Published in final edited form as:

Biochem Biophys Res Commun. 2017 June 03; 487(3): 573–579. doi:10.1016/j.bbrc.2017.04.092.

A novel strategy to dissect endogenous gene transcriptional regulation in live cells

Wenqing Yang^{1,2,3}, Siliang Zhang^{2,3,*}, Yi Zhang¹, and Xin Huang^{2,3}

¹Department of Obstetrics and Gynecology, Xiangya Hospital of Central South University, Changsha, Hunan Province 410005, China

²Magee-Womens Research Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA

³Women's Cancer Research Center, University of Pittsburgh Cancer Institute, Pittsburgh, PA 15213, USA

Abstract

Gene transcription is a central tenet of biology, traditionally measured by RT-PCR, microarray, or more recently, RNA sequencing. However, these measurements only provide a snapshot of the state of gene transcription and only represent an overall readout of complex transcriptional networks that regulate gene expression. In this report, we describe a novel strategy to dissect endogenous gene transcription regulation in live cells by knocking in a reporter gene, *EGFP*, under the control of the endogenous gene promoter, using the *ARID1A* gene as an example. The *ARID1A* gene, encoding a subunit of the ATP-dependent chromatin remodeling complex SNF/SWI, has recently been identified as a tumor suppressor in multiple cancers. Despite studies that elucidate the mechanism of *ARID1A*'s tumor suppressor function, little is known of the genes/events that regulate *ARID1A* expression. Using the HEK293 cells as a model, we discovered novel aspects of *ARID1A* transcription regulation in response to cell cycle progression, DNA damage, and microRNAs, exemplifying the potential of our strategy in providing new insight to the mechanism of gene transcription regulation. This strategy can be generalized to essentially any gene of interest, making it a powerful tool for the study of gene expression heterogeneity, especially in cancer cells, and a robust readout for high-throughput screening of agents that modulate gene transcription.

Keywords

CRISPR-Cas9; *ARID1A*; Transcription Regulation

Corresponding Authors: Xin Huang, Ph.D., Tel: 412-641-3803, Fax: 412-641-6156, huangx2@upmc.edu, Yi Zhang, M.D., Ph.D., Tel: 86-0731-89753717, zhangyi20080903@163.com.

*Present Address: Department of Radiation Oncology, The Third Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang Province 150081, China

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Gene transcription, the process of decoding genetic information from DNA into RNA, is critical for maintaining proper cellular function. Cells have evolved a number of diverse mechanisms to delicately regulate this process; these include transcriptional factors, promoters, enhancers, epigenetic modifications on DNA and histones, and chromatin structures [1]. To study transcriptional regulation of a gene, typically, a reporter gene, such as the genes encoding LacZ, green fluorescent protein (GFP), or luciferase, is placed downstream of the promoter of the gene of interest, and the signal from the reporter gene, either color, fluorescence, or bioluminescence, serves as the readout for indirect measurement of transcriptional regulators [2]. However, the reporter gene approach has several major drawbacks: 1. The need to obtain a complete promoter, which is difficult to define, in the first place, for most genes; 2. The effects of enhancers that are far from the promoter or of epigenetic factors, including chromatin structure, on gene expression cannot be readily assessed; 3. The dynamics of gene expression change under various conditions and are very difficult to monitor; 4. Exogenous plasmids need to be transfected into cells. Therefore, a better approach to monitoring gene expression is highly desirable.

SWItch/sucrose non-fermentable (SWI/SNF) is an ATP-dependent chromatin remodeling complex that disrupts DNA-histone interactions [3]. BAF250a, a protein encoded by the AT-rich interactive domain 1A (*ARID1A*) gene, is one of the accessory subunits of the SWI/SNF complex [4]. Recent data from the cancer genome atlas project demonstrated that *ARID1A* is one of the most frequently mutated genes in human cancers [5], suggesting a universal tumor suppressor function of *ARID1A*. Currently, essentially all studies have focused on elucidating the downstream pathways regulated by *ARID1A*, with tremendous success [6,7,8,9]. However, how *ARID1A* is regulated remains unknown. Therefore, identifying factors that regulate *ARID1A* gene expression is essential for dissecting this complex transcriptional regulatory network and for better understanding of its tumor suppressor function.

In the study reported here, we developed a strategy to monitor endogenous gene transcription regulation by knocking in the open reading frame (ORF) of enhanced GFP (EGFP) into the endogenous promoter of gene of interest using the clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) protein 9 (CRISPR-Cas9). Using *ARID1A* as an example, we monitored real-time *ARID1A* transcription in live cells throughout cell cycle progression, under the treatment of a DNA damage agent, and under regulation of microRNAs (miRNAs). Our data clearly demonstrate the potential of this strategy in tracking endogenous gene expression, which could be a powerful tool for examining the regulatory mechanisms of gene expression and for high-throughput screening of agents that modulate gene transcription.

Materials and Methods

General methods

All methods described in this manuscript were carried out in accordance with the approved guidelines of the University of Pittsburgh.

Plasmid construction and validation

The DNA oligos encoding the single guide RNA (sgRNA) targeting *ARID1A* exon 2 were annealed and cloned into the PX459 plasmids as previously reported [10]. The donor plasmids used for *EGFP* (pUC19-*ARID1A* 5' Arm-*EGFP*-*ARID1A* 3' Arm, Supplementary Dataset 1) knockin at *ARID1A* exon 2 were constructed using the pUC19 plasmid backbone. Specifically, Integrated DNA Technologies (IDT, Coralville, IA) synthesized 1-kb DNA fragments homologous to the 5' and 3' of the genomic site of the reporter gene integration. The *EGFP*ORF was amplified from the pEGFP-N1 plasmid (Clontech, Mountain View, CA). The pUC19 plasmid was digested with *Sma*I (New England BioLabs, Ipswich, MA). The 5' homologous arm, *EGFP*ORF, 3' homologous arm, and *Sma*I-digested pUC19 were then ligated by the Gibson Assembly Cloning Kit (New England BioLabs).

Results

Establishing a model that monitors endogenous *ARID1A* gene transcription

Cells are dynamic systems that are constantly responsive to endogenous and environmental cues; this is achieved largely through regulating the expression of genes that control these responses. However, currently, it remains a challenge to dissect the complex regulatory network of gene transcription in live cells directly. We reasoned that, when a reporter gene, such as *EGFP*, is placed under the control of endogenous promoter, the transcriptional activity of target gene would be reflected by the fluorescence intensity, which could be readily measured and tracked longitudinally in live cells, allowing otherwise unachievable real-time monitoring of endogenous gene expression.

Using *ARID1A* as an example (Fig. 1A), HEK293 cells were co-transfected with an *EGFP* donor plasmid that introduces the *EGFP*ORF fused in-frame with *ARID1A* and the PX459 plasmid with a sgRNA against *ARID1A* exon 2. Individual EGFP-positive cells were sorted into 96-well plates by fluorescence-activated cell sorting (FACS) and cultured for 3 weeks to isolate EGFP-positive clones. However, an unexpected finding was that a large number of the original EGFP-positive single clones became EGFP-negative after 3 weeks of cell culture, suggesting that these cells do not harbor *EGFP* knockin. To identify the source of false fluorescence, we first collected the sorted EGFP-positive cells as a population. Surprisingly, following a week of cell culture, most of the EGFP-positive cells lost their EGFP expression (Fig. 1B, middle panel). These cells were then sorted again into 96-well plates for clonal expansion. However, after the second round of sorting, every sorted EGFP-positive clones remained positive (Fig. 1B, lower panel). Our data demonstrate that the fluorescence from the majority of EGFP-positive cells after the first sorting may actually come from the donor plasmid. Therefore, double FACS sorting is essential to ensure the selection of true *EGFP* knockin clones.

We randomly selected six EGFP-positive clones to validate correct *EGFP* insertion. PCR primers that amplify both the wild type *ARID1A* allele and the *EGFP*-containing allele indicated that all six clones had *EGFP* knockin at the expected locus (Supplementary Figure 1). The PCR product was subjected to DNA sequencing and the results demonstrated a

precise insertion of the *EGFPORF* at the beginning of *ARID1A* exon2 in all six clones (Fig. 1C).

Tracking endogenous *ARID1A* transcription during cell cycle progression

The *ARID1A* gene is known to regulate many aspects of cellular function [11,12,13]. However, how *ARID1A* expression itself is regulated remains largely elusive. As a tumor suppressor, *ARID1A* has been demonstrated to be an essential gene for cell cycle arrest [11,12,13], but whether cell cycle progression also influences the transcription of the *ARID1A* gene remains unknown. The establishment of *EGFP*-knockin cell lines provides an ideal model to address this question.

The *EGFP*-labeled cells were synchronized by serum deprivation at the G0/G1 phase, by aphidicolin at the S phase, and by nocodazole at the G2/M phase (Fig. 2A). The fact that the majority of cells only have the *EGFP* gene inserted at one of the *ARID1A* alleles worked in our favor, allowing us to use the wild type *ARID1A* allele as a control to examine the robustness of the *EGFP* reporter in reflecting endogenous *ARID1A* transcription. Flow cytometry was used to profile the cell cycle in unsynchronized cells and in cells treated with aphidicolin or nocodazole. Interestingly, the *EGFP* intensity was found to fluctuate with cell cycle progression (Fig. 2B), with maximal and minimal *EGFP* expression in the M and G0 phases, respectively. We next examined whether the *EGFP* intensity fluctuation at different phases of the cell cycle reflects corresponding transcriptional activity at the *ARID1A* locus by RT-qPCR.

Three sets of RT-qPCR primers were used, one specific for the *EGFP* gene, one for *ARID1A* exons 1 and 2 that is specific to the wild type *ARID1A* gene, and one for the last two exons of *ARID1A*, exons 18 and 19, which measures the transcription from both wild type and *EGFP* knockin alleles. Remarkably, the RT-qPCR data from all three sets of primers are highly consistent and change of *EGFP* expression during the cell cycle closely mirrored the fluctuation of endogenous *ARID1A* expression (Fig. 2C), suggesting that transcription of the *EGFP* reporter faithfully reflected that of the endogenous *ARID1A* gene. These results are consistent with data from flow cytometry experiment, indicating that our strategy for tracking endogenous *ARID1A* transcription is reliable.

A unique advantage of placing *EGFP* under the control of endogenous *ARID1A* promoter is that it allows real-time monitoring of *ARID1A* transcription in live cells that responds to regulatory or environmental perturbations. To demonstrate this, we performed live-cell fluorescence imaging using the *EGFP* knockin cells at different phases of cell cycle. Remarkably, the intensity of *EGFP* in live cells faithfully reflects the transcriptional change of the *ARID1A* gene during the cell cycle (Fig. 2D, lower panel). As a comparison, we also imaged fixed *EGFP* knockin cells that were cultured on chamber slides. It is clear that the *EGFP* intensity fluctuates during cell cycle progression (Fig. 2D, upper panel), consistent with the results from flow cytometry, RT-qPCR, and live-cell imaging.

Taken together, our data clearly demonstrate the feasibility and reliability of tracking endogenous gene transcription with live-cell imaging by placing a fluorescent gene downstream of an endogenous gene promoter.

Tracking endogenous *ARID1A* transcription during DNA damage response

Defects in DNA damage repair are common in cancer cells [14]. Recently, it has been shown that *ARID1A* deficiency can impair the DNA damage checkpoint [9], thus providing direct evidence that supports its role as a tumor suppressor. Many other tumor suppressor genes, especially the ones in the *TP53* pathway, are markedly upregulated upon DNA damage [15]. Thus, we examined whether our strategy could be used as a quick and convenient approach to assess regulation of *ARID1A* transcription upon DNA damage.

The EGFP-labeled HEK293 cells were treated with 6 μ M camptothecin (CPT), an S-phase specific DNA topoisomerase I inhibitor causing DNA damage [16], for 0, 1, 4, and 8 hrs in S-phase cells arrested by aphidicolin. We examined *ARID1A* transcription by measuring cellular fluorescence intensity with flow cytometry. CPT induced a robust DNA damage response after only 0.5 hr treatment, indicated by TP53 serine 15 phosphorylation and induction of CDKN1A (p21) (Fig. 3A). Interestingly, EGFP intensity started to decrease at 0.5 hr, with a marked drop at 4 hr, measured by flow cytometry (Fig. 3B), which is also reflected by live-cell imaging (Fig. 3C). Similarly, when we used the aforementioned three pairs of RT-qPCR primers to measure *EGFP* and *ARID1A* transcription, the result was consistent with flow cytometry and live-cell imaging data (Fig. 3D). Again, the transcription kinetics of the reporter *EGFP* following DNA damage closely mimicked that of endogenous *ARID1A*. Interestingly, ARID1A and EGFP protein levels followed the same trend (Fig. 3A), suggesting that transcription regulation is likely to be the main mechanism regulating *ARID1A* response to DNA damage.

miRNA regulation of the *ARID1A* gene expression

One theoretical advantage of generating the *ARID1A* allele using our strategy is the ability to embed the *EGFP* reporter within an otherwise full-length transcript, thereby ensuring that the *EGFP* reporter remains under all the transcriptional and posttranscriptional regulatory control elements of the endogenous locus. One such example is miRNA regulation of gene expression. Despite a highly mutated tumor suppressor gene, surprisingly, only a few studies have been published investigating miRNA regulation of *ARID1A* [17,18,19]. Therefore, the EGFP-labeled cell line serves as a perfect model in which to examine the effect of miRNA on *ARID1A* expression.

To date, only three miRNAs, miR-31, miR-221, and miR-222, have been shown to regulate *ARID1A*. However, when we tried to transfect constructs that overexpress miR-221 and miR-222, in contrast to the published report [17], we failed to detect any change of ARID1A, measured by the intensity of cellular fluorescence, Western blotting, and RT-qPCR. By searching computational prediction of the miRNAs that regulate *ARID1A* using TargetScan (www.targetscan.org) software [20], we selected two miRNAs, miR-144 and miR-9 (Supplementary Figure 2), for validation using the EGFP-labeled HEK293 cells.

Transfection of the EGFP-labeled cells with synthetic miR-144 or miR-9 mimic reduced EGFP expression, as measured by fluorescence microscopy (Fig. 4A) or flow cytometry analysis (Fig. 4B). Furthermore, Western blotting data demonstrated that the reduction of ARID1A-EGFP fusion protein mirrored that of endogenous ARID1A (Fig. 4C). Because

miRNAs can regulate their target genes by inhibiting protein translation or through mRNA degradation [21,22], we examined the mechanism of miR-144 and miR-9 regulation of *ARID1A*. By measuring *ARID1A* mRNA levels following transfection of miR-144 or miR-9 mimics by RT-qPCR, it is clear that miRNA-mediated mRNA degradation is likely to be the main mechanism of ARID1A downregulation (Fig. 4D). Collectively, these results indicate that the EGFP reporter remains under appropriate posttranscriptional 3' UTR regulation and our strategy ought to be generalizable for high-throughput screening of miRNAs that regulate any gene of interest.

Discussion

Using the *ARID1A* gene as an example, we have developed a strategy for real-time monitoring of endogenous gene transcription in live cells. By placing an *EGFP* gene under the transcriptional control of the *ARID1A* gene promoter, we demonstrated the potential application of this strategy by tracking the response of endogenous *ARID1A* transcription to cell cycle progression, DNA damage, and microRNA regulation, using cellular EGFP intensity as a simple readout in live cells.

Transcription is the central dogma of biology. Recently, nuclear-localized RNA-targeting Cas9 tagged with GFP was used to track RNA expression in the presence of mRNA-targeting sgRNAs [23], providing a new path to study gene transcription in live cells. Approaching this question from a different angle, we reasoned that placing an *EGFPORF* under the control of endogenous gene promoter would allow us to identify factors that regulate gene transcription by simply monitoring the cellular fluorescence signal. As demonstrated by *ARID1A* as an example, this strategy can be used to dissect transcriptional regulation of any gene of interest, especially in high-throughput screening assays to identify transcriptional and posttranscriptional regulatory factors.

A unique advantage of our strategy is that gene transcriptional activity can be monitored in real time in live cells at the single-cell level, allowing detection of the heterogeneity of gene transcription responding to regulatory signals. Because of genome instability, cancer cells are genetically highly heterogeneous [24], which has a profound impact on almost every aspects of clinical behavior of tumors [25,26], especially on response to therapy [27]. This is clearly demonstrated by the large variation of cell-to-cell EGFP intensity, a simple readout for endogenous *ARID1A* transcriptional activity in individual cells, following CPT treatment (Fig. 3C). Thus, our strategy should be highly valuable in monitoring the response and elucidating the mechanisms of resistance of cancer cells to therapy.

ARID1A is known to regulate cell cycle progression and DNA damage response [9,11,12], but how *ARID1A* transcription is regulated during these processes remains largely unknown. Our study reveals an unexpected role of cell cycle and DNA damage in regulating *ARID1A* transcription. Interestingly, kinetics of ARID1A expression during the cell cycle was reported in mouse embryos in which the cells have highest *ARID1A* expression at G0/G1, but lowest at S and G2/M [28]. The discrepancy between the published report and our results (Fig. 2) may be due to the cell lines used. The HEK293 cells used in our study are immortalized cells that may behave differently from normal mouse preosteoblasts used in

the published study [28] during cell cycle, but how these responses are related to *ARID1A*'s role as a tumor suppressor remains to be further studied.

It is well known that miRNAs are involved in almost every aspect of cell biology [29]. However, due to the short “seed region” required for miRNA regulation of its target genes, identifying miRNA targets has remained a challenge [30]. Our strategy provides a solution to this challenge, in which miRNA regulation of a target gene can be assessed in an endogenous setting. The tests with miR-9 and miR-144 demonstrated the robustness and power of this strategy. In addition, although miR-144 has four predicted target sites in the 3' UTR of *ARID1A* compared to only one for miR-9 (Supplementary Figure 2), it was less powerful than miR-9 in regulating *ARID1A* (Fig. 4D), raising an interesting question on what are the determining factors that govern miRNA regulation of target genes. Thus, our study clearly demonstrates that cellular EGFP intensity can be used as a simple readout for a high-throughput genome-wide miRNA screening assay to identify miRNAs or potentially other genetic/epigenetic factors that regulate the transcription/translation of any gene of interest.

In summary, our data demonstrate that the strategy we have detailed here was successful in tracking endogenous *ARID1A* transcription in real time in live cells. We expect this strategy to be extremely useful for discovering transcriptional regulators of genes of interest and highly valuable for developing assays to identify therapeutic interventions that can modulate the transcriptional activity of an intended target gene.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the NCI-funded RPCI-UPCI Ovarian Cancer(grant number SPOR – P50 CA159981), the American Cancer Society (ACS) (grant number RSG-12-188-01-RMC), and the Liz Tilberis Scholars Award (grant number OCRF 258940)

References

1. Lee, Tong I., Young, Richard A. Transcriptional Regulation and Its Misregulation in Disease. *Cell*. 2013; 152:1237–1251. [PubMed: 23498934]
2. Massoud TF, Gambhir SS. Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes & Development*. 2003; 17:545–580. [PubMed: 12629038]
3. Euskirchen G, Auerbach RK, Snyder M. SWI/SNF Chromatin-remodeling Factors: Multiscale Analyses and Diverse Functions. *Journal of Biological Chemistry*. 2012; 287:30897–30905. [PubMed: 22952240]
4. Wilson BG, Roberts CWM. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer*. 2011; 11:481–492. [PubMed: 21654818]
5. Wu JN, Roberts CWM. ARID1A Mutations in Cancer: Another Epigenetic Tumor Suppressor? *Cancer Discovery*. 2013; 3:35–43. [PubMed: 23208470]
6. Chandler RL, Damrauer JS, Raab JR, et al. Coexistent ARID1A–PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. *Nat Commun*. 2015; 6:6118. [PubMed: 25625625]

7. Guan B, Gao M, Wu CH, et al. Functional Analysis of In-frame Indel ARID1A Mutations Reveals New Regulatory Mechanisms of Its Tumor Suppressor Functions. *Neoplasia*. 2012; 14:986–993. [PubMed: 23097632]
8. Watanabe R, Ui A, Kanno S-I, Ogiwara H, et al. SWI/SNF Factors Required for Cellular Resistance to DNA Damage Include ARID1A and ARID1B and Show Interdependent Protein Stability. *Cancer Research*. 2014; 74:2465–2475. [PubMed: 24788099]
9. Shen J, Peng Y, Wei L, et al. ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. *Cancer Discovery*. 2015; 5:752–767. [PubMed: 26069190]
10. Ran FA, Hsu PD, Wright J, et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protocols*. 2013; 8:2281–2308. [PubMed: 24157548]
11. Nagl NG, Patsialou A, Haines DS, et al. The p270 (ARID1A/SMARCF1) Subunit of Mammalian SWI/SNF-Related Complexes Is Essential for Normal Cell Cycle Arrest. *Cancer Research*. 2005; 65:9236–9244. [PubMed: 16230384]
12. Nagl NG, Zweitzig DR, Thimmapaya B, et al. The c-myc Gene Is a Direct Target of Mammalian SWI/SNF-Related Complexes during Differentiation-Associated Cell Cycle Arrest. *Cancer Research*. 2006; 66:1289–1293. [PubMed: 16452181]
13. Zeng Y, Liu Z, Yang J, et al. ARID1A is a tumour suppressor and inhibits glioma cell proliferation via the PI3K pathway. *Head & Neck Oncology*. 2013; 5:6.
14. Hoeijmakers JHJ. DNA Damage, Aging, and Cancer. *New England Journal of Medicine*. 2009; 361:1475–1485. [PubMed: 19812404]
15. Riley T, Sontag E, Chen P, Levine A. Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol*. 2008; 9:402–412. [PubMed: 18431400]
16. Del Bino G, Lassota P, Darzynkiewicz Z. The S-phase cytotoxicity of camptothecin. *Experimental Cell Research*. 1991; 193:27–35. [PubMed: 1995300]
17. Yang Y, Zhao X, Li HX. MiR-221 and miR-222 simultaneously target ARID1A and enhance proliferation and invasion of cervical cancer cells. *European Review for Medical and Pharmacological Sciences*. 2016; 20:1509–1515. [PubMed: 27160122]
18. Lu WC, Liu CJ, Tu HF, et al. miR-31 targets ARID1A and enhances the oncogenicity and stemness of head and neck squamous cell carcinoma. *Oncotarget*. 2016; 7:57254–57267. [PubMed: 27528032]
19. Wang N, Zhou Y, Zheng L, Li H. MiR-31 is an independent prognostic factor and functions as an oncomir in cervical cancer via targeting ARID1A. *Gynecologic Oncology*. 2014; 134:129–137. [PubMed: 24793973]
20. Lewis BP, Burge CB, Bartel DP. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell*. 2005; 120:15–20. [PubMed: 15652477]
21. Baek D, Villen J, Shin C, et al. The impact of microRNAs on protein output. *Nature*. 2008; 455:64–71. [PubMed: 18668037]
22. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*. 2010; 466:835–840. [PubMed: 20703300]
23. Nelles DA, Fang MY, O'Connell MR, et al. Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell*. 2016; 165:488–496. [PubMed: 26997482]
24. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature*. 1998; 396:643–649. [PubMed: 9872311]
25. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine*. 2012; 366:883–892. [PubMed: 22397650]
26. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*. 2012; 12:323–334. [PubMed: 22513401]
27. Schmitt MW, Loeb LA, Salk JJ. The influence of subclonal resistance mutations on targeted cancer therapy. *Nat Rev Clin Oncol*. 2016; 13:335–347. [PubMed: 26483300]
28. Flores-Alcantar A, Gonzalez-Sandoval A, Escalante-Alcalde D, et al. Dynamics of expression of ARID1A and ARID1B subunits in mouse embryos and in cells during the cell cycle. *Cell and Tissue Research*. 2011; 345:137–148. [PubMed: 21647563]

29. Mendell, Joshua T., Olson, Eric N. MicroRNAs in Stress Signaling and Human Disease. *Cell*. 2012; 148:1172–1187. [PubMed: 22424228]
30. Bartel DP. MicroRNAs: Target Recognition and Regulatory Functions. *Cell*. 2009; 136:215–233. [PubMed: 19167326]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

- A novel strategy was developed to dissect endogenous gene transcriptional regulation
- Gene expression heterogeneity can be visually examined in individual live cells
- This strategy provides a simple readout for high-throughput screening assays

Figure 1. Establishing a model that monitors endogenous *ARID1A* gene transcription

A. Schematic view of our knockin strategy. In our study, the ORF of *EGFP* (717 bp) was introduced into the *ARID1A* exon 2 in frame. The ATG start codon of both ORFs is truncated. There are two alternative splicing products of *ARID1A*. The ATG start codon labeled in black is for the longer transcript, while the one in blue is for the shorter transcript. The knockin was purposely placed in the middle of the ATG start codon for the shorter transcript to destroy it, as indicated by a cross, to ensure that the reporter gene does not interfere with normal *ARID1A* gene splicing and to keep the regulation of *ARID1A* transcription by the endogenous promoter intact. The CRISPR-Cas9 target site is indicated by a red asterisk on exon 2. The red bars flanking the reporter ORF in the donor plasmid represent the 1-kb 5' and 3' homologous arms to *ARID1A* genomic locus. B. *EGFP*-positive clones without sorting (top panel). Fluorescence imaging demonstrates that the number of false *EGFP*-positive clones is high when cells are cultured for 1 week following the first round of sorting (middle panel). The red circles highlight false-positive cells. However, every cell among the clones, following the second round of sorting, is *EGFP* positive (lower panel). Scale bar, 200 μ m. C. DNA sequencing chromatograms indicate successful knockin of the *EGFP* gene at the expected integration site. The stop codon of the *EGFP* gene is underlined by a red bar.

Figure 2. Cell cycle regulation of *ARID1A* transcription

A. Successful synchronization of cell cycle by serum deprivation, aphidicolin, and nocodazole as indicated by flow cytometry analysis. Unsynchronized cells were used as a control. B. Regulation of *ARID1A* expression, as indicated by EGFP intensity, is cell cycle-dependent. Cells are synchronized by serum deprivation at G0 phase, aphidicolin at S phase, and nocodazole at M phase. M phase has the strongest EGFP signal, while G0 has the least. The bar in the synchronized cell plot indicates a drastic difference of the EGFP signal between G0 and M phases, which translates into dramatic change of *ARID1A* transcription between these two phases of cell cycle. C. RT-qPCR results demonstrate changes of *ARID1A* levels at different phases of cell cycle. Error bar, standard deviation (sd) from three independent RT-qPCR experiments. *** $p < 0.0005$. D. Fixed fluorescent and live cell imaging of EGFP expression in cells synchronized at different phases of the cell cycle. It is clear that G2/M phase has the strongest EGFP signal, while G0 has the weakest. Interestingly, the EGFP signal is located inside the nucleus, suggesting a cryptic nuclear localization signal in the first 393 aa of the ARID1A protein. Scale bar, 50 μm for the top two panels; 20 μm for the lower panel. At least five randomly selected view fields were imaged and the representative images were shown in the panel.

Figure 3. Regulation of *ARID1A* transcription by DNA damage signal

A. CPT induced robust DNA damage, demonstrated by phosphorylation of serine 15 (s-15) of TP53 (p-p53) and induction of p21 by Western blotting. Endogenous *ARID1A* produced from the wild type allele is inhibited by DNA damage signals, which are also faithfully reflected by downregulation of EGFP reporter. B. Flow cytometry analysis of EGFP expression in HEK293 cells treated by 6 μ M CPT at 0, 1, 4, and 8 hrs. A clear suppression of fluorescence signal can be detected after 4 hours of CPT treatment. C. Fluorescence imaging in live cells treated by 6 μ M CPT at 0, 1, 4, and 8 hrs demonstrates the suppression of EGFP expression. Interestingly, loss of EGFP is seen in nucleoli after 8 hours of CPT treatment. Scale bar, 20 μ m. At least five randomly selected view fields were imaged and the representative images were shown in the panel. D. RT-qPCR results demonstrate inhibition of endogenous *ARID1A* and reporter *EGFP* transcription following DNA damage at

different time points (hrs). Error bar, standard deviation (sd) from three independent RT-qPCR experiments. ** $p < 0.01$; *** $p < 0.0005$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4. Regulation of *ARID1A* transcription by miR-144 and miR-9

A. Fluorescence imaging in live cells demonstrates the suppression of EGFP expression following transfection of miRNA mimics. A scramble mimic was also used as a control. Scale bar, 50 μ m. At least five randomly selected view fields were imaged and the representative images were shown in the panel. B. Flow cytometry analysis confirms the decrease of fluorescence intensity from the cells treated with miRNA mimics. C. Both endogenous *ARID1A* and reporter EGFP proteins are downregulated by miRNA mimics. D. RT-qPCR results demonstrate inhibition of endogenous *ARID1A* and reporter *EGFP* transcription by miRNA mimics, suggesting that miR-9 and miR-144 regulate *ARID1A* primarily through mRNA degradation. Error bar, standard deviation (sd) from three independent RT-qPCR experiments. * p<0.05.