



**UNIVERSIDAD DE CANTABRIA/
EUSKAL HERRIKO UNIBERTSITATEA**

**MÁSTER EN
BIOLOGÍA MOLECULAR Y BIOMEDICINA**

**TRABAJO FIN DE MÁSTER
IMPACTO DE LA DEFICIENCIA DE SWI/SNF EN LA
SEÑALIZACIÓN POR TGF- β EN CÁNCER DE PULMÓN**

**MASTER`S THESIS
IMPACT OF SWI/SNF DEFICIENCY IN TGF- β
SIGNALING IN LUNG CANCER**

Realizada por: Ignacio González Mimbiela
Dirigida por: Ignacio Varela Egocheaga
Instituto de Biomedicina y Biotecnología de Cantabria.

SANTANDER 2025

INDEX

1. ABSTRACT	2
2. INTRODUCTION.....	3
2.1. Lung cancer.....	3
2.1.1. Genetic alterations in lung cancer	3
2.1.2. Lung cancer treatment	4
2.2. SWI/SNF Chromatin Remodeling Complexes	5
2.2.1. SWI/SNF subfamilies	5
2.2.2. SWI/SNF mutations in human cancer	6
2.2.3. SWI/SNF disruption in lung cancer	8
2.2.4. Molecular mechanisms altered in SWI/SNF-deficient human cancers.....	8
2.3. Epithelial-mesenchymal transition (EMT)	9
2.3. 1. TGF- β intracellular pathway.....	10
3. HYPOTHESIS AND OBJECTIVES	12
4. MATERIALS AND METHODS	13
4.1. Cell biology methods	13
4.1.1. Cell lines and culture conditions.....	13
4.1.2. Mycoplasma PCR	13
4.1.3. Transfection of cell cultures.....	14
4.2. Molecular biology methods	15
4.2.1. RNA extraction.....	15
4.2.3. Western Blot	16
4.2.4. RNA-Seq Library Preparation	17
4.3. Statistical analysis.....	19
5. RESULTS	20
5.1. ARID2 silencing using short interfering RNAs (siRNAs)	20
5.2. Effects of ARID2 depletion in TGF-β-Smad signaling.....	21
5.2.1. Expression of TGF- β type 2 receptor (<i>TGFBR2</i>) in ARID2-deficient cells.....	21
5.2.2. Effects of ARID2 depletion in TGF- β -Smad downstream effectors.....	22
5.3. Characterization of ARID2 silencing effects at a transcriptional level.	23
5.3.1. Differentially expressed genes in ARID2-deficient cells.....	23
5.3.2. Effects of ARID2 loss in TGF- β target gene expression.	25
6. DISCUSSION	27
7. CONCLUSIONS.....	29
8. REFERENCES.....	30

1. ABSTRACT

SWI/SNF is a family of chromatin remodeling complexes that uses the energy provided by ATP hydrolysis to modify the level of chromatin compaction regulating processes such as gene expression or DNA damage repair. Large cancer sequencing studies have provided evidence of SWI/SNF disruption in 20% of cancer patients, which suggest a tumoral suppressor role of these complexes. However, the molecular mechanisms underlying its role in tumorigenesis remain largely unknown. Lung cancer remains the main cause of cancer-related deaths worldwide with a mean survival rate below 20% despite the new development of targeted or immunotherapies. Consequently, any new knowledge about the molecular mechanisms involved in lung cancer and its potential implication in therapy is a topic of great interest.

Morphological changes observed in SWI/SNF-deficient cells made us hypothesize that TGF- β pathway might be affected in these cells due to its function in epithelial-mesenchymal transition (EMT).

Here, we analyze the molecular effect of the depletion of a SWI/SNF subunit (ARID2) in several lung cancer cell lines with different mutational backgrounds. Our results suggest that we efficiently produced ARID2 depletion in our cell models using siRNAs. This depletion is accompanied with an alteration of TGF- β pathway evidenced by downregulation of its receptor (TGFBRII). Moreover, this downregulation is not the result of an overactivation of TGF- β signaling, nor by the upregulation of the negative regulator SMAD7; which suggest that SWI/SNF might directly regulate *TGFBRII* expression. Finally, we have characterized, at the transcriptional level, the response of SWI/SNF deficient cells to TGF- β induction.

2. INTRODUCTION

2.1. Lung cancer

Lung cancer is the most frequently diagnosed cancer worldwide, as well as the leading cause of cancer-related deaths, with a 5-year survival rate of only 19%¹.

According to their molecular and histological features, lung cancer tumors can be classified into several subtypes. Currently, the classification includes two broad types: Non-small cell lung carcinoma (NSCLC), which represents the majority, accounting for about 80-85% of the cases, and small cell lung carcinoma (SCLC). Moreover, NSCLC is further subclassified in lung adenocarcinoma (LUAD, 40% of the cases), squamous cell carcinoma (SCC, 15-20% of the cases) and large-cell carcinoma (LCC, 10-15% of the cases)¹.

LUAD is a malignant epithelial tumor with glandular differentiation usually formed in the periphery of the lung. It can present several different histological patterns, alone or intermixed in the same tumor, and it is characterized by mucin production. SCC is a malignant epithelial tumor often arising at a central location in a main or lobar bronchus. It typically shows keratinization, although not always. LCC is a rare lung tumor type characterized by a lack of differentiation features, generally diagnosed after ruling out the other subtypes².

On the other hand, SCLC is the most clearly associated with smoking habits. It is a neuroendocrine carcinoma characterized by a high proliferative rate and early development of metastasis. It has a poor prognosis, with a 5-year survival rate around 6%³.

2.1.1. Genetic alterations in lung cancer

Lung cancer is characterized by a high rate of somatic mutations and genomic rearrangements, with a great amount of passenger alterations. This makes it difficult to identify the critical driver events.

TP53 (46%), *KRAS* (33%) and *EGFR* (14%) are the most frequently mutated genes in LUAD, followed by *BRAF* (10%), *PIK3CA* (7%), *MET* (7%), *SET2* (9%), *ARID1A* (7%), among others. Interestingly, *KRAS* and *EGFR* mutations are mutually exclusive. LUAD is characterized by alterations in signaling pathway mediated by tyrosine-kinase receptors (RTKs), such as MAP/ERK (76%) or PI3K-mTOR (25%). Additionally, dysregulation in cell cycle control (64%), oxidative stress (22%) and chromatin remodeling pathways are also common (49%)⁴.

SCC displays a high rate of copy number alterations, amplifications and deletions. *TP53* is the most frequently mutated gene (81%). Other common alterations include *CDKN2A*, *PTEN*, *PI3KCA*, *KEAP1*, *NOTCH* or *RB1*⁵.

Regarding SCLC, inactivating *TP53* and *RB1* mutations are found in 90% and 65% of all cases, respectively⁶. Mutations in other tumor suppressor genes like *PTEN*, *NOTCH* and *CRBBP* have been also described, as well as amplification of *MYC* and *FGFR1*. These tumors are characterized by large chromosomal rearrangements and a high mutation burden, as well as a remarkable intratumoral heterogeneity, which promotes rapid tumor progression and acquisition of resistance to therapies³.

2.1.2. Lung cancer treatment

Lung cancer treatment depends on the subtype and stage of the tumor, as well as the presence or absence of specific genetic alterations. Three main approaches are typically used: surgical resection, unspecific treatments such as chemotherapy and radiotherapy, and targeted therapies, like immunotherapy or oncogene pathway inhibitory drugs.

For early-stage localized NSCLC tumors, surgical resection is the optimal procedure, often combined with chemotherapy to increase the efficiency. If surgery is not possible or it is contraindicated, the preferred treatment is stereotactic radiotherapy, alone or combined with cisplatin-based chemotherapy⁷.

In more advanced-stage NSCLC tumors, the approach depends on the histology and genetic features. Squamous tumors, which lack actionable alterations, are usually treated with platinum-based chemotherapeutic agents (cisplatin, carboplatin), alone or in combination with radiotherapy or immune checkpoint inhibitors (ICIs), such as anti-PD-L1 antibodies (pembrolizumab, nivolumab). On the other hand, non-squamous tumors display driver mutations in oncogenes and benefit from specifically targeted drugs, which produce fewer secondary effects. In particular, tyrosine kinase inhibitors (TKIs) against *EGFR* (erlotinib, afatinib, or osimertinib), *BRAF* (dabrafenib), or *ALK* (alectinib) are used regularly⁷.

Regardless of the multiple therapeutic alternatives and the development of new targeted drugs, the acquisition of resistance mechanisms remains an issue. Therefore, new understanding of the molecular aspects involved in lung cancer behavior could be of great use.

2.2. SWI/SNF CHROMATIN REMODELING COMPLEXES

SWI/SNF complexes are regulators of nucleosome positioning, using the energy obtained from ATP hydrolysis to transiently break nucleosome-DNA contacts and mobilize nucleosomes along the DNA by inducing them to slide, as well as directly catalyze their ejection, addition or exchange (Figure 1)⁸. Thus, they modify chromatin structure and consequently modulate the accessibility of specific genomic regions to the transcriptional machinery and other DNA-binding proteins, playing key roles in gene expression regulation, cell differentiation, DNA damage repair and genomic stability^{9,10}.

SWI/SNF was originally described in *Saccharomyces cerevisiae* through two independent screenings intended to identify alterations in genes causing defects in the mating-type switching (SWI) and sucrose-fermentation (SNF) pathways^{11,12}. Later on, it was shown that many of these genes worked together forming a complex involved in the regulation of transcription, which was named SWI/SNF.

2.2.1. SWI/SNF subfamilies

SWI/SNF family of chromatin remodelers, also known as BRGM1/BRM associated factor (BAF) complexes, are large macromolecular assemblies (1-1.5 MDa) composed by 8-12 subunits and conserved across species⁸. To date, three distinct SWI/SNF subfamilies have been identified in mammals, known as canonical BAF (cBAF), polybromo-associated BAF (PBAF) and non-canonical BAF (ncBAF)¹³. All complexes share the so-called core subunits, which include SMARCC1/2, SMARCD1, ACTB, ACTL6A/B, BCL7A/B/C, highly conserved across species, as well as one of two mutually exclusive ATPases: SMARCA4 or SMARCA2. Additionally, each complex contains several other accessory subunits that provide them with a distinct identity and contribute to their assembly, targeting and regulation of lineage-specific gene networks^{14,15}. Thus, SMARCC2, SMARCD2/3, SMARCE1, and SMARCB1 subunits are common to cBAF and PBAF. However, cBAF contains ARID1A or ARID1B, as well as DPF1, DPF2 or DPF3, whereas PBAF incorporates ARID2, PRBM1, PHF10 and BRD7A. On the other hand, ncBAF does not include neither SMARCB1 nor ARID subunits, and has GLTSCR1 or GLTSCR1L and BRD9 instead (Figure 1)⁹.

Altogether, 29 genes encode SWI/SNF complex subunits in humans, some of which display lineage-restricted expression. Moreover, many of these genes can undergo alternative splicing, resulting in different isoforms. Hence, it is estimated that up to 1000 different SWI/SNF could exist in mammals, controlling gene expression in specific cellular contexts^{9,14}.

Not only do each subfamily display different composition, but also different location profiles across the genome. For instance, it has been described that cBAF complexes are predominantly located at enhancers, whereas PBAF act mainly in promoters, and ncBAF can be found in both¹⁶. Recruitment of SWI/SNF complexes to specific genomic sites is driven by the presence of DNA-binding domains and reader domains that enable the recognition of posttranslational modifications (PTMs) on histone tails, including bromodomains, chromodomains and PHD domains¹⁷. Additionally, other domains facilitating protein interactions, as well as ATPase activity, are found in accessory and catalytic subunits.

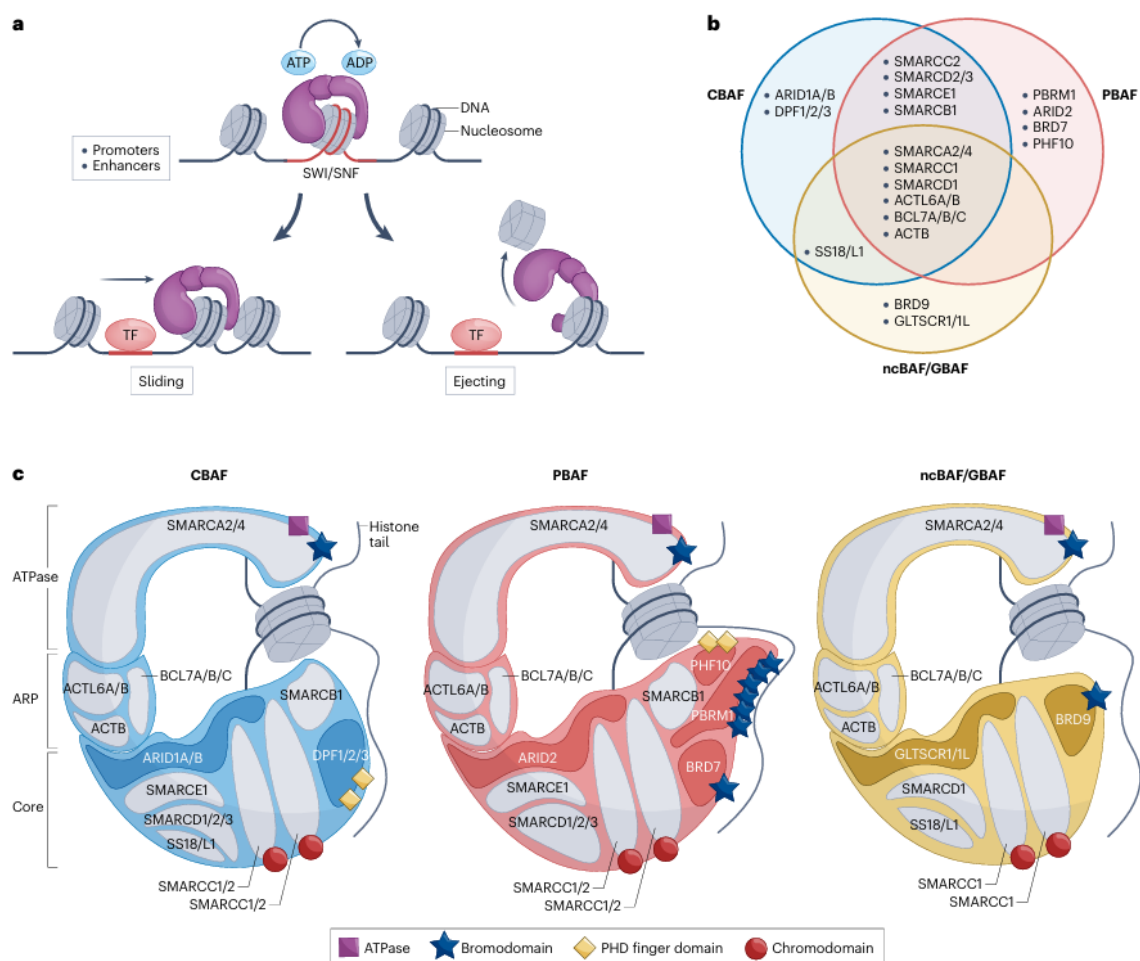


Figure 1 | Functions, composition and structure of SWI/SNF complex subfamilies.
Obtained from Malone and Roberts, 2024⁹.

2.2.2. SWI/SNF mutations in human cancer

There is a vast evidence of alterations of chromatin regulators and chromatin structure in cancer development. The main hypothesis is that these alterations affect the ability of cancer cells to acquire phenotypic plasticity through transcriptional reprogramming,

which is one of the hallmarks of cancer. Among the different families of chromatin remodelers, SWI/SNF are the most frequently mutated in cancer¹⁴.

SWI/SNF was first described to be involved in oncogenesis after the discovery that almost all cases of rhabdoid tumor, a rare pediatric tumor developed in brain soft tissue with a notably bad prognosis, present a biallelic inactivation of *SMARCB1*¹⁸. Recently, cancer genome sequencing studies have highlighted the presence of SWI/SNF mutations in a broad group of tumor types^{8,9,14,19}.

SWI/SNF mutations are ubiquitous across different cancer types, indeed, more than 20% of all tumors exhibit mutations in one or more SWI/SNF subunits, which places SWI/SNF as the second most common alteration in cancer just behind *TP53*. Out of the 29 genes that encode SWI/SNF subunits, at least 9 are recurrently mutated in malignancy⁹. Interestingly, subunit aberrations are not randomly distributed across cancer types, but rather display patterns of association with specific tumors (Figure 2)¹⁴.

Catalytic and accessory subunits are the most frequently affected, whereas core subunits exhibit a lower mutation rate. Most of these are nonsense, deletion and frameshift mutations, which suggest a tumor-suppressor role of SWI/SNF complexes^{10,14}. Interestingly, most of the recurrently mutated genes encoding SWI/SNF subunits are not present in all three subfamilies, and these alterations do not fully inactivate SWI/SNF function, as it has been observed that residual complexes are essential for the survival of SWI/SNF-deficient cancer cell lines⁹.

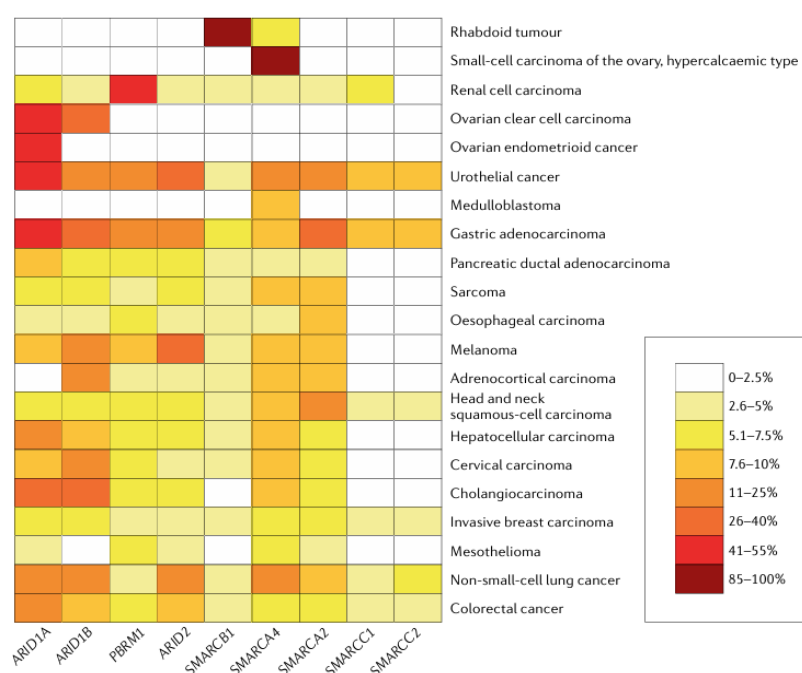


Figure 2 | Frequency of SWI/SNF subunit mutations in different cancer types. Obtained from Mittal and Roberts, 2020¹⁴.

2.2.3. SWI/SNF disruption in lung cancer

The role of SWI/SNF subunits mutations in lung cancer development has been extensively described over the last decades. For instance, it was found that 30% of NSCLC cell lines exhibit reduced expression of the catalytic subunits SMARCA2 and SMARCA4, which is associated with a lower survival-rate in comparison with those with normal expression levels²⁰. Moreover, further studies have confirmed the correlation between the deficiency of SMARCA2 and SMARCA4 with poor prognosis^{21,22}. Similarly, another study analyzing 106 cases of NSCLC revealed a decrease in ARID1A expression in about 35% of the cases and found a strong correlation between its downregulation and malignancy indicators, including a poor differentiation state and higher TNM scores²³. In the same line of findings, previous results of Varela's lab reported the reduction of ARID2 protein levels in 20% of the cases of a 139 lung cancer patients cohort, and, additionally, proved that ARID2 deficiency results in changes in chromatin structure, leading to a transcriptional reprogramming that induces proliferation and metastasis in both *in vivo* and *in vitro* lung cancer models²⁴.

Among the different types of lung cancer, SWI/SNF mutations occur more commonly in NSCLC, specifically in LUAD. In particular, about 20% of LUAD cases harbor a mutation in at least one gene coding for SWI/SNF subunits, being *SMARCA4* (8%), *ARID1A* (6%), *ARID2* (6%), and *SMARCA2* (3%) the most significantly mutated²⁵.

2.2.4. Molecular mechanisms altered in SWI/SNF-deficient human cancers

Regardless of the huge amount of evidence supporting a link between SWI/SNF alterations and cancer, the molecular events underlying its contribution to malignancy are not fully understood. Currently, the main hypothesis is that SWI/SNF deficiency promotes tumorigenesis either by impairing of DNA repair pathways and compromising genomic stability, or via the activation of canonical cancer-promoting pathways.

SWI/SNF participation in DNA damage repair has been extensively described, which suggest that SWI/SNF disruption could induce tumorigenesis due to a promotion of genomic instability²⁶. For instance, it has been reported that ARID1A, SMARCA4 and SMARCA2 are recruited to double strand breaks (DSBs), where SMARCA2/4 modulate the machinery access and assembly²⁷, and ARID1A helps in DSB end resection²⁸. SWI/SNF also participate in other DNA damage repair pathways, such as nucleotide excision repair (NER), where SMARCB1 interacts with XPC to recruit ATM²⁹.

Additionally, SWI/SNF complexes also participate maintaining genomic stability during mitosis. In this context, ARID1A is essential to guarantee genomic integrity in the G2-M

checkpoint²⁸, while SMARCA4 is involved in the decatenation of sister chromatids during the last steps of mitosis³⁰. Thus, the loss of any of these subunits could result in cells progressing through the cell cycle even with DNA damage, or dividing with uneven distribution of genetic material to the daughter cells. Similar consequences were observed in a renal cancer model, in which the absence of PRBM1 caused an impairment in p53-mediated DNA damage response and cell cycle repression³¹.

On the other hand, SWI/SNF has been shown to interact with cancer-associated genes, including *TP53*, *MYC* or *RB1*²⁵, which suggest that defects in these complexes could promote tumorigenesis through the activation of canonical cancer pathways. Furthermore, it has been observed that mutations in SWI/SNF elements and canonical cancer promoting genes are not often found in the same tumors, for example, SWI/SNF disruption and *TP53* or *PTEN* mutations seem to be mutually exclusive in several cancer types¹⁹. Similarly, SMARCA4 downregulation does not occur simultaneously with *EGFR* mutations or *MYC* amplifications in lung cancer³², and its loss is synthetic lethal with CDK4/6 inhibition³³. These mutual exclusivity patterns seem to indicate a redundancy in the consequences of the mutations, suggesting that SWI/SNF impairment is enough to promote tumorigenesis.

Indirect associations between cancer pathways and SWI/SNF disruption have also been discussed. For instance, previous results from Varela's lab showed that ARID2 is necessary for the expression of metastasis inhibitor *MTT1* and adhesion molecule *SDK1* by maintaining an open chromatin structure around their enhancers²⁴.

2.3. EPITHELIAL-MESENCHYMAL TRANSITION (EMT)

Epithelial-mesenchymal transition (EMT) is a process by which epithelial cells lose their characteristic phenotype and acquire mesenchymal features and behavior. During EMT, epithelial cells lose their specialized cell-contacts and apical-basal polarity, reshape their cytoskeleton and undergo transcriptional reprogramming³⁴. Altogether, these changes result in increased cell motility, enabling their migration and integration in surrounding tissue or, in some cases, distant locations. EMT is essential during development, and contributes to other physiological processes such as wound healing or fibrosis, and is complemented by another process known as mesenchymal-epithelial transition (MET) during tumor metastasis³⁵.

Several studies have reported correlations between the loss of SWI/SNF subunits and EMT across different types of cancer, including ARID2 in liver cancer³⁶ or ARID1A in breast, pancreatic, renal and ovarian cancer³⁷. In the context of lung cancer, loss of the

catalytic subunits SMARCA4 or SMARCA2 has been found to induce EMT, marked by morphological changes, as well as a downregulation in the expression of epithelial genes and an increase in mesenchymal markers³⁸. Moreover, previous unpublished findings from Varela's lab showed that SWI/SNF deficiency in lung cancer cell lines is associated to a change in cell morphology, together with a switch in the transcriptional landscape towards the activation of mesenchymal gene expression.

These effects are likely due to an alteration in TGF- β signaling pathway, which has been identified as a key regulator of EMT^{34,35}.

2.3.1. TGF- β intracellular pathway

Members of the Transforming Growth Factor β (TGF- β) family are ubiquitous secreted proteins that activate intracellular pathways involved in the regulation of cell growth and differentiation. The human genome has 33 genes encoding TGF- β family members. TGF- β peptides are produced in an immature form, with a long pro-polypeptide and a signal polypeptide required for secretion. Following enzymatic cleavage of these, the mature polypeptide is released and acts as a homo- or heterodimer activating membrane receptors. TGF- β 1 homodimer is considered the TGF- β signaling mechanism prototype³⁹.

The main or canonical TGF- β signaling pathway relies on a family of intracellular proteins called Smads. Eight different Smads, divided in three different classes, have been established: Receptor activated Smads (R-smads), which include Smad2, Smad3, Smad5 and Smad8; Inhibitory Smads (I-Smads), such as Smad6 and Smad7, and a common Smad4. TGF- β signals through a heteromeric complex of two type I and two type II transmembrane serine-threonine kinase receptors. Upon TGF- β binding, type II receptors bind to and phosphorylate the type I receptors in a conserved Gly/Ser-rich sequence, resulting in the activation of their kinase activity. Type I receptors phosphorylate Smad2 and Smad3, which form a trimeric complex with Smad4 and translocate to the nucleus, where they associate with DNA binding transcription factors in order to modulate target gene expression³⁹. Among the genes regulated we find I-Smads. In particular, Smad2/Smad3 induce the expression of Smad7, which inhibits TGF- β signaling by preventing receptor-mediated activation of R-Smads, as well as mobilizing Smurf E3 ubiquitin-ligases to receptors in order to target them for proteasomal degradation, thus establishing a negative feedback loop⁴⁰.

On the other hand, TGF- β can also activate Smad-independent intracellular pathways, including MAP Kinase (MAPK), Rho-like GTPase signaling pathways, and phosphatidylinositol-3-kinase (PI3K)/AKT pathways³⁹.

TGF- β signaling is involved in the regulation of numerous cellular functions and pathways, including cell growth and differentiation, modulation of immune responses or apoptosis, among others. Throughout cancer development, TGF- β has been shown to play two antagonistic roles: In normal and early-stage tumor cells, it acts as a tumor suppressor. However, as the tumor progresses, cells acquire the ability to escape the suppressive effects of TGF- β and shift their response towards the activation of a proliferative and invasive program⁴¹. In this context, TGF- β has been characterized as a key regulator of EMT³⁵.

3. HYPOTHESIS AND OBJECTIVES

We hypothesize that SWI/SNF might play an important role in maintaining cellular differentiation/identity, likely by regulating TGF- β signaling. The main objective of this project is to characterize molecular and transcriptional alterations as the result of transient inhibition of a SWI/SNF subunit, ARID2, in cellular models of human lung cancer.

The objectives of this work are the following:

1. To generate ARID2 knock-down lung cancer cellular models
2. To explore the consequences of ARID2 loss on TGF- β signaling pathway and its correlation with the presence of common oncogenic mutations.
3. To identify genes whose expression depends on ARID2-containing SWI/SNF complexes, especially those involved in TGF- β signaling.

4. MATERIALS AND METHODS

4.1. Cell biology methods

4.1.1. Cell lines and culture conditions

Human cancer cell lines A549 (ATCC® CRM-CCL-185™), NCI-H460 (ATCC® HTB-177™), and NCI-H1437 (ATCC® CRL-5872™) were used in this work. NCI-H460 and NCI-H1437 cell lines were grown in RPMI 1640, and A549 in DMEM. In both cases, culture media were supplemented with 10% fetal bovine serum (FBS), 2 µg/mL of ciprofloxacin (CPX), and 10 µg/mL of gentamycin. All cell lines were maintained at 37°C and 5% CO₂ conditions.

Cells were subcultured in order to keep them below 80% confluence. To this end, growth media was aspirated, cells were washed with PBS 1X and then detached with 0.25% Trypsin-EDTA at 37°C. Finally, medium was added to neutralize the activity of trypsin and cells were collected.

In order to have a broad representation of the different mutational backgrounds present in patients, a diverse set of human lung cancer cell lines were used. NCI-H460 and A549 harbor inactivating mutations in *KRAS*: Q61H and G12S, respectively, whereas NCI-H12437 has an inactivating mutation in *TP53*. A549 and NCI-H460 cell lines derive from primary tumors of lung adenocarcinoma and large cell lung carcinoma, respectively, whereas NCI-H1437 was obtained from a pleural effusion metastasis of a patient with NSCLC.

Table 1 | Summary of the cell lines used.

Cell line	Tumor type	Growth medium	Mutations
A549	Lung adenocarcinoma	DMEM +10% FBS	KRAS pG12S
NCI-H460	Large cell lung carcinoma	RPMI 1640 + 10% FBS	KRAS pQ61H
NCI-H1437	NSCLC	RPMI 1640 + 10% FBS	TP53 pR267P

4.1.2. Mycoplasma PCR

In order to test mycoplasma contamination, 500 µL-1mL of cell culture supernatant from cells with a minimum of 48 hours was collected and centrifugated to discard detritus. Then, PCR mix was prepared to a final volume of 20 µL with 12.5 µL of NZYTaq Master Mix 2X, 0.4 µL of each forward and reverse primers (table 2), 10.7 µL of nuclease-free H₂O and 1 µL of DNA (PCR protocol is detailed in table 3). PCR products were resolved in a 1% agarose gel and revealed with Sybr-safe and revealed using Gel Doc™ XR+

System (Bio-Rad, # 1708195EDU). Myco-positive samples displayed a band around 480 bp.

Cells that tested positive were treated with 6 µg/mL of CPX and 10 µg/mL for 7 days, after which supernatant was harvested and tested again.

Table 2 | Primers used in mycoplasma PCR.

Primer	Sequence
Forward	5'- GGCGAATGGGTGAGTAACACG -3'
Reverse	5'- CGGATAACGCTTGCGACCTATG -3'

Table 3 | Mycoplasma PCR protocol.

Step	Temperature	Time	Cycles
DNA denaturation	95°C	3 min	x30
DNA denaturation	95°C	30 sec	
Primers annealing	55°C	30 sec	
Extension	72°C	30 sec	
Final extension	72°C	5 min	

4.1.3. Transfection of cell cultures

Twenty-four hours after plating, cells were transfected with a short interfering RNA (siRNA) targeting *ARID2* to achieve gene silencing. In order to enhance the silencing efficiency, two siRNAs targeting different sites of *ARID2* were used. A non-targeted siRNA was used as a control.

For each well of a six-well dish, two solutions were prepared with a total volume of 150 µL of MEM growth medium, one of them with 9 µg of Lipofectamine™ 2000 (Invitrogen, #11668027) and the other one with 100 pmol of the siRNA molecule. After 5 minutes at RT, both solutions were mixed together. After 20 minutes at RT, cells' growth medium was replaced and the final mix was added in a drop-wise manner in each well.

Twenty-four hours after the transfection, medium was changed. Silencing efficiency was monitored via analyzing *ARID2* mRNA and protein levels by RT-qPCR and Western Blot.

4.2. Molecular biology methods

4.2.1. RNA extraction

RNA was purified with NZY Total RNA Isolation kit (NZYtech, #MB13402), following the protocol established by the kit. Cells were harvested as previously explained and placed in eppendorf tubes. Subsequently, eppendorf tubes were centrifugated at 1,200 rpms for 4 minutes and 4°C and supernatant was discarded, pellets were washed with PBS 1X and centrifugated again. PBS 1X was discarded and pellets were stored at -80°C until RNA extraction.

β-Mercaptoethanol was added to each pellet and 20G syringes were used for cell lysis. Then, samples were transferred to spin columns and centrifugated at 11,000g for 1 minute. To reduce the presence of DNA and increase the yield, a digestion with DNase I was performed. Before the elution, several washes with different buffers were done. Finally, columns were placed in sterile eppendorfs and 30-50 µL of RNase-free water were added. Columns were centrifugated and the follow-through containing the RNA was collected and stored at -80°C. RNA concentration was quantified using a Nanodrop spectrophotometer (Thermo Fischer Scientific, ND-2000).

4.2.2. Reverse Transcription quantitative PCR (RT-qPCR)

Reverse transcription reaction was performed using PrimeScript™ RT reagent Kit (Takara, #RR037A), following the manufacturer's instructions. A reaction mix was prepared to a final volume of 10 µL with 2 µL of 5X PrimeScript Buffer, 0.5 µL of PrimeScript RT Enzyme Mix I, 0.5 µL of 50 µM Oligo dT Primer (25 pmol), 2 µL of 100 µM Random 6 mers (200 pmol) and total RNA in RNase-free H₂O. The reaction mix was incubated for 15 minutes at 37°C, after which the reverse transcriptase was inactivated by heating the samples were heated at 85°C for 5 seconds.

Quantitation of mRNA levels was performed in StepOnePlus™ Real-Time PCR system (Applied Biosystems) using PowerUp SYBR Green Master Mix 2X (Applied Biosystems, #A25777). Twenty µL of reaction mix were prepared by adding 10 µL of PowerUp SYBR Green Master Mix 2x, 1.8 µL of 10 µM forward and reverse primers (800 nM), 5.4 µL nuclease-free H₂O, and 1 µL of cDNA. PCR program is detailed in table 5.

Expression of the housekeeping gene β-Actin was used as an internal control. Normalization of gene expression was performed as detailed:

$$\Delta Ct = Ct_{(\text{gene of interest})} - Ct_{(\text{housekeeping gene})}$$

$$\Delta\Delta Ct = \Delta Ct_{(\text{siARID2})} - \Delta Ct_{(\text{siControl})}$$

Relative expression: $2^{-\Delta\Delta Ct}$

Table 4 | Nucleotide sequence of primers used in RT-qPCR.

Gene	Primer	Sequence
ARID2	Forward	5'- CAGCCCATAACTTTGACGCA -3'
	Reverse	5'- TGGTGCAATTCCATCTTCCT -3'
TGFB2	Forward	5'- CTGCAGCATCACCTCCATCTGT -3'
	Reverse	5'- CTTGGGGTCATGGCAAACCTGTC -3'
SMAD7	Forward	5'- CCCTCCTTACTCCAGATACGCG -3'
	Reverse	5'-CCCAGGGGGCCAGATAATTCGT -3'
ACTB	Forward	5'-CCCAGCACAATGAAGATCAA-3'
	Reverse	5'-CGATCCACACGGAGTACTTG-3'

Table 5 | RT-qPCR program

Temp	Time	Cycles	Melting curve
50°C	2 min	Hold	X40
95°C	2 min	Hold	95°C 15 sec
95°C	15 sec	X40	60°C 1 min
60°C	1 min		95°C 15 sec

4.2.3. Western Blot

Protein extraction from cell cultures

Cells were washed with PBS1x and then lysed by adding 200 uL of a mix containing RIPA buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 % NP-40, 1 mM Sodium Orthovanadate, 1 mM NaF) and a phosphatase inhibitors cocktail. Cell lysis was performed by mechanical disruption of the samples using a cell scraper. Total lysates were collected in eppendorf tubes and sonicated for one cycle of 30 seconds at 50% amplitude. Following sonication, samples were centrifuged at 12,000 rpm for 15 minutes and 4°C to remove debris. Finally, supernatants were collected and stored at -80°C. Protein concentration was quantified using Qubit™ Protein Assay Kit.

Western Blot analysis

50 ug of total protein lysates were diluted in 5X Laemli buffer (250 mM Tris pH 6.8, 20% β-mercaptoethanol, 10% SDS, 0.25% bromophenol blue, and 50% glycerol), heated for 5 minutes at 95°C and resolved by SDS-PAGE in 8% polyacrylamide gels at 130V and 180V (for stacking and running gels, respectively). Once the electrophoresis finished, proteins were transferred to nitrocellulose membranes using a wet transfer blot at 400

mA for 1 hour and 30 minutes using the transfer buffer (1X Tris-Glycine, 20% methanol and 0,1% SDS).

Membranes were blocked for 1 hour at room temperature (RT) in either a 4% BSA, or a 10% solution in TTBS (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Tween 20), for phosphorylated and regular proteins, respectively. Then, membranes were blotted overnight with the primary antibodies (diluted in TTBS and 1% BSA) at 4°C.

Next day, membranes were washed three times in TTBS and incubated for 1 hour at RT with the secondary antibody, conjugated with Horseradish Peroxidase (HRP) (antibodies are given in table 6). Finally, membranes were washed three times in TTBS and blots were developed using Enhanced Chemiluminescence (ECL) (Solution 1: 10% Tris HCL 1M pH 8.5, 1% luminol, 0.45% coumaric acid; Solution 2: 10% Tris HCL 1M pH 8.5, 0.6% H₂O₂) and Amersham ImageQuant 800 (Cytiva).

Table 6 | Antibodies used in Western Blot analysis.

	Antibody	Dilution	Reference
Primary antibodies	Anti-ARID2	1:1,000	Santa Cruz Biotech: E-3. SC-166117
	Anti-pSMAD2/3	1:500	Cell Signaling #8828
	Anti-Actin	1:1,000	Santa Cruz Biotech: I-19. SC-1616
Secondary antibodies	Goat anti-mouse	1:10,000	Santa Cruz Biotech: SC-2005
	Goat anti-rabbit	1:10,000	Santa Cruz Biotech: SC-2004
	Rabbit anti-goat	1:10,000	Santa Cruz Biotech: SC-2768

4.2.4. RNA-Seq Library Preparation

For RNA-Seq library preparation, 3 µg of RNA of each sample were used. First, in order to increase the proportion of total mRNA present in each sample, an enrichment step was performed using NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, #E7490). Next, mRNA was randomly fragmented by heating the samples at 94°C for 15 minutes.

cDNA was generated in two different steps. The first strand was generated using PrimeSript RT Reagent Kit (Takara, #RR037A) heating the samples at 37°C for 15 minutes, and then at 85°C for 5 seconds to inactivate the reaction. Then, the remaining RNA were degraded by adding RNase HI (Thermo Fisher Scientific, #EN0201) to the mixture and incubating at 15°C for 2 hours. Then, the second strand was generated by adding T4 DNA polymerase (Thermo Fisher Scientific, #EP0062). Subsequently, samples were incubated at 15°C for 5 minutes before adding 5µL of EDTA 0.5M pH 8.0.

Double-stranded cDNA fragments were purified with Agencourt AMPure XP (Beckman Coulter, #A63881).

DNA fragments polished to obtain blunt ends and phosphorylated using an end-repair kit (Thermo Fisher Scientific, #K0771) containing T4 DNA polymerase for the 5'→3' extension; klenow fragment with 5'-3' nuclease activity for removing the 3' overhangs; and T4 polynucleotide kinase (PNK) for phosphorylation of 5' ends. Next, DNA fragments were adenylated on their 3' ends with Klenow Fragment exo- (Thermo Fisher, EP0422). Then, DNA fragments ligation with paired-end adapters, which had been previously generated through hybridization of two complementary oligonucleotides (sequences are given in table 7), was performed using T4 DNA ligase (Thermo Fisher, #EL0014) overnight at 18°C.

Table 7 | Sequence of P5 and P7 primers used to generate paired-end adapters.

Adapter	Sequence
Adapt_Truseq_P5a	5'-CTACACGACGCTCTTCCGATCT-3'
Adapt_Truseq_P5b	5'-GATCGGAAGAGCGTCGTGTAG-3'
Adapt_Truseq_P7a	5'-GATCGGAAGAGCACACGTCTG-3'
Adapt_Truseq_P5b	5'-CAGACGTGTGCTCTTCCGCTCT-3'

Finally, DNA fragments were amplified by PCR with Phusion high fidelity DNA polymerase (Thermo Fisher Scientific, #F530L) to increase the yield and to add specific sequences to identify the reads from each sample. Different combinations of P5 and P7 indexes were used in each sample.

Table 8 | Nucleotide sequence of P5 and P7 oligonucleotides used for RNA-seq library preparation. Specific 8 bp sequences used to identify each sample are highlighted in bold.

Primer	Sequence
D525	AATGATACGGCGACCAACGAGATCTACAC ACCTCCA AACTCTTTCCCTAC ACGACGCTCTTCCGATC
D701	CAAGCAGAAGACGGCATACGAGAT CGAGTA AATGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
D702	CAAGCAGAAGACGGCATACGAGAT TCTCCG AGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
D703	CAAGCAGAAGACGGCATACGAGAT AATGAGCG GTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
D704	CAAGCAGAAGACGGCATACGAGAT GGAATCTC GTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
D706	CAAGCAGAAGACGGCATACGAGAT ACGAATTC GTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
D707	CAAGCAGAAGACGGCATACGAGAT AGCTTCAG GTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT

D708	CAAGCAGAAGACGGCATAACGAGAT GCGCATT AGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
D709	CAAGCAGAAGACGGCATAACGAGAT CATAGCCG GTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
D710	CAAGCAGAAGACGGCATAACGAGAT TTGCGCGG AGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
D711	CAAGCAGAAGACGGCATAACGAGAT GCGCGAG AGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
D712	CAAGCAGAAGACGGCATAACGAGAT CTATCGCT GTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT
D713	CAAGCAGAAGACGGCATAACGAGAT TTGGAGGT GTGACTGGAGTTCAGACGT GTGCTCTTCCGATCT

Finally, the resulting PCR products were purified with Agencourt AMPure XP (Beckman Coulter, #A63881) and eluted in 20 µL of TE buffer. Concentrations were quantified using Quibit dsDNA BR assay kit, and samples were analyzed using High Sensitivity D1000 ScreenTape Assays (Agilent, #5067-5582/83) to assess the aspect of the libraries.

4.2.5. RNA-Seq data analysis

Raw sequencing data in fastq format was aligned against the human reference genome (GRCh38 version) using hisat2⁴². SAM files were then converted to BAM, cleaned, sorted and indexed using SAMTOOLS⁴³. Raw gene expression estimation from BAM files was estimated using HTSeq2 tool⁴⁴ with Ensembl database transcripts (version 110). DESeq2⁴⁵ was used to identify gene expression differences and ggplot and pheatmap libraries were used to generate graphical representations.

4.3. Statistical analysis

Statistical differences between two groups of biological replications in RT-qPCR were estimated using t-test. DESeq2 packages uses its proper statistical analysis assuming a negative binomial distribution in the gene expression data.

5. RESULTS

5.1. ARID2 silencing using short interfering RNAs (siRNAs)

In order to explore the effects of ARID2 reduction, we used short interfering RNAs to knock down ARID2 expression in three distinct lung cancer cell lines with different mutational backgrounds: A549 and NCI-H460, harboring mutations in *KRAS* (Q61L and G12D, respectively), and NCI-H1437, harboring inactivating mutations in *TP53*. As it is shown in Figure 3, three days after siRNA transfection, RT-qPCR and Western Blot analysis demonstrated a significant reduction of ARID2 at both mRNA and protein level.

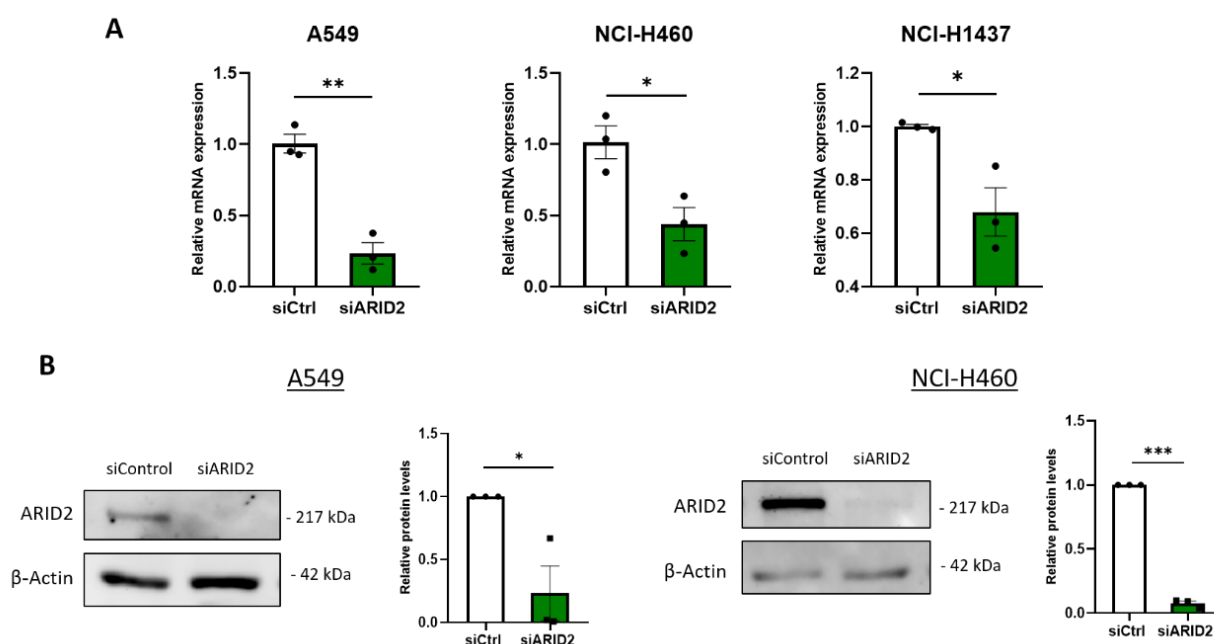


Figure 3 | Knock-down validation of A549, NCI-H460 and NCI-H1437 cell lines by siRNA transfection. **A.** Relative *ARID2* mRNA expression measured by RT-qPCR. **B.** Relative ARID2 protein levels measured by Western Blot. β-Actin was used as a control. Graph values correspond to the mean ± SEM (error bars) of independent experiments quantification. Statistical significances were assessed using Student t-test and are indicated by asterisks: *p-value < 0.05; **p-value < 0.01, ***p-value < 0.0001.

As expected, our results showed a successful reduction of ARID2 in all three lines at both transcript (A549: 0.23 ± 0.075 , p-value = 0.0015; NCI-H460: 0.43 ± 0.116 , p-value = 0.0246; NCI-H1437: 0.67 ± 0.09 , p-value = 0.0244) (Figure 3A) and protein levels (A549: 0.23 ± 0.218 , p-value = 0.0245; NCI-H460: 0.075 ± 0.031 ; p-value < 0.001) (Figure 3B).

5.2. Effects of ARID2 depletion in TGF- β -Smad signaling

5.2.1. Expression of TGF- β type 2 receptor (*TGFBR1*) in ARID2-deficient cells.

Once we confirmed that siRNAs transfection efficiently downregulated ARID2 expression, we studied TGF- β intracellular signaling pathway in control and ARID2-deficient cells. For this purpose, we measured *TGFBR1* transcript expression by RT-qPCR.

The analysis revealed decreased *TGFBR1* mRNA levels in both siARID2 and siControl cells following TGF- β 1 induction when compared with non-induced siControl condition (Fig.4). We hypothesize that this observation might be the result of an intrinsic negative feedback mechanism induced by the in activation of the pathway.

Interestingly, although no statistically significant differences were achieved, siARID2 cells showed a downregulation tendency in *TGFBR1* expression in all cell lines (A549: 0.72 ± 0.053 , $P=0.094$; NCI-H460: 0.43 ± 0.064 , $P=0.144$; NCI-H1437: 0.78 ± 0.079 , $P=0.282$), which suggest that ARID2 subunit might be necessary to facilitate *TGFBR1* transcription. Remarkably, *KRAS*-mutated NCI-H460 cell line exhibited the strongest reduction in *TGFBR1* levels.

Surprisingly, *TGFBR1* mRNA levels are downregulated after TGF- β induction, which suggests that *TGFBR1* is downregulated in a negative feedback loop, although we did not observe significant differences in *TGFBR1* mRNA levels between siARID2 (A549: 0.69 ± 0.153 ; NCI-H460: 0.52 ± 0.206 ; NCI-H1437: 0.34 ± 0.02) and siControl conditions (A549: 0.79 ± 0.053 ; NCI-H460: 0.43 ± 0.064 ; NCI-H1437: 0.32 ± 0.055), regardless of the cell line.

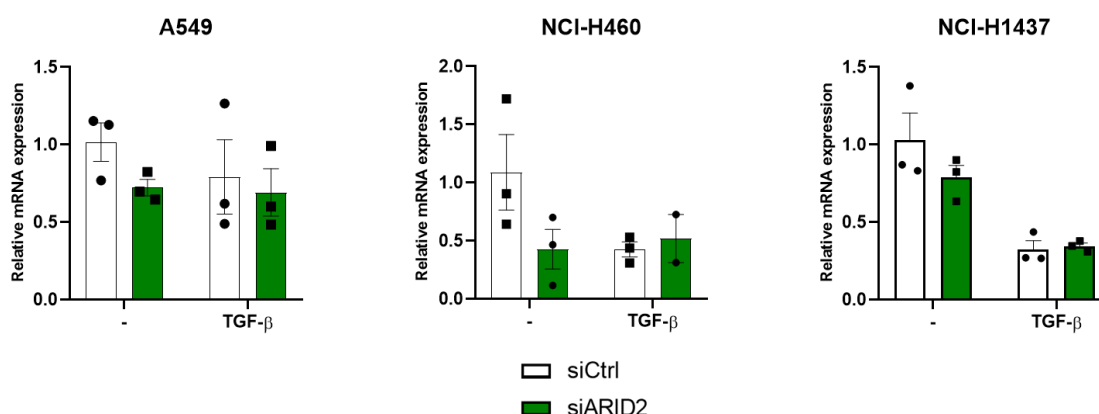


Figure 4 | *TGFBR1* mRNA expression levels in non-induced (-) and TGF- β 1- induced A549, NCI-H460 and NCI-H1437 cell lines. β -Actin (*ATCB*) expression was used as a control. Graph values correspond to the mean $\pm$ SEM of independent experiments quantification. Statistical significances were assessed using Student t-test.

5.2.2. Effects of ARID2 depletion in TGF- β -Smad downstream effectors.

To further characterize the effects of ARID2 depletion in TGF- β signaling, we studied the pathway activation in the different conditions by analyzing the downstream phosphorylation cascade. For this purpose, cells were lysed 40 minutes after TGF- β 1 addition, the moment around which the Smad2/3 phosphorylation reaches its peak. Then, levels of phosphorylated Smad2/3 (pSmad2/3) were assessed by Western Blot analysis.

As expected, TGF- β 1 induction resulted in a significant increase in Smad2/3 phosphorylation in comparison to untreated cells. However, we did not appreciate significant changes in pathway activation in ARID2-deficient cells when compared to control condition in either model (Figure 5A).

Among many other target genes, the trimeric TF complex formed by SMAD2, SMAD3 and SMAD4 induces the expression of the inhibitory Smads, which include SMAD6 and SMAD7. These block the signaling pathway, establishing a negative feedback. In particular, Smad7 is a well characterized inhibitor of type I and type II receptors. As a member of SWI/SNF complexes, ARID2 modulates gene accessibility and thus controls gene expression in different cellular contexts. Hence, we wondered whether ARID2 plays any role in modulation of *SMAD7* gene expression. To elucidate this, we assessed the levels of *SMAD7* transcript in control and ARID2-deficient cells in presence or absence of TGF- β 1 ligand.

In line with the previous result, *SMAD7* transcript levels were markedly increased in both siControl and siARID2 TGF- β -induced cells 2 hours after induction, in comparison to the untreated condition, indicating a successful activation of the route. Furthermore, similar levels of *SMAD7* mRNA were observed in controls and siARID2-transfected cells, indicating that ARID2 does not affect *SMAD7* expression (Figure 5B).

Moreover, untreated siARID2 cells did not display any signal of pathway activation, neither by SMAD2/3 phosphorylation nor SMAD7 upregulation. This seems to indicate that the *TGFBRII* downregulation observed following ARID2 silencing is not directly caused by an overactivation of the pathway and its intrinsic negative feedback, supporting the hypothesis that *TGFBRII* might be a direct target of ARID2-containing SWI/SNF complexes.

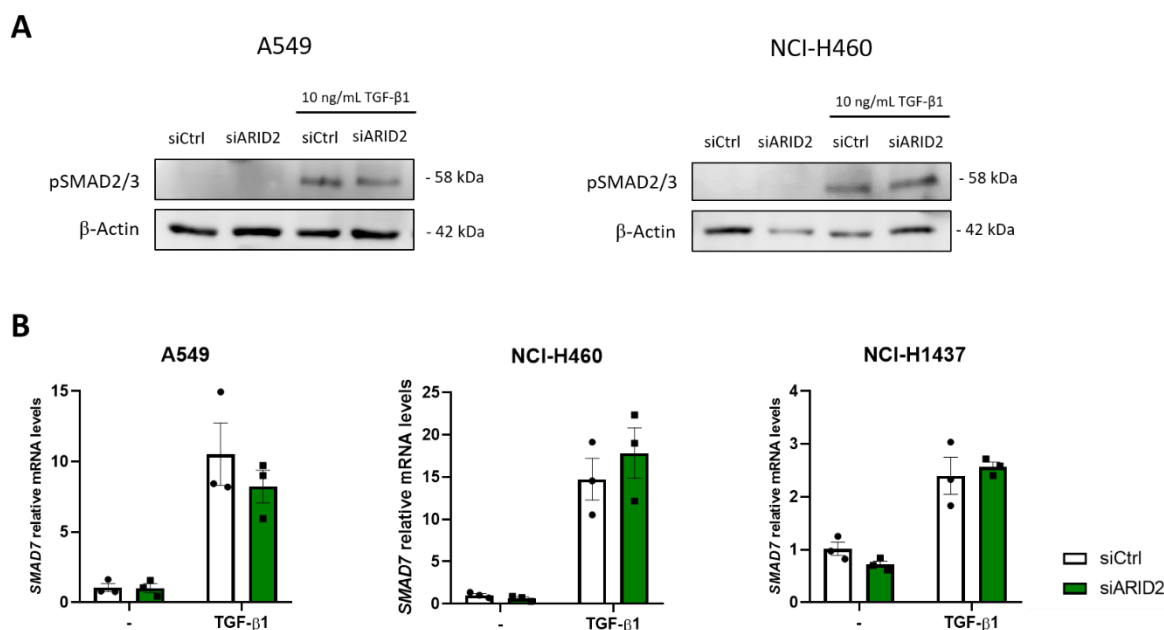


Figure 5 | Effects of ARID2 silencing in TGF- β -Smad pathway downstream effectors.
A. Representative Western Blot analysis of pSmad2/3 levels in siControl and siARID2 cells with and without TGF- β 1 induction. **B.** Representative *SMAD7* mRNA levels in non-induced (-) and TGF- β 1-induced siControl and siARID2 cells. Graph values correspond to the mean $\pm$ SEM (error bars) of independent experiments quantification. β -Actin was used as a control. Statistical significances were assessed using Student t-test.

5.3. Characterization of ARID2 silencing effects at a transcriptional level.

In order to have a broader view of the effects of *ARID2* depletion, we performed RNA-seq analysis of cells from all four different conditions. Unfortunately, due to our limited time, we were only able to perform the analysis of NCI-H460 cell line.

5.3.1. Differentially expressed genes in ARID2-deficient cells

First, we measured the number of differentially expressed genes (DEGs) in siARID2 in comparison to siControl condition. The analysis revealed a similar quantity of upregulated and downregulated genes in siARID2 cells (Figure 6A). This result indicates that ARID2 activity is required for both activation and repression of transcription of different target genes.

Next, to further characterize the main gene networks and cellular pathways affected by ARID2 disruption, we performed a functional analysis using Gene Set Enrichment Analysis (GSEA), using the Hallmark and the Gene Ontology – Biological Function databases.

The analysis revealed alterations in many diverse gene networks with multiple diverse functions. Among those with negative enrichment scores, the analysis identified TGF- β signalling, supporting our hypothesis that ARID2-deficiency is accompanied by a downregulation of TGF- β . Additionally, genes involved in DNA repair mechanisms were also found negatively regulated (Figure 6B).

On the other hand, gene networks involved in EMT were found enriched in ARID2-deficient cells. In the same line of findings, lung-cell differentiation markers and apical junctions showed negative enrichment scores (Figure 6B). These results suggest that ARID2 might be required for the expression of lung cell lineage-specific gene networks, as well as to sustain the expression of epithelial markers.

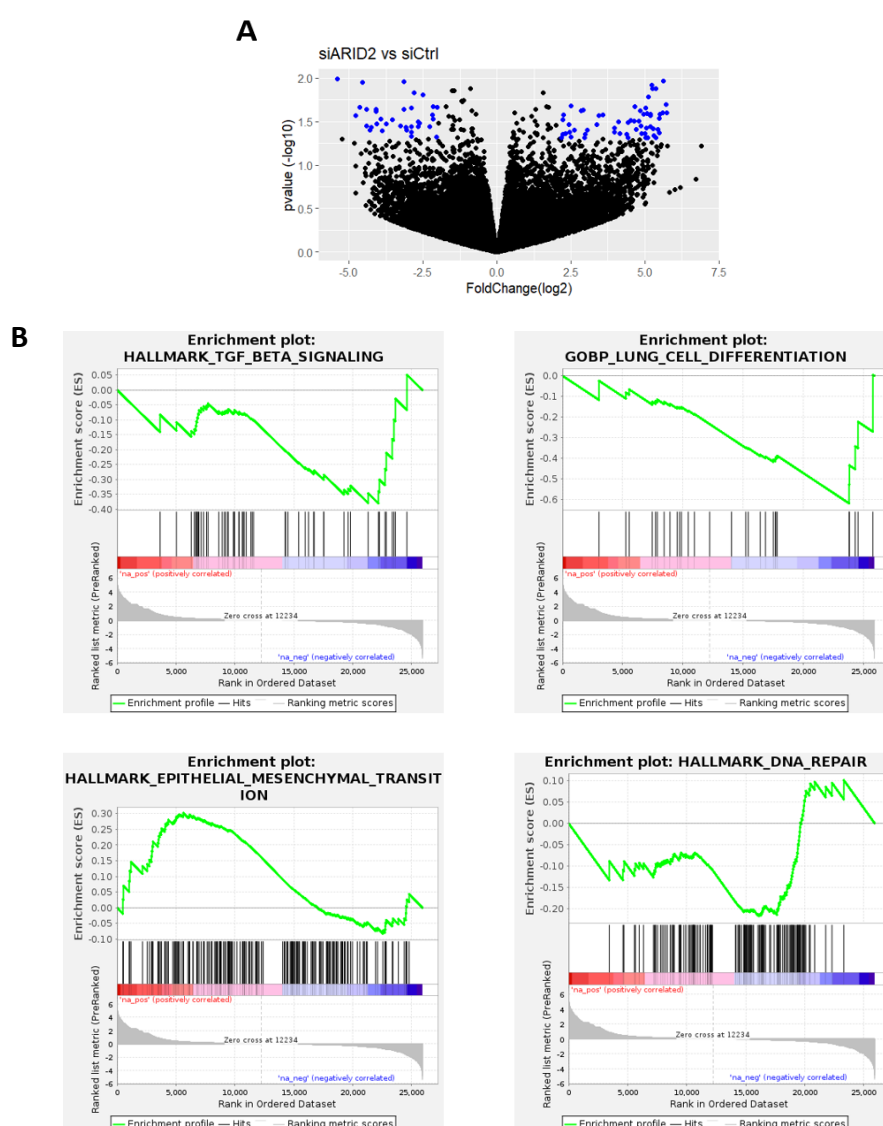


Figure 6 | **A.** Volcano plot representation of upregulated and downregulated genes in siARID2 cells compared to siControl. Genes with a p-value < 0.05 and an absolute log₂fold change > 2 are colored in blue. **B.** Gene Set Enrichment Analysis of several gene networks and pathways differentially regulated in siARID2 cell in comparison to siControl.

5.3.2. Effects of ARID2 loss in TGF- β target gene expression.

To further study ARID2 regulation of TGF- β signaling, we analyzed the number of differentially expressed genes (DEGs) in TGF- β 1-induced siControl and siARID2 cells in comparison to untreated controls. Again, results showed a similar number of upregulated and downregulated genes in both conditions.

Next, in order to elucidate whether ARID2 subunit regulates the expression of TGF- β target genes, we analyzed those genes that showed significant expression changes after TGF- β induction in siControl cells that did not showed differences after ARID2 knock-down (Figure 7B-C). This analysis allowed us to identify a total of 30 genes.

Among the most significantly downregulated genes in siARID2, we found several pseudogenes and non-protein-coding genes, including *LINC01609*, *c21orf91-ot1*, or *PITPNM2-AS1*, although we have not yet found their biological significance. Interestingly, *MT1G* (encoding Metallothionein 1G), which has been described as a tumor suppressor and EMT inhibitor in different cancer types^{46,47}, was downregulated in ARID2-deficient cells. Regarding the upregulated genes, it is worth mentioning *MYCL*, an oncogene from Myc family that is recurrently mutated in SCLC⁴⁸.

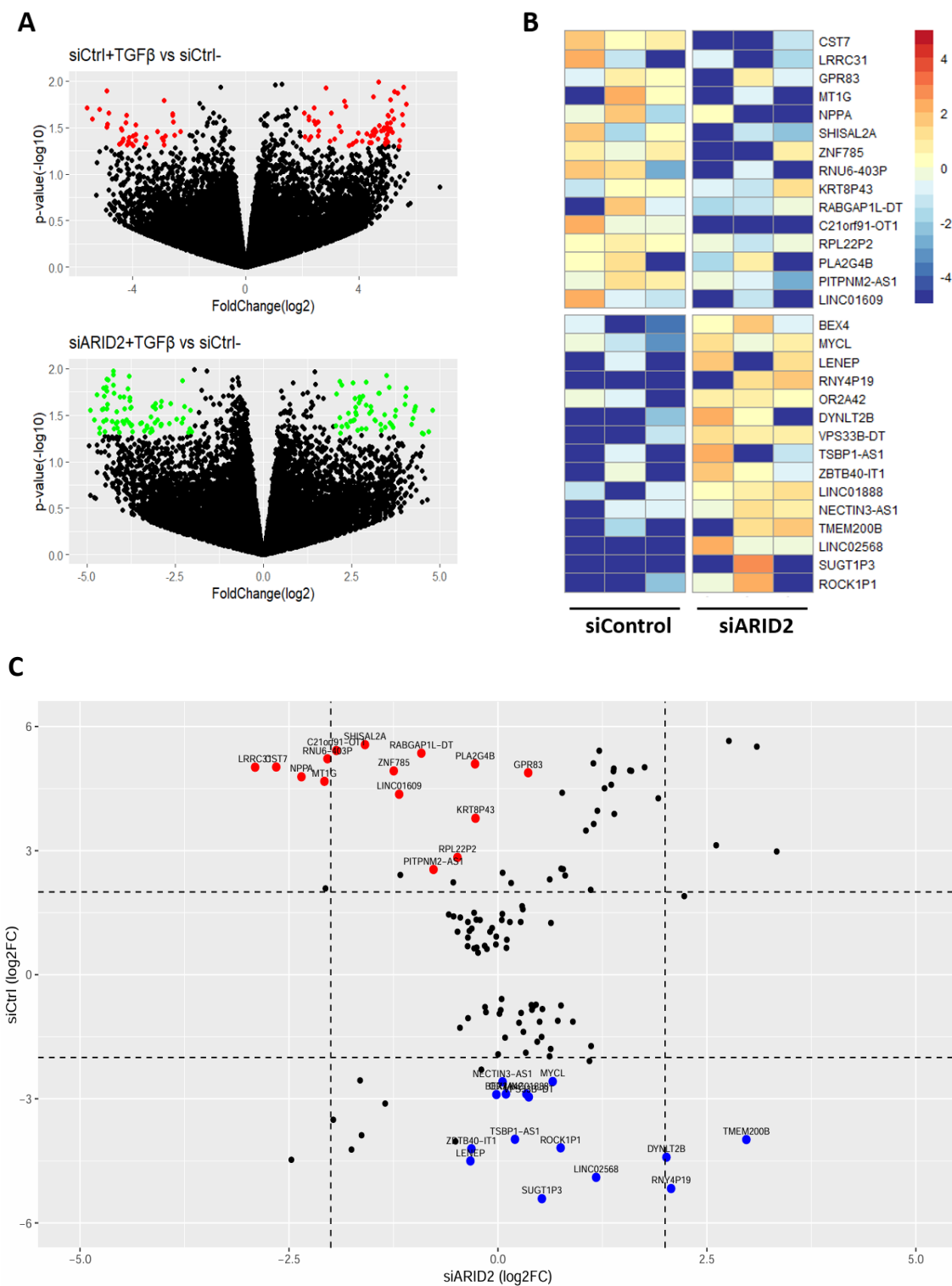


Figure 7 | **A.** Representative volcano plot of DEGs in siControl and siARID2 cells after TGF-β1 induction. Genes with a p-value < 0.05 and an absolute log2FoldChange > 2 are colored in red (siControl) or green (siARID2). **B.** Heatmap and **C.** Scatter plot representation of the most significant DEGs between TGF-β1-induced siControl and siARID2 cells. Gene expression differences are represented according to the log2 of their fold change.

6. DISCUSSION

In the present work, we successfully generated ARID2 knock-down cellular models in different human lung cancer cell lines, where ARID2 downregulation was detected both at mRNA and protein levels. ARID2 depletion resulted in an alteration of TGF- β intracellular pathway, evidenced by a downregulation of *TGFBR1* expression. This finding was consistent across the three different cell lines regardless of their different mutational backgrounds, which indicates that it is not dependent on the activation of other oncogenic signals. We proposed two possible mechanisms to explain decreased *TGFBR1* levels: 1) ARID2-deficient cells exhibit increased basal TGF- β pathway activation which exerts a negative feedback loop, or 2) ARID2 is directly responsible of maintaining *TGFBR1* expression. Our results did not reveal markers of TGF- β signaling activation in siARID2 cells, neither by Smad2/3 phosphorylation nor by *SMAD7* expression, which points against the negative feedback model. Consequently, our findings support the hypothesis that *TGFBR1* gene is a direct target of ARID2 transcriptional regulation, and therefore, its expression is impaired by SWI/SNF complexes lacking ARID2 subunit.

Supporting the hypothesis of a suppression of TGF- β signaling through the downregulation of *TGFBR1* expression, GSEA results revealed a negative score in the gene networks involved in TGF- β signaling. Nevertheless, we were not able to observe differences in the activation status of the intracellular downstream effectors between siARID2 and siControl conditions following induction with TGF- β 1, measured by phosphorylation of Smad2 and Smad3, as well as *SMAD7* mRNA expression. These results seem to indicate that the *TGFBR1* reduction observed in ARID2-deficient cells is not enough to prevent the activation of the Smad pathway, since the remaining levels of receptor could be sufficient to maintain the signal transduction. However, it is possible that ARID2 is required for the expression of TGF- β target genes, which would explain the negative enrichment scores observed by GSEA. Importantly, signaling through TGF- β -Smad pathway displays cycles of activation and inactivation due to a dynamic regulation mediated by negative feedback loops, such as the one induced by Smad7. Here, we only analyzed the effects of a 2-hour induction with TGF- β 1 ligand. Therefore, it would be interesting to analyze the effects at different time points in order to have a wider view of ARID2 loss consequences. On the other hand, TGF- β also signals through multiple Smad-independent mechanisms, and it is possible that, even though the remaining *TGFBR1* levels are able to sustain Smad downstream activation, they are not sufficient to maintain the crosstalk with other pathways, thus causing an imbalance between the activation of Smad and non-Smad signaling routes.

Despite the fact that the reduction in *TGFBR1* expression levels was only moderate, RNA-seq analysis of siControl and siARID2 cells previously induced with TGF- β 1 revealed several significant DEGs in between the two conditions, which seems to confirm the hypothesis that ARID2 is involved in the expression of TGF- β target genes. Among others, we identified a significant downregulation of *MT1G* in ARID2-deficient cells. *MT1G* encodes Metallothionein 1G, a protein involved in metal ion homeostasis, providing zinc/copper for enzymes and transcription factors, as well as protecting against oxidative stress⁴⁹. MT1G has been found mutated and has been associated with suppressor tumoral roles in different cancer types, such as gastric carcinoma, in which its overexpression was shown to inhibit growth and EMT through the regulation of PI3K/Akt pathway⁵⁰. Similarly, we found a significant downregulation of *MYCL* after ARID2-deficiency. *MYCL* is a proto-oncogene belonging to the MYC family of transcription factors. It is worth mentioning that overexpression and amplification of *MYCL* have been recurrently described in small cell lung cancer, playing a key role in malignant transformation and acquisition of therapeutic resistance⁴⁸.

Finally, besides TGF- β pathway defects, GSEA revealed several other differentially regulated gene networks in ARID2-deficient cells. In particular, gene networks involved in EMT showed positive enrichment scores. Remarkably, loss of ARID2 also correlated with a downregulation in gene sets involved in lung-cell lineage specific differentiation pathways, as well as and apical junctions. Taken together, these results suggest that ARID2 is essential for the maintenance of a lung epithelial cell phenotype.

Here, we propose a model whereby loss of ARID2 leads to the acquisition of a malignant and invasive phenotype in lung tumour cells. This is likely the consequence of an suppression in the function of TGF- β signaling pathway, facilitating the loss of epithelial lung cell differentiation status and thus resulting in enhanced EMT.

7. CONCLUSIONS

1. ARID2 depletion in lung cancer cells results in a downregulation of *TGFBRII* mRNA levels.
2. The described downregulation of *TGFBRII* was not the consequence of an intrinsic negative feedback loop induced by TGF- β activation, suggesting that *TGFBRII* expression is directly regulated by ARID2.
3. Despite *TGFBRII* downregulation in ARID2-deficient cells, no differences in TGF- β -Smad2/3 phosphorylation were observed after induction with TGF- β .
4. ARID2 loss is associated with a transcriptional program characterized by an enrichment in EMT gene networks, as well as a downregulation of genes involved in TGF- β signaling and lung epithelial cell differentiation markers.
5. Several genes were differentially expressed in ARID2 deficient cells and control conditions upon TGF- β 1 induction, indicating that ARID2 regulates the expression of certain TGF- β target genes.

8. REFERENCES

1. Schabath, M. B. & Cote, M. L. Cancer progress and priorities: Lung cancer. *Cancer Epidemiology Biomarkers and Prevention* 28, 1563–1579 (2019).
2. Rodriguez-Canales, J., Parra-Cuentas, E. & Wistuba, I. I. Diagnosis and molecular classification of lung cancer. *Cancer Treat Res* 170, 25–46 (2016).
3. Rudin, C. M., Brambilla, E., Faivre-Finn, C. & Sage, J. Small-cell lung cancer. *Nat Rev Dis Primers* 7, (2021).
4. Collisson, E. A. *et al.* Comprehensive molecular profiling of lung adenocarcinoma: The cancer genome atlas research network. *Nature* 511, 543–550 (2014).
5. Hammerman, P. S. *et al.* Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489, 519–525 (2012).
6. George, J. *et al.* Comprehensive genomic profiles of small cell lung cancer. *Nature* 2015 524:7563 524, 47–53 (2015).
7. Duma, N., Santana-Davila, R. & Molina, J. R. Non–Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clin Proc* 94, 1623–1640 (2019).
8. Wilson, B. G. & Roberts, C. W. M. SWI/SNF nucleosome remodellers and cancer. *Nature Reviews Cancer* 2011 11:7 11, 481–492 (2011).
9. Malone, H. A. & Roberts, C. W. M. Chromatin remodellers as therapeutic targets. *Nature Reviews Drug Discovery* 2024 23:9 23, 661–681 (2024).
10. Shain, A. H. & Pollack, J. R. The Spectrum of SWI/SNF Mutations, Ubiquitous in Human Cancers. *PLoS One* 8, (2013).
11. Stern, M., Jensen, R. & Herskowitz, I. Five SWI genes are required for expression of the HO gene in yeast. *J Mol Biol* 178, 853–868 (1984).
12. Neigeborn, L. & Carlson, M. Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. *Genetics* 108, 845–858 (1984).
13. Mashtalir, N. *et al.* Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell* 175, 1272–1288.e20 (2018).
14. Mittal, P. & Roberts, C. W. M. The SWI/SNF complex in cancer — biology, biomarkers and therapy. *Nature Reviews Clinical Oncology* 2020 17:7 17, 435–448 (2020).
15. Alver, B. H. *et al.* The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat Commun* 8, (2017).
16. Michel, B. C. *et al.* A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nature Cell Biology* 2018 20:12 20, 1410–1420 (2018).
17. Mashtalir, N. *et al.* Chromatin landscape signals differentially dictate the activities of mSWI/SNF family complexes. *Science* (1979) 373, 306–315 (2021).

18. Versteeg, I. *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* 394, 203–206 (1998).
19. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet* 45, 592–601 (2013).
20. Fukuoka, J. *et al.* Chromatin Remodeling Factors and BRM/BRG1 Expression as Prognostic Indicators in Non-Small Cell Lung Cancer.
21. Bell, E. H. *et al.* SMARCA4/BRG1 is a novel prognostic biomarker predictive of cisplatin-based chemotherapy outcomes in resected non-small cell lung cancer. *Clinical Cancer Research* 22, 2396–2404 (2016).
22. Matsubara, D. *et al.* Lung cancer with loss of BRG1/BRM, shows epithelial mesenchymal transition phenotype and distinct histologic and genetic features. *Cancer Sci* 104, 266–273 (2013).
23. Zhang, Y. *et al.* ARID1A is downregulated in non-small cell lung cancer and regulates cell proliferation and apoptosis. *Tumor Biology* 35, 5701–5707 (2014).
24. Moreno, T. *et al.* ARID2 deficiency promotes tumor progression and is associated with higher sensitivity to chemotherapy in lung cancer. *Oncogene* 2021 40:16 40, 2923–2935 (2021).
25. Monterde, B. & Varela, I. Role of SWI/SNF chromatin remodeling genes in lung cancer development. *Biochem Soc Trans* 50, 1143–1150 (2022).
26. Brownlee, P. M., Meisenberg, C. & Downs, J. A. The SWI/SNF chromatin remodelling complex: Its role in maintaining genome stability and preventing tumorigenesis. *DNA Repair (Amst)* 32, 127–133 (2015).
27. Rother, M. B. & van Attikum, H. DNA repair goes hip-hop: SMARCA and CHD chromatin remodellers join the break dance. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372, (2017).
28. Shen, J. *et al.* ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. *Cancer Discov* 5, 752–767 (2015).
29. Ray, A. *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* 29, 6206–6219 (2009).
30. Dykhuizen, E. C. *et al.* BAF complexes facilitate decatenation of DNA by topoisomerase II α . *Nature* 497, 624–627 (2013).
31. Feng, H. *et al.* PBAF loss leads to DNA damage-induced inflammatory signaling through defective G2/M checkpoint maintenance. *Genes Dev* 36, 790–806 (2022).
32. Romero, O. A. *et al.* MAX inactivation in small cell lung cancer disrupts MYC-SWI/SNF programs and is synthetic lethal with BRG1. *Cancer Discov* 4, 293–303 (2014).

33. Romero, O. A. *et al.* The tumour suppressor and chromatin-remodelling factor BRG1 antagonizes Myc activity and promotes cell differentiation in human cancer. *EMBO Mol Med* 4, 603–616 (2012).
34. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* 15, 178–196 (2014).
35. Xu, J., Lamouille, S. & Derynck, R. TGF- β -induced epithelial to mesenchymal transition. *Cell Res* 19, 156–172 (2009).
36. Jiang, H. *et al.* Chromatin remodeling factor ARID2 suppresses hepatocellular carcinoma metastasis via DNMT1-Snail axis. *Proc Natl Acad Sci U S A* 117, 4770–4780 (2020).
37. Xu, S. *et al.* ARID1A restrains EMT and stemness of ovarian cancer cells through the Hippo pathway. *Int J Oncol* 65, (2024).
38. Matsubara, D. *et al.* Lung cancer with loss of BRG1/BRM, shows epithelial mesenchymal transition phenotype and distinct histologic and genetic features. *Cancer Sci* 104, 266–273 (2013).
39. Derynck, R. & Budi, E. H. Specificity, versatility, and control of TGF- β family signaling. *Sci Signal* 12, (2019).
40. Yan, X., Xiong, X. & Chen, Y. G. Feedback regulation of TGF- β signaling. *Acta Biochim Biophys Sin (Shanghai)* 50, 37–50 (2018).
41. Massagué, J. TGF β in Cancer. *Cell* 134, 215 (2008).
42. Zhang, Y., Park, C., Bennett, C., Thornton, M. & Kim, D. Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Res* 31, 1290–1295 (2021).
43. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* 10, (2021).
44. Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* 38, 2943–2945 (2022).
45. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 1–21 (2014).
46. Xu, G., Fan, L., Zhao, S. & Ouyang, C. MT1G inhibits the growth and epithelial-mesenchymal transition of gastric cancer cells by regulating the PI3K/AKT signaling pathway. *Genet Mol Biol* 45, (2022).
47. Wang, Y. *et al.* MT1G serves as a tumor suppressor in hepatocellular carcinoma by interacting with p53. *Oncogenesis* 8, (2019).
48. Massó-Vallés, D., Beaulieu, M. E. & Soucek, L. MYC, MYCL, and MYCN as therapeutic targets in lung cancer. *Expert Opin Ther Targets* 24, 101–114 (2020).
49. Coyle, P., Philcox, J. C., Carey, L. C. & Roife, A. M. Metallothionein: The multipurpose protein. *Cellular and Molecular Life Sciences* 59, 627–647 (2002).

50. Xu, G., Fan, L., Zhao, S. & Ouyang, C. MT1G inhibits the growth and epithelial-mesenchymal transition of gastric cancer cells by regulating the PI3K/AKT signaling pathway. *Genet Mol Biol* 45, (2022).