HHS Public Access

Author manuscript

Immunohorizons. Author manuscript; available in PMC 2022 March 07.

Published in final edited form as:

Immunohorizons. ; 5(10): 802–817. doi:10.4049/immunohorizons.2100083.

Deficiencies in the DNA Binding Protein ARID3a Alter Chromatin Structures Important for Early Human Erythropoiesis

Joshua Garton[1], **Malini Shankar**[2], **Brittany Chapman**[2], **Kira Rose**[3], **Patrick M. Gaffney**[4], **Carol F. Webb**[2,3,5,*]

[1]Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK, 73019, USA

[2]Departments of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK, 73104, USA

[3]Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, 73104, USA

[4]Genes and Human Disease Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, 73104, USA

[5]Cell Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, 73104, USA

Abstract

ARID3a is a DNA-binding protein important for normal hematopoiesis in mice, and for in vitro lymphocyte development in human cultures. ARID3a knockout mice die in utero with defects in both early hematopoietic stem cell populations and erythropoiesis. Recent transcriptome analyses in human erythropoietic systems revealed increases in ARID3a transcripts implicating potential roles for ARID3a in human erythrocyte development. However, ARID3a transcript levels do not faithfully reflect protein levels in many cells, and the functions and requirements for ARID3a protein in those systems has not been explored. We used the erythroleukemic cell line K562 as a model to elucidate functions of ARID3a protein in early human erythropoiesis. ARID3a knockdown of hemin stimulated K562 cells resulted in lack of fetal globin production and modifications in gene expression. Temporal RNA-seq data link ARID3a expression with the important erythroid regulators *Gata1*, *Gata2* and *Klf1*. Ablation of ARID3a using CRISPR-Cas9 further demonstrated it is required to maintain chromatin structures associated with erythropoietic differentiation potential. These data demonstrate the ARID3a protein is required for early erythropoietic events and provide evidence for the requirement of ARID3a functions for proper maintenance of appropriate chromatin structures.

[*]To whom correspondence should be addressed: Carol F. Webb, Tel: 405-271-4188, carol-webb@ouhsc.edu, Mailing Address: Department of Medicine, OUHSC, 800 Research Park, Ste. 419, Oklahoma City, OK 73104.
Present address-Kira Rose - Department of Chemistry, University of Central Oklahoma, Oklahoma City, OK, 73034, USA

CONFLICT OF INTEREST

The authors declare no competing financial interests.

INTRODUCTION

The intricate network of transcription factors (TFs) that drive hematopoiesis and erythropoiesis, including development of specific subpopulations, have not been fully elucidated (1). The erythroid-specific globin gene cluster is regulated by an upstream cis-regulatory region, the locus control region (LCR) (2, 3), that binds to specific transcription factors allowing accessibility to the embryonic and adult globin in a developmentally controlled fashion (4-6). Decreased levels of hemoglobin gene transcripts and maturational arrest of erythroid lineages in thalassemia's are associated with defects in expression of the transcription factors GATA1, GATA2 and KLF1, as well as with changes in chromatin accessibility of enhancer regions for the globin genes (7), indicating the importance of these factors in erythropoiesis. Transcripts for ARID3a (A+T rich binding protein 3a) were previously identified to be increased significantly throughout primitive erythropoiesis in a mouse model (8) implicating ARID3a as a potential regulator of hematopoietic reprogramming in combination with GATA1 and GATA2 through motif associations (9-11), possibly through binding to distal enhancer regions (12). However, others found discrepancies in protein versus transcript levels during human erythropoiesis, particularly for GATA1 (13), and we found that ARID3a transcripts do not always correlate with protein levels in mature hematopoietic cell subsets (14). Recently, co-IP LC-MS data, using the megakaryoblastic cell line CMK, revealed that both ARID3a and GATA1 acted in concert for proper regulation of megakaryopoiesis (9), but it is unclear if ARID3a is required for early erythropoiesis in human cells. Therefore, it is critical to assess requirements for individual transcription factors at the protein level during hematopoietic events.

ARID3a was orignally discovered for its ability to increase immunoglobulin transcription in B cells (15-17) and is a member of a large family of proteins, many of which have important roles as epigenetic regulators (18). Modulation of ARID3a levels in cord blood results in skewing of lineage fate decisions (13) and changes in the developmental plasticity within HSCs (19). ARID3a can both suppress and enhance individual gene expression in a cell type-specific fashion (20, 21). We previously reported thatARID3a-deficient mice die in utero at E12.5 due to failed erythropoiesis (22), suggesting that ARID3a could be important for erythropoiesis in mice. However, ARID3a knockout embryos also exhibit a 90% reduction in hematopoietic stem cell numbers (22). It is currently unclear if failed erythropoiesis resulted directly from requirements for ARID3a during erythropoiesis or from earlier hematopoietic progenitor defects.

Our earlier data indicated that the human monomyelocytic cell line K562 constitutively expresses ARID3a protein (23). This human cell line has been used, for some time, as a model for erythrocyte and myeloid lineage development and can be induced with exogenous stimuli to differentiate into erythroid cells that express high levels of embryonic and fetal globin genes (13, 24-27). Therefore, we used this model system to determine if ARID3a protein is required for human globin gene expression and to explore how ARID3a might mediate alterations in gene expression in those cells. Furthermore, we used temporal transcriptome analyses and integrated chromatin accessibility data from ATAC-seq to determine how ARID3a affects gene expression patterns during induction of erythroid

lineage differentiation. These results identify new lineage-specific functions for ARID3a in human erythropoiesis.

MATERIALS AND METHODS

Cell Culture and Transfection

K562 (ATCC® CCL-243™) cells were plated in triplicate at $1x10^5$ per well in 6-well plates with RPMI 1640 + 7.5% fetal calf serum overnight at 37°C prior to treating cells with 0.04 mM hemin (Sigma), as previously described (28). Cells were harvested at 24, 48 and 72 hours and viabilities were assessed using trypan blue exclusion. To evaluate erythroid lineage differentiation, cells were stained with benzidine to detect globin expression, as reported previously (29). Briefly, cells were resuspended in 25μL phosphate-buffered saline (PBS) and stained at a 1:1 ratio with benzidine solution made with 30% fresh hydrogen peroxide. At least 200 cells were evaluated per replicate. Lentivirus expressing shRNA specific for ARID3a, or an unrelated control shRNA, both of which co-express green fluorescent protein (GFP) allowing visualization of infected cells, were purchased from Genecopoeia, Inc,Rockville,MD and used at a multiplicity of infection of 0.6 to 1.0, as previously described (30). The ARID3a sequence targeted was GCAGTTTAAGCAGCTCTA from exon 2, and does not react with other ARID family members (30). K562 cells were infected with virus 30 minutes to 3 hours prior to stimulation with hemin in the presence of 8 μg/ml polybrene following our previous work (30). Lentivirus transfection efficiency was assessed via GFP expression using a Zoe Fluorescent Imager, BioRad, on day two and was typically >70%.

Flow Cytometry

Antibodies to the transferrin receptor, an erythroid precursor marker, CD71 APC-Cy7 (Biolegend Cat #33410) and the glycophorin A erythrocyte marker CD235a PE-Cy7 (Biolegend Cat #349112) were used for surface staining to evaluate erythroid lineage differentiation. Appropriate isotype controls from Bio Legend were used for gating. Myeloid lineage detection was evaluated using surface markers CD24 APC (Biolegend Cat # 311118) and CD33 PE-Cy5 (Biolegend Cat # 303406). Following surface marker staining, cells were fixed with fixation buffer (Bio Legend Cat # 420801), permeabilized with Foxp3/ Transcription Factor Staining Buffer Set (Invitrogen eBioscience Cat # 00552300) and stained for ARID3a with goat antihuman ARID3a peptide-specific antibody, as we described previously (31). Donkey anti-goat IgG PE (Invitrogen Cat# Pl31860) was used as the secondary antibody. Data were collected on a Stratedigm S1200Ex and data post processing and analysis was performed using FlowJo (Tree Star) software version 10.

RNA-seq and Analyses

Total RNA from triplicate samples treated with and without hemin, ARID3a shRNA, and/or scrambled shRNA was isolated using NucleoSpin RNA XS kits (Macherey-Nagel, Cat # 740902.50). RNA concentrations were measured with an Impen Nanophotometer. RNA integrity numbers were obtained using an Agilent 2200 TapeStation. Library construction was performed as described previously (14). Briefly, the Ovation RNA-Seq v2 (NuGEN Technologies) kit was used to generate sequencing libraries. Paired-end (2 x 50bp)

sequencing was performed on a NovaSeq platform. Fastq files were demultiplexed and sequencing adapters were removed using Cutadapt (32). Briefly, we created a Bowtie (33) index based on the UCSC knownGene (34) transcriptome, and aligned paired-end reads directly to this index using Bowtie2. The average sequence depth was 21M reads with an average alignment of 83% mapping to the hg38 genome assembly. Next, we ran RSEM v1.3.0 (35) using default parameters to obtain transcript per million (TPM) values for each gene. Genes with expression values of TPM > 1 in half of the samples were retained, leaving 11,869 transcripts for downstream analyses. Differential gene expression was analyzed using DESeq2 v3.5 (36). Differentially expressed genes (FDR < 0.05) with fold changes ≥ 2 were used for Ingenuity Pathway Analysis (Qiagen) (IPA). Hierarchical clustering (Euclidean) was performed on differentially expressed genes (FDR value adjusted < 0.05) and heatmaps were generated with the pHeatmap package in R. Principal component analysis (PCA) was performed in R using the prcomp function.

ARID3a knockout

Genome editing of ARID3a was performed via CRISPR/ Cas9 mutation of the K562 cell line contracted through Synthego (Redwood City, CA). Briefly, modified guide RNA ARID3a-932711 (5'-CCTCGTAAGTCCAGTCGCCG-3' [TGG]-PAM) targeting exon 3 was chosen to be specific for ARID3a. A bulk knockout sample of greater than 70% knockout was then single cell sorted via flow cytometry for isolation of homozygous ARID3a knockout clones. Sixty-six clones visually confirmed to have only one cell per well after sorting were allowed to grow, and 61 clones were screened by flow cytometry for ARID3a protein expression. Eight of 19 clones selected by flow cytometry were then selected as being wild type, or potentially homozygous knockout, and levels of ARID3a expression were confirmed by Western blotting using a commercial ARID3a antibody (mouse monoclonal IgG Catalog# sc-398367, Santa Cruz Biotech). Homozygous colonies and wild type colonies were used for ATAC-seq analyses.

Western Blotting

For protein detection, total cell extracts from $1x10^6$ cells were resuspended in 50μL of Laemmli sample buffer containing 5% 2-mercaptoethanol. Following 5 min boiling at 90 °C, 10μL of extract was loaded onto pre-cast Mini-PROTEAN TGX (BioRad Cat# 456-1093) gel and transfer was done for 1 h on nitrocellulose 0.2μm (BioRad). Membranes were blocked in 1% gelatin in TBST for 1 h at room temperature, as previously described (37). Blots were probed overnight for ARID3a and actin with mouse anti-ARID3a and rabbit anti-β-actin, respectively. Following incubation with primary antibody, blots were washed three times for 10 mins with TBST and probed with secondary antibody for 1 h at room temperature. The secondary antibody for ARID3a was goat anti-mouse IgG and rabbit anti-goat IgG for actin. Blots were then washed three times for 10 min with TBST. Proteins were detected using the AP conjugate substrate kit (BioRad Cat # 170-6432).

ATAC-seq and analyses

ATAC-seq libraries were generated from wild type and ARID3a$^{-/-}$ K562 clones treated with or without hemin. Duplicate samples of 30,000 cells were washed in cold PBS, pelleted by centrifugation and lysed using cold lysis buffer (10mM Tris-HCl, pH 7.4, 10mM

NaCl, 3mM $MgCl_2$, 0.1% IGEPAL CA-630). Nuclei were collected by centrifugation and the pellet was resuspended in 50μL transposase reaction mix (25μL 2x TD buffer, 2.5μL transposase (Illumina Nextera FC121-1030 TDE1 and TD buffer) and 22.5μL nuclease-free water). The transposition reaction was incubated at 37°C for 30 min. Samples were cleaned using a MiniElute kit (Qiagen) following manufacturer's protocol and eluted in 10μL buffer EB. Library construction was done by PCR in a reaction mix containing 25 μL 2x NEBNext PCR master mix (New England Biolabs), 10μL transposed sample, and 5μL primer mix (1.25μM each of Nextera XT adapter1 and adapter2 primer mix). PCR conditions were 72°C for 5 min, 98°C for 30s, and 11 cycles (98°C for 10s, 68°C for 30s, 72°C for 1 min), ending with 72°C for 5 min. PCR amplified sequencing libraries were cleaned with AMPure Beads (Beckman Coulter). Library quality was determined by analysis on an Agilent TapeStation.

For each sample, 25-99 million 50bp paired-end reads were obtained on an Illumina NextSeq sequencer. All data processing steps were performed within the Partek Flow Genomics Analysis software. Fastq files were processed and both sequencing primers and Nextera transposase adapters were removed using Cutadapt. Trimmed reads were aligned to the hg38 GRCh38 reference genome using Bowtie2v2.2.5 with parameters –very-sensitive -X 2000 (33). Low quality (-Q 30) and duplicate reads, and reads mapping to the ENCODE project blacklist, mtDNA and rDNA genes were removed. MACS2 (38) was used to call peaks for duplicate samples using the parameters -q 0.05 --nolambda --slocal 1000 --llocal 10000 -m 5 50 --shift 0 --extsize 200 --fe-cutoff 1.0.The ATAC peaks of pooled replicate samples were annotated to genomic regions such as transcription start sites (TSS), introns, and exons using RefSeq version 89. Each peak was annotated in relation to these genomic elements and may have multiple gene annotations. DESeq2 v3.5 was run on the ATAC peaks to identify differential chromatin accessibility in wild type (n=8) vs ARID3a$^{-/-}$ (n=8) samples with an FDR cutoff of 0.05 (36).

Statistics

Data for viability, benzidine stain and time course were plotted and all statistical analyses were performed using Prism (Graphpad) version 7. A one-way ANOVA was used for comparisons of multiple groups, followed by Tukey posttest for multiple comparison corrections. All statistical tests and corresponding *P* values are stated in the figure legends. *P* values <0.05 were considered significant.

RESULTS

ARID3a knock down inhibits globin production and expression of erythrocyte markers.

The human K562 erythroleukemia cell line was treated *in vitro* with hemin for five days to allow visible production of red, fetal hemoglobin-producing cells (Figure 1A). While cells treated with an irrelevant control shRNA (Figure 1A, right) resembled those treated with hemin only, cells which received ARID3a shRNA showed no obvious red cells (Figure 1A, middle). Time course analyses indicated near maximal production of globin was achieved in hemin treated cells by day three of treatment, and reduced numbers of globin-producing cells were apparent as early as day one after the inhibition of ARID3a (Figure 1B). Robust inhibition of globin expression was also observed in ARID3a-inhibited samples on day three

in multiple separate experiments (Figure 1 C). Viabilities and cell numbers (not shown) were equivalent on day three in all cultures (Figure 1D), suggesting ARID3a inhibition did not cause cell death or inhibition of cell division. These data suggests that ARID3a is necessary for globin production.

To further assess the effects of ARID3a inhibition on erythroid differentiation in this model, we performed flow cytometry to evaluate the presence of known erythrocyte markers (5). Erythrocyte lineage markers CD71 (TFRC) and CD235a (GYPA) were enhanced as expected by hemin treatment, while cells treated with ARID3a-specific shRNA, with or without hemin stimulation, more closely resembled unstimulated cells with respect to expression of these surface markers (Figure 1E). Similarly, while hemin stimulation resulted in increased expression of monocyte marker CD33 and CD24, cells treated with ARID3a-specific shRNA more closely resembled untreated cells (Figure 1F). The control shRNA stained cells expressed surface markers similar to those of hemin treated cells. Flow cytometric analyses of intracellular ARID3a protein levels on day three of culture confirmed that ARID3a inhibition resulted in less ARID3a protein (Figure 1G). Results for three independent experiments are quantified in Fig. 1H. Together, these data suggest that the ARID3a protein is required for early erythrocyte lineage differentiation in hemin stimulated K562 cells.

ARID3a inhibition of hemin stimulated cells results in down regulation of genes associated with erythroid differentiation.

To further explore the block in erythroid differentiation in ARID3a inhibited samples, a time course RNA-seq experiment was performed over three days in K562 cells treated with and without hemin and with or without ARID3a inhibition. Triplicate samples from each treatment condition were sequenced and differential expression analyses were performed to identify genes affected by ARID3a inhibition. Principal component analysis revealed that untreated samples cluster away from hemin treated samples on both days 2 and 3 (Figure 2). ARID3a-inhibited samples clustered more closely to untreated samples by day 3. Day 2 ARID3a-inhibited samples were closely clustered between untreated and hemin treated samples (Figure 2), indicating perturbed differentiation at this early time point.

Differential expression analyses were performed on untreated vs hemin-stimulated cells at each time point (n=3, FDR < 0.05, FC > ±1.5) to identify genes important for erythroid lineage differentiation. There were 846, 2,228, and 1,055 differentially expressed genes (DEGs) on days one, two and three, respectively, in hemin-treated compared to untreated controls. Hierarchical clustering of the DEGs highlights the differences in transcriptomes over time and shows temporal expression of genes induced or repressed by hemin treatment (Figure 3). Observations confirmed the differential expression of key genes involved in erythroid lineage differentiation (8), such as the hemoglobin genes alpha- *(HBA1, HBA2)* and β-globin *(HBG1, HBG2), NFE2*, *TAL1* and *KLF1*, although they were not all differentially expressed on all three days (Figure 3). These data agree with previous transcriptome analyses of hemin-stimulated K562 cells using microarrays (28, 39, 40), validating our model system. Pathway analyses identified erythrocyte development among the top pathways (Table 1). Transcription, mRNA splicing, nuclear export, and autoimmune

pathways were enriched in hemin-treated samples by day two, as well as histone and nucleosome processes. These pathways were also enriched at day three indicating that most of the changes at the level of transcription were evident by day two.

To identify genes affected by ARID3a inhibition, we performed differential expression analysis on triplicate hemin-stimulated samples treated with a scramble shRNA or ARID3a shRNA with focus on day two (Figure 4A). ARID3a suppressed cultures revealed strong attenuation of hemin-induced transcriptional activation of GATA2, HEMGN, LDB1 and ZFP361. Quantification of transcripts per million (TPM) values for select erythroid genes show effects of ARID3a inhibition (Figure 4B). The expression of the erythroid differentiation marker CD71 (TFRC) examined in Figure 1 was also significantly reduced upon ARID3a inhibition. Additionally, both α- and β-globin genes were significantly reduced within the first two days, suggesting that the majority of changes in gene expression occurred within the first two days.

Further analyses of differential gene expression at day two by Venn diagram indicates overlapping DEGs affected by each treatment condition (Figure 5A). Hemin induction affected more genes than were affected by ARID3a inhibition. These analyses identified 227 overlapping DEGs upregulated by hemin treatment and downregulated by ARID3a inhibition (Figure 5A). Pathway analyses of this gene list identified SLE signaling, mRNA processing (GO:0006396), RNA splicing (GO:0008380) and chromatin binding (GO:0003682) pathways (Figure 5B). Identification of over-represented transcription factor binding sites within −2kbp to +500bp of the promoters of the 227 genes induced by hemin and repressed by ARID3a shRNA showed significant enrichment in genes with binding sites forYY1, PAX3, SIX6, ATF1, and ARID3B. Essential TFs for erythropoiesis (*GATA1, GATA2, KLF1, NFE2*) and their cofactors/mediators (*MED1* and *LDB1*) were all inhibited on day two in samples treated with hemin and shRNA. A list of the top 35 most significantly differentially expressed genes affected by hemin stimulation and those repressed by ARID3a inhibition is given in Table 2. Among the top DEGs repressed by ARID3a, the majority were either transcription factors, micro RNAs, or other small nuclear RNAs involved in splicing. In addition, when the 227 genes affected both by hemin and ARID3a were analyzed for inferred protein associations by Ingenuity Pathway analyses, networks with functions related to cell cycle, cell death and survival, and cell morphology were identified. Interestingly, these analyses identified ARID3a in the top network as a transcription factor associated with FOS and YY1 using both hemin stimulated and ARID3a shRNA inhibited DEGs (Figure 5D-E). Together, these data suggest that ARID3a is important for appropriate gene regulation of factors required for early erythropoiesis.

Chromatin accessibility is altered in ARID3a KO K562 cells

Due to the large number of genes affected by ARID3a inhibition, including histone subunits, we hypothesized that, like other ARID family members, ARID3a might function to maintain important chromatin domains for erythropoiesis. CRISPR-Cas9 gene editing of K562 cells was used to generate clones with biallelic inactivation of *ARID3a* for ATAC-seq analyses (Figure 6) to explore this hypothesis. Single guide RNAs (sgRNAs) targeting exon 3, which codes for the extended DNA-binding domain specific to ARID3 family members, were

used to generate genomic deletions of *ARID3a* (Figure 6A). Bulk deleted clones were then single cell sorted and individual clones were screened for deletion by PCR (not shown), flow cytometry and western blotting (Figure 6B-C). Two clones (BH and AL) selected via reduced intracellular staining were confirmed to exhibit no detectable ARID3a protein via western blotting. In addition, flow cytometry indicated that hemin stimulation of the ARID3a knockout clones did not induce expression of the surface markers CD71 and CD235a (Figure 6D), two gene products that were previously identified as being reduced by ARID3a shRNA inhibition in Figures 1 and 4, further validating those data.

These two knockout clones and two wild type clones were split into duplicate cultures and were used with and without hemin treatment for ATAC-seq analyses. Mixed model analysis was performed using Partek Genomics Suite and identified 504 genomic regions with differential chromatin accessibility in ARID3a WT vs ARID3a KO K562 cells with and without hemin treatment (Figure 7A), and 271 were mapped near genes (−1000bp to +100bp of any TSS) (Figure 7B). A large percentage of differentially accessible sites were also intergenic regions (Figure 7B). Unsupervised hierarchical clustering of all 504 regions indicated that wild type and knockout accessible regions grouped together irrespective of hemin treatment, and showed both increased and decreased regions of accessibility associated with ARID3a deficiency (Figure 7C). Alterations in chromatin accessibility associated with ARID3a KO were particularly evident in large intragenic regions as demonstrated in Figure 7D, and may represent enhancers. A list of the top 10 regions that were most significantly up- or down-regulated between wild type and ARID3a KO clones is shown in Figure 7E, and eight of those are intragenic regions of unknown function. Homer analysis of the promoter regions of genes with increased chromatin accessibility identified NFATC1, ZIC2, and GATA3 motifs (Figure 7F). The same analysis was performed on regions with decreased accessibility and identified enrichment of ARNT:AHR, MTF1, and TEAD motifs (Figure 7G). Pathway analysis of DARs increasing in accessibility in ARID3a KO cells compared to WT cells identified erythropoietin-mediated signaling and thrombin as the top pathways, consistent with roles for ARID3a in erythropoietic functions (Table 3). B cell receptor signaling was also identified as an enriched pathway, consistent with previous data indicating multiple roles for ARID3a in B lymphocyte development and function (31).

Comparison of the 158 DEGs affected by ARID3a inhibition (Figure 4) with the 271 DARs associated with specific genes (Figure 7) only identified 9 genes present in both data sets (Figure 8 A, B) where ARID3a promoter accessibility might be directly correlated with alterations in gene transcription. GO analysis on the 9 overlapping genes revealed associations with TNF-mediated signaling, protein folding, erythrocyte differentiation and protein degradation (Figure 8C). Many of the ARID3a-associated DARs did not map near gene transcription start sites (Figure 7), but may involve regions that function distally as enhancers as occurs with ARID3a in the immunoglobulin heavy chain locus (17, 41, 42). Together, these data suggest that ARID3a-associated effects on transcription are likely to be mediated through multiple mechanisms not limited to alterations in promoter accessibility.

ARID3a deficiency directly affects chromatin regions associated with globin gene regulation

The globin locus control region (LCR) is an intergenic region located far upstream of the fetal globin genes that is critical for developmental regulation of those genes (43). This LCR revealed reduced accessibility in ARID3aKO clones compared to wild type clones, particularly after hemin treatment (Figure 9A). Data from ENCODE identified ARID3a binding sites via ChIP-seq (44) (Figure 9A), in many cases with overlapping GATA1, GATA2, and TAL1 TF binding sites. The erythroid-specific TF loci for NFE2 and TAL1 were also significantly less accessible in ARID3a^{-}KO clones than in wild type cells (Figure 9B-C). Quantification of ATAC peaks from the 5 hypersensitive (HS) sites within the globin LCR region and within the NFE2 and TAL1 loci are indicated in Figure 9D and E. Together, these data suggest that ARID3a is required to maintain appropriate chromatin configurations for globin gene expression and erythropoietic differentiation in K562 cells.

DISCUSSION

In this study, we demonstrated that ARID3a protein is required for hemin-induced, erythrocyte lineage differentiation of the human K562 model cell line. Knockdown of ARID3a with shRNA resulted in a visible reduction in globin production and downregulation of erythroid lineage surface markers CD235a (GYPA) and CD71 (TFRC). RNA-seq analysis revealed that ARID3a is necessary for the expression of genes important for hemin-induced differentiation. Furthermore, our data reveal that ARID3a deficiency results in alterations in chromatin landscapes that contribute to erythropoiesis, and suggest that ARID3a, like other ARID family members, is important for epigenetic regulation of chromatin accessibility. These data indicate a previously unappreciated role for ARID3a in human erythropoiesis and are the first to document genomic sites altered by depletion of ARID3a in this process.

We previously found *ARID3a* knockout embryos exhibited profound defects in erythropoiesis at day 12.5 of gestation (22). Consistent with our observations, Kingsley et al., found ARID3a transcripts were enriched 22-fold in primitive erythropoiesis in the mouse (8). Additional studies in human erythroid progenitors indicated that the *ARID3a* gene locus was differentially methylated during early erythropoiesis with higher expression in fetal erythroblasts (45). More recently, the *ARID3a* locus was identified to be differentially methylated in human primary basophilic erythroblasts (46). Further, others linked ARID3a with the erythroid master regulators, GATA1 and TAL1, in human erythropoietic studies that determined transcription factor landscapes of enhancers in erythroid progenitors (6, 47), but these studies did not directly examine effects due to ARID3a. Our data confirm downregulation of globin genes (*HBA1, HBA2,* and *HBZ*) in ARID3a deficient cells and reveal that ARID3a deficiency leads to blocks in differentiation and repression of key erythroid-specific TFs (*GATA1, GATA2, KLF1, NFE2*) and genes encoding critical cofactors (*MED1, LDB1, CCAR1*) for hemoglobin expression. Alpha-globin genes (*HBA1, HBA2,* and *HBZ*), components of mediator complexes important for erythropoiesis (*MED1*), and histone subunits were among the 227 genes induced by hemin and down-regulated by ARID3a, suggesting ARID3a has a role in globin transcription, or mediates cofactor binding

to transcription start sites (TSS) of erythroid-specific genes through epigenetic mechanisms. Our studies extend previous data that were limited to transcript analyses, and definitively demonstrate a requirement for ARID3a protein in human erythroid development in this model cell line.

We identified 158 differentially regulated genes associated with ARID3a inhibition at day two in this system (Figure 4). These genes showed significant enrichment of the transcription factor binding sites, GATA1 and GATA2. Both GATA1 and GATA2 are critical regulators of erythropoiesis (48-51). GATA2 is expressed in erythroid precursors (52), and as GATA1 levels increase, GATA2 is replaced by GATA1 at many sites throughout the genome, a process called GATA switching (53, 54). Studies of enhancer turnover in $CD34^+$ cells suggested that ARID3a could be associated with the GATA2-to-GATA1 switch, raising the possibility that ARID3a could be involved in epigenetic modifications in those cells (8). ENCODE data of K562 cells showed considerable overlap between GATA1, GATA2, and ARID3a binding sites in many genes important for erythropoiesis, suggesting ARID3a may function with those factors, either as a transcription factor or as an epigenetic regulator mediating opening/closing of chromatin in enhancer/promoter regions. Knockdown of ARID3a with shRNA in GATA1-mutated cells revealed a block in both megakaryocytic and erythroid differentiation and revealed 65% of predicted ARID3a binding sites in K562 cells overlap with GATA1 sites (9). Indeed, the globin locus and other differentially regulated genes exhibit close proximity of binding sites for ARID3a, GATA1, GATA2, and TAL1 (Figure 9). This raises the possibility that ARID3a could be a part of the transcription machinery, which also contains GATA1, to drive erythroid-specific gene programs. Moreover, the GATA switch is mediated by positioning of polycomb subunits EZH2 (55). *EZH2* was significantly downregulated upon inhibition of ARID3a in our studies and our unpublished data suggest it may interact directly with ARID3a in K562 cells. Further studies will be needed to explore if ARID3a is important for the GATA switch through regulation of, or interactions with EZH2 and other associated transcription factors.

Our ATAC data revealed that ARID3a deficiency results in both increases and decreases in chromatin accessibility in both coding and non-coding regions of the genome. Indeed, unsupervised hierarchical clustering of hemin stimulated and unstimulated KO and WT clones suggest that hemin stimulation did not dramatically alter chromatin landscapes (Figure 7C). Rather, the presence or absence of ARID3a defined the majority of the chromatin alterations. These data suggest that ARID3a functions to establish chromatin landscapes necessary for erythroid differentiation as observed by decreases in chromatin accessibility in erythroid-specific enhancer regions in ARID3a KO clones, including the LCR that is essential for the developmentally controlled expression of embryonic, fetal, and adult globin genes. EHMT1 adds repressive H3K9me2 marks to the LCR region (56), and showed 2-fold increased accessibility in ARID3a KO clones. EHMT1 also adds repressive histone marks to H3K9me2 at the γ-globin locus in human adult erythroid cells, thereby reducing expression of both γ-globin and fetal globin (56). Future studies will be required to how ARID3a contributes to these effects. However, our data suggest that ARID3a may participate in mediation of multiple epigenetic events necessary for erythropoiesis, and that these events may require context-dependent transcriptional activities.

It is likely that ARID3a functions in coordination with other epigenetic factors to mediate its effects. While ARID3a contains additional DNA-binding specificity not observed in other ARID family members, it does not contain obvious epigenetic regulatory domains associated with many of the other ARID family members (57). Identifying the proteins associated with ARID3a will likely be necessary to fully understand its functions, and those proteins are likely to interact in cell type specific fashions, and perhaps through activation of enhancers. Of the 504 differentially accessible regions identified by ATAC-seq, 233 of these regions were located in intergenic regions with unknown function. Moreover, there were 12 intergenic regions among the top 20 most differentially accessible regions (Figure 7). The importance of intergenic regions is emphasized by a study on the super-enhancer-derived RNA, alncRNA-EC7/Bloodlinc, which is required for terminal erythropoiesis and red blood cell production (42). Our ATAC data show reduced accessibility of this enhancer region. It is not currently possible to distinguish which of these intergenic regions with altered chromatin accessibility directly contribute to erythropoietic functions, versus other hematopoietic events. For example, some of these regions may be important in other hematopoietic cells where ARID3a is linked to disease activity, such as lupus (27, 58). Histone subunits (*HIST1H2BN*, *HIST1H3B, HIST1H3H*, and *HIST1H4J*), chromatin remodelers (*SAT2B*), heme biosynthesis enzymes (*ALAS1*), and genes implicated in Systemic Lupus Erythematosus (SLE) signaling pathways (*PPP1R15A* and *TROVE2*) were repressed by ARID3a shRNA on day two. Thus, it is not possible from these data alone to elucidate the specific functions of ARID3a-regulated regions in this or other cell types.

Limitations of this study include the use of the K562 transformed cell line which may not faithfully mimic all aspects of fetal globin expression in primary erythropoietic progenitors. In addition, while our data suggest that ARID3a is required for normal transcription levels and to maintain normal chromatin configurations in these cells, our data do not suggest that ARID3a alone is sufficient to mediate all of the alterations observed. It is likely that additional proteins associate with ARID3a to mediate these effects. In addition, some of the ARID3a-associated effects we observed may be indirect effects due to alterations of expression of other transcription factors and/or epigenetic regulators.

Understanding how hemoglobin expression and erythropoiesis are regulated is critical for the development of new therapeutics for diseases such as sickle cell disease and thalassemia's. Further elucidation of how ARID3a functions in erythropoiesis and other hematopoietic events could lead to development of new therapeutic agents for blood disorders. Together, these data expand our knowledge of the importance of ARID3a in hematopoiesis, and particularly in erythroid lineage development, and define new functional and regulatory roles for ARID3a.

ACKNOWLEDGEMENT

We thank the Clinical Genomics Core Facility at Oklahoma Medical Research Foundation, the Flow Cytometry Core Facility at OUHSC, and the Stephenson Cancer Center at OUHSC for core support. We also thank Ken Jones for helpful discussions.

FUNDING

This work was supported by the National Institutes of Health [AI123951 and AI118836 to C.F.W., T32 AI007633 to J.W.G.].

DATA AVAILABILITY

RNA-seq and ATAC-seq data are publicly available through the GEO NCBI database at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131649 under the accession number GSE131649. For original data, contact C. Webb at carol-webb@ouhsc.edu. The following secure token has been created to allow review of record GSE131649 while it remains in private status: **ofilcsesddgfvaf**

References

1. Xavier-Ferrucio J, and Krause DS. 2018. Concise Review: Bipotent Megakaryocytic-Erythroid Progenitors: Concepts and Controversies. Stem Cells 36: 1138–1145. [PubMed: 29658164]
2. Forrester WC, Thompson C, Elder JT, and Groudine M. 1986. A developmentally stable chromatin structure in the human beta-globin gene cluster. Proc Natl Acad Sci U S A 83: 1359–1363. [PubMed: 3456593]
3. Thein SL 2018. Molecular basis of beta thalassemia and potential therapeutic targets. Blood Cells Mol Dis 70: 54–65. [PubMed: 28651846]
4. Patrinos GP, de Krom M, de Boer E, Langeveld A, Imam AM, Strouboulis J, de Laat W, and Grosveld FG. 2004. Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. Genes Dev 18: 1495–1509. [PubMed: 15198986]
5. Pimkin M, Kossenkov AV, Mishra T, Morrissey CS, Wu W, Keller CA, Blobel GA, Lee D, Beer MA, Hardison RC, and Weiss MJ. 2014. Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. Genome Res 24: 1932–1944. [PubMed: 25319996]
6. Huang J, Liu X, Li D, Shao Z, Cao H, Zhang Y, Trompouki E, Bowman TV, Zon LI, Yuan G-C, Orkin SH, and Xu J. 2016. Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. Developmental Cell 36: 9–23. [PubMed: 26766440]
7. Rivella S 2009. Ineffective erythropoiesis and thalassemias. Current opinion in hematology 16: 187–194. [PubMed: 19318943]
8. Kingsley PD, Greenfest-Allen E, Frame JM, Bushnell TP, Malik J, McGrath KE, Stoeckert CJ, and Palis J. 2013. Ontogeny of erythroid gene expression. Blood 121: e5–e13. [PubMed: 23243273]
9. Alejo-Valle O, Weigert K, Bhayadia R, Ng M, Emmrich S, Beyer C, Schuschel K, Ihling C, Sinz A, Flasinski M, Issa H, Regenyi E, Labuhn M, Reinhardt D, Yaspo M-L, Heckl D, and Klusmann J-H. 2021. The megakaryocytic transcription factor ARID3A suppresses leukemia pathogenesis. bioRxiv: 2021.2004.2026.440795.
10. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, and Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods 10: 1213. [PubMed: 24097267]
11. Gomes AM, Kurochkin I, Chang B, Daniel M, Law K, Satija N, Lachmann A, Wang Z, Ferreira L, Ma'ayan A, Chen BK, Papatsenko D, Lemischka IR, Moore KA, and Pereira CF. 2018. Cooperative Transcription Factor Induction Mediates Hemogenic Reprogramming. Cell Rep 25: 2821–2835 e2827. [PubMed: 30517869]
12. Chaudhri VK, Dienger-Stambaugh K, Wu Z, Shrestha M, and Singh H. 2020. Charting the cis-regulome of activated B cells by coupling structural and functional genomics. Nature Immunology 21: 210–220. [PubMed: 31873292]
13. Gillespie MA, Palii CG, Sanchez-Taltavull D, Shannon P, Longabaugh WJR, Downes DJ, Sivaraman K, Espinoza HM, Hughes JR, Price ND, Perkins TJ, Ranish JA, and Brand M. 2020. Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. Molecular Cell.

14. Ratliff ML, Garton J, Garman L, Barron MD, Georgescu C, White KA, Chakravarty E, Wren JD, Montgomery CG, James JA, and Webb CF. 2019. ARID3a gene profiles are strongly associated with human interferon alpha production. J Autoimmun 96: 158–167. [PubMed: 30297159]
15. Webb CF, Das C, Coffman RL, and Tucker PW. 1989. Induction of immunoglobulin mu mRNA in a B cell transfectant stimulated with interleukin-5 and a T-dependent antigen. J Immunol 143: 3934–3939. [PubMed: 2480380]
16. Webb CF, Das C, Eaton S, Calame K, and Tucker PW. 1991. Novel protein-DNA interactions associated with increased immunoglobulin transcription in response to antigen plus interleukin-5. Mol.Cell.Biol 11: 5197–5205. [PubMed: 1922039]
17. Herrscher RF, Kaplan MH, Lelsz DL, Das C, Scheuermann R, and Tucker PW. 1995. The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. Genes Dev. 9: 3067–3082. [PubMed: 8543152]
18. Kortschak RD, Tucker PW, and Saint R. 2000. ARID proteins come in from the desert. Trends Biochem.Sci 25: 294–299. [PubMed: 10838570]
19. Ratliff ML, Templeton TD, Ward JM, and Webb CF. 2014. The Bright Side of Hematopoiesis: Regulatory Roles of ARID3a/Bright in Human and Mouse Hematopoiesis. Frontiers in immunology 5: 113. [PubMed: 24678314]
20. Garton J, Barron MD, Ratliff ML, and Webb CF. 2019. New Frontiers: ARID3a in SLE. Cells 8.
21. Ratliff ML, Garton J, James JA, and Webb CF. 2020. ARID3a expression in human hematopoietic stem cells is associated with distinct gene patterns in aged individuals. Immunity & Ageing 17: 24. [PubMed: 32905435]
22. Webb CF, Bryant J, Popowski M, Allred L, Kim D, Harriss J, Schmidt C, Miner CA, Rose K, Cheng HL, Griffin C, and Tucker PW. 2011. The ARID Family Transcription Factor Bright Is Required for both Hematopoietic Stem Cell and B Lineage Development. Mol Cell Biol 31: 1041–1053. [PubMed: 21199920]
23. Nixon JC, Rajaiya JB, Ayers N, Evetts S, and Webb CF. 2004. The transcription factor, Bright, is not expressed in all human B lymphocyte subpopulations. Cell Immunol 228: 42–53. [PubMed: 15203319]
24. Lozzio BB, Lozzio CB, Bamberger EG, and Feliu AS. 1981. A Multipotential Leukemia Cell Line (K-562) of Human Origin. Proceedings of the Society for Experimental Biology and Medicine 166: 546–550. [PubMed: 7194480]
25. Andersson LC, Jokinen M, and Gahmberg CG. 1979. Induction of erythroid differentiation in the human leukaemia cell line K562. Nature 278: 364–365. [PubMed: 570644]
26. Lozzio BB, and Lozzio CB. 1979. Properties and usefulness of the original K-562 human myelogenous leukemia cell line. Leukemia Research 3: 363–370. [PubMed: 95026]
27. Rutherford T, Clegg JB, Higgs DR, Jones RW, Thompson J, and Weatherall DJ. 1981. Embryonic erythroid differentiation in the human leukemic cell line K562. Proceedings of the National Academy of Sciences 78: 348–352.
28. Addya S, Keller MA, Delgrosso K, Ponte CM, Vadigepalli R, Gonye GE, and Surrey S. 2004. Erythroid-induced commitment of K562 cells results in clusters of differentially expressed genes enriched for specific transcription regulatory elements. Physiological Genomics 19: 117–130. [PubMed: 15252187]
29. Mizuta S, Minami T, Fujita H, Kaminaga C, Matsui K, Ishino R, Fujita A, Oda K, Kawai A, Hasegawa N, Urahama N, Roeder RG, and Ito M. 2014. CCAR1/CoCoA pair-mediated recruitment of the Mediator defines a novel pathway for GATA1 function. Genes to Cells 19: 28–51. [PubMed: 24245781]
30. An G, Miner CA, Nixon JC, Kincade PW, Bryant J, Tucker PW, and Webb CF. 2010. Loss of Bright/ARID3a Function Promotes Developmental Plasticity. Stem Cells 28: 1560–1567. [PubMed: 20680960]
31. Ratliff ML, Mishra M, Frank MB, Guthridge JM, and Webb CF. 2016. The Transcription Factor ARID3a Is Important for In Vitro Differentiation of Human Hematopoietic Progenitors. J Immunol 196: 614–623. [PubMed: 26685208]

32. Martin M 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 17: 3.
33. Langmead B, and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9: 357–359. [PubMed: 22388286]
34. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Giardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, and Kent WJ. 2011. The UCSC Genome Browser database: update 2011. Nucleic Acids Res 39: D876–882. [PubMed: 20959295]
35. Li B, and Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323. [PubMed: 21816040]
36. Love MI, Huber W, and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 550. [PubMed: 25516281]
37. Rajaiya J, Nixon JC, Ayers N, Desgranges ZP, Roy AL, and Webb CF. 2006. Induction of Immunoglobulin Heavy-Chain Transcription through the Transcription Factor Bright Requires TFII-I. Molecular and cellular biology 26: 4758–4768. [PubMed: 16738337]
38. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, and Liu XS. 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9: R137. [PubMed: 18798982]
39. Rutherford TR, Clegg JB, and Weatherall DJ. 1979. K562 human leukaemic cells synthesise embryonic haemoglobin in response to haemin. Nature 280: 164–165. [PubMed: 95354]
40. Smith RD, Malley JD, and Schechter AN. 2000. Quantitative analysis of globin gene induction in single human erythroleukemic cells. Nucleic acids research 28: 4998–5004. [PubMed: 11121491]
41. Webb C, Zong RT, Lin D, Wang Z, Kaplan M, Paulin Y, Smith E, Probst L, Bryant J, Goldstein A, Scheuermann R, and Tucker P. 1999. Differential regulation of immunoglobulin gene transcription via nuclear matrix-associated regions. Cold Spring Harb Symp Quant Biol 64: 109–118. [PubMed: 11232275]
42. Lin D, Ippolito GC, Zong RT, Bryant J, Koslovsky J, and Tucker P. 2007. Bright/ARID3A contributes to chromatin accessibility of the immunoglobulin heavy chain enhancer. Mol Cancer 6: 23. [PubMed: 17386101]
43. Strouboulis J, Dillon N, and Grosveld F. 1992. Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. Genes & Development 6: 1857–1864. [PubMed: 1383089]
44. Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, Hitz BC, Gabdank I, Narayanan AK, Ho M, Lee BT, Rowe LD, Dreszer TR, Roe G, Podduturi NR, Tanaka F, Hong EL, and Cherry JM. 2016. ENCODE data at the ENCODE portal. Nucleic acids research 44: D726–D732. [PubMed: 26527727]
45. Lessard S, Beaudoin M, Benkirane K, and Lettre G. 2015. Comparison of DNA methylation profiles in human fetal and adult red blood cell progenitors. Genome Med 7: 1. [PubMed: 25606059]
46. Bartholdy B, Lajugie J, Yan Z, Zhang S, Mukhopadhyay R, Greally JM, Suzuki M, and Bouhassira EE. 2018. Mechanisms of establishment and functional significance of DNA demethylation during erythroid differentiation. Blood Adv 2: 1833–1852. [PubMed: 30061308]
47. Weiss MJ, Keller G, and Orkin SH. 1994. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. Genes Dev 8: 1184–1197. [PubMed: 7926723]
48. Fujiwara Y, Browne CP, Cunniff K, Goff SC, and Orkin SH. 1996. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. Proc Natl Acad Sci U S A 93: 12355–12358. [PubMed: 8901585]
49. Pevny L, Simon MC, Robertson E, Klein WH, Tsai SF, D'Agati V, Orkin SH, and Costantini F. 1991. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature 349: 257–260. [PubMed: 1987478]

50. Tsai FY, and Orkin SH. 1997. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. Blood 89: 3636–3643. [PubMed: 9160668]

51. Pevny L, Lin CS, D'Agati V, Simon MC, Orkin SH, and Costantini F. 1995. Development of hematopoietic cells lacking transcription factor GATA-1. Development 121: 163–172. [PubMed: 7867497]

52. Leonard M, Brice M, Engel JD, and Papayannopoulou T. 1993. Dynamics of GATA transcription factor expression during erythroid differentiation. Blood 82: 1071–1079. [PubMed: 8353273]

53. Bresnick EH, Martowicz ML, Pal S, and Johnson KD. 2005. Developmental control via GATA factor interplay at chromatin domains. J Cell Physiol 205: 1–9. [PubMed: 15887235]

54. Bresnick EH, Lee HY, Fujiwara T, Johnson KD, and Keles S. 2010. GATA switches as developmental drivers. J Biol Chem 285: 31087–31093. [PubMed: 20670937]

55. Xu J, Shao Z, Li D, Xie H, Kim W, Huang J, Taylor JE, Pinello L, Glass K, Jaffe JD, Yuan GC, and Orkin SH. 2015. Developmental control of polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. Mol Cell 57: 304–316. [PubMed: 25578878]

56. Renneville A, Van Galen P, Canver MC, McConkey M, Krill-Burger JM, Dorfman DM, Holson EB, Bernstein BE, Orkin SH, Bauer DE, and Ebert BL. 2015. EHMT1 and EHMT2 inhibition induces fetal hemoglobin expression. Blood 126: 1930–1939. [PubMed: 26320100]

57. Wilsker D, Patsialou A, Dallas PB, and Moran E. 2002. ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. Cell Growth Differ. 13: 95–106. [PubMed: 11959810]

58. Ward JM, Rose K, Montgomery C, Adrianto I, James JA, Merrill JT, and Webb CF. 2014. Disease activity in systemic lupus erythematosus correlates with expression of the transcription factor AT-rich-interactive domain 3A. Arthritis Rheumatol. 66: 3404–3412. [PubMed: 25185498]

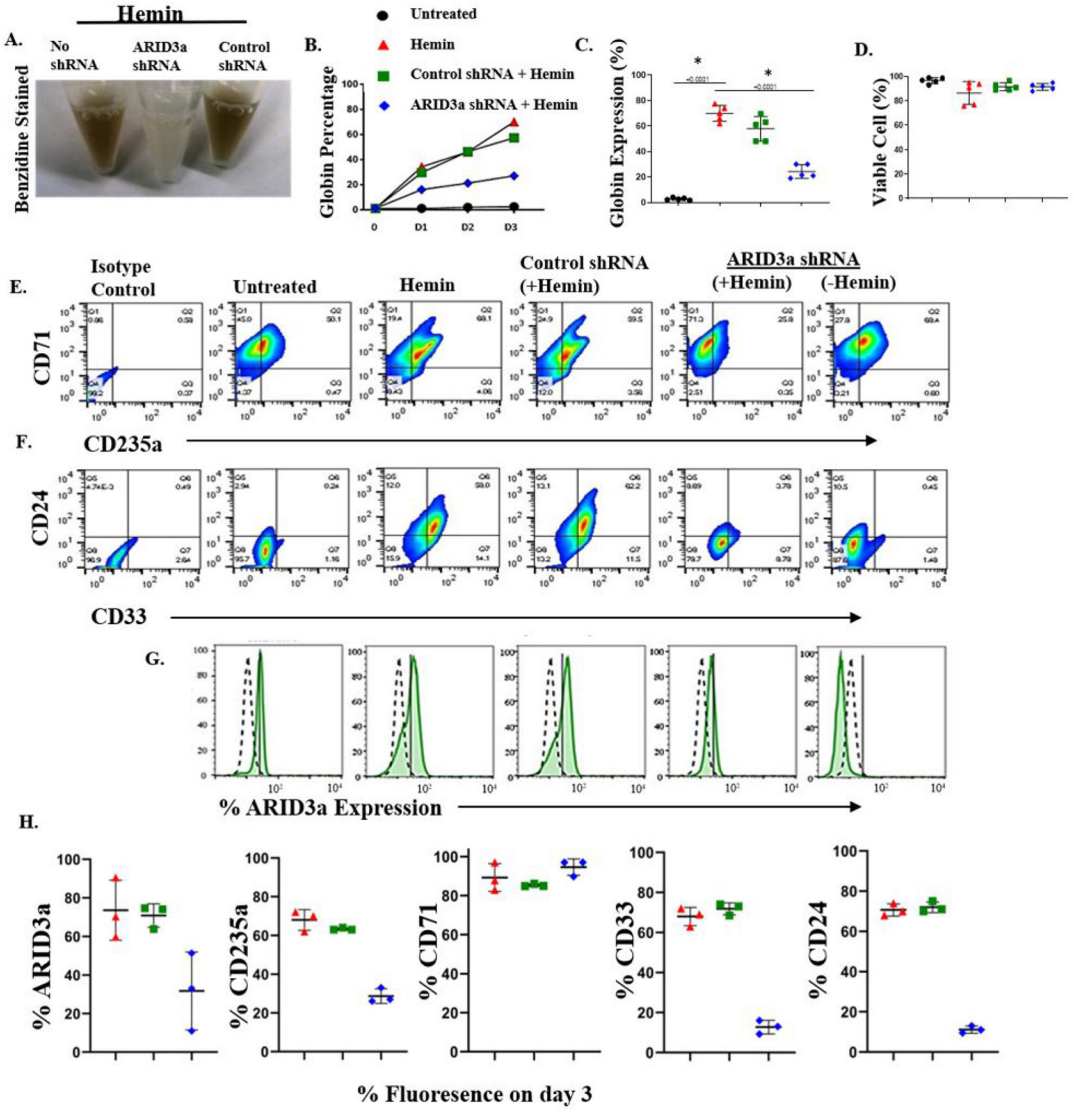


Figure 1. ARID3a is required for hemin-induced fetal globin production and erythroid maturation.

K562 cells were treated with hemin with and without prior transduction of cells with lentivirus expressing ARID3a-specific shRNA or unrelated shRNA control virus. (A) Hemin-stimulated K562 were stained with benzidine to visualize dark-colored globin production. (B) A representative time course experiment (N=3) shows percentages of globin producing cells counted microscopically. (C) Cumulative data from 5 individual experiments show percentages of benzidine positive cells on day three of culture ($P< 0.0001$, one-way ANOVA). (D) Percentages of viable K562 cells counted via trypan blue exclusion on day three are shown. Flow cytometry of cells stimulated as indicated on day 3 of culture shows surface staining of proteins associated with erythrocyte (E) and monocyte (F) lineage differentiation. (G) Flow cytometric histograms presented in normalized mode indicate numbers of cells expressing intracellular ARID3a (shaded peaks) compared to isotype controls (dotted lines). Solid vertical lines depict peak intensities of control unstimulated

cells for comparison. (H) Percentages of ARID3a and the surface markers shown in (E) and (F) are presented for 3 experiments.

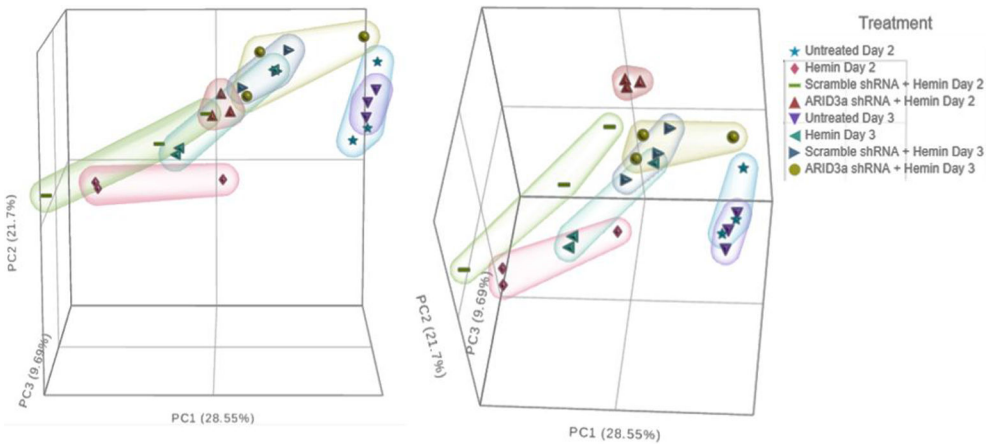


Figure 2. ARID3a inhibition alters transcription profiles of hemin-induced K562 cells.

A principal component analysis of K562 cells treated with hemin for 2 or 3 days, with and without ARID3a inhibition is shown from two different 3-dimensional views. Individual dots represent triplicate cultures sequenced on days 2 and 3. Cell treatments and days are labeled.

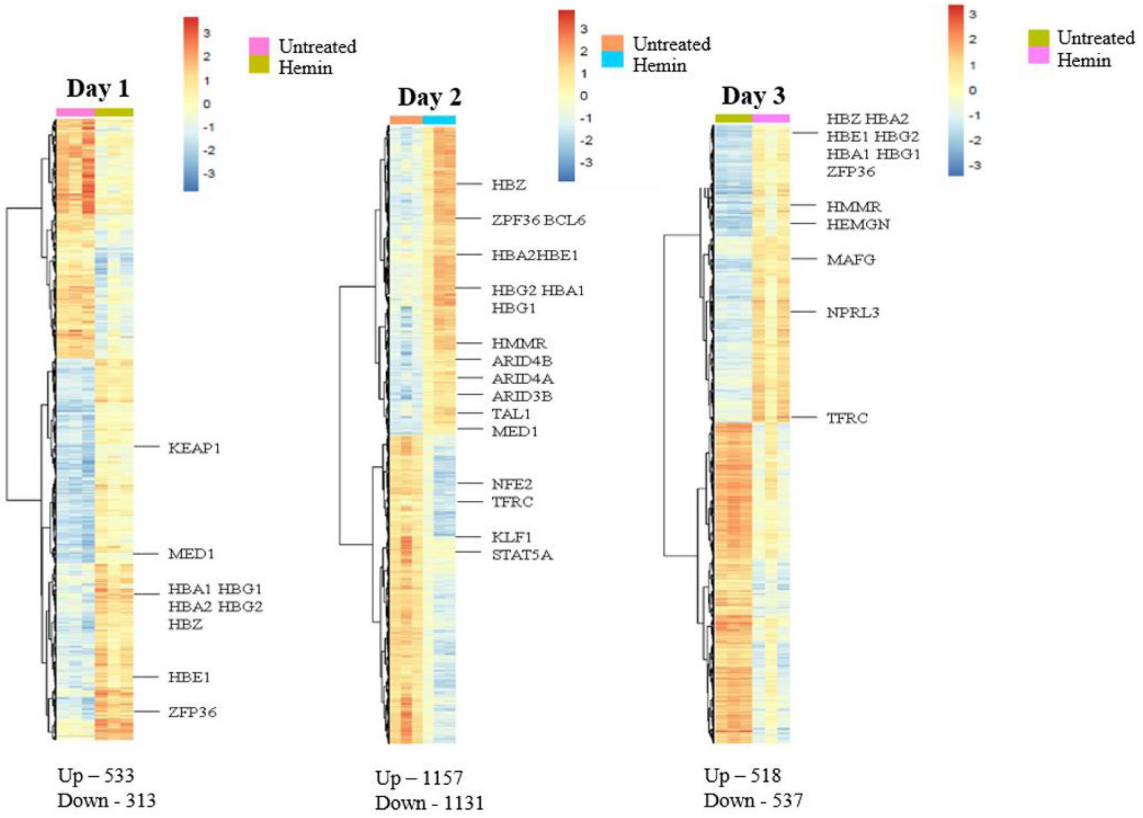


Figure 3. Hemin induces differential gene expression of globin-associated genes within three days.

Heatmaps of differentially expressed genes from triplicate cultures of untreated vs hemin-treated cells are shown at three time points (days 1-3) after hemin stimulation. Numbers of differentially expressed genes (FDR < 0.05, FC ≥ ±1.5, n = 3) are indicated below the heatmaps. Select genes previously associated with erythrocyte differentiation are indicated.

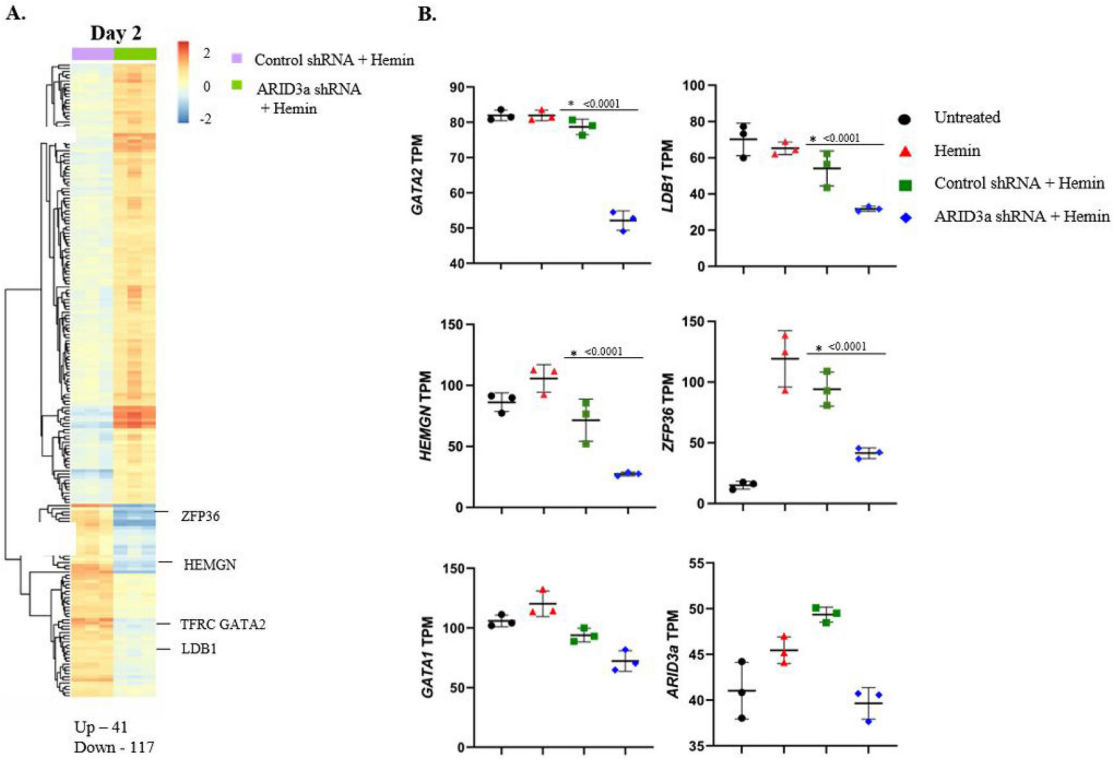


Figure 4. ARID3a inhibition alters hemin-induced gene expression.

(A) A hierarchical clustered heatmap of differentially expressed genes in hemin treated cells with and without ARID3a inhibition from triplicate cultures (each column) are shown at day 2. Total numbers of differentially expressed genes (FDR < 0.05, FC ≥ ±1.5, n = 3) are given at the panel bottoms and select genes are indicated. (B) Transcripts per million (TPM) values of key erythroid genes quantify the effects of both hemin stimulation and ARID3a inhibition. FDR values are displayed above conditions to indicate statistical significance determined by DESeq2.

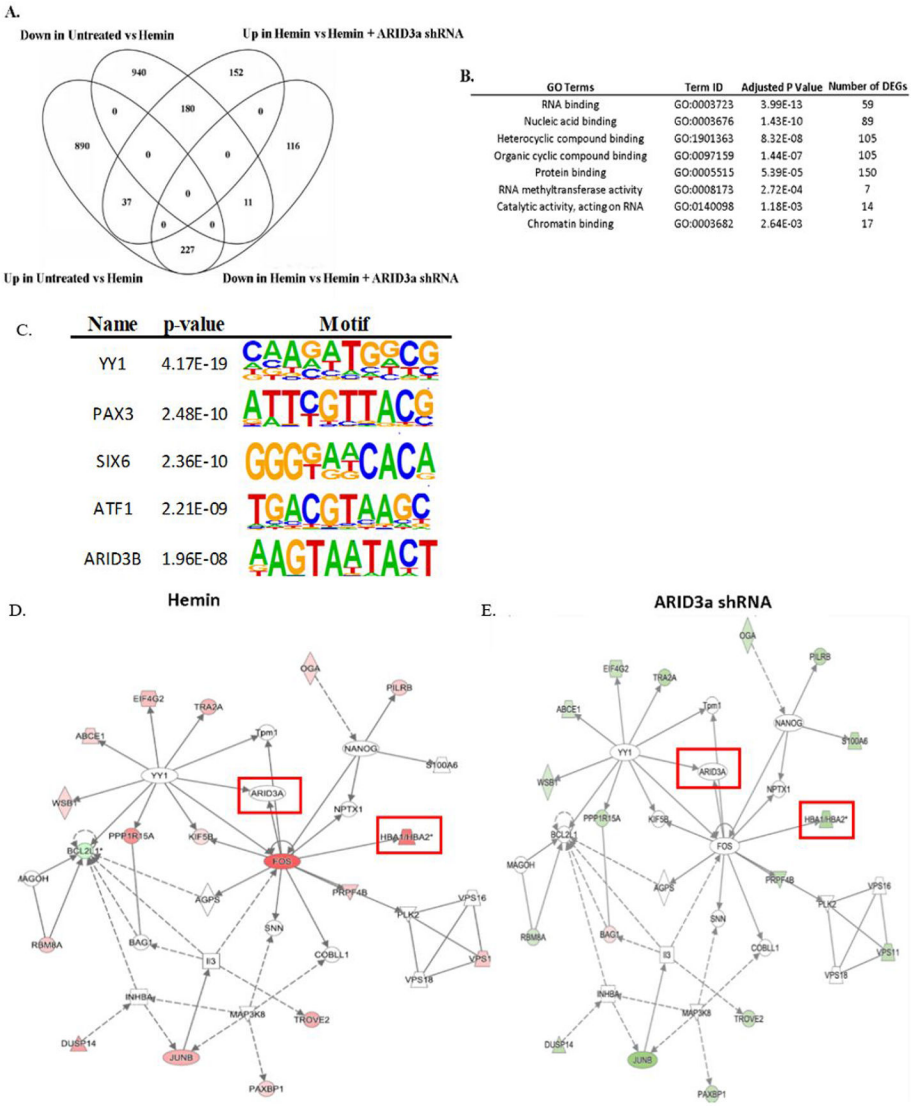


Figure 5. ARID3a is required for expression of a number of genes and pathways important for erythropoiesis.

(A) A Venn diagram indicates numbers of differentially expressed genes on day two of treatment with and without hemin and ARID3a (FDR < 0.05, FC ≥ 1.5, n=3). (B) GO analyses indicates pathways important for the 227 genes that are affected by ARID3a inhibition and hemin induction. (C) The most highly represented transcription factor binding motifs of the 227 overlapping genes are shown. (D, E) Network analyses of the 227 differentially expressed genes in (A) reveal related genes. Log2 fold-change values from the untreated vs hemin comparison were overlaid onto the top network identified by IPA using the 227 overlapping genes. Red color indicates upregulated genes and green color indicates downregulated genes

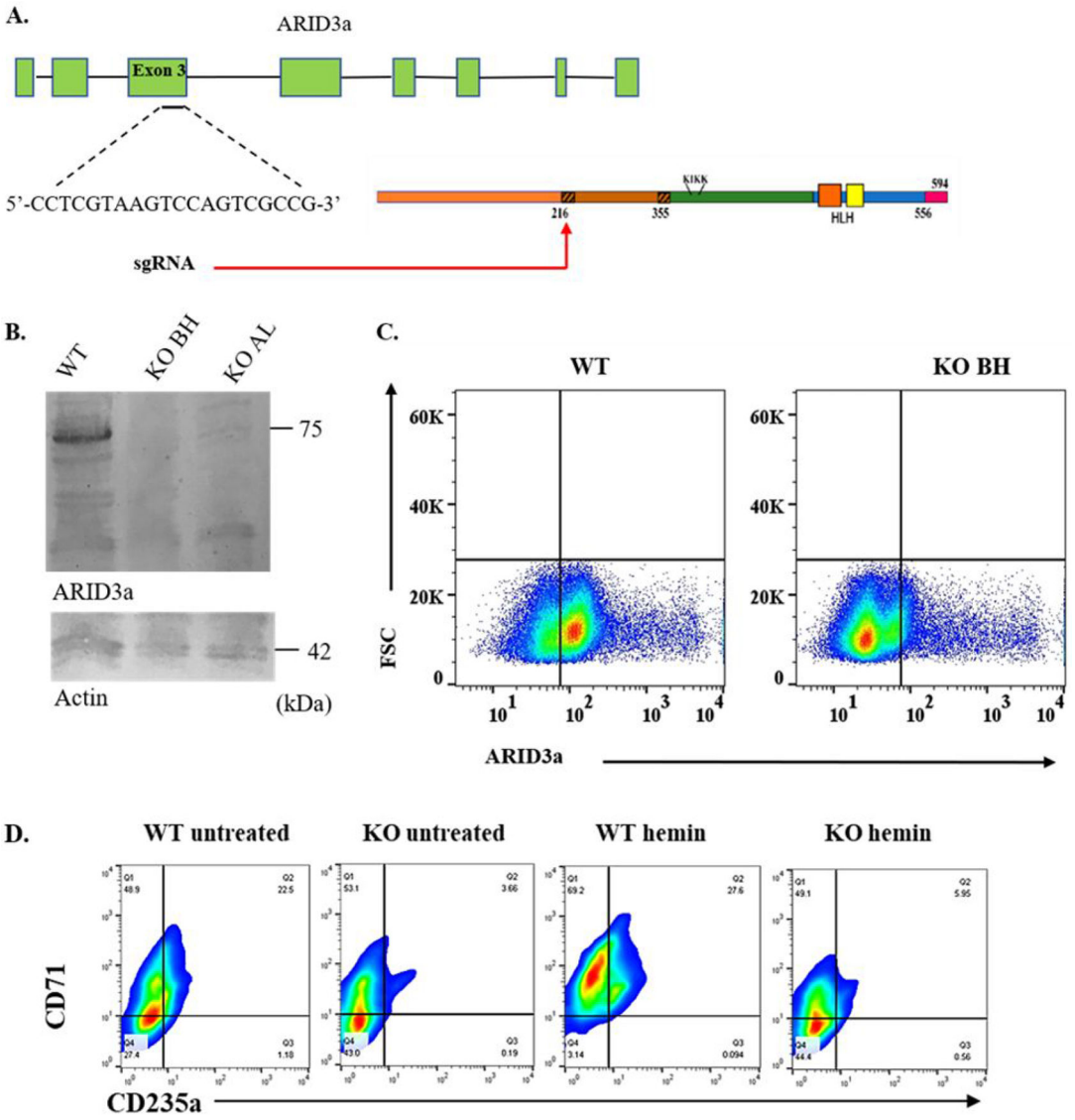
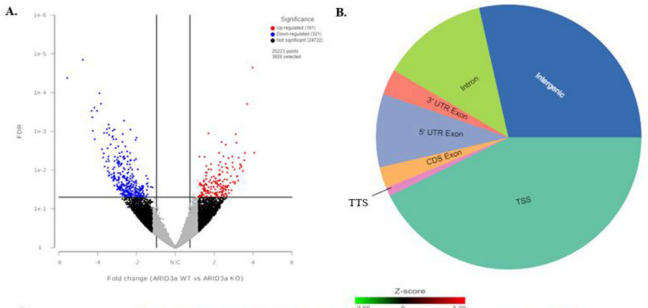


Figure 6. Homozygous ARID3a knockout clones K562 were generated.

(A). Schematic diagram of the sgRNA cut-site that causes deletion of part of the extended-ARID DNA-binding domain used for CRISPR-Cas9 deletion. (B) Single cell sorted K562 clones were analyzed for protein expression of ARID3a by western blotting of a wild type (WT) and two KO clones, BH and AL using 100,000 cells per lane. Actin was used to confirm protein loading in each lane (lower panel). (C) Flow cytometry for intracellular ARID3a confirmed knockout of ARID3a, as shown for clone BH versus the wild type (WT) clone. (D) Representative flow cytometry data from hemin stimulated and untreated clones were used to evaluate erythroid lineage markers in KO clone BH vs the WT clone. Quadrant gates were set according to isotype controls



E.

ARID3a WT vs ARID3a KO			
Region	Gene	FDR	Log_2 fold change
chr6:66948656-66949661		8.90E-05	3.97
chr8:33722158-33723339		8.16E-07	3.79
chr6:49336692-49337120		6.92E-06	3.71
chr13:87124107-87124797		4.55E-04	3.64
chr13:86882294-86883175		8.16E-03	3.38
chr2:57776607-57777370		2.48E-04	3.30
chr4:92181874-92183520		8.16E-03	3.26
chr13:80359555-80360601		6.92E-06	3.02
chr1:243591899-243592874	AKT3	3.69E-05	3.01
chr2:21245411-21245831		2.67E-04	2.94
chr1:240398746-240399716	FMN2	1.27E-03	-2.80
chr2:152054752-152055885	CACNB4	4.55E-04	-2.85
chr2:108824491-108825504	CCDC138	4.55E-04	-2.96
chr4:152264402-152265639	.	5.73E-04	-3.00
chr5:149060307-149061326	SH3TC2	4.55E-04	-3.31
chr8:24222397-24223110	.	2.57E-05	-3.40
chr10:85901808-85902686	GRID1	9.29E-05	-3.43
chr9:83271681-83272154	FRMD3	5.51E-05	-3.66
chr6:113426255-113427950	.	9.16E-03	-3.68
chr4:89299164-89299989	GPRIN3	1.30E-05	-3.88

F.

Name	p-value	Motif
NFATC1	2.36E-10	
ZIC2	2.19E-09	
GATA3	2.01E-09	

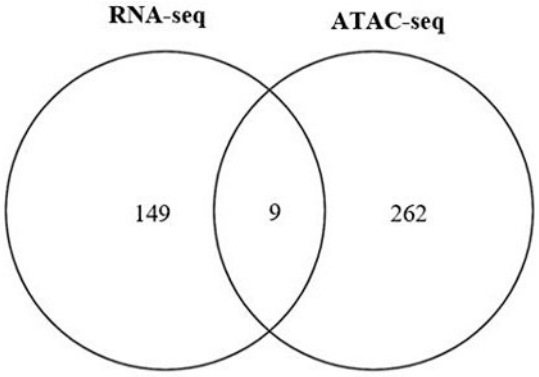
G.

Name	p-value	Motif
ARNT:AHR	2.03E-08	
MTF1	1.97E-08	
TEAD	1.95E-08	

Figure 7. ARID3a is required to maintain chromatin landscapes.

Two homozygous ARID3a KO and WT clones with and without hemin treatment were subjected to ATAC-seq analyses in duplicate cultures. (A) A volcano plot shows differentially accessible regions (DARs) between wild type and ARID3a KO clones (FDR < 0.05, log FC > 2). Dark red dots represent individual genes with FDR < 0.05 (y-axis) and log FC > 2 (x-axis). (B) The genomic distribution of differentially expressed ATAC peaks in WT versus ARID3a KO clones is shown. The human genome was portioned into seven bins relative to RefSeq genes. TSS, transcriptional start site; TTS, transcriptional termination site; CDS, coding sequence; UTR, untranslated region. (C) Unsupervised hierarchical clustering of data from duplicate KO and WT clones with and without hemin generate a heatmap of DARs that reveal both increased and decreased accessible regions. (D) ATAC peaks for the two most differentially accessible regions (which are intragenic) are shown (E) The top 10 upregulated and downregulated DARs are shown. (F) Homer analysis shows enrichment of potential transcription factor binding sites within DARs with increased or decreased accessibility in WT versus KO clones.

A.



B.

Gene
HSP90AB1
HSPA1A
HSPA1B
PCBP1-AS1
PHGDH
PPP1R15A
RBM33
SEL1L
VPS37B

C.

GO Terms	Term ID	Adj.P Value
Regulation of TNF-mediated signaling	GO:1903265	1.03E-05
Protein folding	GO:0042026	2.76E-05
Regulation of erythrocyte differentiation	GO:0045648	6.92E-05
Protein ubiquitination	GO:0031396	1.52E-04

Figure 8. Overlap analysis of DEGs from RNA-seq and DARs from ATAC-seq data.

(A) A Venn diagram indicates the number of common genes when comparing the 158 DEGs from Figure 4 with DARs identified as being within 1000 bp of the transcription start site of a known gene. (B) The 9 genes identified as significant in both RNA- and ATAC-seq data are listed. (C) Pathways identified from GO analyses with those 9 genes are shown.

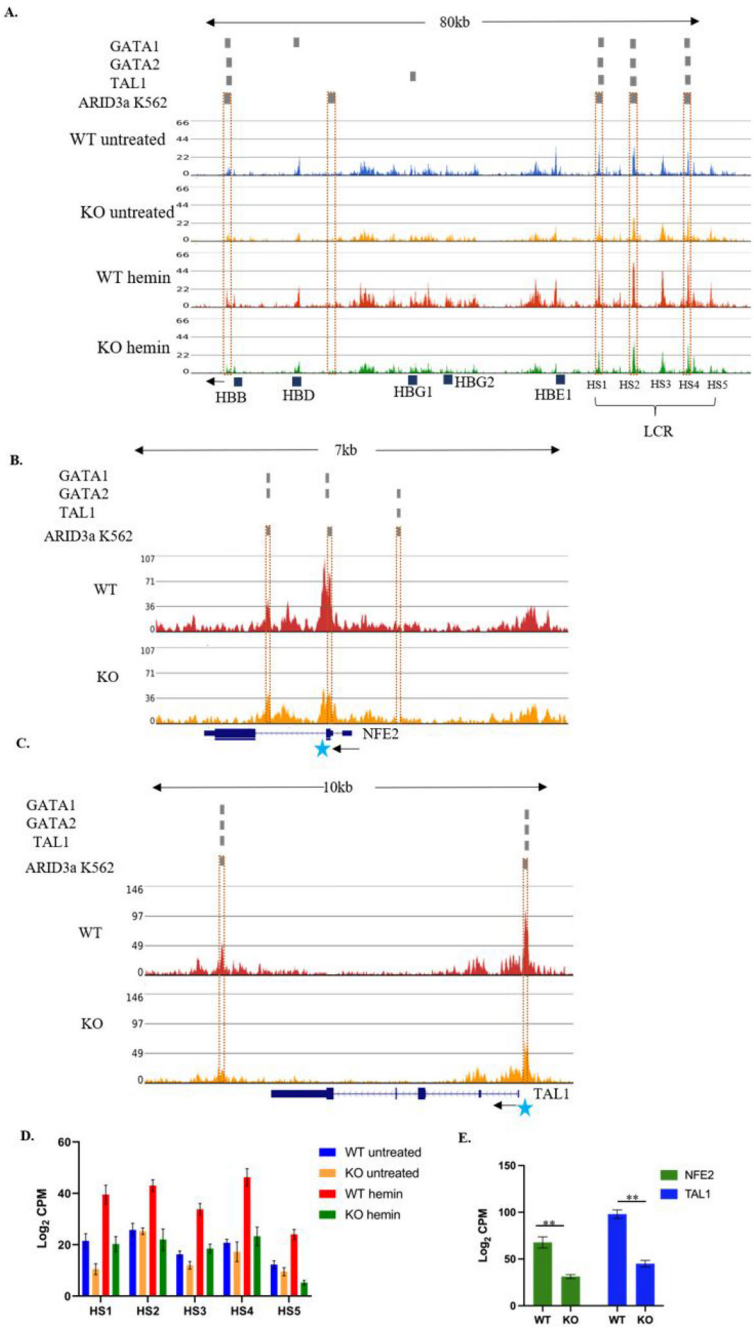


Figure 9. ARID3a inhibition alters chromatin accessibility of regulatory regions important for erythropoiesis.

(A). Chromatin accessibility peaks of the β-globin locus and the LCR in WT and ARID3aKO clones stimulated with and without hemin are shown (2 clones per group, were analyzed in duplicate, FDR < 0.05, log FC > 2). Gray bars indicate transcription factor binding sites for ARID3a, GATA1, GATA2, and TAL1 identified by ENCODE in K562 cells. Dotted rectangles show ARID3a binding sites in all 4 conditions. Chromatin accessibility peaks for two other transcription factors required for erythropoiesis NFE2 (B) and TAL1 (C) are shown for WT and ARID3a KO clones. Log2 counts per million (CPM)

values were quantified for effects due to ARID3a depletion in the LCR region (D), and in sites near the NFE2 and TAL1 genes indicated in (B) and (C) by blue stars (E) stars (E) FDR < 0.005, **.

Table 1.

GO analysis of DEGs induced by hemin

GO Terms	Term ID	Adjusted P Value	Number of DEGs
Erythrocyte differentiation	GO:0030218	1.69E-06	86
Erythrocyte homeostatsis	GO:0034101	5.53E-06	95
Regulation of transcription, DNA-templated	GO:0006355	2.96E-07	45
mRNA splicing	GO:0000398	4.21E-11	41
Nuclear export	GO:0051168	1.32E-06	121

Table 2.

Most significantly differentially expressed genes on Day 2

Untreated Vs Hemin			Hemin Vs Hemin + ARID3a shRNA		
Gene	**Log_2 Fold Change**	**Adj P Value**	**Gene**	**Log_2 Fold Change**	**Adj P Value**
TXNIP	5.24	6.25E-15	*SH3BGR*	2.29	4.02E-07
OSGIN1	3.68	3.63E-16	*HSPA5*	2.05	7.51E-10
HBZ	3.52	6.64E-16	*HERPUD1*	1.83	4.88E-11
AKR1C1	3.01	1.89E-14	*ARG2*	1.75	2.36E-09
SQSTM1	2.99	1.22E-24	*ALDH1A1*	1.61	7.07E-07
MCM5	2.74	1.07E-18	*NFE4*	1.43	3.92E-12
HBA2	2.54	2.68E-15	*SEC24D*	1.40	3.58E-07
HBE1	2.47	3.58E-21	*BEX2*	1.29	4.11E-06
NQO1	2.41	2.22E-32	*CREG1*	1.28	2.85E-06
HBG2	2.12	2.33E-15	*LCP1*	1.22	4.33E-07
GCLM	2.08	5.83E-19	*SERPINH1*	1.20	5.91E-06
PPP1R15A	2.05	2.15E-13	*RTN4*	1.16	3.14E-08
FTL	2.02	3.63E-16	*ACOT13*	1.15	6.60E-07
HBA1	1.98	3.73E-13	*TDP2*	1.11	1.49E-07
HBG1	1.94	1.02E-11	*TPM4*	1.11	4.53E-08
TXNRD1	1.52	9.58E-12	*CTSD*	1.11	1.78E-08
FTH1	1.47	4.70E-12	*PGD*	1.05	3.26E-06
CREM	1.41	1.28E-14	*ACTB*	1.02	2.38E-06
TXN	1.20	6.53E-14	*CTSB*	1.02	1.39E-06
PRKCSH	−1.18	3.70E-16	*SND1*	0.96	8.32E-07
FKBP2	−1.44	1.47E-11	*COPA*	0.89	2.56E-07
UCA1	−1.52	6.71E-13	*GSR*	0.88	2.85E-06
SERPINH1	−1.60	1.47E-11	*ANXA5*	0.88	1.47E-10
PDIA3	−1.63	2.29E-19	*SEC61A1*	0.80	1.54E-06
DNAJB11	−1.69	4.11E-14	*CALU*	0.80	1.71E-09
AC068631.2	−1.95	9.98E-12	*SSX1*	0.77	1.23E-07
HSP90B1	−1.97	3.99E-19	*NQO2*	0.73	5.36E-06
HERPUD1	−1.99	2.64E-14	*PSMC1*	0.68	1.35E-06
NMU	−2.01	3.65E-11	*SEM1*	0.61	4.33E-07
HYOU1	−2.07	7.01E-19	*TNNI3*	−1.04	8.32E-07
PDIA6	−2.19	2.54E-17	*VARS*	−1.06	2.82E-07
MANF	−2.24	8.93E-16	*LINC01029*	−1.10	1.26E-06
PDIA4	−2.35	1.43E-14	*MARCKSL1*	−1.14	4.84E-07
HSPA5	−2.35	7.91E-14	*HEMGN*	−1.95	7.51E-10
CALR	−2.57	9.34E-17	*TXNIP*	−4.15	9.81E-11

Table 3.

GO analysis of differentially accessible regions identified by ATAC-seq

GO Terms	Term ID	Adj.P Value	Number of DEGs
Erythropoietin-mediated signaling	GO:0038162	7.75E-06	10
Thrombin signaling	GO:0015057	6.30E-05	26
B cell receptor signaling	GO:0050852	1.11E-04	21
Regulation of NFAT signaling	GO:0070884	3.34E-04	3
PI3K/AKT signaling	GO:0014065	2.74E-03	6