Supplemental methods

Cell culture conditions

The leukemia cell lines used in this study are listed in Supplementary Table 1. Cell lines SKNO1, TANOUE, REH, and MOLT16 were obtained from Leibniz Institute DSMZ, Germany. Other cell lines were obtained from ATCC (Manassas, VA). The cell lines were maintained in a humidified incubator with 5% CO_2 at 37°C. HL60/S4, 697(EU3), CEM, MOLT4, MOLT16, and TANOUE were grown in RPMI 1640 (Corning, Pittston, PA) supplemented with 10% heat-inactivated (HI) FBS (Gibco, Gaithersburg, MD). KASUMI1 and REH were grown in RPMI 1640 supplemented with 20% HI FBS. SKNO1 was grown in RPMI 1640 supplemented with 10% HI FBS and 10 ng/mL GM-CSF (Stem Cell Technologies, Vancouver, Canada). NALM6 was grown in RPMI supplemented with 20% HI FBS, GlutaMAX, and 2-mercaptoethanol (Gibco, Gaithersburg, MD). SEM and K562 were grown in IMDM (Gibco, Gaithersburg, MD) supplemented with 10% HI FBS. Cells were passaged in 1 to 3 ratio every 3-4 days. The absence of mycoplasma contamination was routinely verified using the MycoAlert PLUS Mycoplasma Detection Kit (Lonza, Walkersville, MD).

Immunostaining and flow cytometric assays

For Western blotting, cell pellets were lysed using RIPA Buffer (Sigma, St. Louis, MO) supplemented with Halt Protease Inhibitor Cocktail (Thermo Scientific, Waltham, MA). Protein levels were quantified using the DC Protein Assay Reagents Package (Bio-Rad, Hercules, CA), following manufacturer's protocol, and a Thermo Scientific NanoDrop

2000c. Samples were diluted with RIPA Buffer, and NuPAGE LDS Sample Buffer (Life Technologies Carlsbad, CA) supplemented with 0.5 M 2-mercaptoethanol (Sigma, St. Louis, MO). Samples were run on 10% SDS-PAGE gels and transferred to Immobilon-P (Millipore IPVH00010) PVDF membranes. Membranes were developed with either 20X LumiGLO Reagent and 20X Peroxide for CL-XPosure X-ray film or WesternSure Premium Chemiluminescent Substrate for LI-COR C-DiGit blot scanner (Thermo Scientific, Waltham, MA). Membranes were stripped and reused following the abcam protocol using the harsh stripping buffer. For assessment of myeloid differentiation, HL60/S4 cells were stained with CD11b-PE antibodies BioLegend (San Diego, CA). Cell proliferation was analyzed by staining with eBioscience Ki67-PE antibody using the Foxp3/Transcription Factor Staining Buffer Set (Invitrogen, CA), following Protocol B for Intracellular (Nuclear) proteins staining. Cell viability was assessed using 0.5% 7-aminoactinomycin D (7-AAD) (Via-Probe™, BD Bioscience, NJ) and 2% FBS in DPBS (Corning, Pittston, PA). For the cell cycle assessment, cells were stained with bromodeoxyuridine (BrdU) using the BrdU-FITC Flow Kit (BD Pharmingen, San Jose, CA). The expression of aldehyde dehydrogenase (ALDH) enzyme was assessed with ALDEFLUOR Kit (Stem Cell Technologies, Vancouver, Canada). Flow cytometric analysis was performed on BD FACSCalibur flow cytometer and data were analyzed using FlowJo V10 software. The antibodies are listed in Supplementary Table 2.

Plasmid preparation, transfections, and transductions

The lentivirus constructs for constitutive and doxycycline-inducible expression of short-hairpin (sh) RNAs are listed in Supplementary Table 3. The constitutive expression of

shRNAs against IGF2BP1 mRNA (shIGF2BP1 sequence 1 (SH1)), shIGF2BP3 and shControl were obtained from Sigma (St. Louis, MO) and allow for the puromycin selection in mammalian cells. The doxycycline-inducible shIGF2BP1 (sequences 2 and 3 (SH2 and SH3)) and scrambled shControl were obtained from GE Dharmacon (Lafayette, CO) and contained GFP reporter. Lentiviral vectors for doxycycline-inducible and constitutive overexpression of GFP, IGF2BP1, ALDH1A1, HOXB4, and MYB are listed in Supplementary Table 4. IGF2BP1-eGFP full length cDNA was kindly provided by Dr. Joel Yisraeli (The Hebrew University of Jerusalem, Israel). Plasmid DNAs were isolated using the EndoFree Plasmid Maxi Kit (Qiagen, Germantown, MD) according to manufacturer's protocol. Lentiviral particles were made using 10 µg pPAX, 7.5 µg pMDG and 10 µg of expression vector as previously described[1]. Retrovirus was assembled with pGag-Pol and pVSV-G plasmids. Vectors for mammalian overexpression were electroporated using Lonza 4D-Nucleofector and Amaxa P3 Primary Cell Kit (Lonza V4XP-3012). Cells were selected on with following doses of antibiotics: 1-10 µg/mL puromycin, 500-800 µg/mL Geneticin (G418), 200 µg/mL hygromycin B, or sorted on a BD FACSAria III sorter. Lines transduced with inducible shRNA were induced with 2.5 µg/mL doxycycline for 4-5 days before collection for RNA or protein analysis.

Chemical compounds

The IGF2BP1 inhibitor BTYNB was obtained from Cayman Chemical. Cells were treated with BTYNB at indicated concentration for 2-4 days. Other compounds were obtained from Sigma, Tocris, and MP Biomedicals. For drug treatments, constitutively expressed shIGF2BP1 and shControl were transduced and selected with 1-10 µg/ml of puromycin

for 5-7 days prior to 24-hour treatment with indicated concentrations of compounds. Cell viability was assessed 24 hours after each treatment. For IGF2BP1 overexpression experiments, IGF2BP1-GFP or eGFP-control cells were sorted and expanded prior to drug application. Three technical and three biological replicates were conducted for each treatment. Data are presented as a mean value of three biological replicates (n=3), with error bars indicating the standard error of mean (± SEM). HL60/S4 differentiation was induced with 1 μM all-trans retinoic acid (ATRA) for 4 days. For detailed information about chemicals used in this study, see Supplementary Table 5.

RNA isolation and gene expression analysis

RNA was isolated using the RNeasy Mini Kit (Qiagen, Germantown, MD), followed by DNase I treatment with DNA-free DNA Removal Kit (Invitrogen, Carlsbad, CA). cDNA was made with the iScript cDNA Synthesis Kit, and quantitative PCR (qPCR) was performed using iTaq Universal SYBR Green Supermix on Bio-Rad C1000 Touch Thermal Cycler CFX96 Real Time System (Bio-Rad, Hercules, CA). qPCR reactions were assembled with two technical replicates, and three biological replicates were performed for each experiment. qPCR data are presented as a mean of three biological replicates (n=3), with error bars indicating the standard error of mean (± SEM). Two pairs of primers targeting the same gene of interest were used to ensure data consistency. Primer's sequences and catalog numbers are listed in Supplementary Table 6.

RNA sequencing analysis of 697 (EU3) ALDH$^+$ subpopulation with and without IGF2BP1 knockdown was performed in four biological replicates (n=4). Each biological replicate represents *de novo* selection of ALDH$^+$ cells and independent doxycycline treatments. RNA isolation and DNase I treatments were conducted as described above. QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina was used for library preparation, and 3'Fwd-mRNAseq was performed on an Illumina sequencer. The compressed fastq files from RNA sequencing were preprocessed for quality control with FastQC, trimmed and filtered with bbmap. The sequences were aligned with STAR software. The SAM tools program was used for indexing the bam file. We then generated gene expression counts with HTSeq-count and detected differential gene expression with DESeq2 in the R program. These software tools provided a list of differentially expressed genes, P-values, false discovery rate (FDR), q-value, and normalized expressions values. Differentiated genes then were utilized for gene ontology (GO) enrichment annotation with cluster Profiler package in R and KEGG pathway analysis with STRING (http://string-db.org/).

Colony-forming cell (CFC) assay

One thousand leukemia cells were seeded for one 3 mL test of MethoCult H4034 Optimum medium according to instructions (STEMCELL Technologies, Vancouver, Canada). After 10-14 days, colonies were independently counted by two people and imaged using the IncuCyte S3 Live-Cell Analysis System (Essen BioScience Inc., Arbor, MI). Three independent assays were routinely conducted for each cell line. CFC data represent mean values with error bars standard error of mean (± SEM).

In vivo experiments

Non-obese diabetic/severe combined immunodeficient gamma (NSG) mice were obtained from Jackson Laboratory. All procedures were carried out in accordance with the Institutional Animal Care and Use Committee (IACUC) guidelines. Mice were bred and maintained under pathogen-free conditions at an American Association for the Accreditation of Laboratory Animal Care accredited animal facility at Pennsylvania State University and housed in accordance with the procedures outlined in the Guide for the Care and Use of Laboratory Animals under an animal study proposal approved by the Animal Care and Use Committee at Pennsylvania State University. For the engraftment experiments, $1x10^3$ -$1x10^6$ cells were injected into tail veins of non-irradiated 6-10 week-old female mice in 100 µL of DPBS per mouse. For inducible shRNA experiments, mice in experimental groups were fed doxycycline food. No blinding was applied to mice experiments. Routinely, each *in vivo* experiment was performed with three technical replicates (three mice per group) and independently repeated two to three times for each cell line. The biological replicates were conducted with the *de novo* transduced, puromycin or GFP selected/sorted cells, and with the efficient IGF2BP1 knockdown verified by western blot. Xenotransplantation in mice was detected according to previously published protocols[2]. Terminally sick mice were euthanized at humane endpoints per IACUC protocol guidelines.

Cross-linking Immunoprecipitation (CLIP)

The cross-linking immunoprecipitation was performed according to the previously published CLIP protocols [3, 4] with the minor modifications. Briefly HL60/S4, 697(EU3), and K562 cells were cultured for 3 days and then fixed with 3% formaldehyde in PBS for 10 min and lysed by sonication (10 pulses for 10 seconds). Immunoprecipitation was performed overnight at 4°C using IGF2BP1 antibody- (MBL, Woburn, MA, Catalog# RN007P) coated Dynabeads (Life technology). After collection by magnetic separation and washing 5 times with lysis buffer, the beads along with RNA-protein complexes were reverse cross-linked, and RNA was extracted using Trizol reagent (Invitrogen) according to the manufacturer's instructions. Isolated RNA was treated with RNase-free DNase I (Thermo Scientific) to remove traces of genomic DNA. First-strand cDNA was generated using iScript cDNA synthesis kit from BioRad according to manufacturer's instruction followed by real-time qPCR using iTaq Universal SYBR green kit (Bio-Rad). CLIP assay was conducted in three biological replicates. Each biological replicate was set with three technical replicates, therefore, nine samples were processed for each cell line.

Photoactivatable Ribonucleoside-Enhanced Crosslinking an Immunoprecipitation

PAR-CLIP was performed with the off-beads ligation protocol as previously described[5]. Briefly, anti-IGF2BP1 antibody (MBL, Woburn, MA) was used on 3 x 10^8 4-SU treated and fixed K562 cells. The radiolabeled, ligated, and reversed transcribed samples were PCR amplified by a pilot PCR with 10 cycles and another 17 cycles. The amplified samples were concentrated and size selected with BluePippin 3% agarose gel cassette

(Sage Science). Library concentrations were measured using Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific) and D1000 Screen Tape TapeStation System (Agilent). The purified samples were sequenced on the Illumina HiSeq3000 with 50 cycles of single-read (SR).

Statistical Analysis

Data were analyzed using GraphPad Prism (GraphPad Software Inc., San Diego, CA, USA) and R programming language version 3.4.4 (R Foundation, Vienna, Austria). A log-rank (Mantel-Cox) test was used to determine p values in Kaplan-Meier survival curves comparison. One-way ANOVA models were used to analyze the differences between 3 or more groups. For two-group analysis, two-sample Student's or Welch's t-tests were used. All tests were two-sided, and values with *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001were considered statistically significant.

Bioinformatics Data Deposition

Gene expression profiling data by high throughput sequencing GEO ID: GSE138704 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138704

PAR CLIP data GEO ID: GSE138063

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE138063

Supplemental Methods References:

1. Elcheva I, Brok-Volchanskaya V, Slukvin I. Direct Induction of Hemogenic Endothelium and Blood by Overexpression of Transcription Factors in Human Pluripotent Stem Cells. *J Vis Exp* 2015 Dec 3; (106)**:** e52910.

2. Notta F, Doulatov S, Dick JE. Engraftment of human hematopoietic stem cells is more efficient in female NOD/SCID/IL-2Rgc-null recipients. *Blood* 2010 May 6; **115**(18)**:** 3704-3707.

3. Knuckles P, Vogt MA, Lugert S, Milo M, Chong MM, Hautbergue GM*, et al.* Drosha regulates neurogenesis by controlling neurogenin 2 expression independent of microRNAs. *Nature neuroscience* 2012 Jun 17; **15**(7)**:** 962-969.

4. Vogt M, Taylor V. Cross-linked RNA Immunoprecipitation. *Bio-protocol* 2013 2013/03/05; **3**(5)**:** e398.

5. Benhalevy D, McFarland HL, Sarshad AA, Hafner M. PAR-CLIP and streamlined small RNA cDNA library preparation protocol for the identification of RNA binding protein target sites. *Methods* 2017 Apr 15; **118-119:** 41-49.