Facultad de **Medicina**

GRADO EN CIENCIAS BIOMÉDICAS

TRABAJO FIN DE GRADO

Escribir Título TFG en Castellano

Caracterización del papel de las deficiencias en SWI/SNF en la progresión del cáncer de pulmón.

Escribir Título TFG en Inglés

Characterization of the role of SWI/SNF deficiencies in the progression of lung cancer.

Autor/a: Paula Díez Vicente

Director/a: Ignacio Varela Egocheaga

Co-director/a:

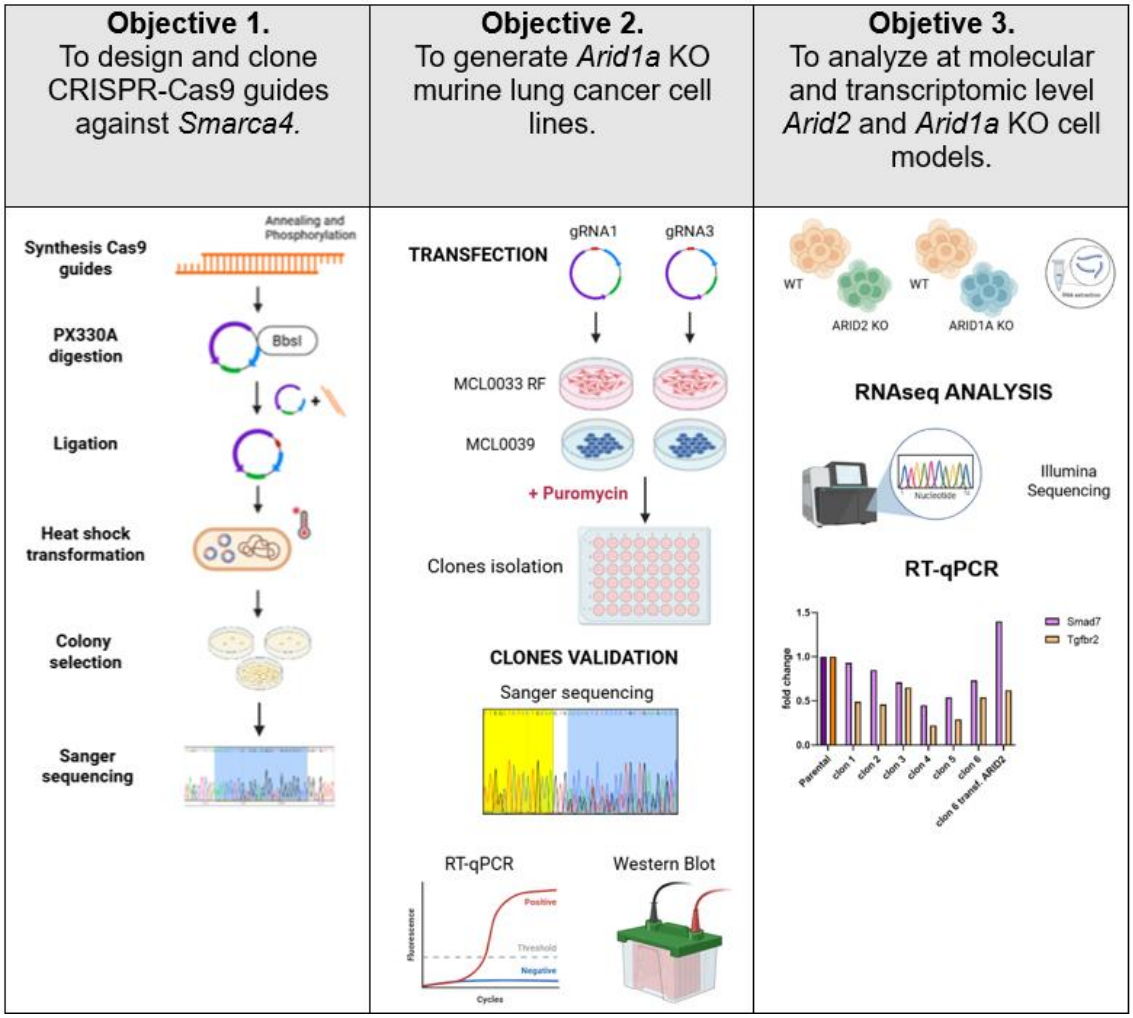
Co-director/a:

Santander, 8 junio 2025

Index

Characterization of the role of SWI/SNF deficiencies in the progression of lung cancer

GRAPHICAL ABSTRACT



AUTHORS

Paula Díez Vicente, Ignacio Varela Egocheaga.

Correspondence: Ignacio.varela@unican.es

Trabajo de Fin de Grado
Grado en Ciencias Biomédicas · Facultad de Medicina
2024 – 2025

Characterization of the role of SWI/SNF deficiencies in the progression of lung cancer

Paula Díez Vicente[1], Ignacio Varela Egocheaga[1*]

[1]Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC). Universidad de Cantabria – CSIC, Santander, Spain.

* Correspondence: Ignacio.varela@unican.es

RESUMEN

SWI/SNF es una familia de complejos remodeladores de la cromatina implicados en la regulación de la expresión génica, la reparación del DNA y la diferenciación celular. Se han encontrado alteraciones inactivantes en distintas subunidades de estos complejos en más del 20 % de los cánceres humanos, especialmente en el cáncer de pulmón no microcítico (NSCLC); donde se asocian a mal pronóstico, mayor invasividad y resistencia a las terapias convencionales. En este contexto, una comprensión más profunda de los mecanismos moleculares que sustentan la función supresora tumoral de estos complejos puede revelar nuevas dianas terapéuticas y ofrecer nuevas oportunidades de tratamiento para los pacientes.

El presente estudió consistió en la generación y caracterización de modelos celulares murinos deficientes en *Arid1a* y *Arid2*, dos de las subunidades más frecuentemente mutadas en adenocarcinoma, junto con *Smarca4.* Posteriormente, se realizaron análisis moleculares y transcriptómicos a partir de librerías de RNA-seq obtenidas de clones knockout y wildtype, con el objetivo de identificar rutas moleculares afectadas por estas deficiencias y potenciales dianas transcripcionales en el contexto del cáncer de pulmón.

ABSTRACT

SWI/SNF is a family of chromatin remodeling complexes involved in gene expression regulation, DNA repair and cellular differentiation. Its role as tumor suppressor has been described reporting inactivating mutations in SWI/SNF subunits in over 20% of human malignancies - particularly in NSCLC - where such alterations are associated with poor prognosis, increased invasiveness and resistance to conventional therapies. In this context, a deeper understanding of the molecular mechanisms underlying the tumor suppression roles of these complexes may unveil new therapeutic targets and provide novel treatment opportunities for patients.

The present study involved the generation and molecular characterization of murine lung cancer cell models deficient in *Arid1a* and *Arid2*, two of the most frequently mutated SWI/SNF subunits in lung adenocarcinoma, along with *Smarca4*. Transcriptomic analyses were subsequently performed using RNA-seq libraries obtained from both knockout and wild-type clones, in order to identify molecular pathways altered after the loss of these subunits as well as gene targets of them in the context of lung cancer.

KEYWORDS

SWI/SNF complex, lung adenocarcinoma, *Arid2*, *Arid1a*, *Smarca4*, gene expression, RNAseq libraries.

INTRODUCTION

Lung cancer has been the leading cause of cancer-related mortality in the current century due to its high prevalence, in both men and women, combined with its significant lethality. It is the most prevalent cancer in men, following by prostate cancer; and the second most common in women. The majority of cases are diagnosed at an advanced stage of the disease when metastasis has already occurred; resulting in a poor prognosis with an overall ratio of mortality-to-incidence of 0.87.[1] According to the International Agency for Research on Cancer, the overall survival rate for both sexes was below 27% in 2022. This makes it a global health issue and highlights the urgent need for further research into the development of novel therapies aimed at improving long-term survival.

Concerning the histological properties, lung cancer can be divided in two main different entities: non-small cell lung cancer (NSCLC), which is highly aggressive and accounts for approximately 85% of the new lung cancer cases and small cell lung cancer (SLCL).[2]

SCLC is a high-grade neuroendocrine carcinoma that represents about 15% of all lung cancer cases. It has been associated with tobacco consumption and primary tumors usually appear in central zones of the lungs. Regarding the genomic profile, SCLC is characterized by a dual inactivation of the tumor suppressors *TP53* and *RB*, although inactivating mutations of *NOTCH* family genes were also observed in 25% of the SCLC patients of this study.[3] These alterations result in an exceptionally aggressive phenotype, which leads to an undesirable prognosis with two thirds of the new cases having metastasis at initial diagnosis.[4] An additional problem is that a combination of chemotherapy and radiotherapy remains the available treatment option for advanced-stage cases, which account for the vast majority of all diagnoses (87%). [5]

NSCLC has been described as a group of heterogenous diseases. It can be histopathologically classified into two predominant subtypes: adenocarcinoma (ADC) and squamous cell carcinoma (SCC). ADC is the most common subtype, accounting for approximately 50% of all lung cancer cases. It frequently harbors activating mutations in *EGFR* and *KRAS*, as well as gene rearrangements involving *ALK*, *ROS1*, and *RET*, particularly among non-smokers. SCC, more closely linked to tobacco exposure, is characterized by early and frequent loss of *TP53*, amplification of *FGFR1* and alterations in the PIK3CA-PTEN pathway. These molecular events are often detectable in premalignant lesions and have emerged as potential therapeutic targets.[2,6]

Currently, therapeutic options for NSCLC patients with advanced disease depend on the presence of actionable driver mutations. Patients with *EGFR* mutations or *ALK* rearrangements receive targeted therapy as first-line treatment, while those lacking such alterations or with unknown status continue to receive platinum-based doublet chemotherapy. In such cases, immune checkpoint inhibitors (ICIs) represent an additional therapeutic option following disease progression. [2,7]

Despite the progress made in molecular targeted therapies over the past decades, drug resistance to third-generation EGFR-TKI inhibitors, such as osimertinib, commonly arises in patients after 1–2 years of treatment [8] and five-year survival rates (15.9%) have barely improved.[9] This highlights the urgent need for continued research into novel strategies and druggable targets to address these aggressive cancers effectively.

High-throughput tumor sequencing in patients has uncovered a significant prevalence of mutations in genes involved in chromatin structure regulation. In particular, recent studies highlight the SWI/SNF family role as a tumor suppressor, with specific nonsense,

frameshit and deletion mutations of individual SWI/SNF subunits found in numerous cancer types. Mutations affecting this family subunits have been reported in over 20% of human malignancies. [10]

Switch/sucrose nonfermentable (SWI/SNF) complex is a family of multisubunit chromatin remodeler proteins that contribute to transcriptional activation using the energy of ATP hydrolysis. These highly evolutionarily conserved complexes bind to gene promoters or enhancers and make chromatin more accessible either by sliding nucleosomes or by promoting their ejection, allowing transcription machinery to bind to the DNA. This process is involved in transcription regulation, DNA repair and lineage differentiation, among other cellular functions.[11–13]

Three SWI/SNF subfamilies have been identified in mammalian cells based on their composition: the canonical BAF complex (Cbaf), the polybromo-associated BAF (PBAF) and the non-canonical BAF complex (ncBAF). Each subfamily is composed of 8 to 12 subunits, which are encoded in 29 different genes in humans. All complexes consist of an ATPase subunit, either SMARCA4 or SMARCA2, which are mutually exclusive; a set of considerably conserved core subunits, as SMARCC1/BAF155, SMARCC2/BAF170 and SNF5/SMARCB1; and several accessory subunits. These accessory subunits are thought to stabilize the complex or bind to different targets and provide complexes a different identity or tissue specificity. For instance, ARID1A and ARID1B are specific to cBAF and also mutually exclusive paralogues whereas ARID2 and PBRM1 appear only in the PBAF, which preferentially binds to promoters. The ncBAF subfamily, which was the last to be discovered, contains GLTSCR1 or GLTSCR1L subunits instead of ARID proteins. It is also characterized by BRD9 subunits, whose bromodomain allows the complex to bind to the DNA. [12,14,15]

Genetic alterations typically occur primarily in the catalytic and accessory subunits, whereas core ones rarely present mutations. An exception is the loss-of-function mutation of *SNF5/SMARCB1* in the pediatric malignant rhabdoid tumors, the first evidence of the contribution of SWI/SNF deficiencies to oncogenesis. The role of *SNF5* in MRT has been well characterized since its discovery in 1998. Biallelic mutations of it leads to alterations of transcription, disrupting Hedgehog, WNT and MYC oncogenic pathways. However, the mechanisms by which mutations in other SWI/SNF subunits promote tumorigenesis remain in some measure understood.[10]

Mutations in other SWI/SNF subunits have been also correlated with disease. For instance, *PBRM1* is altered in close to 40% of clear cell renal carcinoma cases. [15] Additionally, this recent study conducted by Elwy, A. E. *et al*. reveals that reduced *PBRM1* expression levels (<50%) are strongly associated with invasiveness, high density of immune infiltrate and tumor progression. A significant descend in 3-year overall survival (OS) and disease – free survival was also observed. [16]

Moreover, some cancer-associated SWI/SNF mutations can create dependencies on other subunits, [15] suggesting that the oncogenic contribution of SWI/SNF mutations results from anomalous assembly of the remaining subunits. For instance, SMARCA2 dependency has been reported in *SMARCA4*-mutated cell lines (H1299 and A549), where SMARCA2 is essential for complex assembly and cell proliferation.[17] Similar dependencies have been documented between ARID proteins. Emi Sato and coworkers demonstrated that ARID1B may be a therapeutic target in *ARID1A*-deficient ovarian clear cell carcinomas (OCCC). They saw that ARID1B is vital for the proper cBAF

complex assembly in this type of malignancy. [18] This phenomenon also occurs between non-paralogue subunits. This study identifies BRD9-containing complex, also known as ncBAF complex, crucial for the survival of *SMARCB1*-deficient cells in malignant rhabdoid tumors (RT), compensating the loss of *SMARCB1* by incorporating other SWI/SNF subunits.[19] Consequently, targeting SMARCA2, ARID1B and BRD9 respectively could be a promising synthetic lethality strategy for treating these aggressive tumors.

In lung cancer, molecular alterations of the SWI/SNF family are commonly found in NSCLC. It has been seen that these alterations not only contribute to tumor progression but they also weaken DNA repair cell capacity. Varela's group research is currently focused on *ARID1A*, *ARID2* and the ATP-ase *SMARCA4*, due to the fact that these are the genes more frequently mutated. [14]

ARID1A deficiency is found in approximately 5-10% of lung cancer cases. It has been linked with high risk of metastasis and poor prognosis.[20] In terms of treatment sensitivity, resistance to EGFR-TKIs induced by ARID1A loss has been described in patients suffering from NSCLC, although the mechanisms behind remain uncertain.[21]

SMARCA4 appears to be altered in 10% of NSCLC and in 8% of lung adenocarcinoma total cases.[14,22] Truncating alterations in *SMARCA4*, as well as co-mutations in *SMARCA4* and *KRAS*, have been associated with reduced sensitivity to immunotherapy and KRAS G12C inhibitors, larger tumor size and an increased metastasis propensity to the brain and liver; ultimately resulting in decreased overall survival in NSCLC patients compared to those with *SMARCA4* wild-type tumors.[23] In this study, Lingling Xu and coworkers have seen that *SMARCA4*-deficient NSCLC patients have worse prognosis than those with alterations in *EGFR*, *ALK* or *ROS1*. They do neither respond well to conventional chemotherapy. [24]

Moreover, loss of ARID2 protein expression was observed in the 20% of a cohort of lung cancer patients. This loss could be attributed to loss-of-function mutations and is associated with increased proliferative potential and invasiveness, as demonstrated both *in vitro* and *in vivo*. [25]

Apart from that, deficiency in the SWI/SNF chromatin remodeling complex has been associated with increased sensitivity to immunotherapy. Tumors harboring SWI/SNF mutations may exhibit a more immunogenic tumor microenvironment characterized by increased T-cell infiltration and elevated PD-L1 expression; suggesting that these alterations may serve as predictive biomarkers for immunotherapy responsiveness. [11,12] In particular, inactivating alterations in subunits such as *ARID1A* or *SMARCA4* can affect interferon signaling, antigen presentation pathways as well as immune cell recruitment. [26,27]

In summary, lung cancer is still responsible for millions of deaths every year worldwide and although recent immune and targeted therapies have shown promising results, the limited efficacy of these in many cases as well as the development of treatment resistance, indicates the needs to find new therapeutic opportunities for the patients of this dramatic disease. In this context, a better understanding of the molecular mechanisms behind the tumor suppression roles of SWI/SNF complexes offers new opportunities to better understand at the molecular level lung tumors as well for the design of new antitumoral therapies.

RESULTS

We hypothesized that a better knowledge of the molecular mechanisms regulated by SWI/SNF could increase our molecular knowledge about lung cancer progression as well as propose new targets to improve lung cancer patients' therapy. According to that, we propose to generate and characterize at the molecular level SWI/SNF deficient lung cancer cell models. This general aim is divided in the three following concrete objectives:

1. To design and clone CRISPR-Cas9 guides against *Smarca4*.
2. To generate *Arid1a* knockout (KO) murine lung cancer cell lines.
3. To analyze at the molecular level *Arid2* and *Arid1a* KO cell models.

Design of CRISPR-Cas9 guides and cloning for *Smarca4* disruption.

As Varela's lab is interested in studying the effect of inactivating different SWI/SNF subunits, they had already generated constructs expressing guide RNAs against *Arid1a* and *Arid2*. In this first objective of my TFG, I was in charge of generated the constructs containing guide RNAs against *Smarca4.* We aimed at generating different constructs containing three gRNAs into pX330A plasmid, which expresses the Cas9 and ampicillin and puromycin resistance genes. The sgRNAs sequences are provided in Supplementary Table 1 and the plasmid's structural map is illustrated in Figure 1A.

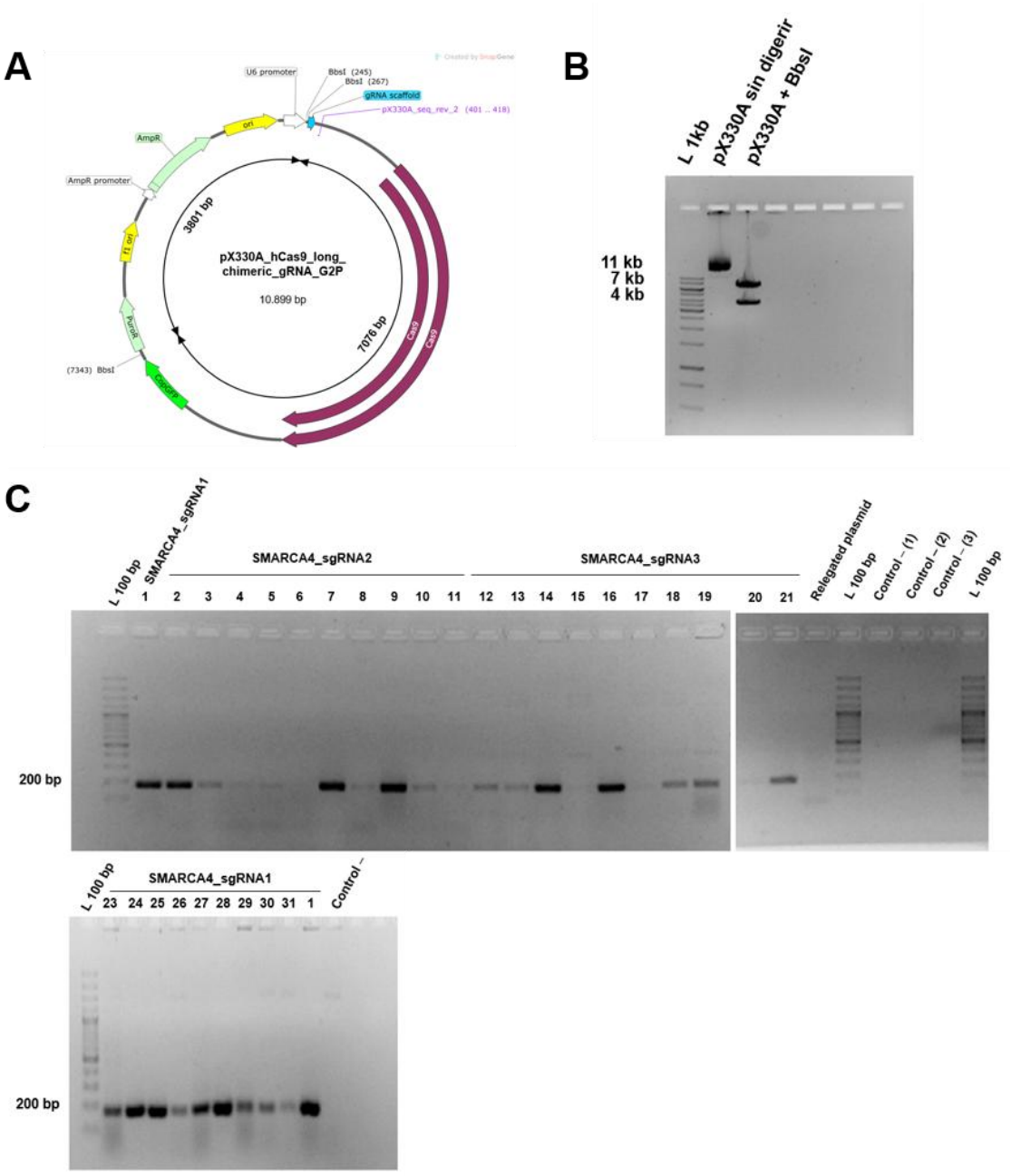


Figure 1. Design of CRISPR-Cas 9 guides and cloning for *Smarca4* disruption.

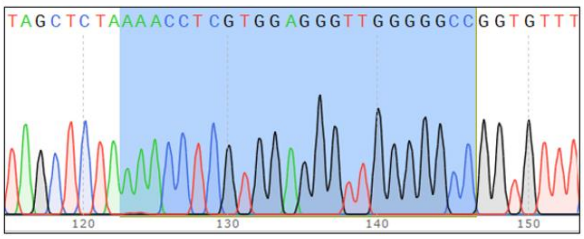
(A) Map of pX330A plasmid. (**B**) Digestion of the plasmid with the BbsI restriction enzyme. **(C)** Colony PCR screening of *E.Coli* transformants harboring ligation products of BbsI-digested pX330A plasmid and three gRNAs targeting *Smarca4*.

The pX330A plasmid was successfully digested with fast digest restriction enzyme BbsI, which has three recognition sites within the plasmid. The band pattern obtained following digestion is shown in Figure 1B.

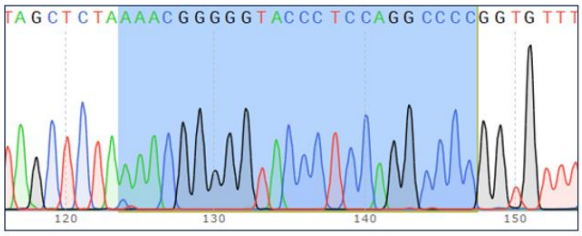
gRNA insertion sequences were generated by the phosphorylation and hybridization of synthetical complementary oligonucleotides, designed to generate compatibly cohesive sequences.

Following ligation and transformation, positive colonies were initially screened by PCR and subsequently confirmed by Sanger sequencing. Chromatograms confirming gRNA insertion in the plasmids are shown in Figure 2. For the PCR screening, we employed the forward oligonucleotide used for insert synthesis in each case as primer forward, and the pX330A_seq_rev2 as the reverse. This strategy allowed us to distinguish colonies that had successfully incorporated the insert (the guide) at the restriction site – yielding an amplification product of 173 bp – from those containing the relegated plasmid backbone, which did not produce any amplification (Figure 1C). The primer sequence is detailed in Supplementary Table 1B.

A



B



C

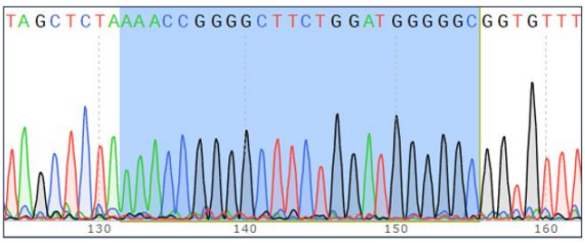


Figure 2. Confirmation of gRNA insertion by Sanger Sequencing in selected bacterial colonies.

(A) Base chromatography confirming that gRNA1 was correctly inserted in the plasmid. Colony 28 was selected as representative of positive clone. **(B)** Base chromatography confirming that gRNA2 was correctly inserted in the plasmid. Colony 7 was selected as representative of positive clone. **(C)** Base chromatography confirming that gRNA3 was correctly inserted in the plasmid. Colony 16 was selected as representative of positive clone.

The gRNA1 transformation was repeated using the remaining ligation product, as the initial transformation yielded low efficiency. This second attempt resulted in a higher number of colonies.

The complete procedure for cloning and transformation is detailed in methods section as well as DNA extraction protocol.

Generation of *Arid1a* KO clones

MCL0039 y MCL0033 Red Fluc cells were transfected by lipofection with two different constructs containing two different Cas 9 guides targeting *Arid1a* gene, gRNA1 and gRNA3. Transfection efficiency at 24 hours post-transfection was notably lower in the MCL0033 Red Fluc cell line compared to MCL0039, as shown in Figure 3. Subsequently, cells that had incorporated the plasmid were positively selected by a 48h-treatment with puromycin, as described in methods section. The construct used for transfection had been previously generated in Varela's lab following the same protocol employed for cloning the *Smarca4* Cas9 guides.

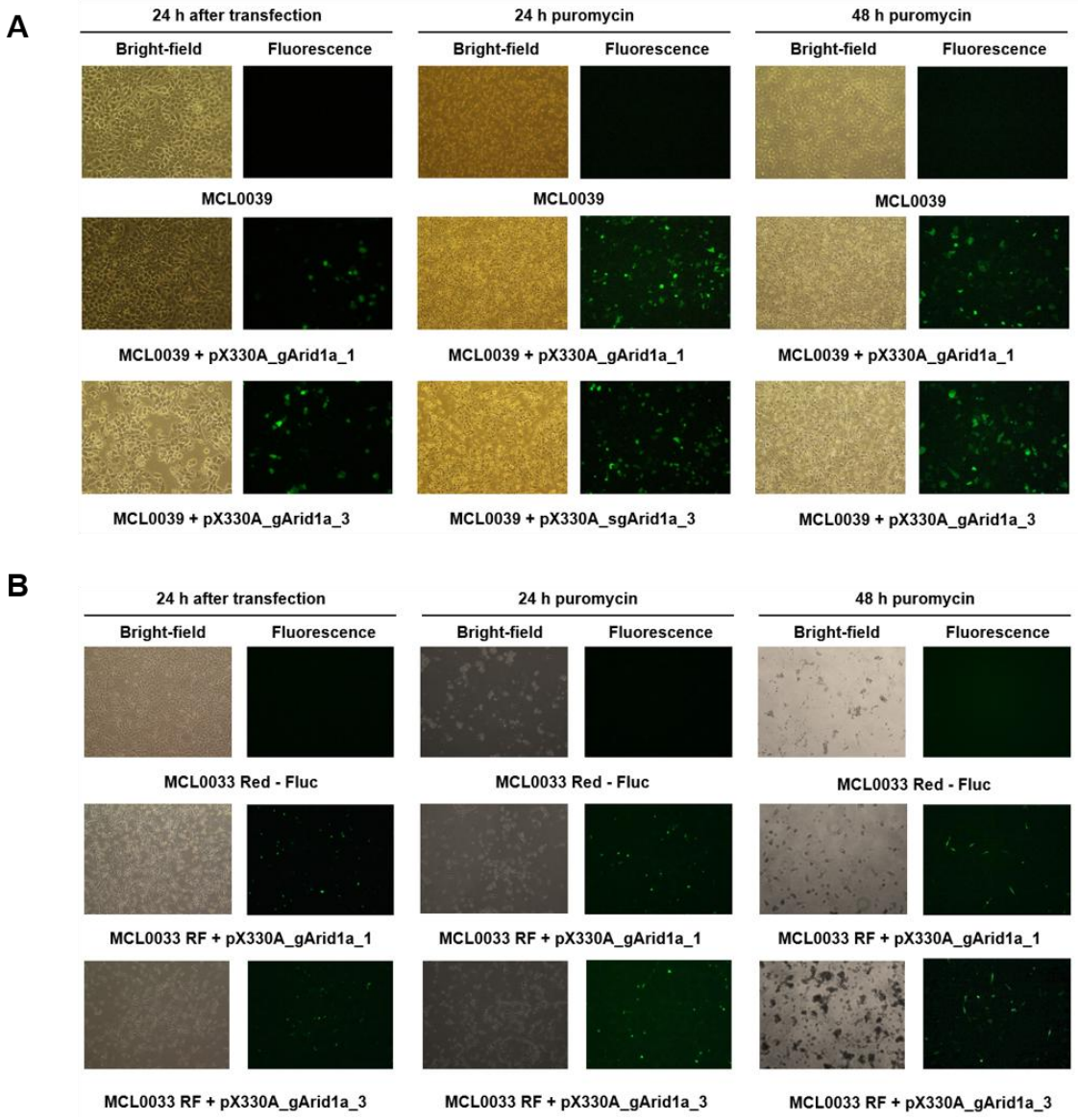


Figure 3. Transfection efficiency of MCL0039 and MCL0033 Red – Fluc cell lines with pX330A_gArid1a_1 and pX330A_gArid1a_3.

(A) Representative fluorescence microscopy pictures showing transfection efficiency in MCL0039 cell line after transfection with the constructs pX330A_gArid1a_1 and pX330A_gArid1a_3. **(B)** Representative fluorescence microscopy pictures showing transfection efficiency in MCL0033 Red - Fluc cell line after transfection with the constructs pX330A_gArid1a_1 and pX330A_sgArid1a_3.

At this point, cells were counted and three new plates were seeded with 2000 cells/plate for each transfection, to allow individual clones to grow separated from one another.

In the case of MCL0039 cell line, 46 clones were isolated 20 days later using cloning cylinders and plated in a 48-well plate to let them grow. 32 clones were successfully expanded. Results obtained by Sanger sequencing showed that clones A1, A2, A4.2, B3.2, B5.2 and C1.2 had incorporated indels as shown in Supplementary Figure 1. Unfortunately, sequencing of the clones transfected with the gRNA1 construct could not be successfully completed, most likely due to the presence of nonspecific amplification bands when amplifying the target region of the *Arid1a* gene in those cases.

On the other side, we faced some difficulties with MCL0033 Red – Fluc cell line. The efficiency of the initial transfection was very low, particularly with the pX330A_gArid1a_3 construct, for which virtually no difference was observed compared to the untransfected control. Consequently, we decided to repeat the procedure. In this occasion, transfection efficiency at 24 hours post-transfection remained somewhat lower in the MCL0033 Red Fluc cell line compared to MCL0039, as shown in Figure 3. However, we chose to proceed with the experiment, as the transfected cells would subsequently be subjected to selection, as detailed in Methods section. Finally, I was only able to isolate and expand 5 clones. Results obtained by Sanger sequencing showed that clone A2 carries a 14 bp deletion in the exon 5 of Arid1a and the mutation seems to be homozygous, as shown in Supplementary Figure 2. In contrast, clones A3 and A4 didn't show any evidence of mutation.

Molecular characterization of *Arid2* KO clones.

Following similar procedures as described above, *Arid2* KO clones had been established in Varela's lab. With the aim of validating whether they had ceased of expressing *Arid2* and they are effectively *Arid2* KO clones, we performed RT-qPCR and Western Blotting.

First, *Arid2* mRNA expression levels were measured by RT-qPCR. A graphical comparison between parental cells and potential knockout clones, including one re-transfected with human *ARID2*, is shown in Figure 4A. Oligonucleotides targeting different exons of the gene were designed in order to detect either the absence of mRNA or the expression of transcripts containing the deletion engineered in the cells, as the Cas9 guides used were specifically directed to exon 5.

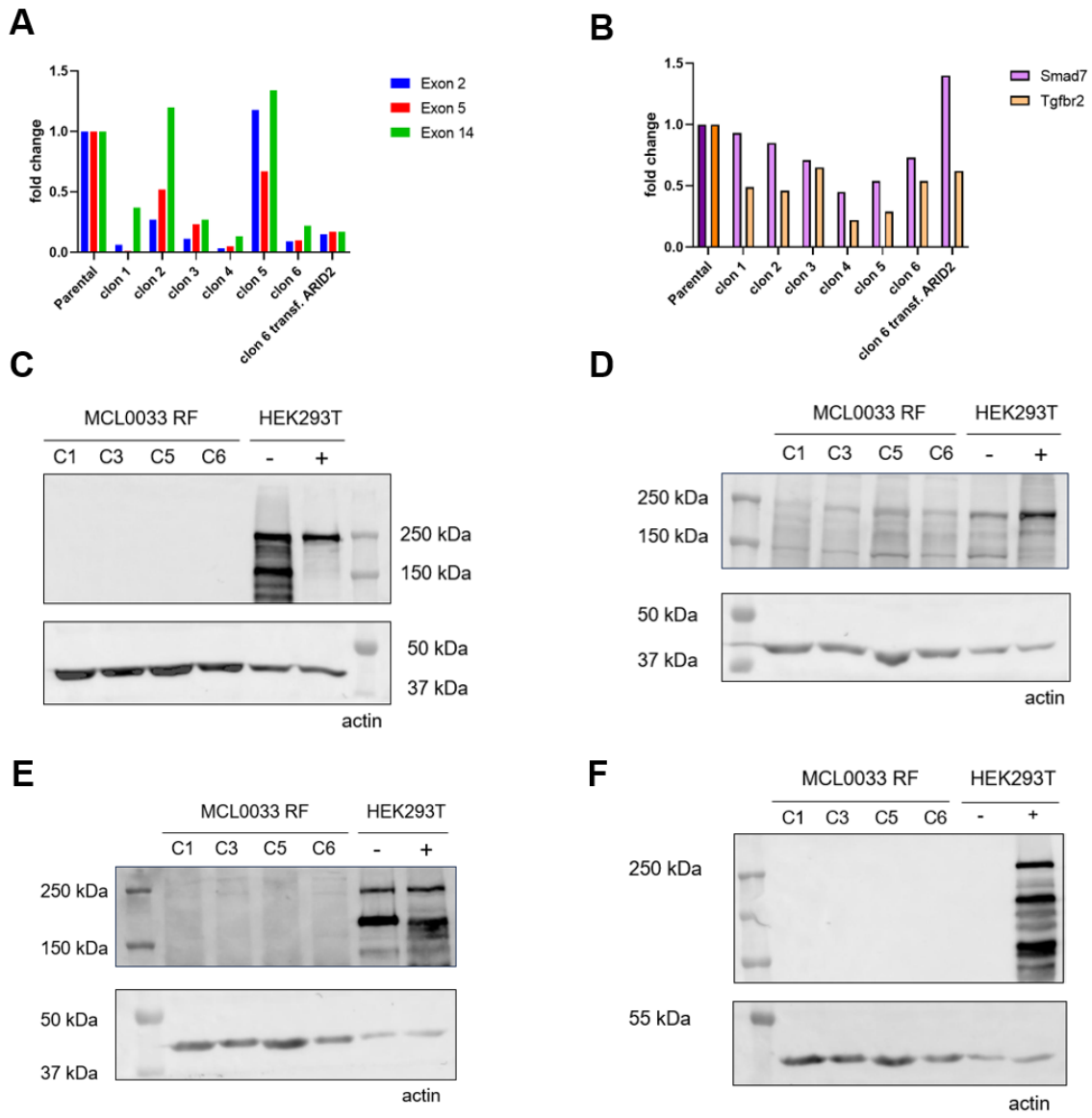


Figure 4. *Arid2* WT and KO clones comparison at mRNA and protein levels.

(A) Bar representation of exons 2, 5 and 14 of *Arid2* expression levels in MCL0033 Red – Fluc potential *Arid2* KO clones measured by RT-qPCR. **(B)** Bar representation of *Smad7* and *Tgfbr2* expression levels in MCL0033 Red – Fluc potencial *Arid2* KO clones measured by RT-qPCR. **(C)** Western blot membrane after incubation with Mouse anti-ARID2 E3 antibody (sc-166117, Santa Cruz Biotechnology) diluted 1:1000 and Goat anti-βactin I-19 antibody (sc-1616, Santa Cruz Biotechnology) diluted 1:1000. **(D)** Western blot membrane after incubation with Rabbit anti-ARID2 NBPI antibody (#26615, Novus Biologicals) diluted 1:1000 and Goat anti-βactin I-19 antibody (sc-1616, Santa Cruz Biotechnology) diluted 1:1000. **(E)** Western blot membrane after incubation with Rabbit anti- ARID2 (A302-229A, Thermo Fisher Scientific) diluted 1:2000 and Goat anti-βactin I-19 antibody (sc-1616, Santa Cruz Biotechnology) diluted 1:1000. **(F)** Western blot membrane after incubation with Mouse anti-V5 Tag antibody (MCA1360GA, Bio-Rad) diluted 1:1000 and Goat anti-βactin I-19 antibody (sc-1616, Santa Cruz Biotechnology) diluted 1:1000.

Parental cells exhibited uniform expression across all assessed exons, as it was set as endogenous control. In contrast, several clones exhibited markedly reduced *Arid2* expression, consistent with a potential KO phenotype. Clones 1, 3, 4 and 6 showed a near-complete loss of mRNA containing exons 2, 5 and 14. In particular, clone 4 displayed expression levels close to zero for the three mentioned exons, suggesting a homozygous loss of *Arid2*.

Clone 2 exhibited a more heterogeneous expression pattern, as exon 14 is expressed at higher levels than in parental cells, while exons 2 and 5 were significantly downregulated. This suggests that the transcript which is being produced contains exon 14 but lacks exons 2 and 5. A similar pattern was observed in clone 5, although in this case only exon 5 was downregulated. This is consistent with the design of the CRISPR-Cas9 guide RNA, which targets exon 5, and may indicate that exon 5 is being skipped during *Arid2* transcription.

Importantly, clone 6 transfected with an *ARID2*-expressing construct ("clon 6 transf. ARID2") did not exhibit a restoration of *Arid2* expression. This can be attributed to the use of a construct encoding the human *ARID2* gene, whereas the RT-qPCR primers were designed to specifically detect the murine transcript. As a result, the human *ARID2* mRNA is likely not amplified in our assay.

Additionally, previous results from Varela's group suggest that SWI/SNF deficiency may affect TGF-β signaling. They have been observed a reduction in TGF-β receptor type II expression levels in human cell models lacking functional SWI/SNF complexes; accompanied by morphological alterations. Notably, some of the knockout clones acquired a more mesenchymal-like morphology, suggesting that these cells may be undergoing an epithelial-to-mesenchymal transition (EMT) or related process. To investigate this in our murine knockout models, we performed RT-qPCR of *Smad7*, a negative regulator of the canonical TGF-β signaling pathway, and Tgf-β receptor type II (*Tgfbr2*).

As you can see in Figure 4B, clones in which *Arid2* levels are reduced exhibited a consistent downregulation of *Tgfbr2* compared to the parental cell line. This reduction was observed across multiple clones. However, contrary to our expectations, *Smad7* levels were also downregulated in these clones. Although they are preliminary results, this may indicate that the impairment of the TGF-β pathway is not mediated through the negative regulator *Smad7*, but rather involves an alternative mechanism.

Moreover, we observed an increased expression of *Smad7* in clone 6 transfected with the *ARID2*-expressing construct compared to untransfected clone 6, which may be linked to a partial restoration of ARID2 function.

We also performed Western blot assays for MCL0033 Red-Fluc clones 1, 3, 5 and 6 in order to assess Arid2 expression levels. Multiple trials were carried out using different antibodies; however, none of them enabled us to detect a band that could be clearly attributed to Arid2 (~240 KDa) in the murine samples (Figures 4C, 4D and 4E). As controls, we included two HEK293T cell samples, one of which was transfected with a plasmid overexpressing the human ARID2.

Initially, we were optimistic about detecting murine Arid2 using the anti-ARID2 NBPI antibody (Figure 4D). However, upon comparison with the positive control, we observed

that the band present in the clones from MCL0033 Red – Fluc cell line appeared to be slightly lower in molecular weight. This discrepancy led us to conclude that the signal was more likely due to nonspecific detection rather than the presence of the target protein.

Furthermore, we performed an additional Western Blot assay in order to confirm that the band observed around 240 KDa (Figures 4C, 4D and 4E) in the HEK293T cell line transfected with a construct encoding the ARID2 protein fused to a V5 tag corresponded with the expressed fusion protein. As shown in Figure 4F, upon incubation of the membrane with the Mouse anti-V5 Tag antibody, signal was detected exclusively in the transfected sample and not in the others. These results confirm that the construct is being expressed.

Transcriptomic profile analysis after *Arid2* or *Arid1a* deficiency.

Aiming to evaluate the downstream impact of *Arid2* and *Arid1a* loss on gene expression, total RNA was extracted from potential KO and WT clones for each gene and RNA-seq libraries were built up as described in Methods section. Subsequently, sequencing of the samples was performed using the Illumina platform.

Sequenced reads were aligned to the mouse reference genome (GRCm39) and differential gene expression analysis between sample groups was carried out using the DESeq tool. Genes of interest were defined as those exhibiting statistically significant differential expression between the two groups (adjusted p-value < 0.05) and a fold change greater than fourfold ($|\log_2 \text{FoldChange}| > 2$).

This analysis revealed that 43 genes were significantly differentially expressed between KO and WT clones, regardless of the cell line. Among this, 37 genes were downregulated whereas 6 exhibited an upregulated expression.

Figure 5A shows a volcano plot highlighting the genes with upregulated expression in red and those downregulated in blue. These genes are also displayed in a heatmap (Figure 5B). *S100a14, Aldh1a1, Ptpn6* and *F13a1* are some of the genes that have a clearer differential expression pattern between KO and WT clones. While *S100a14* and *Aldh1a1* are downregulated in KO clones compared to WT, *F13a1* appears to be upregulated following *Arid2* loss.

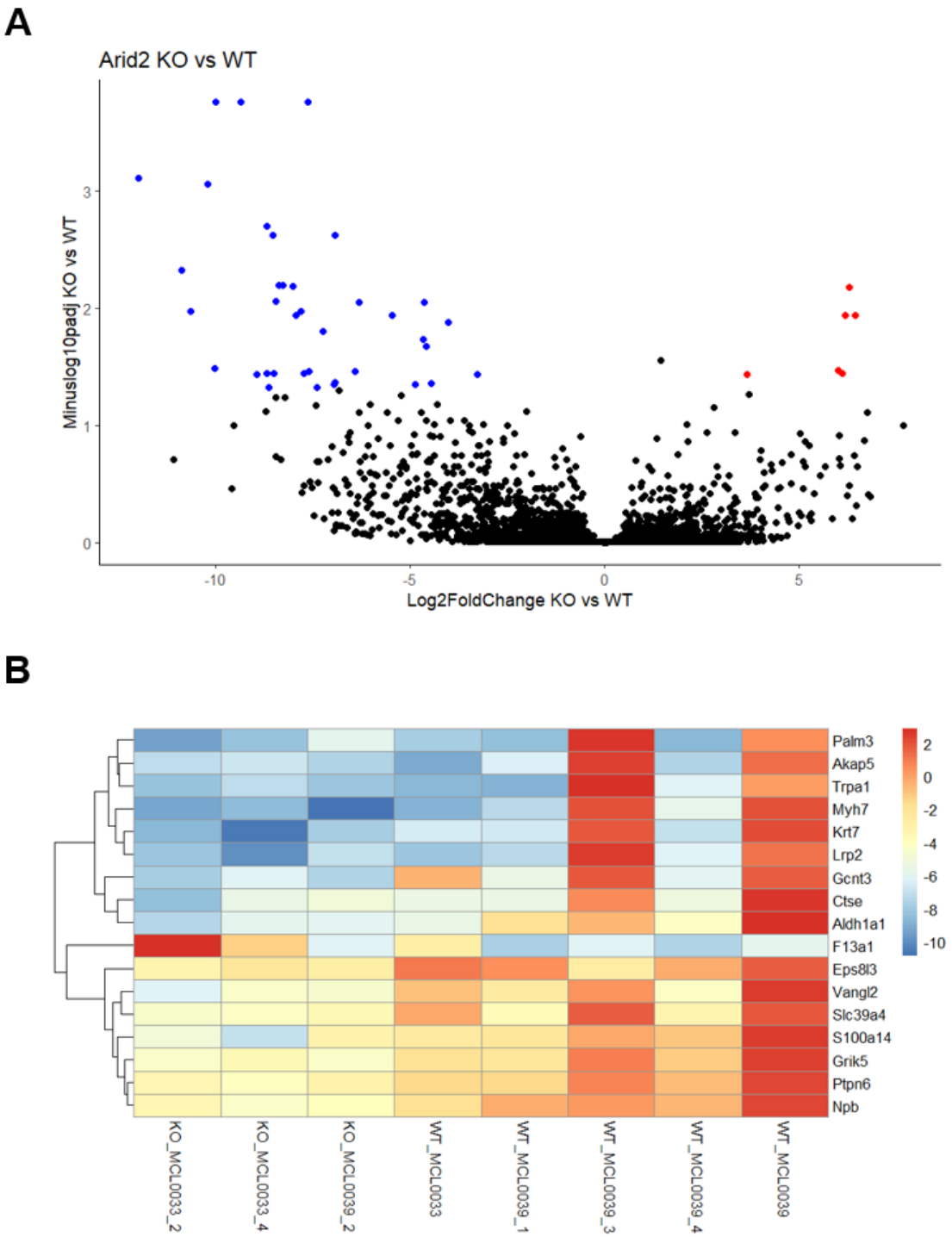
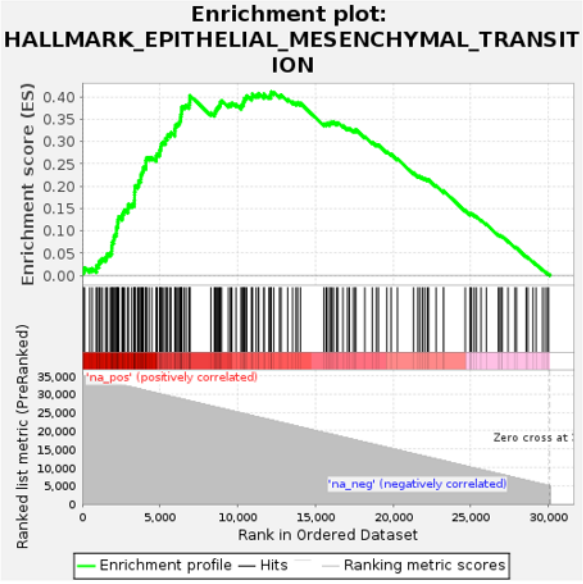


Figure 5. *Arid2* WT and KO clones comparison at mRNA and protein levels.

(A) Volcano plot representation of genes showing a statistically significant different gene expression when comparing knockout clones towards wildtype control clones. Upregulated genes are shown in red and downregulated ones in blue. **(B)** Heatmap representing 17 genes with significant differential expression between samples. Differential expression was considered as significant when the absolute value of log2foldchange is higher than 2 and the p-adjusted value was lower than 0.05.

Afterwards, we performed a Gene Set Enrichment Analysis (GSEA) in order to determine whether a predefined set of genes that share common biological functions showed statistically significant and concordant expression differences between the two phenotypes - KO and WT. For this analysis, we used two databases: the "mouse-ortholog hallmark gene sets (MH)" and the "ontology gene sets (M5)." In both cases, the most enriched pathway in the KO condition was the epithelial–mesenchymal transition (EMT), a key process in cancer progression and metastasis (Figure 6).

A



B

	SYMBOL	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	Sfrp1	1	34999.000	0.0075	Yes
2	Col12a1	94	34905.000	0.0120	Yes
3	Mmp3	147	34852.000	0.0177	Yes
4	Col5a3	480	34519.000	0.0141	Yes
5	Cap2	654	34344.000	0.0157	Yes
6	Thbs2	912	34086.000	0.0144	Yes
7	Wnt5a	914	34084.000	0.0217	Yes
8	Grem1	1059	33939.000	0.0242	Yes
9	Mfap5	1077	33921.000	0.0310	Yes
10	Lox	1192	33806.000	0.0344	Yes
11	Fstl1	1289	33708.000	0.0385	Yes
12	Sgcd	1504	33493.000	0.0386	Yes
13	Col7a1	1568	33429.000	0.0437	Yes
14	Fbn1	1625	33372.000	0.0490	Yes
15	Col16a1	1804	33193.000	0.0502	Yes
16	Crlf1	1888	33109.000	0.0545	Yes
17	Aplp1	1893	33104.000	0.0615	Yes
18	Fbn2	1908	33089.000	0.0682	Yes
19	Pthlh	1930	33067.000	0.0746	Yes
20	Dpysl3	2021	32976.000	0.0787	Yes
21	Prrx1	2033	32964.000	0.0854	Yes
22	Thbs1	2086	32911.000	0.0908	Yes

Figure 6. Gene Set Enrichment Analysis (GSEA).

(A) Enrichment plot for the HALLMARK_EPHITELIAL_MESENCHYMAL_TRANSITION gene set. The plot displays the enrichment score (ES), hits and ranking metric scores across the ordered database. **(B)** Core enrichment genes from the epithelial–mesenchymal transition (EMT) pathway identified by GSEA. The table shows genes contributing most significantly to pathway enrichment in KO versus WT clones. Columns indicate gene symbol, position in the ranked list, rank metric score, running enrichment score (ES), and core enrichment status.

DISCUSSION

The primary objective of the present study is to identify gene networks and signaling pathways regulated by SWI/SNF chromatin remodeling complexes in the context of lung adenocarcinoma. To achieve this, we generated murine cellular models deficient in *Arid2* and *Arid1a* and RNAseq libraries from both models in order to analyse their transcriptomic profile.

First of all, while the successful generation and confirmation of gRNA-containing constructs targeting *Smarca4* was achieved, challenges in transformation efficiency, particularly for sgRNA1, initially hampered progress, requiring optimization through repeated transformation steps.

Secondly, the generation of *Arid1a* KO clones revealed discrepancies between cell lines, with MCL0039 showing higher transfection and clone recovery rates than MCL0033 Red-Fluc. This difference may be attributable to cell-type-specific susceptibility to lipofection or differential DNA repair dynamics post-transfection. If knockout clones aren't obtained, future experiments should consider optimizing transfection conditions for each cell line (adjusting DNA/lipofectamine ratios, cell density at the moment of transfection, incubation times, etc.). Another viable option could be exploring alternative transfection strategies, especially for more resistant lines as MCL0033 Red-Fluc.

Molecular characterization of *Arid2* KO clones yielded compelling evidence of *Arid2* transcript disruption; which is consistent with a KO phenotype. However, as described before, the expression pattern was more heterogeneous in clones 2 and 5. These clones appeared to produce incomplete mRNA transcripts that include exon 14 in clone 2 and both exons 2 and 14 in clone 5, respectively. This observation may indicate the occurence of a phenomenon known as *exon skipping.*[28] When a deletion or mutation is introduced in exon 5, the splicing machinery may bypass the damaged exon in order to preserve transcript integrity and enable the production of a functional protein. In this case, as exon 5 is composed of 219 bp, its exclussion would mantain the reading frame, potentially allowing the synthesis of a partially functional protein. Further experiments will be necessary to confirm this hypothesis.

Apart from that, a consistent reduction of *Tgfbr2* was observed along all the clones, following previous data of Varela's lab. This finding could link with an initial strong activation of Tgfβ pathway upon the induction of *Arid2* KO; followed by a subsequent pathway inhibition through negative feedback regulation, which is what we observed. At first, we hypothesized that this negative feedback could be mediated by Smad7; however, our results showed that its expression levels are downregulated compared to the parental cell line. In conclusion, experimental data suggests that SWI/SNF deficiency leads to the activation of some Tgfβ transcriptional targets and the pathway is subsequently repressed through Smad7-independent regulation.

Unfortunately, our attemps to confirm *Arid2* loss by Western blotting were inconclusive due to the lack of antibody specificity for the murine protein. This technical limitation affected validation and could be resolved in future experiments by producing a specific antibody, something that is currently being considered in Varela's lab.

In our differential expression analysis, we found that 43 genes were significantly differentially expressed between KO and WT clones, regardless of the cell line. Some of these genes have been already correlated with cancer progression. For instance, *S100a14* has been associated with higher proliferation and migration rates in NSCLC,

as it could be indirectly implicated in the tumor effects promoved by *Casc9*.[29] Moreover, Takashi Yokohama *et al*. have seen that *Aldh1a1* expression induction with TGF-β in A459 lung cancer cell line enhaced doxorubicin resistance. [30] However, it is true that there is considerable variation in the expression of some genes within the same group. For this reason, RNAseq analysis of aditional samples of *Arid2* KO clones and *Arid1a* ones generated in the present study is currently being performed. This will enable us to develope a more comprehensive and consistent conclusion, as well as to assess if genes showing significant expression changes in *Arid2* - deficient models are related to those differentially expressed in *Arid1a* - deficient ones.

To conclude, I consider that it would be interesting to perform cell migration and proliferation or treatment sensitivity assays which could provide more comprehensive insights into the tumor-suppresor roles of these chromatin remodelers. In addition, we could inyect them into immunocompetent murine *in vivo* models in order to measure their tumoral capacity.

METHODS

Plasmids' digestion by restriction digestion enzymes

All reactions were performed by incubating the DNA plasmid with the appropriate Fast Digest restriction enzyme (Thermo Scientific) and the 10x FastDigest Buffer (Thermo Scientific, #B64) in a final volume of 20 μl at 37ºC for 30 minutes. 6x Loading Buffer (Thermo Scientific, #R0611) was subsequently added in order to inactivate the enzyme. This also enabled the visualization of the digestion product on a 1x agarose gel to assess whether it had been successfully digested.

Plasmid purification

1,8 volumes of AMPure XP beads (Beckman Coulter, A63881) were added to the sample and then incubated for 5 minutes to allow the beads to bind to the DNA. Sample was next placed in the magnetic rack for 2 minutes. After supernatant was discarded, we performed two washes with 200 μl of ethanol 70%. In order to make sure that ethanol was completely removed, samples were incubated at 37ºC until it was completely evaporated. DNA was finally eluted in 20 μl of Milli Q Water and concentration was quantified in the Nanodrop2000.

CRISPR-Cas9 guides design

Single guide RNAs (sgRNAs) for CRISPR-Cas9 on the *Mus musculus Smarca4* gene (ENSEMBL id: ENSMUSG00000040647) were designed using the Benchling online program (Benchling Inc., San Francisco, CA). The mouse genome assembly GRCm39 (mm39) was selected as the reference. The full coding sequence and exon structure of *Smarca4* were retrieved and analyzed to identify suitable CRISPR target sites.

Benchling was used to identify candidate sgRNA sequences adjacent to NGG protospacer adjacent motifs (PAMs), compatible with the CRISPR-Cas9 system. To maximize gene disruption efficiency, target sites were preferentially located within early exons shared across all annotated transcript isoforms and with lengths not divisible by three (non-triplet exons) in order to prevent potential rescue of the open Reading frame by exon skipping during mRNA splicing.

Guide selection was based on a combination of predicted on-target activity and off-target specificity, determined by whole-genome sequence alignments. Only guides with an efficiency score ≥ 50 and specificity score ≥ 70 were considered. Finally, the top three sgRNAs were selected for experimental validation. They were synthesized as complementary oligonucleotides with overhangs compatible with BbsI restriction sites.

DNA cloning and amplification by bacterial transformation

gRNAs were constructed by the annealing and phosphorylation of synthetic oligonucleotides. Next, pX330A plasmid was digested using BbsI (BpiI) fast digest restriction enzyme (Thermo Scientific, #ER1011), which recognizes GAAGAC (2/6)^ restriction sites and leaves 3'-GGTG and 5'-GTTT sticky ends in this case. After ligation (20 min at RT) of the gRNAs in the pX330A plasmid using T4 DNA ligase (Thermo Scientific, #EL0011), 10 μl of the construct was transformed in DH5α *E.coli* competent cells, following heat-shock protocol (30 min on ice, 20 sec at 42ºC and 2 min on ice). After the heat-shock, bacteria were incubated for 1 hour approximately at 225 rpm and 150 μl were plated in LB-agar plates containing 100 μM of ampicillin. Plates were then

incubated overnight (ON) at 37ºC. Selected colonies were checked first by PCR and then separately grown for 24 hours in the same LB-agar plates. Positive colonies in the PCR were then inoculated and grown ON in 5 ml of LB-agar medium with ampicillin 0,1 mg/ml so as to amplify plasmid concentration. Plasmidic DNA from the bacterial culture was extracted with the Miniprep GeneJET kit (Thermo Scientific, #K0503) following manufacturer instructions.

Cell culture

MCL0033 (*Kras* mut, *Tp53* WT) and MCL0039 (*Kras* mut, *Tp53* KO) murine epithelial cell lines were obtained from Munich Technical University (Rad's group). MCL0033 was then stably transduced with the fluorescent protein Red Fluc previously in Varela's lab.

Both lines were cultured in DMEM (Corning, 10-017-CV) and supplemented with 10 % fetal bovine serum (FBS) (Gibco, 10270-106), ciprofloxacin (2 μg/ml) and gentamicin (10 μg/ml). Cells were maintained at 37ºC in a humified atmosphere containing 5% of CO2 and at a cell confluence between 30-70% idealistically.

Plates were treated with collagen 5 μg/cm2. Collagen stock (3060 μg/ml) was diluted in acetic acid 0.02M, added to the plate and incubated during 1 hour at RT. Before seeding the cells, they were washed twice with 1x phosphate buffered saline (PBS) (Gibco 10010023) in order to eliminate acetic acid rests as well as possible.

In order to detach adherent cells, they were first washed with PBS and then treated with 0.25% Trypsin-EDTA (Gibco, 25200-072). D10F culture media was eventually added to neutralize trypsin and let the cells attach again.

Transfection

For generating *Arid1a* KO clones, all transfections were carried out by lipofection using Lipofectamine™ 2000 reagent (Thermo Scientific, #11668019) following manufacturer instructions, when cells having a confluence of approximately 60-70%.

The constructs containing the Cas9 and the gRNAs (gRNA1 and gRNA3) and the lipofectamine were first incubated separately in MEM culture media (Corning, 10-010-CM) (5 min, RT). Next, DNA plasmids were mix with the transfection reagent and incubated again (20 min, RT). Transfection was then performed in MCL0039 and MCL0033 Red Fluc cell lines, which were then maintained in DMEM not supplemented with FBS.

24 hours after transfection, we change the culture media to D10F (DMEM, 10% FBS, ciprofloxacin (2 μg/ml), gentamicin (10 μg/ml)) and transfected cells were selected with puromycin (1 mg/ml). The dose of puromycin used was 5 μg for each milliliter of culture media. Selective conditions were maintained during 48 hours.

Western Blotting

Protein lysates were prepared in RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 1mM Sodium Orthovanadate, 1mM NaF) by sonication consisting of 10 cycles 30'' ON and 30'' OFF. Protein concentration was quantified using Qubit Protein Assay Kit (Life Technologies, Q33212). Samples were then prepared for the electrophoresis in 6x Laemmli buffer (250mM Tris pH 6.8, 20% β-mercaptoethanol, 10%

SDS, 0.25% bromophenol blue, and 50% glycerol) and boiled for 5 minutes at 95ºC to denaturalize the proteins before running them in 8% polyacrylamide gels prepared in advance (Resolving gel – 2.3 ml NFW, 1.3 ml 30% acrylamide mix, 1.3 ml 1.5 M Tris (pH 8.8), 50 µl 10% SDS, 50 µl 10% ammonium persulfate, 3 µl TEMED; Stacking gel – 680 µl NFW, 170 µl 30% acrylamide mix, 130 µl 1.0M Tris pH (6.8), 10 µl 10% SDS, 10 µl 10% ammonium persulfate, 1 µl TEMED) under a chemical fume hood.

Samples were run in running buffer (900 ml H2O Elix, 100 ml 10X Tris-Glycine, 10 ml 10% SDS) at 135V for 5 minutes to allow their alignment withing the stacking gel. After this initial period, the voltage was increased to 180V during 1 hour approximately. They were then transferred during 2 hours and by wet transference to nitrocellulose membranes using transfer buffer with 20% methanol.

Afterwards, membranes were blocked using 10% milk solution in TTBS (2% 1M TRIS-HCl pH 7.5, 3% 5M NaCl + 0.05% Tween 20) for 1 hour at RT. Next, they were incubated at 4ºC overnight with primary antibodies mouse anti-ARID2 E-3 (Santa Cruz Biotechnology, sc-166117) diluted 1:1000, rabbit anti-ARID2 NBPI (Novus, #26615) diluted 1:1000, rabbit anti- ARID2 (A302-229A, Thermo Fisher Scientific) diluted 1:2000, mouse anti-V5 Tag antibody (MCA1360GA, Bio-Rad) and goat anti-βactin I-19 (Santa Cruz, sc-1616) diluted 1:1000 as well.

Membranes were washed with TTBS the following day and incubated again 1 hour at RT with secondary antibodies (goat anti-mouse, goat anti-rabbit and rabbit anti-goat, respectively) diluted 1:10000. All antibodies were prepared in TTBS 1% BSA.

Membranes were developed using an Enhanced Chemiluminescence (ECL) system. A 1:1 mixture of the two ECL solutions was applied directly onto the membranes. ECL solution 1 consisted of 10% Tris-HCl 1 M pH 8.5, 1% luminol, and 0.45% p-coumaric acid in distilled water. ECL solution 2 contained 10% Tris-HCl 1 M pH 8.5 and 0.06% hydrogen peroxide in distilled water. The reagents were added in the specified order, mixing after each addition, and all solutions were protected from light and stored at 4°C. The prepared ECL mix retains activity for approximately one month under these conditions.

Chemiluminescent signal was detected using an Amersham ImageQuant 800 imaging system.

RNA extraction

Total RNA was purified using NZY Total RNA Isolated Kit (NZYTech, MB13402) and following the manufacturer instructions. Cell lysates were homogenized with a 20G syringe after adding lysis buffer and β-mercaptoethanol as reducing agent in order to reduce viscosity.

Expression analysis by RT-qPCR

First, cDNA was generated by reverse transcription (RT) using the PrimeScript™ RT-PCR Kit (Takara, # RR014B) following the manufacturer's protocol. It is crucial to add the same amount of RNA for each sample, by default 1 µg. To each sample, 2 µl of 5X PrimeScript Buffer (for Real Time), 0.5 µl of PrimeScript RT Enzyme Mix I, 0.5 µl of Oligo dT Primer (50 µM) and 0.5 µl of Random 6 mers (100 µM) were added, and the volume was adjusted to 10 µl with RNAse free water.

Quantitative polymerase chain reaction (qPCR) was performed using forward and reverse specific primers for each target gene at 300 nM, PowerUp SYBR Green Master Mix 2x (Thermo Scientific) and 1 μl of cDNA per reaction in a final volume of 10 μl. We prepared 3 replicates for each sample. A table with the sequences of qPCR primers is provided in Supplementary Table 1.

Target genes expression was then analyzed by the ΔΔCt method and normalized to levels of beta actin mRNA expression, which was used as housekeeping gene.

RNAseq libraries generation

First of all, mRNA was enriched using NEBNextPoly(A) mRNA Magnetic Isolation Module (New England Biolabs, #E7490) and fragmented by heating at 94ºC for 15 minutes. First strand cDNA was generated adding PrimeScriptTM RT Enzyme Mix (Takara, RR037B) in a total volume of 20 μl (15' at 37ºC, 5'' at 85ºC, hold at 4ºC).
Next, RNAseH (Thermo Scientific, #EN0201), DNA Polymerase I (Thermo Scientific, #EP0042) and T4 DNA polymerase were added in order to generate the second strand (dsDNA). EDTA (0,5M Ph 8) was then added to inactivate T4 DNApol enzyme activity.

After DNA was purified using 1,8 volumes of Agentcourt AMPure XP beads (Beckman Coulter, A63881), we performed adapters hybridization and phosphorylation using PNK Enzyme (Thermo Scientific, #EK0032) and T4 DNA Ligase Buffer 10x (Thermo Fisher Scientific). Resulting fragments were repaired by an End-Repair Reaction Mix (Fast DNA End Repair Kit, Thermo Scientific, K0771) incubating for 5 min at 20ºC and then adenylated on their 3' ends using Klenow Fragment exo- (Thermo Scientific, #EP0422). Synthetic P5 and P7 adapters (Supplementary Table 2) were finally ligated to the resulting fragments using T4 DNA ligase (Thermo Scientific, #EL0014) after DNA purification.

Next, a PCR with Phusion High-Fidelity DNA polymerase (Thermo Scientific, #F530L) was performed to add unique Illumina indexes (Supplementary Table 2) in order to identify the reads from each sample in subsequent analysis. DNA was finally purified (0,8x Agentcourt AMPure XP beads (Beckman Coulter, A63881)) and eluted in 20 μl of low TE buffer (Thermo Scientific, #12090015).

Size distribution of the generated libraries was analyzed in the High Sensitivity D1000 ScreenTape Assays (Agilent, 5067-5584). Finally, equimolecular quantities of resulting samples were mixed to obtain between 30 and 60 million of paired-ends per sample. The NGS sequencing was accomplished in Illumina platforms.

RNAseq data analysis

Sequenced reads were aligned to the mouse reference genome (GRCm39) using Hisat2 alignment program. Transcript-level raw counts obtained for each library were imported into Rstudio and processed with the DESeq2 package. Individual count files for each library were read and concatenated column-wise to generate a single gene-by-sample matrix. A sample annotation table was then created, specifying each library's genotype as either wildtype (WT) or knockout (KO) and the factor levels were explicitly ordered so that WT was set as the reference condition. Differential expression analysis was performed with DESeq. Finally, the results were filtered as significant when fulfilling the conditions described in results section.

AGRADECIMIENTOS

En primer lugar, me gustaría agradecer a mi director Ignacio Varela, por su orientación y compromiso a lo largo del desarrollo de este proyecto.

También quiero dar las gracias al resto de miembros del grupo de investigación: Rosa Blanco, Raquel Marqueño, Natalia Sánchez y Nacho González. Su colaboración, así como las conversaciones compartidas, han hecho que esta experiencia haya sido muy enriquecedora y agradable. Y, en especial, a David Martín, por guiarme día a día en la realización de mis experimentos y por su implicación y cercanía. Gracias chicos por hacerme sentir parte del equipo desde el primer día.

No quiero olvidarme de mis compañeros de carrera, con los que he compartido tantas risas y algún que otro llanto a lo largo de estos cuatro años. Por último, gracias a mi familia y amigos, en especial a mis padres y mi hermana Irene, por confiar en mí cada día, por guiarme y apoyarme, pero siempre respetando mis propias decisiones.

REFERENCES

1. Mao, Y., Yang, D., He, J., and Krasna, M.J. (2016). Epidemiology of Lung Cancer. Surg Oncol Clin N Am *25*, 439–445. https://doi.org/10.1016/j.soc.2016.02.001.

2. Gridelli, C., Rossi, A., Carbone, D.P., Guarize, J., Karachaliou, N., Mok, T., Petrella, F., Spaggiari, L., and Rosell, R. (2015). Non-small-cell lung cancer. Nat Rev Dis Primers *1*, 15009. https://doi.org/10.1038/nrdp.2015.9.

3. George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., et al. (2015). Comprehensive genomic profiles of small cell lung cancer. Nature *524*, 47–53. https://doi.org/10.1038/nature14664.

4.Rudin, C.M., Brambilla, E., Faivre-Finn, C., and Sage, J. (2021). Small-cell lung cancer. Nat Rev Dis Primers *7*, 3. https://doi.org/10.1038/s41572-020-00235-0.

5. Pešek, M., and Mužík, J. (2018). [Small-cell lung cancer: epidemiology, diagnostics and therapy]. Vnitr Lek *63*, 876–883.

6. Kadara, H., Scheet, P., Wistuba, I.I., and Spira, A.E. (2016). Early Events in the Molecular Pathogenesis of Lung Cancer. Cancer Prevention Research *9*, 518–527. https://doi.org/10.1158/1940-6207.CAPR-15-0400.

7.Lahiri, A., Maji, A., Potdar, P.D., Singh, N., Parikh, P., Bisht, B., Mukherjee, A., and Paul, M.K. (2023). Lung cancer immunotherapy: progress, pitfalls, and promises. Mol Cancer *22*, 40. https://doi.org/10.1186/s12943-023-01740-y.

8. Du, X., Yang, B., An, Q., Assaraf, Y.G., Cao, X., and Xia, J. (2021). Acquired resistance to third-generation EGFR-TKIs and emerging next-generation EGFR inhibitors. The Innovation *2*, 100103. https://doi.org/10.1016/j.xinn.2021.100103.

9. Ettinger, D.S., Akerley, W., Borghaei, H., Chang, A.C., Cheney, R.T., Chirieac, L.R., D'Amico, T.A., Demmy, T.L., Govindan, R., Grannis, F.W., et al. (2013). Non-small cell lung cancer, version 2.2013. J Natl Compr Canc Netw *11*, 645–653; quiz 653. https://doi.org/10.6004/jnccn.2013.0084.

10. Hohmann, A.F., and Vakoc, C.R. (2014). A rationale to target the SWI/SNF complex for cancer therapy. Trends in Genetics *30*, 356–363. https://doi.org/10.1016/j.tig.2014.05.001.

11. Shi, Y., and Shin, D.S. (2023). Dysregulation of SWI/SNF Chromatin Remodelers in NSCLC: Its Influence on Cancer Therapies including Immunotherapy. Biomolecules *13*, 984. https://doi.org/10.3390/biom13060984.

12. Mittal, P., and Roberts, C.W.M. (2020). The SWI/SNF complex in cancer — biology, biomarkers and therapy. Nat Rev Clin Oncol *17*, 435–448. https://doi.org/10.1038/s41571-020-0357-3.

13. Wilson, B.G., and Roberts, C.W.M. (2011). SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer *11*, 481–492. https://doi.org/10.1038/nrc3068.

14. Monterde, B., and Varela, I. (2022). Role of SWI/SNF chromatin remodeling genes in lung cancer development. Biochemical Society Transactions *50*, 1143–1150. https://doi.org/10.1042/BST20211084.

15. Malone, H.A., and Roberts, C.W.M. (2024). Chromatin remodellers as therapeutic targets. Nat Rev Drug Discov *23*, 661–681. https://doi.org/10.1038/s41573-024-00978-5.

16. Elwy, A.E., Nassar, M.I., Shaban, S.H., and Elsaba, T.M. (2025). The prognostic impact of PBRM1 immunohistochemical expression and its association with CD3 + and CD8 + immune cells in patients with renal cell carcinoma: A retrospective study. Pathology - Research and Practice *268*, 155863. https://doi.org/10.1016/j.prp.2025.155863.

17. Wilson, B.G., Helming, K.C., Wang, X., Kim, Y., Vazquez, F., Jagani, Z., Hahn, W.C., and Roberts, C.W.M. (2014). Residual complexes containing SMARCA2 (BRM) underlie the oncogenic drive of SMARCA4 (BRG1) mutation. Mol Cell Biol *34*, 1136–1144. https://doi.org/10.1128/MCB.01372-13.

18. Sato, E., Nakayama, K., Razia, S., Nakamura, K., Ishikawa, M., Minamoto, T., Ishibashi, T., Yamashita, H., Iida, K., and Kyo, S. (2018). ARID1B as a Potential Therapeutic Target for ARID1A-Mutant Ovarian Clear Cell Carcinoma. IJMS *19*, 1710. https://doi.org/10.3390/ijms19061710.

19. Wang, X., Wang, S., Troisi, E.C., Howard, T.P., Haswell, J.R., Wolf, B.K., Hawk, W.H., Ramos, P., Oberlick, E.M., Tzvetkov, E.P., et al. (2019). BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. Nat Commun *10*, 1881. https://doi.org/10.1038/s41467-019-09891-7.

20. Jin, F., Yang, Z., Shao, J., Tao, J., Reißfelder, C., Loges, S., Zhu, L., and Schölch, S. (2023). ARID1A mutations in lung cancer: biology, prognostic role, and therapeutic implications. Trends Mol Med *29*, 646–658. https://doi.org/10.1016/j.molmed.2023.04.005.

21. Sun, D., Teng, F., Xing, P., and Li, J. (2021). ARID1A serves as a receivable biomarker for the resistance to EGFR-TKIs in non-small cell lung cancer. Mol Med *27*, 138. https://doi.org/10.1186/s10020-021-00400-5.

22. Zhang, M., Dong, Y., Meng, R., and Zhang, D. (2025). SMARCA4 Deficiency in Lung Cancer: From Signaling Pathway to Potential Therapeutic Targets. Genes Chromosomes & Cancer *64*, e70022. https://doi.org/10.1002/gcc.70022.

23. Manolakos, P., Boccuto, L., and Ivankovic, D.S. (2024). A Critical Review of the Impact of SMARCA4 Mutations on Survival Outcomes in Non-Small Cell Lung Cancer. JPM *14*, 684. https://doi.org/10.3390/jpm14070684.

24. Xu, L., Xu, X., Wu, P., Ye, W., Zhao, J., Yang, J., Yao, Y., Chen, M., Wang, X., Wang, A., et al. (2025). Clinical characteristics and prognostic analysis of patients with SMARCA4-deficient lung cancer. Technology and Health Care *33*, 1014–1020. https://doi.org/10.1177/09287329241296242.

25. Moreno, T., Monterde, B., González-Silva, L., Betancor-Fernández, I., Revilla, C., Agraz-Doblas, A., Freire, J., Isidro, P., Quevedo, L., Blanco, R., et al. (2021). ARID2 deficiency promotes tumor progression and is associated with higher sensitivity to chemotherapy in lung cancer. Oncogene *40*, 2923–2935. https://doi.org/10.1038/s41388-021-01748-y.

26. Xu, H., Chen, H.-C., Yang, L., Yang, G., Liang, L., Yang, Y., Tang, H., Bao, H., Wu, X., Shao, Y., et al. (2023). Mutational landscape of SWI/SNF complex genes reveal correlation to predictive biomarkers for immunotherapy sensitivity in lung adenocarcinoma patients. ESMO Open *8*, 101585. https://doi.org/10.1016/j.esmoop.2023.101585.

27. Zhu, G., Shi, R., Li, Y., Zhang, Z., Xu, S., Chen, C., Cao, P., Zhang, H., Liu, M., Pan, Z., et al. (2021). ARID1A, ARID1B, and ARID2 Mutations Serve as Potential Biomarkers for Immune Checkpoint Blockade in Patients With Non-Small Cell Lung Cancer. Front Immunol *12*, 670040. https://doi.org/10.3389/fimmu.2021.670040.

28. Mou, H., Smith, J.L., Peng, L., Yin, H., Moore, J., Zhang, X.-O., Song, C.-Q., Sheel, A., Wu, Q., Ozata, D.M., et al. (2017). CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. Genome Biol *18*, 108. https://doi.org/10.1186/s13059-017-1237-8.

29. Zhao, W., Chen, T., and Zhao, Y. (2020). Upregulated lncRNA CASC9 Contributes to Progression of Non-Small Cell Lung Cancer Through Inhibition of miR-335-3p and Activation S100A14 Expression. OTT *Volume 13*, 6027–6036. https://doi.org/10.2147/OTT.S249973.

30. Yokoyama, T., Saitoh, M., and Miyazawa, K. (2025). TGF -β Enhances Doxorubicin Resistance and Anchorage-Independent Growth in Cancer Cells by Inducing *ALDH1A1* Expression. Cancer Science, cas.70109. https://doi.org/10.1111/cas.70109.

Trabajo de Fin de Grado
Grado en Ciencias Biomédicas · Facultad de Medicina
2024 – 2025

Supplementary information

Characterization of the role of SWI/SNF deficiencies in the progression of lung cancer

Paula Díez Vicente[1], Ignacio Varela Egocheaga[1]

UC | Universidad de Cantabria

A

Guide	Sequence 5' → 3'
Smarca4_sgRNA1_F	CACCGGCCCCCAACCCTCCACGAG
Smarca4_sgRNA1_R	AAACCTCGTGGAGGGTTGGGGGCC
Smarca4_sgRNA2_F	CACCGGGGCCTGGAGGGTACCCCC
Smarca4_sgRNA2_R	AAACGGGGGTACCCTCCAGGCCCC
Smarca4_sgRNA3_F	CACCGCCCCCATCCAGAAGCCCCG
Smarca4_sgRNA3_R	AAACCGGGGCTTCTGGATGGGGGC

B

Primer	Sequence 5' → 3'
pX330A_seq_rev2	CTGCAGAATTGGCGCACG

C

Gene	Sequence 5' → 3'
Arid2 (exon 2)	F: AGAGGGTGAGAGCAGGACC R: CTGGTGTAGAGCCCGTGAAG
Arid2 (exon 5)	F: GCCAGTCATCCTTCCACGAT R: GCCAGTCATCCTTCCACGAT
Arid2 (exon 14)	F: GGTGTTCTTGGAAGGCTGGT R: GCCGCTTCACTCCTATGACA
Smad7	F: CCTCCTCCTTACTCCAGATACCC R: CCCAGGGGCCAGATAATTCG
Tgfβr2	F: CGCTGCATATCGTCCTGTGG R: CCATTGTCGCTGGCCATGAC
β-Actine	F: GCTGTATTCCCCTCCATCGTG R: TGGGGTACTTCAGGGTCAGG

Table S1. Oligonucleotides sequences used for cloning and RT-qPCR assays.

(A) Oligonucleotides sequences used to generate the three gRNAs introduced in pX330A plasmid. (**B**) pX330A_seq_rev2 primer sequence 5'-3. **(C)** Primer sequences used in RT-qPCR targeting *Arid2*, *Smad7*, *Tgfβr2* and *β-Actine.*

A

Adapter	Sequence
Adapt_TruSeq_P5a	CTACACGACGCTCTTCCGATCT
Adapt_TruSeq_P5b	GATCGGAAGAGCGTCGTGTAG
Adapt_TruSeq_P7a	GATCGGAAGAGCACACGTCTG
Adapt_TruSeq_P7b	CAGACGTGTGCTCTTCCGCTCT

B

Index	Sequence
D705	CAAGCAGAAGACGGCATACGAGAT**TTCTGAAT**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D504	AATGATACGGCGACCACCGAGATCTACAC**GGCTCTGA**ACACTCTTT CCCTACACGACGCTCTTCCGATCT
D505	AATGATACGGCGACCACCGAGATCTACAC**AGGCGAAG**ACACTCTTT CCCTACACGACGCTCTTCCGATCT
D506	AATGATACGGCGACCACCGAGATCTACAC**TAATCTTA**ACACTCTTTC CCTACACGACGCTCTTCCGATCT
D507	AATGATACGGCGACCACCGAGATCTACAC**CAGGACGT**ACACTCTTT CCCTACACGACGCTCTTCCGATCT
D508	AATGATACGGCGACCACCGAGATCTACAC**GTACTGAC**ACACTCTTT CCCTACACGACGCTCTTCCGATCT
D509	AATGATACGGCGACCACCGAGATCTACAC**AACGTGAT**ACACTCTTT CCCTACACGACGCTCTTCCGATCT
D510	AATGATACGGCGACCACCGAGATCTACAC**AAACATCG**ACACTCTTT CCCTACACGACGCTCTTCCGATCT

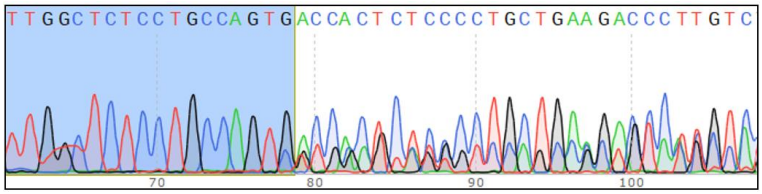
C

Index	Sequence
D524	AATGATACGGCGACCACCGAGATCTACAC**ACAGCAGA**ACACTCTT TCCCTACACGACGCTCTTCCGATCT
D701	CAAGCAGAAGACGGCATACGAGAT**CGAGTAAT**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D702	CAAGCAGAAGACGGCATACGAGAT**TCTCCGGA**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D703	CAAGCAGAAGACGGCATACGAGAT**AATGAGCG**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D704	CAAGCAGAAGACGGCATACGAGAT**GGAATCTC**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D706	CAAGCAGAAGACGGCATACGAGAT**ACGAATTC**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
D707	CAAGCAGAAGACGGCATACGAGAT**AGCTTCAG**GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT

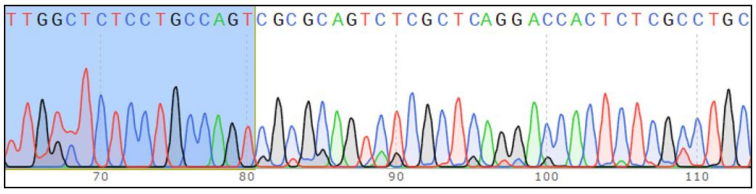
Table S2. Adapters and Illumina Index sequences used for RNAseq libraries generation.

(A) P5 and P7 adapter sequences ligated to cDNA strands. **(B)** Index sequences added to to cDNA strands after 3' ends adenylation and 5' ends repair in order to differenciate the seven samples of MCL0033 Red – Fluc *Arid2* KO clones. **(C)** Index sequences added to cDNA strands after 3' ends adenylation and 5' ends repair in order to differenciate the six samples of MCL0039 *Arid1a* KO clones.

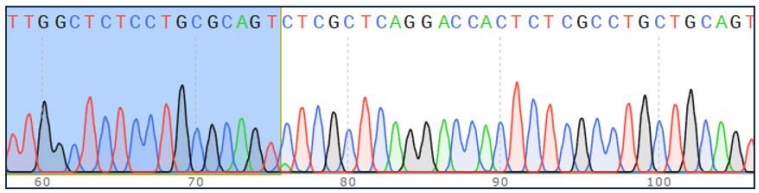
A



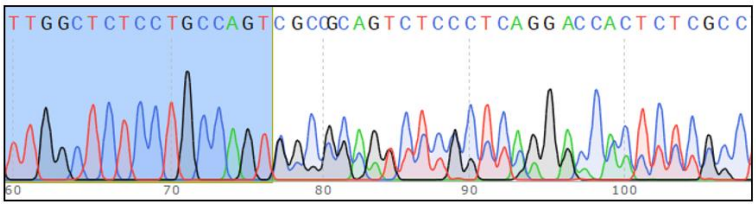
B



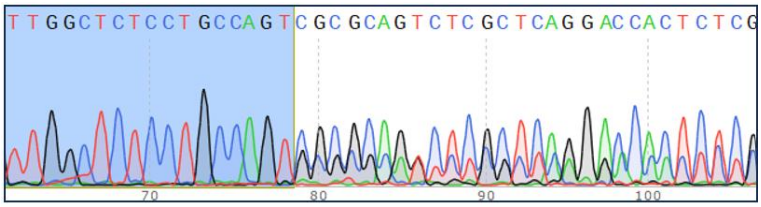
C



D



E



F

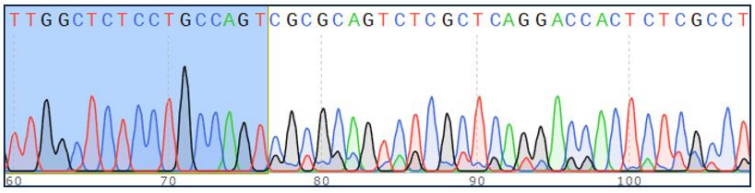


Figure S1. Sanger sequencing chromatograms of the target region in exon 5 of *Arid1a* in MCL0039 *Arid1a* potential KO clones.

(A) Chromatogram showing that clon MCL0039 A1 carries a heterozygous 19 bp deletion in exon 5 of *Arid1a*. **(B)** Chromatogram showing that clon MCL0039 A2 carries a 2 bp deletion in exon 5 of *Arid1a.* **(C)** Chromatogram showing that clon MCL0039 A4.2 carries a 10 bp deletion in exon 5 of *Arid1a.* **(D)** Chromatogram showing that clon MCL0039 B3.2 carries a 2 bp deletion in exon 5 of *Arid1a.* **(E)** Chromatogram showing that clon MCL0039 B5.2 carries a 2 bp deletion in exon 5 of *Arid1a* **(F)** Chromatogram showing that clon MCL0039 C1.2 carries a 2 bp deletion in exon 5 of *Arid1a.*

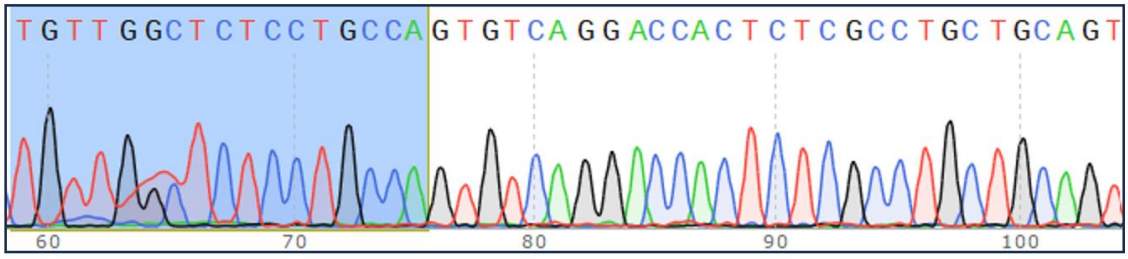


Figure S2. Sanger sequencing chromatograms of the target region in exon 5 of *Arid1a* in MCL0033 Red-Fluc *Arid1a* potential KO clones.

(A) Chromatogram showing that clone MCL0033 Red - Fluc A2 carries a 14 bp homozygous deletion in exon 5 of *Arid1a*.