ARID4B is critical for mouse embryonic stem cell differentiation towards mesoderm and endoderm, linking epigenetics to pluripotency exit

Nihal Terzi Cizmecioglu1,*, Jialiang Huang2, Ezgi G. Keskin1, Xiaofeng Wang3, Idil Esen1, Fei Chen2, and Stuart H. Orkin1,3,*

From the 1Department of Biological Sciences, Faculty of Arts and Sciences, Middle East Technical University, Ankara, Turkey, the 2State Key Laboratory of Cellular Stress Biology, Innovation Center for Cell Signaling Network, School of Life Sciences, Xiamen University, Xiamen, Fujian, China, the 3Geisel School of Medicine, Dartmouth University, Hanover, New Hampshire, USA, and the 4Howard Hughes Medical Institute, Dana Farber/Boston Children’s Cancer and Blood Disorders Center, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA

Edited by Patrick Sung

Distinct cell types emerge from embryonic stem cells through a precise and coordinated execution of gene expression programs during lineage commitment. This is established by the action of lineage specific transcription factors along with chromatin complexes. Numerous studies have focused on epigenetic factors that affect embryonic stem cells (ESC) self-renewal and pluripotency. However, the contribution of chromatin to lineage decisions at the exit from pluripotency has not been as extensively studied. Using a pooled epigenetic shRNA screen strategy, we identified chromatin-related factors critical for differentiation toward mesodermal and endodermal lineages. Here we reveal a critical role for the chromatin protein, ARID4B. Arid4b-deficient mESCs are similar to WT mESCs in the expression of pluripotency factors and their self-renewal. However, ARID4B loss results in defects in up-regulation of the meso/endodermal gene expression program. It was previously shown that Arid4b resides in a complex with SIN3A and HDACS 1 and 2. We identified a physical and functional interaction of ARID4B with HDAC1 rather than HDAC2, suggesting functionally distinct Sin3a subcomplexes might regulate cell fate decisions. Finally, we observed that ARID4B deficiency leads to increased H3K27me3 and a reduced H3K27Ac level in key developmental gene loci, whereas a subset of genomic regions gain H3K27Ac marks. Our results demonstrate that epigenetic control through ARID4B plays a key role in the execution of lineage-specific gene expression programs at pluripotency exit.

During early embryonic development, a series of differentiation and cleavage events lead to the formation of distinct cell types that later form the organism. The emergence of various cell types is a complex process that requires a precisely timed mechanism for successful development. Embryonic stem cells (ESCs) provide an in vitro model for studying early cell fate decisions. ESCs self-renew limitless in vitro. Because they have the capacity to form all cell types (pluripotency), they can be directed to desired lineages under the guidance of specific cytokines.

Cell fate decisions are executed by changes in gene expression. Whereas the gene expression program of a particular lineage is being established, unrelated programs are simultaneously extinguished. The chromatin environment plays a critical role in regulating the timing and the level of gene expression. The ESC-specific gene expression program is stabilized by the interactions of core pluripotency transcription factors and chromatin complexes (1–3). The plasticity of ESC differentiation potential is reflected in an open chromatin structure. Progressively during differentiation, ESCs undergo reorganization of chromatin, architecture and genomic topology (4–9). Alterations in the chromatin environment of ESCs, therefore, may impact lineage commitment dynamics.

Studies have identified chromatin factors regulating the ESC self-renewal and pluripotency (10–18). It is becoming increasingly clear that the chromatin architecture and histone modifications at the ESC stage can affect cell fate specification and differentiation kinetics at later stages (17, 19). However, a comprehensive study of the epigenetic regulators subsequent to the loss of self-renewal and pluripotency has been lacking. Therefore, we sought to determine the role of chromatin factors in an unbiased manner during meso/endodermal lineage commitment. To accomplish this goal, we have monitored the expression of the first lineage-specific master transcription factors to enable a more precise look at the chromatin-related requirements at cell fate decisions. Our approach departs from previous reports focusing on epigenetic effects on ESC characteristics (10, 12).

Results

Functional RNAi screen identifies candidate chromatin factors required for endoderm and mesoderm commitment

We used a pooled shRNA library screen to identify epigenetic factors that impact mouse embryonic stem cell differentiation toward mesoderm and endoderm (Fig. 1a). The shRNA library consisted of 5 previously validated shRNAs per gene targeting ~300 chromatin-related proteins. A Brachyury-GFP; Foxa2-hCD4 reporter mESC line was transduced at low multiplicity of infection with the pooled shRNA library, enabling single shRNA knockdown per cell. After puromycin selection of

This article contains supporting information.

*For correspondence: Stuart H. Orkin, Stuart_Orkin@dfci.harvard.edu; Nihal Terzi Cizmecioglu, nterzi@metu.edu.tr.

This is an Open Access article under the CC BY license.
transduced mESCs, the starting (day 0) population of shRNAs was determined by DNA sequencing. Thereafter, the reporter line was directed toward mesoderm or endoderm. On day 5 of differentiation, shRNAs in the top 5% of differentiated cells (for mesoderm: highest BRACHYURY expression, for endoderm: highest BRACHYURY and FOXA2 expression, lowest SSEA1 expression) as well as the bottom 5% of undifferentiated cells (lowest BRACHYURY and/or FOXA2 expression and highest SSEA1 expression) were determined by cell sorting and DNA sequencing. The analysis was performed by comparing the enrichment of shRNAs in day 5 differentiated cells to day 5 undifferentiated cells or day 0 starting population (Fig. 1, a and b). We found that the loss of chromatin factors more frequently led to the differentiated phenotype with variable strength (Fig. 1b). However, this was to be expected for the screening design as the differentiation efficiency of WT cells was ∼70% as calculated by BRACHYURY-positive cell population on day 5 (Fig. 1c). Consistent with published data (20–26), depletion of several members of PcG and TrX complexes affected mESC differentiation (Fig. 1b). shRNAs that were depleted at least 2-fold in differentiated cell pools versus undifferentiated cell pools were selected as potential candidates and further validated by single shRNA knockdown experiments (Fig. 1c). Observed differentiation defects were similar for mesoderm and endoderm lineages (data not shown). This observation suggests that under the conditions of this screen the candidate chromatin factors might impact a common mesendodermal cell population that gives rise to both lineages.

Figure 1. ARID4B loss leads to meso/endodermal differentiation defects. a, design of the shRNA screen. b, waterfall plot of shRNAs ranked by log2 of the enrichment score in differentiated over undifferentiated cells. Negative controls are in red (not visible since their enrichment score is close to zero) and positive controls (PcG complex members) are in black. c, endoderm differentiation efficiency is plotted as % BRACHYURY-positive cells on day 5 of differentiation. Negative control: nontargeting shRNA. d, flow cytometry data for endoderm differentiation. Negative control; nontargeting shRNA. e–l, RT-qPCR of selected transcripts during endoderm differentiation time course in WT, arid4bΔ, or arid4bΔ cells that re-express human ARID4B.
The role of ARID4B in mESC lineage commitment

ARID4B is essential for successful mESC differentiation toward endoderm and mesoderm

The ARID family protein ARID4B was chosen for in-depth study as its knockdown led to compromised mesoderm and endoderm differentiation. ARID family proteins exhibit DNA-binding activity with little or no sequence-specificity and display diverse functions in development and disease progression (27, 28). Arid4b and related Arid4a proteins contain a Tudor domain and a chromobase domain that recognizes methylated histones (29). In the adult tissues Arid4b expression is restricted to testis and important for spermatogonial development (30–32). Reactivation of expression has been reported in cancer (33–37). Deficiency of ARID4A and ARID4B results in a decrease in repressive chromatin modifications in the Prader-Willi/Angelman imprinting cluster (38). In our experiments knockdown of Arid4b with two independent shRNAs severely compromised differentiation of reporter mESCs toward mesodermal or endodermal lineages (Fig. 1, c and d), and SSEA1 remained high (Fig. S1, a and b), even with a modest decrease in the Arid4b level (Fig. S1c).

ARID4B is reported to be a component of the Sin3a corepressor complex (39, 40). Through its several protein interaction domains, SIN3A serves as a scaffold for histone deacetylases HDAC1/2 and several other proteins that regulate HDAC function and activity (41). Although the Sin3a complex was originally classified as a transcriptional repressor, more recent evidence suggests a role in transcriptional activation (42–44). In addition to Arid4b, knockdown of other members of the Sin3a complex, including Phf12, Mbd4, and Phf21a, lead to defects in commitment of mESCs to mesoderm and endoderm (Fig. 1c, Fig. S1, a and b).

To validate shRNA knockdown findings, we deleted the Arid4b gene in mESCs with CRISPR/Cas9 (Fig. S1d). Arid4b deleted mESCs expressed Oct4 and Nanog at similar levels to WT mESCs (Fig. 1, e and f, Fig. S1, e and f). Oct4 and Nanog expression was suppressed with similar kinetics in WT cells during endoderm or mesoderm commitment. Moreover, Arid4b-deleted mESCs failed to express Brachyury, Foxa2, or Sox17 during endoderm differentiation (Fig. 1, g–i) or mesoderm differentiation (Fig. S1g). Upon extension of endoderm differentiation from 5 to 8 days, we observed markedly reduced expression of Brachyury and Foxa2 in Arid4b-deleted cells (Fig. S1, h and i). Importantly, expression of human ARID4B in Arid4b-deleted mESCs rescued endoderm differentiation defect (Fig. 1, j–l, Fig. S1j). Due to this differentiation defect, we refer to arid4bΔ cells that are exposed to the same differentiation protocol as WT cells as “meso/endoderm directed” rather than “arid4bΔ meso/endoderm cells.”

Hdac1 and Hdac2 exert different roles in lineage commitment

Given the reported presence of HDAC1 and HDAC2 (45) in Arid4b/Sin3a corepressor complexes, we tested whether the differentiation defect upon ARID4B loss is phenocopied by loss of histone deacetylase activity. First, we used a Class I HDAC inhibitor, Merck 60, which is selective toward HDAC1 and HDAC2 with IC50 of 1 and 8 nm, respectively. Histone deacetylation has key functions in maintaining a balance between self-renewal and differentiation (46–51). To prevent confounding effects of Merck 60 treatment at the ESC stage, we limited its use only to the differentiation phase. We assessed endoderm/mesoderm differentiation efficiency upon inhibitor treatment in the Brachyury-GFP, Foxa2-hCD4 reporter mESCs. Increasing concentrations of Merck 60 treatment was associated with elevated histone 3 acetylation (Fig. S2a). BRACHYURY and FOXA2 expression was reduced upon Merck 60 treatment during endoderm differentiation (Fig. S2b). However, SSEA1 levels were unchanged in DMSO or Merck 60-treated cells. Similar results were obtained for Merck 60 treatment during mesoderm differentiation (Fig. S2c).

To resolve ambiguities from inhibitor treatment, we generated independent CRISPR/Cas9-mediated Hdac1 or Hdac2 deletions in mESCs (Fig. 2, a and b). Similar to Arid4b-deleted cells, Hdac1-deleted mESCs fail to express Brachyury, Foxa2, or Sox17 during endoderm differentiation, whereas Hdac2 deletion had no evident effect (Fig. 2, c–e). Mesoderm differentiation was also defective in hdac1Δ cells (Fig. 2f). On the other hand, Nanog suppression during differentiation followed with similar kinetics in WT, hdac1Δ, and hdac2Δ cells (Fig. S2, d–f). These results are consistent with a critical role of HDAC1, but not HDAC2, in early embryogenesis (52, 53). In essence, HDAC1 loss phenocopies aspects of ARID4B deficiency.

We next asked whether the loss of ARID4B or HDAC1 affected neuroectodermal lineage commitment. In contrast to mesoderm or endoderm differentiation, the loss of ARID4B or HDAC1 failed to affect commitment toward neuroectodermal lineage, as evidenced by the expression of Sox1, Pax6, or Jag1 marker genes (Fig. 2, g and h, Fig. S2g). We conclude that the function of ARID4B is essential for meso/endodermal commitment and dispensable for neuroectodermal lineage.

Although both HDAC1 and HDAC2 are present in the Sin3a complex, it is interesting that only Hdac1 deletion phenocopies Arid4b deletion. We tested whether this result might be due to a preferential physical interaction between ARID4B and HDAC1. We performed coimmunoprecipitation using Arid4b complex, it is interesting that only Arid4b successfully immunoprecipitated SIN3A. ARID4B and SIN3A peaks were unchanged in DMSO or Merck 60-treated cells. Similar results were obtained for Merck 60 treatment during mesoderm differentiation (Fig. S2c).

Next, we performed proximity ligation assay (PLA) to detect intact complex composition. ARID4B and SIN3A peaks coincided in high molecular weight fractions. We observed a preferential physical interaction between ARID4B and HDAC1 or HDAC2. This technique utilizes a pair of oligonucleotide-bound antibodies to enable continuous DNA synthesis only if epitopes are in close proximity (40 nm) and is used for intracellular visualization of protein–protein interactions. Consistent with previous results, we observed more ARID4B-HDAC1 interactions than ARID4B-HDAC2 interactions in mESCs (Fig. 2k). The majority of interactions colocalized with 4,6-diamidino-2-phenylindole, consistent of the subcellular localization and function of these interactions.
proteins. A greater number of ARID4B-HDAC1 interactions were not because of differences in abundance, because HDAC1 and HDAC2 were expressed at similar levels in mESCs (Fig. S2h). These results suggest that the observed mesodermal and endodermal differentiation defect of ARID4B deficiency is associated with loss of HDAC1 activity in Sin3a complex.

**arid4bΔ and hdac1Δ cells exhibit similar global histone modification profile**

Next, we investigated the global chromatin profile of endoderm committed WT, arid4bΔ, hdac1Δ, and hdac2Δ cells. To this end, we performed a quantitative analysis of histone post-translational modifications by MS, which allowed for an unbiased examination of histone modifications, as well as their combinational constitution in each cell type. The results were normalized to WT and clustered using the Euclidean distance metric (Fig. 2l). WT cells clustered away from endoderm-differentiated cells, regardless of the genotype (Fig. S2i). These observations are consistent with the similarities in phenotype of ARID4B and HDAC1 loss.

Given the differentiation defect of arid4bΔ cells, it is possible that the arid4bΔ cells maintain an ESC stage histone modification profile. To test this, we compared the global histone modification profile of WT ESCs to those of endoderm-differentiated WT, arid4bΔ, hdac1Δ, and hdac2Δ cells. Interestingly, WT ESCs clustered away from endoderm-differentiated cells, regardless of the genotype (Fig. S2j). These observations support a model in which arid4bΔ cells do not remain as ESCs...
during differentiation but are unable to successfully execute commitment to endoderm or mesoderm lineages.

**arid4bΔ cells fail to up-regulate meso/endodermal gene expression program**

To further investigate the role of ARID4B in mESC lineage commitment, we conducted RNA expression profiling of WT or arid4bΔ cells directed toward mesoderm or endoderm. Hierarchical clustering of the samples showed that RNA-seq retained high reproducibility in replicates (Fig. S3a). Compared with WT cells, arid4bΔ cells showed a reduction in the expression of primitive streak and endodermal markers (Fig. 3a). Comparative analysis of transcriptomes revealed 171 genes were significantly down-regulated (fold-change > 2, adjusted p value < 0.01) in endoderm-directed arid4bΔ cells and 35 genes were up-regulated. We validated the expression of a larger set of lineage specific genes (Fig. 3, b–j). Gene set enrichment analyses (GSEA) demonstrated that the loss of ARID4B was associated with reduced representation of pathways related to proper lineage commitment and embryonic development (Fig. 3, k–m). Signaling pathways activated in stem cell differentiation were down-regulated in arid4bΔ cells (Fig. 3, n–o). On the other hand, type I interferon pathway and cellular viral defense response pathways were strongly activated in arid4bΔ cells (Fig. S3, b and c).

**Chromatin landscape is altered upon Arid4b loss in lineage commitment**

To interrogate changes in the chromatin structure of differentiating mESCs in arid4bΔ cells, we performed ChIP for the histone marks H3K4me3, H3K27me3, and H3K27Ac. We...
compared ChIP-seq intensities of these chromatin marks between WT and arid4bΔ cells using a quantitative algorithm called MAnorm (54). H3K27Ac signal was up-regulated in arid4bΔ mesoderm- or endoderm-differentiated cells (Fig. 4a, Fig. S4a). There was a small but notable change in H3K27me3 levels as well (Fig. 4b, Fig. S4b). In contrast, H3K4me3 peaks were largely unchanged (Fig. 4c, Fig. S4c). Further analysis of H3K27Ac signal revealed the increase to be in regions distal, rather than proximal, to the transcription start site (TSS) (Fig. 4, d and e, Fig. S4, d and e).

Genes responsible for a specific biological process might be coregulated through chromatin changes. Therefore, we used Genomic Regions Enrichment of Annotations Tool (GREAT) to identify biological processes enriched for each chromatin mark (55). Consistent with previous results, genes that lose H3K27Ac and H3K4me3 signal, and genes that gain H3K27me3 signal in mesoderm-directed arid4bΔ cells were strongly enriched in pathways related to embryonic development, pattern specification, and differentiation (Fig. S4, f–h).

H3K27 acetylation is observed in active enhancers. Super-enhancers (SE) are large clusters of enhancers that are marked by broad H3K27Ac and high concentration of transcription activators. They define cell identity by regulating the expression of key cell fate genes (56–58). Given the essential role of ARID4B in mesodermal and endodermal commitment, we assessed whether H3K27Ac changes in arid4bΔ cells correlate with any changes in SEs. We found that the number of SEs is greater in arid4bΔ cells as compared with mesoderm-differentiated WT cells (Fig. 4, f and g). There was a similar increase in the number of SEs in endoderm-differentiated arid4bΔ cells (Fig. 4, h and i). The changes in the number of SEs in arid4bΔ cells might underlie the cell fate defects. Next we analyzed the
genes and pathways enriched in SEs using the GREAT database. We found that SEs unique to endoderm-differentiated WT mESCs were enriched in morphogenetic and developmental processes as well as regulation of transcription (Fig. S4i). No pathways were enriched in common SEs or arid4bΔ unique enhancers. It should also be noted that many of the common SEs exhibited increased H3K27Ac mark in arid4bΔ cells. These results indicate that developmental genes critical for endoderm development might fail to acquire H3K27Ac mark in arid4bΔ cells.

Next, we investigated a possible correlation between changes in chromatin landscape and gene expression. Using SitePro analysis, we found that the genes that are down-regulated in mesoderm-directed WT (blue) and arid4bΔ (red) cells show increased H3K27Ac signal around their TSS. Interestingly, higher H3K4me3 modification around TSS accompanied the H3K27me3 mark in these genes (Fig. 5, a–c). Genes that are down-regulated in endoderm-directed arid4bΔ cells had higher H3K27me3 and H3K4me3, and a pronounced decrease in H3K27Ac than WT (Fig. 5, d–f). On the other hand, up-regulated genes exhibited higher H3K4me3 and H3K27Ac mark around TSS in arid4bΔ cells (Fig. 5, d–f, Fig. S5, d–f). These results indicate that the alterations in H3K27 rather than H3K4 are associated with changes in the gene expression program observed in Arid4b deficiency.

We compared the distribution and the intensity of chromatin marks of WT and arid4bΔ cells using Integrative Genomics Viewer (IGV). Genes required for the establishment of meso/endodermal lineage (such as Bry, Eomes, MixI1, Fgf8, Foxa2, Gsc, Hoxa1, Hoxb1, and Lhx1) and ESC specific genes (Oct4 (Pou5f1), Nanog) in mesoderm-directed WT and arid4bΔ cells. y axes of WT and arid4bΔ tracks are set to the same data range.

Figure 5. Chromatin changes relevant to lineage markers in arid4bΔ cells. a–c, Sitepro analysis of H3K27Ac (a), H3K27Me3 (b), and H3K4Me3 (c) ChIP-seq on transcriptionally down-regulated genes in mesoderm-directed WT (blue) and arid4bΔ (red) cells. x axis, average signal profile; y axis, relative distance from the center (TSS). d–f, Sitepro analysis of H3K27Ac (d), H3K27Me3 (e), and H3K4Me3 (f) ChIP-seq on transcriptionally up-regulated genes in mesoderm-directed WT (blue) and arid4bΔ (red) cells. x axis, average signal profile; y axis, relative distance from the center (TSS). g, Integrative Genomics Viewer visualization of ChIP-seq tracks for selected lineage specific genes (Bry, Eomes, MixI1, Fgf8, Foxa2, Gsc, Hoxa1, Hoxb1, and Lhx1) and ESC specific genes (Oct4 (Pou5f1), Nanog) in mesoderm-directed WT and arid4bΔ cells. x axis, average signal profile; y axis, relative distance from the center (TSS).
higher H3K27me3 and lower H3K27Ac throughout the gene loci in arid4bΔ cells (Fig. 5g, Fig. 5g). However, pluripotency genes Oct4 (Pou5f1) and Nanog were more strongly marked with H3K4me3 and H3K27Ac in arid4bΔ cells (Fig. 5g, Fig. 5g, last two columns).

Discussion

Prior analysis of the role of the Sin3a complex in ESC biology has led to apparently conflicting findings. Sin3a knockout results in embryonic lethality around E3.5 and 6.5 (59, 60). However, loss of the highly related SIN3B protein is lethal only later during development (61). Although Arid4a knockout mice are viable, Arid4b knockout mice die between E3.5 and 7.5 (38). Hdac1 knockout mice are similarly embryonic lethal, whereas Hdac2 deletion is viable (52, 62, 63). Although both HDAC1 and HDAC2 independently interact with SIN3A, it is unclear whether these proteins function within the same complex or are present in alternate Sin3a complexes (45). Taken together with the previous findings, our results point to a unique role of a SIN3A, HDAC1, and ARID4B containing complex in ESC biology and differentiation. We found that, similar to Arid4b deletion, the deletion of Hdac1 but not Hdac2, prevents mesoderm and endoderm differentiation. Our findings support previous reports on the role of HDAC1 in ESCs (50, 53, 64). Moreover, we observe ARID4B interaction with SIN3A and HDAC1, but not with HDAC2, despite considerable Hdac2 expression in these cells. Recently, an ESC-specific variant Sin3a complex was identified, supporting the notion that the composition of the Sin3a complex may vary among cell types and during cell differentiation (65).

A genetic perturbation of a member of a protein complex may lead to formation of residual complexes with different functional outcomes, as recently described for SWI/SNF complexes (66). It is possible that the aberrant interactions in arid4bΔ ESCs might be because of the function of ARID4B-less Sin3a complex rather than the complete loss of Sin3a complex function.

In endoderm-directed arid4bΔ cells, transcripts for 41 genes were up-regulated and 170 genes were down-regulated more than 2-fold (adjusted \( p < 0.05 \)). Similarly, for mesoderm differentiation, transcripts for 39 genes were up-regulated and 308 genes were down-regulated in arid4bΔ cells. Although these genes represent both direct and indirect targets of the Sin3a complex, the observation that a majority of genes are down-regulated upon ARID4B loss is consistent with a role of the Sin3a complex in transcriptional activation. Indeed, evidence from yeast, Drosophila, and mammals reveal that the histone deacetylation by the Sin3a complex has a fine-tuning function for transcribed genes (42–44, 59, 69–79).

Our ChIP-seq experiments revealed critical changes associated with the loss of ARID4B during meso/endodermal differentiation, exemplified by modification at H3K27. Downregulated genes, many of which have key developmental roles, have H3K4me3 around their TSS in arid4bΔ cells, suggesting ARID4B loss does not compromise MLl complex function. However, high H3K27me3 modification accompanies H3K4me3 and there is very little transcriptional output.

The role of ARID4B in mESC lineage commitment

These results suggest that the loss of ARID4B function might alter H3K27me3 deposition or removal in lineage-specific genes upon differentiation and might prevent their transcriptional up-regulation.

On the other hand, we observed elevated H3K27Ac mark and SEs in a subset of genes unrelated to ESC differentiation. SEs harbor a dense population of master regulators of cell fate and the Mediator complex components along with many chromatin factors (56–58). It is possible that the aberrant H3K27Ac-high SE regions in arid4bΔ cells may compete for and sequester away some of these factors required for the chromatin reorganization and transcription of ESC differentiation genes.

Remodeling of the ESC cell cycle is coincident with exit from pluripotency (80, 81). Even though there appears to be a link between these two events, the notion that the change in cell cycle is directly linked to differentiation has been challenged (82, 83). ARID4A has a unique LXCXE motif that mediates interaction with pRB (27). ARID4A recruits the Sin3a corepressor complex (and thus HDAC1) to pRB targets for transcriptional suppression (84–86). This enables cell cycle control through the G1 phase. Interestingly, ARID4B lacks the LXCXE motif and is not predicted to interact with pRB. We also did not detect changes in the number of cycling ESCs or the distribution among cell cycle phases in arid4bΔ ESCs (data not shown). It is conceivable that a change in the composition of the Sin3a complex in arid4bΔ cells might indirectly affect the cell cycle. Similar changes in chromatin complex architecture and function are observed for chromatin remodeling complexes (66, 87–91).

Our Arid4b knockdown and knockout experiments resulted in protein deficiency at the ESC stage, whereas differentiation defects were observed later. Even though arid4bΔ ESCs are similar to WT ESCs on the basis of pluripotency marker expression and cell cycle analyses, we cannot rule out the possibility that the differentiation defect in arid4bΔ cells originates already at the ESC stage. A more detailed analysis of the transcriptomic changes observed in ESCs and throughout the differentiation time course is needed to identify precisely when and where ARID4B function is critical.

Experimental procedures

mESC culture and differentiation

For pooled shRNA screening, a previously established reporter mESC line was used (92) (shared by G. Keller).

mESCs were cultured in the mESC medium (Dulbecco’s modified Eagle’s medium (Life Technologies) supplemented with 15% fetal calf serum (Life Technologies), 0.1 mM β-mercaptoethanol (Sigma), 2 mM L-glutamine (Life Technologies), 0.1 mM nonessential amino acid (Life Technologies), 1% nucleoside mix (Sigma), 1,000 units/ml of recombinant leukemia inhibitory factor (LIF, Chemicon), and 50 units/ml of penicillin/streptomycin (Life Technologies)) on mouse-irradiated fibroblasts (CF-1, Thermo) and gelatinized tissue culture dishes.

Previously established protocols were adapted for mesoderm and endoderm differentiation (92, 93). Briefly mESC plates
The role of ARID4B in mESC lineage commitment

(39x411) (clone number 40146449).

Generation of Arid4b rescue mESCs

using conventional PCR and validated by Western blotting.

ml of ascorbic acid, 4.5

the Broad Institute for sequencing.

Genomic DNA was isolated from sorted cells and was sent to

on BD Aria (DFCI Flow Cytometry Core Facility). Library

tom 5% of undifferentiated cells (lowest BRACHYURY and/or

highest BRACHYURY and FOXA2 expression) as well as bot-

mesoderm: highest BRACHYURY expression, for endoderm:

day 5 of differentiation, the top 5% of differentiated cells (for

mESC sample was taken as the starting shRNA population. At

differentiations were performed as explained above. Day 0

were allowed to recover for 2 days. Mesoderm and endoderm

were collected and dissociated using Accutase (Sigma). Single

cells were replated at 500,000 cells/6-cm Petri dish for endo-

derm and 750,000 cells/6-cm Petri dish for mesoderm in differ-

entiation medium supplemented with cytokines (Activin A (75

ng/ml) for endoderm; Activin A (1 ng/ml), BMP4 (1 ng/ml), and

Wnt3a (3 ng/ml) for mesoderm).

Generation of CRISPR deletion mESCs

Paired single guide RNAs were designed to limit off-target

cleavage and delete critical coding exons of the selected candidate

genes. mESC deletions were performed as previously described

for MEL cells (94, 95). mESC clones were screened

using conventional PCR and validated by Western blotting.

Generation of Arid4b rescue mESCs

Full human ARID4B cDNA was purchased from Dharmacon

(clone number 40146449). HARID4B ORF was amplified with

Ascl and Xbal restriction sites and cloned into pEF1-α-FlagBio

plasmid (96). arid4bΔ mESCs were electroporated with 10 μg

of plasmid using a Bio-Rad electroporator. Clones were screened

with Western blotting using anti-FLAG antibody.

shRNA screen and analysis

A list of epigenetic factors was prepared through literature,

chromatin-related domain homology search, and other database

searches. shRNA selection and library production was done through the Broad Institute the RNAi Consortium.

Brachyury-GFP; Foxa2-hCD4 reporter mESCs were trans-
duced by centrifugation at 2000 rpm at 37 °C for 2 h in serum-

free mESC medium that contains 4 μg/ml of Polybrene. The

transduced cells were immediately washed and plated in con-

ventional mESC medium on a gelatinized tissue culture dish.

Transductions were performed at >200 cells/shRNA to allow

for adequate library representation. After 24 h, transduced cells

were selected using 1 μg/ml of puromycin for 3-4 days. mESCs

were allowed to recover for 2 days. Mesoderm and endoderm

differentiations were performed as explained above. Day 0

mESC sample was taken as the starting shRNA population. At
day 5 of differentiation, the top 5% of differentiated cells (for

mesoderm: highest BRACHYURY expression, for endoderm:

highest BRACHYURY and FOXA2 expression) as well as bot-

tom 5% of undifferentiated cells (lowest BRACHYURY and/or

FOXA2 expression and highest SSEA1 expression) were sorted

on BD Aria (DFCI Flow Cytometry Core Facility). Library

transductions were performed in three independent replicates.

Genomic DNA was isolated from sorted cells and was sent to

the Broad Institute for sequencing.

The analysis of the shRNA screen results were done using

the average of the shRNAs for each gene as well as the

Weighted Sum method on the GENE-E program developed by

the Broad Institute. Day 5 shRNA representation was com-
pared with day 0 mESC shRNA representation. Additionally,

day 5 differentiated to undifferentiated comparison was also

performed. The genes with less than three scored shRNAs were

eliminated from analyses. Genes that are depleted at least 2-

fold compared with the day 0 or day 5 undifferentiated popula-
tion were selected as candidates. Of these candidate genes, the

ones that show up in only one of the three biological replicates

were eliminated. Known Polycomb and Trithorax group pro-

ten proteins were also discarded from further study. The final list of
candidate genes were tested one by one with three independent

shRNAs in mESCs for differentiation toward mesoderm and

endoderm.

Flow cytometry

mESCs or differentiated cells were dissociated into single

cells and stained with anti-SSEA1-Alexa Fluor 647 (eBio-
science, 51-8213) and anti-human CD4-PE (eBioscience, 12-
0049). Flow cytometry was performed on BD Fortessa and an-
alyzed on FlowJo software. Cell sorting was done in DFCI Flow

Cytometry Core Facility on BD FACSAria cell sorters.

RT-qPCR and RNA-seq

Cells were collected and resuspended in TRIzol (Thermo, 15596018). RNA was extracted using Qiagen RNeasy plus kits

accompanying protocols. The concentration of puri-

fied RNA samples was tested on Nanodrop. Equal amounts of total RNA (250 ng to 1 µg) was converted into cDNA using

an iScript cDNA synthesis kit (Bio-Rad, 1708890). qPCR was

performed with primers listed in Table 1 and Table 2.

Flowjo (Bio-Rad) using Bio-Rad CFX96 and CFX384

machines according to the manufacturer’s protocols.

For RNA-seq, genomic DNA was eliminated in a column
during RNA extraction using DNase (Qiagen, 79254). The quality of the RNA samples was tested on an Agilent BioAna-

lyzer (DFCI CCCB Core Facility). Libraries were prepared

using New England Biolabs reagents (NEBnext ultra direc-
tional RNA library prep kit (E7420S), NEBnext rRNA depletion

kit (E6310S), and NEBnext multiplex olligos for Illumina sequenc-
ing (E7335S)). The concentrations of library cDNA samples were

analyzed using Qubit. Sequencing was performed using Illumina

HiSeq2000.

Western blotting

WT and arid4bΔ were grown and lysed directly in 2×

Laemmli buffer (Bio-Rad) including β-mercaptoethanol at

95 °C for 10 min. After centrifugation to remove cell debris,
equal amounts of cell lysate were loaded on 12% polyacrylamide
gel. Primary antibodies (Hdac1 (06-720; Millipore), Hdac2 (sc-
7899; Santa Cruz), Arid4b (A302-233A; Bethyl), α-H3Ac (06-599;
Millipore), actin (Mab1501; Millipore)) and horseradish peroxi-
dase-conjugated secondary antibodies were used for detection

(Table 2).
Table 1
Primer sequences used in this study

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-actin-F-qPCR</td>
<td>ATGAAGATCCCTGAGGAGCGA</td>
</tr>
<tr>
<td>B-actin-R-qPCR</td>
<td>TACATGCTCAGGAGGAGCAC</td>
</tr>
<tr>
<td>Oct4-F-qPCR</td>
<td>CTGAGGCGCCAGCCGAGGAGCACGAG</td>
</tr>
<tr>
<td>Oct4-R-qPCR</td>
<td>CTGAGGCGCCAGCCGAGGAGCACGAG</td>
</tr>
<tr>
<td>Nanog-F-qPCR</td>
<td>AATGTCAGCAGGAGGAGCACGAG</td>
</tr>
<tr>
<td>Nanog-R-qPCR</td>
<td>CCGGAGCAGGAGGAGCAGGAGCAGGAGCAGG</td>
</tr>
<tr>
<td>Eomes-F-qPCR</td>
<td>GTTCCTGAGGAGGAGCAGGAGCAGGAGCAGG</td>
</tr>
<tr>
<td>Eomes-R-qPCR</td>
<td>TGTTTTCGTGGAAGTGGTTCTGGC</td>
</tr>
<tr>
<td>Cxcr4-F-qPCR</td>
<td>GTAGAGGTTGACAGTGTAGAT</td>
</tr>
<tr>
<td>Cxcr4-R-qPCR</td>
<td>GGCTGTAGAGCGAGTGTTGC</td>
</tr>
<tr>
<td>Gsc-F-qPCR</td>
<td>ACCATCTTCACCGAGGAGCACAGCAGCAGCAG</td>
</tr>
<tr>
<td>Gsc-R-qPCR</td>
<td>CTTGGCTCGGCGGTTCTTAAAC</td>
</tr>
<tr>
<td>Pax6-F-qPCR</td>
<td>AAGCATGCTCAGTGGCT</td>
</tr>
<tr>
<td>Pax6-R-qPCR</td>
<td>ATGAAGATCCTGACGAGCCAGGAGCGA</td>
</tr>
<tr>
<td>Sox17-F-qPCR</td>
<td>GCCAAAGACGAACGCAAGCGGT</td>
</tr>
<tr>
<td>Sox17-R-qPCR</td>
<td>TCATGCCCTCACCCTGGCTTC</td>
</tr>
<tr>
<td>Foxa2-F-qPCR</td>
<td>TGGTCACTGGGGACAAGGGAA</td>
</tr>
<tr>
<td>Foxa2-R-qPCR</td>
<td>GGTCTCGGGAAAGCAGTGGC</td>
</tr>
<tr>
<td>Nanog-F-qPCR</td>
<td>ATGAAGTGCAAGCGGTGGCAG AAA</td>
</tr>
<tr>
<td>Nanog-R-qPCR</td>
<td>TTGGTACCTGGGAGGAGCAGGAGCAGG</td>
</tr>
<tr>
<td>Oct4-F-qPCR</td>
<td>ACTTGGACGGAAACTGACAC</td>
</tr>
<tr>
<td>Oct4-R-qPCR</td>
<td>CTGTAGGGAGGGCTTCGGCCTTAAAC</td>
</tr>
<tr>
<td>B-actin-R-qPCR</td>
<td>ATGAAGATCCTGACGAGCCAGGAGCGA</td>
</tr>
<tr>
<td>B-actin-F-qPCR</td>
<td>TACATGCTCAGGAGGAGCAC</td>
</tr>
<tr>
<td>HDAC2-F-qPCR</td>
<td>ATGCAGAGATTTAACGTCGGAG</td>
</tr>
<tr>
<td>HDAC2-R-qPCR</td>
<td>TCTGAGAAGTGAGGAACTTGGC</td>
</tr>
<tr>
<td>HDAC1-F-qPCR</td>
<td>AAGGAGGAGAAGCCAGAAGCG</td>
</tr>
<tr>
<td>HDAC1-R-qPCR</td>
<td>AGGTCTGAGTCTTGCTGTGT</td>
</tr>
<tr>
<td>FD1-F-qPCR</td>
<td>ATGCAGAGATTTAACGTCGGAG</td>
</tr>
<tr>
<td>FD1-R-qPCR</td>
<td>TCTGAGAAGTGAGGAACTTGGC</td>
</tr>
<tr>
<td>Histone 3K27ac</td>
<td>Western blotting</td>
</tr>
<tr>
<td>Histone 3K27me3</td>
<td>Western blotting</td>
</tr>
<tr>
<td>Histone 3K4me3</td>
<td>Western blotting</td>
</tr>
<tr>
<td>Histone 3K27/4me</td>
<td>Western blotting</td>
</tr>
<tr>
<td>SSEA-1</td>
<td>Protein interaction</td>
</tr>
<tr>
<td>CD4-PE</td>
<td>Flow Cytometry</td>
</tr>
<tr>
<td>β-Actin</td>
<td>Western blotting</td>
</tr>
</tbody>
</table>

Table 2
Antibodies used in this study

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Company</th>
<th>Catalog number</th>
<th>Experiment used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hdac1</td>
<td>Millipore</td>
<td>06-720</td>
<td>Western blotting</td>
</tr>
<tr>
<td>Hdac1</td>
<td>Santa Cruz</td>
<td>sc-81598</td>
<td>Proximity ligation assay</td>
</tr>
<tr>
<td>Hdac2</td>
<td>Santa Cruz</td>
<td>sc-7899</td>
<td>Western blotting</td>
</tr>
<tr>
<td>Hdac2</td>
<td>Santa Cruz</td>
<td>sc-9959</td>
<td>Proximity ligation assay</td>
</tr>
<tr>
<td>Sin3a</td>
<td>Active Motif</td>
<td>39865</td>
<td>Western blotting</td>
</tr>
<tr>
<td>Arid4b</td>
<td>Bethyl</td>
<td>A302-233A</td>
<td>Immunoprecipitation, Western blotting, proximity ligation assay</td>
</tr>
<tr>
<td>Histone 3K27ac</td>
<td>Active Motif</td>
<td>39133</td>
<td>ChIP</td>
</tr>
<tr>
<td>Histone 3K27me3</td>
<td>Active Motif</td>
<td>39155</td>
<td>ChIP</td>
</tr>
<tr>
<td>Histone 3K4me3</td>
<td>Active Motif</td>
<td>07-473</td>
<td>ChIP</td>
</tr>
<tr>
<td>h3A3c</td>
<td>Millipore</td>
<td>06-599</td>
<td>Western blotting</td>
</tr>
<tr>
<td>SSEA-1</td>
<td>eBioscience</td>
<td>50-8813</td>
<td>Flow Cytometry</td>
</tr>
<tr>
<td>CD4-PE</td>
<td>eBioscience</td>
<td>12-0049</td>
<td>Flow Cytometry</td>
</tr>
<tr>
<td>β-Actin</td>
<td>Millipore</td>
<td>Mab1501</td>
<td>Western blotting</td>
</tr>
</tbody>
</table>

Co-immunoprecipitation

Nuclear extracts were prepared from WT (C9) and arid4bΔ mESCs using the Universal Magnetic CoIP Kit (Active Motif, catalog number 54002) according to the manufacturer’s protocol. For co-immunoprecipitation, kit protocol was followed. 400 μg of nuclear extract was incubated with 5 μg of anti-Arid4b (A302-233A; Bethyl) antibody. After immunoprecipitation and washes, beads were boiled in 2× Laemmlı buffer (Bio-Rad) supplemented with β-mercaptoethanol at 95 °C for 10 min.

Glycerol sedimentation assay

WT (C9) mESCs were grown and glycerol sedimentation assay was performed as previously described (67).

The role of ARID4B in mESC lineage commitment

PLA was performed using DuoLink In Situ Red Starter Kit (Sigma, DUO92101) using Arid4b (rabbit, 1:250) and Hdac1 (mouse, 1:250) or Hdac2 (mouse, 1:250) primary antibodies on WT (C9) mESCs according to the manufacturer’s protocol. Confocal microscopy (Leica TCS SP8) imaging was performed at Bilkent University UNAM Laboratories.

Histone proteomics

Quantitative analysis of histone post-translational modifications was performed in collaboration with Dr. Jacob Jaffe of the Broad Institute Proteomics Platform. WT, arid4bΔ, hdac1Δ, and hdac2Δ mESCs as well as endoderm-directed cells were collected and processed to isolate histones. The procedure was completed as described in Ref. 97. The enrichment results for each modification in knockout cells were normalized to the WT counterpart and visualized using Morpheus tool of the Broad Institute.

ChIP sequencing

ChIP was performed as previously described (96) using the following antibodies: H3K27ac (Active Motif, 39133), H3K27me3 (Active Motif, 39155), and H3K4me3 (Millipore, 07-473). ChIP-seq libraries were prepared using the NEBNext ChIP-seq library kit (E6240S) and NEBnext multiplex ologs for Illumina sequencing (E7335S) according to the manufacturer’s protocol. The concentrations of library cDNA samples were analyzed using Qubit. Sequencing was performed using Illumina HiSeq2000.

RNA-seq data analysis

RNA-seq reads were aligned to the reference mouse genome mm10 using STAR (98) with default parameters. Aligned reads were counted in the genomic transcripts annotations from GenomicFeatures (99), using Rsamtools (Morgan M, 2016). DESeq2 (100) used for differentially expressed gene analysis was performed with the threshold at an adjusted p value 0.01 and fold-change 2.

ChIP-seq data analysis

ChIP-seq reads were aligned to the mm10 reference genome using Bowtie2 (101) with default parameters. Duplicate reads were removed using PICARD tools (RRID:SCR_006525).
### Table 4
Summary of ChIP-seq data (GSE153634) generated in this study

<table>
<thead>
<tr>
<th>Sample</th>
<th>Raw reads</th>
<th>Aligned reads</th>
<th>Aligned reads rmdup</th>
<th>Peak</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K27ac_endo_Arid4b</td>
<td>81,963,501</td>
<td>76,581,910</td>
<td>9,778,716</td>
<td>35,884</td>
<td>GSM464859</td>
</tr>
<tr>
<td>H3K27ac_endo_wt</td>
<td>47,539,489</td>
<td>45,868,810</td>
<td>9,368,799</td>
<td>14,216</td>
<td>GSM464855</td>
</tr>
<tr>
<td>H3K27ac_meso_Arid4b</td>
<td>49,307,346</td>
<td>45,368,018</td>
<td>18,533,497</td>
<td>12,496</td>
<td>GSM4648467</td>
</tr>
<tr>
<td>H3K27me3_meso_wt</td>
<td>70,967,345</td>
<td>69,144,743</td>
<td>32,912,724</td>
<td>16,660</td>
<td>GSM464858</td>
</tr>
<tr>
<td>H3K27me3_meso_Arid4b</td>
<td>56,464,721</td>
<td>55,101,389</td>
<td>23,397,909</td>
<td>10,326</td>
<td>GSM464854</td>
</tr>
<tr>
<td>H3K4me3_meso_wt</td>
<td>53,621,397</td>
<td>52,172,143</td>
<td>24,627,773</td>
<td>6,150</td>
<td>GSM4648463</td>
</tr>
<tr>
<td>H3K4me3_meso_Arid4b</td>
<td>51,901,859</td>
<td>50,643,552</td>
<td>31,218,452</td>
<td>9,863</td>
<td>GSM4648466</td>
</tr>
<tr>
<td>H3K27me3_endo_wt</td>
<td>52,521,226</td>
<td>51,353,318</td>
<td>28,569,202</td>
<td>26,852</td>
<td>GSM4648466</td>
</tr>
<tr>
<td>H3K27me3_endo_Arid4b</td>
<td>57,662,288</td>
<td>56,239,928</td>
<td>24,042,979</td>
<td>21,752</td>
<td>GSM4648465</td>
</tr>
<tr>
<td>H3K27ac_meso_wt</td>
<td>51,507,812</td>
<td>50,454,506</td>
<td>24,627,773</td>
<td>6,150</td>
<td>GSM4648466</td>
</tr>
<tr>
<td>H3K27ac_meso_Arid4b</td>
<td>51,872,178</td>
<td>50,792,076</td>
<td>29,504,382</td>
<td>25,272</td>
<td>GSM464857</td>
</tr>
<tr>
<td>H3K4me3_meso_wt</td>
<td>52,918,141</td>
<td>51,792,208</td>
<td>31,218,452</td>
<td>26,852</td>
<td>GSM4648466</td>
</tr>
<tr>
<td>H3K4me3_meso_Arid4b</td>
<td>52,918,141</td>
<td>51,792,208</td>
<td>31,218,452</td>
<td>26,852</td>
<td>GSM4648466</td>
</tr>
<tr>
<td>Input_endo_Arid4b</td>
<td>51,803,055</td>
<td>50,792,076</td>
<td>32,502,578</td>
<td>21,752</td>
<td>GSM4648465</td>
</tr>
<tr>
<td>Input_endo_wt</td>
<td>50,660,867</td>
<td>49,790,616</td>
<td>24,042,979</td>
<td>12,496</td>
<td>GSM4648466</td>
</tr>
<tr>
<td>Input_endo_Arid4b</td>
<td>51,901,859</td>
<td>50,643,552</td>
<td>32,502,578</td>
<td>21,752</td>
<td>GSM4648465</td>
</tr>
<tr>
<td>Input_endo_wt</td>
<td>52,918,141</td>
<td>51,792,208</td>
<td>32,502,578</td>
<td>21,752</td>
<td>GSM4648465</td>
</tr>
<tr>
<td>Input_endo_Arid4b</td>
<td>56,060,867</td>
<td>54,790,616</td>
<td>40,859,382</td>
<td>22,292</td>
<td>GSM464853</td>
</tr>
<tr>
<td>Input_endo_wt</td>
<td>52,918,141</td>
<td>51,792,208</td>
<td>39,111,202</td>
<td>14,216</td>
<td>GSM4648464</td>
</tr>
<tr>
<td>Input_endo_wt</td>
<td>56,640,634</td>
<td>55,521,376</td>
<td>44,735,823</td>
<td>35,884</td>
<td>GSM464859</td>
</tr>
</tbody>
</table>

MACS2 (102) was used for ChIP-seq peaks calling MACS2 with the following parameters (–nomodel –keep-dup 1 –extsize = 146 -q 0.01). Peaks were filtered using the consensus excludable ENCODE blacklist (The ENCODE Project Consortium, 2012). MANorm (54) was used for determining differential ChIP-seq peaks between WT and KO as previously described (103) with the threshold of M-value 1 and FDR 0.01.

**Data availability**

Data have been deposited in the Gene Expression Omnibus with accession numbers GSE153633) and GSE153634) (Tables 3 and 4).

**Acknowledgments**—We thank Gordon Keller for providing the Bry-GFP; Foxa2-hCD4 mESC line and mesoderm differentiation protocol. We thank Valerie Gouon-Evans for endoderm differentiation protocols. We thank Davide Seruggia for the preparation of ChIP-seq libraries. We thank members of the Orkin lab for help with experimental protocols, sharing reagents, and critical comments.

**Author contributions**—N. T. C. and S. H. O. conceptualization; N. T. C., E. G. K., and X. W. data curation; N. T. C., J. H., and E. G. K. formal analysis; N. T. C. and S. H. O. supervision; N. T. C., J. H., E. G. K., X. W., and I. E. methodology; N. T. C. writing-original draft; N. T. C. project administration; N. T. C., E. G. K., and S. H. O. visualization; N. T. C. resources; S. H. O. funding acquisition.

**Funding and additional information**—S. H. Orkin is an investigator of the Howard Hughes Medical Institute (HHMI).

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: ESC, embryonic stem cell; HDAC, histone deacetylase; PLA, proximity ligation assay; TSS, transcription start site; shRNA, short hairpin RNA; SE, Super-enhancers; qPCR, quantitative PCR.
functions of JmjD2b/Kdm4b and JmjD2c/Kdm4c in mouse embryonic stem cell identity. *Mol. Cell.* **33,** 32–48 CrossRef Medline


The role of ARID4B in mESC lineage commitment
decommissioning by LSD1 during embryonic stem cell differentiation. Nature 482, 221–225 CrossRef
The role of ARID4B in mESC lineage commitment


