The Role of ARID1A in Oncogenic Transcriptional (de)Regulation in Colorectal Cancer

Dissertation

for the award of the degree

“Doctor of Philosophy”
Division of Mathematics and Natural Sciences

of the Georg-August-Universität Göttingen

within the doctoral program

“Molecular Biology”

of the Georg-August University School of Science (GAUSS)

submitted by

Madhobi Sen

from New Delhi, India

Göttingen, 2018

Thesis Committee

1. Prof. Dr. Steven A. Johnsen, Department of General, Visceral and Pediatric Surgery, University Medical Center, Göttingen

2. Prof. Dr. Matthias Dobbelstein, Institute of Molecular Oncology, University Medical Center, Göttingen

3. Dr. Melina Schuh, Department of Meiosis, Max Planck Institute for Biophysical Chemistry, Göttingen

Members of the Examination Board

1. Referee: Prof. Dr. Steven A. Johnsen, Department of General, Visceral and Pediatric Surgery, University Medical Center, Göttingen

2. 2nd Referee: Prof. Dr. Matthias Dobbelstein, Institute of Molecular Oncology, University Medical Center, Göttingen

Further members of the Examination Board

1. Dr. Melina Schuh, Department of Meiosis, Max Planck Institute for Biophysical Chemistry, Göttingen

2. Prof. Dr. Heidi Hahn, Department of Human Genetics, University Medical Center, Göttingen

3. Prof. Dr. Jürgen Wienands, Institute for Cellular and Molecular Immunology, University Medical Center, Göttingen

4. Dr. Ufuk Günesdogan, Department of Developmental Biology, University of Göttingen

Table of Contents

List of Abbreviations

°C	degree Celsius
µ	micro
ac	Acetylation
AP1	Activator protein 1 family of transcription factors
APC	Adenomatous polyposis coli
APS	Ammonium persulfate
ARID1A	AT-rich interacting domain containing protein 1A
ATP	Adenosine triphosphate
ATAC	Assay for Transposase-Accessible Chromatin
BAF	BRG1or BRM-associated factors
BET	Bromodomain and extra-terminal domain containing proteins
BGP	ß-Glycerophosphate
BSA	Bovine serum albumin
Bp	Base pairs
bZIP	Basic leucine zipper domain
C57BL/6J	C57 black 6 mice
CCLE	Cancer Cell Line Encyclopedia
CDK	Cyclin dependent kinase
CDX2	Caudal-related homeobox protein 2
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP sequencing
CIMP	CpG Methylator Phenotype
CIN	Chromosomal instability
cm	Centimeter
CO_2	Carbon dioxide
COAD	Colon adenocarcinoma
CpG	Cytosine phosphate Guanine
CRC	Colorectal cancer
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats/Cas9 mediated gene editing
CTD	C-terminal domain
DAI	Disease activity Index
DEPC	Diethylpyrocarbonate

DMEM	Dulbecco modified Eagle's minimal essential medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	deoxyribonucleotide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EGS	Ethylene glycol bis (succinimidyl succinate)
EMT	Epithelial to mesenchymal transition
ENCODE	Encyclopedia of DNA Elements
ERK	Extracellular signal-regulated kinases
eRNA	Enhancer RNA
EZH2	Enhancer of zeste homolog 2
FBS	Fetal bovine serum
Floxed	Flanked by LoxP
GAP	GTPase activating protein
GFP	Green fluorescent protein
GO	Gene ontology
GREAT	Genomic Regions Enrichment of Annotations Tool
gRNA	Guide RNA
GSEA	Gene set enrichment analysis
GTP	Guanosine triphosphate
h	Hour
HAT	Histone acetyl transferase
HCC	Hepatocellular carcinoma
HDAC	Histone Deacetylase
H2Bub1	Monoubiquitinated histone 2B
H3	Histone 3
H4	Histone 4
H3K4me1	Histone 3 monomethylated at position lysine 4
H3K4me3	Histone 3 trimethylated at position lysine 4
H3K27ac	Histone 3 acetylated at position lysine 27
H3K27me3	Histone 3 trimethylated at position lysine 27
HP1	Heterochromatin protein 1
HRP	Horseradish peroxidise
HSC70	Heat shock 70kDa protein

IAA	Iodacetamide
IgG	Immunoglobulin G
IHC	Immunohistochemistry
IL	Interleukin
kDa	kilo Dalton
KRAS	Kirsten rat sarcoma viral oncogene homolog
KD	Knockdown
KO	Knockout
L	Liter
m	milli (10^{-3})
M	Molar
MAPK	Mitogen activated protein kinase
Me	Methylation
MEF	Mouse embryonic fibroblast
min	Minute
MNase	Micrococcal Nuclease
mRNA	Messenger RNA
MSI	Microsatellite instability
n	nano (10^{-9})
n	N number, sample size
NGS	Next generation sequencing
NEM	N-ethylmaleimide
NP-40	Nonidet P40
OEC	Ovarian endometroid cancer
OCCC	Ovarian clear cell carcinoma
PARP	poly ADP ribose polymerase
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
pH	Measurement of acidity or alkalinity of a solution
PIC	Protease inhibitor cocktail
PIC	Pre-initiation complex
PRC2	Polycomb Repressive Complex 2
qRT-PCR	Quantitative real-time PCR
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
SAM	S-adenosyl methionine

SCID	Severe combined immunodeficiency
SDS	Sodium dodecylsulfate
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
sec	Second
siRNA	Small interfering RNA
SHO	SCID hairless outbred
SWI/SNF	SWItch/Sucrose Non-Fermentable
TAD	Topologically associated domain
Taq	Thermus aquaticus
TCGA	The cancer genome atlas
TEMED	Tetramethylethylenediamine
TF	Transcription factor
TNFα	Tumor necrosis factor α
Tris	Tris(hydroxymethyl)aminomethane
TSS	Transcription start site
U	Unit (enzyme activity)
V	Voltage
WT	Wildtype

1. Abstract

The deregulation of epigenetic modulation has been well established as a common occurrence in cancer. However, the extent of its involvement in the development and progression of cancer was underscored by genome- and exome-wide sequencing studies over the last several years, which revealed a close association between the epigenome and the pathogenesis of cancer. A class of chromatin regulators, the ATP-dependent chromatin remodellers (particularly subunits of the mammalian BAF complex) are among the most frequently mutated genes across cancer types. These factors showed an alteration frequency in over 20% of all cancers. Among the BAF complex subunits, defects in *ARID1A* (AT-rich interactive domain-containing protein 1A) are the most frequently found and widespread across many cancer types. Mutations in *ARID1A* are usually frameshift or nonsense mutations which lead to a loss of its expression in these tumors. In fact, *ARID1A* is also one of the most frequently mutated chromatin regulators in human colorectal cancer (CRC). Several studies in cell culture and mouse models have shown that loss of ARID1A leads to increased proliferation and tumorigenesis in several cancer types, indicating a tumor suppressive function. Very interestingly, a study described the pivotal role of ARID1A in driving CRC in which its inactivation alone led to the formation of invasive adenocarcinomas in mice.

Surprisingly, in contrast to the expected tumor suppressive role of ARID1A in CRC, we observed that the knockout (KO) of *ARID1A* in CRC cell lines leads to impaired proliferation. Moreover, subcutaneous xenografts in SCID mice using human *ARID1A* KO CRC cells did not form more aggressive tumors than their wildtype counterparts. Also, the generation of several mouse models in the literature of *Arid1a* deletion revealed that it can have oncogenic functions. These results indicate a context-dependent role of ARID1A in cancer. We observed an impairment in proliferation in two of the four cell lines in which we performed *ARID1A* KO. Strikingly, these cell lines harbor the *KRAS*^{G13D} mutation. Therefore, we sought to explore the transcriptional role that ARID1A plays downstream of this pathway. To uncover this, we utilized several publicly available ChIP-seq, mRNA-seq and ATAC-seq datasets and generated our own ChIP-seq dataset for ARID1A in the CRC cell line HCT116. We observed a substantial co-localization of the BAF complex with AP1 transcription factors, such as JunD, that act downstream of the MEK/ERK signaling pathway, suggesting cooperation between these factors. Analysis of the sites at which ARID1A binds showed an enrichment of AP1 transcription factor binding sequences. Most sites co-occupied by ARID1A/AP1 are distal to transcription start sites. Therefore, it is likely that these transcriptional regulators

functionally interact at enhancers to elicit gene expression changes in CRC. To examine this, we next explored the MEK/ERK pathway. We identified some targets that are co-localized at distal regulatory sites for genes and are downregulated by *ARID1A* KO, Trametinib (a MEK/ERK pathway inhibitor) treatment and depletion of JunD. Strikingly, the occupancy of JunD and the acetylation of H3K27 (often an active enhancer mark) was also reduced at these distal regulatory sites upon Trametinib treatment and *ARID1A* KO. Conversely, the occupancy of ARID1A was reduced upon Trametinib treatment or JunD depletion. Thus, these regulatory regions are targets of the MEK/ERK pathway (through AP1) and are dependent on ARID1A as a co-factor. This effect does not seem to be mediated by the known chromatin remodeller functions of ARID1A since the accessibility of chromatin is not affected upon its loss. Thus, we were able to show that ARID1A is required for regulation of *KRAS* mutation-driven CRC by acting as a co-factor with AP1 transcription factors (TFs) which are downstream of the MEK/ERK pathway at distal regulatory elements. Importantly, this enables the identification of a strategy to stratify CRC by *KRAS* mutation status and to target the BAF complex in CRCs that are particularly dependent on this pathway.

2. Introduction

2.1 Epigenetics

The astounding organizational complexity and accuracy in living systems is mediated by the vast amount of information contained in their genetic code. This information is present as a sequence of DNA in the cells, which is inherited over cellular and organismal generations. Cells, however, are posed with a few challenges when dealing with this genetic information. This enormous amount of genetic material must be contained within the constraints of the nucleus (in human cells 2 meters of DNA must be condensed in a space of about 6 µM diameter). Therefore, DNA is a highly condensed structure which also serves to protect it against damaging agents. This compaction however, presents a major challenge for accessibility of the DNA sequences. This information must be accessed so that the genetic material can be replicated as well as expressed. A selectively fine-tuned spatial and temporal expression of genes is what determines the fate of a cell. This expression is carefully orchestrated by several mechanisms which ensure cellular homeostasis. Thus, it is not surprising that in some crucial steps, even a small mis-regulation can disrupt homeostasis completely and manifest itself as diseases such as cancer.

2.1.1 The Nucleosome and Higher Order Chromatin Structure

The large amount of DNA that eukaryotic cells possess is condensed within the constraints of the nucleus with the help of several proteins which package the DNA into higher degrees of organization. This complex of DNA and associated proteins is called chromatin. The basic unit of the chromatin is the nucleosome which consists of around 147 base pairs of DNA wrapped around a histone octamer (Hewish et al., 1973, Richmond et al., 1984). This octamer contains two each of the basic histone proteins H2A, H2B, H3 and H4. These are the canonical histones coupled most often to incorporation after replication of the DNA. These proteins are highly conserved across species, and among the most abundant proteins in the cell (Albig et al., 1997). At certain loci on the genome, however, variants of these histones can be present which modulate the structural and functional fates of these regions. For example, H2AX in its phosphorylated form is a histone variant which is incorporated into sites of the genome where double stranded breaks occur (Rogakou et al., 1998) and helps in the assembly of the repair machinery at these sites. The N-termini of histone proteins are quite unstructured and extend out of the nucleosome. These are the sites of a variety of post translational modifications (Strahl and Allis, 2000, Review).

Interactions between nucleosomes, for example, *via* the H1 histone linker and other architectural proteins, facilitate looping and coiling of the chromatin to further condense DNA into higher order structures (Thoma et al., 1977). While the conventional model proposes that the nucleosome wrapped chromatin (the 11nM fiber) is packaged into a 30nM fiber, recent literature suggests that this does not occur and instead structures such as loops, hubs and stacks involving several nucleosomes are formed (Ou et al., 2017). While packaging ensures that the chromatin fits into the nucleus, the challenge of accessibility for processes like replication, transcription and repair still remains. Therefore, the chromatin must be a dynamic structure which allows access to DNA sequences as and when required. This is a highly regulated process involving several mechanisms. The structure of the folded chromatin itself also contributes to this regulation. Overall, long range interaction studies suggest that the chromatin organizes into two compartments known as compartment “A” and “B”, where A is associated with more transcriptionally active regions and B with transcriptionally inactive regions (Lieberman-Aiden et al., 2009). The more inactive parts of the genome are packaged into domains associated with the nuclear lamina and called Lamin Associated Domains (LADs). These regions consist of mostly highly condensed “heterochromatin” which can be constitutively silenced or be amenable to some regulation (Lieberman-Aiden et al., 2009). Chromatin folding in compartment A is more dynamic and interactions between regions that are at large linear distances from each other form Topological Associated Domains (TADs). These interactions are often between the enhancers and promoters of genes and their placement in the same TAD plays a major role in the regulation of gene expression (de Laat et al., 2013, Dixon et al., 2012). These domains are sometimes established by architectural proteins such as CTCF (CCCTC binding factor) and cohesin (Pombo and Dillon, 2015, Review). The regulation of the structure and function of chromatin, its effect on gene expression and the consequence of this in determining cell fate are studied within the field of epigenetics.

2.1.2 The Mediators of Epigenetic Regulation: Histone Modifying Enzymes and Histone Marks

Epigenetic regulation of gene expression, which plays major roles in processes like development and differentiation, is mediated by a number of proteins. Notably, there are enzymes that carry out reversible but stable modification of DNA such as methylation at CpG dinucleotides, which are associated with transcriptional repression (Issa, 2004, Review). However, apart from these, two major classes of enzymes that facilitate or impede gene expression are the histone modifying enzymes and the chromatin remodellers. The histone modifying enzymes can be further classified into three categories, writers, erasers

and readers. Firstly, writers are enzymes that post-translationally modify (most often) the N-terminal tails of the histone proteins. These modifications include acetylation or methylation of lysine residues, phosphorylation of serine or threonine residues, ubiquitylation, sumoylation etc. For example, histone acetyltransferases (HATs) such as p300 act as transcriptional co-activators and catalyze the acetylation of lysine 27 on histone 3 (H3K27ac) and methyl transferases such as EZH2 catalyze the methylation of histones. Secondly, erasers are enzymes that can remove these modifications thus making histone modification a reversible process and presenting an opportunity for regulation. For example, the acetylation placed by HATs can be removed by histone deacetylases (HDACs). Lastly, the presence of certain modifications allows docking of factors (called readers) to the chromatin through domains that recognize these modifications. Domains such as bromodomains recognize and bind to acetylation whereas chromodomains recognize methylation. A histone modification or a set of defined modifications (referred to as the histone code) can determine the fate of a particular region of the chromatin (Strahl and Allis, 2000, Review). For instance, trimethylation of histone H3 at lysine 27 (H3K27me3) is a repressive mark associated with inactive gene promoters and enhancers, whereas H3K27ac (acetylation at the same position) is associated with active gene promoters and enhancers. H3K4me1 (monomethylation of histone H3 at lysine 4) is a marker of primed enhancers and H3K4me3 (trimethylation at the same position) is a mark of active promoters (Shlyueva et al., 2014, Review). These modifications lead to the recruitment of certain proteins or complexes, which can recognize specific marks, for example, bromodomain containing proteins such as BRD4 which recognize acetylated lysines. This further regulates the expression of these regions. On the other hand, modifications like H3K9me3 can have long range effects in which they recruit the heterochromatin protein HP1 and inactivate constitutively large areas of the chromatin (Wreggett et al., 1994). As mentioned before, this process is reversible and can be modulated by different stimuli. A summary of histone modifiactions and their consequences are provided in Figure 1.

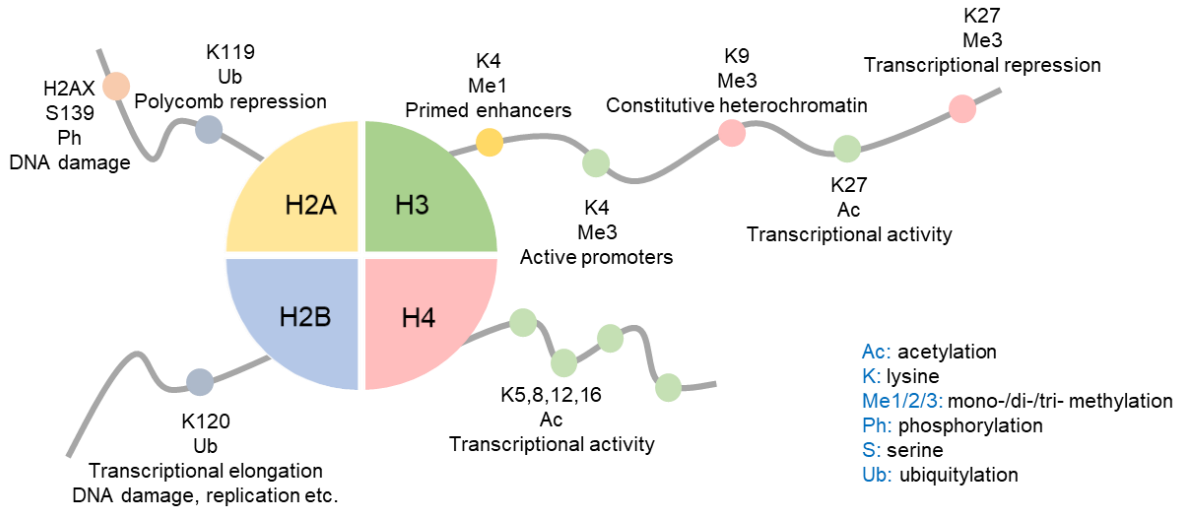


Figure 1: Post-translation histone tail modifications define the fate of the associated chromatin. Some of the most common histone tail modifications on each histone protein are presented. Some of the most interesting modifications that define the transcriptional fate of the chromatin are placed on histone H3. Methylation at lysine 4 can be associated with primed enhancers or active promoters. Acetylation at lysine 27 is an active mark whereas trimethylation at the same site is repressive. Histone H4 carries some acetylations that can be recognized by bromodomain containing proteins. H2A and H2B, both undergo mono-ubiquitylation whereas phosphorylation of the histone variant H2AX is important for DNA damage response.

2.1.3 Chromatin Remodellers

Chromatin remodellers on the other hand, are multi subunit complexes that use the energy of ATP to alter DNA-nucleosome interactions (Clapier et al., 2017, Review). These reposition nucleosomes along the DNA or expose a certain section of the DNA, previously wrapped around a nucleosome, making the DNA accessible to other regulatory proteins. Remodellers, with the help of histone chaperones, can also facilitate the mobilization and ejection of nucleosomes from the DNA and catalyze the exchange of the canonical histone core proteins for histone variants. Apart from their ATPase domains, chromatin remodellers also contain reader domains (for example the SWI/SNF complex contains bromodomains) that recognize histone modifications, which help in their recruitment to the chromatin (Clapier et al., 2017, Review). Moreover, some remodeller complexes can also contain histone modifying enzymes that are coupled to their action (for example the NuRD complex that contains the histone deacetylase HDAC1/2). The large variety of subunits with varied domains in several different combinations give rise to four major classes of functionally distinct chromatin remodellers: SWI/SNF, ISWI, CHD, and INO80, the roles of which have been determined in yeast. The major role of the CHD and ISWI complexes is to assemble the nucleosomes immediately after the replication of DNA and they are associated with

repression by forming tightly packed nucleosome clusters (Torigoe et al., 2011). The SWI/SNF remodellers on the other hand, are associated with gene activation as they facilitate nucleosome sliding and eviction (Whitehouse et al., 1999). These remodellers can make regions in the genome accessible to transcription factors and DNA repair enzymes. Histone exchange with variants is carried out by the INO80 family of remodellers (Mizuguchi et al., 2004). Given their importance in modulating chromatin dynamics, it is not surprising that these factors play a major role in the regulation of gene expression and are often deregulated in cancer.

2.2 The Transcriptional Machinery

Apart from chromatin modifiers, transcription factors which recognize and bind to specific DNA sequences are crucial regulators of gene expression or transcription. Eukaryotic transcription consists of three major steps: 1) Initiation, which involves the assembly of the initiation complex, 2) Elongation, which involves the transcription of the entire mRNA and 3) Termination, which involves the 3' polyadenylation of the mRNA and its disengagement from RNA polymerase. The basal transcriptional machinery consists of a core promoter just upstream of the transcription start site (TSS) where the Pre-Initiation Complex (PIC) (consisting of the general transcription factors that recognize the promoter sequences) assembles. The initial transcription factors such as TFIID and TFIIB recruit RNA Polymerase II (RNA Pol II), which catalyzes the synthesis of mRNA from a DNA template. RNA Pol II carries out multiple rounds of abortive transcription during which short transcripts of about 10 nucleotides are released. Promoter escape occurs when a transcript of sufficient length is formed, and the polymerase can transcribe the gene beyond the first few nucleotides. This is facilitated by the ATPase and kinase activities of the TFIIH general transcription factor which phosphorylates the C-Terminal Domain (CTD) of RNA Pol II at serine 5 (Watson et al., 2013, Molecular Biology of the Gene, 7th ed.). However, even after this, proximal to the promoter, the polymerase is often stalled. In the elongation step, the recruitment of elongation factors such as the P-TEFb (Positive Transcription Elongation Factor b) complex which contains the Cyclin Dependent Kinase CDK9, leads to the phosphorylation of serine 2 of the CTD of RNA Pol II. P-TEFb also phosphorylates and releases the NELF-DSIF complex, which is a negative regulator of elongation, thus releasing the polymerase which is promoter-proximally paused and allowing transcription to proceed (Peterlin and Price, 2006, Review). The phosphorylations of the CTD of RNA Pol II act as cues for the recruitment of 5' capping factors, splicing factors and 3' polyadenylation factors (Watson et al., 2013). These steps ensure the termination of transcription and formation of a mature mRNA.

The above paragraph describes the basal transcription at promoters responsive to the general transcription machinery, studied mainly in *in vitro* experiments. *In vivo,* transcription is a much more diverse and complicated process involving the regulation and integration of external and internal cues. At the promoter itself, several co-regulators and co-factors are recruited. However, the transcription through promoters themselves is basal and does not present many opportunities for regulation. Gene expression, in large part, is regulated by regulatory elements called enhancers that could be at very large linear distances from their target promoters (Lettice et al., 2003).

Each cell type must express a particular set of genes to establish its identity. It does so by activating cell type specific enhancers that can activate the expression of target genes. Enhancers are defined by certain characteristics such regions marked by certain histone modifications. Inactive enhancers are usually compacted and marked by H3K27me3 while primed enhancers bear the activating H3K4me1 mark. Poised enhancers bear the H3K27me3 along with the H3K4me1 mark requiring another step of regulation of removing the H3K27me3 to activate the enhancers. Active enhancers are nucleosome free, adjacently marked by H3K27ac and usually transcriptionally active, that is, bound by RNA polymerase II (Heinz et al, 2015, Review). Enhancer selection and activation is mediated by lineage dependent (LDTF) or pioneer transcription factors that recognize short DNA sequences and can bind to compacted chromatin. These act in concert with numerous other co-regulators and signal dependent transcription factors (TFs activated by a signaling event) to produce an integrated response to extra- or intra-cellular cues. Co-regulators can range from factors such as histone modifying enzymes that modulate the epigenetic landscape further, multi-subunit complexes such as chromatin remodellers which help in opening chromatin and the mediator complex, which through its multiple domains also help in protein-protein interaction and thus DNA looping (Heinz et al., 2015, Review). The collaborative and hierarchical binding of all these transcriptional regulators is what drives cell type specific gene expression (Figure 2). At active enhancers, acetylated H4 recruits the bromodomain containing protein BRD4 (Dey et al., 2003) which recruits P-TEFb, and as explained above, phosphorylates the CTD of RNA Pol II (Marshall et al., 1996), allowing the expression of enhancer RNAs (eRNA). While the role of eRNAs is still elusive, their expression has been correlated with target gene expression changes (De Santa et al., 2010, Kim et al., 2010). The interactions between the enhancers and target promoters is an area of intensive research and is believed to enhance transcription of the target genes by increasing the local concentrations of factors that could act at many steps of transcription, such as the pause-release of RNA Pol II. The interaction of enhancers and promoters could be mediated by chromatin looping and may be facilitated by interactions of the factors

present at the two elements. Moreover, eRNAs and other small RNAs are also believed to potentially mediate this interaction (Heinz et al., 2015, Review). Therefore, correct spatial and temporal gene expression is a very complex process with many layers of regulation involving the folding of the chromatin and interactions of gene promoters with distal regulatory elements. Any mis-regulation in enhancer mediated regulation of gene expression can alter cell identity and cause disruption of cellular homeostasis.

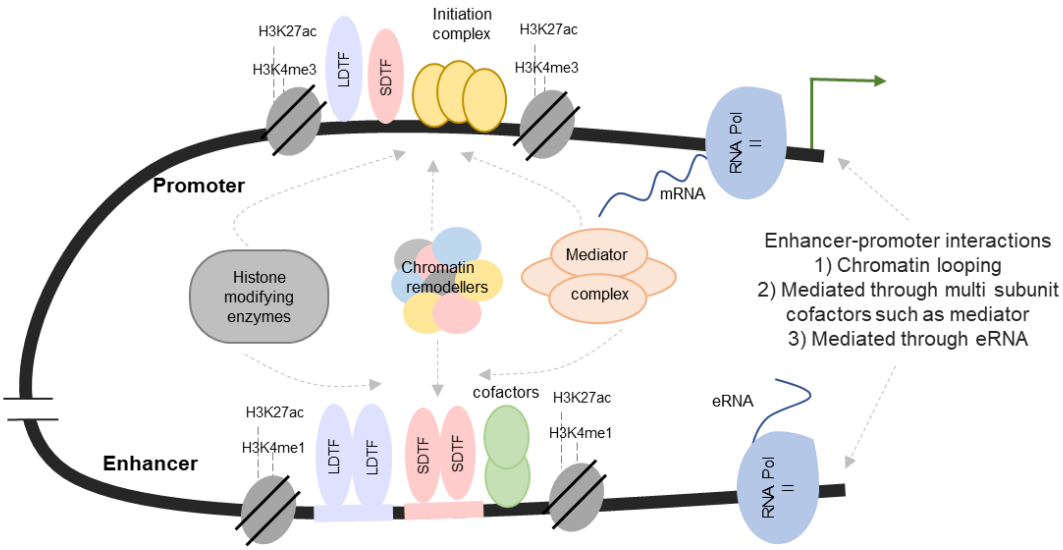


Figure 2: Potential mechanisms of transcriptional regulation by enhancers. The general transcriptional machinery does not explain the fine-tuned regulation of gene expression to define cell type identity. Out of the numerous potential enhancers present in the genome, cells activate those that are recognized by their lineage determining factors (LDTFs). The coordinated binding of signal determining transcription factors (SDTFs) in response to signals further recruits co-regulators and co-factors that alter the epigenetic landscape of the enhancer allowing fine-tuned modulation. The interaction between the enhancers and promoters is still not very well understood. However, certain hypotheses exist. It is postulated that chromatin looping could occur through the activity of architectural proteins, through large multi-subunit complexes, such as chromatin remodellers and mediator or even through eRNAs that are transcribed at the enhancers. Interaction of the enhancer and promoter increases local concentrations of transcriptional regulators at the promoter of the target gene. This helps to ultimately fine-tune the expression of the gene. Figure adapted from Heinz et al., 2015.

2.3 Epigenetic Mis-regulation in Cancer

While it is not surprising that deregulation of factors affecting chromatin structure and function could have consequences in cancer, the extent of their involvement was made clear when next generation sequencing studies revealed that chromatin modifiers are mutated in a large proportion of cancers. Unlike other genetic drivers of cancer which are most often associated with only one or a few types of cancer, these mutations are present in all types of cancers (Kadoch et al., 2013). Especially in the cases where these are driver mutations, changes in chromatin regulation leads to disruption in other tumorigenic genes. Moreover, epigenetic changes caused by alterations in chromatin modifiers are reversible

and therefore these present an interesting opportunity to develop therapies. Perturbations in cancer occur at all the layers of epigenetic regulation. At the fundamental level of histones, it is known that in 80% of pediatric diffuse intrinsic pontine gliomas, the histone H3 (mostly in its variant form H3.3) bears a K27M missense mutation which predicts poor prognosis (Schwartzentruber et al., 2012).

At the level of DNA modifications, it is known that methylated repeat regions in the genome can be lost, to widespread hypomethylation which can activate aberrant gene expression. However, hypomethylation and hypermethylation can also occur in a locus specific manner. The p16 tumor suppressor locus is often hypermethylated in colon, lung and breast cancer leading to the loss of its expression (Merlo et al., 1995). Similarly, oncogenic loci such as RAS can be hypomethylated leading to their overexpression (Nishigaki et al., 2005). Furthermore, the enzymes that catalyze these reactions, the DNA methyltransferases (DNMTs) are sometimes altered in cancer as well.

At the level of histone modifications, the balance of histone acetylation is maintained by two classes of enzymes, the histone acetyltransferases (HATs) and the histone deacetylases (HDACs). Histone acetylation, most often, plays an important role in gene activation and therefore a disruption in this balance can be detrimental. Various chromosomal translocations, deletions and mutations have been reported in these enzymes. The p300/CBP HAT is inactivated in a large proportion of lymphomas (Pasqualucci et al., 2011). Reports also suggest overexpression of HDACs occur in cancer leading to deacetylation over the genome (Bennett and Licht, 2017, Review). One of the most intensively studied group of epigenetic inhibitors that have been developed are HDAC inhibitors. For example, these inhibitors have been shown to reactivate the tumor suppressor p21 in some cancers (Richon et al., 2000).

The methylation of histones on the other hand, can lead to gene activation (for example H3K4me3) or repression (for example H2K27me3) depending on the context. Histone methyl transferases can be of two major classes, lysine methyl transferases (KMTs) or protein arginine methyl transferases (PRMTs). Both classes show alterations in different cancers (Bennett and Licht, 2017, Review). PRMTs have been shown to be overexpressed in non-small-cell lung cancer, lymphoma and leukemia (Bennett and Licht, 2017, Review). This deregulation can promote the oncogenic c-Myc pathway and regulate epithelial to mesenchymal transition (EMT) in breast cancer (Yao et al., 2014). The methyltransferase subunit of the repressive PRC2 complex, EZH2 (enhancer of zeste-homolog 2), which catalyzes the H3K27me3 modification, seems to be altered in a context-dependent fashion. While it is overexpressed and associated with EMT in breast cancer, prostate cancer and

lymphomas (Varambally et al., 2002, Kleer et al., 2003), it can also undergo loss of function mutation in malignancies of the myeloid origin (Ernst et al., 2010). EZH2 inhibitors have been developed by several pharmaceutical companies and are being studied individually and in combination with other inhibitors (Bennett and Licht, 2017, Review). Similarly, the histone lysine demethylases (KDMs) such as lysine (K)-specific histone demethylase 1A (LSD1) are associated with prostate cancer through an interaction with the androgen receptor (Metzger et al., 2005). Inhibitors of LSD1 were tested in clinical trials for small-cell lung cancer but the study was terminated (https://clinicaltrials.gov/).

While the roles of the writers and erasers have been discussed above, readers, which have no apparent enzymatic functions, also seem to be perturbed in various cancers. Among these are BRD2 which is overexpressed in certain lymphomas and translocations of BRD3 and 4 are involved in midline carcinomas (Yan et al., 2011). Inhibitors such as JQ1 which target the acetylated lysine recognizing bromodomains of these proteins have been developed and have shown some promising results in *in vivo* studies (Fillipakopolous et al., 2010).

Lastly and most recently, genome- and exome-wide sequencing studies have revealed that mutations in the human SWI/SNF complex are among the most frequent mutations in cancer. Subunits of this complex are mutated in over 20% of all cancers with varying roles in different cancers (Kadoch et al., 2013, Shain et al., 2013). The most frequently mutated subunit is ARID1A (AT-rich interactive domain containing protein 1A) which has been extensively described in the literature as a tumor suppressor (Guan et al., 2011, Chandler et al, 2015, Mathur et al., 2017). Numerous synthetic lethal, druggable targets have been discovered in the past few years in ARID1A-deficient tumors and the relevance of this wide spectrum of ARID1A loss in cancer is still being actively researched for clinically relevant findings. Apart from these, several microRNAs and non-coding RNAs (ncRNA) have also been shown to play roles in cancer (Schmitt and Chang, 2016, Review). In this section, we have illustrated only a few examples of how epigenetics is known to be mis-regulated in cancer, suggesting the profound impact of epigenetic factors on cancer.

2.4 The SWI/SNF (BAF) Complexes

The mammalian SWI/SNF (BAF-BRG1-associated factors) complexes are large, multi-subunit, chromatin remodellers that utilize the energy of ATP hydrolysis to mobilize, slide and evict nucleosomes which lead to alterations in gene transcription (Clapier, 2017, Review). These complexes were initially discovered in yeast as genes that affect the mating-type switching (SWI) (Neigeborn et al., 1984) and sucrose fermentation (Sucrose Non-Fermenting - SNF) phenotypes (Stern et al., 1984). Their link to chromatin was

established when mutations in histones and chromatin-related proteins suppressed the phenotypes elicited by mutated SWI/SNF genes (Hirschhorn et al., 1992). Functional genetics have shown that the BAF complexes play important roles in lineage specification and development *in vivo*. For example, they act as transcriptional regulators in the development of T cells, (Chi et al., 2002; Wang et al., 2011), hepatocytes (Gresh et al., 2005), embryonic stem cells (Gao et al., 2008; Ho et al., 2009) and cardiac cells (Lickert et al., 2004). This is possible through the combinatorial assembly of the subunits into various complexes that can be expressed in a cell-type specific manner or targeted for different developmental programs (Kadoch et al., 2016, Review). For example, the BAF complex subunit BAF53a is replaced with BAF53b (which alters its gene regulatory function) in the shift from neural stem cells to post mitotic neurons. Moreover, BAF53b is specific to the neuron specific nBAF complex (Lessard et al., 2007). The two major human homologs of the BAF complexes are the BAF-A and PBAF complexes (Kwon et al., 1994, Wang et al., 1996). The BAF-A complex is referred to as the BAF complex from here on in this thesis for simplicity.

In vitro, only four core subunits are required to dissociate nucleosomes from the DNA on a chromatin template. These include the mutually exclusive ATPases SMARCA2 or SMARCA4 (BRG1) and core subunits SMARCB1, SMARCC1 and SMARCC2 that enhance catalytic activity (Phelan et al., 1999). The exact role of the other subunits is not very well understood, but as alluded to above, functional genetics and their perturbation in cancer show that these have important roles *in vivo.* One major possibility for their importance would be for functional specificity. Recruitment of these chromatin remodellers to the chromatin *via* specific protein-protein interactions, leading to the binding of additional transcriptional regulators, could define cell fate.

As a chromatin remodeller, the BAF complex plays an important role in regulation of gene expression and was originally thought to be associated mainly with the promoter, TSS and other 5' regions of genes (Shema-Yaacoby et al., 2013, Raab et al., 2015). However, recent reports have suggested that the BAF complex is targeted to enhancer regions playing a crucial role in enhancer-mediated gene regulation. This has been shown in various systems from mouse embryonic fibroblasts to colorectal cancer cells thereby implicating the BAF complex in widespread enhancer regulation (Mathur et al., 2017, Vierbuchen et al., 2017, Trizzino et al., 2017, Alver et al, 2017). Numerous ways of gene regulation by the BAF complex have been postulated. First, it can bind several transcription coactivator/repressor complexes or histone modifications (through its bromodomain containing subunits such as BRD9), thus regulating the chromatin structure at these sites. For example, ARID1A, a subunit of the BAF complex can bind to nuclear hormone receptors through its C-terminal

domain and stimulate their transcriptional activity (Wu and Roberts, 2013). Another study showed that the BAF complex is recruited to monoubiquitylated histone H2B (Shema-Yaacoby et al., 2013) and can regulate gene expression at these sites. Further, the BAF complex, after binding to the chromatin can recruit other regulatory factors and histone modifying enzymes and mediate promoter-enhancer interaction by facilitating looping of the chromatin (Wu & Roberts, 2013, Review).

In addition to regulation of gene expression, the BAF complex also has implications in other biological processes such as cell cycle progression, DNA repair and histone monoubiquitylation. For example, the BAF complex interacts with the RB protein to repress E2F responsive genes during cell cycle progression (Roberts and Orkin, 2004, Review) and has also been shown to interact directly with p53 and to facilitate p53 mediated transcription (Guan et al., 2011). Additionally, the BAF complex is also recruited to DNA double stranded breaks where it plays a role in repair (Wu and Roberts, 2013, Review) and has been shown to have E3 ubiquitin ligase function as well (Li et al., 2010). Thus, the BAF complex has many diverse functions and its deregulation is associated with tumorigenesis. Indeed, the BAF and PBAF complexes (common and exclusive subunits are shown in Figure 3) have garnered a lot of attention over the last few years due to the high frequency of mutations in their subunits in a wide range of cancers.

The most frequently occurring as well as the most widespread mutation is in the gene for ARID1A (AT-rich interactive domain containing protein 1A) (Kadoch et al.,2013). ARID1A is a subunit of only the BAF complex and is not essential for chromatin remodelling *in vitro*. However, its mutational rates in cancer and functional studies point towards its importance in gene regulation and have prompted much research in deciphering its relevance in cancer. While the next section describes the role of several BAF complex subunits in cancer, the rest of this thesis mainly focusses on ARID1A.

2.5 The BAF Complex and Cancer

Deregulation of the BAF complex in tumorigenesis was initially identified when inactivating mutations of SMARCB1 were found to be the sole drivers of highly aggressive rhabdoid tumors (Rorke et al., 1996). Mouse models with heterozygous deletions in Snf5 (the homolog of SMARCB1) developed highly aggressive and metastatic tumors, establishing the role of Snf5 as a tumor suppressor (Roberts et al., 2000, Klochelndler-Yeivin, 2000, Guidi et al., 2001). On the other hand, an oncogenic role of the BAF complex was described in synovial carcinoma in which a fusion of the SS18 subunit with the SSX gene drives tumorigenesis by directing oncogenic transcription by the BAF complex (Middeljans et al., 2012, Kadoch et al., 2013). Subsequently, genome- and exome-wide sequencing studies

revealed that subunits of the BAF complex are mutated in around 20% of all human cancers (Kadoch et al., 2013, Shain et al., 2013). These mutations span across several subunits of the BAF and PBAF complexes as well as across several cancer types (Figure 3). For example, PBRM1 is mutated in 40% of renal clear-cell carcinomas (Varela et al., 2011), ARID2 in 18% of hepatocellular carcinomas (Li et al., 2011), ARID1B in liver, breast and pancreatic cancer and SMARCA4 in 11% of non-small cell lung cancers (Imielinski et al., 2012). Among the BAF complex subunits, defects in ARID1A are the most widespread across cancer types as well as the most frequently occurring. Mutations in ARID1A are most often nonsense or frameshift mutations leading to nonsense mediated decay (Kadoch et al., 2016, Review). These mutations are by far the most recurrent across all types of cancer, making ARID1A the most studied tumor suppressor subunit of the BAF complex. It is known to be mutated in around 50% of ovarian clear-cell carcinomas (OCCC) (Jones et al., 2010, Wiegand et al., 2010) but also in subsets of several other cancers including 7-10% of colorectal cancers (Wu et al., 2014, Review) which is of particular interest in this project. However, despite intensive research, the molecular mechanisms by which ARID1A (as part of the BAF complex) regulates tumorigenesis remain quite elusive. Several studies have revealed that ARID1A-deficient cells become dependent on certain other factors or pathways to maintain tumorigenesis (explained in detail in the next section). Therefore, identifying these dependencies and targeting them in ARID1A-deficient cells presents a striking opportunity for developing therapies.

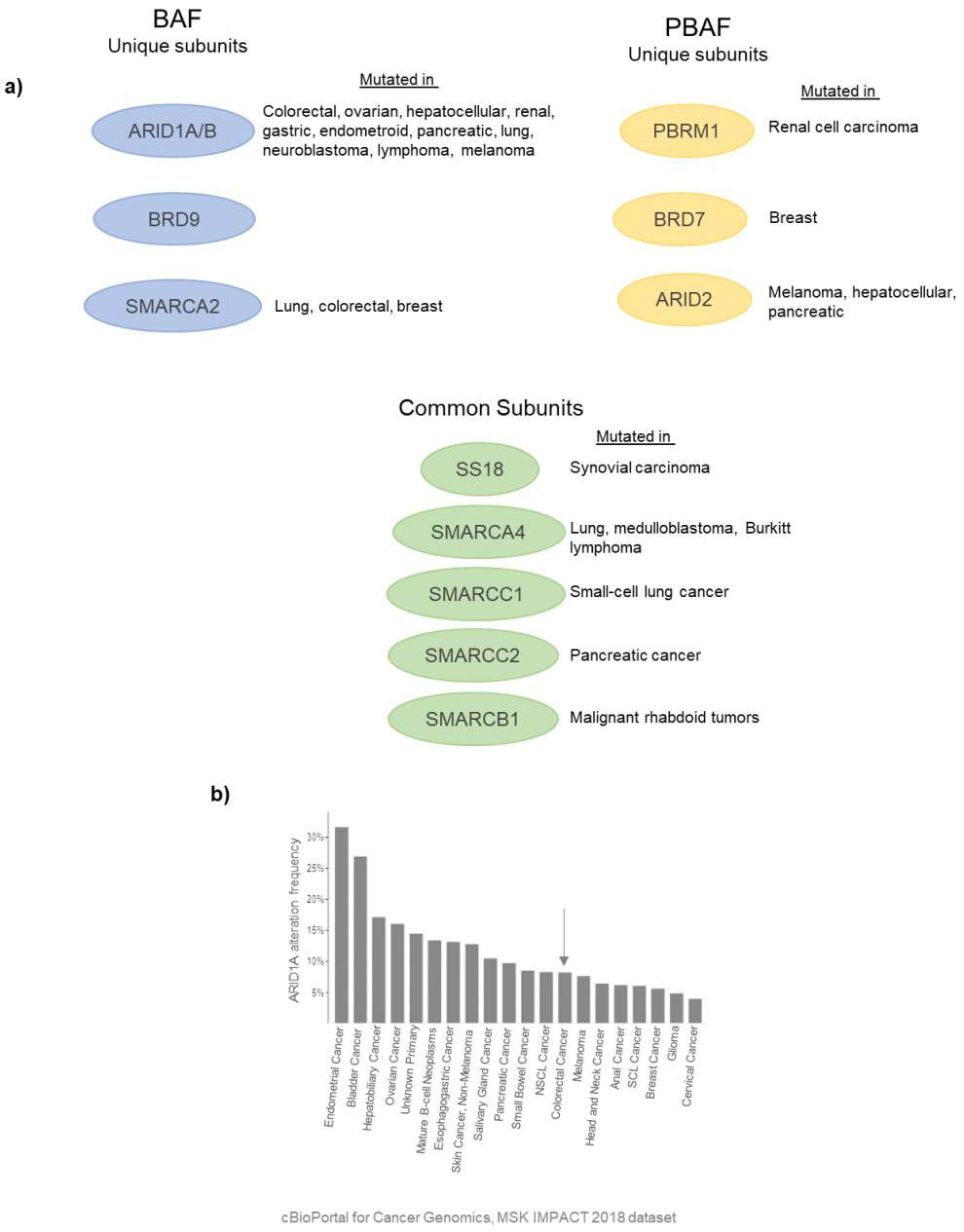


Figure 3: The BAF complex and cancer. The major BAF complexes, BAF-A(BAF) and PBAF, share many common subunits but also possess subunits that that are exclusive to each complex. Many of these subunits are mutated in a wide range of cancers as depicted in (a). The most frequently mutated subunit ARID1A is mutated across cancer types and in around 10% of colorectal cancers (b). A recent dataset, generated using the cBioPortal for Cancer Genomics tool (Gao et al., 2013, Cerami et al., 2012) shows the alteration frequency of ARID1A in various cancers including colorectal cancer.

2.6 Synthetic Lethality Reports Involving ARID1A

Many therapeutic approaches are based on inhibiting the function of oncogenic factors. Since ARID1A expression is lost in tumors, inhibiting its function is not appropriate. Moreover, as described above, ARID1A plays important roles in many pathways, and thus induction of synthetic lethality would be a suitable alternative approach. Synthetic lethality is defined as cell death caused by the deficiency of two dependent factors where, either deficiency alone does not cause the cell death. Therefore, in ARID1A-deficient cells, identifying these dependencies and targeting them presents a striking opportunity for developing therapy. Several studies have revealed that ARID1A-deficient cells become dependent on certain other factors or pathways to maintain tumorigenesis. However, there is limited understanding of the mechanisms underlying these relationships. An initially described vulnerability of ARID1A deficient cells was the residual BAF activity, specifically through ARID1B. Indeed, ARID1B depletion leads to impaired proliferation (Helming et al., 2014, Mathur et al., 2017). However, ARID1B is not druggable and therefore this vulnerability cannot currently be used clinically. Nonetheless, an indirect mechanism to target this vulnerability has been reported in OCCC cells in which inhibition of the BET proteins led to a downregulation of ARID1B and sensitized ARID1A-mutant cells to BET inhibition (Berns et al., 2018).

In OCCC, ARID1A mutations often coexist with PI3K/AKT pathway mutations and ARID1A-deficient tumor cells are dependent on constitutive activity of the PI3K pathway and so perhaps also sensitive to its inhibition. Indeed, ARID1A-deficient breast cancer and OCCC cell lines were shown to be sensitive to PI3K/AKT inhibition (Samartzis et al., 2014) uncovering a potential targeted therapy for tumors with ARID1A deficiency. Another study showed that ARID1A deficiency sensitizes *in vivo* and *in vitro* models to PARP inhibition (PARP stands for poly ADP ribose polymerase which plays a role in DNA repair) (Shen et al., 2015). This would be possible in tumors with ARID1A mutations which compromise its role in DNA repair pathways. This study also indicates that ARID1A interacts with the DNA damage checkpoint kinase ATR. In a screen for synthetic lethal targets for ATR inhibitors, ARID1A was found to be a candidate partner (Williamson et al., 2016). A separate drug screen identified Dasatinib which targets the Src/Abl kinases as particularly effective in ARID1A-deficient tumors (Miller et al., 2016). Furthermore, ARID1A loss has been linked to accumulation of oxidative stress in cells. Along with that, sensitivity to reactive oxygen species (ROS) inducing agents in ovarian and lung cancer has been described (Kwan et al., 2016). It has also been reported that ARID1A and wildtype p53 act in concert to regulate p53 target gene expression (Guan et al., 2011). The stabilization of wildtype p53 has been

suggested to rescue the growth of *ARID1A* mutant cells (Meijer et al., 2013). A study that linked ARID1A expression to driving the expression of the pro-inflammatory cytokine IL6 in OCCC, suggests that IL6 depletion therapy would be another such target (Chandler et al., 2015). Another repressive function of ARID1A is repression of the HDAC6 promoter. Upon loss of ARID1A, HDAC6 is overexpressed which deacetylates p53 and abrogates its pro-apoptotic functions, suggesting a heightened sensitivity of ARID1A-deficient cells to HDAC6 inhibition (Bitler et al., 2017).

A study in 2013 showed that SMARCB1-deficient rhabdoid tumors are sensitive to the inhibition of the histone methyl transferase EZH2 by a small molecule inhibitor called EPZ6438 (Knutson et al., 2013). However, it was not known whether this dependency on EZH2 for proliferation extends to other subunits of the BAF complex. Interestingly, a subsequent study revealed that EZH2 inhibition causes regression of ARID1A mutated OCCC tumors *in vivo* (Bitler et al., 2015) and this synthetic lethality was explained by the antagonistic relationship between the BAF complex and the PRC2 complex (of which EZH2 is a subunit). This synthetic lethal interaction was extensively studied, and resistance mechanisms were described in several cells. *ARID1A* and *KRAS* mutated cells seemed to be resistant to EZH2 inhibition (Kim et al., 2015). Most recently, several of these vulnerabilities were tested in colorectal cancer cell lines. While an enhanced sensitivity to ATR inhibition and PARP inhibition in *ARID1A*-deficient cells was found, it seemed that EZH2 inhibition was not effective in the HCT116 colorectal cancer system (it is KRAS mutant) (Wu et al., 2018). This study also described a new vulnerability found in their screen. This was the inhibition of AURKA, a player in cell cycle progression. Wu et al., 2018 proposed that, this was explained by yet another repressive role of ARID1A wherein AURKA levels were kept in check in the ARID1A-proficient condition by ARID1A.

Thus, it is evident that mutation status of ARID1A is being harnessed extensively to develop novel strategies for cancer therapies. However, as described later in this section 2.8, the role of ARID1A in various cancers is far from ubiquitous. Therefore, in order to discover effective therapeutic strategies, it is very important to consider the context in which ARID1A loss occurs before stratifying patients in clinical trials based on the many synthetic lethal targets identified.

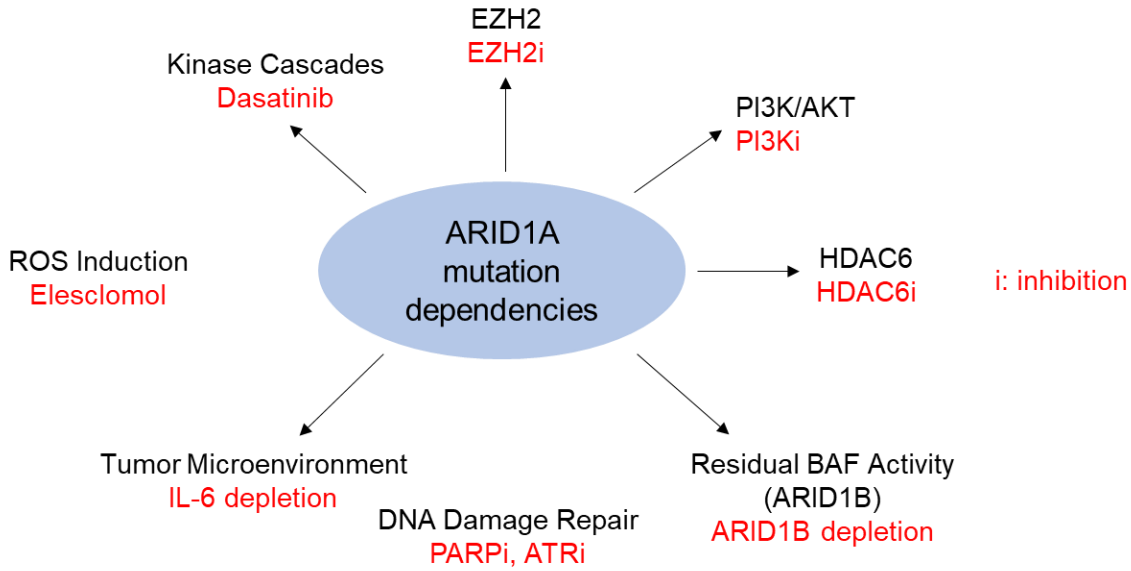


Figure 4: A summary of the ARID1A mutation dependent synthetic lethalities described in section 2.6.

2.7 Targeting EZH2 in ARID1A-Deficient Cancers and Competition with the PRC2 Complex

The synthetic lethality between ARID1A and EZH2 has been an important subject of research in the last few years. An antagonistic relationship between them was described in Drosophila by Kennison et al. in 1988. EZH2 (enhancer of zeste homolog 2) is the catalytic subunit of the Polycomb repressive complex (PRC2) which is associated with transcriptional silencing of genes. As a histone methyltransferase, EZH2 is responsible for the trimethylation of lysine 27 of histone H3 using the co-factor S-adenosyl methionine (SAM) and establishing repressive marks on the chromatin in some cases by recruiting DNA methyltransferases which methylate CpG nucleotides on DNA (Viré et al., 2005). Gain of function mutations for EZH2 have been observed in many cancers which lead to repression of Polycomb target genes (Bennett and Licht, 2017, Review). Perhaps the most well-known example of the antagonism between the BAF complex and PRC2 complex was demonstrated in several studies that revealed a stable repression of the tumor suppressor Ink4a/ARF locus in rhabdoid tumors. In these tumors *SMARCB1*-mutated BAF complexes were unable to oppose the gene repression patterns set by the PRC2 complex (Wilson et al., 2010). In a more recent study, it was shown that loss of SMARCB1 reduces the levels of the BAF complex at typical enhancers, but a small amount of residual BAF complexes can still bind to super-enhancers that drive oncogenesis (Wang et al., 2017). Therefore, Wang et al., 2016 propose that this imbalance between ARID1A and EZH2 blocks

differentiation and drives cancer. Similarly, Kadoch et al., provided evidence of the BAF complex directly evicting the PRC1 complex (another polycomb repressive complex), more efficiently in its SS18-SSX oncogenic form. It was also found that ARID1A-mutated OCCC cells were sensitive to EZH2 inhibition (Bitler et al., 2015). This effect was explained by the antagonistic relationship between ARID1A and EZH2 at ARID1A-EZH2 target genes. This study revealed PIK3IP1 (a negative regulator of the PI3K/AKT pathway) as the target gene associated with this synthetic lethality (Bitler et al., 2015). This can be explained by the fact that ARID1A-deficient tumors are no longer able to activate expression of PIK3IP1 (and thus it is repressed by EZH2) and the PI3K/AKT pathway is constitutively activated. As mentioned before, AKT pathway overexpression and ARID1A deficiency act synergistically to drive tumorigenesis. Inhibition of EZH2 in this case relieves the repression on the PIK3IP1 promoter and thus reduces proliferation. Therefore, it seems that mutations in the BAF complex subunits alter its targeting to the chromatin which leads the disruption of the balance between the BAF complex and PRC2 complex and can contribute to tumorigenesis.

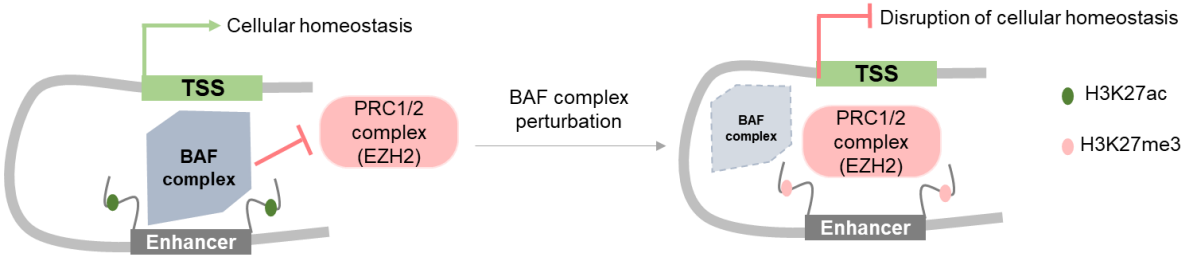


Figure 5: A model of the balance between the BAF complex and PRC2 complex. Under normal cellular circumstances, the BAF complex is able to efficiently oppose PRC2 from sites that regulate the expression of genes important in development, differentiation and tumor suppression (left). When the BAF complex is mutated, it is mis-targeted and the PRC2 can now take its place and suppress these regions by trimethylating H3K27. This loss of balance disrupts cellular homeostasis (right).

2.8 Mouse Models of *Arid1a*-driven Cancer

As described in section 2.5, early models of BAF complex deficient mice revealed the importance of this complex in driving cancer. *Snf5* knockout mice developed highly aggressive tumors, whereas *Brg1 and Brm* (SMARCA4 and SMARCA2) deficient mice showed some predisposition to tumor development (perhaps due to their redundant functions) (Bultman et al., 2000). One of the first mouse models which unravelled a context dependent role of the BAF complex subunit *Brg1* was described by von Figura et al., 2014. The loss of *Brg1* promoted *Kras*-driven neoplastic transformation in the pancreatic duct cells, however its loss also inhibited the progression of *Kras*-driven pre-cancerous lesions

in acinar cells. This already suggested that even though BAF complex component expression is lost in tumors, this has very context-dependent effects.

With the acceptance of ARID1A as a tumor suppressor, several research groups tried to model *ARID1A*-deficient cancer in mice. While these models confirmed its tumor suppressive role in various cancers, some striking observations were made which indicated that ARID1A could also play an oncogenic role. In the model for ovarian cancer Guan et al., 2011 conditionally knocked out *Arid1a* from the mouse ovarian epithelium. As mentioned before, in ovarian and endometroid tumors, mutations in *ARID1A* often co-occur with *PIK3CA* or *PTEN* mutations and are therefore dependent on an increased PI3K-signaling. Therefore, Guan et al., in addition to their *Arid1a* knockout model, also created double knockouts for *Arid1a* and *Pten*. Interestingly, they observed that the loss of either gene alone did not cause any lesions in the ovarian epithelium. Rather, a majority of the double knockouts developed undifferentiated tumors. Similarly, Chandler et al., showed that the simultaneous deletion of *Arid1a* and activation of *Pik3ca* also led to the formation of ovarian clear-cell carcinoma (OCCC) in mice. Zhai et al., 2014, analysed human ovarian endometroid cancer (OEC) in which they found that *ARID1A* mutations co-occur with activating mutations of the PI3K/AKT and Wnt signaling pathways. They tested the loss of *Arid1a* in a *Pten* and *Apc* deleted background. While *Pten/Apc* deletion led to the formation of poorly differentiated OECs in mice, the added deletion of *Arid1a* caused the tumor cells to undergo differentiation and attain a more epithelial phenotype. These models clearly suggest both oncogenic and tumor suppressive roles for *Arid1a* in OCCC and OEC depending on the mutational background in which its loss occurs.

Sun et al., 2017, modelled *Arid1a* loss in liver cancer showing that it has oncogenic and tumor suppressive roles in a time and context dependent manner. While the deletion of *Arid1a* from otherwise wildtype mice protects against DEN (diethylnitrosamine) and carbon tetrachloride induced hepatocellular carcinoma (HCC), overexpression of *Arid1a* accelerated tumor initiation through a mechanism that enhanced oxidative stress in the liver. This was consistent with findings that showed a negative correlation between *Arid1a* expression and survival in HCC patients (Uhlen et al., 2017). Further, Sun et al., deleted *Arid1a* from already existent tumors revealing that a loss of *Arid1a* at this stage leads to enhanced metastasis. Therefore, in the context of late loss, Arid1a is tumor suppressive. In a third model, for pancreatic cancer, Livshits et al., 2018 showed that the loss of Arid1a, in the context of the oncogenic program driven by *Kras* mutation, significantly enhances the formation of pre-cancerous lesions. However, its re-expression in the lesions does not rescue the phenotype suggesting a time specific role for Arid1a loss. Moreover, they also

revealed that the concomitant embryonic loss of Arid1a and mutation of *Kras* does not lead to the formation of lesions, again suggesting the need for pre-existing *Kras* activation.

Lastly, in a striking model of colorectal cancer (CRC), Mathur et al., 2017, showed that the deletion of *Arid1a* alone, sporadically from the body was enough to initiate the formation of invasive adenocarcinomas originating in the colon. These tumors resembled a human subtype of CRC called MSI-positive quite strongly, both in terms of mucinous phenotype and immune cell infiltration. *Arid1a* is one of the most significantly mutated chromatin regulators in CRC (cBioportal for Cancer Genomics, Gao et al., 2013, Cerrami et al., 2012) and this study presented a way to model human CRC. Interestingly, however, the same study, also modelled CRC by deleting *Arid1a* in the background of a very common tumor suppressor *Apc*. *Apc* mutations occur in about 80% of human CRCs and Apc^{Min} mice, which have a heterozygously mutated *Apc* allele, usually develop polyps in their intestine. Unexpectedly, while polyps were found in the Apc^{Min} mice, when *Arid1a* was deleted from these mice, fewer tumors were detected. Any tumors that were detected escaped *Arid1a* deletion and retained its expression suggesting that *Arid1a* was required for *Apc* mutation mediated CRC.

These models suggest that while *Arid1a* is tumor suppressive in many cases, it can also be oncogenic. Thus, it is very important to consider the context in which the loss of ARID1A occurs in order to better understand the mechanisms that drive tumorigenesis to ultimately design precision therapies.

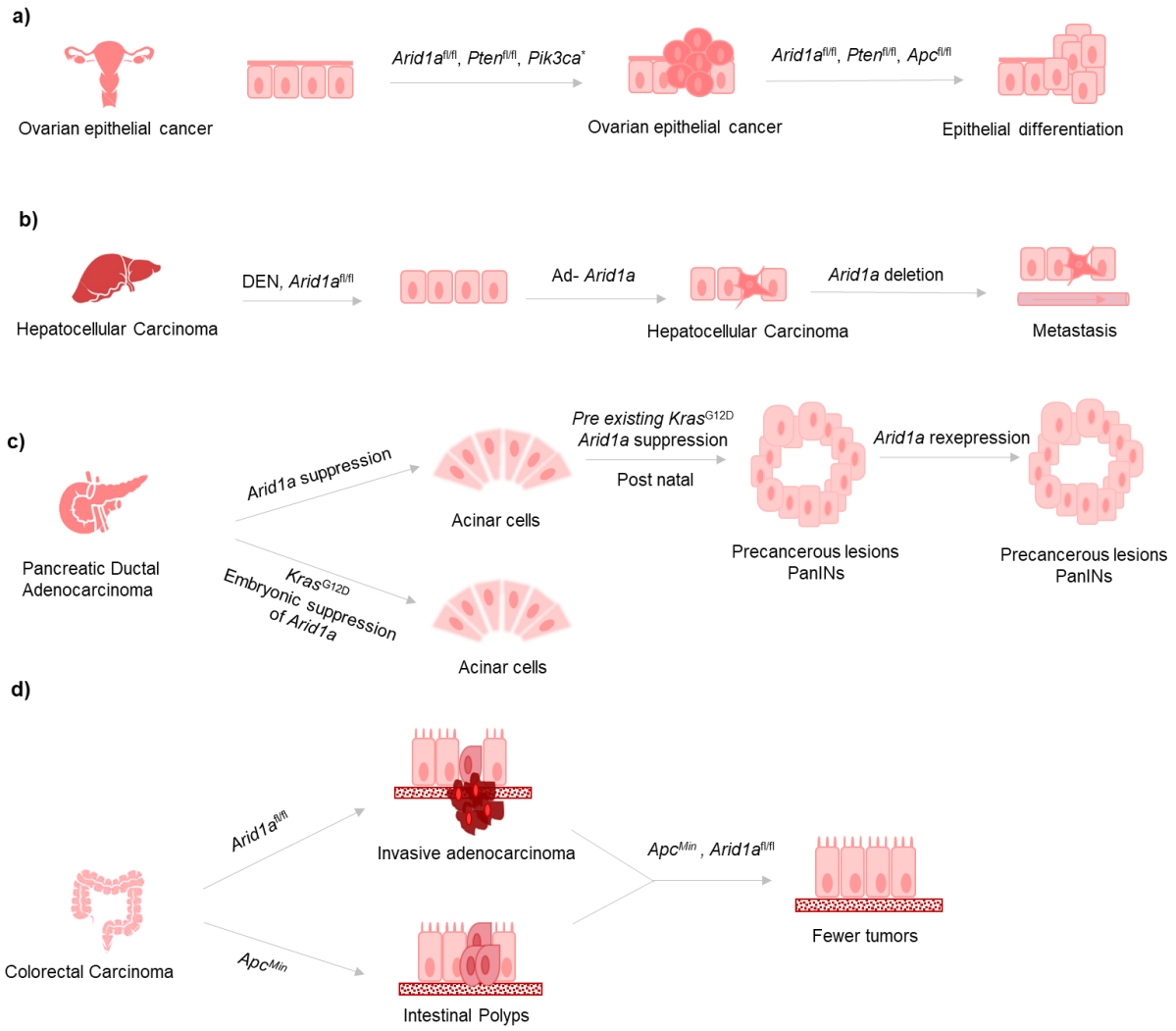


Figure 6: Mouse models describing the loss of ARID1A. More recent literature suggests that the loss of ARID1A can be either tumor suppressive or oncogenic depending on the context in which it occurs

a) In the case of ovarian epithelial cancer, deletion of *Arid1a* in the commonly used *Pten* $^{fl/fl}$ background leads to the formation of tumors, however if a *Apc* deletion also occurs in the same model, these tumors show epithelial differentiation and these mice show longer survival.
b) In hepatocellular carcinoma models induced by Diethylnitrosamine (DEN) *Arid1a* loss has a protective role and overexpression of *Arid1a* leads to the formation of tumors. Further in the context of late tumor stage deletion of *Arid1a*, metastasis occurred.
c) In the pancreas, the suppression of *Arid1a* alone is not tumorigenic, however in the context of *Kras* G12D mutation, it significantly enhances the formation of precancerous lesions called pancreatic intraepithelial neoplasia (PanINs). *Arid1a* has a very time specific role in this case, as its re-expression in this system does not revert the severity of the PanIN lesions. Also, the suppression of *Arid1a* simultaneously with *Kras* mutations in the embryonic stage is not tumorigenic.
d) In colorectal cancer *Arid1a* plays a pivotal role wherein its inactivation alone leads to the formation of invasive adenocarcinomas. However, this inactivation in the background of *Apc* inactivation leads to fewer tumors than either deletion alone.

2.9 The Genetic and Epigenetic Basis of Colorectal Cancer

ARID1A is mutated in around 10% of colorectal cancers and is among the most frequently mutated chromatin regulators in this type of cancer (cBioPortal for Cancer Genomics, Gao et al., 2013, Cerrami et al., 2102). Moreover, a striking mouse model exhibited the importance of ARID1A in colorectal cancer formation (Mathur et al., 2017) and so the aim of our study was to explore further the role of ARID1A in colorectal cancer. Colorectal cancer (CRC) is the third most commonly occurring cancer, which globally represents 10% of all cancer cases (World Cancer Report, WHO, 2014). This arises from a progressive accumulation of alterations in genetic and epigenetic mechanisms that drive the normal intestinal epithelium to form benign adenomas, which ultimately lead to malignant CRC. Initial mutations in genes like *APC* (the most frequently mutated gene in CRC occurring in about 80% of CRCs) can be inherited or occur sporadically (Kinzler and Vogelstein, 1996). Defective APC leads to a loss of control of the Wnt-signaling pathway, which controls the proliferation of the intestinal stem cells (Reya and Clevers, 2005, Review). After the initiation of adenoma formation, its development is supported by several other mutations which play roles at different stages of progression towards colorectal carcinoma. For instance, *KRAS* mutations, which occur in around 40% of CRCs, facilitate the initial stages of adenoma-carcinoma transition by activating the MEK/ERK pathway (Fearon, 2011, Review). Subsequently, mutation in TP53 (which acts as a checkpoint in cell cycle progression by monitoring DNA damage) allows the proliferation of cells with even highly damaged DNA (Vousden and Prives, 2009, Review). Mutations that affect the TGFβ pathway (such as its downstream effector *SMAD4*) and PI3K signaling pathway (such as its negative regulator *PTEN*) are also commonly found (Fearon, 2011, Review). Finally, mutation of genes involved in metastasis complete the malignant transformation (Fearon, 2011, Review). This sequential accumulation of mutations was proposed by Fearon and Vogelstein in 1990. More recently, though, it has been proposed that these mutations do not occur sequentially, rather as more independent pathways of driving CRC, giving rise to molecular subtypes of CRC (Issa, 2008).

In addition to the mutations described above, commonly occurring mutations in the gene *MLH1* (which plays a role in the DNA mismatch repair pathway) can lead to hypermutability (a phenotype known as microsatellite instability or MSI), which may affect crucial oncogenes or tumor suppressor genes like the ones mentioned above (Fearon, 2011). This also gives rise to the CpG island methylator phenotype (CIMP) subtype of CRC (Toyota et al., 1999). Methylated DNA can be used as a biomarker for CRC in precancerous lesions (Okugawa et al., 2015). In fact, interestingly, immunohistochemical studies of human CRC samples

revealed ARID1A is lost in early adenocarcinomas and is associated with microsatellite instability (Lee et al., 2016). Moreover, a subset of CRCs also shows chromosomal instability (CIN) leading to aneuploidy (Lengauer et al., 1997). Thus, it is evident that CRC is a heterogenous disease defined by diverse genetic and epigenetic alterations. Conventionally, CRC has been treated by surgery, radiotherapy and chemotherapy based on nucleoside depletion (5 FU, Folinic acid) and inhibition of DNA synthesis (oxaliplatin, irinotecan) (Gustavsson et al., 2015, Review). However, the definition of pathological and molecular subtypes and subsets of CRC with specific mutations would present an opportunity for targeted therapy. Several targeted therapies have already been developed and tested in clinical trials, individually and in combination with conventional therapy, over the past few years (https://www.clinicaltrials.gov/).

Most recently, genome wide studies have revealed that epigenetic factors such as chromatin remodellers are mutated in a subset of colorectal cancers. At which stage during tumor progression these mutations are important (or whether they are bystander mutations) is unclear. However, as shown by the mouse models of *Arid1a*-deficient CRC, it seems *Arid1a* loss drives CRC in a pathway independent from *Apc*-mutation driven CRC and therefore its inactivation must be an early driver event (Mathur et al, 2017). The potential of uncovering novel therapeutic targets and prognostic biomarkers by exploring the role of ARID1A in colorectal cancer made it an attractive target to study.

2.10 The Wnt and MEK/ERK Signaling Pathways

As elucidated by the mouse models described in section 2.8, the presence of the commonly occurring *KRAS* and *APC* mutations seem to play interesting roles in ARID1A-mediated pancreatic and colorectal cancer. It is therefore important to understand the signaling cascades in which they participate. In the absence of a Wnt ligand, the transcription factor β-catenin is phosphorylated by the destruction complex (which consists of APC) and targeted for proteasomal degradation. However, in the presence of a Wnt ligand, the destruction complex is sequestered to the Wnt receptor and β-catenin is no longer degraded, allowing it to accumulate and translocate to the nucleus. In the nucleus it acts in concert with the TCF family of transcription factors to activate the expression of genes involved in proliferation and. The Wnt pathway is a major signaling cascade that drives the self-renewal of the intestinal stem cells (Clevers and Nusse, 2012, Review). Therefore, the mutations of APC that inactivate it are important drivers of CRC as this allows uncontrolled Wnt signaling to occur in the absence of regulation of β-catenin.

The mitogen activated protein kinase (MAPK) signaling pathways are a group of cascades that relay extracellular stimuli (such as growth factors, cytokine and cellular stresses) to the intracellular transcriptional machinery. Generally, receptor tyrosine kinases are auto phosphorylated on receiving a stimulus. The phosphorylated tyrosines act as docking sites for adaptor proteins such as GRB2, which in turn recruit the guanine nucleotide exchange factor (GEF) SOS. SOS can exchange the GDP on the monomeric G-protein Ras for GTP, thus activating it. Ras is coded by three highly homologous genes, *KRAS*, *NRAS* and *HRAS*, and while all three are found to be mutated in cancer, *KRAS* is most significantly studied. The activation of Ras initiates the cascade of phosphorylation in which, Ras activates Raf (a MAPK kinase kinase), which phosphorylates and activates MEK (MAPK kinase) which then phosphorylates and activates ERK (MAPK). Activation of the kinases usually occurs through phosphorylation of serine and threonine residues in the activating loops of these proteins. Two other MAPK pathways, the JNK (c-Jun N-terminal kinases) and the p38MAPK pathway also act in response to extracellular signals and employ these proteins as MAP kinases in their cascades. While these three operate through separate modules maintained by scaffold proteins, some degree of crosstalk occurs to integrate and amplify the signal. Negative regulation of these factors is carried by feedback loop mechanisms and GTPase activating proteins (GAPs) which attenuate the hydrolysis of GTP bound to Ras thereby inactivating it. Point mutations in *RAS* (as the ones present in CRC) prevent this activity, thereby remaining in a constitutively GTP bound and active state (Whitmarsh, 1996, Review).

The MEK/ERK pathway has been the most extensively studied in cancer. Upon phosphorylation ERK has been reported to accumulate in the nucleus (Chen et al., 1992), where it can phosphorylate and activate many substrates. These targets include other protein kinases (such as RSK), cytoskeletal proteins and transcription factors among others, and phosphorylation occurs *via* interaction with conserved docking sites (Sharrocks et al., 2000). For example, JunD, a transcription factor contains both a D domain and DEF domain which are required for recognition and phosphorylation by ERK (Vinciguerra et al., 2003). JunD is a member of the AP1 transcription factor family which consist of the Fos and Jun proteins. These are DNA binding proteins which contain leucine zippers that can bind to DNA through their bZIP domain. Upon activation, these factors homo- or hetero-dimerize and bind to the DNA directly, recruit further transcriptional regulators and modulate gene expression (Wagner, 2001, Review). A clear example of this is the induction of the AP1 factors themselves. For example, in colorectal cancer cell lines, the activity of the MEK/ERK pathway induces the expression of cJun (which is one of the immediate early genes of this pathway in many systems) and FosL1 (Vial et al., 2003). The activation of this cascade

ultimately leads to the transcription of genes involved in cell growth, proliferation and differentiation (Yoon and Seeger, 2006). Due to the central role of this pathway in crucial processes, it is perturbed in many diseases including cancer. As mentioned before, this perturbation occurs mostly through the activating mutations in *KRAS*, however, mutations in *BRAF* are also very common (cBioPortal for Cancer Genomics). Therefore, several pharmaceutical companies have used this opportunity to design inhibitors against these kinases. The most well-known example is the use of the BRAF inhibitor for the treatment of melanoma. Several MEK/ERK inhibitors have also been developed and are in numerous clinical trials (https://www.clinicaltrials.gov/).

FosL1 and JunD are the most highly expressed AP1 factors in the cell line primarily used in this study. The above description is an extremely simplified summary of the MEK/ERK pathway which mentions the factors relevant to this study. It is important to keep in mind that this pathway integrates a variety of extracellular signals to elicit a variety of intracellular responses, and hundreds of proteins are involved in this process.

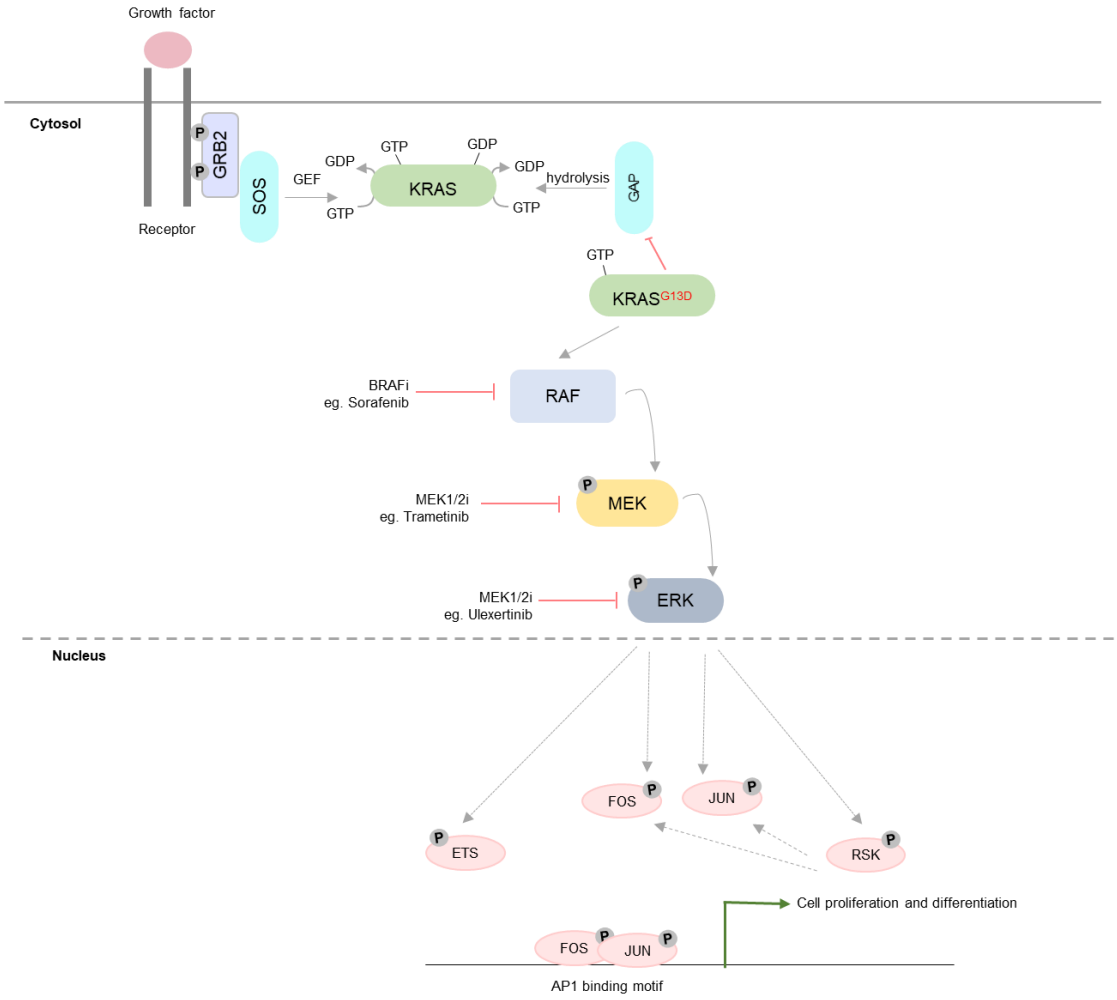


Figure 7: The MEK/ERK Pathway. Upon receiving a stimulus, the receptor tyrosine kinases, undergo autophosphorylation. These sites act as docking sites for the adaptor protein GRB2 which recruits the guanine nucleotide exchange factor (GEF) SOS. SOS exchanges GDP for GTP on RAS, thus activating it and initiating the kinase cascade. The negative regulation of RAS is maintained in part by GAPs (GTPase activating protein) which enhance the hydrolysis of GTP on Ras, thus, inactivating it. In the case of Ras mutations, this activity is inhibited, and it remains constitutively active. Active Ras activates Raf which phosphorylates MEK, which in turn phosphorylates ERK. Phosphorylated ERK has many targets in the cytosol and nucleus. Among these are the Ribosome s6 kinases (RSK) which phosphorylate and activate further proteins such as the AP1 transcription factors. The activation of the AP1 factors allows their dimerization and binding on DNA sequences on the genome where they modulate gene expression. The involvement of this pathway in many diseases including cancer has led to the development of several inhibitors that target different stages of the cascade.

2.11 ARID1A, AP1 and Enhancers

The previous sections illustrated the importance of ARID1A and AP1 factors downstream of the MEK/ERK pathway in colorectal cancer. However, the mechanisms by which they modulate gene expression are still quite unclear. Recent literature and advances in chromatin immunoprecipitation of BAF complex subunits have shed considerable light on the occupancy of these complexes on the genome and their role at these regions. While initial studies reported that the ARID1A is bound at promoter regions (Raab et al., 2015), several studies have reported that the BAF complex is targeted to enhancers. These include reports in erythrocytes (Hu et al., 2011) and oligodendrocyte differentiation (Yu et al., 2013). More recently, enhancer mediated gene regulation has been implicated in the tumor suppressive functions of the BAF complex. In colorectal cancer, it was shown that ARID1A deficient cells lose activity of BAF occupied enhancers. The H3K27ac mark from these enhancers is lost upon ARID1A deletion and expression of the target genes is also downregulated (Mathur et al., 2017). This was also shown in endometroid cancer cell lines (Lakshmikrishnan et al., 2017)

Similarly, in *Smarcb1* deleted mouse embryonic fibroblasts, H3K27ac is lost globally from enhancers. Smarcb1 re-expression leads to increased levels of H3K27ac, p300, BRD4 and mediator in the chromatin fraction suggesting the increased activity of enhancers. (Alver et al. 2017). Very interestingly, in both studies, enhancers that lose H3K27ac are associated with AP1 binding motifs. Intriguingly, a report which studied the role of lineage determining transcription factors in defining the differentiation of mouse embryonic fibroblasts (MEFs) also found a link between the BAF complex and AP1 transcription factors (Vierbuchen et al., 2017). AP1 motifs are very short DNA sequences of about 6-10 nucleotides and are therefore present in large numbers on the genome. AP1 factors have been extensively implicated in enhancer selection during development and differentiation (Madrigal and Alasoo, 2018, Review). Vierbuchen et al., 2017, showed that AP1 TFs were important in

the response to the differentiation stimulus in MEFs, and that BAF binding at these enhancers increased upon stimulation. *In silico*, they were able to show that a loss of AP1 binding sites leads to a loss of BAF binding. At the same time, Trizzino et al., showed that there was a high correlation between ARID1A, H3K27ac and ATC-seq (which assesses chromatin accessibility) data suggesting again that the BAF complex is localized at enhancers. They also found an AP1 binding motif at these enhancers and that the loss of ARID1A does not affect chromatin accessibility at regions strongly bound by ARID1A. This was contrary to the study in the colorectal cancer cell line HCT116 which implied that the loss of ARID1A caused a loss of accessibility and decreased enhancer activity (Kelso et al., 2017).

Thus, it is evident that the BAF complex together with AP1 transcription factors and perhaps other co-regulators, plays an important role in enhancer-mediated gene regulation in differentiation and tumorigenesis. It is, however, important to note that conclusions about ARID1A in several of these reports in colorectal cancer were made based on SMARCA4 and SMARCC1 occupancy due to the lack of chromatin immunoprecipitation data for ARID1A. Thus, ARID1A as a mediator of enhancer-driven regulation and a highly recurrent mutated gene in cancer, presents a very attractive target to explore oncogenic programs driven by enhancers.

2.12 Aims of this Project

Colorectal cancer is the third most commonly occurring cancer and represents 10% of cancers cases. CRC incidence and mortality rates are rapidly rising in many middle and low-income countries, whereas in developed countries (where the occurrence of CRC is the highest), decreasing trends have been observed. By 2030, 2.2 million new cases and 1.1 million deaths are predicted to occur (Arnold et al., 2017). The prevention and treatment of CRC has improved with the advancements of screening methods such colonoscopy and development of novel therapies. However, it is very important to study the underlying mechanisms that drive this disease in order to develop mechanism-based targeted therapy. ARID1A is one of the most frequently mutated chromatin regulators in colorectal cancer. Given the differential roles of ARID1A in driving tumorigenesis in different cancers and its crucial regulatory role at enhancers, our primary goal was to focus on elucidating the mechanisms by which ARID1A controls transcription in colorectal cancer. Moreover, the mouse model created by its deletion seems to point towards a driver role for ARID1A loss in colorectal cancer. One of the major aims of our study was to determine the genome-wide occupancy of ARID1A in the colorectal cancer setting. Subsequently, based on that knowledge, we wanted to compare these results with the existing studies to draw

conclusions about enhancer-mediated gene regulation by ARID1A and the implications of its loss in colorectal cancer, to be better equipped to design potential therapies.

2.13 Preliminary Results

One of the major aims of this study was to study the effect of ARID1A loss in colorectal cancer in the context of other common perturbations. To select the most appropriate cell lines for our study, we made use of the mutation data available on cBioPortal for Cancer Genomics, more specifically, the information available in the Cancer Cell Line Encyclopedia (CCLE) dataset. This is an extensive characterization of 947 cancer cell lines where mutation data is derived from parallel sequencing of greater than 1600 genes and mass spectrometric genotyping (Barretina et al., 2012). Initially, we wanted to characterize the CRC cell lines for the presence of BAF complex subunits Figure 8(a,b). These figures were also presented in my master's thesis in 2016, however, these are very relevant to the additional experiments in this study as well and are further discussed in subsequent sections.

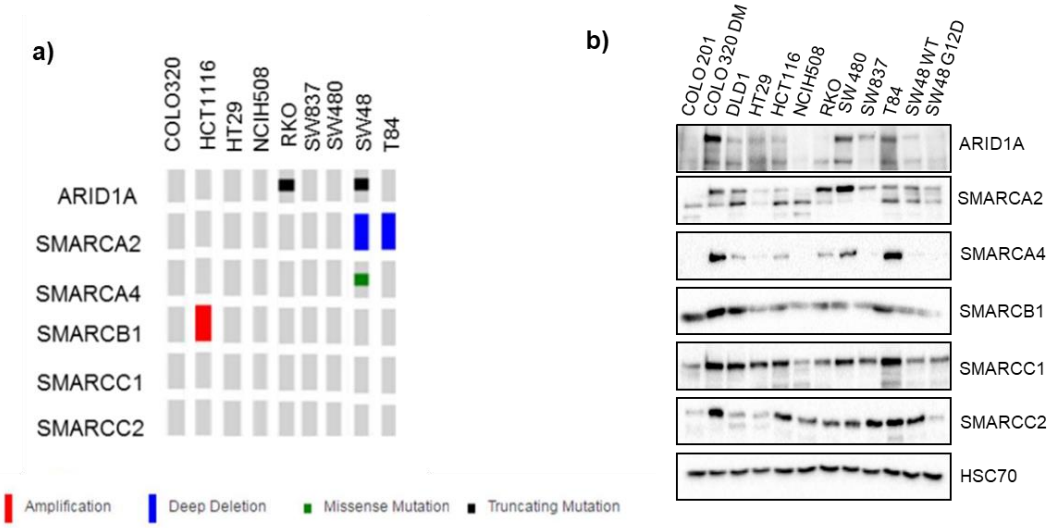


Figure 8: Preliminary data, linked to Figure 11 in section 5.2. The genetic status of various BAF complex subunits in several colorectal cancer cell lines generated using the cBioPortal for Cancer Genomics database (a) (from master's thesis, Madhobi Sen, 2016). The protein expression levels of these components in 12 different colorectal cancer cell lines determined by western blot (b). HSC70 was used as a loading control (from master's thesis, Madhobi Sen, 2016).

3.Materials

3.1 Equipment

Table 1: List of equiptment

Agarose gel chamber	Apogee Electrophoresis, Baltimore, USA
Balance	Sartorius AG, Göttingen, Germany
Bead bath “isotemp”	Fisher Scientific, Waltham, USA
Bioanalyzer 2100	Agilent Technologies, Santa Clara, USA
Biological Safety Cabinet “Safe 2020”	Thermo Fisher Scientific, Waltham, USA
Bioruptor® Pico sonication device	Diagenode, Seraing, Belgium
Bioruptor® Plus sonication device	Diagenode, Seraing, Belgium
Celigo S Cell Imaging Cytometer	Nexcelom Bioscience LLC, Lawrence, USA
Centrifuge (Megafuge 8R)	Thermo Fisher Scientific, Waltham, USA
Centrifuge 4°C (5417R)	Eppendorf AG, Hamburg, Germany
CFX Connect™ Real-Time PCR Detection System	Bio-Rad Laboratories GmbH, Munich, Germany
ChemiDocTM MP System	Bio-Rad Laboratories GmbH, Munich, Germany
Counting chamber (Neubauer)	Brand GmbH & Co. KG, Wertheim, Germany
Freezer -150°C	Ewald Innovationstechnik, Bad Nenndorf, Germany
Freezer -20°C	Liebherr GmbH, Biberach, Germany
Freezer -80°C “New Brunswick™Innova®”	Eppendorf GmbH, Wesseling- Berzdorf, Germany
Gel Imager “Gel iX imager”	Intas Science Imaging, Göttingen, Germany
Ice-machine B100	Ziegra, Isernhagen, Germany
Incubator (cell culture) “Hera cell 150i”	Thermo Fisher, Waltham, USA
Inverted Routine Microscope “Eclipse TS100”	Nikon GmbH, Dűsseldorf, Gemany
Liquid nitrogen tank LS4800	Worthington Industries, Theodore, USA
Magnetic stirrer “IKA®RCT-basic”	IKA®-Werke GmbH & Co. KG, Staufen, Germany
Mini-PROTEAN Tetra Cell Electrophoresis and electroblotting unit	Bio-Rad Laboratories GmbH, Munich, Germany
Mr. Frosty® Cryo Freezing Container	Thermo Fisher Scientific, Waltham, USA
Nanodrop DS-11 “Spectrophotometer”	DeNovix Inc., Wilmington, USA
NucleofectorTM 2b electroporation device	Lonza Bioscience, Switzerland
Personal Computer OPTIPLEX 7020	Dell, Round Rock, USA
pH-meter “WTW-720” InoLab® Series	WTW GmbH, Weilheim, Germany
Pipette Aid® portable XP	Drummond Scientific Co., Broomall, USA
Pipettes “Research plus” Series	Eppendorf AG, Hamburg, Germany
PowerPac™ HC High-Current Power Supply	Bio-Rad Laboratories GmbH, Munich, Germany
Qubit® 2.0 Fluorometer	Invitrogen GmbH, Karlsruhe, Germany
Refrigerator, 4°C	Liebherr GmbH, Biberach, Germany
Roller mixer	A.Hartenstein GmbH,Würzburg, Germany
Scanner (Epson Perfection V700)	Seiko Epson Corporation; Nagano, Japan

Shaker	A.Hartenstein GmbH,Würzburg, Germany
T100™ Thermal Cycler	Bio-Rad Laboratories GmbH, Munich, Germany
ThermomixerC	Eppendorf AG, Hamburg, Germany
Vacuum pump	Vacuubrand GmbH + Co Kg, Wertheim, Germany
Vortex mixer	Scientific Industries, Inc., Bohemia, USA

3.2 Consumables

Table 2: List of consumables

96-well Multiplate PCR plate white	Bio-Rad Laboratories GmbH, Munich, Germany
Amersham™ Protran™ 0.45 µM nitrocellulose Transfer Membrane	GE Healthcare Europe GmbH, Munich, Germany
Cell scraper (16 cm, 25 cm)	Sarstedt AG & Co., Nümbrecht, Germany
Cellstar 6- 12- and 24 well cell culture plate	Greiner Bio-One GmbH, Frickenhausen, Germany
Cellstar tissue culture dish 100×20 mm	Greiner Bio-One GmbH, Frickenhausen, Germany
Cellstar tissue culture dish 145×20 mm	Greiner Bio-One GmbH, Frickenhausen, Germany
Cellstar tubes, 15mL and 50 mL	Greiner Bio-One GmbH, Frickenhausen, Germany
Cryo Tube™ Vial (1.8 mL)	Thermo Fisher Scientific, Waltham, USA
Disposable Safety Scalpel	FEATHER Safety Razor Co., Osaka, Japan
FACS tube with cell strainer cap (12 x 75 mm)	BD Biosciences, Bedford, USA
Gel blotting paper (Whatman)	Sartorius AG, Göttingen, Germany
Injekt-F Syringes (1 mL)	B. Braun, Melsungen, Germany
Microtube 1.5 mL, 2 mL	Sarstedt AG & Co., Nümbrecht, Germany
Multiply PCR Microtube strip (8 x 0.2 mL)	Sarstedt AG & Co., Nümbrecht, Germany
NORM-JECT Syringes of different volumes	Henke Sass Wolf GmbH, Tuttlingen, Germany
Pipette filter tips	Sarstedt AG & Co., Nümbrecht , Germany
Pipette tips	Greiner Bio-One GmbH, Frickenhausen, Germany
Pipettes, serological	Sarstedt AG & Co., Nümbrecht, Germany
Reaction tubes (1.5 mL, 2 mL)	Sarstedt AG & Co., Nümbrecht, Germany
Sealing tape for qPCR plates	Bio-Rad Laboratories, Hercules, USA
Syringe filter, 0.2µm	Sartorius AG, Göttingen, Germany

3.3 Chemicals and Reagents

Table 3: List of chemicals, reagents and cell culture reagents

Acetic acid	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Agarose	GeneOn GmbH, Ludwigshafen am Rhein, Germany
Albumin Fraction V (BSA)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Ammonium persulfate (APS)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Ammonium sulfate $(NH_4)_2SO_4$	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Aprotinin	Carl Roth GmbH & Co. KG, Karlsruhe, Germany

β-Glycerolphosphate (BGP)	Sigma-Aldrich Co., St. Louis, USA
Bromophenol blue	Sigma-Aldrich Co., St. Louis, USA
Calcium Chloride ($CaCl_2$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Chloroform	Merck Millipore, Darmstadt, Germany
Colorless co-precipitant	Ambion, Waltham, USA
Co-precipitant Pink	Bioline, Luckenwalde, Germany
Crystal violet	Merck Millipore, Darmstadt, Germany
Diethylpyrocarbonate (DEPC)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Dimethyl sulfoxide (DMSO)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
di-Sodium hydrogen phosphate	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
DL-Dithiothreitol (DTT), BioUltra, ≥99.0%	Sigma-Aldrich Co., St. Louis, USA
DMEM	GIBCO®, Invitrogen GmbH, Darmstadt, Germany
DMEM/F12	GIBCO®, Invitrogen GmbH, Darmstadt, Germany
dNTPs	Jena Bioscience GmbH, Jena, Germany
Disuccinimidyl glutarate (DSG)	Thermo Fisher Scientific, Waltham, USA
ethylene glycol bis(succinimidyl succinate) (EGS)	Thermo Fisher Scientific, Waltham, USA
Ethanol absolute	Merck Millipore, Darmstadt, Germany
Ethylenediaminetetraacetic acid (EDTA)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Fetal Bovine Serum (FBS)	Thermo Scientific HyClone, Logan, USA
Formaldehyde (37%)	Sigma-Aldrich Co., St. Louis, USA
Glutaraldehyde (25%)	Sigma-Aldrich Co., St. Louis, USA
Glycerol	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Glycine	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Guaiac resin	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
HD Green® DNA stain	Intas Science Imaging GmbH, Göttingen, Germany
Hydrochloric acid (HCl)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Hydrogen peroxide solution (H_2O_2), 30%	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Iodacetamide	Sigma-Aldrich Co., St. Louis, USA
Isopropanol	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Leupeptin	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Lithium chloride (LiCl), 8M	Sigma-Aldrich Co., St. Louis, USA
Magnesium chloride ($MgCl_2$)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
McCoy's 5A(modified)	GIBCO®, Invitrogen GmbH, Darmstadt ,Germany
MEM	GIBCO®, Invitrogen GmbH, Darmstadt,Germany
Methanol	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
N-ethylmaleimide (NEM)	Sigma-Aldrich Co., St. Louis, USA
Nickel chloride ($NiCl_2$)	Sigma-Aldrich Co., St. Louis, USA
N-Lauryl sarcosine	Sigma-Aldrich Co., St. Louis, USA
Nonidet™ P40 (NP-40)	Sigma-Aldrich Co., St. Louis, USA
Opti-MEM	GIBCO®, Invitrogen GmbH, Darmstadt , Germany
Paraformaldehyde (16%)	Electron microscopy sciences, Hatfield, USA
PBS tablets	GIBCO®, Invitrogen GmbH, Darmstadt, Germany
Pefabloc SC Protease Inhibitor	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
pH solutions (pH 4.01, 7.01, 10.01)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Ponceau S solution	Sigma-Aldrich Co., St. Louis, USA
Penicillin-Streptomycin solution	Sigma-Aldrich Co., St. Louis, USA
PMSF	Calbiochem, VWR International GmbH,Darmstadt, Germany

Protein-A Sepharose CL-4B	GE Healthcare, Uppsala, Sweden
Roti®-Phenol/Chloroform/Isoamyl alcohol	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Rotiphorese® Gel 30	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
RPMI 1640	GIBCO®, Invitrogen GmbH, Darmstadt , Germany
Sepharose™ CL-4B	GE Healthcare, Uppsala, Sweden
Skim milk powder	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Sodium acetate	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Sodium azide	AppliChem GmbH, Darmstadt , Germany
Sodium chloride (NaCl)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Sodium deoxycholate	AppliChem GmbH, Darmstadt , Germany
Sodium dodecylsulfate (SDS)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Sodium fluoride (NaF)	AppliChem GmbH, Darmstadt , Germany
Sodium hydroxide (NaOH)	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
SYBR Green I	Roche Diagnostics GmbH, Mannheim , Germany
Tamoxifen (>99%)	Sigma-Aldrich Co., St. Louis, USA
TEMED	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
TNFα	R&D Systems, Minneapolis, USA
Tris	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Triton X-100	AppliChem GmbH, Darmstadt , Germany
TRIzol® Reagent	Invitrogen GmbH, Karlsruhe , Germany
Trypsin-EDTA (0.05%)	GIBCO®, Invitrogen GmbH, Darmstadt , Germany
Tween-20	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
α,α-Trehalose Dihydrate	AppliChem GmbH, Darmstadt, Germany

3.4 Inhibitors

Table 4: List of inhibitors

Inhibitor	Target	Source	Cat. No.
Dasatinib	Src, Abl, c-Kit	Selleckchem	S1021
EPZ6438	EZH2	Selleckchem	S7128
Olaparib	PARP1/2	Selleckchem	S1060
Trametinib	MEK1/2	Biomol GmbH	871700-17-3

3.5 Kits

Table 5: List of kits used

Bioanalyzer DNA High sensitivity kit	Agilent Technologies, Santa Clara, USA
ChIP-IT High Sensitivity® kit	Active Motif, La Hulpe, Belgium
CATS mRNA-seq Kit (with polyA selection) for library preparation v1	Diagenode, Seraing, Belgium
Cell Line NucleofectorTM Kit R, V, L	Lonza Bioscience, Switzerland
Immobilon Western Chemiluminescent HRP Substrate	Millipore, Billerica, USA
Lipofectamine™ RNAiMAX	Invitrogen GmbH, Karlsruhe, Germany
MicroPlex Library Preparation Kit v2 for ChIP-seq library preparation	Diagenode, Seraing, Belgium
Qubit dsDNA HS assay	Invitrogen GmbH, Karlsruhe, Germany

3.6 Ladders

Table 6: List of molecular weight ladders

Gene Ruler^{TM} DNA-Ladder	Fermentas GmbH, St. Leon-Rot, Germany
PageRulerTM Prestained Protein Ladder	Fermentas GmbH, St. Leon-Rot, Germany

3.7 Enzymes

Table 7: List of enzymes

Micrococcal Nuclease (MNase)	New England Biolabs, Frankfurt am Main, Germany
Proteinase K	Invitrogen GmbH, Karlsruhe, Germany
Reverse Transcriptase (M-MuLV)	New England Biolabs, Frankfurt am Main, Germany
RNase inhibitor	New England BioLabs GmbH, Frankfurt am Main, Germany
RNase A	Macherey-Nagel GmbH & Co. KG, Düren, Germany
Taq DNA Polymerase	Prime Tech, Mink, Belarus

3.8 Software

Table 8: List of software used

Software/Tool	Source
cBioPortal	http://www.cbioportal.org/
Celigo S Imaging Cytometer Environment	Nexcelom Bioscience LLC, Lawrence, USA
CFX Manager Software 3.1 for qPCR cycler	Bio-Rad Laboratories, Hercules, USA
EnrichR	http://amp.pharm.mssm.edu/Enrichr/
Epson Scanner Software	Seiko Epson Corporation; Nagano, Japan
ImageJ	https://imagej.nih.gov/ij/
Galaxy	http://usegalaxy.org/
Gene Venn	http://genevenn.sourceforge.net/
GraphPad Prism	GraphPad Prism version 4 for Windows, GraphPad Software, Inc., San Diego, USA
Gene Set Enrichment Analysis (GSEA)	http://software.broadinstitute.org/gsea/index.jsp
Image Lab 5.2	Bio-Rad Laboratories, Hercules, USA
Integrative Genome Browser (IGV)	http://software.broadinstitute.org/software/igv/
Microsoft Excel, Word, PowerPoint	Microsoft, Redmond, USA
MIT CRISPR design software	http://crispr.mit.edu/
Morpheus	https://software.broadinstitute.org/morpheus/
OncoLnc	http://www.oncolnc.org/
ReMap/1.2	Integration of ChIP-seq datasets, prediction of colocalizing transcription regulators
Primer designing tool NCBI/Primer-BLAST	https://www.ncbi.nlm.nih.gov/tools/primer-blast/
uEYE Cockpit	IDS Imaging Development Systems GmbH, Obersulm, Germany

3.9 Tools for Analysis of Sequencing Data

All of the analyses presented in this thesis were performed on the GWDG High Performance Computing Cluster or on the Galaxy server.

Table 9: List of bioinformatics tools used

Tool	Purpose
BOWTIE2/2.2.6 bowtie2 --very-sensitive	Mapping fastq to reference genome Output: BAM
CUFFLINKS/2.2.1 cufflinks	Estimation of transcript abundances
CUFFLINKS/2.2.1 cuffdiff	Differential expression of RNA-seq data
DEEPTOOLS/2.4.0 bamCoverage -e 200 --ignoreDuplicate	Generation of ChIP-seq track to view on the genome browser Output: Bigwig
DEEPTOOLS/2.4.0 ComputeMatrix --reference-point	Preparation of data for plotting a heatmap or a profile of given regions
DEEPTOOLS/2.4.0 plotProfile	Plotting aggregate profiles for ChIP-seq data
DEEPTOOLS/2.4.0 plotHeatmap	Plotting heatmaps for RNA-seq data
FASTQC/0.11.4 fastqc	Quality check of fastq files
FASTX/0.0.14 fastx_trimmer -f 12	Trimming of fastq files
Genomic Regions Enrichment of Annotations Tool (GREAT)/3.0.0	Prediction of functions of cis-regulatory regions
HOMER/4.10 findMotifs.pl	Motif analysis
MACS2/2.1.1.20160309 macs2 callpeak --broad-cutoff 0.05	Peak calling Output: Bed
SRATOOLKIT/2.8.2 fastq-dump	Extraction of fastq from sra
SAMTOOLS/1.4 samtools view samtools sort samtools index	Generating, sorting and indexing BAM files

3.10 Publicly Available Datasets

Table 10: Publicly available datasets in the HCT116 cell line used

ChIP-seq and ATAC-seq	GEO Accession	Reference
SMARCC1 (WT and *ARID1A* KO) SMARCA4 (WT and *ARID1A* KO) H3K27ac (WT and *ARID1A* KO)	Series GSE71510	Mathur et al., 2017
H3K27me3 (WT and *ARID1A* KO) ATAC-seq (WT and *ARID1A* KO)	Series GSE101966	Kelso at al., 2017
TCF7L2	GSM782123	ENCODE

BRD4	GSM2058664	Baranello et al., 2016
CTCF	GSM1224650	ENCODE
FOSL1	Series GSE32465	ENCODE
JUND	Series GSE32465	ENCODE
RNA-seq		
WT and ARID1A KO	Series GSE71511	Mathur et al., 2017
Hi-C		
HCT116-RAD21-mAC	Series GSE104333	Rao et al., 2017

3.11 Primers
Genotyping PCR

Table 11: List of genotyping PCR primers used

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Reference
ARID1A knockout	ACTGGAAGAAGACAAAAGTGC	CTGCTGCTCCAGACAAAGAA	Jacobe Rapp

qRT-PCR

All primers designed for this study were ordered from Sigma-Aldrich Co.

Table 12: List of qRT-PCR primers used

Gene	Forward Primer (5'-3')	Reverse Primer (3'-5')	Reference
ADAM19	GGAGCCTGGATGGACAAGAG	GTGGATGCTTTTCTCTCACGG	This project
ALCAM	GGAAATATGAAAAGCCCGATGGC	ACATCGTCGTACTGCACACTT	This project
ARID1A Exon 2	GCAAGATGAGACCTCAGCCA	CCATAAGGAGGAATCTGCTGTGT	Jacobe Rapp
ARID1A Exon 5	TCCTCATACCTCCCCTCACC	TGAGCGAGACTGAGCAACAC	This project
ARID1A Exon 11	GGTGGATTGACTCAGGTCAACA	AAAGATGTCTGGGGGAGGGT	This project
CD44	GGACAAGTTTTGGTGGCACG	GGTTATATTCAAATCGATCTGCGCC	This project
CXCL2	GAAAGCTTGTCTCAACCCCG	TGGTCAGTTGGATTTGCCATTTT	Kosinsky et al., 2018
EGFR	AGTATTGATCGGGAGAGCCG	TCGTGCCTTGGCAAACTTTC	This project
EMP1	CTGGGACCCTTCAGAACTCTCT	TGGACCACAAAGATACCAGCC	This project
EREG	CTCTGCCTGGGTTTCCATCTTC	TCACTGGACTCTCCTGGGATAC	This project
F3	ACCTGGAGACAAACCTCGGA	TCCCGGAGGCTTAGGAAAGT	This project
GAPDH	ATGGGGAAGGTGAAGGTCG	GGGGTCATTGATGGCAACAATA	This project
IL6	ATTCAATGAGGAGACTTGCCTGG	TTGGGTCAGGGGTGGTTATTG	This project
JAG1	ACCCCCTGTGAAGTGATTGAC	CTGACTCTTGCACTTCCCGT	This project
NFKB1	GCACCCTGACCTTGCCTATT	GCTCTTTTTCCCGATCTCCCA	This project
RELB	AGCGGAAGATTCAACTGGGC	TGTCATAGACGGGCTCGGAA	Upasana Bedi, 2014
S100A14	CTCCTGTCTTGTCTCAGCGG	TGAGCATCCTCTGCGTTGG	This project
SMURF2	TGGGAAGAAAGGAGAACCGC	ATATTCGGATGCCGGTCGTG	This project

ChIP-qPCR

Enh: enhancer, TF: transcription factor (ARID1A/JunD positive), ac: H3K27ac positive

Table 13: List of ChIP-qPCR primers used

Gene	Forward Primer (5'-3')	Reverse Primer (3'-5')	Reference
AXIN2_TSS_TF	GGCCTGCCAACTTCAAAGGG	ATCAATGGTGAGTGCCGAGG	Vijaya Kari
EREG_enh_TF	AGCAAGGTCAAAATAAACCGTATGC	AAGTGGTTGCCCAACAGTCA	This project
EREG_enh_ac	ATCATCTGTGTTATATCACCTGGCA	TTTCTTGTCTGGTGGCATTGGT	This project
F3_enh_TF	CACCGACGAGATTGTGAAGGA	CCGAGGTTTGTCTCCAGGTAAG	This project
F3_enh_ac	TGCTTCCGAGTTGGCTGTAG	CTCAGCCCACTAACCGTCTTT	This project
JAG1_enh_TF	TGCCCTAGATAGAGAAGGGATGAA	AATCGCAAACTTTCGGACACAC	This project
JAG1_enh_ac	GATACGCCTTCGCTGCATCA	CGCAAACTTTCGGACACACTC	This project
OLIG2_negative	GTCACCAACGCTCCCTGAAAT	CTGCACGCGGGTACCTATAAT	Najafova et al., 2016
SMURF2_enh_TF	GAAGTGACTGCAGACGTGGA	ACTCATCCCCCAGGAATGGA	This project

3.12 siRNAs

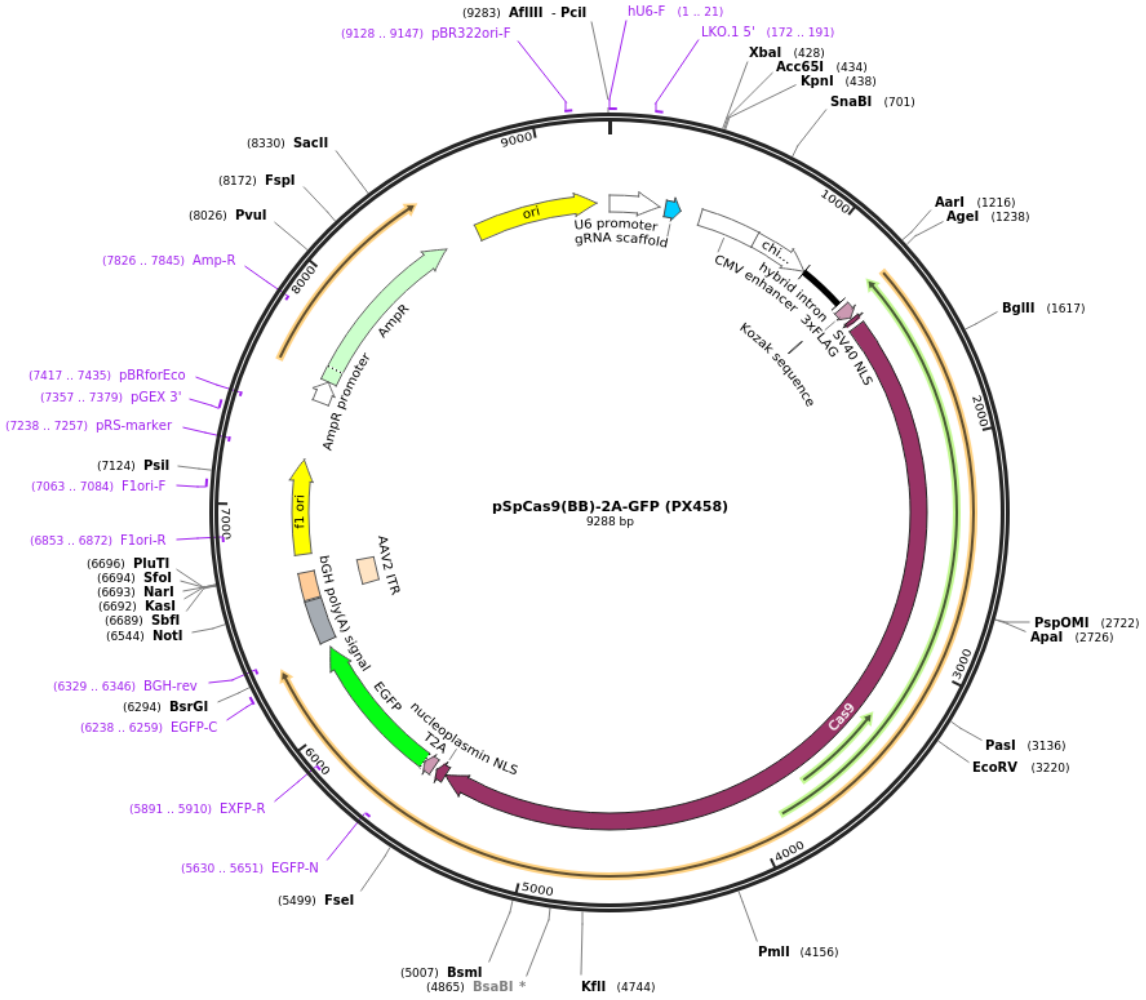
Table 14: List of siRNAs used

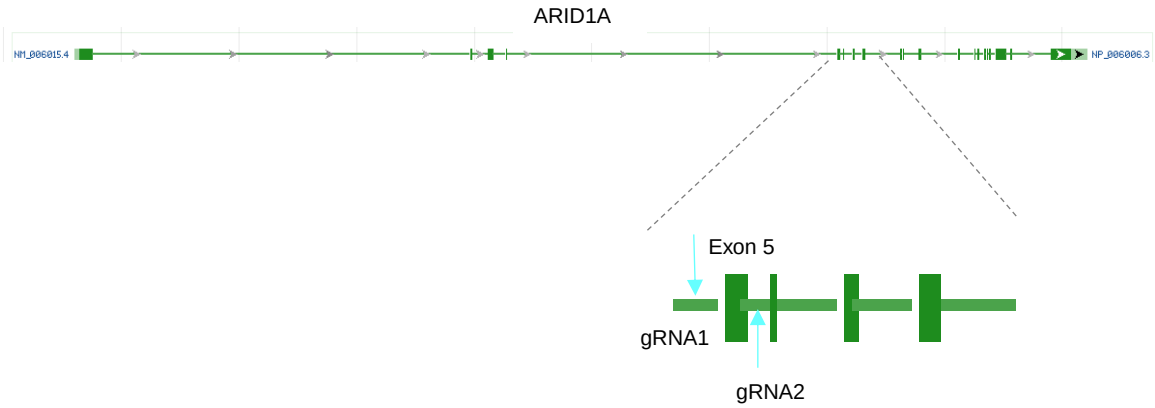
Target Gene	SiRNA sequence (5'-3')	Source	Cat.No.
siGENOME Non-Targeting siRNA #5	UGGUUUACAUGUCGCUAA	Dharmacon	D-001210-05-20
siGENOME siRNA SMARTPool JUND	CGCUCAAGGACGAGCCACA GAAACACCCUUCUACGGCG CCGACGAGCUCACAGUUCC CCGGCAGCAUGAUGAAGAA	Dharmacon	D-003900-08-D-00390010
ON-TARGET plus siRNA ARID1A		Dharmacon	LU-017263-00-0005
#1	GAAUAGGGCCUGAGGGAAA		
#2	AGAUGUGGGUGGACCGUUA		
#3	GCAACGACAUGAUUCCUAU		
#4	GGACCUCUAUCGCCUCUAU		
siGENOME siRNA TCF7L2		Dharmacon	D-003816-01, 03,04,17
#1	GAUGGAAGCUUACUAGAUU		
#2	AAAGUGCGUUCGCUACAUA		
#3	UCACGCCUCUUAUCACGUA		
#4	ACACUUACCAGCCGACGUA		
siGENOME siRNA BRD2		Dharmacon	D-004935-01,02,04,18
#1	AGAAAGGGCUCAUCGCUUA		
#2	GAAAAGAUAUUCCUACAGA		
#3	GAUGAAGGCUCUGUGGAAA		
#4	UUAGAGAGCUUGAGCGCUA		
SiGENOME siRNA BRD3			
#1	CCAAGAGGAAGUUGAAUUA		D-004936-02,03,04,17
#2	AAUUGAACCUGCCGGAUUA		
#3	GGAGAGCUCUUCGGACUCA		
#4	CGGCUGAUGUUCUCGAAUU		
siGENOME siRNA BRD4		Dharmacon	D-004937-03,02
#3	UAAAUGAGCUACCCACAGA		
#4	GAACCUCCCUGAUUACUAU		

3.13 CRISPR/Cas9 mediated knockout of ARID1A

Table 15: Details for plasmids and guide RNAs used in CRISPR/Cas9 mediated knockout

Vector name	pSpCas9(BB)-2A-GFP (PX458)	Addgene #48138 Reference: Ran et al., 2013
Bacterial resistance	Ampicillin	
Promoter	hU6	
Name of gRNA	hARID1A-3	hARID1A-4
Sequence of insert (sense)	CACCGCAGTGTTTCACTCGTTGCC	CACCGGACTGCCCCCAGTAATATTA
Sequence of insert (antisense)	AAACGGGCAACGAGTGAAACACTGC	AAACTAATATTACTGGGGGCAGTCC
Targeted Gene	hARID1A-Ex5-A	hARID1A-Ex5-B
NCBI reference sequence	NM_006015.4	
Resulting modification	Knock out	
Restriction enzyme	BbsI (BpiI)	


(9283) AflIII - PciI
hU6-F (1 .. 21)
(9128 .. 9147) pBR322ori-F
LKO.1 5' (172 .. 191)
XbaI (428)
Acc65I (434)
KpnI (438)
SnaBI (701)
(8330) SacII
(8172) FspI
(8026) PvuI
(7826 .. 7845) Amp-R
AarI (1216)
AgeI (1238)
BglII (1617)
(7417 .. 7435) pBRforEco
(7357 .. 7379) pGEX 3'
(7238 .. 7257) pRS-marker
(7124) PsiI
(7063 .. 7084) F1ori-F
(6853 .. 6872) F1ori-R
(6696) PluTI
(6694) SfoI
(6693) NarI
(6692) KasI
(6689) SbfI
(6544) NotI
PspOMI (2722)
ApaI (2726)
(6329 .. 6346) BGH-rev
(6294) BsrGI
(6238 .. 6259) EGFP-C
PasI (3136)
EcoRV (3220)
(5891 .. 5910) EXFP-R
(5630 .. 5651) EGFP-N
(5499) FseI
PmlI (4156)
(5007) BsmI
(4865) BsaBI *
KflI (4744)
ori
U6 promoter
gRNA scaffold
chi...
CMV enhancer
hybrid intron
3XFLAG
SV40 NLS
Kozak sequence
AmpR
AmpR promoter
ori f1
AAV2 ITR
bGH poly(A) signal
EGFP
nucleoplasmin NLS
T2A
Cas9
pSpCas9(BB)-2A-GFP (PX458)
9288 bp


ARID1A
NM_006015.4
NP_006006.3
Exon 5
gRNA1
gRNA2

gRNA

Exon 5

Exon 5 (out of frame, complete knockout)

CCTGGGCAACGAGTGAAACACTGTCTCAAAAAAAAAAAATTTTTTTTTTTTAATAAAAATAGTATCATGACTAAAGAACGTGT
GTGATGTATTTGCTCTTGGTTGTTTAAGGAAAATGCTAAGCAAGTAGTAGGATTATTGAAAGTAGAATCTTTCTGCCTAAT
ATTACTAATCCATGTTCTTATATATATGTTCTAGGATCTATCTGGTTCAATAGATGACCTCCCCATGGGGACAGAAGGAG
CTCTGAGTCCTGGAGTGAGCACATCAGGGATTTCCAGCAGCCAAGGAGAGCAGAGTAATCCAGCTCAGTCTCCTTTCT
CTCCTCATACCTCCCCTCACCTGCCTGGCATCCGAGGCCCTTCCCCGTCCCCTGTTGGCTCTCCCGCCAGTGTTGCTC
AGTCTCGCTCAGGACCACTCTCGCCTGCTGCAGTGCCAGGTACCCTCAAGTGCTGGGCTTTAGGGAGAAGGTGACTG
CCCCCAGTAATATTAAGGGGGAA

Figure 9: The plasmid map of the pSpCas9(BB)-2A-GFP (PX458) plasmid. The plasmid used to transfect Cas9 and the guide RNAs targeting exon 5 of the ARID1A gene (figure from Addgene). The sequence represents the region around exon 5 (marked in gray) of ARID1A, and the position of the guide RNAs targeting regions flanking exon 5. The green letters represent the PAM sequences.

3.14 Primary Antibodies

Table 16: List of primary antibodies

Antibody	Species	Clone	Source	Cat. No.	Dilution	
					WB	ChIP
ARID1A	Rabbit	D4A8U	CST	12354	1:1000	4-5µg
ARID1A	Mouse	PSG3	Merck Millipore	04-080	-	4µg
ARID1A	Mouse	PSG3	Santa Cruz	sc-32761	-	4µg
ARID1A	Mouse	PSG3	Santa Cruz	sc-32761X	-	4µg
ERK	Rabbit	Polyclonal	Santa Cruz	sc-94	1:1000	-
JUND	Rabbit	Polyclonal	Santa Cruz	sc-74	1:1000	1.5µg
H3K27ac	Rabbit	Polyclonal	Diagenode	c15910196	-	2µg
H3K27me3	Rabbit	Polyclonal	Diagenode	c15310069	1:1000	2µg
HSC70	Mouse	B-6	Santa Cruz	sc-7298	1:30000	-
IgG	Rabbit	Polyclonal	Abcam	ab46540	-	2-4 µg
pERK	Rabbit	Polyclonal	CST	9101	1:500	-
SMARCA2	Rabbit	D9E8BXP®	CST	11966	1:1000	-
SMARCA4	Rabbit	A52	CST	3508	1:1000	-
SMARCB1	Rabbit	D9C2	CST	8745	1:1000	-

SMARCC1	Rabbit	D7F8S	CST	11956	1:1000	-
SMARCC2	Rabbit	D8O9V	CST	12760	1:1000	-

3.15 Secondary Antibodies

Table 17: List of secondary antibodies

Antibody	Cat. No.	Dilution	Source
Goat Anti-Mouse IgG-HRP	sc-2004	1:10,000	Santa Cruz
Goat Anti-Rabbit IgG-HRP	sc-2005	1:10,000	Santa Cruz

3.16 Human Colorectal Cancer Cell Lines

Table 18: List of cell lines used

Cell Line	Tissue	Disease	Source
COLO 201	Colon; derived from metastatic site: ascites	Colorectal adenocarcinoma	ATCC® (CCL-224™)
COLO 320DM	Colon	Dukes' type C colorectal adenocarcinoma	ATCC® (CCL-220™)
DLD1	Colon	Dukes' type C colorectal adenocarcinoma	ATCC® (CCL-221™)
EGI1	Extrahepatic bile duct	Cholangiocarcinoma	Scherdin et al., 1987
HCT116	Colon	Colorectal carcinoma	ATCC® (CCL-247™)
HT-29	Colon	Colorectal adenocarcinoma	ATCC® (HTB-38™)
L3.6pl	Pancreas to liver selection in nude mice	Pancreatic cancer	Bruns et al., 1999
NCI-H508	Cecum	Colorectal adenocarcinoma	ATCC® (CCL-253™)
RKO	Colon	Colorectal carcinoma	ATCC® (CRL-2577™)
SW48	Colon	Dukes' type C colorectal adenocarcinoma	ATCC® (CCL-231™)
KRAS (G12D/+) SW48	Colon	Dukes' type C colorectal adenocarcinoma	Horizon Discovery (HD 103-011)
SW480	Colon	Colorectal adenocarcinoma	ATCC® (CCL-228™)
SW837	Rectum	Rectal adenocarcinoma	ATCC® (CCL-235™)
T84	Colon; derived from metastatic site: lung	Colorectal carcinoma	ATCC® (CCL-248™)

TFK1	Extrahepatic bile duct	Cholangiocarcinoma	Saijyo et al., 1995

3.17 Cell Culture Media and Cell Culture Source

Table 19: List of cell lines and their cell culture media

Cell line	Culture Medium	Provided by
COLO 201	RPMI 1640	Jessica Eggert, Dept. of Surgery, UMG, Göttingen
COLO 320DM	RPMI 1640	Jessica Eggert, Dept. of Surgery, UMG, Göttingen
DLD1	RPMI 1640	Jessica Eggert, Dept. of Surgery, UMG, Göttingen
EGI1	DMEM	Elisabeth Hessmann, Dept. of Gastroenterology, UMG, Göttingen
HCT116	McCoy's 5A (modified)	Vijaya Kari, Dept. of Surgery, UMG, Göttingen
HT-29	McCoy's 5A (modified)	Jessica Eggert, Dept. of Surgery, UMG, Göttingen
L3.6pl	MEM	Vivek Mishra, Dept. of Surgery, UMG, Göttingen
NCI-H508	RPMI 1640	Jessica Eggert, Dept. of Surgery, UMG, Göttingen
RKO	DMEM F12	Jessica Eggert, Dept. of Surgery, UMG, Göttingen
SW48	McCoy's 5A (modified)	Jessica Eggert, Dept. of Surgery, UMG, Göttingen
KRAS (G12D/+) SW48	McCoy's 5A (modified)	Jessica Eggert, Dept. of Surgery, UMG, Göttingen
SW480	RPMI 1640	Robyn Kosinsky, Dept. of Surgery, UMG, Göttingen
SW837	DMEM/F12	Robyn Kosinsky, Dept. of Surgery, UMG, Göttingen
T84	DMEM /F12	Jessica Eggert, Dept. of Surgery, UMG, Göttingen
TFK1	DMEM	Elisabeth Hessmann, Dept. of Gastroenterology, UMG, Göttingen

4. Methods

Cell Culture

4.1 Cell Culture

The cell lines used for this project were cultured in media mentioned in section 3.17 and supplemented with 10% FBS, 1% glutamine and 100 U/mL penicillin/streptomycin. All cells were tested to be mycoplasma free. The cells were maintained at 37°C and 5% CO_2 and were passaged every 3-4 days. The cells were frozen in a freezing medium containing 42% FBS, 8% DMSO and 50% penicillin/streptomycin free medium.

4.2 siRNA mediated knockdown

The cells were resuspended in antibiotic free medium. The transfection mix for each reaction contained 500 µL OptiMEM medium with 5 µL RNAiMax transfection reagent and 1.5 µL 20 µM siRNA. This was left undisturbed at room temperature for 20 min. 20 µM non-targeting siRNA was used as a negative control. The siRNA sequences are provided in section 3.12. The transfection mix was added to the wells of a 6-well plate followed by the seeding of 180,000-250,000 cells to ensure maximum interaction between the cells and the siRNA-lipofectamine complex. Antibiotics were replenished the next day. RNA and protein were harvested after 48 h.

4.3 CRISPR/Cas9 mediated knockout

CRISPR mediated gene editing was used to generate *ARID1A* knockout colorectal cancer cell lines to mimic ARID1A-deficient cancer cells. Four cell lines namely HT29, HCT116, DLD1 and COLO320DM were used. Guide RNAs (gRNAs) targeting the two introns flanking exon 5 of the *ARID1A* gene were designed and off-target binding effects were minimized based on scores obtained on the MIT CRISPR design software. The gRNAs were designed such that the number of nucleotides between the positions that are targeted was not a complete reading frame. The excision of this fragment resulted in an out of frame deletion causing a loss of the protein. The guide RNAs were cloned into a pSpCas9(BB)-2A-GFP plasmid using the BbsI restriction enzyme. 4-6 µg of plasmid was transfected into the cells by electroporation using a Lonza Nucleofector (the kits and programs for each cell line are mentioned below). Briefly, 3-5 million cells were resuspended in the appropriate kit buffer along with the plasmid in a maximum volume of 110 µL. 48 h after transfection the cells were sorted as 192 single cell clones by fluorescence assisted cell sorting based on GFP-positivity. The single cell clones were expanded and screened for ARID1A loss by genotyping PCR and western blot. A single clone for each cell line was used for all

experiments. The details about the plasmid and gRNA sequences are provided in section 3.13.

Table 20: Electroporation kits, protocols and transfection efficiency

Cell line	Kit	Program	Transfection Efficiency
COLO320DM	R	T-001	57.7%
DLD1	L	T-020	70.6%
HCT116	V	D-032	75.9%
HT29	R	W-017	39.2%

4.4 Stimulation with TNFα

Recombinant TNFα was reconstituted in its vehicle 0.2% BSA. It was diluted to 10ng/mL in cell culture medium. The cells were treated with TNFα or vehicle for 30 min or 6h after which RNA was harvested.

4.5 Inhibitor Treatment

Increasing doses of inhibitor were prepared in DMSO. These were then diluted in a large volume of cell culture medium to bring the concentration of DMSO to 1:1000. Inhibitor treatment was renewed every 48 h. As controls, the cells were either treated with 1:1000 diluted DMSO or left untreated. The list of all inhibitors used in this project is provided in section 3.4.

4.6 Cell Proliferation Assay

5000-7000 cells were seeded in each well of a 24-well plate in duplicates or triplicates for each condition. Cell proliferation was assessed by measuring confluence using a Celigo S Cell Imaging Cytometer every 24 h. Readings were recorded for 5-7 days and relative confluence against days was plotted. The confluence in each well on each day was normalized with the confluence in that well on day 1.

4.7 Crystal Violet Staining

Different conditions were set up in a 6 or 24-well format. The cells were grown for 5-7 days. Once confluent, the wells were washed with 1X PBS. The cells were fixed for 10 min in 99% methanol and stained with 1% crystal violet prepared in 2% ethanol. The plates were scanned, and the staining intensity was assessed using the ImageJ software using the 16-bit image setting. The mean values for intensity were used to plot proliferation curves.

Molecular Biology

4.8 DNA extraction and genotyping PCR

Cells were lysed in lysis buffer (10 mM Tris/HCl, 400 mM NaCl, 2 mM EDTA, 2% SDS) containing 10µg/ml proteinase K to digest proteins. These were incubated overnight at 65°C. The DNA was precipitated with isopropanol and the DNA pellet was washed twice in 70% ethanol. The isolated DNA was resuspended in 50-200 µL of ddH_20. 100ng DNA was used for each PCR reaction. Each PCR reaction was conducted in a volume of 25 µL containing 2.5 µL 10X buffer B, 2.5 µL 2mM dNTPs, 2 µL 25mM $MgCl_2$, 1 µL 10 µM forward and 10 µM reverse primers and 0.15 µL Taq polymerase. The reaction allowed 15 min of initial denaturation followed by 35 cycles of 30 s at 95°C for denaturation, 45 s at 60°C for annealing and 60 s at 72°C for elongation. Final elongation was carried out at 72°C for 10 min. The PCR products were run on a 1 % agarose gel at 100 V for 45 min and visualized using an INTAS imager.

4.9 Protein Extraction

The cells were washed once with PBS. RIPA Lysis Buffer (1X PBS 0.5% sodium deoxycholate (w/v), 0.1% SDS (w/v), 1% NP-40 (v/v) containing protease inhibitors (100µM β-glycerophosphate disodium salt hydrate (BGP), 100 µM N-Ethylmaleimide, 100 µM Pefabloc, 1 µM Aprotinin/Leupeptin) was added to each well of the plate. The wells were scraped, and the samples were sonicated for 10 min using a Bioruptor with 30 s ON/OFF cycles.

4.10 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western Blot

Samples for SDS-PAGE were prepared by mixing protein extracts with 6X Lämmli sample buffer (375 mM Tris/HCl, 10% SDS, 50% glycerol, 0.06% Bromophenol blue, 600mM DTT) which were then denatured at 95°C for 5 min. Equal quantities of protein were resolved on 7-10% SDS polyacrylamide gels (375 mM Tris-HCl (pH 8.8), 0.1% SDS, 0.1% APS, 0.04% TEMED).The SDS PAGE was run at 20 mA for 1.5-2 h in running buffer (25 mM Tris, 192 mM glycine, 0.01% SDS). The separated proteins were transferred onto a nitrocellulose membrane at 100V for 1.5 h in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol) at 4°C. The transfer was verified by Poncaeu S staining. The membrane was blocked with 5% (w/v) milk or BSA in Tris buffered saline-Tween20 (TBST). The membrane was incubated with the primary antibody at 4°C overnight. On the next day, it was washed thrice in TBST for 10 min each and then incubated with the secondary antibody for 1 h at room

temperature. The washing step was repeated and the HRP substrate was added. The nitrocellulose membrane was viewed under a ChemiDoc™ imager.

4.11 RNA Extraction and Quality Check

All gene expression experiments were performed in triplicate. The cells were washed twice with PBS and lysed by adding 500 µL of TRIzol® reagent. The wells were scraped, and the cells were resuspended and transferred to 1.5 mL tubes. Following this, 100 µL of chloroform was added and the tubes were shaken vigorously for 20 s. These were then centrifuged at 15000 g at 4°C for 30 min. The aqueous phase was transferred to new tubes and an equal volume of isopropanol was added. Isopropanol precipitation occurred at -20°C overnight. Next day, the tubes were centrifuged at 15000 g at 4°C for 30 min. and the pellet was washed in 70% ethanol twice. The supernatant was removed, and the pellet was air dried. The RNA was resuspended in 40 µL DEPC water. The quality of the RNA isolated was checked by running the samples on a 0.8% agarose formaldehyde gel, after denaturation at 70°C.

4.12 cDNA Synthesis

For cDNA synthesis 1µg of total RNA was mixed with 2 µL of 15 µM random primers and 4 µL of 2.5 mM dNTP mix and incubated 5 min at 70°C. 4 µL of reverse transcription master mix containing 2 µL 10X reaction buffer, 10 units of RNase inhibitor, 25 units of reverse transcriptase and 1.625 µL of DEPC water were added to each sample. cDNA synthesis took place at 42°C for 1h followed by enzyme inactivation for 5 min at 95°C. Finally, the samples were diluted to 50 µL volume in water.

4.13 Quantitative Real Time PCR (qRT-PCR)

1-2 µL of ChIP or cDNA in a reaction volume of 25 µL was used for quantitative real-time PCR. Each reaction contained 75 mM Tris-HCl (pH 8.8), 20 mM $(NH4)_2SO_4$, 0.01% Tween-20, 3 mM $MgCl_2$, 200 µM dNTPs, 0.5 U/reaction Taq polymerase, 0.25% Triton X-100, 1: 80,000 SYBR Green I, 300 mM Trehalose and 30 nM primers. Standard curves were used for quantification. The expression of a gene in each cDNA sample was normalized to the expression of the housekeeping gene GAPDH, the mRNA levels of which should remain unaffected by any of the treatments. ChIP DNA samples were normalized to their input and expressed as % input. For ChIP-qPCR, primers were designed using sequences obtained from IGV tracks. Primers were designed using NCBI primer blast and primer sequences are indicated in section 3.11. The programs used for qRT-PCR were as follows:

Table 21: qPCR conditions

	ChIP		mRNA levels		
	95°C	2 min	95°C	5 min	
45X	95°C	10 s	95°C	10 s	40X
	60°C	30 s	62°C	30 s	

Chromatin Immunoprecipitation (ChIP)

4.14 ChIP using MNase shearing

This protocol was adapted from Raab et al., 2015. DNA-DNA crosslinking in confluent HCT116 cells was performed using 0.3% formaldehyde (in PBS). The plates were incubated for 30 min at 4°C on a rocker. 1.5 mL 2M glycine was added to quench the formaldehyde and the plates were incubated for 5 min at room temperature. The plates were then washed 3 times with cold PBS. They were then scraped in 1.5 mL PBS containing 1mM PMSF. These were then centrifuged at 13000 g for 1 min. The supernatant was removed. The pellets were resuspended in 1 mL swelling buffer (25 mM HEPES, 1.5 mM $MgCl_2$, 10 mM KCl, 0.1%NP-40, 1mM PMSF) containing PIC (Protease inhibitor cocktail: 100 µM β-glycerophosphate disodium salt hydrate (BGP), 100 µM N-Ethylmaleimide, 100 µM Pefablock, 1 µM Aprotinin/Leupeptin, 1 mM $NiCl_2$, 1 µM iodoacetic acid) and they were incubated for 10 min at 4°C on a rotating wheel. The solution was mixed by pipetting 5-10 times and the nuclei were then pelleted at 2000 rpm for 7 min at 4°C. The pellet was then resuspended in 5 mL of sucrose buffer A (11% sucrose, 15 mM HEPES pH 7.9, 60 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 0.5 mM PMSF, PIC). 5mL sucrose buffer B (30% sucrose, 15 mM HEPES pH 7.9, 60 mM KCl, 2 mM EDTA, 0.5mM EGTA, 0.5 mM PMSF, PIC) was layered over sucrose buffer A to create a density gradient and spun at 3000 rpm for 10 min. The pellet was washed in 10 mL NUC buffer (15 mM HEPES, pH 7.9, 60 mM KCl, 15 mM NaCl, 0.32 M sucrose, 0.5 mM PMSF, PIC) by spinning at 2000 rpm for 7 min. The pellet was resuspended in 1 mL NUC buffer containing 3.3 µL 1M $CaCl_2$ / mL. The samples were then incubated at 37°C for 5 min to pre-warm. 0.5 µL MNase was added and the reactions were incubated for 15 min at 37°C. The reaction was stopped by adding 100 µL 0.5M EGTA. The samples were stored on ice for 5 min. 1.1 mL 2X lysis buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine, PIC) was added to the samples which were then passed 5 times through a 20G needle and then 5 times through a 23G needle. 1% triton X-100 was added and the samples were spun at 13000 g for 15 min at 4°C. The supernatant was stored on ice; the

pellet was resuspended in 1 mL 1X lysis buffer and incubated for 1.5 h at 4°C on a rotating wheel. These were then spun at 13000 g for 15 min at 4°C again and the supernatant was combined with the first supernatant. 10% volume was saved as input. The rest of the lysate was added to bead conjugated antibodies (ARID1A antibodies tested are listed in section 3.14) which were then rotated overnight at 4°C. The next day, the beads were washed 6 times with RIPA wash buffer (50 mM HEPES-KOH pH 7.6, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-deoxycholate) and once with TE buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA) containing 50mM NaCl. The immunoprecipitated DNA was extracted by phenol chloroform extraction.

4.15 ChIP using sonication for shearing

Confluent cells in 15 cm plates were crosslinked for 20 min by adding 1% formaldehyde in PBS. Glycine of final concentration of 125 mM was added for 5 min to quench the formaldehyde. The cells were washed twice with cold PBS and scraped in 1.5 mL of nuclear preparation buffer (150 mM NaCl, 20 mM EDTA pH 8.0, 50 mM Tris-HCl pH 7.5, 0.5% NP-40, 1%Triton X-100, 20 mM NaF, PIC). The nuclear pellet was isolated from the lysate by centrifugation at 12,000 g for 1 min and the pellet was resuspended 150-300 µL sonication buffer-1 (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.5%/0.1% SDS (w/v), PIC) and incubated at 4°C on a rotating wheel for 15 min. The SDS content was diluted to 0.25%/0.05% SDS using 150-300 µL sonication buffer-2 (20 mM EDTA, 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40 (v/v), NaF 20 mM). The samples were sonicated in a Bioruptor Pico for 20-30 cycles with 30 s ON/OFF cycles. The soluble chromatin was obtained by centrifugation at 12000 g for 10 min and pre-cleared with 100µL of 50% slurry of Sepharose beads for 1 h. The chromatin was then diluted in dilution buffer (20 mM EDTA, pH 8.0 50 mM Tris-HCl, 150 mM NaCl, 1% (v/v) NP-40, 20 mM, NaF 0.5% (w/v), sodium deoxycholate). Following this, 100 µL of chromatin extract was diluted up to 500 µL with IP buffer (20 mM EDTA, 50 mM Tris-HCl pH 8.0, 150 mM NaCl 1% (v/v) NP-40, 20 mM NaF 0.5% (w/v) sodium deoxycholate, 1% (w/v) SDS, PIC) and incubated overnight with primary antibodies (ARID1A antibodies tested are listed in section 3.14). Immunoglobulin bound complexes were precipitated by adding 30 µL of 50% slurry of Protein-A/G for 2 h. Following incubation, the samples were centrifuged at 2,000 g for 2 min. The beads were washed with ice-cold IP buffer once, wash buffer (0.5 M LiCl, 1% (v/v) NP-40, 1% (w/v) sodium deoxycholate, 20 mM EDTA, 10 mM Tris-HCl (pH 8.5), 20 mM NaF) twice, IP buffer once more and finally TE (10 mM Tris-HCl (pH 8.0), 1 mM EDTA) buffer. The immunoprecipitated DNA was extracted by phenol chloroform extraction.

Variations of the above two protocols were made. Changes in crosslinking conditions (by adding dual crosslinker DSG), SDS concentration in the buffers were made.

4.16 Dual Crosslinking with EGS

ARID1A ChIP was performed as described in Zirkel et al., 2018. Briefly, HCT116 cells that were 80-90% confluent in 15 cm plates were cross linked for 20 min in 15mM EGS, 20 min in 2mM EGS at room temperature and then in 1% paraformaldehyde for 40 min at 4°C. The samples were then processed using the Active Motif ChIP-IT High Sensitivity Chromatin immunoprecipitation kit according to manufacturer's instructions. The samples were sonicated in a Bioruptor Pico for 15 cycles (30 s ON/OFF) to obtain 200-500bp fragments. 10 µg sheared chromatin was used for immunoprecipitation overnight with 5 µL of ARID1A antibody (cell signaling) at 4°C. After incubation with protein G agarose beads provided in the kit for 2 h, the beads were washed, and the chromatin was de-crosslinked at 65°C overnight. DNA was extracted by following the manufacturer's instructions.

4.17 ChIP for H3K27ac, JunD and TCF7L2

For H3K27ac, JunD and TCF7L2 ChIPs, confluent HCT116 WT and HCT116 *ARID1A* KO cells in 15 cm plates were crosslinked for 20 min by adding 1% formaldehyde in phosphate buffered saline. 125 mM glycine was added for 5 min to quench the formaldehyde. The cells were harvested in nuclear preparation buffer (150 mM NaCl, 20 mM EDTA pH 8.0, 50 mM Tris-HCl pH 7.5, 0.5% NP-40, 1%Triton X-100, 20 mM NaF, PIC). The nuclear pellet was isolated and resuspended 150-300 µL sonication buffer-1 (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.5%/0.1% SDS (w/v), PIC). The SDS content was diluted to 0.5% SDS using 150-300 µL sonication buffer-2 (20 mM EDTA, 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40 (v/v), NaF 20 mM). The samples were sonicated in a Bioruptor Pico for 15 cycles (30 s ON/OFF). The soluble chromatin was pre-cleared in sepharose. The chromatin was then diluted in dilution buffer (20 mM EDTA, pH 8.0 50 mM Tris-HCl, 150 mM NaCl, 1% (v/v) NP-40, 20 mM, NaF 0.5% (w/v), sodium deoxycholate). Following this, 100 µL of chromatin extract was diluted up to 500 µL with IP buffer (20 mM EDTA, 50 mM Tris-HCl pH 8.0, 150 mM NaCl 1% (v/v) NP-40, 20 mM NaF 0.5% (w/v) sodium deoxycholate, 1% (w/v) SDS, PIC) and incubated overnight with primary antibodies (section 3.14). Immunoglobulin bound complexes were precipitated by adding Protein A/G sepharose. The beads were washed as described before, and the chromatin was de crosslinked at 65°C overnight and DNA was isolated by phenol-chloroform extraction.

4.18 ChIP Western Blot

The steps of chromatin immunoprecipitation were followed as explained above. 40 µL RIPA buffer and 10 µL 6X Lammli buffer were added to the beads. Before loading on a gel, 12 µL

β-mercaptoethanol was added to the samples and the proteins were denatured 95°C for 15 min. Western blot analysis was performed as explained earlier.

4.19 Phenol Chloroform DNA Extraction

The steps of immunoprecipitation were followed as explained above followed by DNA isolation. DNA was isolated by phenol chloroform extraction. For this, 50 µL of 10 mM Tris, pH 8.0 containing 0.2 µg/ µL RNase A was added to the beads and incubated for 30 min at 37°C. 50 µL of 2X sonication buffer 1 (50mM Tris pH 8.0, 10mM EDTA, 1% SDS) and 1µL proteinase K was added and incubated at 800 rpm overnight at 65°C. The samples were centrifuged at 2000 g for 2 min at room temperature and the supernatant was stored. 100 µL of 10 mM Tris pH 8.0 was added on the beads and incubated at 800 rpm for 10 min at 65°C. The samples were centrifuged at 5000 g for 2 min and the supernatant was combined with the earlier supernatant. 10 µL 8M LiCl and 4 µL linear polyacrylamide was added to the supernatant followed by addition of 200 µL phenol/chloroform/isoamylalcohol (25:24: 1). The samples were centrifuged at 15000 g for 2 min and the aqueous phase was transferred to 1.5 mL low binding tubes. The organic phase was back extracted by adding 200 µL 10mM Tris pH 8.0, 0.4M LiCl and the second aqueous phase was added to the first one. The DNA was precipitated by adding 1 mL ethanol and incubating for more than 2 h at -80°C.The samples were then centrifuged at 15000 g for 30 min at 4°C.The samples were then rinsed with 70% ethanol. The pellets were dried, and the DNA was resuspended in 40 µL water. DNA concentration was measured by Qubit® dsDNA HS assay kit.

Animal experiments

4.20 Xenografts

For xenografts, SCID Hairless Outbred mice were used. 1 million HT29 WT or HT29 *ARID1A* KO (2 different clones) were prepared in Matrigel and then inserted in syringes. The WT cells were injected subcutaneously on the right flank of 11 mice, 1 clone of the HT29 KO cells were injected in the left flank of 5 mice and another clone on the left flank of 6 mice to rule out the effects of clonal variation. Tumor size was measured every few days for 21 days. Upon reaching 1 cm diameter, the mice were sacrificed and tumors were harvested. These were weighed, and their volume measured. Sections of the tumors were snap frozen and stored for further analysis.

4.21 Genetic Model

Mice of the C57BL/6J background were used. Mice with various genotypes were injected with 5% tamoxifen to induce the knockout of *ARID1A* from the colon. The expression of the Cre ER^{T2} fusion was placed under the control of the colon specific gene *Cdx2.* The generation of the genetic mouse model was carried out by Dr. Robyn Kosinsky and Dr. Florian Wegwitz, Dept. of Surgery, UMG, Göttingen. For assessing the inflammation that might be caused due to the loss of ARID1A, we measured the disease activity index (DAI). For this, body weight, stool consistency and a blood occult test every week for 6 months was measured. Scores for the various parameters were assigned as follows, weight loss: 0-1% (0), 1-5% (1), 5-10% (2), 10-15% (3), >15% (4), stool consistency: normal (0), soft (1), very soft (2). To assess intestinal bleeding, stool guaiac test was performed. In this test, the presence of heme in blood, acts as a catalyst for the oxidation of alpha-guaiaconic acid to quinone, which is detectable as a blue color. The intensity of the blue colour is scored as: no blue staining (0), weak staining (1), medium (2) and strong blue staining (3). After 6 months, the mice were sacrificed, and the colons were examined, harvested, snap frozen and stored for further analysis.

Next Generation Sequencing (NGS)

4.22 mRNA-seq and ChIP-seq library preparation and sequencing

Total RNA from the HCT116, DLD1 and COLO320DM cell lines (WT, *ARID1A* KO or *ARID1A* KD) were extracted in triplicates and confirmed to be of good quality by gel electrophoresis. Libraries for poly(A) mRNA-seq were prepared using the Capture and Amplification by Tailing and Switching" (CATS) RNA-seq library preparation kit according to manufacturer's protocol using 50 ng total RNA as the starting material. Briefly, polyadenylated mRNA was selected based on binding to oligo d(T) beads. 10 ng of mRNA was fragmented, end repaired and polyadenylated at the 3' end. Next, the poly d(T) 3' Illumina adaptor initiated cDNA synthesis. On reaching the 5' end, template switching occurred and the 5' adaptor was incorporated. The adaptors contain the P5 and P7 sequences required to cluster on the Illumina flow cell as well as the indices that are used for identifying samples when multiplexed. The library was then PCR amplified and purified using 0.9X AMPure® XP beads. The quality of the libraries was determined by running them on a Bioanalyzer and assessing the fragment length and quantity of the libraries. A 2 nM

pool of all the mRNA-seq libraries were sequenced (50bp single-end sequencing) by Diagenode, Seraing, Belgium on the HiSeq 2500.

ChIP-seq libraries were prepared using the Diagenode Microplex library preparation kit according to manufacturer's instructions. Briefly, double stranded DNA was end repaired and made blunt ended. Next the Microplex stem loop adaptors were attached on the 5' ends leaving a nick at the 3' end. The stem loop adaptors were then cleaved and the 3' ends of the genomic DNA were extended to add the Illumina indices through a high-fidelity amplification. The libraries were purified using 1X AMPure® XP beads. The quality of the libraries was determined by running them on a Bioanalyzer and assessing the fragment length and quantity of the libraries. A 2 nM pool of all the ChIP-seq libraries were sequenced (50bp single-end sequencing) by the Transcriptome and Genome Analysis Laboratory of the University of Göttingen on the HiSeq 4000.

Bioinformatic analysis

4.23 mRNA-seq data processing

Fastq files were obtained from the sequencing facility or downloaded from the NCBI GEO database (accession numbers are provided in section 3.10). The quality of the sequencing was determined using FASTQC. These were then trimmed using specific trimming conditions suggested by Diagenode. This is due to the differences in the CATS protocol such as template switching and artificial poly(A) tailing. The following command was used:

```
cutadapt --trim-n -a GATCGGAAGAGCACACGTCTG -a AGAGCACACGTCTG | cutadapt -u 3 -a A{100} --no-indels -e 0.16666666666666666 - | cutadapt -O 8 --match-readwildcards -g GTTCAGAGTTCTACAGTCCGACGATCSSS -m 18 -o -
```

The trimmed fastq files were then mapped to the hg19 version of the human genome using Bowtie2 under the TopHat module using the --very sensitive end-to-end setting. The abundances of the various transcripts in the obtained BAM files were estimated by CuffLinks and differential expression analysis between different conditions was carried out using CuffDiff. For further analysis in differential expression, those genes which showed q value ≤ 0.05, log2FC≥ 0.7 or ≤-0.7 for HCT116 and log2FC≥ 0.6 or ≤-0.6 for DLD1 and COLO320DM were used. Only those genes that showed a considerable FPKM (Fragments Per Kilobase of transcript per Million mapped reads), determined based on average FPKM values in the parental cell line were considered in further analysis.

4.24 ChIP-seq data processing

Fastq files were obtained from the sequencing facility or downloaded from NCBI GEO database (accession numbers are provided in section 3.10). The quality of the sequencing was determined using FASTQC. The first 12-13bp from the 5' end were trimmed using FASTX trimmer where necessary. The trimmed fastq files were then mapped to the hg19 version of the human genome using Bowtie 2 using the --very sensitive end-to-end setting The BAM files obtained were sorted and indexed using SAMTOOLS. These files were converted to BigWig format using BamCoverage to visualize on the IGV genome browser. The normalize to 1X setting was used and an extension length of 200bp was used. Peaks were called using the MACS2 software and BED files were obtained. The --nomodel setting was used and the FDR q value to call peaks was set at ≤ 0.05. Broad peaks were called for histone modifications and JUND and narrow peaks for the other transcription factors.

4.25 Functional analysis and integration of ChIP-seq and RNA-seq data

The GeneVenn online tool was used to create Venn diagrams between gene lists. Processed RNA-seq data was subjected to Gene Set Enrichment Analysis (GSEA) using the 'gene set' setting. Significant enrichments (FDR q value ≤ 0.05) in the MSigDB C2 (curated gene sets) and c6 (oncogenic signatures) databases were determined. Gene Ontology (GO) and pathway enrichment analysis were performed using the EnrichR software using lists of downregulated genes as input. HOMER was used to find motifs enriched in ChIP-seq datasets. Scrambled sequences of the input file were used as background. The Genomic Regions Enrichment of Annotations Tool (GREAT) was used to find region-gene associations using the default basal plus extension setting. The ReMap analysis tool was used to find colocalizing transcription regulators using a BED file as input. Heatmaps and aggregate plots were created using the reference point mode of the computeMatrix deepTools tool followed by plotProfile or plotHeatmap. Profiles and heatmaps were plotted at the center of the peaks provided in the input BED file +/- 5kb.

Statistical Analyses

Graphs in this study were designed on Microsoft Excel or GraphPad Prism version 4. For calculating statistical significance to compare parameters between different conditions, the unpaired t-test was used. P-values were determined by this test and significance depicted

as ***p ≤ 0.001, **p ≤ 0.01, *p ≤ 0.05. Statistical tests for the analysis of NGS data were performed by the indicated software.

5. Results

5.1 *ARID1A* and *KRAS* mutations tend to be mutually exclusive in colorectal adenocarcinoma

Given that the subunits of the BAF complex are mutated so frequently in cancer, research interest in the relevance of these mutations has rapidly increased. ARID1A, the most commonly mutated subunit, has been extensively described in the literature as a tumor suppressor in various cancers (Guan et al., 2011 Mathur et al., 2017, Chandler et al., 2015, Sun et al., 2017, Livshits et al., 2018). Its loss alone was shown to be a driver of colon cancer by Mathur et al. in 2017, in mice, in which a sporadic loss of *Arid1a* from an otherwise wildtype background led to the formation of invasive adenocarcinomas originating in the colon. However, as mentioned in the introduction, other reports have shown that the loss of ARID1A, in some contexts, can be oncogenic, for example in oxidative stress driven liver cancer (Sun et al., 2017).

ARID1A is a frequently mutated gene in colorectal cancer (CRC) (Cancer Genome Atlas, 2012) associated with the mucinous and microsatellite instable subtypes (Cajuso et al., 2014, Mathur et al., 2017). Therefore, initially, we wanted to obtain an overview of *ARID1A* mutations in CRC and their correlation with other commonly occurring mutations in the available patient data. For this we used the cBioPortal database for a comprehensive analysis of patient and cell line mutation data (Gao et al., 2013, Cerami et al., 2012). We observed that in the eight colorectal cancer patient datasets represented, (which recruited between 72 and 1134 patients), *ARID1A* is mutated in up to 12% of cases (Figure 10a). Most *ARID1A* mutations are missense (many of which are nonsense mutations) or truncating mutations (Figure 10b), which could lead to a loss of the expression of the protein. Indeed, it has been reported that mutations in *ARID1A* lead to the loss of the protein from tumor samples originating from various organs, detected *via* immunohistochemistry (IHC) (Wang et al., 2004, Maeda et al., 2010, Wiegnand et al., 2011, Guan et al., 2011, Jones et al., 2012).

ARID1A is also one of the most frequently mutated chromatin regulators in CRC, with only the mutation rates of KMT2B and D being slightly higher (Figure 10c). These mutation rates are comparable to those of well-described oncogenes and tumor suppressors (Figure 10d). Taken together, genetic perturbations in chromatin regulators are represented in a large fraction of CRC patients. Conceivably, chromatin regulators which play a very important role in the transcription of genes, could be an appropriate target for cancer cells to hijack. Thus,

we sought to elucidate the role that ARID1A plays in CRC. While the loss of ARID1A alone leads to the formation of invasive adenocarcinomas, its role in the context of other frequently occurring mutations remains elusive. When we looked at patient survival based on *ARID1A* mRNA expression using the TCGA colon adenocarcinoma (COAD) data, we observed no significant difference between low and high mRNA expressers of ARID1A (Figure 10e). However, in both this dataset, patients are not classified based on differences in their mutational background. Therefore, it seems that early ARID1A loss drives colon cancer, as described by Mathur et al., 2017. However, its loss in the context of other commonly occurring mutations might have a different outcome. As alluded to in the same study, Arid1a loss in the context of *Apc* mutations is not tumor suppressive. Therefore, we explored if there was any correlation between mutations in *ARID1A* and those in other common oncogenes and tumor suppressors. Very interestingly, in the TCGA Colorectal Adenocarcinoma dataset, we found that *ARID1A* mutations are mutually exclusive with mutations in the *KRAS* gene (Figure 10f). As described in the introduction, mutations in KRAS are activating and lead to the hyperactivation of the MEK/ERK pathway. Significant mutual exclusivity is also observed with *TP53* mutations and a tendency towards mutual exclusivity with *APC* mutations. This suggests that ARID1A plays a role in tumorigenesis driven by mutations in these CRC-relevant factors.

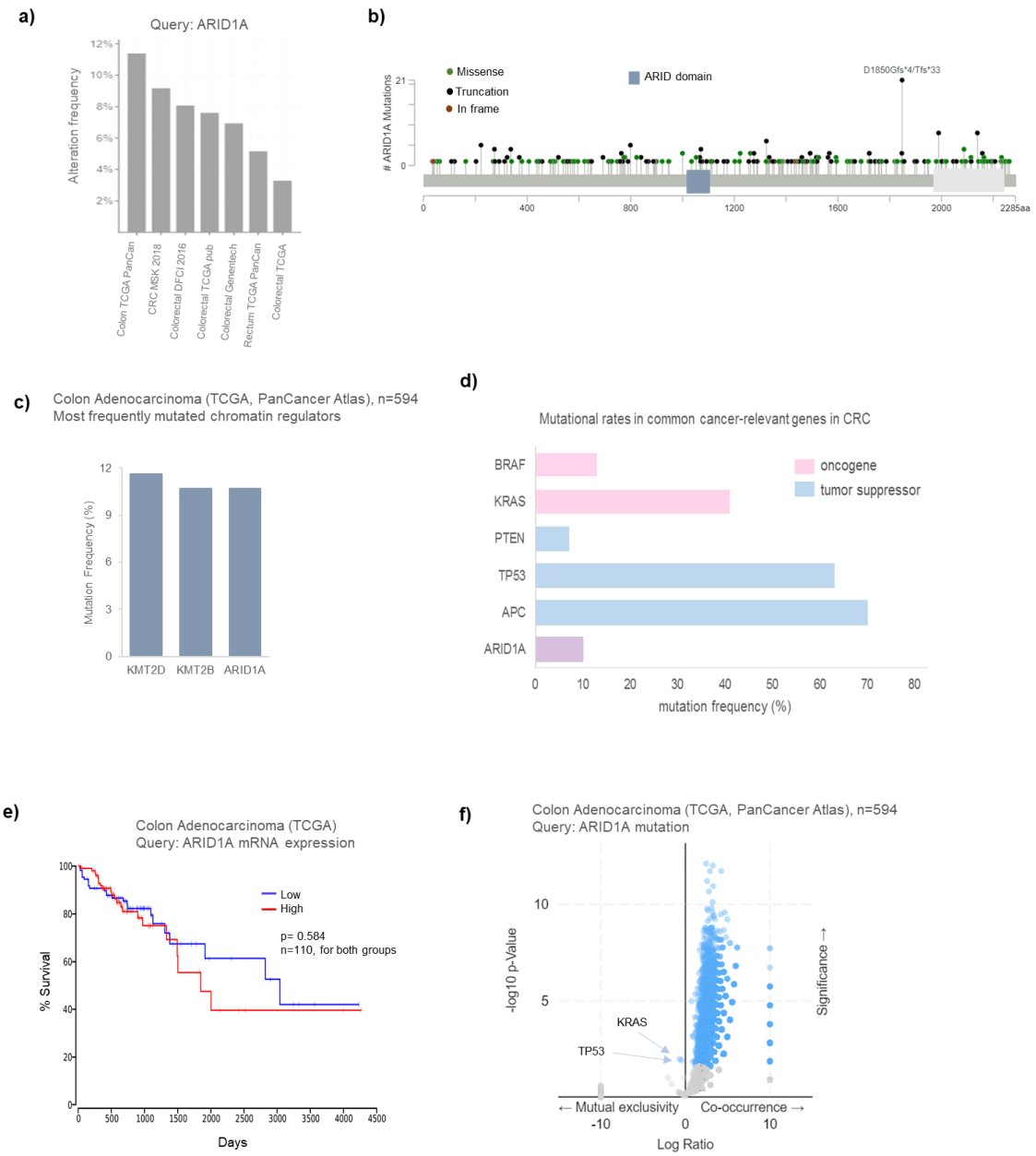


Figure 10: ARID1A status in CRC patient data. The mutation frequency of *ARID1A* in all the colorectal cancer datasets available on the cBioPortal for Cancer Genomics database. ARID1A is mutated in up to 12% of CRCs (a). The mutational landscape on the *ARID1A* gene generated using the cBioPortal tool (b). Most of the mutations are truncating or missense mutations that lead to the loss of the functional protein. *ARID1A* is one of the most frequently mutated chromatin regulators in the TCGA COAD dataset (c). *ARID1A* mutation rates as compared to the most commonly occurring mutations in colorectal cancer in all the colorectal cancer datasets available in the cBioPortal database (d). Kaplan-Meier plot depicting the % survival of the upper and lower quartile of ARID1A mRNA expressers in the TCGA colon adenocarcinoma (COAD) dataset shows that there is no correlation between ARID1A expression and survival (e). The plot was generated using the OncoLnc tool (Anaya et al., 2016). On analysis of the relationship between *ARID1A* mutations and other commonly occurring mutations we observe that *KRAS* mutations are significantly mutually exclusive in the TCGA COAD dataset (f). All mutation data related plots were plotted using the cBioPortal tools (Gao et al., 2013, Cerami et al., 2012)

5.2 Choice of Model system

Since the publication of a series of studies describing the tumor suppressive role of ARID1A, there have been numerous publications that have described targets that are synthetic lethal to the loss of ARID1A (Bitler et al., 2015, Shen et al., 2015, Samartzis et al., 2014, Miller et al., 2016, Kim et al., 2015 among others). These publications are all aimed at describing a specific vulnerability of ARID1A-deficient cells. These vulnerabilities were most often targetable by a small molecular inhibitor and have been introduced in section 2.6. Initially, we created ARID1A-deficient systems by siRNA mediated knockdown (KD) or CRISPR/Cas9-mediated knockout (KO) and tested several of the synthetic lethalities. In our CRC cell line systems (as well as some cholangiocarcinoma and pancreatic cancer cell lines), these vulnerabilities did not seem to exist. These results are presented in Supplemental Figure 1.

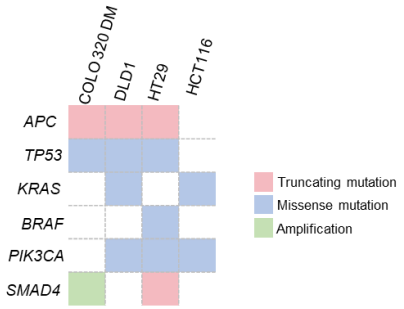
The colorectal cancer cell lines tested in this project were: COLO201, COLO320DM, DLD1, HCT116, HT29, NCIH-508, RKO, SW480, SW837, T84, $SW48^{WT}$ (SW48 cells that are wildtype for KRAS) and $SW48^{G12D}$ (SW48 cells that have a heterozygous knockin of a KRAS activating mutation (G12D)). To choose appropriate models from these cell lines, we determined the levels of the subunits of the BAF complex. These results were presented in my master's thesis in 2016 and are presented in the Introduction in this thesis (Figure 8 (a,b))

According to the CCLE data, RKO and SW48 have truncating mutations in ARID1A. T84 and SW48 harbor SMARCA2 deletions and SMARCB1 is amplified in HCT116. A missense mutation for SMARCA4 was seen in SW48 (Figure 11a). This information was compared to the mutations described for the available cell lines in the NCI-60 Human Tumor Cell Lines Screen (Reinhold et al., 2012). Certain inconsistencies were observed between the two. For example, in HCT116 cells there was a truncating mutation in SMARCA4 and a missense mutation in SMARCA2 in the NCI-60 screen which was not represented in the CCLE data. Similarly, an amplification of SMARCB1 in HCT116 cells was not observed in the NCI-60 screen. Therefore, these were used to attain a general overview of mutational status to determine appropriate cell lines for experiments; however, their accuracy could not be tested. The protein expression of certain BAF complex subunits was determined by western blot analysis (Figure 11b). The HCT116, COLO320DM, DLD1, HT-29, SW480 and T84 cell lines expressed all the tested components, ARID1A, SMARCA2, SMARCA4, SMARCB1, SMARCC1 and SMARCC2. SMARCA4 expression was absent in COLO201, NCIH-508, SW837 and SW48G12D. ARID1A expression was absent in COLO201 and NCIH-508.

Even though the CCLE data revealed truncating mutations in ARID1A for RKO and SW48, both these cell lines expressed ARID1A, though to a lesser extent than the other ARID1A expressing cell lines. These could be non-functional protein products that are still expressed despite the truncating mutation.

Four cell lines which expressed all the tested subunits of the BAF complex and showed variations in commonly occurring mutations were chosen for further experiments. While COLO320DM, HT29, DLD1 were mutant for *TP53* and *APC*, HCT116 was wildtype for these genes (Figure 11c). Since we were most interested in the relationship between *KRAS* mutations and *ARID1A* mutations, we included 2 cell lines with G13D mutations (HCT116 and DLD1) in *KRAS* and two cell lines that were wildtype for *KRAS* (COLO320DM and HT29). It is, however, important to note that, the HT29 cell line harbors a mutation in the *BRAF* gene which is also a player in the MEK/ERK pathway. However, the consequence of this mutation is not always the same as that of the KRAS mutation (Morkel et al., 2015), even though it leads to defects in the same pathway. Lastly, we narrowed down on the HC116 system as our main model system in this study due to the vast amount and range of Next Generation Sequencing (NGS) data publicly available in this system. Not only is a very large number of ChIP-seq datasets available in this system, a lot of ARID1A research has been done in this system, making datasets in the *ARID1A* knockout (KO) available. Among these are, BAF complex subunit ChIP-seq in the wildtype and *ARID1A* KO system, several important transcription factors such as TCF7L2 (downstream of the Wnt pathway), AP1(FosL1, JunD, downstream of the MEK/ERK pathway) and BRD4 (an interesting regulator of enhancer activity). Since we were studying a chromatin remodeller, it was also highly advantageous to have access to ATAC-seq data in the HCT116 *ARID1A* wildtype (WT) and KO systems. Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) assesses the accessibility of chromatin, based on the cleavage of accessible DNA by a hyperactive transposase and attachment of sequencing adaptors at the ends of the cleaved DNA. A comprehensive list of the sequencing data used in this study can be found in Figure 11d. Therefore, the HCT116 cell line presented us with a system in which an integrated investigation could be performed and thus several experiments were performed in this system.

c)



d) Publicly available datasets in the **HCT116** system used in this study

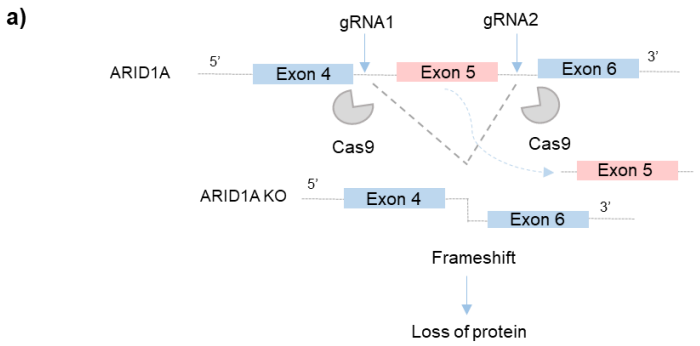
ChIP-seq and ATAC-seq	**GEO Accession**	**Reference**
SMARCC1 (WT and *ARID1A* KO) SMARCA4 (WT and *ARID1A* KO) H3K27ac (WT and *ARID1A* KO)	Series GSE71510	Mathur et al., 2017
H3K27me3 (WT and *ARID1A* KO) ATAC-seq (WT and *ARID1A* KO)	Series GSE101966	Kelso at al., 2017
TCF7L2	GSM782123	ENCODE
BRD4	GSM2058664	Baranello et al., 2016
CTCF	GSM1224650	ENCODE
FOSL1	Series GSE32465	ENCODE
JUND	Series GSE32465	ENCODE
RNA-seq		
WT and *ARID1A* KO	Series GSE71511	Mathur et al., 2017
Hi-C	Series GSE104333	Rao et al., 2017

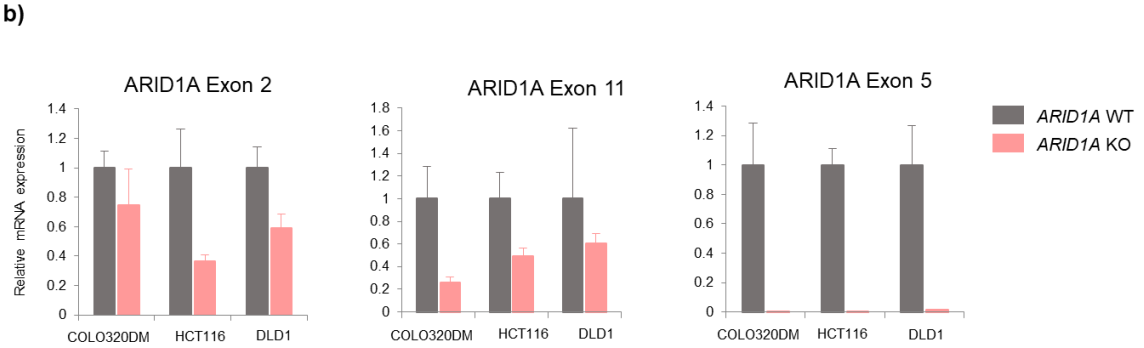
Figure 11 (linked to Figure 8): Choice of *in vitro* human CRC model systems. Four cell lines namely, COLO320DM, DLD1, HT29 and HCT116 were selected based on expression of all BAF complex subunits (refer to Figure 10(a,b) in section 2.12) and a variety in the mutational background of commonly occurring mutations for further experiments (c). This table was generated using the cBioPortal database, ExPASy Cellosauraus database and information from Ahmed et al., 2013. The HCT116 system was chosen as the primary system in our study due to the large amount of publicly available data generated in this system which is very useful for an integrated analysis. The datasets available in the HCT116 system and used in this study are listed along with their NCBI GEO accession numbers in (d). All other sequencing datasets were generated by us for this study. cBioPortal for Cancer Genomics tools (Gao et al., 2013, Cerami et al., 2012).

5.3 The ARID1A protein is completely lost after CRISPR/Cas9-mediated knockout of the *ARID1A* gene

When we looked at patient survival based on *ARID1A* mRNA expression using the TCGA colorectal adenocarcinoma dataset, we observed no significant difference between low and high expressers of *ARID1A*. This was contrary to the established tumor suppressive role of ARID1A described in CRC. However, since these patients harbor a variety of other mutations, the outcome of ARID1A loss in these contexts could vary. It seems early ARID1A loss drives colon cancer, however its occurrence in the context of other commonly occurring mutations might have a different outcome. To explore this, we used *in vitro* systems to mimic the loss of ARID1A in cancer. We generated 4 *ARID1A* KO colorectal cancer cell lines, HCT116, HT29, COLO320DM, DLD1 (which show variations in mutational background, Figure 10c). We continued to use the HCT116 system as our primary system due to the reasons explained in the previous section. We used CRISPR/Cas9 mediated gene editing to knock out ARID1A. We targeted two introns flanking exon 5 of *ARID1A* with two guide RNAs. We ensured that the number of base pairs deleted would yield an out of frame

product on transcription of the truncated gene (Figure 12a). We transfected the four cell lines with a plasmid containing the gRNAs and Cas9. We sorted the cells that were GFP-positive as single cell clones. We then expanded these clones for each cell line and screened them for KO of *ARID1A* by genotyping PCR, qRT-PCR for mRNA and western blot for protein. At the transcript level, we observed that the KO of exon 5 does not lead to a block in transcription of the gene, as primers for *ARID1A* designed at exon 2 and exon 11 still show some mRNA products in the KO cells of three of the four tested cell lines (Figure 12b). However, qRT- PCR with primers designed at exon 5 showed a complete loss of the transcript (Figure 12b). The production of a truncated product probably leads to nonsense mediated decay, or degradation of the truncated or non-functional protein and therefore a complete loss of the protein from the cells, as can be seen for all four cell lines in Figure 12c. Thus, we were able to produce model systems to study the lack of ARID1A in the context of other commonly occurring mutations.





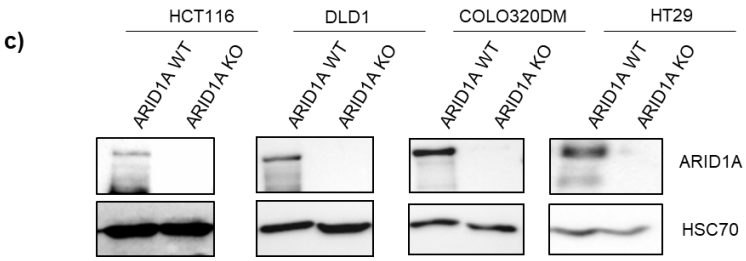


Figure 12: CRISPR/Cas9-mediated deletion of ARID1A in colorectal cancer cell lines. CRISPR/Cas9 mediated gene editing was used to knock out *ARID1A* from four colorectal cancer cell lines with varying mutational backgrounds. For this two guide RNAs flanking exon 5 of *ARID1A* were used (exact sequences and genomic positions are provided in section 3.13). The region flanked, once deleted ensures a frame shift leading to the loss of the protein from the cells (a). The gRNAs and Cas9 containing plasmids were transfected by electroporation and the cells were FACS sorted into single cell clones. The GFP-positive clones were screened for KO by genotyping PCR (not shown), qRT-PCR and by western blot. A single cell clone for each cell line was selected for further experiments. As can be seen in (b) the deletion of exon 5 did not lead to a complete loss of the *ARID1A* mRNA in 3 of the tested cell lines. qRT-PCRs against exon 2 showed minor depletion whereas qPCRs against exon 11 showed higher but still not complete loss. Only qRT-PCRs against exon 5 showed a complete depletion (b). At the protein level, ARID1A was lost from all four cell lines confirming the knock out and providing a system that mimics the human condition (c). qRT-PCRs were run in biological triplicates and technical duplicates.

5.4 ARID1A loss leads to an impairment in proliferation of *KRAS* mutant CRC cell lines

After obtaining the ARID1A-deficient systems, we sought to characterize them phenotypically. As can be seen in Figure 12c, the knockout of *ARID1A* led to the complete loss of the protein from the cells mimicking the human condition where ARID1A expression is lost from the tumors. Surprisingly, contrary to its described tumor suppressive role, we observe that in two cell lines (HCT116 and DLD1), the knockout of *ARID1A* led to a severe impairment in proliferation. In the other two cell lines (COLO320DM and HT29) proliferation was not affected by the loss of ARID1A. Interestingly, the cell lines which showed defects in proliferation (as well as changes in morphology), harbored mutations for *KRAS* (G13D) whereas the cell lines which showed no defect were WT for *KRAS*. Consistently, *KRAS* mutations occur in ~40% of CRC (cBioPortal for Cancer Genomics) and are mutually exclusive with ARID1A mutations (Figure 10f). As can be seen in Figure 13(a,b), the proliferation and morphology of the *KRAS* WT cells COLO320DM and HT29 does not change upon knockout of *ARID1A* (Figure 13(c,d)). The proliferation was estimated by measuring the confluence of the cells (equal numbers seeded) every day for 5 days. In the case of the HCT116 cell line, the *ARID1A* KO cells were smaller and had more filopodia-like projections on their surface (Figure 13c). The morphology of the *ARID1A* KO cells was very similar to that of the *ARID1A* KO cells described by Mathur et al., 2017, who used a completely different strategy to knock out the gene. This implied to us that the change is probably specific to ARID1A loss. Moreover, these cells showed a 21.4% reduction in relative confluence and their proliferation was severely impaired (Figure 13c). Similarly, the DLD1 cell line with *ARID1A* KO seemed to grow in more compact colonies. Moreover, these cells too, showed a significant proliferation defect and their relative confluence was reduced by 12.2% (Figure 12d).

Interestingly, the cell lines which showed a proliferation defect also harbored Wnt pathway mutations. The HCT116 cell line harbors a *CTNNB1* (encoding β-catenin) mutation whereas the DLD1 cell line harbors an *APC* mutation (cBioPortal for Cancer Genomics). Defects in this pathway are the most common perturbation in CRC and these cancers are most often highly dependent on this pathway. This is in line with the proposed oncogenic role of ARID1A in Wnt-defective CRCs (Mathur et al., 2017, 2018). Interestingly, the HT29 cell line also harbors a mutation in *APC* and does not seem to proliferate less in the absence of ARID1A, which suggests that it might not be completely dependent on the Wnt pathway. We explored this dependency at the transcriptional level, by looking at the colocalization of ARID1A with the downstream effector of the Wnt pathway, TCF7L2 and gene expression changes upon its KD. These results are presented in detail in Supplemental Figure 2.

Thus, in the context of other commonly occurring mutations, ARID1A is not tumor suppressive. We indicated that CRCs that are *KRAS* mutated require the expression of ARID1A to maintain tumorigenic properties such as proliferation. This was consistent with our results that showed that *KRAS* and *ARID1A* mutations are mutually exclusive events in colorectal adenocarcinoma (TCGA dataset). Together, this suggests that *ARID1A* and *KRAS* mutations employ different pathways in driving tumorigenesis and ARID1A is required in MEK/ERK signaling mediated CRC cell lines.

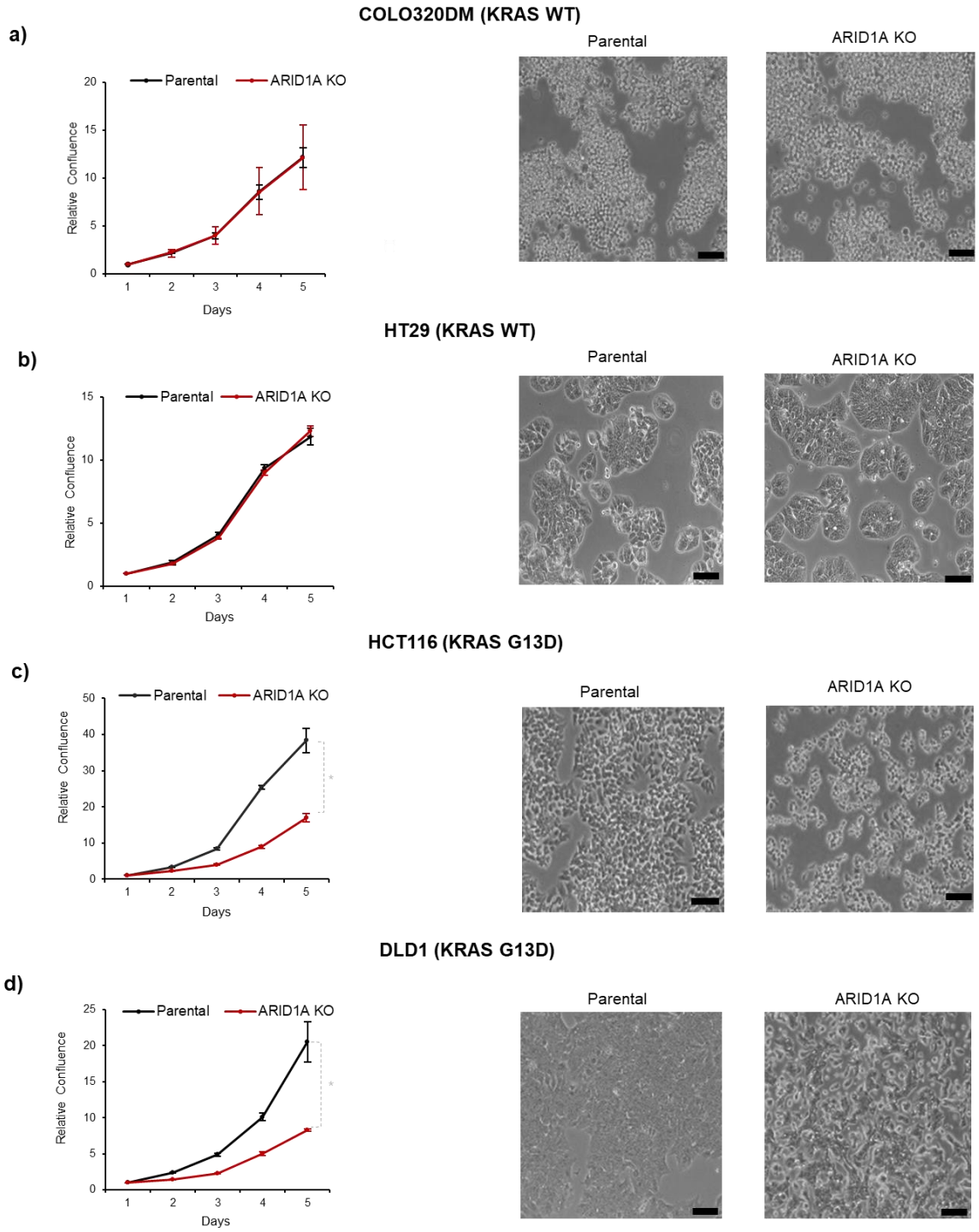


Figure 13: Phenotypic changes upon *ARID1A* knock out. Exon 5 of the *ARID1A* gene was knocked out using CRISPR/Cas9 gene editing as previously described. A single cell clone lacking ARID1A for each cell line was selected for further experiments. These clones were expanded, and their phenotype observed. Morphologically, The COLO320DM and HT29 *ARID1A* KO clones were identical to their parental lines (a,b), while the HCT116 and DLD1 *ARID1A* KO lines showed morphological changes. While the HCT116 *ARID1A* KO cells were smaller and showed filopodia like projections (c), the DLD1 KO cells seemed to grow in more compact colonies (d). Proliferation assays was carried out over 5 days to measure difference in proliferation of the *ARID1A* KO cells. While the KRAS WT cell lines (COLO320DM, HT29) showed no proliferation changes upon KO of *ARID1A (a,b)*, the *KRAS* G13D mutated cell lines (HCT116 and DLD1) showed significant impairment in proliferation upon KO of *ARID1A (c,d)*. n for proliferation assay= 2, representative of 3 independent experiments, * $p<0.05$ unpaired t-test, scale bar=100 µM.

5.5 ARID1A loss in *in vivo* models in the context of the tumor microenvironment

Since most of the literature until recently described ARID1A as a tumor suppressor, we wanted to ensure that the oncogenic effect we observed in our tested cell lines was not an artefact of an *in vitro* system, where lack of the tumor microenvironment had yielded a different result. Therefore, initially, we used HT29 *ARID1A* KO cells to inject subcutaneously as xenografts in SCID Hairless outbred mice, as these cells showed no proliferation defect. This experiment was carried out with Dr. Florian Wegwitz, Dr. Robyn Kosinsky and Sabine Bolte, University of Göttingen. Two clones of *ARID1A* KO cells were used to negate clonal variations. The genotyping PCRs of the two clones show that the *ARID1A* gene was truncated (Figure 14a). *ARID1A* KO HT29 cells were injected into the left flank of 11 mice. *ARID1A* WT HT29 cells were injected into the right flank of all 11 mice. 5 million cells in Matrigel were initially injected and the growth of the tumor was monitored over the next 21 days. Even within the context of a microenvironment, we observed that the *ARID1A* KO cells did not form more aggressive tumors than their WT counterparts. A representative picture with equally sized tumors (harvested from the same mouse) is shown in Figure 14b. On analysis of the harvested tumors, final weight and volume of the *ARID1A* KO tumors did not differ significantly from their WT counterparts. Moreover, the growth over time also did not differ (Figure 14c).

In another *in vivo* model, we used the Cre-ER^{T2} LoxP system. The *Arid1a* gene was flanked by LoxP sites (floxed) and the Cre recombinase ER^{T2} fusion protein was placed under the control of the colon specific C*dx2* promoter. The conditional knockout was generated by the administration of Tamoxifen which is an agonist for the ER^{T2} receptor that activates the Cre recombinase. Moreover, some mice in our cohort had an *Apc* loss background, where the *Apc* gene was also either heterozygously or homozygously floxed. We monitored the health of these mice over 6 months at the end of which no tumors were found in any genotype. All these experiments were performed with the help of Dr. Robyn Kosinsky. The results are presented in Supplemental Figure 3.

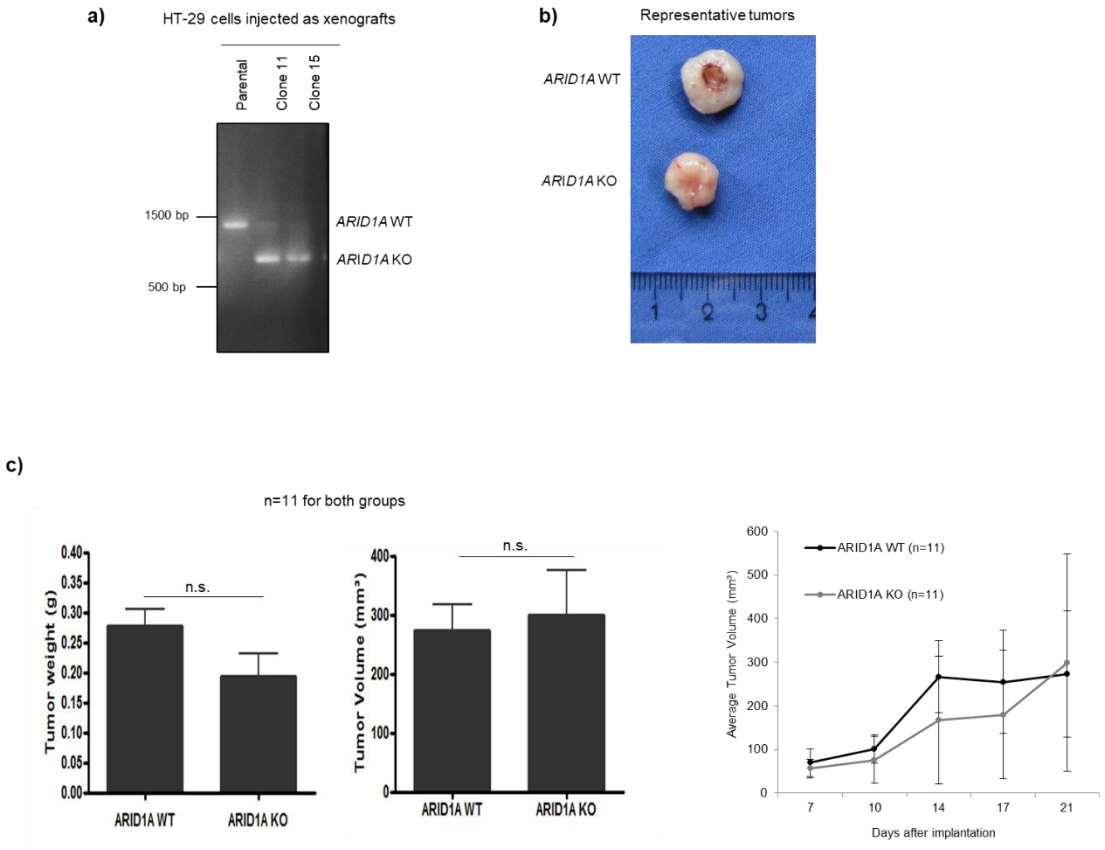
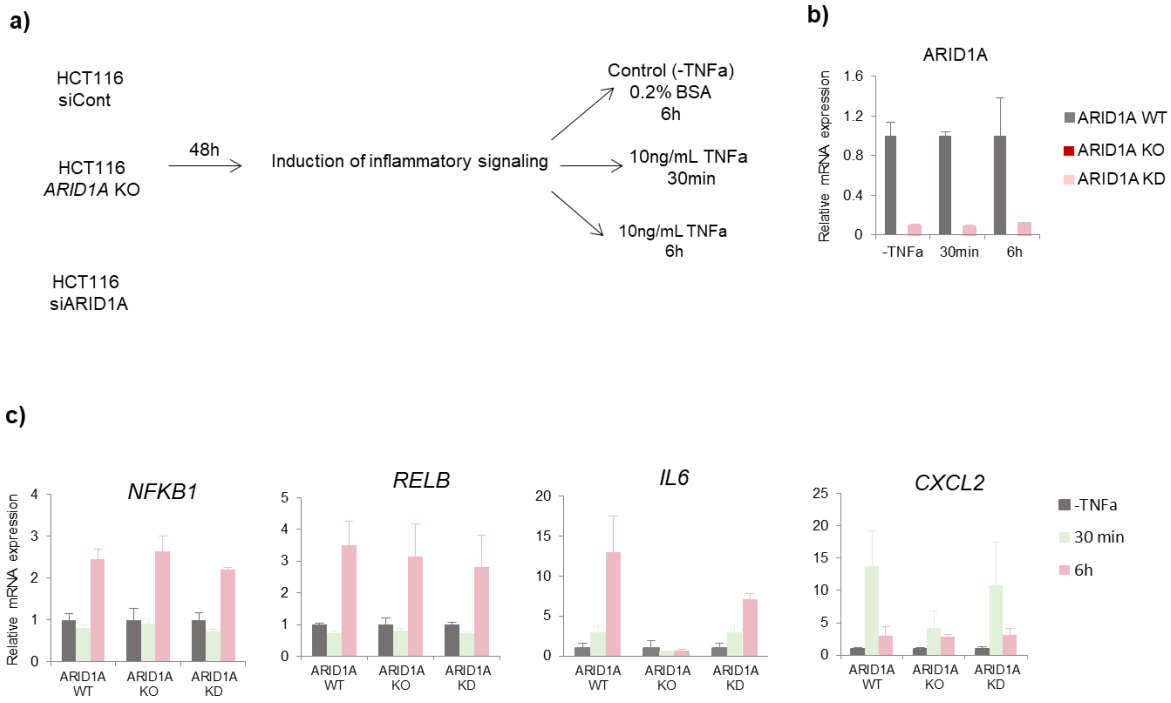


Figure 14: *In vivo* model for ARID1A loss. HT29 cells were injected subcutaneously into SCID Hairless Outbred mice as xenografts. HT29 *ARID1A* WT cells were injected on the right flank of the mouse whereas *ARID1A* KO cells were injected on the left flank. Two KO clones were used to rule out clonal variations. Genotyping PCRs show that these clones showed a KO of deleted exon (a). The growth of the tumors was measured for 3 weeks, every 3-7 days. The mice were sacrificed when the tumors reached a size of 1cm X 1cm or more (b). The tumor volume and weight were measured. We observed no significant differences in final tumor weight and volume between the WT and *ARID1A* KO groups (c). Moreover, tumor growth over time between the *ARID1A* WT and KO tumors did not differ significantly. n=11, right WT, left *ARID1A* KO, n.s. not significant $p>0.05$, unpaired t-test.

5.6 ARID1A loss dampens the inflammatory environment elicited by NFκB signaling

Chandler et al., 2016, published a study which showed a direct relationship between the tumor suppressive role of ARID1A and the inflammatory tumor microenvironment in the ovarian cancer system. In this study, it was proposed that ARID1A represses the promoter of the pro inflammatory cytokine *IL6*. IL6 is known to play many roles in cancer including metastasis of CRC (Rokavec et al., 2014). Thus, when ARID1A is lost, IL6 is expressed in an uncontrolled manner and creates a pro-inflammatory environment which supports oncogenesis. Expanding on this, we aimed to further explore this phenomenon in our system. To this end, we used the HCT116 system (WT, *ARID1A* KO and *ARID1A* KD) and treated them with the inflammatory cytokine TNFα which induces IL6 to create an inflammatory environment. We treated the cells for 30 min or 6 h to elicit early and late gene expression responses (Figure 15a). Figure 15b shows the relative loss of *ARID1A* in the KDs and KOs. Minimal levels of *ARID1A* are expressed after knockdown. On stimulation with TNFα we found that NFKB pathway was induced as seen by the induction of *NFKB1* and *RELB* (Figure 15c). We observed that the induction of *IL6* was strongest after 6 h in the WT condition and contrary to the published study, this response was dampened upon the loss of ARID1A. A similar effect was seen with the early response inflammatory cytokine *CXCL2*. *CXCL2* is expressed downstream of the NFκB pathway and is induced after 30 min as seen in Figure 15c. However, in this case as well, the induction of *CXCL2* was dampened upon the loss of ARID1A. Therefore, it seems, even in the context of the tumor



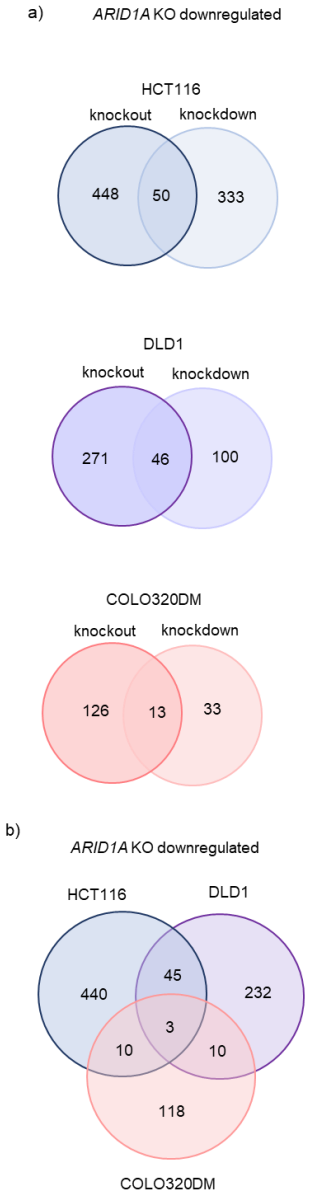
microenvironment, ARID1A does not elicit a heightened inflammatory network and is still oncogenic in our model of colorectal cancer.

Figure 15: ARID1A loss in the context of the tumor microenvironment. To mimic the inflammatory signaling in tumors, HCT116 cells were treated with TNFα. HCT116 WT, *ARID1A* KO and cells with *ARID1A* depletion (siRNA mediated KD) were treated with vehicle or 10ng/mL TNFα for 30 min or 6h (a). The loss of ARID1A from the cells occurred as expected, with complete loss of *ARID1A* mRNA in the KOs and a substantial depletion in the KDs. (b) TNFα stimulation was confirmed by increased *NFKB1* mRNA levels after 6h (c). The increase in *NFKB1* levels was similar in all three conditions. The activation of *RELB* consistently showed a similar trend (c). The activation of *IL6* was strongest in the WT condition at 6h and its levels were dampened in the KO and KD conditions (c). *CXCL2*, an early response gene, showed activation at 30 min, however, in this case as well, the activation was dampened in the KO condition and remained unchanged in the KD condition. qRT-PCRs run in biological triplicates and technical duplicates.

5.7 The depletion of ARID1A has varied effects in colorectal cancer cell lines

To elucidate the apparent oncogenic role that ARID1A plays in colorectal cancer, we performed RNA-sequencing to determine gene expression patterns in three colorectal cancer cell lines with ARID1A depletion (either knockdown or knockout). We included cell lines WT and mutant for *KRAS*. For the HCT116 WT and *ARID1A* KO we used a publicly available dataset. We were mainly interested in the genes that are downregulated by the loss of ARID1A. Since ARID1A is part of a chromatin remodeller and associated with active gene transcription (Wu et al. 2013, Mathur et al., 2017), these are likely to be more direct targets. Genes that are upregulated by the loss of ARID1A are likely due to secondary effects. However, as described in the Introduction, repressive functions of the BAF complex have been defined (Chandler et al., 2015, Wu et al., 2018) Initially, we overlapped the genes downregulated upon *ARID1A* KD and *ARID1A* KO in each cell line. The overlap of these genes in each cell line was surprisingly small, between 9% and 30% (Figure 16a). This could be due to the persistent loss of ARID1A in the KOs, leading to more long term secondary effects which are different from the immediate transient effects. We decided to continue our analysis with the comparison of the WT vs. the KO data, as this is also more representative of the clinical condition of persistent loss. The overlap of genes downregulated by *ARID1A* KO was very small between the three cell lines. However, very interestingly, the responses of the *KRAS* mutant cell lines (HCT116 and DLD1) to the loss of ARID1A were more similar to each other compared with the WT *KRAS* cell line (COLO320DM). This can be seen in the overlap of the genes downregulated by *ARID1A* KO in the three cell lines. There is a much higher overlap of genes between the HCT116 and DLD1 cell lines than any of these cell lines with the *KRAS* WT COLO320DM (Figure

16b). This was consistent with our proliferation assays which showed a dependency of the *KRAS* mutant cell lines to ARID1A. Moreover, we performed gene ontology (GO) for the sets of genes downregulated in each cell line. Figure 16c contains the 10 most significantly enriched GO terms in these gene sets. As expected, in the HCT116 cell line, among the enriched terms were “epidermal growth factor receptor signaling pathway” and “regulation of epithelial cell proliferation” (Figure 16c). While the former once again indicates the involvement of ARID1A in the MEK/ERK pathway, the latter suggests the ARID1A plays a role in the regulation of cell proliferation, which could explain the proliferation defect that we observe in the *ARID1A* KOs of this cell line. In the DLD1 cell lines, most of the enriched GO terms were involved in cytoskeletal structure, function and cellular movement. This could perhaps explain the change we observed in the morphology of the *ARID1A* KO DLD1 cells. The mis-regulation of genes involved in the cytoskeletal arrangement perhaps leads to the more compact colonies seen in the DLD1 cell line with *ARID1A* KO (Figure 16c). In the COLO320DM cell line, several GO terms related to cardiac cell differentiation and the Wnt pathway were enriched (Figure 16c). This cell line (like for proliferation) differed from the other two cell lines. However, fascinatingly, the Wnt pathway is also a very important player in cardiac differentiation and ARID1A perhaps plays a role in this pathway. Overall, we saw the gene expression changes induced by the KO and KD of *ARID1A* vary quite significantly between cell lines and depends on the depletion method used. However, the two *KRAS* mutant cell lines are more similar to each other than to the *KRAS* WT cell line. Furthermore, this provided another clue towards the involvement of ARID1A in the MEK/ERK pathway.



c)

HCT116 *ARID1A* KO downregulated GO Terms
regulation of transcription from RNA polymerase II promoter (GO:0006357)
response to endoplasmic reticulum stress (GO:0034976)
positive regulation of transcription from RNA polymerase II promoter (GO:0045944)
epidermal growth factor receptor signaling pathway (GO:0007173)
hippo signaling (GO:0035329)
positive regulation of transcription, DNA-templated (GO:0045893)
intrinsic apoptotic signaling pathway (GO:0097193)
intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress (GO:0070059)
regulation of epithelial cell proliferation (GO:0050678)
regulation of fat cell differentiation (GO:0045598)

DLD1 *ARID1A* KO downregulated GO Terms
positive regulation of chemotaxis (GO:0050921)
regulation of chemotaxis (GO:0050920)
actin filament bundle assembly (GO:0051017)
actin filament bundle organization (GO:0061572)
platelet degranulation (GO:0002576)
regulation of actin cytoskeleton organization (GO:0032956)
extracellular matrix organization (GO:0030198)
positive regulation of locomotion (GO:0040017)
negative regulation of tumor necrosis factor-mediated signaling pathway (GO:0010804)
regulation of cytoskeleton organization (GO:0051493)

COLO320 *ARID1A* KO downregulated GO Terms
positive regulation of cardiac muscle cell differentiation (GO:2000727)
positive regulation of cardiocyte differentiation (GO:1905209)
regulation of cardiac muscle cell differentiation (GO:2000725)
hair cell differentiation (GO:0035315)
positive regulation of cell differentiation (GO:0045597)
cGMP-mediated signaling (GO:0019934)
regulation of relaxation of cardiac muscle (GO:1901897)
regulation of Wnt signaling pathway, planar cell polarity pathway (GO:2000095)
regulation of leukocyte degranulation (GO:0043300)
cAMP catabolic process (GO:0006198)

Figure 16: Gene expression changes upon *ARID1A* KO and KD. We performed RNA-sequencing in three colorectal cancer cell lines with *ARID1A* KO or KD. For the HCT116 cell line we only performed RNA-seq in the *ARID1A* KD condition and used a publicly available dataset for the HCT116 *ARID1A* KO system (Mathur et al., 2017). There was a moderate overlap of genes downregulated by *ARID1A* KO or KD in the different cell lines (a). However, the overlap of genes downregulated by *ARID1A* KO between the cell lines was not very significant. The HCT116 and DLD1 (*KRAS* mutant) cell lines show a greater overlap (b). We looked for Gene Ontology terms that are enriched in the set of genes downregulated by ARID1A KO in each cell line using the EnrichR software (Chen et al., 2013, Kuleshov et al., 2016). The biological processes enriched varied in the three cell lines quite significantly (c). In the HCT116 cell line, cancer relevant pathways such as EGFR signaling, and epithelial cell proliferation were enriched. In the DLD1 cell line, pathways related to extracellular matrix and cytoskeleton were enriched. In the COLO320DM cell line the Wnt pathway and pathways related to cardiac cell differentiation were enriched. n=3 for RNA-seq (performed for this study), n=2 (publicly available dataset), log2FC ≤ -0.7 for HCT116, ≤ -0.6 for DLD1 and COLO320DM, q ≤ 0.05.

5.8 ARID1A is involved in the downstream transcriptional network of the MEK/ERK pathway

Since we hypothesized that ARID1A plays a role in the regulation of the MEK/ERK pathway, we wanted to know whether this regulation was at the signaling level or the transcriptional level. Initially, to confirm our hypothesis, we ran gene set enrichment analyses on the set of genes downregulated upon *ARID1A* KO in the two *KRAS* mutant cell lines, HCT116 and DLD1. In the test for oncogenic signatures we saw an enrichment of gene sets involved in the MEK/ERK pathway for both cell lines (Figure 17(a,b)). In the HCT116 cell line, the EGFR_UPv1_UP (genes that are upregulated upon an overexpression of the MEK/ERK pathway receptor EGFR) gene set was enriched in the WT cells. Moreover, the MEK_UPv1_UP (genes that are upregulated upon an overexpression of MEK) gene set was also enriched in the WT cells. Additionally, in the DLD1 cell line the KRAS300_UPv1_UP (genes upregulated upon overexpression of KRAS) were enriched in the WT condition. Together, these suggested that upon the loss of ARID1A, the genes that are expressed as a result of hyperactive MEK/ERK signaling are no longer expressed, thus suggesting that ARID1A is required for the expression of these genes. Subsequently, we also performed a pathway enrichment for the set of *ARID1A* KO downregulated genes using the EnrichR software. In both cell lines we found genes involved in the MEK/ERK pathway to be enriched. Interestingly, the pathway “AP-1 transcription network” was enriched in both cell lines (Figure 17c). AP-1 transcription factors are activated upon MEK/ERK signaling and act downstream of this pathway. This suggested that ARID1A perhaps plays a role at the transcriptional regulation level of this pathway. We confirmed this by western blot for pERK in the *ARID1A* WT and KO cells. If the regulation was at the signaling level (through an indirect mechanism), the activating phosphorylation of ERK would be reduced upon the knockout of ARID1A. Since this is not the case in both cell lines (figure 16d), and ARID1A is a chromatin regulator, we hypothesized that ARID1A regulates the MEK/ERK pathway at the transcriptional level.

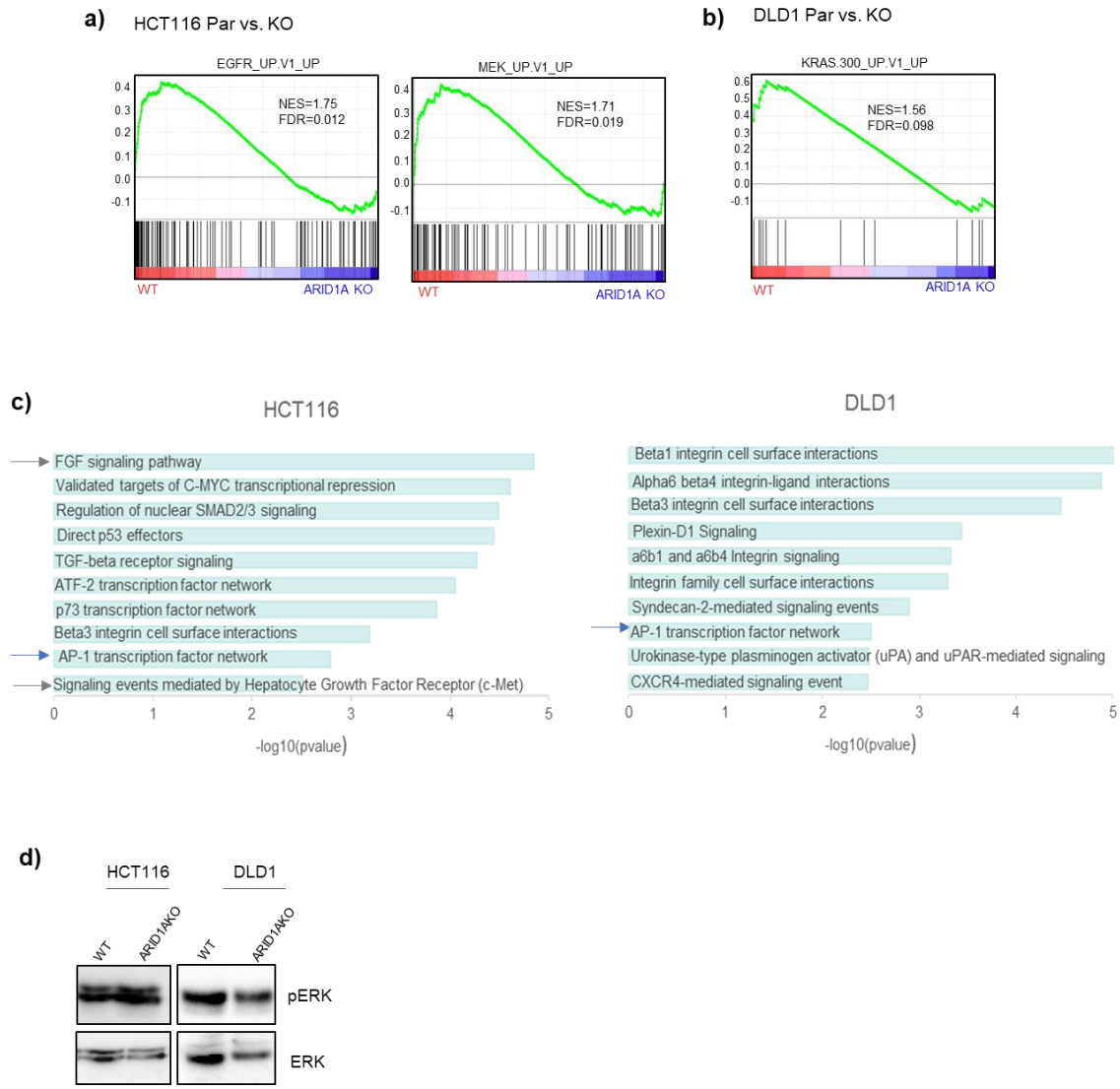


Figure 17: The genes downregulated upon *ARID1A* KO are involved in the MEK/ERK pathway. On further analysis of the RNA-seq data by gene set enrichment analysis (GSEA) or pathway enrichment using the EnrichR software, it was seen that the genes downregulated by *ARID1A* KO are enriched in the MEK/ERK pathway (a, b). In the HCT116 cell line GSEA shows that genes upregulated upon EGFR or MEK overexpression are enriched in the WT condition (a). Similarly, in the DLD1 cell line genes upregulated upon KRAS overexpression are enriched in the WT condition (b). Pathway enrichment analysis revealed that genes downregulated by *ARID1A* KO are involved in MEK/ERK related signaling cascades and its downstream AP-1 transcription network in both cell lines (c). This effect is indeed at the level of transcription and not *via* a disruption of the signaling pathway, as can be seen by the unchanged levels of pERK in the *ARID1A* KO system (d).

5.9 The *ARID1A* KO cells are slightly more sensitive to MEK1/2 inhibition

Since we observed a dependency of CRC cell lines (driven by the MEK/ERK pathway) on ARID1A, we wanted to test the sensitivity of these cell lines to the inhibition of this pathway by using a MEK1/2 inhibitor, Trametinib. Trametinib is FDA approved for the treatment of *BRAF* mutated melanoma, non-small cell lung cancer and aplastic thyroid cancer. In combination with other drugs, it is also in several Phase I/II clinical trials for colorectal cancer (https://clinicaltrials.gov). We measured the sensitivity of the four CRC cells by measuring the proliferation-associated intensity of crystal violet staining after 5 days of continuous Trametinib treatment. While all four cell lines were quite sensitive to Trametinib treatment, we expected that the cell lines that were *KRAS* mutant (or *BRAF* mutant) would be more sensitive to this treatment as they were more dependent on the MEK/ERK pathway. However, this was not the case. The half maximal inhibitory concentration (IC_{50}), that is the concentration at which confluence of the cells is reduced by half compared to the control condition, ranged from ~2.5 nM to ~100 nM. Expectedly, the BRAF mutant HT29 cell line was most sensitive to Trametinib treatment, followed by the KRAS mutant HCT116 cell line. Surprisingly, the KRAS mutant DLD1 cell line was the most resistant to the treatment with Trametinib (Figure 18a). This could be due to a cell line specific resistance mechanism. To test the efficiency of the inhibitor at attenuating the MEK/ERK signaling, we detected the levels of pERK (phosphorylated ERK) after treatment in the WT and the *ARID1A* KO HCT116 cell line. ERK is phosphorylated at threonine 202 and tyrosine 204 in its activating loop upon the induction of the kinase cascade of the MEK/ERK pathway, and its levels indicate the activation of the pathway. Therefore, as we can see in Figure 18b, even very low doses of Trametinib reduce the levels of pERK very significantly thus blocking this pathway. We next sought to determine the sensitivity of the *ARID1A* KOs to Trametinib treatment compared to the WT cells. All four cell lines showed an increased sensitivity to Trametinib upon the KO of *ARID1A* (Figure 18c). This could be due to the hypothesized cooperative role that ARID1A plays with the AP1 factors downstream of the MEK/ERK pathway leading to an additive reduction in proliferation upon loss of transcriptional activity by both factors.

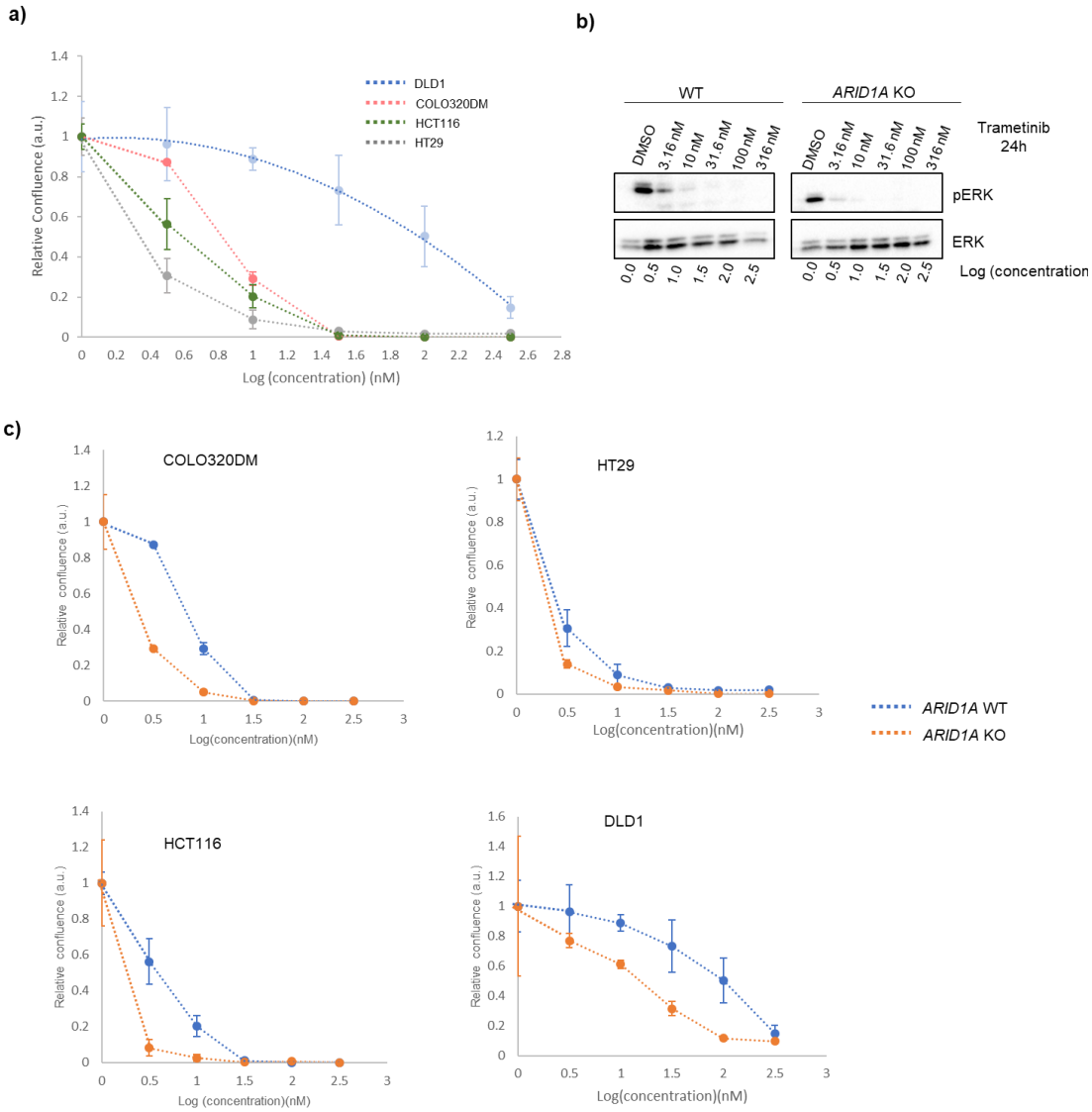


Figure 18: ARID1A KO cells are more sensitive to MEK1/2 inhibition by Trametinib. The responsiveness of the cell lines indicated to the MEK1/2 inhibitor Trametinib was tested. These proliferation assays are based on crystal violet staining after 5 days of inhibitor treatment. The HT29 cell line is most sensitive to Trametinib treatment with an IC_{50} of 2.5 nM while the DLD1 cell line was least sensitive to it with an IC_{50} of 100 nM (a). The efficacy of Trametinib at attenuating the MEK/ERK pathway was determined by the levels of pERK (phosphorylated ERK, at Thr202/Tyr204 of the activating loop) in the HCT116 WT and *ARID1A* KO system. As can be seen in (b) Trametinib very effectively reduced the levels of pERK. Moreover, in all four cell lines the *ARID1A* KO cells were slightly more sensitive to Trametinib treatment (c).

5.10 Successful chromatin immunoprecipitation for ARID1A

Since ARID1A plays a role in the MEK/ERK signaling mediated transcription network, we needed to know at what positions in the genome ARID1A was localized. For this, we had to perform chromatin immunoprecipitation for ARID1A in the HCT116 cell line. This was not trivial, as ARID1A is a member of such a large multi subunit complex which does not bind directly to DNA, and therefore very difficult to immunoprecipitate along with its associated DNA. In Figure 19a are some of the conditions under which we tried to chromatin immunoprecipitate ARID1A.These conditions differed mainly in the crosslinking agent, concentration and time used, the use of an additional protein-protein cross linker and the shearing method (agent, time and concentration of SDS in the buffers). While the immunoprecipitation of ARID1A seemed to be possible with all the antibodies we had (data not shown), we detected no DNA in the pull down. On adapting a protocol described by Zirkel et al., 2018, we were able to pull down ARID1A bound DNA. This protocol, uses the protein-protein crosslinker EGS to crosslink the subunits of the complex and other interacting proteins, before crosslinking DNA and protein using methanol-free paraformaldehyde. Moreover, all further processing steps are carried out using the Active Motif high sensitivity ChIP kit. This method yielded DNA in the pull down which we sequenced. As can be seen in Figure 19b, at an exemplary locus on the genome, the sequencing is of good quality. It colocalizes consistently with ChIP-seq tracks of other BAF complex subunits, SMARCC1 and SMARCA4 (public dataset). Moreover, there is a strong colocalization with active mark H3K27ac (public dataset). Also, what can be observed is that many of the ARID1A binding sites are distal from the transcription start site of the exemplary gene SMURF2 shown here. Finally, on a more global level, the overlap of the three BAF complex subunits shows high overlap but also independent binding (Figure 19c). This is highly dependent on the quality of the ChIP-seq datasets and the number of peaks obtained from it. However, this suggests that there are BAF complexes independent of ARID1A (for example those containing its mutually exclusive partner, ARID1B) and that ARID1A could also bind to the genome independently of these subunits (perhaps in the presence of the other core subunit SMARCA2). However, the large independent area for ARID1A is also due to the quality of the ChIP-seq dataset and the much higher number of peaks obtained in this experiment.

a)

	% of formaldhyde used for crosslinking	2mM DSG for dual crosslinking	Crosslinking time (formaldehyde)	Shearing method	% of SDS in ChIP buffers
Condition 1	0.3%	-	30 min	MNase treatment	--
Condition 2	1.0%	-	20 min	Sonication (30 cycles)	0.5%
Condition 3	1.0%	-	20 min	Sonication (30 cycles)	0.1%
Condition 4	0.2%	+	20 min	Sonication (30 cycles)	0.01%
Condition 5	0.4%	+	20 min	Sonication (30 cycles)	0.01%
Condition 6	0.4%	-	20 min	Sonication (30 cycles)	0.01%
Condition 7	1% glutaraldehyde	SUPERase.in to maintain RNA-protein interactions	10 min	Sonication (30 cycles)	0.1%
Condition 8	1% paraformaldehyde	15mM EGS 2mM EGS	40 min	Sonication (15 cycles)	-

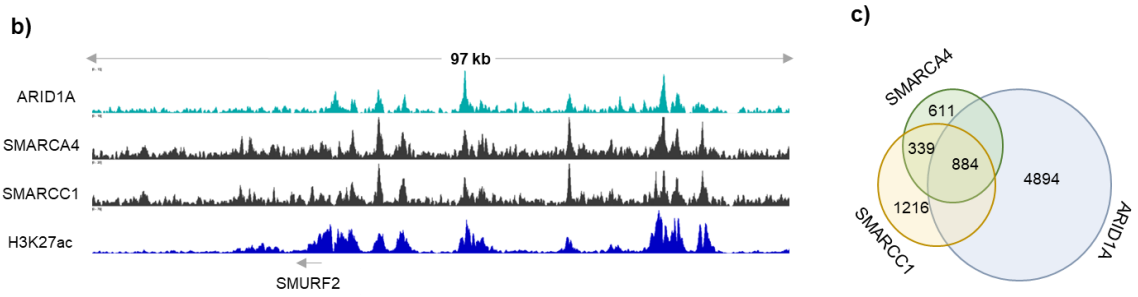


Figure 19: Optimization of ChIP-seq for ARID1A. The different conditions used for the optimization of chromatin immunoprecipitation for ARID1A. The main variations occurred in DNA-DNA crosslinking agent, time and temperature, the presence of a protein-DNA cross linker and the concentration of SDS in the buffers. DNA sequenced was for the ChIP experiment described in condition 8. As shown at this exemplary locus, the ARID1A peaks coincide very well with the other BAF complex subunits, SMARCA4 and SMARCC1 as well as with the active histone mark H3K27ac in the HCT116 system (b). On a more global level we see a large overlap of the ARID1A peaks with the other BAF complex subunits (c).

5.11 Changes in the epigenetic landscape upon the knockout of ARID1A

Once we had an ARID1A ChIP-seq dataset to work with, we were able to make conclusions about ARID1A-deficient systems based on the localization of ARID1A and not based on the occupancy of other BAF complex subunits. We were first interested in determining the effect the knockout of *ARID1A* has on the epigenetic landscape at the regions where it binds. Fortunately, several public datasets for histone modifications exist in the HCT116 WT and *ARID1A* KO systems which we could analyze. We analyzed the datasets for H3K27ac (an active histone mark present at enhancers and promoters), H3K27me3 (a repressive mark), H3K4me1 (a mark for primed enhancers), H3K4me3 (an active mark present at promoters) and ATAC-seq (assesses chromatin accessibility). We plotted the aggregate plots in Figure 20 on scales determined by plotting the signal of a particular experiment on its own binding

sites. We did so to obtain meaningful information when we next plotted the signal of each of these datasets on the ARID1A bound sites. This avoids detecting effects that are not very relevant. For example, the H3K27me3 at ARID1A bound sites if plotted on its own scale would yield a profile, with a scale up to 2. This information is not meaningful as the H3K27me3 signal goes up to 70 as seen on its own binding sites and therefore a signal of 2 is probably background. Interestingly, we found that the H3K27ac, H3K4me1 and H3K4me3 marks reduce in the *ARID1A* KO at a global level, irrespective of the presence of ARID1A (Figure 20 (a,d,e)), suggesting a regulation of enzymes that catalyze these modifications by ARID1A. Strikingly, while the H3K4me1 and H3K4me3 marks did not change substantially at the ARID1A bound sites, H3K27ac reduced dramatically at these sites (Figure 20a). As described in the introduction, a balance between the BAF complex and EZH2 containing PRC2 complex has been described in the literature (Wang et al., 2017, Kim et al., 2015). According to this, when H3K27ac reduces upon ARID1A loss, H3K27me3 at these sites should increase due to the PRC2 now occupying these regions. However, we observed, that while H3K27me3 does not change upon the knockout of *ARID1A*, at ARID1A bound sites, this mark is hardly present in the WT condition and does not increase upon the loss of ARID1A (Figure 20b). Finally, we expected the accessibility of chromatin to decrease upon the deletion of a chromatin remodeller subunit. Interestingly, the average signal of ATAC-seq is higher at ARID1A bound sites than the average signal of all ATAC-seq peaks indicating that the chromatin is more accessible in the presence of the BAF complex. However, the accessibility does not change upon the depletion of ARID1A, which was very surprising, both globally and at ARID1A bound sites (Figure 20c).

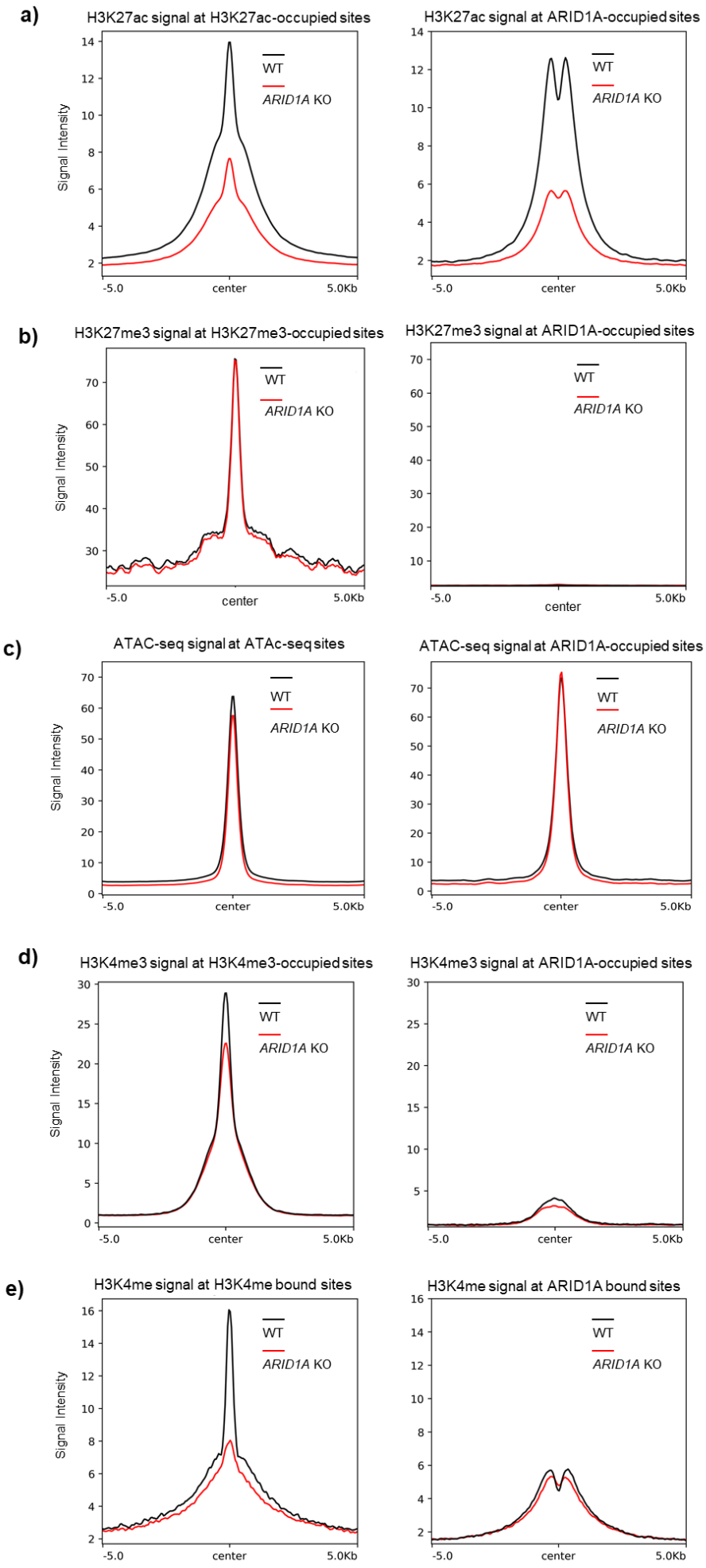

a)
H3K27ac signal at H3K27ac-occupied sites
H3K27ac signal at ARID1A-occupied sites
b)
H3K27me3 signal at H3K27me3-occupied sites
H3K27me3 signal at ARID1A-occupied sites
c)
ATAC-seq signal at ATAc-seq sites
ATAC-seq signal at ARID1A-occupied sites
d)
H3K4me3 signal at H3K4me3-occupied sites
H3K4me3 signal at ARID1A-occupied sites
e)
H3K4me signal at H3K4me bound sites
H3K4me signal at ARID1A bound sites
Signal Intensity
WT
ARID1A KO
-5.0
center
5.0Kb

Figure 20: Changes in epigenetic landscape upon *ARID1A* KO. We generated ChIP-seq data for ARID1A and on the basis of that could make conclusions about the effect its loss has on the epigenetic landscape based on its own binding. Each aggregate plot was plotted using the reference-point mode of the ComputeMatrix tool followed by the plotProfile tool. Each histone modification or ATAC-seq was plotted first on its own scale, (left panel) to determine a meaningful scale. The signal of the same histone modification was plotted on the ARID1A binding sites, using the same scale. In the case of H3K27ac, H3K4me3 and H3K4me, there was a global reduction after the depletion of ARID1A irrespective of the binding of ARID1A (a, d, e). The H3K27ac mark was also significantly reduced at the ARID1A bound sites (a) whereas in the case of H3K4me3 and H3K4me, there was only a slight reduction (d, e). The H3K27me3 mark was unaffected upon knockout of ARID1A and was hardly even present at ARID1A bound sites (b). The accessibility of chromatin assessed by ATAC-seq, surprisingly increased slightly upon *ARID1A* KO at ARID1A bound sites (c).

5.12 ARID1A binds at regions distal from the transcription start sites and colocalizes with AP1 transcription factors

It has been reported that BAF complex plays a role at mainly enhancers (Mathur et al., 2017, Vierbuchen et al., 2017, Alver et al., 2017) where it is involved in regulation of gene expression. Therefore, to start with we wanted to know what kind of regions in the genome ARID1A binds to. For this we used the Genomic Regions Enrichment of Annotation Tool (Cory et al., 2010) to calculate how far from the transcription start sites (TSS) each ARID1A binding site was. We found that a large majority of ARID1A binding sites are 5-500 kb from the TSSs (Figure 21a) which was consistent with literature indicating the localization of the BAF complex at distal regulatory elements even in other systems. Subsequently, we wanted to know what other transcription regulators ARID1A colocalizes with. The ReMap tool (Cheneby et al., 2018) runs based on a database of binding sites for 485 transcriptional regulators using Public (GEO or ArrayExpress) or ENCODE datasets. A consistent and integrated analysis of the data from different sources and across systems yields an atlas of regulatory regions for these transcription regulators in human cells. The annotation tool overlaps these regions with the user's regions of interest (in our case, regions bound by ARID1A) and reveals what other transcription regulators could bind at the same sites. This information is of course limited by the data that is available and thus not comprehensive. However, upon ReMap analysis, very interestingly, we observed that many of the transcription factors that colocalized with ARID1A were AP1 family members (Figure 21b). This indicated to us once again that ARID1A is involved in the transcriptional network that is induced by the MEK/ERK signaling pathway. Furthermore, on motif analysis using the HOMER software we found that the ARID1A-bound sites were very significantly enriched in AP1 DNA binding motifs (Figure 21c). In this analysis, the sequences represented by the regions bound are overlapped with putative DNA binding motifs for a database of sequence

specific DNA binding proteins. Finally, we used the publicly available ChIP-seq datasets for FosL1 and JunD (the two most highly expressed AP1 factors in the HCT116 system) and looked at their binding at ARID1A occupied regions. Each line in the heatmaps represents a region bound by ARID1A, ordered in descending order of ChIP-seq signal. The heatmaps are plotted at the center of the region and extend 5kb up and downstream of the center. As can be seen the intensity of signal for FosL1 and JunD at these regions indicates a strong colocalization of these factors with ARID1A (Figure 21d).

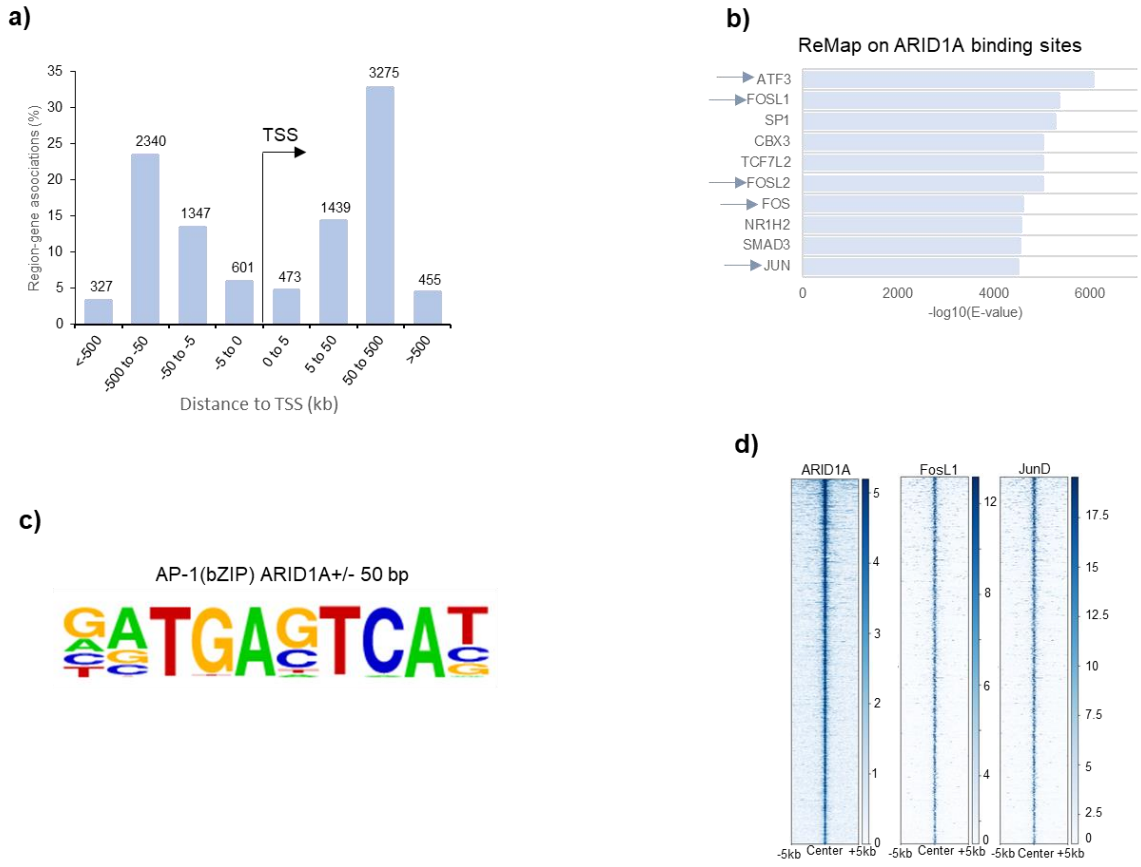


Figure 21: ARID1A binds at mainly distal regions and AP1 binding sites. Initial analysis of ARID1A bound peaks using GREAT v3.0.0 analysis (Cory et al., 2010) showed that most ARID1A bound sites are located between 5-500 kb from transcription start sites (a). ReMap v1.2 analysis (Cheneby et al., 2018) revealed that ARID1A binding coincides with the binding of several AP1 transcription factors indicated by the blue arrows (b). Moreover, motif analysis by the HOMER tool revealed the sequences that are bound by ARID1A are enriched in AP1 DNA binding motifs (c). A strong colocalization of the two most abundant AP1 factors, FosL1 and JunD, in the HCT116 system is also seen at ARID1A bound sites. The heatmaps were plotted using ComputeMatrix and plotHeatmap tools on the Galaxy server. All the heatmaps are plotted from highest to lowest signal at the center of ARID1A bound peaks +/- 5kb (d).

5.13 ARID1A-occupied distal regulatory elements are also bound by AP1 transcription factors

Since a majority of ARID1A-bound sites were distal to the TSSs, we sought to explore the distal regulatory functions of ARID1A at these sites. Therefore, in the next steps, we identified ARID1A-bound enhancers and determined what kind of genes these enhancers could regulate. Additionally, we also sought to know whether these sites are also dependent on the AP1 transcription factors. Having this knowledge would enable us to further validate that distal regulatory elements, that are coregulated by ARID1A and AP1 factors, are important for the modulation of genes that are expressed in *KRAS* mutation driven CRC cell lines. To identify regions that are bound by ARID1A that could qualify as active enhancers, we overlapped the H3K27ac occupied regions in the HCT116 cell line with the ATAC-seq enriched regions. While H3K27ac is a putative active enhancer (and TSS) mark, ATAC-seq determines the regions of the genome that are accessible to cleavage by a hyperactive Tn5 transposase. The accessibility of the chromatin could suggest transcriptional activity as well. We overlapped the ARID1A bound regions with regions that were accessible and marked by H3K27ac (most ARID1A-bound regions). Finally, we subtracted any regions from this overlap that represent TSSs. This analysis yielded 3061 regions which we defined as ARID1A bound 'enhancers' (Figure 22a). Conceivably, the regions that were identified were enriched in the Jun-AP1 binding motif (Figure 22b) and ReMap analysis revealed a colocalization of several AP1 transcription factors including FosL1 and JunD (Figure 22c). Finally, in the heatmaps, which were plotted at the center of these regions, the signal of the FosL1 and JunD was high, indicating their expected colocalization (Figure 22d). Thus, we identified ARID1A-bound enhancers in the HCT116 cell line which were seen to be also bound by FosL1 and JunD, suggesting a cooperative role between these factors.

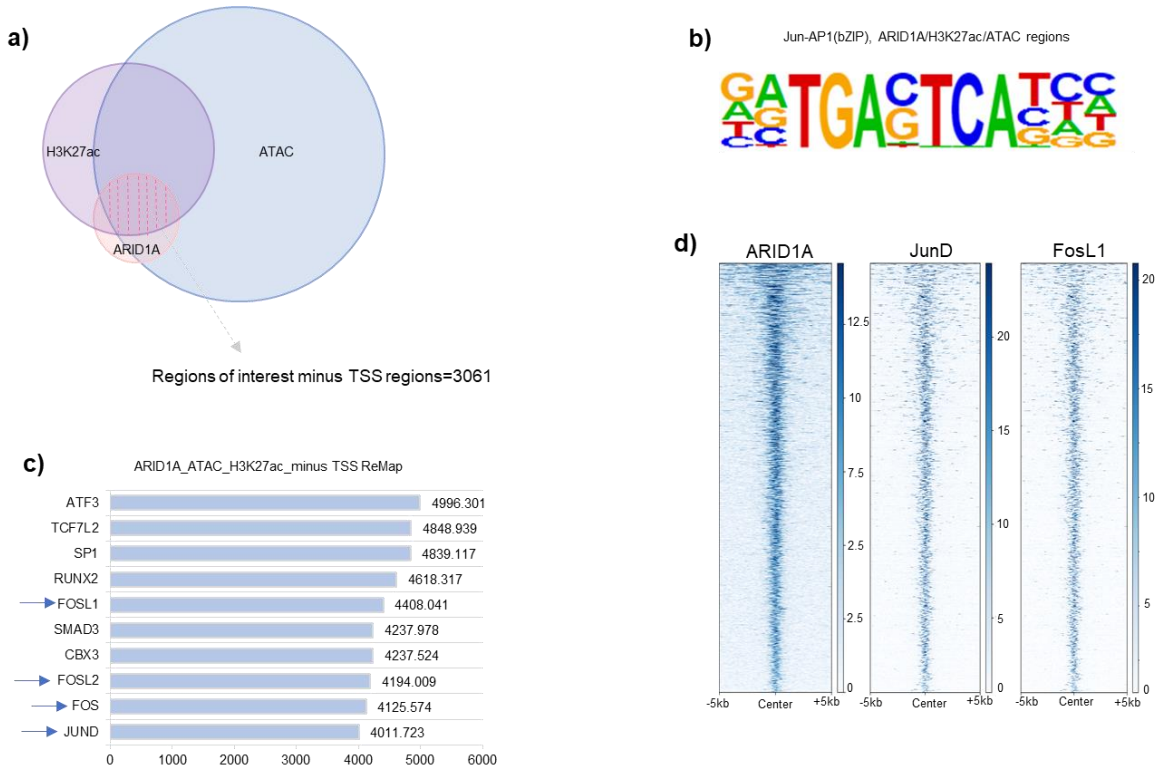


Figure 22: ARID1A and AP1 colocalize at ARID1A-bound enhancers. ARID1A bound distal regulatory elements (enhancers) were identified based on ARID1A binding, H3K27ac occupancy, and openness based on ATAC-seq data. Any Transcription Start Site (TSS) regions were subtracted (a). Motif analysis on these sites revealed an AP1 DNA binding motifs (b). ReMap analysis revealed colocalization with several AP1 transcription factors including FosL1 and JunD (c). Moreover, FosL1 and JunD colocalize strongly with ARID1A at these regions. The heatmaps were plotted using ComputeMatrix and plotHeatmap tools on the Galaxy server. All the heatmaps are plotted from highest to lowest signal of ARID1A at the center of the regions of interest +/- 5kb (d).

5.14 The genes associated with the identified ARID1A-occupied enhancers are regulated by its loss

Next, to find genes that are associated with these regions we used the GREAT tool. GREAT annotates function to regulatory domains based on genes that are proximal to the input regions. It assigns a regulatory domain to each gene based on the user's parameters and then overlaps these regions with the input genomic regions. It is however limited by assigning region-gene associations based only on linear distance on the genome. As is known, the three-dimensional structure of the genome plays a major role in enhancer-TSS interactions and enhancers can often be at a very large linear distance from their target genes, which is not accounted for in this tool. However, to get an idea of the kind of genes that are associated with the ARID1A bound enhancers that we had we identified, we used this tool. We found 3218 genes associated with the 3061 regions we identified. To focus only on the genes that could potentially be regulated by these enhancers, we looked at the

overlap of the associated genes with genes that are actually downregulated by the knockout of *ARID1A*. We identified 223 such genes (Figure 23a), several of which were known targets of the MEK/ERK pathway such as DUSP6 and several of which were cancer relevant genes. For the sake of simplicity, we chose to do further experiments on three genes that are potential MEK/ERK targets. Using these genes as examples, we wanted to validate that targets of the MEK/ERK pathway are often dependent on ARID1A acting as a co-factor at regulatory elements. While we show the results for only these three genes, several other genes in the 223 identified targets were tested and seen to follow similar trends.

The three selected genes were *EREG, F3* and *JAG1*. EREG (Epiregulin) is a ligand for EGFR (epidermal growth factor receptor) and hence interesting to study in the context of *KRAS* mutations. There has been extensive research on EREG expression in colorectal cancer patient samples and its properties as a prognostic marker in various therapies. One study suggests that low EREG expression is associated with better overall survival (Kuramochi et al., 2012). F3 (tissue factor III) encodes a glycoprotein receptor for coagulation factor VII, generally, initiating the blood coagulation cascades. However, it has also been implicated in cancer metastasis because deletion of a metastasis associated enhancer related to F3 has been found to block metastasis (Morrow et al., 2018). JAG1 (Jagged1) is a ligand for the notch receptor and is involved in cardiovascular development (Loomes et al., 1999). After selecting these genes to perform further experiments, we checked for the expression of these genes in *KRAS* WT and G13D mutated cell lines using the CCLE data (explained in section 5.2) visualized using the Morpheus software (https://software.broadinstitute.org/morpheus). Very interestingly, the expression of the three tested genes was higher in the *KRAS* G13D mutated cell lines as compared to the WT cell lines (Figure 23b), suggesting that the expression of these genes is dependent on the MEK/ERK signaling pathway and its downstream transcriptional network. To confirm this, we knocked down the AP1 factor JunD in the HCT116 system and measured the expression of these genes. While EREG and JAG1 were significantly downregulated, the expression of F3 was not affected (Figure 23c). This could be due to compensatory roles of other AP1 factors. Therefore, we used a MEK1/2 inhibitor, Trametinib (referred to as Tram in the figures), to attenuate the pathway and activation of the AP1 factors through this pathway. By doing this, we were able to observe a significant downregulation of all three genes (Figure 22c). Furthermore, as anticipated, upon knockout of *ARID1A* in both the *KRAS* G13D mutated cell lines HCT116 and DLD1, we observe that these genes are significantly downregulated (Figure 23d). When the HCT116 *ARID1A* KO cells are treated with Trametinib, there is further downregulation in the case of EREG and not much change in the case of F3 and JAG1. These results indicate that ARID1A plays a role in the

transcription of these MEK/ERK pathway target genes, perhaps *via* regulation through an interplay with distal regulatory elements associated with the AP1 factors identified in Figure 21a.

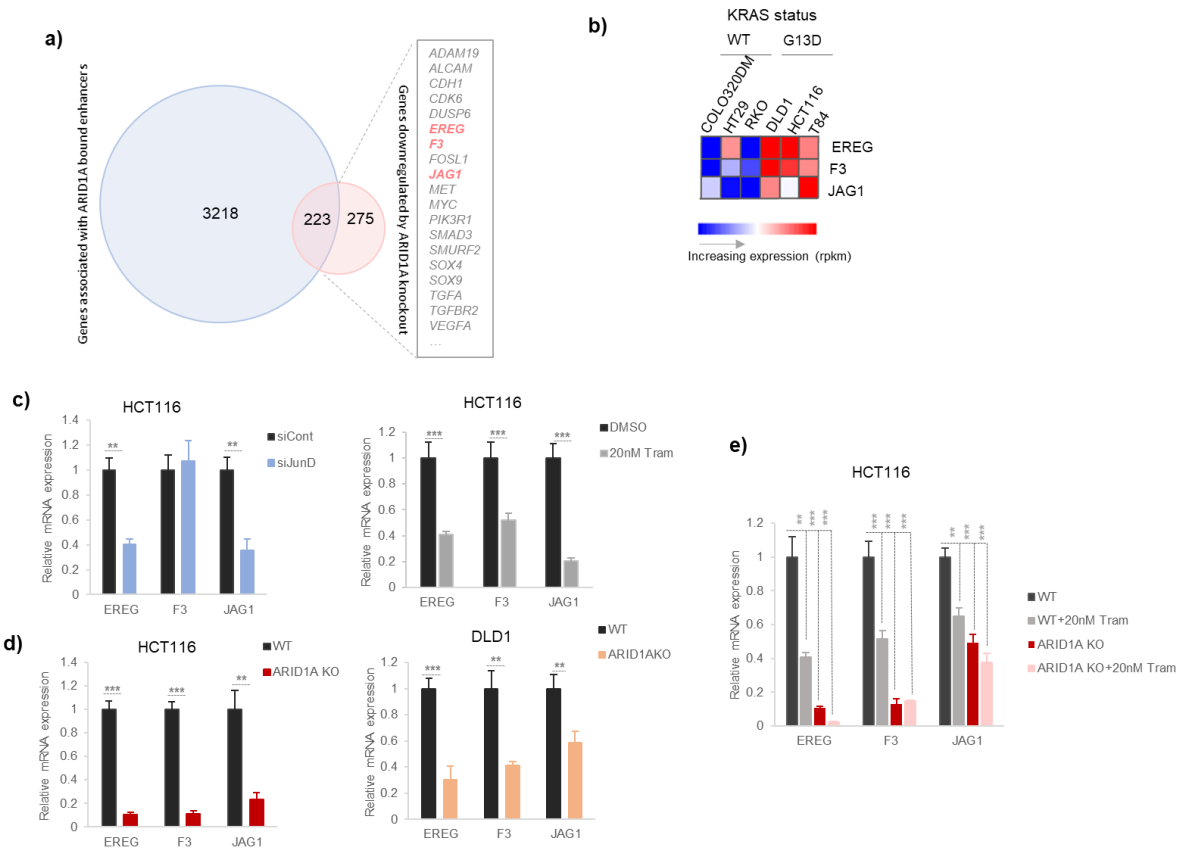


Figure 23: Integration of ChIP-seq and RNA-seq data. To identify genes that could be regulated by ARID1A bound enhancers, we identified genes associated with these regions (using GREAT analysis). We overlapped these genes with genes downregulated by *ARID1A* KO. 223 such genes were determined, many of which were colorectal cancer relevant (a). We focused on three genes to study further mechanisms namely EREG, F3 and JAG1 (a). On analyzing expression data available on Morpheus, we saw that these genes are expressed much higher in KRAS mutated colorectal cancer cell lines as compared to their WT counterparts suggesting their regulation by the MAPK pathway. These were chosen based on regulation upon ARID1A KO in both the HCT116 and DLD1 cell lines. Moreover, these genes were downregulated upon treatment with 20nM MEK inhibitor Trametinib showing a dependency on the AP1 transcription factors. The knockdown of JunD did not lead to the downregulation of F3. Error bars represent the standard deviation between three biological replicates. Significance calculated using unpaired t-test, * p<0.05, ** p<0.01, ***p<0.001.

5.15 The H3K27ac signal reduces very significantly at ARID1A-bound enhancers upon deletion of *ARID1A*

To explain the loss of expression of ARID1A/AP1 enhancer target genes, we performed a global analysis at all these sites to determine if the openness of chromatin at these sites was affected by the loss of ARID1A or if the active histone mark H3K27ac was changed. It

is conceivable that the knockout of a chromatin remodeller subunit would have effects on the accessibility of chromatin. However, very surprisingly, the accessibility of chromatin as assessed by ATAC-seq is not changed at these sites upon the loss of ARID1A (Figure 24a). In fact, on a more global analysis of the two ATAC-seq datasets (in the WT and *ARID1A* KO), we observe that the loss of ARID1A does not decrease the accessibility of chromatin genome-wide (except for a select few loci). In fact, accessibility is even gained at many loci, that are generally bound by ARID1A in the WT condition. Subsequently, we looked at the H3K27ac signal at these sites upon the *ARID1A* KO. As seen in figure 20a, there is a global genome-wide loss of this mark upon the loss of ARID1A and this is also true for the identified ARID1A/AP1 enhancers (Figure 24b). The loss of H3K27ac could indicate the inactivity of an otherwise active region, thus explaining the loss of gene expression of the target genes of the enhancers that lose this mark.

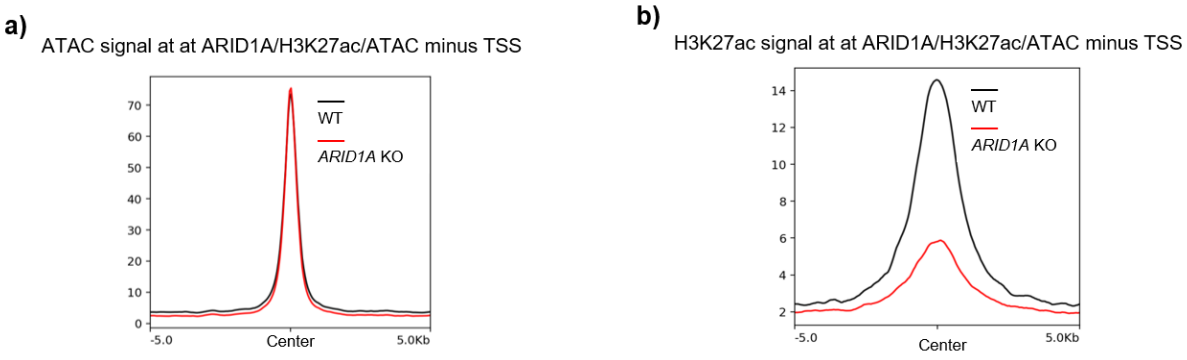


Figure 24: Loss of H3K27ac at ARID1A bound enhancers upon *ARID1A* KO. The signal of ATAC and H3K27ac in the WT and ARID1A KO conditions were plotted on the center of the ARID1A bound distal regulatory elements (+/- 5kb) in aggregate plots. The ATAC-seq signal remained unchanged upon *ARID1A* KO (a) and whereas there was a drastic reduction in H3K27ac levels in the *ARID1A* KO system at these regions (b). The aggregate profiles were plotted using ComputeMatrix and plotProfile tools on the Galaxy server.

5.16 A closer look the genomic regions surrounding *EREG, F3* and *JAG1*

To look more closely at the three target genes we identified, we examined the regions around their TSSs to find the potential ARID1A/AP1 dependent enhancers. For this we analyzed the ChIP-seq tracks for ARID1A, JunD, H3K27ac (WT and *ARID1A* KO), ATAC-seq and 3D genome interaction assessed by CTCF binding and calling of TAD boundaries, in the HCT116 system. The potential enhancers we identified are marked by gray arrows in Figure 25. These enhancers were intragenic in the case of *EREG* and *JAG1* (within intronic sequences of the adjacent gene *SLX4P*) and intergenic in the case of F3. All three potential enhancers were occupied by ARID1A and JunD. Moreover, H3K27ac decreased

significantly at these sites upon the knockout of ARID1A. The accessibility of chromatin at these sites measured by ATAC-seq did not change (or even increased) upon the knockout ARID1A. Furthermore, we called TAD boundaries that define regions that are within one Topologically Associated Domain (TAD) and thus have a higher chance of interacting (TAD boundary calling was performed by Xin Wang, University of Göttingen). All the enhancers we identified fall in the same TAD as the transcription start site (TSS) of their target genes thus allowing for interactions between these elements. It is interesting to note that not all TAD boundaries are occupied by the boundary factor CTCF and there are many sites on the genome where CTCF is bound irrespective of TAD boundaries. To conclusively prove that there is an actual physical interaction between these elements, however, further chromatin conformation capture experiments need to be performed.

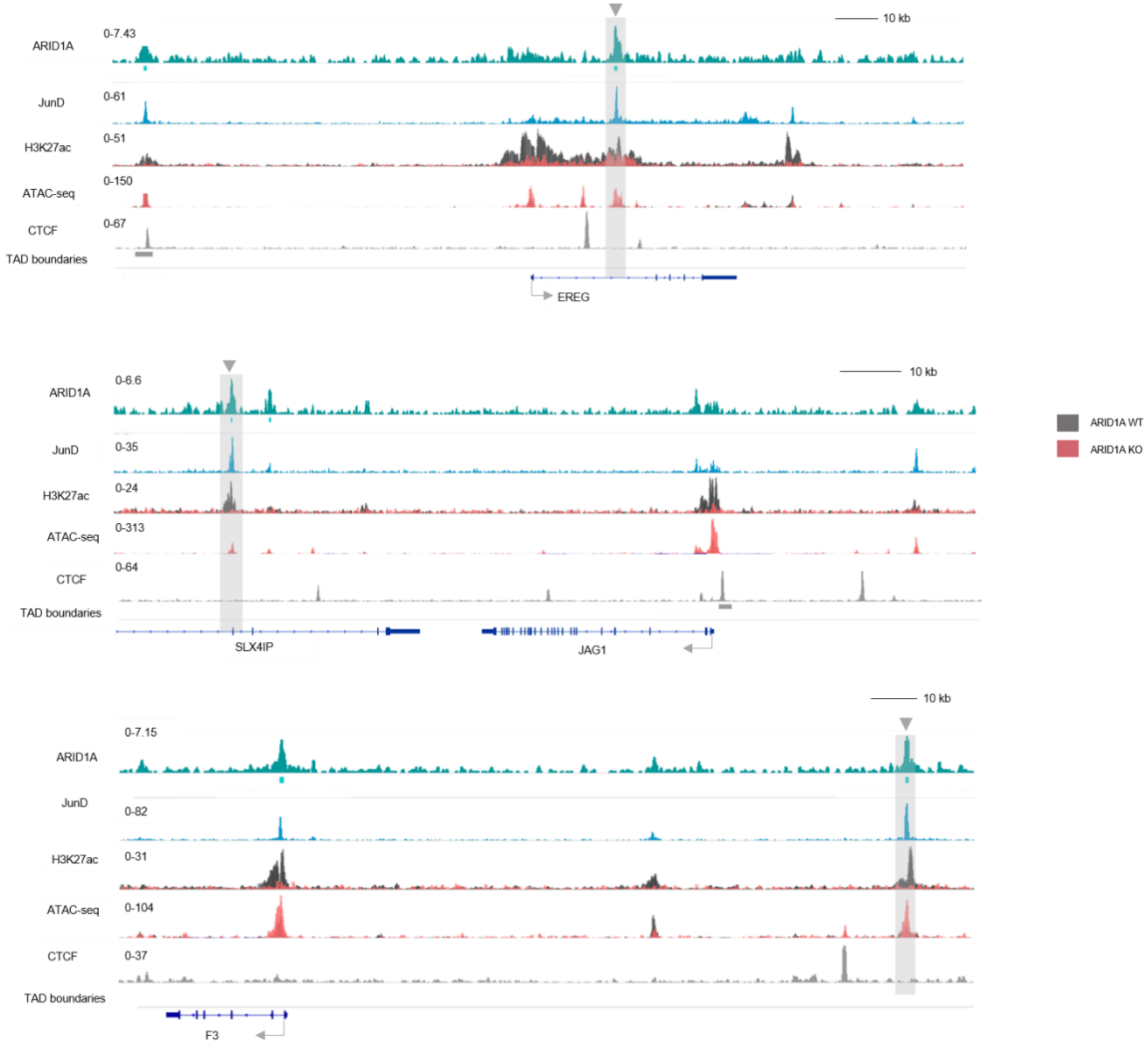


Figure 25: ChIP-seq tracks around the regions and genes of interest. ChIP-seq tracks at regions around the genes of interest viewed on the IGV genome browser. As can be seen, distal regulatory regions, either intra or inter genic, (marked by gray arrows) are occupied by ARID1A and JunD. H3K27ac at these sites is reduced upon *ARID1A* KO. ATAC-seq signal remained unchanged upon *ARID1A* KO. The regulatory regions identified were within TAD boundaries allowing for potential interaction with the TSS.

5.17 The binding of ARID1A and JunD at the ARID1A/AP1 enhancers are interdependent

To investigate the enhancers identified for *EREG, F3* and *JAG1* more closely, we obtained the sequence of the genomic DNA of these enhancer regions. We designed primers for these and performed ChIP-qPCRs under several conditions. All ChIP-qpCR signals were considered to be significantly over the background signal. Depending on the experiment, the background signal was either the average signal of the ChIP for IgG, or the average signal for qPCR at a negative site (for example a region that is methylated at H3K27). Firstly, we examined the occupancy of H3K27ac in the HCT116 WT, *ARID1A* KO and 20nM Trametinib treated conditions. Consistent with the previous results, H3K27ac at these enhancers were reduced significantly upon knockout of ARID1A (Figure 26a). Interestingly, at these sites, H3K27ac was also lost upon treatment with 20nM Trametinib (that is attenuation of the MEK/ERK pathway) suggesting that the attenuation of the downstream transcriptional network could block the acetylation of H3. Next, we determined the occupancy of JunD in the three same conditions. As expected, the occupancy of JunD was reduced significantly on treatment with 20nM Trametinib (Figure 26b). However, very strikingly, the occupancy of JunD was also reduced significantly from these enhancers upon the knockout of *ARID1A* suggesting that its presence is required for JunD (Figure 26b). This was not due to a downregulation in the expression of JunD upon *ARID1A* KO as can be seen in the western blot in Figure 26c. Finally, we also performed ChIP-qPCRs for ARID1A in HCT116 WT, 20nM Trametinib treated and siRNA depleted JunD conditions. Here we found that fascinatingly, the converse of the above described results is also true. The binding of ARID1A is reduced upon the depletion of JunD from these regions (either by Trametinib treatment or by siRNA mediated knockdown) (Figure 26d). A clear trend was observed for all three enhancer regions; however, significance could not be obtained (p values were between 0.08 and 0.1). This could be due to the large experimental variations that occur during a ChIP experiment and the low number of biological replicates (n=2). Having said that, there is a clear indication that ARID1A and JunD binding at these enhancers is dependent on the presence of both factors. Moreover, H3K27ac at these enhancers is also dependent on the binding of these factors.

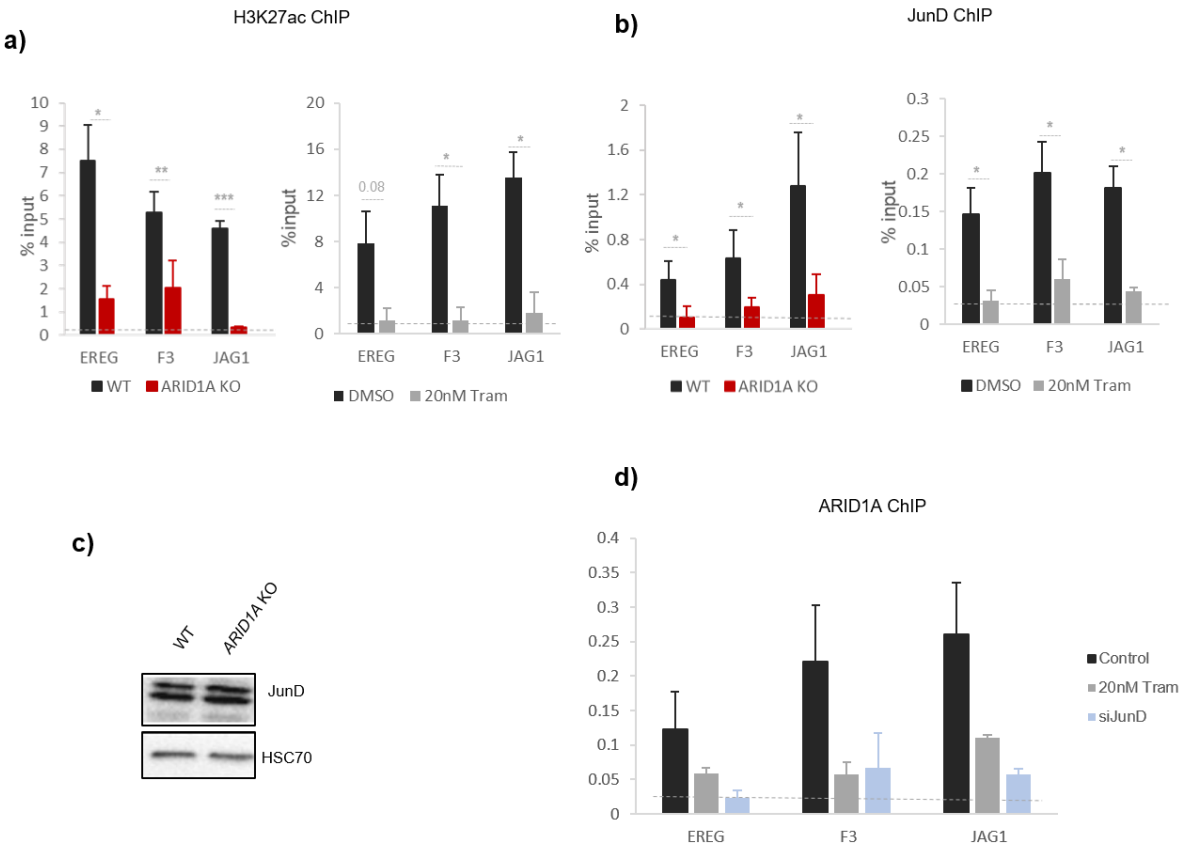


Figure 26: ARID1A and JunD binding at enhancers is dependent on the presence of both co-factors and the loss of either leads to a loss of H3K27ac from these enhancers. ChIP-qPCR was performed for H3K27ac and JunD in the ARID1A WT and KO systems at the enhancers identified in Figure 21a. Upon KO of ARID1A both H3K27ac (as expected) and JunD occupancy are reduced at these sites. The same ChIPs were performed in the WT and cells treated with 20nM Tram for 24h. Upon Trametinib treatment occupancy of JunD (as expected) and H3K27ac was reduced (a, b). This effect was not due to a change of expression of JunD in the *ARID1A* KO system. Moreover, the converse was also true, upon treatment with 20nM Trametinib or JunD knockdown, the occupancy of ARID1A was reduced (d). This effect was not due to a change of expression of JunD in the ARID1A KO system. The dotted lines represent the average background signal calculated based on IgG signal or signal from a negative site. Error bars represent the standard deviation between 2-3 biological replicates. Significance calculated using the unpaired t-test, * p<0.05, ** p<0.01, ***p<0.001.

6. Discussion

The importance of epigenetic regulators in cancer has been widely recognized, and much of the current research is focussed on deciphering the role of these regulators in driving oncogenic programs in cancer cells. As described in the previous sections, subunits of the BAF complex are among the most frequently mutated genes in cancer. While the primarily accepted role of the BAF complex is in chromatin remodelling, it is becoming increasingly clear that it is also involved in several other processes within the cell. Perhaps the most interesting is its role as a transcriptional co-factor, independent of its chromatin remodelling functions. It has been shown that ARID1A containing BAF complexes occupy enhancer regions and their loss from these regions causes a decrease in enhancer activity (Mathur et al., 2017, Vierbuchen et al., 2017, Lakshmikrishnan et al., 2017). In this study, we explored the role of ARID1A in enhancer mediated transcriptional control of gene expression in colorectal cancer cells. We found that cell lines that are *KRAS* mutated are especially dependent on ARID1A. In the absence of ARID1A, the proliferation of these cells is severely impaired. Furthermore, we confirmed that ARID1A itself is indeed mainly localized at enhancers in our colorectal cancer system. Moreover, it acts as a co-factor at regions also bound by the AP1 transcription factors which act downstream of the MEK/ERK pathway formatting a transcription network. We showed that the loss of ARID1A leads to a disruption of this transcriptional network at enhancers as both H3K27ac and the AP1 factor JunD are lost from these sites. This is accompanied by a downregulation expression of the associated target genes.

6.1 ARID1A acts as a co-factor with AP1 factors downstream of the MEK/ERK pathway

In this study, we explored the transcriptional role of ARID1A at distal regulatory regions (enhancers) that are controlled by the AP1 transcription factors downstream of the MEK/ERK pathway. We found that *KRAS* mutant colorectal cancer cell lines (specifically the G13D mutation) were particularly dependent on the presence of ARID1A. Our model for this mechanism is as follows. When there is constitutive MEK/ERK signaling due to mutant *KRAS*, the AP1 factors are phosphorylated and activated. These localize at some common target gene promoters but also at distal regulatory elements. It appears that many of these enhancers are co-occupied by ARID1A. Upon the loss of ARID1A, the binding of JunD (perhaps with other AP1 factors) is reduced from these enhancers. Moreover, the H3K27ac mark at these sites is also considerably diminished (Figure 26). And finally, the expression of the target genes that we identified was significantly downregulated. We hypothesize that the loss of the expression of the target genes is due to the loss of activity of their enhancers.

The loss of the active mark could perhaps be due to the impairment in recruitment of a histone acetyl transferase, upon the disassembly of the transcription network at these sites (Figure 25). The BAF complex has been shown to be important for the recruitment of p300 HAT in the MEF system (Alver et al., 2017) and so it is conceivable that this would occur in this system too. We were however, not able to conclusively prove this, due to technical difficulties in performing ChIP-qPCR for p300. We also have another interesting hypothesis to explain the loss of acetylation based on deacetylation by a histone deacetylase recruited by BRD4 which also colocalizes at the enhancers we identified. The tests for this hypothesis are preliminary and are presented in Supplemental Figure 4.

Thus, we believe that in *KRAS* mutant CRC, ARID1A loss attenuates oncogenicity, by disrupting a subset of the transcriptional networks downstream of this perturbation. While the loss of JunD is not very relevant in colorectal cancer, we used it as a model to disrupt the same transcription network. We did this by either performing siRNA mediated knockdown of JunD or treating with Trametinib which blocks its upstream activation. The loss of JunD from the chromatin also leads to a loss of ARID1A from these sites as well as a striking reduction in H3K27ac. This is the case both by siRNA mediated knockdown and treatment with Trametinib and could to some extent explain the sensitivity of the HCT116 cells to Trametinib (even though Trametinib has more widespread effects on proliferation related genes) and moreover the additive effect on proliferation upon ARID1A KO and Trametinib treatment. Thus, in this context, ARID1A is ***not*** tumor suppressive, rather it plays a supportive role for the AP1 transcription factors that act downstream of the commonly hijacked MEK/ERK pathway.

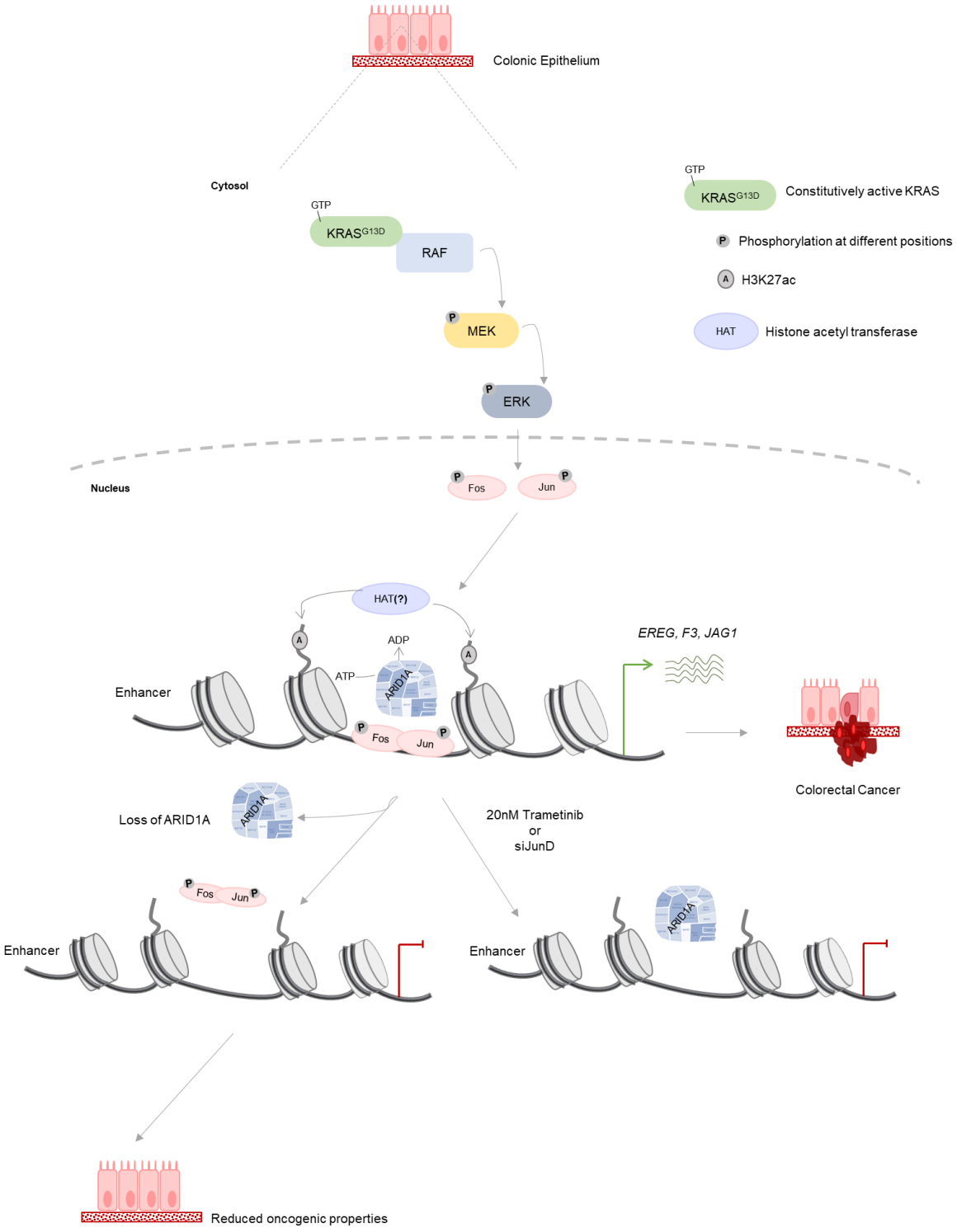

Colonic Epithelium
Cytosol
GTP
KRASG13D
RAF
MEK
ERK
Nucleus
Fos
Jun
HAT(?)
ADP
ATP
ARID1A
Enhancer
EREG, F3, JAG1
Colorectal Cancer
Loss of ARID1A
20nM Trametinib
or
siJunD
Reduced oncogenic properties
GTP
KRASG13D
Constitutively active KRAS
Phosphorylation at different positions
H3K27ac
HAT
Histone acetyl transferase

Figure 27: Model. In this study we present a model of *KRAS* mutated colorectal cancer cells that are dependent on ARID1A for oncogenicity. In the ARID1A WT condition, the *KRAS* G13D mutation leads to a constitutively active KRAS and hence constant activation of the phosphorylation cascade of the MEK/ERK pathway. This leads to an activation on the AP1 transcription factors (among which are the Fos and Jun family members). At several distal regulatory sites where these factors bind, ARID1A is also present. These sites are also marked by the active mark H3K27ac and their potential target genes are expressed. Upon the loss of ARID1A (which is the case in ~10% of colorectal cancers), the binding of JunD specifically, but perhaps also other AP1 factors, is lost. H3K27ac is also lost and the expression of the target genes is lost. In this case, the proliferation of the cells is severely impaired. Conversely, upon the depletion of JunD from these sites, ARID1A binding is lost and H3K27ac is also reduced. BAF complex structure adapted from Swetansu et al., 2016.

6.2 Role of Epigenetic Modulators in Colorectal Cancer

While it is not surprising that cancer cells would hijack epigenetic mechanisms that are crucial for gene expression regulation, a striking theme emerging from recent genome and exome-wide sequencing studies is the extent of perturbations in histone modifiers and chromatin remodellers in all types of cancer. These studies highlight the need to explore mechanisms of epigenetic deregulation in cancer because they may provide excellent opportunities to develop therapies based on molecular mechanisms that are often reversible. A well described epigenetic mechanism in colon cancer is perturbations in DNA methylation. However, beyond the methylation of DNA itself, ARID1A is among the most frequently mutated chromatin regulators in colorectal cancer. As shown in Figure 10, it is mutated at rates similar to *KMT2B* and *D*. The KMT2 family of genes encode histone methyltransferases that catalyze H3K4 methylation. These, along with the BAF complex subunits, are among the most frequently mutated genes across cancer types. We found that in the colorectal cancer datasets available, these mutations tend to co-occur significantly as shown in Figure 28. A significant proportion of CRC patients have mutations in both these epigenetic modulators (cBioPortal for Cancer Genomics).

It is also interesting to note that genetic disorders involving ARID1A (Coffin-Siris syndrome) (Tsurusaki et al., 2012) and KMT2D (Kabuki syndrome) (Ng et al., 2010), show similar phenotypes in terms of developmental delay and facial dysmorphisms (Izumi, 2016).While the consequences of *KMT2* and *ARID1A* mutations in patients with varied mutational backgrounds is not very clear, the co-inactivation of these regulators presents an interesting model of how disruption in the epigenetic landscape driven by these cooperative changes could drive cancer. The KMT2 factors have important roles in transcriptional control *via* enhancers. These enzymes (specifically KMT2 C and D) are part of complexes that place the H3K4me1 mark at enhancers to prime them. These are very important steps in lineage determination and differentiation (Lee et al., 2013, Rao et al., 2015, Review). As we have

shown in section 5.17, the loss of ARID1A from colorectal cancer cells leads to the loss of activity of certain enhancers. It is possible that in patients where the H3K4 methyl transferases are also inactivated, the activation of some crucial enhancers is hindered and normal gene regulatory networks are disrupted. Indeed, it has been shown that variations in H3K4me1 defined enhancers in primary colorectal cancer tissues are predictive of the colorectal cancer transcriptome (Akhtar-Zaidi et al., 2012). Therefore, it is conceivable that the two most perturbed epigenetic regulators in cancer would act in concert to reprogram enhancers that drive tumorigenesis. The dysregulation of enhancers by the perturbation of these factors needs to be explored further to develop mechanism-based therapies.

Gene A	Gene B	Neither	A Not B	B Not A	Both	Log Odds Ratio	p-Value	Adjusted p-Value ▲	Tendency
ARID1A	KMT2D	3028	163	181	101	2.339	<0.001	**<0.001**	Co-occurrence Significant
KMT2C	KMT2D	3047	144	187	95	2.375	<0.001	**<0.001**	Co-occurrence Significant
KMT2B	KMT2D	3122	69	227	55	2.395	<0.001	**<0.001**	Co-occurrence Significant
ARID1A	KMT2C	3030	204	179	60	1.605	<0.001	**<0.001**	Co-occurrence Significant
ARID1A	KMT2B	3123	226	86	38	1.809	<0.001	**<0.001**	Co-occurrence Significant
KMT2B	KMT2C	3141	93	208	31	1.616	<0.001	**<0.001**	Co-occurrence Significant

Figure 28: Mutations in two most perturbed epigenetic modulators in cancer co-occur. Analysis of all the colorectal cancer datasets on the cBioPortal for Cancer Genomics database revealed that ARID1A is among the most frequently altered chromatin regulator in colorectal cancer. It is mutated at rates similar to the KMT2 family of histone methyl transferases. The mutations in these factors co-occur significantly. It is interesting to note that though these factors are unrelated, they could both act in defining and activating enhancers, and therefore their inactivation could act in concert to disrupt enhancer mediated gene regulatory networks.

6.3 ARID1A at Enhancers

Most conclusions about the enhancer functions of ARID1A are based on the occupancy of other BAF complex subunits such as SMARCC1 and SMARCA4 (Mathur et al., 2017, Kelso et al., 2017). While this is much more information than was available earlier (due to the difficulty of performing chromatin immunoprecipitation for the BAF complex), it still does not account for all ARID1A-containing BAF complexes. We were able to determine the genome-wide occupancy of ARID1A in the HCT116 cells. We obtained 5778 peaks in the ChIP-seq data and it was of relatively good quality. This was also one of the first ChIP-seq datasets to have been generated for ARID1A (Raab et al., 2015 performed ChIP-seq for ARID1A in HepG2 cells, however the binding sites for ARID1A differed quite significantly from our dataset). The experiment for determining the occupancy of ARID1A was quite challenging because it is difficult to crosslink the BAF complex stably on the chromatin (perhaps because of its large multi-subunit nature). Moreover, since it does not have a sequence

specific binding motif, it is probably recruited to the chromatin by various factors making its interaction with the chromatin unstable. While we tried many different methods to try and optimize the ChIP for ARID1A, we were finally able to obtain good results using an adaptation of the protocol described by Zirkel et al., 2018. This protocol employs a protein-protein crosslinker first, followed by protein-DNA crosslinking by paraformaldehyde. The protein-protein crosslinking perhaps ensures that the complex remains intact and bound to any other recruiting factors.

On obtaining genome-wide occupancy data for ARID1A, we were able to show that ARID1A acts as a co-factor at enhancers that are occupied by AP1 transcription factors. These transcription factors are most often activated by the MEK/ERK pathway and act to transactivate downstream targets. While the expression of the target genes we defined was affected by the loss of ARID1A, we showed that this was probably not due to its chromatin remodelling activity. This is because chromatin accessibility, as assessed by ATAC-seq, does not change at ARID1A-bound sites upon its loss. What is disrupted though is the transcriptional machinery present at the enhancers. Upon the loss of ARID1A, JunD occupancy and H3K27ac are reduced. The converse is also true, where a depletion of *JUND* (or active JunD) *via* siRNA-mediated knockdown or Trametinib treatment results in a reduction of ARID1A occupancy and H3K27ac. The expression of genes that are associated with these enhancers is also downregulated by the loss of ARID1A and active JunD. It is important to keep in mind that associated genes are defined by linear proximity on the genome. While the occupancy of factors at enhancers and changes in gene expression upon the perturbations of these factors are good indicators of these regions having regulatory functions, the 3D structure of the genome needs to be considered in order to pin-point enhancer-promoter interactions. Using HiC data (Rao et al, 2018) for the HCT116 cell line, we were able to show that the enhancer-promoter pairs that we identified (*EREG, F3, JAG1*) lay within the same Topologically Associated Domains (TADs) making their interactions more probable. However, further experiments are required to show the direct interaction between the enhancer and promoters of these genes.

The broader question of how enhancer activity is regulated and how this affects target genes remains. The loss in enhancer activity can occur by a disruption of the transcriptional machinery present at enhancers. This machinery is responsible for recruiting further histone modifying enzymes that change the enhancer landscape to a more active state, which can then recruit further factors that fine-tune gene regulation. The interaction between the enhancer and promoter itself occurs *via* chromatin looping which is probably enhanced by physical proximity in three dimensions. The interaction of the enhancer and promoters increases the local concentrations of transcriptional regulators (such as transcription factors

and chromatin modifying enzymes) which enhances the transcription driven by the basal machinery present at the gene promoter. Roles for multi-subunit complexes such as the mediator and enhancer RNA transcribed by RNA Pol II in mediating enhancer-promoter interactions have been proposed (Heinz et al., 2015, Review). It is therefore also conceivable that BAF complexes occupying enhancers (with their multiple subunits with multiple domains) could mediate enhancer-promoter interaction *via* transcription factors present at these two sites. As described, the BAF complex through its multiple subunits interacts with various important transcriptional regulators such as TP53 (Guan et al., 2011), MYC (Rahman et al., 2011), CTNNB1 (Barker et al., 2001), BRD4 (Rahman et al., 2011) to name just a few. Therefore, the BAF complex could be playing a role as a transcriptional co-factor in addition to its chromatin remodelling functions.

We and others have shown that enhancer activity is attenuated upon the loss of ARID1A. This is assessed by the significant reduction of the active enhancer mark H3K27ac at regions bound by the BAF complex. The BAF complex has no known acetyltransferase activity; however, it has been shown to physically interact with the histone acetyltransferase p300 (Ogiwara et al., 2011). Alver et al., 2017 showed that re-expression of *Smarcb1* in MEFs increased the p300 levels, along with other enhancer components BRD4 and subunits of the mediator complex in the chromatin fraction This suggested that the BAF complex recruits p300 to the chromatin. Similarly, through interactions with several other proteins, the BAF complex could recruit histone acetyl transferases.

In another attempt to explain the loss of H3K27ac at ARID1A/AP1 bound enhancers, we came up with an interesting hypothesis. Since the acetylation of histones is also regulated by histone deacetylases (HDACs), we looked at regulators that localize at ARID1A-bound sites. Apart from the AP1 factors, BRD4 was found to localize significantly at these sites. BRD4 is known to interact both with the BAF complex and with the repressive NuRD complex (Rahman et al., 2011), through its subunit CHD4. Therefore, we hypothesized that in the absence of the BAF complex, BRD4 could recruit the NuRD complex to enhancers. Histone Deacetylase 1/2 (HDAC1/2) are members of the NuRD complex that could potentially deacetylate H3K27 and thus reduce enhancer activity. We tested this hypothesis in some preliminary experiments by depleting the BET proteins, BRD2,3 and 4, in *ARID1A* KO cells to prevent the deacetylation of chromatin. We did not, however, observe a rescue in gene expression of the target genes. Perhaps this competitive interaction does not exist, or the rescue of acetylation is not enough to activate the enhancer in the absence of ARID1A. Therefore, further experiments need to be performed to test this hypothesis.

What is not very clear is why the loss of ARID1A results in a global loss of H3K27ac. While the explanation given above is plausible at the sites where the BAF complex is bound, it is more difficult to explain this effect at sites where it is not bound. On checking the mRNA levels of the HATs p300 and CBP in the *ARID1A* WT and KO systems, we observed no differences. Even though we were not able to get completely convincing results, it seems that the protein levels of these factors are also not affected by ARID1A loss (data not shown). Therefore, there might be other effects that explain this. For example, the activity of the HATs might be affected or other lesser known HATs for H3K27ac might be regulated by the loss of ARID1A. However, it is clear that the BAF complex mediates enhancer activity *via* interactions with many transcriptional regulators.

6.4 The role of the BAF complex in Wnt-signaling mediated transcriptional regulation

One of the most interesting observations made by Mathur et al., 2017 was that the simultaneous inactivation of two tumor suppressors, *Apc* and *Arid1a*, led to the formation of fewer tumors than the inactivation of either protein alone. Moreover, the few tumors that were formed retained Arid1a expression, suggesting that it is required for tumorigenesis driven by *Apc* inactivation in mice. As described in a review by the same author (Mathur and Roberts, 2018) and mentioned in section 2.8, ARID1A drives invasive colorectal adenocarcinomas in the absence of mutations in colon cancer-relevant genes. Interestingly, Holik et al., 2014 described a genetic mouse model in which deletion of *Smarca4* from the intestinal epithelium attenuated Wnt signalling-mediated target gene expression. The link between the BAF complex and Wnt-signaling has been described before in other contexts. SMARCA4 has been shown to interact with β-catenin and drive the transcription of Wnt target genes in human cell lines (Barker et al., 2001). This interaction has also been shown to play important roles in liver regeneration (Li et al., 2018), blood vessel development (Curtis et al., 2012) and cardiac development (Bevilacqua et al., 2014, Review) where the Wnt pathway is an important player (Tian et al., 2010, Review). Interestingly in vascular endothelial cells, Curtis et al., 2012 proposed an antagonistic relationship between SMARCA4 and CHD4 at Wnt target genes, similar to the hypothesis we presented in the previous section. Also, our RNA-seq data in the COLO320DM cell line revealed the deregulation of many genes involved in cardiac development and the Wnt pathway upon the KO of *ARID1A*. Conversely, there has also been a report of ARID1B containing BAF complexes repressing Wnt-target promoters (Vasileiou et al., 2015). Therefore, it is evident that the BAF complex plays a role in transcriptional regulation mediated by Wnt-signaling. To explore this further in the colorectal cancer setting, we looked for factors in our analysis that could be involved in this pathway. Strikingly, we found that TCF7L2, the downstream

effector of the Wnt pathway, colocalizes strongly with ARID1A at enhancers. Moreover, most of this colocalization is at enhancer regions. To decipher this network further, we tried to perform chromatin immunoprecipitation for TCF7L2 in the *ARID1A* WT and KO conditions. While we obtained high signal for TCF7L2 at typical Wnt-target promoters such as *AXIN2*, we detected no signal at the enhancers also bound by ARID1A, even though the publicly available ChIP-seq data showed very strong binding of TCF7L2 at these regions. This could be because of the more stable binding of TCF7L2 directly to DNA at the Wnt-target promoters, whereas binding at enhancers is perhaps through interactions with other factors and hence more transient. We suspect this to be true because the binding motifs at the TCF7L2 bound enhancers (co-bound by ARID1A) were still for AP1 factors and not for TCF7L2, suggesting that TCF7L2 does not directly bind to DNA at these sites. Furthermore, when we knocked down TCF7L2 in the HCT116 cell line, we did not observe a downregulation of genes that are downregulated by ARID1A. This suggested that, either the role of TCF7L2 is compensated by other TCF factors or that it is not an essential player the enhancers we tested. However, the BAF complex does seem to play a role downstream of the Wnt signaling pathway in colorectal cancer.

6.5 ARID1A: tumor suppressor or oncogene in colorectal cancer?

One of the major topics that we addressed during the course of this project was the suggested tumor suppressive role of ARID1A in colorectal cancer. ARID1A mutations most often lead to a loss of the protein from the cells. Moreover, unlike in any other cancer, ARID1A has been shown to play a driver role in colorectal cancer. Sporadic inactivation of ARID1A from the entire organism, in an otherwise wildtype background, leads to the formation of invasive adenocarcinomas originating in the colon of mice (Mathur et al., 2017). This points towards the utmost importance of this protein for preventing tumorigenesis. This is consistent with its expression being lost in patient tumors and with prior research that showed tumor suppressive roles for ARID1A in other cancer types (Guan et al., 2011, Chandler et al., 2015, Sun et al., 2017, Livshits et al., 2018). If the loss of ARID1A (at ***any*** step in the transition from the normal epithelium to invasive adenocarcinoma) was tumor suppressive, it could be that the patients who lacked ARID1A expression might show poorer survival rates. However, surprisingly, we found that in colorectal cancer, there is no correlation between the expression of ARID1A and the survival of patients. The mutations in ARID1A are most often missense or frameshift mutations, which do not necessarily lead to a loss in mRNA expression (as we also saw in our CRISPR KO cells). The production of a truncated product leads to nonsense-mediated decay, or degradation of the non-functional protein and therefore a complete loss of the protein from the cells. Thus, while mRNA expression is perhaps not the most appropriate way to stratify patients, protein levels

would be a good indicator to assess the prognostic value of ARID1A in predicting survival. However, since, we did not find a correlation between *ARID1A* expression and survival, we questioned whether the path of tumorigenesis mediated by ARID1A loss was context dependent in some respects. First, it would be meaningful to decipher the timepoint (during the transformation) at which ARID1A loss acts as a driver of colorectal cancer. Second, it would also be important to know how the existing mutational background contributes to the loss of ARID1A. That is, in the context of already existing perturbations, perhaps the loss of ARID1A is redundant or might even protect against tumorigenesis. Another interesting point to address would be the consequences of heterozygous mutations and compensatory effects of the mutually exclusive subunit ARID1B. While these are questions that are still not answered, this study contributes to the evidence mounting from several studies that point to context dependent functions of ARID1A in driving tumorigenesis.

The mouse models introduced in section 2.8 clearly illustrated these context dependent functions. While in pancreatic cancer, late loss of Arid1a is described as a passenger event, which does not affect tumorigenesis (Livshits et al., 2018), the late loss of Arid1a in liver cancer was shown to promote metastasis (Sun et al., 2017). In colorectal cancer, Arid1a loss is a driver event. However, in the context of existing *Apc* mutations, *Arid1a* loss prevents *Apc* mutation-driven colorectal cancer (Mathur et al., 2017). In contrast, in ovarian and pancreatic cancer, co-occurring mutations of *Pten* and *Kras* respectively are required for Arid1a to function as a tumor suppressor (Guan et al., 2011, Livshits et al., 2018). Moreover, in this study, we showed that the loss of ARID1A severely impaired the proliferation of *KRAS* mutant colorectal cancer cell lines. This, along with the data from Mathur et al., suggests that in the context of commonly occurring mutations in *APC* and *KRAS*, ARID1A is actually ***required*** to maintain tumorigenesis.

In analysis of human colorectal cancer patient samples, it has also been difficult to correlate ARID1A loss with prognosis or stage specificity. Lee et al., in 2016, observed that in in early colon adenocarcinomas associated with microsatellite instability, ARID1A is lost in about 10% of cases. However, this study which analysed 552 early stage human colorectal cancer samples also showed that ARID1A mutation is associated with factors that predict poor prognosis. Moreover, the frequency of *ARID1A* mutation increased in higher stage tumors suggesting that ARID1A loss occurs as tumor formation progresses. However, analysis of the consequence of the loss at this late stage was not performed. As mentioned previously, late stage loss of ARID1A could either protect against tumorigenesis or be completely inconsequential. In analyses of patient material for pancreatic cancer, renal clear cell carcinoma, endometroid and breast cancer, ARID1A loss was associated with aggressive,

undifferentiated and late stage tumors (Park et al., 2015, Mamo et al., 2012, Zhang et al., 2018).

Therefore, it is evident that ARID1A is ***not*** simply the tumor suppressor that it was initially described to be. Its role varies depending on the context of the organ in which it occurs as well as the time at which it occurs within the same organ. Specifically, in colorectal cancer, the loss of ARID1A alone showed an unprecedented driver role (Mathur et al., 2017). However, as we showed, in the context of other perturbations such as *KRAS* mutations, its loss is disadvantageous for cancer cells.

6.6 Phenotypic changes

Contrary to the initial literature describing the roles of ARID1A as a tumor suppressor, we observed a severe impairment of proliferation upon the depletion of ARID1A in colorectal cancer cell lines. Unlike ovarian cancer models, in which ARID1A loss alone is not sufficient to drive tumorigenesis (Guan et al., 2011, Chandler et al., 2015), colorectal cancer models show that the loss of ARID1A alone is enough to drive cancer (Mathur et al., 2017). However, in contrast with rhaboid tumors (where *SMARCB1* inactivation alone causes tumor formation and is the only perturbation to occur in that system), defects in the BAF complex are not the only perturbations that occur in CRC. Many other (more frequently occurring) mutations are observed in CRC which often play driver roles. Therefore, as mentioned in section 2.8, the loss of ARID1A can elicit different phenotypic changes depending on the context. For example, in our experiments, we observed that in colorectal cancer cells with different mutational backgrounds, the effect of ARID1A deletion is varied. When comparing phenotypic changes depending on the *KRAS* status of the cell lines we tested, we observed that the *KRAS* mutant cell lines show a severe defect in proliferation and changes in morphology. On the other hand, the cell lines which were wildtype for *KRAS* showed no phenotypic changes. It is important to keep in mind that we compared ARID1A deficiency in cell lines stratified by the occurrence of one other mutation (*KRAS*). It is likely that the rest of the mutational background of each cell line elicits different responses to the loss of ARID1A. However, the *KRAS* background was an important parameter to consider, as mechanistically, we were able to show that ARID1A is required for transcriptional regulation downstream of the KRAS pathway.

We were initially surprised with the impairment in proliferation we observed, as until then, ARID1A had been described in the literature as a bona fide tumor suppressor. Therefore, to rule out artefacts of the *in vitro* system, we explored the loss of ARID1A in two *in vivo* contexts. However, our results were supported as we also did not see a tumor suppressive role in the *in vivo* systems. In our xenograft model, we injected *KRAS* wildtype HT29 cells

with or without KO of *ARID1A* into SHO mice (performed with Dr. Florian Wegwitz and Dr. Robyn Kosinsky, University of Göttingen). After 3 weeks, tumors were formed in both conditions without any significant differences. Also, in our conditional knockout model for Arid1a loss, we did not see any tumor formation after 6 months (performed with Dr. Robyn Kosinsky, University of Göttingen). However, we can draw limited conclusions from this experiment as we did not ascertain whether the knockout of *Arid1a* had been efficient. The conditional knockout, based on the expression of the Cre-ER^{T2} downstream of the colon-specific *Cdx2* promoter, was generated to produce colon specific genetic models. Other intestinal models have resulted in the formation of small intestinal tumors which is unlike what occurs in the human condition where more colorectal tumors are formed (Hinoi et al., 2007). Therefore, we wanted to model CRC based on *Arid1a* loss using this system. However, past experiments in our group have shown the inefficiency of the model where despite expression of *Cdx2*, there was no knockout of the targeted gene (data from Dr. Robyn Kosinsky and Lorenz Chua). Therefore, while it would be very interesting to model CRC based on *Arid1a* loss in a genetically engineered mouse model, perhaps another promoter such as the CAC-Cre (carbonic anhydrase-Cre) should be used. Perhaps even oncogenic drivers such as *KRAS* should be used to model colorectal cancer. Haigis et al., 2008 described one such model. Moreover, it would be extremely informative to know how the knockout of *Arid1a* at different time points in the context of various background mutations affects the progression and metastasis of colorectal cancer.

Having considered these observations, it seems that *KRAS* mutant colorectal cancer cell lines are particularly dependent on ARID1A and its loss impairs proliferation. While we were able to mechanistically show that the loss of ARID1A affects transcriptional regulation of MEK/ERK target genes *via* enhancer mediated regulation, we could not identify a master regulator for proliferation that is perturbed by the disrupted enhancers. The three genes that we studied (*EREG, F3 and JAG1*) all have an interesting role in cancer as explained in section 5.14; however, their loss cannot explain the proliferation defect. EREG (Epiregulin) is a ligand for EGFR (epidermal growth factor receptor). There has been extensive research on EREG expression in colorectal cancer patient samples and its properties as a prognostic marker in various therapies. Kuramochi et al., 2012, suggested that low EREG expression is associated with better overall survival. F3 (tissue factor III) encodes a glycoprotein receptor for coagulation factor VII, generally, initiating the blood coagulation cascades. However, it has also been implicated in cancer metastasis because deletion of a metastasis associated enhancer at the F3 locus has been found to impair metastatic outgrowth (Morrow et al., 2018). JAG1 (Jagged1) is a ligand for the notch receptor and is involved in cardiovascular development (Loomes et al., 1999) which, as described subsequently, has

an interesting link with the BAF complex. The effect on proliferation is probably a combinatorial effect of the pathways deregulated by *ARID1A* loss. However, ARID1A also has other cellular functions that could play a role in eliciting this response to its loss. For example, ARID1A has been shown to play roles in E2F and p53 mediated transcription, DNA replication and repair and in telomere maintenance, all of which could affect proliferation (Roberts and Orkin, 2004, Review). While these roles have been described in various other systems, they could also apply to the colorectal cancer cell line system. In fact, in preliminary experiments, we have observed that the deletion of *ARID1A* from the HCT116 system causes a decrease in replication fork progression (performed by Josephine Choo, University of Göttingen).

The morphological changes observed in the HCT116 and DLD1 system can be explained by the gene signatures that were seen to be affected upon the knockout of *ARID1A*. In the case of HCT116 several epithelial cell programs are affected. Mathur et al., 2017 reported a loss of the epithelial marker E-cadherin in HCT116 *ARID1A* KO cells, which we reproduced (data not shown). This might partially explain the more mesenchymal-like morphology we observe for the *ARID1A* KO cells which display filopodia-like projections. Similarly, the knockout of ARID1A in the DLD1 cell line most significantly affected the pathways that were involved with the cytoskeletal machinery, which explains how a deregulation would cause a change in morphology. Many of these pathways were actin filament related. The actin cytoskeleton not only plays a role in cell morphology and motility but also plays major roles in signaling scaffolds and during the cytokinesis step of mitosis (Carpenter, 2000, Review). An integrated dysregulation of any of these processes could play a part in the phenotype observed in the DLD1 cells upon the knockout of *ARID1A*.

Overall, it seems that the loss of ARID1A is particularly detrimental to *KRAS* mutated colorectal cancer cells. While, the transcriptional dependency of the MEK/ERK on ARID1A is amply clear, the proliferation may be affected by a combination of some of the consequences of ARID1A loss.

6.7 Relation with the MEK/ERK Pathway

The impairment of proliferation upon the KO of *ARID1A* that we observed was common to *KRAS* mutant cells (HCT116 and DLD1). While this indicated that *ARID1A* is perhaps required for the tumorigenic properties mediated by *KRAS* mutation, downstream analysis confirmed our hypothesis. This would suggest that in human cancer, *KRAS* mutations and *ARID1A* mutations employ different pathways of tumor initiation and progression, as it would be disadvantageous for cancer cells that are driven by the activation of *KRAS* to lose ARID1A expression. Therefore, at least in early cancer samples, these mutations should

occur independently of one another. To explore this, we analyzed the colorectal adenocarcinoma datasets available on the cBioPortal for Cancer Genomics database which had more than 300 samples. This analysis yielded some interesting results that are presented in Figure 29. All the colorectal adenocarcinoma datasets showed that *KRAS* and *ARID1A* mutations are mutually exclusive or show a tendency towards mutual exclusivity. In a large dataset of metastatic colorectal cancer, we observed that there is a minor co-occurrence between the two mutations. This could indicate that in the metastatic stage of colorectal cancer, the loss of ARID1A, while it occurs, might be a passenger event and therefore redundant. This is similar to what was observed in the progression of PDAC in the mouse model described in section 2.8. The re-expression of Arid1a in tumors formed by *Kras* activation and *Arid1a* deletion had no additional consequences. Therefore, in later stages perhaps, the loss of Arid1a does not have too much consequence and hence can co-occur with other mutations. Further analyses of patient material stratified by stage and mutational background need to be performed to confirm this hypothesis. However, these initial findings certainly suggest a cooperative role of ARID1A in KRAS driven transcriptional programs.

During this project we were able to show that many MEK/ERK driven programs downstream of a *KRAS* mutation were dependent on ARID1A. Upon its loss, genes induced by this pathway were downregulated. Further downstream, we found that ARID1A cooperates with the AP1 factor JunD at enhancers. We are not the first to show that ARID1A containing BAF complexes are important in AP1 driven enhancer regulation. Mathur et al, 2017 and Vierbuchen et al., 2017 showed this in colorectal cancer and in lineage determination respectively. However, we were able to elucidate a transcription network involving ARID1A and the AP1 transcription factor JunD which acts downstream of the MEK/ERK pathway. The loss of either factor led to the disruption of the transcriptional machinery at these sites; that is, the loss of ARID1A led to a reduction of JunD binding, while a loss of JunD led to a loss of ARID1A binding. Moreover, attenuation of the MEK/ERK pathway (with a MEK1/2 inhibitor, Trametinib), which would prevent the activation of JunD, also led to the loss of ARID1A binding. Moreover, all the perturbations just described also led to a reduction of H3K27ac from these enhancers suggesting a loss in enhancer activity. Indeed, the expression of genes associated with these enhancers is downregulated upon any of these perturbations.

The MEK/ERK signaling pathway is highly complicated and involves many proteins. In this project, we explored the role of the two most highly expressed AP1 factors, FosL1 and JunD (based on our RNA-seq data), in the HCT116 system. ChIP-seq data is also available for these two factors in the HCT116 system. While we observed a strong colocalization of these factors with ARID1A, this does not rule out the possibility of other AP1 factors being involved, as ARID1A localizes on an AP1 DNA binding motif. In any kind of colocalization analysis, we are limited by the information and data that are already available. Upon the knockdown of *JUND* and *FOSL1* we obtained quite variable effects on gene expression suggesting the involvement of other AP1 factors in this network. Genes that were potential targets of the enhancers bound by ARID1A/AP1 were not necessarily downregulated by the knockdown of either *JUND* or *FOSL1* (data not shown). This could partially be due to the redundant functions of the AP1 functions and compensation by other AP1 factors such as cJun, JunB and FosL2 which can be activated by the MEK/ERK pathway. We overcame this obstacle by using Trametinib which would block the activation of the AP1 factors at an upstream step. In the HCT116 cell line, upon treatment with Trametinib, most of the target genes were downregulated confirming that these enhancers are indeed targets of the MEK/ERK pathway. Furthermore, we confirmed that the effects we observed were at the level of transcription by checking for the activation of the MEK/ERK pathway upon the loss of ARID1A. Even though the EGFR receptor is downregulated upon the knockout of ARID1A, this has no effect on the activation of the pathway, as the levels of phosphorylated ERK do not decrease upon the knockout of ARID1A. Therefore, ARID1A regulates the transcriptional network induced by the MEK/ERK pathway.

Colon Adenocarcinoma Dataset	Number of Samples	Log Odds Ratio	Adjusted p-value	Tendency
DFCI, Cell Reports, 2016	619	-0.223	0.320	Mutual exclusivity
TCGA (PanCancer Atlas), 2018	594	-0.761	0.007	Mutual exclusivity
TCGA (Provisional), 2016	640	-0.559	0.171	Mutual exclusivity
MSKCC, metastatic colorectal cancer, 2018	1134	0.022	0.499	Co-occurence

Figure 29: Mutual exclusivity of *KRAS* and *ARID1A* mutations in colon adenocarcinomas and metastatic colorectal cancer. Analysis of *KRAS* and *ARID1A* mutations in patient samples reveals that while these mutations appear to be mutually exclusive in colorectal adenocarcinomas, there is a slight tendency towards co-occurrence in metastatic colorectal cancer suggesting different paths of tumor initiation by the two tumor suppressors.

6.8 Sensitivity to Trametinib Treatment

We used the inhibitor Trametinib to look at the effects of an attenuated MEK/ERK signaling in our study. Trametinib is a highly selective allosteric inhibitor of MEK1/2 that binds adjacent to the ATP binding site on MEK preventing its phosphorylation by RAF and hence its activation (Gilmartin et al., 2011). This would suggest that cancer cells that are particularly dependent on the activity of MEK/ERK pathway would be particularly sensitive to this inhibitor. Indeed, Trametinib was approved for therapy of *BRAF* mutated melanomas. However, due to development of resistance mechanisms, the combination of Trametinib with a BRAF inhibitor is now being used (Flaherty et al., 2012).

In colon cancer, Yamaguchi et al., 2011 reported that cell lines with *BRAF* and *KRAS* mutations were very sensitive to Trametinib while those which were wildtype for these factors were more resistant. As can be seen in Figure 18b, a concentration as low as 10 nM of Trametinib is enough to abrogate the phosphorylation of ERK completely in the HCT116 cell line. Moreover, when exploring the Genomics of Drug Sensitivity in Cancer database, we found this to be true for a larger panel of colorectal cancer cells lines. In our own screen, we tested the sensitivity of the four colorectal cancer cell lines we were studying to Trametinib. We tested this in both the *ARID1A* WT and KO conditions. As expected, the *BRAF* mutant cell line HT29 was extremely sensitive to Trametinib. However, surprisingly, we found that the *KRAS* mutant cell line DLD1 was more resistant to Trametinib treatment than the *KRAS* and *BRAF* wildtype cell line COLO320DM. Moreover, in a preliminary study in Figure 18 we showed that all four cell lines were slightly more sensitive to Trametinib when lacking ARID1A. However, on a larger scale analysis of cell lines (on the Genomics of Drug Sensitivity in Cancer database), we saw that *ARID1A* mutation status does not render the cells more sensitive to Trametinib treatment. Therefore, the sensitivity of ARID1A mutated (and perhaps more appropriately ARID1A-deficient) cells to the inhibition of MEK1/2 needs to be confirmed on a larger scale. In our mechanistic model, we propose a transcriptional network consisting of ARID1A and AP1 factors at enhancers downstream of the MEK/ERK signaling pathway. Treatment with Trametinib disrupts this network by reducing the occupancy of JunD, ARID1A and the H3K27ac on the histones. Therefore, in ARID1A-proficient cells, the effect of Trametinib treatment could partially be explained by this disruption. Of course, Trametinib also modulates the more well-known targets of the MEK/ERK pathway which have defined roles in proliferation. In ARID1A-deficient systems, Trametinib treatment elicits a similar or slightly better response. This is conceivable as the

enhancer network we have described is already disrupted in the deficient system and a further disruption would yield a slight additive effect as compared to the proficient system.

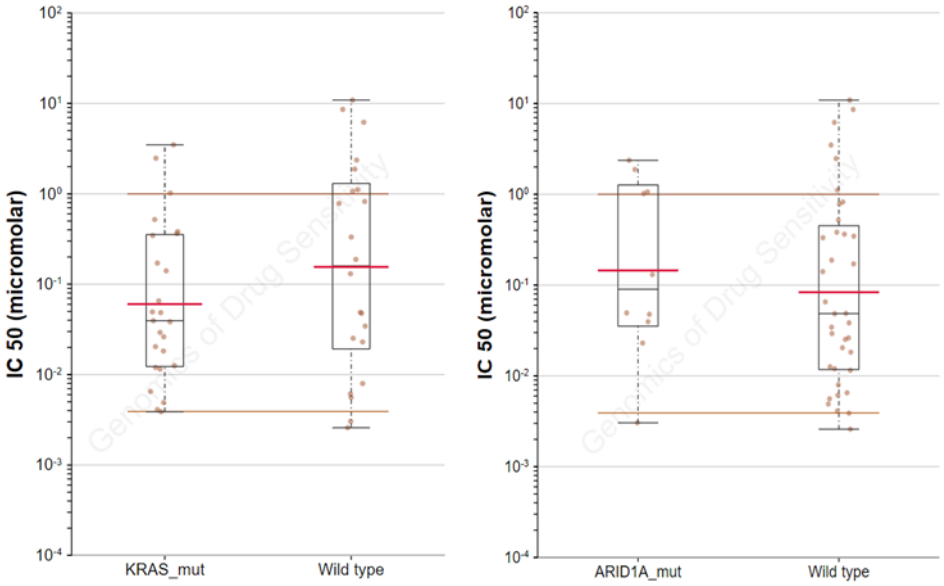


Figure 30: Sensitivity of colorectal cancer cell lines to Trametinib. KRAS mutated colorectal cancer cell lines are more sensitive to Trametinib treatment than their wildtype counterparts as they are especially dependent on the activity of the MEK/ERK pathway. On the other hand, ARID1A mutated colorectal cancer cell lines display a slightly lower sensitivity to Trametinib as compared to their wildtype counterparts. This could potentially be explained by an already disrupted transcriptional network at the ARID1A/AP1 enhancers in the absence of ARID1A. These figures were generated on the Genomics of Drug Sensitivity Database (Yang et al., 2013).

6.9 Inflammation Dampening

As a chromatin remodeller, the BAF complex has been described to have activating functions in gene expression regulation (Clapier, 2017, Review). However, several reports have described a repressive function for the BAF complex, in which it binds to the promoter of target genes to repress their activity (Chandler et al., 2015, Wu et al., 2018, Bitler et al., 2017). Chandler et al., 2015 demonstrated that ARID1A containing BAF complexes repress the pro-inflammatory cytokine interleukin 6 (IL6) promoter. They were able to show that in the ARID1A-depleted system, the expression of IL6 is heightened. This creates a proinflammatory microenvironment which promotes the growth of the tumor.

In line with this, in our initial experiments, we wanted to look at the role of ARID1A in inflammatory signaling in the colorectal cancer system. The inflammatory tumor microenvironment has been shown to play a very important role in colorectal cancer

progression. To test this effect, we treated HCT116 cells with the Tumor Necrosis Factor alpha (TNFα) in ARID1A-proficient and deficient conditions. We observed that the inflammatory signaling was impaired in the KO setting, as assessed by the levels of the early response gene of this pathway, CXCL2. Similarly, the induction of *IL6* expression by TNFα was dampened upon the KO of ARID1A. While this was contrary to what was published, it fits with our model of an oncogenic function for ARID1A in colorectal cancer.

Upon examining the occupancy of ARID1A at the *IL6* promoter in the HCT116 system, we found very little occupancy (Figure 31). Similarly, near the *IL6* promoter, no potential ARID1A bound enhancers were found. Therefore, it seems that contrary to the published study, in the HCT116 system, ARID1A perhaps plays an activating role at the *IL6* promoter or enhancer, which would explain the dampening of its expression upon ARID1A loss. A chromatin immunoprecipitation experiment for ARID1A after the induction of the signaling pathway by TNFα would confirm whether it directly regulates the expression of IL6 by binding to and activating a regulatory element in our system, or if this is mediated *via* an indirect mechanism.

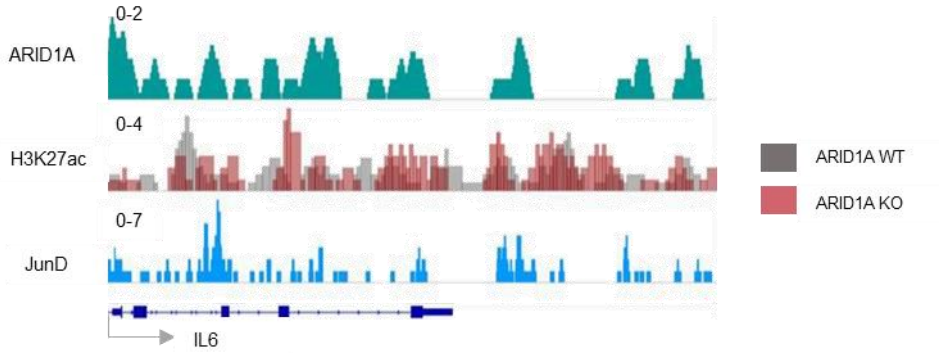


Figure 31: ARID1A, JunD and H3K27ac at the *IL6* promoter. The promoter of Il-6 in the HCT116 system, unlike the ovarian cancer system is not occupied by ARID1A and therefore the repressive function described in the literature in not applicable in this system.

Another potentially promising line of inquiry in this regard is with the E3 ubiquitin ligase associated function of ARID1A described by Li et al., 2010. Recently our group showed that the loss of the H2B ubiquitylating enzyme RNF40, dampens the inflammatory response elicited by TNFα treatment (Kosinsky et al., 2018). Perhaps ARID1A regulates inflammation as a part of a complex that can carry out H2B ubiquitylation. However, whether or not the H2B ubiquitylation levels are even affected by the loss of ARID1A needs to be verified in order to further explore this mechanism.

6.10 Synthetic lethality

Over the last few years several vulnerabilities for BAF complex defective cancers have been described. Some oncogenic functions of BAF complex subunits have been described such as in synovial carcinoma (Kadoch et al., 2013) and in intestinal cancers (Mathur et al., 2017, Holik et al., 2014). However, several subunits were described as tumor suppressors which means their expression in cancer is lost and they cannot be targeted directly. Therefore, synthetic lethality was seen as a suitable method to target these tumors. Moreover, many of the identified vulnerabilities could be targeted using small molecule inhibitors. While these have shown some promise, it has become increasingly clear that the context of the tumor must be considered carefully before applying these therapies to ARID1A-deficient cancers.

One of the best described synthetic lethalities is based on residual activity of the BAF complex. Large scale synthetic lethality screens revealed ARID1B to be the top hit for ARID1A deficient cancers (Helming et al., 2014). Though ARID1B is not druggable, reports have shown that depletion of ARID1B reduced the tumorigenic effect of ARID1A loss (Helming et al., 2014, Mathur et al., 2017). In an indirect mechanism, it was shown that BET inhibition downregulated ARID1B and consequently made ARID1A-deficient systems more sensitive to treatment (Berns et al., 2017). BET inhibitors, however, affect the expression of several other genes, as BET proteins are a major class of regulators of enhancer activity. When we tested this vulnerability in our system we observed that the depletion of ARID1B also led to an impairment in proliferation of the HCT116 cells suggesting once again the oncogenic role of the BAF complex in at least the HCT116 cell line. Moreover, depletion of both components had an additive effect. ARID1A and ARID1B containing BAF complexes are probably targeted to enhancers redundantly and uniquely, which would explain the additive effect. Previously, this synthetic lethality was also mechanistically explained based on accessibility of the chromatin (Kelso et al., 2017). ARID1B was shown to play a role in maintaining chromatin accessibility only in an ARID1A-deficient condition. However, this conclusion was based on SMARCA4 and SMARCC1 binding. As we know, SMARCA4 is a mutually exclusive subunit and there are BAF complexes that contain SMARCA2 instead. Therefore, it was important to draw conclusions about the effect of ARID1A loss on accessibility of the chromatin, based on ARID1A occupancy. When we analyzed the ATAC-seq data ourselves, we found that upon the loss of ARID1A, there were very few regions in the genome that became inaccessible. Rather, overall, there was no significant change in openness of chromatin. Moreover, at regions actually bound by ARID1A, there was also no change in accessibility. This points to the fact that while the BAF complex has chromatin remodelling functions, the loss of ARID1A does not prevent this activity. From our results, it appears that ARID1A containing BAF complexes act as co-factors, by recruiting other

components of the transcriptional machinery at enhancers to ensure correct spatial and temporal gene expression. Therefore, to know that relationship between the two mutually exclusive subunits, ARID1A and ARID1B, we would also need to know the occupancy of ARID1B in order to determine their unique and redundant functions in maintaining enhancer activity downstream of the MEK/ERK pathway.

Another well-known synthetic lethality described in the literature is that with EZH2. It has been shown that *Smarcb1* depleted tumors in mice are rescued to some extent by depletion of *Ezh2* (Roberts et al., 2000). As explained in section 2.7 this is due to the antagonism between the BAF complex and the PRC2 complex. While this antagonism seems to be true in the context of development, in recent years, the small molecule inhibitor of EZH2 has been also been proposed as a synthetic lethal therapy for ARID1A mutant tumors. Bitler et al., 2015 tested the efficacy of the inhibitor on various cell lines and also tested an *in vivo* model. They found that ARID1A-deficient models were highly sensitive to the inhibition of EZH2. This was explained by the removal in repression set by EZH2 at BAF-bound sites (in the absence of ARID1A) by the inhibitor. Moreover, it was proposed that cells with mutation of *KRAS* were able to overcome this vulnerability due to some non-enzymatic functions of EZH2 (Kim et al., 2015). We tested both the sensitivity and resistance in colorectal cancer and cholangiocarcinoma cell lines (where ARID1A mutations are also common). We found that, in these two systems, irrespective of the *KRAS* status, ARID1A-depleted cells were not more sensitive to EZH2 inhibition. Moreover, interestingly, we observed that upon treatment with the EZH2 inhibitor, the proliferation of all the tested cell lines increased. Liu et al., 2017 showed that the knockout of EZH2 in mice leads to inflammation in the colon (a risk factor for CRC). We also hypothesized that this could be due to the lift in repression of some proliferative genes. We also proposed a potential explanation why this synthetic lethal relationship was not validated in our system. When we looked at the H3K27me3 signal at ARID1A bound sites, there was close to no signal at these sites. Moreover, upon the KO of *ARID1A*, the signal of H3K27me3 remained the same, suggesting that even in the context of a defective BAF complex, EZH2 does not methylate these regions and therefore is not specifically sensitive to its inhibition.

Since ARID1A-bound sites do not seem to be trimethylated upon the loss of ARID1A, it seems that the antagonism between the BAF complex and PRC2 complex is likely to not exist in this context. Thus, while EZH2 inhibition has shown promise in *Smarcb1*-mutated rhabdoid tumors and *ARID1A*-mutated ovarian clean cell carcinoma, this relationship does not appear to be applicable in colorectal cancer cells.

6.11 Future Directions

It is evident that ARID1A-containing BAF complexes play a role in enhancer mediated gene regulation and this role is beyond their chromatin remodelling functions. It is also clear that these complexes act downstream of several oncogenic signaling pathways to modulate gene expression *via* enhancers and thus promote oncogenic programs in certain contexts. Thus, in the future, it would be important to decipher the gene regulatory functions of the BAF complex. 3D chromatin interaction experiments will be required to determine the relationship between BAF bound enhancers and their potential target promoters. One potentially productive way to show this at the individual gene level would be to inactivate the enhancers identified by tethering a repressive KRAB domain on them and then look at gene expression changes. Moreover, looking at changes in the 3D architecture of the chromatin after the loss of ARID1A might also be enlightening. Additionally, the effects ARID1A loss on the stability of the BAF complex and its targetability to different regions of the chromatin needs to be studied more clearly.

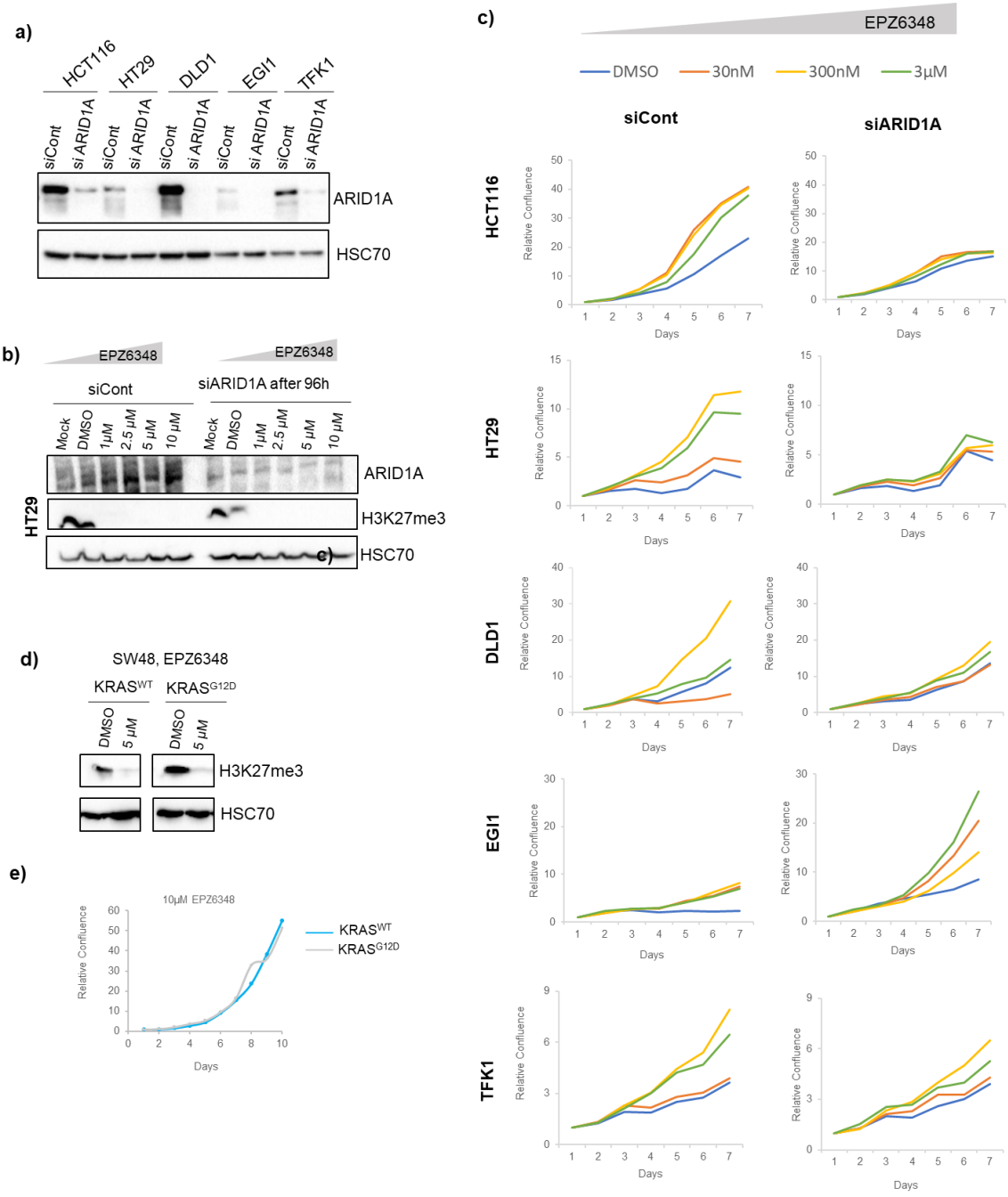
In a more clinical perspective, while this study has uncovered more information about ARID1A-mutated colorectal cancers, some pertinent questions remain. For example, the stage and dose dependency of ARID1A loss in CRC is still not completely clear. However, as mentioned, since its loss drives certain oncogenic programs, it would be interesting to explore the possibility of targeting the BAF complex itself in these contexts. While depletion of *ARID1B* in HCT116 cells has shown a similar trend in proliferation as the loss of ARID1A, the depletion of proteins is not an appropriate approach clinically. Recently, the development of Proteolysis Targeting Chimera (PROTACs) have made the degradation of proteins *in vivo* possible. These molecules consist of two recognition domains, one which recognizes the protein of interest and the other which engages an E3 ubiquitin ligase that tags the protein for proteasomal degradation (Cermakova et al., 2018, Review). Indeed, promising results have been observed in synovial carcinoma, in which SS18-SSX containing oncogenic BAF complexes are degraded by a small molecule degrader of BRD9 (also a subunit of the BAF complex). This leads to a regression of tumors *in vivo* (Brien et al., 2018). Thus, it would be very exciting to test whether PROTAC-mediated degradation of the BAF complex in *KRAS* (and perhaps *APC*) mutated colorectal cancer yields therapeutic benefit. This would allow development of a therapy for a large fraction of colorectal cancers based on solid molecular mechanisms.

6.12 Concluding Remarks

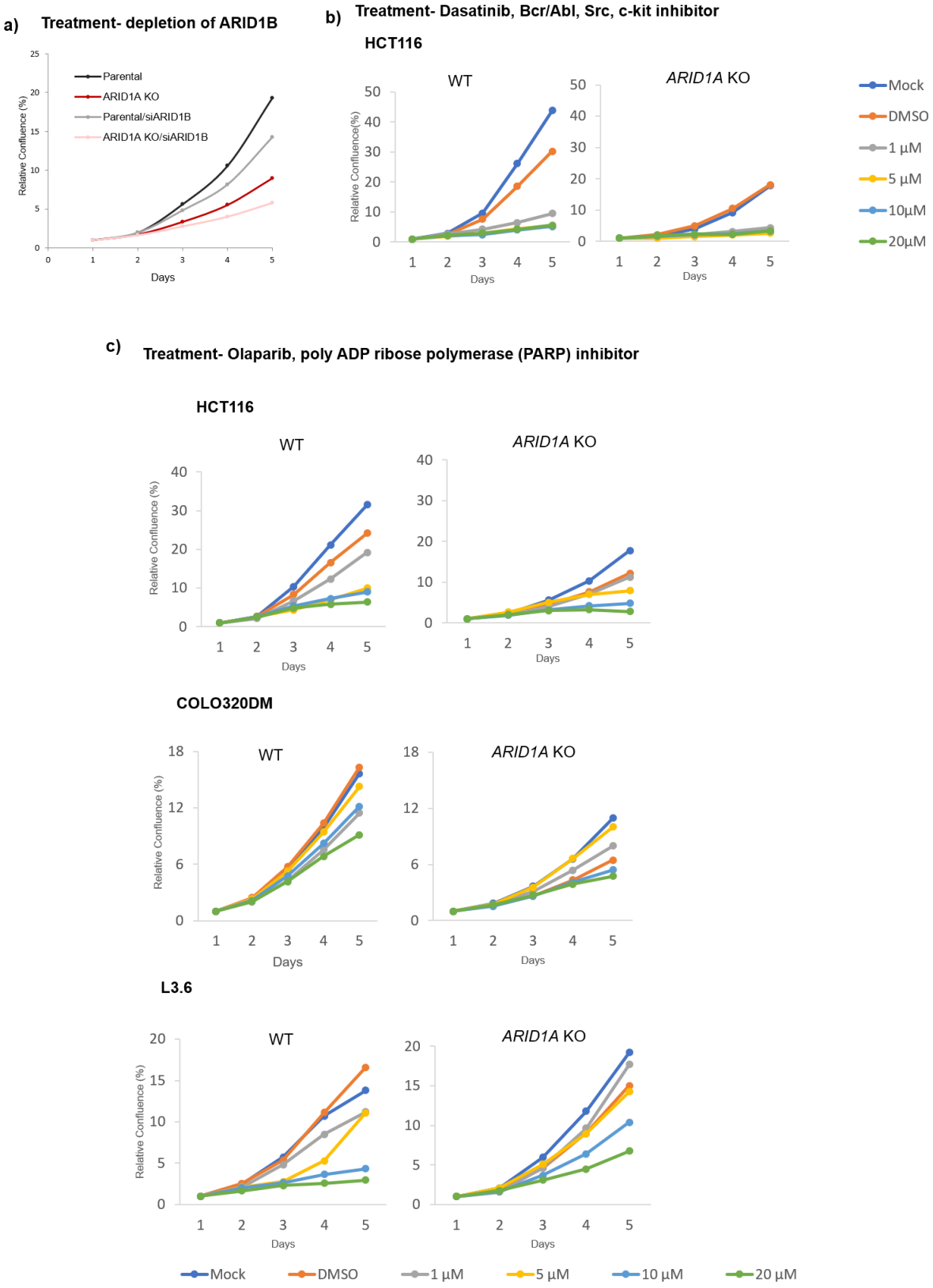
Taken together, we were able to add to the existing knowledge about the context-dependent functions of ARID1A in cancer. While we confirmed that ARID1A is oncogenic in specific contexts of colorectal cancers we were able to show mechanistically that *KRAS*-mutated colorectal cancer cells are especially dependent on the presence of ARID1A in the cells. In cells that are *KRAS*-mutated, upon the loss of ARID1A, the transcription network at enhancers consisting of ARID1A and the AP1 transcription factors is disrupted and deregulation of gene expression occurs. This is accompanied by an impairment in proliferation of these cells. Moreover, like much other literature we also showed that ARID1A has roles beyond its chromatin remodelling activity, in this case as a transcriptional cofactor. Further, we expanded on already existing knowledge about the interplay of ARID1A and AP1 transcription factors by showing its relevance in the development of a subset of colorectal cancers. We even suggest the possibility of the BAF complex being a targetable entity in these cancers. On exploring this further, we might gain important insights into the mechanisms by which ARID1A-containing BAF complexes regulate enhancers as well as how this knowledge can be used to therapeutic benefit.

7. Supplemental Figures

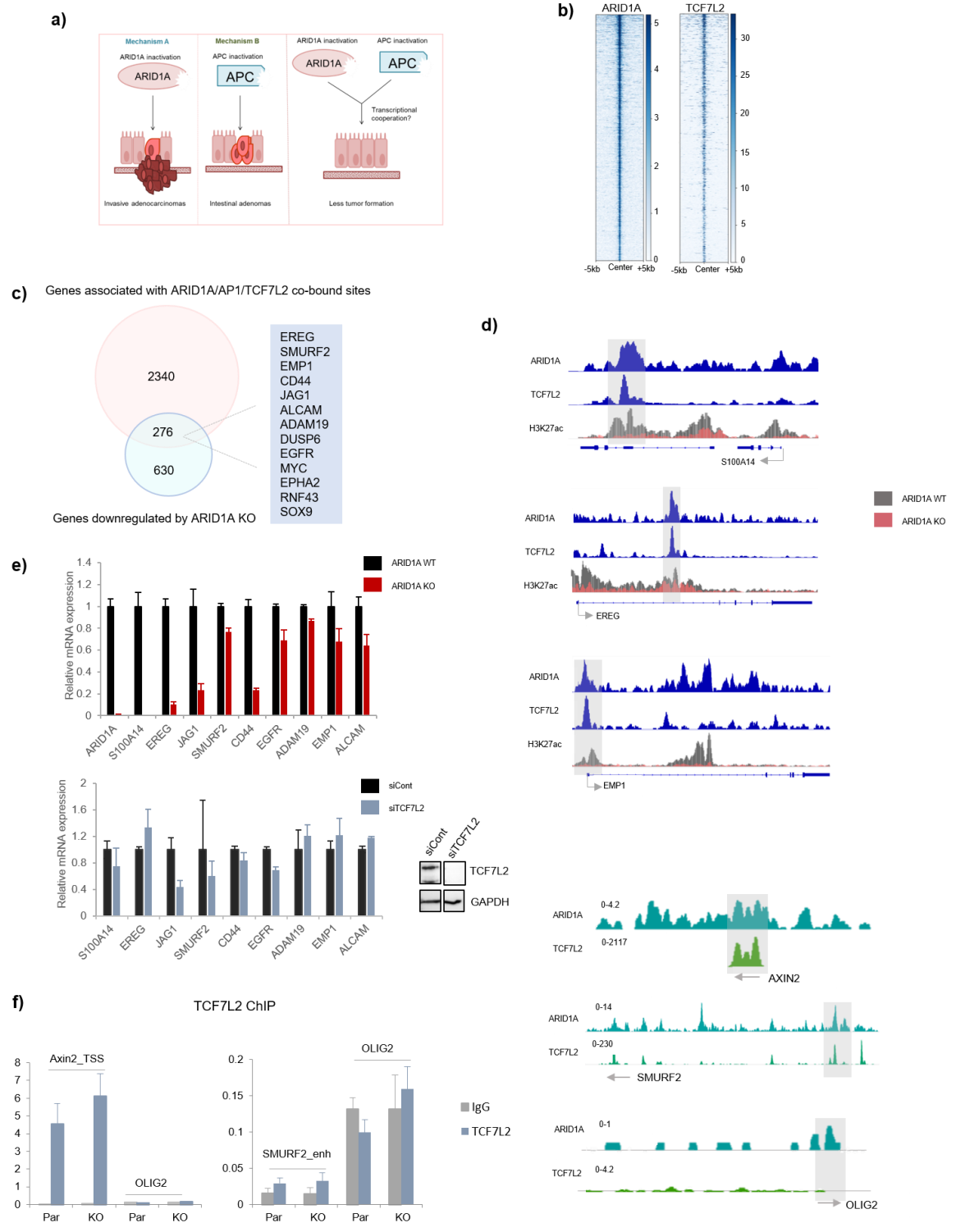
7.1 Supplemental Figure 1a



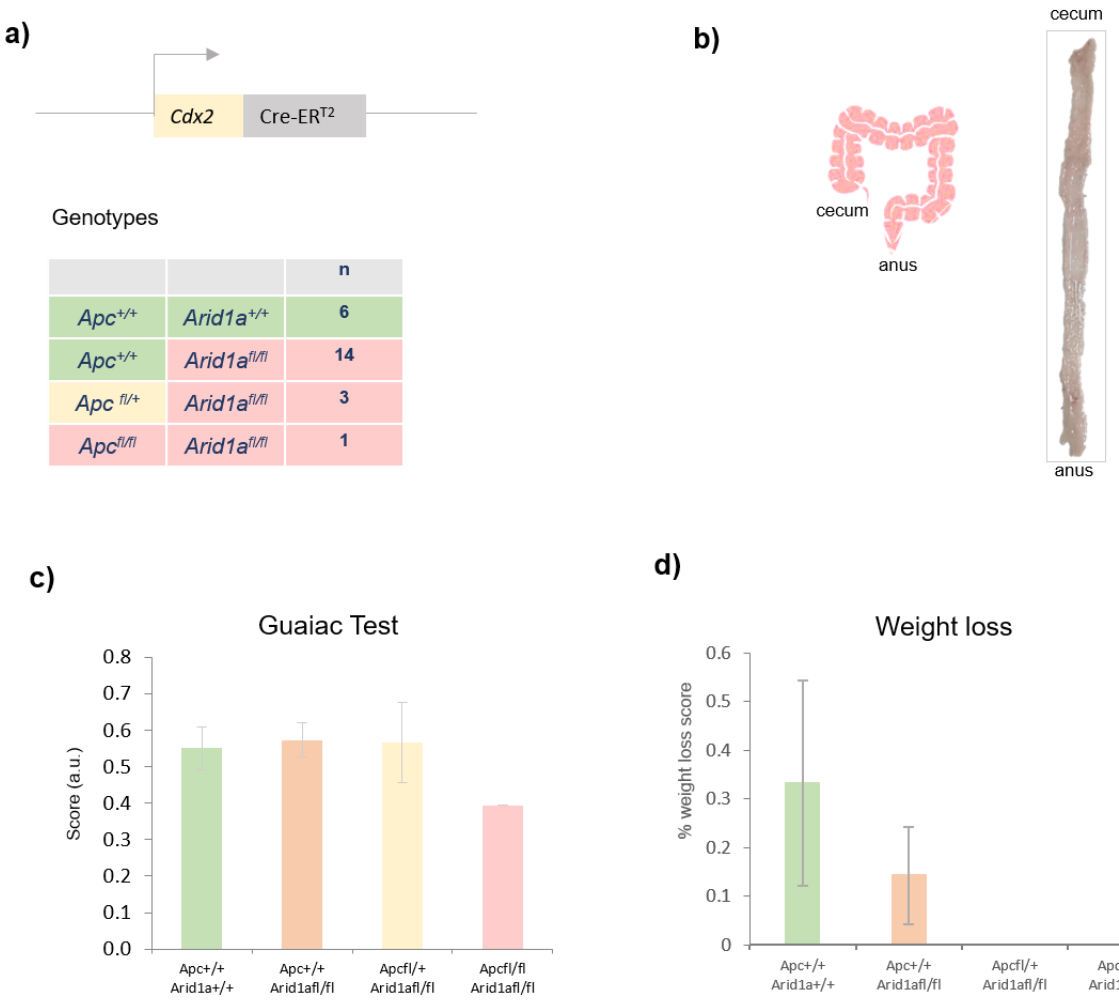
7.1 Supplemental Figure 1b



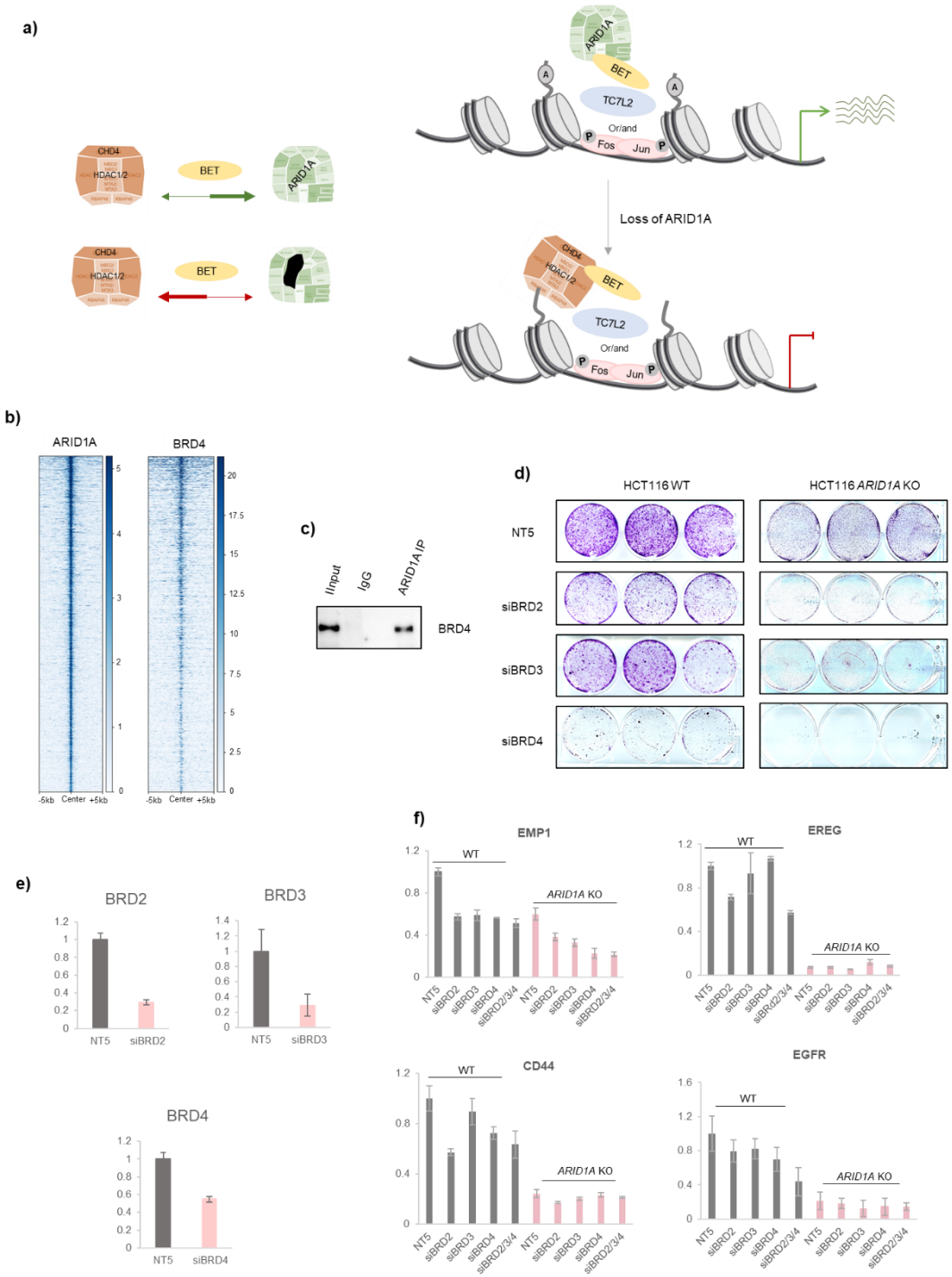
7.2 Supplemental Figure 2



7.3 Supplemental Figure 3



7.4 Supplemental Figure 4



7.5 Supplemental Figure Legends

Supplemental Figure 1a: Testing published synthetic lethal targets in ARID1A-deficient systems. With the burst of research in the role of ARID1A as a tumor suppressor, many studies have focussed on finding synthetic lethal targets in ARID1A-deficient systems. One of the major synthetic lethalities was described for EZH2 wherein ARID1A-deficient cells were more sensitive to the inhibition of this H3K27 methyltransferase enzyme by a small molecule inhibitor EPZ6438. To test this in colorectal cancer cells and cholangiocarcinoma cells where ARID1A mutations are also prevalent (EGI1 and TFK1), we depleted ARID1A from these cells using siRNA mediated knockdown (a). In the HT29 cell line we validated that the inhibitor depleted levels of H3K27me3 by inhibiting EZH2 (b). Contrary to the published findings, in all the cell lines tested we found that the ARID1A-deplepeted cells were not more sensitive to the inhibition of EZH2 (c). Strikingly, all the cell lines proliferated at even higher rates in the EPZ treated conditions as compared to the control treatment. The synthetic lethality with EZH2 has been explained by the competitive interplay of the BAF complex and PRC2 complex (of which EZH2 is a part) at common target sites. However, we observed in the HCT116 cell line that ARID1A bound sites hardly harbor the H3K27me3 mark even in the *ARID1A* KO system (shown in Figure 19b). Another study showed that the sensitivity to EPZ in ARID1A-deficient cells is abrogated in the presence of a KRAS mutation. We also compared this in the isogenic CRC cell lines SW48 and SW48^{G12D} which harbors a KRAS mutation. While the KRAS mutant cells had a much higher level of H3K27me3 which was completely diminished by the inhibitor, the treatment did not have any differential effect in the proliferation of the two cell lines (these were partially performed during my master's thesis, 2016). Together, these suggested that this synthetic lethality and its associate resistance mechanisms is not valid in CRC cell lines.

Supplemental Figure 1b: Testing published synthetic lethal targets in ARID1A-deficient systems. With the burst of research in the role of ARID1A as a tumor suppressor, many studies have focussed on finding synthetic lethal targets in ARID1A-deficient systems. One non-targetable synthethic lethality is the mutually exclusive subunit or ARID1A, ARID1B. We depleted ARID1B in the HCT116 WT and *ARID1A* KO cells by siRNA mediated knockdown. The loss of ARID1B alone led to an impairment in proliferation and when this occurred in the context of *ARID1A* KO, it was further reduced (a). This confirmed that the loss of BAF complex subunits in the HCT116 system does not enhance its oncogenicity and the loss of both mutually exclusive subunits has an additive effect on proliferation. Furthermore, we tested two other reported synthetic lethalities. In (b) we tested the BCL/ABL, SRC inhibitor Dasatinib in HCT116 WT and KO cells. While we used very high concentrations of the drug, we saw no differential effects. Similarly, we tested the PARP inhibitor (c) Olaparib in two CRC cell lines, HCT116 and COLO320DM and one pancreatic cancer cell line, L3.6. In this case as well we observed no synthetic lethality between ARID1A depletion and inhibition of PARP, in any of the cells we tested.

Supplemental Figure 2: ARID1A is required for Wnt signaling driven colorectal cancer. As shown by Mathur et al., in APC mutation driven colon cancer, ARID1A plays a crucial role and in its absence no tumors are formed (a). We used the β-catenin mutated CRC cell line HCT116 to explore this further. This mutation has the same effect as the APC mutation, that is, hyperactivation of the Wnt pathway. We found that the KO of *ARID1A* impairs proliferation in this cell line (Figure 12c) suggesting its requirement in the Wnt pathway. Very interestingly we observed a very strong colocalization between ARID1A and TCF7L2 indicating that these work at a subset of similar target sites (b). These sites were also occupied the AP1 transcription factors and they sites are mostly distal regulatory elements. Subsequently, we determined the genes that are associated with the binding of these transcriptional regulators and overlapped them with the set of genes downregulated upon *ARID1A* KO (c). As can be seen in (d), at three exemplary loci, regulatory elements are occupied by TCF7L2 and ARID1A. As in the case of AP1 co-occupied sites, here too, the H3K27ac reduces considerably upon the KO *ARID1A.*On examining the gene regulation by qPCR, we observed that while many genes were downregulated by *ARID1A* KO, these targets were most often not affected by the depletion of TCF7l2 (e). Since TCF7L2 is a DNA binding transcription factor, we assumed that the loss of this factor would lead to the loss of expression of these genes. Furthermore, we performed ChIP-qPCR for TCF7L2 (f) in the WT and *ARID1A* KO cells to determine if the binding of TCF7L2 is lost upon ARID1A loss. While the experiment worked, as can be seen for qPCRs for the positive site AXIN2 promoter (a direct Wnt pathway target) and negative site, OLIG2, we could not detect any signal for our sites of interest. As seen at the exemplary locus SMURF2 which showed the highest signal in ChIP-seq among our sites of interest (10X lower than AXIN2), the signal in qPCR was lower than that of the OLIG2. This could be due to indirect binding of TCF7L2 at distal regulatory elements (as compared to the

Supplemental Figure 3: *In vivo* genetic mouse model for Arid1a loss. A genetic model to study the loss of *Arid1a* from the mouse colonic epithelium was created. (a) In this model, the Cre-ER^{T2}(Cre recombinase fused to a mutant estrogen receptor which is inducible by tamoxifen) was placed under the control of the colon specific gene *Cdx2* promoter. To study this in the background of the commonly occurring *Apc* mutation we used four genotypes wherein *Arid1a* or *Apc* or both were flanked by LoxP sites (a). Weight was monitored, and Guaiac test was performed every week for 6 months to assess the effects of Arid1a loss in the *Apc* WT and deleted background. After 6 months the mice were sacrificed, and the colons were harvested. No tumors were found in any genotype as can be seen in a representative image (b). Consistently, Guaiac tests which measures the intensity of intestinal bleeding showed no difference between the WT and KO groups (c). All genotypes suffered very minimal weight loss (d).

Supplemental Figure 4: Hypothesis: competitive interaction of the BET proteins with the BAF complex and the NuRD complex. In another hypothetical model we sought to explain the loss of H3K27ac from the ARID1A/TCF7L2/AP1 bound enhancers upon the knockout of *ARID1A.* It has been shown that some of the Bromodomain and Extra-terminal domain proteins (BET) interact with both the BAF complex and the NuRD complex, another chromatin remodeller which is repressive and consists of the histone deacetylase HDAC1/2. In the model depicted in (a), we hypothesized that in the cell the BET proteins competitively interact with the BAF and NuRD complexes and recruit them to TCF7L2/AP1 enhancers. In the WT condition, this competition favours the interaction with the BAF complex, and the correct transcriptional machinery assembles, and the target genes are transcribed. When ARID1A is lost, perhaps the BET proteins now interact with the NuRD complex bringing it to the TCF7L2/AP1 enhancers. Its deacetylase activity removes the acetylation from H3K27, silencing the target genes. We performed a few initial experiments to test this hypothesis. Firstly, we checked the interaction of the BET protein BRD4 with ARID1A in the HCT116 system. There was a strong colocalization genome wide at ARID1A binding site (b). Moreover, we were able to prove this colocalization by coimmunoprecipitation (c). We looked at the effects on proliferation upon depletion of the 3 BET proteins BRD2,3 and 4 expressed in the HCT116 system. The impairment in proliferation upon *ARID1A KO* can also be observed on crystal violet staining (d). Moreover, both the WT and KO cells respond similarly to the depletion of these proteins. While BRD4 loss was the most disadvantageous for them, BRD3 loss reduced proliferation but at a subtler level (d). Thus, differential effects in the transcription network as explained above did not manifest in proliferation differences. This could be due to compensatory mechanisms and several other independent roles of the BET proteins. To explore deeper, we looked at the transcriptional changes in some of the target genes upon BRD2, 3, and 4 (individually and together) knockdown, in the WT and *ARID1A KO* condition. We predicted that the depletion of the BET proteins might rescue the gene expression attenuated by the recruitment of HDAC1/2. While the knockdowns worked with moderate effects (e), we could not rescue gene expression of target genes downregulated by the KO of *ARID1A* (f). The interplay of the BET proteins through their interactions with these chromatin remodellers is very interesting and would be fascinating to investigate further. BAF and NurD complex structure adapted from Swetansu et al., 2016.

8. References

Ahmed, D., Eide, P. W., Eilertsen, I. A., Danielsen, S. A., Eknæs, M., Hektoen, M., Lind, G. E., Lothe, R. A. (2013). Epigenetic and genetic features of 24 colon cancer cell lines. Oncogenesis, *2*.

Akhtar-Zaidi, B., Cowper-Sal-lari, R., Corradin, O., Saiakhova, A., Bartels, C. F., Balasubramanian, D., Myeroff, L., Lutterbaugh, J., Jarrar, A., Kalady, M. F., Willis, J., Moore, J. H., Tesar, P. J., Laframboise, T., Markowitz, S., Lupien, M., Scacheri, P. C. (2012). Epigenomic enhancer profiling defines a signature of colon cancer. Science *336*, 736-9.

Albig, W. and Doenecke, D. (1997). The human histone gene cluster at the D6S105 locus. Hum. Genet. *101*, 284–294.

Alver, B.H., Kim, K.H., Lu, P., Wang, X., Manchester H.E., et al. (2017). The SWI/SNF chromatin remodeling complex is required for maintenance of lineage specific enhancers. Nat. Commun. *8*, 14648

Anaya, J. (2016). OncoLnc: linking TCGA survival data to mRNAs, miRNAs and lncRNAs.PeerJ CompSci.

Arnold, M., Sierra, M.S., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.(2017) Global patterns and trends in colorectal cancer incidence and mortality. Gut *66*, 4, 683-91

Baranello, L, Wojtowicz, D., Cui, K., Deaiah, B.N., et al.,RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. (2016). Cell *165*, 357–371.

Barker, N., Hurlstone, A., Musisi, H., Miles, A., Bienz, M., & Clevers, H. (2001). The chromatin remodelling factor Brg-1 interacts with beta-catenin to promote target gene activation. The EMBO journal, *20*,17, 4935-43.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. a, Kim, S.,Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity Supp. Nature, *483*, 7391, 603–7.

Bennett, R.L., and Licht, J.D. (2018). Targeting Epigenetics in Cancer. Annu. Rev. Pharmacol. Toxicol. *58*, 187–207.

Berns, K., Caumanns, J.J., Hijmans, E.M., Gennissen, A.M.C, Severson, T.M., Evers, B., Wisman, G.B.A., Jan Meersma, G., Lieftink, C., Beijersbergen, R.L., Itamochi, H., van der Zee, A.G.J., de Jong, S., Bernards, R. (2018). ARID1A mutation sensitizes most ovarian clear cell carcinomas to BET inhibitors. Oncogene, *37*, 33, 4611-4625.

Bevilacqua, A., Willis, M. S., & Bultman, S. J. (2013). SWI/SNF chromatin-remodeling complexes in cardiovascular development and disease. Cardiovascular pathology *23*, 2, 85-91.

Bitler, B.G., Aird, K.M., Garipov, A., Li, H., Amatangelo, M., Kossenkov, A. V., Schultz, D.C., Liu, Q., Shih, I.M., Conejo-Garcia, J.R., et al. (2015). Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. Nat. Med. *21*, 231–238.

Bitler, B.G., Fatkhutdinov, N., and Zhang, R. (2015). Potential therapeutic targets in *ARID1A* -mutated cancers. Expert Opin. Ther. Targets *19*, 1419–1422.

Bitler, B.G., Wu, S., Park, P.H., Hai, Y., Aird, K.M., Wang, Y., Zhai, Y., Kossenkov, A. V., Vara-Ailor, A., Rauscher, F.J., et al. (2017). ARID1A-mutated ovarian cancers depend on HDAC6 activity. Nat. Cell Biol. *19*, 962–973.

Brien, G. L., Remillard, D., Shi, J., Hemming, M. L., Chabon, J., Wynne, K., Dillon, E. T., Cagney, G., Van Mierlo, G., Baltissen, M. P., Vermeulen, M., Qi, J., Fröhling, S., Gray, N. S., Bradner, J. E., Vakoc, C. R., … Armstrong, S. A. (2018). Targeted degradation of BRD9 reverses oncogenic gene expression in synovial sarcoma. eLife, *7*.

Bruns, C.J., Harbison, M.T., Kuniyasu, H., Eue, I., and Fidler, I.J. (1999). In vivo selection and characterization of metastatic variants from human pancreatic adenocarcinoma by using orthotopic implantation in nude mice. Neoplasia *1*, 50–62.

Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., et al. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. Molecular Cell *6*, 1287–1295.

Cajuso, T., Hänninen, U.A., Kondelin, J., Gylfe, A.E., Tanskanen, T., Katainen, R., et al. (2014) Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B, ARID2 and ARID4A in microsatellite unstable colorectal cancer. International Journal of Cancer *135*, 611–623.

Carpenter, C.L. (2000). Actin cytoskeleton and cell signaling. Crit. Care Med. *4,* 94-9.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404.

Cermakova, K. and Hodges, C.H. (2018). Next-Generation Drugs and Probes for Chromatin Biology: From Targeted Protein Degradation to Phase Separation. Molecules *23*,8.

Chandler, R.L., Damrauer, J.S., Raab, J.R., Schisler, J.C., Wilkerson, M.D., Didion, J.P., Starmer, J., Serber, D., Yee, D., Xiong, J., et al. (2015). Coexistent ARID1A-PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. Nat. Commun. *6*, 1–14.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *128*, 14.

Chen, R.H., Sarnecki, C., Blenis, J.(1992). Nuclear localization and regulation of erk- and rsk-encoded protein kinases. Mol Cell Biol. *12*, 915–927

Chèneby, J., Gheorghe, M., Artufel, M., Mathelier, A., & Ballester, B. (2017). ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic acids research, *46*.

Chi, T.H., Wan, M., Zhao K., Taniuchi I., Chen L., Littman D.R., Crabtree G.R. (2002). Reciprocal regulation of CD4/CD8 expression by SWI/SNF-like BAF complexes. Nature *418*, 195–199.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat. Rev. Mol. Cell Biol. *18*, 407–422.

Clevers, H., Nusse, R.(2012). Wnt/β-Catenin Signaling and Disease. Cell *149*, 1192-1205.

Curtis, C. D., and Griffin, C. T. (2012). The chromatin-remodeling enzymes BRG1 and CHD4 antagonistically regulate vascular Wnt signaling. Molecular and cellular biology *32*, 7, 1312-20.

De Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499–506.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., Muller, H., Ragoussis, J., Wei, C. L., Natoli, G. (2010). A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. PLoS Biol. *8*.

Dey, A., Chitsaz, F., Abbasi, A., Misteli, T. & Ozato, K. (2003). The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. PNAS *100*, 8758–8763.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012).

Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Ernst T., Chase A.J., Score J., Hidalgo-Curtis C.E., Bryant C., et al. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat. Genet. *42*, 722–26.

Fearon, E. R. (2011). Molecular Genetics of Colorectal Cancer. Ann. Rev. of Path.: Mechanisms of Disease, *6*,1, 479–507.

Fearon, E.R. (2011). Molecular Genetics of Colorectal Cancer. Annu. Rev. Pathol. Mech. Dis. *6*, 479–507.

Fearon, E.R., Vogelstein, B., (1990). A genetic model for colorectal tumorigenesis. Cell *61* ,5.

Filippakopoulos, P., Qi, J., Picaud, S., et al. (2010). Selective inhibition of BET bromodomains. Nature *468*, 7327, 1067-73.

Flaherty, K., Robert, C., Hersey, P., Nathan, P., Garbe, C., Milhem, M., et al. (2012) METRIC Study roup. Improved survival with MEK inhibition in BRAF-mutated melanoma. N Engl J Med *367,*107–14.

Gao X., Tate P., Hu P., Tjian R., Skarnes W.C., Wang Z. (2008) ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. PNAS *105*, 6656–6661.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O.,chultz, N.(2013) Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal Complementary Data Sources and Analysis Options. Sci. Signal *6*, 269, 1–20.

Gilmartin, A.G., Bleam, M.R., Groy, A., Moss, K.G., Minthorn. E., Kulkarni, S.G., et al., (2011). GSK1120212 (JTP-74057) Is an Inhibitor of MEK Activity and Activation with Favorable Pharmacokinetic Properties for Sustained In Vivo Pathway Inhibition. Clin. Cancer Res. *17*, 5, 989-1000.

Gresh, L., Bourachot B., Reimann A., Guigas B., Fiette L., Garbay S., Muchardt C., Hue L., Pontoglio M., Yaniv M., et al. (2005). The SWI/SNF chromatin-remodeling complex subunit SNF5 is essential for hepatocyte differentiation. EMBO J. *24*, 3313–3324.

Guan, B., Wang, T.L., and Shih, I.M. (2011). ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers. Cancer Res. *71*, 6718–6727.

Gustavsson, B., Carlsson, G., Machover, D., Petrelli, N., Roth, A., Schmol,l H.J., Tveit K.M., Gibson F.(2015) A Review of the Evolution of Systemic Chemotherapy in the Management of Colorectal Cancer. Clin Colorectal Cancer *14*,1-10.

Haigis, K.M., Kendall, K.R., Wang, Y., et al. (2008). Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. Nat Genet. *40*, 600-8.

Heinz, S., Romanoski, C.E., Benner, C., Glass, C.K. (2015). The selection and function of cell type specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 3, 144-154

Helming, K.C., Wang, X., Roberts, C.W.M. (2014). Vulnerabilities of mutant SWI/SNF complexes in cancer. Cancer Cell *26,* 309–17

Hewish,D.R. and Burgoyne, L.A. (1973). The digestion of chromatin DNA at regularly spaced sites by a nuclear dexoyribonucelase.Biochem. and Biophy. Research Comm. *52*, 2, 504-510.

Hinoi, T., Akyol, A., Theisen, B.K., Ferguson, D.O., Greenson, J.K., Williams, B.O., et al. (2007) Mouse model of colonic adenoma-carcinoma progression based on somatic Apc inactivation. Cancer Res. *67*, 9721–30.

Hirschhorn, J.N., Brown, S.A., Clark, C.D., Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes Dev. *6*, 12A, 2288–98.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke H.A., Young R.A.(2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–47.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. PNAS *106*, 5187–5191.

Holik, A. Z., Young, M., Krzystyniak, J., Williams, G. T., Metzger, D., Shorning, B. Y., & Clarke, A. R. (2014). Brg1 loss attenuates aberrant wnt-signalling and prevents wnt-dependent tumourigenesis in the murine small intestine. PLoS genetics *10*, 7.

Hota, S. K., & Bruneau, B. G. (2016). ATP-dependent chromatin remodeling during mammalian development. Development *143*, 16, 2882-97.

Hu, G., Schones, D.E., Cui, K., Ybarra, R., Northrup, D., et al. (2011). Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. Genome Res. *21*, 10, 1650–58

Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., et al. (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell *150*, 1107–20.

Issa, J.P. (2008). Colon Cancer:It's CIN or CIMP. Clin. Cancer Res. *14*, 19, 5939-40.

Issa, J. (2004). GpG mthylator phenotype in cancer. Nature Reviews Cancer *4*, 988–993.

Izumi, K. (2016). Disorders of Transcriptional Regulation: An Emerging Category of Multiple Malformation Syndromes. Molecular syndromology, *7*, 5, 262-273.

Jones, S., Wang, T.L., Shih, Ie M., Mao, T.L., Nakayama, K., Roden, R., et al. (2010). Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. Science *330*, 228–31.

Kadoch, C., Copeland, R.A., and Keilhack, H. (2016). PRC2 and SWI/SNF Chromatin Remodeling Complexes in Health and Disease. Biochemistry *55*, 1600–1614.

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat. Genet. *45*, 592–601.

Kadoch, C., and Crabtree, G. R. (2013) Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma, Cell 153, 71-85.

Kelso, T.W.R., Porter, D.K., Amaral, M.L., Shokhirev, M.N., Benner, C., and Hargreaves, D.C. (2017). Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. Elife *6*.

Kennison, J.A., Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. PNAS *85*, 8136–40.

Kim, K., Jung, H.Y., Oh, M., Cho, H., Lee, J., Lee, H.J., Jang, S., and Lee, M.S. (2015a). Loss of ARID1A Expression in Gastric Cancer : Correlation with Mismatch Repair Deficiency and Clinicopathologic Features. *15*, 201–208.

Kim, K.H., Kim, W., Howard, T.P., Vazquez, F., Tsherniak, A., Wu, J.N., Wang, W., Haswell, J.R., Walensky, L.D., Hahn, W.C., et al. (2015b). SWI/SNF-mutant cancers depend on catalytic and non-catalytic activity of EZH2. Nat. Med. *21*, 1491–1496.

Kim, T.K. Gray J.M., Costa A.M., Bear D.M., Wu J., Harmin D.A., Laptewicz M., Barbara-Haley K., Kuersten S., Markenscoff-Papadimitriou E., Kuhl D., Bito H. Worley P.F., Kreiman G., Greenberg M.E. (2010).Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kinzler, K. W. & Vogelstein, B.(1996). Lessons form hereditary colorectal cancer. Cell *87*, 159–170.

Kleer, C.G., Cao, Q., Varambally, S., Shen, R.,Ota, I., et al. (2003). EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. PNAS 100, 11606–11.

Klochendler-Yeivin, A, Fiette, L, Barra, J., Muchardt, C., Babinet, C., Yaniv, M. (2000). The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. EMBO Rep.*1*, 6, 500–6.

Knutson, S. K., Warholic, N. M., Wigle, T. J., Klaus, C. R., Allain, C. J., Raimondi, A., Keilhack, H. (2013). Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. PNAS *110,* 19, 7922–7.

Kosinsky, R.L., Chua, R.L, Qui, M., Saul, D., Mehlich, D., Ströbel, P., Schildhaus, H.U., Wegwitz, F., Faubion, W.A., and Johnsen, S.A. (2018). Loss of RNF40 Decreases NF-κB Activity in Colorectal Cancer Cells and Reduces Colitis Burden in Mice. J. Crohn's Colitis, *jjyy165*.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., McDermott, M.G., Monteiro, C.D., Gundersen, G.W., Ma'ayan A. (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research.

Kuramochi, H., Nakajima, G., Kaneko, Y., Nakamura, A., Inoue, Y., Yamamoto, M., Hayashi, K. (2012) Amphiregulin and epiregulin mRNA expression in primary colorectal cancer and corresponding liver metastases. BMC Cancer *12*, 88.

Kwan, S.Y., Cheng, X., Tsang, Y.T., Choi, J.S., Kwan, S.Y., Izaguirre, D.I., Kwan, H.S., Gershenson, D.M., Wong, K.K. (2016). Loss of ARID1A expression leads to sensitivity to ROS-inducing agent elesclomol in gynecologic cancer cells. Oncotarget *7,* 56933–56943.

Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E., and Green, M. R. (1994). Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. Nature *370*, 477-481.

Lakshminarasimhan, R., Andreu-Vieyra, C., Lawrenson, K., Duymich, C.E., Gayther S.A., Liang G., Jones P.A. (2017). Down-regulation of ARID1A is sufficient to initiate neoplastic transformation along with epigenetic reprogramming in non-tumorigenic endometriotic cells. Cancer Letters *401*, 11–19

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665.
Li, X.S., Trojer, P., Matsumura, T., Treisman, J.E., and Tanese, N. (2010). Mammalian SWI/SNF-A Subunit BAF250/ARID1 Is an E3 Ubiquitin Ligase That Targets Histone H2B. Mol. Cell. Biol. *30*, 1673–1688.

Lee, L. H., Sadot, E., Ivelja, S., Vakiani, E., Hechtman, J. F., Sevinsky, C. J., Shia, J. (2016). ARID1A expression in early stage colorectal adenocarcinoma: an exploration of its prognostic significance. Human Patholog *57*, 97-104.

Lengauer, C., Kinzler, K.W., Vogelstein, B.(1997). Genetic instability in colorectal cancers. Nature *386*, 623–627

Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Staahl, B. T., Wu, H., Aebersold, R., Graef, I. A., and Crabtree, G. R. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. Neur*on 55*, 201-215.

Lettice, L. A., Heany, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Elgar, G., Hill, R.E., de Graaff, E. (2003)A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.

Li, M., Zhao, H., Zhang, X., Wood, L.D., Anders, R.A., Choti, M.A., et al. (2011) Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. Nat Genet. *43*, 828–9.

Li, X.S., Trojer, P., Matsumura, T., Treisman, J.E., Tanese, N. (2010). Mammalian SWI/SNF-A subunit BAF250/ARID1 is an E3 ubiquitin ligase that targets histone H2B. Mol. Cell Biol. *30* 1673–1688.

Lickert, H., Takeuchi, J.K., Von Both, I., Walls, J.R., McAuliffe, F., Adamson, S.L., Henkelman, R.M., Wrana, J.L., Rossant, J., Bruneau, B.G. (2004). Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. Nature *432*, 107–112.

Liu, Y., Peng, J., Sun, T., Li, N., Zhang, L., Ren, J., Yuan, H., Kan, S., Pan, Q., Li, X., Ding, Y., Jiang, M., Cong, X., Tan, M., Ma, Y., Fu, D., Cai, S., Xiao, Y., Wang, X.,Qin, J. (2017). Epithelial EZH2 serves as an epigenetic determinant in experimental colitis by inhibiting TNFα-mediated inflammation and apoptosis. Proceedings of the National Academy of Sciences of the United States of America, *114*, 19.

Livshits, G., Alonso-Curbelo, D. Morris, J.P. 4th, Koche, R., Saborowski, M., Wilkinson, J.E., Lowe, S.W. (2018). Arid1a restrains Kras-dependent changes in acinar cell identity. Elife 7.

Loomes, K.M., Underkoffler, L.A., Morabito, J., Gottlieb, S., Piccoli, D.A., Spinner, N.B., Baldwin, H.S., Oakey, R.J. (1999) The expression of Jagged1 in the developing mammalian heart correlates with cardiovascular disease in Alagille syndrome. Hum. Mol. Genet. *8*, 2443–2449.

Madrigal, P., and Alasoo, K. (2018). AP-1 Takes Centre Stage in Enhancer Chromatin Dynamics. Trends Cell Biol. *28*, 509–511.

Maeda, D., Mao, T.L., Fukayama, M., et al. (2010) Clinicopathological Significance of Loss of ARID1A Immunoreactivity in Ovarian Clear Cell Carcinoma. International Journal of Molecular Sciences *11*, 5120–5128.

Mamo, A., Cavallone, L., Tuzmen, S., et al. (2012). An integrated genomic approach identifies ARID1A as a candidate tumor-suppressor gene in breast cancer. Oncogene *31*, 16, 2090–2100.

Marshall, N.F., Peng, J., Xie, Z., Price, D.H., and Chem, J.B. (1996). Control of RNA Polymerase II Elongation Potential by a Novel Carboxyl-terminal Domain Kinase. JBC *271*, 27176–27183.

Mathur, R. (2018). ARID1A loss in cancer: Towards a mechanistic understanding. Pharmacol. Ther. *190*, 15–23.

Mathur, R., Alver, B.H., San Roman, A.K., Wilson, B.G., Wang, X., Agoston, A.T., Park, P.J., Shivdasani, R.A., and Roberts, C.W.M. (2017). ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. Nat. Genet. *49*, 296–302.

Mathur, R., and Roberts, C.W.M. (2018). SWI/SNF (BAF) Complexes: Guardians of the Epigenome. Annu. Rev. Cancer Biol. *2*, 413–427.

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature Biotechnology, *28*, 5, 495-501.

Meijer, A., Kruyt, F.A., van der Zee, A.G., Hollema, H., Le P., ten Hoor, K.A., et al. (2013). Nutlin-3 preferentially sensitises wild-type p53-expressing cancer cells to DR5-selective TRAIL over rhTRAIL. Br J Cancer. *109,* 10, 2685–95.

Merlo, A.,Herman, J.G., Mao, L., Lee, D.J., Gabrielson, E., et al. (1995). 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat.Med. *1*, 686–92

Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., et al. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature *437,* 436–39.

Middeljans, E., Wan, X., Jansen, P.W., Sharma, V., Stunnenberg, H.G., Logie, C., (2012). SS18 together with animal-specific factors defines human BAF-type SWI/SNF complexes. PLoS One *7*.

Miller, R.E., Brough, R., Bajrami, I., Williamson, C.T., McDade, S., Campbell, J., Kigozi, A., Rafiq, R., Pemberton, H., Natrajan, R., et al. (2016). Synthetic Lethal Targeting of ARID1A-Mutant Ovarian Clear Cell Tumors with Dasatinib. Mol. Cancer Ther. *15*, 1472–1484.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.H., Sen, S., Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science *303*, 343–348.

Morkel, M., Riemer, P., Bläker, H., & Sers, C. (2015). Similar but different: distinct roles for KRAS and BRAF oncogenes in colorectal cancer development and therapy resistance. Oncotarget *6*,25, 20785-800.

Morrow, J. J., Bayles, I., Funnell, A., Miller, T. E., Saiakhova, A., Lizardo, M. M., Bartels, C. F., Kapteijn, M. Y., Hung, S., Mendoza, A., Dhillon, G., Chee, D. R., Myers, J. T., Allen, F., Gambarotti, M., Righi, A., DiFeo, A., Rubin, B. P., Huang, A. Y., Meltzer, P. S., Helman, L. J., Picci, P., Versteeg, H. H., Stamatoyannopoulos, J. A., Khanna, C., Scacheri, P. C. (2018). Positively selected enhancer elements endow osteosarcoma cells with metastatic competence. Nature Medicine *24*, 2, 176-185.

Najafova, Z., Tirado-Magallanes, R., Subramaniam, M., Hossan, T., Schmidt, G., Nagarajan, S., Baumgart, S.J., Mishra, V.K., Bedi, U., Hesse, E. et al. (2016). BRD4 localization to lineage-specific enhancers is associated with a distinct transcription factor repertoire. Nucleic Acids Res.

Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., Lee, C., Turner, E. H., Smith, J. D., Rieder, M. J., Yoshiura, K., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D. A., Bamshad, M. J., … Shendure, J. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature genetics, 42, 9, 790-3.

Neigeborn, L., Carlson, M. (1984). Genes affecting the regulation of *SUC2* gene expression by glucose repression in *Saccharomyces cerevisiae*. Genetics *108*, 4, 845–58.

Nishigaki, M., Aoyagi, K., Danjoh, I., Fukaya, M., Yanagihara, K., et al. (2005). Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays. Cancer Res. *65*, 2115–24

Ogiwara, H., Ui, A., Satoh, H., Yokomi, I., Yokota, J., Kohno, T. (2011). Histone acetylation by CBP and p300 at double-strand break sites facilitates SWI/SNF chromatin remodeling and the recruitment of non-homologous end joining factors. Oncogene *30*, 2135–2146.

Okugawa, Y., Grady, W. M., & Goel, A. (2015). Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. Gastroenterology *149*, 5, 1204–1225.

Ou, H.D., Phan, S., Deernick, T.J., Thor, A., Elisman, M.H., O'Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science *357*, 6349.

Park, J.H., Lee, C., Suh, J.H., et al. (2015) Decreased ARID1A expression correlates with poor prognosis of clear cell renal cell carcinoma. Hum Pathol., 46, 3, 454–460.

Pasqualucci, L., Dominguez-Sola, D., Chiarenza, A., Fabbri, G., Grunn, A., et al. (2011). Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature *471*, 189–95

Peterlin, B.M., and Price, D.H. (2006). Controlling the Elongation Phase of Transcription with P-TEFb. Mol. Cell *23*, 297–305.

Phelan, M.L., Sif, S., Narlikar, G.J., Kingston, R.E. (1999). Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. Mol. Cell *3*, 2, 247–53.

Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: Players and mechanisms. Nat. Rev. Mol. Cell Biol. *16*, 245–257.

Raab, J. R., Resnick, S., & Magnuson, T. (2015). Genome-Wide Transcriptional Regulation Mediated by Biochemically Distinct SWI/SNF Complexes. PLoS Genetics *11*, 12, 1–26.

Rahman, S., Sowa, M. E., Ottinger, M., Smith, J. A., Shi, Y., Harper, J. W., & Howley, P. M. (2011). The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. Molecular and cellular biology *31*, 13, 2641-52.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols *8*,11, 2281-2308.

Rao, S.S.P., Huang, S.C., Hilaire, B.G.S., Engreitz, J.M., Perez, E.M., KiefferKwon K.R., Sanborn A.L., Johnstone S.E., Bascom G.D., Bochkov I.D., et al. (2017). Cohesin loss eliminates all loop domains. Cell *171*, 305–309.

Rao, R.C., and Dou, Y. (2015). Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. Nature Reviews. Cancer *15*, 6, 334-46.

Reinhold, W. C., Sunshine, M., Liu, H., Varma, S., Kohn, K. W., Morris, J., Pommier, Y. (2012). CellMiner: A web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. Cancer Research *72,* 14, 3499–3511.

Reya, T., Clevers, H. (2005). Wnt signaling in stem cells and cancer. Nature *434*, 7035, 843-50.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., Klug, A. (1984).Stucture of the nucleosome core particle at 7 Å resolution. Nature *311*, 532-537.

Richon, V.M., Sandhoff, T.W., Rifkind, R.A., Marks, P.A. (2000). Histone deacetylase inhibitor selectively induces p21WAF1 expression and gene-associated histone acetylation. PNAS *97*, 10014–19.

Roberts, C.W., Galusha, S.A., McMenamin, M.E., Fletcher, C.D., Orkin, S.H. (2000). Haploinsufficiency of *Snf5* (integrase interactor 1) predisposes to malignant rhabdoid tumors in mice. PNAS *97*,25, 13796–800.

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J. Biol. Chem. *273*, 5858–5868.

Rokavec, M., Öner, M. G., Li H., Jackstadt, R., Jiang, L., Lodygin, D., Kaller, M., Horst, D., Ziegler, P. K., Schwitalla, S. et al. (2014). IL-6R/STAT3/miR-34a feedback loop promotes EMT-mediated colorectal cancer invasion and metastasis. J. Clin. Invest. *124*, 1853-1867.

Rorke, L.B., Packer, R.J., Biegel, J.A. (1996) Central nervous system atypical teratoid/rhabdoid tumors of infancy and childhood: definition of an entity. J Neurosurg. *85,* 56–65.

Saijyo, S., Kudo, T., Suzuki, M., Katayose, Y., Shinoda, M., et al. (1995) Establishment of a new extrahepatic bile duct carcinoma cell line, TFK-1. Tohoku J Exp Med. *177*, 61-71.

Samartzis, E.P., Gutsche, K., Dedes, K.J., Fink, D., Stucki, M., and Imesch, P. (2014). Loss of ARID1A expression sensitizes cancer cells to PI3K- and AKT-inhibition. Oncotarget *5*, 5295–5303.

Scherdin, U G.M., Klouche, M. (1987) In vitro interaction of a difluoromethyl-ornithine (DFMO) and human recombinant interferon-alpha (rIFN-alpha) on human cancer cell lines. Immunobiology *175* 1-143.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T.W., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A.K., Tönjes, M., et al. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226–231.

Shain, A.H., and Pollack, J.R. (2013). The Spectrum of SWI/SNF Mutations, Ubiquitous in Human Cancers. PLoS One *8*.

Sharrocks, A.D., Yang, S.H., Galanis, A.(2000). Docking domains and substratespecificity determination for MAP kinases, *Trends Biochem. Sci*. 25, 448–453.

Shema-yaacoby, E., Nikolov, M., Haj-yahya, M., Siman, P., Allemand, E., Yamaguchi, Y., Muchardt, C., Urlaub, H., Brik, A., Oren, M., et al. (2013). Resource Systematic Identification of Proteins Binding to Chromatin-Embedded Ubiquitylated H2B Reveals Recruitment of SWI / SNF to Regulate Transcription. Cell Reports *4*, 601–608.

Shen, J., Peng, Y., Wei, L., Zhang, W., Yang, L., Lan, L., Kapoor, P., Ju, Z., Mo, Q., Shih, I.M., et al. (2015). ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. Cancer Discov. *5*, 752–767.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: From properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272–286.

Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. Nat. Rev. Genet. *19*, 453–467.

Stern, M., Jensen, R., Herskowitz, I. (1984). Five SWI genes are required for expression of the HO gene in yeast. J. Mol. Biol. *178*, 4, 853–68.

Strahl, B.D., Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Sun, X., Chuang, J.C., Kanchwala, M., Wu, L., Celen, C., Li, L., Liang, H., Zhang, S., Maples, T., Nguyen, L.H., Wang, S.C., Signer, R.A., Sorouri, M., Nassour, I., Liu, X., Xu, J., Wu, M., Zhao, Y., Kuo, Y.C., Wang, Z., et al. (2016). Suppression of the SWI/SNF component Arid1a promotes mammalian regeneration. Cell Stem Cell *18*, 456–466.

Sun, X., Wang, S. C., Wei, Y., Luo, X., Jia, Y., Li, L., et al. (2017). Arid1a Has Context-Dependent Oncogenic and Tumor Suppressor Functions in Liver Cancer. Cancer Cell, *32*, 5, 574–589.

Thoma, F., Koller, T., Klug, A., (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J. Cell Biol. *83*, 2, 403-427.

Tian, Y., Cohen, E. D., Morrisey, E. E. (2010). The importance of Wnt signaling in cardiovascular development. Pediatr. Cardiol. *31*, 342–348.

Torigoe, S. E., Urwin, D. L., Ishii, H., Smith, D. E. & Kadonaga, J. T. (2011). Identification of a rapidly formed nonnucleosomal histone-DNA intermediate that is converted into chromatin by ACF. Mol. Cell *43*, 638–648.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., Issa, J.P.J. (1999) CpG island methylator phenotype in colorectal cancer. PNAS *96*, 8681-6.

Trizzino, M., Barbieri, E., Petracovici, A., Licciulli, S., Zhang, R., and Gardini, A. (2018). The Tumor Suppressor ARID1A Controls Global Transcription via Pausing of RNA Polymerase II Article The Tumor Suppressor ARID1A Controls Global Transcription via Pausing of RNA Polymerase II. Cell Reports *23*, 3933–3945.

Tsurusaki,Y., Okamoto, Y., Matsumoto, N. et al. (2012) Mutations of the subunits of the SWI/SNF complex cause Coffin-Siris syndrome. Nature Genetics *44*, 376-378.

Uhlen, M, Zhang, C, Lee, S, Sjöstedt E, Fagerberg, L, Bidkhori, G, Benfeitas, R, Arif, M, Liu, Z, Edfors, F, Sanli, K, Feilitzen, von K, Oksvold, P, Lundberg, E, Hober, S, Nilsson, P, Mattsson, J, Schwenk, JM, Brunnström, H, Glimelius, B, Sjöblom, T, Edqvist, P-H, Djureinovic, D, Micke, P, Lindskog, C, Mardinoglu, A, Ponten, F.(2017). A pathology atlas of the human cancer transcriptome. Science, *357*

Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., et al. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature *419*, 624–29.

Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C.K., Stephens, P., et al. (2011) Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature.*469*, 539–42.

Vasileiou, G., Ekici, A. B., Uebe, S., Zweier, C., Hoyer, J., Engels H. (2015). Chromatin-remodeling-factor ARID1B represses Wnt/beta-catenin signaling. Am. J. Hum. Genet. *97*, 445–456.

Vial, E. (2003). Elevated ERK-MAP kinase activity protects the FOS family member FRA-1 against proteasomal degradation in colon carcinoma cells. J. Cell Sci. *116*, 4957–4963.

Vierbuchen, T., Ling, E., Cowley, C.J., Couch, C.H., et al. (2017). AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection Article AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. Mol. Cell *68*, 1067–1082.

Vinciguerra, M., Vivacqua, A., Fasanella, G., Gallo, Cuozzo, C., Morano, A., Maggiolini, M., Musti, A.M. (2004) Differential phosphorylation of c-Jun and JunD in response to the epidermal growth factor is determined by the structure of MAPK targeting sequences, J. Biol. Chem. *279*, 9634–9641.

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, Fuks, F. (2005). The Polycomb group protein EZH2 directly controls DNA methylation. Nature, *439*, 7078, 871–874.

von Figura, G., Fukuda, A., Roy, N., Liku, M.E., Morris Iv, J.P., Kim, G.E., Russ, H.A., Firpo, M.A., Mulvihill, S.J., Dawson, D.W., et al. (2014). The chromatin regulator Brg1 suppresses formation of intraductal papillary mucinous neoplasm and pancreatic ductal adenocarcinoma. Nat. Cell Biol. *16*, 255–267

Vousden, K.H., Prives, C. (2009). Blinded by light: the growing complexity of p53. Cell *137*, 413-431.

Wang, W., Côté, J., Xue, Y., Zhou, S., Khavari, P.A,. Biggar, S.R., Muchardt, C., Kalpana G.V., Goff, S.P., Yaniv, M., Workman, J.L., Crabtree, G.R. (1996). Purification and biochemical heterogeneity of mammalian SWI/SNF complex. EMBO J *15*, 19, 5370-5382.

Wang, X., Lee, R.S., Alver, B.H., Haswell, J.R., Wang, S., et al. (2017). SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. Nat. Genet. *49*, 289–95.

Wang, X., Werneck, M.B.F., Wilson, B.G., Kim, H.J., Kluk, M.J., Thom, C.S., Wischhusen, J.W., Evans, J.A., Jesneck, J.L., Nguyen, P., et al.(2011). TCR-dependent transformation of mature memory phenotype T cells in mice. J Clin Invest. *121*, 3834–3845.

Wang, X., Lee, R.S., Alver, B.H., Haswell, J.R., Wang, S., Mieczkowski, J., Drier, Y., Gillespie, S.M., Archer, T.C., Wu, J.N., et al. (2017). SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. Nat. Genet. *49*, 289–295.

Watson et al., (2013). Molecular Biology of the Gene, 7th edition, Cold Spring Harbor Laboratory Press

Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M. D. & Owen-Hughes, T. (2003). Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. Mol. Cell. Biol. *23,* 1935–1945.

Whitmarsh, A.J. (2007). Regulation of gene transcription by mitogen-activated protein kinase signaling pathways. Biochim. Biophys. Acta - Mol. Cell Res. *1773*, 1285–1298.

Whitmarsh, A.J., and Davis, R.J. (1996). Transcription factor AP-1 regulation by mitogen-activated protein kinase signal transduction pathways. J. Mol. Med. *74*, 589–607.

Wiegand, K.C., Shah, S.P., Al-Agha, O.M., Zhao, Y., Tse, K., Zeng, T., et al. (2010). ARID1A mutations in endometriosis-associated ovarian carcinomas. N Engl J Med. *363*, 1532–43.

Williamson, C.T., Miller, R., Pemberton, H.N., Jones S.E., Campbell, J., Konde, A., Badham, N., Rafiq, R., Brough, R., Gulati, A., Ryan, C.J., Francis, J., Vermulen, P.B., Reynolds, A.R., Reaper, P.M., Pollard, J.R., Ashworth, A., Lord, C.J.(2016). ATR inhibition as a synthetic lethal therapy for tumours deficient in ARID1A. Nature Commun. *13*, 7, 13837.

Wilson, B. G., Wang, X., Shen, X., McKenna, E. S., Lemieux, M. E., Cho, Y. J., Koellhoffer, E. C.,Pomeroy, S. L., Orkin, S. H., and Roberts, C. W. (2010). Epigenetic antagonism between polycomband SWI/SNF complexes during oncogenic transformation, Cancer Cell *18*, 316-328.

Wreggett, K.A., Hill, F., James, P.S., Hutchings, A., Butcher, G.W., Singh, P.B. (1994). A mammalian homologue of *Drosophila* heterochromatin protein 1 (HP1) is a component of constitutive heterochromatin. Cytogenet. Cell Genet *66*, 99-103.

Wu, C., Lyu J., Yang, E.J., Zhang, B., Shim J.S., (2018). AURKA-CDC25C axis to induce synthetic lethality in ARID1A-deficient colorectal cancer cells. Nat. Commun. *9*, 3212

Wu, J. N., & Roberts, C. W. M. (2013). ARID1A mutations in cancer: Another epigenetic tumor suppressor? Cancer Discovery, *3*, 1, 35–43.

Yamagishi, M., and Uchimaru, K. (2017). Targeting EZH2 in cancer therapy. Curr. Opin. Oncol. *29*, 375–381.

Yamaguchi, T., Kakefuda, R., Tajima, N., Sowa, Y., Sakai, T. (2011). Antitumor activities of JTP-74057 (GSK1120212), a novel MEK1/2 inhibitor, on colorectal cancer cell lines in vitro and in vivo. Int J Oncol. *39*, 1, 23-31

Yan, J., Diaz, J., Jiao, J., Wang, R., You, J. (2011). Perturbation of BRD4 protein function by BRD4-NUT protein abrogates cellular differentiation in NUT midline carcinoma. J. Biol. Chem. *286*, 27663–75.

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J. A., Thompson, I. R., Ramaswamy, S., Futreal, P. A., Haber, D. A., Stratton, M. R., Benes, C., McDermott, U., Garnett, M. J. (2012). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic acids research, *41*.

Yao, R., Jiang, H., Ma, Y., Wang, L., Wang, L., et al. (2014). PRMT7 induces epithelial-to-mesenchymal transition and promotes metastasis in breast cancer. Cancer Res. *74*, 5656–67

Yu, Y., Chen, Y., Kim B., Wang H., Zhao C., et al. (2013). Olig2 targets chromatin remodelers to enhancers to initiate oligodendrocyte differentiation. Cell *152*, 248–61

Zhai, Y., Kuick, R., Tipton, C., Wu, R., Sessine, M., Wang, Z., Baker, S.J., Fearon, E.R., and Cho, K.R. (2015). Arid1a inactivation in an Apc and Ptendefective mouse ovarian cancer model enhances epithelial differentiation and prolongs survival. J. Pathol. 238, 21–30.

Zhang, L., Wang, C., Yu, S., Jia, C., Yan, J., Lu, Z., & Chen, J. (2018). Loss of ARID1A Expression Correlates with Tumor Differentiation and Tumor Progression Stage in Pancreatic Ductal Adenocarcinoma. Technology in cancer research & treatment *17*.

Zirkel, A., Nikolic, M., Sofiadis, K., Mallm, J., Brackley, C.A., and Gothe, H. (2018). HMGB2 Loss upon Senescence Entry Disrupts Genomic Organization and Induces CTCF Clustering across Cell Types Resource HMGB2 Loss upon Senescence Entry Disrupts Genomic Organization and Induces CTCF Clustering across Cell Types. Mol. Cell *17*, 70, 730-744.

Acknowledgements

The last four years have been an immense learning experience – scientifically and beyond. I came partially prepared with experiences that I had in the 21 years before I got to Göttingen. While I am very grateful to all the scientific input and encouragement I have got over the past four years, for me at least this PhD was much more than that. Biologically speaking our cells are living systems that function at high levels of accuracy, however beyond that definition what really makes us 'alive' is our interaction with human beings. That defines who we become and what we choose to do. And so, everything that have I achieved up to this point is only a reflection of the circumstances that I was fortunate to operate in and these circumstances were in a large part created by people. I want to use this opportunity to thank the many people who made it possible for me to do this. This will be long so please bear with me. I will live up to my reputation of telling the most detailed stories about people.

I have to begin by thanking my supervisor Prof. Steve Johnsen. His enthusiasm about science and research created a great learning environment in the lab. At the same time, his enthusiasm in general made the lab a friendly place to be in. I know this has not been an easy project, so I am grateful for his time and effort to come up with new ideas all the time and for three years of continuous financial support. His generosity extended beyond the typical student-mentor relationship. He made sure that we participated in relevant conferences besides involving us in grant application and paper review processes. This exposed me to the cutting edge of the field and developed in me a sense of critical thinking. I will always remember how he quietly slipped me a copy of 'Darm mit Charme', which is a humorous account of the on goings in our intestines, because I had once mentioned that I loved reading.

Prof. Matthias Dobbelstein, who was my mentor in the program gave me some very useful advice on how I should go about choosing a research field for my PhD when I started my master's here. I am very grateful for his interest and input at my thesis committee meetings and at joint seminars the third floor of the GZMB. I am also very thankful for his encouragement and support in the somewhat difficult last stretch of my PhD and I hope to continue interacting with him in the future.

Dr. Melina Schuh, a member of my thesis advisory committee, has been an inspiration as a very successful young woman scientist. I am very grateful to her for her input during the thesis committee meetings. I will always remember that despite being very busy she always carefully read my report before she came to these meetings.

I would like to thank the International Max Planck Research School and the Molecular Biology Program for financially supporting me during the beginning of my PhD. Moving to Germany which seemed daunting at first was made extremely easy because of the efforts of Steffen and Kerstin. They have helped me with any problem that I had through this time, from Kerstin translating every letter that I brought to her to Steffen carrying my bike to a repair shop when I crashed it into a bus during my first week in Germany. I have immensely enjoyed immersing myself into the German Christmas concert culture with Steffen. This is much more and beyond their excellent organization of the program which to say the least keeps it functioning at a highly efficient level. To this end, we have been fortunate to participate in some wonderful PhD seminars, retreats and career fairs organized brilliantly by them.

It is impossible to go through a PhD without an extremely supportive and cohesive work environment. This is not only very important for the many frustrations that PhD students go through but also for celebrating the smallest of joys. In this aspect, the Johnsen lab has been one of the best. I am lucky to have worked with many colleagues who have become friends. Added to that, was the group of Matthias Dobbelstein as neighbors, who gave us their scientific inputs, reagents (in desperate times) and a very friendly atmosphere to work in. I started my master's here with very limited lab experience and therefore I really have to thank all the members of the GZMB third floor for showing me something or the other. Everyone was very kind and willing to help whenever I asked.

As a rotation student, Robyn's project was my first exposure to experimental science. She was very patient as she taught me to do things that I had only read about. Later, she spared a lot of her time to conduct the mice experiments for my project. I am very thankful to her for not judging me when I was afraid of the mice (even though she absolutely loves animals) and instead patiently helping me to partially get over that fear. I enjoyed our many conversations at the mouse facility. I am also very grateful to her for having proofread my thesis with the greatest eye for detail and at very short notice. She has the reputation of catching the smallest mistakes for a reason. It was also her feedback about writing during my rotation that gave me an idea of how a scientific report should be written and so for that too I want to say thank you.

I have to thank Florian, who first worked with the generation of the *Arid1a* mouse model in our lab. Subsequently, he too was extremely patient with me when he helped me get over my fear of mice, taking time from his busy schedule to train with me. I am grateful to him for

his suggestions at lab meeting presentations and whenever we happened to talk about my project in the corridor. In addition, I just have mention the great music he always put on in the lab that made tedious experiments much more enjoyable.

I want to thank Vijaya, who was always helpful with her tricks in several techniques, Zeynab for having suggestions for every problem that we faced and her encouragement when things didn't work, Evan and Iga for the fun movie nights and fun conversations, Hannah for her sense of humor (and also the music), Jana for her constant support and her friendship and Oliver for the many BAF complex conversations, the lunches and his friendship.

I would like to thank the past members of the Johnsen group. Wanhua explained to me the details of a ChIP experiment, something I would struggle with over the next years. Anusha was always willing to help whenever I had a doubt in the early stages of my PhD. Thanks also to Simon for the interesting lunch conversations and his subsequent advice on life in industry and Sabine for helping with the mouse work as well as every German document I brought her.

On a special note I also want to thank the members of the 'small office'. This was my favourite place in the lab and the place where several extremely interesting scientific and non-scientific conversations took place. I want to thank Feda for the many experiments she showed me how to do in the beginning of my time in the lab even though it was not her responsibility and for her friendship. I want to thank Xin for teaching me the very basics of Bioinformatics and conducting some of the analysis himself and Ana, for offering to help me at every step of the way. Both Xin and Ana also provided very constructive feedback on my thesis. Finally, as Xin puts it Nicole is truly the 'best' Nicole. Her presence in our office just made it a more pleasant place. I am grateful to her prepping the plasmids that were used in this project, but beyond that for handling much of my German issues, for always being interested in our various cultures, hobbies and thus starting many exciting chats.

The rotation students who worked with me during this project, were also very helpful. Fereshteh and Shyam's experiments helped to further my project and they taught me how I could improve my teaching skills.

What I will take with the most happiness from Göttingen are the wonderful friends I made during my time here. In many direct and indirect ways, they contributed greatly to my PhD and wellbeing as a PhD student.

Jojo, with whom even seven-hour train rides are one never-ending conversation, lent me her ear, her hand and her head depending on what I needed when without question. She has been a major support through these few years and for that I will always be grateful. She

not only performed a few experiments for this project but constantly boosted the small office morale by her extremely positive presence. We went through similar experiences of dealing with adulthood and I was very fortunate to find someone on the same page with whom I could discuss many things ranging from science to careers to other aspects of life with. Figuring out together what our opinions on many matters are led to conversations that I hope will continue for a very long time. Of course, she also commented on this thesis after carefully reading it and understanding it even though it was not exactly her topic. It is needless to say that her friendship had an extremely positive effect on my life as a PhD student and in general. I am sure it will continue to do so.

My conversations with Lorenz began with science and the similar projects we worked on, but I think we have by now talked about everything under the sun. And somehow there's still more. From ranting about our initial struggles with being foreigners who did not understand German, to being angry about many subtle biases that we noticed in society and then feeling regret at our inability to articulate this properly. I am so thankful to him for being a great friend while he was in Göttingen (for every meal he cooked on bad experiment days) but more importantly for continuing to be a great friend even after all this time and change. He too has had a very positive effect on my life in these important years.

Cathie, Marija and Franzi have been solid support and just a lot of fun to hang out with over the last four years in Göttingen. Conversations with them about the about the responsibility of the scientific community to society and in the role of women in science have helped me develop a better sense of what it means to be part of this. I have learnt from each of them and they have been in their own ways inspiring women. I also want to thank Cathie for our weekly sports (but actually therapy) sessions. It was so good to find someone so relatable.

Xin is one of the most genuine and positive people I've ever met, something we all could do with a bit more of. Though he constantly teases me (I am no better), this has made working in the lab so much fun because it's like having a brother around. I want to thank him for his help with all my Bioinformatics and computer related issues but also for being a very good friend. At crucial moments he has done everything he could to help me. I cannot even count the number of times he (and Alice) have cooked for me when I have returned to Göttingen at odd hours. I will always remember how he did everything he could so that I wouldn't fall sick and could visit my immunocompromised aunt when we went to the AACR conference in Chicago. As my office neighbour, discussing science with him is so much fun too and when the science is too much it's always fun to go on our Google Earth travels.

I also want to thank Ana for the many fun times in and outside our office. For showing up at my flat with bags of frozen food when I was sick, for calling my internet provider about 20

times until my problem was solved, for her company and good nature in and outside the lab. I am very grateful to her offering to help me with experiments anytime she noticed a stressful situation and for giving me feedback on every presentation, email and report that I showed her.

Finally, Vivek has been a great friend and source of advice on many things from the time that he was here in Göttingen until now. I'm so glad that we have been in touch all this while and that we continue to be friends. I'm thankful to him for always looking out for me, his help with teaching me how to do some experiments and in general being a great person to talk to about my PhD and about many other things. I am very grateful for his always sound and thought-out advice.

Several things I learnt during my school and college days helped me during my PhD. The professors of the Biochemistry bachelor's program at Sri Venkateswara College, University of Delhi gave me some exceptional training to pursue a PhD in the life sciences. In particular, I would like to mention Dr. Nandita Narayanasamy whose lectures often started out technically and ended with philosophical debates about the working of the brain. She really seeded in me the curiosity to find out more about many topics. Dr. Latha Narayanan's lectures with the immaculate sketches she drew on the blackboard made me imagine for the first time, what the cell could look like.

I am grateful for the experiences I had at school. My interest in Biology started in a ninth-grade biology lesson explaining the partitioning of the four chambers of the human heart to separate oxygenated and deoxygenated blood. That was the day I decided that I had to continue to explore the functioning of living systems. Our school in its special way taught us to be humble, not take ourselves too seriously and to be collaborative and not competitive. I think these are three major traits that any scientist should possess, and I wish that it were more prevalent.

It was also at school that I made some of my closest friends. Ruchi and Megha who have since school pursed completely different lines of study, have been constants in my life since I was ten years old. Even though we speak once every few months, the detail to which they have followed my PhD and my life in Göttingen is amazing. In different worlds, we've grown up together, often going through very similar experiences. Both of them have given me some very useful advice at crucial times and I am very grateful for their friendship.

Varun and Rohini, who also went on to pursue their PhDs in science were great people to share PhD experiences and future plans with. Varun listened to and looked through many aspects of my project, even though it was probably not that interesting for him. He was one

of the people who was always excited to hear about ARID1A since the beginning and probably the only person outside the lab who knows me who also knows ARID1A.

Finally, I want to thank my family. I cannot even begin to write about them without being overwhelmed by the number of things that I have to thank them for but here is a try.

I wish my grandfather was still alive to see this day. He was so excited when I started my PhD and having learnt how to use email at a late age always joked that I should send him my PhD at grandfather@heaven.com if he was no longer around when it happened. I really wish I could do that because in some senses, he would be the happiest to see this. He continued to learn new things until the day he died, and I think we can all learn something from that.

My aunt and uncle Chandan Sen and Siddhartha Sen have been in touch with me in great detail over all these years both in terms of work and other things. They and my cousin Gautam (with whom my brother and I grew up) gave me a home away from home in England that I could visit anytime while I was in Germany which I did and enjoyed several times. They have always been interested in what was happening in Göttingen and gave some very useful advice about work and other things at crucial times. They will also probably the first people that I will practice my defense on.

My aunt and uncle June Hampson and Bhaskar Sengupta know every detail of my PhD and have been in close touch giving very helpful advice at important times. Even through chemotherapy my aunt knit me gloves and scarves to protect me from the cold German winter. They also very extensively proofread my thesis. My aunt caught with her eye for detail every single typing error and improved the structure of many sentences. My uncle had a look at it multiple times and gave me some very solid inputs on how to improve my scientific writing. Constructive feedback is such a useful thing and I am so grateful that I got so much from them. This is only a minor detail of how involved they have been in my life.

I have to thank Bina and Shubhadra didi, who despite not having the good fortune to receive education themselves; understood the importance of it and made sure that during exams at school and college, I focussed on studying.

Lastly, I have to thank my parents and brother. Even though I constantly tell my brother that he is stupid, I just go around copying him. He made my first CV and has gone through every single subsequent one telling me very accurately what the person I'm sending it to wants to see. It was great to have his extra seven years of experience in the scientific PhD/Postdoc life to know about anything that I wanted to ahead of time. Also, he has answered all the

stupidest computer, math and life related questions that I had for him at extremely odd hours of the day.

My parents, I've come to realize more and more are exceptions as far as parents go but in my humble opinion, what parents should be. They've been more friends though. I am extremely grateful to them for taking care of my education in the best possible way. Much more importantly though, I'm grateful to them for raising us as confident, honest and open-minded individuals, if I say so myself. From the beginning they made us aware of the biases that exist in society and how important it was never to misjudge anyone because of those prejudices. This has completely shaped how we relate to other human beings and broadened the range of interactions that are possible for us. I suspect that this is the reason I have seven-page acknowledgement section. In line with that, they have never let their own interests come in the way of anything we did, while giving us detailed and precise advice even for the tiniest issues in our lives. They had nothing to do with science, having worked with English Literature all their lives. This only exposed us to many different things and made home a very interesting place to grow up which it continues to be.

I have tried to keep this acknowledgement exhaustive, because I genuinely think it is very important to thank people for the positive roles they have played in my (academic) life, in any small measure. And so, if you, whoever you are, have read up until here, thank you very much for that interesting conversation, for the phone call, for that smile in the corridor, for that reassurance, for the help, for that piece of advice. Even the tiniest thing matters and can change someone's day and I am very grateful for all of you for giving me the circumstances that have helped me become who I am. I hope that that I will be able to use all that I've learnt to contribute in even a small measure meaningfully to science and society.