EXPRESSION OF ARID3B IN BREAST CANCER CELL LINES


A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY


BY


NİHAN ÖZDEMİRLER


IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY


DECEMBER 2014

Approval of the thesis:

**EXPRESSION OF ARID3B IN BREAST CANCER CELL LINES**

submitted by **NİHAN ÖZDEMİRLER** in partial fulfillment of the requirements for the degree of **Master of Science in Biology Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences**    _____

Prof. Dr. Orhan Adalı
Head of Department, **Biology**    _____

Assoc. Prof. Dr. A. Elif Erson-Bensan
Supervisor, **Biology Department, METU**    _____

**Examining Committee Members:**

Assoc. Prof. Dr. Mesut Muyan
Biology Department, METU    _____

Assoc. Prof. Dr. A. Elif Erson-Bensan
Biology Department, METU    _____

Assoc. Prof. Dr. Tülin Yanık
Biology Department, METU    _____

Assist. Prof. Dr. Mehmet Somel
Biology Department, METU    _____

Assist. Prof. Dr. Bala Gür Dedeoğlu
Institute of Biotechnology, Ankara University    _____

**Date:**          15.12.2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name:    NİHAN ÖZDEMİRLER

Signature            :

# ABSTRACT

EXPRESSION OF ARID3B IN BREAST CANCER CELL LINES

Özdemirler, Nihan

M.S., Department of Biology

Supervisor : Assoc. Prof. Dr. A. Elif Erson-Bensan

December 2014, 67 pages

The ARID, A-T Rich Interaction Domain, is a helix-turn-helix motif based domain which binds to DNA by both sequence-specific and non-sequence specific manners. The ARID protein family is found in protozoa, green algae, higher plants, fungi and metazoans, while it is not seen in archea or eubacteria. ARID domain harboring genes are involved in variety of biological processes including embryonic development, cell lineage gene regulation, chromatin modification and cell cycle control.

ARID3B, a member of ARID family, is highly conserved throughout evolution. It is homologous to retinoblastoma binding proteins, mouse Bright and Drosophila dead ringer protein. ARID3B has several roles in cellular processes; histone methylation, chromatin structure modification, transcriptional activation and possibly in development. Given its known roles, ARID3B has been linked to cellular transformation and tumor progression, and was shown to be deregulated in various different cancers. A possible link between ARID3B and breast cancer was suggested by our group. We reported miR-125b, a tumor suppressor in breast cancers to target and downregulate ARID3B mRNA. In this study, we aimed to contribute to the efforts to better understand ARID3B function in breast cancers by analyzing its expression in breast cancer cell lines. Our results confirmed the immunohistochemistry (IHC) studies performed in our lab on patient samples where a positive correlation between ARID3B expression and ER positivity was detected. Interestingly, we also detected a negative correlation between ARID3B and ERBB2 positivity, also confirming the IHC stud-

ies. Future studies will be required to mechanistically confirm these observations and their significance in terms of breast carcinogenesis.

Keywords: Breast cancer, ARID3B, ER$\alpha$, ERBB2

# ÖZ

ARID3B'NİN MEME KANSERİ HÜCRE HATLARINDA İFADE EDİLMESİ

Özdemirler, Nihan

Yüksek Lisans, Biyoloji Bölümü

Tez Yöneticisi    : Doç. Dr. A. Elif Erson-Bensan

Aralık 2014 , 67 sayfa

ARID, adenin ve timin bazlarınca zengin etki bölgesi, heliks sarmal yapıya sahip, DNA'ya sekansa bağlı ya da sekanstan bağımsız yollarla bağlanabilen bir protein bölgesidir. ARID ailesi, bazı protozoalarda, alglerde, yüksek bitkilerde, mantarlarda ve metazoanlarda bulunmasına karşın, arkelerde ve öbakterilerde görülmemektedir. ARID bölgesine sahip proteinler embriyonik gelişim, kromatin modifikasyonu ve hücre döngüsü kontrolü gibi çeşitli biyolojik süreçlerde görev almaktadır.

ARID ailesi arasında, ARID3B evrimsel süreçte büyük ölçüde korunmuştur. Retinoblastoma proteinine bağlanan proteinlerle, farede bulunan Bright ile meyve sineğinde bulunan dead ringer proteinlere homologtur. ARID3B; histon metilasyonu, kromatin yapı değişikliği, transkripsiyonel aktivasyon ve muhtemelen gelişim gibi çeşitli hücresel mekanizmalarda görev almaktadır. Bilinen görevlerinin yanında, ARID3B hücresel dönüşüm ve tümör gelişiminde rol oynadığı bilinmektedir ve çeşitli kanserlerde regülasyonunun değiştiği gösterilmiştir. ARID3B ve meme kanseri arasındaki muhtemel ilişki grubumuz tarafından daha önce gösterilmişti. Meme kanserinde tümor supresör olan mir-125b'nin ARID3B mRNA'sını hedeflediği ve regülasyonunu inhibe ettiğini göstermiştik. Bu çalışmada ise, ARID3B'nin meme kanserindeki rolünü, meme kanseri hücre hatlarındaki ekspresyonuna bakarak incelemeyi hedefledik. Laboratuvarımızda yürütülen başka bir çalışmada hasta örnekleri ile yapılan immunohistokimya sonucuna göre, ARID3B ve ER anlatımı arasında gösterilen pozitif iliş-

kiyi desteklemektedir. Bunun yanı sıra, immunohistokimya ile gösterilen ARID3B ve ERBB2 anlatımı arasındaki negatif ilişki, sonuçlarımızca desteklenmektedir. Bu gözlemleri mekanistik bir sonuca bağlayacak ve meme kanseri gelişiminde önemini gösterecek çalışmaların yapılması da gerekmektedir.

Anahtar Kelimeler: Meme kanseri, ARID3B, ER$\alpha$, ERBB2

*To my family*

# ACKNOWLEDGMENTS

# TABLE OF CONTENTS

xii

# LIST OF TABLES

# LIST OF FIGURES

FIGURES

# LIST OF ABBREVIATIONS

BLAST          Basic Local Alignment Search Tool

bp          Base Pairs

cDNA          Complementary Deoxyribonucleic Acid

DEPC          Diethyl Pyrocarbonate

DMSO          Dimethyl Sulfoxide

DNA          Deoxyribonucleic Acid

DNase I          Deoxyribonuclease I

dNTP          Deoxyribonucleotide Triphosphate

DTT          Dithiothreitol

EDTA          Ethylenediaminetetraacetic Acid

g          Centrifuge gravity force

GAPDH          Glyceraldehyde 3-Phosphate Dehydrogenase

NCBI          National Center for Biotechnology Information

miRNA          microRNA

mRNA          Messenger RNA

PCR          Polymerase Chain Reaction

PBS          Phosphate Buffered Saline

RISC          RNA-induced Silencing Complex

RNA          Ribonucleic Acid

RNase          Ribonuclease

rpm          Revolution Per Minute

RT-PCR          Reverse Transcription Polymerase Chain Reaction

TBE          Tris-Boric acid-EDTA

# CHAPTER 1

# INTRODUCTION

## 1.1  ARID Family

AT-rich interactive domain (ARID) containing proteins are highly conserved, DNA binding proteins with helix-turn-helix domain which binds to the major groove of the DNA [33]. ARID is named after identification of a novel protein highly similar to murine protein Bright and Drosophila gene product dead ringer (dri). Bright is a B-cell specific transactivator and generally binds AT-rich domains on DNA [11]. On the other hand, Dri protein was also found to bind AT-rich sequences on DNA [10]. Not only in mouse or *Drosophila melanogaster*, but also in other organisms including mammals, fungi and plants; ARID domain is seen with its several cellular functions such as differentiation, cell cycle control, cellular growth and embryonic development. ARID is about 100 amino acid long domain and present in 15 annotated human proteins. ARID domain containing proteins are classified in several subclasses and these subclasses are represented in relation to their sequence identities. 15 human ARID family proteins present which are shown in the Figure 1.1, including their alternate names and amino acid lengths [37]. In general, ARID family proteins can be classified as follows; ARID1, ARID2, ARID3, ARID4, ARID5, JARID1, JARID2.

ARID1 subfamily has two members, ARID1A and ARID1B. Both have orthologs in *Drosophila melanogaster* OSA and *S.cerevisiae* Swi1p. Both proteins are SWI/SNF like chromatin remodeling proteins. ARID1A and ARID1B are also components of SWI/SNF chromatin remodeling complexes in mammalians. ARID1A and ARID1B have 80% identical ARID region and 50% identical full length amino acid sequence.
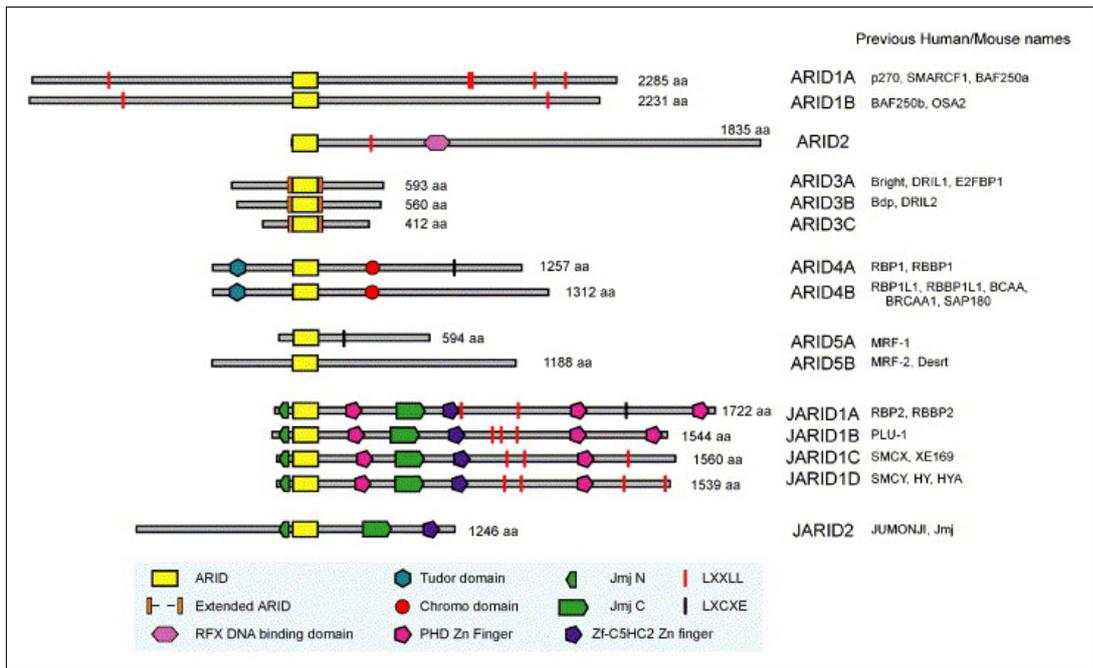
Figure 1.1: AT-rich interaction domain (ARID) containing human proteins. Database analyses reveal 15 different human proteins containing ARID with nearly 70% to 85% homology on their ARID sequences. [37]

ARID1A has several LXXLL motifs which are potential nuclear-hormone receptor-binding sites, as well as glutamine-rich regions, ARID1B does not possess these motifs or regions [35].

ARID2 subfamily has only one identified member which has an ortholog in *Drosophila* BAP170 (Brahma associated protein 170). ARID2 and BAP170 possess a DNA binding domain called RFX. RFX is named after a protein named *R*egulatory *F*actor *X*, binds X box of MHC class II genes to control their expression. In addition, BAP170 is a component of SWI/SNF-like complex, PBAP (Polybromo-associated Brm). The presence of BAP170 in the PBAP complex supports the evidence that ARID-containing subunits are crucial for SWI/SNF like chromatin-remodeling complexes [23]. It was also shown that human ARID2 may act as a new partner of cyclin A1-CDK1 complex using yeast triple-hybrid method, by the high expression level in testis [6].

ARID3 subfamily has three members named ARID3A, ARID3B and ARID3C. These three members have 80% amino acid sequence identity in their ARID domains. *Droso-*

2

*phila* ortholog for ARID3 subfamily is DRI. Together with conserved ARID domain; *Drosophila* ortholog DRI seems to have the highest similarity with ARID3 subfamily among other ARID subfamilies as having extended ARID region. N and C terminal sequences of ARID3 proteins form alpha helices which known as extended ARID (e-ARID). Extended ARID (e-ARID) spans about 40 amino acid residues of the ARID consensus sequence and it is specific for ARID3A, ARID3B and ARID3C [15]. The sequence of e-ARID can be identical up to 70% among both DRI and ARID3 proteins. The C terminus of e-ARID was shown to have DNA binding capability [27]. e-ARID also has a protein motif named REKLES, which is not found in any non-ARID proteins. REKLES has two subdomains named REKLES$\alpha$ and REKLES$\beta$. Both subdomains have important roles in nucleocytoplasmic shuttling of ARID3A. REKLES$\alpha$ is responsible for import of ARID3A while and REKLES$\beta$ is responsible for export of ARID3A. In addition to that, REKLES$\beta$ is responsible for binding of ARID3A to nuclear matrix association regions (MARs) [20].

ARID4 subfamily has two members, ARID4A and ARID4B. These members have 74% identity in their ARID domain and up to 40-50% sequence identity in their full length sequence. Both proteins contain Tudor domain and a chromodomain, while ARID4A has an additional LXCXE, pRB binding motif. Although the exact function of Tudor domain is not yet clear, it is present in RNA-binding proteins and also in oncogenes like Staphylococcal nuclease domain-containing protein 1 (SND1). Tudor domain is found in several proteins which colocalize with ribonucleoprotein or single stranded DNA-associated components within nucleus and the membrane of mitochondria and kinetochores [36]. Chromodomains are responsible for binding to methylated lysine residues and take role in transcriptional repressors. Accordingly, studies showed that ARID4A is repressor for E2F transcription through pRB. The ortholog of ARID4 family in *Drosophila* is CG7274 which contains ARID and Tudor domains [37].

ARID5 subfamily contains two members, ARID5A and ARID5B. They only have 70% identity in their ARID domains. Both proteins can bind to similar AT-rich sequences of transcriptional modulator of human cytomegalovirus major immediate-early promoter; as a result they repress transcription [14]. Interestingly, ARID5B knockout mice were reported to have lipid metabolism defects [22].

3

JARID1 subfamily has four members, JARID1A, JARID1B, JARID1C, JARID1D. These have 80% identity in their ARID domains as well PHD, zinc finger C5HC2 domains and LXXLL and Jmj motifs. On the other hand, JARID1A has LXCXE motif and has nuclear receptor-mediated transcription enhancer ability. The *Drosophila* ortholog is Little Imaginal Discs (LID) protein of JARID1 subfamily [4].

JARID2 subfamily has only one member. JARID1 and JARID2 have 25% identity in ARID domain, however the other regions have 80% identity. This lowest identity in ARID domain resulted JARID2 to be classified as a different subfamily. *Drosophila* ortholog is CG3654. As a result of knockout mice studies, JARID2 has important role in development of cardiac and muscle cells and also it represses the cyclin D1 in cardiac myocytes [36].

Previous studies have primarily showed that, ARID domain has preferential capability to bind to AT-rich sequences. By taking consideration of this difference, it was concluded that ARID subfamilies can be divided into two different categories according to their specificity to AT-rich sequences: AT-rich specific proteins and nonspecific proteins [37]. Among ARID subfamilies it was shown that only ARID3 and ARID5 subfamilies have AT-rich sequence binding capability, although the others have some degree of affinity but not sequence specificity. For example, ARID1 and ARID5 have similar contacts with DNA, but it was not sequence specific for ARID1 subfamily [15]. By taking consideration of this finding in the light of previous findings ARID3 and ARID5 are classified as AT-rich specific proteins, while ARID1A can be an example for AT-rich nonspecific proteins. In addition, according to nuclear magnetic resonance solution structure results, ARID3 subfamily has shown DNA major groove contacts through e-ARID portion of ARID domain. Figure 1.2 represents the structure of ARID in *Drosophila* ARID3-like, human ARID1A, human ARID5B and Yeast Swi1p [37].

As explained previously, ARID domain is a DNA binding domain with its $\alpha$-helix structure. The recognition helix; one helix of either helix-turn-helix or helix-loop-helix structure contacts with the major groove of DNA. According to the secondary structure prediction softwares ARID regions have at least six $\alpha$-helices for contacting with DNA. For two ARID domain containing proteins ARID3B (previous name, Dri)

4
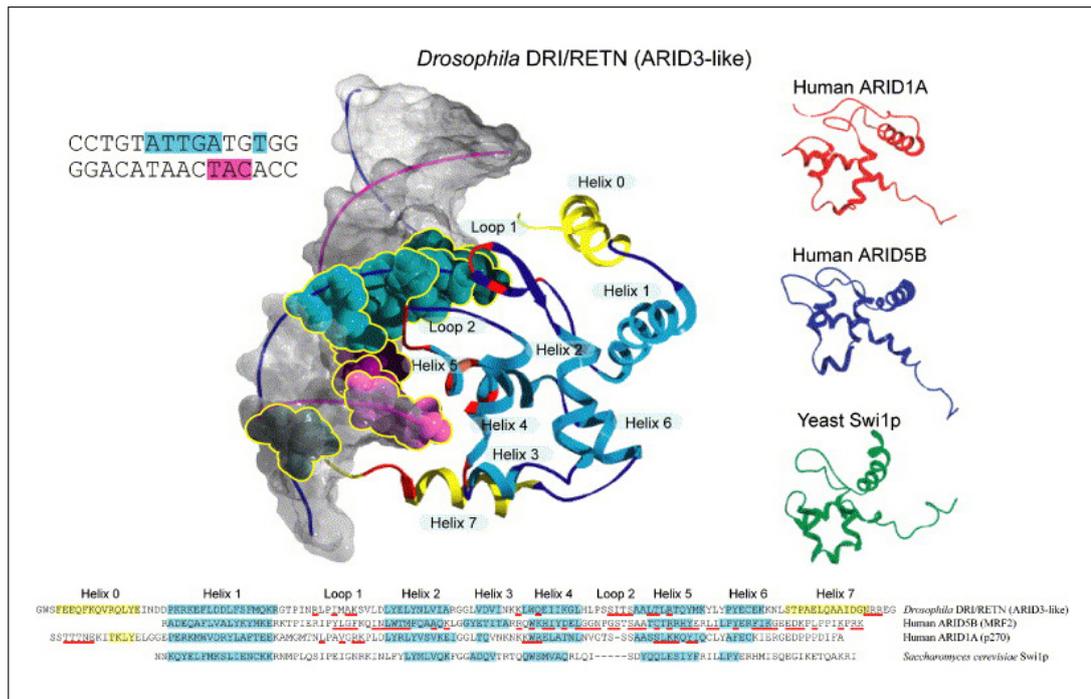
Figure 1.2: Drosophila ARID3-like, human ARID1A, human ARID5B, Yeast Swi1p NMR Solution Structures. Light Blue: ARID helices, Yellow: e-ARID, Dark Blue regions in ribbons: loops or turns, Dark blue arrows: beta sheets in DRI ARID, Red: DNA contacted-proteins, Blue/pink: nucleotides within DNA duplex make protein contact [37]

and ARID5B (previous name, MFR2) are shown in Figure 1.3. Both structures have common features while some distinct differences exist. ARID3B has seven helices with one $\beta$-sheet, on the other hand, ARID5B has six helices and a loop between two helices where $\beta$-sheet is located in Dri. The 3D structure of helices H2-H6 is the same however, there is no complete superimposition. DNA contact and recognition were predicted to be done through helix H5 and the preceding helices, the other residues are for minor groove and phosphate backbone contact [36].

According to recent studies, DNA binding activity of ARID domains can be regulated by other cellular processes. In undifferentiated Tera 2 (human embryonal carcinoma) and THP-1 (human acute monocytic leukemia) cells hCMV (human cytomegalovirus) enhancer is repressed. ARID5B was isolated as a repressor of hCMV enhancer in
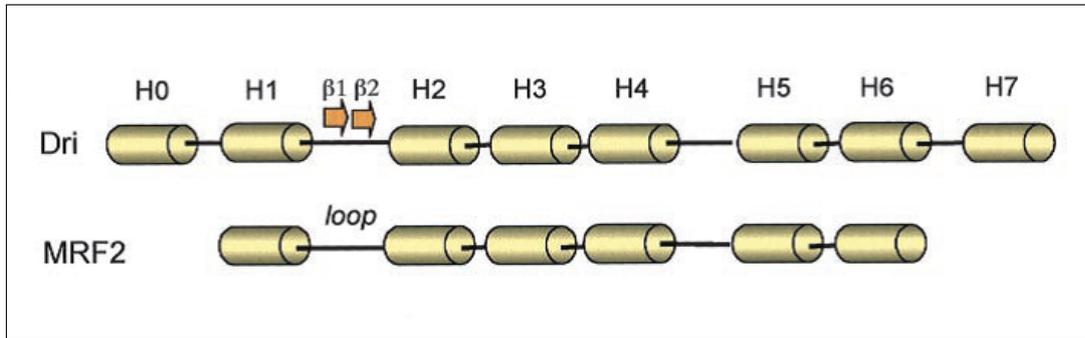
Figure 1.3: Helical structures of ARID domain representation in ARID3B (Dri) and ARID 5B (MRF2) proteins. Cylinders show helices and arrows show $\beta$-sheets in Dri. Similar loop positions detected for ARID5B (MRF2) [36]. "H" represents helical structures.

these cells. However, retinoic acid induced differentiation of Tera 2 and THP-1 resulted in reduced ARID5B DNA binding activity, so hCMV enhancer activation [36]. Thus, it can be concluded that ARID5B can work as either repressor or activator in different concepts.

The distribution of the ARID proteins within the tissues also varies. Table 1.1 summarizes the human ARID proteins according to their tissue distributions and amino acid lengths together with NCBI protein accession numbers [36, 37].

## 1.2 ARID3B and Its Isoforms

ARID3B is a member of the AT-rich interaction domain (ARID) family which is highly conserved throughout evolution. ARID3B was previously known as BDP (Bright-Like Protein) and maps to 15q24. A pseudogene for ARID3B also exists on chromosome 1p31 [25].

ARID3B has several roles in cellular processes; histone methylation, chromatin structure modification, transcriptional activation and involvement in development [36]. ARID3B was shown to be overexpressed in both mice neuroblastoma [18] and ovarian cancers [5]. Deletion of Arid3b in mice results in cardiovascular and craniofacial malformations, limb bud defects, poor angiogenesis and even embryonic lethality [18].

6

Table 1.1: ARID proteins and expression profiles.

| ARID protein | Previous name | Amino acid length | Accession numbers | Tissue distribution | Reference |
|---|---|---|---|---|---|
| ARID1A | p270 | 2285 aa | NP_006006.3 | Broad | [36] |
| ARID1B | pKIAA1235 | 1711 aa | NP_059989.1 | Broad | [36] |
| ARID2 | pKIAA1557 | 1835 aa | NP_689854.2 | Broad | [37] |
| ARID3A | Bright,DRIL1 | 593 aa | NP_005215.1 | Restricted, in mature B cells, not in T cells or immature B cells. In mouse tissues, expression in testis | [36] |
| ARID3B | Bdp,DRIL2 | 560 aa | NP_006456.1 | Broad, especially in placenta, testis, leukocytes | [36] |
| ARID3C | Bright-like | 412 aa | NP_001017363.1 | Broad | [37] |
| ARID4A | RBP1,RBBP1 | 1257 aa | NP_002883.2 | Broad | [36] |
| ARID4B | RBP1L1 | 1311 aa | NP_057458.4 | Restricted, In normal tissues expression present in testis, in cancer tissues expression in breast, ovary,lung,colon and pancreas | [36] |
| ARID5A | MRF1 | unknown | NP_997646.1 | Not reported | [36] |
| ARID5B | MRF2 | 1188 aa | NP_115575 | Broad | [36] |
| JARID1A | RBP2,PBBP2 | 1722 aa | NP_005047.1 | Broad | [36] |
| JARID1B | PLU-1 | 1544 aa | NP_006609.3 | Restricted, In normal tissues expression present in testis, up regulated in breast cancer tissues. | [36] |
| JARID1C | SMCX | 1560 aa | NP_004178.1 | Broad | [36] |
| JARID1D | SMCY | 1538 aa | NP_004644.2 | Broad, specific for males | [36] |
| JARID2 | jumonji | 1266 aa | NP_004964.2 | Restricted, brain, heart, skeletal muscle, kidney and thymus | [36] |

ARID3B gene transcribes a 4253 bp mRNA which has 9 exonic and 8 intronic regions. ARID3B protein has 560 amino acid, its molecular weight is 61 kDa and exclusively located in nucleus. It was shown that the localization of 61 kDa (full length) protein and 28 kDa isoform may vary; the former is mainly present in nucleus while the latter is located in nucleus as well but once overexpressed its localization mainly shifts to cytoplasm and mitochondria. According to Cowden et al. [17], this was explained by the lack of nuclear localization signal of 28 kDa isoform and its requirement for another protein for translocation to nucleus. As a result of the overexpression of this isoform, the concentration of "chaperon" protein was predicted to be limited for nuclear transportation. It was also experimentally shown that; 28 kDa protein has only 4th exon and some specific part from intron 4. ARID3B can also be seen in cytosol and plasma membrane as well as nuclear in different cells like other ARID subfamily members such as ARID3A and ARID3C once sumoylated [17]. Although ARID3B does not have the same region as ARID3A and ARID3C for sumoylation, the mechanistic explanation for this pattern remains to be eluciated [17].

## 1.3   ARID3B in Cancer

Several studies revealed altered expression of ARID proteins in different cancers. For example, JARID1B is upregulated in breast tumor cells, while poorly expressed in colon cancer cell lines [36]. ARID4B expression is very high in breast, ovary, colon and pancreatic carcinomas [36]. It is known that, SWI/SNF complexes as well as ARID1A are commonly deregulated in tumor cells. In cervical carcinoma and breast cancer cell lines the expression of ARID1A is dramatically decreased. ARID3A also binds to E2F transcription factor and it rescues Ras-induced senesence in primary murine fibroblasts to give them oncogenic property [36].

ARID3B is expressed during development and loss of ARID3B results in embryonic lethality [25]. ARID3B is known to bind to hyphophosphorylated pRB which supports its importance in cell cycle regulation [25] Hence, it is not surprising that ARID3B expression is found to be deregulated in cancers. For example, ARID3B expression and neuroblastoma progression are highly correlated. ARID3B with its oncogenic activity transforms MEFs in combination with MYCN [18]. Interest-

8

ingly, it was also noted that in neural cells, amyloid precursor intracellular domain (AICD) induces ARID3B transcription. This is important because of the fact that in Alzheimer disease, amyloid precursor protein is cleaved to amyloid precursor intracellular domain (AICD) and $\beta$-amyloid. The relationship between induction of ARID3B through amyloid precursor intracellular domain (AICD) in Alzheimer disease and ARID3B protein levels remain to be elucidated [24].

Another study confirmed with the knockout mice experiments that, ARID3B protein expression is higher in fetal brain compared to adult brain, which enhances the understanding of the importance of the involvement of ARID3B in neuronal development. Within the same study, it was also shown that ARID3B protein expression is high in stratified squamous epithelium, while it is lower in basal cells. The authors speculated that ARID3B may act as tumor suppressor by promoting differentiation as expression is changed according to differentiation state [31].

On the other hand, in normal and malignant tissues of esophagus and stomach; there is negative correlation between ARID3B expression levels and cancer progression. It was also speculated by the authors that, ARID3B has dual roles; acts as either tumor suppressor or oncogene, depending the concentration of ARID3B and target gene [31].

Another recent study used ovarian cancer cell lines as a model and showed that, ARID3B induces apoptosis once overexpressed by inducing pro-apoptotic genes such as Caspase 10, Caspase 7 and BIM while inhibiting pro-survival genes Bcl-2, Bcl-xL and XIAP. In addition to that, the same study demonstrated through a transcription factor, polyomavirus enhancer activator 3 protein (PEA3) binding to ARID3B promoter, EGFR modestly activates ARID3B expression [17].

During morphogenesis of multi-cellular organisms epithelial-to-mesenchymal transition (EMT) is a crucial process and it is required for normal developmental processes. However, during tumor invasion and metastasis, it has been known that EMT has several important roles [3]. There are limited studies to support the role of ARID3B in EMT. It was shown that *Arid3b* is needed for mouse embryonic mesenchymal cells development. *Arid3b* null mice were early embryonic lethal and phenotypes of these mutants have three common characteristics as follows; neural tube defects, small

branchial arches and cardiovascular system defects. According to these results, it was concluded that the reason of embryonic lethality is the increased apoptosis in cranial mesenchymal cells in *Arid3b* null mutants [32].

ARID3B has also been shown to be regulated by microRNAs. miR-125a and miR-125b belong to the same family of microRNAs with identical seed sequence. miR-125a targets ARID3B in ovarian cancer cell lines; OVCA433 and DOV13. Overexpression of miR-125a resulted decrease in protein levels of ARID3B and also it was shown that it induced mesenchymal-to-epithelial transition. On the other hand, if miR-125a is transcriptionally repressed; ARID3B promotes mesenchymal phenotype and disease progression is favored [5].

Using breast cancer cell lines MCF7 and MCF10A as model it was also shown that miR-125b targets ARID3B in breast cancer cells. There is negative correlation in expression of miR-125b and ARID3B and by supporting phenotypical changes; ARID3B may be a possible regulator of cell motility [1].

## 1.4 Aim of the Study

As an evolutionarily conserved protein, ARID3B was shown to be targeted by miR-125b in our laboratory [1]. Given that miR-125b is predicted to function as a tumor suppressor in breast cancer, deregulated ARID3B expression is also of interest to better understand breast tumorigenesis.

Therefore in this study, we aimed to characterize ARID3B expression in breast cancer cells to contribute to the understanding of how ARID3B may have role in breast cancer.

# CHAPTER 2

# MATERIALS AND METHODS

## 2.1 ARID3B Coding Sequence Cloning

ARID3B nucleotide sequence was retrieved from National Center for Biotechnology Information (NCBI), with the accession number; NM_006465.2. 1683 bp full coding sequence was determined to be amplified and cloned into pcDNA 3.1 (-) (Life Technologies). pcDNA 3.1 (-) vector map is given in Appendix Figure D.1. By selecting the restriction enzyme sites, which were also present in the vector to-be-used, the primer pair was designed. For the forward primer, Kozak sequence, was also added to improve expression levels. The designed forward and reverse primers are shown in Figure 2.1.

Forward: 5' ATAAGAAT   GCGGCCGC   AAAATGG   GAGCCACTTCAG   3'
                        *NotI Restriction*   *Kozak seq.*   *Primer*

Reverse: 5'   GC   GGATCC   TCAGAGGGACCAGCTGGTG   3'
                    *BamHI Restriction*   *Primer*
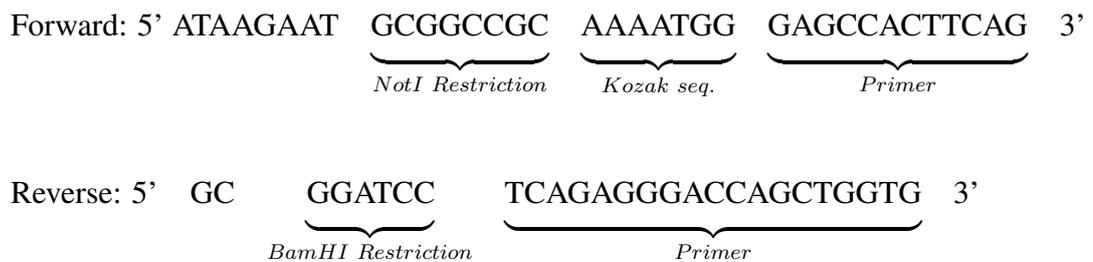
Figure 2.1: Forward and Reverse Primers for ARID3B Coding Sequence

## 2.2 Mammalian Cell Culture Conditions and Treatments

MCF10A was obtained from ATCC (LGC Standards GmbH, Germany). MDA-MB-231, MDA-66 (with permission from Dr. F. Gannon) [2] and MCF7 cell lines were a

kind gift from Dr. U. H. Tazebay (Gebze Institute of Technology, Kocaeli, Turkey). SKBR3 and MDA-MB-361 cell lines were a kind gift from Dr. I. Yulug (Bilkent University, Ankara, Turkey). Non-tumorigenic immortal MCF10A cells were grown in DMEM/F12 medium with 5% horse serum, 100 mg/ml Epidermal Growth Factor (EGF), 1 mg/ml hydrocortisone, 1 mg/ml cholera toxin, and 10 mg/ml insulin. MDA-MB-231, MDA-MB-361 and MCF7 cells were grown in DMEM with 10% Fetal Bovine Serum (FBS). MDA-66 cells were maintained in MDA-MB-231 medium with 0.4 mg/mL Hygromycin (Roche Applied Sciences, Germany). SKBR3 cells were grown in MDA-MB-231 medium with 1% non-essential amino-acids. All media contained 1% Penicillin/Streptomycin. All cell lines were grown as monolayers and were incubated at $37\,°C$ with 95% humidified air and 5% $CO_2$.

AG825 (Tocris Bioscience) was used to inhibit ERBB2. SKBR3 and MDA-MB-361 cells were seeded (nearly 100,00 cells) 24 hours prior to treatment in 6-well plates. After 24h, 50 $\mu$M AG825 final concentration or DMSO were added to the cells in 2 mL medium.

## 2.3   RNA Isolation and complementary DNA Synthesis

Cells were grown in T75 cell culture flasks to 70% confluency. Growing medium was sucked off from the flask and 8 mL of Trizol reagent (Guanidium thiocyanate) from Invitrogen (Cat# 15596-018) was used to lyse the cells. For complete dissociation of nucleoprotein complexes, cells were incubated at room temperature for 5 minutes and lysates were transferred to 15 mL sterile tubes. 1.6 mL chloroform was added to the sterile tubes and shaken for 15 seconds by hand. After 3 minutes incubation at room temperature, samples were centrifuged at 4700 g for 20 minutes at $8\,°C$. After centrifugation, phases were separated and RNA remained in the aqueous phase with an approximate volume of 5 mL. To precipitate RNA, 4 mL of isopropanol was added to each 5 mL aqueous phase containing RNA. The samples were again incubated at RT for 15 minutes and centrifuged at 4700 g at $4\,°C$ for 20 minutes. As a result, RNA formed a gel-like pellet at the bottom of tube. Supernatant was removed and RNA pellet was washed by 75% ethanol by gentle vortex. To remove ethanol, samples were centrifuged again at 4700 g for 7 minutes at $4\,°C$. Without disturbing the RNA

pellet, ethanol was removed from the tube and samples were allowed to dry at room temperature for 10 minutes. RNA pellet was suspended in 20-50 $\mu$L RNase-free water, and quantified using NanoDrop 2000c (Thermo, USA) to determine purity and quantity of RNA. RNA was stored at $-80\,°$C for future use.

To continue with complementary DNA (cDNA) synthesis DNase I treatment was performed. All isolated RNAs were treated with Deoxyribonuclease I (DNase I) from Fermentas (Cat # EN05-21) in order to obtain DNA-free RNA. Reaction mix is listed in Table 2.1.

Table 2.1: DNase I reaction set-up (Fermentas)

| RNA (1 $\mu$g/$\mu$L) | 10 $\mu$L |
|---|---|
| 10x Reaction Buffer | 10 $\mu$L |
| DNase I (1 u/$\mu$L) | 10 $\mu$L |
| Ribonuclease Inhibitor (40 u/$\mu$L) | 5 $\mu$L |
| DEPC-dH2O | 65 $\mu$L |
| Total Volume | 100 $\mu$L |

Dnase treatment was done at $37\,°$C in water bath for 60 minutes. After while, equal amount of Phenol: Chloroform: Isoamyl Alcohol (25: 24: 1) was added to stop the reaction (100 $\mu$L). Tube was vortexed for 30 seconds and incubated on ice for 10 minutes. Samples were centrifuged at 14000 g for 20 minutes at $4\,°$C. The upper phase containing RNA was taken into a fresh tube. According to volume of upper phase, 3X volume 100% ice cold ethanol and 1/10X volume 3M NaAc was added to samples and samples were incubated at $-20\,°$C overnight to precipitate RNA. After overnight incubation, samples were centrifuged for 30 minutes at 14000 g at $4\,°$C. After discarding the supernatant, samples were washed with 70% cold ethanol and again centrifuged for 10 minutes at 14000 g at $4\,°$C. Finally, pellet was dissolved in 20-50 $\mu$L RNase-free dH$_2$O. RNA was quantified by NanoDrop 2000c (Thermo, USA) to determine purity and quantity. Lack of DNA in RNA samples was further confirmed by PCR using housekeeping GAPDH primers. Next, RevertAid First Strand cDNA Synthesis Kit from Fermentas (Cat# K1632) was used for first strand cDNA synthesis. Random hexamer primers were used in the reaction. The reaction mix is listed in Table 2.2.

Table 2.2: cDNA synthesis reaction set-up (Fermentas)

| RNA | 1 $\mu$g (1-2 $\mu$L) |
|---|---|
| **Primer (random hexamer or oligodT)** | 1 $\mu$L |
| **dNTP** | 2 $\mu$L |
| **Nuclease-free water** | Adjust according to RNA amount |
| **Total Volume** | 12 $\mu$L |
| Briefly centrifuge, incubate at 70 °C for 5 minutes, chill on ice. | |
| **5X Reaction Buffer** | 4 $\mu$L |
| **Ribolock RNase inhibitor** | 1 $\mu$L |
| Briefly centrifuge, incubate 37 °C for 5 minutes for oligodT and 25 °C for 5 minutes for random hexamer | |
| **RevertAid RT Enzyme** | 1 $\mu$L |
| **Total Volume** | 20 $\mu$L |
| Incubate at 25 °C for 10 minutes for oligodT; 42 °C for 60 minutes for random hexamer, stop the reaction at 70 °C for 10 minutes and chill on ice. | |

## 2.4 Polymerase Chain Reaction Conditions using MCF7 cDNA for ARID3B Coding Sequence Primers

Using forward and reverse cloning primers PCR reaction was set-up using Pfu polymerase as described in Table 2.3 and Table 2.4. 50 $\mu$L of the final PCR products of the reaction were electrophoresed on a 2% agarose gel at 100V and photographed under UV light.

1683 bp product (full length ARID3B) was cut out from the agarose gel and the interested amplicon was extracted using Zymoclean Gel DNA Recovery Kit, Zymo Research. DNA fragment was extracted from the gel using blade and placed in 1.5 mL microcentrifuge tube. 3 volumes of Agarose Dissolving Buffer (ADB) was added to each volume of agarose. The mixture was incubated at 55 °C for 10 minutes using heat block. The melted solution was transferred to Zymo-Spin™ Column in a Collection Tube. The tube was centrifuged at 14,000 g for 60 seconds. The flow-through was discarded. 200 $\mu$L DNA Wash Buffer added to the column and centrifuged at 14,000 g for 60 seconds, this step was performed twice. 8 $\mu$L RNase-free water was added to the column, with clean collection tube and centrifuged at 14,000 g for 60 seconds to collect DNA. The quantification was performed using NanoDrop 2000c (Thermo, USA) to determine purity and quantity of DNA.

14

Table 2.3: Polymerase Chain Reaction mixture to amplify coding sequence of ARID3B

| 1X Master Mix | |
|---|---|
| 10X Buffer without MgSO$_4$ | 5 $\mu$L |
| dNTP | 5 $\mu$L |
| Forward primer | 10 $\mu$L |
| Reverse primer | 10 $\mu$L |
| MgSO$_4$ | 4 $\mu$L |
| Molecular grade dH$_2$O | 12,75 $\mu$L |
| DMSO | 1,5 $\mu$L |
| *Pfu* Polymerase | 0,5 $\mu$L |
| cDNA (MCF7) | 1,25 $\mu$L |
| **Total Volume** | **50 $\mu$L** |

Table 2.4: Polymerase Chain Reaction conditions for ARID3B Coding Sequence

| Heated lid on 105 °C | | |
|---|---|---|
| 94 °C | 2:00 min | |
| 95 °C | 0:30 min | |
| 64 °C | 0:30 min | } 35 cycles |
| 72 °C | 3:30 min | |
| 72 °C | 10:00 min | |

## 2.5 Linearization of pcDNA Vector and Ligation Reaction

pcDNA Empty vector (4 $\mu$g) and ARID3B CDS DNA (4 $\mu$g) were double digested with Bam HI (Thermo Scientific) and NotI (Thermo Scientific) enzymes as described in Table 2.5, the ratios and the most suitable buffer for the reaction was obtained from manufacturer's database. The reaction was incubated at $37\,^{\circ}$C, overnight. To check the efficacy of the reaction, the reaction products were electrophoresed on a 2% agarose gel at 100V and photographed under UV light. The agarose gel extraction was performed as described above using Zymoclean Gel DNA Recovery Kit, Zymo Research.

Table 2.5: Double Digestion Reaction mixture, ARID3B PCR product and pcDNA Empty Vector

| ARID3B PCR Product (40 ng/$\mu$L) | | pcDNA-Empty Vector (200 ng/$\mu$L) | |
|---|---|---|---|
| DNA | 110 $\mu$L | Plasmid | 20 $\mu$L |
| Bam HI Buffer | 15 $\mu$L | Bam HI Buffer | 15 $\mu$L |
| NotI | 8 $\mu$L | NotI | 8 $\mu$L |
| Bam HI | 4 $\mu$L | Bam HI | 4 $\mu$L |
| dH2O | 13 $\mu$L | dH2O | 106 $\mu$L |
| **Total Volume** | **150 $\mu$L** | **Total Volume** | **150 $\mu$L** |

ARID3B insert and linearized vector were ligated by T4 DNA ligase enzyme. By adding 5 $\mu$L of the DNA to 2 $\mu$L of T4 DNA ligase buffer with 0.1 $\mu$g of pcDNA vector and 1 unit of T4 DNA ligase, ligation reaction was performed for 10 minutes at $22\,^{\circ}$C with a total volume of 20 $\mu$L.

## 2.6 Transformation of competent *E.coli* cells

Competent *E.coli* cells were transformed according to Molecular Cloning: A Laboratory Manual [30] 50 $\mu$L chemically competent *E.coli* cells were thawed on ice. 2 $\mu$L of ligation reaction product was added to competent cells and incubated for 30 minutes on ice. Cells were incubated for 45 seconds at $42\,^{\circ}$C to heat shock and then were incubated on ice for 5 minutes. 500 $\mu$L LB medium was added to each reaction mix. Cells were grown at $37\,^{\circ}$C for 1 hour with 200 rpm shaking. LB-Agar plates containing 100 $\mu$g/mL ampicillin were prepared and 250 $\mu$L cell suspension was spread over

these plates and incubated overnight at $37\,^{\circ}\text{C}$.

The colonies were selected and for the ligation confirmation purposes colony PCR was performed using ARID3B forward and reverse primers. After confirming the efficacy of the previous ligation experiment with PCR, the plasmids were isolated using Roche High Pure Plasmid Isolation kit (Cat# 11754777001). For further confirmation, the plasmids with inserts sent to sequencing.

## 2.7 TA Cloning

p-GEMT vector is a linearized vector with single deoxythymidine 3' end, so that the recircularization of the PCR product is inhibited while insertion, the vector map is shown in Appendix Figure D.2. Before inserting PCR product to the p-GEMT vector, the ligation procedure will be explained in the following section, addition of single deoxyadenosine overhangs at each 3' end of the PCR product was performed using Taq polymerase. This step is necessary before cloning as this will increase the efficiency of the cloning. The reaction mixture was incubated at $72\,^{\circ}\text{C}$ for 30 minutes. Table 2.6 shows the ingredients and the volumes of the mixture.

Table 2.6: TA Cloning Reaction Mixture

| TA Cloning Mixture | |
|---|---|
| Taq Buffer with KCl without $MgCl_2$ | 1 $\mu$L |
| $MgCl_2$ | 1 $\mu$L |
| dATP | 0,2 $\mu$L |
| MCF7 (PCR product) | 7,6 $\mu$L |
| Taq polymerase | 0,2 $\mu$L |
| **Total** | **10 $\mu$L** |

## 2.8 Ligation the products to p-GEMT and p-GEMT Easy Vector Systems

After TA cloning; ligation using T4 DNA Ligase (Thermo Scientific) was performed. The eluted DNA was used with the following ratios of vector; 1:1, 1:3, 1:7.

17

The calculation was done using the following formula:

$$\frac{ng\ of\ vector * kb\ size\ of\ insert}{kb\ size\ of\ vector} * insert:\ vector\ molar\ ratio = ng\ of\ insert$$

The concentration and quality of insert was calculated using NanoDrop 2000c (Thermo, USA). 35,1 ng/$\mu$L of PCR product was obtained. The insert size was 1683 base pair and the size of the vector was 3.0 kb.

In example to get 1:1 ratio,

$$\frac{50\ ng\ of\ vector\ *\ 1.683kb\ insert}{3.0\ kb\ vector} * 1:1 = 28\ ng$$

to be used.

In the light of this information the set-up was prepared as shown in Table 2.7. The ligation reactions were incubated in the refrigerator at $4\,^{\circ}$C, overnight.

Table 2.7: T4 DNA Ligase Reaction Mixture

| **1:1 ratio** | 2 $\mu$L Vector | 2 $\mu$L insert |
|---|---|---|
| **1:3 ratio** | 1 $\mu$L Vector | 3 $\mu$L insert |
| **1:7 ratio** | 0,5 $\mu$L Vector | 3,5 $\mu$L insert |
| | 2X Ligation Buffer - 5 $\mu$L | |
| | T4 DNA Ligase – 1 $\mu$L | |
| **Total** | **10 $\mu$L** | |

## 2.9 DH5$\alpha$ Transformation, Selection of Colonies and Sequencing of Inserts

The following day, ligation reactions were transformed to DH5$\alpha$ competent cells using the protocol according to Molecular Cloning: A Laboratory Manual [30]. The competent cells were thawed in ice for 5 minutes. 2 $\mu$L ligation products were mixed with the competent cells and incubated for 30 minutes in ice. Cells were exposed to heat shock for exactly 1 minute at $42\,^{\circ}$C and were incubated on ice for 5 minutes. 900 $\mu$L SOC medium was added to each reaction mix. Cells were grown at $37\,^{\circ}$C for 1.5 hours with shaking 200 rpm. Then the tubes were centrifuged at 3000 rpm for 10 minutes. The pellet was resuspended in 100 $\mu$L SOC medium without vortex. Spread 40 $\mu$L X-gal and 4 $\mu$L IPTG on each agar plate containing 25 mL LB Agar and 25

18

$\mu$L (100 $\mu$g/ml) Ampicillin. After the chemicals dried add 100 $\mu$L resuspended ligates on each plate and incubate at $37\,^{\circ}$C, overnight. After blue-white selection of the colonies, insert containing colonies were selected and PCR was set up to check the presence of the insert with insert specific primers. After confirmation of the presence of the insert, ARID3B, restriction digestion was performed to both p-GEMT with ARID3B and pcDNA empty vector as previously described on Section 2.5 for further experiments. Later, the colonies also sent to sequencing to further confirm the presence of correct ARID3B full coding sequence.

## 2.10 Linearization of p-GEMT-Easy-ARID3B Vector and pcDNA 3.1(-) Empty vector

To express ARID3B in mammalian cell lines the mammalian expression vector pcDNA 3.1 (-) should be used. Thus, the ARID3B construct within p-GEMT-Easy vector was cut and ligated to pcDNA 3.1 (-) empty vector. Table 2.8 shows the double-digestion reaction of both vectors.

Table 2.8: Double digestion set up of pcDNA EV and p-GEMT with ARID3B insertion

| pcDNA EV (131 ng/$\mu$l) - (4 $\mu$g) | | p-GEMT with inserted ARID3B (100 ng/$\mu$l) - (4 $\mu$g) | |
|---|---|---|---|
| Buffer Bam HI | 15 $\mu$l | Buffer Bam HI | 15 $\mu$l |
| Bam HI | 4 $\mu$l | Bam HI | 4 $\mu$l |
| NotI | 8 $\mu$l | NotI | 8 $\mu$l |
| pcDNA EV | 30.5 $\mu$l | p-GEMT-ARID3B | 40 $\mu$l |
| dH$_2$O | 83 $\mu$l | dH$_2$O | 93.5 $\mu$l |
| **TOTAL** | 150 $\mu$l | **TOTAL** | 150 $\mu$l |

## 2.11 Ligation of ARID3B to pcDNA 3.1 (-) Empty Vector

The double digested products were extracted from the gel using Agarose Gel Extraction kit, Zymogen, quality and quantity were also confirmed by NanoDrop 2000c (Thermo, USA). The ligation reaction set-up was performed according to Table 2.9 and selected colonies were tested for the presence of ARID3B using PCR with ARID-

3B specific primers.

Table 2.9: Ligation set up of ARID3B to pcDNA EV using T4 DNA Ligase

| | Insert | Vector | Buffer | Ligase | dH$_2$O | Total |
|---|---|---|---|---|---|---|
| **Rxn 1 (1:1)** | 1.8 $\mu$l | 1 $\mu$l | 1 $\mu$l | 1 $\mu$l | 5.2 $\mu$l | 10 $\mu$l |
| **Rxn 1 (1:3)** | 5.4 $\mu$l | 1 $\mu$l | 1 $\mu$l | 1 $\mu$l | 1.6 $\mu$l | 10 $\mu$l |
| **No insert** | - | 1 $\mu$l | 1 $\mu$l | 1 $\mu$l | 8 $\mu$l | 10 $\mu$l |
| **No ligase** | 1.8 $\mu$l | 1 $\mu$l | 1 $\mu$l | - | 6.2 $\mu$l | 10 $\mu$l |

## 2.12  ARID3B Coding Sequence Amplification and Cloning to p-GEMT-Easy Vector

MCF7 breast cancer cell line was maintained as instructed in the ATCC (LGC Standards GmbH, Germany) guideline.  Using the cell line as template, cDNA was obtained as previously explained in Section 2.3 and reaction was set up using coding sequence specific primers, the primer sequences are shown in Appendix Table C.1. The RNA isolation was performed using Trizol reagent according to manufacturers' instructions and concentration of RNA was measured using NanoDrop 2000c (Thermo, USA). After DNase treatment, cDNA synthesis (3$\mu$g) was performed as explained previously.

Using MCF7 cDNA as template, ARID3B coding sequence was cloned using ARID3B coding sequence specific primers, using Pfu Polymerase (Thermo Scientific) for high fidelity, the conditions are shown in Appendix Table C.1. The desired amount of PCR product was obtained according to the conditions shown in Table 2.3 and Table 2.4, previously.

Only colony #2 and #5 were selected and plasmid isolation was performed using High Pure Plasmid Isolation Kit (Roche), according to manufacturers' instructions. The concentration and purity of the isolates were determined with NanoDrop 2000c (Thermo, USA) and continued with double digestion of both colonies to check the insertion of ARID3B, the reaction was shown in Table 2.8.

After reconfirmation of the whole sequence for ARID3B coding sequence within

pcDNA 3.1 (-) by digestion, 1683 bp ARID3B and pcDNA 3.1 (-) empty vector were ligated using T4 DNA Ligase (Thermo Scientific) as manufacturers instructed. The colony numbers were as shown below in Table 2.10 with different insert: vector ratios and also colony PCR was performed to check the presence of ARID3B within the colonies.

Table 2.10: Transformation results of pcDNAEV and ARID3B ligations to XL1-Blue competent cells

|  | **Expected number** | **Observed number** |
|---|---|---|
| **5 $\mu$l transformant 1:1** | Few colonies | >30 |
| **5 $\mu$l transformant 1:3** | Few colonies | >30 |
| **1 $\mu$l transformant 1:1** | Few colonies | >15 |
| **1 $\mu$l transformant 1:3** | Few colonies | >20 |
| **No insert** | 0 | 11 |
| **No ligase** | 0 | 6 |
| **Positive control** | Many colonies | Many colonies |

Two colonies to confirm the insertion size were cut and sent to sequencing. The double digestion was performed using Bam HI and NotI enzymes as described in Table 2.11.

Table 2.11: Double digestion of ARID3B inserted pcDNA 3.1 (-) construct #1 and ARID3B inserted pcDNA 3.1 (-) construct #2 with restriction enzymes Bam HI and NotI.

| **construct #1 (117 ng/ $\mu$l)** | | **construct #2 (163.7 ng/ $\mu$l)** | |
|---|---|---|---|
| Buffer Bam HI | 9.5 $\mu$l | Buffer Bam HI | 2 $\mu$l |
| Bam HI | 2 $\mu$l | Bam HI | 2 $\mu$l |
| NotI | 1 $\mu$l | NotI | 1 $\mu$l |
| pcDNAARID3B-1 | 2 $\mu$l | pcDNAARID3B-2 | 3 $\mu$l |
| dH$_2$O | 4.5 $\mu$l | dH$_2$O | 12 $\mu$l |
| **TOTAL** | 20 $\mu$l | **TOTAL** | 20 $\mu$l |

The products were incubated with the enzymes within 10X Bam HI buffer as manufacturers instructed, following the incubation the reaction end products were run in 1% agarose gel. The expected sizes should be as follows, ARID3B coding sequence is 1683 bp, ARID3B inserted pcDNA 3.1 (-) construct is 7083 bp and pcDNA 3.1 (-)

empty vector (when ARID3B coding sequence is removed) is 5400 bp.

## 2.13    Stable Transfection of MCF10A cells with pcDNAARID3B

MCF10A cells were seeded in 6 well plate 24-hours before the transfection to allow reach 80% confluency, which means nearly 2.4 x $10^6$ cells per well. TurboFect Transfection Reagent (Thermo Scientific) was used in 3:2 (TurboFect: plasmid) ratio. For one plate, 4 $\mu$g DNA was diluted within 400 $\mu$L Optimem, 6 $\mu$L transfection reagent was added and samples were incubated for 15-20 minutes at room temperature. Whole mixture was added drop wise to each well, gently rocked, cells were kept with the transfection medium for 24 hours at $37\,^\circ$C, 5% $CO_2$. Later, cells were grown in G418 containing medium for 10-15 days to select pcDNA 3.1(-) with ARID3B containing cells as either monoclonal or polyclonal. The whole reaction set up was performed as shown in Table 2.12:

Table 2.12: Turbofect Transfection Set-up with ARID3B Coding Sequence constructs using pcDNA as vector.

| For 4 $\mu$g | construct #1 (163.7 ng/$\mu$L) | construct #2 (117.6 ng/$\mu$L) | construct #3 (140.6 ng/$\mu$L) | pcDNA EV |
|---|---|---|---|---|
| **DNA** | 24.4 $\mu$L | 34 $\mu$L | 28.4 $\mu$L | 30.3 $\mu$L |
| **Turbofect** | 6 $\mu$L | 6 $\mu$L | 6 $\mu$L | 6 $\mu$L |
| **Optimem** | 365.6 $\mu$L | 360 $\mu$L | 365.6 $\mu$L | 363.7 $\mu$L |

For untransfected well, 400 $\mu$L Optimem was added. Next day, 500 $\mu$g/ mL G418 antibiotic was added using literature information [16]. G418 stock solution is 50 mg/ mL and following calculation was done to get appropriate concentration:

500 $\mu$g/ mL x 4 mL = 50000 $\mu$g/ mL x V

V= 40 $\mu$L G418 for one well

After all untransfected cells died, which took 5 days, G418 concentration was dropped to its half; to 250 $\mu$g/ mL.

## 2.14 Real Time Polymerase Chain Reaction for ARID3B, Empty Vector Transfected MCF10A cells

Polyclonal colonies were selected from empty vector and ARID3B containing pcDNA transfected MCF10A cells RNA isolation and cDNA synthesis were performed as explained in Section2.3. Real time PCR was performed using these cDNAs to check the expression levels of ARID3B in empty vector and ARID3B containing pcDNA transfected cells. The reaction mixture was used as shown in Table 2.13 and the reaction conditions were shown in Table 2.14.

Table 2.13: Real Time PCR reaction mixture with FastStart Universal SYBR Green Master (Roche)

| 1X Master Mix | |
|---|---|
| FastStart Universal SYBR Green Master | 10 $\mu$L |
| Forward primer (5 $\mu$M) | 1.2 $\mu$L |
| Reverse primer (5 $\mu$M) | 1.2 $\mu$L |
| Molecular grade dH$_2$O | 3.6 $\mu$L |
| cDNA (diluted) | 4 $\mu$L |
| **Total Volume** | **20 $\mu$L** |

Table 2.14: Polymerase Chain Reaction conditions for ARID3B Coding Sequence

| 95 °C | 10:00 min | |
|---|---|---|
| 94 °C | 0:30 min | |
| 58 °C | 0:30 min | } 40 cycles |
| 72 °C | 0:30 min | |
| 55 − 90 °C | 00:05 min | |

$\Delta\Delta$Ct analysis was performed as previously described [7] and graphical representation was generated using Graph Pad Prism 8. The final products were run on 2% agarose gel at 100V and visualized under UV.

## 2.15 Protein Isolation and Western Blotting

Cells were grown until they reach the optimal confluency. Total protein isolation was done by the following procedure; the proliferating cells were washed two times with phosphate-buffered saline (PBS) and all cells were collected with total protein lysis

buffer; 150mM NaCl, 50 mM Tris, 1% Triton X-100, 0.5% Sodium deoxcholate, 0.1% SDS and a tablet of complete protease inhibitor complex (Roche) or M-PER Mammalian Protein Extraction Kit (Thermo Scientific) was used. Cells were centrifuged at 14,000 g for 15 minutes at $4\,°C$.

Supernatant was collected and protein concentrations were determined using BCA protein assay (Pierce). For cytoplasmic protein isolation cells were washed two times with phosphate-buffered saline (PBS) and all cells were collected with cytoplasmic protein lysis buffer; 20 mM hydroxyethyl piperazineethanesulfonic acid (HEPES), 10 mM KCl, 0.1 mM ethylenediaminetetraacetic acid, 1 mM Dithiothreitol (DTT), 1.5 mM $MgCl_2$, 0.5 mM NaCl, 10% NP-40, 25% glycerol and a tablet of complete protease inhibitor complex (Roche). Supernatant was collected and protein concentrations were determined using BCA protein assay (Pierce).

For nuclear protein isolation cells were washed two times with phosphate-buffered saline (PBS) and all cells were collected with nuclear protein lysis buffer; 20 mM hydroxyethyl piperazineethanesulfonic acid (HEPES), 10 mM KCl, 0.1 mM ethylenediaminetetraacetic acid, 1 mM Dithiothreitol (DTT), 1.5 mM $MgCl_2$ and a tablet of complete protease inhibitor complex (Roche) or NE-PER Nuclear and Cytoplasmic Extraction Kit (Thermo Scientific) was used. Supernatant was collected and protein concentrations were determined using BCA protein assay (Pierce).

Proteins (50 $\mu$g) were then denatured in 6X Laemmli buffer (30% 2-mercaptoethanol, 12% SDS, 0.012% bromophenol blue, 60% Glycerol, 0.375 M Tris) at $100\,°C$ for 5 minutes. Proteins were resolved by an 8% SDS-polyacrylamide gel and were transferred to PVDF membranes (Roche). Membranes were blocked for 1 hour at room temperature in Tris Buffer Saline Tween (TBST), (137 mM NaCl, 20 mM Tris, pH: 7.6, 0.1% Tween 20) with 5% non-fat dry milk (Bio-Rad) for ARID3B detection, The membranes were incubated for 16 h at $4\,°C$ with; rabbit polyclonal anti-ARID3B antibody (1:1500, Abcam) in 5% non-fat dry milk-0.1% TBST. The membranes were washed three times with 0.1% TBST, and incubated for 1 hour with peroxidase conjugated mouse anti-rabbit antibody (1:2000, Santa Cruz) in 5% non-fat dry milk-0.1% TBST. After washing three times with 0.1% TBST, antigen-antibody complexes were visualized with the enhanced chemiluminescence kit (Pierce) by exposure to X-ray

films (Kodak). Later, the blots were stripped and hybridized with monoclonal goat anti-mouse actin beta antibody (1:1000, Santa Cruz) in 5% non-fat dry milk - 0.1% TBST. Antigen-antibody complexes were visualized as described above. Anti-actin beta (ACTB) antibody detected a 42 kDa band.

## 2.16   Densitometric Analysis of Western Blots

Western Blotting Results were visualized either using X-ray films (Kodak) or Bio-Rad Image Lab Software. Densitometric analyses of the results were performed according to band intensities of proteins using Image J (NIH) software. The fold changes were calculated using the following equation by normalizing protein of interest to housekeeping proteins.

$$Fold\ change = \frac{Treatment(\frac{ARID3B}{ACTB})}{Control(\frac{ARID3B}{ACTB})}$$

Treatment stands for AG825 treated cells, while control stands for DMSO treated cells. Both groups were hybridized with anti-ARID3B antibody to check ARID3B levels and anti-actin beta as internal loading control. Data was analyzed using two way ANOVA following by Sidak's multiple comparison test using Graph Pad Prism 8 software.

# CHAPTER 3

# RESULTS AND DISCUSSION

## 3.1   ARID3B Isoforms and Pseudogene

To start our expression analysis, we first performed bioinformatics analysis to determine potential isoforms of ARID3B. Using National Center for Biotechnology Information (NCBI) and University of California Santa Cruz Genome Browser (UCSC) databases molecular weights, exon-intron boundaries and also binding regions of particular antibodies which were used within this study are schematized as shown in Figure 3.1, Figure 3.2, Figure 3.3 and Figure 3.4. There are 4 isoforms, compared to the full length ARID3B (61 kDa),are possibly produced by alternative splicing to generate 48 kDa, 28 kDa, 23 kDa and 14 kDa sizes.

Among these isoforms, only 61 kDa isoform has been experimentally shown to induce apoptosis in ovarian cancer cell lines when overexpressed. Overexpression of 61kDa ARID3B has been associated with upregulation of pro-apototic BIM, induction of Tumor Necrosis Factor alpha (TNF$\alpha$) and TNF related apoptosis inducing ligand (TRAIL) induced apoptosis [17].

In addition to several isoforms, a pseudogene of ARID3B on chromosome 1p31 was reported earlier [25]. To determine if ARID3B pseudogene was expressed, RT-PCR was performed as shown in Figure 3.5. The nucleotide comparison between ARID3B full length sequence and pseudogene is shown in Appendix Table E.1.

According to the above results, we did not detect a band that may have indicated the expression of the pseudogene.
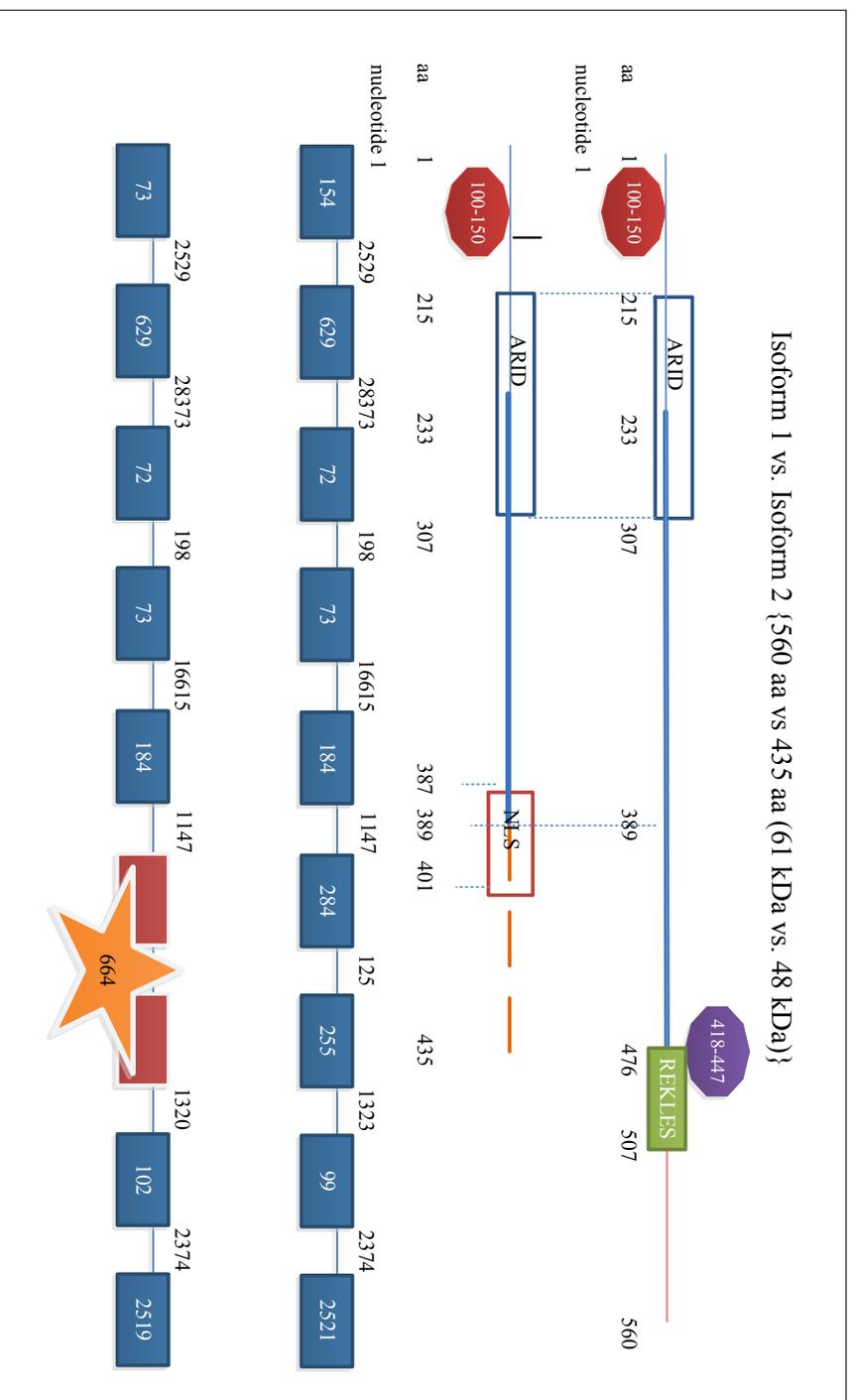
Figure 3.1: 560 and 435 amino acid ARID3B isoforms with their nucleotide lengths.

● : N-terminus antibody binding site. ● : C-terminus antibody binding site. ■ : Exon — : Intron ★ : isoform 2 Specific intron
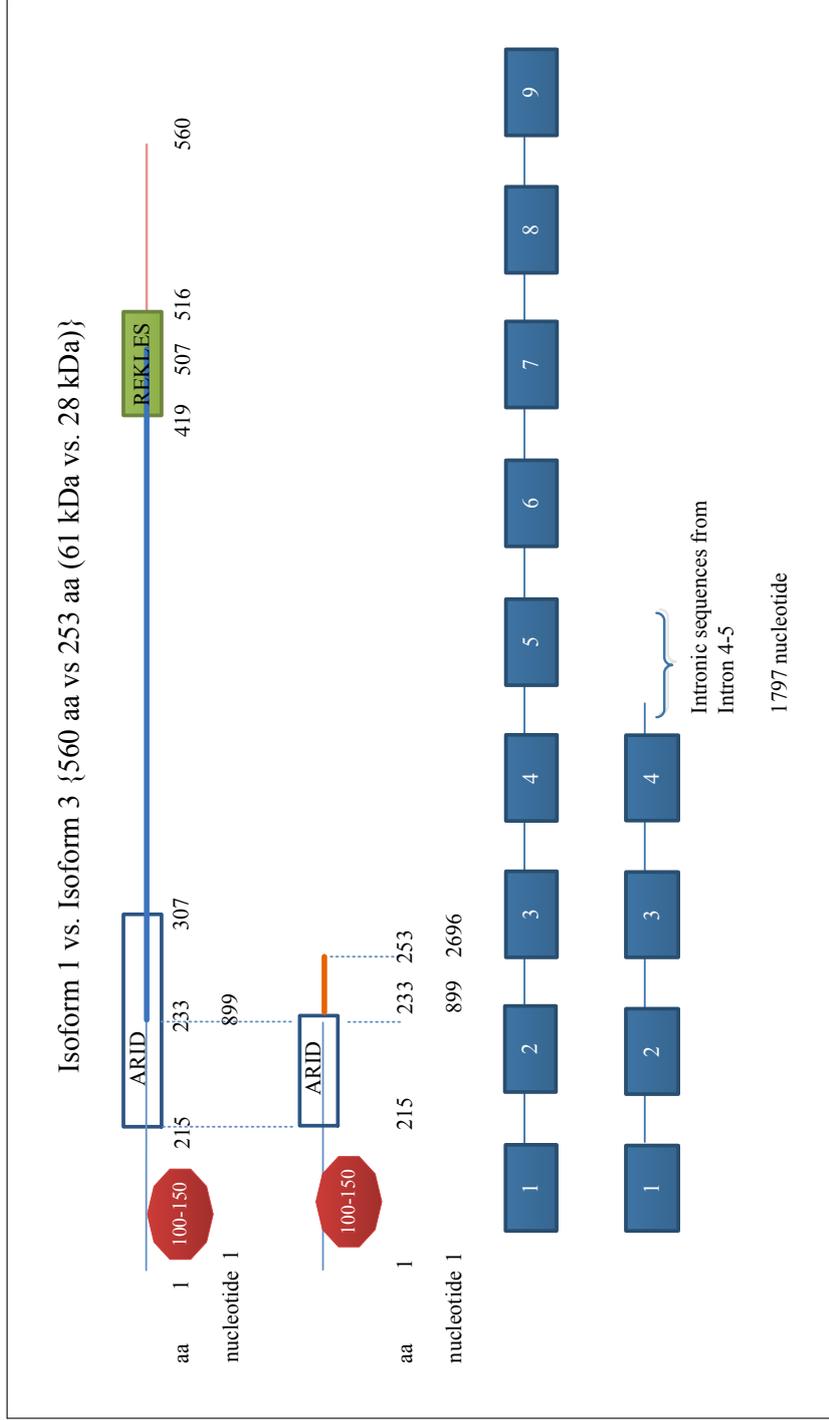
Figure 3.2: 560 and 253 amino acid ARID3B isoforms with their nucleotide lengths.

● : N-terminus antibody binding site.  ● : C-terminus antibody binding site.  ▬ : Exon  ── : Intron
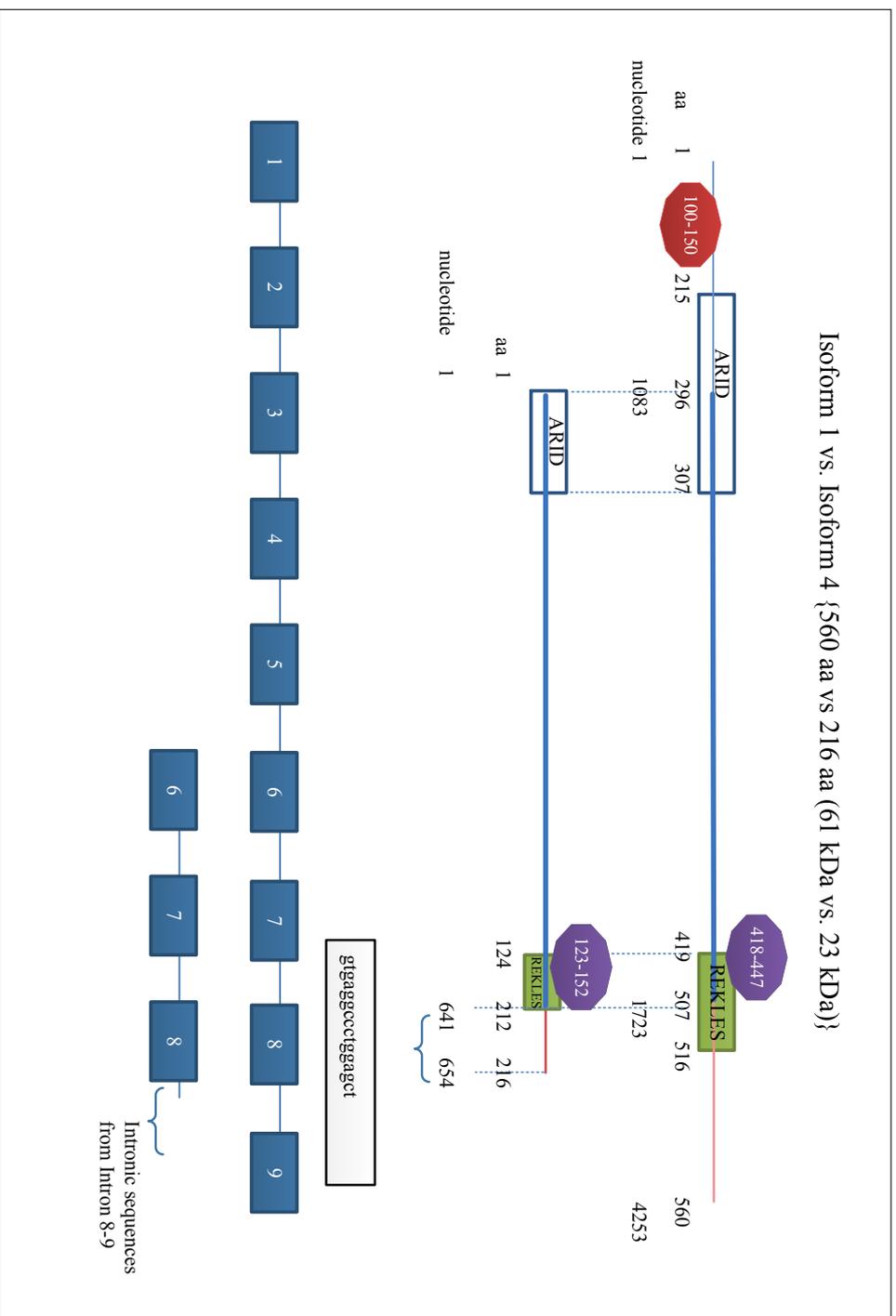
Figure 3.3: 560 and 216 amino acid ARID3B isoforms with their nucleotide lengths.

● : N-terminus antibody binding site. ● : C-terminus antibody binding site. ▬ : Exon ── : Intron
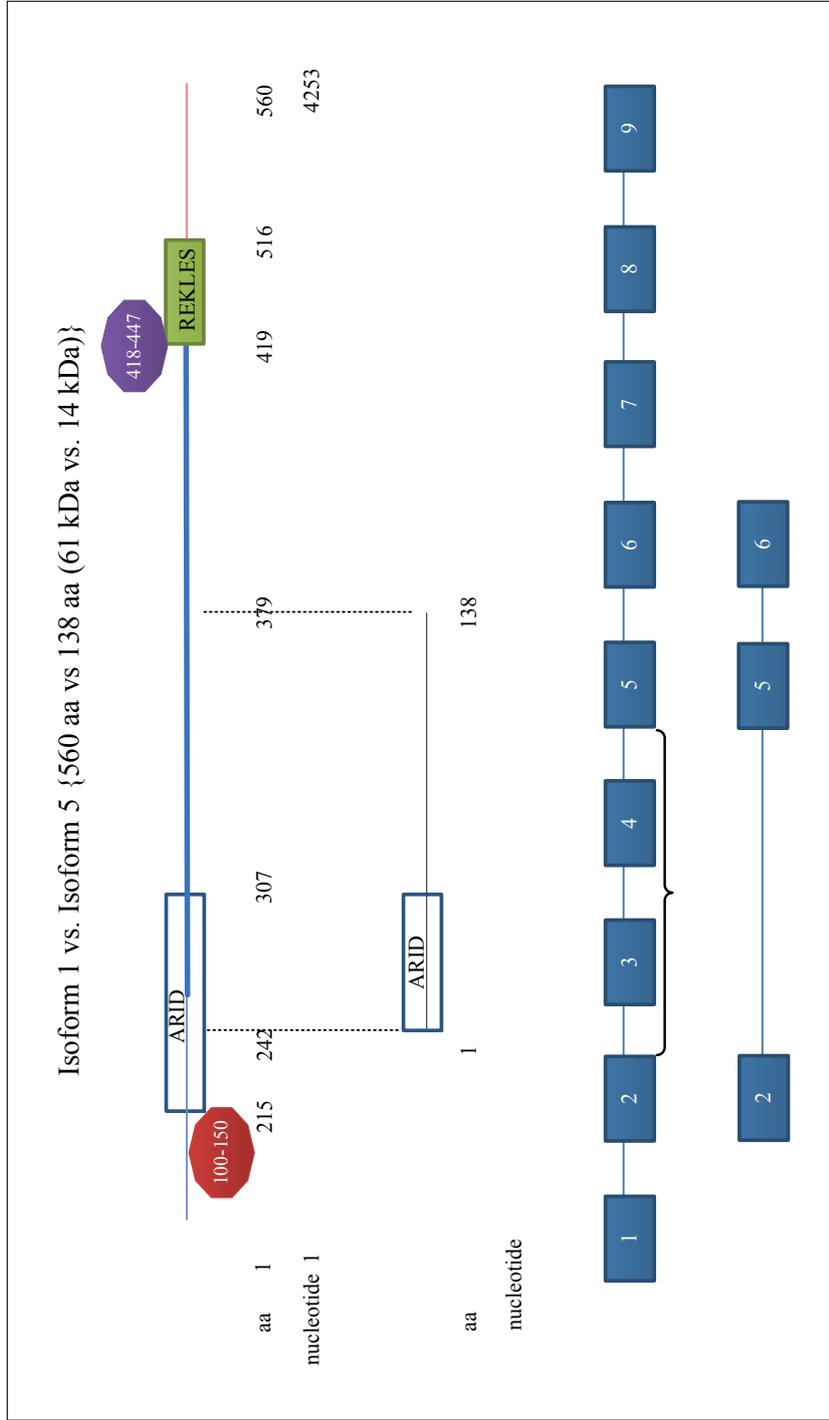
Figure 3.4: 560 and 138 amino acid ARID3B isoforms with their nucleotide lengths.

● : N-terminus antibody binding site. ● : C-terminus antibody binding site.
● : N-terminus antibody binding site. ■ : Exon ─── : Intron
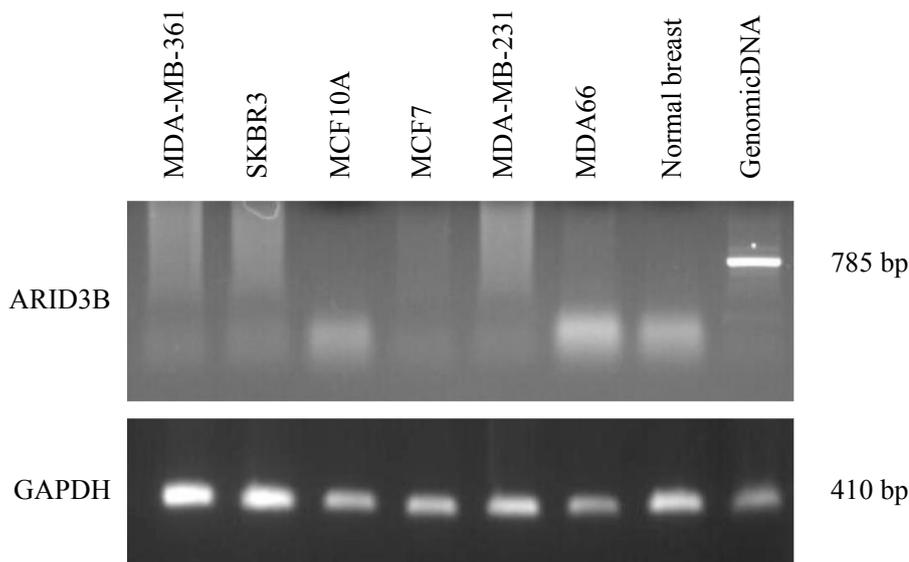
Figure 3.5: Pseudogene ARID3B expression analysis by polymerase chain reaction using pseudogene specific primers with several different breast cancer cell line cDNAs, normal breast cDNA and genomic DNA. GAPDH primers were also used to show the quality of cDNAs.

## 3.2 ARID3B Coding Sequence Cloning to pcDNA 3.1 (-) Empty Vector

To investigate the role of ARID3B, we planned to initially clone the coding sequence of ARID3B into a mammalian expression vector for transfection experiments. The ARID3B coding sequence was previously provided to our laboratory in pBluescript plasmid by Numata et al. [25], this construct was used as template and coding sequence specific primers, with NotI and Bam HI restriction sites, 1683 base pair length coding sequence was amplified. PCR product was isolated using Agarose Gel Extraction Kit (Roche) and the concentration was calculated using NanoDrop 2000c (Thermo, USA). pcDNA 3.1 (-) was used for cloning, the vector map is shown in Appendix Table D.1. Cloning was performed as explained described earlier. The cloning result is shown in Appendix Figure G.1.

The colonies were then sent to the sequencing to Central Laboratory, METU for further sequence confirmation. According to the sequencing results as shown in Fig-

ure 3.6 and Figure 3.7, extra CAG nucleotides were detected which are not present within the full coding sequence as in NCBI database. Several other colonies yielded the same result. The original template was also sequenced to find the extra CAG nucleotide. Because such a sequence is not reported in the database, MCF7 cells were chosen to amplify ARID3B coding sequence.

```
Query  1601  AGATCAGGATCAACGGCAGGG-A-A-GACAGAGCAGAGGCCTCGGCTGCAGCACTGAACC  1657
             |||||||||||||||||||| | |  ||||||||||||||||||||||||||||||||||
Sbjct   368  AGATCAGGATCAACGGCAGGGCAGAAGACAGAGCAGAGGCCTCGGCTGCAGCACTGAACC   427
```

Figure 3.6: Colony# 2 ARID3B-F3 primer reading result, blasted with ARID3B coding sequence on NCBI nucleotide database

```
Query   367  AGATCAGGATCAACGGCAGGGCAGAAGACAGAGCAGAGGCCTCGGCTGCAGCACTGAACC   426
             ||||||||||||||||| ||  |||||||||||||||||||||||||||||||||||||
Sbjct  1400  AGATCAGGATCAACGGCA-GG--GAAGACAGAGCAGAGGCCTCGGCTGCAGCACTGAACC  1456
```

Figure 3.7: Colony# 3 ARID3B-F3 primer reading result, blasted with ARID3B coding sequence on NCBI nucleotide database

## 3.3 ARID3B Coding Sequence Amplification and Cloning to p-GEMT-Easy Vector

cDNA synthesis was performed as previously explained and the RT-PCR result of MCF7 cDNA and confirmation with housekeeping GAPDH primers are shown in Appendix Figure F.1.

In order to check if any CAG present within MCF7 cDNA template before continuing any further experiments, PCR product was sent to sequencing. According to the sequencing results, there was no CAG insertion at the suspected region. Figure 3.8 shows the sequencing confirmation for the absence of CAG.

Then, ARID3B CDS insert was ligated to p-GEMT-Easy, the whole procedure was explained previously in Section 2.5–2.9. To observe the presence of ARID3B CDS within the colonies, colony PCR was performed using ARID3B CDS primers, the result is represented in Figure 3.9.
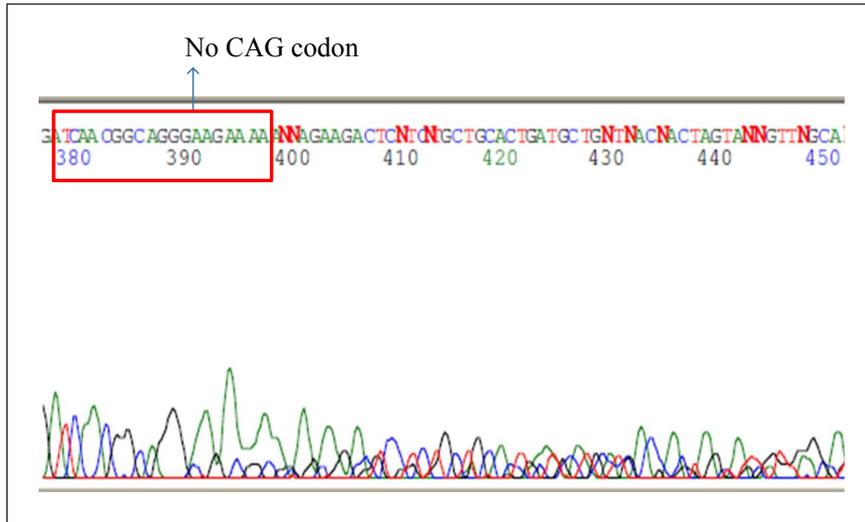
Figure 3.8: Sequencing Result for MCF7 templated ARID3B coding sequence for confirmation of CAG absence
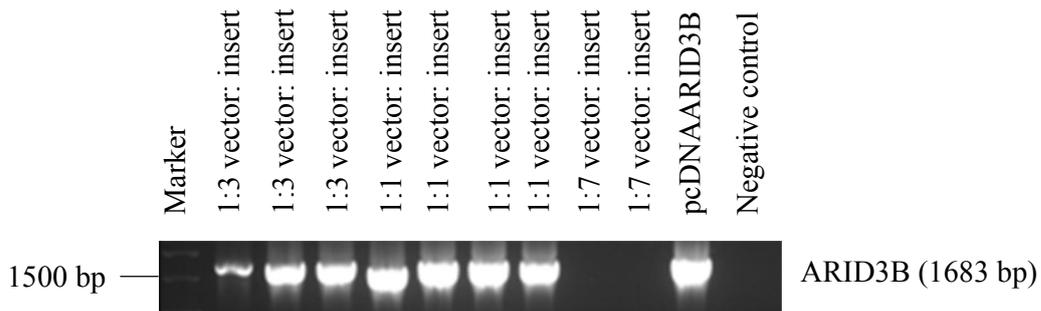


Figure 3.9: Colony PCR with ARID3B CDS primers after PGEMT-Easy Ligation reaction. M: DNA Ladder, Gene Ruler (Fermentas); 1-2-3: 1:3 vector: insert ratio; 4-5-6-7: 1:1 vector: insert ratio; 8-9: 1:7 vector: insert ratio; 10: ARID3B inserted pcDNA as positive control; 11: negative control

Insertion was confirmed by PCR and then sent to sequencing for further sequence confirmation. Figure 3.10 represents the PCR result of digestion for ARID3B extraction from p-GEMT-Easy and digestion of pcDNA empty vector. Figure 3.11 represents double digestion of constructs after ligation to pcDNA 3.1 (-).

With the preliminary confirmation using enzyme digestion of the products, the samples were sent to the Central Laboratory, METU for further sequence confirmation.

34
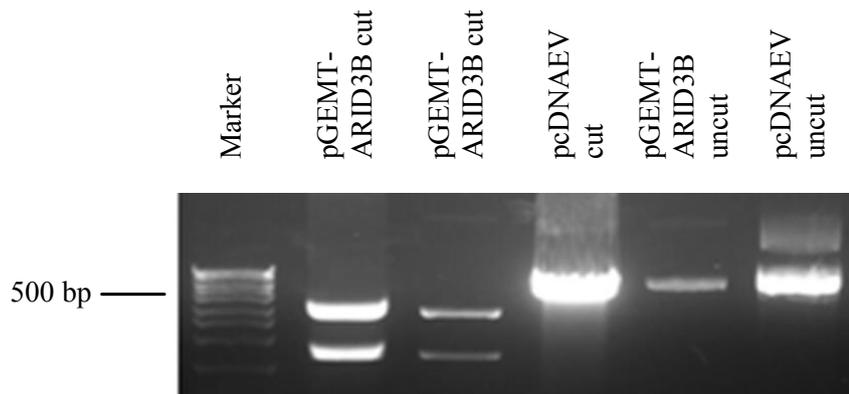
Figure 3.10: Double Digestion of pcDNA EV and pGEMT-ARID3B. M: DNA Ladder, Gene Ruler (Fermentas); 1: pGEMT-ARID3B cut, 2: pGEMT-ARID3B cut, 3: pcDNAEV cut, 4: pGEMT-ARID3B uncut, 5: pcDNAEV uncut.

pGEMT empty: 3015 bp, pGEMT-ARID3B 4698bp, ARID3B: 1683bp, pcDNAEV: 5400bp
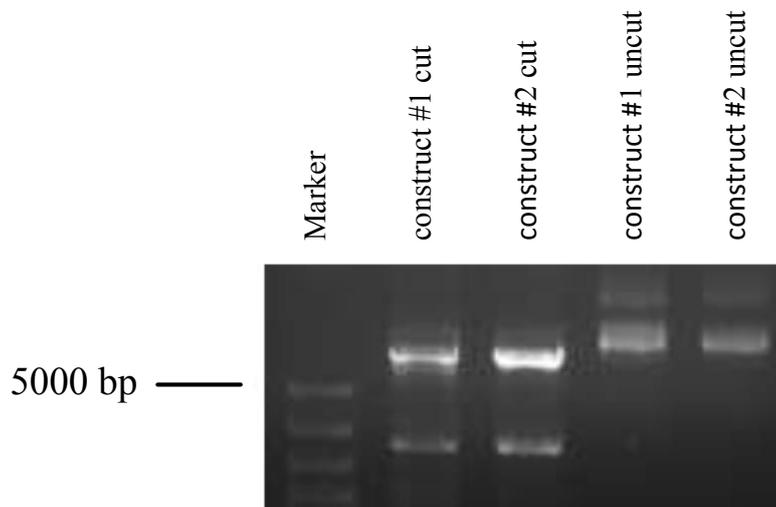


Figure 3.11: Double Digestion of construct #1 and construct #2. M: DNA Ladder, Gene Ruler (Fermentas); 1: construct #1 cut, 2: construct #2 cut, 3: construct #1 uncut, 4: construct #2 uncut.

Sizes, ARID3B: 1683 bp, ARID3B inserted pcDNA 3.1 (-): 7083bp, pcDNA-digested 5400 bp

## 3.4 Expression of ARID3B within ARID3B inserted pcDNA 3.1(-) transfected MCF10A cells

Following transfection of ARID3B inserted pcDNAconstructs in MCF10A cells and selection of stable transfectants, RT-PCR was performed to detect overexpression of ARID3B in transfected MCF10A cells.

In Figure 3.12 relative expression of ARID3B transfected cells compared to empty vector transfected cells is shown. The reaction condition and primer sequences are shown on Section 2.14 and Table C.1. Unfortunately, there was no significant increase in ARID3B expression in transfected cells compared to vector only transfected cells.

## 3.5 Protein Expression Level Detection of ARID3B within ARID3B transfected MCF10A cells

To further check if there is any protein level increase in total lysates of ARID3B transfected cells compared to empty vector transfected cells, Western Blotting was performed as shown in Figure 3.13.

ARID3B have several isoforms as described earlier. The anticipated result was to visualize overexpressed 61kDa (full length isoform) together with endogenous 48 kDa and 28 kDa isoforms, as N-terminus ARID3B antibody was used in this assay as indicated in Section 2.15.

The protein levels of ARID3B of ARID3B inserted pcDNA 3.1 (-) transfected MCF10A lysates and pcDNA-empty vector transfected lysates were the same, the transfections were repeated twice. However, we failed to generate an ARID3B overexpressing model system.

Overexpression experiment was repeated one more time with MCF10A cell line, transiently with the same constructs. Figure 3.14 shows the transient transfection results for 24 and 48 hours. When compared to controls, transient transfection also did not cause ARID3B overexpression. There are some possible explanations for this result, either cells could not tolerate ARID3B overexpression, or there are some compen-
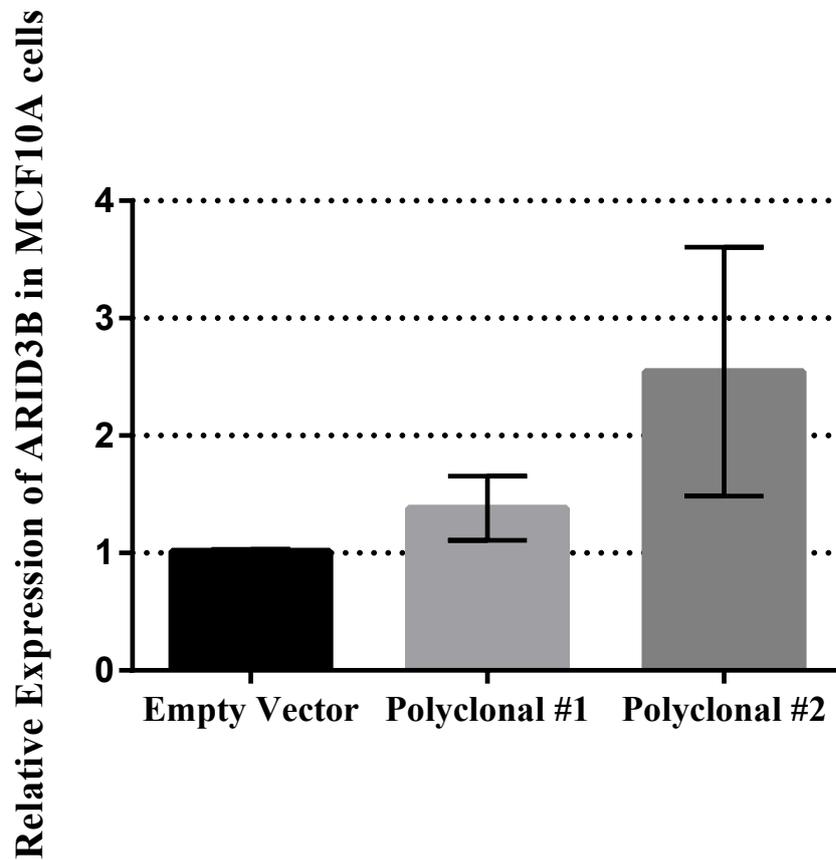
Figure 3.12: Expression of ARID3B within ARID3B transfected MCF10A cell line. Reaction efficiency incorporated $\Delta\Delta Ct$ formula [7] was used for quantification. Normalization performed for Empty vector transfected cells. Error bars represents the standard deviation of two independent experiments with total six technical replicates. There was no significant increase of ARID3B expression in ARID3B transfected MCF10A cells compared to empty vector transfected cells. (t test, p> 0.05)

MCF10A-ARID3B-Empty vector
MCF10A-ARID3B-1
MCF10A-ARID3B-2
MCF10A-ARID3B-mixed

ARID3B
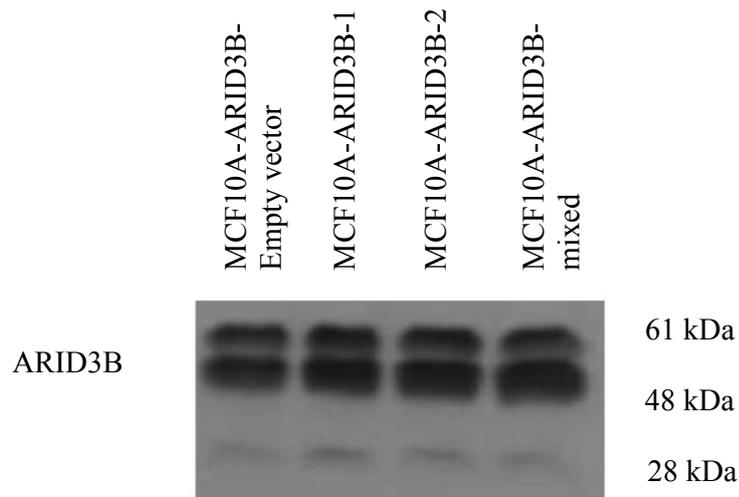
61 kDa
48 kDa
28 kDa

Figure 3.13: ARID3B protein expression in ARID3B transfected and empty vector transfected MCF10A cell lines detected by Western Blotting. 61kDa and 48 kDa isoforms were recognized, 28 kDa isoform was also present as lower faint band.



24 hours                    48 hours

MCF10A-ARID3B
MCF10A-Empty vector
MCF10A-untransfected
MCF10A-ARID3B
MCF10A-Empty vector
MCF10A-untransfected

ARID3B
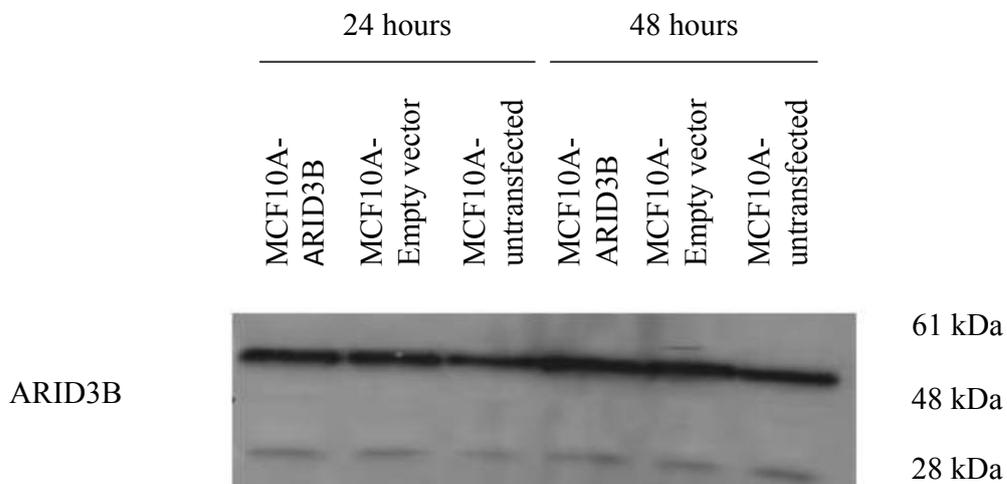
61 kDa
48 kDa
28 kDa

Figure 3.14: ARID3B expression in MCF10A-ARID3B, empty vector transfected and untransfected MCF10A cell lines detected by Western Blotting at 24 hours and 48 hours time points. 61 kDa and 28 kDa isoforms were present, but 48 kDa isoform was not recognized.

satory mechanisms to counteract the ARID3B overexpression. It is also possible that our transfection based approach was simply not efficient enough.

Apart from MCF10A ARID3B overexpression experiment MDA-MB-361 cell line was used to overexpress ARID3B with the same vector constructs. However, there was no overexpression observed in this cell line either (data not shown).

## 3.6   ARID3B Protein Expression in Breast Cancer Cell lines

Another study from our laboratory investigated expression of ARID3B in breast cancer patient samples using immunohistochemistry. Analysis of 63 patients with invasive breast carcinoma samples resulted 89% positive correlation between nuclear ARID3B staining and ER-staining. On the other hand, there was a negative correlation between ARID3B staining and ERBB2 staining; 63% nuclear ARID3B stained samples were ERBB2 (-) patients. ARID3B stainings was observed in nuclear lysates of the samples as well as weak cytoplasmic and membranous stainings. In order to verify patient data with breast cancer cell lines, different breast cancer cell lines that were representatives of ER positive and ERBB2 positive status were selected.

To delineate the ARID3B protein expression within different contexts, four different cell lines were used. Using an N-terminus antibody which detects the 100-150 amino acid region and recognizes 61 kDa, 48 kDa and 28 kDa isoforms, ARID3B presence was detected using nuclear and cytoplasmic lysates. The aim of using nuclear and cytoplasmic lysates was to also show particular localization of ARID3B as nuclear ARID3B staining was correlated with estrogen receptor (ER) positivity.

Compared to MCF10A (triple negative) immortalized non-tumorigenic cells, MCF7 (ER$\alpha$ positive) cells showed increased nuclear 61 kDa and 48 kDa ARID3B expression. This situation may suggest there may be a relationship between ARID3B and ER$\alpha$ presence. To test this hypothesis, MDA66 cells (ER$\alpha$ positive) and MDA-MB-231 cells together with MCF10A and MCF7 cells were used in Western Blotting with N-terminus ARID3B antibody to check the relationship between two proteins, in Figure 3.15a. 28 kDa isoforms was barely detectable on the films. MDA66 (ER$\alpha$ positive) and MCF7 (ER$\alpha$ positive) cells showed increased ARID3B expression com-
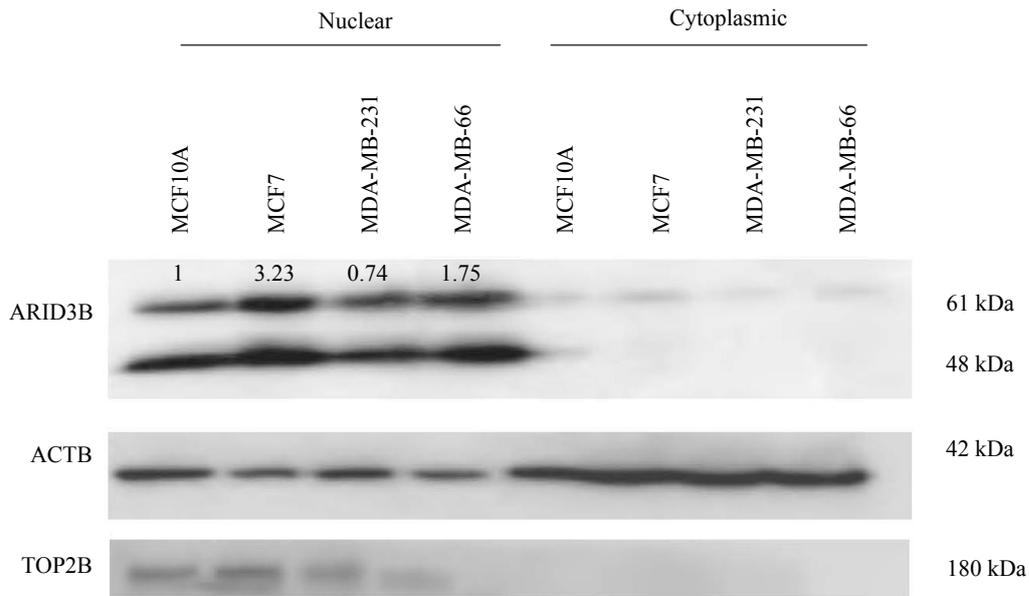
(a) ARID3B protein expression in MCF10A, MCF7, MDA-MB-231, MDA66 nuclear and cytoplasmic lysates detected by Western Blotting. MDA66 (ER$\alpha$ positive) and MCF7 (ER$\alpha$ positive) cells showed increased ARID3B expression compared to MCF10A (ER$\alpha$ negative) and MDA-MB-231 (ER$\alpha$ negative) cells, with ACTB and TOP2B loading controls.
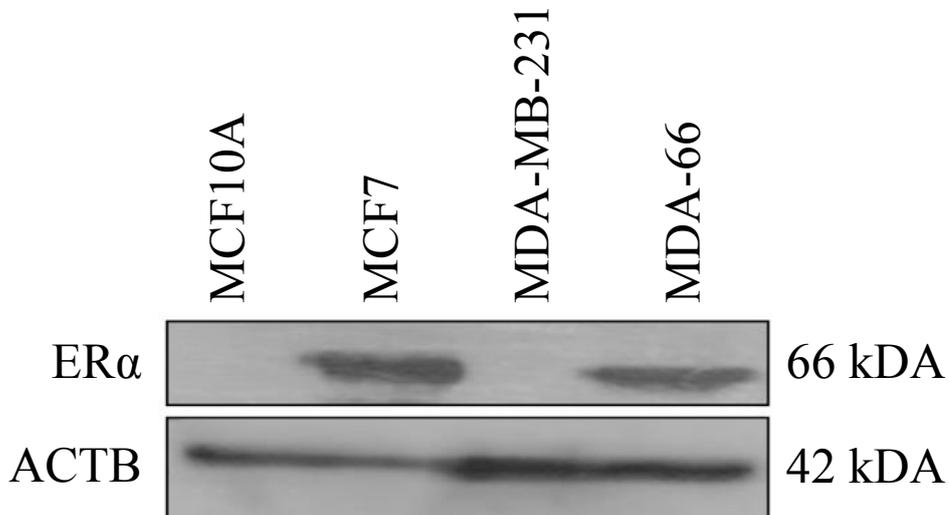


(b) ER$\alpha$ expression was also checked in MCF10A, MCF7, MDA-MB-231, MDA66 total lysates.

Figure 3.15: ARID3B and ER$\alpha$ protein expression in MCF10A, MCF7, MDA-MB-231, MDA66 nuclear, cytoplasmic and total lysates with ACTB, TOP2B loading controls.

pared to MCF10A (ER$\alpha$ negative) and MDA-MB-231 (ER$\alpha$ negative) cells as in the same fashion with patient samples results. The equal loading control was performed with anti-ACTB antibody and nuclear lysates confirmed and discriminated from cytoplasmic lysates with anti-TOP2B antibody. The Image J (NIH) software was used to quantify the bands, as described in Section 2.16. According to analysis; densitometric values are as follows, shown on Figure 3.15a and normalization performed on MCF10A; MCF10A (densitometric value: 1), MCF7 (densitometric value: 3.23), MDA-MB-231 (densitometric value: 0.74), MDA-66 (densitometric value: 1.75). It is worth pointing out that there was anti-ACTB staining in nuclear lysates as shown in Figure 3.15a. Recent studies showed that, actin may have crucial functions as being cytoskeletal protein within nucleoplasm [26]. In addition, ACTB and actin related proteins were also shown as component of chromatin remodelling complexes [26]. And last but not least, ACTB was shown as part of an important component for RNA polymerase I [8, 28], II [12, 21] and III [13] to function in transcription process.

In addition, it was observed there might be a negative relationship between ARID3B and ERBB2 expression as was shown in patient samples. To check this, ERBB2 overexpressing breast cancer cells were used to perform Western blot analyses. MCF10A (triple negative), and ERBB2 positive breast cancer cell lines; BT474, MDA-MB-361 and SKBR3 total proteins were used as shown in Figure 3.16. Equal loading was confirmed with anti-ACTB antibody. According to the result, BT474 and MDA-MB-361 cell lines indeed lacked 61 kDa ARID3B, which supports the negative correlation between ERBB2 and ARID3B expression; however SKBR3 cell line did have 61 kDa ARID3B isoform, as shown in Figure 3.16. The latter result was actually consistent with the other data from the laboratory with patient IHC results. According to this, 63% ERBB2 positive cases showed negative correlation with ARID3B expression. However, the remaining showed ARID3B expression as was shown in SKBR3 cell line.

## 3.7 ARID3B expression in ERBB2 positive cell lines

ERBB2 is well known to be a member of signalling pathways for cellular proliferation and survival. Together with ERBB2, estrogen receptor (ER) is also a player
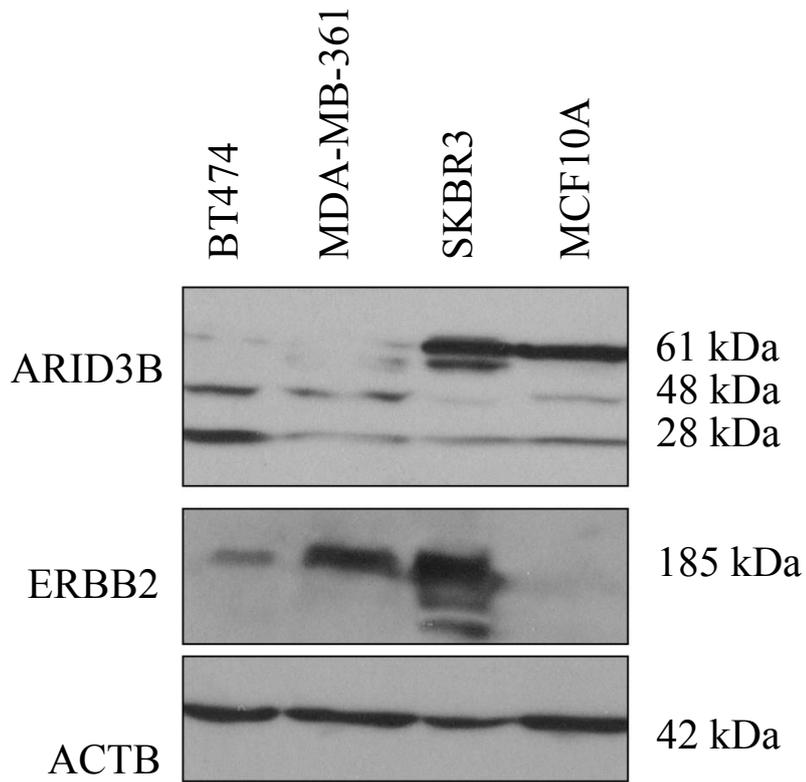
Figure 3.16: ARID3B expression in total lysates of ERBB2 positive BT474, MDA-MB-361, SKBR3 breast cancer derived cell lines and ERBB2 negative non-tumorigenic immortalized breast cancer cell line MCF10A detected by Western Blotting.

of cellular proliferation and survival pathways. Several breast cancer therapies are designed to target ERBB2 and/or ER, however it is also challenging due to crosstalk between both pathways. Several studies showed that there is negative correlation between ERBB2 and ER expression in breast cancer [19]. For example, p44/42 MAPK activation results in degradation of ER. It was also shown that, inhibition of HER receptor layer via trastuzumab, lapatinib, perfuzumab and trastuzumab-emtansine therapies either single or combined, can lead reactivation/restoration of ER signalling in ERBB2 positive breast cancer cell lines [9].

This kind of relationship might be present in breast cancer derived cell lines, especially in SKBR3 with presence of both ARID3B and ERBB2 being different from the other breast cancer derived cell lines. Thus, we asked whether ARID3B expression would be affected from ERBB2 inhibition. To test the effect of ERBB2 on ARID3B expression, we treated SKBR3 cells with AG825, a known inhibitor of ERBB2 [34].

As shown in Figure 3.17 and Figure 3.18, upon 50 $\mu$M AG825 treatment for 24 hours, ARID3B expression was lost compared to the DMSO treated samples. A similar loss was seen in 48 hours treatment.
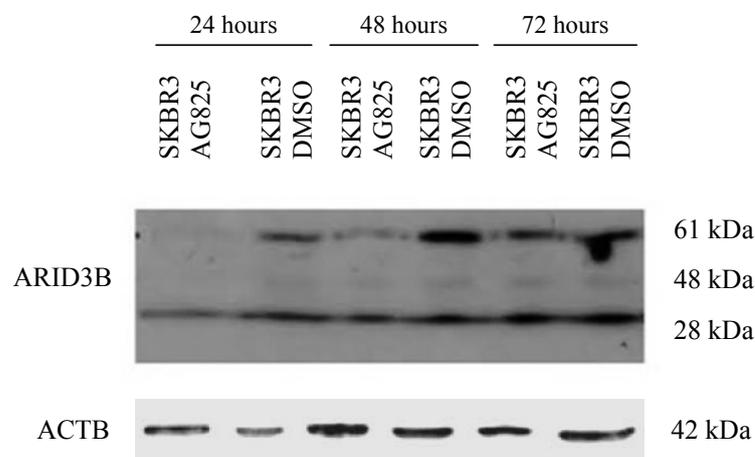


Figure 3.17: ARID3B expression in total lysates of ERBB2 inhibitor drug treated and DMSO added as negative control SKBR3 breast cancer derived cell lines detected by Western Blotting. Experiment was repeated three times, analysis for three replicates are graphed using GraphPad Prism 8 software.Normalization was done according to ACTB and using Image J (NIH) relative densities were calculated.
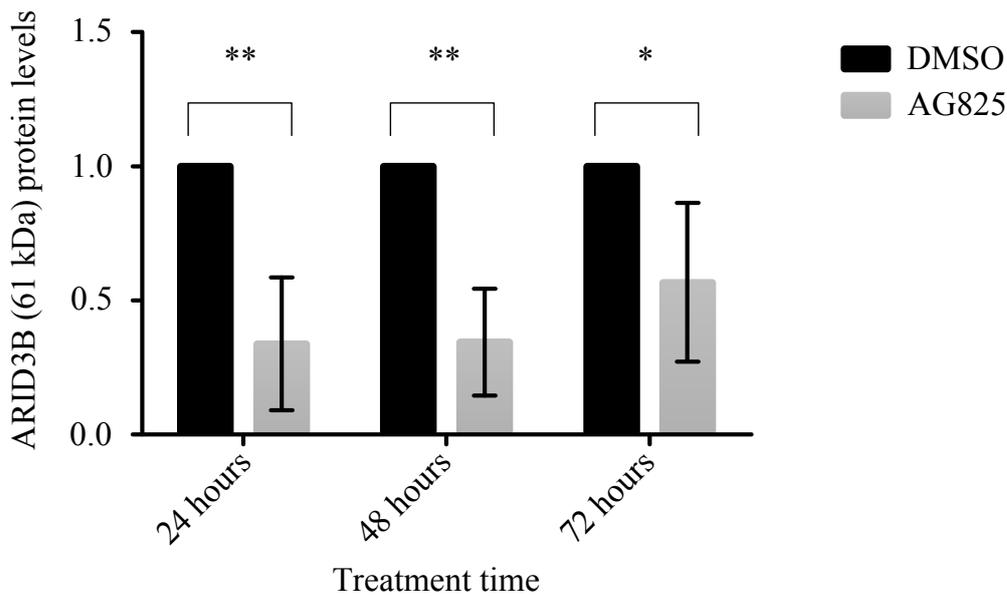
Figure 3.18: ARID3B (61kDa) protein expression levels among AG825 and DMSO treated SKBR3 cells in different time points, 24 hours, 48 hours and 72 hours. This graph was generated with three biological replicates. Densitometry analysis performed using Image J (NIH) programme and normalized to ACTB. Data was analyzed by two way ANOVA followed by Sidak's multiple comparison test. (**) indicates statistical significance ($p < 0.001$)

As there might be a crosstalk between ER$\alpha$ and ERBB2 in SKBR3 breast cancer derived cell line, we checked the presence of ER$\alpha$ protein on ER$\alpha$ negative SKBR3 cells with western blotting. Although SKBR3 is ER$\alpha$ negative cell line, the drug treatment may affect the upstream pathways and ER$\alpha$ expression may be restored.

However, as shown in Figure 3.19 there was no expected restoration on ER$\alpha$ levels. The size of the protein should be seen as 66 kDa as in MCF7 and MDA66 cell lines, but there were no bands present at this level. The lower bands are non-specific bands.

These results suggested that there might be an effect of ERBB2 inhibition on ARID3B protein level decrease in SKBR3 cell line. However, there is no relationship between ERBB2 treatment and ARID3B relationship in favor of ER$\alpha$ protein activation in SKBR3 cells.

44
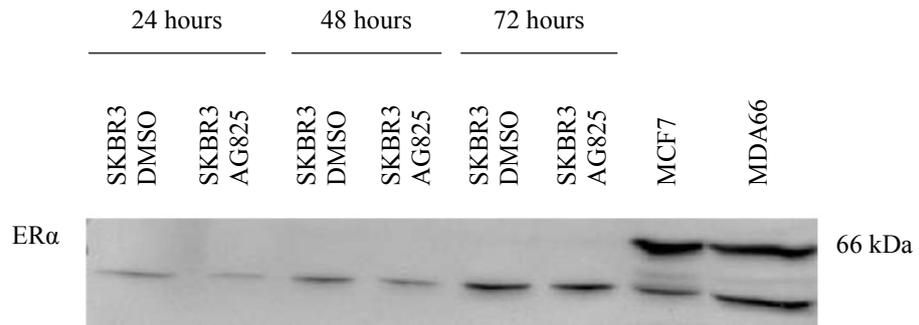
Figure 3.19: ER$\alpha$ expression in total lysates of ERBB2 inhibitor drug treated and DMSO added as negative control SKBR3 breast cancer derived cell lines detected by Western Blotting. MCF7 and MDA66 cell lines were used as positive control for ER$\alpha$ expression.

In summary, while we detected a negative correlation between ERBB2 and ARID3B expression in 2 out 3 cell lines, ERBB2 positive and ARID3B positive breast cancer cell lines and patients exist. It will be of future interest to investigate the relationship between ARID3B and ERBB2.

# CHAPTER 4

# CONCLUSION

The aim of this study was to investigate expressional characteristics of ARID3B in breast cancer.

ARID3B has several isoforms according to literature and bioinformatics analysis was performed to outline all isoforms including their genomic intron, exon boundaries and protein sizes among isoforms. As a result of this bioinformatics analysis, isoforms were visualized by schematic representation for the first time in literature. In addition to this, there is also pseudogene presence of ARID3B. The expressional analysis within breast cancer cell lines was also performed to see whether the pseudogene was expressed. Based on our primers, there was no transcript in breast cancer cell lines.

Next, we wanted to investigate the effect of ARID3B overexpression MCF10A and MDA-MB-231 breast cancer cell lines transfected with pcDNA-ARID3B construct. However, we failed to generate an overexpression model. There may be some compensatory mechanisms to hinder overexpression due to intrinsic feedback mechanisms. It is also possible that viral induction of ARID3B expression may yield better results.

Then, in the light of immunohistochemistry results of patient samples of another study from our laboratory, ARID3B protein expression was tested using different breast cancer cell lines. According to immunohistochemistry results, a potential relationship between ER positivity and nuclear ARID3B expression was detected. ER (+) and ER (-) breast cancer cell lines were used to check ARID3B protein expression levels. As a result, ER (+) cell lines such as MCF7, MDA 66 showed up to two fold stronger

61 kDa nuclear protein expression compared to ER (-) cell lines like, MCF10A and MDA-MB-231. This finding correlates with the immunohistochemistry results with patient samples.

Next, possible relationship between ERBB2 and ARID3B expression was investigated. Immunohistochemistry results revealed a 63% correlation between ARID3B expression and ERBB2 negativity. In parallel with the patient samples results, it was detected that 2 out of 3 ERBB2 positive cell lines lacked ARID3B full length expression, while SKBR3 (ERBB2 positive cell line) had expression of ARID3B full length.

Considering a possible crosstalk between ERBB2 and ER$\alpha$ we investigated whether any outcome on ARID3B expression and presence and/or absence of ERBB2 can be observed. To support this hypothesis, ERBB2 was inhibited in cell line SKBR3 (ER$\alpha$ negative, ERBB2 positive). As a result of this experiment, AG825 drug treated lysates showed decreased ARID3B full length protein expression. Interestingly, it was observed that ARID3B full length protein expression was restored in 72 hours compared to 48 hours and 24 hours. We think this may be due to several reasons, endogenous activation or inactivation other signalling cascades may result with re-activation of ERBB2 or because we kept the drug for 72 hours, it may have lost its effect.

Given the potential cross talk between ERBB2 and ER, ER$\alpha$ activation was checked in AG825 treated SKBR3 cell lines. No reactivation of ER$\alpha$ was detected. Effects of ARID3B expression loss upon inhibition of ERBB2 in SKBR3 cells and the fact that ERBB2 positive other two cell lines had ARID3B expression, the connection between ERBB2 and ARID3B seems complex and needs further investigation.

It would be interesting to extend the findings by checking protein levels of ARID3B isoforms besides full length isoform.

# REFERENCES

[1] S. Akhavantabasi, A. Sapmaz, S. Tuna, and A. E. Erson-Bensan. miR-125b targets ARID3B in breast cancer cells. *Cell Structure and Function*, 37(1):27–38, 2012.

[2] B. H. Akman, T. Can, and A. E. Erson-Bensan. Estrogen-induced upregulation and 3'-UTR shortening of CDC6. *Nucleic Acids Res.*, 40(21):10679–10688, Nov 2012.

[3] C. C. Alves, F. Carneiro, H. Hoefler, and K.-F. Becker. Relative malignant potential of human breast carcinoma cell lines established from pleural effusions and a brain metastasis. *Frontiers in Bioscience*, 14:3041–3050, 2009.

[4] S. W. Chan and W. Hong. Retinoblastoma-binding protein 2 (Rbp2) potentiates nuclear hormone receptor-mediated transcription. *Journal of Biological Chemistry*, 276(30):28402–28412, 2001.

[5] K. D. Cowden Dahl, R. Dahl, J. N. Kruichak, and L. G. Hudson. The epidermal growth factor receptor responsive miR-125a represses mesenchymal morphology in ovarian cancer cells. *Neoplasia*, 11(11):1208–1215, Nov 2009.

[6] S. Diederichs, N. Bäumer, P. Ji, S. K. Metzelder, G. E. Idos, T. Cauvet, W. Wang, M. Möller, S. Pierschalski, J. Gromoll, M. G. Schrader, H. P. Koeffler, W. E. Berdel, H. Serve, and C. Müller-Tidow. Identification of interaction partners and substrates of the Cyclin A1-CDK2 complex. *Journal of Biological Chemistry*, 279(32):33727–33741, 2004.

[7] S. Fleige, V. Walf, S. Huch, C. Prgomet, J. Sehm, and M. Pfaffl. Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR. *Biotechnology Letters*, 28(19):1601–1613, 2006.

[8] N. Fomproix and P. Percipalle. An actin–myosin complex on actively transcribing genes. *Experimental Cell Research*, 294(1):140 – 148, 2004.

[9] M. Giuliano, M. Trivedi, and R. Schiff. Bidirectional crosstalk between the estrogen receptor and human epidermal growth factor receptor 2 signaling pathways in breast cancer: Molecular basis and clinical implications. *Breast Care*, 8(4):256–262, 2013.

[10] S. L. Gregory, R. D. Kortschak, B. Kalionis, and R. Saint. Characterization of the dead ringer gene identifies a novel, highly conserved family of sequence-specific DNA-binding proteins. *Mol. Cell. Biol.*, 16(3):792–799, Mar 1996.

[11] R. F. Herrscher, M. H. Kaplan, D. L. Lelsz, C. Das, R. Scheuermann, and P. W. Tucker. The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes Dev.*, 9(24):3067–3082, Dec 1995.

[12] W. A. Hofmann, L. Stojiljkovic, B. Fuchsova, G. M. Vargas, E. Mavrommatis, V. Philimonenko, K. Kysela, J. A. Goodrich, J. L. Lessard, T. J. Hope, P. Hozak, and P. de Lanerolle. Actin is part of pre-initiation complexes and is necessary for transcription by RNA polymerase II. *Nat. Cell Biol.*, 6(11):1094–1101, Nov 2004.

[13] P. Hu, S. Wu, and N. Hernandez. A role for $\beta$-actin in RNA polymerase III transcription. *Genes & Development*, 18(24):3010–3015, 2004.

[14] T. Huang, T. Oka, T. Asai, T. Okada, B. W. Merrills, P. N. Gertson, R. H. Whitson, and K. Itakura. Repression by a differentiation-specific factor of the human cytomegalovirus enhancer. *Nucleic Acids Research*, 24(9):1695–1701, 1996.

[15] J. Iwahara, M. Iwahara, G. W. Daughdrill, J. Ford, and R. T. Clubb. The structure of the dead ringer–DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA. *The EMBO Journal*, 21(5):1197–1209, 2002.

[16] E. Johnson, D. D. Seachrist, C. M. DeLeon-Rodriguez, K. L. Lozada, J. Miedler, F. W. Abdul-Karim, and R. A. Keri. HER2/ErbB2-induced breast cancer cell migration and invasion require p120 catenin activation of Rac1 and Cdc42. *J. Biol. Chem.*, 285(38):29491–29501, Sep 2010.

[17] S. Joseph, V. E. Deneke, and K. D. Cowden Dahl. ARID3B induces tumor necrosis factor alpha mediated apoptosis while a novel ARID3B splice form does not induce cell death. *PLoS One*, 7(7):e42159, 07 2012.

[18] K. Kobayashi, T. Era, A. Takebe, L. M. Jakt, and S. Nishikawa. ARID3B induces malignant transformation of mouse embryonic fibroblasts and is strongly associated with malignant neuroblastoma. *Cancer Res.*, 66(17):8331–8336, Sep 2006.

[19] G. Konecny, G. Pauletti, M. Pegram, M. Untch, S. Dandekar, Z. Aguilar, C. Wilson, H. M. Rong, I. Bauerfeind, M. Felber, H. J. Wang, M. Beryt, R. Seshadri, H. Hepp, and D. J. Slamon. Quantitative association between HER-2/neu and steroid hormone receptors in hormone receptor-positive primary breast cancer. *J. Natl. Cancer Inst.*, 95(2):142–153, Jan 2003.

[20] R. D. Kortschak, P. W. Tucker, and R. Saint. ARID proteins come in from the desert. *Trends Biochem. Sci.*, 25(6):294–299, Jun 2000.

[21] A. Kukalev, Y. Nord, C. Palmberg, T. Bergman, and P. Percipalle. Actin and hnRNP U cooperate for productive transcription by RNA polymerase II. *Nat. Struct. Mol. Biol.*, 12(3):238–244, Mar 2005.

[22] M. H. Lahoud, S. Ristevski, D. J. Venter, L. S. Jermiin, I. Bertoncello, S. Zavarsek, S. Hasthorpe, J. Drago, D. de Kretser, P. J. Hertzog, and I. Kola. Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. *Genome Research*, 11(8):1327–1334, 2001.

[23] L. Mohrmann, K. Langenberg, J. Krijgsveld, A. J. Kal, A. J. R. Heck, and C. P. Verrijzer. Differential targeting of two distinct SWI/SNF-related drosophila chromatin-remodeling complexes. *Molecular and Cellular Biology*, 24(8):3077–3088, 2004.

[24] T. Müller, C. G. Concannon, M. W. Ward, C. M. Walsh, A. L. Tirniceriu, F. Tribl, D. Kögel, J. H. Prehn, and R. Egensperger. Modulation of gene expression and cytoskeletal dynamics by the amyloid precursor protein intracellular domain (AICD). *Molecular Biology of the Cell*, 18(1):201–210, 2007.

[25] S.-i. Numata, P. P. Claudio, C. Dean, A. Giordano, and C. M. Croce. Bdp, a new member of a family of DNA-binding proteins, associates with the retinoblastoma gene product. *Cancer Research*, 59(15):3741–3747, 1999.

[26] I. A. Olave, S. L. Reck-Peterson, and G. R. Crabtree. Nuclear actin and actin-related proteins in chromatin remodeling. *Annual Review of Biochemistry*, 71(1):755–781, 2002. PMID: 12045110.

[27] A. Patsialou, D. Wilsker, and E. Moran. DNA-binding properties of ARID family proteins. *Nucleic Acids Research*, 33(1):66–80, 2005.

[28] V. V. Philimonenko, J. Zhao, S. Iben, H. Dingova, K. Kysela, M. Kahle, H. Zentgraf, W. A. Hofmann, P. de Lanerolle, P. Hozak, and I. Grummt. Nuclear actin and myosin I are required for RNA polymerase I transcription. *Nat. Cell Biol.*, 6(12):1165–1172, Dec 2004.

[29] M. Riaz, F. Elstrodt, A. Hollestelle, A. Dehghan, J. Klijn, and M. Schutte. Low-risk susceptibility alleles in 40 human breast cancer cell lines. *BMC Cancer*, 9(1):236, 2009.

[30] J. Sambrook and D. W. Russell. *Molecular Cloning: A Laboratory Manual*. Number v. 1 in Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, third edition, 2001.

[31] S. J. Samyesudhas, L. Roy, and K. D. C. Dahl. Differential expression of {ARID3B} in normal adult tissue and carcinomas. *Gene*, 543(1):174 – 180, 2014.

[32] A. Takebe, T. Era, M. Okada, L. M. Jakt, Y. Kuroda, and S.-I. Nishikawa. Microarray analysis of PDGFR$\alpha$+ populations in {ES} cell differentiation culture

identifies genes involved in differentiation of mesoderm and mesenchyme including {ARID3b} that is essential for development of embryonic mesenchymal cells. *Developmental Biology*, 293(1):25 – 37, 2006.

[33] J. A. Tidwell, C. Schmidt, P. Heaton, V. Wilson, and P. W. Tucker. Characterization of a new ARID family transcription factor (Brightlike/ARID3C) that co-activates Bright/ARID3A-mediated immunoglobulin gene transcription. *Molecular Immunology*, 49(1–2):260–272, 2011.

[34] C. M. Tsai, A. Levitzki, L. H. Wu, K. T. Chang, C. C. Cheng, A. Gazit, and R. P. Perng. Enhancement of chemosensitivity by tyrphostin AG825 in high-p185(neu) expressing non-small cell lung cancer cells. *Cancer Res.*, 56(5):1068–1074, Mar 1996.

[35] X. Wang, N. G. Nagl, S. Flowers, D. Zweitzig, P. B. Dallas, and E. Moran. Expression of p270 (ARID1A), a component of human SWI/SNF complexes, in human tumors. *International Journal of Cancer*, 112(4):636–642, 2004.

[36] D. Wilsker, A. Patsialou, P. B. Dallas, and E. Moran. ARID proteins: A diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. *Cell Growth Differ*, 13(3):95–106, 2002.

[37] D. Wilsker, L. Probst, H. M. Wain, L. Maltais, P. W. Tucker, and E. Moran. Nomenclature of the ARID family of DNA-binding proteins. *Genomics*, 86(2):242 – 251, 2005.

[38] R. D. Zhang, I. J. Fidler, and J. E. Price. Relative malignant potential of human breast carcinoma cell lines established from pleural effusions and a brain metastasis. *Invasion and Metastasis*, 11(4):204–15, 1991.

# APPENDIX A

# MAMMALIAN CULTURE PROPERTIES

The table below shows the breast cancer cell lines properties in regards to; source, tumor type, tumor subtype, ER-PR-ERBB2 gene status, used in this study [38, 29].

Table A.1: Properties of Cell Lines

| Cell Lines | Subtype | ER Status | PR Status | ERBB2 Status | Source | Tumor Type |
|---|---|---|---|---|---|---|
| MDA-MB-231 | Basal B | - | - | - | Pleural Effusion | Metastatic Adenocarcinoma |
| MDA66 | Basal B | + | - | - | Pleural Effusion | Metastatic Adenocarcinoma |
| MCF10A | Basal B | - | - | - | Reduction Mammoplasty | Fibrocystic Disease |
| MCF7 | Luminal | + | + | - | Pleural Effusion | Metastatic Adenocarcinoma |
| MDA-MB-361 | Luminal | + | + | + | Brain metastasis | Metastatic Adenocarcinoma |
| SKBR3 | Luminal | - | - | + | Pleural Effusion | Metastatic Adenocarcinoma |
| BT474 | Luminal | + | + | + | Primary | Ductal Carcinoma |

# APPENDIX B

# BACTERIAL CULTURE MEDIUM

**LB**

Yeast Extract 5 g

Tryptone 10 g

NaCl 10 g

1N NaOH 1mL

All the components are mixed and total volume is completed to 1 L with $dH_2O$, pH is adjusted to 7.4 and autoclaved for sterilization.

**LB Agar**

Yeast Extract 5 g

Tryptone 10 g

NaCl 10 g

1N NaOH 1 mL

Agar 15 g

All the components are mixed and total volume is completed to 1 L with $dH_2O$, pH is adjusted to 7.4 and autoclaved for sterilization.

**IPTG Stock Solution (0.1M)**

1.2 g IPTG

Add water up to 50 mL final volume, sterilize by filtration and store at $4\,^{\circ}\mathrm{C}$.

**X-Gal (2 mL)**

100 mg 5-bromo-4-chloro-3-inodyl-$\beta$-galactoside

Dissolve in 2 mL N,N'dimethyl-formamide. Cover with aluminum foil and store at $-20\,^{\circ}\mathrm{C}$.

**SOC Medium**

2.0 g Tryptone

0.5 g Yeast Extract

1 mL 1M NaCl

0.25 mL 1M KCl

1 mL 2M $Mg^{2+}$ (final 20 mM)

1 mL 2M glucose (final 20 mM)

Add tryptone, yeast extract, NaCl, KCl to 97 mL distilled water, autoclave and cool down. Add 2M $Mg^{2+}$ and 2M glucose (final 20mM), make it 100 mL with sterile distilled water, the pH of the solution should be 7.0.

# APPENDIX C

# PRIMERS AND PCR CONDITIONS

Table C.1: Primers and PCR Conditions

| Primer | Sequence | Annealing temperature |
|---|---|---|
| GAPDH | Forward:<br>5'- GGGAGCCAAAAGGGTCATCA-3' | 56 °C |
| | Reverse:<br>5'- TTTCTAGACGGCAGGTCAGGT -3' | |
| ARID3B-F1 | 5'ATAAGAATGCGGCCGCAAAATGGAGCCACTTCAG-3' | 72 °C |
| ARID3B-F2 | 5'- TGCCAAGCTGTATGAACTGG-3' | 57.3 °C |
| ARID3B-F3 | 5'-TCTTTGGCTACTCACCTGCT-3' | 56 °C |
| ARID3B-R | 5'-GCGGATCCTCAGAGGGACCAGCTGGTG-3' | 72.6 °C |
| ARID3B-R1 | 5'-ATCCCCAGCAACCACTCTCA-3' | 59.5 °C |
| ARID3B-Psuedogene_F | 5'-TGCCTTCACCCTCAGCTCA-3' | 58.8 °C |

# APPENDIX D

# VECTOR MAPS



Figure D.1: The map of pcDNA 3.1 (-) 5.4 kb (Life Technologies)

| T7 | ↓ | 1 start |
|---|---|---|
| ApaI | | 14 |
| AatII | | 20 |
| SphI | | 26 |
| BstZI | | 31 |
| NcoI | | 37 |
| BstZI | | 43 |
| NotI | | 43 |
| SacII | | 49 |
| EcoRI | | 52 |

| SpeI | 64 |
|---|---|
| EcoRI | 70 |
| NotI | 77 |
| BstZI | 77 |
| PstI | 88 |
| SaII | 90 |
| NdeI | 97 |
| SacI | 109 |
| BstXI | 118 |
| NsiI | 127 |
| | 141 |

↑ SP6

Figure D.2: Map of P-GEMT Easy 3.015 kb (Promega)

# APPENDIX E

# PSEUDOGENE SEQUENCE COMPARISON WITH ARID3B

# FULL LENGTH

Table E.1: BLAST results of ARID3B pseudogene and mRNA sequences

```
Query   17    GCCCCGGCTGTGGCCCCCGGCTGCGGAGGAGTCCGAGACGCAGCTGCCGCGCCGGGGCCT   76
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Sbjct   1     GCCCCGGCTGTGGCCCCCGGCTGCGGAGGAGTCCGAGACGCAGCTGCCGCGCCTGGGCCT   60

Query   77    GAGCTGCCGCCTCCTCCGCCGCCCGAAAACCCGGAGTGCCCCGCACAGGTTTAGACCCAG   136
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   61    GAGCGGCCGCCTCCTCCGCCACCCGAAAACCCGGAGTGCCCCGCACAGGTTTAGACCCAG   120

Query   137   TGTGCCTGGTGGGCTGTGCCCAGGTTCAGAGTCATGCCACTCTGTGGGTGAAGCTTGAGG   196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   121   TGTGCCTGGTGGGCTGTGCCCAGGTTCAGAGTCATGCCACTCTGTGGGTGAAGCTTGAGG   180

Query   197   CAAAAATGGAGCCACTTcagcagcagcagcagcagcagcagcaacaacag----------   246
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Sbjct   181   CAAAAATGGAGCCACTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAACAGCAGCAGCAAC   240

Query   247   --aagcagcCACACCTGGCTCCTCTGCAGATGGATGCCAGAGAGAAGCAGGGCCAGCAGA   304
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   241   AGAAGCAGCCACACCTGGCTCCTCTGCAGATGGATGCCAGAGAGAAGCAGGGCCAGCAGA   300

Query   305   TGAGAGAAGCCCAGTTCTTGTATGCCCAAAAGCTGGTCACACAGCCGACTCTCCTTTCCG   364
              |||||||||||||||||||||||||||||||||||||||| |||||||||| |||||||
Sbjct   301   TGAGAGAAGCCCAGTTCTTGTATGCCCAAAAGCTGGTCACGCAGCCGACTCTCTTTTCCG   360

Query   365   CCACAGCTGGGAGACCTTCTGGCAGCACTCCCTTAGGTCCCTTAGCCAGAGTTCCACCCA   424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   361   CCACAGCTGGGAGACCTTCTGGCAGCACTCCCTTAGGTCCCTTAGCCAGAGTTCCACCCA   420

Query   425   CCGCAGCAGTGGCCCAAGTGTTTGAACGGGGCAACATGAACTCAGAGCCTGAGGAAGAGG   484
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   421   CCGCAGCAGTGGCCCAAGTGTTTGAACGGGGCAACATGAACTCAGAGCCTGAGGAAGAGG   480

Query   485   ACGGAGGTTTGGAAGATGAGGATGGGGATGATGAAGTTGCAGAGGTGGCTGAGAAAGAAA   544
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Sbjct   481   ACGGAGGTTTGGAAGATGAGGACGGGGATGATGAAGTTGCAGAGGTGGCTGAGAAAGAAA   540

Query   545   CCCAGGCTGCTTCAAAATATTTTCATGTGCAGAAAGTAGCTCGCCAAGATCCCAGAGTGG   604
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   541   CCCAGGCTGCTTCAAAATATTTTCATGTGCAGAAAGTAGCTCGCCAAGATCCCAGAGTGG   600

Query   605   CACCCATGTCCAATCTACTTCCAGCACCAGGGCTCCCACCACATGGACAACAAGCTAAAG   664
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   601   CACCCATGTCCAATCTACTTCCAGCACCAGGGCTCCCACCACATGGACAACAAGCTAAAG   660

Query   665   AAGACCATACCAAAGATGCTTCCAAGGCCTCACCTTCTGTCTCCACAGCAGGACAGCCGA   724
              ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||| ||
Sbjct   661   AAGACCATACCAAAGATGCTTCCAAGGCCTCATCTTCTGTCTCCACAGCAGGACAGCTGA   720

Query   725   ACTGGAATCTGGATGAGCAGCTCAAGCAGAATGGTGGTTTGGCCTGGAGTGATGATGCAG   784
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   721   ACTGGAATCTGGATGAGCAGCTCAAGCAGAATGGTGGTTTGGCCTGGAGTGATGATGCAG   780

Query   785   ATGGAGGCCGGGGAAGAGAGATCTCTCGAGATTTTGCCAAGCTGTATGAACTGGACGGTG   844
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   781   ATGGAGGCCGGGGAAGAGAGATCTCTCGAGATTTTGCCAAGCTGTATGAACTGGACGGTG   840

Query   845   ATCCTGAAAGGAAAGAGTTCCTGGATGACCTCTTCGTCTTTATGCAGAAGAGGGGGACCC   904
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Sbjct   841   ATCCTGAAAGGAAAGAGTTCCTGGATGACCTCTTCGTCTTTATGCAGAAGAGGCGGACCC   900
```

# Table E.1: BLAST results of ARID3B pseudogene and mRNA sequences (cont'd)

```
Query   905   CCATCAACCGAATCCCCATCATGGCCAAACAGATCCTGGACCTGTACATGCTGTATAAGC   964
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   901   CCATCAACCGAATCCCCATCATGGCCAAACAGATCCTGGACCTGTACATGCTGTATAAGC   960

Query   965   TGGTGACCGAGAAGGGAGGCCTGGTGGAGATCATCAACAAGAAGATCTGGAGGGAGATCA   1024
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Sbjct   961   TGGTGACCGAGAAGGGGGGCCTGGTGGAGATCATCAACAAGAAGATCTGGAGGGAGATCA   1020

Query   1025  CCAAAGGCCTAAACCTGCCCACATCCATCACCAGCGCCGCCTTCACCCTCA-----GGAC   1079
              |||||||||||||||||||||||||||||||||           ||||||||||    ||||
Sbjct   1021  CCAAAGGCCTAAACCTGCCCACATCCATCACCAGCGCTGCCTTCACCCTCAGCTCAGGAC   1080

Query   1080  GCAGTACATGAAGTATCTGTATGCCTATGAGTGTGAGAAGAAAGCCTTGAGTTCCCCAGC   1139
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1081  GCAGTACATGAAGTATCTGTATGCCTATGAGTGTGAGAAGAAAGCCTTGAGTTCCCCAGC   1140

Query   1140  CGAGCTCCAGGCAGCAATTGATGGCAACCGCAGGGAGGGCCGGCGGCCCAGCTACAGCTC   1199
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Sbjct   1141  CGAGCTCCAGGCAGCAATTGATGGCAACCGCAGGGAGAGCCGGCGGCCCAGCTACAGCTC   1200

Query   1200  CTCCCTCTTTGGCTACTCACCTGCTGCGGCTACTGCTGCTGCCGCTGCCGGGGCCCCTGC   1259
              ||||||||||||||||||||||||| |||||||||||||||||||||||| |||||||||
Sbjct   1201  CTCCCTCTTTGGCTACTCACCTGCTGCGGCTACTGCTGCTGCCGCTGCCGAGGCCCCTGC   1260

Query   1260  CCTTCTCTCCCCACCCAAGATCCGCTTTCCCATCCTTGGGCTTGGCTCCAGCAGTGGTAC   1319
              |||  || |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1261  CCTCCTCTCCCCACCCAAGATCCGCTTTCCCATCCTTGGGCTTGGCTCCAGCAGTGGCAC   1320

Query   1320  CAATACCAGTAGCCCTCGGATATCCCCAGCAACCACTCTCAGGAAAGGTGATGGAGCCCC   1379
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Sbjct   1321  CAATACCAGTAGCCCTCGAATATCCCCAGCAACCACTCTCAGGAAAGGTGATGGAGCCCC   1380

Query   1380  AGTGACAACAGTGCCTGTGCCAAATCGTCTGGCTGTGCCCGTGACCTTGGCAAGCCAGCA   1439
              ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1381  AGTGACAACAGTGGCTGTGCCAAATCGTCTGGCTGTGCCCGTGACCTTGGCAAGCCAGCA   1440

Query   1440  GGCTGGTACTCGGACCGCCGCACTGGAGCAGCTGCGGGAGCGGCTGGAGTCAGGGGAGCC   1499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1441  GGCTGGTACTCGGACCGCCGCACTGGAGCAGCTGCGGGAGCGGCTGGAGTCAGGGGAGCC   1500

Query   1500  TGCTGAGAAGAAGGCATCGAGGCTGTCTGAGGAGGAGCAGCGCCTGGTGCAGCAGGCCTT   1559
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1501  TGCTGAGAAGAAGGCGTCGAGGCTGTCTGAGGAGGAGCAGCGCCTGGTGCAGCAGGCCTT   1560

Query   1560  CCAGCGCAACTTTTTCAGCATGGCACGGCAGCTCCCCATGAAGATCAGGATCAACGGCAG   1619
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1561  CCAGCGCAACTTTTTCAGCATGGCACGGCAGCTCCCCATGAAGATCAGGATCAACGGCAG   1620

Query   1620  GGAAGACAGAGCAGAGGCCTCGGCTGCAGCACTGAACCTGACCACGAGTAGCATTGGGAG   1679
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1621  GGAAGACAGAGCAGAGGCCTCGGCTGCAGCACTGAACCTGACCACGAGTAGCATTGGGAG   1680

Query   1680  CATTAACATGTCTGTGGACATCGATGGCACCACCTATGCAGGTGTGCTGTTTGCCCAGAA   1739
              |  ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Sbjct   1681  CGTTAACATGTCTGTGGACATCGATGGCACCACCTACGCAGGTGTGCTGTTTGCCCAGAA   1740

Query   1740  GCCTGTGGTCCACCTCATCACGGGGTCTGCTCCCCAGAGCCTCGGcagcagcgccagcag   1799
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1741  GCCTGTGGTCCACCTCATCACGGGGTCTGCTCCCCAGAGCCTCGGCAGCAGCGCCAGCAG   1800

Query   1800  cagcagcagcTCTCACTGTTCACCAAGTCCTACCTCATCCCGGGGCACCCCCAGCGCAGA   1859
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1801  CAGCAGCAGCTCTCACTGTTCACCAAGTCCTACCTCATCCCGGGGCACCCCCAGCGCAGA   1860

Query   1860  GCCCTCCACCAGCTGGTCCCTCTGATGGGCAGGACCCAGCTTCCCACTTGCCACTCTCCT   1919
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Sbjct   1861  GCCCTCCACCAGCTGGTCCCTCTGATGGGCAGGACCCAGCTTTCCACTTGCCACTCTCCA   1920

Query   1920  GTCGAGAGTGAAGGAAGTTGATGCACAGAATTTACCTCATCTCACAGAGCCCACGTCGAC   1979
              ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Sbjct   1921  GTCGAGAGTGAAGGAAGTTGATGCACAGAATTTACCTCATCTCACGGAGCCCACGTCGAC   1980

Query   1980  GAGCACCATTTGGCCAGACATTGAGAGTTTGGACGTGTCCGTCTGTCCAGGCTCCATTCA   2039
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   1981  GAGCACCATTTGGCCAGACATTGAGAGTTTGGACGTGTCCGTCTGTCCAGGCTCCATTCA   2040

Query   2040  GGTCCTGCTGTACTCTGGGGGCAGGGAGAGGCTAGAGGGGCAGGACAGACAGCGTTGTCC   2099
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2041  GGTCCTGCTGTACTCCGGGGGCAGGGAGAGGCTAGAGGGGCAGGACAGACAGCGTTGTCC   2100

Query   2100  AATCAGGGATTGTCCTGGAGAACTGGGTGGGGGTCTGTGGGTGTCCAGCTTCCTCCAGGG   2159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   2101  AATCAGGGATTGTCCTGGAGAACTGGGTGGGGGTCTGTGGGTGTCCAGCTTCCTCCAGGG   2160

Query   2160  TTCCCCAGCCACCTCCCAGCTCAGGGCACAGTGTATCGACAATCTGTCAGCCGACGCAGG   2219
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||  ||||
Sbjct   2161  TTCCCCAGCCACCTCCCAGCTCGGGGCACAGTGTATCGACAATCTGTCAGCCGACACAGG   2220
```

```
Query  2220  GCTGGAGGCCCTCCATTTCCCTCCCCTTTTAATTTATTTCCCTGACCCTGCCTTCTGCCA  2279
             |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Sbjct  2221  GCTGGAGGCCCTCCATTTCCCTCCCCTTTTAATTTATTTCCCTGCCCCTGCCTTCTGCCA  2280

Query  2280  TGACCCCAGGGCCACTAGTCTCTTCCCCTGTTTTTAAGTCCCATGTGGGGCTGCCAGGAG  2339
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2281  TGACCCCAGGGCCACTAGTCTCTTCCCCTGTTTTTAAGTCCCATGTGGGGCTGCCAGGAG  2340

Query  2340  CCAAGAGCATGCCTTCATGCCCTTCTCTGCAGAGCATGGGATGGGGCCCTTCCAGCCCAC  2399
             |||||||||||||||||||||||||||||||| |||||| || ||| |||||||||||||
Sbjct  2341  CCAAGAGCATGCCTTCATGCCCTTCTCTGCTGAGCAT-GGTTGGTGCCCTTCCAGCCCAC  2399

Query  2400  TTTGCCTACTGTTAGGTTCCTGGGGTACGGGTCCCTCAAACACAAGTCTTGAATCAGTCC  2459
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2400  TTTGCCTACTGTTAGGTTCCTGGGGTACGGGTCCCTCAAACACAAGTCTTGAATCAGTCC  2459

Query  2460  TCAGGGCCCCAGATCCCCCTTCTAATAGTCTGAAGGAAGAGGCTGTGAGGGCAGTTCAGG  2519
             |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Sbjct  2460  TCAGGGCCCCAGATCCCCCTTCTAATGGTCTGAAGGAAGAGGCTGTGAGGGCAGTTCAGG  2519

Query  2520  GCCAGGGCCCACAGGGCATCTGCTGATGGGAAGGCAGAGCCTCGGGGCTGCCCAGCCCTG  2579
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2520  GCCAGGGCCCACAGGGCATCTGCTGATGGGAAGGCAGAGCCTCGGGGCTGCCCAGCCCTG  2579

Query  2580  GCACCTGCTTCACACGGACCATTCTCTGACCGCCCCCGCCAAACCTCATGCTCCCCACTT  2639
             |||||||||||||||||||||||||||||| ||||||| |||||||||||||||||| ||
Sbjct  2580  GCACCTGCTTCACACGGACCATTCTCTGGCCGCCCCC-CCAAACCTCATGCTCCCCGCTT  2638

Query  2640  GCCCTTATTGGCCTGCTTTCCCAAAAACAGCCCACTCAtcctttccccagccctcttcag  2699
             ||||||||| |||||||||||||||||||||||||||| |||||||||||||||||||||
Sbjct  2639  GCCCTTATTGTCCTGCTTTCCCAAAAACAGCCCACTCATCCTTTCCCCAGCCCTCTTCAG  2698

Query  2700  ccctccccagtcctccccagccttcttcagccttcttcagc-----tcagccctcttctA  2754
             ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Sbjct  2699  CCCTCCCCAGTCCTCCCCAGCCTTCTTCAGCCTTCTTCAGCCTTCTTCAGCCCTCTTCTA  2758

Query  2755  TTTCTTATGGGGCCATCTTGGCCTCACCATCCCCACACGTGCCGATGTCAGGTTCATTGA  2814
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2759  TTTCTTATGGGGCCATCTTGGCCTCACCATCCCCACACGTGCCGATGTCAGGTTCATTGA  2818

Query  2815  AATACCTCAGATTTGTAATTGTTGATGTTTGAGTCTTCAGGAAGTATTTGCATCTCTGAA  2874
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2819  AATACCTCAGATTTGTAATTGTTGATGTTTGAGTCTTCAGGAAGTATTTGCATCTCTGAA  2878

Query  2875  ATCttttttatatgtgttttgctgactttgtgttttttttttattttaaaattttATCGT  2934
             |||||||| ||||||||||||||||| |||||||||||| ||||||||||||| ||| ||
Sbjct  2879  ATCTTTTTTATATGTGTTTTGCTGAC-----TTTTTTTTTTATTTTAAAATTTTTATTGT  2933

Query  2935  AG--CACCAAAGAAATGAAAACAGATCACCCCCAAAGCCAAGCAAGTGATTGGGTAACCC  2992
             |  |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Sbjct  2934  TGTACACCAAAGAAATGAAAACAGATCACCCCCAAAGCCAAGCAAGTGATTGGGTACCCC  2993

Query  2993  GGCCCCTCTGGCCCTTCCCTCAAGCCCAGTGAGGAGGTGGGTGGAGGCATGCAGGGGAGA  3052
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2994  GGCCCCTCTGGCCCTTCCCTCAAGCCCAGTGAGGAGGTGGGTGGAGGCATGCAGGGGAGA  3053

Query  3053  CGACTCAGGGATCGCGGGGGCGGGGAGGTGGAGTGGAGAGGCCCGCTCTCACACCGGGTC  3112
             |||||||||||||| |||||||||||||||||||||||||||||||||||||| |||||
Sbjct  3054  CGACTCAGGGATCACGGGGGCGGGGAGGTGGAGTGGAGAGGCCCGCTCTCACACTGGGTC  3113

Query  3113  GGGCCCTGAAACTGTACCAGTTCTGAAGACCGTTTTCTGCCACACCCCCACCCCCAGCTG  3172
             |||||||||||| |||||||||||||||| |||||||||||||||||||||||||||||
Sbjct  3114  GGGCCCTGAAACCGTACCAGTTCTGAAGAGCGTTTTCTGCCACACCCCCACCCCCAGCTG  3173

Query  3173  ACCAGGGAGGCAGAGGACACCTCCCCATCACTCAGCATTCCCAGCTCAGGGCACAGGCAG  3232
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3174  ACCAGGGAGGCAGAGGACACCTCCCCATCACTCAGCATTCCCAGCTCAGGGCACAGGCAG  3233

Query  3233  GCTGGGGTCCTCGCTGGCCTTGCCAGAGGTTGAGCCGCCCCAGGGTATGAGGAGATGAAT  3292
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3234  GCTGGGGTCCTCGCTGGCCTTGCCAGAGGTTGAGCCGCCCCAGGGTATGAGGAGATGAAT  3293

Query  3293  AACTCCACAGCTCCTCCTGGACCCTGCGCGGGAGCAGGCAGCTCCTGTGCTGTAAAGAAA  3352
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3294  AACTCCACAGCTCCTCCTGGACCCTGCGCGGGAGCAGGCGGCTCCTGTGCTGTAAAGAAA  3353

Query  3353  ATTGATTCGCTTGCGGCTCACCTCAGTCGAGGAAGCCCTGGAATGTTCAGCAGAACACCA  3412
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3354  ATTGATTCGCTTGCGGCTCACCTCAGTCGAGGAAGCCCTGGAATGTTCAGCAGAACACCA  3413

Query  3413  CCACTGTGACATGGGGCTGTGGCAGTGGGAGACACCGCGCGTGGCGTGGGTGTGTGTGGC  3472
             ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Sbjct  3414  CCACTGTGACATGGGGCTGTGGCAGTGGGAGACACCGCATGTGGCGTGGGTGTGTGTGGC  3473

Query  3473  AGCGTGGCTCGCATTCAGTCCTGTGTAGGAGAGGAAAGGGAATCAAAAGCTTGAGCACAA  3532
             ||||||||||| ||||||| |||||||||||||||||||||||||||||||||||||||
Sbjct  3474  GGCGTGGCTCGCACTCAGTCCTGTGTAGGAGAGGAAAGGGAATCAAAAGCTTGAGCACAA  3533
```

## Table E.1: BLAST results of ARID3B pseudogene and mRNA sequences (cont'd)

```
Query  3533  ACAGATACCTCAGCCCGGTAAGCTGCAGCTCTGCTCAACTGCGTCTTCTCTCAGCCCTCc  3592
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3534  ACAGATACCTCAGCCCGGTAAGCTGCAGCTCTGCTCAACTGCGTCTTCTCTCAGCCCTCC  3593

Query  3593  acacacactcacccccactcccacacacatacacacacacacacacaTACCCCATCTCCC  3652
             ||||||  |||||||||||||||||||||  ||||||||||||||  || ||||||||||
Sbjct  3594  ACACACGCTCACCCCCACTCCCACACACACACACACACACACACACACCATCTCCC      3653

Query  3653  GAGGGCTGACCTCCTCTGGCCTCAGCCTGCAGGGTGTGGGCAGAGAAGGGCATCTGGGAC  3712
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3654  GAGGGCTGACCTCCTCTGGCCTCAGCCTGCAGGGTGTGGGCAGAGAAGGGCATCTGGGAC  3713

Query  3713  GTGGTGCCAGTGAGGAGCCCAGTTGGCTGGCACTGGGCCCATTTGAAGGTGTCTCAGACA  3772
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3714  GTGGTGCCAGTGAGGAGCCCAGTTGGCTGGCACTGGGCCCATTTGAAGGTGTCTCAGACA  3773

Query  3773  TTTGGCCAGTATGTCTTTCTCAGGGGTTTGGTCACAAAGGATGGACTCTTCCCACCCAGA  3832
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3774  TTTGGCCAGTATGTCTTTCTCAGGGGTTTGGTCACAAAGGATGGACTCTTCCCACCCAGA  3833

Query  3833  GGATGCAGGGAAAGCACACTGTGTCTTTCCGGTCATTGGATCCTCTCCCTTTCCCCAGGC  3892
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3834  GGATGCAGGGAAAGCACACTGTGTCTTTCCGGTCATTGGATCCTCTCCCTTTCCCCAGGC  3893

Query  3893  AGCTCGCCTGGCCACACCGTTGGAGTGAACCCTCACTGCCCTCAAGGACAACAGCAGGGT  3952
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3894  AGCTCGCCTGGCCACACCGTTGGAGTGAACCCTCACTGCCCTCAAGGACAACAGCAGGGT  3953

Query  3953  GTCACCCAGAGCCCGATGAGGGGTGCCAGCAGGTGCCCCCACGAGTGGGGCCTTGGCCCA  4012
             ||||||||||||||||| |||||||| |||||||||||||||||||||||||||||||||
Sbjct  3954  GTCACCCAGAGCCGGATGAGGGATGCCAGCAGGTGCCCCCACGAGTGGGGCCTTGGCCCA  4013

Query  4013  AGCAGAGCTTCCCCTGAAGGTGCCATCAGCCAGGGCAGCTCTGTCCCCTCCTGCTCTCCA  4072
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4014  AGCAGAGCTTCCCCTGAAGGTGCCATCAGCCAGGGCAGCTCTGTCCCCTCCTGCTCTCCA  4073

Query  4073  TCTTATATGTATTCTAACCAGGAAAAATGTGATAGCACATGGGTAGCCTAGGCAGTGAGT  4132
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct  4074  TCTTATATGTATTCTAACCAGGAAAAATGTGATAGCACATGGGTAGCCTAGGCAGTGAAT  4133

Query  4133  AAATACCTCAGATGTCCTCCTGCA  4156
             ||||||||||||||||||||||||
Sbjct  4134  AAATACCTCAGATGTCCTCCTGCA  4157
```

# APPENDIX F
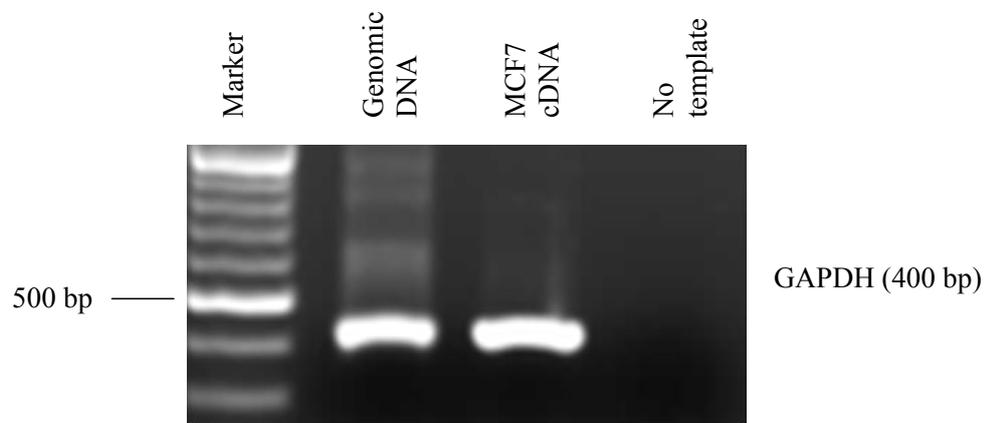
## CDNA SYNTHESIS CONFIRMATION



Figure F.1:  MCF7 cDNA synthesis PCR with GAPDH primers after DNase treatment.  M: DNA Ladder, Gene Ruler (Fermentas); Lane 1:  Genomic DNA, positive control; Lane 2: MCF7 cDNA; Lane 3: No template control

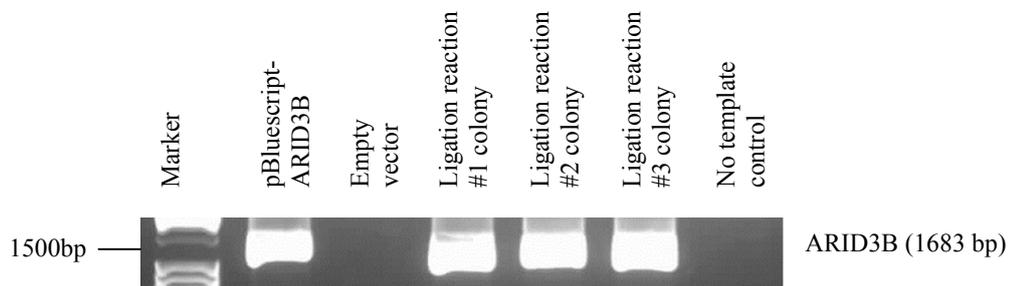# APPENDIX G

## COLONY PCR AFTER T4 DNA LIGASE



Figure G.1: Colony PCR using ARID3B CDS primers after T4 DNA Ligase Reaction. M: DNA Ladder, Gene Ruler (Fermentas); Lane 1: pBluescript-ARID3B, positive control; Lane 2: Empty vector, negative control; Lane 3: Ligation reaction #1 colony Lane 4: Ligation reaction #2 colony Lane 5: Ligation reaction #3 colony Lane 6: No template control