Identification of GTF2IRD1 as a novel transcription factor essential for acute myeloid leukemia

Yaser Heshmati[1], Gözde Türköz[1], Marios Dimitriou[1], Aditya Harisankar[1], Johan Boström[2], Huan Cai[1], Nadir Kadri[1], Mikael Altun[2], Hong Qian[1], Julian Walfridsson#[1]*

[1] Center for Hematology and Regenerative Medicine, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

[2] Research Division of Translational Medicine and Chemical Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden.

Running head: GTF2IRD1 is essential for AML

***Corresponding author:** Julian Walfridsson, Julian.walfridsson@ki.se.

Word count: abstract: 196, main text: 3967.

Tables and figures: 6 main figures and 2 supplemental files.

Abstract

Acute myeloid leukemia (AML) is an aggressive blood malignancy characterized by clonal accumulating of immature myeloid progenitors in the bone marrow and peripheral blood. Transcription factors are the most frequently mutated and dysregulated genes in AML and they have critical roles in AML pathogenesis and progression. In this study, we performed large-scale RNA interference screens in *MLL-AF9* transformed AML cells and identified *GTF2IRD1* as a novel transcription factor essential for the survival of various types of myeloid leukemic cells *in vitro* and *in vivo*, but not for primary normal hematopoietic cells. Inhibition of *GTF2IRD1* reduced the frequency of primary childhood and adult AML cells, including cell populations enriched for leukemia-initiating cells. In animal models for AML, inhibition of *GTF2IRD1* significantly delayed the disease progression. Inhibition of *GTF2IRD1* caused an accumulation of quiescent AML cells in the G0 phase of the cell cycle but caused minor effects in apoptosis. In line with this, RNA sequencing analysis revealed a significant downregulation of E2F targets as a consequence of inhibition of GTF2IRD1. Taken together, we identified GTF2IRD1 as a transcription factor with a selective importance in AML and our findings may contribute to new therapeutic inventions for the disease.

Introduction

Acute myeloid leukemia (AML) is a phenotypically and genetically heterogeneous type of blood cancer, characterized by clonal expansion and loss of differentiation ability of immature myeloid progenitors, which outcompete normal cells in the bone marrow and peripheral blood (1). AML is the most common type of acute leukemia (comprising 80% of all leukemic cases in adults and 15-20% in children) with an incidence rate of 3-5 cases per 100,000 population in United States (2). With the currently approved therapeutic protocols, AML is cured only in 35-40% of patients under 60 years old, and in 5-15% of the patient over 60 years old with a median survival of 5-10 months (3, 4). Identification of novel target genes and mechanisms that selectively control AML cell growth and disease progression may facilitate development of new treatment approaches that specifically targets the AML cells with less harmful effects on normal cells.

In AML genomes, transcription factors (TFs) are among the most frequently mutated and dysregulated genes, including chromosomal translocations in CBF, RARα, RUNX1 and the HOX gene family, but also point mutations targeting for example C/EBPα, P53 and RUNX1, are common aberration in AML genomes (5,6). TFs and their coactivators were early recognized for their involvement in AML development as oncogenic drivers (7). Thus, in order to improve the clinical outcome for patients with AML, it is important to gain further understanding of the molecular mechanisms promoting AML pathogenesis and progression governed by TFs.

GTF2IRD1 belongs to the TF II-I family and was first discovered in yeast (8). It is characterized by multiple helix-loop-helix like I-repeat domains (9). It can function as either an activator or repressor depending on the cell type, target gene and its context (9-11). GTF2IRD1 is perhaps best known for its implicated role in Williams-Beuren syndrome, where GT2IRD1 and the homolog GTF2I are part of a microdeletion, causing symptoms such as cardiovascular disease, a distinctive craniofacial appearance, and intellectual disability and hypersociability (12). In addition, GTF2IRD1 has been reported to have an important role in different types of cancer, including colorectal cancer (13-15), pancreatic cancer (16) and breast tumor formation (17,18). However, the role of GT2IRD1 in normal and malignant hematopoiesis is largely unknown.

In this study, we used shRNA-based RNA interference (RNAi) screens at a large scale, to identify novel TFs that were required for cell growth of *MLL-AF9* transformed AML cells. We

discovered GTF2IRD1 as being critical for AML cell growth but not primary normal hematopoietic cells. In syngeneic and patient derived xenograft mouse models for AML, inhibition of GTF2IRD1 significantly delayed disease onset and prevented maintenance of patient-derived leukemia-initiating cells (LICs). Furthermore, we have shown that GTF2IRD1 function as a transcriptional repressor and inhibition of the TF resulted in accumulation of quiescent AML cells , which was consistent with downregulation of E2F targets, revealed by RNA sequencing (RNA-seq) analysis.

Results

Identification of transcription factors required for cell growth of AML cells, using large-scale shRNA-based RNAi screens

To identify novel TFs that are required for cell growth of AML cells carrying the *MLL-AF9* translocation, we performed shRNA-based screens on the NOMO-1, THP-1 and genetically defined mouse AML cells, all carrying the MLL-AF9 fusion oncogene. The barcoded lentiviral RNAi library consisted of 27,500 shRNAs targeting approximately 5,000 annotated genes enriched for signaling pathways and the majority of annotated TFs (Cellecta Inc.). The AML cells were transduced by viral transfer in a pooled format as described in Figure 1A and the cells were harvested for genomic DNA extraction at an initial time point and after ten cell divisions. Effects in cell growth of the shRNA-targeted genes was determined by next generation sequencing of PCR amplified barcodes, originating from the genomic DNA of the AML cells. By calculating the ratio of the isolated barcoded shRNAs at the initial time point with ten cell divisions at population level, 648 target genes with at least a five-fold reduction, were overlapping between all three cell lines used in the screens (Figure 1B, Supplementary Table 1). With a focus on TFs, 31 of the in total 648 hits represented annotated TFs (Figure 1C). Among the identified 31 TFs, several have previously been reported to have a role in AML, including MEIS1, IKZF1/2, EZH1, HES1 and MYC, suggesting the validity of the screens, whereas 10 of the 31 TFs have not been reported to have a role in AML (Figure 1D). GTF2IRD1 was one of the TFs with highest effects in reduced cell proliferation upon inhibition. Due to its remarkable overexpression in subpopulations enriched for leukemia-initiating cells compared to normal counterpart cells (19), and its important role in different cancer types (13-18), we decided to focus on GTF2IRD1 to validate its functional effects and the underlying mechanism.

Inhibition of *GTF2IRD1* prevents expansion of AML cell lines and delays disease onset in a mouse model for AML.

To validate the findings from the screen, we performed cell growth competition assays by first co-culturing mouse *MLL-AF9* AML cells transduced with an shRNA vector targeting *Gtf2ird1* co-expressing RFP, with control cells expressing a negative scrambled shRNA and GFP. Efficient inhibition of *Gtf2ird1* mRNA in mouse *MLL-AF9* AML cells (Figure 2A), and flow cytometry analysis revealed a marked reduction in number of live AML *Gtf2ird1* inhibited cells, as compared to the negative control cells (Figure 2B and C). Likewise, a significant reduction in mRNA levels upon inhibition of *GTF2IRD1* in the human THP-1 and NOMO-1 AML cells expressing the *MLL-AF9* translocation (Figure 2D), prevented cell growth of both cell lines (Figure 2E and F), indicating that *GTF2IRD1* is essential for different AML cells carrying the *MLL-AF9* translocation. Similarly, inhibition of *GTF2IRD1* in non *MLL*-rearranged leukemic cells, including HL-60, K-562 or NB-4 cells (Figure 2G), significantly reduced cell growth of these cells *in vitro,* showing that *GTF2IRD1* is important also for AML cells not carrying the *MLL-AF9* translocation (Figure 2H and I).

To investigate the role of *Gtf2ird1* in AML disease progression, we transplanted mouse *MLL-AF9* AML cells transduced with two individual *Gtf2ird1* shRNAs or a negative scramble control shRNA, into recipient mice. Inhibition of *Gtf2ird1* was robust as compared to the control (Figure 2J). Mice that received control cells (Sc shRNA) develop AML within 18-21 days after transplantation (Figure 2J). However, in sharp contrast, mice that received AML cells transduced with the two individual shRNAs survived significantly longer that the control group (Figure 2K), showing that inhibition of *Gtf2ird1* delay disease onset in an immunocompetent transplantation mouse model for AML.

***GTF2IRD1* is not essential for normal hematopoietic cell growth and function.**

We then asked whether inhibition of *GTF2IRD1* has an effect on the expansion and differentiation of normal mouse hematopoietic stem and progenitor cells (HSPCs). Dynamic flow cytometric analysis of the HSPCs using an *in vitro* cell growth competition assays, showed no significant difference in expansion between *Gtf2ird1* inhibited cells as compared to control cells transduced with a non-targeting scramble vector, during 7 days of culturing (Figure 3A-C). Furthermore, neither total number of colonies nor mixed myeloid-erythroid CFUs (CFU-GEM) nor granulocyte/macrophage CFUs (CFU-GM), were significantly affected in terms of colony numbers from cells transduced with shRNA against *Gtf2ird1* as compared to cells carrying non-targeting negative control vector (Figure 3D and E). This indicated that *Gtf2ird1* was not essential for proliferation, survival or differentiation of mouse HSPCs *in vitro*.

Next, we investigated the importance of *Gtf2ird1* in engraftment and performed intravenous transplantation of sorted ckit$^+$ cells transduced with *Gtf2ird1* or a non-targeting negative control vector, into lethally radiated recipient mice. Flow cytometric analysis of BMs from the transplanted mice revealed minor changes in the engraftment of mice transplanted with *Gtf2ird1* inhibited cells compared to the control group, suggesting that *Gtf2ird1* is not critical for engraftment of HSCs in the bone marrow (Figure 3F).

To investigate the importance of *GTF2IRD1* in normal human hematopoietic cell proliferation and survival, we tested effects in cell growth of umbilical cord blood cells (UCBs) transduced with shRNA vectors against *GTF2IRD1* compared to co-cultured control cells (Figure 3G). Dynamic flow cytometric analysis of the mixed cells showed non-significant differences in expansion in suspension of the UCBs transduced with an shRNA against GTF2IRD1 compared to the control cells (Figure 3H and I). To explore the importance of *GTF2IRD1* in primary human hematopoietic cells in a more long-term perspective, we again transduced UCBs with *GTF2IRD1 specific* shRNAs and control vectors, and maintained the UCBs in a competitive setting on stromal cell feeder layers for 3 weeks. Flow cytometric analysis showed that the number of live CD45 positive (CD45$^+$) UCBs were not significantly changed between the *GTF2IRD1* targeted cells and control cells (Figure 3J). Consistent with the transplantation studies in the syngeneic mice in Figure 3F, intravenous tail injections of sorted human CD34 positive (CD34$^+$) UCBs into humanized NSG-SGM3 recipient mice, showed non-significant differences in engraftment between the *GTF2IRD1* targeted cells compared to the control cells (Figure 3K). Together, these data suggested that *GTF2IRD1* is not essential for normal hematopoietic viability and differentiation *in vitro* as well as *in vivo*.

CRISPR-Cas9 mediated disruption of *GFT2IRD1* prevents expansion of AML cells

To rule out possible off-target effect associated to the shRNA-mediated loss of function experiments, we conducted CRISPR-Cas9 mediated gene editing experiments in combination with cell growth competition assays. THP-1 and MV4-11 AML cells which constitutively expressed the Cas9 nuclease were transduced with a guide RNA (gRNA) vector homologous to the coding region of human *GTF2IRD1* (Figure 4A). The *GTF2IRD1* gRNA co-expressing the fluorescent marker (iRFP670) caused a significant disruption of *GTF2IRD1*, using a non-targeting control gRNA expressing BFP (Figure 4B). By mixing the cells and monitoring the relative ratio of the cells with flow cytometry, we observed a strong reduction of live cells

transduced with *GTF2IRD1* gRNA compared to control cells, in both THP-1 and MV4-11 AML cells as compared to the control cells (Figure 4C, D, E).

Inhibition of *GTF2IRD1* reduces the frequency of primary adult and childhood AML cells, including LICs *ex vivo*.

Having demonstrated a selective importance of *GTF2IRD1* in various AML cell lines and disease progression in animal models for AML, we wanted to investigate the role of *GTF2IRD1* in primary adult and childhood AML samples, including enriched LICs. For that purpose, the patient samples from both adults and children were transduced with either *GTF2IRD1* targeting vectors or negative control vectors (used in Figure 2D-H) and maintained in a co-culture system with stromal cells. Although one of the samples (”Adult AML 7”) did not show a significant difference due to high variation of the cells in the negative control sample, flow cytometric analysis of *GTF2IRD1* targeted cells caused a consistent reduction in maintenance of primary $CD45^+$ AML bulk cells isolated from adult patients, compared to the cells transduced with a control vector (Figure 5A). Similarly, the $Lin^-CD34^+CD38^-$ subpopulation enriched for LICs, revealed a significant reduction in the frequency of cells upon inhibition of *GTF2IRD1* with the exception of one sample (“Adult AML 5”), where the variation in the negative control cells was too large to show a statistically significant difference (Figure 5B). To investigate if the importance of *GTF2IRD1* was conserved in primary childhood AML cells, we conducted shRNA-based inhibition of *GTF2IRD1* in two childhood AML samples as described in Figure 5A and B. Consistent with the AML samples from adults, inhibition of *GTF2IRD1* significantly reduced the numbers of $CD45^+$ bulk as compared to the control cells Figure 5C. We also observed a reduction of $Lin^-CD34^+CD38^-$ childhood AML cells for one of the samples upon inhibition of *GTF2IRD1* (“Childhood AML 1”), whereas the reduction of cells in the second sample (“Childhood AML 2”) was not statistically significant (Figure 5D).

Importantly, intrafemoral injection of five adult and one childhood primary AML samples, with and without shRNA-mediated inhibition of *GTF2IRD1* into humanized NSG-SGM3 mice, with the exception of one sample (“Adult AML 8“), showed a significant decrease in the total number of human $CD45^+$ cells in the BM of the recipient mice as compared to the control cells. Together, these studies suggested that *GTF2IRD1* is essential for maintenance of both primary AML cells from both adults and children, including the LICs and for engraftment in mice.

Downregulation of *Gtf2ird1* as a transcriptional repressor blocks AML cells in the G0 phase of the cell cycle without causing apoptosis.

Next, we asked whether the observed essential role for *Gtf2ird1* in AML cells was associated to a function in cell cycle progression or cell apoptosis. Analysis of mouse *MLL-AF9* AML cells 72 hours after *Gtf2ird1* inhibition with significant mRNA reduction (Figure A) showed a strong accumulation of cells in the G0 phase of the cell cycle and a subsequent reduction in G1, S and G2-M phases, in comparison to the negative control cells (Figure 6B). In contrast, the percentage of early and late number of apoptotic cells was relatively low (between 1-3%) (Figure 6C), suggesting that the main cellular consequence to inhibition of *Gtf2ird1* was a resulting in accumulation of cells in the G0 phase of the cell cycle.

RNA profile pattern under *GTF2IRD1* suppression show a correlation to E2F targets.

To understand the molecular mechanism by which *GTFF2IRD1* was essential for AML cell growth, we performed differential gene expression profiling with RNA sequencing (RNA-seq) analysis with and without shRNA-mediated loss of function of THP-1 AML cells, using an shRNA against *GTFF2IRD1* and a negative control vector expressing a non-targeting shRNA (Figure 6D). Using statistical (FDR<0.1) and a two-fold change cutoff, 343 genes were found to be up regulated and 443 genes were down regulated a consequence of inhibition of GTF2IRD1 (Figure 6E). To identify the dysregulated gene sets associated with the cellular response upon the *GTF2IRD1* inhibition we performed gene set enrichment analysis (GSEA) (20, 21). This did not reveal any significantly altered pathways (KEGG, Biocarta, Wikipathways) or gene ontology. On the contrary, the expression changes in response to the inhibition of *GTFF2IRD1* compared to the hallmark gene sets were significantly correlated with the gene signature of E2F targets data sets (FDR Q-value=0.0007 and normalized enrichment score [NES]= 1.8.), (Figure 6F and G), suggesting that *GT2IRD1* is required for AML cell growth via regulation of E2F targets.

Discussion

This study investigates the role of GTF2IRD1 in AML and identifies *GTF2IRD1* as being required for cell growth of various AML cells *in vitro* and *in vivo*, including primary AML cells and LICs from adults and children, but not normal hematopoietic cells. The essential role of *GTFIRD1* in AML is associated to cell cycle progression of cells from G0 into S phase.

Despite the genetic and epigenetic heterogeneity in primary AML samples (22), our results highlight a consistent importance of *GTF2IRD1* in maintenance of primary childhood and adult AML cells with mixed genotypes *in vitro* and *in vivo,* including for enriched populations of LICs that drive the disease (23). Although GTF2IRD1 has previously not been reported to have a function in leukemia, it is overexpressed in AML hematopoietic stem and progenitor cells as compared to normal counterpart blood cells (19). In addition, there is a growing body of evidence that GTF2IRD1 has a functional important role in additional cancer types, including colorectal cancer (13-15), pancreatic cancer (16), and breast cancer (17,18). It is therefore possible that GTF2IRD1 has a more general role in cancer than previously anticipated.

In the present study, we also show that the essential function of *GTF2IRD1* in AML is selective. Although further work is needed to address the underlying mechanisms to the observed leukemia-specific role, neither is normal primary hematopoietic cells significantly affected by functional inhibition of *GTF2IRD1* nor is the reconstitution of the hematopoietic cells in animal models. Consistent with this, *Gtf2ird1* homozygote knockout mice is viable and display phenotypic features restricted to craniofacial abnormalities and cognitive defects and a mild growth deficiency (24, 25, 26). One possible explanation to the selective importance is the determined overexpression of GTF2IRD1 in subpopulations of AML cells as compared to normal counterpart cells (19).

In our study, inhibition of *Gtf2ird1* in AML cells results in reduction of cell expansion via a mechanism that involve a cell cycle arrest and that quiescent cells are resistant to loss of function of *Gtf2ird1*, manifested by an accumulation of the cells in the G0 phase of the cell cycle. Consistent with this, GSEA analysis revealed a significant correlation to downregulation of E2F targets upon inhibition of *GTF2IRD1*, and E2F effectors have a principal role in regulating the cell cycle of cells (27). The link of *GT2IRD1* to cell cycle regulation has also been reported for colon cancer (28). Unfortunately, we were not able to find ChIP grade antibodies for GTF2IRD1 and could therefore not determine direct binding targets.

Overall, we show for the first time that inhibition of GTF2IRD1 prevent disease progression and cell growth of primary AML cells but not normal hematopoietic cells. Chemical targeting of TFs via proteasome-mediated proteolysis is an approach that has proven to be a remarkable approach for the treatment of various leukemia's (29). In this way, our findings might contribute to increased understanding of AML and provide therapeutic targets that may facilitate clinical

solutions for the disease.

Material and Methods

Cell culture, cell lines and primary cells

The human AML cell lines; THP-1, NOMO-1, MV4-11, K-562 HL-60 and NB-4 (all purchased from DSMZ) were cultured in RPMI with 10% FBS, 100 U/mL penicillin and 100 ug/mL streptomycin at 37 °C, 5% CO2. The murine *MLL-AF9* AML cells (30) were cultured in the same condition with 10 ng/mL rmIL-3 (R&D systems). Culturing of normal murine HSPCs and human $CD34^+$ UCBs cells has been described previously (31).

Cell growth assays of primary AML samples

Isolation and culturing of the AML samples was carried out as previously explained (30). Briefly, transduced AML cells were cultured on MS-5 cells (DSMZ) in Myelocult media H5100 (StemCell Technologies Inc.) with rhIL-6, rhIL-3, rhFl3/Flk-2 ligand, rhTPO, and rhSCF and rhG-CSF (Stemcell technology), at a concentration of 20 ng/mL.

Vectors, molecular cloning and knockout analysis

The inducible shRNA and gRNA vectors were generated as described before (30). gRNAs were designed by Optimized CRISPR Design - MIT (http://crispr.mit.edu) provided by the Zhang laboratory and cloned into the inducible gRNA expression vectors. The constitutive lentiviral-based Cas9-mCherry vector was kindly provided by Marco Herold (31). CRISPR/Cas9 growth assay was explained in previous article (30). All oligos used in the study are listed in supplementary Table 2.

Transfection and Transduction

Transfection and transduction were performed as previously described (30).

Large-scale shRNA screens

Large scale shRNA screens were described in detail in a previous study (30). Briefly, around 3×10^8 cells (THP-1, NOMO-1 and murine AML) were transduced with pooled barcoded lentiviral shRNA libraries targeting genes playing role in signaling pathway (Module 1, Cellecta Inc). The transduction efficiency was kept below 25% to minimize the number of cells with multiple integrations of the vectors. Half of the cells were collected for an initial input control (T0), 48 hours after selection with puromycin (2 μg/mL) (Invitrogen) and the remaining

cells were harvested 10 cell divisions. The barcodes were amplified from genomic DNA and the purified barcodes were used for next generation sequencing (HiSeq 2000, Illumina).
The ratios of the barcodes from the two screens were calculated by first dividing the normalized abundance of each barcode after T10, the abundance at the initial time point (T0) separately for each cell line.

Flow cytometric analysis and sorting

Flow cytometry and sorting was performed as previously described (29). Primary AML cells were harvested and stained with human CD45 and cells were analyzed by a high-throughput automated plate reader (BD LSRFortessa). Dead cells were excluded using the Near-IR Live/Dead marker (Invitrogen). For the cell growth competition assays, cells washed with PBS were stained with Near-IR Live/Dead marker in a 96-well plate then a high-throughput automated plate reader was used to detect the absolute number of live cells.
To determine the level of engraftment of human AML cells, the isolated BMs stained with human anti-CD45 and Near-IR Live/Dead marker to detect live cells. All FACS Analysis was performed by FlowJo Version 9.3.3 software (TreeStar). All antibodies used in the study are listed in Supplementary Table 3.

RNA extraction and quantitative real time PCR (qPCR)

RNA was extracted using RNeasy Micro or Mini Kit (Qiagen), following the manufactures instructions. cDNA synthesis was performed with 500 ng of RNA by using cDNA synthesis kit (Fermentas). qPCR was performed using ABI Thermal Cycler and qPCR primers were purchased from Eurofins (primers are presented in Supplementary Table 4).

AML mouse models

All the mice were kept in a pathogen-free animal facility in Karolinska Institutet, Huddinge, Sweden. The C57BL/6 wild type mice and the NOD-scid IL2Rgnull-3/GM/SF, NSG-SGM3 mice were purchased from The Jackson Laboratory. All of the transplanted mice were investigated daily for symptoms of leukemogenesis and disease progression was monitored by complete blood tests.
The mouse MLL-AF9 induced AML cells and AML mouse model was generated as previously described (30, 31). The generated MLL-AF9 AML cell line was used for further transplantation experiments or *in vitro* experiments.

For the Kaplan-Meier survival analysis, wild type CD45.2 C57BL/6 mice were used for transplantation of murine MLL-AF9 AML cells. 100,000 AML cells transduced with lentiviral shRNA vectors, selected with puromycin (2 µg/mL) and were intravenously transplanted into non-irradiated recipient mice. After disease onset, the mice were euthanized.

The experiments using humanized NSG-SGM3 mice were performed on animals aged six to ten weeks. Recipient mice were subjected to sub-lethal irradiation (220 cGy) and six to 12 hours post irradiation the primary childhood AML patient BM samples were transplanted via intra-femur injection using doses of 100,000-500,000 cells/mouse. BM samples were collected from the bones of euthanized primary recipient mice 6-8 weeks post transplantation and the primary leukemic cells were analyzed with flow cytometric analysis for donor cell engraftment (See Supplementary Table 4 for antibodies).

Cell cycle and apoptosis analysis

The procedure of cell cycle assay and apoptosis assay were explained before (30).

RNA-Sequencing

Total RNA was extracted from mouse MLL-AF9 cells by RNeasy Micro Kit (Qiagen), 72 hours after transduction of the cells, or by selection with puromycin. The process was described in previous study (30).

Western Blotting

Total proteins were prepared from cells using RIPA Buffer with Protease Inhibitor Cocktail (sc24948, Santa Cruz Biotechnology). Protein quantified using the Bradford assay (Bio-Rad). 20 ug protein was prepared in NuPAGE reducing agent (NP0007, Invitrogen). Proteins were separated in 4-15% Tris Glycine gels (Biorad) and transferred to nitrocellulose membranes (1704158, Biorad) by using Trans Blot Turbo Blotting System (170-4155, Biorad).

Nitrocellulose membranes were blocked with 5% nonfat dried milk (170-6404, Biorad) with 1X PBS (D8537, Sigma) and 0.2% Tween 20 (P9416, Sigma) before incubating with the primary polyclonal antibodies EHMT1 and A5441, Sigma used to detect human EHMT1 and Beta actin respectively. After incubating with secondary antibodies, the detection of proteins was performed by using enhanced chemiluminescence kit (32106, Thermo Scientific) and the image was captured by a LI-COR Odyssey Fc Imager system. Quantification of the relative intensity of the bands were performed by using Image Studio Lite (LI-COR Biosciences).

Statistical analysis

Statistical analysis was done with GraphPad Prism version 6 and P-values of <0.05 were considered statistically significant. GSEA analysis was done using log2 transformed expression values and the Molecular Signature Data Base, MSigDB (http://software.broadinstitute.org/gsea/msigdb/annotate.jsp), applying "Compute Overlaps" tool and "Hallmark gene sets", "Reactome gene sets".

Ethics Statement

All experiments were done with the ethical standards and according to the Declaration of Helsinki and according to national and international guidelines and has been approved by the authors' institutional review board. The primary AML childhood BM samples were obtained from Uppsala biobank (reg.nr 827) after approval from NOPHO - Nordic Society of Pediatric Hematolgy and Oncology and the local ethical committee at Stockholm (ethical permit: 2017-1059-32). All animal experiments were approved by the local ethics committee in Linköping, Sweden and the Swedish board of agriculture (Ethical permit number; ID 530). UCB samples were obtained from the Karolinska Hospital (Stockholm, Sweden) with informed consent of the parents.

Acknowledgements

We thank Tim Somervaille for kindly providing the MLL-AF9 vector and Marco Herold for kindly providing the Cas9-mCherry vector. We also would like to thank NOPHO and Josefine Palle for providing the primary childhood AML cells as well as Cecilia Götherström and the National Cord Blood bank at Karolinska University Hospital for providing the UCBs. We would like to acknowledge the MedH Core Flow Cytometry facility (Karolinska Institutet), supported by KI/SLL, for providing cell sorting services, cell analysis services, technical expertise and scientific input. We also would like to thank the Affymetrix core facility at Neo, BEA, Bioinformatics and Expression Analysis, which is supported by the board of research at the Karolinska Institute and the research committee at the Karolinska hospital.

Author contributions

All authors have read and provided input into the manuscript. Y.H. designed and performed most experiments, analyzed data and wrote the manuscript. G.T. deisgned and performed experiments, animal work, cell growth assays and RNA sequencing preparation. N.K. M.D. provided input into the study and helped with long-term expansion of hematopoietic cells and

flow cytometric analysis. A.H. performed analysis of RNA-Seq data and the shRNA screens. J.B. cloned and designed the inducible gRNA and shRNA vectors. H.C. helped with the RNA-seq analysis. M.A. designed the gRNA and shRNA vectors and provided input in the project. H.Q. helped to design the studies, drafting the manuscript, analysis of the experiment and supervision of Y.H. J.W. Conceived the project, supervised the study, analyzed the data, interpreted result and wrote the manuscript. All authors participated in editing of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Funding

This work was supported by the Wallenberg Foundation, The Swedish Research council, The Swedish Cancer Society, Swedish Childhood Cancer Fund, Magnus Bergwalls Foundation, Karolinska Institutet, Åke Wibergs Foundation, Radiumhemmets forskningsfonder, Dr Åke Olsson Foundation for Hematological Research.

Figure legends

Figure 1. Largescale RNAi screens identified GTF2IRD1 as a novel transcription factor with a potential role in AML cell growth.

A. Experimental scheme for identification of target genes by performing large-scale shRNA screens in AML cells. **B**. Venn diagram showing overlapping target genes between mouse *MLL-AF9* AML and human THP-1 and NOMO-1 AML cell lines that caused >five-fold reduction in expansion of the cells in the screens after ten cells divisions. **C.** Venn diagram showing overlap between the 648 identified target genes from the screens, to all annotated TFs (Pfam database). **D.** The table represent the 38 known and unknown TFs identified from the screens.

Figure 2. GTF2IRD1 inhibition impairs mouse and human leukemia cell growth and prolongs survival in mice with AML.

A. Real time qPCR analysis of mRNA levels of *Gtf2ird1* assessed in mouse AML cells after transduction with vectors targeting *Gtf2ird1* or control (Sc) shRNAs. The mRNA levels were normalized to *Gapdh*. The data is presented as mean ±S.E.M., ****$p<0.001$, (unpaired t-test).

B. Representative flow cytometric analysis of the cell growth competition assay in (C), where the mouse AML cells transduced with lentiviral vectors expressing shRNA against *Gtf2ird1* and RFP, or negative control vectors expressing scrambled shRNA and GFP (Sc), were mixed and propagated for 15 days. **C**. Line chart of the cell growth competition assay where the relative ratio of the RFP^+ *Gtf2ird1* inhibited cells and the control cells was dynamically monitored by flow cytometric analysis. Percentages of live RFP^+ cells transduced with *Gtf2ird1* shRNA (red line) were normalized to 100% based on the Day 1 numbers. **D**. Real time qPCR analysis of mRNA levels of *GTF2IRD1* in THP-1 and NOMO-1 cells after transduction with vectors targeting *GTF2IRD1* or control (Sc) shRNAs, (normalized to UBC) in cells transduced with Sc control (black bar) or GTF2IRD1 shRNA (grey bar) as indicated. The data is presented as mean ±S.E.M., ****$p<0.001$, (unpaired t-test). **E**. Representative flow cytometric analysis of THP-1 cells transduced with lentiviral vectors expressing shRNA against *GTF2IRD1*, or control negative control vectors expressing scrambled shRNA and GFP (Sc). The transduced cells were used in the cell growth competition assay in (F) and propagated for 15 days **F**. Line chart showing the relative ratio of the live cells and the control GFP^+ cells dynamically monitored by flow cytometric analysis, at the indicated time points for 15 days after mixing the cells. Percentages of live RFP^+ cells transduced with *GTF2IRD1* shRNA were normalized to 100% based on the Day 1 values and shown in bar graphs compared to Sc control shRNA. **G**. Real time qPCR analysis of mRNA levels of *GTF2IRD1* in HL-60, K562 and NB4 cells after transduction with vectors targeting *GTF2IRD1* or control (Sc) shRNAs, (normalized to UBC) in cells transduced with Sc control (black bar) or GTF2IRD1 shRNA (grey bar) as indicated. The data is presented as mean ±S.E.M., ****$p<0.001$, (unpaired t-test). **H**. Representative flow cytometric analysis of HL-60, K-562 and NB-4 cells, cells transduced with lentiviral vectors expressing shRNA against *GTF2IRD1*, or control negative control vectors expressing scrambled shRNA and GFP (Sc). The transduced cells were monitored as described in (E). **I**. Line chart showing the relative ratio of the live cells and the control GFP^+ cells dynamically monitored by flow cytometric analysis, at the indicated time points for 15 days after mixing the cells. Percentages of live RFP^+ cells transduced with *GTF2IRD1* shRNA were normalized to 100% based on the Day 1 values and shown in bar graphs compared to Sc control shRNA. **J.** Real time analysis of *Gtf2ird1* mRNA level (normalized to Gapdh) in *MLL-AF9* derived AML cells transduced with control (Sc) or two independent *Gtf2ird1* shRNAs before transplanting into sublethally radiated mice. **K**. Kaplan-Meier survival curves are shown for mice transplanted with mouse *MLL-AF9* AML cells (200,000 cells/mouse), transduced with shRNA

against *Gtf2ird1* or negative control vectors. n = 10 in each cohort. The data is presented as mean ±S.E.M., *p<0.05, ***p<0.005, ****p<0.001, (unpaired t-test).

Figure 3. *GTF2IRD1* suppression is not essential for cell growth of normal hematopoietic cells *in vitro* and *in vivo*.

A. Real time qPCR analysis of mRNA levels of *Gtf2ird1* assessed in mouse HSPCs after transduction with vectors targeting *Gtf2ird1* or control (Sc) shRNAs. The mRNA levels were normalized to *Gapdh*. The data is presented as mean ±S.E.M., ***p<0.005, (unpaired t-test). **B**. Representative flow cytometry analysis charts showing percentage of live RFP^+ *Gtf2ird1* inhibited mouse HSPCs (*Gtf2ird1* shRNA-RFP) relative to GFP^+ control cells (Sc shRNA-GFP) at the indicated days after mixing the cells for the *in vitro* growth competition assay in (C). **C**. Line charts depicting the percentage of live RFP^+ cells normalized to 100% with GFP^+ cells after mixing the HSPCs, determined by flow cytometry analysis at the indicated time points. **D** and **E**. Clonogenic progenitor assays of sorted *Gtf2ird1* inhibited or control HSPCs where equal cell numbers were plated in in methylcellulose and total colonies were counted after seven days CFU-GM and CFU-GEMM (D), or after ten days (E), of clonal growth. Counts of different colony types are shown as mean ±SEM of n = 3, ns = non-significant, (unpaired t-test). **F**. Line charts showing the percentage of engraftment in the BM of recipient mouse after transplantation of sorted *Gtf2ird1* RFP^+ inhibited CD45.1 and GFP^+ control cells into a congenic transplantation model. Live RFP^+ cells were normalized to 100% after mixing the HSPCs, determined by flow cytometry analysis at after 2, 5, 8 and 12 weeks post transplantation for total donor $CD45.1^+$ cells. The data is represented as the mean ±SEM, n=6. **G**. Real time qPCR analysis of mRNA levels of *GTF2IRD1* in $CD34^+$ enriched cells after transduction with vectors targeting *GTF2IRD1* (*GTF2IRD1* shRNA, grey bar) or control shRNAs, (Sc control (black bar), normalized to UBC. The data is presented as mean ±S.E.M., **p<0.01, (unpaired t-test). **H**. Representative flow cytometry plots of percentage of live RFP^+ *GTF2IRD1* inhibited human HSPCs (*GTF2IRD1* shRNA-RFP) relative to GFP^+ control cells (Sc shRNA-GFP) at the indicated days after mixing the cells for the *in vitro* growth competition assay in (I). **I**. Line chart of cell growth competition assay where $CD34^+$ CB cells transduced with shRNAs against *GTF2IRD1* were mixed and expanded in suspension with supplemented cytokines for 15 days. The number of viable cells were determined by flow cytometric analysis and the data is shown as mean ±SEM of n = 3, ns = non-significant (unpaired t-test). **J.** Quantification of long termed expanded live normal $CD34^+$ UCBs transduced with *GTF2IRD1* targeting vectors or negative scramble control vectors after culturing the cells in suspension for 3 weeks. The number of

human $CD45^+$ UCBs was determined by flow cytometric analysis after three weeks of culturing. The data is represented as mean ±SEM of n = 3, ns = non-significant, (unpaired t-test). **K.** Box plot showing the percentage of live $CD45^+$ UCBs transduced with *GTF2IRD1* targeting shRNA or scramble control vector, eight weeks after transplantation of 50,000 cells into humanized NSG-SGM3 recipient mice by intra femural injection. The BM of the xenografted mice was analyzed for human $CD45^+$ engraftment after transplantation by flow cytometry. Results are shown as mean ±SEM of n = 3 mice per experiment, ns = non-significant, (unpaired t-test).

Figure 4. CRISPR-Cas9 gene editing confirmed that *GTF2IRD1* is essential for expansion of human leukemic cells.

A. Schematic image of gRNA design and targeting location at exon 5 of *GTF2IRD1*. The gRNA target sequence is highlighted in bold letters and the PAM sequence is highlighted in blue. **B**. Bar diagram showing the level of gene disruption 72 hours after transduction of THP-1 cells, for the gRNA used in C-E, determined by digital PCR analysis. The data is shown as mean ±SEM of n = 3, (unpaired t-test) ****p<0.001. **C.** Representative flow cytometry analysis of percentage of THP-1 AML cells stably expressing the Cas9 nuclease and the mCherry reporter transduced with a gRNA vector expressing gRNA against *GTF2IRD1* (iRFP-670^+) or control gRNA (BFP^+), at the day of mixing the cells and after 16 days of co-culturing for the cell growth competition assay shown in (C). **D** and **E**. Line chart of cell growth competition assay showing the relative abundance of *GTF2IRD1* (iRFP-670^+) THP-1 cells (C) or NOMO-1 cells (D), or control gRNA (BFP^+) cells, determined by flow cytometric analysis at the indicated time points. The data is shown as mean ±SEM of n = 3, ns = non-significant (unpaired t-test).

Figure 5. Inhibition of *GTF2IRD1* significantly reduce the frequency of primary adult and childhood AML cells, *ex vivo* and in a xenograft mouse model for AML.

A and **B.** Bar diagram showing the frequency of patient derived $CD45^+$ bulk cells form seven adult AML patient samples (A), or $Lin^-CD34^+CD38^-$ enriched LICs from five adult AML patient samples, transduced with *GTF2IRD1* shRNA (grey bars) compared to Sc shRNA (black bars). AML patient samples were cultured in long term on MS-5 stromal cells. The data is shown as mean ±SEM of n = 3, ns = non-significant, **p<0.01, ***p<0.005, ****p<0.001 (unpaired t-test). **C** and **D**. Bar diagram of the absolute number of patient derived $CD45^+$ bulk cells form two childhood AML patient samples (C), or $Lin^-CD34^+CD38^-$ enriched LICs from the two childhood AML patient samples, transduced with *GTF2IRD1* shRNA (grey bars)

compared to Sc shRNA (black bars). AML patient samples were cultured in long term on MS5 stromal cells. The data is shown as mean ±SEM of n = 3, ns = non-significant, *p<0.05. **E**. Box plot of the frequency of engrafted primary AML patient $CD45^+$ cells transduced with GTF2IRD1 or Sc shRNA, isolated from the BM of recipient NSG-SGM3 mice after eight weeks. The frequencies of the engrafted cells were determined by flow cytometric analysis and each individual dot represents level of engraftment in an individual mouse. *p<0.05, **p<0.01, ***p<0.005 (unpaired t-test).

Figure 6. inhibition of GTF2IRD1 arrest AML cells in the G0 phase of the cell cycle and downregulation of E2F targets

A. Real time qPCR analysis of mRNA levels of *Gtf2ird1* assessed in mouse *MLL-AF9* AML cells after transduction with vectors targeting *Gtf2ird1* or control (Sc) shRNAs. The mRNA levels were normalized to *Gapdh*. The data is presented as mean ±S.E.M., ***p<0.005, (unpaired t-test). **B**. Bar graphs illustrating flow cytometric quantification of mouse *MLL-AF9* AML cells transduced with *Gtf2ird1* shRNAs, or a negative control vector, stained with Ki-67 and DAPI, at 72 hours after transduction. The data is presented as mean ±S.E.M., n=3, **p<0.01, ***p<0.005, ****p<0.001 (unpaired t-test). **C**. mouse *MLL-AF9* AML cells transduced with *Gtf2ird1* shRNA, or a negative control vector, stained with Ki-67 and DAPI, at 72 hours after transduction. The data is presented as mean ±S.E.M., ns = non-significant, **p<0.01, (unpaired t-test), n=3. **D**. Western blot of protein extract from THP1 cells with shRNA-based inhibition of GTFF2IRD1 and control cells transduced with a negative control vector (sc), that were used for the RNA-seq analysis in Figure E-G. **E**. MA plot of RNA-seq analysis. Indicated genes represent top 20 leading edge genes **F**. GSEA enrichment plot of hallmark genes, showing the significant correlation to E2F targets upon GTFF2IRD1 inhibition in THP1 cells. NES score and FDRq scores are indicated in the diagram. **G**. Leading edge analysis of top 20 most significant genes affected as a consequence of biological triplicate experiments of GTF2IRD1 inhibition in THP1 cells.

References

1. Estey E, Dohner H. Acute myeloid leukaemia. Lancet. 2006;368(9550):1894-907.
2. Yamamoto JF, Goodman MT. Patterns of leukemia incidence in the United States by subtype and demographic characteristics, 1997-2002. Cancer Causes Control. 2008;19(4):379-90.
3. Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an

international expert panel, on behalf of the European LeukemiaNet. Blood. 2010;115(3):453-74.
4. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. N Engl J Med. 2015;373(12):1136-52.
5. Kelly LM, Gilliland DG. Genetics of myeloid leukemias. Annu Rev Genomics Hum Genet. 2002;3:179-98.
6. Khan I, Eklund EE, AL Gartel. Therapeutic Vulnerabilities of Transcription Factors in AML. Mol Cancer Ther. 2021 Feb;20(2):229-237.
7. Look AT. Oncogenic transcription factors in the human acute leukemias. Science. 1997;278(5340):1059-64.
8. O'Mahoney JV, Guven KL, Lin J, Joya JE, Robinson CS, Wade RP, et al. Identification of a novel slow-muscle-fiber enhancer binding protein, MusTRD1. Mol Cell Biol. 1998;18(11):6641-52.
9. Perrotti D, Cesi V, Trotta R, Guerzoni C, Santilli G, Campbell K, et al. BCR-ABL suppresses C/EBPalpha expression through inhibitory action of hnRNP E2. Nat Genet. 2002;30(1):48-58.
8. Preudhomme C, Sagot C, Boissel N, Cayuela JM, Tigaud I, de Botton S, et al. Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: a study from the Acute Leukemia French Association (ALFA). Blood. 2002;100(8):2717-23.
9. Masuda T, Zhang X, Berlinicke C, Wan J, Yerrabelli A, Conner EA, et al. The transcription factor GTF2IRD1 regulates the topology and function of photoreceptors by modulating photoreceptor gene expression across the retina. J Neurosci. 2014;34(46):15356-68.
10. Ring C, Ogata S, Meek L, Song J, Ohta T, Miyazono K, et al. The role of a Williams-Beuren syndrome-associated helix-loop-helix domain-containing transcription factor in activin/nodal signaling. Genes Dev. 2002;16(7):820-35.
11. Tantin D, Tussie-Luna MI, Roy AL, Sharp PA. Regulation of immunoglobulin promoter activity by TFII-I class transcription factors. J Biol Chem. 2004;279(7):5460-9.
12. Kozel BA, Barak B,Kim CA, Mervis CB, Osborne LR, Porter M, Pober BR. Williams syndrome. Nat Rev Dis Primers. 2021 Jun 17;7(1):42.
13. Pan Z, He Y, Zhu W, Xu T, Hu X, Huang P. A Dynamic Transcription Factor Signature Along the Colorectal Adenoma-Carcinoma Sequence in Patients With Co-Occurrent Adenoma and Carcinoma. Front Oncol. 2021 May 21;11:597447.
14. Nambara S, Masuda T, Kobayashi Y, Sato K, Tobo T, Koike K, Noda M, Ogawa Y, Kuroda Y, Ito S, Eguchi H, Sugimachi K , Mimori K. GTF2IRD1 on chromosome 7 is a novel oncogene regulating the tumor-suppressor gene TGFβR2 in colorectal cancer. Cancer Sci. 2020 Feb;111(2):343-355.
15. Anuraga G, Tang W-C, Phan NN, Khoa Ta HD, Liu Y-H, Wu Y-F, Lee K-H, Wang C-Y. Comprehensive Analysis of Prognostic and Genetic Signatures for General Transcription Factor III (GTF3) in Clinical Colorectal Cancer Patients Using Bioinformatics Approaches. Curr Issues Mol Biol. 2021 Apr 27;43(1):2-20.
16. Zhuang H, Zhang C, Hou B. GTF2IRD1 overexpression promotes tumor progression and correlates with less CD8+ T cells infiltration in pancreatic cancer. Biosci Rep. 2020 Sep 30;40(9):BSR20202150.
17. Huo Y, Su T, Cai Q, Macara IG. An In Vivo Gain-of-Function Screen Identifies the Williams-Beuren Syndrome Gene GTF2IRD1 as a Mammary Tumor Promoter. Cell Rep. 2016 Jun 7;15(10):2089-2096.
18. Cicatiello L, Mutarelli M, Grober OMV, Paris O, Ferraro L, Ravo M, Tarallo R, Luo S, Schroth GP, Seifert M, Zinser C, Chiusano ML, Traini A, Bortoli M, Weisz A. Estrogen

receptor alpha controls a gene network in luminal-like breast cancer cells comprising multiple transcription factors and microRNAs. Am J Pathol. 2010 May;176(5):2113-30.
19. Bagger FO, Sasivarevic D, Sohi SH, Laursen LG, Pundhir S, Sønderby CK, Winther O, Rapin N, Porse BT. BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. Nucleic Acids Res. 2016 Jan 4;44(D1):D917-24.
20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50.
21. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics, Volume 27, Issue 12, 15 June 2011, Pages 1739–1740.
22. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser MH, Thol F, Bolli N, M.D, Gundem G, Van Loo P, Martincorena I, Ganly P, Mudie L, McLaren S, O'Meara S, Raine K, Jones DR, Teague JW, Butler AP, Greaves MF, Ganser A, Döhner K, Schlenk RF, Döhner H, Campbell PJ. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016 Jun 9; 374(23): 2209–2221.
23. Bonnet, D. and J.E. Dick, Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med, 1997. 3(7): p. 730-7.
24. Tassabehji M, Hammond P, Karmiloff-Smith A, Thompson P, Thorgeirsson SS, Durkin ME, Popescu NC, Hutton T, Metcalfe K, Rucka A, Stewart H, Read AP, Maconochie M, Donnai D. GTF2IRD1 in craniofacial development of humans and mice. Science. 2005 Nov 18;310(5751):1184-7.
25. Young EJ, Lipina T, Tam E, Mandel A, Clapcote SJ, Bechard AR, Chambers J, Mount H.T J, Fletcher PJ, Roder JC, Osborne LR. Genes Brain Behav. 2008 Mar; 7(2): 224–234.
26. Enkhmandakh B, Makeyev AV, Erdenechimeg L, Ruddle FH, Chimge NO, Tussie-Luna MI, et al. Essential functions of the Williams-Beuren syndrome-associated TFII-I genes in embryonic development. Proc Natl Acad Sci U S A. 2009;106(1):181-6.
27. Giacinti C, Giordani A. RB and cell cycle progression. Oncogene. 25, pages 5220–5227 (2006).
28. Nambara S, Masuda T, Kobayashi Y, Sato K, Tobo T, Koike K, Noda M, Ogawa Y, Kuroda Y, Ito S, Eguchi H, Sugimachi K, Mimori K. GTF2IRD1 on chromosome 7 is a novel oncogene regulating the tumor-suppressor gene TGFβR2 in colorectal cancer. Cancer Sci. 2020 Feb;111(2):343-355.
29. Bhagwat AS, Vakoc CR. Targeting Transcription Factors in Cancer. Trends Cancer. 2015 Sep 1;1(1):53-65.
30. Heshmati Y, Turkoz G, Harisankar A, Kharazi S, Bostrom J, Kamali Dolatabadi E, et al. The chromatin-remodeling factor CHD4 is required for maintenance of childhood acute myeloid leukemia. Haematologica. 2018.
31. Heshmati Y, Kharazi S, Türköz G, Chang D, Kamali Dolatabadi E, Boström J, Krstic A, Boukoura T, Wagner E, Kadri N, Månsson R , Altun M, Qian H , Walfridsson J. The histone chaperone NAP1L3 is required for haematopoietic stem cell maintenance and differentiation. Sci Rep. 2018 Jul 25;8(1):11202.
32. Aubrey BJ, Kelly GL, Kueh AJ , Brennan MS, O'Connor L , Milla L, Wilcox S , Tai L, Strasser A, Herold MJ. An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo. Cell Rep. 2015 Mar 3;10(8):1422-32.
33. Somervaille TC, Cleary ML. Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. Cancer Cell. 2006;10(4):257-68.

Supplemental Table 2. All oligos and gRNAs used in this study.

Target site	Application	Forward oligo (5´>3´)	Reverse oligo (5´>3´)
m*Gtf2ird1*	qPCR	TCGGATGTGTACCTGCTGCA	TCCCTGAGCAGCCTCTCATA
h*GTF2IRD1*	qPCR	CCAAAGACACCACGAAGCTGGA	CCACAGTCTTCAGACATGCTGC
h*UBC*	qPCR	CTGGAAGATGGTCGTACCCTG	GGTCTTGCCAGTGAGTGTCT
m*Hprt1*	qPCR	GAAAAGGACCTCTCGAAGTGTTG	CACTAATGACACAAACGTGATTCAAA
dHTS	1st PCR barcodes	TCTCTGGCAAGCAAAAGACGGCATA	TGCCATTTGTCTCGAGGTCGAGAA
dGex	2nd PCR barcodes	CAAGCAGAAGACGGCATACGAGA	AATGATACGGCGACCACCGAGA
GexSeqN	HT seq. barcodes	ACAGTCCGAAACCCCAAACGCACGAA	

Supplemental Table 3. All antibodies used in this study.

Antibody	Conjugated	Clone	Source
Anti-mouse TER-119	Purified	TER-119	Biolegend
Anti-mouse CD3	Purified	17A2	Biolegend

Anti-mouse/Human CD45R/B220	Purified	RA3-6B2	Biolegend
Anti-mouse Ly-6G/Ly-6C (Gr-1)	Purified	RB6-8C5	Biolegend
Anti-mouse NK1.1	Purified	PK136	Biolegend
Anti-mouse CD11B	Purified	M1/70	Biolegend
Anti-mouse CD16/32	Purified	93	Biolegend
Anti-mouse CD117 (c-kit)	APC-Cy7	2B8	Biolegend
Ki-67	FITC	B56	BD
Anti-human CD20	PE-Cy5	2H7	Biolegend
Anti-human CD4	PE-Cy5	RPA-T4	Biolegend
Anti-human CD8	PE-Cy5	RPA-T8	Biolegend
Anti-human CD2	PE-Cy5	RPA-2.10	Biolegend
Anti-human CD56	PE-Cy5	HCD56	Biolegend
Anti-human CD235a	PE-Cy5	HIR2	Biolegend
Anti-human CD3	PE-Cy5	HIT3a	Biolegend
Anti-human CD19	PE-Cy5	HIB19	Biolegend
Anti-human CD34	APC	581	Biolegend
Anti-human CD38	PE-Cy7	HB7	BD
Anti-human CD45	BV786	HI30	BD

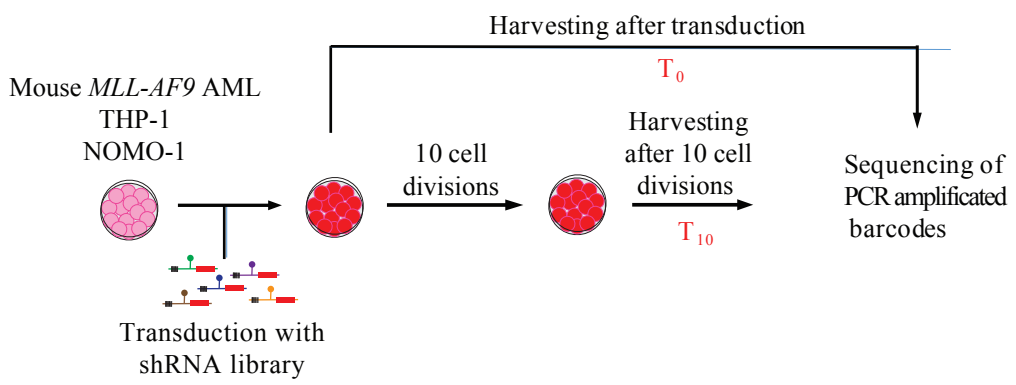
Supplemental Table 4. Primary AML patient samples related to Figure 5.

Sample	Type	ID	MLL Translocation	Experiment	Figure
AML 1	Adult	2009-022	Non MLL	*Ex vivo*	5A, B

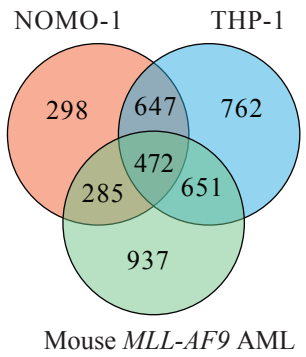
AML 2	Adult	2011-031	Non MLL	Ex vivo and *in vivo*	5A, B and E
AML 3	Adult	2011-055	Non MLL	*Ex vivo* and *in vivo*	5A, B and E
AML 4	Adult	2008-23	Non MLL	*Ex vivo*	5A, B
AML 5	Adult	2008-27	Non MLL	*Ex vivo*	5A, B
AML 6	Adult	2012-22	Non MLL	*Ex vivo*	5A
AML 7	Adult	2012-14	Non MLL	*Ex vivo*	5A
AML 8	Adult	2011-48	Non MLL	*In vivo*	5E
AML 9	Adult	2012-24	Non MLL	*In vivo*	5E
AML 10	Adult	2014-16	Non MLL	*In vivo*	5E
AML 1	Childhood	03-215	Non MLL	*Ex vivo*	5C and D
AML 2	Childhood	06-013	Non MLL	*Ex vivo*	5C and D
AML 3	Childhood	03-316	Non MLL	*In vivo*	5E

Figure 1

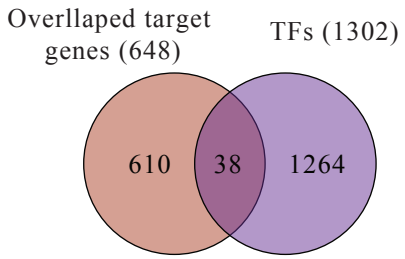
A



B



C



D

TFs	Role in AML
MAX	Unknown
CREB3L3	Unknown
GTF2IRD1	Unknown
ISL1	Unknown
NEUROG1	Unknown
FOXO4	Unknown
SSRP1	Unknown
FOXD3	Unknown
FOXA2	Unknown
MYF5	Unknown
E2F5	Known
IKZF2	Known
YY1	Known
HES1	Known
FLI1	Known
IKZF3	Known
SP1	Known
MYC	Known
MLXIPL	Known
SNAI1	Known
KEAP1	Known
MXI1	Known
PRDM1	Known
SRY	Known
STAT6	Known
HOXB4	Known
FHL2	Known
IKZF1	Known
REL	Known
MEIS1	Known
NANOG	Known

Figure 2

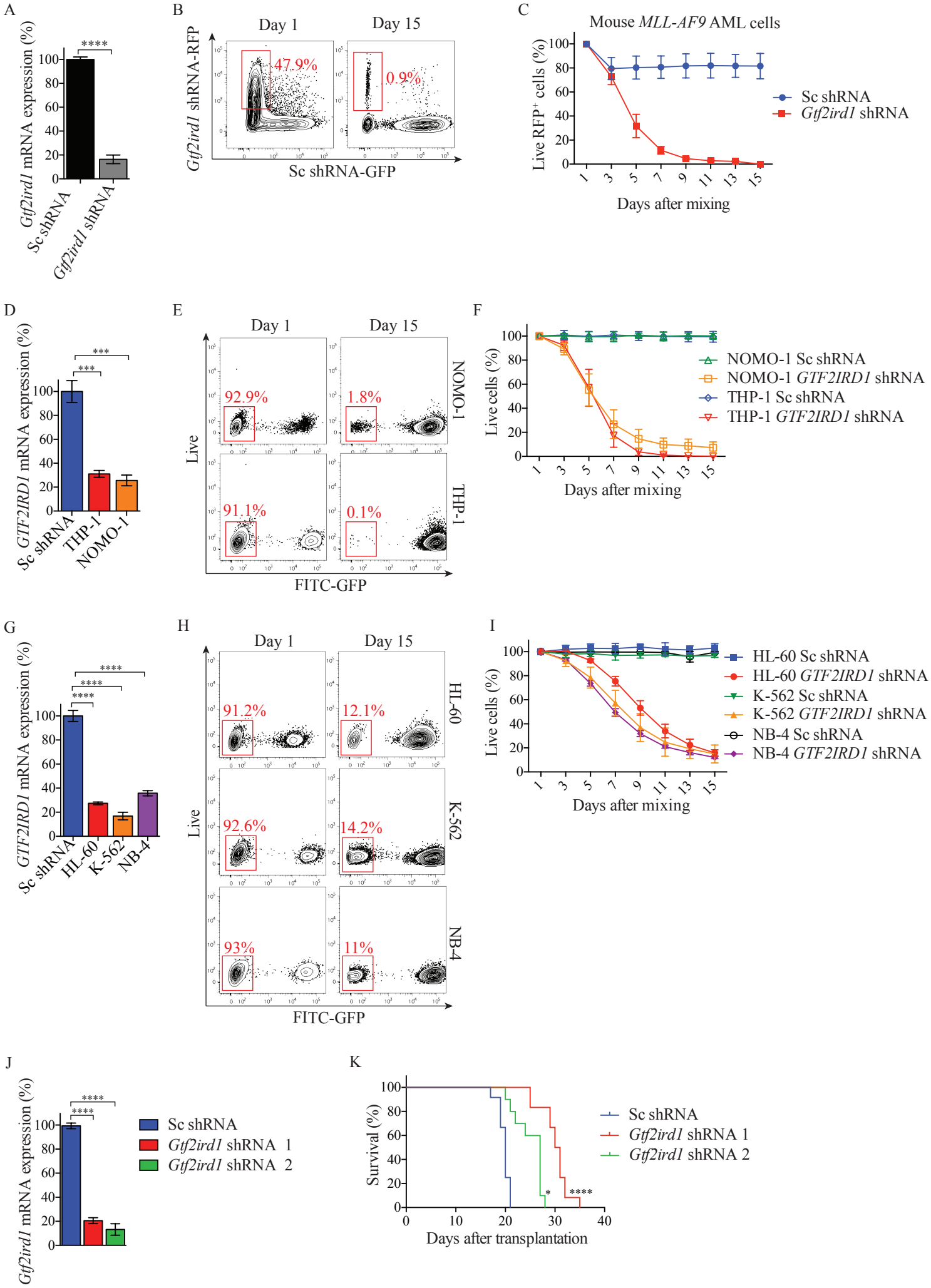

A
Gtf2ird1 mRNA expression (%)
****
Sc shRNA
Gtf2ird1 shRNA
B
Day 1
Day 15
Gtf2ird1 shRNA-RFP
47.9%
0.9%
Sc shRNA-GFP
C
Mouse MLL-AF9 AML cells
Live RFP+ cells (%)
Sc shRNA
Gtf2ird1 shRNA
Days after mixing
D
GTF2IRD1 mRNA expression (%)
***
***
Sc shRNA
THP-1
NOMO-1
E
Day 1
Day 15
Live
92.9%
1.8%
NOMO-1
91.1%
0.1%
THP-1
FITC-GFP
F
Live cells (%)
NOMO-1 Sc shRNA
NOMO-1 GTF2IRD1 shRNA
THP-1 Sc shRNA
THP-1 GTF2IRD1 shRNA
Days after mixing
G
GTF2IRD1 mRNA expression (%)
****
****
****
Sc shRNA
HL-60
K-562
NB-4
H
Day 1
Day 15
Live
91.2%
12.1%
HL-60
92.6%
14.2%
K-562
93%
11%
NB-4
FITC-GFP
I
Live cells (%)
HL-60 Sc shRNA
HL-60 GTF2IRD1 shRNA
K-562 Sc shRNA
K-562 GTF2IRD1 shRNA
NB-4 Sc shRNA
NB-4 GTF2IRD1 shRNA
Days after mixing
J
Gtf2ird1 mRNA expression (%)
****
****
Sc shRNA
Gtf2ird1 shRNA 1
Gtf2ird1 shRNA 2
K
Survival (%)
Sc shRNA
Gtf2ird1 shRNA 1
Gtf2ird1 shRNA 2
*
****
Days after transplantation

Figure 3

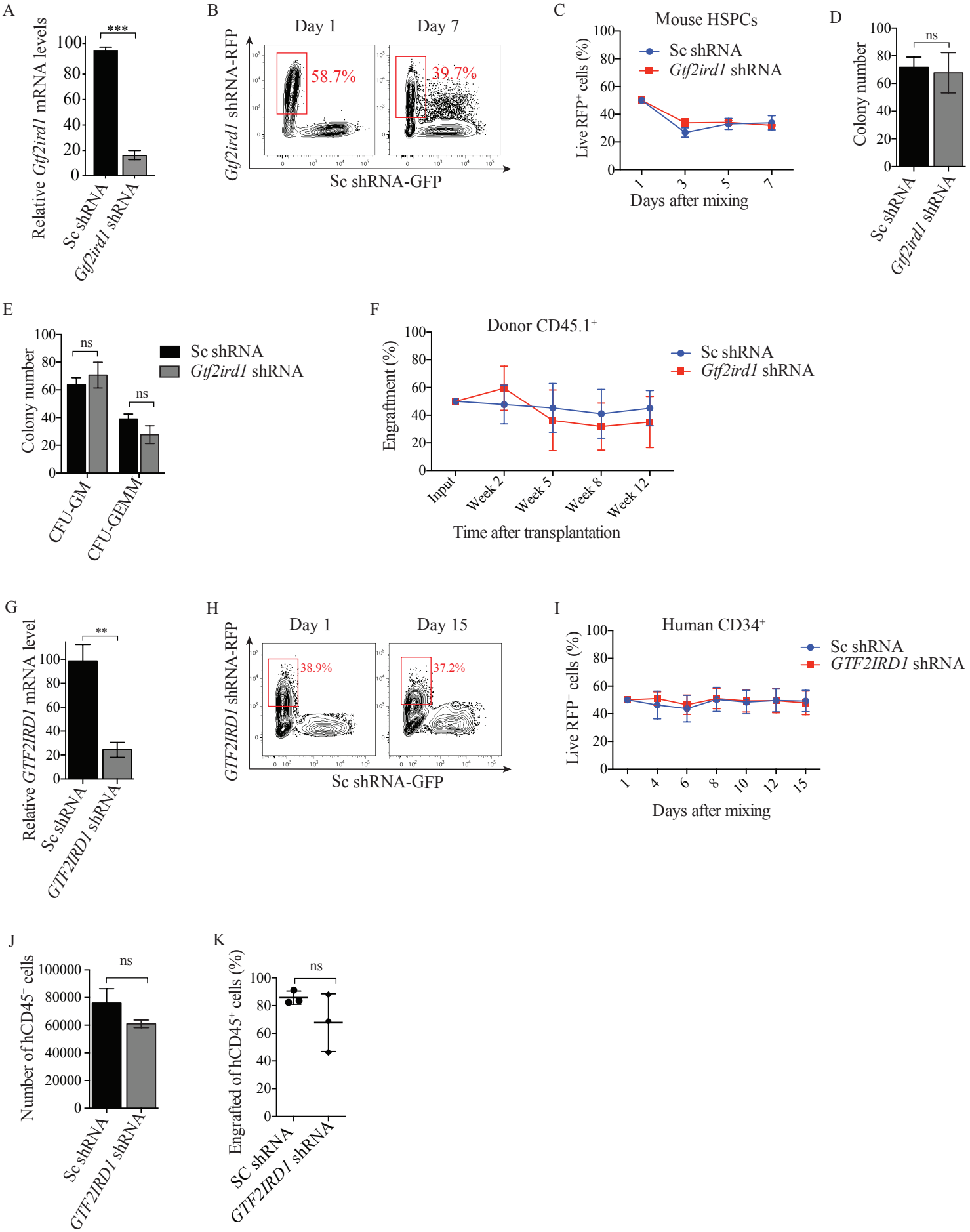

A
Relative Gtf2ird1 mRNA levels
***
Sc shRNA
Gtf2ird1 shRNA
B
Day 1
Day 7
Gtf2ird1 shRNA-RFP
58.7%
39.7%
Sc shRNA-GFP
C
Mouse HSPCs
Live RFP+ cells (%)
Sc shRNA
Gtf2ird1 shRNA
Days after mixing
D
Colony number
ns
Sc shRNA
Gtf2ird1 shRNA
E
Colony number
ns
ns
Sc shRNA
Gtf2ird1 shRNA
CFU-GM
CFU-GEMM
F
Donor CD45.1+
Engraftment (%)
Sc shRNA
Gtf2ird1 shRNA
Input
Week 2
Week 5
Week 8
Week 12
Time after transplantation
G
Relative GTF2IRD1 mRNA level
**
Sc shRNA
GTF2IRD1 shRNA
H
Day 1
Day 15
GTF2IRD1 shRNA-RFP
38.9%
37.2%
Sc shRNA-GFP
I
Human CD34+
Live RFP+ cells (%)
Sc shRNA
GTF2IRD1 shRNA
Days after mixing
J
Number of hCD45+ cells
ns
Sc shRNA
GTF2IRD1 shRNA
K
Engrafted of hCD45+ cells (%)
ns
SC shRNA
GTF2IRD1 shRNA

Figure 4

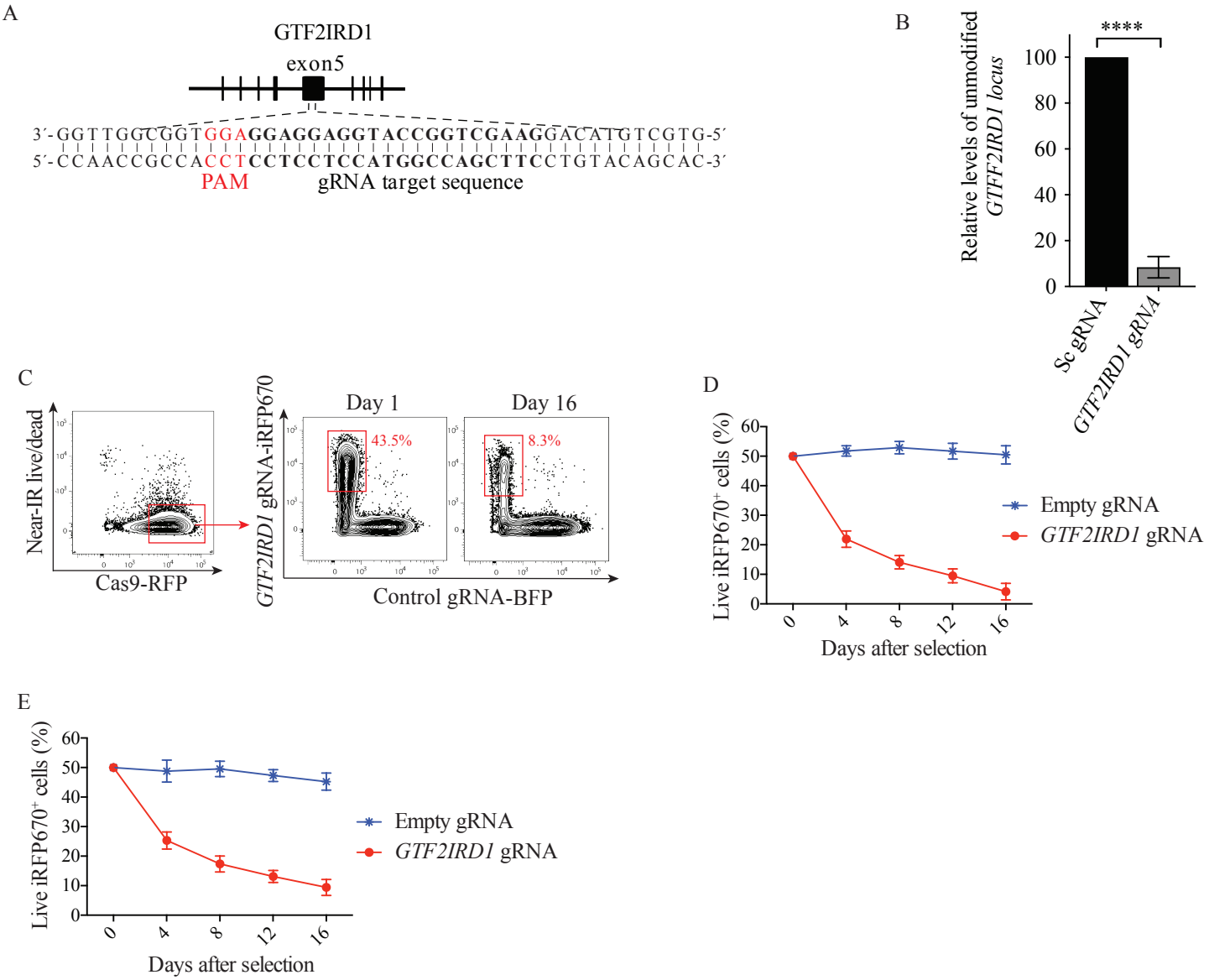
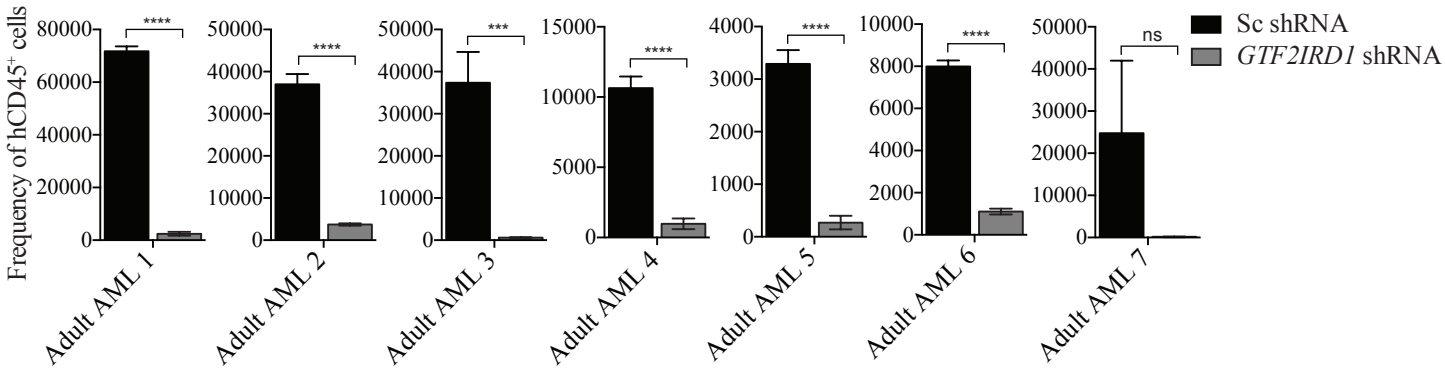
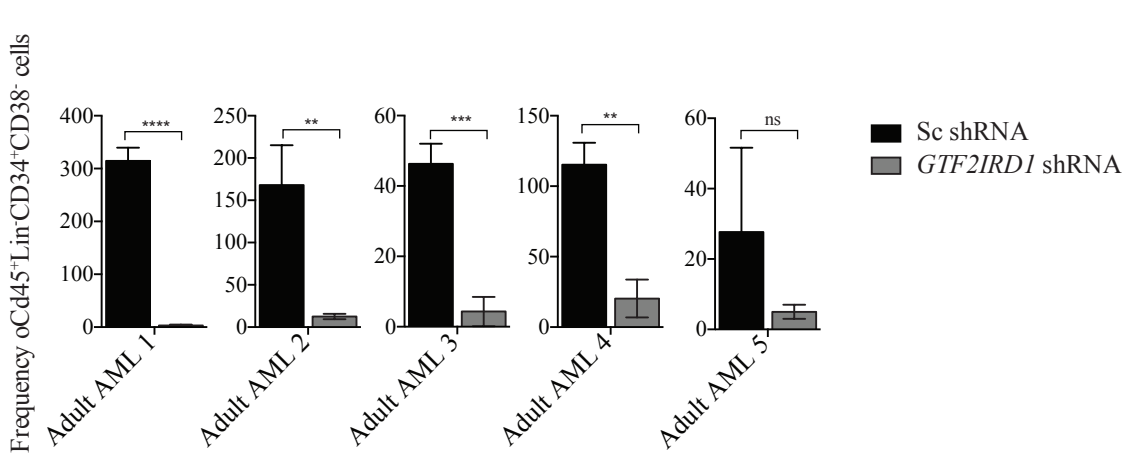

A
GTF2IRD1
exon5
3′-GGTTGGCGGTGGAGGAGGAGGTACCGGTCGAAGGACATGTCGTG-5′
5′-CCAACCGCCACCTCCTCCTCCATGGCCAGCTTCCTGTACAGCAC-3′
PAM
gRNA target sequence
B
Relative levels of unmodified GTFF2IRD1 locus
****
100
80
60
40
20
0
Sc gRNA
GTF2IRD1 gRNA
C
Near-IR live/dead
Cas9-RFP
GTF2IRD1 gRNA-iRFP670
Day 1
Day 16
43.5%
8.3%
Control gRNA-BFP
D
Live iRFP670+ cells (%)
60
50
40
30
20
10
0
0
4
8
12
16
Days after selection
Empty gRNA
GTF2IRD1 gRNA
E
Live iRFP670+ cells (%)
60
50
40
30
20
10
0
0
4
8
12
16
Days after selection
Empty gRNA
GTF2IRD1 gRNA

Figure 5

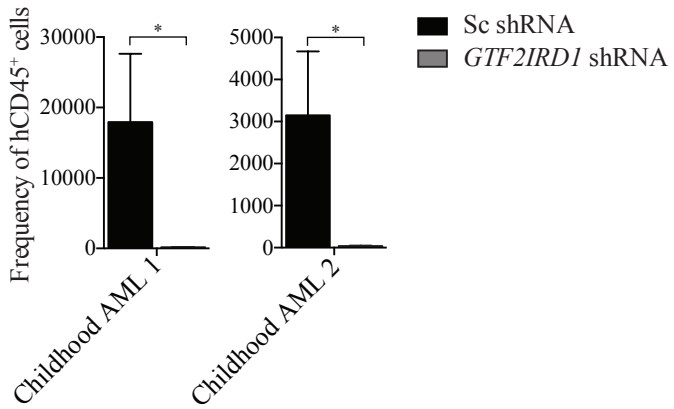
A



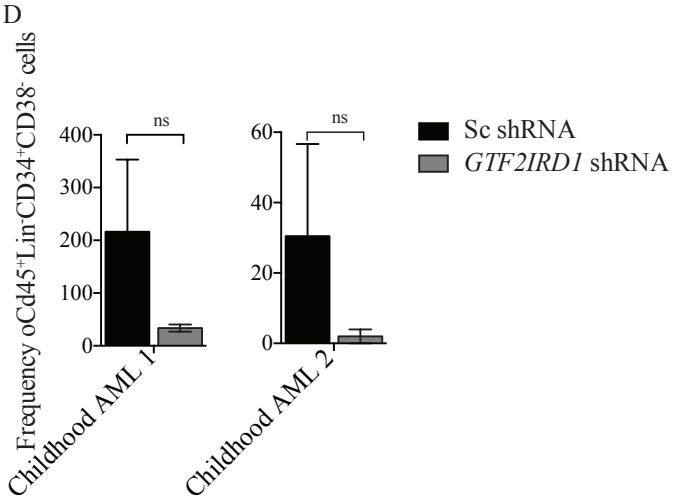
B



C



D



E

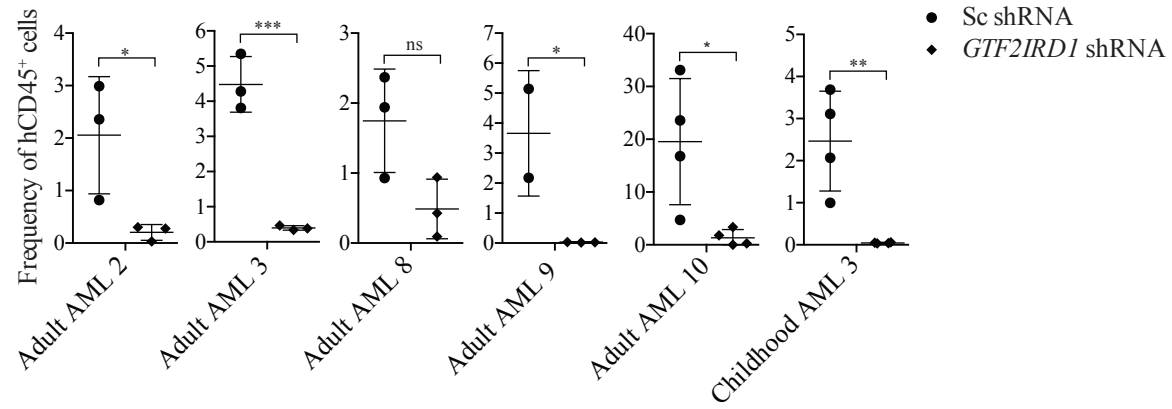
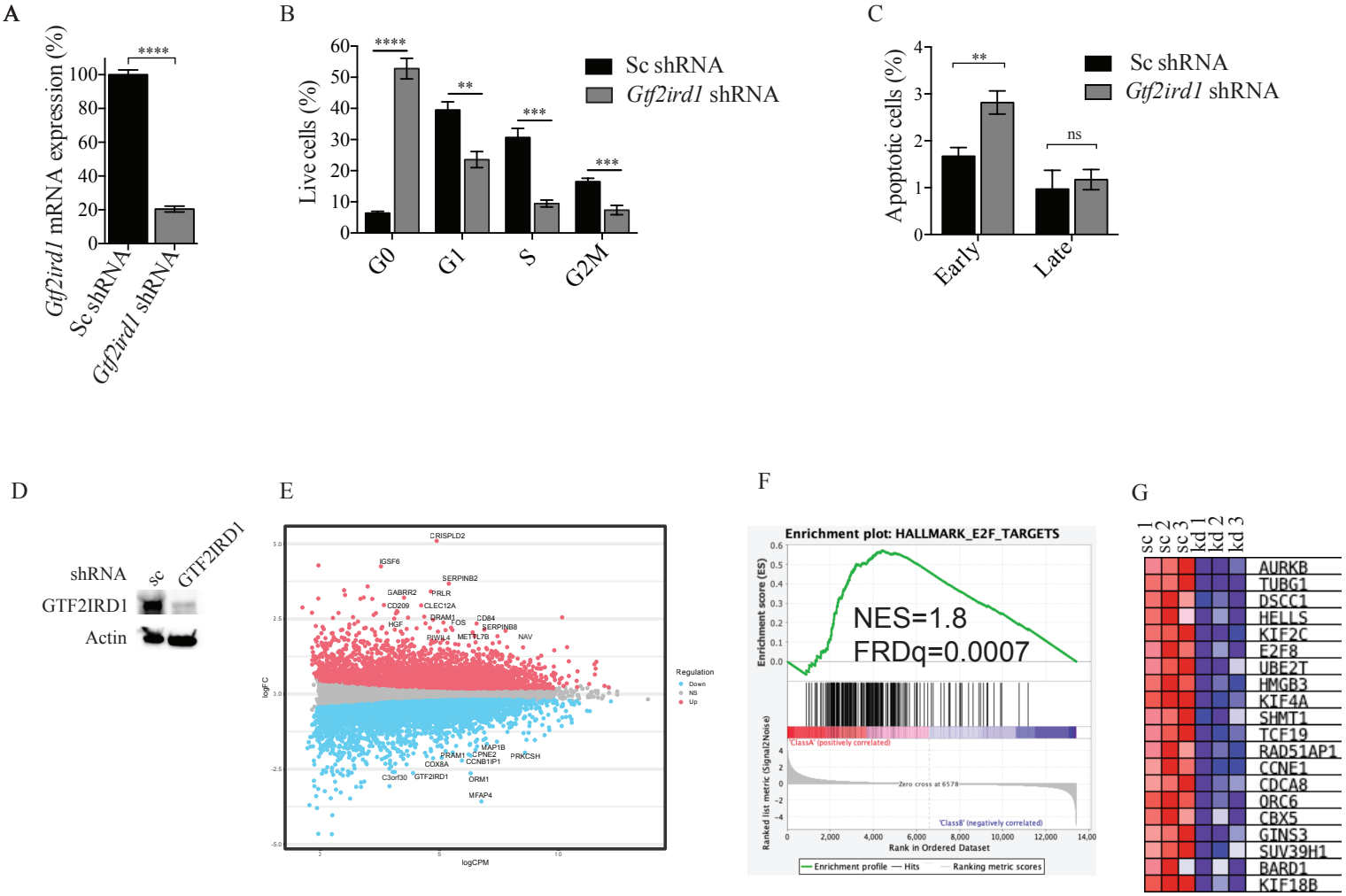


Figure 6


A
Gtf2ird1 mRNA expression (%)
Sc shRNA
Gtf2ird1 shRNA
****
B
Live cells (%)
Sc shRNA
Gtf2ird1 shRNA
G0
G1
S
G2M
****
**
***
***
C
Apoptotic cells (%)
Sc shRNA
Gtf2ird1 shRNA
Early
Late
**
ns
D
shRNA
sc
GTF2IRD1
GTF2IRD1
Actin
E
logFC
logCPM
Regulation
Down
NS
Up
F
Enrichment plot: HALLMARK_E2F_TARGETS
NES=1.8
FRDq=0.0007
Enrichment score (ES)
Ranked list metric (Signal2Noise)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
G
sc 1
sc 2
sc 3
kd 1
kd 2
kd 3
AURKB
TUBG1
DSCC1
HELLS
KIF2C
E2F8
UBE2T
HMGB3
KIF4A
SHMT1
TCF19
RAD51AP1
CCNE1
CDCA8
ORC6
CBX5
GINS3
SUV39H1
BARD1
KIF18B