as "integral component of membrane" but with no subcellular assignment (108). The number of peptides quantified for each protein is shown followed by the fold-change (FC) comparing the average S/N ratio from the two JK-1 WT samples to the average S/N ratio from the two Δ*BSG*-knockout clones or the average S/N ratio from the two Δ*CD44*-knockout clones or the S/N ratio from the cRBC sample. The normalized S/N ratio for each protein in JK-1 cells (WT, Δ*BSG*, or Δ*CD44*) or cRBCs is shown. The PM_CS_XC_ShG_2 peptides worksheet shows all 237 identified plasma membrane proteins annotated as CS, PM, ShG, or XC. The Mapping_Existing_Proteomes worksheet shows the presence or absence of the 677 proteins identified via surface proteomics with existing published RBC proteomes (indicated by PMID numbers).