

# Identification of molecular targets of SWI/SNF alterations in cancer development

Master's thesis

**BEATRIZ MONTERDE MARTÍNEZ**

**Supervisor:** Ignacio Varela Egocheaga

Instituto de Biomedicina y Biotecnología de Cantabria. Universidad de Cantabria-CSIC.  
Santander. Spain.

Master in Molecular Biology and Biomedicine

Course 2017-18



## ABBREVIATIONS

<b>293T/17</b>	HEK 293T clone 17
<b>ATAC-Seq</b>	Assay for Transposase Accessible Chromatin sequencing
<b>ATP</b>	Adenosine triphosphate
<b>BAF</b>	BRG1-associated factor
<b>bp</b>	Base pair
<b>BSA</b>	Bovine Serum Albumin
<b>cDNA</b>	Complementary DNA
<b>CHD</b>	Chromodomain Helicase DNA-binding
<b>ChIP-Seq</b>	Chromatin Immunoprecipitation Sequencing
<b>Ct</b>	Cycle threshold
<b>DMEM</b>	Dulbecco's Modified Eagle Medium
<b>DSB</b>	DNA double-strand breaks
<b>EDTA</b>	Ethylenediaminetetracetic acid
<b>FBS</b>	Fetal bovine serum
<b>GSEA</b>	Gene set enrichment analysis
<b>HDAC</b>	Histone Deacetylase
<b>HEPES</b>	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
<b>ICLs</b>	Interstrand crosslinks
<b>IFN</b>	Interferon
<b>INO80</b>	Inositol auxotroph 80
<b>ISWI</b>	Imitation switch
<b>kDa</b>	Kilodaltons
<b>RNA-Seq</b>	Transcriptome sequencing
<b>NER</b>	Nucleotide excision repair
<b>NGS</b>	Next-generation sequencing
<b>NURD</b>	Nucleosome remodelling and deacetylases

<b>PBAF</b>	Polybrome-BRG1 associated factor
<b>PBS</b>	Phosphate-buffered saline
<b>PCR</b>	Polymerase Chain Reaction
<b>PEI</b>	Polyethylenimine
<b>qRT-PCR</b>	Reverse transcription and quantitative polymerase chain reaction
<b>qPCR</b>	Quantitative polymerase chain reaction
<b>rcf</b>	Relative centrifugal force
<b>shARID1A</b>	Short hairpin RNA against <i>ARID1A</i> gene
<b>shRNA</b>	Short hairpin RNA
<b>SDS</b>	Sodium dodecyl sulfate
<b>SDS-PAGE</b>	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
<b>SEM</b>	Standard error of mean
<b>SWI/SNF</b>	SWItch/Sucrose Non Fermentable
<b>SWR1</b>	SWi2/snf2-Related
<b>tGFP</b>	Turbo Green Fluorescent Protein
<b>tRFP</b>	Turbo Red Fluorescent Protein
<b>TBS-T</b>	Tris Buffer Saline-Tween 20
<b>TGFβ2</b>	Transforming growth factor beta 2

# CONTENTS

<b>1. Abstract</b> .....	<b>1</b>
<b>2. Introduction</b> .....	<b>2</b>
2.1. Cancer genomics .....	2
2.2. Chromatin remodelling complexes .....	3
2.3. Role of chromatin remodelling complexes in cancer .....	4
2.4. Influence of chromatin remodelling complexes in transcriptional regulation ..	6
2.5. Role of chromatin remodelling complexes in DNA repair .....	7
2.6. Mutual exclusivity between SWI/SNF subunits and other genes.....	8
2.7. Therapeutic exploitation of SWI/SNF mutations .....	9
2.8. Previous data obtained in our laboratory .....	9
<b>3. Working hypothesis and objectives</b> .....	<b>11</b>
<b>4. Materials and methods</b> .....	<b>12</b>
4.1. Cell lines and culture conditions .....	12
4.2. Generation of stably-transduced cell lines.....	12
4.3. Induction of shRNA and turboRFP expression.....	13
4.4. FACS sorting of stably-transduced cell lines.....	13
4.5. RNA isolation and qRT-PCR.....	14
4.6. Western Blot .....	15
4.7. cDNA NGS libraries.....	15
4.8. RNA-Seq analysis.....	17
<b>5. Results</b> .....	<b>18</b>
5.1. Generation of stably-transduced cell lines.....	18
5.2. FACS sorting of stably-transduced cells.....	19
5.3. Effect of <i>ARID1A</i> -deficiency on SK-OV-3 cells.....	21
5.3.1. Validation of <i>ARID1A</i> knock-down .....	21
5.3.2. Transcriptional alterations after <i>ARID1A</i> knock-down.....	23

<b>5.4. Effect of <i>ARID1A</i>-deficiency in Caco-2 cells .....</b>	<b>26</b>
5.4.1. Validation of <i>ARID1A</i> and <i>ARID1B</i> knock-down.....	26
5.4.2. Transcriptional alterations after <i>ARID1A</i> knock-down.....	29
<b>5.5. Generation of A549 stably-transduced cell lines .....</b>	<b>33</b>
5.5.1. Validation of <i>ARID1A</i> and <i>ARID1B</i> knock-down.....	33
<b>6. Discussion and on-going work.....</b>	<b>35</b>
<b>7. Conclusions .....</b>	<b>40</b>
<b>8. References .....</b>	<b>41</b>

## 1. Abstract

SWI/SNF chromatin remodelling complex has been described to be altered in nearly 20 % of all human tumour types, which places it among the most broadly mutated molecular systems in human cancer, just after *TP53*. However, the molecular mechanism underlying its involvement in tumour progression remains elusive. Among the different subunits of the complex, *ARID1A* has been identified as one of the most frequently mutated genes in several human malignancies, such as gynaecological and intestinal tumours. Therefore, the main purpose of this Master's thesis is to get a further insight into the transcriptional alterations resulted from *ARID1A*-deficiency in different cellular contexts.

In order to achieve this goal, we have generated stably-transduced cell lines for a doxycycline-inducible vector that directs the expression of different shRNAs targeting *ARID1A* in different human cancer cell lines. After the verification of an effective *ARID1A* knock-down, RNA-Seq experiments revealed both shared and tissue specific molecular pathways altered in the different cell lines. Among them, it should be highlighted an upregulation of genes belonging to proliferative pathways, as well as a downregulation of genes involved in apoptosis, which suggests an augmentation in their oncogenic capacities. What is more, there was also an upregulation of genes involved in cell migration, which might imply a potential increase in their metastatic capacities. Finally, gene set enrichment analyses showed a significant upregulation of genes related to the immune response. These results might help to clarify the molecular pathways underlying the role of *ARID1A* alteration in tumour progression and they could also suggest new therapeutic opportunities for SWI/SNF-deficient tumours.

## **2. Introduction**

### **2.1.Cancer genomics**

Tumour cells arise from normal cells that undergo a neoplastic transformation in part due to the accumulation of somatic mutations. The majority of these mutations are called passengers because they are accumulated due to the intrinsic genomic instability typical of tumour cells, but actually they do not confer them any selective advantage. On the contrary, the so-called driver mutations occur in a specific subset of genes, referred as cancer genes, which alteration is likely to contribute to cancer development because they provide a competitive advantage in terms of proliferation, migration or adaptation.

In the last years, there has been an increase in the number of next-generation sequencing (NGS) methods available, what has contributed to complete our understanding of the genome, the transcriptome and the epigenome of any organism or cell<sup>1</sup>. Furthermore, the gradually decrease of its cost has contributed to its current widespread use, as well as to the discovery of key genes involved in different human diseases, including cancer<sup>1,2</sup>.

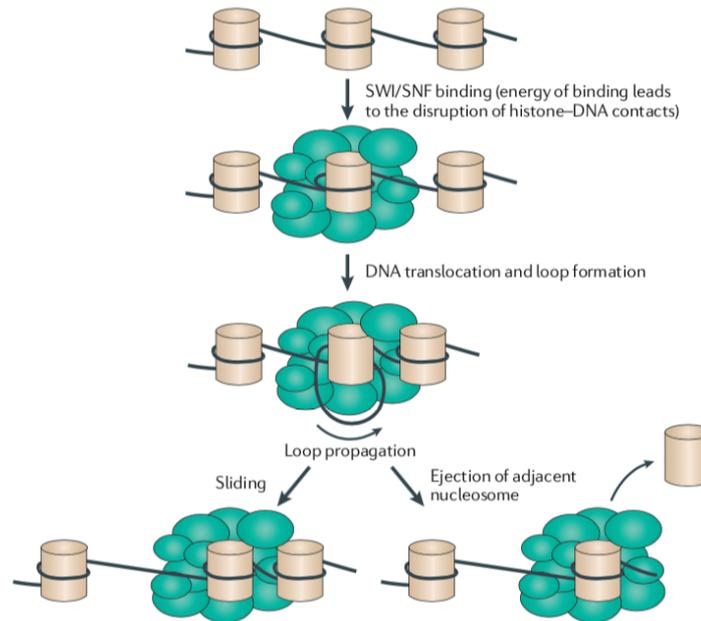
The quantitative nature of NGS techniques has opened a whole plethora of new applications besides the identification of the sequence of specific regions. For example, transcriptome sequencing (RNA-Seq) is a powerful tool used to study gene expression. Similarly, the identification of target DNA regions for specific molecular modifications (DNase treatment, chromatin immunoprecipitation, transposon mediated tagmentation, and so on) can provide biological information of relevance about the function and structure of specific DNA sequences.

Moreover, the information provided by different NGS techniques is complementary. For instance, Assay for Transposase Accessible Chromatin sequencing (ATAC-Seq), which is a technique used to assess opened chromatin regions, is likely to correlate with RNA-Seq results assuming that transcriptionally active genes are located on open regions of the DNA. Additionally, it can be combined with Chromatin Immunoprecipitation Sequencing (ChIP-Seq) to identify which regulatory factor is responsible for the observed change in the structure of DNA.

The next step that should be addressed in the near future is the implementation of NGS into the clinical practice. It is expected that the application of both NGS techniques and the information provided by them might result in a general improvement of cancer patients' outcome, as it would allow the choice of the most appropriated targeted therapy according to the molecular characteristics of tumour cells. What is more, it may also predict resistance to treatments and reveal the presence of neoantigens (antigens only expressed in tumour cells), which are ideal candidates to design targeted therapies<sup>3</sup>.

## 2.2. Chromatin remodelling complexes

Chromatin remodelling complexes use the energy provided by ATP hydrolysis to disrupt nucleosome-DNA contacts, move nucleosomes along DNA and catalyse the ejection, the insertion or the exchange of nucleosomes<sup>4,5</sup> (figure 1). Consequently, they modify the accessibility of specific regions of DNA to the enzymatic transcriptional machinery, what could potentially lead to changes in gene expression.



**Figure 1. Mechanism of remodelling of SWI/SNF complex.** This process can be divided in several steps: SWI/SNF binding, disruption of histone-DNA contacts, creation of a loop of DNA that propagates around the nucleosome and repositioning of DNA with respect to the nucleosome (sliding). This process can also result in the ejection of an adjacent nucleosome. Imagen taken from *Wilson and Roberts* (2011)<sup>4</sup>.

According to their subunit composition and biochemical activity, chromatin remodelling complexes can be divided in four major families: SWI/SNF, INO80/SWR1, ISWI and NURD/CHD<sup>4</sup>.

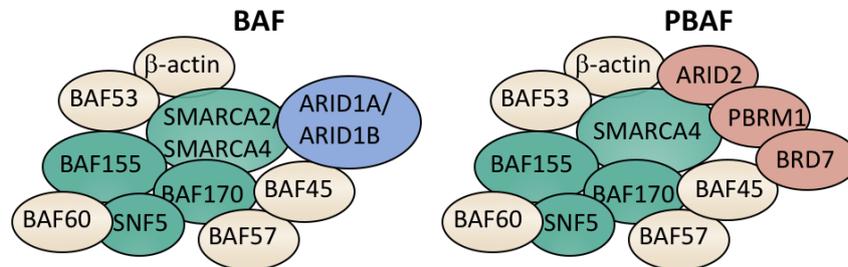
It has been previously demonstrated that SWI/SNF is the most frequently mutated chromatin remodelling complex in different types of human cancers<sup>6</sup>. The molecular pathways behind the contribution of the alteration in this complex to tumorigenesis still remain elusive<sup>4</sup>, although part of its role is supposed to be mediated through changes in the transcriptional program of the cell.

SWI/SNF is a highly related family of multisubunit complexes which is comprised of 9-12 subunits and it is evolutionary conserved from yeast to mammals<sup>4</sup>. It was firstly identified in *Saccharomyces cerevisiae* after the performance of two genetic screenings aimed at identifying mutations in genes affecting the mating-type switching (SWI) and the sucrose fermentation (SNF) pathways<sup>4,7</sup>.

Mammalian SWI/SNF complexes are composed of one of the two mutually exclusive catalytic ATPase subunits (SMARCA2 or SMARCA4), a set of highly conserved core subunits (SNF5, BAF155 and BAF170), which are present in all complexes, and variant or accessible subunits (ARID1A,

ARID1B, ARID2, PBRM1 and BRD7), that are thought to contribute to the targeting, the assembly and the regulation of lineage-specific functions of the complexes<sup>4</sup>.

In turn, two major subclasses of SWI/SNF chromatin remodelling complexes have been described in mammals relying on the variant subunits included: the BRG1-associated factor (BAF) complex (also known as SWI/SNF-A) and the polybrome-BRG1 associated factor (PBAF) complex (usually referred as SWI/SNF-B) (figure 2).



**Figure 2. Mammalian SWI/SNF complexes.** SWI/SNF chromatin remodelling complexes are composed of evolutionarily conserved core subunits (green) and variant subunits (yellow). ARID1A and ARID1B (blue) are unique to BAF complexes, whereas ARID2, PBRM1 and BRD7 (red) are only part of PBAF complexes. Imagen adapted from *Wilson and Roberts (2011)*<sup>4</sup>.

Some subunits that are common to BAF and PBAF complexes are encoded by genes that produce different isoforms by alternative splicing. Moreover, they belong to gene families that often display differential lineage-restricted expression, what means that some subunits are only expressed in some specific tissues<sup>8</sup>. It is therefore likely that a large number of different SWI/SNF complexes exist in mammals and contribute to regulate lineage and tissue specific gene expression<sup>4</sup>.

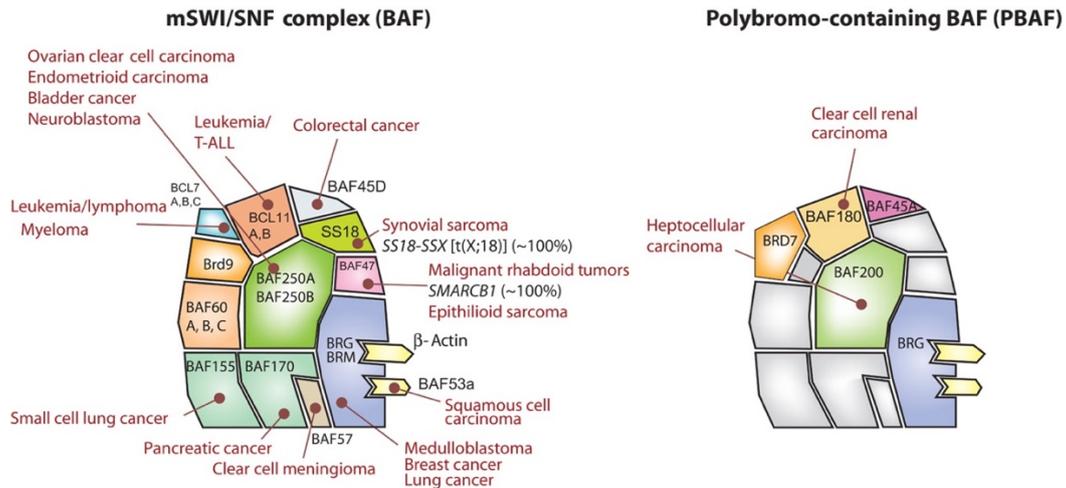
Mammalian SWI/SNF complexes allow the recruitment of both activating and repressing factors to specific DNA regions. For example, they recruit histone deacetylases (HDAC), which remove the activating acetyl marks from histone tails and contribute to gene repression<sup>4</sup>.

### 2.3. Role of chromatin remodelling complexes in cancer

Recent sequencing studies of human cancers have demonstrated that genes involved in the establishment of chromatin structure, and particularly genes encoding subunits of the SWI/SNF chromatin remodelling complex, are frequently mutated<sup>4</sup>.

For instance, *SNF5* is inactivated via biallelic mutations in 98 % of malignant rhabdoid tumours, which is a rare childhood cancer<sup>4,8</sup>. Similarly, *PBRM1* is inactivated in 41 % of clear cell renal cell carcinoma patients, what makes it the second most commonly mutated gene after *VHL*<sup>4,8,9</sup>. Likewise, *ARID1A* is mutated in 50 % of ovarian clear cell carcinomas<sup>4</sup>, which is one of the most lethal subtypes of ovarian cancer, in 35 % of endometrioid carcinomas<sup>4</sup>, in 9.4 % of colorectal and in 8.2 % of lung cancers<sup>6</sup>. *ARID2* has been found inactivated in 18.2 % of hepatocellular carcinomas<sup>10</sup> and it is also a significantly mutated gene in melanoma<sup>11</sup>.

Intriguing, although the tumour-suppressive mechanism of the complex might be general, the alteration of each subunit tends to be manifested in an apparently tissue specific manner<sup>8</sup> (figure 3), which might indicate tissue-specific functions of each subunit.



**Figure 3. Schematic overview of SWI/SNF subunit alterations identified across different types of human tumours revealed by exome sequencing.** The different subunits of BAF and PBAF complexes are represented in colours. Each subunit is attached to the main tumour type on which it has been reported as mutated. Image taken from *Kadoch et al. (2015)*<sup>8</sup>.

Genes encoding components of the SWI/SNF chromatin remodelling complex are among the most commonly and broadly mutated cancer genes<sup>12</sup>, following a mutational profile similar to well-known tumour suppressor genes, such as *TP53*. In particular, a recent study performed by *Shain and Pollack* (2013)<sup>13</sup> proved that the average frequency of SWI/SNF mutations across all tumour types (19 %) approached the one of *TP53* (26 %).

Among the different subunits of the SWI/SNF complex, *ARID1A* is the most frequently mutated gene in human cancer, and is often inactivated by frameshift or truncating mutations that are distributed along the length of the gene, resulting in lack of protein expression<sup>8,14</sup>.

As the majority of *ARID1A* mutations are loss of function, it is hypothesized to be a tumour suppressor gene. What is more, nearly all human samples tested present heterozygous mutations. This finding implies that the inactivation of a single *ARID1A* allele is enough to drive tumorigenesis and suggests that it might function as a negative-dominant tumour suppressor<sup>8</sup>.

However, inactivation of both *Arid1a* alleles in mice is necessary for the development of tumours in the colon, what probes that *Arid1a* functions as a recessive tumour suppressor in that particular animal model<sup>15</sup>. Nevertheless, in some hepatocellular carcinoma patients, *ARID1A* was overexpressed in primary tumours, but not in metastatic lesions<sup>12</sup>. Although conditional deletion of *Arid1a* in mouse liver protects against tumour initiation, it accelerates progression and metastasis in established tumours. Consequently, *Arid1a* has context-dependent oncogenic and tumour-suppressive functions in the development of some tumours<sup>12</sup>.

## 2.4. Influence of chromatin remodelling complexes in transcriptional regulation

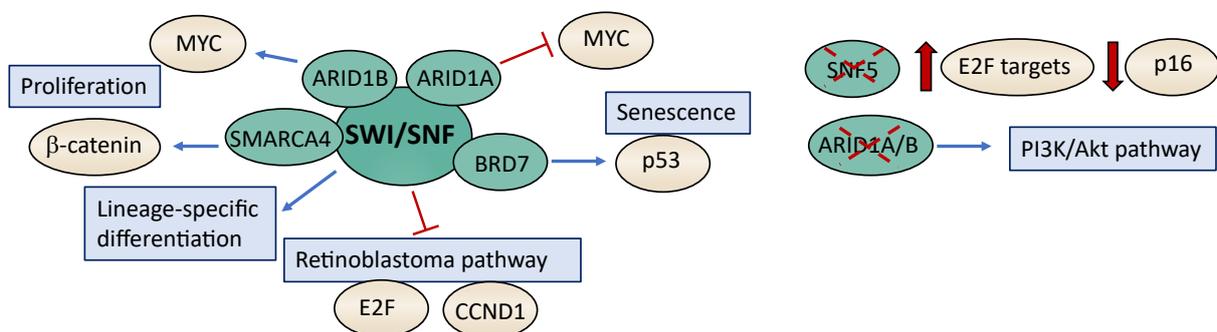
SWI/SNF complex has been postulated to be involved in the differentiation of many lineages, regarding that under physiological conditions it contributes to the processes of neurogenesis, myogenesis, adipogenesis, osteogenesis and hematopoiesis<sup>4</sup>. The interaction with tissue-specific transcription factors could explain how the SWI/SNF complex is able to promote differentiation in such different cellular contexts.

On the other hand, SWI/SNF complex is able to interact directly with genes belonging to signalling pathways already known to be dysregulated in cancer. Therefore, defects in any of its subunit might promote tumour formation through the activation of canonical cancer pathways.

Thus, it has been described that SWI/SNF complex is able to both promote or repress the expression of the oncogene *MYC*, depending on its subunit composition. During cell differentiation, SWI/SNF complexes containing ARID1A can repress directly *MYC* transcription<sup>4</sup>. Therefore, its inactivation may result in the aberrant activation of MYC programs, what could provide an explanation to tumour formation.

Similarly, *BRD7* has been found frequently deleted in a subset of breast tumours that express wild-type *TP53*, what suggest that there might be exclusivity between mutations in these genes<sup>4</sup>. Although it remains to be characterized in further detail, it has been postulated that BRD7 may act as a cofactor for TP53 in the oncogene induced senescence process<sup>4</sup>.

Furthermore, rhabdoid tumours harbouring *SNF5* inactivating mutations are characterized by the overexpression of E2F target genes and the downregulation of p16<sup>INK4A</sup><sup>4</sup>. Additionally, it has been observed that SWI/SNF complex binds RB, facilitating the repression of its target genes, *E2F* and *CCND1*. Collectively, both observations suggest that the alteration of SWI/SNF complex might result in an impaired RB tumour-suppressor pathway and consequently the control upon cell cycle might be less restrictive.



**Figure 4. Schematic overview of the influence of SWI/SNF in transcriptional regulation.** Activation pathways are represented in blue, whereas repression pathways are indicated in red. Imagen adapted from *Wilson and Roberts (2011)*<sup>4</sup>.

It has also been reported that loss of function of SWI/SF subunits (most commonly *ARID1A* or *ARID1B*) is frequently accompanied by an activating mutation in the PI3K/Akt pathway<sup>8</sup>. What is more, it has been proved that both *ARID1A* inactivation and *PI3K* activation are required to give rise to ovarian clear cell carcinoma in model animals<sup>16</sup>.

In the same way, it has been observed that  $\beta$ -catenin is also activated in those medulloblastomas on which SWI/SNF is altered. In relation to this observation, it is known that  $\beta$ -catenin interacts directly with the catalytic subunit SMARC4, regulating its ability to activate the expression of genes involved in cell proliferation<sup>8</sup>.

Finally, transcriptional regulation mediated by SWI/SNF is even more intricate taking into account the distinct biochemical forms of the complex and its interaction with transcriptional factors and co-regulators. Regarding this, it has been proposed that ARID1B and ARID2 function coordinately to modulate the expression of hundreds of genes, depending on its interaction with factors that activate (BRCA1 or MAX) or repress (TEAD4 or TCF7L2) transcription<sup>17</sup>. By contrast, ARID1A seems to have opposing effects to ARID2 in transcriptional regulation<sup>17</sup>.

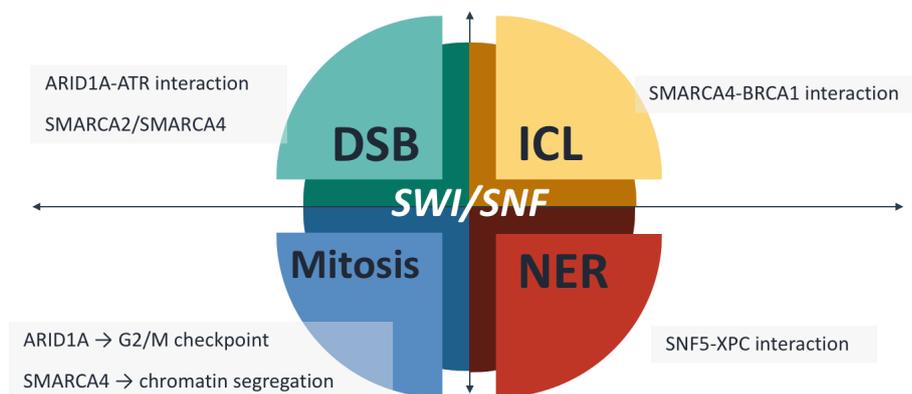
## **2.5. Role of chromatin remodelling complexes in DNA repair**

Some subunits of the SWI/SNF complex are involved in DNA repair mechanisms. As a result, the alteration of this complex is likely to promote genomic instability, what suggests a second potential mechanism of tumorigenesis, in addition to transcriptional dysregulation.

ARID1A is recruited to DNA double-strand breaks (DSB) through its interaction with ATR (a kinase that controls cellular response to DNA damage) and facilitates DNA DSB end resection<sup>18</sup>. Furthermore, ARID1A is required for G<sub>2</sub>-M DNA damage checkpoint<sup>18</sup>. Consequently, *ARID1A* loss might allow cells to progress through the cell cycle, even in case they had accumulated DNA damage.

The catalytic subunits of the SWI/SNF complex are also recruited to DSB through their bromodomains, where they increase chromatin accessibility to allow the proper assembly of those protein complexes required for DSB repair<sup>19</sup>. Furthermore, it has been proved that SMARCA4 takes part in the interstrand crosslinks (ICLs) damage repair through its direct interaction with BRCA1<sup>20</sup>.

SWI/SNF complex is also needed in other DNA repair mechanisms, such as the nucleotide excision repair (NER). Thus, SNF5 associates with XPC (a factor that recognizes damage induced by UV radiation) and is required for the recruitment of ATM to the damage site and for its activation by phosphorylation<sup>21</sup>. This evidence reinforces the role of SWI/SNF complex in maintaining genomic integrity.



**Figure 5. Schematic overview of the contribution of SWI/SNF complex to the maintenance of genomic integrity.** The figure shows the main DNA repair mechanisms on which SWI/SNF complex takes part, including the response to double-strand breaks (DSBs), DNA interstrand crosslinks (ICLs) damage and nucleotide excision repair (NER), as well as its role during mitosis to ensure a proper chromosome segregation.

Lastly, SMARCA4 is involved in the decatenation of newly replicated sister chromatids and exerts a key role for appropriate chromosome segregation after mitosis. Consequently, *SMARCA4* loss gives rise to spontaneous anaphase bridges<sup>22</sup>, what may result in an inequitable distribution of chromosomes, providing an additional source of genomic instability.

## 2.6. Mutual exclusivity between SWI/SNF subunits and other genes

There is increasing scientific evidence proving that mutations in genes encoding components of the SWI/SNF complex tend to be mutually exclusive. For instance, in colorectal cancer *ARID1A* is usually mutated in samples where neither *SMARCA4*, *ARID2*, *SMARCA2* nor *ARID1B* are mutated<sup>5</sup>. These experimental observations could be explained either by a synthetic lethality relationship between different subunits or by a loss of selective pressure to inactivate SWI/SNF complex after the first subunit mutation, what would imply that these subunits may display redundant functions.

In agreement with the first hypothesis, it has been postulated that *ARID1A* and *ARID1B* present a synthetic lethal relationship in the colorectal cancer cell line HCT116, after realizing that *ARID1A*<sup>-/-</sup> cells had a prominent decrease in cell proliferation when downregulating the expression of *ARID1B*<sup>23</sup>. Moreover, non-small-cell lung carcinoma cell lines lacking the expression of *SMARCA4* formed less colonies and were more prone to cellular senescence after the downregulation of *SMARCA2*<sup>24</sup>, what suggests that mutations in both catalytic subunits are lethal.

Similarly, there might be exclusivity between mutations in different SWI/SNF subunits and genes belonging to pathways involved in cancer development. For instance, it has been demonstrated that SWI/SNF and *TP53* mutations do not usually concur in the same colorectal tumour neither in the same ovarian clear-cell carcinoma<sup>25</sup>. Likewise, mutations in genes encoding SWI/SNF subunits and *PTEN* are mutually exclusive in the same tumour types<sup>6</sup>.

## 2.7. Therapeutic exploitation of SWI/SNF mutations

Considering that almost 20 % of all human cancers contain defects in chromatin remodelling complexes<sup>4,26,27</sup>, any new knowledge about the molecular mechanism behind their involvement on cancer development, as well as in how to exploit these deficiencies for cancer treatment, would benefit to thousands of cancer patients worldwide.

Consequently, it might be worthy to exploit the synthetic lethality relationship observed between some subunits of the complex in order to design specific therapeutic approaches, as it is the case of *SMARCA2* inhibition in *SMARCA4*-deficient tumours<sup>24</sup>.

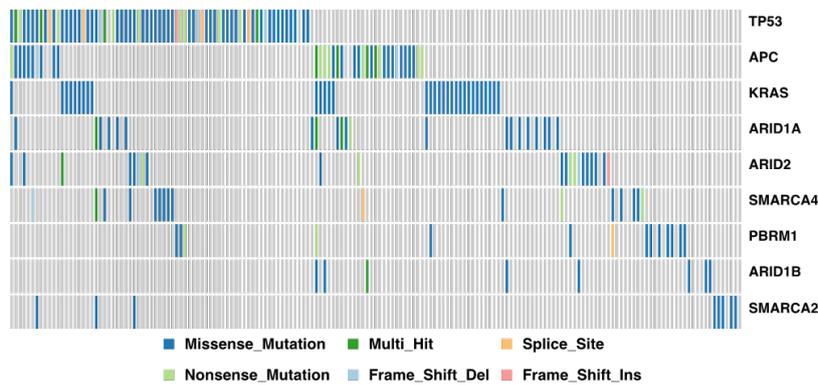
Another approach that could be further investigated, considering the involvement of SWI/SNF complex in DNA repair, could be the treatment with DNA-damaging agents (such as, etoposide or cisplatin) or with inhibitors targeting key enzymes of the DNA repair mechanisms (like ATR or PARP1). In relation to this approach, *ARID1A*-deficient tumours seem to be much more sensitive to DSB-inducing drugs, such as PARP inhibitors<sup>18</sup>. Furthermore, it has been proved that the combined use of BRD4 and PARP inhibitors is lethal in different cancer cell lines<sup>28</sup>.

Finally, it has been recently reported that the murine melanoma B16F10 cell line increase its sensitivity to T cell-mediated cytotoxicity after mutating different subunits of the BAF complex (in particular, *Arid2*, *Pbrm1* and *Brd7*)<sup>29</sup>. These results seem to indicate that tumours harbouring mutations in the SWI/SNF complex might respond better to immunotherapy. In agreement with this hypothesis, mice bearing tumours induced with *Pbrm1*-deficient B16F10 cells were more strongly infiltrated by cytotoxic T cells, developed smaller tumours and had an improved survival in comparison with control cells<sup>29</sup>. Additionally, whole exome sequencing of metastatic clear cell renal carcinomas from human patients has verified that inactivation of *PBRMI* is associated with better clinical outcome from immune checkpoint therapy<sup>30</sup>.

## 2.8. Previous data obtained in our laboratory

During the last years, in our laboratory we have performed targeted sequencing of 250 genes involved in the regulation of chromatin structure, over a cohort of more than 700 human cancer samples from different tumour types. The information provided by this analysis has allowed the identification of the most frequently mutated SWI/SNF components, which are the accessory (*ARID1A*, *ARID1B*, *ARID2* and *PBRMI*) and catalytic subunits (*SMARCA2* and *SMARCA4*) of the complex.

Additionally, our data showed exclusivity between the alteration in different SWI/SNF subunits and between alterations in the SWI/SNF complex and in well-known cancer genes, like *TP53*, *KRAS* or *APC* (figure 6).



**Figure 6. Mutual exclusivity behaviour of SWI/SNF mutations.** The representation of the main mutations observed in the tumour samples analysed in our laboratory proved mutual exclusivity between SWI/SNF subunits, and between this complex and other cancer genes. Each vertical rectangle represents an independent patient and each row a gene. Rectangles have been coloured according to the type of mutation reported<sup>31</sup>.

Interestingly, we have also noticed that *ARID2* is mutated in nearly 15 % of lung malignancies, and that its inactivation correlates with worse prognosis, what supports the potential role of *ARID2* as a *bona fide* tumour suppressor gene and as a prognostic biomarker in lung cancer<sup>32</sup>.

In order to get further insight into the molecular mechanism by which *ARID2* alteration gives rise to lung cancer development, we generated A549 cells where *ARID2* expression could be modulated in an inducible manner. The performance of RNA-Seq experiments on *ARID2*-deficient cells revealed a downregulation of genes involved in cellular adhesion (*NPNT*, *CNTNAP2*, *FAT3*, *FNI* or *VCAN*), what might explain the reported increase in migration and invasion capacities. Additionally, we observed downregulation of tumour suppressor genes (*RPS6K2*, *TNFSF10*, *ISM1* or *LDLRAD4*), together with upregulation of protumoral and antiapoptotic genes (*HOXB1*, *BCL2A1* or *RCVRN*). This combined transcriptional program alteration implies an increase in the oncogenic capacities of *ARID2*-deficient cells, what could provide an explanation to the poor clinical outcome observed in mutated patients<sup>32</sup>.

### **3. Working hypothesis and objectives**

Regarding that collectively subunits belonging to the SWI/SNF chromatin remodelling complex have been described to be altered in almost one out of five human tumours, any new knowledge about the molecular mechanisms associated with these defects might be exploited for the design of new antitumoral therapies that could potentially benefit to thousands of cancer patients worldwide.

Therefore, the general aim of this Master's thesis is the identification of the molecular pathways that are altered as a result of inactivating different subunits of the SWI/SNF complex in distinct cellular contexts.

In order to achieve this goal, we have proposed the following concrete objectives:

1. Generation of stably-transduced cell lines from different tumour types that inducibly direct the expression of shRNAs targeting different subunits of the SWI/SNF complex.
2. Identification by RNA-Seq of both cell-independent and cell-specific molecular pathways altered after the knock-down of different SWI/SNF subunits.
3. Validation of the transcriptional alterations by qRT-PCR.

## 4. Materials and methods

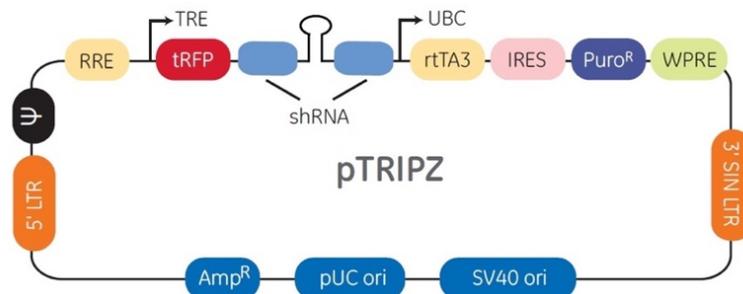
### 4.1. Cell lines and culture conditions

Human cancer cell lines SK-OV-3 (ATCC<sup>®</sup> HTB-77<sup>™</sup>), Caco-2 (ATCC<sup>®</sup> HTB-37<sup>™</sup>), A549 (ATCC<sup>®</sup> CRM-CCL-185<sup>™</sup>) and HEK 293T/17 (ATCC<sup>®</sup> CRL-11268<sup>™</sup>) were purchased from the American Type Culture Collection.

SK-OV-3, A549 and HEK 293T/17 cell lines were maintained in DMEM (Sigma, #RNBG4725) supplemented with 10 % of fetal bovine serum (FBS) (HyClone, # SH3007103), 30 µg/ml of gentamycin and 2 µg/ml of ciprofloxacin at 37 °C in a humidified atmosphere containing 5 % of CO<sub>2</sub>. Caco-2 cell line was maintained in the same culture conditions, but its growth medium was additionally supplemented up to 20 % of FBS, according to ATCC's culture specifications.

### 4.2. Generation of stably-transduced cell lines

Stably-transduced cell lines were generated using doxycycline-inducible pTRIPZ constructs (Dharmacon) targeting different subunits of the SWI/SNF complex. In particular, constructs V3\_343789, V3\_343791 and V3\_41001 were used for *ARID1A* knock-down, whereas constructs V3\_306688, V3\_306689 and V3\_306691 were selected to downregulate *ARID1B*. The empty vector (#RHS4750) was used to generate a control to which refer the results obtained in each cell line.



**Figure 7. Detailed vector map of pTRIPZ.** Schematic representation of the elements of the pTRIPZ lentiviral vector. This vector contains the sequence of an shRNA that interferes with the expression of a target gene. shRNA expression is controlled by an inducible promoter (TRE), activated by the addition of doxycycline. Additionally, it contains a reporter gene, which is a red fluorescent protein (turboRFP, represented as tRFP) that is transcribed in the same transcription unit as the shRNA, and a eukaryotic selection marker, which is a gene that confers resistance to the antibiotic puromycin. <http://dharmacon.horizondiscovery.com/uploadedFiles/Resources/ptripz-inducible-lentiviral-manual.pdf>

Lentiviruses harbouring the sequence of shRNAs were generated in the HEK 293T clone 17 (293T/17) packaging cell line. 293T/17 cells were seeded at 20 % of confluency and transfected with 2 µg of a given pTRIPZ construct, in addition to 1.5 µg of psPAX2 and 0.5 µg of pMD2.G plasmids, needed for a correct packaging of the lentiviruses. Polyethylenimine (PEI) (1 mg/ml, Sigma-Aldrich,

#4087191) was used as the transfection reagent in a 4:1 proportion (4  $\mu$ l of PEI/ 1  $\mu$ g of DNA). The next day the supernatant containing recombinant lentiviral particles was collected, filtered through a 0.45  $\mu$ m pore size filter (Millipore) and mixed in proportion 1:1 with the corresponding medium used to maintain cells that were going to be infected, adding polybrene up to a final concentration of 8  $\mu$ g/ml. In some cases, an additional centrifugation step (134 rcf, for 45-60 minutes) was carried out to force the process of infection. Additionally, in order to achieve a higher infection titre, the same procedure was repeated the following day.

Transduced cells were selected adding puromycin to the medium, reaching a final concentration that ranged from 0.5 to 5  $\mu$ g/ml, depending on the sensitivity of each cell line (0.5  $\mu$ g/ml for SK-OV-3 cells, 2  $\mu$ g/ml for A549 cells and 5  $\mu$ g/ml for Caco-2 cells). The minimum amount of antibiotic required to efficiently kill all non-transduced cells had been previously assessed in each cell line. Puromycin was refreshed daily during the selection process, which lasted at least 8 days. Additionally, a negative control, which consisted of non-infected cells treated with puromycin, was parallelly used to determine the end of the selection process. Once the stable cells had been successfully selected, puromycin dose was reduced to a maintenance dose, which ranged from one half to one tenth of the selection dose.

#### **4.3. Induction of shRNA and turboRFP expression**

With the aim of reaching an effective knock-down of *ARID1A* or *ARID1B*, the expression of the corresponding shRNAs was induced adding doxycycline to the medium (final concentration 1  $\mu$ g/ml) for at least 5 days. Considering that the half-life of this inductor in culture is 24 hours, it was refreshed daily in order to avoid fluctuations in the level of expression of the shRNAs. The expression of turboRFP, which could be detected between 8-16 hours after induction, was followed in a fluorescence microscope (Nikon Eclipse TS100) and quantified by flow cytometry (MACSQuant® VYB).

#### **4.4. FACS sorting of stably-transduced cell lines**

In order to enrich the proportion of transduced cells, a cell sorting step was performed using a FACS-Aria II cell sorter (Becton Dickinson) and the PE channel. Once the cells transduced with shRNAs or empty vectors had been selected, the expression of turboRFP was induced adding 1  $\mu$ g/ml of doxycycline for at least 16 hours. Cells were harvested, washed twice in phosphate-buffered saline (PBS), counted and re-suspended in 0.2  $\mu$ M filtered sorting buffer containing 250 ml of D-PBS ( $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  free), 3.75 ml of 1 M HEPES stock solution (final concentration 15 mM), 2.5 g of BSA in 250 ml (final concentration 1%), 2.5 ml of Penicillin/ Streptomycin (100 U/ml of each) and 1 ml of 0.5 M stock solution of EDTA (2 mM). Finally, cell suspensions of  $5 \times 10^6$  cells/mL were prepared and filtered using a nylon mesh with a pore size of 70  $\mu$ m.

The top 5 % of the cells expressing the highest levels of turboRFP were purified using the sorter. For proper cell recovery after the sorting process, the cells were collected in polypropylene tubes containing DMEM supplemented with 50 % of FBS to prevent dry out and cell death. Finally, the surviving cells were seeded and expanded in DMEM complete growth medium.

#### 4.5. RNA isolation and qRT-PCR

Total RNA was isolated and purified using Extract Me Total RNA Kit (Blirt) according to the manufacturer's instructions. RNA concentrations were measured in an ND-2000 Nanodrop and their spectres were checked to verify the absence of contaminating chemical compounds. To accomplish the reverse transcription reaction, Takara PrimeScript cDNA Synthesis kit (Takara Bio, #6110A) was used following the manufacturer's instructions. mRNA expression of the corresponding genes was measured by qRT-PCR using Luminaris Color HiGreen qPCR Master Mix (Thermo Fisher Scientific, #K0391) with the StepOnePlus™ Real-Time PCR system (Applied Biosystems).  $\beta$ -actin was used as housekeeping gene and the  $\Delta\Delta C_t$  method<sup>33</sup> was used for quantification and comparison. In order to assess significant gene expression differences between cells transduced with shRNAs or empty vectors, a one-tailed unpaired t-Test was performed assuming equal variances and a 95 % confidence interval. The complete list of the primers used for the qRT-PCR reactions can be found in table 1.

Primers used to validate transcriptional alterations	
HsRT-ACTB-Fwd	CCCAGCACAAATGAAGATCAA
HsRT-ACTB-Rev	CGATCCACACGGAGTACTTG
HsRT-ARID1A-Fwd	CCAGTAAGGGAGGGCAAGAA
HsRT-ARID1A-Rev	AGAGCTCCTTCTGTCCCAT
HsRT-ARID1B-Fwd	GCGCAACAAAGGAGTCACC
HsRT-ARID1B-Rev	GCCCATGCCATACAACACTGAG
HsRT-ATR-Fwd	GATCAGCTTACTGCCGTTCC
HsRT-ATR-Rev	GTTCCCATAGGACCCATTCC
HsRT-BCL2-Fwd	CAATCACGCGGAACACTTGA
HsRT-BCL2-Rev	AGATTGATGGGATCGTTGCC
HsRT-COL4A2-Fwd	GAAGTTTGATGTGCCGTGTG
HsRT-COL4A2-Rev	TCCTTTACGTCCCTGCAGTC
HsRT-PIK3CD-Fwd	GCTGGAGTTCGACATCAACA
HsRT-PIK3CD-Rev	CGGTCTTAAGCTGGTCCTTG
HsRT-SMARCA2-Fwd	GCGACCTGGAGAAGGATGTC
HsRT-SMARCA2-Rev	GCCGGGCACTCTTAAACACT
HsRT-THBS1-Fwd	TGCCATCCGCACTAACTACA
HsRT-THBS1-Rev	ATCAACAGTCCATTCCCTCGTT

Table 1. 5' → 3' sequence of primers designed to validate transcriptional alterations.

#### **4.6. Western Blot**

Cells were washed twice in PBS, lysed in a buffer containing 50 mM Tris-HCl pH 8.0 and 2 % SDS and further disrupted through a disposable 20G needle (Henke Sass Wolf, #4710009025). Protein concentrations were determined by Qubit™ Protein Assay Kit (Life Technologies, #Q33212). 50 µg of total protein lysate was separated by SDS-PAGE in 8 % polyacrylamide gels and transferred to nitrocellulose membranes using Trans-Blot® Turbo™ Transfer System (Bio-rad, #1704158).

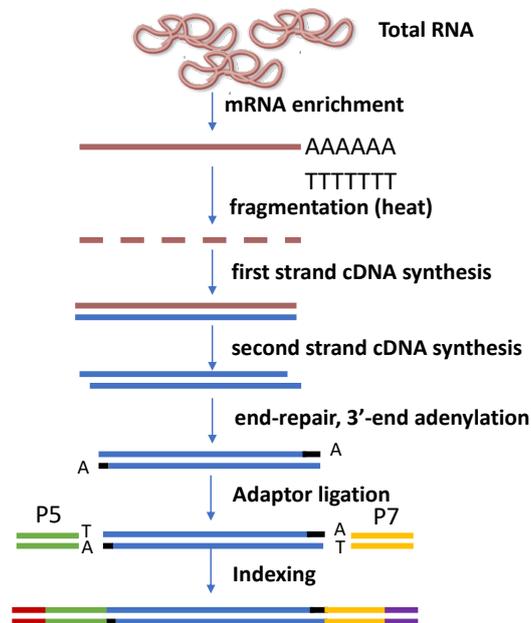
Subsequently, membranes were washed with TBS-T buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0,1% Tween 20) and blocked for 1 hour at room temperature, using 5 % of non-fat milk (Skim Milk Powder, Sigma Aldrich, #70166) solution in TBS-T as blocking agent. Membranes were then incubated with primary antibodies anti-ARID1A (Abcam, ab176395) and anti-Actin (I-19, SC-1616, Santa Cruz), diluted 1:250 in TBS-T containing 1 % of BSA, at 4 °C overnight. Then, membranes were carefully washed in TBS-T three times. Primary antibodies were detected by incubating the membranes with goat anti-rabbit or donkey anti-goat secondary antibodies (LI-COR Biotechnology) conjugated to IRDye 800CW (#926-32211) or IRDye 680RD (#926-68074) respectively at 1:5,000 dilutions and incubated for 1 hour at room temperature. Finally, antibody signals were visualized using Odyssey® Clx Imaging System (LI-COR Biotechnology).

#### **4.7. cDNA NGS libraries**

For RNA-Seq library preparation, 10 µg of total RNA was treated with DNase (TURBO DNA-free™ Kit, # AM1907) to remove any gDNA contamination. Additionally, its quality was measured using RNA ScreenTape® (4200 TapeStation Instrument - Agilent Genomics). Then, an enrichment step was performed using the NEBNext® Poly(A) mRNA Magnetic Isolation Module (#E7490S) to increase the proportion of mRNA. mRNA was randomly fragmented by heating the samples 15 minutes at 94 °C. The first strand of cDNA was generated using the PrimeScript First Strand cDNA Synthesis kit (Takara, #6110A), incubating the samples for 15 minutes at 37 °C, followed by a 5 second step at 85 °C to inactivate the polymerase. The synthesis of second strand of cDNA was initiated by the addition of DNA polymerase I (Thermo Fisher Scientific, #EP0042) and the remaining molecules of RNA were degraded adding RNase HI (Thermo Fisher Scientific, # EN0202) to the previous mix, which was incubated for 2 hours at 15 °C. T4 DNA polymerase (Thermo Fisher Scientific, # EP0062) was added to each sample to extend and complete the synthesis of the second strand and an incubation step of 5 minutes at 15 °C was performed before adding 5 µl of EDTA 0.5 M pH 8.0 to each sample.

The double-stranded cDNA fragments obtained were purified using Agencourt AMPure XP (Beckman Coulter, #082A63881). Sequencing libraries were prepared through a series of enzymatic steps, which included end-repair (Fast DNA End Repair Kit, Thermo Fisher Scientific #K0771) and 3'-end adenylation (Klenow Fraggmnet Exo, Thermo Fisher Scientific, #EP0422) of the resultant

fragments, PE adaptor construction through the hybridization of phosphorylated complementary synthetic oligonucleotides, PE adaptor ligation (T4 DNA Ligase, Thermo Fisher Scientific, #EP0062) and PCR indexing amplification (Phusion high fidelity DNA polymerase, Thermo Fisher Scientific, #F530L) (figure 8).



**Figure 8. Schematic overview of RNA-Seq library preparation.** Library preparation method begins with an initial isolation step of mRNA from total RNA samples, which is then submitted to random fragmentation by heat and cDNA synthesis. Once double stranded DNA fragments are generated, a series of enzymatic steps are performed, including DNA fragments blunting and Illumina adaptor ligation adding specific sequences to identify the reads from each sample.

Size distribution of the RNA-Seq libraries obtained was analysed using DNA 1000 kit or D1000 ScreenTape Assay (4200 TapeStation Instrument - Agilent Genomics), depending on their concentration, which was quantified using the Qubit® dsDNA BR Assay Kit (Life Technologies #Q32851).

Before sequencing the RNA-Seq libraries, a qPCR was performed to determine the quality of the preparations. A fragment that contained the complete length of the PE Illumina adaptors was used as a control to determine the efficiency of the adaptor ligation in the preparations. Finally, equimolecular amounts of each sample were mixed to reach a concentration of 10 µg/ml in a final volume that ranged between 10 and 20 µl. The sequencing of the resultant pool was carried out in a High-Seq® platform (Illumina). A minimum of 70 million 75 paired-end reads was generated from each of the 3 replicates that had been independently prepared for each sample group (shEmpty or shARID1A in different cell lines).

#### 4.8. RNA-Seq analysis

Paired-end reads from RNA-Seq were aligned to the human reference genome (hg19) using Tophat2. The number of reads of each gene was quantified using HTSeq. Differentially expressed genes between *ARIDIA*-deficient cells and their respective control were identified using DESeq, requiring a minimum of 3 counts for a gene in more than two independent samples. The results obtained were filtered with the following criterion: absolute  $\log_2\text{FoldChange} > 0.5$  and adjusted p-value  $< 0.05$ . The set of genes that accomplished both conditions, which were considered to be dysregulated after *ARIDIA* knock-down, were manually reviewed and grouped in pathways using KEGG Mapper. Finally, gene set enrichment analysis (GSEA)<sup>34</sup> was performed using the normalized counts of all genes expressed in the different conditions tested, using a defined set of genes (c5.all.v6.1.symbols.gmt, Gene ontology) in order to identify significant enrichment at the molecular pathway level.

## 5. Results

### 5.1. Generation of stably-transduced cell lines

Taking advantage of the doxycycline-inducible lentiviral constructs provided by Dharmacon, we have been able to efficiently generate stably-transduced cell lines that express shRNAs targeting *ARID1A* or *ARID1B* (for further details look up table 2). Additionally, a control consisting of an empty lentiviral vector (#RHS4750) was also generated in each cell line.

Cell line	Knock-down gene	Identification code of the shRNA
SK-OV-3	ARID1A	V3_410041
		V3_343791
Caco-2	ARID1A	V3_410041
		V3_343789
		V3_343791
	ARID1B	V3_306688
		V3_306689
		V3_306691
A549	ARID1A	V3_410041
		V3_343789
		V3_343791
	ARID1B	V3_306688
		V3_306689
		V3_306691

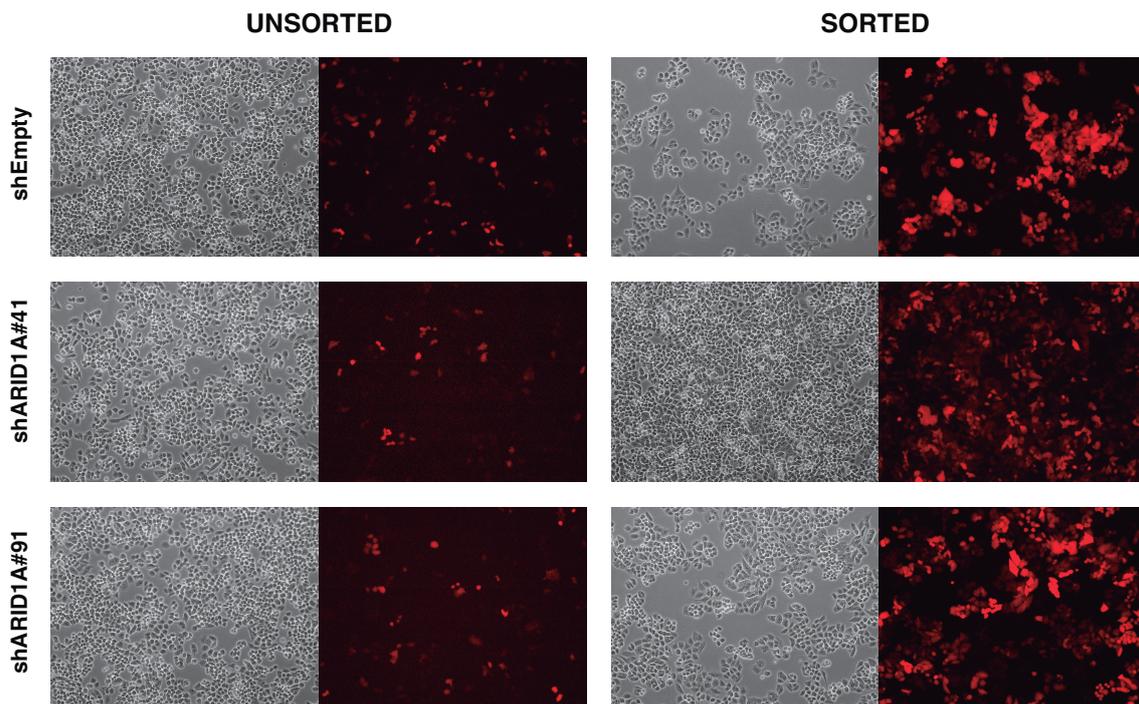
**Table 2. Set of stable cell lines generated in our laboratory.** This table reports the complete list of stably-transduced human cancer cell lines generated, the target gene and the identification code of the different shRNAs tested.

The above indicated cell lines had been chosen according to the bibliographic evidences on which the presence of SWI/SNF mutations have been demonstrated. Following this criterion, *ARID1A* has been downregulated in an ovarian adenocarcinoma (SK-OV-3), a colorectal adenocarcinoma (Caco-2) and a lung carcinoma (A549) cell lines, whereas *ARID1B* has been knocked-down in a colorectal adenocarcinoma (Caco-2) and a lung carcinoma (A549) cell lines.

It should be noticed that in every cell line, more than one shRNA for each particular gene has been tested in order to assess which one is more efficient and also to confirm the specificity of the effect. Furthermore, both *ARID1A* and *ARID1B* have been knocked-down in more than one cell line to identify tissue-independent and tissue-specific transcriptional alterations.

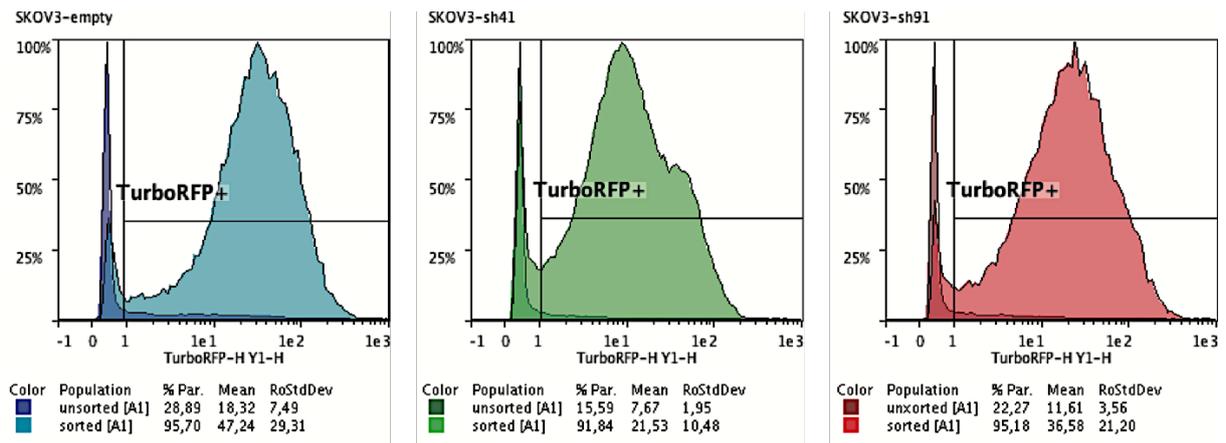
## 5.2. FACS sorting of stably-transduced cells

As a first ascertainment of the efficiency of our experimental approach, we monitored the expression of the reporter gene on each cell line by fluorescence microscopy. As it can be seen in the left panel of figure 9, we observed a highly variably number of turboRFP-expressing cells in each culture. Therefore, we decided to perform a cell sorting step in order to isolate the 5 % cell fraction that expressed the highest levels of the fluorescent marker. Cell sorting resulted in a fraction turboRFP-expressing cells higher than 90 %, as it can be estimated by looking at the images showed in the right panel of figure 9.



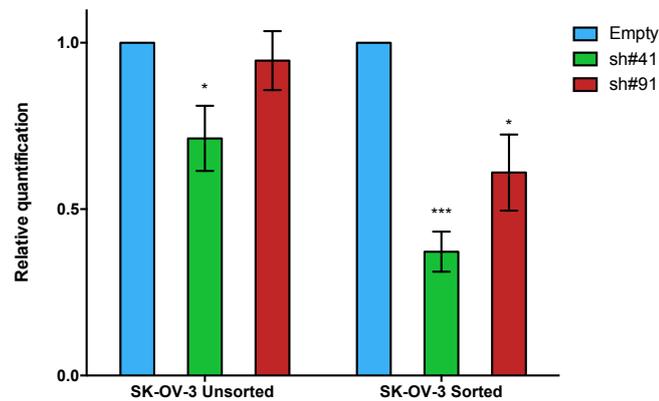
**Figure 9. Verification of the generation of stably-transduced cell lines measured through the expression of the reporter gene.** Representative microscopic images showing turboRFP expression of unsorted (left) and sorted (right) SK-OV-3 cells knock-down for different shRNAs targeting *ARID1A* (V3\_410041, shARID1A#41; V3\_343791, shARID1A#91) treated with 1  $\mu\text{g/ml}$  of doxycycline for 4 days. Fluorescence microscope images were taken with an equal exposure time of 600 ms.

The results provided by fluorescence microscopy images revealed differences at the number of turboRFP-expressing cells between unsorted and sorted cell lines, but in order to confirm those qualitative differences observed, we also quantified by flow cytometry the level of expression of turboRFP. As it can be distinguished in figure 10, on average 22.4 % of unsorted cells expressed the red fluorescent protein, in contrast with the 94.2 % of turboRFP-expressing cells after cell sorting.



**Figure 10. Histograms representing turboRFP expression in SK-OV-3 stable cell lines.** The level of expression of the reporter gene in each stable cell line has been overlapped before and after cell sorting, using different tonalities of the same colour.

Finally, we performed a qRT-PCR on both unsorted and sorted stable cell lines to verify if this approach had actually succeeded at decreasing the mRNA expression of *ARID1A*. As it is shown in figure 11, the higher expression of turboRFP on sorted cells correlated with a further knock-down of *ARID1A*.

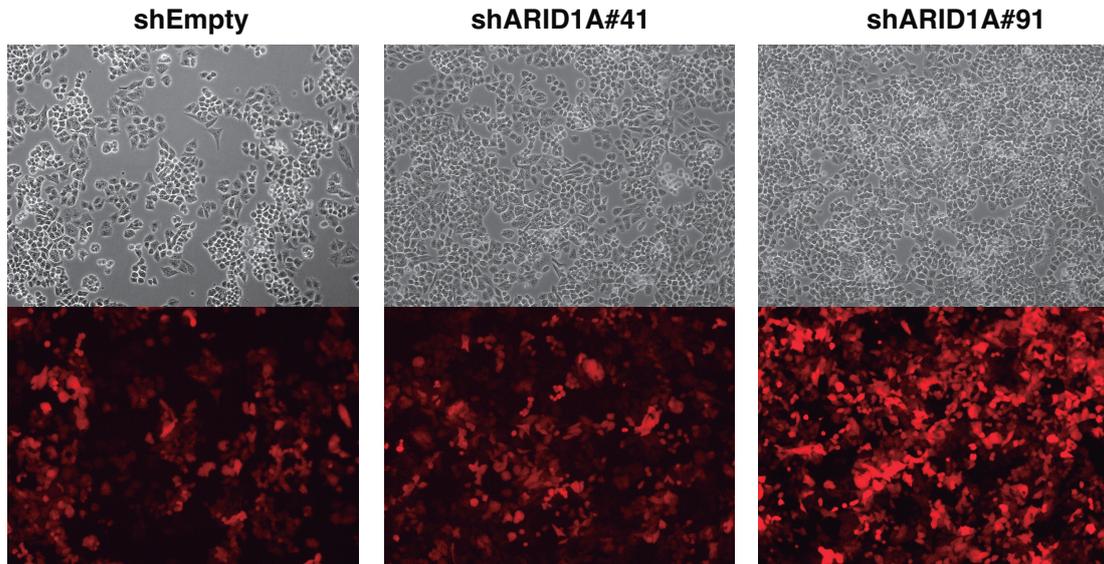


**Figure 11. *ARID1A* expression in SK-OV-3 stable cells lines knock-down for *ARID1A* before and after cell sorting.** qRT-PCR for *ARID1A* in unsorted (left) and sorted (right) SK-OV-3 cells stably-transduced with different shRNAs targeting *ARID1A* (V3\_410041, shARID1A#41; V3\_343791, shARID1A#91) treated 7 days with 1  $\mu$ g/ml of doxycycline. The relative quantification is represented as a mean  $\pm$  standard error of mean (SEM) of three independent experiments. The results of the one-tailed unpaired t-Test are represented with asterisks with the following criterion: \* p-value < 0.05, \*\* p-value < 0.01 and \*\*\* p-value < 0.001.

### 5.3. Effect of *ARID1A*-deficiency on SK-OV-3 cells

#### 5.3.1. Validation of *ARID1A* knock-down

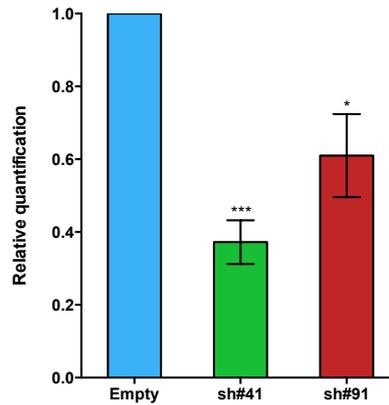
Once *ARID1A*-deficient SK-OV-3 cell lines had been selected and sorted, we confirmed the effective insertion of the pTRIPZ allele inducing the expression of turboRFP with doxycycline. As it is shown in figure 12, the majority of cells expressed the reporter gene.



**Figure 12. Verification of the generation of SK-OV-3 stably-transduced cell lines measured through the expression of the reporter gene.** Representative microscopic images showing turboRFP expression of sorted SK-OV-3 cells stably-transduced with different shRNAs targeting *ARID1A* (V3\_410041, shARID1A#41; V3\_343791, shARID1A#91), after adding 1  $\mu\text{g}/\text{ml}$  of doxycycline for 9 days. Fluorescence microscope images were taken with an equal exposure time of 600 ms.

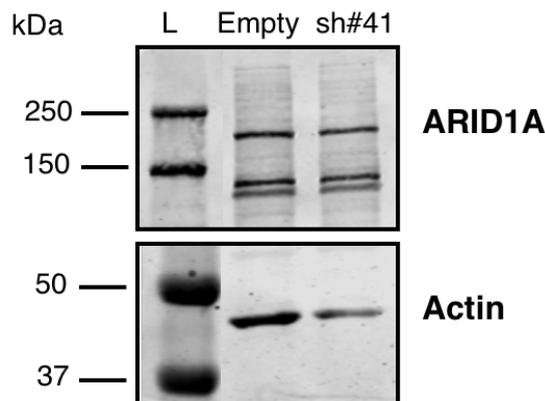
Moreover, we also performed a qRT-PCR to verify an effective knock-down of *ARID1A* at mRNA level. Thus, *ARID1A* expression was compared between cells expressing shRNAs against cells expressing the empty vector, after having treated both groups with doxycycline. Additionally, three independent biological replicates were prepared for each stable cell line to perform the statistical test for differential expression.

We observed that both shRNAs tested significantly downregulated the expression of *ARID1A* (figure 13). Considering that shARID1A#41 achieved a further decrease, we selected this stable cell line to prepare RNA-Seq libraries.



**Figure 13.** *ARID1A* expression at mRNA level in SK-OV-3 stably-transduced cells lines. qRT-PCR for *ARID1A* in sorted SK-OV-3 cells knock-down for different shRNAs targeting *ARID1A* (V3\_410041, sh#41V3\_343791, sh#91) treated with 1  $\mu\text{g/ml}$  of doxycycline for 7 days. The results of the one-tailed unpaired t-Test are represented with asterisks with the following criterion: \* p-value < 0.05, \*\* p-value < 0.01 and \*\*\* p-value < 0.001.

Finally, we performed a Western Blot analysis to prove if the repression of *ARID1A* had been achieved at the protein level. Unfortunately, as it is shown in figure 14, we failed to see clear significant differences in terms of ARID1A production. This result could probably be explained considering that *ARID1A* expression at mRNA level has just been reduced up to a 40 % of the basal expression (figure 13). Nevertheless, the high molecular weight of the protein of interest (> 200 kDa) that complicates a correct separation and transfer of the bands, together with the low specificity of the available antibodies and the variability in the total protein loaded in each lane precludes a robust conclusion. Therefore, further research is needed to definitively demonstrate ARID1A-deficiency at the protein level.

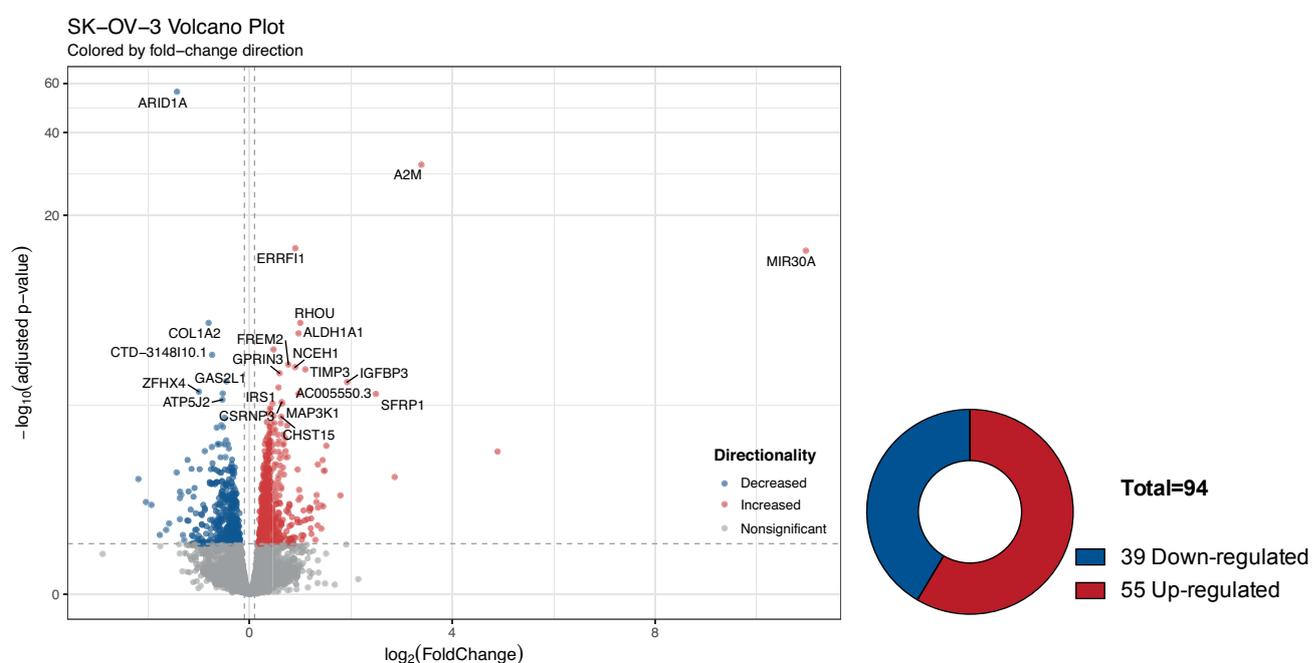


**Figure 14.** *ARID1A* expression at protein level in SK-OV-3 knock-down cells. Western blot experiment measuring ARID1A production in SK-OV-3 cells transduced with shRNA V3\_410041 targeting *ARID1A* (sh#41) or the empty vector (empty), after 17 days of treatment with 1  $\mu\text{g/ml}$  of doxycycline.  $\beta$ -actin (42 kDa) has been used as the loading control. L: protein ladder (PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa, #26619).

### 5.3.2. Transcriptional alterations after *ARID1A* knock-down

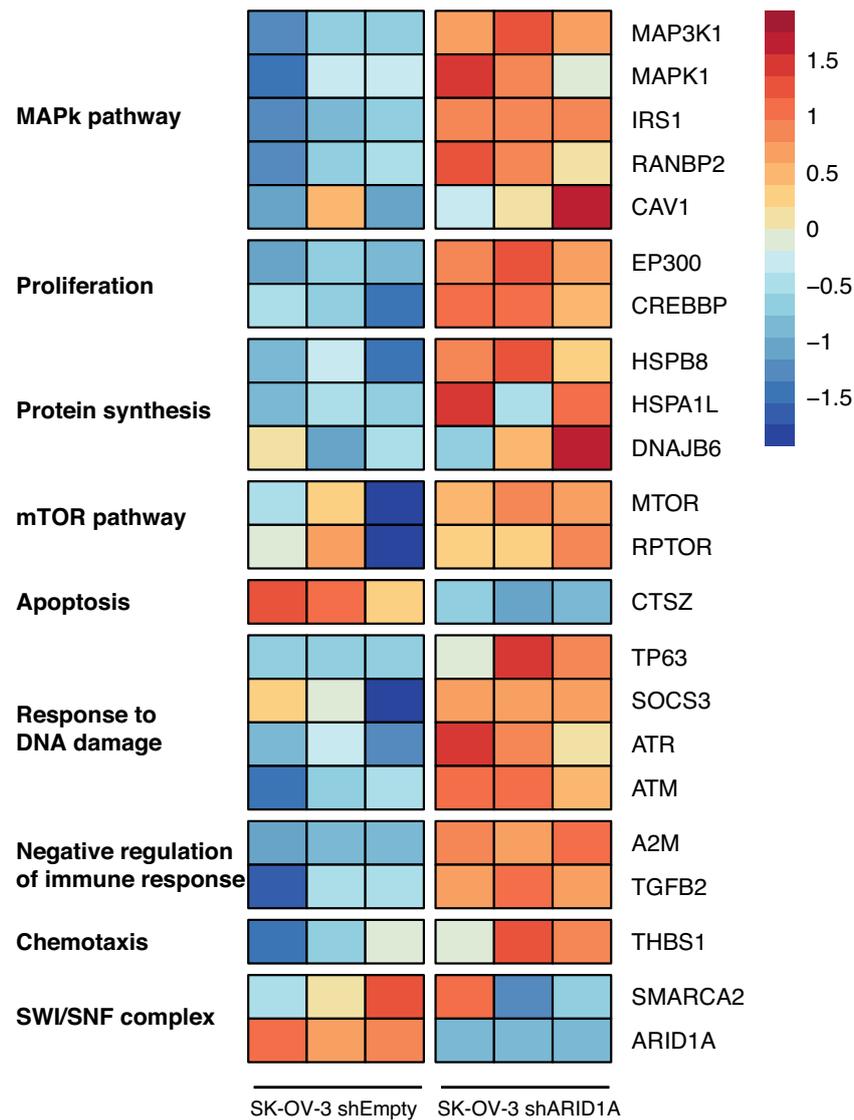
We performed RNA-Seq experiments in three *ARID1A*-deficient SK-OV-3 cell lines that had been independently treated with doxycycline for 7 days. After applying several filtering criteria (detailed in “RNA-Seq Analysis” section), we found that *ARID1A*-downregulation was accompanied with the dysregulation of 94 genes. It should be pointed out that the differential expression analysis was carried out without adjusting the data to a negative binomial distribution because there was too much dispersion between some biological replicates, which impaired the identification of some transcriptional alterations.

From the total amount of genes whose expression had significantly changed, nearly 60 % (55/94) were up-regulated and the remaining 40 % (39/94) were down-regulated (figure 15).



**Figure 15. Differentially expressed genes in *ARID1A*-deficient SK-OV-3 cells.** Volcano plot representing the  $-\log_{10}(\text{adjusted } p\text{-value})$  versus  $\log_2(\text{FoldChange})$  of genes expressed in SK-OV-3 cells using the ggplot2 package. Those genes that had been considered to be dysregulated after the downregulation of *ARID1A* are represented in red (up-regulated) or blue (downregulated), whereas those genes that have not changed significantly are represented in grey. Additionally, a doughnut chart representing the proportion of genes up- and downregulated has been graphed.

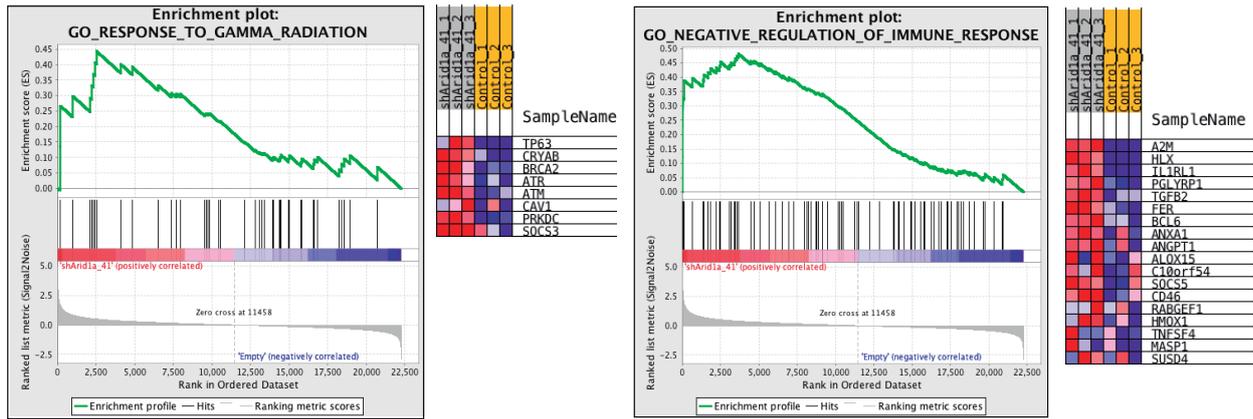
The analysis of the data provided by the RNA-Seq experiment (figure 16) revealed an upregulation of genes belonging to the MAPk (*MAP3K1*, *MAPK1*, *IRS1*, *RANBP2* and *CAVI*) and mTOR pathways (*MTOR* and *RPTOR*). Moreover, genes involved in proliferation (*EP300* and *CREBBP*) and in protein synthesis (*HSPB8*, *HSPAIL* and *DNAJB6*) were also upregulated, which suggest an enhanced proliferative phenotype in *ARID1A*-deficient cells.



**Figure 16. Transcriptional alterations of *ARID1A*-deficient SK-OV-3 cells.** Heatmap representing a selection of differentially expressed genes in *ARID1A*-deficient SK-OV-3 cells (n=3) and grouped according to their respective molecular pathway. Gene expression differences between shARID1A and shEmpty cells have been represented according to the log<sub>2</sub> of the fold change and have been indicated by colours, which range between dark red (overexpression) to dark blue (downregulation).

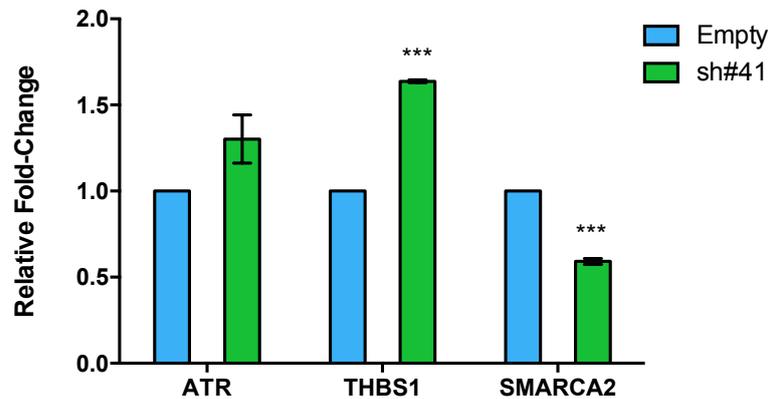
We also detected an upregulation of genes involved in DNA repair damage (*TP63*, *SOCS3*, *ATR* and *ATM*), which could be correlated with an increase in genomic instability. Additionally, some genes related to the negative regulation of immune response (*A2M* and *TGFB2*) have been upregulated in *ARID1A*-deficient cells, which might point to a higher resistance to immunotherapy.

Complementary, GSEA analysis identified significant enrichment in some interesting pathways, among which it should be highlighted the response to gamma radiation (which could be linked to the response to DNA damage) and the negative regulation of the immune response (figure 17).



**Figure 17. Molecular pathways significantly altered in *ARIDIA*-deficient SK-OV-3 cells.** GSEA analysis has shown significant enrichment of genes involved in the response to gamma radiation and in the negative regulation of the immune response. GSEA results have been represented according to the enrichment score (ES): a positive value indicates enrichment in *ARIDIA*-deficient cells (red), whereas a negative value implies an enrichment in control cells. Heatmaps of significantly enriched genes from each pathway are also represented; genes highlighted in red are overexpressed in *ARIDIA*-deficient cells, whereas those represented in blue are downregulated.

Finally, we validated by qRT-PCR the expression of some genes that have been found dysregulated in the RNA-Seq to ascertain the veracity of the results obtained by this technique. As it is shown in figure 18, the expression of *THBS1* and *SMARCA2* showed a significant alteration after *ARIDIA*-downregulation, whereas *ATR* showed a non-significant tendency, probably due to the high variability observed between the biological replicates. Therefore, further experiments are needed to confirm this change.



**Figure 18. Validation of the transcriptional alterations observed in *ARIDIA*-deficient SK-OV-3 cells.** Bar graph representing the results of the qRT-PCR validation of a selection of differentially expressed genes following *ARIDIA*-downregulation in SK-OV-3 cells. The relative fold-changes are represented as mean  $\pm$  SEM of three independent experiments. The results of the one-tailed unpaired t-Test are represented with asterisks with the following criterion: \* p-value < 0.05, \*\* p-value < 0.01 and \*\*\* p-value < 0.001.

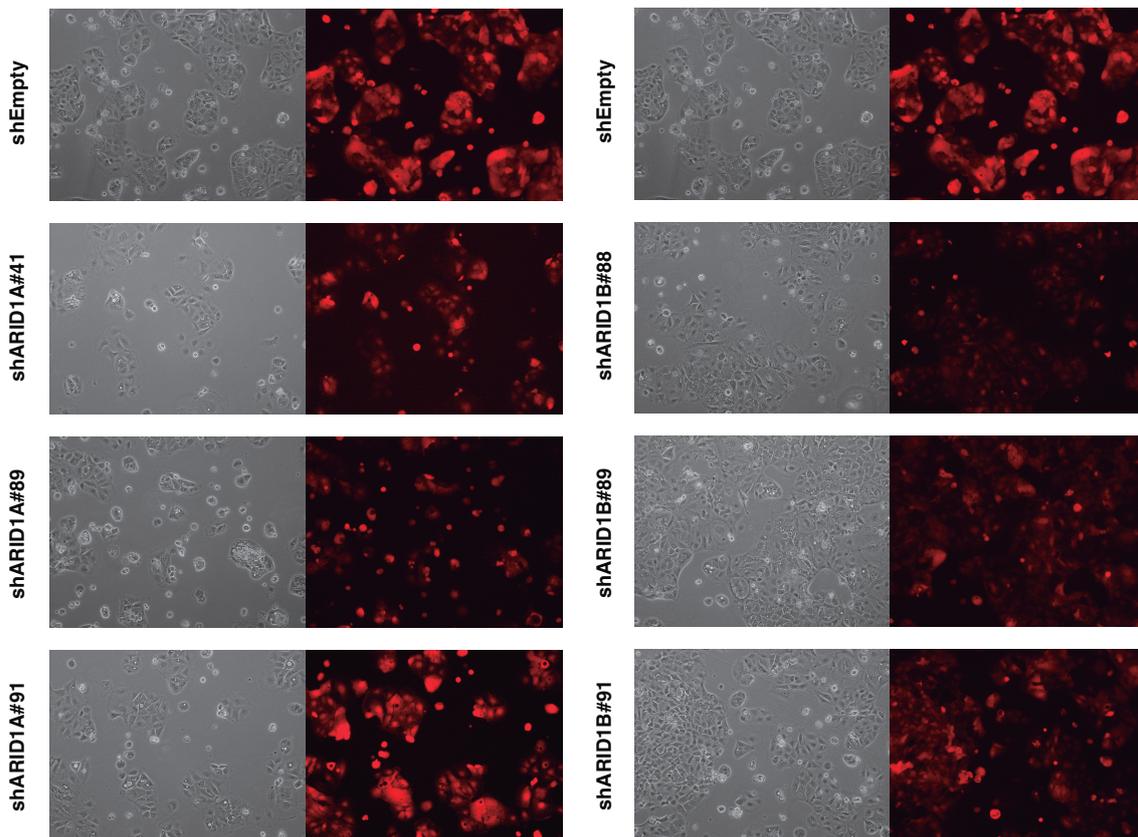
#### 5.4. Effect of *ARID1A*-deficiency in Caco-2 cells

Regarding the further downregulation of *ARID1A* that had been achieved after sorting SK-OV-3 cells, we decided to equally sort Caco-2 stable cell lines in order to increase the proportion of transduced cells.

##### 5.4.1. Validation of *ARID1A* and *ARID1B* knock-down

Similarly to SK-OV-3 stable cell lines, we followed the infection process in Caco-2 cells by checking the expression of turboRFP after inducing its expression with doxycycline.

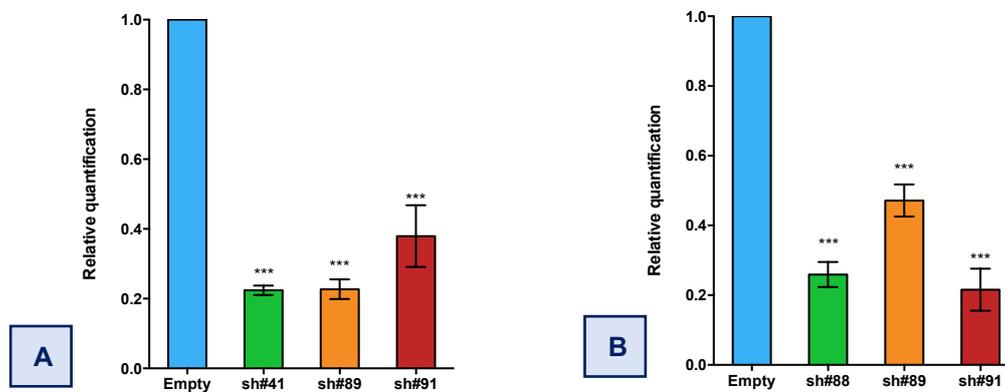
Taking into account that the images shown in figure 19 were taken with the same exposure time, it is possible to compare the level of expression of turboRFP in the different shRNAs tested. Thus, we have observed that there are some stable cell lines on which the transduction process seems to have gone better in terms of the emission of red fluorescence, like the shARID1A#91. Therefore, it could be assumed that this stable cell line is expressing higher levels of the corresponding shRNA.



**Figure 19. Verification of the generation of Caco-2 stably-transduced cell lines measured through the expression of the reporter gene.** Representative microscopic images showing turboRFP expression of sorted Caco-2 cells stably-transduced with different shRNAs targeting *ARID1A* (V3\_410041, shARID1A#41; V3\_343789, shARID1A#89, V3\_343791, shARID1A#91) and *ARID1B* (V3\_306688, shARID1B#88, V3\_306689, shARID1B#89; V3\_306691, shARID1B#91), after adding 1  $\mu\text{g/ml}$  of doxycycline for 2 days. Fluorescence microscope images were taken with an equal exposure time of 600 ms.

Even if we had seen that all the stable cell lines generated expressed the reporter gene after the addition of doxycycline, a qRT-PCR was additionally performed to verify that those cells were actually expressing the shRNAs and also to determine which one was more efficient at silencing the expression of the corresponding target gene at mRNA level.

To that end, the level of expression of both *ARID1A* and *ARID1B* in stable cell lines expressing shRNAs was compared to their respective control (cells expressing the empty vector), after treating all knock-down and control cells with doxycycline. Three independent biological replicates were prepared for each stable cell line to perform the statistical test for differential expression. As it can be seen in figure 20, the results provided by the qRT-PCR showed that all the shRNAs tested had downregulated *ARID1A* or *ARID1B* up to different levels.



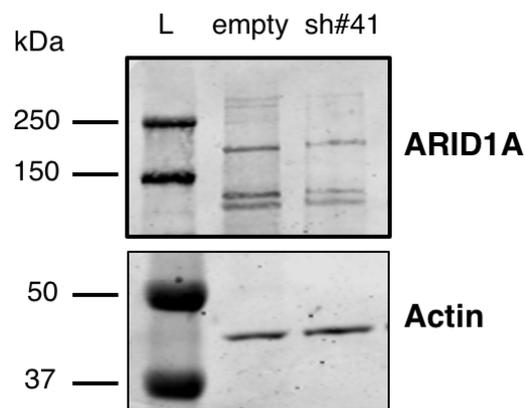
**Figure 20. *ARID1A* and *ARID1B* mRNA expression in Caco-2 stably-transduced cells lines.** A. qRT-PCR for *ARID1A* in sorted Caco-2 cells stably-transduced with different shRNAs targeting *ARID1A* (V3\_410041, sh#41; V3\_343789, sh#89, V3\_343791, sh#91) treated with 1  $\mu$ g/ml of doxycycline for 6 days. B. qRT-PCR for *ARID1B* in sorted Caco-2 cells stably-transduced with different shRNAs targeting *ARID1B* (V3\_306688, sh#88, V3\_306689, sh#89; V3\_306691, sh#91) treated with 1  $\mu$ g/ml of doxycycline for 6 days. The results of the one-tailed unpaired t-Test are represented with asterisks with the following criterion: \* p-value < 0.05, \*\* p-value < 0.01 and \*\*\* p-value < 0.001.

Curiously, although we had seen that shARID1A#91 Caco-2 cells expressed the highest level of turboRFP, when we checked the downregulation of *ARID1A* at the mRNA level, it was the least efficient in silencing its target gene. Therefore, the expression of the reporter gene does not actually correlate precisely with the downregulation achieved at the mRNA level of the target gene. It should be highlighted that the emission of red fluorescence is an indicative of the efficiency of the transduction process, but it is the specific sequence of the shRNA which determines if it is going to silence effectively its target gene.

Furthermore, the statistical analysis performed proved that the shRNA V3\_410041 (sh#41), similarly to what had happened in SK-OV-3 cells, had been the most efficient in downregulating *ARID1A*, even if it was expressed at lower levels than the other shRNAs. Therefore, we chose this stable cell line to prepare the RNA-Seq libraries.

The statistical analysis also revealed that the shRNA V3\_306691 (sh#91) had silenced *ARID1B* in a more efficient way, and accordingly it was selected to prepare the RNA-Seq libraries.

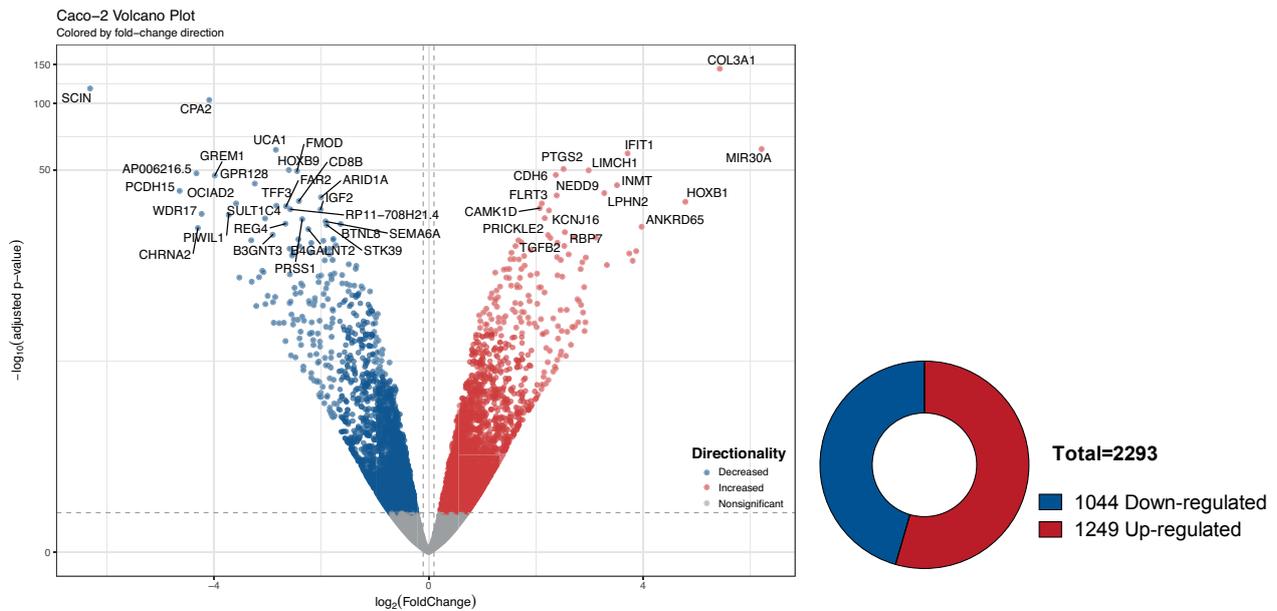
Lastly, we performed a Western Blot analysis in Caco-2 cells to confirm *ARID1A*-deficiency at the protein level. As it can be seen in figure 21, *ARID1A*-deficient Caco-2 cells showed a clear reduction in *ARID1A* protein production. Therefore, the decrease achieved at the mRNA level actually correlated with an efficient knock-down at the protein level.



**Figure 21. *ARID1A* expression at protein level in Caco-2 knocked-down cells.** Western blot experiment measuring *ARID1A* production in Caco-2 cells transduced with shRNA V3\_410041 targeting *ARID1A* (sh#41) or the empty vector (empty), after 17 days of treatment with 1  $\mu\text{g/ml}$  of doxycycline.  $\beta$ -actin (42 kDa) has been used as the loading control. L: protein ladder.

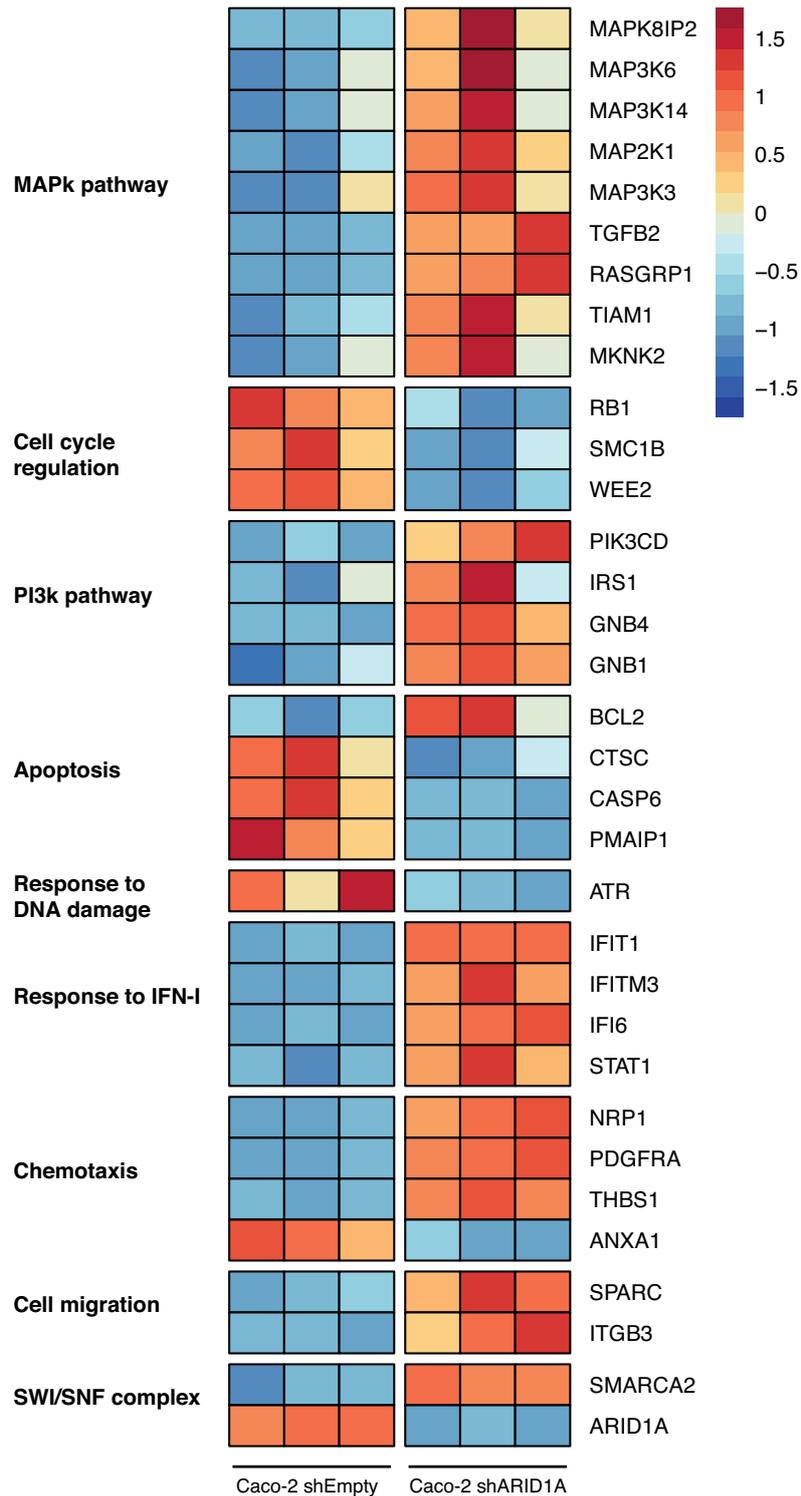
#### 5.4.2. Transcriptional alterations after *ARID1A* knock-down

As in the case of SK-OV-3 cells, we performed RNA-Seq experiments in three *ARID1A*-deficient Caco-2 cell lines. Once having applied the same filtering criterion than in *ARID1A*-deficient SK-OV-3 cells, we found a total of 2,293 genes dysregulated with an approximately equal proportion of genes upregulated (54.4 %, 1249/2293) and downregulated (45.5 %, 1044/2293) (figure 22).



**Figure 22. Differentially expressed genes in *ARID1A*-deficient Caco-2 cells.** Volcano plot representing the  $-\log_{10}(\text{adjusted } p\text{-value})$  versus  $\log_2(\text{FoldChange})$  of genes expressed in Caco-2 cells using the ggplot2 package. Those genes that had been considered to be dysregulated after the downregulation of *ARID1A* are represented in red (up-regulated) or blue (downregulated), whereas those genes that have not changed significantly are represented in grey. Additionally, a doughnut chart representing the proportion of genes up and downregulated has been graphed.

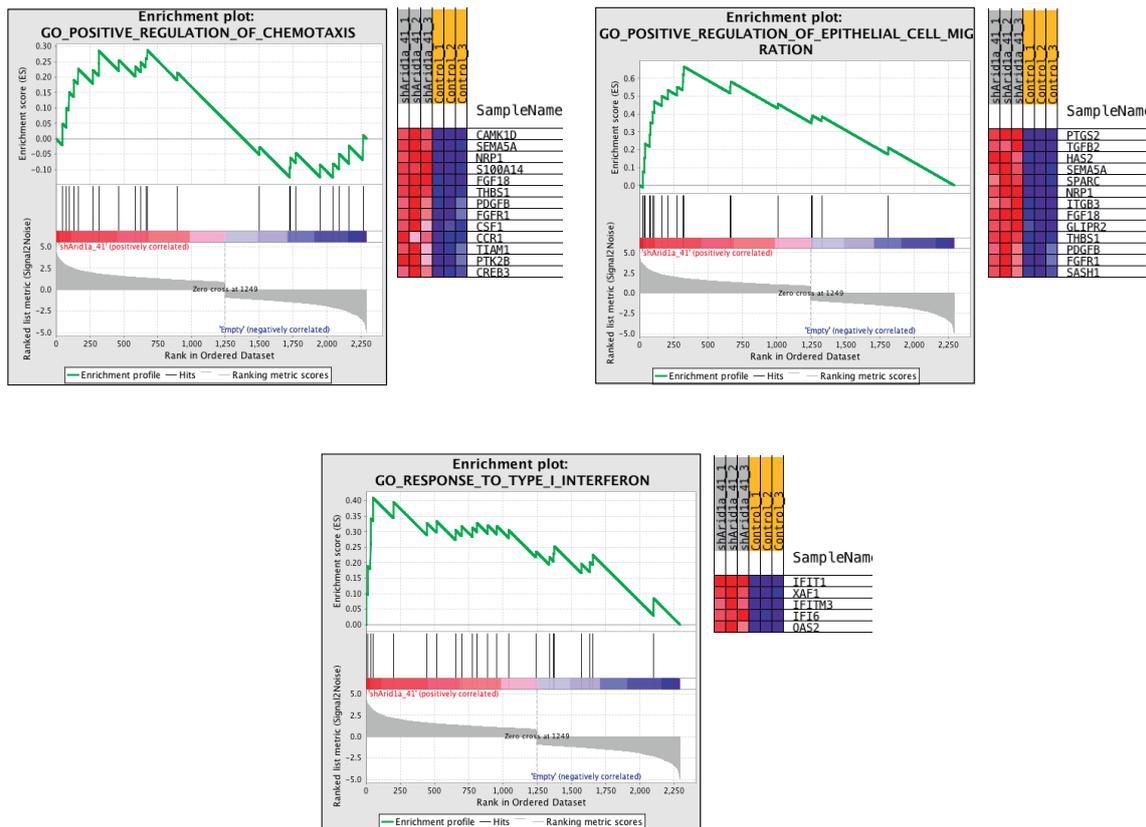
RNA-seq results (figure 23) indicated that the downregulation of *ARID1A* in Caco-2 cells was accompanied with an upregulation of genes related to the MAPK (*MAPK8IP2*, *MAP3K6*, *MAP3K14*, *MAP2K1*, *MAP3K3*, *TGFB2*, *RASGRP1*, *TIAMI*, and *MKNK2*) and PI3K pathways (*PIK3CD*, *IRS1*, *GNB4* and *GNB1*), as well as a downregulation of genes involved in the control of cell cycle (*RBI*, *SMC1B* and *WEE2*) and apoptosis (*BCL2*, *CTSC*, *CASP6* and *PMAIP1*). Collectively, these results might point to a more proliferative transcriptional programme in *ARID1A*-deficient cells.



**Figure 23. Transcriptional alterations of *ARID1A*-deficient Caco-2 cells.** Heatmap representing a selection of differentially expressed genes in *ARID1A*-deficient Caco-2 cells (n=3) and grouped according to their respective molecular pathway. Gene expression differences between shARID1A and shEmpty cells have been represented according to the log2 of the fold change and have been indicated by colours, which range between dark red (overexpression) to dark blue (downregulation).

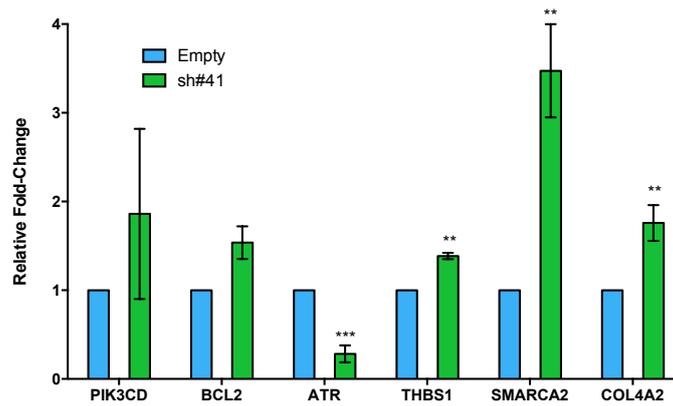
Furthermore, we also noticed a dysregulation of genes that seems to be altered specifically in this cell line, like those involved in chemotaxis (*NRP1*, *PDGFRA*, *THBS1* and *ANXA1*) and cell migration (*SPARC* and *ITGB3*), which might imply an increase in metastatic potential.

Remarkably, RNA-Seq revealed an upregulation of genes involved in the response to type I interferon (*IFIT1*, *IFITM3*, *IFI6* and *STAT1*), what suggest, as in the results obtained in SK-OV-3 cells, a potential role of these genes as biomarkers to predict response to immunotherapy. Additionally, GSEA analysis validated a significant enrichment of the above discussed pathways (figure 24).



**Figure 24. Molecular pathways significantly altered in *ARID1A*-deficient Caco-2 cells.** GSEA analysis showed significant enrichment of genes involved in the positive regulation of chemotaxis and epithelial cell migration, as well as in the response to type I interferon. GSEA results have been represented according to the enrichment score (ES): a positive value indicates enrichment in *ARID1A*-deficient cells (red), whereas a negative value implies an enrichment in empty-cells. Heatmaps of significantly enriched genes in each pathway are also represented with the following legend: genes highlighted in red are overexpressed in *ARID1A*-deficient cells, whereas those represented in blue are downregulated.

Finally, we validated by qRT-PCR the expression of some genes (*PIK3CD*, *BCL2*, *ATR*, *THBS1*, *SMARCA2* and *COL4A2*) that have been found dysregulated in the RNA-Seq results. We observed that these genes have changed in the same direction by both techniques (figure 25).



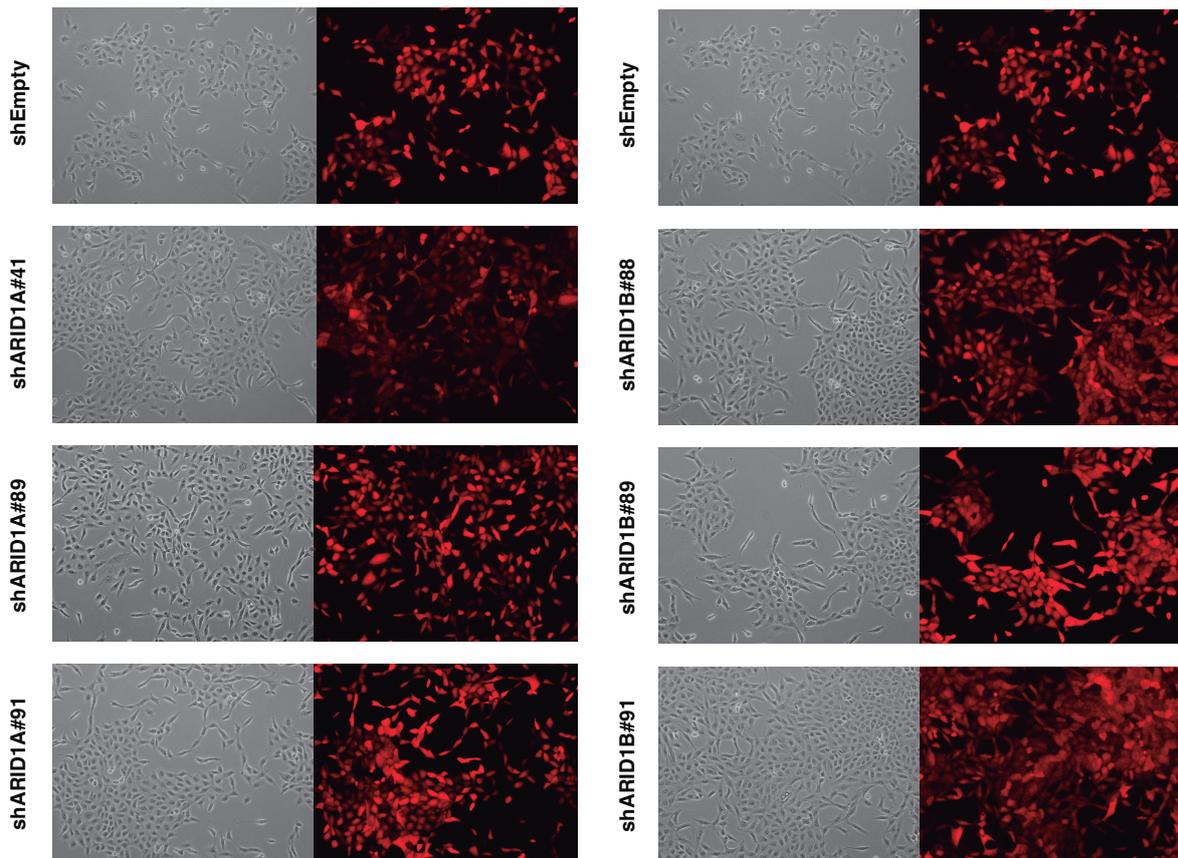
**Figure 25. Validation of transcriptional alterations observed in ARID1A-deficient Caco-2 cells.** Bar graph representing the results of the qRT-PCR validation of a selection of differentially expressed genes following *ARID1A*-downregulation in Caco-2 cells. The relative fold-changes are represented as mean  $\pm$  SEM of three independent experiments. The results of the one-tailed unpaired t-Test are represented with asterisks with the following criterion: \* p-value < 0.05, \*\* p-value < 0.01 and \*\*\* p-value < 0.001.

Whereas these results are significant in the case of *ATR*, *THBS1*, *SMARCA2* and *COL4A2*, we saw clear tendencies in *PIK3CD* and *BCL2* that did not achieve statistical significance, probably due to a high variability among the replicates. Thus, further experimentation is needed to validate these changes.

## 5.5. Generation of A549 stably-transduced cell lines

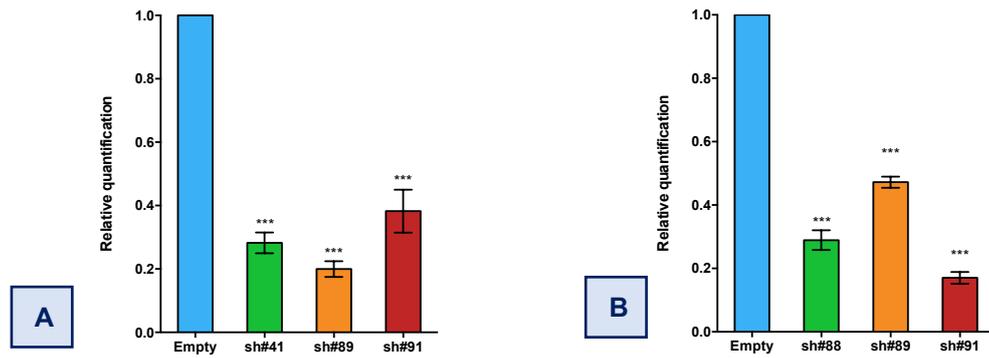
### 5.5.1. Validation of *ARID1A* and *ARID1B* knock-down

We have also generated stably-transduced A549 cell lines knock-down for *ARID1A* and *ARID1B*. Like in the previous sections, we have checked the expression of turboRFP in the different stable cell lines generated. As it can be distinguished in figure 26, the high level of expression of the reporter gene is particularly striking, considering that these cells had not been sorted. Therefore, these results seem to indicate that A549 cells have been better transduced than SK-OV-3 or Caco-2 cells.



**Figure 26. Verification of the generation of A549 stably-transduced cell lines measured through the expression of the reporter gene.** Representative microscopic images showing the expression of turboRFP of A549 cells stably-transduced with different shRNAs targeting *ARID1A* (V3\_410041, shARID1A#41; V3\_343789, shARID1A#89, V3\_343791, shARID1A#91) and *ARID1B* (V3\_306688, shARID1B#88, V3\_306689, shARID1B#89; V3\_306691, shARID1B#91), after adding 1  $\mu\text{g/ml}$  of doxycycline for 5 days. Fluorescence microscope images were taken with an equal exposure time of 600 ms.

Furthermore, when we looked up the downregulation of both *ARID1A* and *ARID1B* by performing a qRT-PCR (figure 27), we realised that an effective downregulation of both genes had been achieved and that it was comparable with the one reached in sorted Caco-2 cells. Therefore, we decided that the cell sorting step was not essential in this cell line.



**Figure 27. *ARID1A* and *ARID1B* mRNA expression in A549 stably-transduced cells lines.** **A.** qRT-PCR for *ARID1A* in A549 cells stably-transduced with different shRNAs targeting *ARID1A* (V3\_410041, sh#41; V3\_343789, sh#89, V3\_343791, sh#91) treated with 1  $\mu$ g/ml of doxycycline for 5 days. **B.** qRT-PCR for *ARID1B* in A549 cells stably-transduced with different shRNAs targeting *ARID1B* (V3\_306688, sh#88, V3\_306689, sh#89; V3\_306691, sh#91) treated with 1  $\mu$ g/ml of doxycycline for 5 days. The results of the one-tailed unpaired t-Test are represented with asterisks with the following criterion: \* p-value < 0.05, \*\* p-value < 0.01 and \*\*\* p-value < 0.001.

Regarding the results of the statistical analysis, we selected the shRNAs V3\_343789 (sh#89) and V3\_306691 (sh#91) to prepare the RNA-Seq libraries because they had further silenced *ARID1A* and *ARID1B*, respectively.

## 6. Discussion and on-going work

Regarding the overall results obtained after *ARIDIA* knock-down in different cancer cell lines, it is particularly impressive how different are the transcriptional alterations reported, in terms of total number of genes dysregulated (94 genes in SK-OV-3 cells against 2,293 in Caco-2 cells). These differences could in part be explained considering that *ARIDIA* has been downregulated more efficiently in Caco-2 stable cell lines at both mRNA (*ARIDIA* expression on knock-down cells has been reduced up to 22.4 %, figure 20, whereas in SK-OV-3 knock-down cells its expression has only decreased up to a 37.2 %, figure 13) and protein levels (where we only reported a clear repression in Caco-2 knock-down cells, figure 14 and figure 21). Nevertheless, a more prominent role of *ARIDIA* in regulating gene expression on Caco-2 cells cannot be completely ruled out.

Interestingly, in spite of the differences reported in the number of genes dysregulated, we have observed that *ARIDIA*-downregulation in different cellular contexts results in the alteration of both shared and tissue-specific molecular pathways.

Among the tissue-independent molecular pathways altered, it should be highlighted an upregulation of genes associated with proliferation or improved survival (like those belonging to the MAPK, mTOR and PI3K/Akt pathways), as well as a downregulation of genes involved in apoptosis and the regulation through cell cycle progression (figure 16 and figure 23), which collectively suggest a transcriptional change towards a more proliferative phenotype.

This overlap has not been observed at the gene level but at the pathway level instead, which might suggest that these genes are not direct targets of *ARIDIA* regulation, as it seems to be the result of a higher proliferative rate. Therefore, in order to prove the presumably higher proliferative potential of *ARIDIA*-deficient cells, we are planning to do proliferation assays. Regarding this, in our laboratory we have performed similar experiments with *ARID2*-deficient A549 cells, so all the protocols are already set up.

Additionally, we have observed that even if some pathways are dysregulated in both cell lines, their alterations lead to opposite effects, as it is the case of genes implicated in the response to DNA damage, as well as those related to the immune response. This is not completely surprising as it has been reported that *ARIDIA* might play different, and even opposite roles, depending on the cellular context<sup>12</sup>.

It is particularly interesting that *ARIDIA*-deficiency in SK-OV-3 cells might result in a more harmful environment, considering that there has been an upregulation of genes in charge of the response to DNA damage. Additionally, GSEA analysis has confirmed that the DNA repair molecular pathway is significantly enriched in *ARIDIA*-deficient cells (figure 17).

Therefore, this result opens up the possibility of studying the effect of *ARID1A* alteration on DNA repair mechanisms after the induction of DNA damage. Considering the direct interactions that has been described between *ARID1A* and *ATR*<sup>18</sup>, it is plausible to expect that *ARID1A*-downregulation might result in an impaired response to DSB, what might in turn result in an increase of genomic instability. On the other hand, in *ARID1A*-deficient Caco-2 cells *ATR* has been significantly downregulated, although the significance of these result should be carefully considered as only a single gene of a particular pathway has been significantly altered. In order to further investigate this effect, we have decided to study the accumulation of DNA damage foci using gamma-H2AX immunostaining in *ARID1A*-deficient cells treated with DNA damaging agents, like etoposide or cisplatin.

The dysregulation of genes related to the immune response observed in both cell lines might imply different sensitivities to immunotherapy approaches. In relation to this, *ARID1A* deficient SK-OV-3 cells presented an upregulation of genes that belong to the negative regulation of the immune system, which could be linked to the response of these tumours to immunotherapy. However, in the case of *ARID1A*-deficient Caco-2 cells there has been an upregulation of genes involved in the response to type I interferon (IFN), which seems to be a promising preliminary result because there is solid evidence that supports a positive correlation between the expression of IFN-stimulated genes and a favourable disease outcome in several tumour types<sup>35</sup>.

Additionally, it has been reported that type I interferon signalling might be involved in therapeutic effect of some monoclonal antibodies, such as ipilimumab (anti-CTLA4) or nivolumab (anti-PD1)<sup>35</sup>. And also, mice bearing tumours induced by the murine colon carcinoma cell line CT26 showed a decrease in the intratumoral abundance of T<sub>Reg</sub> and an accumulation of T helper 17 after injecting vectors encoding type I IFN<sup>35</sup>, what confirms the role of type I IFN in cancer immunosurveillance.

Interestingly, it is known that the expression of some genes belonging to the response to type I IFN, such as *IFITM3*, relies specifically on BAF complexes. Therefore, we could assume that the *IFITM3* overexpression measured by RNA-Seq is a direct effect of *ARID1A*-downregulation. However, in opposition to our results, *Yan et al. (2005)*<sup>36</sup> found that the expression of *IFITM3* was strongly inhibited in *ARID1A*-deficient HeLa cells at the protein level. This observation reinforces the hypothesis that *ARID1A* may display different regulatory programmes, depending on its interaction with tissue-specific activators or repressors.

Finally, in both *ARID1A*-deficient cell lines transforming growth factor beta 2 (TGFβ2) has been found upregulated, which is particularly interesting because the increased expression of this molecule is associated with poor prognosis in many cancers<sup>37</sup>. In particular, high expression of TGFβ1 in tumour cells correlates with worse prognosis and metastases in colorectal cancer<sup>38</sup>. Moreover, perturbations of its signalling pathway are essential for cell proliferation and cell invasion<sup>38</sup>, what concurs with our RNA-Seq results. This tumour-promoting function of TGFβ could be explained considering that it affects the

recognition and destruction of tumour cells through the regulation of immune cell function and it also stimulates angiogenesis<sup>38</sup>. Additionally, it has been reported that the inhibition of TGF $\beta$  gives rise to an enhanced infiltration of cytotoxic T cells and T helper into primary tumour and the metastases in model animals<sup>39</sup>.

Collectively, these results had encouraged us to further study the association between SWI/SNF impairment and the response to immunotherapy. For that, our laboratory has recently acquired commonly used T-cell and NK cell lines and we are planning to co-culture them with *ARIDIA*-deficient cells in order to investigate potential changes in their sensitivity to immunotherapy.

On the other hand, *ARIDIA*-deficiency leads to transcriptional alterations that seems to be tissue-dependent. For example, Caco-2 knock-down cells showed an overexpression of genes involved in cell migration and chemotaxis. A significant enrichment in these molecular pathways has also been proved by GSEA analysis (figure 24). Therefore, these results might suggest a higher metastatic potential following *ARIDIA*-downregulation in this cell line, although specific invasion and migration assays should be performed in order to verify this hypothesis.

To complete our understanding of the transcriptional alterations resulted from *ARIDIA* deficiency, we are also sequencing *ARIDIA*-deficient A549 cells. Presumably the results obtained will help to clarify which alterations are tissue-independent and which ones seem to be tissue-specific, considering that the downregulation of *ARIDIA* achieved at mRNA level in A549 cells is similar to the one reached in Caco-2 cells.

It is also worthy to mention that in our laboratory we have also characterized the transcriptional changes resulted from *ARID2*-deficiency in the A549 cell line. Additionally, we have also prepared the RNA-Seq libraries of *ARID1B* knock-down A549 cells. Collectively, the comparison of the results obtained after the downregulation of the three different ARIDs in the same cell line might shed some light on the shared and specific targets of each ARID, which besides could indicate if they have synergic or opposing effects in terms of transcriptional regulation.

In relation to the qRT-PCR validations, our results should be carefully considered as they have been obtained using the same RNA that was used to prepare the RNA-Seq libraries. Therefore, to clearly demonstrate if those genes are systemically dysregulated after *ARIDIA* knock-down, qRT-PCR validations should be repeated in independently generated *ARIDIA*-deficient cells and also in stable cell lines generated with a different shRNA targeting *ARIDIA*. Ideally, we should also validate tissue-specific transcriptional alterations, such as genes involved in DNA repair in SK-OV-3 cells, genes related to chemotaxis and cell migration in Caco-2 cells and genes related to immune response in both cell lines.

Strikingly, the expression of SMARCA2 has changed in opposite directions in both cell lines after *ARID1A* knock-down. It has been significantly increased in Caco-2 cells by both RNA-Seq (figure 23) and qRT-PCR (figure 25), which is particularly interesting because according to “The Human Protein Atlas” these cells do not express this catalytic subunit.

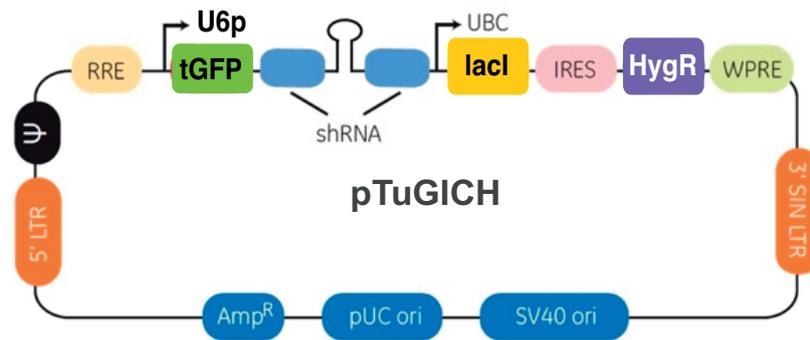
On the contrary, in SK-OV-3 knock-down cells SMARCA2 has been found significantly downregulated by RNA-Seq (figure 16) and qRT-PCR (figure 18). In this case, considering that SMARCA2 and ARID1A are part of the same protein complex, *ARID1A* -downregulation might give rise to an excess of SMARCA2 molecules. Therefore, regarding that there would not be enough molecules of ARID1A to assemble and form BAF complexes, the cell could decrease the synthesis of SMARCA2 as a compensatory mechanism.

Although the results obtained seems to be contradictory, this finding is particularly noteworthy because it has not been previously described that *ARID1A* is able to transcriptionally regulate the expression of the catalytic subunits of the complex. However, further research will be needed to determine if SMARCA2 is a direct target of ARID1A-containing BAF complexes or if its transcriptional alteration is an indirect consequence of *ARID1A*-downregulation.

In order to reinforce the data obtained from the RNA-Seq experiments, in addition to the above-mentioned assays, it would be advisable to validate the repression of ARID1A at the protein level. In relation to this, we have performed a preliminary Western Blot where we have seen a clear reduction of ARID1A production in Caco-2 cells (figure 21). Unfortunately, that has not been the case of SK-OV-3 stable cells, where we have not seen any significant decrease in ARID1A expression due to problems in the lane total protein load (figure 14). The high molecular weight of ARID1A (250 kDa) together with the lack of really specific antibodies against this protein complicate to a greater extent the performance of Western Blots. The presence of multiple bands might represent different ARID1A isoforms, but a lack of specificity of the primary antibody cannot be completely ruled out. In relation to this, we have tested two different antibodies in these experiments and we have obtained similar results.

Finally, we are also planning to study the potential synthetic lethality among different subunits of the SWI/SNF complex or with another canonical pathways known to be dysregulated in cancer. In order to accomplish this objective, we have developed a new inducible lentiviral vector (pTuGICH, figure 28), which has been modified to allow its combined used with pTRIPZ constructs.

In first place, we have replaced the doxycycline-inducible promoter by and IPTG-inducible promoter (U6p), which has been modified to be expressed in human cells. Secondly, we have change the reporter gene, which is now a green fluorescent protein (turboGFP, *represented as tGFP*), to allow the visual marking of the pTuGICH-expressing cells. Finally, we have also replaced the eukaryote selection marker by a gene that confers resistance to the antibiotic hygromycin.



**Figure 28. Detailed vector map of pTuGICH.** Representation of the main elements of the pTuGICH lentiviral vector.

Considering that both pTRIPZ and pTuGICH constructs have different promoters, reporter genes and selection markers, it would be possible to generate stable cell lines on which two genes could be independently downregulated.

The overall results obtained in this Master's thesis have allowed the identification of potential pathways altered following SWI/SNF alteration. We could take advantage of these deficiencies to design new antitumoral therapies and additionally this knowledge could lead to the identification of prognostic factors that might help to better define the subset of patients expected to benefit from different antitumoral treatments.

## 7. Conclusions

1. We have efficiently generated stably-transduced SK-OV-3, Caco-2 and A549 cell lines that inducibly express shRNAs for both *ARID1A* and *ARID1B*, efficiently repressing the expression of both genes
2. *ARID1A* knock-down in different cell lines results in the alteration of both shared and tissue specific molecular pathways.
3. *ARID1A*-deficiency is accompanied with an upregulation of genes associated with proliferation and a downregulation of genes involved in apoptosis in SK-OV-3 and Caco-2 cell lines.
4. *ARID1A*-deficiency in SK-OV-3 cells is accompanied with an upregulation of genes involved in DNA-repair, which suggests an increase in genomic instability.
5. *ARID1A*-deficiency in Caco-2 cells is associated with the overexpression of genes involved in chemotaxis and cell migration, which might imply a potential increase in their metastatic capacities.
6. *ARID1A*-deficiency in both SK-OV-3 and Caco-2 cells alters the expression of genes associated with different sensitivities to immunotherapy, which is likely to open new therapeutic opportunities for cancer patients.

## 8. References

1. Illumina. An introduction to Next-Generation Sequencing Technology. [https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina\\_sequencing\\_introduction.pdf](https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf)
2. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
3. Berger, M. F. & Mardis, E. R. The emerging clinical relevance of genomics in cancer medicine. *Nat. Rev. Clin. Oncol.* **15**, 353–365 (2018).
4. Wilson, B. G. & Roberts, C. W. M. SWI/SNF nucleosome remodellers and cancer. *Nat. Rev. Cancer* **11**, 481–492 (2011).
5. Gonzalez-Perez, A., Jene-Sanz, A. & Lopez-Bigas, N. The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. *Genome Biol.* **14**, r106 (2013).
6. 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).
7. Tang, L., Nogales, E. & Ciferri, C. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Prog. Biophys. Mol. Biol.* **102**, 122–128 (2010).
8. Kadoch, C. & Crabtree, G. R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* **1**, e1500447 (2015).
9. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).
10. Li, M. *et al.* Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat. Genet.* **43**, 828–829 (2011).
11. Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251–263 (2012).
12. Sun, X. *et al.* Arid1a Has Context-Dependent Oncogenic and Tumor Suppressor Functions in Liver Cancer. *Cancer Cell* **32**, 574–589.e6 (2017).
13. Shain, A. H. & Pollack, J. R. The Spectrum of SWI/SNF Mutations, Ubiquitous in Human Cancers. *PLOS ONE* **8**, e55119 (2013).
14. Otto, J. E. & Kadoch, C. A Two-Faced mSWI/SNF Subunit: Dual Roles for ARID1A in Tumor Suppression and Oncogenicity in the Liver. *Cancer Cell* **32**, 542–543 (2017).
15. Mathur, R. *et al.* ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nat. Genet.* **49**, 296–302 (2017).
16. Chandler, R. L. *et al.* Coexistent ARID1A–PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. *Nat. Commun.* **6**, 6118 (2015).
17. Raab, J. R., Resnick, S. & Magnuson, T. Genome-Wide Transcriptional Regulation Mediated by Biochemically Distinct SWI/SNF Complexes. *PLOS Genet.* **11**, e1005748 (2015).
18. Shen, J. *et al.* ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. *Cancer Discov.* **5**, 752–767 (2015).
19. Rother, M. B. & van Attikum, H. DNA repair goes hip-hop: SMARCA and CHD chromatin remodellers join the break dance. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372**, (2017).

20. Wang, H. *et al.* BRCA1/FANCD2/BRG1-Driven DNA Repair Stabilizes the Differentiation State of Human Mammary Epithelial Cells. *Mol. Cell* **63**, 277–292 (2016).
21. 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).
22. Dykhuizen, E. C. *et al.* BAF complexes facilitate decatenation of DNA by topoisomerase II $\alpha$ . *Nature* **497**, 624–627 (2013).
23. Kelso, T. W. R. *et al.* Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. *eLife* **6**, (2017).
24. Oike, T. *et al.* A synthetic lethality-based strategy to treat cancers harboring a genetic deficiency in the chromatin remodeling factor BRG1. *Cancer Res.* **73**, 5508–5518 (2013).
25. Wu, S., Fatkhutdinov, N. & Zhang, R. Harnessing mutual exclusivity between TP53 and ARID1 A mutations. *Cell Cycle* **16**, 2313–2314 (2017).
26. Wang, X., Haswell, J. R. & Roberts, C. M. W. Molecular Pathways: SWI/SNF (BAF) complexes are frequently mutated in cancer—mechanisms and potential therapeutic insights. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **20**, 21–27 (2014).
27. Hohmann, A. F. & Vakoc, C. R. A rationale to target the SWI/SNF complex for cancer therapy. *Trends Genet. TIG* **30**, 356–363 (2014).
28. Sun, C. *et al.* BRD4 Inhibition Is Synthetic Lethal with PARP Inhibitors through the Induction of Homologous Recombination Deficiency. *Cancer Cell* **33**, 401–416.e8 (2018).
29. Pan, D. *et al.* A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. *Science* **359**, 770–775 (2018).
30. Miao, D. *et al.* Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science* **359**, 801–806 (2018).
31. Thaidy Moreno Rodriguez. Molecular Characterization of the role of chromatin remodeling complexes in tumor progression (PhD thesis, Universidad de Cantabria, 2017).
32. Moreno *et al.* Recurrent ARID2 somatic mutations are associated with poor prognosis, increased metastatic potential and can be exploited for therapy in lung cancer (manuscript under review).
33. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2 $^{-\Delta\Delta CT}$  Method. *Methods* **25**, 402–408 (2001).
34. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
35. Zitvogel, L., Galluzzi, L., Kepp, O., Smyth, M. J. & Kroemer, G. Type I interferons in anticancer immunity. *Nat. Rev. Immunol.* **15**, 405–414 (2015).
36. Yan, Z. *et al.* PBAF chromatin-remodeling complex requires a novel specificity subunit, BAF200, to regulate expression of selective interferon-responsive genes. *Genes Dev.* **19**, 1662–1667 (2005).
37. Visan, I. Targeting TGF- $\beta$  in cancer. *Nat. Immunol.* **19**, 316 (2018).
38. Ikushima, H. & Miyazono, K. TGF $\beta$  signalling: a complex web in cancer progression. *Nat. Rev. Cancer* **10**, 415–424 (2010).
39. Tauriello, D. V. F. *et al.* TGF $\beta$  drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature* **554**, 538–543 (2018).