12-2021

V-Myb Avian Myel oblast Viral Oncogene Homolog Like 1 (MYBL1) Knockdown and its Role in a Triple Negative Breast Cancer

Nabras Mahmoud Abdulrahman

V-MYB AVIAN MYEL OBLAST VIRAL ONCOGENE HOMOLOG LIKE 1 (MYBL1) KNOCKDOWN AND ITS ROLE IN A TRIPLE NEGATIVE BREAST CANCER

THESIS

Presented in Partial Fulfillment of the Requirements for

the Degree Master of Science in the Graduate School

of Texas Southern University

By

Nabras Mahmoud Abdulrahman, A.S, B.S.

Texas Southern University

2021

Approved by

Dr. Audrey Player
Chairperson, Thesis Committee

Dr. Gregory Maddox
Dean, The Graduate School

Approved by

Dr. Audrey Player	10/29/2021
Chairperson, Thesis Committee	Date
Dr. Marian Hillar	10/29/2021
Committee Member	Date
Dr. Desirée Jackson	10/29/2021
Committee Member	Date
Dr. Aladdin Sleem	10/29/2021
Committee Member	Date

V-MYB AVIAN MYEL OBLAST VIRAL ONCOGENE HOMOLOG LIKE 1 (MYBL1) KNOCKDOWN AND ITS ROLE IN A TRIPLE NEGATIVE BREAST CANCER

By

Nabras Mahmoud Abdulrahman, M.S.

Texas Southern University, 2021

Professor Audrey Player, Ph.D., Advisor

Triple Negative Breast Cancer (TNBC) is defined as negative for three genes, estrogen receptor (ESR), progesterone receptor (PR) and Human epidermal growth factor receptor (HER2-neu) genes. Previous data show the V-Myb Avian Myel oblast Viral Oncogene Homolog Like 1 (MYBL1) gene is over-expressed in Triple negative breast cancer cell line (MDA-MB231). MYBL1 belongs to the MYB family of genes which are transcription factors and proto-oncogenes which are associated with cell cycle regulation, apoptosis, and differentiation, all of which are key events associated with cancers. It could be that MYBL1 contributes to these same processes in TNBC. Instead of studying MYBL1's contribution to several of these processes, we were mainly concerned with identifying genes that were either directly or indirectly affected by down-regulation of MYBL1 gene. Utilizing a gene silencing approach helps to identify genes that cooperate with MYBL1 in the signaling processes in cancer. Although the focus of our laboratory is TNBC, there are two parts to this current study, one that examines MYBL1 in luminal cancers cell line (MCF7) and one that examines MYBL1 in TNBC, designated Part 1 and Part 2, respectively.

For Part 1, we performed analyses of MCF7 (Luminal breast cancer cell line) receptor positive cells where estrogen receptor gene was silenced; and another MCF7 preparation where cMYB gene was silenced. Both datasets were obtained from Gene Expression Omnibus (GEO).

These datasets were chosen because even though they were neither TNBC or directly involved MYBL1 as the primary target, comparative analyses of both datasets showed MYBL1 knock-down (KD). We reasoned that even under these conditions, genes either directly or indirectly associated with MYBL1 might be identified. For Part 2 of this study, short hairpin RNA (shRNA) lentiviral transduction was used to down-regulate MYBL1 in MDA MB231 TNBC cells. A substantial number of reliable differentially expressed genes were identified here. Overall, genes recognized as associated with MYBL1 in the MCF7 luminal preparations (Part 1) were drastically different from genes identified as associated with MYBL1 in the TNBC KD study (Part 2). In both datasets, we identify novel genes that appear to be coordinately expressed with MYBL1 in breast cancers. This study led to identification of candidate genes that might be important towards the study of characterizing MYBL1 expression in TNBC. Two of these genes, transcription factor 19 (TCF19) and Kinesin-like protein (kinesin family member 18B) (KIF18B) have been experimentally validated. MYBL1 is a strong candidate gene to study for its contribution to the development of TNBC. Continued analyses of these genes and their relationship to MYBL1 should lead to a better understanding of signaling processes in breast cancers.

Keywords: Breast Cancer, Triple Negative, MYBL1 knock-down.

TABLE OF CONTENTS

	Page
LIST OF TABLES……....	iv
LIST OF FIGURES …….……	v
LIST OF ABBREVIATIONS…….....	vii
VITA……	xi
ACKNOWLEDGMENTS……	xii
CHAPTER	
1. INTRODUCTION……	1
2. LITERATURE REVIEW…….	10
3. DESIGN OF THIS STUDY…….....	22
4. RESULTS AND DISCUSSION…….....	30
5. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS……...	54
REFERENCES……..	56

LIST OF TABLES

Table		Page
1.	List of Primer Sequences for Genes Used in this Study……	24
2.	Initial List of Genes Identified as Common between the MCF7 ESR 1 shRNA Compared to MCF7 cMYB siRNA …….……	34
3.	MYBL1 shRNA Viral Particle Samples (sequences) Obtained from Origene for Testing (labeled A-D) ……	38
4.	Shows the Most Important Change that Affected by MYBL1 KD….....	43
5.	Original List of 13 Genes Analyzed for Pattern of Expression in Part 2 Conditions …….....	51
6.	Data from MYBL1 Knock-down in the MDA MB231 Cells E2F4 and TP53 Were Down-regulated …….....	52

LIST OF FIGURES

Figure		Page
1.	Summary of the Incidence of Breast Cancers Based on Molecular Signatures …….	4
2.	Diagram Demonstrating the Anatomical Location of the Progenitor Cells Associated with Luminal Cancer Compared to Basal-like TNBC…..	4
3.	Gorbatenko Data Analyses of Breast Cancer Subtypes for Expression of HER2, cMYB, MYBL1 and MYBL2 Expression in Patient Samples….	15
4.	Hierarchical Clustering (HC) Analyses to Demonstrate MYBL1 Expression in Patients (Maire dataset)……....	17
5.	Experiment Outline of Part 1 and Part 2 of this Current Study ….…….	17
6.	Experiment Outline of Part 2 of this Study..…..……....	20
7.	Summary of shRNA Procedure when MDA MB231 Cells were Transduced with hRNA.……	20
8.	The Affymetrix Microarray Genechip…….	21
9.	Hierarchical Cluster of Maire Patient Data-Set …….…….....	35
10.	Microarray Analyses of Select Genes in Player (2021) Cell Line Dataset. Plot of Transcript Levels……..	35
11.	PCR Performed to Experimentally Screen Select Genes.……	36
12.	Analyses of shRNA Scramble Sequence and MYBL1 shRNA Particles to Screen for Sequence that More Effectively Decreases MYBL1 Expression in MDA MB231 Cells ……	39
13.	Transcript and Protein Analyses of MYBL1 LVA shRNA Preparation to Demonstrate Knock-down Efficiency……..……	40
14.	Clarion Microarray Gene-chip Scatter Plot to Demonstrate Differential Gene Expression ……	42

15. Gene Ontology to Determine Gene Set Enrichment in Dataset……. 43

16. Final List of Candidate Genes Based on Differential Expression Analyses of MYBL1 shRNA Compared to Scrambled Control. …….. 44

17. STRING™ Analyses Demonstrating Significantly Relevant Relations Between MYBL1, MYBL2, TCF19 and KIF18b. …… 46

18. PCR and Western Experimental Gel Analysis of Candidate Genes Using Cell Line Samples…… 48

19. Analysis of Candidate Genes in Patient Samples……... 49

20. STRING™ Protein: Protein Interaction Determines a Relationship between E2F4, E2F7, MYBL1 and MYBL2 with TP53 in Cell Cycle Signaling ……..…… 52

LIST OF ABBREVIATIONS

(ATCC®)	American Type Culture Collection
AMV	Avian Myeloblastosis Virus
Anti-HER2	Antibody therapy targeted treatments
ATF3	Activating transcription factor 3
BCA	Bicinchoninic acid kit
BL1	Basal-like types 1
BL2	Basal-like types 2
CA	California
CCNB1	Cyclin B1
cDNA	Complementary DNA
DEG	Differentially Expressed Genes
DBD	DNA binding domain
dT	Dynabeads® Oligo
DCIS	Ductal Carcinoma in Situ
DDX58	DExD/H-Box Helicase 58
DMEM	Dulbecco's Modified Eagle Minimum essential media
dNTPs	Deoxyribonucleic triphosphate
DUSP7	Dual Specificity Phosphatase 7
E2F4	Elongation factor 4
E2F6	Elongation factor 6
E2F7	Elongation factor 7
ESR1	Estrogen Receptor
EtBr	Ethidium Bromide
FBXL13	F-Box and Leucine Rich Repeat Protein 13
FBXO8	F-Box Protein 8

FBXO32	F-Box Protein 32
G1	GAP 1 phase
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GEO	Gene Expression Omnibus
Her2	Human epidermal growth factor receptor
HRP	Horseradish peroxidase
IBC	Inflammatory Breast Cancer
IDC	Invasive Ductal Carcinoma
ILC	Invasive Lobular Carcinoma
IM	Immunomodulatory
KD	Knock-down
KIF18B	Kinesin-like protein (kinesin family member 18B)
Ki67	Nuclear protein Ki67
LCIS	Lobular Carcinoma in Situ
Luminal A	Hormone-receptor positive (estrogen-receptor positive, progesterone-receptor positive HER2 negative).
Luminal B	Hormone-receptor positive (estrogen-receptor positive, progesterone-receptor positive and HER2 positive).
LAR	Luminal androgen receptor
LIN9	LIN9-DREAM MuvB core complex component
LIN37	LIN37-DREAM MuvB core complex component
LIN52	LIN52 -DREAM MuvB core complex component
LIN54	LIN-54DREAM MuvB core complex component
Linoo673	Long Intergenic Non-Protein Coding RNA 673
MA	Massachusetts
mRNA	Messenger RNA
miRNAs	MicroRNA
MAF1	Repressor of RNA polymerase III transcription MAF1
M	Mesenchymal group

MSL	Mesenchymal stem-like () subtype
MCF7	Luminal breast cancer cell line
MDA-MB231	Triple negative breast cancer cell line
MCF10A	Receptor negative non-tumor cell lines
ul	microliter
ug	microgram
MOPs	3-(N-morpholino) propane sulfonic acid buffer
MYB	Proto-oncogene, transcription factor
MYBL1	V-Myb A vian Myel oblast Viral Oncogene Homolog Like 1
MYBL2	Myb-related protein B
NCBI	National Center for Biotechnology Information
NFIB	Nuclear Factor I B
NFE2L3	Nuclear Factor, Erythroid 2 Like 3
NE	Nebraska
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PR	Progesterone receptor
PRDM5	PR/SET Domain 5
PSPH	Phosphoserine Phosphatase
RANGAP1	Ran GTPase Activating Protein 1
RBBP4	Retinoblastoma Binding Protein 4
RBL1	Retinoblastoma-Like 1 Protein
RBL2	Retinoblastoma-Like 2 Protein
RIPA	Radioimmunoprecipitation assay
RMA	Robust Multi Array
RNA	Ribonucleic acid
RNAi	RNA interference
RPM	Revolutions per minute

SANT	Swi3, Ada2, N-Cor, and TFIIIB
shRNA	short hairpin RNA
siRNA	Small interfering RNA
SLC25A1	Solute Carrier Family 25 Member 1
S phase	Synthesis Phase
TAQ	Polymerase thermostable DNA polymerase I
TBE	Tris/Borate/EDTA
TCF19	Transcription factor 19
TNBC	Triple Negative Breast Cancer
TFDP2	Transcription Factor Dp protein 2
TFDP1	Transcription Factor Dp protein 1
TP53	Tumor protein p53
UTSW	University of Texas Southwest Core Facility
ZN75A	Zinc Finger Protein 75a
ZN436	Zinc Finger Protein 436

VITA

2016 …….……	A.S., Science Houston Community College Houston, Texas
2018 …….……	B.S., Biology Texas Southern University Houston, Texas
2019-2021……	Graduate Teaching Assistant Department of Biology College of Science, Engineering, and Technology Texas Southern University Houston, TX
2021……..……	Graduate Teaching Assistant Department of Chemistry College of Science, Engineering, and Technology Texas Southern University Houston, TX
Major Field……	Biology

ACKNOWLEDGEMENTS

All praises and thanks belong to ALLAH Almighty for giving me the capacity and persistence to complete this study. Nothing can be done without ALLAH's favor, support, and mercy. "I trust the journey because the Lord of the Worlds is the Best Planner", and every aspect of our existence is completely under his control!

I thank and dedicate all my works to my love, teacher, influencer, to that man who had a purest heart and a soul of angel, who is not here physically but always present in my thoughts and heart, to my father " Mahmoud Abdulrahman". To the father who was not a scholar or a minister, but a simple worker, who worked hard to see his children had obtained the best education, and from there my ambitious has been started. Thank you for being who you were a true, honest, strong, and kind simple man with great ideas. My father as I have promised, since the first day you left, to make you proud and glad in your grave. I'm renewing my promise to continue working more diligently towards my next chapter until achieve a milestone.

Endless thank and dedication to the woman whose her breath giving me the blessing and strength in this life, to my "Mother", no matter what words I say to thank her and how much I say it, it will never do her justice.

Many greetings and thanks to my backbone, my brothers and sisters for support and engagement. Special thanks to my special sister Noura Abdulrahman, I could not ask for a better sister, friend, and support, through our life together. Through the good and bad times, stick together as a team, and support one another.

I would like to thank my advisor Dr. Audrey Player for her time and guidance. I'm thankful to her to all knowledge that I gain throughout the entire thesis project. I express my gratitude for giving me a privilege as a first student who has started the project with her.

I would like to express my warmest gratitude to my committee members Dr. Aladdin Sleem, Dr. Desirée Jackson, and Dr. Marian Hillar for the consideration, time, and valuable comments. Thank you all wholeheartedly for your help and sincere sight.

I would like to extend my honest appreciation to my two exemplary professors Drs. Warren Williams and Daryl Wilkerson for all their numerous unforgettable support and kind feeling throughout whole my journey in both undergraduate and graduate levels.

CHAPTER 1

INTRODUCTION

BREAST CANCER TYPES:

The progression of adult-onset cancers is a long process involving genetic changes that happen over an extended period of time. These changes are specifically related to alterations in cell growth, programmed cell death and ultimately cellular differentiation so as to define the disease as a cancer (as opposed to some other disease such as diabetes or muscular dystrophy) [1]. For breast cancers, the initial and subsequent changes occur in cells associated with breast tissues. For many breast cancers the progenitor cell types have not been identified and characterized. But one can speculate that the progenitors and patterns leading to pathogenesis are different in each of different breast cancer types.

Breast and other cancers tend to be heterogenous and genetically complex. As evidence, a brief description of the 'types of breast cancers' are summarized below. Although there are many different types of breast cancer, most are rare (see below). As a general description, benign samples are in situ and non-life threatening, and the invasive cancers metastasize to distal organs and are thereby malignant. The recurring cancers recur at the same or different locations. Similar to other cancers, breast cancers are described and diagnosed based on pathological diagnoses and molecular subtype analyses [2] (https://www.breastcancer.org/symptoms/types). Cancers characterized based on their

molecular signatures were often identified using gene expression microarray studies. The microarray experimental platform has been instrumental in defining the gene expression profiles and identifying potential therapeutic targets useful for treating different cancers.

Detailed List of the Different Types of Breast Cancers Based on Pathological Diagnoses:

Ductal Carcinoma in Situ (DCIS): DCIS is a non-invasive carcinoma, that originates in the milk ducts of the breast. Although the cancers do not metastasize, DCIS patients have an increased risk of developing invasive cancers later.

Invasive Ductal Carcinoma (IDC): IDC is the common type of breast cancer, detected in ~80% of breast cancers. IDC is an invasive, infiltrating ductal cancers, occurring in the milk ducts of patients.

Tubular Carcinoma of the Breast: Tubular carcinomas are a subtype of the IDC cancers. They have a tube-shaped tubular configuration, but they also originate in the milk ducts. These tumors occur in 8%-27% of patients.

Medullary Carcinoma of the Breast: Medullary carcinomas are rare invasive ductal carcinomas. They appear as soft, fleshy masses resembling the brain's medulla. They occur in about 3% of patients.

Mucinous Carcinoma of the Breast: Mucinous Cancers are rare cancers that also begin in the milk ducts. The cancers are defined as abnormal cells 'floating in pools of mucin'.

Papillary Carcinoma of the Breast: Papillary cancers are also rare (~1-2%) invasive carcinomas. The cancers have well defined borders with small finger-like projections.

Cribriform Carcinoma of the Breast: Cribriform cancers are also rare invasive cancers. The cancers are low grade (i.e., look normal), but appear to have 'holes or display cribriform-like configurations.

Invasive Lobular Carcinoma (ILC): ILC cancers are invasive cancers that begin in the lobules of the breast. They are the second most common type of cancers occurring in ~20% of patients.

Inflammatory Breast Cancer (IBC): IBC is a rare very aggressive, invasive cancer. The cancers have the appearance of 'sheets' instead of lumps, which makes them difficult to treat and detect.

Lobular Carcinoma in Situ (LCIS): LCIS is defined as an abnormal cell growth in the lobules; often referred to as a neoplasia instead of carcinomas. The cancers remain in the breast but are diagnostic for future invasive cancers.

Male Breast Cancer: Breast cancer in men is rare, occurring at a rate of less than 1%. The cancers tend to be small, but they can also be invasive.

Paget's Disease of the Nipple: Paget's disease is a rare breast cancer involving cancer around the nipple and ducts draining towards the nipple.

Phyllodes Tumors of the Breast: Phyllodes cancers present as tumor cells that grow in a leaf-like configuration. The cancer occurs in less than 1% of breast cancer patients.

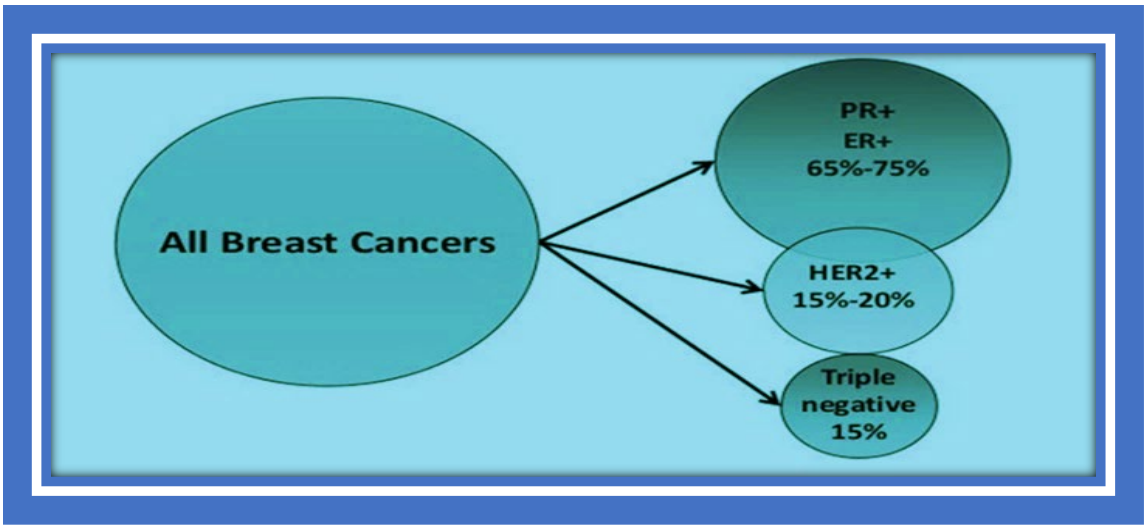


Figure 1: **Summary of the Incidence of Breast Cancers Based on Molecular Signatures (Burstein, Goldhirsch, St Gallen 2007)**

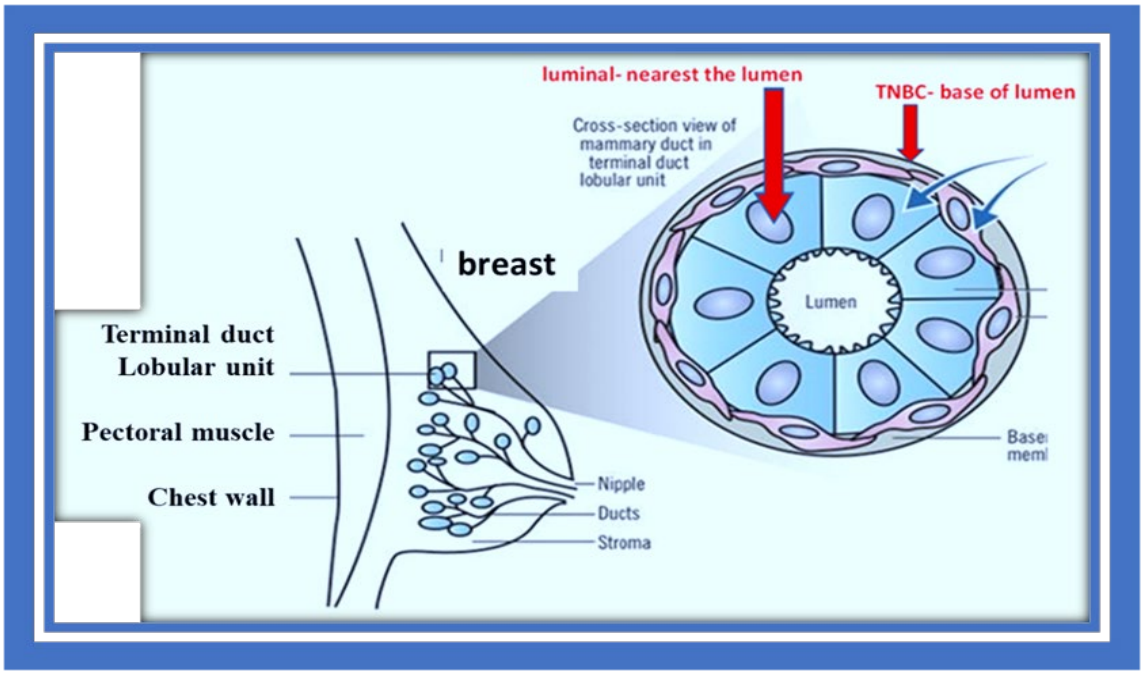


Figure 2: **Diagram Demonstrating the Anatomical Location of the Progenitor Cells Associated with Luminal Cancer Compared to Basal-like TNBC [3]**

List of Breast Cancer Types Based on Molecular Analyses:

The incidence and description of breast cancers characterized based on their molecular signatures are presented in Figure 1, and a more detailed description is given below [2] (https://www.breastcancer.org/symptoms/types). The more common breast cancer designations are often referred to as luminal or Triple negative/basal-like subtypes. An explanation describing the anatomical distinction between luminal compared to basal-like/triple negative is given in Figure 2 [3].

Molecular Subtypes of Breast Cancer (Luminal A/B; triple negative/basal-like; HER2-enriched; Normal-like): *Luminal A* cancers are defined by molecular analyses to be estrogen receptor and progesterone positive and HER2 negative. *Luminal B* are defined as estrogen receptor and progesterone positive and HER2 positive. The prognosis is slightly worse than Luminal A. *Triple negative/basal-like* are estrogen receptor, progesterone and HER2 negative. Although TNBC and basal-like are defined together, (and share substantial over-lap based on gene expression analyses), data show that there are molecular differences between the subtypes. A more detailed description of TNBC is given below. *HER2-enriched* cancers are estrogen receptor and progesterone negative and HER2 positive. *Normal-like* are similar to Luminal A estrogen receptor and progesterone positive and HER2 negative but have the appearance of low grade with low nuclear protein Ki67 (Ki67) levels.

Triple negative breast cancer: These cancers are estrogen receptor, progesterone negative and HER2 negative. They represent ~15% of breast cancers. They are not dependent on hormone regulation, have low 5-year survival, and lack suitable targeted therapies. Data suggest that their progenitor cells are associated with the basement region of the ducts, compared to luminal cancers which border the luminal breast regions (Figure 2). Even though TNBC are characterized based on receptor-negative status, Lehmann et al., [4] show that TNBCs can be further divided into six sub-categories based on molecular signatures and clustering patterns identified using microarray. These data emphasize the heterogeneity and complexity of the cancers. Lehmann grouped TNBC into cancers designated as Luminal androgen receptor (LAR) or molecular apocrine cancers, two basal-like types (BL1 and BL2), an immunomodulatory group, a mesenchymal (M) group and a mesenchymal stem-like (MSL) subtype. The LAR sub-category includes estrogen receptor samples with over-expression of androgen receptor, and other genes involved in hormonal regulation. The basal type 1 (BL1) includes over-expression of genes involved in pathways related to Ribonucleic acid (RNA) polymerase processes, cell cycle regulation and cell division. Samples characterized as Basal type 2 (BL2) genotype are thought to be of myoepithelial origin, including genes associated with growth factor signaling processes, gluconeogenesis, and glycolysis. Cells characterized in the immunomodulatory (IM) category appear similar to medullary breast cancers and are enriched in genes involved in immune signaling pathways, natural killer cell pathway, cytokine signaling, and antigen identification and processing. The M and MSL groups are enriched in genes associated with cell motility, proliferation, mesenchymal-like differentiation, extra-cellular matrix

proteins. Collectively, the complexity of the breast cancer subtype data explains (in part) why patients differ in their response to therapies.

Information Related to Breast Cancers Aligned with Our Research Goals:

There is a 3% chance that a woman will die from breast cancer and less than 1% chance for men. In 2020, an estimated 276,480 new cases of invasive breast cancer will be diagnosed in women and 2820 new cases in men (https://www.breastcancer.org). Of the 276,480 cases, approximately10-20% are receptor-negative, TNBC. Considering all breast cancer types, approximately 43,170 women and 522 men diagnosed with breast cancer are estimated to die in 2020. Conventional hormone therapy (e.g. tamoxifen) and anti-HER2 antibody therapy targeted treatments (e.g. trastuzumab) exist for receptor positive cancers, but because TNBC lack these targeted genes, therapies like tamoxifen and trastuzamab are not effective [5]. TNBC patients have few therapeutic options, having to basically rely upon chemotherapy and radiation therapies [6]. As a result, there is a need to further characterize TNBC and identify genes key to the tumorigenic process. Once novel genes have been identified, they can be considered as targeted strategies to combat breast cancers and ultimately used as therapies to save lives.

In recent years, MYBL1 is looking like a good candidate to study for its possible contribution to the pathogenesis in TNBC. MYBL1 is a putative oncogene and a key gene involved in regulation of cell proliferative, differentiation and apoptosis, processes which are hallmarks of cancers. Data from our laboratory and others show that the MYBL1 gene is differentially expressed in TNBC and other cancers [7] [8]. Although the goal of the current studies is to provide insights into signaling processes related to the MYBL1 gene in TNBC cells, we also designed experimental approaches to identify genes associated with

MYBL1 in datasets generated using receptor-positive cancer cells. Two receptor-positive knockdown studies were retrieved from Gene Expression Omnibus (GEO) [9] based on their experimental objectives.

Both studies were designed with the goal of identifying genes affected by silencing the cMYB gene. cMYB belongs to the MYB family of genes which also includes MYBL1 and MYBL2. Data show that the MYB family of genes are co-expressed in a variety of different cancer cell lines and patient samples [10] [11] [12]. If cMYB is selectively silenced in the two GEO datasets, then MYBL1 (and MYBL2) should also be down-regulated, and If MYBL1 is down-regulated, we expect genes directly and indirectly affected by MYBL1 are coordinately affected. Consistent with previous observations, cMYB, MYBL1 and MYBL2 were each down-regulated in the GEO datasets. So, we proposed that these studies could be used to identify genes affected by down-regulation of MYBL1 and results of these data would lead to a better understanding of MYBL1 related processes in breast cancers. The GEO datasets were used in Part 1 of the current study and a separate analysis (Part 2) was designed to determine processes specifically related to MYBL1 expression in TNBC samples.

It's estimated that over 92 different breast cancer and non-tumorigenic cell lines are available [13]. The cell lines are characterized by their molecular signatures, invasive potential, site of origin and morphological properties just to note some characteristics. Often Luminal cell types (i.e., MCF7) are compared to TNBC (i.e., MDA MB231) because the two molecular subtypes represent receptor positive *compared to* receptor negative genotypes. Like many other studies, for the project described in this document, Luminal

type cancer cells were chosen for comparison to TNBC cells and samples. Both are cancers, but they are divergent with respect to receptor status and morphological characteristics.

Use of the MCF7 cell line data compared to the MDA MB231 datasets also allow for comparison to well documented published data. For many of the experiments outlined in the current project, data generated using MCF7 cells were compared to data generated using MDA MB231 cells.

CHAPTER 2

LITERATURE REVIEW

BACKGROUND

Discovery of the MYB Family Genes

The focus of this study is to better further characterize MYBL1 in breast cancers. MYBL1 belongs to the MYB family of genes which include cMYB, MYBL1 and MYBL2. The c-MYB gene was the first of the MYB family members to be discovered based on its similarity to the Avian Myeloblastosis Virus (AMV) v-myb gene sequence [14]. AMV is a highly oncogenic chicken leukemia virus which is capable of transforming immature hematopoietic cells. Compared to c-MYB, AMV has truncations in both N-terminal and C-terminal regions and contains 16 intragenic point mutations. The MYBL1 and MYBL2 genes were identified later after a Complementary DNA (cDNA) library was screened using the c-MYB sequence (11). Data show that cMYB, MYBL1 and MYBL2 demonstrate both similar and different patterns of expression under various conditions. Unique to MYBL1, the gene is a master regulator in the meiosis phase of the cell cycle in testis (12), and thus demonstrates high levels in normal testes undergoing spermatogenesis. Although not entirely unique to MYBL1, the gene regulates expression of small RNAs which control epigenetic and post-transcriptional silencing processes [15].

Comparative Analyses of MYB Family Sequences as Explanation of Their Shared Gene and Unique Gene Expression Properties

The MYB genes share both similarities and differences upon comparison of their protein sequences. The genes share an N-terminal DNA binding domain (DBD), an internal transactivation domain and C-terminal regulatory domain [16, 17]. The domains are particularly important for directing DNA binding and protein: protein interactions with co-activators, transcription factors and other proteins. The DBD located at the N-terminus is highly conserved in the family of genes, and the transactivation and C-terminal regulatory regions are less conserved and functions as transactivation domains. The positions of these regions can be demonstrated using the NCBI Web CD Search Tool [18].

There is greater than 90% similarity between the DBD in MYB family of genes. Each contains the SANT (Swi3, Ada2, N-Cor, and TFIIIB) domain, which is critically important as a DNA-chromatin recognition sequence, allowing for chromatin remodeling, and subsequent transcriptional regulation. This region implicates epigenetic regulation of MYBL1. The SANT domain is defined by binding to SW13, ADA2, N-CoR and TFIIIB genes which in turn bind proteins with MYB-like G-C rich motifs. The DBD region is involved in both activation and transcriptional repressor complexes [16].

Because the MYB family of proteins share incredible homology in their conserved DNA binding domain, on some level they interact with some of the same genes. Rushton et showed that MYB family genes recognize some of the same reporter gene constructs in *in vitro* transfection studies [11] . Compared to the N-termini, the transactivation and C-terminal region are less well conserved in MYB family proteins. Data show the N-termini

of c-MYB are involved in intra and inter negative regulator functions and truncation of the C-terminal regions result in tumors [19].

Some investigators suggest that differences in the C-terminal region is a key reason the MYB genes can display different signaling mechanisms and ultimate functions. Differences in the domains coupled with the different overall sequences of the MYB family genes, emphasize the complex signaling network of the MYBL1 genes.

Expression of MYBL1 in Various Tissues and its Relationship to the Cell Cycle Signaling and Tumorigenesis

Relationship of MYBL1 to cell cycle: Because the MYB family of genes are putative oncogenes [20] and function in processes key to dysregulation of the cell cycle and ultimately to tumor progression, they are likely candidates to study in search of targeted therapies. The c-MYB gene was previously considered as a target for luminal breast cancer therapy because (a) the gene is highly expressed in these tissues, (b) it is critical to growth and a key gene in cell cycle signaling, and (c) the gene leads to tumors in *in vitro* studies [21]. Until several years ago, small molecules and c-MYB RNA interference (RNAi) were examined as targeted therapeutic approaches [22] in pre-clinical studies. These studies have since been discontinued, but it could be that MYBL or even MYBL2 are more suitable candidates to study for a more prominent role in cancers. Less is known about MYBL1 because there are fewer studies related to MYBL1 compared to c-MYB. Some of the first experiments performed on characterization of MYBL1 involved analyses of the gene in cell cycle signaling, so the genes involvement in cell cycle signaling is well documented. Ziebold et al., [23] performed one of the earliest experiments linking MYBL1 to the cell cycle processes. Their studies show that MYBL1 is involved in GAP 1 phase (G1) to

Synthesis (S) phase progression and activation is via phosphorylation of the C-terminus by cyclin dependent kinases. Studies by Marharmati et al., [24] show that MYBL1 cooperates with MYC to mediate progression to S phase in smooth muscle cells.

Emphasizing the complexity of the involvement of MYBL1 in the human signaling mechanisms, the cell cycle can be regulated by various interactions between Elongation factor 4 (E2F4), Elongation factor 5 (E2F5), LIN9 complex DREAM MuvB core complex component (LIN9), Lin-37 DREAM MuvB Core Complex Component (LIN37), Lin-52 DREAM MuvB Core Complex Component(LIN52), Lin-54 DREAM MuvB Core Complex Component (LIN54), MYBL1, MYBL2, Retinoblastoma-Like 1 Protein (RBL1), Retinoblastoma-Like 2 Protein (RBL2), Retinoblastoma Binding Protein 4 (RBBP4) , Transcription Factor Dp1 (TFDP1) and Transcription Factor Dp2 (TFDP2) proteins which collectively are designated as DREAM or LINC complexes [25, 26]. Progression from S phase is dependent on the formation of subcomplexes between LIN9, LIN37, LIN52, LIN54 and MYBL2. Of the genes listed above, E2F4, LIN37 and TFDP2, MYBL2 and MYBL1 (of course) were dysregulated in data described in the current knockdown study. Description of E2F4 in the current project will be addressed in the Results section of this document.

Relationship of MYBL1 to tumorigenesis: Studies show the MYBL1 gene in tumors is activated by amplifications, rearrangements, and translocation events [7, 27, 28] ; often translocations involve fusion with the Nuclear Factor I B (NFIB) gene. Although the precise mechanism has not been defined in all cases, data show MYBL1 dysregulation in Burkitt's' lymphoma [29, 30], Cutaneous Adenocystic Carcinoma [31], triple negative breast adenoid cystic carcinoma [27] and TNBC in Player et al.,studies [8]. Arsura et

al.,[29] showed that MYBL1 downstream interactions with MYC gene promotes G1 to S phase transition and increased survival in lymphomas. Their data demonstrate the relationship between MYBL1, cell signaling, apoptosis and cancers.

Liu et al., [21] examined 181 patient breast cancers via microarray and performed. Supervised Network Analyses with the goal of determining prognostic significance of c-MYB in receptor positive cancers. Along with cMYB, the authors determined the relevance of MYBL1 in the different breast cancers. Utilizing multivariant analysis methods, Liu et al. identified MYBL1 and 9 other genes associated with poor prognosis in receptor positive samples. Four of the 9 genes identified by Liu et al. were identified as differentially expressed in the current receptor negative MYBL1 knockdown dataset including Repressor of RNA polymerase III transcription MAF1 (MAF1), Dual Specificity Phosphatase (DUSP7), and Solute Carrier Family 25 Member 1(SLC25A1), with Ran GTPase Activating Protein 1(RANGAP1) data not shown. Gorbatenko et al., [32] examined MYBL1 expression in basal compared to normal and luminal breast samples. Gorbatenko et al., found higher levels of MYBL1 in basal tumors (Figure 3 [arrow]). Note, MYBL2 and MYBL1 demonstrated similar expression in basal cells. Basal-like cells are closely related to TNBCs (based on gene expression analyses and morphology). Player et al., observed a similar pattern of expression in TNBC patient

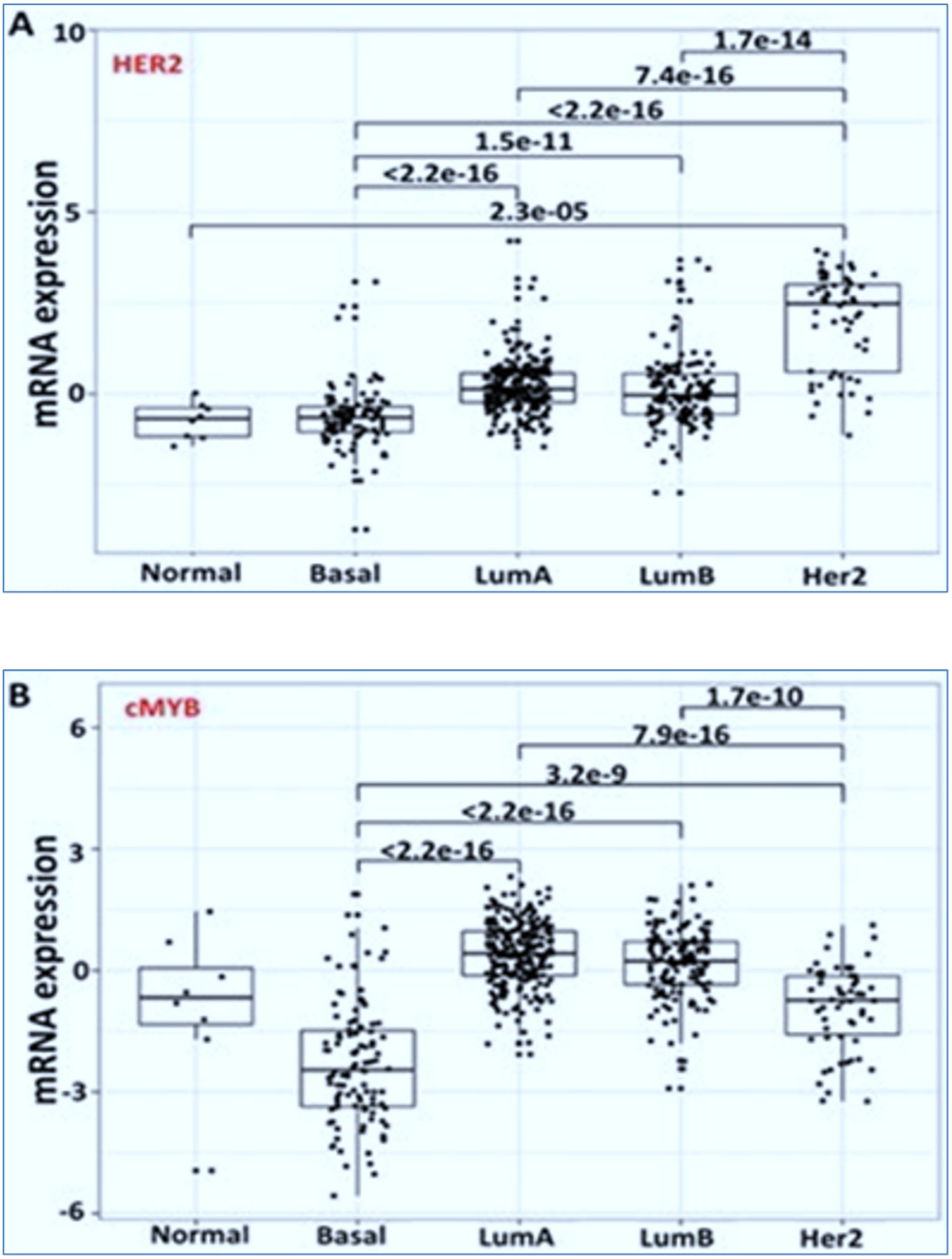


Figure 3 continued

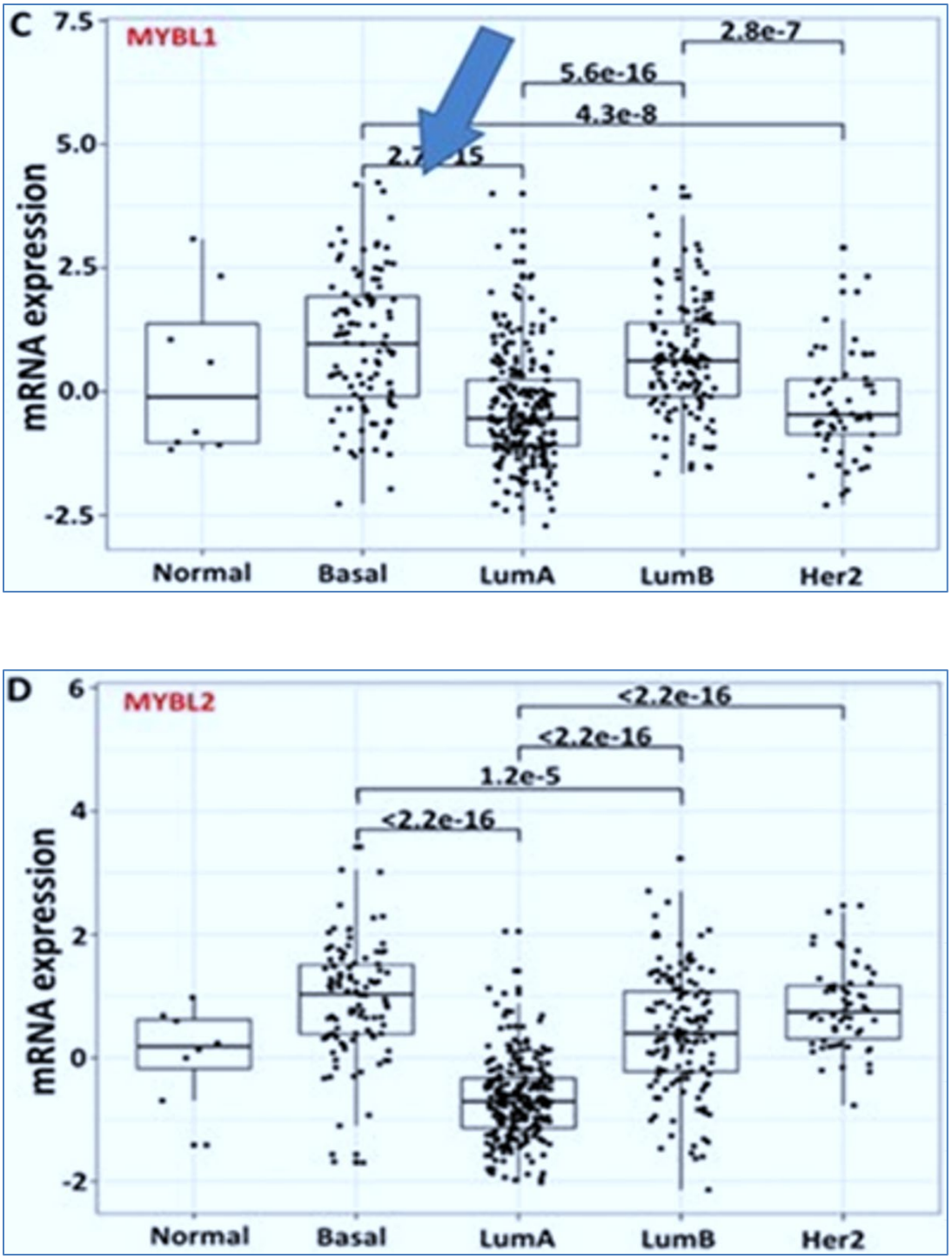


Figure 3: **Gorbatenko Data [32] Analyses of Breast Cancer Subtypes for Expression of HER2, cMYB, MYBL1 and MYBL2 Expression in Patient Samples**

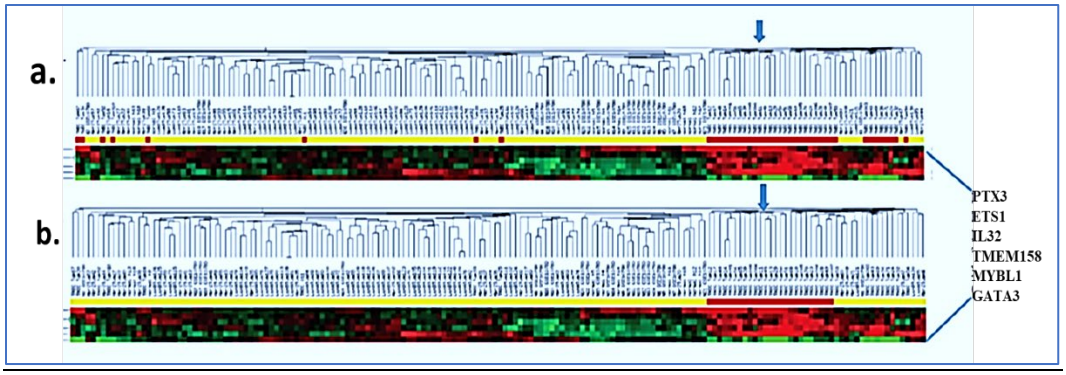


Figure 4: **Hierarchical Clustering (HC) Analyses to Demonstrate MYBL1 Expression in Patients (Maire dataset).**

(a) **Red bar=TNBC patients (region under arrow). Note that most of red bar is cluster together (with most of the TNBC 'clustering together') based on these 6 genes.**

(b) **these are the same TNBCs omitting the genes that do not cluster together. Red region under the bar indicates over-expression of genes. There is a single line of green corresponding to down-regulation of GATA3 which is known to be down regulated in TNBC. 168 patients across top of HC [8].**

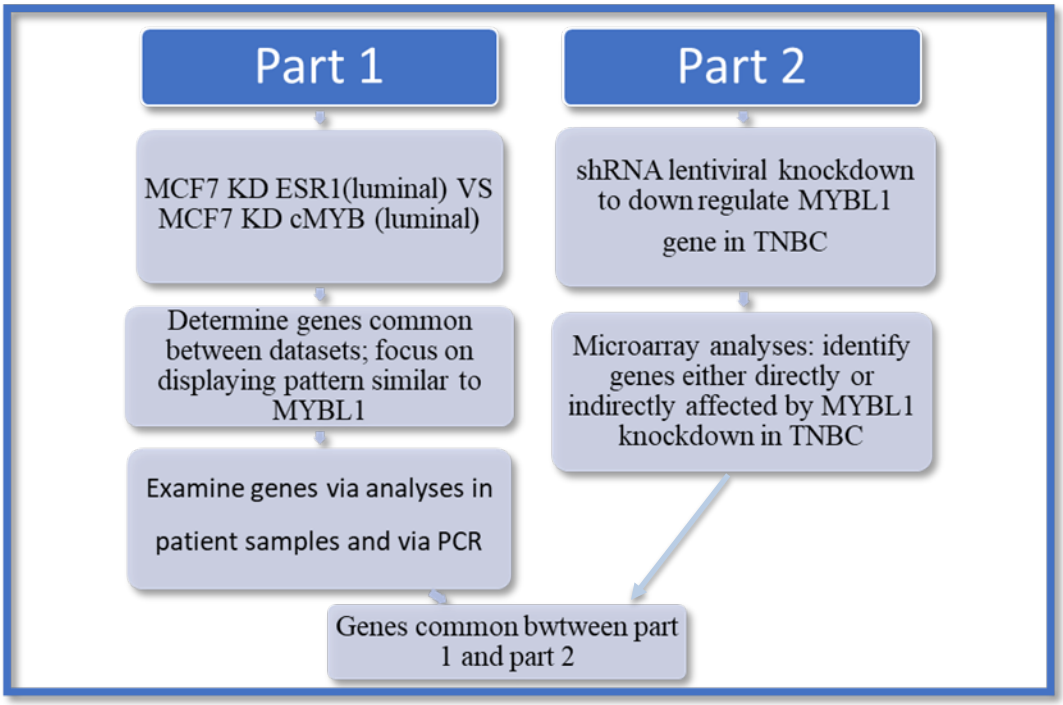


Figure 5: **Experiment Outline of Part 1 and Part 2 of Current Study**

Part 1 Summarized

When we began this study, we performed an extensive search of GEO in an attempt to identify previously published datasets where MYBL1 was silenced in TNBC; none existed. However, we did find 'gene silencing' datasets where MCF7 luminal cell line was the host cell and MYBL1 was *indirectly affected by the knock-down process.* DNA microarray is commonly used to identify differentially expressed genes associated with these processes. So, in addition to searching for knock-down studies, we included filters that included 'microarray analyses' as the experimental platform. Dr Audrey Player has generated a substantial number of breast cancer microarray datasets, so studies in GEO that use microarrays would allow direct comparison to result in her datasets. Two different studies were identified in GEO. One study involved shRNA silencing of estrogen receptor (ESR1) in MCF 7, following by microarray analyses [33], and the other involved small interfering RNA (siRNA) silencing of cMYB gene in MCF7 cells followed by custom microarray analyses [34]. Both GEO datasets involved analyses of the MCF7 luminal breast cancers, and both experiments lead to substantial down-regulation of cMYB, MYBL1 and MYBL2; therefore, the datasets could be utilized to identify genes affected by MYBL1 down-regulation.

As background, untreated MCF7 cells express all three MYB genes, and it has been shown that ESR1 (if expressed, as it is in MCF7), can regulate co-expression of cMYB, MYBL1 and MYBL2 genes [35]. Related to the cMYB siRNA experiments, cMYB was directly targeted for silencing, and as expected MYBL1 and MYBL2 were coordinately down-regulated. This is consistent with other studies that show that when the MYB family of genes are co-expressed, they tend to be co-regulated to some extent; so, it is not

surprising that knock-down of cMYB will lead to down-regulation of MYBL1 and MYBL2. The objective of the current research is to characterize MYBL1. We understand that analyses of MCF 7 datasets will lead to identification of MYBL1 related genes that characterize luminal cancers. Still, we proposed that cross comparison of genes identified as differentially expressed in the ESR1 knockdown compared to the cMYB knockdown could lead to discovery of genes related to MYBL1 regulated processes across cell lines (i.e., in luminal cancers and TNBC).

Part 2 Summarized

For Part 2, shRNA lentiviral was utilized to target MYBL1 in MDA MB231 cells. An outline of the shRNA procedure is given in Figure 6 [36]. For Part 2, MDA MB231 cells were transduced with shRNA (Figure 7), screened and successful knock-downs were analyzed via microarray (Figure 8)
https://tools.thermofisher.com/content/sfs/brochures/activity2_structure_function.pdf.
Genes identified as differentially expressed in Part 1 and Part 2 were then compared in an effort to identify MYBL1 related genes. We proposed that this approach could lead to discovery MYBL1 related genes (and processes) common to both luminal and TNBC.

Part 2 Experimental Outline

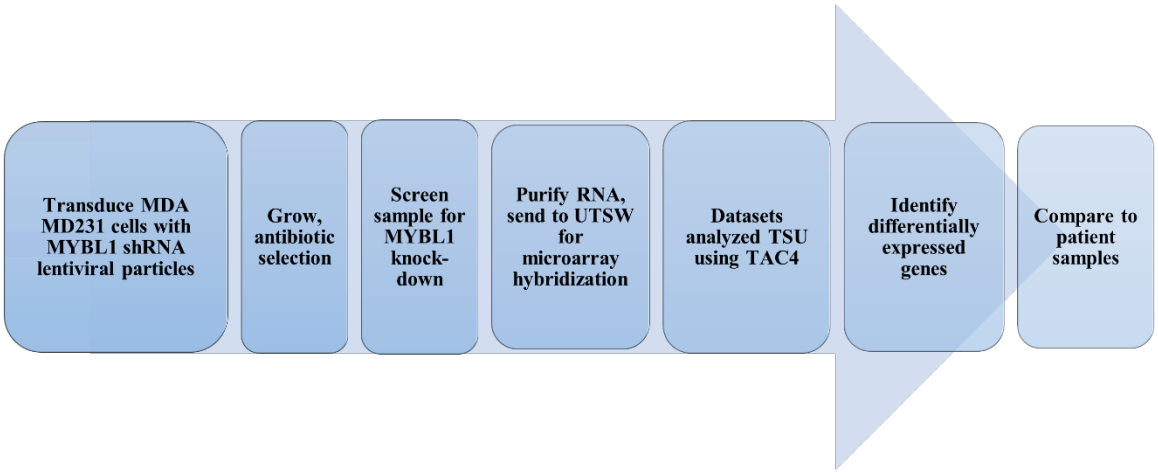


Figure 6: Experiment Outline of Part 2 of This Study

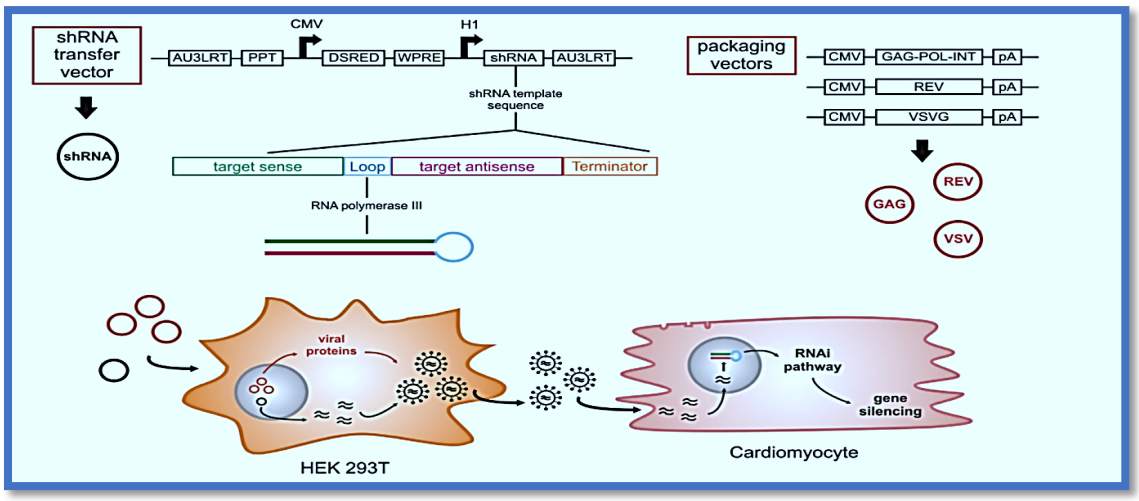


Figure 7: Summary of shRNA Procedure

HEK cells are transfected with the transgene of interest, and all cis-elements required for RNA production and packaging. Packaged particles are generated, then harvested) and sent to customer for transduction experiments [36]. In the figure above viral particles generated by HEK293T cells are used to transduce cardiomyocyte cells.

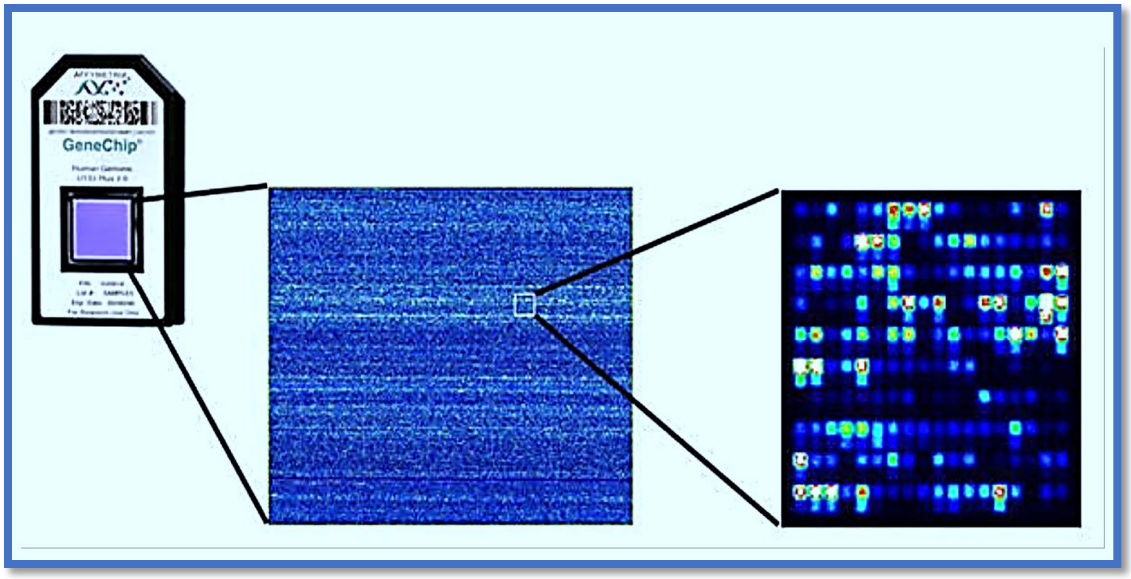


Figure 8: **The Affymetrix Microarray Genechip**

Immobilized sense strands complementary to the target sequence are on the genechip. Gene-chips containing either 56,000 transcripts or 186,000 transcripts, splice variants and small RNAs displayed as blue dots in the middle figure. aRNA (or small RNA) is labeled and hybridized to the genechip. High copy number corresponds to an intense to intense probe-set signals; displayed as lighter spots in the far-right magnified pane. Control sequences at known concentrations are pane. Control sequences at known concentrations are also supplied on the gene-chip and used for determination of copy number [37].

CHAPTER 3

DESIGN OF THE STUDY

Materials and Methods

Cell Lines

MCF7 (receptor positive luminal cells), MDA MB231 (receptor negative TNBC cells) and MCF10A (receptor negative non-tumor) cell lines were utilized for the study. All cell lines were purchased from American Type Culture Collection (ATCC®) atcc.org (Manassas, VA, USA). The cells were cultured in Dulbecco's Modified Eagle Minimum essential media (DMEM) supplemented with 1% penicillin and 10% serum in a 37° C incubator with 5% CO2 as suggested. The cells were fed twice weekly and harvested once they obtained 80-90% confluency using a 0.25% trypsin solution (Millipore, Sigma, St. Louis, MO, USA). The GEO breast cancer patient datasets included Maire et al., [38] (GSE65216), Perou et al., [34] (GSE21371), and Muthukaruppan et al., [33] (GSE37820).

Ribonucleic Acid (RNA) Isolation

The cell lines utilized in this study were grown to approximately 90% confluence in T75 dishes and harvested by adding Triazol to the culture dish and processed as suggested ThermoFisher Scientific (Waltham, Massachusetts). As summary, whole cells were lysed using 1 milliliter of Triazol, followed by addition of 200ul of chloroform. Mixtures were centrifuged at 12,000 Revolutions Per Minute (RPM)

to separate the aqueous top layer. The aqueous layer was removed, placed in a clean tube and 500ul of 95% ethanol was added. This mixture was placed at -20°C for 30 minutes, removed and centrifuged for another 30 minutes at 10,000 RPM. The RNA pellet was collected and 20ul of clean water was added; to ensure mixing, the sample was heated at ~60°C for 1 minute.

One microliter (ul) was removed for spectrophotometry analyses of the 260/280 absorbance ratio. Suitable RNA values ranged between 1.8-2.0 ratio. An aliquot of the RNA was also removed for gel electrophoresis. Up to 1 microgram (ug) of RNA was added to RNA sample buffer, heated at ~60°C for 1 minute and separated on 1% agarose gel. The RNA gel contained 1 gram of agarose, 1x 3-(N-morpholino) propane sulfonic acid (MOPs) buffer, 2µl ethidium bromide and 7% formaldehyde.

Generating Complementary DNA (cDNA)

The cDNA was generated using iScript cDNA kit (Bio-Rad, Hercules CA, USA). One microgram (ug) of the total RNA is added to a mixture of 5x buffer, reverse transcriptase, random hexamers, The Dynabeads® Oligo dT mixture, deoxyribonucleic triphosphate acids (dNTPs) and water to 20 microliters. The mixture was placed at 45°C for one hour, followed by 85°C for 3 minutes inactivation of the reverse transcriptase. The mixture was cooled, and 80ul of water was then added. The resulting cDNA was stored at -20°C until ready for polymerase chain reactions.

Primer Generation

The primer3 [39] program (http://bioinfo.ut.ee/primer3-0.4.0/) was used to generate the primer sequences for each of the target genes. Nucleotide sequences corresponding to each of the target genes was obtained from Affymetrix NetAffx [37]

(http://www.affymetrix.com/estore/analysis/index.affx). For each gene, the resulting amplicon sizes were designed so that the amplicon ranged from 200 – 300 nucleotides.

The University of California Santa Cruz Genome database [40] (https://genome.ucsc.edu/) was used to test the alignment of the primer-sets to precise gene loci. The primers sequences were submitted to and synthesized by IDTDNA.com (Coralvill,e Iowa). The primer sequences for the genes used in this study are listed in Table 1. The ATF3 and PRDM5 primers were generated by did not perform well following analyses.

Table 1: List of Primer Sequences for Genes Used in This Study

Gene	Left Primer	Right Primer	Size of Amplicon
MYBL1	AAGTCTGGGCTTATTGGACATAA	TGCAAGTATGGCTGCTACATG	202bp
ATF3	GCTCTTTTCCCCACCTGTCT	TTGGAAATCAAGGCACCATT	203bp
DDX58	GACAGAACTGCAGCCATGAC	AGGGGTACAAGCGATCCATG	277bp
E2F6	TCTACCTAATTGCTGTCTCCCA	CTGGAGAGAGGGCAAGGTAC	242bp
E2F7	CGCTAGACTTGGATATTTTGGGT	CCCCGACTCTTGTACCAGAA	199bp
FBXL13	GCCATTACCTGCACATTTTGGA	CAAACCAACGTGGAGGGTCA	246bp
FBXO8	ACCTCACTAGCCCTCATGTG	GGTTGCACACTGAAGCTTGA	271bp
FBXO32	AGATCCGCAAACGATTAATTCT	AGGGGGACCCTTCTGAAGT	273 bp
Linoo673	TGAAGGCTCGTGTTCTCCAT	AAGTGCCACAACCCCTGATA	220bp
NFE2L3	TGGGCAAAAGCGATTAAGGG	ACACTGTAGCTCCTATGGCA	280bp
PSPH	TGTAGAGCTGCTGGGAGAAC	ACAAAACTCTCCAGGAAATCGA	258bp
ZN75A	TCTCTTCACAGTAGCAGGCT	CCCTAGTGGTGGTTAACAAGAT	228bp
GAPDH	TCCCTGAGCTGAACGGGAAG	GGAGGAGTGGGTGTCGCTGT	217bp
MYBL2	GAGGGGGTCTGTGAATCTGA	CCATCCTAAGCAGGGTCTGA	265bp
KIF18b	GCTCTTTTCCCCACCTGTCT	TTGGAAATCAAGGCACCATT	203bp
TCF19	TCTTAGGGGAAGGGGAGAGA	GTCACAGCCATCACACTGGT	266bp

Polymerase Chain Reaction

The polymerase chain reaction (PCR) was used to confirm the microarray data and also demonstrate differential expression between genes in different samples. PCR was performed using the cDNA products. Each PCR reaction contained 2ul of forward and reverse primers (generated to be specific for the particular gene), 2ul of cDNA (generated above), 10ul of 2x thermostable DNA polymerase I TAQ polymerase master mix (which contained TAQ enzyme, dNTPs and TAQ buffer; Life Technologies, Carlsbad, California) and water to 20ul. The samples were place in PCR quality tubes, placed in the Bio-Rad Thermal Cycler (Hercules California) and processed for (a) 5 minutes at 95ºC degrees (b) then 30-32 cycles for 30 seconds at 95ºC, followed by 30 seconds at 58ºC degrees, then 30 seconds at 78ºC.

Gel Electrophoresis

A 2% agarose gel was utilized for analysis the PCR products. The gel was prepared using 2 grams of agarose, added to 100ml of 1X Tris/Borate/EDTA (TBE) buffer. This solution was microwaved, cooled, and 1ul Ethidium Bromide (EtBr) added, before pouring into the gel chamber. Ten microliters of the PCR product and 2ul of sample buffer were loaded onto the gel for electrophoresis.

Densitometer

The gels were viewed on the Licor, and the intensity of amplicons were analyzed using software available on Licor. All the values were normalized compared to the control.

shRNA Knockdown of MYBL1 in MDA MB231 Cells

The MYBL1 shRNA Lentiviral particles and the scramble control particles were purchased from Origene (Cat # TL303089V; Rockville Md, USA). Four MYBL1 target specific particles (packaged from the pGFP-C-shLenti vector; labeled LVA, LVB, LVC, LVD) were supplied by Origene and screened for their efficiency to suppress expression of the MYBL1 transcript. The lentiviral particles were transduced into MDA MB231 TNBC cells (at a MOI of 10:1) and screened to determine the sequence most effective at down-regulating the MYBL1 transcript. MDA MB231 cells were incubated with the targeted or scrambled viral particles for 72 hours in the presence of polybrene (sc-134220; Santa Cruz Biotechnology, Dallas TX, USA) in complete cell culture media. Lentiviral particles were removed, and fresh media was added to the cells. The transduced cells were selected following growth in 1ug/ml puromycin (CAS 53792; Santa Cruz Biotechnology, Dallas TX, USA). LVA particles, identified by the TCTGATCCTGTAGCATGGAGTGACGTTAC sequence, demonstrated the most significant down-regulation of MYBL1 mRNA, as a result this preparation was used for the future experiments. Cells transduced with LVA, and the scrambled control sequence were maintained in the presence of puromycin.

Processing of GEO Datasets

Experimental results from thousands of different types of experiments are searchable and available in GEO available via National Center for Biotechnology Information (NCBI, Bethesda, Maryland). Two datasets were found, examined for quality, and extracted for use in this current project. The MCF7 ESR1 shRNA data set was GSE37820. The MCF7 cMYB siRNA dataset was GSE21371.

The datasets were loaded into the GEO2R online analysis tool, supplied by NCBI, which allowed for differential gene expression analyses. Using GSE37820 as an example, (a) MCF7 microarrays from scrambled shRNA preparations (control) were compared to (b) MCF microarrays from ESR1 shRNA preparations. T-test – like analyses were performed to compare transcript levels *across* thousands of genes. Genes displaying a difference in transcript levels are considered differentially expressed. The data are sorted based on fold change and statistical significance. When processing the GEO knock-down dataset, very few genes were identified as differentially expressed, so a decision was made to 'relax' the fold change stringency, prior to selection. The analyses conditions were also relaxed if independent datasets showed evidence of differential expression. The datasets can be further compared using MolbioTools (http://www.molbiotools.com/listcompare.html) which allows for gene lists comparisons.

Microarray and Data Analyses Continued

RNA purified from MDA MB231 preparations transduced with either scrambled or LVA shRNA were shipped overnight to the University of Texas Southwest Core Facility (UTSW; Dallas Texas, USA). The aRNA was prepared by the UTSW core facility and hybridized to the Affymetrix Clarion microarray gene-chip which includes ~186,000 probe-sets (i.e., transcripts, splice variants, siRNA and SNORna). The CEL intensity files were made available to our laboratory, and data analyses was performed (at TSU) using the Affymetrix TAC 4.0 software (Thermo Fisher Scientific, Waltham Mass). CEL intensity results were normalized using functions in Robust Multi Array (RMA) program and the differentially expressed genes were generated following Limma Bioconductor analyses.

Probe-sets that displayed at least a 4-fold difference in expression between the MYBL1 LVA preparation compared to the scrambled sequence were selected for analyses. A lower differential expression threshold (of 2x) was used for selection of the microRNA (miRNAs), as the miRNAs were also screened via TargetScan analyses [41] to identify their predicted nucleotide targets. The p-values were not generated by the TAC 4.0 program as two Clarion microarrays were hybridized and subsequently processed to identify the differentially expressed genes. Differentially expressed candidate genes were also examined using Gene Ontology [42] analyses and by comparison to the gene expression patterns in both untreated cell lines and patient samples. Transcript plots and analyses were performed using Microsoft Excel. Comparisons between genes on the various differentially expressed gene-lists were performed utilizing the Molbio-tools (http://www.molbiotools.com/listcompare.html). Protein: protein interactions were performed using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING™) program [43]. STRING™ analyses are based on interrogation of millions of data generated experimentally, theoretical, and published.

Western Blotting

Log phase cells were washed with cold phosphate buffered saline (PBS), placed on ice and scraped using cold Radioimmunoprecipitation assay (RIPA) lysis buffer (Santa Cruz Biotechnology, Dallas TX, USA). The supernatants were mixed in the cold room and centrifuged to collect solubilized protein preparations. Protein concentrations were determined using the Pierce Bicinchoninic Acid Kit (BCA) (as suggested by ThermoFisher Scientific, Waltham Massachusetts). Up to 50ug/lane of the protein was calculated for loading to a gel lane. Protein was mixed with sample buffer, heated and loaded for

electrophoresed using protocols available at Novus Biologics (Novus Biologics Littleton, Colorado). The protein preparations were transferred to nitrocellulose and probed for detection of specific proteins using antibody concentrations and incubation times as recommendation of the suppliers.

Antibodies: Actin was used at a $1:10^4$ dilution (NB600-501SS; Novus Biologicals LLC, Littleton, Colorado). MYBL1 was used at a 1:500 dilution (sc-514682; Santa Cruz Biotechnology, Santa Cruz, California). MYBL2 was used at a 1:500 dilution (sc-81192; Santa Cruz Biotechnology, Santa Cruz, California). TCF19 was used at a 1:100 dilution (sc-390923; Santa Cruz Biotechnology, Santa Cruz California). KIF18b was used at a dilution of 1:1000 (A303-982A; Bethyl Laboratories, Montgomery, Texas). Secondary horseradish peroxidase conjugated Anti mouse (HAF007; R and D Systems, Minneapolis Minnesota) and Anti Rabbit (NBP-2-30348H; Novus Biologicals LLC, Centennial, Colorado) antibodies were used at a dilution of 1:4000. Western blotting results were visualized with the Clarity Western ECL substrate (Bio-Rad, Hercules, California) on a LI-COR digital imaging system (LI-COR Biotechnology, Lincoln, Nebraska).

CHAPTER 4
RESULTS AND DISCUSSION

Results generated by the microarray platform should only serve as a starting point towards identifying potential candidates identified as differentially expressed genes. Microarray data must be validated. As validation of the expression patterns, gene expression levels should be compared to independent datasets and patient samples, and transcript levels experimentally validated via PCR and protein analyses via western or immunohistochemistry at a minimum. If a gene demonstrates the same pattern across all analyses, it can be considered a reliable candidate for further study. In summary, the purpose of these analyses is to identify genes that demonstrate a reliable pattern of differential gene expression, across all the analyses, not just one experiment or one method of analyses.

The main focus of our laboratory revolves around characterizing TNBC. These cancers have poor 5-year survival, are over-represented in women of African-descent, and unlike receptor-positive patients, there are few therapeutic options for TNBC patients [44]. Although several genes are being studied as possible therapeutics targets, to date, none are being used as consistent therapy. TNBC patients are generally treated with the less specific chemotherapy and radiation therapy. As a result, we are studying MYBL1 with the objective of identifying novel genes that are enriched in these cancers, so that the genes might be considered for more extensive studies; and if reliable, even as future therapy in TNBC.

As mentioned, this study is partitioned into Part 1 and Part 2. Both parts involve analysis of gene expression when MYBL1 is down-regulated with the goal of identify genes either directly or indirectly affected by down-regulation of MYBL1.

The difference between the two sections are the host cells and the primary gene-targeted-for-silencing. In Part 1, the host cell is the MCF7 luminal cancer, and the gene targets are ESR1 or cMYB. For Part 2 the host cell is the MDA MB231 cells, and the target gene is MYBL1, which is more directly aligned with the focus of our laboratory. In designing these experiments, we recognized that our approach would allow identification of genes associated with luminal cancers (Part 1) and TNBC (Part 2). Towards the end of this section, we speculate as to which family of genes might be associated with MYBL1 in both luminal and TNBC. As expected, because the target cancer cells are so different, there is little to no over-lap between MYBL1 associated genes in luminal and TNBC samples.

Results from Part 1 Studies

Reliable differentially expressed genes can only be identified by examining the genes across different experimental platforms in search for a consistent pattern of expression. For the Part 1 study, thirteen genes were initially identified as common between the MCF7 ESR1 silencing datasets and the MCF cMYB silencing datasets (Table 2). Thirteen genes were chosen following comparisons between the ESR1 and cMYB datasets, however upon further analyses the number was even smaller. Based on their levels of differential expression, MYBL1, DExD/H-Box Helicase 58 (DDX58), Elongation factor 7 (E2F7), Long Intergenic Non-Protein Coding RNA 673 (Lin00673), F-Box and Leucine Rich Repeat Protein 13 (FBXL13), Phosphoserine Phosphatase (PSPH), and Activating

transcription factor (ATF3) demonstrated significant differential transcript levels across both silencing experiments as determined by their fold-change on the microarray platform.

All of the genes were not significantly differentially expressed, still they were analyzed for their expression in patient samples (Figure 9). The Zinc Finger Protein 436 (ZNF436) and Zinc Finger Protein 75a (ZNF75A) genes were not in the patient datasets, so they were not represented in the Hierarchical cluster analyses. Judging from the cluster analyses, consistent with the results in Table 2, MYBL1, E2F7 and PSPH display the strongest differential expression (with E2F7 and PSPH demonstrating a pattern similar to MYBL1). This was determined by comparing the intensely of the red color in the cluster analyses and the 'clustering' between MYBL1, E2F7 and PSPH. The genes in Table 2 were also examined for their expression in Dr Player's 2010 cell line dataset (not published). Again, E2F7 demonstrated a pattern of expression similar to the MYBL1 profile in the cell line datasets (Figure 10). The Dr Player's 2010 cell line dataset and other independent breast cancer datasets were used for comparison to the GEO datasets during the gene-selection process. Dr Player's 2010 dataset show low transcript levels and small (although significant) differences between gene expression, which likely (in part) contributes to the poor performance of the so-called 13-gene list.

For PCR analyses, ZNF436 was not in the cell line dataset, and we could not generate quality PCR primers for ATF3 and PRDM5, so they were excluded from analyses. Nonetheless, judging from the cell line data, E2F7 demonstrated a pattern of expression consistently over-expressed with MYBL1 in the TNBC and MCF7 cells (Figure 11). This was determined by visual inspection. A more reliable conclusion must involve densitometer. These data suggested that our initial selection was either poor, or a more

stringent selection method should have been used. The E2F7 gene did appear somewhat promising, but further studies must be performed. E2F7 belongs to the E2F family of transcription factor genes.

The E2F family of genes are involved in p53 regulated cell cycle signaling [45-47]. And previous data show E2F7 over-expression in breast cancer [48]. We are not suggesting this is the case, here, because the current data only serves as an observation at this point. If our current data prove to be reliable (i.e., upon repeat experimental analyses), the data can serve to describe MYBL1 regulation of (or with) E2F7 under our experimental conditions. Further analyses of E2F and other genes are included in the sections below. Early in the analyses, the PSPH gene appeared to be a promising gene candidate, but the PCR results were not reproducible, so the gene was not considered further. This is the purpose of the screening process, which is to eliminate genes that do not reproducibly validate.

Table 2: **Initial List of Genes Identified as Common Between the MCF7 ESR1 shRNA Compared to MCF7 cMYB siRNA. The Significance cut-off is >2-fold with p Value <0.05.**

Gene Description	Symbol	Fold Difference	P Value
V-Myb A vian Myel oblast Viral Oncogene Homolog Like 1	MYBL1	2.14	0.000193
Activating transcription factor 3	ATF3	2.1	0.00935
DExD/H-Box Helicase 58	DDX58	2	0.0037
Elongation factor 6	E2F6	1.2	0.0274
Elongation factor 7	E2F7	3.6	2.29E-03
F-Box and Leucine Rich Repeat Protein 13	FBXL13	2.5	0.0124
F-Box Protein 8	FBXO8	1	0.00248
F-Box Protein 32	FBXO32	1	5.76E-01
Long Intergenic Non-Protein Coding RNA 673	Linoo673	2.1	0.0117
Nuclear Factor, Erythroid 2 Like 3	NFE2L3	1	1.71E-02
PR/SET Domain 5	PRDM5	1	0.000906
Phosphoserine Phosphatase	PSPH	6.9	3.72E-01
Zinc Finger Protein 75a	ZN75A	1.3	0.0283
Zinc Finger Protein 436	ZN436	1	0.632

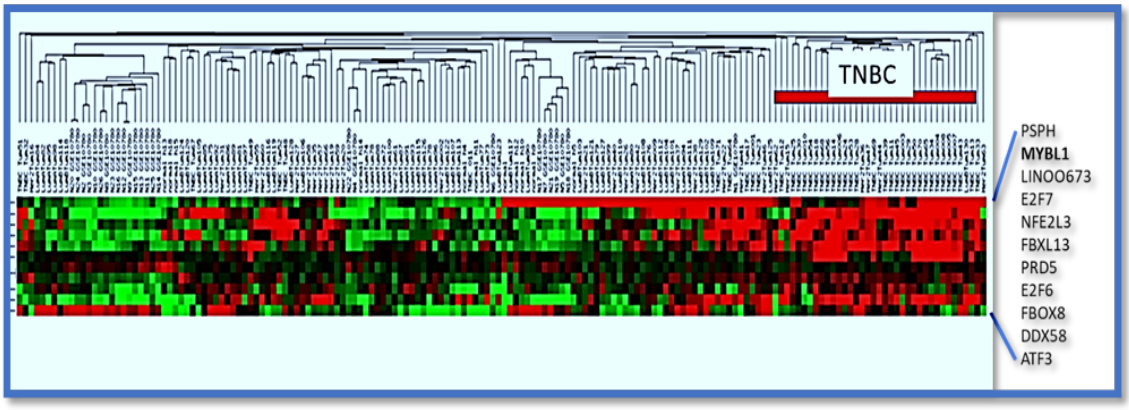


Figure 9: **Hierarchical Cluster of Maire Patient Data-set**

Select genes were analyzed for their expression pattern in patient samples. Not all genes were present in the patient dataset. Red bar across the top designates TNBC patient samples 'clustering together' based on gene expression. Red regions within the plot represents higher level of gene expression. Green indicates lower gene expression levels. Note the region under the bar (i.e., TNBC) shows high gene expression. PSPH, MYBL1, LINOO673, E2F7, NFE2L3, FBXL13, PRDM5, E2F6, FBX08, DDX58 and ATF3 genes are examined in this patient dataset.

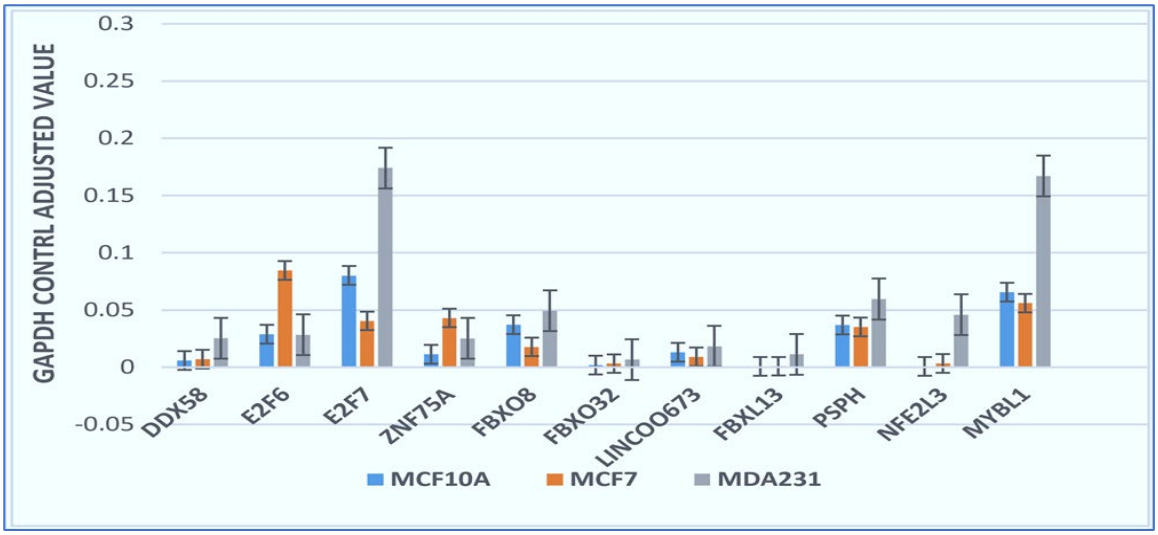


Figure 10: **Microarray Analyses of Select Genes in Player (2010) Cell Line Dataset. Plot of Transcript Levels**

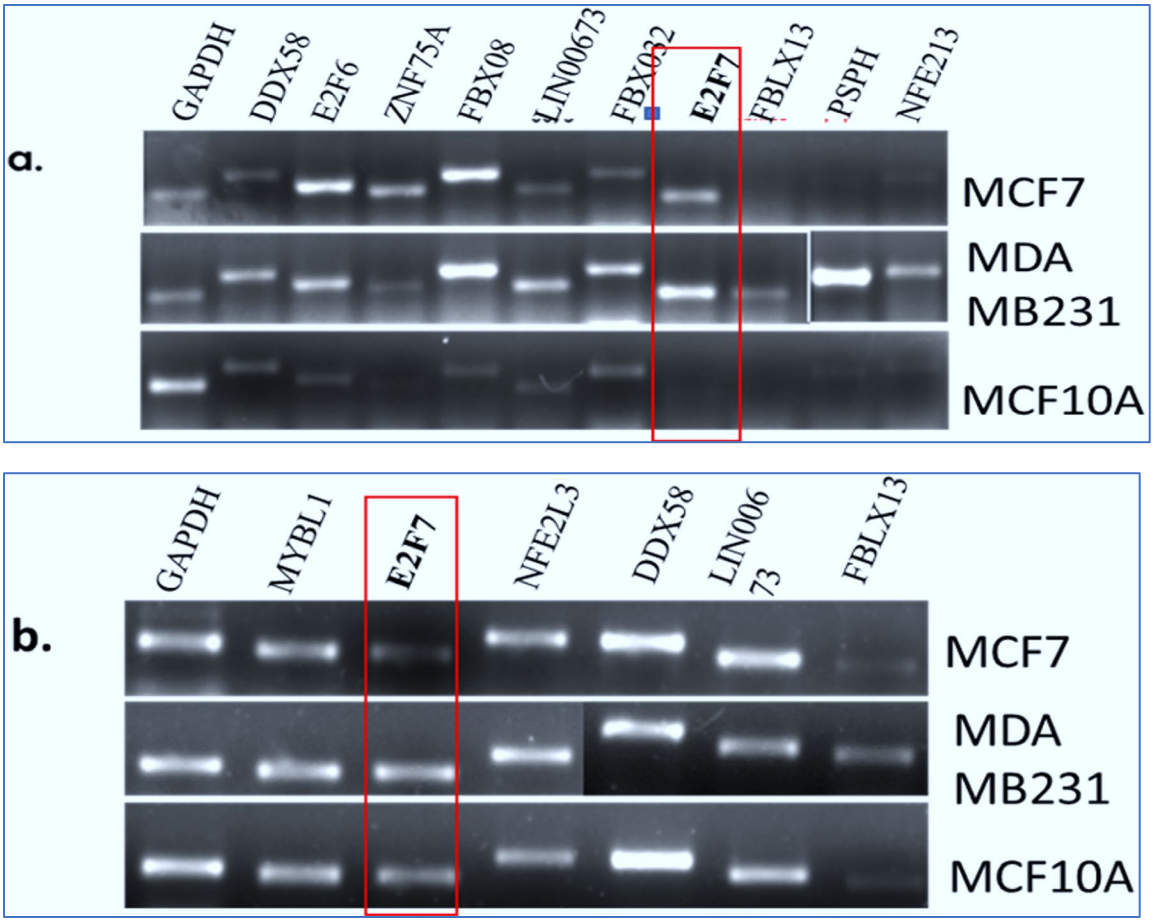


Figure 11: PCR Performed to Experimentally Screen Select Genes

(a) Select genes were analyzed in 3 cell lines.

(b) Certain genes analyzed in the top panel were selected for repeat analyses. Only the E2F gene appeared to demonstrate a reproducible pattern of expression similar to that observed for MYBL1.

As mentioned, we initially selected 13 genes common to ESR1 and cMYB knockdown studies in MCF7. But upon inspection (a) all were not truly differentially expressed (b) all did not validate or (c) were not represented in particular samples, or (d) gene primers were not successfully generated (like with ATF3 and PRDM5). So, the list of initial genes decreased from 13—to—11—to—10—to—6—1 gene, the E2F7 gene.

Please note that this pattern of screening is common to DNA microarray. The microarray is a useful experimental platform, but it is only used as a tool to identify potential candidate genes. Even with this explanation, a good portion the responsibility lies with our selection process. Only genes displaying substantial different expression should have been initially selected.

Genes were only considered for further study if they displayed a consistent experimental pattern, like that observed for E2F7. E2F7 appeared over-expressed with MYBL1, suggesting co-expression and/or co-regulation. The E2F family of genes are classified as transcription factors involved in transcription and proliferation events, so it will not be surprising if MYBL1 is found to associate with E2F7. Lui et al. [48] found that E2F7 gene was over-expressed in breast cancers. In addition, data show a significantly strong relationship between MYBL1 and an E2F family member, E2F4 in cell cycle signaling [49]. Our data will have to be repeated using samples other than those used in this study before we can make definitive statements related to E2F7 (or other E2F family genes) and MYBL1.

Part 2 shRNA Knockdown of Mybli in MDA MB231 TNBC CELLS

Since identifying MYBL1 over-expressed in TNBC, we have been concerned with further characterizing the gene in these cancers. Part 2 of this current project will allow for a better understanding of MYBL1 in TNBC.

Origene guaranteed one lentiviral preparation suitable for knock-down, so all were screened to select for the one suitable for knock-down of MYBL1 gene in MDA MB231 TNBC cells. The sequences for the lentiviral preparations are listed in Table 3. The five MYBL1 shRNA viral particle samples (sequences) were obtained from Origene, including

a scrambled lentiviral control, and shRNAs corresponding to 4 different regions of the MYBL1 nucleotide sequence. The MYBL1 shRNA preparations were labeled A-D. The MYBL1 shRNA preparations were screened to determine which effectively decreased expression of the MBYL1 in MDA MB231 TNBC cells. When LVA, LVB, LVC and LVD were transduced into MDA MB231 TNBC, data show the LVA preparation was most effective at KD of MYBL1 (Figure 12). As a result, the LVA preparation was utilized in subsequent experiments, since it demonstrated substantial downregulation of MYBL1 gene in MDA MB231 cells.

Table 3: Five MYBL1 shRNA Viral Particle Samples (sequences)

MYBL1 Lentiviral	Sequence
TL303089VA	TCTGATCCTGTAGCATGGAGTGACGTTAC
TL303089VB	CTTGTAATGGTGGCAACAGTGAAGCTGTT
TL303089VC	TAGCACTCCACCAGCCATCCTCAGAAAGA
TL303089VD	CAGGCACTCAACTGTTGACTGAAGACATT

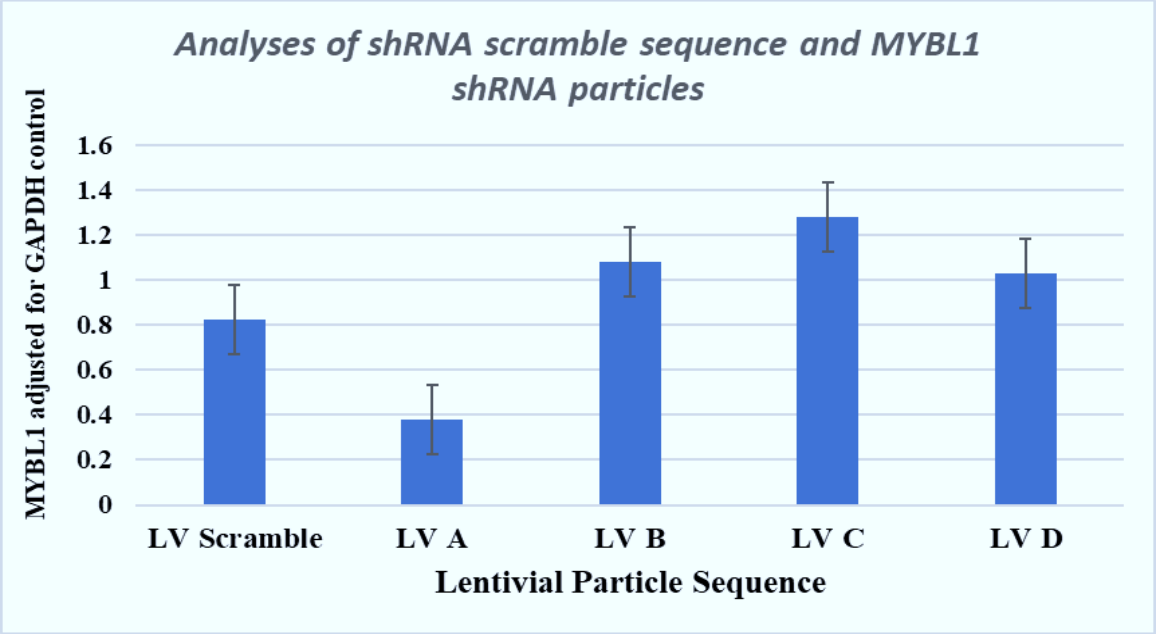


Figure 12: Analyses of shRNA Scramble Sequence and MYBL1 shRNA Particles to Screen for the Sequence that Most Effectively Decreases MYBL1 Expression in MDA MB231 Target Cells

Transcripts and Protein Levels Validation for Both Scrambled and LVA Sequences

To further evaluate the efficiency of the KD, we examined the transcript and protein levels for the scrambled control compared to the LVA sequences in MDA MB231 transduced cells. The scrambled control served as a negative control and the LVA preparation contained the shRNA MYBL1 target sequence. The transcript levels were determined using PCR and protein expression levels were determined using Western blotting analyses. Our data showed that LVA in MDA MB231 cells led to substantial KD of MYBL1 transcript and protein levels compared to the scrambled control sequence (Figure 13). The transcript levels are displayed in the top panel and the protein levels are displayed in the bottom panel). MYBL1 levels in untreated MDA MB231 cells served as positive controls for MYBL1 transcript and protein levels. Substantially high levels of

MYBL1 transcript and protein are expressed in the MDA MB231 cells that were not treated with lentiviral particles.

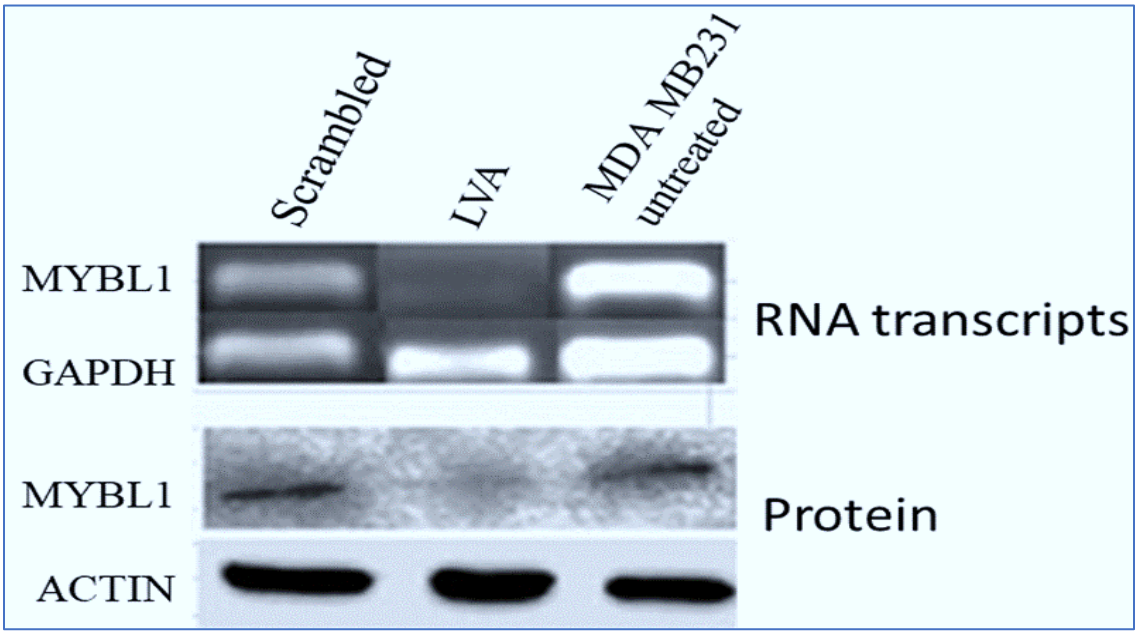


Figure 13: Transcript and Protein Analyses of MYBL1 LVA shRNA Preparation to Demonstrate Knock-down Efficiency

Three MDA MB231 preparations were examined for transcript or protein levels. MDA MB231 preparations containing Scrambled control, MYBL1 LVA or untreated MDA MB231 cells were examined. The transcript analyses are in the top panel and the protein analyses are in the bottom panel.

Gene Microchip Analysis

MDA MB231 cells were transduced with MYBL1 shRNA or scrambled control sequences, and total RNA preparations were shipped to UTSW for Clarion microarray hybridizations. Following microarray, the differentially expressed cells were determined using the TAC4 program downloaded from Affymetrix.com.

A plot of the differentially expressed genes are represented as a Scatter plot in Figure 14. The over-expressed probe-sets are represented in red and down-regulated probe-

sets in green. The probe-sets include transcripts, splice variants and small RNAs. About 6000 probe-sets are represented in the Scatter plot. When the dataset is processed to remove noise or background, the data-set decreases by10x to ~600 differentially expressed probe-sets. Table 4 was extracted from the TAC4 analyses to demonstrate how the gene expression values are displayed by the TAC4 software.. The transcript or probe-set levels are displayed as log 2 values to represent the signal intensity of the particular probe-set. We converted the TAC4 values using an antilog calculator. As example, MYBL1 transcript levels in the scrambled control is 4705 and 1024 in the knock-down sample; this shows a decrease by >4 fold. Background level on the microarray is generally around 200. Note the transcript values for MYBL2. Data presented here are the first to demonstrate regulation of MYBL2 by MYBL1. Except for the antilog values, all of the calculations presented in this figure are performed by TAC4. Even though samples like miR222 and SNORA30 show incredible differential expression, it is impossible to examine all of the probe-sets that appear ‘interesting’. Although ~600 differentially expressed data points are a manageable number to study, additional filters were performed to identify gene-set enrichment. It is beyond the scope of this study to examine a vast number of genes.

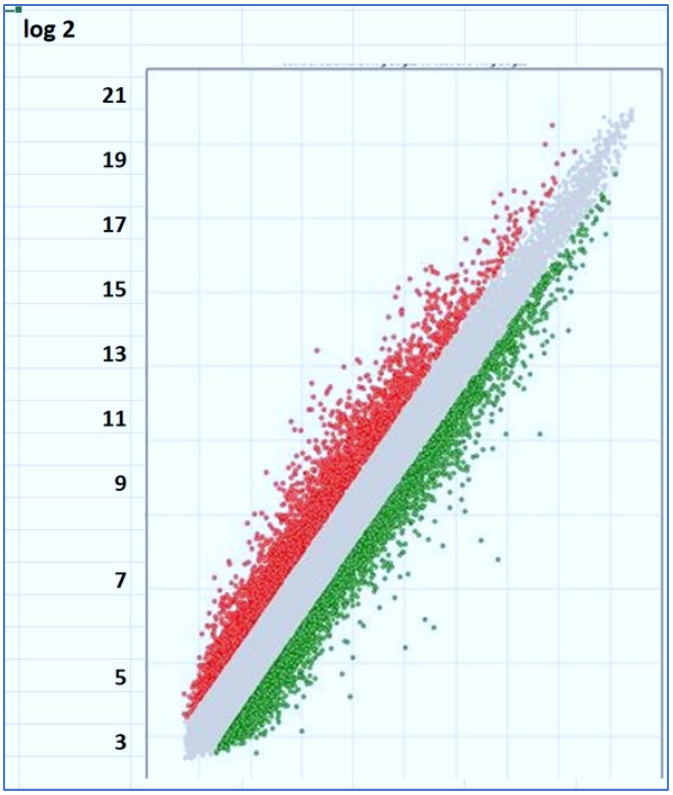


Figure 14: **Clarion Microarray Gene-chip Scatter Plot to Demonstrate Differential Gene Expression**

Differentially expressed probe-sets were selected based on pvalue<0.05 and >2-fold difference between shRNA scramble vs MYBL1 shRNA (LVA). Red represents over-expressed and green represents under-expressed probe-set levels.

Table 4: Example of the Data displayed by TAC4 and Also the Microarray Transcript Levels for MYBL1, MYBL2 and small RNAs in Scrambled Compared to LVA.

Scram	LVA	FOLD CHANGE	GENE
11.87 (3300)	8.7 (415)	-6.45x	MYBL2
12.25 (4705)	10.23 (1024)	-4.06 x	MYPL1
5.41 (42)	11.03 (2000)	+48.89 x	MIR222
6.17 (102)	11.81 (3300)	+49.67 x	SNORA30

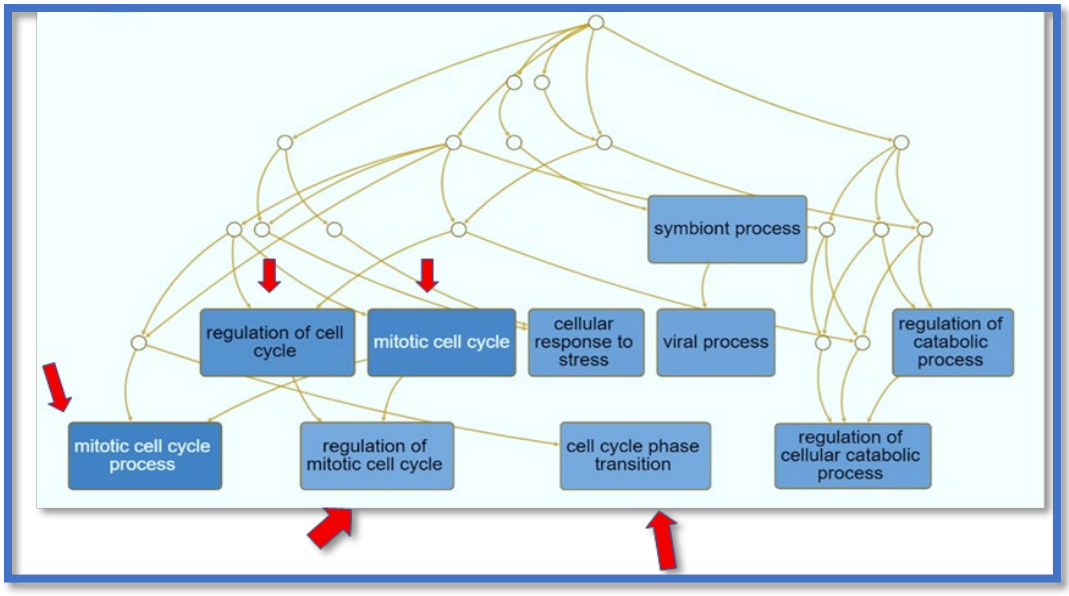
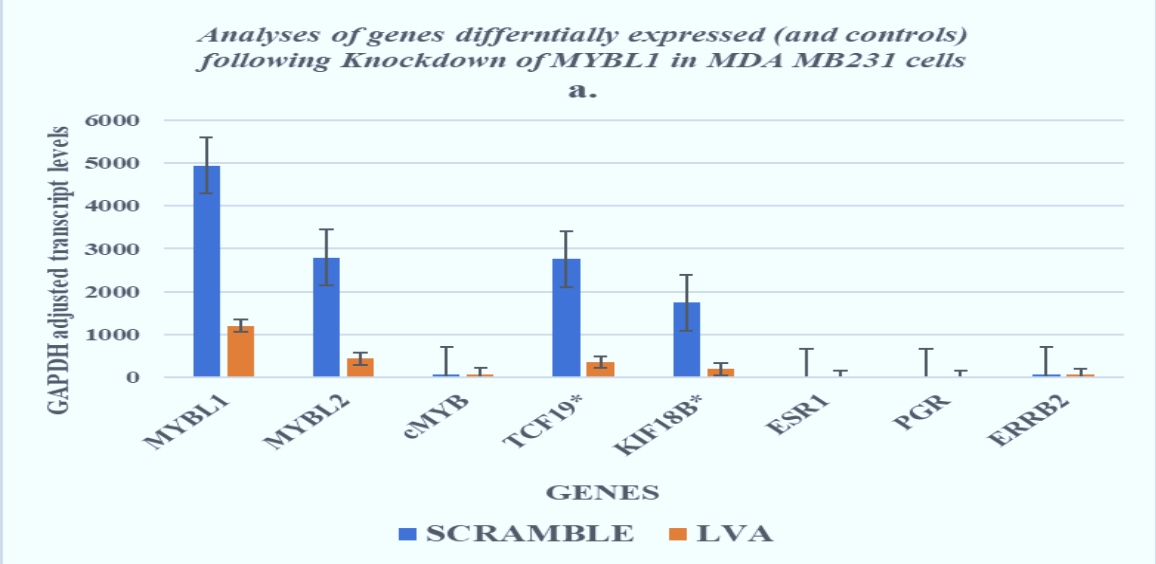
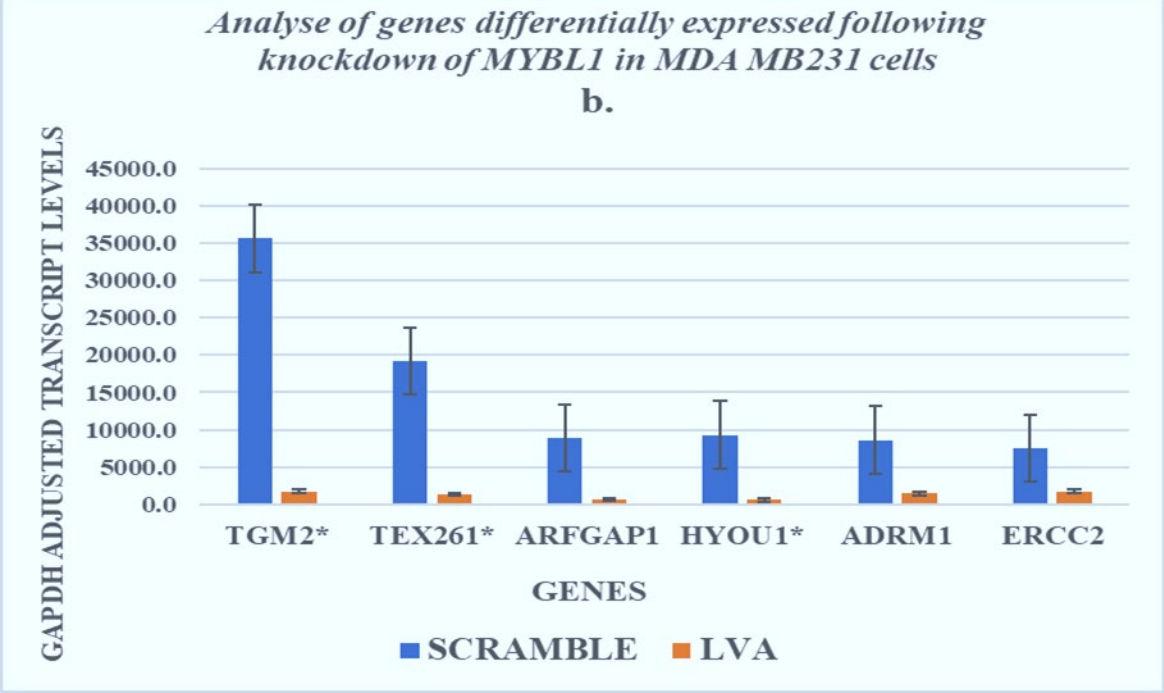


Figure 15: Gene Ontology to Determine Gene Set Enrichment in the Microarray Dataset

A shortened gene list of differentially expressed genes show enrichment of genes enriched for cell cycle signaling processes. These genes are plotted in Figure 16.

After filtering based on >4-fold expression, the shortened gene list was examined using Gene Ontology and genes either directly or indirectly involved in cell signaling were identified (Figure 15). This shortened dataset included MYBL1, MYBL2, TCF19, KIF18B and other genes. Based on our interest, statistical significance and Gene Ontology, transcript analyses of our gene candidates are plotted in Figure 16.





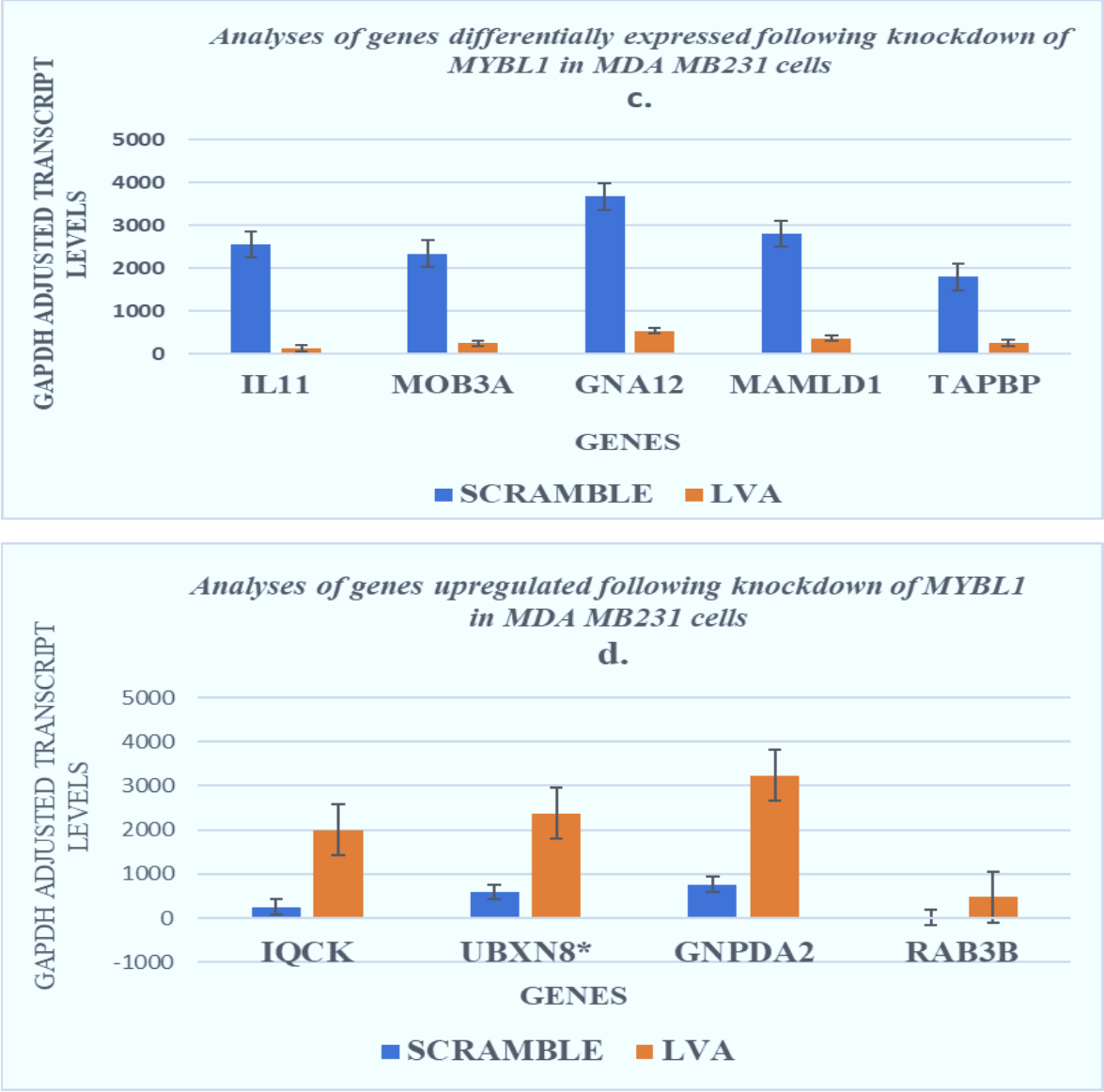


Figure 16: Final list of Candidate Genes Based on Differential Expression Analyses of MYBL1 shRNA Compared to Scrambled Control

Data extracted directly from microarray results. (a-d) Transcript levels are plotted. Blue bars represent transcript levels for the scrambled control and orange bars represent the transcript levels of the MYBL1 shRNA sample

MDA MB231 cells are negative for cMYB, ESR1, PGR and HER2, as a result, these genes served as quality controls for the performance of lentiviral particles and the

microarray (Figure 16a). Thus far the most promising genes in Part 2 of this project are MYBL2, TCF19 and KIF18B. Our laboratory is in the process of examining the relationship between MYBL1 and the genes included in Figure 16b-d. In separate studies, data show MYBL1, MYBL2, TCF19 [50] and KIF18B [51] are involved in cell cycle processes. As validation, STRING™ analyses demonstrate significant protein: protein interactions between the genes (Figure 17). Along with our 4 genes, STRING™ show close relationship with LIN9 and CCNB1 cell cycle signaling genes.

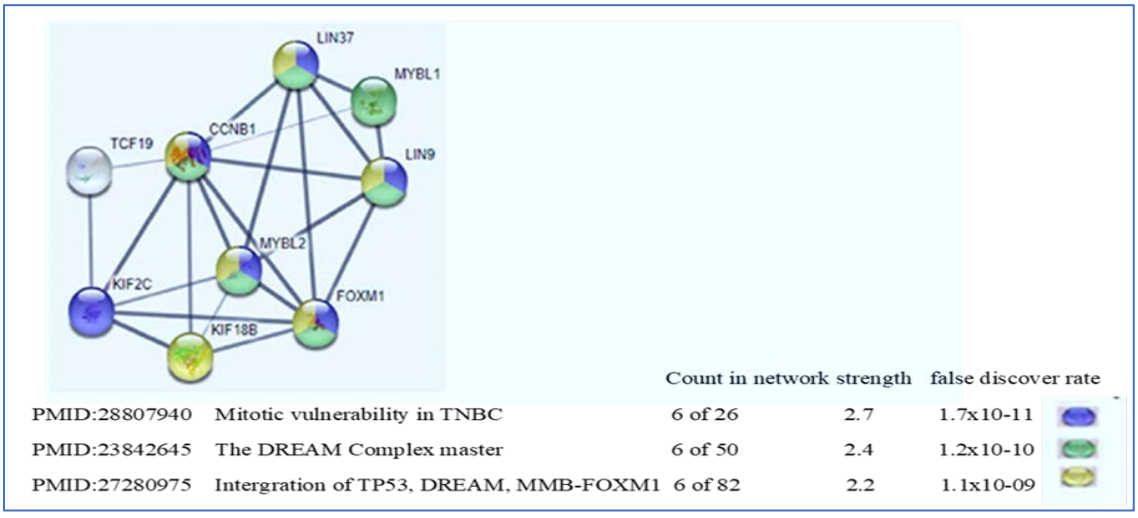


Figure 17: STRING™ Analyses Demonstrating Significantly Relevant Relations Between MYBL1, MYBL2, TCF19 and KIF18b

There are strong relationships between MYBL2 and TCF19 and KIF18b based on published data. MYBL1, MYBL2, TCF19 and KIF18B were entered into STRING™ to determine other genes documented as associated.

MYBL1, MYBL2, TCF19 and KIF18B levels were examined for expression of RNA transcript and protein levels (Figure 18) in untreated MC10A, MCF7 and MDA MB231 cell line preparations. Overall, MYBL1, MYBL2, TCF19 and KIF18B RNA and

protein levels were over-expressed in TNBC compared to the non-tumor MCF10A consistent with the KD data that show co-ordinate expression of the four genes.

KIF18B is over-expressed in MDA MB231. KIF18B is a kinesin motor protein involved in chromosome motility [52], and Wolter et al [53] show that kinesins are regulated by MYBL2. A similar signaling process could occur in TNBC. We can go one step further and speculate that over-expression of KIF18B with MYBL1 and MYBL2 in TNBC contributes to the motility and subsequent metastatic potential of TNBC cells. We are considering experiments that will address this possibility. As for TCF19, both TCF19 [54] and MYBL1 are transcription factors and have been independently shown to be involved in cell cycle signaling events. It could be that MYBL1 (MYBL2) and TCF19 genes cooperate in TNBC cell cycle signaling processes. The differential expression of these genes in TNBC suggest their involvement in the cancers. Whether or not this is a strong driving force must be determined.

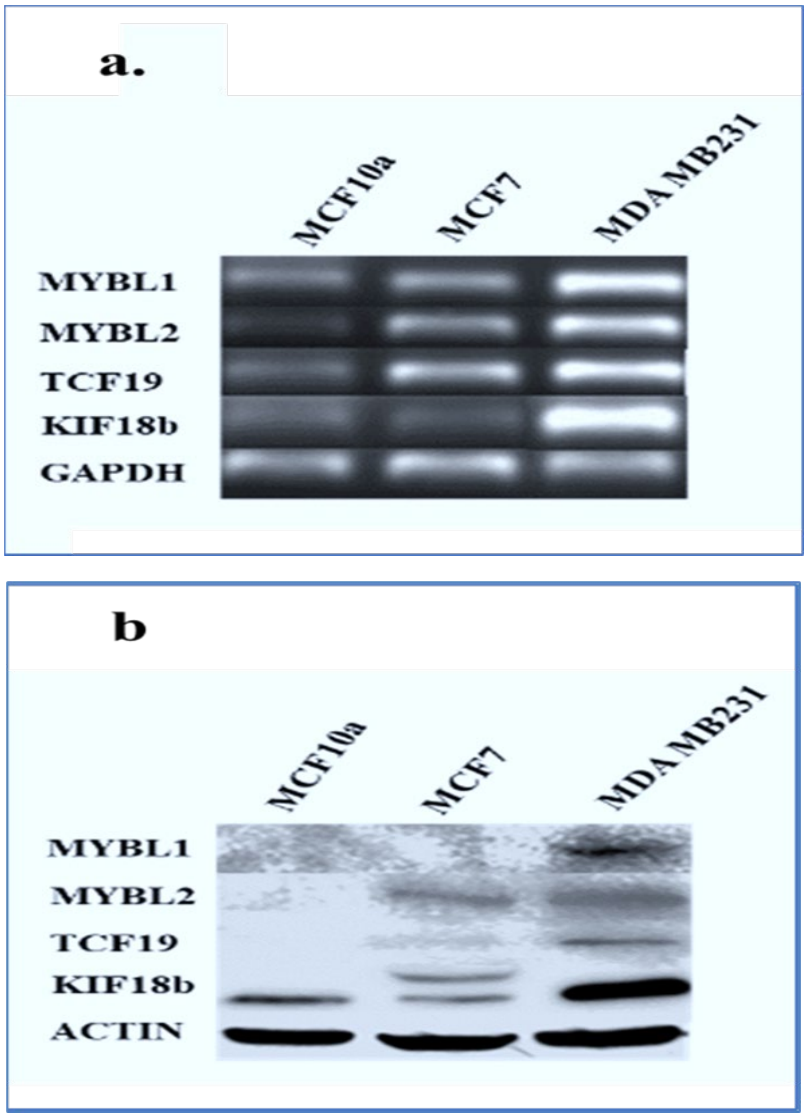


Figure 18: **PCR and Western Experimental Gel Analysis of Candidate Genes Using Cell Line Samples**

(a) transcript analyses of MYBL1, MYBL2, TCF19, KIF18B and GAPDH control in MCF10A (basal non tumor), MCF7 (luminal) and MDA MB231 (TNBC).

(b) protein analyses of MYBL1, MYBL2, TCF19, KIF18B and actin control in MCF10A (basal non tumor), MCF7 (luminal) and MDA MB231 (TNBC).

As further validation, MYBL1, MYBL2, TCF19 and KIF18B transcript levels were examined in GEO (previously published) patient datasets (figure 19). The results were like

observed in the cell line preparations. MYBL1, MYBL2, TCF19 and KIF18B were over-expressed in cancer patient samples compared to the non-tumor transcript samples.

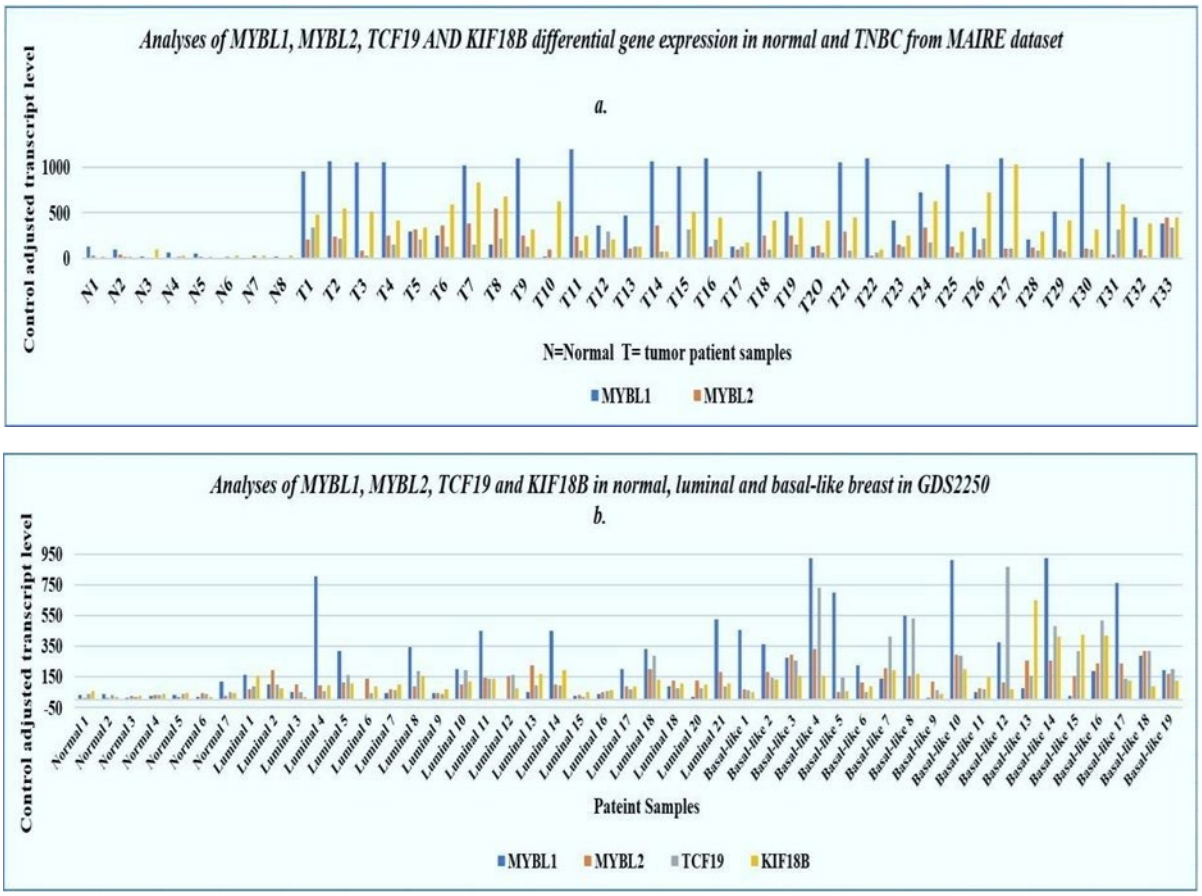


Figure 19: Analysis of Candidate Genes in Patient Samples

(a) Maire TNBC dataset including normal and TNBC patient samples analyzed for MYBL1, MYBL2, TCF19 AND KIF18B.

(b) Analysis of MYBL1, MYBL2, TCF19 and KIF18B in patient transcript datasets. GDS2250 patient dataset retrieved from GEO including normal, luminal, and basal-like TNBC samples.

Analyses of Genes Identified in Part I Compared to Part 2 of the Study

Certain genes identified as common to knock-down of ESR1 in MCF7 and cMYB in MCF7 were examined to determine if the genes were affected by knockdown of MYBL1 in TNBC. The answer is “No” (Table 5).

Of the original gene list, DDX58 and FBX08 were affected by knock-down of MYBL1 in TNBC, but contrary to results observed in the Part 1 study, both genes were up-regulated when MYBL1 was knock-down (in Part 2). DDX58 and FBX08 were affected by MYBL1 KD, but direction of differential expressed differed. In Part 1, DDX58 and FBX08 was down-regulated when MYBL1 was down-regulated. In Part 2, DDX58 and FBX08 is up-regulated with MYBL1 KD which suggests a different pattern of regulation or involvement in different signaling mechanisms.

In Part 1 of the study, E2F7 transcription factor was affected by knock-down of ESR1 in MCF7 and cMYB in MCF7 and followed a pattern similar to MYBL1 in knock-down and untreated cell preparations. Although E2F7 was not affected by MYBL1 knock-down in Part 2, a transcription factor family member, E2F4 was affected by knock-down of MYBL1 in Part 2 (Table 6). STRINGTM bioinformatic analyses show protein: protein interactions between MYBL1, E2F4 and E2F7, so there is experimental evidence of a relationship between MYBL1, E2F4 and E2F7. STRINGTM also show involvement of p53 with these genes, and p53 was also affected by MYBL1 knock-down (Figure 20).

Table 5: Original List of 13 Genes Analyzed for Pattern of Expression in Part 2 Conditions

Gene Description	Symbol	scramble	LVA KD
V-Myb A vian Myeloblast Viral Oncogene Homolog Like 1	MYBL1	12.4	10.3 (-4.06)
Activating transcription factor 3	ATF3	NC	NC
DExD/H-Box Helicase 58	DDX58	10.95	11.98 (+2.05)
Elongation factor 6	E2F6	NC	NC
Elongation factor 7	E2F7	NC	NC
F-Box and Leucine Rich Repeat Protein 13	FBXL13	NC	NC
F-Box Protein 8	FBXO8	11.32	12.9 (-2.97)
F-Box Protein 32	FBXO32	NC	NC
Long Intergenic Non-Protein Coding RNA 673	Linoo673	NC	NC
Nuclear Factor, Erythroid 2 Like 3	NFE2L3	NC	NC
PR/SET Domain 5	PRDM5	NC	NC
Phosphoserine Phosphatase	PSPH	NC	NC
Zinc Finger Protein 75a	ZN75A	NC	NC
Zinc Finger Protein 436	ZN436	NC	NC

Table 6: Data from MYBL1 knock-down in the MDA MB231 Cells

E2F4 and TP53 were Down-regulated. These Data Have Not Been Experimentally Validated.

Scramble Transcript Level LOG 2 Values	MYBL1 shRNA	Fold Change	Gene
12.39	10.79	-3.9x	E2F4
13.12	10.78	-5.07x	TP53

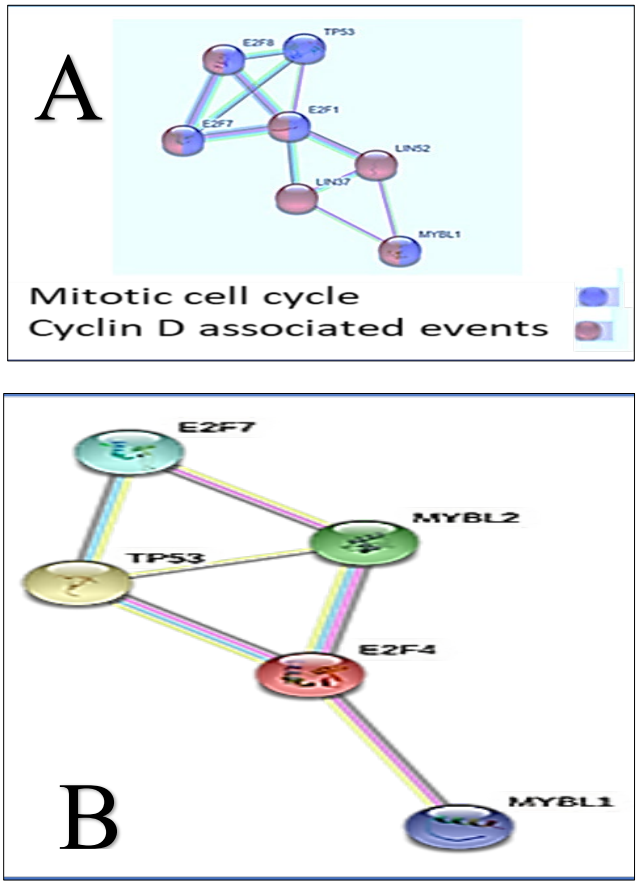


Figure 20: STRING™ Protein

Protein interaction determines a relationship between E2F4 (identified in Part 2 study) and E2F7 (identified in Part 1 study). There is also protein: protein interaction between E2F factors and MYBL1 and MYBL2 with TP53 in cell cycle signaling.

For breast cancers, it could be that E2F7 functions with MYBL1 in MCF7 luminal cells and E2F4 functions with MYBL1 in TNBC. Published data validate co-operation between E2F4 and MYBL1 and MYBL2 in cell cycle processes dream [26, 46, 55], so this part of the statement has been validated. Gene alignment studies show a 33% similarity between E2F4 and E2F7, and common functional domains in both E2F factors (data not shown). The E2F transcription factor data are purely speculative at this point; because even though E2F4 showed knockdown in TNBC, and previous published data show co-expression with MYBL1 and E2F4, the E2E4 microarray data presented here was not experimentally validated. Microarray data presented here show that KD of MYBL1 leads to down-regulation of E2F4, but these data cannot address the mode of regulation. The current study suggests that MYBL1 signaling processes are different in MCF7 and TNBC.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The goal of this study is to (a) identify genes either regulated by or with MYBL1 in breast cancer in MDA MB231 TNBC cells (b) to further characterize TNBC and (c) begin to address the role of MYBL1 in these cancers. Data by other investigators suggest that the MYBL1 gene is involved in tumor progression in different cancers. Because MYBL1 is an oncogene, and associated with cell cycle regulation, apoptosis, and differentiation, we see it as a suitable biomarker to study in TNBC. The current study is only the beginning. We are considering two approaches to further characterize MYBL1 in TNBC. First, we will repeat the KD study as it is described here, but we will validate our panel of genes utilizing additional cell lines and patient sample datasets. Second, the MYBL1 KD studies can be expanded to examine the ability of MYBL1 to produce tumors in a mice model. These studies are instrumental towards validating the role of MYBL1 in tumor development. Following transplantation, if MYBL1 plays a role in tumor development, KD of the gene will prevent tumor formation in the mice model.

For the current study, our data show that the initial hypothesis was not correct to use DEG in ESR1 KD (MCF7) and cMYB KD (MCF7) datasets to examined and identify genes that might be associated with MYBL1. Our data suggest that the MCF7 DEG are not differentially expressed in TNBC. The Luminal cancer transcriptome is very different from the TNBC transcriptome. Except for the E2F family genes, our data show that genes

associated with MYBL1 in Luminal cancers differ from those associating with MYBL1 in TNBC. We did find that MYBL1 KD in TNBC leads to KD of E2F4 and TP53. In addition, STRINGTM analyses show there are associations between E2F4, E2F7, TP53 along with MYBL1 and MYBL2. Much of the data are not experimentally validated, but we will consider future studies to examine the role of E2F4 and E2F7 in TNBC tumors using additional datasets. For future studies will also consider more stringent statistical selection parameters to select for DEG, similar to the conditions used for the MYBL1 KD in TNBC. For example, for KD of MYBL1 in TNBC we utilized a high, statistical cut-off (of 4-fold) for selection of our candidate genes; a more reliable list of candidate genes can be identified under these conditions.

In addition to the genes examined in this thesis, a number of small RNAs, including microRNA(miRNA) and small nucleolar RNA (snoRNAs) were identified as differentially expressed and could be involved in TNBC signaling mechanisms. Our microarray data revealed mir222, snorA30, miR3942, miR4295, miR4418, and miR3661 as strong candidates to consider for further studies. SnorA30 which is up-regulated by 50x and mir222 is up-regulated by 40x and studies suggest mir222 is regulated by MYBL1 [32]. Nineteen genes were selected to study for their possible relationship to MYBL1 in TNBC. Two genes, TCF19 and KIF18B, were selected for more immediate studies. We are currently examining the possibility that MYBL1, MYBL2, TCF19 and KIF18b are cooperating in key signaling processes in TNBC. Each gene was validated bioinformatically using patient datasets and experimentally based on transcript and protein levels in cell lines. We will continue these studies in addition to other studies with a goal towards characterizing MYBL1 in TNBC.

REFERENCES

1. Hanahan, D. and R.A. Weinberg, The hallmarks of cancer. Cell, 2000. 100(1): p. 57-70.
2. Pandya, S. and R.G. Moore, Breast development and anatomy. Clin Obstet Gynecol, 2011. 54(1): p. 91-5.
3. Weigelt, B., F.C. Geyer, and J.S. Reis-Filho, Histological types of breast cancer: how special are they? Mol Oncol, 2010. 4(3): p. 192-208.
4. Lehmann, B.D., et al., Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest, 2011. 121(7): p. 2750-67.
5. Dawood, S., Triple-negative breast cancer: epidemiology and management options. Drugs, 2010. 70(17): p. 2247-58.
6. Bianchini, G., et al., Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol, 2016. 13(11): p. 674-690.
7. Kim, J., et al., MYBL1 rearrangements and MYB amplification in breast adenoid cystic carcinomas lacking the MYB-NFIB fusion gene. J Pathol, 2018. 244(2): p. 143-150.
8. Player, A., et al., Identification of candidate genes associated with triple negative breast cancer. Genes Cancer, 2017. 8(7-8): p. 659-672.

9. Barrett, T., et al., NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res, 2013. 41(Database issue): p. D991-5.

10. Cicirò, Y. and A. Sala, MYB oncoproteins: emerging players and potential therapeutic targets in human cancer. Oncogenesis, 2021. 10(2): p. 19.

11. Rushton, J.J., et al., Distinct changes in gene expression induced by A-Myb, B-Myb and c-Myb proteins. Oncogene, 2003. 22(2): p. 308-13.

12. Ness, S.A., Myb protein specificity: evidence of a context-specific transcription factor code. Blood Cells Mol Dis, 2003. 31(2): p. 192-200.

13. Dai, X., et al., Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. J Cancer, 2017. 8(16): p. 3131-3141.

14. Franchini, G., et al., Structural organization and expression of human DNA sequences related to the transforming gene of avian myeloblastosis virus. Proc Natl Acad Sci U S A, 1983. 80(24): p. 7385-9.

15. Li, X.Z., et al., An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol Cell, 2013. 50(1): p. 67-81.

16. Boyer, L.A., et al., Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes. Mol Cell, 2002. 10(4): p. 935-42.

17. Ko, E.R., et al., A conserved acidic patch in the Myb domain is required for activation of an endogenous target gene and for chromatin binding. Mol Cancer, 2008. 7: p. 77.

18. Marchler-Bauer, A., et al., CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res, 2017. 45(D1): p. D200-d203.

19. Rushton, J.J. and S.A. Ness, The conserved DNA binding domain mediates similar regulatory interactions for A-Myb, B-Myb, and c-Myb transcription factors. Blood Cells Mol Dis, 2001. 27(2): p. 459-63.

20. Zhou, Y. and S.A. Ness, Myb proteins: angels and demons in normal and transformed cells. Front Biosci (Landmark Ed), 2011. 16: p. 1109-31.

21. Liu, L.Y., et al., A supervised network analysis on gene expression profiles of breast tumors predicts a 41-gene prognostic signature of the transcription factor MYB across molecular subtypes. Comput Math Methods Med, 2014. 2014: p. 813067.

22. Liu, X., et al., Reassessing the Potential of Myb-targeted Anti-cancer Therapy. J Cancer, 2018. 9(7): p. 1259-1266.

23. `Ziebold, U. and K.H. Klempnauer, Linking Myb to the cell cycle: cyclin-dependent phosphorylation and regulation of A-Myb activity. Oncogene, 1997. 15(9): p. 1011-9.

24. Marhamati, D.J., et al., A-myb is expressed in bovine vascular smooth muscle cells during the late G1-to-S phase transition and cooperates with c-myc to mediate progression to S phase. Mol Cell Biol, 1997. 17(5): p. 2448-57.

25. Rashid, N.N., H.A. Rothan, and M.S. Yusoff, The association of mammalian DREAM complex and HPV16 E7 proteins. Am J Cancer Res, 2015. 5(12): p. 3525-33.

26. Engeland, K., Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM. Cell Death Differ, 2018. 25(1): p. 114-132.

27. Togashi, Y., et al., MYB and MYBL1 in adenoid cystic carcinoma: diversity in the mode of genomic rearrangement and transcripts. Mod Pathol, 2018. 31(6): p. 934-946.

28. Xu, L.H., et al., MYB promotes the growth and metastasis of salivary adenoid cystic carcinoma. Int J Oncol, 2019. 54(5): p. 1579-1590.

29. Arsura, M., et al., A-myb rescues murine B-cell lymphomas from IgM-receptor-mediated apoptosis through c-myc transcriptional regulation. Blood, 2000. 96(3): p. 1013-20.

30. Golay, J., et al., Expression of A-myb, but not c-myb and B-myb, is restricted to Burkitt's lymphoma, sIg+ B-acute lymphoblastic leukemia, and a subset of chronic lymphocytic leukemias. Blood, 1996. 87(5): p. 1900-11.

31. Kyrpychova, L., et al., Small Subset of Adenoid Cystic Carcinoma of the Skin Is Associated With Alterations of the MYBL1 Gene Similar to Their Extracutaneous Counterparts. Am J Dermatopathol, 2018. 40(10): p. 721-726.

32. Gorbatenko, A., et al., HER2 and p95HER2 differentially regulate miRNA expression in MCF-7 breast cancer cells and downregulate MYB proteins through miR-221/222 and miR-503. Sci Rep, 2019. 9(1): p. 3352.

33. Muthukaruppan, A., et al., Multimodal Assessment of Estrogen Receptor mRNA Profiles to Quantify Estrogen Pathway Activity in Breast Tumors. Clin Breast Cancer, 2017. 17(2): p. 139-153.

34. Thorner, A.R., et al., Potential tumor suppressor role for the c-Myb oncogene in luminal breast cancer. PLoS One, 2010. 5(10): p. e13073.

35. Gonda, T.J., P. Leo, and R.G. Ramsay, Estrogen and MYB in breast cancer: potential for new therapies. Expert Opin Biol Ther, 2008. 8(6): p. 713-7.

36. Brea, M.S., et al., Epidermal Growth Factor Receptor Silencing Blunts the Slow Force Response to Myocardial Stretch. J Am Heart Assoc, 2016. 5(10).

37. Dalma-Weiszhausz, D.D., et al., The affymetrix GeneChip platform: an overview. Methods Enzymol, 2006. 410: p. 3-28.

38. Maire, V., et al., TTK/hMPS1 is an attractive therapeutic target for triple-negative breast cancer. PLoS One, 2013. 8(5): p. e63712.

39. Koressaar, T. and M. Remm, Enhancements and modifications of primer design program Primer3. Bioinformatics, 2007. 23(10): p. 1289-91.

40. Wang, T. and T.S. Furey, Analysis of complex disease association and linkage studies using the University of California Santa Cruz Genome Browser. Circ Cardiovasc Genet, 2009. 2(2): p. 199-204.

41. Agarwal, V., et al., Predicting effective microRNA target sites in mammalian mRNAs. Elife, 2015. 4.

42. Mi, H., et al., PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res, 2019. 47(D1): p. D419-d426.

43. Szklarczyk, D., et al., The STRING^{TM} database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res, 2017. 45(D1): p. D362-d368.

44. Carey, L.A., et al., Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama, 2006. 295(21): p. 2492-502.

45. Kent, L.N. and G. Leone, The broken cycle: E2F dysfunction in cancer. Nat Rev Cancer, 2019. 19(6): p. 326-338.

46. Fischer, M. and G.A. Müller, Cell cycle transcription control: DREAM/MuvB and RB-E2F complexes. Crit Rev Biochem Mol Biol, 2017. 52(6): p. 638-662.

47. Fischer, M., et al., Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. Nucleic Acids Res, 2016. 44(13): p. 6070-86.

48. Liu, Z.L., et al., Expressions and prognostic values of the E2F transcription factors in human breast carcinoma. Cancer Manag Res, 2018. 10: p. 3521-3532.

49. Litovchick, L., et al., Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. Mol Cell, 2007. 26(4): p. 539-51.

50. Zeng, C.X., et al., TCF19 enhances cell proliferation in hepatocellular carcinoma by activating the ATK/FOXO1 signaling pathway. Neoplasma, 2019. 66(1): p. 46-53.

51. Wu, Y., et al., KIF18B promotes tumor progression through activating the Wnt/β-catenin pathway in cervical cancer. Onco Targets Ther, 2018. 11: p. 1707-1720.

52. Ji, Z., et al., KIF18B as a regulator in microtubule movement accelerates tumor progression and triggers poor outcome in lung adenocarcinoma. Tissue Cell, 2019. 61: p. 44-50.

53. Wolter, P., et al., Central spindle proteins and mitotic kinesins are direct transcriptional targets of MuvB, B-MYB and FOXM1 in breast cancer cell lines and are potential targets for therapy. Oncotarget, 2017. 8(7): p. 11160-11172.

54. Zhou, Z.H., et al., TCF19 contributes to cell proliferation of non-small cell lung cancer by inhibiting FOXO1. Cell Biol Int, 2019.

55. Iness, A.N., et al., The cell cycle regulatory DREAM complex is disrupted by high expression of oncogenic B-Myb. Oncogene, 2019. 38(7): p. 1080-1092.