nature immunology

ARTICLES

https://doi.org/10.1038/s41590-021-01079-w

Identification of RIOK2 as a master regulator of human blood cell development

Shrestha Ghosh[1,2], Mahesh Raundhal[1,2,8], Samuel A. Myers[3,4], Steven A. Carr[3], Xi Chen[5], Gregory A. Petsko[6] ✉ and Laurie H. Glimcher[1,2,7] ✉

Anemia is a major comorbidity in aging, chronic kidney and inflammatory diseases, and hematologic malignancies. However, the transcriptomic networks governing hematopoietic differentiation in blood cell development remain incompletely defined. Here we report that the atypical kinase RIOK2 (right open reading frame kinase 2) is a master transcription factor (TF) that not only drives erythroid differentiation, but also simultaneously suppresses megakaryopoiesis and myelopoiesis in primary human stem and progenitor cells. Our study reveals the previously uncharacterized winged helix-turn-helix DNA-binding domain and two transactivation domains of RIOK2 that are critical to regulate key hematopoietic TFs GATA1, GATA2, SPI1, RUNX3 and KLF1. This establishes RIOK2 as an integral component of the transcriptional regulatory network governing human hematopoietic differentiation. Importantly, *RIOK2* mRNA expression significantly correlates with these TFs and other hematopoietic genes in myelodysplastic syndromes, acute myeloid leukemia and chronic kidney disease. Further investigation of RIOK2-mediated transcriptional pathways should yield therapeutic approaches to correct defective hematopoiesis in hematologic disorders.

Anemia is a hallmark of a plethora of hematologic disorders associated with aging, chronic diseases such as renal failure and inflammation, bone marrow failure and myeloid neoplasms[1–3]. Aberrant red blood cell differentiation (erythropoiesis) underlies anemias and can be accompanied by myeloid proliferation. For example, myelodysplastic syndromes (MDS), a heterogeneous group of clonal hematologic disorders, are classically characterized by anemia and myeloproliferation[4]. The average survival time following diagnosis of MDS is 3 years[5] owing to few treatment options, and roughly 20–30% of patients with MDS progress to acute myeloid leukemia (AML)[6]. Hence, the substantial risks of allogeneic bone marrow transplants in elderly patients, together with a dearth of effective US Food and Drug Administration (FDA)-approved drugs make it imperative to revisit the origins of blood cell differentiation (hematopoiesis) defects underlying anemia and myeloproliferation to identify new druggable targets[7,8].

RIOK2 is an atypical serine threonine kinase that plays important roles in the final maturation steps of the pre-40S ribosomal complex to facilitate cytoplasmic translation[9–11]. Our recent study revealed that hematopoietic cell-specific heterozygous deletion of *Riok2* leads to anemia and myeloproliferation in mice[12], reminiscent of MDS-associated phenotypes in patients[6,13]. We also noted reduced *RIOK2* mRNA levels in patients with the del(5q) subtype of MDS as compared to healthy individuals[12,14]. However, the mechanisms underlying RIOK2-driven hematopoietic differentiation remain largely unexplored. We noted the presence of a putative DNA-binding winged helix-turn-helix (wHTH) domain[15] next to the RIO (kinase) domain in RIOK2 (refs. [9,16]). Because wHTH domains form key structural motifs in both prokaryotic and eukaryotic bona fide TFs[17], we asked whether this atypical kinase might also function as a TF in hematopoietic differentiation to regulate lineage fate determination. Remarkably, we discover that RIOK2 is a master transcriptional regulator of hematopoietic lineage commitment and that its ablation drives primary human hematopoietic stem and progenitor cells (HSPCs) toward MDS-associated hematopoietic differentiation defects. Our transcriptomic profiling, structural modeling, chromatin immunoprecipitation assays with sequencing (ChIP–seq), assay for transposase-accessible chromatin with high-throughput sequencing (ATAC–seq) and structure–function domain studies using deletion mutants revealed that RIOK2 regulates specific genetic programs in hematopoiesis via its previously unappreciated wHTH DNA-binding domain and two transactivation domains (TADs). Mechanistically, RIOK2 transcriptionally modifies the expression of key lineage-specific TFs, such as GATA1, GATA2, SPI1, RUNX3 and KLF1 to fine-tune lineage fate determination in primary human hematopoietic stem cells. We further demonstrate that GATA1 and RIOK2 function in a positive feedback loop to drive erythroid differentiation. These discoveries thus present therapeutic opportunities to correct hematopoietic differentiation defects in a range of hematologic disorders. To the best of our knowledge, we present the first evidence of a protein that not only controls 40S ribosome biogenesis governing translation but also functions in the nucleus as a master TF by regulating the expression of key TFs that determine hematopoietic stem cell fate.

Results

RIOK2 regulates differentiation of hematopoietic lineages. In agreement with our previous preclinical study in mouse models[12], we observed that short hairpin RNA (shRNA)-mediated knockdown (KD) of RIOK2 led to a significant decline in erythroid

[1]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [2]Department of Immunology, Harvard Medical School, Boston, MA, USA. [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]La Jolla Institute for Immunology, La Jolla, CA, USA. [5]Department of Molecular & Cellular Biology, Lester and Sue Smith Breast Center, Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA. [6]Department of Neurology, Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [7]Department of Medicine, Harvard Medical School, Boston, MA, USA. [8]Present address: Jnana Therapeutics, Boston, MA, USA. ✉e-mail: gpetsko@bwh.harvard.edu; laurie_glimcher@dfci.harvard.edu

differentiation in two different human erythroleukemia cell lines: TF-1 and K562 (Fig. 1a–d). Partial loss of RIOK2 modestly, but significantly, inhibited progression from the proerythroblast-like stage ($CD71^+CD235a^-$) to the more differentiated stage ($CD71^+CD235a^+$)[18] in both cell lines (Extended Data Fig. 1a,b). Additionally, we observed a concomitant skewing toward megakaryopoiesis in RIOK2-KD cells (Extended Data Fig. 1c,d). To validate this observation and its correlation with the emergence of hematologic disorders, we used healthy donor-derived primary human HSPCs, which are the cells of origin for hematologic neoplasms in patients[19]. Using CRISPR–Cas9 genome editing, we knocked down or knocked out (KO) RIOK2 (guide RNAs targeting LacZ used as control) in primary human HSPCs (Fig. 1e). Loss of RIOK2 in HSPCs not only blocked erythroid lineage commitment; it also increased megakaryopoiesis and myelopoiesis (Fig. 1f,g and Extended Data Figs. 1e,f and 10). Depletion of RIOK2 led to substantial retention of HSPCs in the uncommitted stage (Q4) rather than passive progression to other lineages (Fig. 1f), suggesting that RIOK2 is actively involved in regulating early hematopoietic stem cell differentiation. Analysis of cell pellets in RIOK2-proficient versus RIOK2-deficient HSPCs underscored the significance of RIOK2 in driving erythroid differentiation (Fig. 1h). To further define the direct roles of RIOK2 in myelopoiesis and megakaryopoiesis, we cultured HSPCs in differentiation media that selectively promoted myeloid or megakaryocytic differentiation. Consistently, loss of RIOK2 markedly elevated myelopoiesis and megakaryopoiesis (Fig. 1i and Extended Data Fig. 2a–d). To investigate if RIOK2 functions further upstream in the differentiation pathway, we performed colony-forming unit (CFU) assays using primary human HSPCs. Loss of RIOK2 substantially impaired burst-forming unit-erythroid (BFU-E) and colony-forming unit-erythroid (CFU-E) formation and concomitantly increased megakaryocytic (CFU-Mk) and granulocyte–macrophage (CFU-GM) progenitors (Fig. 1j,k and Extended Data Fig. 2e,f). Quantitative mass spectrometry-based proteomics further validated that loss of RIOK2 impedes expression of erythroid lineage proteins (HBB, HBA1, SPTA1, TFRC and FECH) while promoting the expression of megakaryocytic and myeloid-lineage-specific proteins (ITGA2B, ITGAM, THPO, PPBP and FYB1; Fig. 1l and Extended Data Fig. 2g,h). In agreement with published studies[9–11], gene-set enrichment analysis (GSEA) confirmed defects in ribosome biogenesis upon loss of RIOK2 (Extended Data Fig. 2i). Given that reduced ribosome levels selectively inhibit erythroid differentiation[20], we believe that the kinase activity of RIOK2 in 40S ribosome biogenesis and translation is also required for erythroid differentiation. However, it seemed unlikely that the significant reduction of *RIOK2* mRNA in del(5q) MDS[12,14] could be explained by defective ribosome biogenesis. One explanation for reduced *RIOK2* mRNA levels, however, could be control of *RIOK2* transcription by other factors.

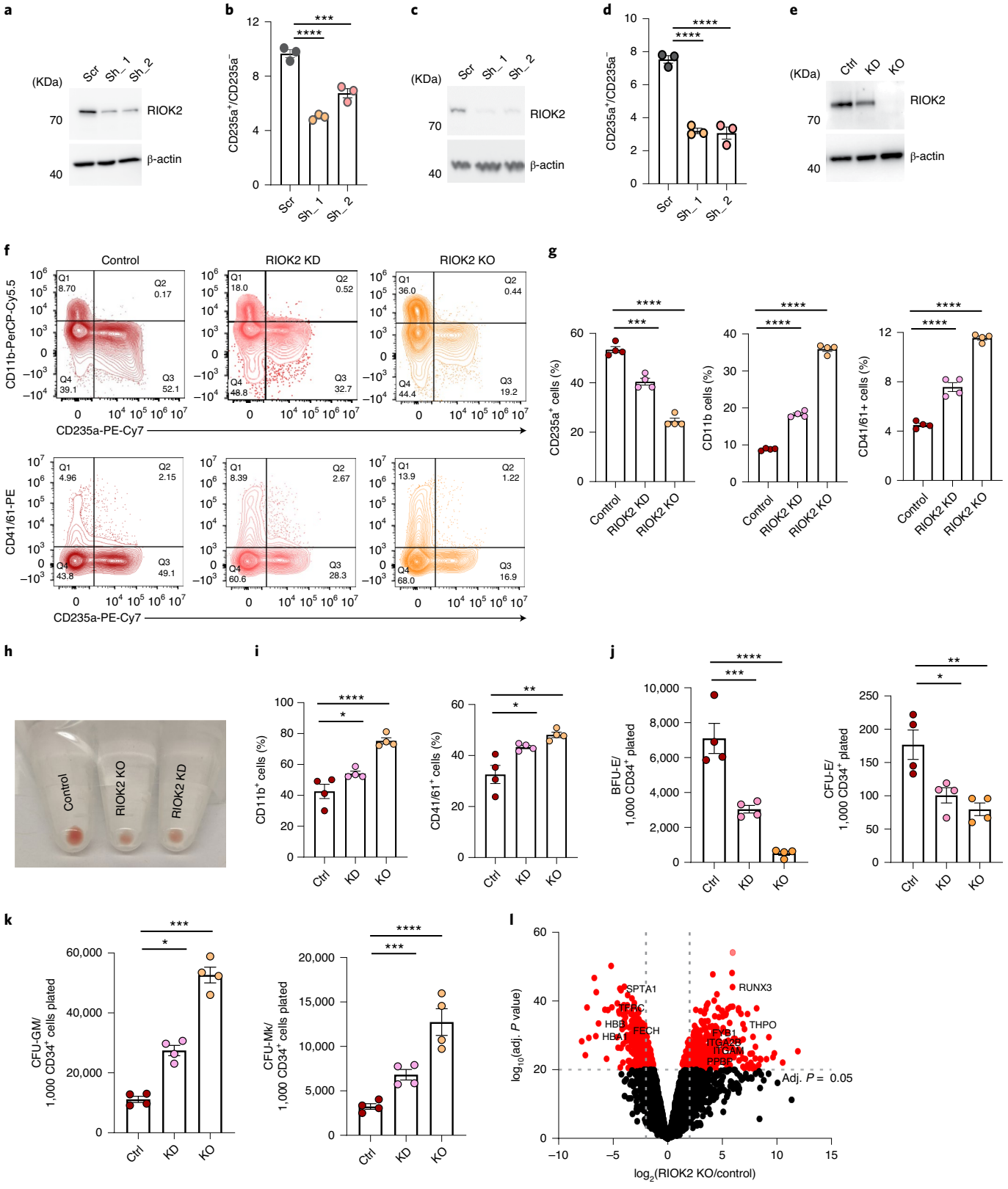
GATA1 and RIOK2 form a positive feedback loop. GATA1 is the most critical known transcriptional regulator of erythropoiesis[21,22]. Given the integral roles of RIOK2 in hematopoietic differentiation, we sought to identify whether RIOK2 itself was regulated by GATA1 in hematopoietic stem cells. Analysis of the promoter region of *RIOK2* indeed revealed a GATA binding motif (Extended Data Fig. 3a). Among GATA family members involved in hematopoiesis (GATA1, GATA2 and GATA3)[23], only GATA1 robustly drove RIOK2 expression in luciferase reporter assays (Extended Data Fig. 3b). ChIP using two different monoclonal antibodies against GATA1 confirmed its binding to the *RIOK2* promoter (Fig. 2a,b). Mutation of the GATA binding site in the *RIOK2* promoter markedly reduced GATA1-driven expression of RIOK2 (Fig. 2c,d). Additionally, we observed a positive correlation between *GATA1* and *RIOK2* mRNA levels when GATA1 was either reduced or overexpressed (Fig. 2e and Extended Data Fig. 3c). Overexpression of GATA1 partially restored *RIOK2* mRNA levels and erythroid differentiation in RIOK2-KD cells (Fig. 2f–h), indicating that defects in erythroid differentiation due to loss of RIOK2 are only partially driven by GATA1. To address this, we compared hematopoietic differentiation patterns upon loss of GATA1 and RIOK2 in primary human HSPCs (Extended Data Fig. 3d,e). Although GATA1 was more potent than RIOK2 in driving erythroid differentiation, RIOK2 was much more potent than GATA1 in regulating myelopoiesis and megakaryopoiesis (Fig. 2i and Extended Data Fig. 3f). Unexpectedly, we also observed decreased GATA1 protein expression upon gradual loss of RIOK2 (Fig. 2j,k). Our findings indicate a positive feedback loop between RIOK2 and GATA1 in driving hematopoietic differentiation; however, it remained unclear how RIOK2 drove GATA1 expression (Extended Data Fig. 3g). These data along with the previously reported dynamic localization of RIOK2 in the nucleus[10] strongly suggested that RIOK2's regulation of lineage commitment could not be wholly explained by its known functions in 40S ribosome biogenesis and cytoplasmic translation.

RIOK2 regulates key hematopoietic lineage-specific transcription factors. To gain deeper mechanistic insights into RIOK2's involvement in hematopoietic differentiation, we performed bulk RNA-sequencing (RNA-seq) analysis in RIOK2 KD and KO primary human HSPCs (Fig. 3a,b). Consistent with full versus partial RIOK2 depletion, the KO versus control group generated more differentially expressed genes than KD versus control group (Extended Data Fig. 4a). We further pursued significantly altered genes overlapping in both KD and KO cells. Gene expression profiling clearly indicated a block in erythroid lineage commitment and a concomitant elevation in myelopoiesis and megakaryopoiesis-associated gene signatures with dose-dependent loss of RIOK2 (Fig. 3a,b). GSEA confirmed this observation (Fig. 3c,d).

Notably, *GATA1* mRNA levels were significantly downregulated upon RIOK2 depletion (Fig. 3a,e), explaining the previously observed diminution in GATA1 protein levels in RIOK2-depleted HSPCs (Fig. 2j,k). Remarkably, loss of RIOK2 also altered the expression profiles of additional key TFs that determine hematopoietic lineage commitment: specifically, *GATA1* and *KLF1* (refs. [24,25]) were downregulated, whereas *GATA2*, *SPI1* and *RUNX3* (refs. [26–28]) were upregulated, among others (Fig. 3a,e). We next investigated if RIOK2-driven expression of the identified TFs promoted erythroid differentiation. Overexpression of RIOK2 partially but significantly

Fig. 1 | RIOK2 drives erythropoiesis and suppresses megakaryopoiesis and myelopoiesis. a, Representative immunoblot showing RIOK2 KD in TF-1 cells; sh_1 and sh_2, shRIOK2_1 and shRIOK2_2. **b**, Ratio of erythroid ($CD235a^+$) to non-erythroid ($CD235a^-$) TF-1 cells after differentiation; scr, scrambled. **c**, Immunoblot showing shRNA-mediated RIOK2 KD in K562 cells. **d**, Ratio of erythroid ($CD235a^+$) to non-erythroid ($CD235a^-$) K562 cells after differentiation. **e**, Immunoblot showing RIOK2 expression after KD and KO of RIOK2 in HSPCs; ctrl, control. **f**, Flow plots depicting erythropoiesis (CD235a), megakaryopoiesis (CD41/CD61) and myelopoiesis (CD11b) in differentiating HSPCs after KD and KO of RIOK2. **g**, Quantification of data presented in **f**. **h**, Cell pellets of RIOK2-sufficient and RIOK2-deficient HSPCs after 14 d of erythroid differentiation. **i**, Selective differentiation of control versus RIOK2 KD and KO HSPCs toward myeloid and megakaryocytic lineages. **j**,**k**, BFU-E, CFU-E, CFU-GM and CFU-Mk cells formed in control versus RIOK2 KD and KO HSPCs. **l**, Volcano plot showing differentially expressed proteins in control versus RIOK2-depleted HSPCs. $n = 4$ independent donors in **g** and **i**–**k**, $n = 3$ technical replicates in **b** and **d**. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, one-way analysis of variance (ANOVA) with Tukey's or Dunnett's correction. Data are represented as the mean ± s.e.m. Data in **a**, **c** and **e** are representative of three independent experiments.

increased *GATA1* mRNA levels and erythroid differentiation in GATA1 KD cells (Fig. 3f,g). Consistent with the established roles of key myeloid-lineage TF PU.1 (encoded by *SPI1*) in blocking erythroid differentiation[29,30], we noted that KD of SPI1 partially rescued erythroid differentiation in RIOK2-KD cells (Fig. 3h,i). RUNX3 is another important hematopoietic TF whose elevated expression in patients with MDS negatively correlates with median survival[27]. In line with this, we observed increased expression of RUNX3 upon RIOK2 deficiency (Fig. 3a,e) and noted as well that KD of RUNX3 partly rescued erythroid differentiation in RIOK2-KD cells (Fig. 3j,k). However, ablation of GATA2, a key hematopoietic TF[31], did not rescue erythroid differentiation in RIOK2-KD cells



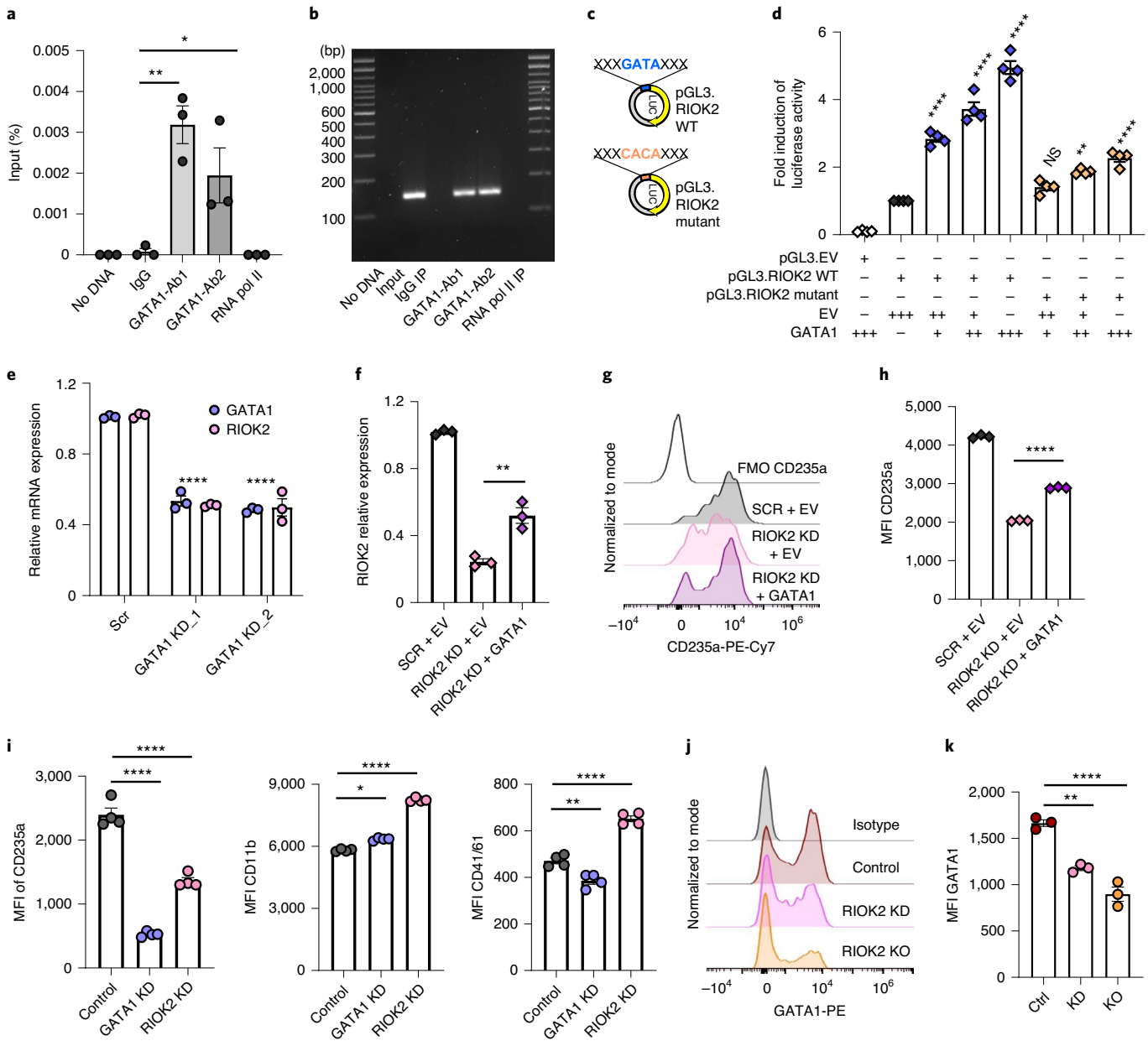


Fig. 2 | RIOK2 and GATA1 function in a positive feedback loop to regulate hematopoietic differentiation. a, Graph showing relative binding of GATA1 to the *RIOK2* promoter using two different monoclonal antibodies against GATA1—negative control anti-rabbit immunoglobulins (IgG) and antibodies against RNA polymerase II. **b**, Agarose gel picture of bands quantified in **a**. **c**, Diagram showing incorporation of WT (WT: GATA) or mutant (CACA) *RIOK2* promoter in basic pGL3.1 luciferase reporter plasmid. **d**, Quantification of WT (WT: GATA) and mutant (CACA) *RIOK2*-promoter-driven luciferase activity in response to a dose-dependent increase in GATA1 expression in HEK293 cells. **e**, Quantitative PCR with reverse transcription (RT–qPCR) assessment of *GATA1* and *RIOK2* mRNA expression (normalized to actin) upon KD of GATA1 using two different CRISPR RNAs (crRNAs). **f**, *RIOK2* mRNA levels after overexpression of GATA1 in RIOK2-KD cells. **g**, Fluorescence-activated cell sorting (FACS) plots showing erythroid differentiation (CD235a expression) in RIOK2-KD cells expressing either empty vector (EV) or ectopic GATA1. **h**, Quantification of data presented in **g**. **i**, Quantification of erythroid (CD235a), myeloid (CD11b) and megakaryocytic (CD41/CD61) progression on day 5 of differentiating HSPCs after KD of GATA1 and RIOK2, respectively. **j**, FACS plot depicting GATA1 protein expression in HSPCs upon KD or KO of RIOK2. **k**, Quantification of data presented in **j**. $n = 3$ independent experiments in **a**, **e**, **f** and **h**; $n = 4$ technical replicates in **d**; $n = 4$ and 3 independent donors in **i** and **k**, respectively. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; NS, not significant, one-way ANOVA with Tukey's or Dunnett's correction. Data are represented as the mean ± s.e.m. Data in **b** are representative of three independent experiments.

(Extended Data Fig. 4b,c). Consistent with the roles of GATA2 to promote megakaryopoiesis in the absence of GATA1 (ref. [32]), KD of GATA2 more robustly restored megakaryopoiesis to baseline levels in RIOK2-KD cells than suppression of SPI1 or RUNX3 (Extended Data Fig. 4d). We conclude that RIOK2 regulates the expression of multiple TFs critical for hematopoietic lineage commitment.

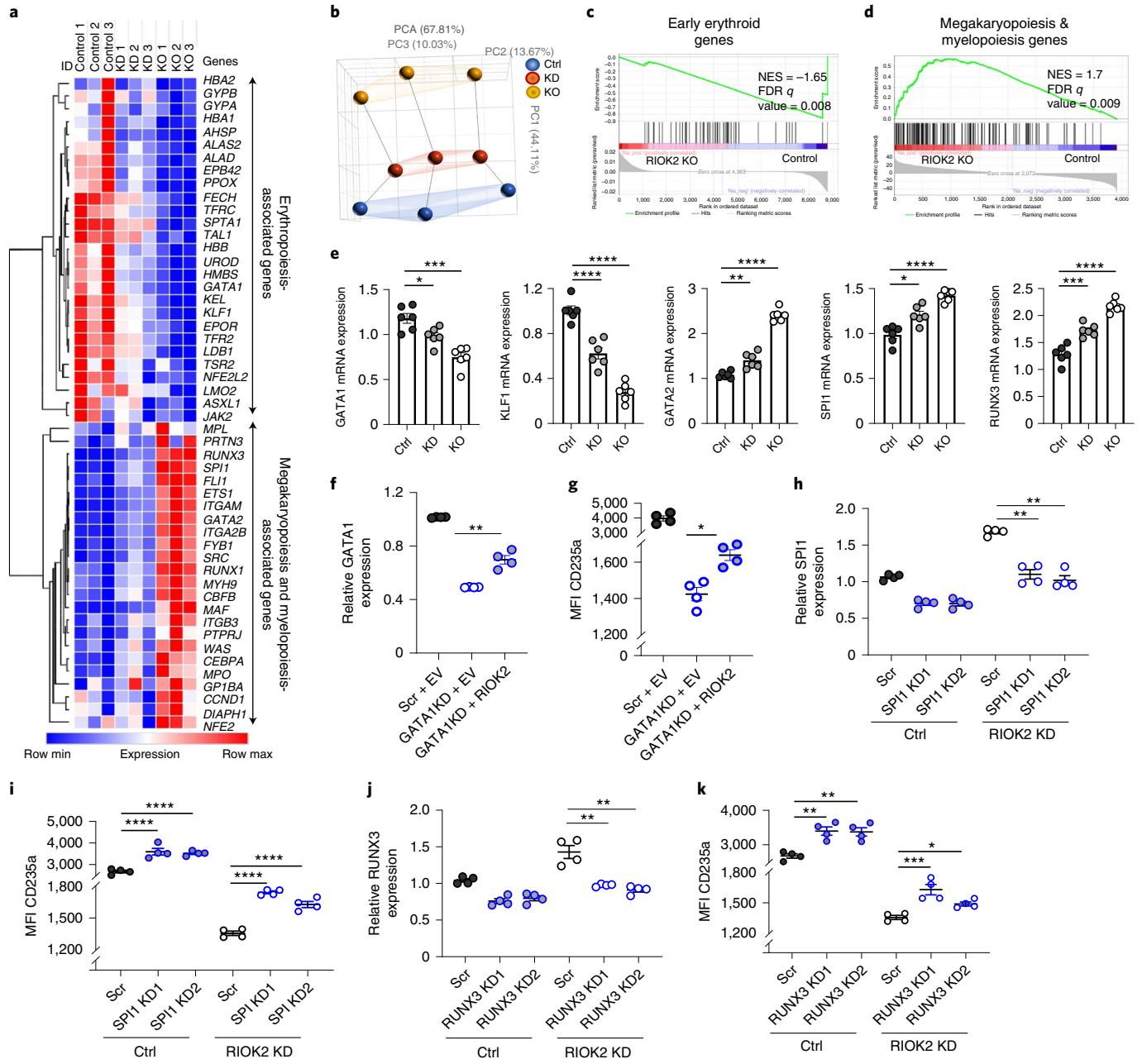


Fig. 3 | RIOK2 regulates the expression of key transcription factors in hematopoiesis. a, Heat map of differentially expressed genes (cutoff, adjusted *P* value < 0.05) from bulk RNA-seq of primary human HSPCs with KD and KO of RIOK2; $n = 3$ donors. **b**, Principal-component analysis (PCA) of data presented in **a**. Lines joining data points indicate individual donors. **c**,**d**, GSEA plots of early erythropoiesis-, megakaryopoiesis- and myelopoiesis-associated genes in RIOK2-depleted versus control HSPCs, respectively. **e**, *GATA1*, *GATA2*, *SPI1*, *RUNX3* and *KLF1* mRNA expression (normalized to actin) in HSPCs upon KD and KO of RIOK2; $n = 6$ independent donors. **f**, *GATA1* mRNA level (normalized to actin) in GATA1 KD cells ectopically expressing EV or RIOK2; SCR, scrambled. **g**, Erythroid progression in GATA1 KD cells with ectopic expression of EV or RIOK2. **h**, *SPI1* mRNA level (normalized to actin) in control versus RIOK2-KD cells after inhibition of SPI1 with two different crRNAs. **i**, Erythroid progression in scrambled versus RIOK2-KD cells with or without KD of SPI1 using two different crRNAs. **j**, *RUNX3* mRNA level (normalized to actin) in control versus RIOK2-KD cells after knocking down RUNX3 with two different crRNAs. **k**, Erythroid progression in scrambled versus RIOK2-KD cells with or without KD of RUNX3 using two different crRNAs; $n = 4$ technical replicates in **f**–**k**. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; one-way ANOVA with Tukey's or Dunnett's correction. Data are represented as the mean ± s.e.m.

RIOK2 binds to transcription start sites of its target genes. Given the notable alteration of transcriptomic profiles in RIOK2-depleted HSPCs as compared to control HSPCs (Fig. 3a), we wondered whether loss of RIOK2 led to global changes in chromatin accessibility. To address this, we performed ATAC–seq in differentiating RIOK2-proficient and RIOK2-deficient primary human HSPCs with comparable mapped reads (Extended Data Fig. 4e). Indeed, loss of RIOK2 led to dramatic reduction in chromatin accessibility at the promoter, intronic and intergenic regions (Fig. 4a). Transcription start site (TSS) and gene plots further indicated an inhibition in chromatin accessibility at the promoter regions of genes in the absence of RIOK2 (Fig. 4b,c; note the maximum signal

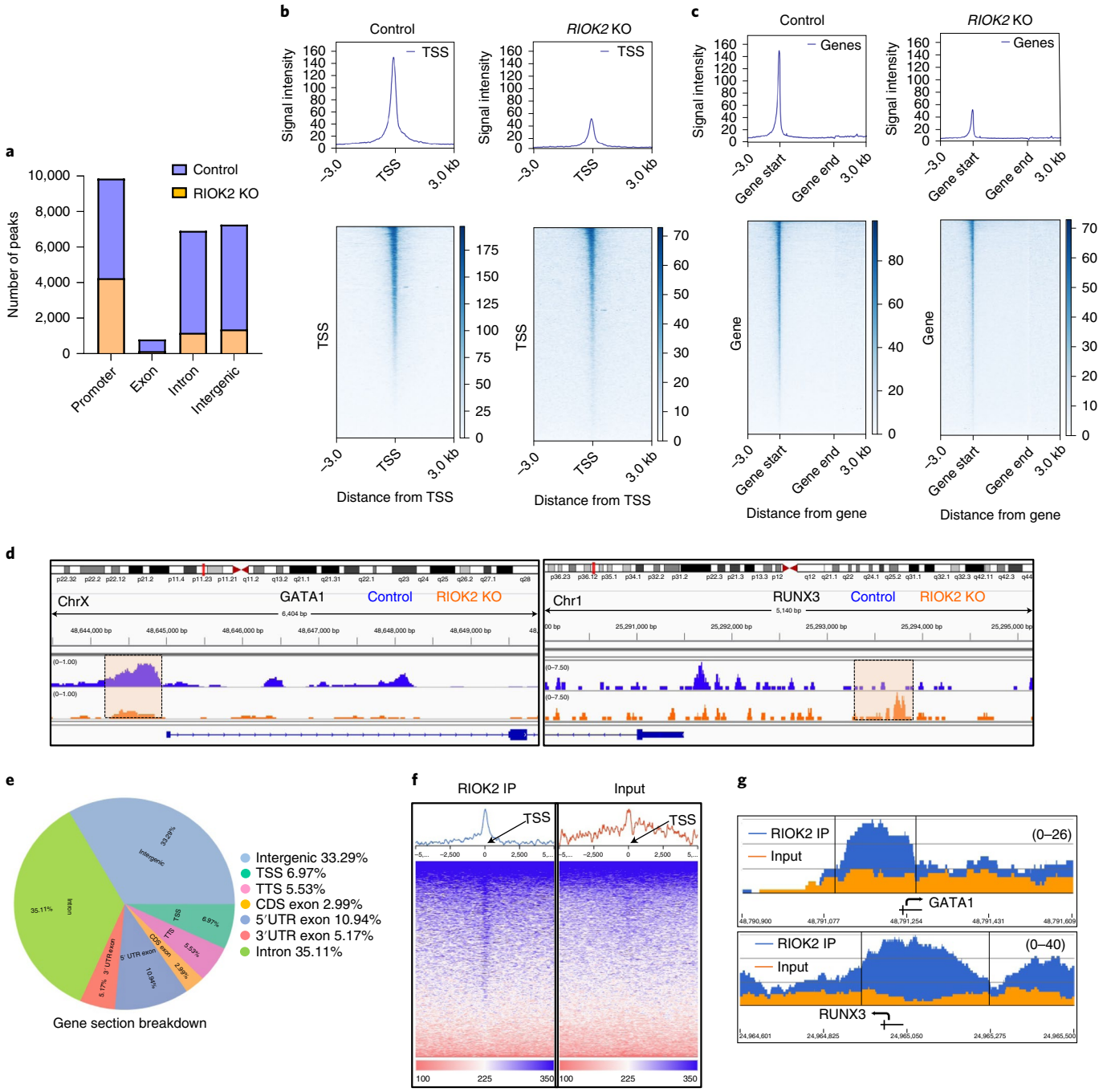


Fig. 4 | RIOK2 binds to the promoter regions of its targets and regulates chromatin accessibility. a, Graph depicting ATAC–seq analysis showing number of peaks detected in control versus RIOK2-KO HSPCs at the promoter, exon, intron and intergenic regions. **b**,**c**, TSS and gene plots depicting chromatin accessibility in control and RIOK2-KO HSPCs at the TSSs of genes and in the entire gene body, respectively. **d**, Chromosome view plots depicting chromatin accessibility at the promoters of *GATA1* and *RUNX3* in control versus RIOK2-KO HSPCs. **e**, Gene-section breakdown map of RIOK2 occupancy in the entire genome. **f**, TSS plots showing enrichment of RIOK2 at the TSSs of genes as compared to input. **g**, Chromosome view plots depicting binding of RIOK2 at the promoters of *GATA1* and *RUNX3*.

intensity of TSS plots in control HSPCs was 175 as opposed to only 70 in RIOK2-depleted HSPCs). Importantly, loss of RIOK2 led to differential chromatin accessibility at the promoter regions of its putative target genes, *GATA1* and *RUNX3* (Fig. 4d). These data indicated an integral role of RIOK2 in transcriptional regulation of genes. Examination of the amino acid sequence of RIOK2 revealed that it harbors a previously unexplored N-terminal wHTH domain[9], which is generally considered to be a structural motif involved in transcription[17,33]. Hence, we asked whether RIOK2 functions as a TF to regulate hematopoietic differentiation. To address this, we performed ChIP using monoclonal antibodies against RIOK2 followed by high-throughput DNA sequencing. Of note, occupancy of RIOK2 was observed at the TSS and 5′ untranslated region (UTR) apart from intergenic and intronic regions (Fig. 4e), analogous to the genome-wide distribution patterns of bona fide TFs[34–36]. We also observed notable enrichment of RIOK2 around TSSs (Fig. 4f

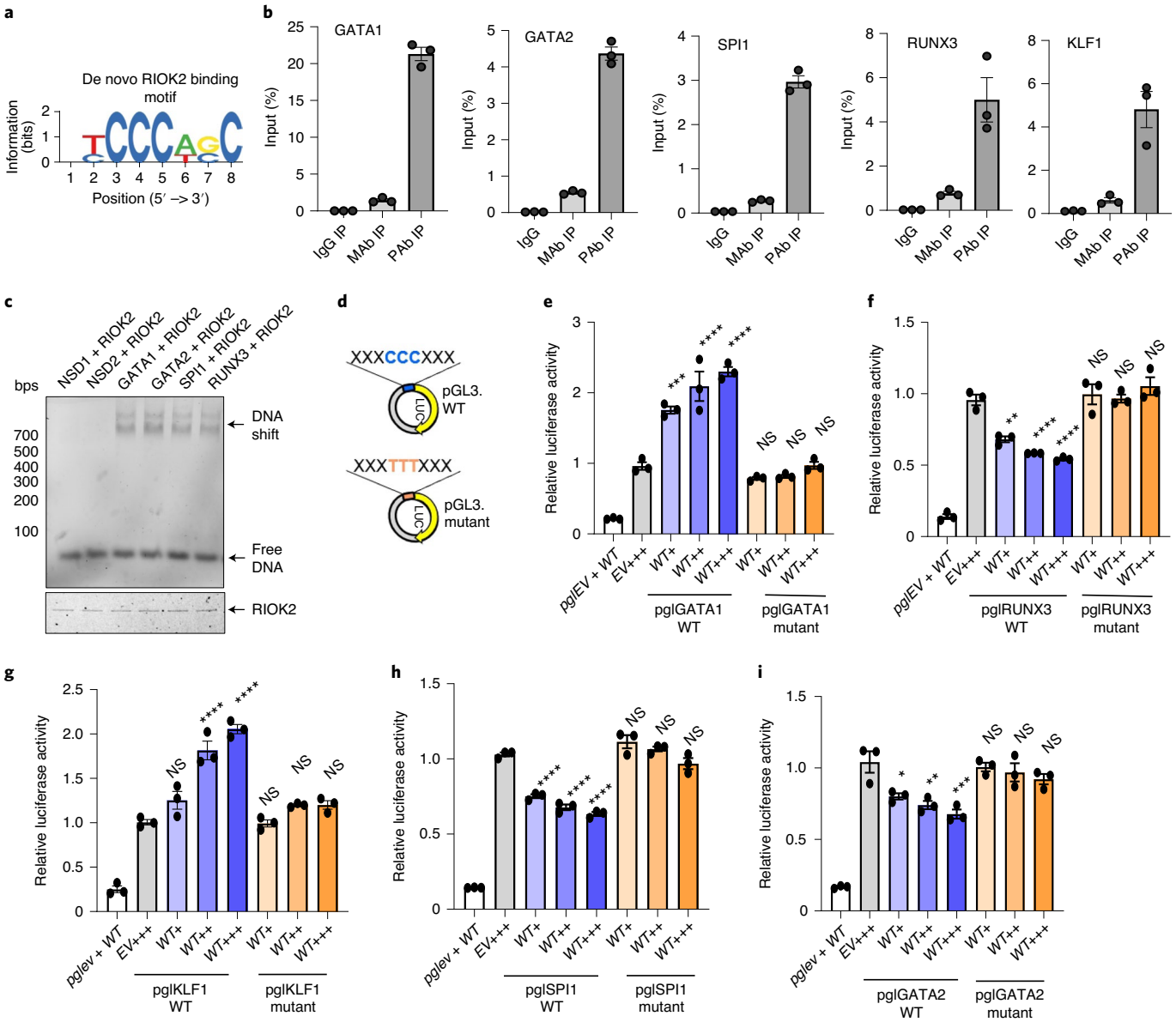


Fig. 5 | RIOK2 binds to a specific de novo nucleotide motif at the promoter regions of its targets to regulate transcription. a, De novo nucleotide motif bound by RIOK2 in the human genome. **b**, Relative binding of RIOK2 to the promoter regions of *GATA1*, *GATA2*, *SPI1*, *RUNX3* and *KLF1* via ChIP using monoclonal antibodies (Mabs) and polyclonal antibodies (Pabs) against RIOK2. **c**, Electrophoretic mobility shift assay showing DNA migration in the presence of recombinant human RIOK2 and duplex DNA incorporating promoter regions of *GATA1*, *GATA2*, *SPI1* and *RUNX3*. NSD1 and NSD2, nonspecific duplex DNA. **d**, Incorporation of WT (WT: CCC) or mutant (MUT: TTT) promoters of RIOK2 targets in basic pGL3.1 luciferase reporter plasmid. **e**–**i**, WT (WT: CCC) and mutant (MUT: TTT) GATA1, RUNX3, KLF1, SPI1 and GATA2 promoter-driven luciferase activity in response to increasing RIOK2 expression, respectively; $n = 3$ technical replicates in **e**–**i**, $n = 3$ independent experiments in **b**. All comparisons done w.r.t. EV control in **e**–**i**. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; one-way ANOVA with Tukey's or Dunnett's correction. Data are represented as the mean ± s.e.m. Data in **c** are representative of four independent experiments.

and Extended Data Fig. 5a). Critically, occupancy of RIOK2 was observed at the promoter regions of its putative target genes (Fig. 4g and Extended Data Fig. 5b).

RIOK2 binds to specific de novo nucleotide motifs. Next, we curated de novo nucleotide-binding motifs of RIOK2 across the entire human genome (Extended Data Fig. 5c) and identified a central cytosine-rich motif in the promoter regions of all TFs that we had observed to be differentially regulated by RIOK2, including GATA1, GATA2, SPI1, RUNX3 and KLF1 (Fig. 5a and Extended Data Fig. 5d). ChIP using both monoclonal and polyclonal antibodies against RIOK2 confirmed its binding to the promoter regions of target genes (Fig. 5b). Next, duplex DNAs were designed to encompass RIOK2-binding motifs in the promoter regions of *GATA1*, *GATA2*, *SPI1* and *RUNX3* (Supplementary Table 2). Electrophoretic mobility shift assay confirmed binding of RIOK2 protein to DNA in vitro (Fig. 5c). Mutation of three central cytosine nucleotides to thymine in the de novo motif was sufficient to partially, but significantly, attenuate RIOK2's DNA-binding affinity (Extended Data Fig. 6a,b). We also noted an increase in the DNA-binding affinity of RIOK2 over a time course (Extended Data Fig. 6c,d). Additionally, increasing protein concentrations of RIOK2 increased its DNA-binding

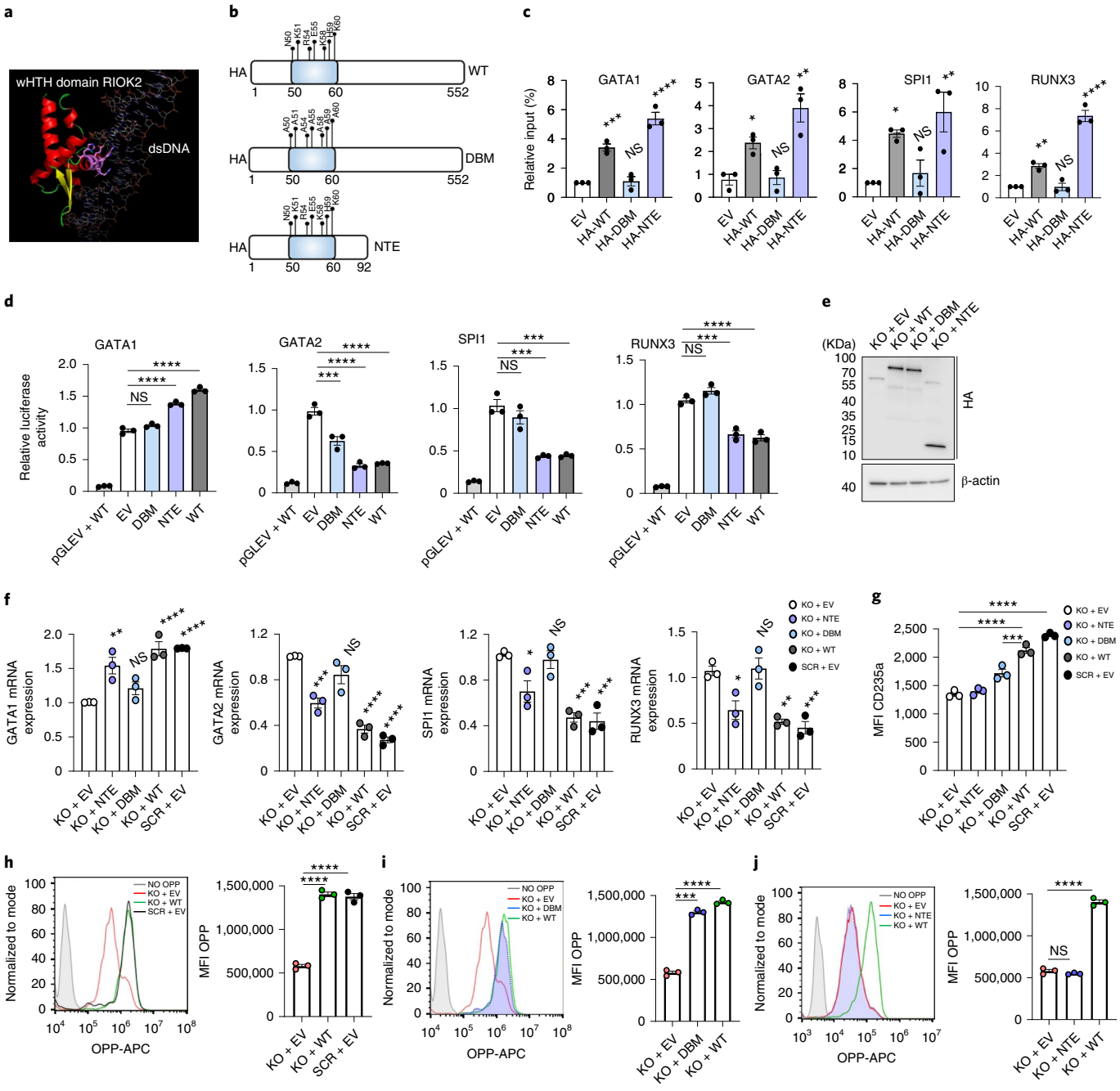


Fig. 6 | The wHTH domain of RIOK2 harbors a DNA-binding domain that is critical in hematopoiesis. a, Structural modeling of the wHTH domain of RIOK2 (residues 1–92) associated with the B-form of double-stranded DNA. α-helices and β-stands are marked in red and yellow, respectively; the helix interacting with DNA is marked in purple. **b**, Diagram showing WT, DBM and NTE plasmids of RIOK2 with an N-terminal HA-tag. **c**, Relative binding affinities of EV, WT, DBM and NTE RIOK2 to promoters of *GATA1*, *GATA2*, *SPI1* and *RUNX3*, assessed by ChIP. **d**, GATA1, GATA2, SPI1 and RUNX3 promoter-driven luciferase reporter activity after expression of EV, WT, DBM or NTE RIOK2. **e**, Immunoblot showing expression of ectopically expressed HA-tagged WT, DBM and NTE RIOK2 in RIOK2-KO HSPCs. **f**, mRNA levels of *GATA1*, *GATA2*, *SPI1* and *RUNX3* (normalized to actin) in RIOK2-KO HSPCs ectopically expressing EV, WT, DBM or NTE RIOK2. **g**, Erythroid progression in RIOK2-KO HSPCs ectopically expressing EV, WT, DBM or NTE RIOK2. **h**, *O*-propargyl-puromycin (OPP) incorporation in scrambled versus RIOK2-KO cells after reconstitution of WT RIOK2 or EV. **i**, OPP incorporation in RIOK2-KO cells ectopically expressing EV, WT or DBM RIOK2. **j**, OPP incorporation in RIOK2-KO cells ectopically expressing EV, WT or NTE RIOK2; $n = 3$ independent experiments in **c**, **f** and **g**; $n = 4$ technical replicates in **d**; $n = 3$ technical replicates in **h**–**j**. All comparisons done w.r.t. EV control in **c** and **d**. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; one-way ANOVA with Tukey's or Dunnett's correction. Data are represented as the mean ± s.e.m. Data in **e** are representative of three independent experiments.

ability in a dose-dependent manner (Extended Data Fig. 6e,f). Furthermore, luciferase reporter assays confirmed that RIOK2 activated *GATA1* and *KLF1* promoters while downregulating *SPI1*, *RUNX3* and *GATA2* promoter activities (Fig. 5d–i). Mutation of the cytosine-rich motifs in the de novo nucleotide-binding sequence markedly impaired the transcriptional activity of RIOK2 (Fig. 5d–i).

Hence, we conclude that RIOK2 functions as a master TF via its ability to bind to a specific DNA motif and regulate the expression of several key hematopoietic TFs.

The DNA-binding domain of RIOK2 is critical in hematopoiesis. Structural modeling of the N-terminal wHTH domain of RIOK2 predicted seven evolutionarily conserved residues, Asn50, Lys51, Arg54, Glu55, Lys58, His59 and Lys60, that could associate with the B-form of double-stranded DNA (Fig. 6a). We generated a DNA-binding mutant (DBM) of RIOK2 by mutating all seven residues to alanine in a lentiviral vector expressing full-length RIOK2 (1–552 residues) with an N-terminal hemagglutinin (HA)-tag (Fig. 6b). We simultaneously generated a vector expressing only the N-terminal extension (NTE, 1–92 residues) of RIOK2, which forms its wHTH domain (Fig. 6b). ChIP revealed significantly compromised binding affinity of DBM RIOK2 to the target gene promoters (Fig. 6c). In contrast, NTE RIOK2 displayed more robust DNA binding as compared to wild-type (WT) RIOK2 (Fig. 6c), likely because its simpler conformation facilitated stronger interactions with DNA. While NTE RIOK2 could regulate the promoter activity of RIOK2's target genes, DBM RIOK2 displayed reduced transactivation (Fig. 6d). While DBM was unable to rescue altered mRNA levels of target genes in the RIOK2-KO setting (such as *GATA1*), NTE RIOK2 largely restored them to levels comparable with WT RIOK2 (Fig. 6e,f). Consistently, DBM RIOK2 did not increase erythroid differentiation as compared to WT RIOK2, demonstrating the relevance of RIOK2's DNA-binding activity in driving erythropoiesis (Fig. 6g and Extended Data Fig. 6g). However, the NTE of RIOK2 also failed to restore impaired erythroid differentiation in RIOK2-KO cells (Fig. 6g and Extended Data Fig. 6g). We speculated that loss of the RIO kinase domain in NTE RIOK2 may have blocked cytoplasmic translation, thus compromising erythroid differentiation. While DBM displayed nascent protein synthesis comparable to WT RIOK2, expression of NTE RIOK2 failed to rescue protein translation in RIOK2-KO cells (Fig. 6h–j). Together, our findings demonstrate that aberration in DNA-binding transcriptional function of RIOK2 is sufficient to impair hematopoiesis even when its translational activity is not affected, thus identifying RIOK2 as a master transcriptional regulator via the DNA-binding properties of its wHTH domain.

RIOK2 encompasses two transactivation domains. Eukaryotic transcription requires both DNA-binding domains and TADs, the latter of which recruit cofactors to initiate transcription. To identify the TADs of RIOK2, we reasoned that α-helices are likely to facilitate binding with co-activators or co-repressors, in part by the flexibility they can confer[37]. As winged helix domains mostly harbor TADs close to their DNA-binding helices[17], we generated N-terminal HA-tagged mutants with deletions in helices forming the wHTH domain of RIOK2 (Fig. 7a,b). Deletion of amino acids 14–29 (ΔTAD1) and 77–90 (ΔTAD2) did not restore mRNA levels of target genes in the RIOK2-KO cells (Fig. 7c,d). ChIP confirmed comparable DNA-binding abilities of ΔTAD1 and ΔTAD2 mutants to WT RIOK2 (Extended Data Fig. 7a), indicating that the loss of transcriptional regulation in ΔTAD1 and ΔTAD2 did not stem from compromised DNA binding. Examination of the crystal structure of residues 2–301 of RIOK2 further suggests that TAD1 (yellow) and TAD2 (pale gray) confer accessibility to the wHTH domain of RIOK2 to potentially bind co-activators or co-repressors (Fig. 7e). Next, we performed co-immunoprecipitation of ectopically expressed HA-tagged WT versus ΔTAD1 and ΔTAD2 RIOK2 using monoclonal anti-HA antibodies followed by mass spectrometry to analyze quantitative spectral counts of the immunoprecipitated proteins. Intriguingly, we observed binding of RIOK2 with the core transcriptional assembly factors at TSSs, such as POLR2A, WDR43 and DDX21 (ref. [38]; Fig. 7f). Loss of either TAD1 or TAD2 compromised these interactions (Fig. 7f), signifying that the TADs of RIOK2 facilitate its interaction with the core transcriptional complex at TSSs. Furthermore, expression of ΔTAD1 or ΔTAD2 neither rescued erythroid differentiation nor restored megakaryocytic and myeloid differentiation to baseline levels in RIOK2-depleted HSPCs (Fig. 7g and Extended Data Fig. 7b). In contrast, deletion of residues 35–43, comprising another helix in the wHTH domain, displayed more potent erythroid differentiation than WT RIOK2 in the KO setting, indicating that this helix likely functions as a transrepressor domain (TRD) of RIOK2 (Extended Data Fig. 7c,d).

Given the integral roles of the DNA-binding domains and TADs of RIOK2 in hematopoiesis (Extended Data Fig. 7e,f), we next assessed the involvement of its kinase activity in hematopoietic lineage commitment. To this end, we generated the kinase-dead mutant of RIOK2 p.Lys123Ala (K123A)[9] with an N-terminal HA-tag. Interestingly, the K123A mutant of RIOK2 did not affect transactivation of its targets, as opposed to the DBM and ΔTAD1 mutants which compromised RIOK2's transactivation functions (Extended Data Fig. 8a). This suggests that the kinase activity of RIOK2 does not play a role in transcriptional functions. Consistently, K123A RIOK2 displayed comparable DNA-binding ability alongside WT, ΔTAD1 RIOK2 and ΔTAD2 RIOK2, whereas the DBM failed to bind DNA (Extended Data Fig. 8b). To confirm the roles of RIOK2's kinase activity in cytoplasmic protein translation, we assessed nascent protein synthesis. The K123A mutant RIOK2 markedly compromised the recovery of protein translation in KO cells as compared to DBM and WT RIOK2 (Extended Data Fig. 8d). Although the kinase-dead K123A mutant impaired erythroid differentiation on a comparable level with DBM and ΔTAD1 RIOK2, it significantly rescued myelopoiesis and megakaryopoiesis (Extended Data Fig. 8e–g). This is consistent with existing literature that reported selective inhibition of erythroid differentiation with blockade of protein translation[20].

Correlation of mRNA expression of *RIOK2* in myelodysplastic syndromes, acute myeloid leukemia and chronic kidney disease. Importantly, we studied the expression levels of *RIOK2* in individuals with hematologic disorders. Corroborating our observation that loss of RIOK2 dampens *GATA1* and *KLF1* mRNA levels while elevating *GATA2*, *SPI1* and *RUNX3* expression (Fig. 3a,e), we noted that mRNA expression of *RIOK2* positively correlated with *GATA1* and *KLF1*, and negatively correlated with *GATA2* and *RUNX3* in individuals with MDS (GSE19429; Fig. 8a). Additionally, expression of *RIOK2* positively correlated with erythroid gene expression and negatively correlated with myeloid and megakaryocytic gene expression (Fig. 8b). We also obtained similar correlations in individuals with AML (GSE131184; Fig. 8c,d). Furthermore, the mRNA expression of *RIOK2* was significantly reduced in individuals with chronic kidney disease (CKD) as compared to healthy controls (GSE37171; Fig. 8e), which also positively correlated with the expression of early erythroid genes (Fig. 8f). These findings strongly suggest that RIOK2-mediated effects on hematopoiesis may be physiologically relevant in MDS, AML and the anemia of CKD. Collectively, our data establish RIOK2 as a master regulator of hematopoietic differentiation (Extended Data Fig. 9).

Discussion

Anemia is one of the most frequently occurring blood disorders endangering the health of roughly one-third of the world population[39]. It is a hallmark of a plethora of hematologic disorders, such as myeloid neoplasms, bone marrow failure syndromes and hematologic malignancies[3]. The morbidities associated with these hematologic disorders are compounded by anemia, thus often outweighing the treatment benefits from allogeneic stem cell transplants leaving only a handful of FDA-approved drugs/therapies for treatment of these disorders[6,40]. Anemia is also frequently diagnosed in patients

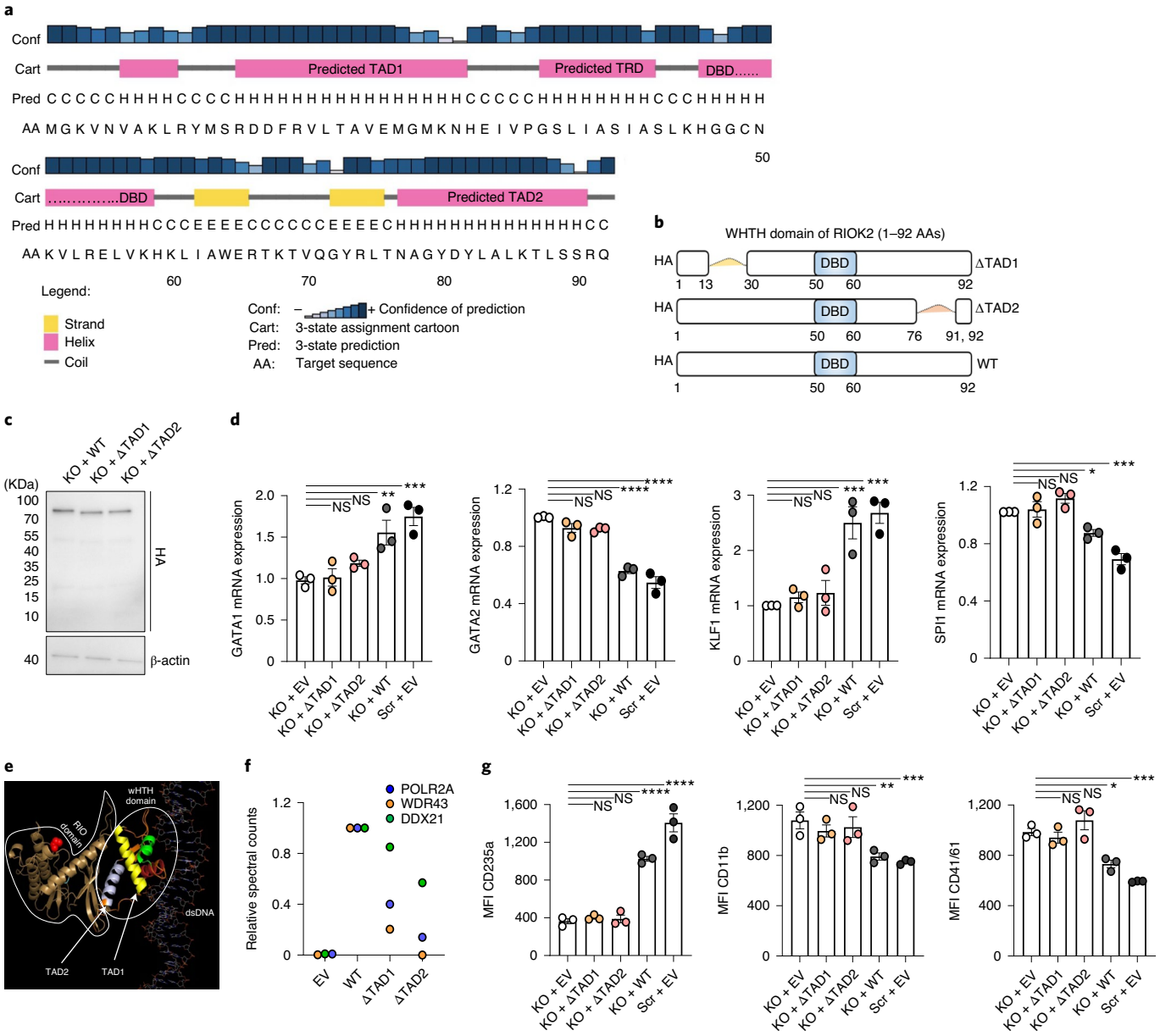


Fig. 7 | The two transactivation domains of RIOK2 facilitate binding with a core transcriptional complex required for hematopoiesis. a, Three-state prediction model showing residues forming helices, strands and coils in the wHTH domain of RIOK2 (1–92 residues). DBD, DNA-binding domain. **b**, Diagram showing WT, TAD1 and TAD2 deleted (ΔTAD1 and ΔTAD2) RIOK2 plasmids with N-terminal HA-tag; AAs, amino acids. **c**, Immunoblot showing ectopically expressed HA-tagged WT, ΔTAD1 RIOK2 and ΔTAD2 RIOK2 in RIOK2-KO HSPCs. **d**, mRNA expression of *GATA1*, *GATA2*, *KLF1* and *SPI1* (normalized to actin) in RIOK2-depleted HSPCs ectopically expressing EV, WT, ΔTAD1 RIOK2 or ΔTAD2 RIOK2. **e**, Crystal structure of residues 2–301 of human RIOK2 (Protein Data Bank (PDB) 6HK6) docked onto the B-form of DNA: 2–92, wHTH domain; 93–289, RIO domain. TAD1, yellow; TAD2, pale gray; DBD, red; RIO domain, chocolate. The end of the RIO domain is marked in red spheres. **f**, Relative spectral counts reflecting binding intensities of HA-WT versus HA-ΔTAD1, HA-ΔTAD2 and EV with POLR2A, WDR43 and DDX21. **g**, Erythroid (CD235a), myeloid (CD11b) and megakaryocytic (CD41/CD61) progression in differentiating HSPCs ectopically expressing EV, WT, ΔTAD1 or ΔTAD2 RIOK2 in RIOK2-KO setting, respectively; $n = 3$ independent experiments in **d** and **g**. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; one-way ANOVA with Tukey's or Dunnett's correction. Data are represented as the mean ± s.e.m. Data in **c** are representative of three independent experiments.

treated with chemotherapy or other cytotoxic agents[41]. Deregulation of transcriptional regulatory networks are key determinants of ineffective hematopoiesis underlying erythroid differentiation defects conducive to anemia[42]. There is thus a dire need to revisit the origins of hematopoietic differentiation defects to identify additional targets for novel therapies in treating anemia. Our discovery of RIOK2 as a key TF in erythroid differentiation opens up many new avenues of investigation. In this regard, stabilization of RIOK2's expression or stimulation of its DNA-binding and transactivation functions present intriguing approaches to reverse anemia.

RIOK2 is reported to be a critical component of pre-40S ribosome biogenesis and cytoplasmic translation[9,11]. Here, we have uncovered an unexpected function of RIOK2 as a master transcriptional regulator driving hematopoietic lineage commitment via its N-terminal DNA-binding wHTH domain that also incorporates its two TADs. However, whether RIOK2 harbors additional TADs and/or TRDs

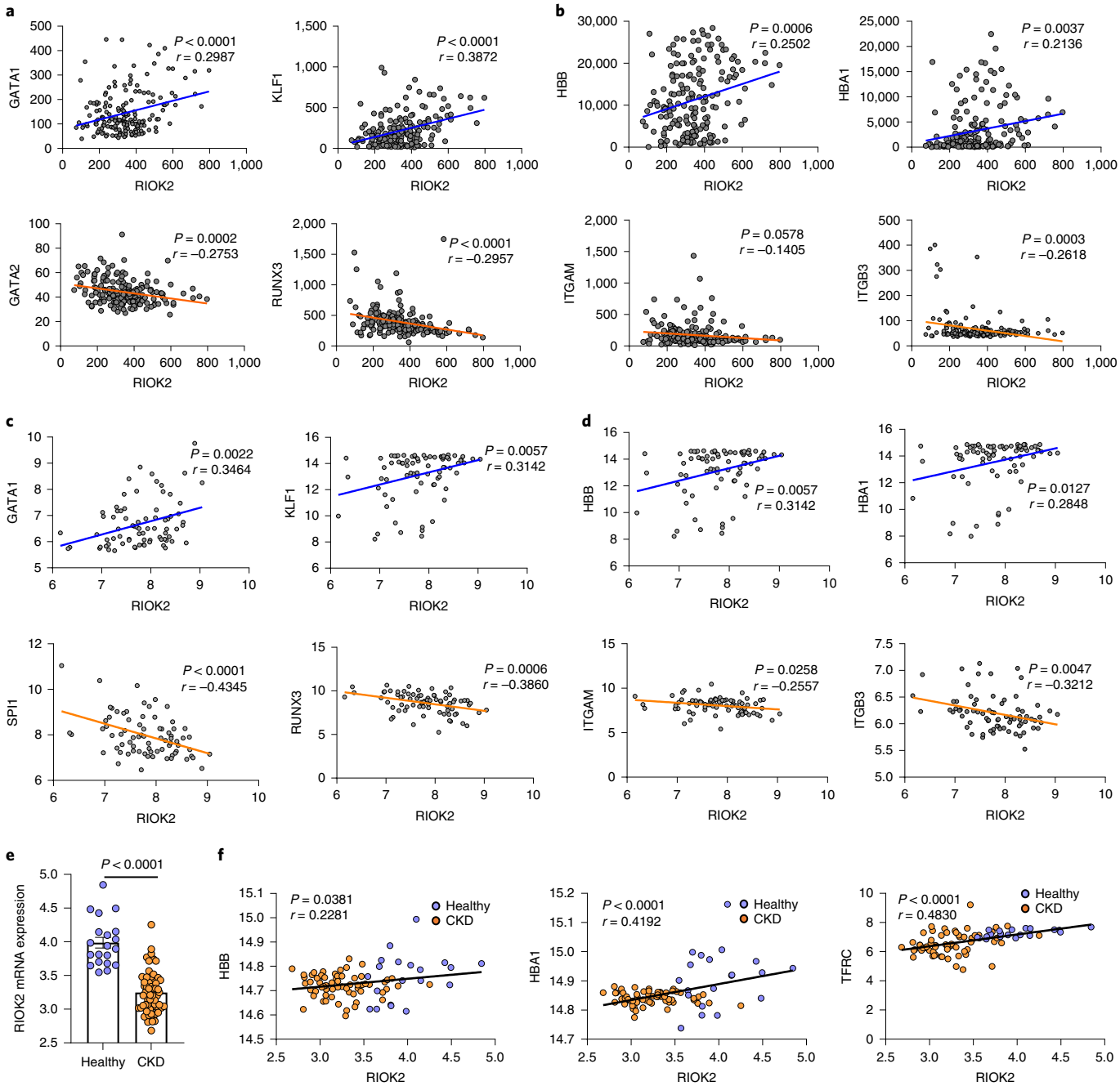


Fig. 8 | mRNA expression of *RIOK2* correlates with its targets and other hematopoietic genes in individuals with hematologic disorders. a, Correlation of the mRNA expression of *RIOK2* with *GATA1*, *KLF1*, *GATA2* and *RUNX3* in MDS patient-derived bone marrow cells; $n = 183$. **b**, Correlation of the mRNA expression of *RIOK2* with *HBB*, *HBA1*, *ITGAM* and *ITGB3* in MDS patient-derived bone marrow cells; $n = 183$. **c**, Correlation of the mRNA expression of *RIOK2* with *GATA1*, *KLF1*, *SPI1* and *RUNX3* in AML patient-derived bone marrow samples; $n = 76$. **d**, Correlation of the mRNA expression of *RIOK2* with *HBB*, *HBA1*, *ITGAM* and *ITGB3* in AML patient-derived bone marrow samples; $n = 76$. **e**, *RIOK2* mRNA expression in the whole blood of healthy controls versus patients with chronic kidney disease (CKD); $n = 20$ healthy controls, $n = 63$ CKD patients. **f**, Correlation of the mRNA expression of *RIOK2* with *HBB*, *HBA1* and *TFRC* in healthy controls ($n = 20$) versus individuals with CKD ($n = 63$). Two-tailed Pearson's correlation performed in **a**–**d** and **f**. Pearson's correlation coefficients (r) and P values are shown in **a**–**d** and **f**; unpaired nonparametric Mann–Whitney test in **e**.

in its C-terminal half, and whether they functionally integrate with the N-terminal wHTH domain to promote RIOK2's transcriptional abilities are still unknown. This study further prompts speculation that RIOK2 may use its kinase and/or ATPase functions to modify its co-activators/co-repressors to differentially regulate transcriptional outcomes of its targets. Nevertheless, an unprecedented involvement in both transcriptional and translational machineries makes RIOK2 a rare protein that warrants further investigation into its possible involvement in other cellular processes. We are hopeful that this study will also lay a foundation to discovering how proteins, like RIOK2, may integrate transcriptional processes with translational outcomes to drive biological functions.

GATA1 is deemed as the master erythroid TF[21]. Our observation that a positive feedback loop between GATA1 and RIOK2 drives

erythroid differentiation presents novel opportunities to target dyserythropoiesis. The hematopoietic lineage-specific expression levels of GATA1 and GATA2 are dynamically controlled by transcriptional programs. While GATA1 is predominantly expressed in erythroid, megakaryocytic and eosinophilic progenitors, GATA2 is highly expressed in HSPCs. There is a dynamic shift from GATA2 to GATA1 expression during erythropoiesis, which is referred to as the GATA switch. GATA1 thus replaces GATA2 at several genomic loci to drive erythropoiesis[43,44]. As RIOK2 promotes GATA1 and inhibits GATA2 expression in primary human HSPCs, we suggest that RIOK2-driven transcriptional networks may play a critical role in the GATA switch. Consistent with this notion, our observations that loss of RIOK2 not only blocks erythroid differentiation but also retains a substantial fraction of HSPCs in their progenitor state, further argue that RIOK2 regulates the GATA switch during erythropoiesis. Additionally, we provide evidence that RIOK2 transcriptionally regulates other key lineage-specific TFs in hematopoiesis, such as SPI1, RUNX3 and KLF1. Our discovery thus encourages remodeling of the existing map of transcriptional networks governing hematopoiesis and positions RIOK2 as an integral component of these regulatory networks dictating lineage fate determination during hematopoietic stem cell differentiation. Further studies will examine how RIOK2 upregulates the expression of some TFs while downregulating others to fine-tune hematopoietic lineage commitment, a feature characteristic of other master transcriptional regulators of cell fate such as T-bet[45].

Identification of RIOK2 as a master transcriptional regulator governing several hematopoietic lineages significantly advances our current understanding of the transcriptomic landscape underlying hematopoietic differentiation and opens up new areas of exploration. Comprehensive analysis of de novo motifs bound by RIOK2 in the entire genome will expand our insights into RIOK2's transcriptome regulatory network. Our findings may lead to new approaches to target these newly identified regulatory networks in hematopoiesis that may be relevant for MDS and other hematologic disorders such as AML and anemia of CKD. Finally, investigation of RIOK2-mediated transcriptional pathways as well as assessment of RIOK2's upstream regulation should provide new targetable options to treat the anemia of aging, chronic kidney and inflammatory diseases as well as other hematologic disorders.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-021-01079-w.

Received: 9 May 2021; Accepted: 25 October 2021;
Published online: 22 December 2021

References

1. Palapar, L. et al. Anaemia and physical and mental health in the very old: an individual participant data meta-analysis of four longitudinal studies of ageing. *Age Ageing* **50**, 113–119 (2021).
2. Lopes, M. B. et al. A real-world longitudinal study of anemia management in non-dialysis-dependent chronic kidney disease patients: a multinational analysis of CKDopps. *Sci. Rep.* **11**, 1784 (2021).
3. Becktell, K. et al. Aplastic Anemia & MDS International Foundation: Bone Marrow Failure Disease Scientific Symposium 2018. *Leuk. Res.* **80**, 19–25 (2019).
4. Saygin, C. & Carraway, H. E. Current and emerging strategies for management of myelodysplastic syndromes. *Blood Rev.* **48**, 100791 (2020).
5. Hong, S. et al. Survival following relapse after allogeneic hematopoietic cell transplantation for acute leukemia and myelodysplastic syndromes in the contemporary era. *Hematol. Oncol. Stem Cell Ther.* **S1658-3876**, 30178–3 (2020).
6. Garcia-Manero, G., Chien, K. S. & Montalban-Bravo, G. Myelodysplastic syndromes: 2021 update on diagnosis, risk stratification and management. *Am. J. Hematol.* **95**, 1399–1420 (2020).
7. Feld, J., Navada, S. C. & Silverman, L. R. Myelo-deception: luspatercept & TGF-beta ligand traps in myeloid diseases & anemia. *Leuk. Res.* **97**, 106430 (2020).
8. Bhatt, V. R. & Steensma, D. P. Hematopoietic cell transplantation for myelodysplastic syndromes. *J. Oncol. Pract.* **12**, 786–792 (2016).
9. Ferreira-Cerca, S. et al. ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit. *Nat. Struct. Mol. Biol.* **19**, 1316–1323 (2012).
10. Zemp, I. et al. Distinct cytoplasmic maturation steps of 40S ribosomal subunit precursors require hRio2. *J. Cell Biol.* **185**, 1167–1180 (2009).
11. Ameismeier, M., Cheng, J., Berninghausen, O. & Beckmann, R. Visualizing late states of human 40S ribosomal subunit maturation. *Nature* **558**, 249–253 (2018).
12. Raundhal, M. et al. Blockade of IL-22 signaling reverses erythroid dysfunction in stress-induced anemias. *Nat. Immunol.* **22**, 520–529 (2021).
13. Pellagatti, A. & Boultwood, J. The molecular pathogenesis of the myelodysplastic syndromes. *Eur. J. Haematol.* **95**, 3–15 (2015).
14. Pellagatti, A. et al. Deregulated gene expression pathways in myelodysplastic syndrome hematopoietic stem cells. *Leukemia* **24**, 756–764 (2010).
15. Brennan, R. G. The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* **74**, 773–776 (1993).
16. LaRonde-LeBlanc, N. & Wlodawer, A. A family portrait of the RIO kinases. *J. Biol. Chem.* **280**, 37297–37300 (2005).
17. Teichmann, M., Dumay-Odelot, H. & Fribourg, S. Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes. *Transcription* **3**, 2–7 (2012).
18. Nakahata, T. & Okumura, N. Cell surface antigen expression in human erythroid progenitors: erythroid and megakaryocytic markers. *Leuk. Lymphoma* **13**, 401–409 (1994).
19. Tothova, Z. et al. Multiplex CRISPR–Cas9-based genome editing in human hematopoietic stem cells models clonal hematopoiesis and myeloid neoplasia. *Cell Stem Cell* **21**, 547–555 (2017).
20. Khajuria, R. K. et al. Ribosome levels selectively regulate translation and lineage commitment in human hematopoiesis. *Cell* **173**, 90–103 (2018).
21. Gutierrez, L., Caballero, N., Fernandez-Calleja, L., Karkoulia, E. & Strouboulis, J. Regulation of GATA1 levels in erythropoiesis. *IUBMB Life* **72**, 89–105 (2020).
22. Gutierrez, L. et al. Ablation of Gata1 in adult mice results in aplastic crisis, revealing its essential role in steady-state and stress erythropoiesis. *Blood* **111**, 4375–4385 (2008).
23. Fujiwara, T. GATA transcription factors: basic principles and related human disorders. *Tohoku J. Exp. Med.* **242**, 83–91 (2017).
24. Ling, T. & Crispino, J. D. *GATA1* mutations in red cell disorders. *IUBMB Life* **72**, 106–118 (2020).
25. Gnanapragasam, M. N. & Bieker, J. J. Orchestration of late events in erythropoiesis by KLF1/EKLF. *Curr. Opin. Hematol.* **24**, 183–190 (2017).
26. Zhang, D. E. et al. Function of PU.1 (Spi-1), C/EBP and AML1 in early myelopoiesis: regulation of multiple myeloid CSF receptor promoters. *Curr. Top. Microbiol. Immunol.* **211**, 137–147 (1996).
27. Yokomizo-Nakano, T. et al. Overexpression of RUNX3 represses RUNX1 to drive transformation of myelodysplastic syndrome. *Cancer Res.* **80**, 2523–2536 (2020).
28. Daw, S. & Law, S. The functional interplay of transcription factors and cell adhesion molecules in experimental myelodysplasia including hematopoietic stem progenitor compartment. *Mol. Cell. Biochem.* **476**, 535–551 (2020).
29. Nerlov, C. & Graf, T. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev.* **12**, 2403–2412 (1998).
30. Rekhtman, N. et al. PU.1 and pRB interact and cooperate to repress GATA-1 and block erythroid differentiation. *Mol. Cell. Biol.* **23**, 7460–7474 (2003).
31. Shimizu, R. & Yamamoto, M. Quantitative and qualitative impairments in GATA2 and myeloid neoplasms. *IUBMB Life* **72**, 142–150 (2020).
32. Huang, Z. et al. GATA-2 reinforces megakaryocyte development in the absence of GATA-1. *Mol. Cell. Biol.* **29**, 5168–5180 (2009).
33. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262 (2005).
34. Zhong, X. et al. HoxA9 transforms murine myeloid cells by a feedback loop driving expression of key oncogenes and cell cycle control genes. *Blood Adv.* **2**, 3137–3148 (2018).
35. Xu, J. et al. Transcriptional silencing of γ-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev.* **24**, 783–798 (2010).
36. Chen, T. W. et al. ChIPseek, a web-based analysis tool for ChIP data. *BMC Genomics* **15**, 539 (2014).
37. Harish, B., Swapna, G. V., Kornhaber, G. J., Montelione, G. T. & Carey, J. Multiple helical conformations of the helix-turn-helix region revealed by NOE-restrained MD simulations of tryptophan aporepressor, TrpR. *Proteins* **85**, 731–740 (2017).
38. Bi, X. et al. RNA targets ribogenesis factor WDR43 to chromatin for transcription and pluripotency control. *Mol. Cell* **75**, 102–116 (2019).

39. Chaparro, C. M. & Suchdev, P. S. Anemia epidemiology, pathophysiology, and etiology in low- and middle-income countries. *Ann. N. Y. Acad. Sci.* **1450**, 15–31 (2019).
40. Feld, J., Belasen, A. & Navada, S. C. Myelodysplastic syndromes: a review of therapeutic progress over the past 10 years. *Expert Rev. Anticancer Ther.* **20**, 465–482 (2020).
41. Groopman, J. E. & Itri, L. M. Chemotherapy-induced anemia in adults: incidence and treatment. *J. Natl Cancer Inst.* **91**, 1616–1634 (1999).
42. Edginton-White, B. & Bonifer, C. The transcriptional regulation of normal and malignant blood cell development. *FEBS J* https://doi.org/10.1111/febs.15735 (2021).
43. Suzuki, M. et al. GATA factor switching from GATA2 to GATA1 contributes to erythroid differentiation. *Genes Cells* **18**, 921–933 (2013).
44. Moriguchi, T. & Yamamoto, M. A regulatory network governing Gata1 and Gata2 gene transcription orchestrates erythroid lineage differentiation. *Int. J. Hematol.* **100**, 417–424 (2014).
45. Szabo, S. J. et al. A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell* **100**, 655–669 (2000).

Methods

Ethics. Our research complies with the ethical regulations stipulated by the Dana-Farber Cancer Institute. Data collection and analysis were not performed blind to the conditions of the experiments. The experimental organization included no randomization.

Primary and secondary cell culture. CD34$^+$ primary human HSPCs were obtained from Fred-Hutchinson Cancer Research center. The cells were thawed and washed with 1× PBS constituting 1% human AB serum, followed by revival in expansion medium comprising StemSpan SFEM II medium supplemented with 1× CC100 (STEMCELL Technologies), 1% penicillin–streptomycin (P/S), 1% glutamine and 10 ng ml^{-1} of the cytokine TPO. The cells were subjected to differentiation medium containing IMDM reconstituted with 3% human AB serum, 2% human AB plasma, 1% P/S, 0.06% heparin solution (STEMCELL Technologies), 1 ng ml^{-1} interleukin-3 (PeproTech), 10 ng ml^{-1} SCF (PeproTech), 200 µg ml^{-1} holo-transferrin (Sigma-Aldrich) and 3 U ml^{-1} erythropoietin (Dana-Farber Cancer Institute pharmacy). For selective differentiation of HSPCs into myeloid or megakaryocytic lineages, HSPCs were cultured in SFEM II media supplemented with myeloid expansion supplement II 100X (02694) or megakaryocyte expansion supplement 100X (02696; both STEMCELL Technologies), respectively. Cells were cultured at a density of 0.1–0.01 × 10^6 cells per ml and incubated at 37 °C with 5% CO_2. Media were changed every alternate day, as required. These protocols are previously described[20].

TF-1 human erythroblast cells were purchased from the American Type Culture Collection (ATCC; CRL-2003) and maintained in RPMI-1640 medium (ATCC 30-2001) containing 10% fetal bovine serum (FBS), 2 ng ml^{-1} GM-CSF and 1% P/S. For differentiation, the cells were cultured in RPMI-1640 medium containing 10% FBS, 5 U ml^{-1} erythropoietin and 1% P/S. K562 human erythroid cells were purchased from the ATCC (CCL-243) and maintained in IMDM medium (ATCC 30-2005) containing 10% FBS and 1% P/S. For differentiation, the cells were cultured in IMDM medium containing 10% FBS, 40 µM hemin and 1% P/S. HEK293FT cells were purchased from the ATCC and maintained in DMEM medium supplemented with 10% FBS and 1% P/S.

Generation of lentiviral vectors and infection. The shRNA constructs targeting human RIOK2 (shRIOK2-1 and shRIOK2-2) and scrambled shRNA were obtained from the Mission shRNA collection (SHCLNG-NM_018343, Sigma-Aldrich). RIOK2 was cloned in the pHAGE-MCS-IRES-ZsGreen lentiviral vector with an N-terminal HA-tag. DBM, transactivation domain deletion 1 and 2 (ΔTAD1 and ΔTAD2), TRD deleted (ΔTRD) and kinase-dead (p.Lys123Ala: K123A) mutants were generated by site-directed mutagenesis. The NTE construct was generated by cloning the first 92 amino acids of RIOK2 with an N-terminal HA-tag.

For lentiviral production, HEK293FT cells were transiently transfected with pVSV-G, pDelta8.9 and RIOK2 vectors using X-tremeGENE HP DNA transfection reagent according to the manufacturer's protocol. Viral supernatant was collected 48 h after transfection. Primary HSPCs or secondary cells were spinfected at a density of 0.1–0.2 × 10^6 cells per well in six-well plates with 8 µg ml^{-1} polybrene (Millipore) at 3,392*g* for 1.5 h at 32 °C and left overnight. The medium was changed the next morning. Lentiviral transduction efficiency reached 60–75% for primary human HSPCs and >95% for secondary cells after 48 h of infection. The positively transduced cells (ZsGreen$^+$) were FACS sorted for further analysis.

CRISPR–Cas9 gene editing. Primary human HSPCs and human erythroblast cells (TF-1) were electroporated using Lonza-Amaxa 4D-nucleofector unit according to the manufacturer's protocols. For ribonucleoprotein formation, tracer RNAs and crRNAs (IDT) were incubated at 95 °C for 5 mins at equal molar ratios and mixed with equal molar concentration of Cas9 peptide (IDT). This was followed by incubation at 37 °C for 15 min and the ribonucleoprotein mix was kept at 4 °C until use within the next 2 h. Cells were electroporated at a density of 0.1–0.2 × 10^6 cells per well of a 16-well electroporation strip. Immediately after electroporation, fresh medium was added to the cells and transferred to 37 °C. Genome editing efficiency was analyzed 48–72 h after electroporation by nucleotide sequencing, RT–qPCR and immunoblotting. The crRNAs used for KD of RIOK2 were GAACGGCGGGTTTCTTACCG and CATTTGTCAACCGATAGCCC, and the crRNAs used for KO of RIOK2 were TGACTTCAGGGTCTTGACCG and TGATTACAATCGTCATGCAG. The crRNA used against LacZ (control) was TTCTCCGCGGGAACAAACGG.

Luciferase reporter assay. For transactivation assays, the promoter regions of *RIOK2*, *GATA1*, *GATA2*, *SPI1*, *RUNX3* and *KLF1* (500–750 bps upstream of the ATG start codon) were cloned in a pGL3.1 basic vector. Luminescence intensities were normalized using co-transfection of Renilla expression vector in HEK293 cells using Lipofectamine 3000 reagent. Dual luciferase assays (Promega) were performed per the manufacturer's instructions. Luminescence intensities were captured using an EnVision Multimode plate reader (Perkin Elmer).

Quantitative RT–PCR. RNA was isolated from cells using the RNeasy Plus Micro kit (Qiagen) following the manufacturer's protocols. Genomic DNA was removed using genomic DNA eliminator spin columns, followed by isolation of total RNA using phenol-free RNeasy MinElute spin columns. Reverse transcription was performed using qScript cDNA Synthesis kit (QuantaBio). Real-time RT–qPCR was performed using QuantStudio 6 RT–PCR system (Applied Biosciences) and PerfeCTa SYBR Green FastMix reaction mixes (QuantaBio). The comparative C_T method has been used for all quantifications using corresponding β-actin mRNA levels for normalization. The primers used for RT–qPCR are listed in Supplementary Table 1.

Immunoblotting. Cells were washed with ice-cold 1× PBS, then resuspended in RIPA lysis buffer (Life Technologies) supplemented with 1× complete protease inhibitor cocktail and 1× phosphatase inhibitor cocktail (Thermo Scientific). Lysis was carried out on a rocker at 4 °C for 15–30 min followed by removal of cellular debris by centrifugation at 14,000*g* for 20 min. The supernatant was mixed with Laemmli buffer and incubated at 95 °C for 10 min. Equal amounts of proteins were resolved by SDS–PAGE. The proteins were then transferred onto PVDF membranes (Thermo Fisher Scientific) followed by blocking and probing with the primary antibodies diluted in fresh blocking buffer at 4 °C overnight: HA rabbit monoclonal antibody at 1:1,000 dilution (C29F4, 3724S, Cell Signaling Technology), β-actin rabbit polyclonal antibody at a dilution of 1:1,000 (3967S, Cell Signaling Technology) and RIOK2 mouse monoclonal antibody at a 1:1,000 dilution (OTI3E11, TA505140, Origene). After incubation overnight, membranes were washed four times with PBST buffer for 5 min each on a rocker at 25 °C, followed by incubation with horseradish peroxidase (HRP)-linked anti-mouse IgG (7076S, Cell Signaling Technology) or HRP-linked anti-rabbit IgG (7076S, Cell Signaling Technology) at a 1:3,000 dilution in fresh blocking buffer for 1 h at 25 °C. After incubation with secondary antibodies, membranes were washed four times with PBST buffer for 5 min each on a rocker at 25 °C, followed by incubation with Pierce Western blotting substrates (Thermo Scientific) mixed at a 1:1 ratio for 2–5 mins at 25 °C. The protein bands were then visualized using the ChemiDoc Touch Imaging system (Bio-Rad).

Immunoprecipitation and mass spectrometry. For collection of cellular lysates, lysis was similarly performed as described above. Around 5% of the supernatant was saved as input, and prewashed monoclonal anti-HA-tagged agarose beads slurry (A2095, Sigma-Aldrich) was added to the rest of the supernatant and incubated at 4 °C on a rocker overnight. The beads were washed with RIPA lysis buffer, resuspended in 2× Laemmli buffer, boiled at 95 °C for 10 min and immediately loaded for SDS–PAGE separation. Electrophoresis was performed until the proteins ran one-sixth of the entire gel size. The gel was stained with Coomassie dye for 1 h at 25 °C, de-stained at 4 °C overnight, and were cut with sterile blades, washed with mass spectrometry-grade acetonitrile (ACN) solution and submitted for mass spectrometry analysis at BIDMC Mass Spectrometry Core. For elution of mutant RIOK2 proteins, whole-cell lysate collection and washing of beads after immunoprecipitation were done in RIPA lysis buffer containing 500 mM NaCl concentration. After washing, the beads were incubated in elution buffer (0.1 M glycine solution, pH 2.0) supplemented with HA peptides (Millipore Sigma, I2149) at 0.2 µg µl^{-1} for 30 min with gentle flicking of tubes every 10 min. The beads were then centrifuged at 2,500*g* to collect the eluted proteins in the supernatant.

Gel-separated protein samples were reduced with 55 mM dithiothreitol, alkylated with 10 mM iodoacetamide (Sigma-Aldrich) and digested overnight with TPCK-modified trypsin/LysC (Promega) at a pH of 8.3. Peptides were extracted, dried out in a SpeedVac and resuspended in 10 µl of 1% ACN:98.9% water and 0.1% formic acid (FA). Next, 3 µl of the digested protein samples were analyzed in positive-ion mode via microcapillary liquid chromatography–tandem mass spectrometry (LC–MS/MS) using a high-resolution hybrid QExactive HF Orbitrap Mass Spectrometer (Thermo Fisher Scientific) via higher-energy collisional dissociation with data-dependent analysis with one MS1 scan followed by eight MS2 scans per cycle (top 8). Peptides were delivered and separated using an EASY-nLCII nanoflow high-performance liquid chromatography (Thermo Fisher Scientific) at 300 nl min^{-1} using self-packed 15 cm length × 75 µm internal diameter C_{18} fritted microcapillary columns. Solvent gradient conditions were 120 min from 3% B buffer to 38% B (B buffer, 100% ACN; A buffer, 0.9% ACN/0.1% FA/99.0% water). MS/MS spectra were analyzed using Mascot Version 2.6 (Matrix Science) by searching the reversed and concatenated human protein database (http://www.ebi.ac.uk/uniprot/database/download.html) with a parent ion tolerance of 18 ppm and fragment ion tolerance of 0.05 Da. Carbamidomethylation of cysteine (+57.0293 Da) was specified as a fixed modification and oxidation of methionine (+15.9949 Da) and deamidation of asparagine/glutamine (+0.984 Da) as variable modifications. Results were imported into Scaffold Q+S 4.11 software (Proteome Software) with a peptide threshold of ~75% and protein threshold of 95%, resulting in a peptide false discovery rate of ~1%. Known contaminants such as keratins, caseins, trypsin and BSA were removed from the analysis.

The immunoprecipitated proteins yielded quantitative spectral counts in mass spectrometry which reflect their binding intensities. The spectral counts of individual proteins were normalized using spectral counts of respective immunoprecipitated HA-tagged versions of RIOK2 and the fold changes with respect to WT RIOK2 were calculated.

Fluorescence-activated cell sorting analysis. For analysis of cellular surface marker expressions via flow cytometry, in vitro cultured cells were washed with PBS and incubated with fixable viability dyes (Tonbo Biosciences) at a 1:800 dilution in PBS for 20 mins at 25 °C in dark. The cells were washed with staining buffer and labeled with fluorochrome-conjugated antibodies diluted (1:100/200) in staining buffer for 30 min at 4 °C: FITC-CD34 (343604, BioLegend), BV421-CD38 (356617, BioLegend)/PE-Cy7-CD38 (356608, BioLegend), APC-CD71 (OKT9, eBioscience), PE-Cy7-CD235 (306620, BioLegend), PE-CD41/CD61 (359806, BioLegend), PerCP-Cy5.5-CD11b (301328, 101228, 393106, BioLegend), AF700-CD42b (303928, BioLegend), APC-CD135 (313307, BioLegend), BV510-Human Lineage (348807, BioLegend), PE-CD68 (333807, BioLegend), BV421-CD45RA (304129, BioLegend), BV421-CD45 (304031, 368521, BioLegend), BV785-CD45 (368527, BioLegend), PerCP-Cy5.5-CD123 (306015, BioLegend), AF700-CD117 (313245, Biolegend), PE-CD36 (336205, BioLegend), BV510-CD34 (343527, BioLegend), FITC-CD15 (301904, BioLegend), APC-CD41 (343709, BioLegend), PE-CD14 (325605, BioLegend) and APC-CD13 (301706, BioLegend). For intracellular staining, cells were labeled for surface markers as mentioned above, followed by fixation and permeabilization using Foxp3/transcription factor staining kit (Thermo Fisher Scientific). PE-conjugated GATA1 rabbit monoclonal antibodies (13353S, Cell Signaling), PE-conjugated HA-tagged rabbit monoclonal antibodies or PE-conjugated control rabbit IgG isotype (5742, Cell Signaling) were used at a 1:100 dilution. All FACS analyses were performed using CytoFLEX Flow Cytometer (Beckman Coulter). Data were analyzed using FlowJo 10.0.7 and plotted using GraphPad Prism. The gating strategies are shown in Extended Data Fig. 10.

Methylcellulose assay. Primary human HSPCs were washed with 1× PBS and mixed with semisolid methylcellulose medium (H4034, STEMCELL Technologies) by brief vortexing. Cells were plated at a density of 1,000 cells per well in a six-well plate followed by incubation in a humidified chamber at 37 °C for 14 d. Imaging of the colonies was performed using EVOS M5000 Imaging system (Thermo Fisher Scientific). The colony-forming cells were then collected by triturating the wells using staining buffer and multicolor flow cytometry was performed as described above and in refs. [46,47]. The gating strategies are shown in Extended Data Fig. 10. The colonies were scored using a STEMvision automated CFU counter. The STEMvision acquisition, analyzer and colony marker software programs were used to acquire images, analyze colonies and mark them, respectively.

Quantitative proteomic analysis. Quantitative proteomic analysis of primary Human HSPCs was performed as described[48]. Lysis of cells was done by adding 10 µl of 8 M urea, 10 mM TCEP and 10 mM iodoacetamide in 50 mM ammonium bicarbonate solution to 1×10^6 cells followed by incubation at 25 °C for 30 min in the dark. Urea was diluted to 1.5 M with ammonium bicarbonate followed by trypsin digestion at 37 °C overnight. The lysate digest was spun directly onto a C18 Stage tip at 3,500*g* through the glass mesh.

On-column TMT labeling was performed next. For resin conditioning, 50 µl methanol was added followed by 50 µl 50% ACN/0.1% FA and equilibration was done using 75 µl 0.1% FA twice. After centrifugation of the entire digest at 3,500*g* for loading, 1 µl of TMT reagent in 100% ACN was added to 100 µl freshly made HEPES (pH 8) and passed over the C18 resin at 350*g*. After washing with 75 µl 0.1% FA twice, peptide elution was performed with 50 µl 50% ACN/0.1% FA followed by a second elution with 50% ACN/20 mM ammonium formate at a pH of 10. Measurement of absorbance at 280 nm was used for estimating peptide concentrations followed by analysis of labeling efficiency of the elution. The samples were then mixed before Stage tip-based fractionation and analysis.

For Stage tip bSDB fractionation, ~20 µg total peptides was loaded using 200-µl pipette tips packed with two punches of sulfonated divinylbenzene (SDB-RPS, Empore) with a 16-gauge needle. Then 25 µl 20 mM NH_4HCO_2, at a pH of 10, was used to perform a pH switch. This was regarded as part of fraction one. This was followed by step fractionation using 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 25%, 42% and 50% ACN concentrations with each fraction collected in individual autosampler vials. They were then dried via vacuum centrifugation and stored at −80 °C until further analysis.

Data acquisition. On-line chromatography was performed by an Easy-nLC 1200 (Thermo Fisher) at a flow rate of 200 nl min^{-1}. Peptide separation was performed using PicoFrit (New Objective) columns with an internal diameter of 75 µm packed with 1.9 µm AQ-C18 material (Dr. Maisch) up to 20 cm at 50 °C. The LC gradient went from 6% B at 1 min to 30% B in 85 mins, then increased to 60% B by minute 94 and 90% by min 95, and finally to 50% B until the end of the run that lasted 110 min. Thermo Scientific Lumos Tribrid was used to perform mass spectrometry. A precursor scanning was done from 350 to 1,800 *m/z* at 60,000 resolution. Then the topmost intense multiply charged precursors within a 2-s window were selected for higher-energy collisional dissociation at 50,000 resolution. Threshold for precursor isolation width was set to 0.7 *m/z* and the maximum MS2 injection time was set at 110 ms for an automatic gain control of 6e4. Dynamic exclusion was set at 45 s and only charge states of two to six were selected for MS2. For data acquisition run, half of each fraction was injected.

Data processing. The Uniprot Human database (12/28/2017) containing common laboratory contaminants was used to search all data with Spectrum Mill (Agilent). For the search, variable modifications of N-terminal protein acetylation, oxidation of methionine, TMT-11plex labels and a fixed modification of carbamidomethylation of cysteine were used. The following cutoffs were used for the search: enzyme specificity was set to trypsin, a maximum of three missed cleavages was used, maximum precursor-ion charge state was six, and the MS1 and MS2 mass tolerance were set to 20 ppm. Using a reverse, decoy database, false discovery rates for peptides and proteins were found to be less than 1%. Identification with at least two distinct peptides and a Spectrum Mill score protein level score of ~20 were set as cutoffs for reporting of proteins.

The Spectrum Mill protein/peptide summary module was used to correct the TMT11 reporter ion intensities in each MS/MS spectrum for isotopic impurities. The afRICA correction method was used which implemented determinant calculations according to Cramer's Rule and general correction factors obtained from the reagent manufacturer's certificate of analysis.

Median values were normalized, and the median absolute deviation-scale dataset was subjected to a two-sample moderated *t*-test. Then Benjamini–Hochberg procedure was implemented to correct for multiple hypothesis testing. An arbitrary cutoff at an adjusted P value < 0.05 was drawn for differentially abundant proteins.

Bulk RNA sequencing. RNA library preparations, sequencing reactions and initial bioinformatic analysis were conducted at GENEWIZ as follows.

For library preparation with polyA selection and HiSeq sequencing, RNA samples received were quantified using a Qubit 2.0 Fluorometer (Life Technologies) and RNA integrity was checked using an Agilent TapeStation 4200 (Agilent Technologies). RNA-seq libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina following manufacturer's instructions (NEB). Briefly, mRNAs were first enriched with Oligo(dT) beads. Enriched mRNAs were fragmented for 15 min at 94 °C. First-strand and second-strand cDNAs were subsequently synthesized. cDNA fragments were end repaired and adenylated at 3′ ends, and universal adaptors were ligated to cDNA fragments, followed by index addition and library enrichment by limited-cycle PCR. The sequencing libraries were validated on the Agilent TapeStation (Agilent Technologies), and quantified by using Qubit 2.0 Fluorometer (Invitrogen) as well as by qPCR (KAPA Biosystems). The sequencing libraries were clustered on one lane of a flowcell. After clustering, the flowcell was loaded on the Illumina HiSeq instrument (4000 or equivalent) according to the manufacturer's instructions. The samples were sequenced using a 2× 150-bp paired-end configuration. Image analysis and base calling were conducted by the HiSeq Control Software. Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into fastq files and de-multiplexed using Illumina's bcl2fastq 2.17 software. One mismatch was allowed for index sequence identification.

Data analysis. The sequence reads were trimmed to remove possible adaptor sequences and nucleotides with poor quality using Trimmomatic v.0.36. The trimmed reads were mapped to the *Homo sapiens* reference genome available on ENSEMBL using the STAR aligner v.2.5.2b to generate BAM files. Unique gene hit counts were calculated by using feature Counts from the Subread package v.1.5.2. Only unique reads that fell within exon regions were counted.

GSEA was performed using Broad Institute's GSEA Software. Sets of erythroid-, megakaryocytic- and myeloid-lineage genes and genes involved in ribosome biogenesis were derived by accounting for the differentially regulated genes that passed the cutoff (adjusted P value < 0.05) in control versus RIOK2-depleted HSPCs. Additionally, canonical pathways such as KEGG, REACTOME and BIOCARTA were consulted.

Assay for transposase-accessible chromatin with high-throughput sequencing. ATAC–seq on snap-frozen control and RIOK2-depleted HSPCs was performed at the Center for Functional Cancer Epigenetics at Dana-Farber Cancer Institute. Briefly, after high-throughput Illumina sequencing of pair-ended reads, the raw reads were mapped to the reference genome (hg19) using alignment software (configure yaml). After alignment, the mapped reads were then normalized (by downsampling) to 4 million reads. The uniquely mapped reads were then piled up. Nd is defined as the number of unique region locations. N1 is defined as the number of unique region locations with only one read. The PBC (a common method to determine sample complexity) is then simply calculated as N1/Nd. ATAC–seq samples with a good PBC score (≥0.90) were analyzed. Peaks were called using the MACS peak calling software. Chromatin accessibility was analyzed using Integrative Genomics Viewer software.

Chromatin immunoprecipitation. ChIP was performed using EZChIP kit (Millipore Sigma) according to the manufacturer's instructions. Around 2–4 µg of primary antibodies was used (GATA1 (D24E4) XP rabbit monoclonal antibody (4589S, Cell Signaling); GATA1 rat monoclonal antibody (N6; SC-265, Santa Cruz Biotechnology); HA rabbit monoclonal antibody (C29F4, 3724S, Cell Signaling Technology); RIOK2 mouse monoclonal antibody (OTI3E11, TA505140, Origene); RIOK2 rabbit polyclonal antibody (NBP130098, Thermo Fisher Scientific)) or

control IgGs. The immunoprecipitated DNA was purified using Spin columns (Millipore Sigma/Qiagen) and used for RT–qPCR using QuantStudio 6 (Applied Biosciences).

ChIP sequencing. ChIP–seq library preparation, HiSeq sequencing and initial bioinformatic analysis were conducted at GENEWIZ as follows. ChIP DNA samples were quantified using a Qubit 2.0 fluorometer (Life Technologies) and the DNA integrity was checked with a 4200 TapeStation (Agilent Technologies). The NEB NextUltra DNA Library Preparation kit was used following the manufacturer's recommendations (Illumina). Briefly, the ChIP DNA was end repaired and adaptors were ligated after adenylation of the 3′ ends. Adaptor-ligated DNA was size selected, followed by clean up, and limited-cycle PCR enrichment. The ChIP library was validated using an Agilent TapeStation and quantified using a Qubit 2.0 Fluorometer as well as real-time PCR (Applied Biosystems). The sequencing libraries were multiplexed and clustered on one lane of a flowcell. After clustering, the flowcell was loaded on the Illumina HiSeq instrument according to the manufacturer's instructions (Illumina). Sequencing was performed using a 2×150-bp paired-end configuration. Image analysis and base calling were conducted by the HiSeq control software. Raw sequence data (.bcl files) generated from Illumina HiSeq were converted into fastq files and de-multiplexed using Illumina's bcl2fastq 2.17 software. One mismatch was allowed for index sequence identification.

For the bioinformatics analysis workflow, sequence reads were trimmed to remove possible adaptor sequences and nucleotides with poor quality at the 3′ end (error rate > 0.01) using CLC Genomics Server 9.0. Trimmed data were then aligned to the reference genome for human GRCh38. During the mapping, only specific alignment was allowed. The TSS plots and chromosome view plots were generated using Partek Flow software. The HOMER algorithm was used to curate de novo nucleotide-binding motifs of RIOK2 across the entire human genome.

Electrophoretic mobility shift assay. Recombinant human GST-tagged RIOK2 (rhRIOK2, Abcam) was incubated with duplex DNA (listed in Supplementary Table 2) at 5:1 molar ratio in a low-salt buffer for 20–30 min at 4 °C and 2% agarose gel was cast and cooled to 4 °C. The DNA–protein mixture was supplemented with 0.22-µm filtered 50% glycerol solution to facilitate loading. The gel was run for 1.5 h at 4 °C, followed by incubation in ethidium bromide solution for 30 min and visualized using ChemiDoc Touch Imaging system (Bio-Rad). ImageJ was used to quantify band shifts.

Nascent protein synthesis in vitro. Measurement of nascent protein synthesis in vitro was performed using Click-iT Plus OPP Alexa Fluor 647 Protein Synthesis Assay kit (Thermo Fisher Scientific) following the manufacturer's instructions. Briefly, cells in culture were supplemented with Click-iT OPP at a final working concentration of 20 µM and incubated for 30 min at 37 °C. The cells were washed with PBS, stained with fixable viability dye, fixed and permeabilized followed by incubation with Click-iT reaction buffer containing Alexa Fluor 647 picolyl azide and incubated at 25 °C for 30 min. The cells were washed and analyzed via flow cytometry.

Structural modeling of RIOK2. The crystal structure of residues 2–301 of human RIOK2 (PDB 6HK6)[49], comprising the N-terminal wHTH domain (residues 2–92) and the kinase domain (residues 93–289), was docked onto B-form DNA by aligning the wHTH domain with the crystal structure of the DNA-binding domain of the diphtheria toxin repressor bound to its DNA operator sequence (PDB 1F5T)[50]. Modeling was done using Modeller software solely to give a crude impression of the relative positions of the various subdomains of the N-terminal part of the protein when RIOK2 binds to DNA because the structure of most of the C-terminal half of the protein is unknown.

Statistical tests. Data are presented as the mean ± s.e.m. Unpaired two-tailed Student's *t*-test was used for comparing two groups. ANOVA with Tukey's/Dunnett's correction or Kruskal–Wallis test with Dunn's correction was used for comparisons among multiple groups, wherever applicable. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in a previous publication[20]. Data distribution was assumed to be normal, but this was not formally tested. Data points were not excluded from the analysis. Microsoft Excel was used to list all numerical values and Prism v8.0/9.0 (GraphPad) was used to perform statistical analyses. Sample size was not predetermined.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Raw data generated from proteomic analysis are deposited in MassIVE under accession code MSV000088238. Data generated from RNA-seq are deposited in the Gene Expression Omnibus (GEO) database under accession code GSE185922. ATAC–seq and ChIP–seq datasets are deposited in the GEO database under accession numbers GSE174714 and GSE174719, respectively. The mRNA correlations in individuals with MDS, AML and CKD have been drawn using publicly available GEO datasets under accession codes GSE19429, GSE131184 and GSE37171. This study did not generate new unique reagents. Requests for any materials (plasmids generated in this study, crRNAs) and reagents should be directed to and will be fulfilled by L.H.G. with appropriate material transfer agreements. Source data are provided with this paper.

References

46. Li, J. et al. Isolation and transcriptome analyses of human erythroid progenitors: BFU-E and CFU-E. *Blood* **124**, 3636–3645 (2014).
47. Kaufman, D. S., Hanson, E. T., Lewis, R. L., Auerbach, R. & Thomson, J. A. Hematopoietic colony-forming cells derived from human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **98**, 10716–10721 (2001).
48. Myers, S. A. et al. Streamlined protocol for deep proteomic profiling of FAC-sorted cells and its application to freshly isolated murine immune cells. *Mol. Cell Proteomics* **18**, 995–1009 (2019).
49. Wang, J., Varin, T., Vieth, M. & Elkins, J. M. Crystal structure of human RIOK2 bound to a specific inhibitor. *Open Biol.* **9**, 190037 (2019).
50. Chen, C. S., White, A., Love, J., Murphy, J. R. & Ringe, D. Methyl groups of thymine bases are important for nucleic acid recognition by DtxR. *Biochemistry* **39**, 10397–10407 (2000).

Acknowledgements

We thank V. Sankaran and R. Voit (Boston Children's hospital) for their valuable input in primary human HSPC culture techniques and for providing HMD-empty vector and HMD-GATA1 expression plasmids; S. Badrinath, N. Mathewson and K. Wucherpfennig for their kind assistance in CRISPR–Cas9 genome editing and for providing EF1α-MCS-IRES-ZsGreen lentiviral expression vector; and M. Katsuyama (Kyoto Prefectural University of Medicine) for generously providing GATA2 and GATA3 overexpression plasmids. We also thank A. Wight for assistance in analyzing RNA-seq results. This work was funed by a Discovery Research Grant, the Edward P. Evans Foundation (to L.H.G.), the National Institutes of Health (NIH) small grant program (1R03HL156574-01 to L.H.G.), institutional funding, the Dana-Farber Cancer Institute (to L.H.G.), The Pussycat Foundation Helen Gurley Brown Presidential Initiative (to S.G.), National Cancer Institute (NCI) Clinical Proteomic Tumor Analysis Consortium grants (NIH/NCI U24-CA210986 and U01 CA214125 to S.A.C.) and NIH grants (R01HL146642 and R37CA228304 to X.C.).

Author contributions

Conceptualization, S.G. and L.H.G.; design, S.G., G.A.P. and L.H.G.; methodology, S.G., S.A.M., S.A.C., X.C. and G.A.P.; data analysis, S.G., M.R., G.A.P. and L.H.G.; funding acquisition, S.G., M.R. and L.H.G.; writing–original draft, S.G., G.A.P. and L.H.G.; supervision, G.A.P. and L.H.G.

Competing interests

L.H.G. is a former Director of Bristol-Myers Squibb and the Waters Corporation and currently serves on the Board of Directors of GlaxoSmithKline Pharmaceuticals and Analog Devices and on the scientific advisory boards of Repare Therapeutics, Abpro Therapeutics and Kaleido Therapeutics. G.A.P. is on the Scientific Advisory Boards of Amicus Therapeutics, MeiraGTx, Annovis Bio, Retromer Therapeutics and Proclara Biosciences and holds equity or stock options in Denali Therapeutics, MeiraGTx, Annovis Bio, Retromer Therapeutics and Proclara Biosciences, companies that are developing therapies for neurodegenerative diseases. M.R. is currently an employee at Jnana Therapeutics. The other authors declare no competing interests.

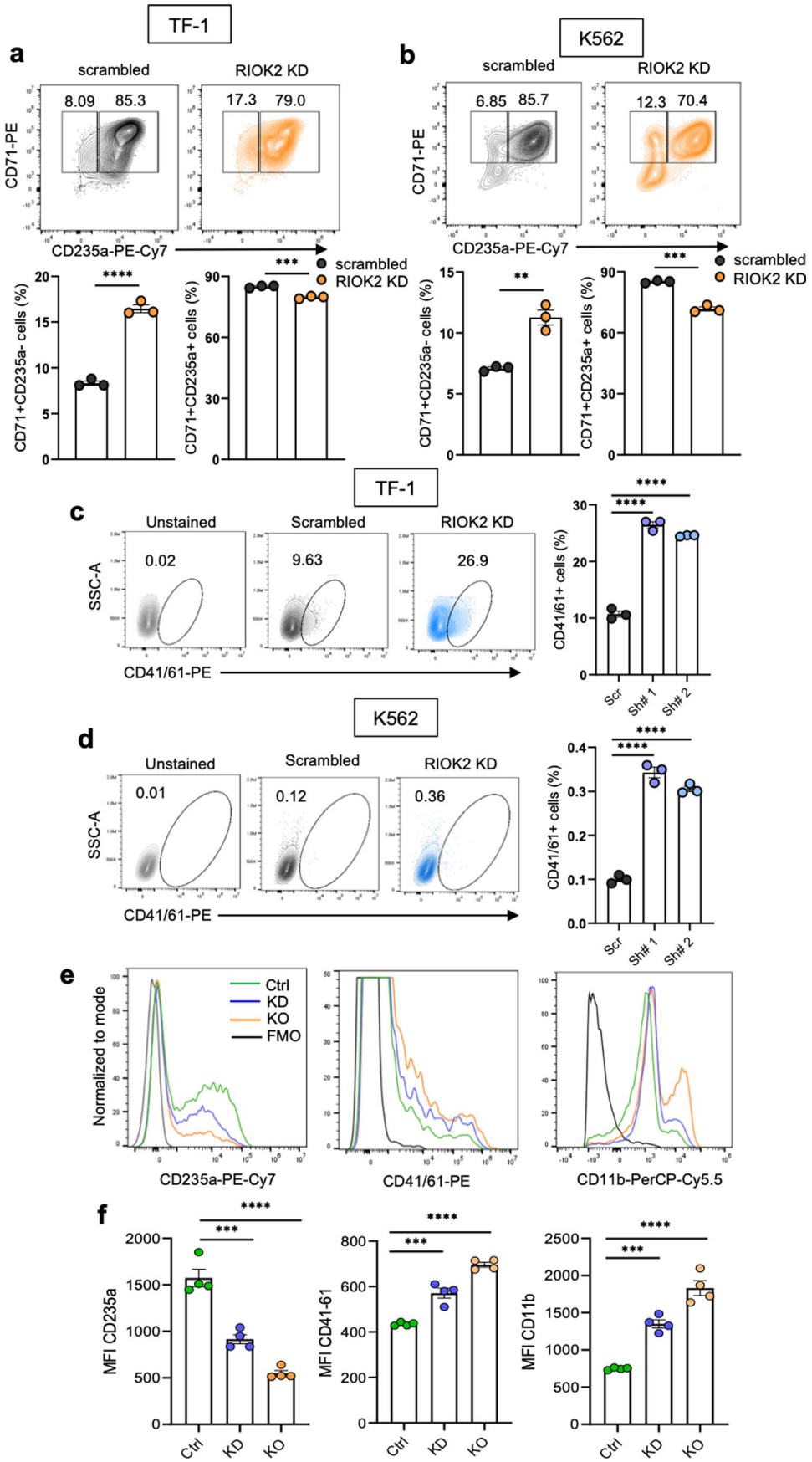
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41590-021-01079-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41590-021-01079-w.

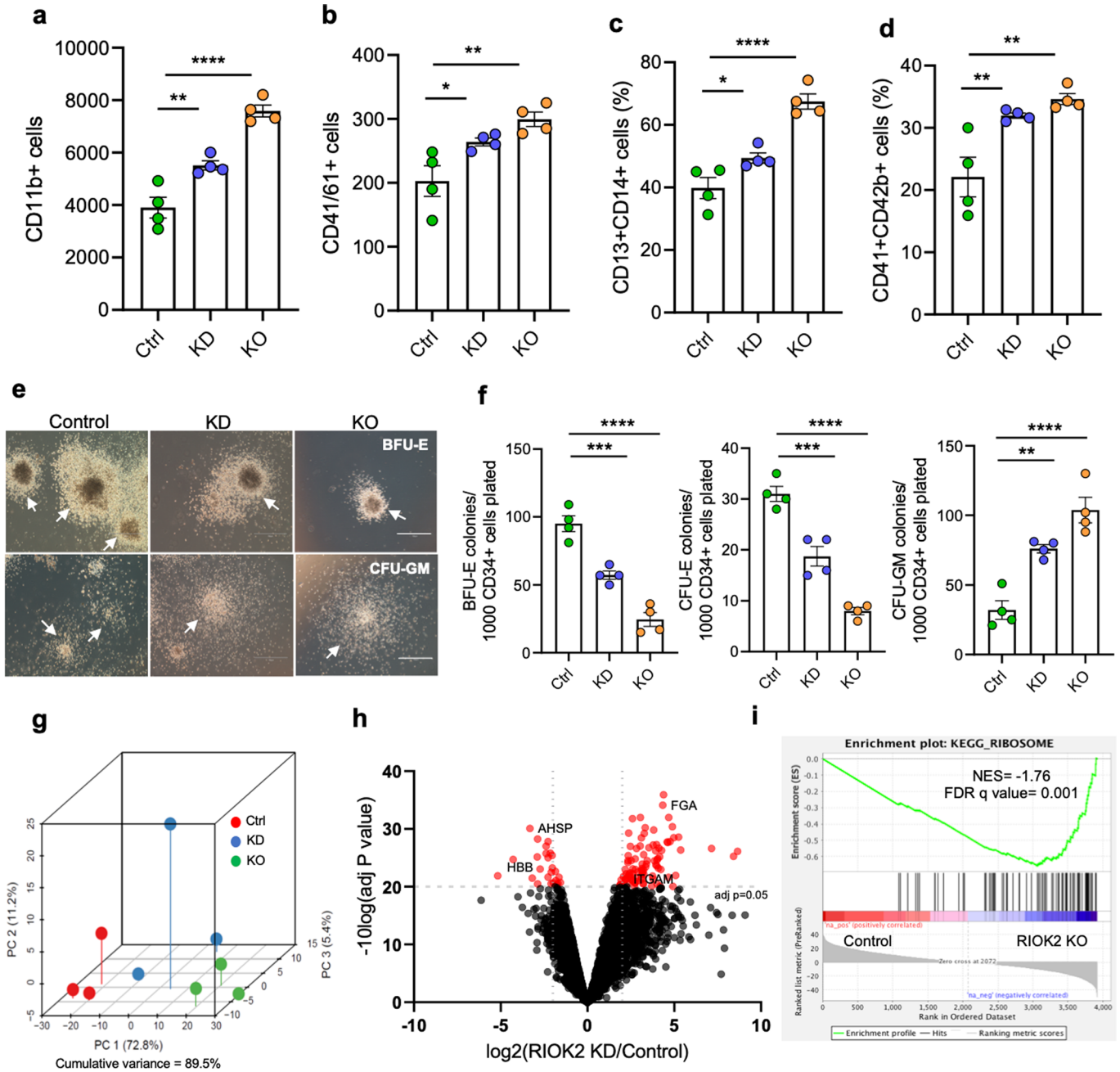
Correspondence and requests for materials should be addressed to Gregory A. Petsko or Laurie H. Glimcher.

Peer review information *Nature Immunology* thanks Kazuhiko Igarashi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. L. A. Dempsey was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

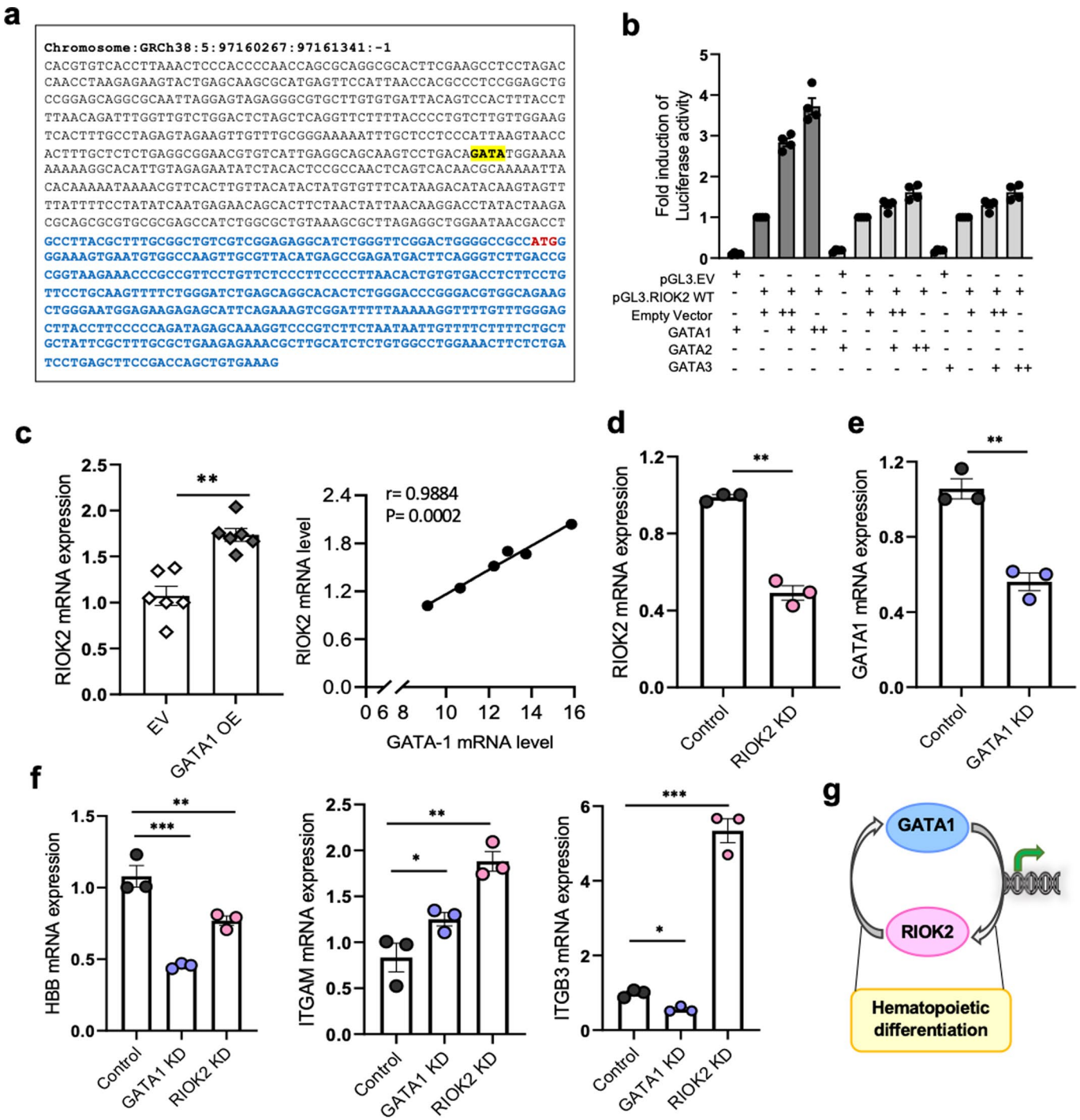


Extended Data Fig. 1 | See next page for caption.

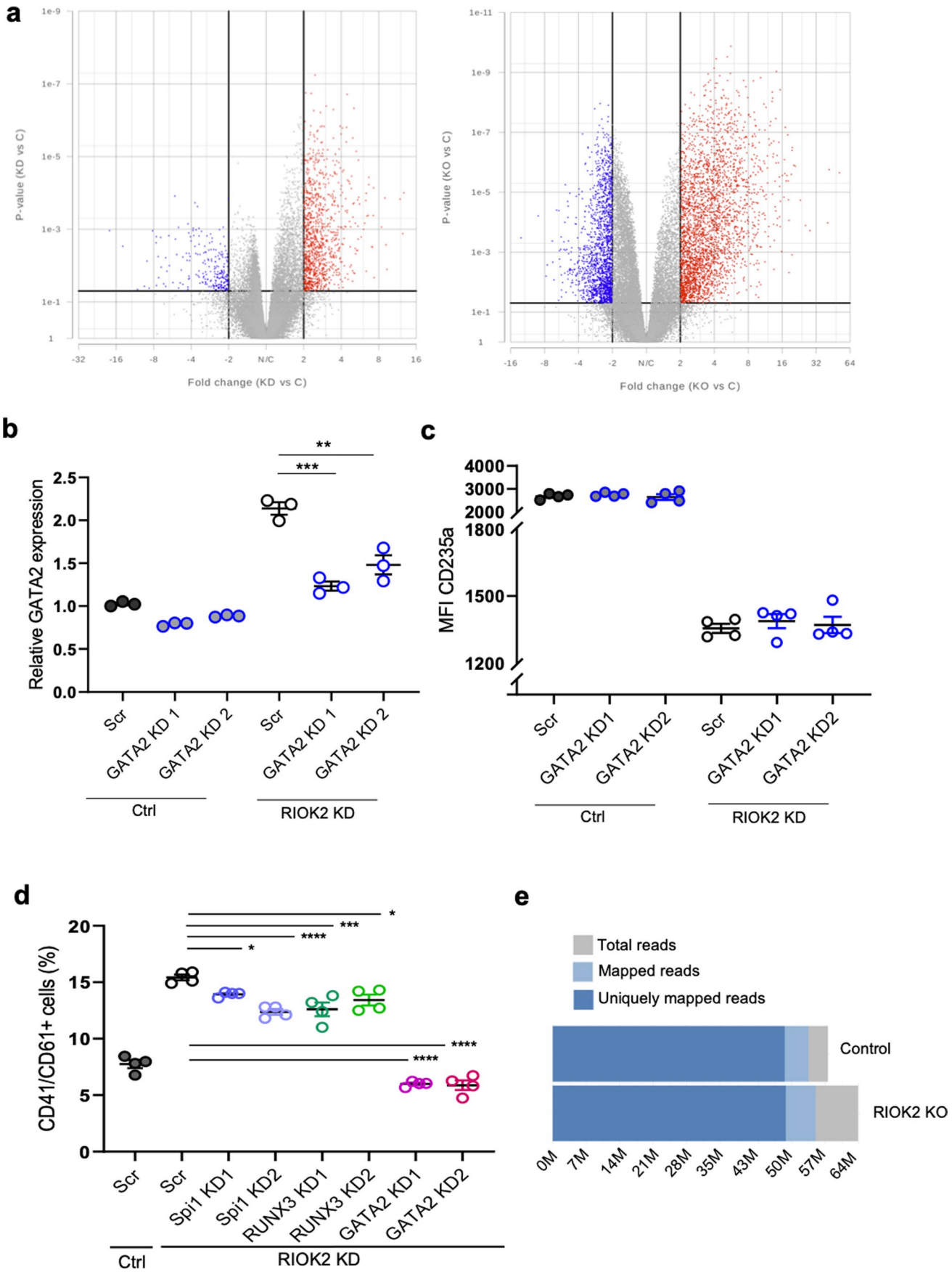
Extended Data Fig. 1 | RIOK2 drives erythropoiesis and concomitantly suppresses megakaryopoiesis and myelopoiesis. (**a,b**) Frequency of CD71 + CD235a- and CD71 + CD235a + population in control vs RIOK2 depleted TF-1 and K562 cells, respectively. (**c,d**) FACS plots and frequency of megakaryocytes (CD41/CD61 +) after RIOK2 knockdown in TF-1 and K562 cells, respectively; sh#1 and 2: shRIOK2 #1 and #2. (**e**) Histogram plots depicting erythroblasts (CD235a), megakaryocytes (CD41/61) and myeloblasts (CD11b) in differentiating HSPCs after KD and KO of RIOK2. (**f**) Quantification of data presented in e, n = 4 independent donors. n = 3 technical replicates in a-d. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, one-way ANOVA with Tukey's or Dunnett's correction in c,d,f; unpaired two-tailed Student's *t*-test in a,b. Data represented as mean ± s.e.m.



Extended Data Fig. 2 | RIOK2 drives erythropoiesis and inhibits megakaryopoiesis and myelopoiesis. (**a,b**) Total no. of myeloid (CD11b +) and megakaryocytic (CD41/61 +) cells after selective differentiation of control (Ctrl) vs RIOK2 KD and KO HSPCs to myeloid and megakaryocytic lineages, respectively. (**c,d**) Myeloid (CD13 + CD14 +) and megakaryocytic (CD41 + CD42b +) cells after selective differentiation of control (Ctrl) vs RIOK2 KD and KO HSPCs to myeloid and megakaryocytic lineages, respectively. (**e**) Images of blast forming unit-erythroid (BFU-E) and colony forming unit-granulocyte macrophage progenitors (CFU-GM) after KD and KO of RIOK2 in HSPCs, scale bar 750 µm. (**f**) Quantification of BFU-E, CFU-E and CFU-GM colonies in control vs RIOK2 KD and KO HSPCs. (**g**) Principal component analysis (PCA) of quantitative proteomic dataset showing control vs RIOK2 KD and KO primary human HSPCs, n = 3 donors. (**h**) Volcano plot showing differentially expressed proteins in control vs RIOK2 KD HSPCs. (**i**) Gene set enrichment analysis (GSEA) plot showing defective ribosome biogenesis in RIOK2 KO vs control HSPCs. n = 4 independent donors in a-d, f. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, one-way ANOVA with Tukey's or Dunnett's correction. Data represented as mean ± s.e.m. Fig. e representative of 4 independent experiments.

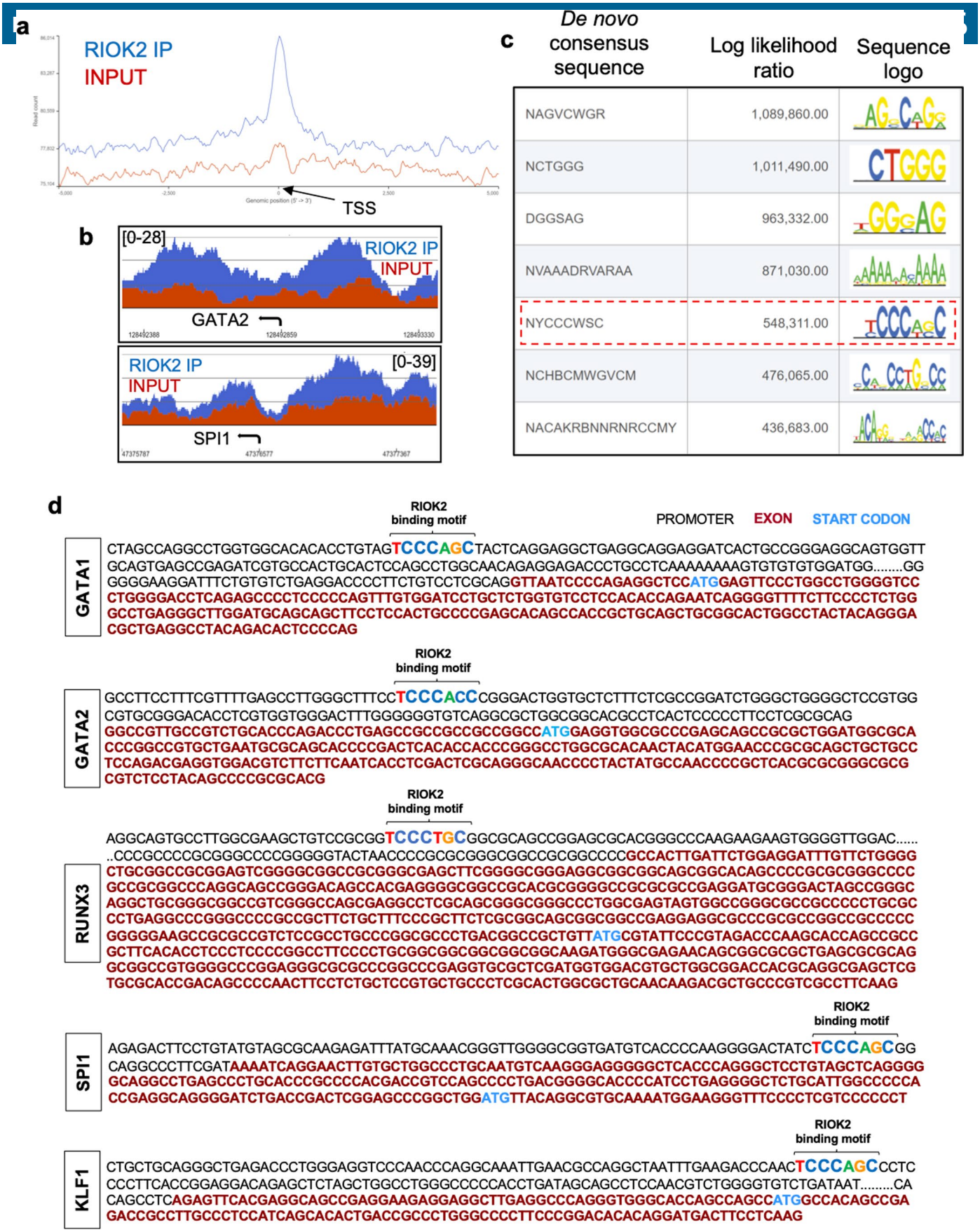


Extended Data Fig. 3 | RIOK2 and GATA1 form a positive feedback loop to regulate hematopoietic differentiation. (**a**) GATA binding motif (yellow) in the promoter region of *RIOK2* (black) followed by the first exon (blue) and ATG start codon (red). (**b**) RIOK2 promoter-driven luciferase activity in response to increasing GATA1, GATA2 and GATA3 expression in HEK293 cells. (**c**) mRNA levels of *RIOK2* (normalized to actin) after GATA1 overexpression (OE) and adjoining co-relation plot; EV: empty vector; Pearson's correlation coefficient (r) and P value are shown. (**d**) *RIOK2* mRNA expression (normalized to actin) in RIOK2 KD vs control cells. (**e**) *GATA1* mRNA expression (normalized to actin) in GATA1 KD vs control cells. (**f**) *HBB*, *ITGAM* and *ITGB3* mRNA expression (normalized to actin) in RIOK2 KD vs GATA1 KD HSPCs, compared to control. (**g**) Schema showing positive feedback loop between GATA1 and RIOK2 regulating hematopoietic differentiation. n = 3 independent donors in d-f; n = 4 and 6 technical replicates in b and c respectively. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, one-way ANOVA with Tukey's correction in f, Unpaired two-tailed Student's *t*-test in fig. c-e. Data represented as mean ± s.e.m.

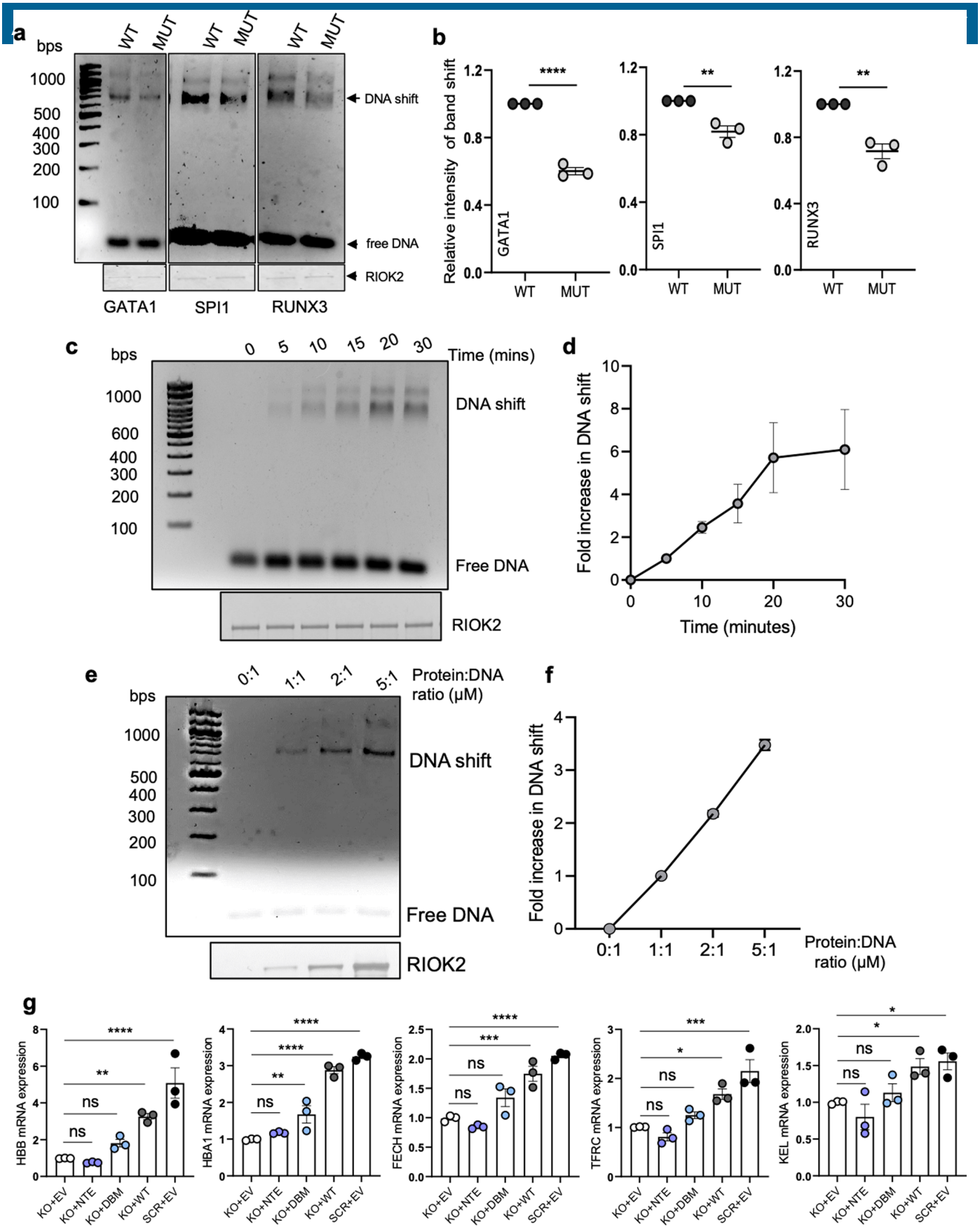


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | RIOK2 regulates expression of transcription factors involved in hematopoietic lineage commitment. (**a**) Volcano plots showing differentially expressed genes (cut off: adjusted P-Value<0.05) in control vs RIOK2 KD and control vs RIOK2 KO HSPCs; red: upregulated genes, blue: downregulated genes. (**b**) GATA2 mRNA level (normalized to actin) in control vs RIOK2 KD cells after suppression of GATA2 with 2 different crRNAs. (**c**) Erythroid progression (CD235a) in scrambled vs RIOK2 KD cells with or without suppression of GATA2 with 2 different crRNAs. (**d**) Megakaryocytic progression (CD41/61+ cells) in scrambled vs RIOK2 KD cells with or without knockdown of SPI1/RUNX3/GATA2 using 2 crRNAs against each. (**e**) Graph showing comparable mapped reads in control and RIOK2 KO HSPCs in ATAC-sequencing. n = 4 technical replicates in b-d. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, one-way ANOVA with Tukey's or Dunnett's correction. Data represented as mean ± s.e.m.

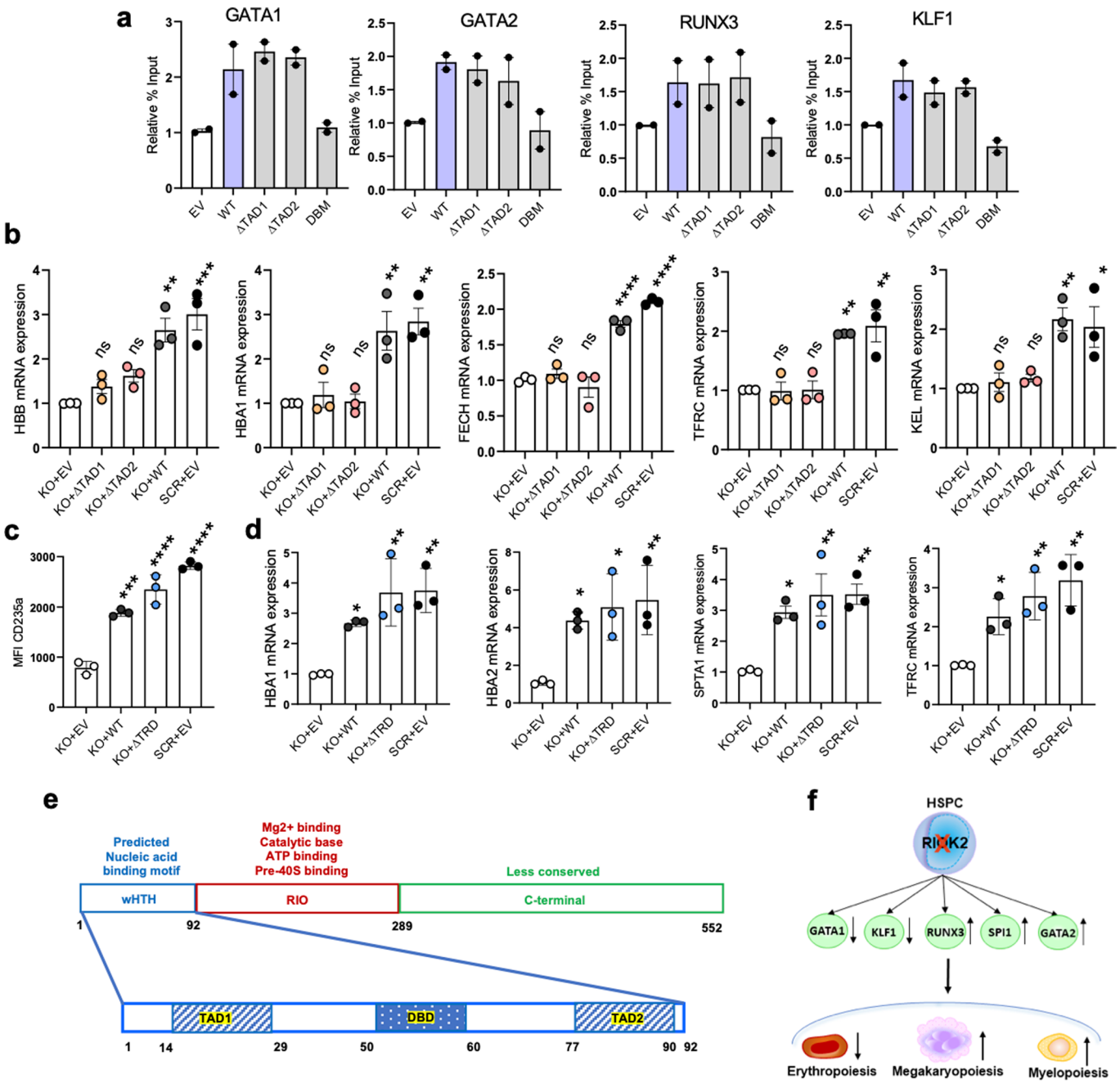


Extended Data Fig. 5 | RIOK2 binds to a specific de novo nucleotide motif in the human genome. (**a**) Relative binding intensity of RIOK2 at the transcription start site (TSS) in RIOK2 immunoprecipitated (RIOK2 IP) vs input sample. (**b**) Chromosome view plots depicting binding of RIOK2 at the promoters of *GATA2* and *SPI1*. (**c**) De novo DNA binding sequences identified in the entire genome via ChIP sequencing using monoclonal antibodies of RIOK2. The sequence highlighted in red was identified in the promoters of RIOK2's target genes. (**d**) Presence of de novo nucleotide binding motif specific for RIOK2 in the promoter regions of its putative target genes: *GATA1, GATA2, RUNX3, SPI1*, and *KLF1*.

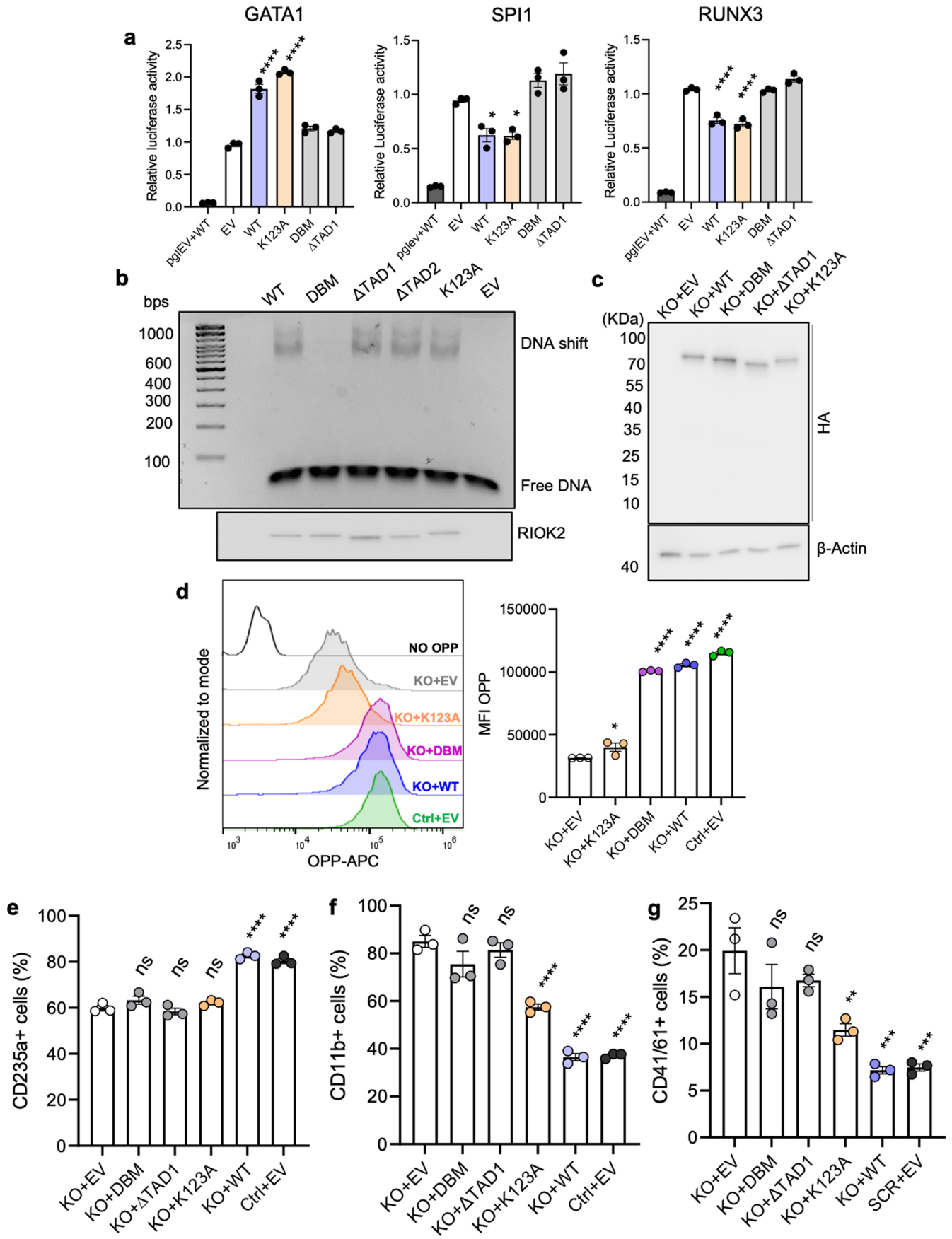


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | RIOK2 binds DNA in vitro. (**a**) Electrophoretic mobility shift assay (EMSA) showing WT (central CCC) or mutant (MUT: central CCC mutated to TTT) DNA migration in the presence of recombinant human RIOK2. (**b**) Quantitative band shift represented in a. (**c,d**) EMSA and adjoining quantification showing increasing DNA binding ability of RIOK2 over a time course; mins: minutes. (**e,f**) EMSA and adjoining quantification showing dose-dependent increase in DNA binding ability of RIOK2 with increasing protein concentration. (**g**) mRNA levels of early erythroid genes (*HBB, HBA1, FECH, TFRC, KEL*: normalized to actin) in RIOK2-KO HSPCs ectopically expressing EV, WT, DBM or NTE RIOK2, SCR: scrambled. n = 3 and 4 per group in d and f respectively. n = 3 experimental replicates in b,g. ** $p < 0.01$, **** $p < 0.0001$, ns: not significant, Unpaired two-tailed Student's *t*-test in b, one-way ANOVA with Dunnett's correction in g. Data represented as mean ± s.e.m. Fig. a,c representative of 3 independent experiments and fig. e representative of 4 independent experiments.

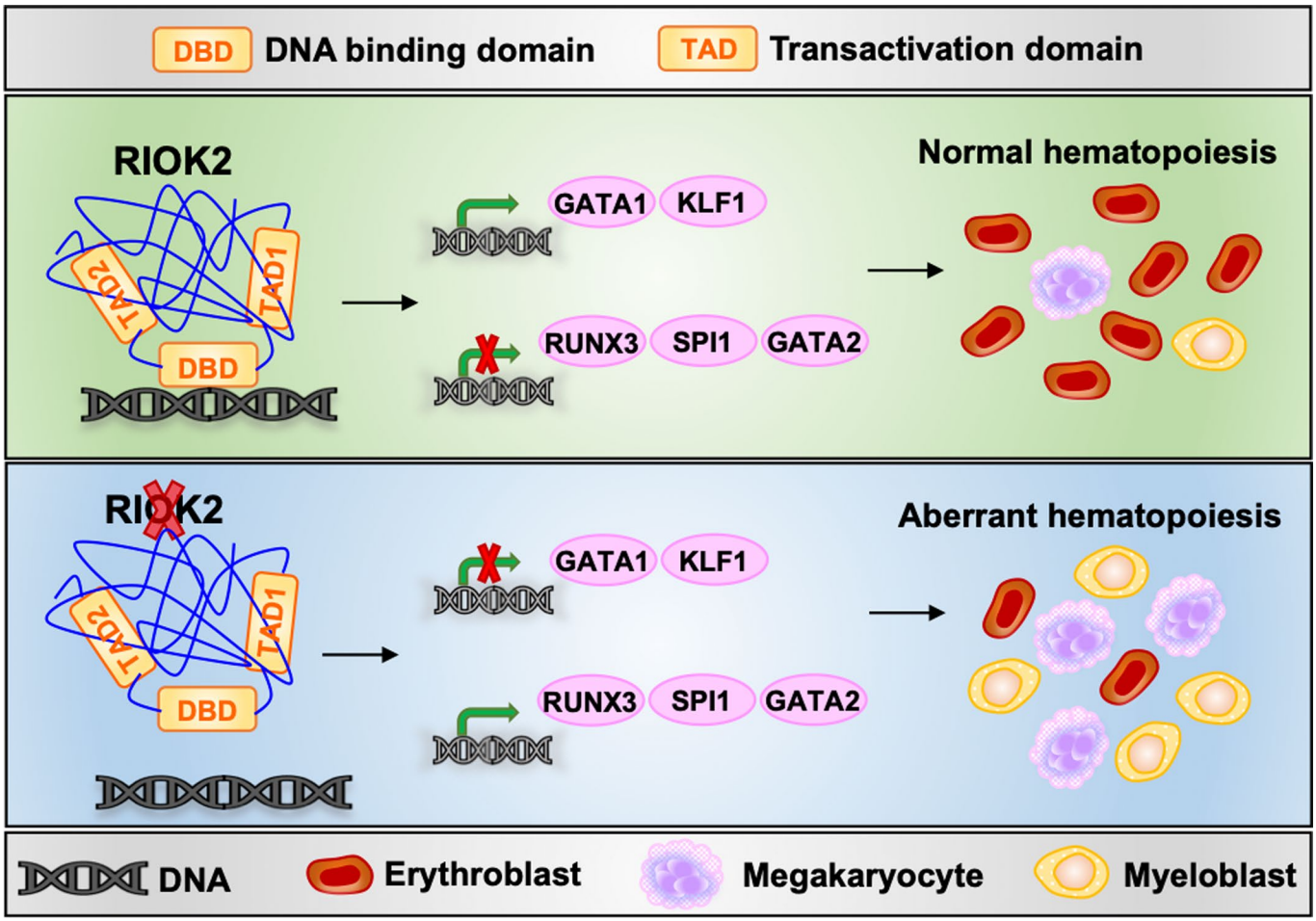


Extended Data Fig. 7 | Characterization of the transactivation (TAD) and transrepressor (TRD) domains of RIOK2. (**a**) ChIP of the promoter regions of *GATA1*, *GATA2*, *RUNX3* and *KLF1* by empty vector (EV) or HA-tagged wild-type (WT), ΔTAD1 (transactivation domain 1 deleted) and ΔTAD2 (transactivation domain 2 deleted) RIOK2 using anti-HA antibodies. (**b**) mRNA levels of early erythroid genes (*HBB*, *HBA1*, *FECH*, *TFRC*, *KEL*: normalized to actin) in RIOK2-KO HSPCs ectopically expressing EV, WT, ΔTAD1 or ΔTAD2 RIOK2, SCR: scrambled. (**c**) Erythroid progression (CD235a) in RIOK2-KO HSPCs ectopically expressing EV, WT or ΔTRD (deletion of transrepressor domain) RIOK2. (**d**) mRNA levels of early erythroid genes (*HBA1*, *HBA2*, *SPTA1*, *TFRC*: normalized to actin) in RIOK2-KO HSPCs ectopically expressing EV, WT or ΔTRD RIOK2. All comparisons done with respect to KO + EV group in fig. b-d. (**e**) Schema showing the 3 known domains of human RIOK2: N-terminal wHTH domain, central RIO domain and C-terminal domain. Positions of the DNA binding domain (DBD) and the transactivation domains (TAD1 and TAD2) of RIOK2 are shown. (**f**) Diagram illustrating RIOK2 as a master transcriptional regulator of key transcription factors (GATA1, KLF1, RUNX3, SPI1 and GATA2) in hematopoiesis. n = 2 per group in a; n = 3 independent donors in b-d. Data represented as mean ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns: not significant, one-way ANOVA with Tukey's or Dunnett's correction in b-d. Data represented as mean ± s.e.m.

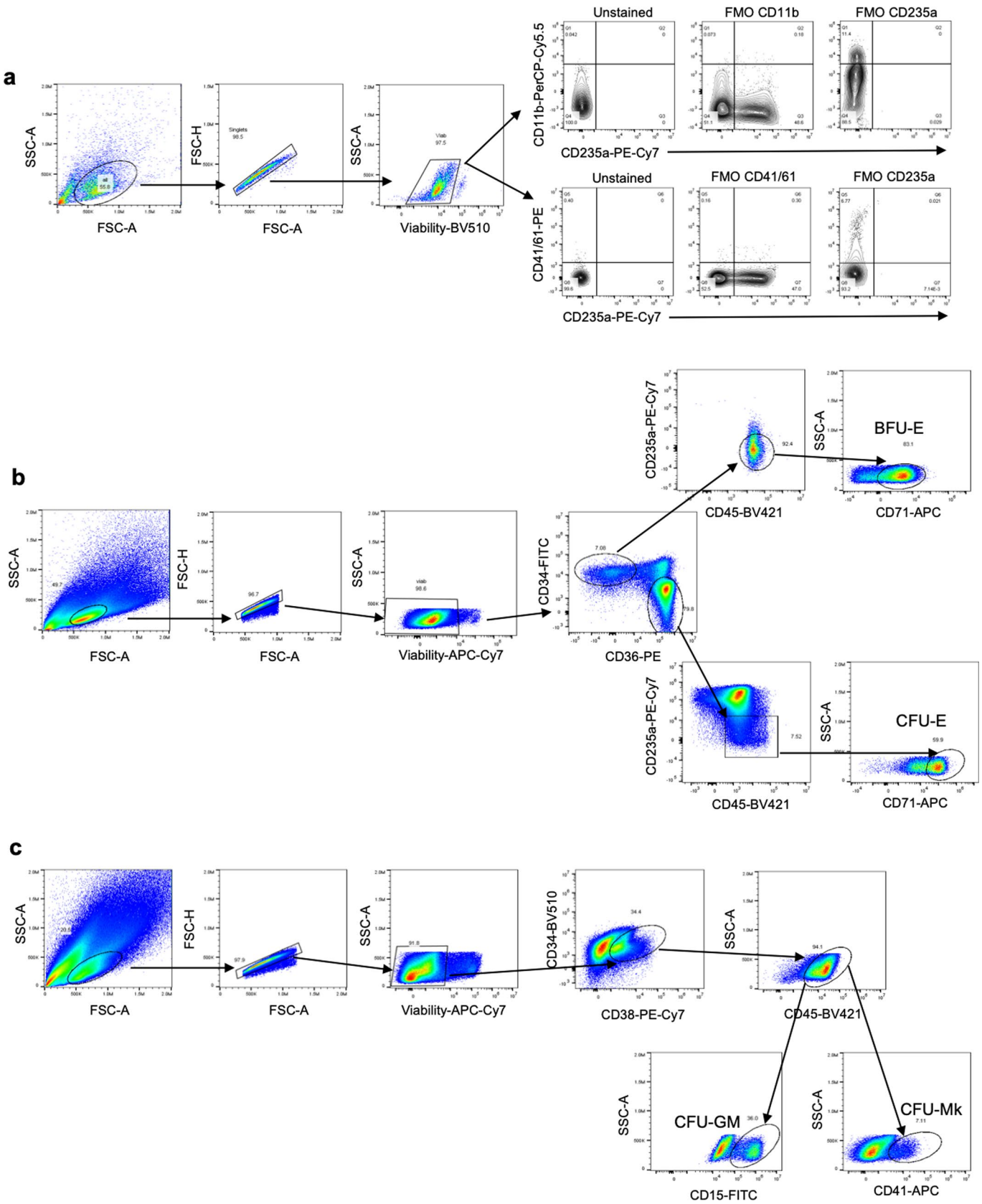


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | The kinase domain of RIOK2 is involved in erythropoiesis but does not affect RIOK2's transcriptional activities. (**a**) GATA1, SPI1 and RUNX3 promoter-driven luciferase reporter activity after expression of EV, WT, K123A, DBM or ΔTAD1 RIOK2; all comparisons done with respect to EV group (**b**) EMSA showing DNA-binding affinities of EV, WT, K123A, DBM, ΔTAD1 and ΔTAD2 RIOK2. (**c**) Western blot showing expression of ectopically expressed HA-tagged WT, DBM, K123A, ΔTAD1 RIOK2 or EV in RIOK2 KO HSPCs. (**d**) O-propargyl-puromycin (OPP) incorporation in control (Ctrl) vs RIOK2 KO cells after reconstitution of WT, DBM, K123A RIOK2 or EV. (**e-g**) Selective progression of erythroid (CD235a), myeloid (CD11b), and megakaryocytic (CD41/61) differentiation in HSPCs ectopically expressing EV, WT, DBM, ΔTAD1 or K123A RIOK2 in RIOK2-KO setting, respectively. All comparisons done with respect to EV in a, and KO + EV group in d-g. n = 3 technical replicates in a, d-g. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns: not significant, one-way ANOVA with Tukey's or Dunnett's correction. Data represented as mean ± s.e.m. Fig. b and c representative of 2 independent experiments.



Extended Data Fig. 9 | Graphical illustration of RIOK2 functioning as a master transcription factor governing hematopoietic differentiation. RIOK2 transcriptionally regulates key TFs in hematopoietic differentiation.



Extended Data Fig. 10 | Flow cytometry plots showing gating strategies. (**a**) erythroid, myeloid and megakaryocytic lineages (**b**) BFU-E and CFU-E formation (**c**) CFU-GM and CFU-Mk formation.

nature research

Corresponding author(s): Laurie H. Glimcher, M.D.
Gregory A. Petsko, D.Phil.

Last updated by author(s): Oct 18, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	CytoFLEX Flow Cytometer (Beckman Coulter), EVOS M5000 Imaging system (ThermoFisher Scientific), Proxeon UHPLC (Chromatography for mass spectrometry-based proteomic analysis), Illumina HiSeq instrument (4000 or equivalent; Life Technologies-RNA sequencing), Quantstudio6 RT-PCR system (Applied Biosciences), STEMvision automated colony counter (STEMCELL Technologies).
Data analysis	GraphPad Prism (v8 and v9); Mascot Version 2.6 (Matrix Science) for MS analysis, bulk RNA and ATAC-seq data were aligned using STAR aligner v.2.5.2b and config.yaml respectively; Partek Flow software (CHIP-seq data analysis), Basepair software (ATAC-seq data analysis: https://www.basepairtech.com/features/); Benjamini-Hochberg procedure (multiple hypothesis testing for mass spectrometry-based proteomic analysis); Modeller software by Andrej Sali v.9.0 (structural modelling), Microsoft excel (included in Microsoft Office 365), ImageJ (1.x), GSEA v.3.0, FlowJo v.10.0.7, STEMvision acquisition, analyzer and colony marker software (STEMCELL Technologies, included in STEMvision™ 21 CFR Part 11 Compliance Software Add-On).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw data generated from proteomic analysis and RNA-sequencing are submitted in MassIVE and Gene Expression omnibus (GEO) databases and their accession

codes are MSV000088238 and GSE185922 respectively. Data generated from ATAC-sequencing and ChIP sequencing are uploaded in the Gene Expression Omnibus (GEO) database and their accession numbers are GSE174714 and GSE174719. The source data for all figures (both Main and Extended) are provided with the paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine sample size. Sample sizes were estimated based on preliminary experiments, with an effort to achieve a minimum of n=3 per group, which proved sufficient to determine reproducible results.
Data exclusions	No data were excluded throughout the studies.
Replication	All experiments were reliably reproduced. All experiments (except for -omics data ie. RNA-seq, ATAC-seq, ChIP-seq and quantitative proteomics) were performed independently at least three times. The proteomics and RNA-seq data were generated from primary human stem and progenitor cells isolated from n=3 adult healthy donors for each experimental set-up, and assessed by Principal Component Analysis (PCA) and hierarchical clustering methods which showed biological replicates from the same experimental groups clustering together. The ATAC (n=2 adult healthy donors for each experimental set-up) and ChIP Seq experiments were replicated twice with 3 internal technical replicates.
Randomization	Not relevant to our study, experiments were performed in primary human stem and progenitor cells isolated from healthy adult donors or human erythroleukemia cell lines.
Blinding	For generating proteomic dataset, mass spectrometry-based analysis was performed by our collaborator blinded to the effects of our gene of interest on hematopoiesis; RNA-seq and CHIP-seq were performed by Genewiz and unbiasedly analyzed; ATAC-seq was performed at a core facility of DFCI and the researchers were blinded. In all other experiments, analysis was objective and did not require blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

All antibodies used in the study were validated, as stated in the manufacturer's website
The following primary antibodies were used:
Western blotting: Primary antibodies-HA rabbit monoclonal antibody at 1:1000 (C29F4, 3724S, Cell Signaling Technology), beta-Actin rabbit polyclonal antibody at 1:1000 (3967S, Cell Signaling Technology), RIOK2 mouse monoclonal antibody at 1:1000 (OTI3E11, TA505140, Origene). Secondary antibodies-HRP-linked anti-mouse IgG (7076S, Cell Signaling Technology) or HRP-linked anti-rabbit IgG (7076S, Cell Signaling Technology) at 1:3000 dilution.

Fluorescence-Activated Cell sorting (FACS) analysis was done using 1:100/200 times dilution of the following antibodies: FITC-CD34 (343604, BioLegend), BV421-CD38 (356617, BioLegend)/ PE-Cy7-CD38 (356608, BioLegend), APC-CD71 (OKT9, eBioscience), PE-Cy7-CD235 (306620, BioLegend), PE-CD41/CD61 (359806, BioLegend), PerCP-Cy5.5-CD11b (301328, 101228, 393106, BioLegend), AF700-CD42b (303928, BioLegend), APC-CD135 (313307, BioLegend), BV510-Human Lineage (348807, BioLegend), PE-CD68 (333807, BioLegend), BV421-CD45RA (304129, BioLegend), BV421-CD45 (304031, 368521, BioLegend), BV785-CD45 (368527, BioLegend), PerCP-Cy5.5-CD123 (306015, BioLegend), AF700-CD117 (313245, Biolegend), PE-CD36 (336205, BioLegend), BV510-CD34 (343527, BioLegend), FITC-CD15 (301904, BioLegend), APC-CD41 (343709, BioLegend), PE-CD14 (325605, BioLegend), APC-CD13 (301706,

BioLegend), PE-conjugated GATA1 rabbit monoclonal antibodies (13353S, Cell Signaling), PE-conjugated HA-tag rabbit monoclonal antibodies, or PE-conjugated control rabbit IgG isotype (5742, Cell Signaling).

Chromatin Immunoprecipitation (ChIP): (GATA-1 (D24E4) XP Rabbit monoclonal antibody (4589S, Cell Signaling); GATA-1 rat monoclonal antibody (N6) (SC-265, Santa Cruz Biotechnology); HA rabbit monoclonal antibody (C29F4, 3724S, Cell Signaling Technology), RIOK2 mouse monoclonal antibody (OTI3E11, TA505140, Origene); RIOK2 rabbit polyclonal antibody (NBP130098, Thermo Fisher Scientific)) or control IgGs (Millipore/ Santa Cruz Biotechnology).

All used antibodies are commercially available.

Validation

All the antibodies used for Western blotting, chromatin immunoprecipitation and flow cytometry are validated for the use of immunoblotting, immunoprecipitation and flow cytometry, as available on the manufacturer's website.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

TF-1 human erythroblast cells from ATCC (ATCC® CRL-2003TM), K562 human erythroid cells from ATCC (ATCC® CCL-243), HEK293FT cells from ATCC, CD34+ primary human hematopoietic stem and progenitor cells (HSPCs) from Fred-Hutchinson Cancer Research center, Seattle, USA.

Authentication

All authenticated and validated cell lines were used in this study, as available on the company's website. Staining for the surface markers (CD235a, CD71, CD11b, CD41/61, CD34 and CD38) was used to define the differentiation status of the primary and secondary cells used.

Mycoplasma contamination

Cell lines were not tested for mycoplasma contamination but no indication of contamination was observed.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

Data generated from ChIP sequencing are uploaded in the Gene Expression Omnibus (GEO) database and the accession number is GSE174719. The raw .bam and .bai files are provided.

Files in database submission

.bam and .bai files.

Genome browser session
(e.g. UCSC)

No longer applicable.

Methodology

Replicates

CHIP-seq experiments were performed on TF-1 cells, and the data obtained from sequencing were validated by CHIP analysis in n=3 experimental replicates.

Sequencing depth

2x150 Paired End (PE) configuration, a total of 30 million reads for each sample.

Antibodies

RIOK2 mouse monoclonal antibody (OTI3E11, TA505140, Origene).

Peak calling parameters

Image analysis and base calling were conducted by the HiSeq Control Software (HCS). Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into fastq files and de-multiplexed using Illumina's bcl2fastq 2.17 software. One mis-match was allowed for index sequence identification.

Data quality

Sequence reads were trimmed to remove possible adapter sequences and nucleotides with poor quality at 3' end (error rate > 0.01) using CLC Genomics Server 9.0. Trimmed data was then aligned to reference genome for human genome GRCh38. During the mapping, only specific alignment was allowed. After investigating the quality of the raw data, sequence reads were trimmed to remove possible adapter sequences and nucleotides with poor quality using Trimmomatic v.0.36. The trimmed reads were mapped to the Homo sapiens reference genome available on ENSEMBL using the STAR aligner v.2.5.2b. The STAR aligner is a splice aligner that detects splice junctions and incorporates them to help align the entire read sequences. BAM files were generated as a result of this step. Unique gene hit counts were calculated by using feature Counts from the Subread package v.1.5.2. Only unique reads that fell within exon regions were counted.

Software

The TSS plots, chromosome view plots and de novo nucleotide binding sequences were generated using Partek Flow software.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	For analysis of cellular surface marker expressions via flow cytometry, in vitro cultured cells were washed twice with PBS and incubated with fixable viability dyes (Tonbo Biosciences) at 1:800 dilution in PBS for 20 mins at RT in dark. The cells were washed once with staining buffer and labelled with required fluorochrome-conjugated antibodies diluted in staining buffer for 30 minutes at 4°C: FITC-CD34 (343604, BioLegend), BV421-CD38 (356617, BioLegend)/ PE-Cy7-CD38 (356608, BioLegend), APC-CD71 (OKT9, eBioscience), PE-Cy7-CD235 (306620, BioLegend), PE-CD41/CD61 (359806, BioLegend), PerCP-Cy5.5-CD11b (301328, 101228, 393106, BioLegend), AF700-CD42b (303928, BioLegend), APC-CD135 (313307, BioLegend), BV510-Human Lineage (348807, BioLegend), PE-CD68 (333807, BioLegend), BV421-CD45RA (304129, BioLegend), BV421-CD45 (304031, 368521, BioLegend), BV785-CD45 (368527, BioLegend), PerCP-Cy5.5-CD123 (306015, BioLegend), AF700-CD117 (313245, Biolegend), PE-CD36 (336205, BioLegend), BV510-CD34 (343527, BioLegend), FITC-CD15 (301904, BioLegend), APC-CD41 (343709, BioLegend), PE-CD14 (325605, BioLegend), APC-CD13 (301706, BioLegend). For intracellular staining, cells were labelled for surface markers as mentioned above, followed by fixation and permeabilization using Foxp3/Transcription factor staining kit (Thermo Fisher Scientific). PE-conjugated GATA1 rabbit monoclonal antibodies (13353S, Cell Signaling), PE-conjugated HA-tag rabbit monoclonal antibodies, or PE-conjugated control rabbit IgG isotype (5742, Cell Signaling) were used at 1:100 dilution.
Instrument	All FACS analyses were performed using CytoFLEX Flow Cytometer (Beckman Coulter).
Software	Data were analyzed using FlowJo v.10.0.7 and plotted using GraphPad Prism v8/v9.
Cell population abundance	Populations of genome-edited cells were analyzed by FACS, cell sorting was not performed. For lentivirally transduced cells, ZsGreen+ cells were sorted and further analyzed. Intracellular staining for Hemagglutinin (HA) was also used to confirm transduction efficiency.
Gating strategy	The starting cell population was gated on SSC-A/FSC-A plot. Singlet cells were identified by FSC-H/FSC-A gating. Viable cells were identified by staining with fixable Zombie dyes. Positive/Negative populations were determined by unstained, FMO and isotype controls.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.