

# UCSF

## UC San Francisco Previously Published Works

### Title

ARID2 deficiency promotes tumor progression and is associated with higher sensitivity to chemotherapy in lung cancer

### Permalink

<https://escholarship.org/uc/item/0dx8n84p>

### Journal

Oncogene, 40(16)

### ISSN

0950-9232

### Authors

Moreno, Thaidy  
Monterde, Beatriz  
González-Silva, Laura  
[et al.](#)

### Publication Date

2021-04-22

### DOI

10.1038/s41388-021-01748-y

Peer reviewed

Published in final edited form as:

*Oncogene*. 2021 April 01; 40(16): 2923–2935. doi:10.1038/s41388-021-01748-y.

## **ARID2 deficiency promotes tumor progression and is associated with higher sensitivity to chemotherapy in lung cancer**

Thaidy Moreno<sup>1</sup>, Beatriz Monterde<sup>1</sup>, Laura González-Silva<sup>1</sup>, Isabel Betancor-Fernández<sup>2</sup>, Carlos Revilla<sup>1,†</sup>, Antonio Agraz-Doblas<sup>1</sup>, Javier Freire<sup>3</sup>, Pablo Isidro<sup>4</sup>, Laura Quevedo<sup>1</sup>, Rosa Blanco<sup>1</sup>, Santiago Montes-Moreno<sup>3</sup>, Laura Cereceda<sup>3</sup>, Aurora Astudillo<sup>4</sup>, Berta Casar<sup>1,6</sup>, Piero Crespo<sup>1,6</sup>, Cristina Morales Torres<sup>5</sup>, Paola Scaffidi<sup>5</sup>, Javier Gomez-Roman<sup>3</sup>, Eduardo Salido<sup>2</sup>, Ignacio Varela<sup>1,\*</sup>

<sup>1</sup>Instituto de Biomedicina y Biotecnología de Cantabria. Universidad de Cantabria-CSIC. Santander, Spain

<sup>2</sup>Departamento de Patología, Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER). Tenerife, Spain

<sup>3</sup>Servicio de Anatomía Patológica y Biobanco Valdecilla. HUMV/IDIVAL. Santander, Spain

<sup>4</sup>Biobanco del Principado de Asturias (BBPA). Hospital Universitario Central de Asturias. Oviedo, Spain

<sup>5</sup>Cancer Epigenetics Laboratory, The Francis Crick Institute. London, UK

<sup>6</sup>Centro de Investigación Biomédica en Red de Cáncer (CIBERONC)

### **Abstract**

The survival rate in lung cancer remains stubbornly low and there is an urgent need for the identification of new therapeutic targets. In the last decade, several members of the SWI/SNF chromatin remodeling complexes have been described altered in different tumor types. Nevertheless, the precise mechanisms of their impact on cancer progression, as well as the application of this knowledge to cancer patient management are largely unknown.

In this study, we performed targeted sequencing of a cohort of lung cancer patients on genes involved in chromatin structure. Additionally, we studied at the protein level the expression of these genes in cancer samples and performed functional experiments to identify the molecular mechanisms linking alterations of chromatin remodeling genes and tumor development.

Remarkably, we found that 20% of lung cancer patients show ARID2 protein loss, partially explained by the presence of *ARID2* mutations. Additionally, we showed that ARID2-deficiency provokes profound chromatin structural changes altering cell transcriptional programmes which

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

<sup>\*</sup>*To whom correspondence should be addressed:* Ignacio Varela, Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria, Albert Einstein 22, 39011 Santander, Spain., Ph. +34 942203931, [ignacio.varela@unican.es](mailto:ignacio.varela@unican.es).

<sup>†</sup>Deceased during the review of the manuscript

### **Conflict of Interest**

The authors disclose no potential conflicts of interest.

bolsters the proliferative and metastatic potential of the cells both *in vitro* and *in vivo*. Moreover, we demonstrated that ARID2 deficiency impairs DNA repair, enhancing the sensitivity of the cells to DNA damaging agents.

Our findings support that *ARID2* is a *bona-fide* tumor suppressor gene in lung cancer that may be exploited therapeutically.

## Keywords

ARID2; Lung Cancer; Next-generation sequencing technologies; SWI/SNF

---

## Introduction

Lung cancer is the major cause of cancer-related deaths worldwide with an average 5-year survival rate below 20% irrespective of the subtype<sup>1</sup>. Consequently, any new knowledge about the molecular mechanisms that drive this disease could have a great impact on the treatment of patients. Recently, large genomic projects have facilitated the identification of major players in this tumor type. Thus, small cell lung cancer (SCLC) which constitutes around 15% of all cases, is mainly driven by mutations in *TP53* and *RBI*, but the role of other genes like *PTEN*, *SLIT2* or *CREBBP* has been also described<sup>2</sup>. Among non-small cell lung cancer (NSCLC), more than half of the cases are adenocarcinomas, where *TP53*, *KRAS*, *EGFR*, *ALK*, *ROS1* and *BRAF* are the main recurrently altered genes<sup>3</sup>, while squamous cell carcinomas (SCC) are genetically more heterogeneous and poor in actionable mutations so far.

Lately, several members of the SWI/SNF family of chromatin remodeling complexes have been identified recurrently altered in different tumor types adding to the accumulated compelling evidence on the role of chromatin structure in cancer development. It is estimated that approximately 20% of all tumors contain alterations in these complexes, a frequency that is only exceeded by mutations in *TP53*<sup>4</sup>. In the case of non-small cell lung cancer, the expression of any of the two mutually exclusive catalytic ATPase subunits (SMARCA2 or SMARCA4) is lost in 30% of the cases where it is associated with worse prognosis<sup>5</sup>. Additionally, *ARID1A*, which encodes for one of the auxiliary subunits of the complex, frequently appears mutated in lung adenocarcinoma<sup>6</sup>.

## Materials and Methods

Detailed protocols can be found in Supplementary Methods.

## Next-generation sequencing

Cancer patient primary tumor samples and, when available, matched corresponding normal samples, were obtained from different tumor biobanks after the corresponding approval of the hospital ethics committees and patient informed consent. A detailed list of the origin and characteristics of each sample can be found in Supplementary Table 1. DNA was extracted using the Agencourt DNAdvance Beckman Coulter kit (Beckman Coulter, USA), fragmented and submitted to end-repair and adenylation, adaptor ligation and PCR indexing

amplification. Target capture was performed using a Sure Select® user-defined probe kit (Agilent Technologies, USA).

For ATAC-Seq libraries, cell nuclei were extracted using cold lysis buffer and submitted to tagmentation (Nextera DNA Library Preparation Kit, Illumina, USA). After purification, adapter sequences were used to complete Illumina sequencing adapters by PCR with Phusion High Fidelity DNA polymerase (Thermo Fisher Scientific, UK).

Total RNA was purified using Extract Me Total RNA Kit (Blirt, USA). Reverse transcription was performed using the Takara PrimeScript cDNA Synthesis kit (Takara Bio Europe, France). Poly-A mRNA was enriched, fragmented, and submitted to cDNA generation using PrimeScript Enzyme for first strand and RNase HI, DNA polymerase I and T4 DNA Polymerase (Thermo Fisher Scientific, UK) for the second strand. Afterwards, genomic libraries were generated as above. Individual mRNA expression was measured by qRT-PCR using Luminaris Color HiGreen qPCR Master Mix (Thermo Fisher Scientific, UK).  $\beta$ -actin was used as housekeeping gene and the Ct method was used for quantification and comparison.

### Sequencing Data Analysis

DNA sequence data was mapped to the human genome (hg19) using BWA 0.7.3<sup>7</sup>. Additionally, Samtools 0.1.18<sup>8</sup>, Picard 1.61 (<http://broadinstitute.github.io/picard/>) and GATK 2.2.8<sup>9</sup> were used for format transformation, cleaning, sorting and indexing of the bam files, marking PCR duplicates and performing indel local realignment. RAMSES<sup>10</sup> and PINDEL 0.2.4d<sup>11</sup> were used for substitutions and small insertion and deletion identification respectively. All *ARID2* mutations were validated by PCR amplification coupled with ultrasequencing at 10,000x coverage. Additionally, a similar orthogonal validation of more than 180 mutations randomly picked showed a near 80% of specificity in the mutation calling. OncodriveFML software was run to detect genes with evidence of positive selective pressure<sup>12</sup>.

ATAC-Seq reads were aligned against the human genome (hg19) using BWA 0.7.3<sup>7</sup>. Accessible regions were identified using MACS 2.1.2<sup>13</sup>. A combined list of all the regions identified in all the samples by MACS, as well as the list of enhancers annotated in the GeneHancer project<sup>14</sup>, were used to identify significant changes in region accessibility in *ARID2*-deficient cells versus control using DESeq2<sup>15</sup>. Region annotation was performed using CHIPSeeker software<sup>16</sup>. BEDTools<sup>17</sup> was used to estimate the overlapping of the identified regions with ENCODE publicly available histone marks ChIP-Seq A549 data and the results were plotted using deeptools v3.3.1<sup>18</sup>. Motif enrichment analysis was performed using HOMER<sup>19</sup>. Finally, alignments were visualized using IGV genome browser<sup>20</sup>.

RNA-Seq data was aligned using Tophat<sup>21</sup> to the human genome (hg19). Differentially expressed genes (DEG) were identified using HTSeq + DESeq2<sup>22,23</sup>. Gene-set Enrichment analysis on Gene Ontology terms was performed using GSEA software<sup>24</sup>. For the analysis of the TCGA database lung adenocarcinoma patients, raw counts for each patient (n=524) were downloaded and normalized using DESeq2 software. Low-*ARID2* expressing lung adenocarcinoma patients (n = 64) were defined as those with a normalized expression less

than the mean minus two standard deviations calculated from the whole cohort. Similarly, high ARID2-expressing lung adenocarcinoma patients (n=78) were defined as those with a normalized expression higher than the mean plus two standard deviations. DEG between two groups were identified using DESeq2.

### Cell Culture and *in vitro* assays

A549, NCI-H1568 and NCI-H460 lung cancer cell lines were obtained from The Francis Crick Institute common and ATCC repositories, authenticated by STR profiling, and tested for mycoplasma. Tetracycline-inducible pTRIPZ constructs V2THS\_74399 (v2), V3THS\_347660 (v3) (*ARID2*), V2THS-283735 (*ARID1A*) and V2THS-11753 (*ARID1B*) were used for stable cell line generation (Dharmacon/GE Healthcare, USA). The empty vector (RHS4750) was used as control. Virus production was performed by transfecting HEK293T/17clone cells with the pTRIPZ and packaging constructs. Infected cells were selected with 1 µg/ml puromycin and isolated by FACS using a FACS-Aria II cell sorter (Becton Dickinson, USA) based on TurboRFP expression after the induction with 1 µg/ml of Doxycycline

Extrapolated growth curves were constructed over a period of fourteen days by serial passaging and cell counting with a hemocytometer or by PrestoBlue® assay (Thermo Fisher Scientific, UK). Cell proliferation was also analyzed using the CellTrace™ CFSE Cell Proliferation Kit (Invitrogen, USA) in cells synchronized by gradual serum deprivation following published protocols<sup>25</sup>. Cells were harvested at 48 hours and subjected to division peak resolution by flow cytometry. The cell proliferation index was analyzed using MODFIT software (Verity, USA). Proliferation index was the sum of the cells in all generations divided by the calculated number of original parent cells.

*In vitro* cell migration assays were performed by using 8-µm pore size transwell chambers (Corning™ Transwell™ Multiple Well Plate) in 24-well plates using 10% FBS in the lower chamber as chemo-attractant. For invasion assays, cells were plated on growth factor-reduced Matrigel (BD Biosciences) pre-coated 8 µm pore transwell chambers. Filters or invasive cells were quantified by fixing chambers in 4% paraformaldehyde for 10 min and staining with crystal violet.

Growth inhibition assays were performed to determine the half maximal inhibitory concentration (IC<sub>50</sub>) values for different antitumoral drugs. Viability after 48 h was determined by PrestoBlue® reagent (Thermo Fisher Scientific, UK). IC<sub>50</sub> value for each drug were determined with Prism software (GraphPad, USA).

### *In vivo* tumorigenesis assays

Animal studies were conducted in compliance with guidelines for the care and use of laboratory animals and were approved by the Ethics and Animal Care Committee of Universidad de Cantabria. For proliferation assays, five million cells in 500 µl of PBS were subcutaneously injected into the flanks of 6-8-week-old female nude mice (Athymic Nude-Foxn1nu, Envigo, UK). 24 days after the injection, mice were euthanizing and tumor tissues were harvested for analyses. For metastasis assays, 2.5 million of cells in 500 µl of PBS with 0,1% BSA, were tail injected into 6-8-week-old female nude mice. Hairpin expression in the

cells was induced with 1 µg/ml of Doxycycline 7 days before injection. After two months, mice were euthanized and tumor tissues were harvested for analyses. In both cases, to keep the hairpin expression, the animals were treated from the day of injection with 2 mg/mL of Doxycycline in the drinking water supplemented with 1% sucrose refreshed every 2-3 days.

### Western blot analysis

Total protein lysates were prepared in RIPA buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 % NP-40, 1 mM Sodium Orthovanadate, 1 mM NaF), separated by SDS-PAGE in 8% polyacrylamide gels and transferred to nitrocellulose membranes. Subsequently, membranes were washed with TBST (50 mM TRIS + 150 mM Sodium chloride + 0,1% Tween 20, pH 7,4) and blocked using 5% non-fat milk solution in TBS for 1 h at room temperature. Membranes were then incubated with primary antibodies anti-ARID2 (E-3, Santa Cruz) and anti-Actin (I-19, Santa Cruz), diluted 1:200 and 1: 1,000 in TBST with 5% (w/v) BSA at 4°C overnight, respectively. Donkey anti-mouse or donkey anti-goat secondary antibodies (LI-COR Biotechnology, USA) conjugated to IRDye 800CW (926-32212) or IRDye 680RD (926-68074) respectively were used as secondary antibodies.

### Immunohistochemistry analysis

For ARID2 detection on paraffin sections, antigen retrieval was performed for 32 minutes at 97 °C in citrate buffer pH 6, incubated with 1:300-1:500 anti-ARID2 antibody (abcam ab113283) and developed with HRP-polymer secondary antibodies (Optiview, Roche). ARID2 expression was evaluated by two pathologists on coded tissue sections, without information about the *ARID2* mutation status. Only surgical pathology cases with enough material, both tumor and non-neoplastic surrounding tissue, were considered for ARID2 immunohistochemistry. A consensus score was reached viewing the slides by two pathologists at a multiheaded scope.

Immunofluorescence was performed in cells fixed with 4% paraformaldehyde in PBS for 15 min at room temperature. The cells were permeabilized with 0.5% Triton X-100 in PBS and blocked with 3% BSA in PBT (PBS containing 0.05% Triton X-100). Finally there were subjected to immunofluorescence staining with ARID2 antibody (E-3, sc-166117 Santa Cruz, USA or A302-230A, Bethyl Laboratories, USA), anti-phospho-Histone H2A. X Ser139 (γH2AX, clone JBW301, Merck Millipore, USA) or anti-53BP1 antibody (H-300, sc-22760, Santa Cruz, USA). Cover slides were incubated with Alexa labeled secondary antibodies and mounted in VECTASHIELD Antifade Mounting Medium with DAPI (Vector Labs, USA). Colocalization of ARID2 with 53BP1 or γH2AX was performed measuring the variation in intensity across the lines drawn using the linescan tool from MetaMorph® (Molecular Devices, USA). For DNA repair assays, cells were treated with 10 µM Etoposide for 1 h. Subsequently, γH2AX foci were quantified using ImageJ software at different recovery times after removing the drug from the media.

### Statistical analysis

In all cases, at least three independent experiments were performed in order to assess the statistical significance of all differences. In figure legends, the specific statistical test performed in each case is indicated. In general, for quantitative variables, a one-tailed t-test

with equal variance was used to identify significant differences between groups. For qualitative variables, a Fisher exact test was used in order to identify significant differences between groups of patients. The different software used for the identification of mutations, gene expression differences, enrichment transcription factor binding sites and gene ontology terms have their own statistical models explained in detailed in the references. When multiple tests were performed the significance is shown corrected for multiple testing.

## Results

### Loss of ARID2 protein expression in 20% of lung cancer patients is partially explained by the presence of *ARID2* mutations

In order to understand better the role of chromatin remodeling complexes in lung cancer development, we performed a genetic screening on the coding sequences of known cancer genes as well as members of the main chromatin remodeling complexes (Suppl. Table 2). We applied targeted next-generation sequencing technologies in a collection of 81 lung cancer cases (40 lung adenocarcinomas, 12 squamous cell carcinomas and 29 small cell carcinomas) (Suppl. Table 1). Interestingly, we found mutations in *ARID2* in 5 of the patients, 3 classified as lung adenocarcinoma and 2 as small cell carcinoma (Figure 1a and Suppl. Table 3). Additionally, to identify lung cancer driver genes, we run OncodriveFML<sup>27</sup> with our data. This software identifies genes with a number and distribution of predicted deleterious mutations higher than expected by chance, evidencing positive selection. *ARID2* ranked second in the list of genes showing significant positive selection after multi-test correction, just below *TP53* (Figure 1b and Suppl. Table 4). To validate these results, we sequenced *ARID2* coding sequences in a second cohort of 144 lung adenocarcinoma cases and found mutations in 12 patients. If we consider all analyzed lung cancer patients, irrespective of the subtype, *ARID2* mutations occur at a frequency of 7.5 % (17/225) (Suppl. Table 3). In the case of lung adenocarcinomas (40 and 144 patients from first and second cohorts respectively), we found *ARID2* non-synonymous mutations in 7% of the samples (13/184), which is near twice the frequency reported in COSMIC database for this tumor type (3.7 %, 83/2241)<sup>28</sup> and ranks *ARID2* among the ten genes most commonly mutated in lung cancer. In concordance with a potential role of *ARID2* as tumor suppressor, many of the identified mutations, clustered at the beginning of the protein sequence, are predicted to generate a premature truncation of the protein (Figure 1c and Suppl. Table 3). Subsequently, to check if the loss of ARID2 function is a common feature in lung cancer, we performed immunohistochemistry analyses in 139 of the studied samples finding loss or low/heterogeneous ARID2 production in approximately 20% of the cases (28/139) (Figure 1d and Suppl. Figure 1). Additionally, loss of *ARID2* expression was significantly more frequent in *ARID2*-mutated patients (6/10 Fisher exact test  $p=0.0098$ ). Interestingly, this was also true for some samples with missense mutations, which suggests that these mutations might interfere with the correct folding or processing of the protein. Indeed, many of the mutations found are predicted to produce deleterious effects in the protein according to SHIFT or Polyphen algorithms (Suppl. Table 3). We observed a complete loss of ARID2 signal in many lung cancer samples suggesting a selective pressure to inactivate both *ARID2* alleles. In addition, some non-mutated samples showed also loss of ARID2 production, suggesting the existence of non-genetic mechanisms that interfere with *ARID2* expression.

Since the presence of normal tissue contamination in the tumor samples as well as the sequencing strategy followed for normal tissue, prevented an estimation of the cellularity or the zygosity of the mutations, we could not determine whether full loss of ARID2 protein is due to genetic loss or silencing of the wild-type allele.

### **ARID2-deficiency increases proliferative and metastatic potential *in vitro* and *in vivo***

In order to check if alterations in *ARID2* could promote lung cancer development, we knocked down the protein in different ARID2-proficient NSCLC cell lines. As it can be observed in Figure 2 a-b, *ARID2* mRNA and protein production was efficiently reduced by two different shRNAs. This reduction was accompanied by an increase in the proliferation of A549 cells, as well as their invasion and migration capacities compared to those cells transduced with the empty vector. Similar results were obtained in NCI-H460 cell line (Figure 2 c-e and Suppl. Figure 2). Moreover, when these cells were injected into immunocompromised mice, they showed a greater capacity to produce tumors *in vivo* (Figure 2 f and Suppl. Figure 3).

RNA-Seq experiments in A549 transduced cell lines showed that loss of ARID2 was accompanied by significant changes in the expression of 1155 genes (366 upregulated and 789 downregulated), that supported the observed phenotypes in the cells (Figure 2g and Suppl. Table 5). Thus, we observed a downregulation of genes involved in cellular adhesion and cell differentiation such as *NPNT*, *CDH6*, *FAT3*, *FNI*, *SOX2* or *SDC2* as well as an upregulation of genes associated with a higher cell-cycle progression such as *CDC45*, *MCM2* or *HIST1H1E*, which could be associated with the increased proliferation, migration and invasion capacities of ARID2-deficient cells. Additionally, we observed downregulation of other tumor suppression genes like *RPS6K2*, *TNFSF10*, *ISM1* or *LDLRAD4* together with upregulation of protumoral and anti-apoptotic genes like *HOXB1*, *BCL2A1* or *RCVRN*. Most of these alterations were not observed when we knocked down ARID1A or ARID1B subunits in the same cells (Suppl. Figure 4a) which indicates a specific gene set regulation by ARID2-containing SWI-SNF complexes. Transcriptional changes in selected genes were further validated by qRT-PCR in independently generated ARID2 knock-down cell lines (Suppl. Figure 4b).

Altogether, these results prove that *ARID2* plays a tumor suppressor function in lung cancer.

### **ARID2-deficiency is accompanied by widespread chromatin changes, specially affecting enhancers**

We hypothesized that gene expression changes observed in ARID2-deficient cells might be the result of changes in SWI/SNF chromatin remodeling activity. To investigate this, we performed ATAC-Seq experiments in ARID2 knocked-down A549 cells. ARID2 loss was accompanied by a general loss of chromatin accessibility with 990 regions that showed a significant loss of chromatin accessibility versus 687 regions that showed increased accessibility (Suppl. Table 6). Interestingly, those regions that lost accessibility in the absence of ARID2 were located distal to gene transcription start sites (Figure 3a) and showed enrichment of enhancer specific H3K4me1 and H3K27ac histone marks according to ENCODE project data. An opposite behavior is observed on those regions that gained



accessibility after ARID2 loss (Suppl. Figure 5). Additionally, AP-1 family transcription factor binding motif, described as abundantly present in enhancers<sup>29</sup>, is highly enriched on those regions that showed less accessibility on ARID2-deficient cells (Figure 3b). In order to explore a special impact of ARID2 loss on enhancers, we analyzed the accessibility of those regions annotated as enhancers in the GeneHancer project<sup>14</sup>. As it can be seen in Figure 3c, the loss of accessibility is significantly more profound in enhancers than in the rest of the genome, as 87 % (1744 of 2001) of the enhancers that showed significant accessibility changes, lose accessibility in ARID2-deficient cells. Additionally, many of the target genes of these enhancers showed significant downregulation in the RNA-Seq data (Figure 3d). All this support that *ARID2* is essential to keep an open chromatin conformation around enhancers which significantly impacts on the transcriptional regulation of specific gene networks (Suppl. Table 6).

### **ARID2 is essential to maintain the expression of the metastasis inhibitor MTSS1 and the adhesion molecule SDK1**

In order to identify ARID2 target genes of broad relevance to lung cancer patients, we compared our RNA-Seq results with differential expression analysis performed on human lung adenocarcinoma patients from TCGA database. Eighteen genes were found upregulated in both ARID2-deficient cells, and in low-ARID2 expressing lung adenocarcinoma patients (Suppl. Table 7). Among them, we found *AREG*, *ERG* or *NGF* growth factors that might explain the higher proliferation capabilities of ARID2-deficient cells (Figure 3e and Suppl. Table 7). In addition, we found 133 genes downregulated in both datasets indicating a main gene expression activating role of ARID2 in this cellular context.

Interestingly, among those genes whose expression rely on ARID2, we found *MTSS1*, a well described metastasis inhibitor<sup>29,29,30</sup>, as well as *SDK1*, involved in cell-cell adhesion. The observed significant reduction of *MTSS1* and *SDK1* expression likely explain the higher invasion capabilities of ARID2-deficient cells. Additionally, we found a significant reduction of chromatin accessibility on two enhancers regulating these genes in GeneHancer database after ARID2 loss<sup>14</sup>. This observation is concordant with the hypothesis that *MTSS1* and *SDK1* expressions are positively regulated by ARID2 by keeping an open chromatin structure at their enhancers (Figure 3f).

### **ARID2 loss impairs DNA damage repair**

The RNA-Seq analysis also revealed deleterious consequences suffered by ARID2-deficient cells that could be exploited therapeutically. Gene-set enrichment analyses (GSEA) on the transcriptional alterations observed in ARID2-deficient cells also showed a significant upregulation of genes involved in DNA damage detection and repair, suggesting a defective DNA damage response (Figure 4a). Supporting a role of ARID2 in DNA repair, analysis of its localization in untransduced A549 cells showed a co-localization with  $\gamma$ H2AX and 53BP1 at the DNA repair foci (Figure 4b and Suppl. Figure 6). In addition to that, ARID2-deficient cells showed a delay in the resolution of DNA damage foci compared to wildtype cells in A549 and NCI-H460 NSCLC cell lines upon treatment with etoposide. Interestingly, this delay was not dependent on TP53 function as we could see a similar delay in ARID2-deficient NCI-H1568 cell line which is TP53 deficient (Figure 4c-d). These observations

indicate that ARID2 deficiency inhibits efficient DNA repair and suggest that its loss may sensitize cells to DNA damaging agents.

### ARID2-deficiency increases cell sensitivity to chemotherapy and veliparib

As platinum-based chemotherapy is widely used for the treatment of lung cancer patients<sup>3</sup>, we first examined ARID2-deficient cell lines sensitivity to cisplatin, as well as etoposide. As it can be seen in Figure 4e and Suppl. Figure 7, in concordance with defective DNA repair in the absence of ARID2, ARID2-deficient A549 and NCI-H460 cells exhibited a higher sensitivity to both compounds compared to control cells.

Additionally, in the last decade, many researchers have described a higher sensitivity of PARP inhibitors in tumors harboring defects in DNA repair mechanisms due to synthetic lethality<sup>32</sup>. Consequently, we checked if this might apply as well to ARID2-deficient cells. As it can be seen in Figure 4e, ARID2 loss led to a higher sensibility to veliparib, a well described PARP inhibitor that is under research in several clinical trials in breast, ovarian and, most importantly, lung cancer. This observation suggests ARID2 deficiency as a useful marker for the stratification of lung cancer patients that may benefit for PARP inhibitor treatment.

## Discussion

Although some evidence of the presence of *ARID2* alterations in lung cancer have been reported previously<sup>33</sup>, the relevance of these alterations for oncogenesis has not been clearly proved. Our results showed an *ARID2* mutation recurrency higher than the one reported in COSMIC database. Additionally, the distribution and predicted impact of the mutations found and our *in-vitro* and *in-vivo* experiments provided compelling evidence of the role of *ARID2* as *bona-fide* tumor suppressor in lung cancer. Supporting this, *ARID2* has been already proposed as cancer driver gene in melanoma and hepatocellular carcinoma<sup>34,35</sup>. We don't have enough data to definitely prove that complete ARID2 activity loss is necessary for tumor progression but several evidences support this idea. First, in many cases we observed a complete loss of ARID2 expression in the human adenocarcinoma samples which suggests a selective pressure to inactivate both *ARID2* alleles. Additionally, we observed that V2THS\_74399 shRNA construct was less efficient in abrogating ARID2 expression than V3THS\_347660 and cells transduced with the former typically showed less pronounced changes than cells transduced with the latter.

The precise molecular mechanisms by which alterations in chromatin remodeling complexes promote cancer development are not sufficiently understood. Interactions with well-described cancer genes like *TP53*, *RB* or *MYC* have been described<sup>36-38</sup>. In addition to this, they play essential roles in the activation of differentiation and the suppression of proliferative programs of many cellular lineages<sup>39</sup>. In this study we described a list of near 200 genes that are specifically deregulated after ARID2 loss in both our cellular model and in lung adenocarcinoma patients from the TCGA database. Some of these genes, like *AREG*, *EREG* or *NGF* growth factors might account for the higher proliferation capabilities of ARID2-deficient cells. In terms of the molecular mechanisms behind this regulation, we show that ARID2-deficiency is associated with widespread chromatin structural changes.

Our results prove that ARID2 is essential to keep an open chromatin structure in enhancer regions in agreement with an important role of different SWI/SNF members in regulating enhancer activity<sup>40,41</sup>. Two of these ARID2-dependent enhancers regulate *MTSS1* and *SKD1* expression that, consequently, showed a significant downregulation in ARID2-deficient cells. This suggests that ARID2 might regulate directly the expression of *MTSS1* and *SKD1*, although further work is necessary to finally confirm this. *MTSS1* is a well described migration and invasion inhibitor, associated with worse prognosis in several tumor types<sup>30,31,42</sup>, and *SKD1* plays important roles in cell-cell adhesion. Their deficiency might well explain the higher migration and invasion capabilities of ARID2-deficient cells.

In addition, we observed an active role of ARID2 in the detection and repair of DNA damage *in vitro* in lung cancer cell lines, as ARID2-deficient cells present important delays in the resolution of DNA damage foci that were not dependent on TP53 activity. In accordance with this view, other members of the SWI/SNF complex have been shown to be involved in different steps of DNA damage repair<sup>43–45</sup>.

Finally, any advance in the possibility of exploiting therapeutically any vulnerability associated to deficiency in SWI/SNF complex genes is of great interest, as approximately 20% of all human cancers are reported to have alterations in this complex. In this study ARID2-deficient cells showed a higher sensitivity to different DNA-damaging therapies, likely as a result of the ARID2 involvement in DNA repair. Considering that platinum-based chemotherapy is still widely used in lung cancer patients with high variable success<sup>3</sup>, our results suggest that ARID2 expression might be explored as a stratification marker for these therapies. Moreover, we show that ARID2 deficiency shows synthetic lethality with PARP inhibition using veliparib, an inhibitor that has shown good results in the treatment of breast cancer<sup>46</sup> and is included in several clinical trials on breast, ovarian and, most importantly, lung cancer. Our results suggest that the stratification of lung cancer patients according to ARID2 expression might improve the efficiency of PARP inhibitors in non-small cell lung cancer. Additionally, a very recent study has shown that ARID2-deficient melanoma cells are particularly sensitive to immunotherapy through alterations in mTORC1 and IFN $\gamma$  pathways<sup>47</sup>. Interestingly, we observed that some downstream response genes in these pathways such as *GBP2*, *GBP3* and *SCD5* are significantly downregulated in ARID2-deficient cells (Suppl. Table 5). All these results support the potential use of *ARID2* expression as a new stratification marker for personalized treatment in lung cancer patients.

In summary, here we present compelling evidence for the role of *ARID2* as tumor suppressor in lung cancer. Although *ARID2* has been proposed as a driver gene in other tumor types<sup>35,48</sup>, little has been reported about the molecular mechanisms underlying this involvement. In this work, we propose that its role in lung cancer is exerted by fostering a specific pro-oncogenic transcriptomic program as a result of changes in chromatin structure around enhancers. Importantly, our results indicate that ARID2-deficiency could be exploited for lung cancer patient treatment.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

In memoriam of Carlos Revilla, we will miss you forever. We would like to thank the support of the funding agencies Spanish Ministerio de Economía y Competitividad, Fundación Ramón Areces, European Research Council, Asociación Española contra el cáncer, Cancer Research UK, UK Medical Research Council, Wellcome Trust, Servicio de Salud del Principado de Asturias, Instituto de Salud Carlos III and Fundación Bancaria Cajastur (specific grant references are included in the funding support section). We will like to thank as well the technical support of the different institutions and common services as well as the patients that agreed to participate in this study. We finally want to thank Dr. Francisco Real, Dr. Roland Rad, Dr. Jose Pedro Vaqué and Dr. Javier Leon for providing critical reagents and advice as well as to all the patients that agreed to participate in this study. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

## Financial Support

I. V. is supported by SAF2012-31627 and SAF2016-76758-R grants from the Spanish Ministerio de Economía y Competitividad (MINECO), by a Fundación Ramón Areces grant and by ERC2014-StG637904 grant from the European Research Council. I. V. has been awardee of the Programa Ramón y Cajal (MINECO, Spain). T. M. has been awardee of the Ayudas para la contratación de investigadores predoctorales (MINECO, Spain). B. M. is awardee of the Ayudas para la formación de profesorado universitario (FPU, Ministerio de Educación y Formación Profesional, Spain). PC laboratory is supported by grant SAF-2015-63638R (MINECO/FEDER, UE); by Centro de Investigación Biomédica en Red de Cáncer (CIBERONC) and by Asociación Española Contra el Cáncer (AECC), grant GCB141423113. BC has been supported by a Retos Jóvenes Investigadores grant SAF2015-73364-JIN (AEI/FEDER, UE) and a grant from Fundación Francisco Cobos. P. S. is supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001152), the UK Medical Research Council (FC001152). HUCA/IUOPA which is jointly financed by Servicio de Salud del Principado de Asturias, Instituto de Salud Carlos III and Fundación Bancaria Cajastur. This research was funded in part by the Wellcome Trust [FC001152]. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

## Data availability

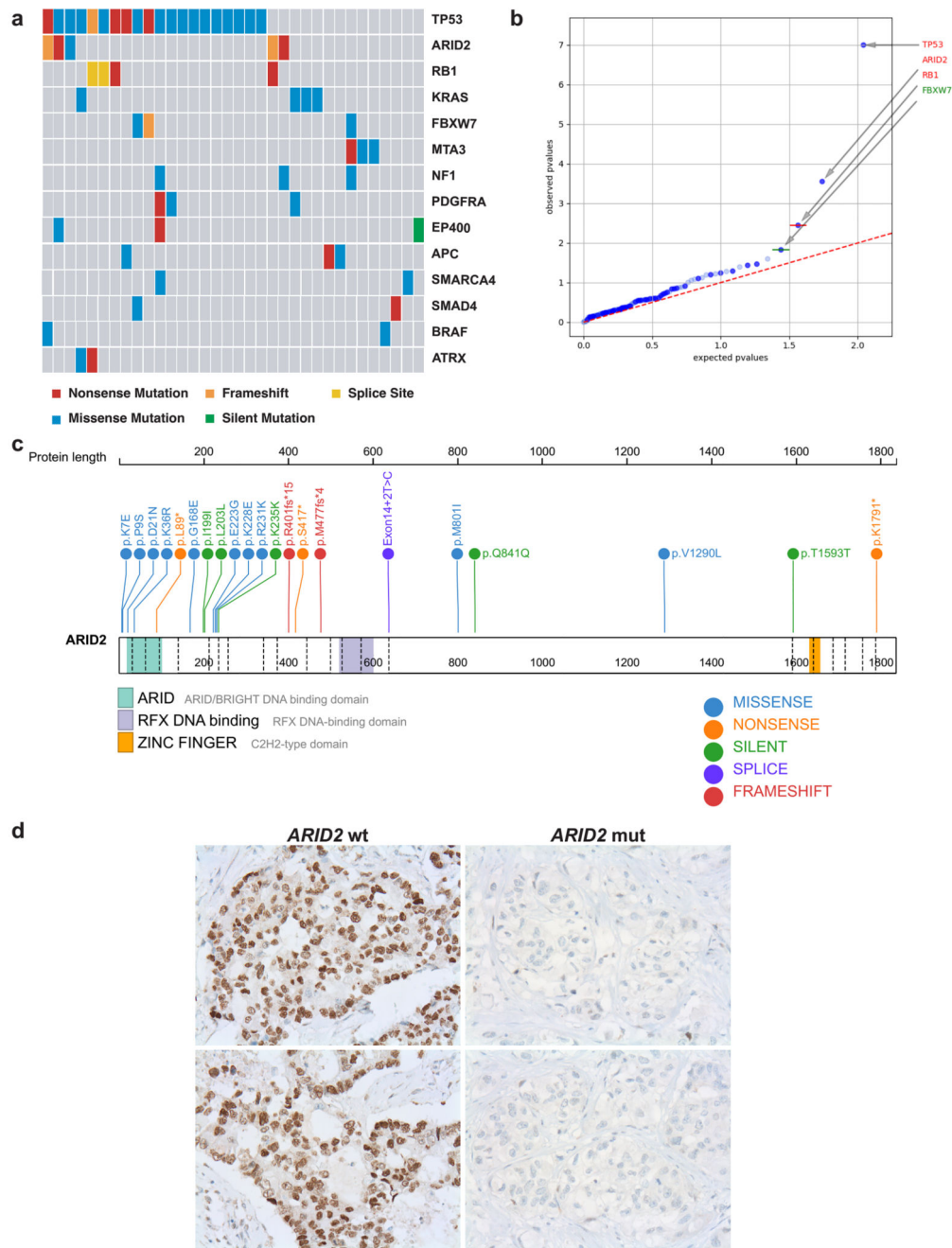
DNA-Seq, ATAC-Seq and RNA-Seq data: <https://www.ebi.ac.uk/ena/data/search?query=PRJEB26936>. All computer code is available upon request.

## References

1. Lovly CM, Carbone DP. Lung cancer in 2010: One size does not fit all. *Nat Rev Clin Oncol*. 2011; 8:68–70. [PubMed: 21278771]
2. Gelsomino F, Rossi G, Tiseo M. MET and Small-Cell Lung Cancer. *Cancers*. 2014; 6:2100–2115. [PubMed: 25314153]
3. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong K-K. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer*. 2014; 14:535–546. [PubMed: 25056707]
4. Masliah-Planchon J, Bièche I, Guinebretière J-M, Bourdeaut F, Delattre O. SWI/SNF Chromatin Remodeling and Human Malignancies. *Annu Rev Pathol Mech Dis*. 2015; 10:145–171.
5. Reisman DN, Sciarrotta J, Wang W, Funkhouser WK, Weissman BE. Loss of BRG1/BRM in human lung cancer cell lines and primary lung cancers: correlation with poor prognosis. *Cancer Res*. 2003; 63:560–566. [PubMed: 12566296]
6. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. *Cell*. 2012; 150:1107–1120. [PubMed: 22980975]
7. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl*. 2009; 25:2078–2079.
9. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20:1297–1303. [PubMed: 20644199]

10. Martínez N, Almaraz C, Vaqué JP, Varela I, Derdak S, Beltran S, et al. Whole-exome sequencing in splenic marginal zone lymphoma reveals mutations in genes involved in marginal zone differentiation. *Leukemia*. 2014; 28:1334–1340. [PubMed: 24296945]
11. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009; 25:2865–2871. [PubMed: 19561018]
12. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol*. 2016; 17 doi: 10.1186/s13059-016-0994-0
13. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol*. 2008; 9:R137. [PubMed: 18798982]
14. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database. 2017; 2017 doi: 10.1093/database/bax028
15. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15:550. [PubMed: 25516281]
16. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015; 31:2382–2383. [PubMed: 25765347]
17. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841–842. [PubMed: 20110278]
18. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016; 44:W160–W165. [PubMed: 27079975]
19. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell*. 2010; 38:576–589. [PubMed: 20513432]
20. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. *Cancer Res*. 2017; 77:e31–e34. [PubMed: 29092934]
21. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013; 14:R36. [PubMed: 23618408]
22. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010; 11:R106. [PubMed: 20979621]
23. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015; 31:166–169. [PubMed: 25260700]
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005; 102:15545–15550. [PubMed: 16199517]
25. Lauand C, Niero EL, Dias VM, Machado-Santelli GM. Cell cycle synchronization and BrdU incorporation as a tool to study the possible selective elimination of Erbb1 gene in the micronuclei in A549 cells. *Braz J Med Biol Res Rev Bras Pesqui Medicas E Biol*. 2015; 48:382–391.
26. McQuin C, Goodman A, Chernyshev V, Kamensky L, Cimini BA, Karhohs KW, et al. CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biol*. 2018; 16:e2005970. [PubMed: 29969450]
27. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol*. 2016; 17 doi: 10.1186/s13059-016-0994-0
28. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017; 45:D777–D783. [PubMed: 27899578]
29. Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res*. 2013; 23:777–788. [PubMed: 23482648]

30. Kayser G, Csanadi A, Kakanou S, Prasse A, Kassem A, Stickeler E, et al. Downregulation of MTSS1 expression is an independent prognosticator in squamous cell carcinoma of the lung. *Br J Cancer*. 2015; 112:866–873. [PubMed: 25625275]
31. Giacobbe A, Compagnone M, Bongiorno-Borbone L, Antonov A, Markert EK, Zhou JH, et al. p63 controls cell migration and invasion by transcriptional regulation of MTSS1. *Oncogene*. 2016; 35:1602–1608. [PubMed: 26119942]
32. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. *Science*. 2017; 355:1152–1158. [PubMed: 28302823]
33. Manceau G, Letouzé E, Guichard C, Didelot A, Cazes A, Corté H, et al. Recurrent inactivating mutations of *ARID2* in non-small cell lung carcinoma. *Int J Cancer*. 2013; 132:2217–2221. [PubMed: 23047306]
34. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat J-P, et al. A Landscape of Driver Mutations in Melanoma. *Cell*. 2012; 150:251–263. [PubMed: 22817889]
35. Li M, Zhao H, Zhang X, Wood LD, Anders RA, Choti MA, et al. Inactivating mutations of the chromatin remodeling gene *ARID2* in hepatocellular carcinoma. *Nat Genet*. 2011; 43:828–829. [PubMed: 21822264]
36. Flowers S, Beck GR, Moran E. Transcriptional Activation by pRB and Its Coordination with SWI/SNF Recruitment. *Cancer Res*. 2010; 70:8282–8287. [PubMed: 20851996]
37. Tordella L, Khan S, Hohmeyer A, Banito A, Klotz S, Raguz S, et al. SWI/SNF regulates a transcriptional program that induces senescence to prevent liver cancer. *Genes Dev*. 2016; 30:2187–2198. [PubMed: 27737960]
38. Romero OA, Torres-Diz M, Pros E, Savola S, Gomez A, Moran S, et al. MAX Inactivation in Small Cell Lung Cancer Disrupts MYC-SWI/SNF Programs and Is Synthetic Lethal with BRG1. *Cancer Discov*. 2014; 4:292–303. [PubMed: 24362264]
39. Wilson BG, Roberts CWM. SWI/SNF nucleosome remodellers and cancer. *Nat Rev Cancer*. 2011; 11:481–492. [PubMed: 21654818]
40. Alver BH, Kim KH, Lu P, Wang X, Manchester HE, Wang W, et al. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat Commun*. 2017; 8:14648. [PubMed: 28262751]
41. Nakayama RT, Pulice JL, Valencia AM, McBride MJ, McKenzie ZM, Gillespie MA, et al. SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. *Nat Genet*. 2017; doi: 10.1038/ng.3958
42. Taylor MD, Bollt O, Iyer SC, Robertson GP. Metastasis suppressor 1 (MTSS1) expression is associated with reduced in-vivo metastasis and enhanced patient survival in lung adenocarcinoma. *Clin Exp Metastasis*. 2018; 35:15–23. [PubMed: 29218652]
43. Lee H-S, Park J-H, Kim S-J, Kwon S-J, Kwon J. A cooperative activation loop among SWI/SNF,  $\gamma$ -H2AX and H3 acetylation for DNA double-strand break repair. *EMBO J*. 2010; 29:1434–1445. [PubMed: 20224553]
44. Niimi A, Chambers AL, Downs JA, Lehmann AR. A role for chromatin remodellers in replication of damaged DNA. *Nucleic Acids Res*. 2012; 40:7393–7403. [PubMed: 22638582]
45. Ray A, Mir SN, Wani G, Zhao Q, Battu A, Zhu Q, et al. Human SNF5/INI1, a Component of the Human SWI/SNF Chromatin Remodeling Complex, Promotes Nucleotide Excision Repair by Influencing ATM Recruitment and Downstream H2AX Phosphorylation. *Mol Cell Biol*. 2009; 29:6206–6219. [PubMed: 19805520]
46. Rugo HS, Olopade OI, DeMichele A, Yau C, van 't Veer LJ, Buxton MB, et al. Adaptive Randomization of Veliparib-Carboplatin Treatment in Breast Cancer. *N Engl J Med*. 2016; 375:23–34. [PubMed: 27406347] Pan D, Kobayashi A, Jiang P, Ferrari de Andrade L, Tay RE, Luoma A, et al. A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. *Science*. 2018;eaao1710.
48. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat J-P, et al. A Landscape of Driver Mutations in Melanoma. *Cell*. 2012; 150:251–263. [PubMed: 22817889]
49. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018; 28:1747–1756. [PubMed: 30341162]

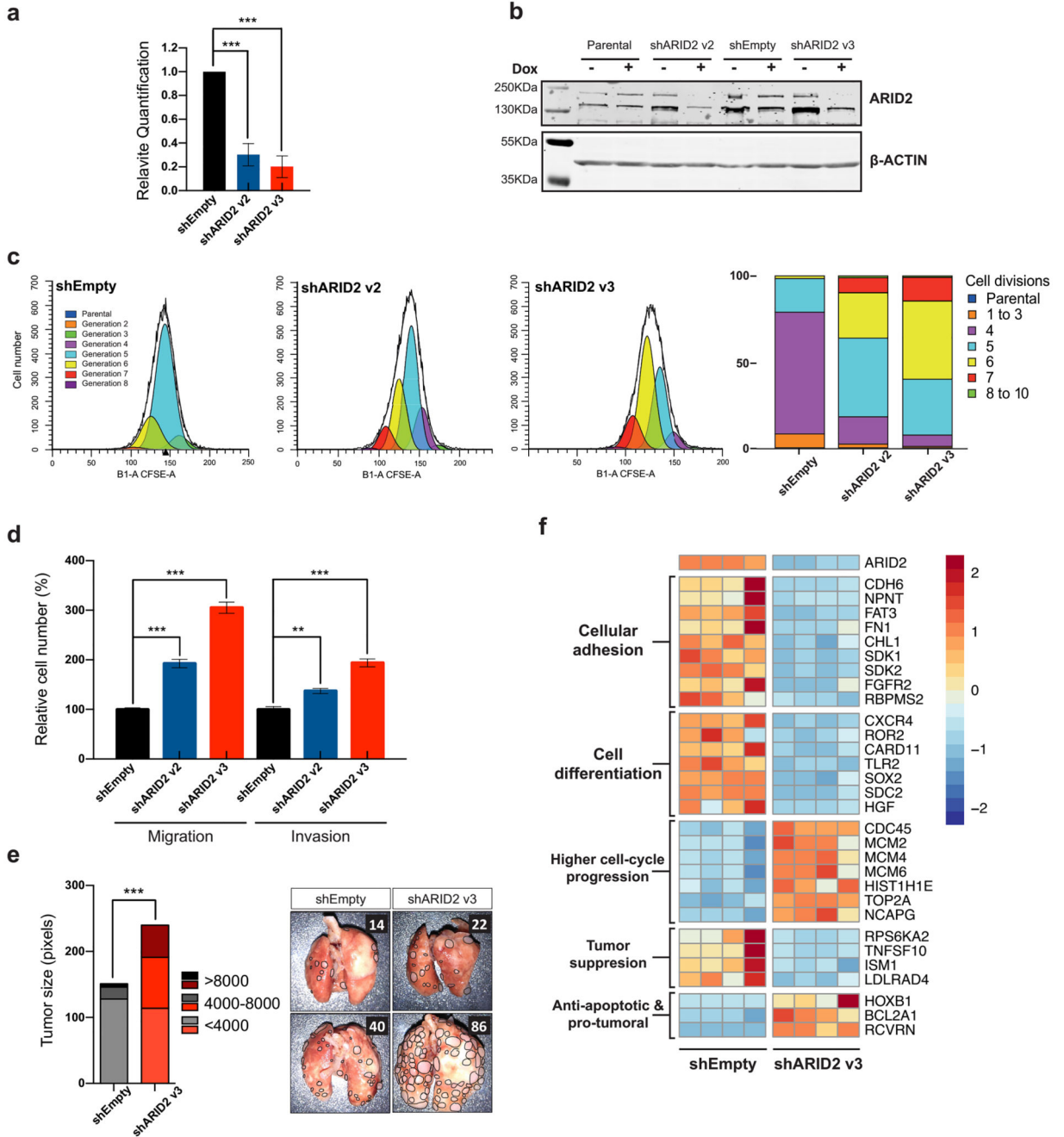


**Figure 1. Frequent *ARID2* mutations associated with protein loss.**

(a) Box representation of the mutated patients for the most significantly mutated genes according to OncodriveFML in the lung cancer cohort generated with Maftools<sup>49</sup>. Each box in the central matrix represents an independent patient. Colored boxes represent mutated patients for the corresponding gene in a color code indicating the type of mutation. (b) Representation of the significance analysis of the functional impact of the mutations found in each gene performed by OncodriveFML. Genes in read showed a q-value < 0.1 after multi-test correction. (c) Visual representation of the location of all identified *ARID2*

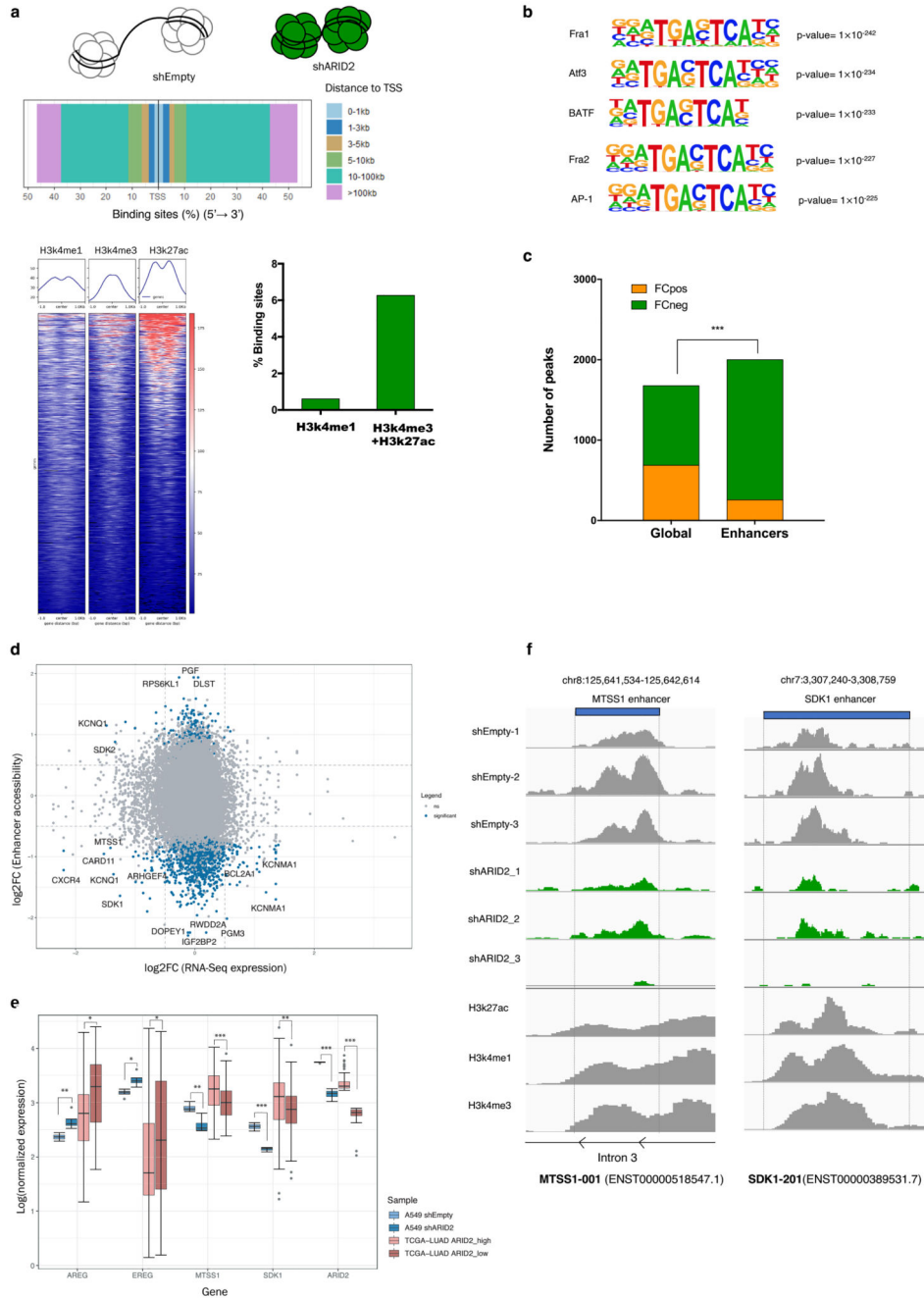
mutations in our discovery and validation lung cancer cohort in relation to the functional protein domains. (d) Representative images of ARID2 immunohistochemistry experiments in two *ARID2*-mutated (right) and two *ARID2*-wildtype (left) lung adenocarcinomas.





**Figure 2. ARID2 deficiency is associated with an increase in oncogenesis *in vitro* and *in vivo*.** (a) Bar representation of ARID2 expression level fold changes measured by qRT-PCR in A549 cells transduced with shARID2 v2 and v3 as well as the empty vector which is used as control. Data is shown as mean  $\pm$  SEM of three independent experiments, relative to control cells A549 Empty vector (black bars), (two-tailed t-test \* $p < 0.05$ , \*\*  $p < 0.01$  and \*\*\*  $p < 0.001$ ). (b) Representative image of a western blot analysis measuring ARID2 protein levels in A549 parental cells as well as those cell lines transduced with ARID2 shRNAs and the empty vector. In all the cases, the results are shown with and without induction of the

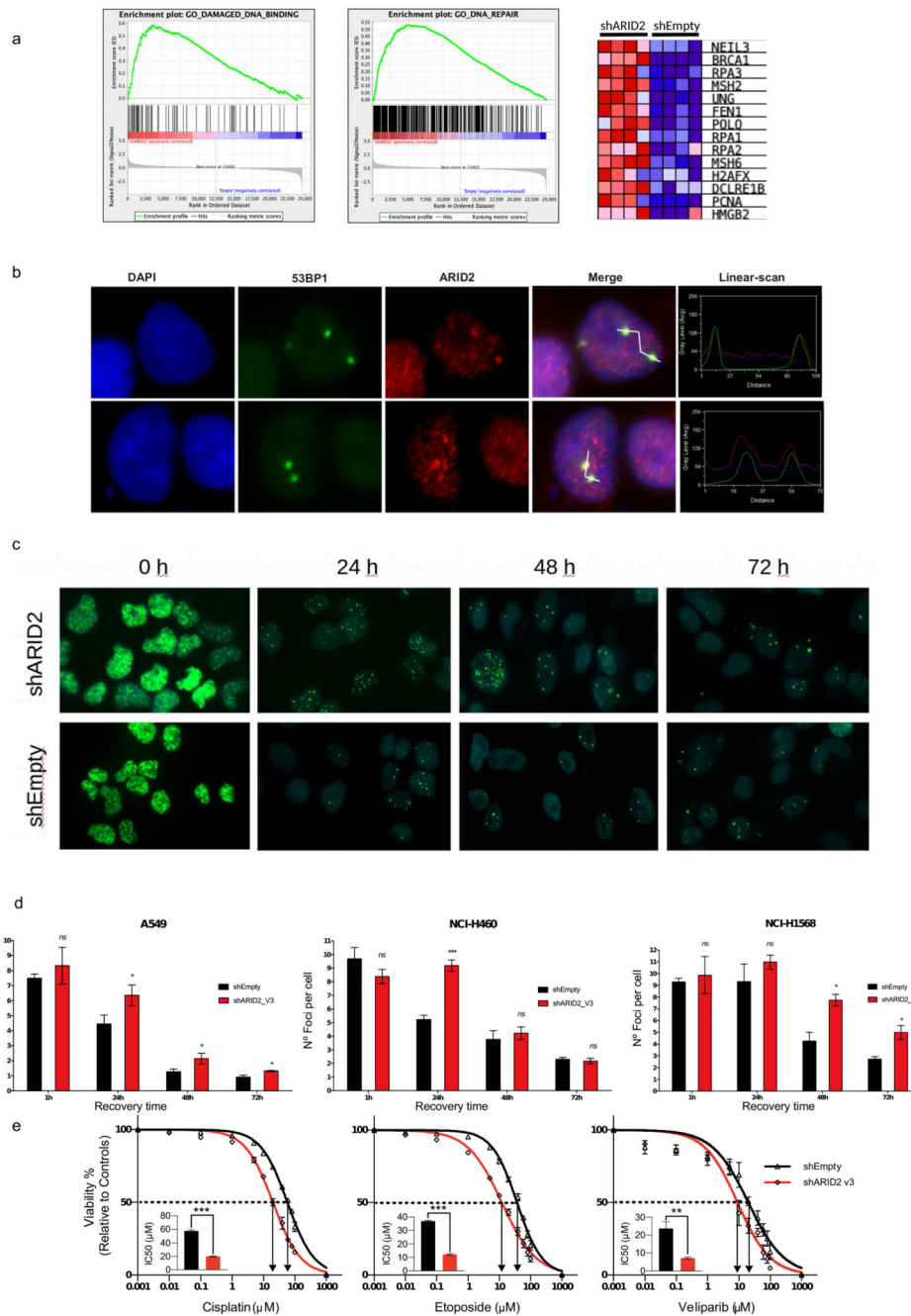
shRNA expression by doxycycline (Dox) treatment. (c) Representative experiment of the number of cell divisions suffered by the cells in 48h estimated by CFSE labelling in A549 cells by flow cytometry. Bar quantification on the number of cells that have suffered each number of cell divisions is represented on the right. (d) Bar representation of quantified cells in destination chamber on migration and invasion assays of A549 cells transduced with two different ARID2 shRNAs (blue and red bars). Data is shown as mean  $\pm$  SEM of three independent experiments, relative to control cells A549 Empty vector (black bars), (two-tailed t-test \* $p < 0.05$ , \*\*  $p < 0.01$  and \*\*\*  $p < 0.001$ ). (e) Representative images of lung metastasis generated in intravenously injected mice with A549 cells transduced either with shEmpty, or shARID2 v3 vectors (n=7 per group). Individual metastasis are delineated in the image and counted (upper right corner numbers). On the left, a quantification of the number and size of the tumors generated in the two groups is shown. (Fisher exact test \* $p < 0.05$ , \*\*  $p < 0.01$  and \*\*\*  $p < 0.001$ ). (f) Heatmap representation of a selection of differently expressed genes in ARID2-deficient A549 cells (n=4) and grouped according to their biological function. Expression differences go from red (upregulation) to blue (downregulation) according to the  $\log_2$  of the fold change.



**Figure 3. Profound chromatin structural changes on enhancers affect gene expression after ARID2 loss.**

(a) Analysis of the genomic regions that significantly lost chromatin accessibility after ARID2 loss in A549 cells. In the upper panel the regions are grouped according to their distance to nearest gene transcription start site (TSS). Below, the intensity of H3K4me1, H3K4me3 and H3k27ac histone marks in each identified region is represented by heatmaps (left). Additionally, the percentage of identified regions that overlap with regions with histone modification marks are represented in a bar graph (right). (b) Enrichment of sequence motifs identified by HOMER in those regions that lost chromatin accessibility

after ARID2-loss in A549 cells. (c) Bar representation of the number of genome-wide (global) regions or enhancer regions which lost (FCneg green bars) or gained (FCpos yellow bars) after ARID2 loss in A549 cells. (Fisher exact test \*\*\*  $p < 0.001$ ) (d) Dot plot representing the correlation between the accessibility changes in enhancer regions and expression changes on the target genes for each enhancer. Blue dots represent enhancers that showed significant accessibility changes in ARID2-deficient A549 cells. (e) Boxplot graph of gene expression differences identified in both our ARID2-deficient A549 cells and in lowly *ARID2* expressing lung adenocarcinoma patients (*ARID2\_low*) versus highly *ARID2* expressing patients (*ARID2\_high*) from TCGA database, (DEseq2 statistical test \* $p < 0.05$ , \*\*  $p < 0.01$  and \*\*\*  $p < 0.001$ ). (f) Visualization, in two described *MTSS1* and *SDK1* enhancers, of read alignments for the different replicates of our ATAC-Seq experiments in ARID2-deficient A549 cells. Additionally, read alignments of ChIP-Seq experiments performed against different histone marks during ENCODE project are also represented.



**Figure 4. ARID2 deficiency affects DNA repair and affects sensitivity to anti-tumor therapies**  
 (a) Results of the Gene-Set Enrichment Analysis (GSEA) from RNA-Seq experiments in ARID2-deficient A549 cells showing enrichment of genes involved in different DNA repair ontologies. On the right, a heatmap representation of the expression of different genes belonging to these ontologies in the different replicates is included. (b) Representative images of immunofluorescence experiments demonstrating colocalization of 53BP1 (green) and ARID2 (red) in A549 cells in DNA damage foci. Colocalization was confirmed on the right through the parallel quantification of red and green signals on a manually selected path

through the image using the LineScan tools from Methamorph software (c) Representative images of DNA-repair foci visualized by H2AX immunofluorescence (green) in transduced nuclei stained with DAPI (blue) at different recovery times after the treatment with etoposide in NCI-H460 cell lines. (d) Bar representation of the foci quantification in each transduced cell line. (d) Bar quantification of the number of foci per cell at different recovery times after DNA damage induced by etoposide, in different ARID2-deficient cell lines, the results are represented as mean  $\pm$  SEM of at least three independent experiments, (two-tailed t-test \* $p < 0.05$ , \*\*  $p < 0.01$  and \*\*\*  $p < 0.001$ ). (e) Representative experiments measuring cell survival to increasing concentrations of cisplatin, etoposide and veliparib on A549 cells transduced with shEmpty (black), or shARID2v3 (red) vectors. Bar graphs represent the calculated IC50 value for each experiment. In all cases, the results are represented as mean  $\pm$  SEM of at least three independent experiments, (two-tailed t-test \* $p < 0.05$ , \*\*  $p < 0.01$  and \*\*\*  $p < 0.001$ ).