The Tumor Suppressor ARID1A Controls Global Transcription via Pausing of RNA Polymerase II

Highlights

- ARID1A binds most active promoters and enhancers in ovarian clear cell carcinoma
- Depletion of ARID1A decreases promoter-proximal paused RNAPII
- ARID1B rescues ARID1A loss and restores physiological RNAPII activity
- A subset of genes is strictly dependent on ARID1A in ovarian clear cell carcinoma

Authors
Marco Trizzino, Elisa Barbieri, Ana Petracovici, ..., Silvia Licciulli, Rugang Zhang, Alessandro Gardini

Correspondence
agardini@wistar.org

In Brief
Trizzino et al. show that ARID1A regulates RNA polymerase II promoter-proximal pausing in ovarian clear cell carcinoma. Conversely, ARID1A depletion has modest effects on accessibility. Following ARID1A loss, upregulation of ARID1B restores physiological polymerase II dynamics, rescuing transcription. However, multiple p53 and estrogen target genes are strictly dependent on ARID1A.

Data and Software Availability
GSE104545
The Tumor Suppressor ARID1A Controls Global Transcription via Pausing of RNA Polymerase II

Marco Trizzino,1,4 Elisa Barbieri,1,4 Ana Petracovici,2 Shuai Wu,1 Sarah A. Welsh,1,3 Tori A. Owens,1 Silvia Licciulli,1 Rugang Zhang,1 and Alessandro Gardini1,5,*

1The Wistar Institute, Gene Expression and Regulation Program, 3601 Spruce Street, Philadelphia, PA 19104, USA
2Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Boulevard, Philadelphia, PA 19104, USA
3Biochemistry and Molecular Biophysics Graduate Group, Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Boulevard, Philadelphia, PA 19104, USA
4These authors contributed equally
5Lead Contact
*Correspondence: agardini@wistar.org
https://doi.org/10.1016/j.celrep.2018.05.097

SUMMARY

AT-rich interactive domain-containing proteins 1A and 1B (ARID1A and ARID1B) are mutually exclusive subunits of the chromatin remodeler SWI/SNF. ARID1A is the most frequently mutated chromatin regulator across all cancers, and ovarian clear cell carcinoma (OCCC) carries the highest prevalence of ARID1A mutations (~57%). Despite evidence implicating ARID1A in tumorigenesis, the mechanism remains elusive. Here, we demonstrate that ARID1A binds active regulatory elements in OCCC. Depletion of ARID1A represses RNA polymerase II (RNAPII) transcription but results in modest changes to accessibility. Specifically, pausing of RNAPII is severely impaired after loss of ARID1A. Compromised pausing results in transcriptional dysregulation of active genes, which is compensated by upregulation of ARID1B. However, a subset of ARID1A-dependent genes is not rescued by ARID1B, including many p53 and estrogen receptor (ESR1) targets. Our results provide insight into ARID1A-mediated tumorigenesis and unveil functions of SWI/SNF in modulating RNAPII dynamics.

INTRODUCTION

AT-rich interactive domain-containing proteins 1A and 1B (ARID1A and ARID1B) are mutually exclusive subunits of the chromatin remodeler complex SWI/SNF (Raab et al., 2015; Ryme et al., 2009). As part of the core SWI/SNF complex, ARID1A and ARID1B are conserved throughout metazoans and expressed across all human tissues. Although the AT-rich interactive domain (ARID) has been shown to bind DNA in vitro (Kortschak et al., 2000; Wilsker et al., 2004), the function of ARID1A and ARID1B in vivo is poorly understood. ARID1A and ARID1B are components of the BAF sub-family of SWI/SNF complexes, in association with the BRG1/BRM ATPases that maintain accessibility at most promoters and enhancers (Bao et al., 2015; Tolstorukov et al., 2013). There are limited genome-wide studies of ARID1A and ARID1B, which suggest substantial overlap of their target loci but functionally distinct roles (Raab et al., 2015). Importantly, ARID1A is the most frequently mutated chromatin regulator across all human cancers (Kadoch et al., 2013). In particular, ovarian clear cell carcinoma (OCCC) carries the highest prevalence of ARID1A mutations (~57%) (Bitler et al., 2015, 2017; Jones et al., 2010). Conversely, ARID1B is rarely found mutated. Importantly, the large majority of ARID1A mutations are frameshift indels or nonsense point mutations resulting in functional loss of the entire protein (Ayhan et al., 2012; Jones et al., 2010).

Recent studies have proposed ARID1A as regulator of H3 acetylation (H3K27ac) (Alver et al., 2017; Mathur et al., 2017) and chromatin accessibility (Kelso et al., 2017) at distal enhancer regions. Using a model of OCCC, we examine the immediate consequences of ARID1A loss on transcription and nucleosome positioning. We find only a marginal role for ARID1A in maintaining accessibility at regulatory elements. Instead, we describe an unanticipated role for this tumor suppressor in regulating widespread RNAPII pausing. The mechanism of polymerase pausing is largely conserved across higher metazoans and represents a pivotal instrument to allow fast, coordinated, and robust transcription at hundreds of loci during cell homeostasis and development (Adelman and Lis, 2012; Levine, 2011). Paused RNA polymerase is centrally regulated by the negative elongation factor (NELF), which prevents productive elongation and determines arrest of the RNAPII holoenzyme within the first 100 bp from the transcription start site (TSS). Depletion of NELF and loss of paused RNAPII result in overall loss of polymerase from the gene locus and transcriptional downregulation (Gilchrist et al., 2010; Williams et al., 2015). Although several complexes that allow coordinated release of paused polymerase have been identified (Gardini et al., 2014; Patel et al., 2013; Rahl et al., 2010), little is known about additional mechanisms and proteins that actively enforce pausing genome-wide (Chen et al., 2015). Here, we identify a
RESULTS

ARID1A Binds Active Enhancers and Promoters in Ovarian Epithelial Cells

Although ARID1A is a constitutive component of the evolutionarily conserved SWI/SNF complex, its genome-wide localization has remained elusive. To profile ARID1A binding in ovarian epithelial cells, we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) in the OCCC-derived RMG-1 cell line (Nozawa et al., 1988) using two different antibodies (Figures 1 and S1A). We assessed the specificity of ARID1A antibodies and its association with SWI/SNF by co-immunoprecipitation (Figure S1B). Alongside ARID1A, we profiled H3K27ac, a chromatin mark that encompasses transcriptionally active regulatory elements. We attributed H3K27ac/ARID1A peaks to promoters when located in the vicinity (<1 kb) of an annotated TSS, whereas distal peaks (>1 kb from the TSS) were classified as enhancers. ChIP-seq revealed that ARID1A is bound to nearly all active regulatory elements in OCCC cells (37,773 enhancers, 13,167 promoters). ARID1A and H3K27ac profiles were highly correlated across all human genes (Figures 1B and 1C), suggesting that ARID1A is broadly distributed across active genes and that its occupancy increases with transcriptional activity (e.g., we found robust ARID1A binding across the top 200 genes expressed in RMG-1; Figure S1C). Altogether, our results are consistent with recent reports suggesting association of SWI/SNF with enhancers (Alver et al., 2017; Mathur et al., 2017), supporting the notion that ARID1A also targets a wide spectrum of active genes in mammalian cells (Mathur et al., 2017; Raab et al., 2015; Tolstorukov et al., 2013).

To further characterize the binding pattern of ARID1A in OCCC, we used assay for transposase-accessible chromatin followed by sequencing (ATAC-seq) to chart DNA accessibility and found, as expected, that all ARID1A target regions are highly accessible (Figure S1D). However, we observed unanticipated differences in the occupancy profiles of ARID1A at enhancers and promoters. Specifically, at most promoters, ARID1A presents a bimodal distribution overlapping the “−1” and “+1” nucleosomes, while the nucleosome-depleted region (NDR) surrounding the TSS is devoid of ARID1A (Figures 1D and 1F). Conversely, ARID1A binding profile at enhancers mirrors that of the NDR defined by ATAC-seq (Figures 1E and 1G).

Loss of ARID1A Determines Widespread Transcriptional Decrease in Absence of Major Accessibility Changes

To recapitulate early events in ovarian carcinogenesis, we depleted ARID1A in OCCC-derived cells. We used the OCCC cell line RMG-1, bearing two wild-type alleles of ARID1A, to generate stable clones with an inducible shRNA against ARID1A under the control of a tetracycline-responsive promoter. In this system, we nearly eliminate chromatin-bound ARID1A within 72 hr (Figures 2A and S1E). Next, we measured changes in accessibility and transcription upon depletion of ARID1A. Previous reports postulate that ARID1A may specifically modulate a subset of enhancers (Kelso et al., 2017), but our ChIP-seq analysis (Figure 1) suggests a widespread role of ARID1A at most protein-coding genes. Notably, ATAC-seq analyses indicated that accessibility at promoters is largely unchanged upon ARID1A knockdown (KD) (Figure 2B). We next examined enhancer regions before and after ARID1A-KD (Figures 2C–2E). We clustered enhancers into two categories on the basis of H3K27ac level: (1) weak enhancers (quartiles 1 and 2 of H3K27ac intensity) and (2) strong enhancers (H3K27ac quartiles 3 and 4). Although we observed loss of accessibility at cis-regulatory elements (Kelso et al., 2017), this effect was largely confined to weak enhancers (Figure 2C) that are poorly bound by ARID1A (Figure 2E). Conversely, accessibility at strong enhancers, highly acetylated and enriched in ARID1A, was slightly affected (Figure 2D). Statistical analysis of differential accessibility (edgeR; Table S1) confirmed that significant changes (false discovery rate [FDR] < 10%) in accessibility were detectable in only ~2% of the tested regions, all of which represented weak enhancers. To further examine the enhancer landscape of OCCC, we performed unbiased motif search and retrieved the same transcription factor (TF) signature (E2F, ELF3, SOX8; Data S1) at both categories of enhancers, with the sole exception of AP-1, which was found only at strong enhancers, suggesting that FOS/JUN TFs drive enhancer activity in RMG-1 regardless of the presence of ARID1A.

Next, we performed RNA sequencing (RNA-seq) analysis (spike-in controlled, two independent replicates) and observed that short-term depletion of ARID1A (72 hr), despite modest or no change in accessibility, elicits widespread transcriptional inhibition (Figures 2F and 2G). We tested the possibility that decreased enhancer activity (ARID1A dependent) may affect global transcription. We identified the active enhancer regions neighboring the 200 most downregulated protein-coding genes and analyzed their accessibility and acetylation status. Remarkably, we found no significant changes in DNA accessibility or H3K27ac levels at these enhancers (Figure S2C), suggesting that ARID1A regulates basal transcription of protein-coding genes independent of its putative role at distal cis-regulatory elements. Furthermore, to exclude off-target or indirect effects, we performed transient depletion of ARID1A using a different short hairpin RNA (shRNA) sequence and rescued its protein levels with exogenous ARID1A (wild-type [WT] RMG-1 cells were co-transfected with shRNAs and either the pLVX backbone vector or the pLVX-ARID1A vector). We assayed expression of candidate genes downregulated in our inducible system and found that their modulation is ARID1A dependent (Figure 2H). Furthermore, we analyzed the effect of short-term (72 hr) ARID1A depletion on two additional cell lines, the ovarian cystadenocarcinoma OVCA429 and the embryonic-derived HEK293T. Although OVCA429 recapitulated the effect observed in RMG-1 cells (Figure S2A), we could not observe major changes in gene expression when performing the same experiment in HEK293T (Figure S2B), suggesting that this phenotype may be specific to ovarian epithelial cells. Taken together, our data suggest that ARID1A modulates widespread transcriptional activation with a mechanism independent of chromatin remodeling.
ARID1A Loss Affects RNAPII-Dependent Transcription

Global variations of steady-state RNA, especially at such a broad scale, may be attributed to dysregulation of RNA processing and stability. To determine whether loss of ARID1A directly affects nascent transcription, we performed global run-on sequencing (GRO-seq; Core et al., 2008; Gardini, 2017) and RNA-seq of the chromatin fraction (CHROM-RNA-seq; Lai et al., 2015). Depletion of ARID1A led to significant decrease (Wilcoxon rank-sum test...
p < 2.2 × 10^{-16}) of nascent transcription (GRO-seq) both at protein-coding genes and intergenic enhancers (Figure 3A). Notably, ARID1A loss did not affect either RNAPII-dependent rRNA genes or RNAPIII-transcribed genes (Figure 3A), while the vast majority of RNAPII-dependent genes were downregulated (Figure 3B).

To investigate the nature of this transcriptional impairment, we used GRO-seq to gauge pausing indexes (PIs) for all transcribed genes (n = 6,281). The PI (Core et al., 2008) is an estimate of the ratio of RNA polymerase II (RNAPII) that is found in pausing conformation versus its elongating form. We observed that pausing of RNAPII (−50/+300 bp from the TSS) diminished across all active genes in the absence of ARID1A (Figure 3D; Wilcoxon rank-sum test p < 2.2 × 10^{-16}). Next, we performed PI analysis using CHROM-RNA-seq, which is highly enriched for primary
transcripts. Coherently, CHROM-RNA-seq revealed a decrease in PIs across RNAPII genes (Figure 3E; Wilcoxon rank-sum test $p < 2.2 \times 10^{-10}$). Collectively, our analysis of nascent and primary transcripts suggests that loss of ARID1A has a direct bearing on real-time transcription and may affect the dynamics of RNAPII.

**ARID1A Loss Causes Global Reduction of Paused RNAPII**

To investigate the impact of ARID1A loss on RNAPII dynamics, we performed ChiP-seq, in two independent biological replicates, before and after depletion of ARID1A. We used antibodies raised against Rpb1 (POLR2A) that recognize all isoforms of RNAPII, irrespective of phosphorylation status. ChiP-seq revealed significant decrease of RNAPII occupancy at most TSSs of protein-coding genes (Wilcoxon rank-sum test $p < 2.2 \times 10^{-10}$; Figures 4A–4D). There is broad consensus that most of the promoter-proximal RNAPII signal represents the paused form of polymerase (Adelman and Lis, 2012; Henriques et al., 2013; Levine, 2011; Williams et al., 2015). After transcription initiation, an early RNAPII elongation complex begins...
Figure 4. Accumulation of Paused RNAPII in OCCC Cells Depends on ARID1A

(A) ARID1A depletion (ARID1A-KD) affects RNAPII levels at the majority of active TSSs. The heatmap shows 5,914 TSS targeted by SWI/SNF, ranked by mean intensity of ARID1A.

(B) Average profiles of RNAPII ChIP-seq (ARID1A-WT, KD) around the TSS of 5,914 genes presented in (A).

(C) UCSC Genome Browser screenshot of RNAPII ChIP-seq profiles (ARID1A-WT, KD) at PEX13 and MYC loci. There is a significant reduction of RNAPII occupancy at the promoter-proximal regions.

(D) Box plots showing read depth of RNAPII ChIP-seq at the TSS of ARID1A target genes. RNAPII occupancy is significantly reduced (p < 2.2 \times 10^{-16}) after depletion of ARID1A, in two independent replicates.

(E) Immunoblot analysis of the chromatin fraction of OCCC cells, before and after ARID1A depletion. Although total RNAPII levels (assayed with two different antibodies, see Experimental Procedures) and phosphorylation of Ser2 remain unchanged, phosphorylation of Ser5 is reduced. Lamin C is shown as loading control. Quantitative immunoblots are shown Figure S3B.

(F) Traveling ratio analysis reveals changes in RNAPII occupancy at proximal promoter versus gene body. Incremental plots demonstrate that ARID1A depletion elicits a reduction of pausing genome-wide.

(G) Parental RMG-1 cells were infected with shRNAs against ARID1A and rescued with exogenously expressed full-length ARID1A. qPCR at three candidate genes shows that downregulation of RNAPII at TSS is significantly reversed (n = 3) by ARID1A overexpression. Error bars represent SD.
mRNA synthesis, only to arrest within the first 30–50 bp. The early elongating RNAPII is highly phosphorylated at Ser5 of the CTD (C-terminal domain) and is under the control of the negative elongation complex, NELF and DSIF. Productive elongation requires the recruitment of pTEFb, which phosphorylates NELF, DSIF, and Ser2 of the CTD to ultimately release RNAPII from the promoter-proximal region (Adelman and Lis, 2012; Levine, 2011). Notably, we observed a sharp reduction of RNAPII at the TSSs of coding genes (Figures 4A–4D and S3A) upon depletion of ARID1A. We performed genome-wide analysis of phospho-RNAPII and observed a significant reduction of initiating and elongating polymerase, but phospho-Ser5 boasted the largest decrease (Figure S3B). Importantly, total levels of RNAPII in chromatin extracts remained unchanged (Figure 4E; quantitative blots in Figure S3C). To further investigate RNAPII dysregulation upon loss of ARID1A, we analyzed the traveling ratio of polymerase to assess changes in the fraction of paused versus elongating RNAPII (Rahl et al., 2010). Depletion of ARID1A elicited a marked increase in traveling ratio at active genes (n = 6,281; Wilcoxon rank-sum test p < 2.2 × 10−16; Figure 4F), consistent with PI analysis of nascent transcription (Figures 3D and 3E) and suggestive of a role for ARID1A in maintaining RNAPII pausing in ovarian epithelial cells. To confirm that ARID1A is directly implicated in regulation of pausing, we performed a rescue experiment in parental RMG-1 cells. Using a different shRNA for ARID1A, we were able to successfully restore RNAPII occupancy upon co-expression of exogenous ARID1A (Figure 4G).

**ARID1A Depletion Does Not Affect Transcriptional Initiation**

Our data suggest that ARID1A loss prevents accumulation of paused RNAPII at the proximal promoter. We hypothesize that ARID1A mediates RNAPII regulation directly at the chromatin level. Accordingly, bulk levels of RNAPII and its elongating form (Ser-2P) do not change in chromatin extracts (Figure 4E; quantitative blots in Figure S3B), whereas the initiation/early elongation form of RNAPII (Ser-5P) diminishes. Furthermore, protein levels of pausing regulators NELF and SPT5 are unchanged. However, widespread dampening of transcription at RNAPII genes may be also attributed to initiation defects and/or impaired loading of the pre-initiation complex, all of which would result in marked loss of RNAPII at the TSS (Figure 4A). First, we analyzed a group of specialized RNA genes (RNAPII-dependent U snRNAs) that are not known to rely on polymerase pausing to regulate their expression. U snRNAs are highly transcribed short genes constitutively expressed and boast the highest levels of RNAPII by ChIP-seq. Our data indicate that depletion of ARID1A does not affect loading and initiation of RNAPII at these genes (Figure S3D). Next, we investigated the ability of ARID1A-depleted cells to properly initiate transcription, by leveraging the CDK9 inhibitor flavopiridol (an inhibitor of pTEFb kinase activity). Flavopiridol prevents transcriptional elongation within hours of treatment and leads to accumulation of initiated RNAPII that is unable to clear the proximal promoter region (Rahl et al., 2010). We treated both ARID1A-WT and ARID1A-KD cells with flavopiridol for 2 hr before performing ChIP-seq of RNAPII. Treatment of ARID1A-depleted cells with flavopiridol completely restored accumulation of RNAPII at the proximal promoter, to an extent comparable with, if not higher than, that of ARID1A-competent cells treated with flavopiridol (Figures 5A–5C). Importantly, treatment with flavopiridol also abolished the global decrease of traveling ratio determined by ARID1A depletion (Figure 5D). Collectively, these data indicate that ARID1A loss hinders accumulation of RNAPII at proximal promoters without affecting the process of initiation or the integrity of the pre-initiation complex.

Next, we assessed changes in the chromatin status of OCCC cells. First, we profiled H3K36me3, a methylation mark that is co-transcriptionally deposited and correlates with transcriptional elongation (Fong et al., 2017; Zhang et al., 2015), to determine if ARID1A may have an additional impact in regulating release of paused RNAPII and downstream elongation. In fact, levels of H3K36me3 are known to mirror productive elongation of RNAPII throughout the gene body. Our analysis suggests that loss of ARID1A does not impair elongation, because average levels of H3K36me3 moderately increase across the gene body of active loci in ARID1A-KD (Figures S4A and S4B). Next, to determine if ARID1A depletion changes the chromatin status of its target genes and enhancers, we performed ChIP-seq of H3K27ac. Previous reports suggest that ARID1A depletion leads to loss of acetylation at enhancers (Alver et al., 2017; Kelso et al., 2017; Mathur et al., 2017). Consistent with these findings, we observed moderate but significant decrease of H3K27ac across enhancers (p < 2.2 × 10−16; Figures S4C and S4D). However, the sharpest decrease of acetylation was found at weak enhancers, which may lose acetylation as a mere consequence of decreased accessibility (Figures 2 and S4C). Furthermore, enhancers associated to the top 200 downregulated genes upon ARID1A depletion did not display changes in acetylation (Figure S2C). We next examined the acetylation status of promoters. Notably, ARID1A target promoters displayed a significant increase of H3K27ac at their +1 nucleosome (Figures S4E and S4F), which was validated by ChIP-qPCR at select loci upon depletion of ARID1A with a different shRNA sequence (Figure S4G).

Collectively, our analysis of the chromatin status further supports the hypothesis that ARID1A directly favors accumulation of paused RNAPII, rather than supporting transcription indirectly through deposition of activating chromatin marks.

**ARID1B Restores Paused RNAPII and Transcription in Ovarian Cancer Cells**

To investigate how ovarian epithelial cells cope with the global defects in transcription that follow ARID1A loss, we examined its paralog, ARID1B. ARID1B represents a specific vulnerability in ARID1A-mutated tumors, and their synthetic lethality has been previously described (Helming et al., 2014). We surmised that OCCC cells could leverage ARID1B to recover from the widespread transcriptional repression that follows loss of ARID1A. Hence, we induced depletion of ARID1A through an extended time-course experiment (days 0–7 of doxycycline [DOX]). Ovarian epithelial cells express low levels of ARID1B and gradually increase its expression in response to the KD of ARID1A (Figures 6A and S4H). Furthermore, ChIP-qPCR showed a gradual and steady increase of ARID1B occupancy at the TSSs
of genes typically bound by ARID1A, peaking at day 7 (Figure 6B). The peak of ARID1B occupancy (day 7) correlated with rescue of promoter-proximal paused-RNAPII (Figure 6B) and restoration of mRNA levels (Figure 6B). Collectively, these data suggest that upregulation of ARID1B compensates for the loss of ARID1A and re-establishes physiological levels of paused RNAPII. To validate the function of ARID1B as a regulator of transcription and RNAPII dynamics in OCCC, we performed a complementation experiment in WT RMG-1 cells. We depleted ARID1A while overexpressing exogenous ARID1B (Figure 6C) and determined that ARID1B is sufficient to rescue paused RNAPII and restore gene expression at a set of candidate genes (Figure 6C). These data suggest that ARID1A loss induces ARID1B upregulation in ovarian epithelial cells, which re-enforces paused polymerase genome-wide. However, the oncogenic mechanism triggered by ARID1A loss remains unclear. Genetic and biochemical evidences suggest that ARID1A and ARID1B are only partly redundant. We hypothesize that a set of genes may be strictly dependent on ARID1A and escape rescue by ARID1B. To determine differences between ARID1A- and ARID1B-regulated transcriptions, we performed RNA-seq on days 0 and 7 of DOX induction. Notably, the global repression of transcription determined by ARID1A depletion (day 3; Figure 2E) is relieved by ARID1B upregulation (day 7; Figure 6D). Nonetheless, we detected 1,817 genes differentially expressed between days 0 and 7 (DESeq2; 906 upregulated, 911 downregulated; FDR < 5%; Table S2). These transcripts represent potential ARID1A-dependent genes that may drive tumorigenesis in OCCC. Ingenuity Pathway Analysis revealed that dysregulated genes are significantly enriched in the following functional categories: cancer (p = 2.48 × 10^{-7}), DNA repair (p = 2.94 × 10^{-12}), cell cycle (p = 1.06 × 10^{-14}), and cell death and survival (p = 4.45 × 10^{-12}). For instance, the PARP-binding protein PARI prevents genomic instability by inhibiting inappropriate homologous recombination (Moldovan et al., 2012), and its expression is not rescued by ARID1B. Similarly, the centrosomal proteins CENPE and CENPK are persistently downregulated at day 7 of DOX induction and may underlie defects in chromosome alignment and segregation that lead to genomic instability.

Last, TP53 (p = 6.32 × 10^{-4}) and the estrogen receptor ESR1 (p = 1.96 × 10^{-9}) were among the top enriched upstream regulators of ARID1A-dependent genes. Notably, p53 is rarely mutated in ARID1A-deficient ovarian tumors, suggesting that the impairment of its function may occur transcriptionally at its genomic targets. Overall, these data suggest that a subset of genes, strictly dependent on ARID1A for their fine-tuned expression, may underlie cell transformation in ovarian epithelial cells.

**DISCUSSION**

**A Model for ARID1A Function in Normal and Malignant Cells**

Here we present a comprehensive genome-wide profiling of ARID1A activity in ovarian epithelial cells and propose a mechanistic model that explains both the physiological role of ARID1A in transcription and the oncogenic consequences of its genetic ablation, which occurs in nearly 60% of clear cell carcinomas (OCCC). In summary, ARID1A targets active promoters
and ensures that physiological levels of paused RNAPII accumulate after initiation. When ARID1A is the predominant ARID component of the BAF complex, its loss results in widespread transcriptional repression due to ineffective pausing. Upregulation of ARID1B relieves the transcriptional crunch, but a set of genes that are strictly dependent on ARID1A do not recover.

Figure 6. Upregulation of ARID1B Restores RNAPII Pausing and Transcription at Most Genes

(A) Immunoblot analysis of ARID1A and ARID1B during a time course of doxycycline (days 3, 5, and 7). ARID1B is increasingly upregulated in response to ARID1A depletion, peaking at day 7. Quantitative immunoblots are shown in Figure S4H.

(B) ChIP-qPCR analysis of RNAPII and ARID1B binding for a panel of ARID1A target genes at 0 hr (CTRL), 3 days, and 7 days of treatment with doxycycline. ARID1B gradually replaces ARID1A at the TSS, while RNAPII proximal-promoter occupancy is restored. Similarly, qRT-PCR for the same genes (fold enrichment compared with CTRL, normalized to 18 S rRNA) suggests that transcription resumes its physiological level at day 7, concurrent with highest occupancy of ARID1B. Error bars represent SD.

(C) Parental RMG-1 cells were transduced with shRNAs against ARID1A and a full-length ARID1B vector. ChIP-qPCR for RNAPII and qRT-PCR at three candidate loci show that ARID1A-dependent transcriptional regulation is complemented (p < 0.0111) by ARID1B (qRT-PCR fold change compared with shLUC, normalized by 18 S rRNA). Error bars represent SD.

(D) Correlation plot (spike-in normalized RNA-seq) for all expressed coding genes shows that global transcriptional downregulation provoked by ARID1A depletion (day 3 of DOX induction, green) is largely rescued by ARID1B upregulation (day 7 of DOX induction, purple). Genes that are persistently dysregulated at day 7 have been functionally evaluated (Ingenuity Pathway Analysis [IPA]) and may represent ARID1A-specific targets.
Distribution of ARID1A and Its Role in Remodeling

First, we provide an accurate profiling of ARID1A genome-wide in ovarian cells. We find that ARID1A targets most active promoters and enhancers and is required to sustain transcription of both protein-coding genes and noncoding enhancer RNAs. Notably, we observe a very distinct profile of ARID1A at promoters (nucleosome associated) and enhancers (centered on the NDR). This finding may reflect different functions of the SWI/SNF complex or underlie a different chromatin organization at these regulatory regions and is in accordance with previous profiling of the catalytic BRG1 subunits in melanoma cells (Laurette et al., 2015).

Previous work suggested a preference for ARID1A and SWI/SNF toward binding enhancer elements (Alver et al., 2017; Mathur et al., 2017; Kelso et al., 2017), but other studies revealed a more diffuse and broad distribution of the ARID1 proteins and other SWI/SNF subunits, similar to our observations (Laurette et al., 2015; Morris et al., 2014; Raab et al., 2015). Although the SWI/SNF complex is a universal chromatin remodeler in eukaryotes, the role for ARID1A in maintaining accessibility at promoters and enhancers in OCCC is marginal, suggesting that its role in nucleosome positioning may be negligible. Previous studies revealed a role for SWI/SNF components (especially BRG1) as modulators of accessibility at enhancers (Bao et al., 2015; Laurette et al., 2015), but we find that the contribution of ARID1A is limited to the weakest enhancer regions, displaying the lowest enrichment for ARID1A by ChIP-seq. Similarly, promoter accessibility is regulated by the BRG1 and SNF subunits of SWI/SNF (Tolstorukov et al., 2013) but is not affected by ARID1A according to our ATAC-seq analysis. Coherently, a previous DNase sequencing (DNase-seq) study suggested that very limited changes in hypersensitive sites occurred in ARID1A knockout (KO) ovarian cells (Kim et al., 2016), however, micrococcal nuclease sequencing (MNase-seq) studies in mouse embryonic stem cells yielded different results (Lei et al., 2015). It is possible that ARID1A contributes to nucleosome remodeling in embryonic stem cells but not in somatic cells. In our system, the most active promoters and enhancers display very limited changes in accessibility after 3 days of ARID1A depletion, despite massive transcriptional repression (assessed by RNA-seq, GRO-seq, and CHROM-RNA-seq). It must also be noted that accessibility may not always reflect changes in nucleosome positioning (Mueller et al., 2017). Our studies may be extended in the future using MNase digestion followed by H3 ChIP, before excluding any role for ARID1A as adjuvant of nucleosome remodeling in somatic cells.

A Role for ARID1A in Regulation of RNAPII

Our data suggest that ARID1A endows the SWI/SNF (BAF) complex with regulatory activity toward RNAPII pausing, determining the physiological rate of transcription for a large fraction of active genes in ovarian epithelial cells. The mechanism of RNAPII pausing has emerged, over the past decade, as a fundamental regulatory step of transcription in higher eukaryotes (Adelman and Lis, 2012). It must be noted that pausing of RNAPII is not simply a brake for transcription, as paused RNAPII in embryonic stem cells is required to maintain cell homeostasis and properly respond to environmental stimuli and cues (Henriques et al., 2013; Shao and Zeitlinger, 2017; Williams et al., 2015). In the absence of properly functioning pausing machinery, such as upon depletion of NELF, thousands of genes were shown to respond with transcriptional attenuation (Williams et al., 2015). In fact, polymerase that fails to pause appears not to be released into productive elongation but is lost from the body of the gene (Gilchrist et al., 2010). These findings are consistent with our observation that failure to accumulate paused RNAPII at the proximal promoter after ARID1A depletion results in general attenuation of transcription. Importantly, we show that protein levels of RNAPII, NELF, and SPT5 are not affected by depletion of ARID1A, and the elongation block induced with flavopiridol fully restores proximal paused RNAPII, thereby excluding defects in the initiation step.

Pausing of RNAPII is governed by protein complexes such as NELF, PAF, Mediator, and Integrator (Chen et al., 2015; Galbraith et al., 2013; Gardini et al., 2014; Goldman et al., 2016; Patel et al., 2013), as well as chromatin structure (Adelman and Lis, 2012). Interestingly, depletion of ARID1A elicited a marked increase of H3K27 acetylation at the +1 nucleosome of most protein-coding genes. Recently, an association between acetylation levels of the +1 nucleosome and the pausing activity of RNAPII was proposed (Boija et al., 2017). In particular, the authors suggested that fine-tuned levels of H3K27ac on the first nucleosome downstream of the TSS are required to maintain a proper balance between paused RNAPII and productive elongation. We speculate that ARID1A may be required for the correct acetylation balance, with increased acetylation resulting from ARID1A loss preventing accumulation of paused RNAPII at the proximal promoter.

Role of ARID1A at Enhancers

ARID1A has been recently proposed to control enhancer acetylation via recruitment of the EP300 acetyltransferase (Alver et al., 2017; Mathur et al., 2017). Our data, however, indicate that the highest occupancy of ARID1A anticorrelates with changes in H3K27ac levels, which are more likely to occur at weak enhancers. Weak enhancers likely constitute inactive or poised enhancers that are not actively regulating growth and homeostasis of RMG-1 cells. We also find that the most active, ARID1A-bound, regulatory elements (strong enhancers) are significantly enriched in AP-1 binding sites. In the context of ovarian epithelial cells, the AP-1 TF may sustain open conformation and acetylation of enhancers with the support of the catalytic activity of SWI/SNF (Vierbuchen et al., 2017), leaving ARID1A dispensable for such role. Nonetheless, we hypothesize that enhancer activity may be compromised in ARID1A-deficient cells, because GRO-seq data revealed widespread transcription attenuation at enhancers. GRO-seq assays detect reads from the enhancer-associated bi-directional noncoding RNAs (eRNAs) (Gardini and Shiekhattar, 2015). Several studies suggest that enhancer activity is functionally linked to transcription of eRNAs and their accumulation at chromatin (Bose et al., 2017; Gardini and Shiekhattar, 2015; Lal et al., 2015). We speculate that transcription attenuation at strong enhancers will affect their prior expression level and may pave the way to cellular transformation.
Intriguingly, enhancers may be subjected to the same RNAPII pausing mechanism enforced at protein-coding genes (Henriques et al., 2018), suggesting a similar role for ARID1A. Future research on the enhancer-promoter network of ovarian epithelial cells will clarify the precise role of ARID1A at these regulatory elements.

**Tumorigenesis Mediated by ARID1A Loss**

Finally, in this work we show that upregulation of ARID1B offsets loss of ARID1A at protein-coding genes and restores physiological transcription, establishing a mechanistic model for the synthetic lethality of ARID1A and ARID1B (Helming et al., 2014). Consistently, ARID1B was reported upregulated in mouse embryonic fibroblasts that lack one or both ARID1A alleles (Helming et al., 2014). However, genetic and biochemical evidence suggest that ARID1A and ARID1B are only partly redundant. First, the two paralogs seem to associate with different sets of TFs (Raab et al., 2015). Second, genetic mouse models suggest that the two proteins are functionally distinct (Celen et al., 2017; Goldman et al., 2016; Krosl et al., 2010; Vasileiou et al., 2015; Zhai et al., 2016). As a matter of fact, we found a subset of 1,817 genes that are not entirely rescued by ARID1B expression and remain dysregulated. Although ARID1B restores RNAPII pausing at the majority of protein-coding genes, some loci may be preferentially modulated by ARID1A and de facto trigger tumorigenesis once normal epithelial cells biallelically lose ARID1A. Importantly, a subset of permanently dysregulated genes are TP53 targets. Given that TP53 is rarely mutated in the two paralogs seem to associate with different sets of TFs (Celen et al., 2017; Goldman et al., 2016; Krosl et al., 2010; Vasileiou et al., 2015; Zhai et al., 2016). As a matter of fact, we found a subset of 1,817 genes that are not entirely rescued by ARID1B expression and remain dysregulated. Although ARID1B restores RNAPII pausing at the majority of protein-coding genes, some loci may be preferentially modulated by ARID1A and de facto trigger tumorigenesis once normal epithelial cells biallelically lose ARID1A. Importantly, a subset of permanently dysregulated genes are TP53 targets. Given that TP53 is rarely mutated in this type of tumor. Last, although our studies were conducted in ovarian cancer cells, ARID1A is the most mutated chromatin modifier in cancers of various origin. Future studies will help determine the role of impaired RNAPII pausing and partial complementation by ARID1B in additional tumour types with recurrent (>30%) ARID1A loss-of-function mutations such as hepatocellular carcinoma, uterine carcinomas, and stomach adenocarcinoma.

**EXPERIMENTAL PROCEDURES**

**Cell Lines**

RMG-1 cells were obtained from JCRB Cell Bank and maintained as described in Supplemental Experimental Procedures. RMG-1 cells were infected with lentivirus carrying the inducible tet-pLKO-shARID1A (sequence: GCCTGATC with pLKO.1-shARID1A (TRCN0000059090) and selected with 2 μg/mL puromycin for 72 hr. For ARID1B rescue experiments, a commercial plasmid was used (Supplemental Experimental Procedures). RMG-1 cells were lentivirally infected with pLVX-ARID1A vector. The pLVX empty vector was used as control.

**Western Blot**

Western blots were performed as described in Supplemental Experimental Procedures.

**Antibodies**

Antibodies used are listed in Supplemental Experimental Procedures.

**ChIP-Seq and ChIP-qPCR Sample Processing**

Samples from different conditions were processed together to prevent batch effects. For each replicate, 10 million RMG-1 cells were used to perform ChIP-seq experiments, as described in Supplemental Experimental Procedures. For flavopiridol experiments, ARID1A-WT and ARID1A-KD cells were treated with flavopiridol (2 μM) for 2 hr before being harvested. For ChIP-qPCR, on day 1, the sonicated lysate was aliquoted into single immunoprecipitations of 2.5 × 10⁶ cells each. On day 3, ChIP eluates and input were assayed by real-time qPCR.

**ChIP-Seq Analyses**

Sequences were aligned to the reference hg19, using Burrows Wheeler Alignment (BWA) tool, with the MEM algorithm (Li, 2013). Detailed settings are described in Supplemental Experimental Procedures.

We called peaks for each individual using MACS2, at 5% FDR, with default parameters (Zhang et al., 2008). Heatmaps were generated with ChASE version 1.0.11 (Younes et al., 2016) and average profiles with seqMINER version 1.3.4 (Zhan and Liu, 2015).

**ATAC-Seq Sample Processing**

Two independent biological replicates of 50,000 cells per condition were processed as described in the original ATAC-seq protocol paper ( Buenrostro et al., 2013).

**ATAC-Seq Analyses**

ATAC-seq data were processed using the same pipeline described for ChiP-seq, with one modification: all mapped reads were offset by +4 bp for the forward strand and −5 bp for the reverse strand. After peak calling, peaks replicated in both of the two ARID1A-WT samples were used for downstream analyses. edgeR (Robinson et al., 2010) was used to test for regions with differential accessibility between ARID1A-KD and ARID1A-WT using a likelihood ratio test in a generalized linear model with the following parameters: glmLRT(tilt,coef = 2), on the basis of quantile-normalized read counts.

**RNA-Seq Sample Processing**

Total RNA-seq was performed on two independent biological replicates per condition. Samples from different conditions were processed together to prevent batch effects. For each sample, 10 million cells were used to extract total RNA and produce RNA-seq libraries (Supplemental Experimental Procedures).

**RNA-Seq Analyses**

 Reads were aligned to hg19 using STAR version 2.5 ( Dobin et al., 2013). Detailed parameters are in Supplemental Experimental Procedures. We filtered barn files on the basis of alignment quality (q = 10) using SAMtools version 0.1.19 (Li et al., 2009). We used the latest annotations obtained from GENCODE to build reference indexes for the STAR alignment. FeatureCounts (Liao et al., 2014) was used to count reads mapping to each gene. RSEM (Li and Dewey, 2011) was used to obtain FPKM (fragments per kilobase of exon per million fragments mapped). We analyzed differential gene expression using DESeq2 (Love et al., 2014), with the following model: design = ~condition, where condition indicates either ARID1A-WT or ARID1A-KD.

**GRO-Seq**

GRO-seq was performed on two independent biological replicates per condition as previously described (Gardini et al., 2014). GRO-seq data were analyzed using the same pipeline described for total RNA-seq.

**Other Statistical and Genomic Analyses**

All statistical analyses and plots were performed using R version 3.3.1 and ggplot2 (F Core Team, 2016; Wickham, 2009). BEDTools version 2.27.1 was used for genomic analyses (Quinlan and Hall, 2010). For testing ratio analysis, TSS regions were extended ~50–~300 bp. Other thresholds were tested (up to +500 bp downstream), with no differences in the results. Pathway analysis was performed with Ingenuity Pathway Analysis Suite (QIAGEN, https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). Motif analyses were performed with Meme-Chip (Bailey et al., 2009).
DATA AND SOFTWARE AVAILABILITY

The accession number for the sequencing data reported in this paper is GEO: GSE104545.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, two tables, and one data file and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.05.097.

ACKNOWLEDGMENTS

This work was supported by grant R01 CA202919 (R.Z.) from the NIH and funding from the W. W. Smith Charitable Trust to A.G. S.A.W. is supported by NIH training grant T32-GM071339. We thank Kavitha Sarma, Christopher D. Brown, Minal Caliskan, and Benjamin Bitler for insightful discussions. Drosophila S2 cells used for spike-in were kindly provided by Maya Capelson (University of Pennsylvania). All genome-wide data were generated and processed at the Wistar Institute with outstanding technical support from all members of the Genomics Facility, especially Dr. Shashi Bala (NIH-National Cancer Institute [NCI]-funded Cancer Center Support Grant [CCSG] P30-CA010815).

AUTHOR CONTRIBUTIONS

M.T., E.B., and A.G. designed the experiments. M.T., E.B., A.P., S.W., S.A.W., T.A.O., and S.L. performed experiments. M.T. performed all computational and statistical analyses. R.Z. provided valuable conceptual advice and support throughout the entire study. M.T. and A.G. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 20, 2017
Revised: March 20, 2018
Accepted: May 30, 2018
Published: June 26, 2018

REFERENCES


