AN INVESTIGATION OF THE MOLECULAR MECHANISMS INVOLVED IN THE GENERATION OF BREAST CANCER

by

JOSEPH TYSON DEANGELIS

TRYGVE O. TOLLEFSBOL, CHAIR
ANDRA FROST
R. DOUGLAS WATSON
DANNY WELCH
THANE WIBBELS

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

BIRMINGHAM, ALABAMA
2010
AN INVESTIGATION OF THE MOLECULAR MECHANISMS INVOLVED IN THE
GENERATION OF BREAST CANCER

JOSEPH TYSON DEANGELIS

BIOLOGY

ABSTRACT

Cell culture models of oncogenesis that use cellular reprogramming to generate a neoplastic cell from a normal cell provide one of the few opportunities to study the early stages of breast cancer development. Human mammary epithelial cells (HMECs) were induced to undergo a neoplastic transformation using defined genetic elements to generate transformed HMECs (THMECs). We conducted proteomic and transcriptomic analysis of three points in the neoplastic progression of breast cancer. First we monitored alterations in the proteome that revealed a progressive trend towards a metastatic state. Next, we identified a set of genes from transcriptomic analysis that displayed this same type of progressive trend and using pathway analysis illustrated that a portion of these genes have been shown to be regulated by hepatocyte nuclear factor 4a (HNF4a). We also illustrated using network analysis that HNF4a is associated with a large number of genes expressed in breast cancer brain metastasis. Lastly, from our transcriptomic data we identified 63 epigenetic modulating enzymes that are differentially expressed following hRASV12 introduction. The differential expression of epigenetic modulators on a global level may be an initial event in oncogenesis. Future studies will be aimed at identifying the exact role of HNF4a in breast cancer metastasis, understanding the effect of drastic alterations in the expression of epigenetic modulators and its role in oncogenesis, and elucidation of the differences at the genomic and transcriptomic levels of THMEC cells and very small embryonic like stem cells.
DEDICATION

I would like to dedicate this dissertation to my family members who have helped me get to where I am at today: Carsyn, Allie, Mom, and Dad.
ACKNOWLEDGMENTS

I would like to thank my mentor Dr. Trygve Tollefsbol for the freedom he allowed me to have and the countless opportunities he provided me, opportunities that few graduate students get. I would like to thank the postdocs in my lab, Dr. Li, Dr. Liu, and Dr. Meeran, who helped me understand the intricate details of my experiments. I would also like to thank the undergraduate students who helped me with my research and kept me on my toes at all times: Natalie, Lila, Su, Anna, Lana, Emily, Woody, and Mike. I would like to thank the faculty and graduate students of the Biology department who have put up with me for the better part of a decade and showed me what science is truly about. I would like to thank my committee members for their guidance during the pursuit of my degree. I would especially like to thank Dr. Wibbels, for giving my first taste of research as an undergraduate, and Dr. Welch, for teaching me what I learned at cancer biology journal club.

I would like to thank the directors and fellows of the Cancer Prevention and Control Training Program. I can’t even begin to describe how valuable the CPCTP has been to me. The insightful career advice, exposure to clinical and public health based research, and the funding are just a few of the things I will be forever grateful for.

Lastly, I would like to thank the women of Komen, especially Kathy Stein, Beverly Laird, and Rebecca Di Piazza. It is your words and actions that have inspired me and illustrated to me what the word hero really means. It is the drive and passion that I learned from Komen that I will forever carry with me.
TABLE OF CONTENTS

Page

ABSTRACT ........................................................................................................................................ iii

DEDICATION ....................................................................................................................................... iv

ACKNOWLEDGMENTS ...................................................................................................................... v

LIST OF TABLES .................................................................................................................................. vii

LIST OF FIGURES .............................................................................................................................. viii

LIST OF ABBREVIATIONS .................................................................................................................. x

GENERAL INTRODUCTION ............................................................................................................... 1

2D DIFFERENCE GEL ELECTROPHORESIS ANALYSIS OF DIFFERENT TIME POINTS DURING THE COURSE OF NEOPLASTIC TRANSFORMATION OF HUMAN MAMMARY EPITHELIAL CELLS ................................................................. 13

IDENTIFICATION OF A POTENTIAL ROLE OF HNF4A IN BREAST CANCER BRAIN METASTASIS .............................................................................................................................. 59

HYPERMETHYLATION AND ONCOGENESIS ..................................................................................... 90

AN OVERVIEW OF EPIGENETIC ASSAYS ....................................................................................... 114

THE EXPRESSION PROFILES OF EPIGENETIC MODULATING GENES UNDERGO DRASTIC ALTERATIONS DURING BREAST TUMORIGENESIS .... 129

GENERAL DISCUSSION ...................................................................................................................... 144

GENERAL LIST OF REFERENCES ...................................................................................................... 148

APPENDIX ........................................................................................................................................ 155

vi
LIST OF TABLES

<table>
<thead>
<tr>
<th>Tables</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D DIFFERENCE GEL ELECTROPHORESIS ANALYSIS OF DIFFERENT TIME POINTS DURING THE COURSE OF NEOPLASTIC TRANSFORMATION OF HUMAN MAMMARY EPITHELIAL CELLS</td>
<td></td>
</tr>
<tr>
<td>1. Gel Assignments for 2-DIGE Experiments</td>
<td>33</td>
</tr>
<tr>
<td>2. Proteomic and Statistical Data of Spots that were Significantly Different in Abundance over the Course of Breast Tumorigenesis</td>
<td>37</td>
</tr>
<tr>
<td>IDENTIFICATION OF A POTENTIAL ROLE OF HNF4A IN BREAST CANCER BRAIN METASTASIS</td>
<td></td>
</tr>
<tr>
<td>1. Molecular and cellular functions associated with gene list BCT1</td>
<td>88</td>
</tr>
<tr>
<td>2. Physiological and developmental functions associated with gene list BCT1</td>
<td>89</td>
</tr>
<tr>
<td>HYPERMETHYLATION AND ONCOGENESIS</td>
<td></td>
</tr>
<tr>
<td>1. Impacts of Hypermethylation on Oncogenesis</td>
<td>106</td>
</tr>
<tr>
<td>THE EXPRESSION PROFILES OF EPIGENETIC MODULATING GENES UNDERGO DRASTIC ALTERATIONS DURING BREAST TUMORIGENESIS</td>
<td></td>
</tr>
<tr>
<td>1. Gene Cluster 1</td>
<td>140</td>
</tr>
<tr>
<td>2. Gene Cluster 2</td>
<td>141</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

**Figure**  | **Page**
---|---

**INTRODUCTION**
1. Neoplastic Transformation .......................................................... 5
2. Effects of DNA methylation ............................................................ 8

**2D DIFFERENCE GEL ELECTROPHORESIS ANALYSIS OF DIFFERENT TIME POINTS DURING THE COURSE OF NEOPLASTIC TRANSFORMATION OF HUMAN MAMMARY EPITHELIAL CELLS**
1. Confirmation of gene transfer and neoplastic transformation .................. 30
2. Two-dimensional difference gel electrophoresis of protein extracts from SHMEC, THMEC 40d, THMEC 80d .......................................................... 34
3. Protein abundance graphs of significantly differentially expressed proteins identified by 2D-DIGE ........................................................................ 39
4. Real-Time PCR analysis of mRNA expression of maspin (SERPINB5) in THMECs and the MCF10AT Metastatic Progression Model. .................... 41
5. Network analysis of differentially expressed proteins ................................ 44
6. Schematic of cell culture experimental design ........................................... 58

**IDENTIFICATION OF A POTENTIAL ROLE OF HNF4A IN BREAST CANCER BRAIN METASTASIS**
1. Experimental overview ........................................................................... 80
2. Graph of SERPINB5 expression in neoplastically transforming human mammary epithelial cells. ........................................................................ 81
3. Schematic of Filters applied to generate gene list BCT1 .......................... 82
4. Molecular network, T1, generated from gene list BCT1 ............................ 83
5. HNF4a and TGFβR1 expression in neoplastically transforming human mammary epithelial cells. ........................................84

6. Molecular network, M1, involved in brain metastasis ..........................85

7. Schematic of network relationships.....................................................86

8. Molecular Supranetwork, SM1, generated from network merge of linked networks..................................................................................87

HYPERMETHYLATION AND ONCOGENESIS

1. Possible cyclic mechanisms by which aberrant methylation patterns are initiated and preserved. .................................................................98

THE EXPRESSION PROFILES OF EPIGENETIC MODULATING GENES UNDERGO DRASTIC ALTERATIONS DURING BREAST TUMORIGENESIS

1. Unsupervised hierarchical clustering of differentially expressed epigenetic modulating genes in neoplastically transforming human mammary epithelial cells ........................................................................................................138

2. Profile plots of differentially expressed epigenetic modulating genes ..........139
LIST OF ABBREVIATIONS

2D-DIGE  two-dimensional differential in gel electrophoresis
ANOVA  analysis of variance
APS  ampicillan-streptomycin-penicillin
B-H  Benjamini-Hotchberg
BVA  Biological Variation Analysis
CDK  cyclin-dependant kinase
CHARM  Comprehensive Arrays for Relative Methylation Analysis
ChIP  chromatin immunoprecipitation
CV  coefficient of variation
DIA  Differential in-Gel Analysis
DMR  region of DNA methylation variation
DNAm  DNA methylation
DNMT  DNA methyltransferase
ECL  enhanced chemiluminescence
ECM  extracellular matrix
EGCG  (-)-epigallocatechin gallate
EMT  epithelial to mesenchymal transition
ExonRMA  Exon Level Robust Multiarray Algorithm
FC  fold change
FDR  false discovery rate
GEO  Gene Expression Omnibus
HAT  histone acetyltransferase
HDAC histone deacetylase
HMEC human mammary epithelial cells
HMT histone methyltransferase
HNF hepatocyte nuclear factor
IEF isoelectric focusing
IPA Ingenuity Pathway Analysis
IS internal standard
MBD methyl binding domain
MBDP methyl binding domain protein
MEGM mammary epithelial growth medium
mirRNA micro RNA
MMP matrixmetalloproteinases
MODY Maturity On-set Diabetes of the Young
MSG metastatic suppressor gene
MSP methylation specific PCR
MS-SSCA methylation sensitive-single strand conformation analysis
NCBI National Center for Bioinformatics
nChIP native ChIP
PML promyelocytic
Q-ChIP quantitative ChIP
RAR Retinoic Acid Receptor
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcriptase PCR</td>
</tr>
<tr>
<td>SAM</td>
<td>S-Adenosylmethionine</td>
</tr>
<tr>
<td>SHMEC</td>
<td>HMEC-SV40-hTERT</td>
</tr>
<tr>
<td>shRNA</td>
<td>short hairpin RNA</td>
</tr>
<tr>
<td>siRNA</td>
<td>small interfering RNA</td>
</tr>
<tr>
<td>SNuPE</td>
<td>single nucleotide primer extension</td>
</tr>
<tr>
<td>TGFβ</td>
<td>transforming growth factor beta</td>
</tr>
<tr>
<td>THMEC</td>
<td>Transformed HMEC</td>
</tr>
<tr>
<td>TRAP</td>
<td>Telomeric Repeat Amplification Protocol</td>
</tr>
<tr>
<td>TSA</td>
<td>trichostatin A</td>
</tr>
<tr>
<td>UniProtKB</td>
<td>UniProt Knowledgebase</td>
</tr>
<tr>
<td>VEGF</td>
<td>vascular epithelial growth factor</td>
</tr>
<tr>
<td>xChIP</td>
<td>conventional ChIP</td>
</tr>
<tr>
<td>φnx</td>
<td>Pheonix gag-pol cells</td>
</tr>
</tbody>
</table>
INTRODUCTION

The Puzzle Problem

Cancer is an enigma. Cancer is an encompassing term for over 100 diseases that affect all races, all ages, both sexes, can occur in nearly every tissue type in the human body and is one of the most complex biomedical problems in existence. To highlight the complexity of the biological nature of cancer, take for example five of the requirements for life: 1) growth and development, 2) respiration and metabolism, 3) reproduction, 4) interaction with the environment and response to stimuli, and 5) genetic adaption. All of these characteristics are displayed by cancers and the only difference between a true living organism and cancer is a lack of a homeostasis. It is this high degree of similarity to a living organism that provides cancers with their intricacy, complexity, variability as well as their resiliency.

For more than three decades, cancer research has been one of the most funded avenues of biomedical research, yet the National Cancer Institute estimates that in 2010 there will be an average of 1500 cancer-related deaths per day.¹ And although we have observed a decline in the incidence and mortality of some cancers, the incidence and mortality of others are on the rise.² What is most interesting is that the observed decline in some cancer types cannot be tied to a specific groundbreaking discovery from basic science researchers, but rather is the result of an increase in prevention-based education and access to healthcare.
So what then is inhibiting the discovery that will significantly reduce the number of cancer related deaths? Some suggest it is due to the estimated 10 year period before a cancer patient can reap the benefits of a bench side discovery, which makes it hard to identify which discoveries are most significant. Others blame a system that hinders the funding of high risk-high reward research. Despite these attempts to rationalize the lack of a groundbreaking discovery that would substantially reduce the number of cancer related deaths, the fact remains that cancer has proven to be too difficult to conquer. At what point do we decide a new approach is necessary?

The nature of scientific research causes researchers to specialize and it is in this unavoidable truth that I believe the true problem that cancer researchers face can be found, a problem which I have entitled “The Puzzle Problem.” Each cancer researcher does his or her part to best understand their individual piece of the giant puzzle known as cancer. Over time each individual cancer researcher follows the path where their data takes them, on a quest to master their own piece of the puzzle. Each individual puzzle piece provides insight into the shape and size of the surrounding pieces, yet in itself lacks the information needed to construct the puzzle as a whole. Some researchers may acquire multiple pieces of the puzzle over time; however, no one researcher has yet to hold enough pieces to construct the entire puzzle. It is in the construction of the entire puzzle and the placement of each piece within that puzzle that the problem arises. In most cases, the placing of each piece into the puzzle is based on assumption, as the fact remains that we simply do not know what cancer truly is. In other words, there’s nothing wrong with each individual piece of the puzzle, it’s how we are fitting them together that is the
problem. This dissertation will involve the investigation of multiple individual puzzle pieces as well as provide an alternative theory for the puzzle as a whole.

**Oncogenesis and Metastasis**

One of the least investigated aspects of cancer is oncogenesis. We will define oncogenesis as the period of time during which a cell acquires the phenotypic characteristics associated with cancer. Although no two cancers are alike, the phenotypic characteristics of cancer can be grouped into six major hallmarks: insensitivity to antigrowth signals, self-sufficiency in growth signals, replicative immortality, evasion of apoptosis, angiogenesis, and invasion and metastasis. A more detailed explanation of the hallmarks of cancer can be found in Chapter 5. Each hallmark is essential to the development of a cancer cell and each can be achieved through any given number of dysregulated pathways. In particular, invasion and metastasis provide cancer cells with the more aggressive features and are more closely associated with short term survival than any of the other hallmarks. Yet, this hallmark is one of the least investigated aspects of cancer because by definition it involves migration of cells within an organism as whole, making it very difficult to accurately investigate metastatic pathways *in vitro*. Over the last decade there has been much debate over the origin of metastatic behavior. Certain cancer cells display a higher metastatic potential than others and the processes that are encouraging metastatic behavior in these cells remain beyond our understanding. Chapters 1 and 2 will involve the analysis of metastasis-linked gene expression during breast oncogenesis. To study the process of breast oncogenesis we will utilize a cell-
culture based model that produces neoplastically transformed breast cancer cells from human mammary epithelial cells.\textsuperscript{5}

**Neoplastic Reprogramming**

In 1999, Weinberg and colleagues published a landmark procedure, which showed that normal, somatic human cells can be induced to undergo a neoplastic transformation through the retroviral-mediated addition of three defined genetic elements: Simian Virus 40 Early Region (\textit{SV40 ER}), exogenous human telomerase reverse transcriptase (\textit{hTERT}), and an activated, oncogenic form of hRAS (\textit{hRASV12}).\textsuperscript{6} Addition of each of the three genes either activates or inactivates specific pathways in a manner similar to how these pathways become dysregulated during oncogenesis in vivo. The addition of \textit{SV40 ER} leads to an inactivation of tumor suppressors p53 and pRB.\textsuperscript{7} Stable expression of exogenous \textit{hTERT}, the catalytic component of telomerase, promotes an increase in telomerase activity and provides the cell with the ability to replicate indefinitely.\textsuperscript{8} Lastly, the introduction of an activated hRAS mimics a sustained supply of growth signals and causes a consistent activation of one of the major pathways responsible for cellular division.\textsuperscript{6} Figure 1 is a schematic of the neoplastic transformation model.

Three years after the initial publication of the transformation protocol, Elenbaas and colleagues illustrated that using the same procedures, Human Mammary Epithelial Cells (HMECs) can be induced to undergo the same type of neoplastic transformation as shown in other cells.\textsuperscript{5} It is now understood that for each cell type, specific genetic changes must occur.\textsuperscript{9} Certain cell lines require activation/inactivation of specific pathways that other cell lines do not need to have altered for successful transformation to
occur. What is truly remarkable about this procedure is that the addition of SV40 ER, hTERT, and hRASV12 to a specific cell line results in the dysregulation of the pathways that specific cell line requires for full transformation.9

Addition of SV40 ER is the first step in the transformation procedure. SV40 ER codes for both large T (LT) and small t (ST) antigens.7 By binding to and inactivating p53, LT serves to disable one of the most effective self defense mechanisms of the cell. Sequestering of p53 prevents p53 from asserting its growth suppressive function as well as initiating cell death mechanisms, such as apoptosis.10 In addition to p53, LT also disrupts the cell cycle regulating abilities of pRb. Essential to the G1→S transition, pRb regulates cell division through its interactions with E2F-1.11 Primarily known as a transcriptional activator, E2F-1 binds to promoter regions of many genes known to play a role in cell division, thus initiating transcription of the downstream gene.12 In normal

Figure 1. Neoplastic Transformation of Human Mammary Epithelial Cells. Stable addition of three genetic elements, SV40ER, hTERT, and hRAS-V12, to Human mammary epithelial cells (HMECs) produces Transformed HMECs (THMECs). THMECs display most if not all of the features of breast cancer.
cells, pRb carries out its role in the G1 checkpoint by binding to and inactivating E2F-1. Once the proper signal is given, pRb is phosphorylated and releases from E2F-1. Free from pRb, E2F-1 is now able to carry out its role as a transcriptional activator. More than 50% of breast cancers display inactivating mutations in the gene which codes for p53 and approximately 25% of breast cancers display mutations in the Rb gene.\textsuperscript{13, 14} As a result, abnormal cellular events which would usually elicit a self-destruct response or cell cycle arrest from the cell, go unnoticed. \textit{Through the addition of the SV40 ER, these mutations are effectively simulated.}

The role of the ST antigen in the transformation process was initially thought to be minor. Recently, this role has been shown to be just as imperative for successful transformation as that of LT. Despite this importance, the function of ST is a simple one: inactivation of protein phosphatase 2a (PP2A).\textsuperscript{7} The dephosphorylating duties of PP2A modulate almost every aspect of cellular biology which becomes dysregulated as a result of oncogenesis. Possibly the most critical of the effects asserted by PP2A inactivation is the stabilization and activation of the potent oncoprotein, c-Myc.\textsuperscript{15} With PP2A no longer able to modulate specific signal transduction pathways, supranormal levels of c-Myc result.

Following stable expression of SV40 ER, the next step in the procedure is the introduction of exogenous hTERT. Stable expression of exogenous hTERT results in an upregulation of telomerase activity and cellular immortalization. Now able to escape replicative senescence, the third and final step of the transformation process is the addition of an activated form of RAS, known as hRASV12. In normal cells, RAS plays a critical role in cell cycle regulation. Through serine/threonine phosphorylation of other
proteins, RAS mediates signal transduction pathways that are initiated by tyrosine kinases receptors. RASV12 carries a mutation that prevents inhibitory signals from regulating RAS mediated pathways. Other functions asserted by addition of an activated form of h-RAS include an increase in both the metastatic and angiogenic potential, as well as an increase in mitochondrial metabolism. Although only 5% of breast cancers display mutations in the hRAS gene, dysregulation of Ras pathways is a common occurrence in breast cancers.

**Epigenetic Mechanisms of Gene Regulation**

Epigenetic mechanisms function as the gate keepers by controlling access to DNA, but have no means of actually initiating the processes of transcription themselves. The two canonical epigenetic mechanisms are DNA methylation and histone modifications. Although the role of these mechanisms in development and disease has been known for some time, the true importance of this role has only recently begun to be revealed. For example, genome-wide DNA methylation patterns have recently been shown to be capable of defining the cellular state. Histone modifications and DNA methylation are closely associated and in most instances cooperate to regulate gene expression. Figure 2 illustrates the effects of DNA methylation. In the past year, groundbreaking research has shown that DNA methylation (DNAm) is important not only to the normal development of an individual cell but also in the onset of a diseased state, such as cancer. Feinberg and colleagues, using a novel epigenetic technique entitled comprehensive high-throughput arrays for relative methylation (CHARM), identified regions of the genome that undergo tissue-specific DNAm
alterations, termed regions of DNAm variation (DMRs).\textsuperscript{19, 23} Although the exact medical significance of DMRs has yet to be ascertained, the fact that the methylome can define a cell in fashion similar to how DNA defines the individual is a landmark discovery in itself. Although we know that the methylome undergoes discrete transformations as an individual cell transitions through the stages of differentiation, the mechanisms that drive these epigenomic transformations from one stage to the next remain beyond our understanding. In order to understand what is driving these epigenomic processes, we must first establish an understanding of DNAm patterns and the histone code at specific stages of oncogenesis. Another approach to investigating epigenetic mechanisms is to

**Figure 2. Effects of DNA Methylation along the 5' promoter region.**

**A.** Hypomethylation of 5' promoters allows binding of both repressors and activators and can elicit either an increase or decrease in transcription. **B.** Site specific hypermethylation can lead to increases or decreases in gene transcription depending on the binding site that becomes hypermethylated. **C.** Dense hypermethylation can lead to the binding of Methyl CpG Binding Domain Proteins.
monitor the expression of epigenetic modulating proteins over the course of tumorigenesis.

The enzymes that modulate epigenetic processes fall into one of three groups: Writers, Erasers, and Readers. Writers, or transferases, are the enzymes responsible for establishing the epigenetic code by transferring functional groups to either nucleotides or amino acids. Epigenetic writers include DNA methyltransferases (DNMTs), histone acetyltransferases (HAT), and histone methyltransferases (HMTs). Alterations in both the expression and activity of the DNMTs can lead to drastic changes in gene expression, and a dysregulation of the DNMTs has been shown to contribute to a growing number of pathologies. Although there are many members of the DNMT family; DNMT1, DNMT3a, and DNMT3b carry out the vast majority of DNA methylation, the DNMT3 family is responsible for de novo methylation. Of the HATs, the MYST family has been shown to play a pivotal role in both cancer and normal human development. Histone methylation also occurs at lysine residues but unlike histone acetylation, which involves the transfer of single acetyl groups, HMTs can transfer up to three methyl groups to a single residue. Histone methylation is carried out in part by the SET family of proteins and can elicit either an increase or decrease in gene transcription. Erasers are those enzymes responsible for removing functional groups from histone tails including histone deacetylases (HDACs) and histone demethylases (HDM). The importance of the latter family of enzymes to cancer has only recently been revealed. Three primary classes (I, II, and III) of HDACs exist where classification is based on similarity to yeast proteins. A DNA demethylase has yet to be identified and as such the erasers are comprised entirely of histone modifying enzymes.
Readers are the enzymes that recognize and bind a specific epigenetic mark and are responsible in large part for linking the two epigenetic processes of DNA methylation and histone modifications. Epigenetic readers are much more diverse than either writers or erasers. Examples of histone readers include polycomb proteins, heterochromatin proteins, and DNMT3L. DNMT3L assists DNMT3A by first recognizing a lack of methylation on H3-K4, prior to de novo methylation by DNMT3A. DNA methylation readers are mainly comprised of methyl binding domain (MBD) proteins but also include enzymes such as UHRF1. By binding to newly methylated DNA, MBDs are a major reason DNA methylation tends to be a more stable, persistent, yet still reversible, alteration than chromatin modifications. MBDs physically block access to large areas of DNA, and can have more of an effect on the binding of transcription factors than methylated cytosines alone.

“-omic” Analyses vs Site-Specific Analyses

We are now in the period that researchers have dubbed the “Post-genomic era.” Although multiple versions of the human genome exist and each of these versions is constantly being updated and modified, we now have the ability to monitor changes in gene expression and protein expression in the context of the entire genome or proteome. Even though these technologies are not without their faults, such as prohibitively high costs and a lack of standardized procedures, they have revolutionized the way biomedical research is conducted, cancer research in particular, and in some cases illustrated that some of science’s biggest assumptions were incorrect. For example, methylomic analysis served to identify CpG shores. The discovery of CpG shores drastically changed
the way we view patterns of DNA methylation, as the most important alterations to the methylome were thought to occur only along CpG islands.\textsuperscript{19}

Site-specific or mechanistic studies were for more than half a century the main way biomedical research was conducted; however, with the advent of “-omic” analyses, the role of site-specific analyses has changed. Site-specific analyses provide a detailed understanding of biological mechanisms, but will never be able to determine the importance of a particular biological change in a genomic context. “-omic” analyses on the other hand, are able to illustrate which changes are important in the context of the entire genome, but lacks the precision to understand a specific biological mechanism in its entirety. With regards to “The Puzzle Problem,” site-specific analyses will serve to better understand the makeup of individual puzzle pieces, and the use of “-omic” analyses will assist in the construction of the puzzle as a whole.

Neoplastic transformation produces transformed breast cells that display many if not all of the characteristics of breast cancer cells. The ability to neoplastically transform normal human cells provides the opportunity to monitor the very early events of tumorigenesis in real time. Careful analysis of not only the presence but also the timing of these early events occurring across the proteome, transcriptome, and methylome may allow elucidation of key transient epigenetic and genetic changes that are only seen during oncogenesis, but are undetectable in fully developed cancers. We chose three points in the neoplastic progression of breast cancer to study alterations in the proteome and transcriptome, with the intent of identifying specific pathways or molecules that play a key role in that acquisition of oncogenic phenotype. Using transcriptomic analysis we
have also attempted to understand how the expression of epigenetic modulating enzymes is altered during oncogenesis.

Through the application of –omic analyses we have found that genes and proteins that display alterations in their expression over the entire course of oncogenesis that was measured are associated with an increase in metastatic potential. We have also identified hepatocyte nuclear factor 4a (HNF4a) as being associated with a portion of these molecules as well as with genes that are expressed at higher levels in breast cancer brain metastases than in breast cancer bone metastases. Lastly, we show that the expression of 63 epigenetic modulating enzymes is drastically altered following the addition of hRASV12. The latter finding suggesting that an alteration in the expression of epigenetic modulating enzymes may precede the alterations to the epigenome commonly observed in breast cancers.
2D DIFFERENCE GEL ELECTROPHORESIS ANALYSIS OF DIFFERENT TIME POINTS DURING THE COURSE OF NEOPLASTIC TRANSFORMATION OF HUMAN MAMMARY EPITHELIAL CELLS

by

JOSEPH TYSON DEANGELIS, YUANYUAN LI, NATALIE MITCHELL, LANDON WILSON, HELEN KIM, TRYGVE TOLLEFSBOL

Submitted to *Journal of Proteome Research*

Format adapted for dissertation
ABSTRACT

Cell culture models of oncogenesis that use cellular reprogramming to generate a neoplastic cell from a normal cell provide one of the few opportunities to study the early stages of breast cancer development. Human mammary epithelial cells (HMECs) were induced to undergo a neoplastic transformation using defined genetic elements to generate transformed HMECs (THMECs). To identify proteins that displayed significantly different levels of abundance at three consecutive time points in oncogenesis over an eighty day period, protein extracts were analyzed by two dimensional difference gel electrophoresis (2D-DIGE). Seven proteins were found to be significantly different in abundance: keratin 1, keratin 7, heat shock protein 4A-like, t-complex protein 1, stathmin, gelsolin, and maspin. Keratin 7 and maspin displayed a linear down-regulation over eighty days. All of these proteins are involved in the maintenance of a metastatic state including cytoskeletal modifications and motility. We conclude that following neoplastic induction, THMECs display an early and progressive increase in metastatic potential. Further investigations into the function and regulatory mechanisms of these proteins will provide an unparalleled understanding of the initial states through which a breast cancer cell transitions following acquisition of the genetic abnormalities required for oncogenesis.

Keywords

2D-DIGE, Neoplastic transformation, Human Mammary Epithelial cells, Maspin, Metastasis, Breast cancer, Ingenuity Pathway Analysis
INTRODUCTION

One of the most important yet least understood aspects of cancer biology is oncogenesis, or the process by which a normal, somatic cell transitions to an oncogenic state. There currently is no method to study this process in vivo; however, a procedure known as neoplastic reprogramming confers the ability to study this process in vitro. First used to generate cancer cells from normal fetal lung fibroblasts, this procedure has also been used to generate breast cancer cells from human mammary epithelial cells (HMECs).1,2 HMECs are normal, somatic cells that display a finite lifespan and lack the ability to form tumors. Neoplastically transformed HMECs display most if not all of the characteristics of breast cancer cells, including the ability to form tumors in nude mice.

Two general types of cellular reprogramming exist, reprogramming to a pluripotent or stem state, and reprogramming to a neoplastic state.1,3 These two groups are not mutually exclusive as they both use some of the same genomic regions to achieve the desired induced state and possess many of the newly acquired traits in common. The therapeutic importance of pluripotent reprogramming is very high, whereas neoplastic programming has little therapeutic value. The true value of neoplastic reprogramming lies in the ability to monitor the mechanisms driving the transition of a normal cell to a neoplastic cell in real-time. Neoplastic reprogramming in human cells using defined genetic elements was first discovered by Weinberg et al.1 In 1999, Weinberg and colleagues published a landmark procedure illustrating that normal human cells can be induced to undergo a neoplastic transformation through retroviral-mediated addition of three defined genetic elements: Simian Virus 40 Early Region (SV40 ER), exogenous human telomerase reverse transcriptase (hTERT), and an activated, oncogenic form of
hRAS (hRAS-V12). Addition of each of the three genes either activates or inactivates specific pathways in a manner similar to how these pathways become dysregulated during oncogenesis in vivo. The addition of SV40 ER inactivates the defense mechanisms of the cell by sequestering p53 and pRB. Stable expression of exogenous hTERT, the catalytic component of telomerase, promotes an increase in telomerase activity and provides the cell with the ability to replicate indefinitely. Lastly, the introduction of activated hRAS mimics a sustained supply of growth signals and causes a consistent activation of one of the major pathways responsible for cellular division and the acquisition of a malignant phenotype.

The ability to neoplastically transform normal human cells provides the opportunity to monitor the very early events of tumorigenesis in real time. For example, we have previously shown that neoplastically transformed human fetal lung fibroblasts display a progressive increase in both mRNA and protein expression of members of the DNA methyltransferase (DNMT) family of enzymes following antibiotic selection of the third construct, hRas-V12. The DNMTs play a pivotal role in cancer biology by modifying the normal patterns of DNA methylation. In fully developed tumors, the progressive nature of the change in expression would not be observed; only a difference in expression would be seen. The key to understanding oncogenesis lies in the ability to identify which genes possess this progressive nature and then to understand what is driving the oncogenesis-related progressive alterations in gene expression.

Two Dimensional Difference in Gel Electrophoresis (2D-DIGE) is one of the best available methods for discovery-based proteomic research; it is a relatively low cost “-omic” assay, that is highly quantitative and extremely accurate. The ability to analyze
individual samples on 12 gels in one experimental run virtually eliminates experimental variance due to minor differences in reagents, conditions, etc. Additionally, inclusion of the same internal standard on each gel confers the ability to quantitatively compare all 24 samples across the 2D gel dataset.

We hypothesized that the expression of proteins, as important to the maintenance of the oncogenic state as the DNMTs, would be progressively upregulated or progressively down regulated in a breast cancer model of oncogenesis. To test this hypothesis, we applied 2D-DIGE to identify differentially expressed proteins in neoplastically transforming HMECs in the first 80 days (d) after selection for \textit{hRas-V12}. Eighteen spots were identified as differentially expressed, with 9 spots each containing a single protein. Of these 9 spots, two spots were duplicates, leaving a total of 7 unique identifications: keratin 1, keratin 7, heat shock protein 4A-like, t-complex protein 1, stathmin, gelsolin, and maspin. These proteins have been shown to be involved in the loss of cellular rigidity and the gain of cellular motility, alterations that are associated with a metastatic phenotype.\textsuperscript{8-14} Of these 7 proteins, two (maspin and keratin 7) displayed a near perfect linear down regulation. \textit{SERPINB5}, the gene that codes for the maspin protein, is considered to be a metastasis suppressor gene (MSG) and the observed linear down-regulation over an 80 d period is consistent with a progressive trend toward a metastatic phenotype. Identification of genes that display a progressive alteration in protein expression confirmed our hypothesis and will provide the basis for future studies involving the regulation and function of both oncogenesis-related proteins.
MATERIALS AND METHODS

Cell Culture

Human Mammary Epithelial Cells (HMECs), at early passage, were obtained from Lonza (Basel, Switzerland) and grown in serum free Mammary Epithelial Growth Medium (MEGM) (Lonza; Basel, Switzerland) without sodium bicarbonate at 37°C and 0.1% CO₂. The omission of sodium bicarbonate allowed HMECs to be grown at a lower CO₂ level, which helped to maintain HMECs in the best possible condition. MEGM was changed daily and HMECs were subcultured at 75% confluency. HMECs at all stages of transformation were cultured according to the cell culture procedures described by Hammond et al. (1984) and Stampfer and Bartley (1985).¹⁵,¹⁶ Transient retrovirus was produced using Phoenix gag-pol 293 Human Embryonic Kidney cells (Φnx) created by Dr. Gary Nolan (Stanford University) and obtained through ATCC (Manassas, VA).¹⁷ Φnx, MCF-7, and MDA-MB-231 cells were grown in DMEM supplemented with 10% heat-inactivated fetal bovine serum and 1% APS (Amphotericin B, Penicillin, Streptomycin). Photographs of HMEC-SV40-hTERT (SHMEC) and Transforming HMECs (THMECs) were taken with a CoolPix5100 (Nikon; Melville, NY) at 40X magnification on an Eclipse TS100 microscope (Nikon; Melville, NY). Four cell lines of the MCF10 Isogenic Breast Cancer Metastatic Progression Model, MCF10A, MCF10AT, MCF10CAa.1α, and MCF10CAd.1, were acquired as a kind gift from Dr. Danny Welch (University of Alabama at Birmingham). Cells were cultured as previously described.¹⁸,¹⁹ MCF10A are the immortalized non-tumorigenic cell line. MCF10AT cells are the tumorigenic, pre-metastatic cell line derived from MCF10A. MCF10CAa.1α and MCF10CAd.1 are the two metastatic clonal lines. Each of the metastatic variants were
created from a separate carcinoma that was generated by tail vein injections of MCF10AT cells into severe combined immunocompromised mice.

**Plasmid Preparation**

Each of the three retroviral plasmids, \textit{pBabe-neo-SV40} (10891), \textit{pBabe-hygro-hTERT} (1773), \textit{pBabe-puro-RASV12} (1768) and the pseudotyping plasmid, \textit{pCL-VESVG} (1733) were obtained from ADDGENE (Cambridge, MA) with each plasmid’s respective ID number in parentheses.\textsuperscript{1} Plasmids were prepared according to the following procedures: Following bacterial culture and subsequent purification, purity of each plasmid was assessed by determination of the A260:A280 ratio. Once purity was verified, plasmid sequence integrity was verified using pBabe sequencing primer forward (5’-CTTTATCCAGCCCTCAC-3’).

**Retroviral Production and HMEC Infection**

Transient retrovirus was produced according to the following procedures: Φnx cells were plated on gelatin-coated 100 mm plates and grown to a confluency of 85%. For each retroviral infection, 15 μg of \textit{pBabe} plasmid and 15 μg of VSV-G plasmid was transfected into Φnx cells using Virapack Transfection Kit (Stratagene; Cedar Creek, TX). At 3 h post-transfection, cells were washed with PBS and fresh media added. At 48 h post-transfection retroviral supernatant was filtered through a .45 μm filter and polybrene added to a final concentration of 4 μg/mL. Retroviral supernatant (2 mL) was then applied to HMECs at 35% confluency. After 3 h, 3 mL of MEGM was added and infection was carried out for 48 h. Cells were washed with PBS at 48 h post-infection to
remove all traces of retrovirus and fresh MEGM was added. Infected HMECs were then allowed to recover from infection for 2 d before beginning selection. For each infection, antibiotic selection was carried out according to Elenbaas et al, 2001.²

Reverse Transcriptase-Polymerase Chain Reaction and Western Blot

RT-PCR was used to confirm expression of SV40-ER and exogenous hTERT (exo-hTERT). Total RNA was harvested using an RNEASY kit (Qiagen; Valencia, CA). Purified RNA was then subjected to DNase1 treatment and RNA repurified. In order to prevent RNA degradation, immediately following DNase1 treatment and subsequent purification, 2 μg RNA was reverse transcribed to cDNA using cDNA First Strand Synthesis kit (Invitrogen; Carlsbad, CA). PCR was performed using primers, specific to SV40 Large T (F: 5’-GCTTTGCAAAGATGGATAAAG, R: 5’-ACTAAACACAGCATGACTC), exo-hTERT (F: 5’-GACACACATTCCACAGGTCG, R: 5’-GACTCGACACCGTGCACCTAC), and Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (F: 5’-GAGAGACCCTCACTGCTG, R: 5’-GATGGTACATGACAAGGTGC) cDNAs.¹ Electrophoresis was carried out at 100V. Bands were visualized under UV light and analyzed with Kodak Digital Science Software (Kodak; Rochester, NY). Western Blotting of protein extracts was used to determine stable expression of hRASV12. Nuclear protein extracts were isolated using NE-PER Nuclear Extraction Kit (Thermo Scientific; Rockford, IL). Protein concentration was determined with the Bradford method of protein quantification using the Bio-Rad Protein Assay (Bio-Rad; Hercules, CA). Nuclear protein extract (20 μg) was separated by electrophoresis at 100 V, transferred to a nitrocellulose membrane, and blocked in 5% dry milk in Tris buffered saline solution
with 1% Tween (TBST) overnight. Primary antibody incubation was carried out overnight at 4° using mouse monoclonal antibodies specific to hRas (sc-53958). Membranes were then probed with goat anti-mouse IgG-HRP (sc-2005). Protein Bands were visualized using enhanced chemiluminescence (ECL) reagents (Thermo Scientific; Rockford, IL). All antibodies used in immunoblotting were obtained from Santa Cruz Biotechnology (Santa Cruz, CA) with catalog numbers listed in parentheses.

**Telomeric Repeat Amplification Protocol (TRAP)**

A TRAP assay using TRAPEze Kit (Millipore Corporation; Billerica, MA) was performed to confirm activation of telomerase activity as a result of stable expression of exo-hTERT as previously described. TRAP assay was performed according to procedures described in the TRAPEze manual. Total protein extracts were harvested using CHAPs lysis buffer. Protein concentration was determined as described above and 250 ng of protein extract was used for each reaction. Heat-inactivated samples were incubated at 85°C for 10 min. Protein extract from MDA-MB-231 cells served as a positive control and one reaction lacking protein extracts served as a negative control. Samples were incubated at 30°C for 30 min, followed by a 33 cycle PCR. PCR conditions were followed according to the kit manual. Reaction products (25 μL) were loaded onto a 12.5% non-denaturing PAGE and electrophoresed for 1.5 h at 300 V. Gels were stained with SYBR Green (Lonza; Basel, Switzerland) and visualized under UV radiation.
**THMEC Xenografts**

Female Balb/c nude mice 4–6 weeks of age were obtained from Charles River Laboratories (Wilmington, MA). THMECs at 40 d post-hRas introduction were injected subcutaneously in the left flank of each mouse. For each injection, 0.1 mL cell culture medium containing 1x 10^6 cells was mixed with an equal volume of Matrigel and the entire 0.2 mL was injected. Mice were sacrificed at 6 weeks post-injection and average tumor volume (mm^3) was calculated as follows: tumor volume (TV) on day 42 = (a x b^3)/2, where a is the largest diameter (in mm) and b is the diameter (in mm) perpendicular to a.

**Experimental Design**

We chose three points in the neoplastic progression to assess proteomic differences. These three time points in the neoplastic progression were as follows: SHMEC, THMEC 40 d post-hRAS, and THMEC 80 d post-hRAS. SHMEC cells are HMECs stably expressing SV40ER and exo-hTERT and represent premalignant breast cells. THMECs are HMECs stably expressing all three genetic elements, SV40ER, exo-hTERT, and hRAS-V12. For each time point, 4 biological replicates were generated for a total of 12 samples. For THMECs two independent samples were generated for each time point (40 d and 80 d) for each of two separate THMEC clonal lines generated from the same SHMEC clonal line. A schematic of the experimental design is shown in Supplementary Figure 1. THMECs at 80 d were derived from the same cell populations as those at 40 d. Four biological replicates of SHMEC were generated from one SHMEC clonal line, which was the same SHMEC cell line used to generate both THMEC cell
lines. Experiments were designed in this fashion to optimize a balance between a sufficient number of biological replicates and a minimalization of variation and sample number. Finally, each sample was dye-swapped for a final n=8 for each time point.

**RNA and Protein Extraction for Real-Time PCR and 2D-DIGE**

RNA, protein, and DNA were extracted using *Tri Reagent (Ambion, Austin, TX)* from HMEC-SV40-hTERT, THMEC 40 d post-hRas transfection, THMEC 80 d post-hRas, MCF10A, MCF10AT, MCF10CAa.1α, and MCF10CAD.1 following the manufacturer’s protocol. For each sample 4 biological replicates were generated as described above. Following the final ethanol wash, the RNA pellet was resuspended in 100% deionized formamide and stored at -80°C. Protein pellets were stored in 100% ethanol at -20°C.

**Preparation of Protein Extracts for 2D-DIGE and CyDye labeling**

Preparation of protein extracts and 2D-DIGE experiments were carried out as previously described by Kim et al. Precipitated protein from each cell sample was resuspended in isoelectric focusing (IEF) buffer and protein concentrations determined using the *2D Quant Kit (GE Healthcare, Piscataway, NJ)*. All three CyDyes were obtained from *GE Healthcare (Piscataway, NJ)*. To eliminate dye-specific bias, all samples were dye swapped by labeling with both Cy3 and Cy5. An internal standard (IS) was generated by pooling 50 μg of each of the 12 samples and labeling all of this mixture with 200 pmol of Cy2. Protein (50 μg) for each sample was labeled according to the
manufacturer’s instructions with 200 pmol of either Cy3 or Cy5. The Cy3 and Cy5 sample for each gel can be viewed in Table 1.

_2D Difference Gel Electrophoresis-1st Dimension Separation_

For each 2D DIGE gel, 50 μg of a Cy3-labeled sample, 50 μg of a Cy5-labeled sample, and 50 μg of the IS were pooled and the final volume brought up to 110 μL with IEF buffer before diluting 1:1 with rehydration buffer (7 M urea/2 M thiourea/ 4% CHAPS/40 mM Tris-HCl, pH 8.8) containing 1% IPGphor ampholytes pI range 4-7 (GE Healthcare; Piscataway, NJ) and 30 mM DTT. Twelve 11.0 cm immobilized pH gradient (IPG) DryStrips pH gradient 4-7 (GE Healthcare; Piscataway, NJ) were rehydrated overnight for 20 h at room temperature (RT) in individual troughs in the IPG strip rehydrating chamber (GE Healthcare; Piscataway, NJ). The choice was made to use a pH 4-7 strip instead of the broader pH 3-10 strip based on previous experience of the authors and preliminary testing prior to the start of data collection. Each DryStrip was placed gel-side down on top of the respective 220 μL sample and overlaid with mineral oil. The following day, DryStrips were laid gel side down and focused on an Ettan IPGphor II (GE Healthcare; Piscataway, NJ) at RT overnight using the following protocol: 6 h at 500 V, 1 h at 1000 V, 2.5 h at 6000 V, 2 h at 6000 V, 20 h at 300 V. Wicks were changed every 30 min for the first 1.5 h. The next day, DryStrips were placed at -80°C.
2D Difference Gel Electrophoresis-2nd Dimension Separation

DryStrips were thawed at RT for 15 min; and then equilibrated in SDS-sample buffer (6 M urea, 75 mM Tris-HCL, pH 8.8, 20% glycerol, 2 % SDS, 65 mM DTT) twice for 15 min each and then once for 15 min in SDS-sample buffer containing freshly add iodoacetamide (135 mM) and a trace of bromophenol blue. Each strip was then placed on top of a 12.5% Criterion precast gel (Bio-Rad; Hercules, CA). Second dimension separation for all twelve gels was carried out in a Dodeca Cell (Bio-Rad; Hercules, CA) at 150 V for 1.5 h at RT until bromophenol blue reached the bottom of the gel. Water chilled to 4°C was circulated through the coils in the cell to maintain a low temperature.

Gel Image Acquisition

Following 2nd dimension separation, each gel was scanned on a Typhoon Trio+ Variable Mode Imaging System (GE Healthcare; Piscataway, NJ), using the specific laser band-pass filters for each dye’s excitation and emission wavelengths. The excitation/emission wavelength combinations were (480 ± 35 nm)/(530 ± 30 nm) for Cy2, (540 ± 25 nm)/(590 ± 35 nm) for Cy3, and (620 ± 30 nm)/(680 ± 30 nm) for Cy5. Each gel was scanned individually and Photo Multiplier Tube (PMT) voltages were adjusted for maximum image quality with minimal signal saturation and clipping. Images were cropped and exported as 16-bit GEL files using ImageQuant TL (GE Healthcare; Piscataway, NJ) and imported into DeCyder Image Analysis (GE Healthcare; Piscataway, NJ). During the scanning process Gel 3 tore (HMEC-SV40- hTERT replicate B-Cy3; THMEC 80 d replicate A-Cy5) and was excluded from image analysis.
**Image and Statistical Analysis**

GEL files were analyzed using the Batch Processor, Difference In-gel Analysis (DIA), and Biological Variation Analysis (BVA) modules within DeCyder Image Analysis Software. Batch processing was accomplished by setting 1250 as the upper spot limit. DIA was performed within individual gels comparing the Cy3 or Cy5 spot pattern to the Cy2 spot pattern. Ratios of Cy3/Cy5 spot volumes on each gel in the 11 gel data set were normalized against the Cy2 spot pattern. Results were obtained as abundance ratios for each spot. The following ratios were used: SHMEC/THMEC 40 d and THMEC 40 d/THMEC 80 d. To determine statistical significance a Student’s t-test was calculated for each ratio with Decyder’s False Discovery Rate applied. Spots determined as significantly different (FDR corrected \( p < 0.05 \)) were then processed for identification.

Expression graphs were generated as follows: the abundance for each spot was normalized to the internal standard (Cy2) on its gel to generate a standard abundance ratio for each spot. Abundance ratios for each spot at each time point were then averaged to generate a mean for SHMEC, THMEC 40 d, and THMEC 80 d.

**Protein Identification**

Following scanning each gel was post-stained with Sypro Ruby (Invitrogen; Carlsbad, CA) for spot picking. Spot picking was carried out using the ProPic (Isogen Life Science; Netherlands) robotic spot picker. Gel plugs were rinsed three times with 1 mL 50% aqueous acetonitrile in 10 mM ammonium bicarbonate, pH 8.0, to remove Sypro Ruby stain. After evaporation of solvent by SpeedVac, each gel plug was rehydrated in 25 μL of 10 mM ammonium bicarbonate, pH 8.0. Trypsin Gold (10 μL of
12.5 μg/μL was added to each gel plug solution, and the mixture agitated overnight for 16 h. The supernatant was removed and plugs were washed twice with 10 mM ammonium bicarbonate. These rinses were combined, evaporated to dryness on a SpeedVac, then reconstituted in 10 μL of 0.1% formic acid. Peptide sequences were determined using an ABI-Sciex 4000 Qtrap (Applied Biosystems: Carlsbad, CA). The list of peptide sequences was exported and searched for matches within the nonredundant NCBI database (9/27/07) using the MASCOT search engine at www.matrixscience.com. Mass accuracy was set at 150 ppm, missed trypsin cuts at 1, Carbamidomethyl (C) fixed modifications, and Oxidation (HW) and Oxidation (M) variable modifications. The “gi” accession numbers obtained from MASCOT were used to search the UniProt Knowledgebase (UniProtKB) to obtain Swiss Protein Database ID numbers for each individual protein.

**Analysis of Maspin mRNA and Protein Expression**

Real-time PCR was used to assess mRNA levels of maspin (SERBINB5) in cell samples. RNA was prepared as described above using Tri Reagent. RNA suspended in formamide was purified using RNEasy kit and resuspended in RNase free water. Purified RNA was reverse transcribed as described above and cDNA was purified using a QiaQuick PCR Purification Kit (Qiagen; Valencia, CA). Purified cDNA (10 ng) was analyzed for expression of SERBINB5 (Hs00985283_m1) using Taqman Assays (Applied Biosystems; Carlsbad, CA) on a MiniOpticon Real-time Thermocycler (Bio-Rad; Hercules, CA). Specific Assay IDs are in parentheses. GAPDH served as the endogenous control. Expression values for THMEC 40 d and THMEC 80 d were calculated using the
ΔΔCt method relative to HMEC-SV40-hTERT expression. Expression values for MCF10CAa.1α and MCF10CAD.1 were calculated using the ΔΔCt method relative to MCF10AT expression. Western blots were performed as described above to assess maspin protein levels in THMEC model and MCF10 model. Maspin (sc-22762) antibodies were obtained from Santa Cruz Biotechnology (Santa Cruz, CA), with catalog number in parentheses. Actin served as the internal control.

Network Analysis using Ingenuity Pathway Analysis Software

A data set of the 7 proteins identified to have significantly different abundance ratios were uploaded into Ingenuity Pathway Analysis software, version 8.5 (released 02-13-2010). Each identifier was mapped to its corresponding object in Ingenuity's Knowledge Base. These molecules, called Network Eligible molecules, were overlaid onto a global molecular network developed from information contained in Ingenuity’s Knowledge Base. A network of Network Eligible molecules was then algorithmically generated based on connectivity. Functional Analysis of a network identified the biological functions that were significant ($p<.05$) to the molecules in the network. The network molecules associated with biological functions in Ingenuity’s Knowledge Base were considered for the analysis. Right-tailed Fisher’s exact test was used to calculate a p-value determining the probability that each biological function and/or disease assigned to that network is due to chance alone.
RESULTS

Generation of Induced Neoplastic Breast Cells

Using the protocols established by Elenbaas et al. we induced human mammary epithelial cells (HMECs) to undergo an oncogenic transformation to produce neoplastically transformed human mammary epithelial cells (THMECs). We first sought to confirm stable expression of all three genetic elements (SV40 ER, exo-hTERT, and hRAS-v12), illustrate an upregulation in telomerase activity, and highlight the tumorigenicity of our THMECs in vivo. It should be noted that we have chosen to use a different terminology than Elenbaas et al. used when referring to neoplastically transformed HMECs. We refer to HMECs expressing SV40 ER and exo-hTERT as SHMECs, whereas Elenbaas et al. referred to their cell line as HMLE. We refer to HMECs expressing all three constructs as THMECs, whereas Elenbaas et al referred to their cell line as HMLER. This was done in order to eliminate any confusion on the clonal origin of our cell lines.

Stable expression of SV40 ER and exo-hTERT was confirmed using reverse transcriptase PCR (RT-PCR). Figure 1A displays the results of RT-PCR. The presence of bands in all THMECs clearly indicates a stable expression of both SV40 ER and exo-hTERT in THMECs. Neither HMECs nor MCF-7 breast cancer cells expressed either SV40 ER or exo-hTERT. A lack of expression of exo-hTERT in MCF-7 cells is especially important as it indicates that the primers specific to exo-hTERT do not amplify endo-hTERT transcripts, as MCF-7 cells have been shown to both express endo-hTERT and display a high degree of telomerase activity. Telomerase activity was analyzed using the telomerase repeat amplification protocol (TRAP). A TRAP assay was performed
Figure 1. Confirmation of gene transfer and neoplastic transformation. A. Expression of SV40 ER, hTERT, and hRAS. RT-PCR of RNA extracted from HMEC, HMEC-SV40-hTERT-hRas (THMEC) 10d-70d, and MCF-7 cells using primers specific for SV40 Large T and exogenous hTERT was performed. GAPDH expression was analyzed to ensure equal loading. MCF-7 cells served as a negative PCR control. B. Upregulation of Telomerase Activity. The Telomeric Repeat Amplification Protocol (TRAP) assay was performed using protein extracts from HMEC, SHMEC, THMEC 10d and positive control, MDA-MB-231. Heat-Inactivated (HI) samples were incubated at 85°C for 10 min to inactivate telomerase. A reaction using 2 μL of DI water instead of protein extract served as a negative control testing for the presence of PCR artifacts. The laddering effect observed in SHMEC, T-HMEC 10d and MDA-MB-231 is indicative of telomerase activity. The black arrow indicates the 36 bp internal control (IC) band. C. Expression of hRAS. Western blot of total protein extracted from HMECs, HMEC-SV40-hTERT (SHMEC), and THMECS 10d, 20d, and 30d using antibodies specific for hRAS. D-E. Images of SHMEC (D) and THMEC 30d (E). Arrow indicates cylindrical, mesenchymal like morphology. All pictures were taken with a CoolPix P5100 (Nikon; Melville, NY) at 40X magnification on an Eclipse TS100 microscope (Nikon). HMECs stably expressing SV40 and hTERT (SHMEC) display typical epithelial morphology. The cell population of THMEC (HMEC-SV40-hTERT-hRASV12) at 30 days post-hRAS transfection was comprised of cells having a mesenchymal-like morphology. F. Analysis of tumorigenicity. Female nude mice were injected subcutaneously in the left flank of each mouse with 1 x 10^6 THMEC (40 days post-hRAS) cells. Arrows indicate the location of the tumor on the corresponding mouse.
using protein extracts harvested from HMECs, HMEC-SV40-hTERT (SHMEC), T-HMECs 10d post-hRAS introduction, and MDA-MB-231 cells (Figure 1B). The presence of a intense laddering effect, as well as the similarity between SHMEC and T-HMECs 10d and the positive control, MDA-MB-231, a breast cancer cell line known to display a high level of telomerase activity, is clearly indicative of a positive result for telomerase activity in SHMEC and T-HMECs 10d. The black arrow indicates the 36 bp internal control (IC) band, that is present in all samples. The presence of telomerase activity as confirmed by a TRAP assay, as well as the presence of exo-hTERT transcripts as confirmed by RT-PCR is indicative of a successful stable addition of the exo-hTERT gene to HMECs-SV40. To confirm a stable introduction of hRASv12, protein extracts were analyzed using a western blot. Figure 1C displays the results of immunoblotting and clearly indicates an increase in hRas expression in THMEC 10d, 20d, 30d compared to HMEC and SHMEC.

Cell Culture Observations and Assessment of Tumorigenicity

Figure 1D and 1E displays photographs taken of SHMECs and THMECs 30d. Figure 1D shows SHMECs displaying a cobblestone-like appearance that is typical of an epithelial morphology. Figure 1E shows THMECs 30d post-hRas. At 30d post-hRas, THMECs have lost both contact inhibition and the cobblestone-like appearance seen in 1D and now display a much more cylindrical phenotype indicated by the arrow in 1E. This observation is supported by the fact that neoplastically transformed HMECs have been shown to undergo an epithelial to mesenchymal transition (EMT). We next assessed THMEC tumorigenicity in vivo, by injecting THMECs at 40 d post-hRas.
introduction into immune-compromised nude mice. We choose 40 d post-hRas because at this point in the neoplastic induction foci formation was a common occurrence. All subcutaneous injections produced tumors (3 independent trials using 5 mice for a total n=15). The mean tumor volume for all 15 mice was 126.22 mm$^3$. A photograph of 3 representative tumors and the corresponding mice they were derived from can be viewed in Figure 1F. Arrows indicate the mice and location from which the tumors were removed. Injection of normal HMECs as a negative control was deemed unnecessary as HMECs are normal cells that lack the ability to form tumors.

Identification of Differentially Expressed Proteins During a Neoplastic Transformation

2D-DIGE was applied to assess differential expression at the proteomic level. A total of 4 protein samples for each of the following time points were analyzed: SHMEC, THMEC 40 d, and THMEC 80 d. We then dye swapped each protein sample to remove any dye-specific bias, for a total n=8 for each time point (4 biological replicates, with 2 technical replicates for each biological replicate). All 24 labeled samples were separated two dimensionally on 12 gels, with each gel also containing an internal standard, generated by combining equal amounts of all proteins) for cross-gel comparison. A schematic of the images generated for each time point can be viewed in Figure 2A and the exact gel assignments can be viewed in Table 1. A Student’s T-Test (performed on both pairs; SHMEC vs THMEC 40 d and THMEC 40 d vs THMEC 80 d) and a One-Way ANOVA were used to identify spots that were significantly (FDR corrected $p<.05$)
<table>
<thead>
<tr>
<th>Gel #</th>
<th>Cy 5</th>
<th>Cy 3</th>
<th>Cy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HMEC-SV40-hTERT(A)</td>
<td>THMEC-40 d(A)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>2</td>
<td>THMEC-40 d(A)</td>
<td>THMEC-80 d(A)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>3</td>
<td>THMEC-80 d(A)</td>
<td>HMEC-SV40-hTERT(B)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>4</td>
<td>HMEC-SV40-hTERT(B)</td>
<td>THMEC-40 d(B)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>5</td>
<td>THMEC-40 d(B)</td>
<td>THMEC-80 d(B)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>6</td>
<td>THMEC-80 d(B)</td>
<td>HMEC-SV40-hTERT(C)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>7</td>
<td>HMEC-SV40-hTERT(C)</td>
<td>THMEC-40 d(C)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>8</td>
<td>THMEC-40 d(C)</td>
<td>THMEC-80 d(C)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>9</td>
<td>THMEC-80 d(C)</td>
<td>HMEC-SV40-hTERT(D)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>10</td>
<td>HMEC-SV40-hTERT(D)</td>
<td>THMEC-40 d(D)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>11</td>
<td>THMEC-40 d(D)</td>
<td>THMEC-80 d(D)</td>
<td>Internal Standard</td>
</tr>
<tr>
<td>12</td>
<td>THMEC-80 d(D)</td>
<td>HMEC-SV40-hTERT(A)</td>
<td>Internal Standard</td>
</tr>
</tbody>
</table>
Figure 2. Two-dimensional difference gel electrophoresis of protein extracts from SHMEC, THMEC 40 d, THMEC 80 d. A. Schematic of images per time point. Green and red gels correspond to Cy3 and Cy5 labeled samples, respectively. Blue arrow indicates the location of maspin (Master Spot #919) on a 2D-Gel image from each time point. B. A representative image of a 2D gel. This image is of Gel 10 Internal Standard labeled with Cy2. Circled spots are those identified as being significantly (FDR corrected p<0.05) differentially expressed between HMEC-SV40-hTERT/THMEC 40 d or HMEC-SV40-hTERT/THMEC 80 d, or ANOVA. Spot numbers for unique IDs correspond to the Master Spot numbers assigned by DeCyder. For each numbered spot, a topographic map of signal intensity is displayed with the area used to calculate intensity shown in yellow.
different in abundance over the course of neoplastic induction, in at least one of the three statistical tests performed. Selection of spots for identification was done by selecting the most significant spots for different trend types. Of the spots that were significantly different, 18 spots were chosen for identification. The location of each spot on a representative 2D gel can be viewed in Figure 2B. Spots identified to be differentially expressed are circled in pink. Of these 18 spots, 9 spots were identified as containing a single protein. These 9 spots are circled and numbered in Figure 2B and shown with their respective topographic peaks outlined in yellow. Additionally, 9 spots contained a mixture of proteins and further analysis is needed to identify which proteins are associated with the observed changes in expression. The 9 spots identified to have a mixture of proteins are circled but not numbered in Figure 2B. Spot numbers correspond to the master spot numbers assigned by Decyder. Topographic maps were exported from Decyder. Of these 9 spots, 2 spots had duplicate identities. In total our methods produced 7 unique protein identifications (gene names are listed in parentheses): Heat Shock Protein 4A like (HSPA4L), Gelsolin (GSN), Keratin 1 (KRT1), T-complex Protein 1 (TCP1), Keratin 7 (KRT7), Maspin (SERPINB5), and Stathmin (STMN1). Gelsolin and maspin each had two spots with individual IDs, most likely corresponding to differing post-translational modifications given the similarity in the migration in the second dimension.

Proteomic data on all 18 spots can be viewed in Table 2, with spots containing a single protein in bold. Relative abundance graphs for the 9 spots containing a single protein are shown in Figure 3. Gelsolin and stathmin increased in abundance during neoplastic induction, whereas heat shock protein 4-like, keratin 1, t-complex protein 1, keratin 7,
and maspin were decreased. For all spots that illustrated statistically significant change, with the exception of T-complex protein, the fold change (FC) >1.20. Three spots displayed a near perfect linear down-regulation in protein abundance over 80 days as indicated by their respective coefficients of determination: keratin 7 ($r^2=0.9913$); maspin ($r^2=0.9947$); maspin ($r^2=0.9997$).
| a   | b         | c | d          | e           | f        | g         | h         | i       | j       | k    |
|-----|-----------|---|------------|--------------|----------|-----------|-----------|---------|---------|------|-----|
| 407 | -1.23     | .0008 | Heat shock 70kDa protein 4-like | O05757    | gi|31541941 | 497      | 13%     | 95479 | 5.63 |
|     | 1.28      | (.00046) | Mitochondrial inner membrane protein | Q16891    | gi|48145703 | 1138     | 29%     | 84027 | 6.08 |
|     |           |     | Gelsolin isoform b | P06396    | gi|38044288 | 462      | 13%     | 80876 | 5.58 |
|     |           |     | Procollagen-lysine, 2-oxoglutarate 5-dioxgenase 3 precursor | O60568    | gi|4505891 | 165      | 7%      | 85302 | 5.69 |
| 473 | 1.28      | (.00015) | Gelsolin isoform b | P06396    | gi|38044288 | 537      | 15%     | 80876 | 5.58 |
|     | 1.34      | (5.80E-05) | Gelsolin isoform b | P06396    | gi|38044288 | 638      | 21%     | 80876 | 5.58 |
| 592 | -1.23     | .012 | Heat shock 70kDa protein 8 isoform 1 | P11142    | gi|5729877 | 1446     | 46%     | 71082 | 5.37 |
|     | (0.18)    |     | Lamin B1 | P20700    | gi|5031877 | 784      | 24%     | 66653 | 5.11 |
| 598 | -1.32     | 4.00E-08 | Heat shock 70kDa protein 8 isoform 1 | P11142    | gi|5729877 | 1611     | 47%     | 71082 | 5.37 |
|     | (3.4E-06) |     | V-type proton ATPase catalytic subunit A | P38606    | gi|291868 | 201      | 12%     | 68677 | 5.35 |
|     |           |     | Lamin B1 | P20700    | gi|5031877 | 183      | 8%      | 66653 | 5.11 |
| 603 | -1.94     | 1.10E-10 | Serum albumin | P02768    | gi|28592 | 151      | 6%      | 71316 | 6.08 |
|     | (4.30E-07) |     | Fragile X mental retardation syndrome-related protein 1 | P51114    | gi|887793 | 90       | 4%      | 69991 | 5.91 |
| 604 | -2.30     | 1.10E-10 | Keratin, type II cytoskeletal 1 | P04264    | gi|11935049 | 440      | 13%     | 66198 | 8.16 |
|     | (4.30E-07) |     | Keratin, type II cytoskeletal 1 | P04264    | gi|11935049 | 440      | 13%     | 66198 | 8.16 |
| 606 | -1.29     | 9.00E-08 | Heat shock 70 kDa protein 8 isoform 1 | P11142    | gi|5729877 | 1775     | 50%     | 71082 | 5.37 |
|     | (5.0 E-07) |     | Heat shock 70 kDa protein 1 | P08107    | gi|4529893 | 743      | 27%     | 70280 | 5.48 |
|     |           |     | Lamin B2 | Q03252    | gi|27436951 | 723      | 23%     | 67762 | 5.29 |
|     |           |     | V-type proton ATPase catalytic subunit A | P38606    | gi|291868 | 291      | 13%     | 68677 | 5.35 |
| 615 | -1.20     | .03 | Heat shock 70 kDa protein 1 | P08107    | gi|4529892 | 1274     | 38%     | 70267 | 5.48 |
|     | (2.1E-06) |     | Plastin 3 | P13797    | gi|7549809 | 590      | 23%     | 71279 | 5.41 |
|     |           |     | Lamin B2 | Q03252    | gi|345758 | 284      | 11%     | 59079 | 5.87 |
|     |           |     | PDC-E2 precursor (AA -54 to -56) | P10515    | gi|35360 | 122      | 3%      | 65806 | 8.03 |
| 651 | -1.17(.016) | .027 | T-complex protein 1 isoform a | P17987    | gi|57863257 | 688      | 28%     | 60819 | 5.80 |
| 749 | 1.22      | (.0017) | Cytosolic non-specific dipeptidase | Q96KP4    | gi|8922699 | 252      | 14%     | 53088 | 5.58 |
|     |           |     | FK506-binding protein 5 | Q13451    | gi|145816 | 187      | 11%     | 50859 | 5.80 |
| 753 | -1.44     | (4.30E-07) | Keratin, type II cytoskeletal 7 | P08729    | gi|67782365 | 1754     | 55%     | 51411 | 5.40 |
|     | -1.30     | (.01) | Keratin, type II cytoskeletal 7 | P08729    | gi|67782365 | 1304     | 44%     | 51411 | 5.40 |
| 754 | -1.30     | (5.00E-07) | Keratin, type II cytoskeletal 7 | P08729    | gi|67782365 | 1304     | 44%     | 51411 | 5.40 |
| 919 | -1.32     | -1.24 | Keratin, type II cytoskeletal 8 | P05787    | gi|181573 | 1618     | 53%     | 53529 | 5.52 |
|     | (.00047) | (.01) | Keratin, type II cytoskeletal 7 | P08729    | gi|67782365 | 1304     | 44%     | 51411 | 5.40 |
|     |           |     | Chain A, The 2.1 A Structure | P36952    | gi|62738363 | 556      | 35%     | 42259 | 5.72 |

Table 2. Proteomic and Statistical Data of Spots that were Significantly Different in Abundance over the Course of Breast Tumorigenesis
<table>
<thead>
<tr>
<th>Master Spot Number</th>
<th>Fold change between SHMEC and THMEC 40 d. Numbers in parentheses correspond to false discovery rate corrected p values from a Student’s T-test. Only statistically significant data is displayed.</th>
<th>Fold change between THMEC 40 d and THMEC 80 d. Numbers in parentheses correspond to false discovery rate corrected p values from a Student’s T-test. Only statistically significant data is displayed.</th>
<th>False discovery rate corrected p values from a One-way ANOVA (SHMEC/THMEC40 d/THMEC80 d). Only statistically significant data is displayed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Bold text indicates spots that contained only one protein</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>b</td>
<td>Master Spot Number Assigned by Decyder</td>
<td>Fold change between SHMEC and THMEC 40 d. Numbers in parentheses correspond to false discovery rate corrected p values from a Student’s T-test. Only statistically significant data is displayed.</td>
<td>Fold change between THMEC 40 d and THMEC 80 d. Numbers in parentheses correspond to false discovery rate corrected p values from a Student’s T-test. Only statistically significant data is displayed.</td>
</tr>
<tr>
<td>c</td>
<td>Protein Name</td>
<td>d</td>
<td>d</td>
</tr>
<tr>
<td>d</td>
<td>Swiss-Prot ID</td>
<td>e</td>
<td>e</td>
</tr>
<tr>
<td>e</td>
<td>“gi” accession # obtained from MASCOT database</td>
<td>f</td>
<td>f</td>
</tr>
<tr>
<td>f</td>
<td>MOWSE score obtained from MASCOT database</td>
<td>g</td>
<td>g</td>
</tr>
<tr>
<td>g</td>
<td>Percent coverage of fragmented peptides vs total peptide length</td>
<td>h</td>
<td>h</td>
</tr>
<tr>
<td>h</td>
<td>Predicted molecular weight in Daltons</td>
<td>i</td>
<td>i</td>
</tr>
<tr>
<td>i</td>
<td>Predicted isoelectric point (pI)</td>
<td>j</td>
<td>j</td>
</tr>
</tbody>
</table>

|            | 920 | -1.29 (0.0015) | -1.27 (0.041) | 2.10E-07 | Chain A, The 2.8 A Structure Of A Tumour Suppressing Serpin | P36952 | gi|62738525 | 640 | 39% | 43071 | 5.72 |
|------------|-----|----------------|--------------|---------|------------------------------------------------------------|-------|----------------|-----|------|--------|------|
|            | 1028| -1.55 (2.30E-05) | -1.43 (.01) | 1.30E-08 | 60S acidic ribosomal protein P0                               | P05388 | gi|4506667 | 362 | 29% | 34423 | 5.71 |
|            |     |                |              |         | L-lactate dehydrogenase B chain                               | P07195 | gi|45557032 | 266 | 21% | 36900 | 5.71 |
|            |     |                |              |         | Anexin VIII                                                | P13928 | gi|178701 | 109 | 11% | 37103 | 5.68 |
|            | 1387| 1.51 (0.0019) | .015  |       | Stathmin 1 isoform b                                        | P16949 | gi|224451142 | 444 | 40% | 19925 | 6.51 |

Bold text indicates spots that contained only one protein
Figure 3. Protein abundance graphs of significantly differentially expressed proteins identified by 2D-DIGE. To the left of each graph is the corresponding gene name and master spot# assigned by Decyder. Protein expression results are displayed as mean log standard abundance ratios. See methods for derivation of protein standard abundance ratios. Error bars reflect S.E.M. Coefficients of Determination ($r^2$) were calculated using Microsoft Excel. A. Heat Shock Protein 4A Like (HSP4AL). B. Gelsolin (GSN). C. Gelsolin (GSN). D. Keratin 1 (KRT1). E. T-Complex Protein 1A (TCP1). F. Keratin 7 (KRT7). G. Maspin (SERPINB5). H. Maspin (SERPINB5). I. Stahmin (STMN1).
Analysis of SERPINB5 mRNA and Protein expression in THMECs and Metastatic Progression Model

To validate our proteomic data, we used western blotting to assess maspin (SERPINB5) protein expression and real-time PCR to assess SERPINB5 mRNA levels at all three time points of the THMEC model that were analyzed in 2D-DIGE (Figure 4A). Consistent with the proteomic data, SERPINB5 expression was progressively down-regulated over the course of tumorigenesis. The down-regulation also displayed the same linear trend as indicated by the coefficient of determination $r^2=0.9849$. SERPINB5 is known as a metastasis suppressor gene (MSG). MSGs suppress characteristics of metastasis such as invasion and motility, but have no effect on cellular division, immortality, or other non-metastatic characteristics of cancer. The progressive down-regulation of SERPINB5 expression illustrates a trend toward a metastatic phenotype. To confirm that SERPINB5 is indeed associated with metastasis, we analyzed mRNA levels of SERPINB5 in three cell lines of the MCF10AT metastatic progression model (Figure 4B). MCF10AT is the pre-metastatic cell line, whereas MCF10CAa.1α. and MCF10CAD.1 are two metastatic clones each generated from separate mice by tail vein injections of MCF10AT. The loss of SERPINB5 expression observed in THMECs can be observed in both metastatic clones relative to their premalignant origin (MCF10AT). SERPINB5 expression in both metastatic cell lines was roughly half of that observed in MCF10AT cells. Real-time data for SERPINB5 can be viewed in Figure 4A and 4B. We also assessed maspin protein levels in both models using western blot analysis. Alterations in the protein expression of maspin in THMECs correlated with the observed changes in maspin mRNA expression. In the metastatic progression model, maspin
Figure 4. Real-Time PCR analysis of mRNA expression of maspin (SERPINB5) in THMECs and the MCF10AT Metastatic Progression Model. **A.** mRNA expression of SERPINB5 in SHMEC, THMEC 40 d, 80 d. **B.** mRNA expression of SERPINB5 in metastatic progression model (MCF10AT, MCF10CAd.1α, and MCF10CAa.1). Expression ratios for THMEC 40 d, THMEC 80 d, MCF10CAd.1α, and MCF10CAa.1 were calculated using the ΔΔCt method relative to the expression of HMEC-SV40-hTERT (for THMECs) and MCF10AT (for MCF10CAd.1α and MCF10CAa.1). SHMEC and MCF10AT expression was set to 1.0. GAPDH which served as endogenous control. Error bars represent SEM. The columns represent mean relative expression. Coefficient of Determination ($r^2$) for maspin in THMEC model was calculated using Microsoft Excel. **C.** Western blot of maspin protein expression in neoplastic progression model and metastatic progression model. Maspin protein expression was down regulated over the course of tumorigenesis in THMECs. Maspin expression was less in premalignant MCF10AT cells than an in immortalized MCF10A cells. Maspin expression was undetectable in both metastatic carcinomas, MCF10CAd.1α and MCF10CAa.1. In both models actin served as internal control.
expression was less in premalignant MCF10AT cells than in immortalized MCF10A cells. Maspin protein expression was undetectable in both metastatic carcinomas, MCF10CAd.1α and MCF10CAa.1. To summarize our findings on SERPINB5 expression, a near identical linear down-regulation over the entire 80 d period was observed in protein spots (#919 and #920), real-time assays, western blotting, and microarray analysis (microarray data not shown) in THMEC model.

**Network and Functional Analysis**

To determine if these seven molecules displayed any commonalities with regard to function and association with other common molecules, we uploaded the 7 gene names of the proteins identified to be differentially expressed into Ingenuity Pathway Analysis (IPA) software. IPA uses a manually curated database of molecule interactions as well as biological functions and pathways to drive its network and functional analysis algorithms. A single network was generated from the gene list that contained all 7 gene names (Figure 5). Molecules are represented by shapes based on their biological role, according to the legend in Figure 5. Direct interactions are represented by a solid line and indirect relationships are indicated by a dashed line. This network is supported by 83 publications, of which we have cited a selection of publications related to the 7 genes identified using proteomic analysis.11,26-31 All 7 molecules were within a 2 molecule connection of one another within the network. The network has two primary nodes: TGFB1 and MYC. Both primary nodes are well known to play a role in breast cancer and have been shown to interact with one another.32,33
We next overlaid relevant biological functions associated with the network that involved the 7 differentially expressed proteins. All biological processes were determined to be statistically significant ($p<.05$) using a Fisher’s Exact test. The biological functions and their associated $p$ values are as follows: cell movement of tumor cells ($p=5.08e^{-5}$), cell movement of breast cell lines ($p=4.1e^{-4}$), polymerization of filaments ($p=2.68e^{-4}$), and a malignant tumor phenotype ($p=3.16e^{-3}$).\textsuperscript{27,34-45} Molecules associated with a particular biological function are connected with a pink line. Those molecules associated with a malignant phenotype are indicated with a pink arrow. The association of $KRT7$ and $SERPINB5$, both shown to be progressively down-regulated over the course of tumorigenesis, with malignancy and cell movement, respectively, exemplifies a trend towards a metastatic phenotype.
Figure 5. Network analysis of differentially expressed proteins. A network is a graphical representation of the molecular relationships between molecules. Molecules are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). All edges are supported by at least one reference from the literature, from a textbook, or from canonical information stored in the Ingenuity Pathways Knowledge Base. The node color indicates the directional change in protein expression of up- (green) or down- (red). Yellow nodes were added in by IPA to build the network and protein expression was not assessed. Nodes are displayed using various shapes that represent the functional class of the gene product. Edges are displayed as solid (direct) or dashed (indirect) to indicate the nature of the relationship. Functional analysis of a network identified the biological functions that were significant \( (p<0.05) \) to the molecules in the network. The network molecules associated with biological functions in Ingenuity’s Knowledge Base were considered for the analysis. Right-tailed Fisher’s exact test was used to calculate a p-value determining the probability that each biological function and/or disease assigned to that network is due to chance alone. The biological functions cell movement of tumor cells \( (p=5.08e-5) \), cell movement of breast cell lines \( (p=4.1e-4) \), and polymerization of filaments \( (p=2.68e-4) \) are connected to the respective molecules with pink lines. Molecules associated with the biological process of a malignant tumor \( (p=3.16e-3) \) phenotype are indicated with pink arrows.
DISCUSSION

The inability to predict when a cell will begin a transformation from a normal state to a neoplastic state *in vivo* makes it impossible to study the initial states that a neoplastic cell transitions through as it acquires oncogenic features. This problem is circumvented through the use of *in vitro* cellular reprogramming. Neoplastic transformation is one method of cellular reprogramming that is capable of producing a neoplastic breast cell from a normal breast epithelial cell, thus enabling the ability to analyze the early cellular events of oncogenesis in real time. Using previously published protocols to generate breast cancer cells from normal somatic breast cells, we analyzed three time points in the neoplastic progression to breast cancer to identify proteins that were progressively differentially expressed. We have seven differentially expressed proteins, two of which displayed a linear progressive loss of abundance over 80 d following selection for *hRAS*. Network analysis revealed a common association with MYC and TGFB1, both implicated in oncogenesis; the latter playing a pivotal role in the generation of breast cancer stem cells. Functional analysis of the IPA-generated network illustrates a strong correlation between the proteins and metastatic state. The progressive nature of the trend towards a metastatic state is exemplified by the linear down regulation of the metastasis suppressor gene (MSG), *SERPINB5*, and the epithelial cytokeratin, *KRT7*. As important as the presence and nature of alterations in expression, are the biological functions that are acquired due to differential expression of these seven proteins. Polymerization of filaments and cellular movement are two biological functions associated with our data set that are both instrumental in the acquisition of a metastatic state.
Stathmin and gelsolin are both implicated in the regulation of the cytoskeleton.\textsuperscript{12, 13} Stathmin affects microtubules by binding and sequestering tubulin subunits.\textsuperscript{12} Gelsolin binds and severs actin filaments.\textsuperscript{13} Thus they both serve to promote cytoskeletal depolymerization. One characteristic of normal, epithelial cells is a static cellular structure inferred by a rigid cytoskeleton. During the acquisition of a metastatic phenotype, this static, rigid structure transitions to a dynamic, pliable structure, more suited for motility. An integral aspect in the gain of cellular motility is the ability to break down and rebuild the cytoskeleton in a highly dynamic fashion. The observed increases in stathmin and gelsolin abundance in THMECs promote the cytoskeletal disassembly, a process that is not required for the rigid state of an epithelial cell. Although stathmin and gelsolin are not solely responsible for the acquisition of metastatic characteristics, their increased abundance increases the metastatic potential of THMECs through the attenuation of the static, rigid cytoskeleton that is characteristic of non-motile cells.

\textit{SERPINB5}, the gene that codes for the maspin protein, is a MSG that has been shown to regulate cellular motility, although the exact mechanisms are poorly understood.\textsuperscript{27, 47} \textit{SERPINB5} is related to the serpin family of extracellular protease inhibitors and although structurally very similar to other members of the serpin family, maspin does not undergo the conformational change characteristic of inhibitory serpins. Originally thought to only function in the extracellular environment, maspin has recently been shown to localize intracellularly, possibly even to the nucleus.\textsuperscript{47} Overexpression of maspin has also been demonstrated to result in a decrease in both cellular motility and metastasis; however, very little is known about the mechanisms by which maspin elicits
its inhibitory effects on these processes. We have clearly shown here that SERPINB5 mRNA and protein level expression is progressively down-regulated in early tumorigenesis and is consistent with a metastatic phenotype. Additionally, with regard to maspin, we have established two cell culture-based progression models in which to begin elucidation of the currently unknown biological function of maspin. With regard to regulation of maspin expression, the association of maspin with TGFβ1 signaling pathways may help to explain the cause for the observed progressive trend in maspin expression. TGFβ1 has been shown to promote the conversion of neoplastically transformed HMECs to a stem-like state in a time-dependant manner, suggesting that maspin may play a role in not only the acquisition of a metastatic phenotype but also in the maintenance of stemness. Future studies will be aimed at assessing the role of maspin in breast cancer stem cells.

Two previous studies have utilized 2D-DIGE to analyze alterations in the proteome in breast cancer models of tumorigenesis.\textsuperscript{48,49} The first study utilized an erbB2 model of breast cancer.\textsuperscript{48} In that study they compared cells over-expressing erbB2, a gene that has been previously shown to be associated with 25-30% of breast cancers, to normal human mammary luminal epithelial cells.\textsuperscript{50} Similar to our study, they identified multiple cytokeratins that were differentially expressed, including keratins 13A, 17, and 19. Interestingly they did not identify cytokeratin 1 or 7 in their study to be differentially expressed, suggesting that the regulation of these two genes in breast cancer occurs independent of the erbB2 pathway. Additionally, the erbB2 over-expressing cells were shown to overexpress maspin, a protein known to inhibit metastatic processes. An overexpression of maspin in these cells suggests that this model is indicative of a less
metastatic variant of breast cancer. The second study compared proteomic differences between a cell line, EM-G3, derived from an infiltrating ductal breast carcinoma and normal mammary epithelial cells.\textsuperscript{49} They also found differences in multiple cytokeratins, including cytokeratin 19. Interestingly, they identified cytokeratin 7 as showing no significant change in protein expression. The differences involving the differential expression of keratin 7 and cytokeratin 19 are most likely due to the different types of breast cancer model systems used. Additionally, an increase in cytokeratin 19 expression occurring with the addition of \textit{SV40ER} prior to generation of SHMEC cells by addition of \textit{exo-hTERT}, would result in an inability to detect the differential expression of cytokeratin 19. The detection of cytokeratin 7 differential expression in our study suggests that the loss of cytokeratin 7 may be associated with breast cancer cells that possess an increased potential to metastasize. Future studies will be aimed at understanding the effect of the loss of cytokeratin 7 on metastatic processes.

The initial upregulation in \textit{GSN} and \textit{STMN1} illustrates an early increase in metastatic potential, whereas the linear nature of the observed trend in maspin and \textit{KRT7} expression illustrates that the increase in metastatic potential is also a progressive event in oncogenesis. We intend to use these proteomic findings, in combination with data from microarray experiments, as the basis for future studies into understanding the loss/gain of the biological functions associated with a metastatic state. Future directions for this project also include methylomic analysis of DNA methylation patterns and further proteomic analysis using large format gels and a broader pH range for separation in the first dimension.
ACKNOWLEDGEMENTS

The authors would like to acknowledge Lana Grinberg, Anna Pendleton, and Su Ngyuen for their help in generating THMECs. We would like to acknowledge the Targeted Metabolics and Proteomics Laboratory members Dr. Mark Cope, Gloria Robinson, and Richie Herring for assistance with the 2D-DIGE protocols. We would like to acknowledge the Tollefsbol lab members who provided informal advice during all aspects of the projects. JTD was supported by an NCI Cancer Prevention and Control Training Program (R25CA047888). This work was also supported in part with grants from Susan G. Komen for the Cure and the NIH (CA129415, TOT, PI). The instrumentation for running, imaging and processing 2D-gels was provided by a NCRR Shared Instrumentation grant (S10 RR16849, HK, PI). The mass spectrometers were purchased from funds provided by NCRR grants (S10 RR11329, RR13795, RR19231, S. Barnes, PI). Operation of the Comprehensive Cancer Center Proteomics-Mass Spectrometry Shared Facility was partially supported by funds from a P30 Core support grant (CA13148-35, E. Partridge, PI) to the UAB Comprehensive Cancer Center. We would like to acknowledge our collaboration with the lab of Dr. Danny Welch (University of Alabama at Birmingham) for providing us with the metastatic variants of the MCF10AT metastatic progression model.
REFERENCES


Generation of Breast Cancer Stem Cells through Epithelial-Mesenchymal Transition.
*PLoS ONE* **2008**, 3, (8), e2888.


Supplementary Figure 1. Schematic of cell culture experimental design. These three time points in the neoplastic progression were as follows: SHMEC, THMEC 40 d post-hRAS, and THMEC 80 d post-hRAS. For each time point, 4 biological replicates were generated for a total of 12 samples. For THMECs two independent samples were generated for each time point (40 d and 80 d) for each of two separate THMEC clonal lines (THMEC2 and THMEC3) generated from the same SHMEC clonal line.
IDENTIFICATION OF A POTENTIAL ROLE OF HNF4A IN BREAST CANCER BRAIN METASTASIS

by

JOSEPH TYSON DEANGELIS,TRYGVE TOLLEFSBOL

In preparation for PLOS ONE

Format adapted for dissertation
ABSTRACT

Background
We had previously shown, using proteomic analysis, that in neoplastically transforming human mammary epithelial cells, proteins that displayed significantly different abundances over the first 80 days of breast tumorigenesis were involved in processes associated with invasion and cellular movement. Two of these proteins displayed a linear trend, suggesting a progressive acquisition of metastatic features. The goal of this study was to identify, from genome-wide gene expression analysis, genes that displayed a progressive trend and to identify common biological functions or molecular interactions of these genes. We also hypothesized that the common molecular interactions observed in vitro would also be present in gene expression data from a previously published breast cancer metastasis clinical study.

Results
We first generated a gene list composed of 127 genes that display a highly significant ($q > .001$), progressive alteration in gene expression in the early stages of breast oncogenesis. Functional analysis of this gene list illustrates a significant ($q > .025$) association with biological functions associated with metastasis including invasion, migration, and adhesion. Network-based molecular interaction analysis identified a common association between our gene list and Hepatocyte Nuclear Factor 4a (HNF4a), a member of the Nuclear Receptor superfamily. Lastly, using clinical gene expression data we illustrate an association between differentially expressed genes associated with breast cancer brain metastasis and HNF4a.
Conclusions

These findings indicate an association of HNF4a with differentially expressed genes in a cell culture model of breast tumorigenesis. Our results also strongly suggest that HNF4a plays a key role in breast cancer metastasis to the brain. These studies could have important implications in novel therapeutic approaches to breast cancer and its metastases.
BACKGROUND

Despite the vast amount of research conducted on cancer as a whole, the initial stages of oncogenesis are not well understood. A lack of understanding of the earliest stages of oncogenesis stems from an inability to identify, \textit{in vivo}, a neoplastic cell immediately after it has departed from normality. As a result, most cancer research has focused on analyzing cancer cells from tumors once they have reached an observable mass. It is believed that an understanding of the earliest stages of oncogenesis, immediately following the acquisition of the genetic abnormalities required to sustain an oncogenic phenotype, will provide insight into the very nature of cancer itself and assist in the identification of novel anti-cancer targets that would go unseen in the later stages of oncogenesis. It is not possible to analyze the early stages of oncogenesis \textit{in vivo}, although a process known as neoplastic transformation provides the ability to study these changes \textit{in vitro}. A novel model of neoplastic transformation was first discovered by Weinberg and colleagues, who used viral-mediated serial gene transfer of \textit{SV40 Early Region} (ER), \textit{hTERT}, and \textit{hRAS-V12} to generate cancer cells from normal fetal lung fibroblasts (1). This process was then later applied to normal human mammary epithelial cells (HMECs) to produce breast cancer cells (2). The neoplastic breast cells display many if not all of the features of actual breast cancers. More recently this same process has been shown to produce breast cancer stem cells over time (3). The transforming growth factor β (TGF β) signalling pathway was shown to be instrumental in promoting this transition to a stem-like state (3).

Using the aforementioned transformation procedure to generate neoplastically transformed fetal lung fibroblasts, we have previously shown that three members of the
DNA methyltransferase (DNMT) family of enzymes display a progressive, or gradual, increase in both mRNA and protein expression following the addition of the final genetic element, \textit{hRAS} (4, 5). More recently we have shown that neoplastically transformed human mammary epithelial cells (THMECs) display a progressive increase in metastatic potential (6). Using proteomic analysis we identified a metastatic suppressor gene, \textit{SERPINB5}, which is progressively down-regulated in the first 80 days following introduction of \textit{hRAS}. Since the progressive trend displayed by the DNMTs and \textit{SERPINB5} is believed to be a key component of the acquisition of the more aggressive features that cancers display, we sought to identify, at the transcriptomic level, which genes display this same type of progressive trend. We were most interested in identifying biological functions modulated by these genes as well as any individual molecular associations shared by these genes.

Using microarray analysis, we first identified a gene list consisting of 127 genes that were significantly differentially expressed over 80 days (d) following \textit{hRAS} introduction and that displayed a progressive, unidirectional change in gene expression. Next, this gene list was uploaded into Ingenuity Pathway Analysis (IPA) software for functional and molecular interaction analysis. Functional analysis revealed a significant association between our gene list and biological functions associated with the acquisition of a metastatic phenotype. Network-based molecular interaction analysis revealed an association between our gene list and Hepatocyte Nuclear Factor 4a (HNF4a).

We then were interested in determining if the association with HNF4a that was observed \textit{in vitro}, would be present in gene expression data from a previously published breast cancer metastasis clinical study. Unexpectedly, we found that that HNF4a was
associated with genes that were expressed at significantly higher levels in breast cancer brain metastasis compared to breast cancer bone metastasis. Genes that were expressed at significantly higher levels in breast cancer bone metastasis compared to breast cancer brain metastasis failed to produce a single molecular interaction network containing HNF4a. The identification of significantly differentially expressed molecules associated with HNF4a in the neoplastic breast cancer model system that lacks the microenvironment of the brain, suggests that the role of HNF4a, with respect to breast cancer brain metastasis, may be involved in driving breast cancer cells to the brain. Future studies will be aimed at determining the exact role of HNF4a in breast cancer brain metastasis. Our study illustrates, for the first time, a link between HNF4a and breast cancer metastasis.

RESULTS AND DISCUSSION

Neoplastically transformed breast cancer cells were produced from human mammary epithelial cells using modifications of the protocols first described by Elenbaas, et al (2). Generation of transforming HMECs (THMECs) which stably express all three genetic elements (SV40ER, hTERT, and hRAS-V12), confirmation of THMEC tumorigenicity, and explanation of experimental design has been previously reported (6). An experimental overview can be viewed in Figure 1. Figure 1 (top) depicts the timeline over which the analysis was performed and indicates when in the neoplastic progression samples were harvested. The three time points (HMEC-SV40ER-hTERT (SHMEC), THMEC 40d post-hRAS, and THMEC 80d post-hRAS) at which samples were collected are termed conditions and can be viewed in Figure 1 (middle). For each condition four
biological replicates were produced. A more detailed explanation of the experimental design can be found in the methods section and in Ref. 6. RNA and protein were sequentially precipitated from each plate and the assays they were used for can be viewed in Figure 1 (bottom). Proteomic analysis has been previously reported (6) and in this study we will focus on transcriptomic analysis.

To identify genes that displayed a progressive change in gene expression we assessed genome-wide gene expression in the aforementioned three conditions (n=4 for each condition) using Affymetrix Human GeneST 1.0 Arrays. GeneST arrays are whole-transcript exon arrays that assess the relative amount of exonic RNA from multiple exons over the full length of the transcript. For each of the ~28,000 genes covered by the GeneST array there is an average of 26 probes per gene, with multiple probes per exon (7). Normalized intensity values for each probe are then summarized at the gene level to generate a single intensity value for each gene. For each gene, 12 individual summarized intensity values, each corresponding to the expression of that particular gene in each of the 12 samples, were analyzed. Prior to gene list generation, SERPINB5 expression was analyzed to ensure quality of the gene expression results. Our previous findings indicated that SERPINB5 expression is progressively down-regulated over 80 d following hRAS introduction and displayed a linear trend (6). The observed change in mRNA expression of SERPINB5 measured by microarray analysis was nearly identical to the previously observed down-regulation of SERPINB5 mRNA and protein expression (6). The changes in SERPINB5 mRNA expression in this system are shown in Figure 2.

Once we had confirmed the validity of our microarray data set, the following quality control and statistical-based filters were applied to enrich for genes that displayed
a progressive upregulation or down-regulation in gene expression over 80 d following introduction of hRAS. A schematic of the enrichment process can be viewed in Figure 3. The first step in the enrichment process was application of the quality control filters, in order to provide more statistical power during multiple test corrections. All genes were first filtered based on the level of expression of each gene relative to all 28,000 genes on the array in all 12 samples. A gene was removed from the data set if the following criterion was not met: For at least one condition, 4 of 4 replicates must have an expression value higher than the overall 20th percentile. The next quality control filter that was applied removed any genes that had a large amount of variation between replicates within each condition and was performed by calculating a coefficient of variation (CV). For each condition, a CV (σ/μ) was calculated for each condition and genes were removed if: \( CV > 0.5 \) for all 3 conditions. Parameters used for quality control filters were the default settings in GeneSprings.

The second step in the enrichment process was to isolate, from the 14,144 genes remaining after application of quality control filters, genes that showed a significant variance in mean expression between the three conditions. A Welch’s ANOVA was applied to each gene, which produced a p-value corresponding to the probability that the observed change in gene expression for that particular gene is due to chance alone. A Benjamini-Hochberg multiple test correction was applied to each p-value to produce an FDR (false discovery rate) corrected q-value corresponding to each p-value. The threshold for significance was set to \( q < 0.001 \). Of the 14,144 genes, 4,428 genes met the cut-off for significance. Next, the 4,428 genes significant at \( q < 0.001 \), were enriched for genes that displayed a fold change (FC) of >1.35 in both pairs of conditions.
(Pair1=µSHMEC → µTHMEC 40d; Pair2=µTHMEC 40d → µTHMEC 80d). The remaining 177 genes were identified to have a FC>1.35 in mean gene expression for both pairs of conditions, from the 4,428 genes significant at q<0.001. Lastly, from these 177 genes, we enriched for genes that had a positive change in both pairs or a negative change in both pairs. The above mentioned filters allowed us to isolate 127 genes that displayed a significant, progressive, unidirectional change in expression over the first 80 d of oncogenesis. This gene list is termed BCT1.

We next sought to identify two separate datasets corresponding to BCT1. We first wanted to identify the biological functions that were significantly associated with gene list BCT1. Secondly, we were interested in identifying individual molecules that displayed interactions with multiple genes from BCT1. Given the stringency of our expression-based enrichment process, we elected to go with a more qualitative based approach for functional analysis. By this we mean that only gene names were used for pathway and network analysis and not any of the actual expression data. This allowed us to focus on individual molecular interactions between genes. A molecular interaction could be transcriptional repression, phosphorylation, binding, complexing, etc. and each interaction is based on at least one publication. Ingenuity Pathway Analysis (IPA) software has quickly become the standard for making biological interpretations of this type. IPA utilizes a manually curated database, known as the Ingenuity Knowledge Base, of molecular interactions to drive its pathway and network algorithms.

Human Gene Organization (HUGO) gene symbols for each of the 127 genes in BCT1 were uploaded into IPA using default parameters. The first biological interpretation performed was biological function analysis. A right-tailed Fisher’s Exact
test was used to calculate a p-value determining the probability that the association between genes in \textit{BCT1} and each biological function in Ingenuity’s Knowledge Base was due to chance alone. A Benjamini-Hochberg multiple test correction was then applied to generate a q-value corresponding to each p-value that has been false discovery rate corrected. The threshold for significance was set at \( q < 0.025 \). Cellular and molecular biological functions significantly associated with \textit{BCT1} can be found in Table 1. Physiological and developmental biological functions can be found in Table 2. Categories, functions, and functional annotations are those assigned by IPA. Genes from \textit{BCT1} that are associated with each functional annotation are listed in each table, as well as the q-value for each functional annotation.

Statistically significant cellular and molecular functions associated with \textit{BCT1} include processes that are associated with a metastatic or malignant phenotype, such as invasion, migration, adhesion, and angiogenesis. Taken collectively, the cellular and molecular functions associated with genes that displayed a progressive, unidirectional change in gene expression, indicate that following addition of hRAS, THMECs are progressively becoming more capable of invading surrounding tissue and altering their three dimensional structure in order to migrate to another location. Interestingly the functional annotation \textit{disassembly of filaments} was shown to be significant (\( q = 0.0199 \)) and is indicated by the presence of CLASP2, ITGA2, ITGAV, SEMA3A, and SGK1 in \textit{BCT1}. This correlation supports our previous proteomic analysis that indicated an increase in cytoskeletal processing due to observed increases in gelsolin and stathmin protein expression (6).
Statistically significant physiological and developmental functions associated with \textit{BCT1} are involved in differentiation, morphogenesis, adhesion, angiogenesis, and invasion. Functional analysis confirmed one of our initial hypotheses that genes that were progressively altered in the early stages of tumorigenesis were involved in the acquisition of the more aggressive characteristics breast cancers display. The association with differentiation and morphogenic pathways supports the findings of Mani, et al. in that neoplastically transformed HMECs are progressing towards a stem-like state (3).

We next asked whether the molecules in \textit{BCT1} displayed any molecular associations in common. The network analysis component of IPA is a very powerful tool for identifying any associations with individual molecules a particular set of genes may share. See Methods for further explanation of network generation. A single molecular network was generated from the genes in \textit{BCT1} and can be viewed in Figure 4. This network was termed \textit{T1}. Red molecules can be found in gene list \textit{BCT1}, all of which were down-regulated during oncogenesis. Purple molecules were added by IPA during the building process. The molecular network was pseudocolored by the authors for viewing purposes; however, no changes were made to the makeup of the network itself. Network \textit{T1} displayed two primary nodes, Hepatocyte Nuclear Factor 4a (HNF4a) and Transforming Growth Factor Beta Receptor 1 (TGFBR1). Using a Fisher’s Exact test, we identified the TGFβ canonical pathway to be statistically ($q < .05$) correlated with network \textit{T1}. Using IPA’s overlay function, the two molecules in \textit{T1} associated with TGFβ signalling (HNF4a and TGFBR1) are connected with red lines.

We then went back to our gene expression data to determine if alterations in HNF4a expression and TGFBR1 could explain the observed changes in the molecules
associated with these two genes. Graphs of HNF4a and TGFB1R mRNA expression can be viewed in Figure 5. HNF4a expression was clearly upregulated in both pairs of conditions \( (q=.002) \); however it failed to meet the statistical significance cutoff for inclusion into list \( BCT1 \). TGFB1R was downregulated in \( Pair1 \) but showed no change in \( Pair2 \). Although TGFB1R expression was highly significant, it failed to meet the cutoff of FC>1.35 in both pairs for inclusion into list \( BCT1 \).

Using this same model system, TGFβ signalling was previously shown to play a major role in the generation of breast cancer stem cells, a process also shown to contribute to the acquisition of metastatic features. TGFβ signalling is highly complex and is involved in numerous pathologies in addition to breast cancer. As such, it is important to understand molecules that are modulating the downstream effects of TGFβ signalling. The observed increase in HNF4a expression suggests that in the early stages of tumorigenesis it functions as a repressor, mediating TGFβ signalling at the transcriptional level. Hepatocyte Nuclear Factor 4a is a member of the Nuclear Receptor superfamily and binds to DNA primarily as a homodimer to regulate transcription of the downstream gene (9). HNF4a has been shown to regulate developmental and differentiation pathways in multiple tissues types, most notably in the liver and kidneys, but also in pancreatic islets. Through interaction with both co-repressor and co-activator complexes, HNF4a has been shown to act as both a transcriptional activator and a transcriptional repressor and to be affected by TGFβ signalling through the interaction with activated SMAD proteins (9, 10). HNF4a expression has also been shown to be regulated by the p53 pathway (11). Disruption in the HNF4a gene has been strongly linked to Maturity Onset Diabetes of the Young (MODY), due to its regulation
of genes involved in glucose metabolism (12). Although HNF4a has been previously demonstrated to be involved in liver and intestinal cancers, its role in breast cancer is relatively unknown and, to our knowledge, it has never previously been linked to breast cancer metastasis. The loss of expression of TGFBR1, a key component in the transduction of extracellular TGFβ signalling, is also likely to contribute to dysregulation of TGFβ signalling. Future studies will be needed to identify exactly how the absence of TGFBR1 affects the TGFβ signalling cascade. TGFBR1 has been previously shown to be regulated by DNA methylation. Methylation of the promoter region of TGFBR1 could explain why the down-regulation of TGFBR1 did not display a progressive trend.

Although the role of HNF4a has been strongly linked to type 2 diabetes, its roles in breast cancer and breast cancer metastasis are unknown. To identify a possible role of HNF4a in breast cancer metastasis, we analyzed global gene expression patterns in metastatic breast cancers using publicly available array data from a previously published study (13). .CEL files from 25 Affymetrix Human Genome U133 Plus 2.0 arrays were downloaded from Gene Expression Omnibus (GEO) (SubSeries = GSE14017; SuperSeries =14020). Each of the 25 arrays corresponded to an individual patient that was diagnosed with metastatic breast cancer. Of the 25 individual cancers that were analyzed, 10 were from breast cancer bone metastases and 15 were from breast cancer brain metastases. Raw .CEL files were downloaded from GEO and uploaded to GeneSprings. Filters were then applied to isolate genes that were significantly differentially expressed at a higher level in brain metastases than in bone metastases. Genes were first filtered based on expression in the 20th percentile. A Student’s T-Test assuming unequal variance was applied to identify genes that were significantly (q < .05).
differentially expressed between bone and brain metastasis. Genes were then filtered based on fold change, where genes with $\text{FC}<1.35$ were removed. Following expression, statistical, and fold change based filtering, we next isolated the 523 genes where $\mu_{\text{BRAIN}} > \mu_{\text{BONE}}$. This list of 523 genes differentially expressed at a higher level in brain metastasis than in bone metastasis and is termed $BCM1$.

$BCM1$ was uploaded to IPA, as described above, for molecular network generation. Sixteen individual networks were generated and termed $M1$-$M16$. One of the molecular networks generated illustrated a high level of connectivity to HNF4a. This network, $M1$, illustrating HNF4a as the most connected node can be viewed in Figure 6. Blue molecules can be found in gene list $BCM1$. Yellow molecules were added in by IPA during the building process. The molecular network was pseudocolored by the authors for viewing purposes; however, no changes were made to the makeup of the network itself. $M1$ illustrates the association of HNF4a with 14 individual molecules that were differentially expressed at higher levels in brain metastases than in bone metastases, where each association is based on at least one publication.

Of the 16 individual networks that were generated, 8 networks, including network $M1$, were linked by at least one molecule. A schematic of linked networks and the number of molecules they shared in common can be viewed in Figure 6. Using IPA’s network merge function, the individual molecules in each of the 8 linked networks were combined to generate a single supranetwork termed $SM1$. Supranetwork $SM1$ can be viewed in Figure 7. Grey molecules are those found in gene list $BCM1$. Multiple well-defined nodes can be observed, two of which, TP53 and HNF4a are outlined in black and shown at higher magnification. The connectivity of HNF4a within network $SM1$ is
comparable to that of the tumor suppressor TP53. Map SM1 clearly illustrates a high degree of association between molecules involved in breast cancer brain metastasis and HNF4a. Interestingly, the directionality of molecular interactions converges on TP53, suggesting that genes differentially expressed at higher levels in brain metastasis serve to modulate the tumor suppressor functions of TP53. This is not the case for HNF4a, whereas the directionality of interactions radiates outward from HNF4a, suggesting a regulatory role for HNF4a in a large number of genes associated with breast cancer brain metastasis. The directionality of HNF4a interactions in combination with the fact that alterations in HNF4a were observed in a cell culture model of breast cancer, that lack the microenvironment of the brain, strongly suggests a driver role for HNF4a with respect to breast cancer brain metastasis.

In summary, we first identified 127 genes shown to have a progressive trend in gene expression following the acquisition of the genetic elements associated with breast tumorigenesis. We illustrated that these functions are associated with metastatic features. It was also shown that a portion of these genes mapped to a molecular network, containing HNF4a and TGFBR1 as primary nodes, which was statistically correlated with the TGFβ signalling pathway. We then isolated 523 genes that were expressed at higher levels in tumors isolated from breast cancers that had metastasized to the brain compared to bone. Using network analysis we identified a strong association of brain metastasis specific genes to HNF4a. Very little is known regarding the role of HNF4a in breast cancer and its role in breast cancer metastasis has never been identified. Our findings suggest that HNF4a may not only be important to the initial acquisition of metastatic features, but may also be involved in either promoting or sustaining brain metastasis.
CONCLUSIONS

In the initial stages of breast oncogenesis, genes displaying a significant progressive change in expression are significantly associated with biological functions associated with metastasis. These genes are also display a common association, identified using IPA’s molecular network analysis tool, with HNF4a based on published literature. It was also shown that HNF4a is associated with genes expressed at higher levels in breast cancer brain metastasis compared to breast cancer bone metastasis.

METHODS

Experimental Design

SHMEC, and THMEC cell lines were generated using the protocols originally reported by Elenbaas et al. (2) For each condition (SHMEC, THMEC 40d, and THMEC 80d), 4 biological replicates were generated for a total of 12 samples. Each condition represented a time point in the neoplastic progression. Additional information regarding generation of SHMEC and THMEC cell lines, study design, and cell culturing can be found in Ref 6. Briefly, for THMECs two independent samples were generated for each time point (40d and 80d) for each of two separate THMEC clonal lines generated from the same SHMEC clonal line. Four biological replicates of SHMEC were generated from one SHMEC clonal line, which was the same SHMEC cell line used to generate both THMEC cell lines.

RNA and Microarray Processing

RNA was extracted as previously reported, and stored in 100% deionized formamide. Storage in formamide provides for long-term storage with little to no
degradation of the RNA, and allowed us to first complete our proteomic experiments prior to starting microarray analysis. The quality of each RNA sample was determined by analysis on the 2100 Bioanalyser (Agilent) prior to RNA labeling. Detailed GeneChip analysis procedures are presented in the Manufacturer’s GeneChip Expression Technical Manual (Affymetrix). Briefly, 100 ng of total RNA from each of 12 samples was used to generate double-stranded cDNA by linear amplification using T7-linked random primers and reverse transcriptase. Subsequently, cRNA was generated by standard methods (Affymetrix) followed by cRNA fragmentation, end label biotinylation and preparation of hybridization cocktail. The arrays were hybridized overnight at 45°C, and then washed, stained, and scanned the next day. .CEL files were generated for each of the 12 samples using Affymetrix GeneChip Command Console.

**Gene List and Expression Graph Generation**

Expression and statistical based filters were applied using the software package GeneSprings v10 (Agilent Technologies, Santa Clara, CA). Briefly, the raw .CEL GeneChip files were uploaded in GeneSprings where intensity values were background-subtracted, variance stabilized, and normalized with ExonRMA method. To generate expression graphs from microarray data, normalized intensity values for each biological replicate were calculated as follows: for a particular gene (normalized intensity value + 16)/(18S RNA normalized intensity value + 16). Mean expression values were then calculated for each time point (SHMEC, THMEC 40d, and THMEC 80d). 18S RNA was used as an endogenous control for all expression graphs. The normalized expression value for SHMEC bio rep C for 18S RNA was considered an outlier and as such expression data from SHMEC bio rep C was excluded from all expression graphs.
**Functional and Network Analysis**

Gene list T1 was imported into Ingenuity Pathway Analysis (IPA) software from GeneSprings. IPA utilizes a manually curated database of functional interactions. Interactions are supported by published information. To rank relevant biological functions, a Fisher’s Exact test was applied and significance cut off set at $q < .025$. For network analysis, each gene identifier was mapped to its corresponding object in the Ingenuity’s Knowledge Base. These molecules, called Network Eligible molecules, were overlaid onto a global molecular network developed from information contained in Ingenuity’s Knowledge Base. A network is a graphical representation of the molecular relationships between molecules. Unique Networks of Network Eligible Molecules were then algorithmically generated based on connectivity. Molecules are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). All edges are supported by at least one reference from the literature, from a textbook, or from canonical information stored in the Ingenuity Pathways Knowledge Base. Nodes are displayed using various shapes that represent the functional class of the gene product. Edges are displayed as solid (direct) or dashed (indirect) to indicate the nature of the relationship.

**COMPETING INTERESTS**

The authors declare they have no competing interests.
AUTHORS' CONTRIBUTIONS

JTD was responsible for study conception, experimental design, all aspects of data collection and interpretation, and manuscript preparation. TOT was responsible for, study conception, experimental design, data interpretations, and manuscript revisions.

ACKNOWLEDGEMENTS

JTD was funded in part by an NCI Cancer Prevention and Control Training Program (R25CA047888). This work was supported in part by funds from the National Cancer Institute (R01 CA129415) and Susan G. Komen for the Cure.

REFERENCES


Figure 1 Experimental Overview. Top. Timeline illustrating the stable addition of hRASV12 to SHMECs (HMECs stably expressing SV40 Early Region and exogenous hTERT). Addition of hRAS to SHMEC cells generated neoplastically transforming HMECs (THMECs). THMECs were harvested at 40d and 80d post-hRAS introduction. Middle. Schematic Representation of the three conditions that were used for microarray analysis. Bottom. RNA and protein were sequentially precipitated from each plate and used for the listed assays.
Figure 2 Graph of *SERPINB5* mRNA expression in neoplastically transforming human mammary epithelial cells. mRNA expression results are displayed as mean normalized expression values relative to 18S RNA expression. See Methods for mRNA expression data. Error bars reflect S.E.M. Coefficient of Determination ($r^2$) was calculated using Microsoft Excel.
Multiple filters were applied to identify genes that displayed a significant progressive unidirectional change over 80 d following hRAS introduction.

**Figure 3** - Schematic of quality control and statistical filters applied to generate gene list $BCT1$. Multiple filters were applied to identify genes that displayed a significant progressive unidirectional change over 80 d following hRAS introduction.
Figure 4 - Molecular network, T1, generated from gene list BCT1.
Network analysis of differentially expressed genes in neoplastically transforming human mammary epithelial cells. Red molecules correspond to genes from BCT1. Purple molecules were added into the network by IPA during the building process. Functional analysis of this network identified the TGFβ canonical pathway as being significantly \( (q < 0.05) \) associated with this network. Molecules associated with TGFβ signalling are connected with red lines. See Methods for further information on network generation.
Figure 5 - HNF4a and TGFβR1 expression in neoplastically transforming human mammary epithelial cells. mRNA expression graph of HNF4a (A.) and TGFBR1(B.). mRNA expressions are displayed as mean normalized expression values relative to 18S RNA expression. See Methods for mRNA expression data. Error bars reflect S.E.M. Coefficient of Determination ($r^2$) for HNF4a expression was calculated using Microsoft Excel.
Figure 6 - Molecular network, $M1$, involved in brain metastasis
Network analysis of genes from $BCM1$ revealed a single network with HNF4a as the most connected node. Blue molecules correspond to genes from $BCM1$. Yellow molecules were added into the network by IPA during the building process. See Methods for further information on network generation.
Figure 7-Schematic of network relationships. Each rectangle corresponds to an individual network generated from network analysis of gene list BCM1. Networks that had one or two molecules in common are connected by a straight line, with the number of molecules in common listed next to each line. The blue networks were all linked together in a single group and were merged together to generate a single supranetwork comprised of the molecules from M1-N8. M1 is outlined in red and contains HNF4a as the most connected node. Unlinked networks are displayed in orange and possessed no molecules in common with one another or with any of the linked networks.
Figure 8 - Molecular Supranetwork, SM1, generated from network merge of linked networks. Using IPA’s network merge function 8 individual networks that were all linked by one or more molecules were combined to generate a single network consisting of all the individual molecules in each of the eight linked networks. Grey molecules correspond to genes from BCM1. White molecules were added into the network by IPA during the building process. Two well defined, highly connected (HNF4a and TP53) nodes are shown at higher magnification. See Methods for further information on generation of merged network.
<table>
<thead>
<tr>
<th>Category</th>
<th>Function</th>
<th>Function Annotation</th>
<th>q-value</th>
<th>Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell-To-Cell Signaling and Interaction</td>
<td>attachment</td>
<td>attachment of endothelial cell lines</td>
<td>0.015</td>
<td>IL1B, ITGA2, ITGAV</td>
</tr>
<tr>
<td>Cell Cycle</td>
<td>cell division process</td>
<td>arrest in cell division process of leukocyte cell lines</td>
<td>0.0199</td>
<td>INHBA, MET, NOTCH1, TFRC</td>
</tr>
<tr>
<td>Cell Morphology</td>
<td>blebbing</td>
<td>blebbing of plasma membrane</td>
<td>0.0199</td>
<td>FST, INHBA</td>
</tr>
<tr>
<td>Cell-To-Cell Signaling and Interaction Cell-To-Cell Signaling and Interaction</td>
<td>activation</td>
<td>activation of fibroblasts</td>
<td>0.0199</td>
<td>IL1B, ITGA2, ITGAV</td>
</tr>
<tr>
<td>Cell-To-Cell Signaling and Interaction Cell-To-Cell Signaling and Interaction</td>
<td>adhesion</td>
<td>adhesion of tumor cells</td>
<td>0.0199</td>
<td>F3, IL1B, ITGA2, ITGAV</td>
</tr>
<tr>
<td>Cellular Compromise</td>
<td>adhesion</td>
<td>adhesion of tumor cells</td>
<td>0.0199</td>
<td>F3, IL1B, ITGA2, ITGAV</td>
</tr>
<tr>
<td>Cellular Compromise</td>
<td>disassembly</td>
<td>disassembly of filaments</td>
<td>0.0199</td>
<td>CLASP2, ITGA2, ITGAV, SEMA3A, SGK1</td>
</tr>
<tr>
<td>Cellular Development</td>
<td>developmental process</td>
<td>developmental process of endothelial cells</td>
<td>0.0199</td>
<td>IL1B, INHBA, ITGA2, MET, NOTCH1, SEMA3A</td>
</tr>
<tr>
<td>Cellular Development</td>
<td>differentiation</td>
<td>differentiation of Langerhans cells</td>
<td>0.0199</td>
<td>ITGAV, ITG6</td>
</tr>
<tr>
<td>Cellular Growth and Proliferation</td>
<td>colony formation</td>
<td>colony formation of prostate cell lines</td>
<td>0.0199</td>
<td>ITGAV, ITG6</td>
</tr>
<tr>
<td>Cellular Movement</td>
<td>invasion</td>
<td>invasion of cytotrophoblastic cells</td>
<td>0.0199</td>
<td>FST, INHBA</td>
</tr>
<tr>
<td>---</td>
<td>migration</td>
<td>migration of prostate cell lines</td>
<td>0.0199</td>
<td>ITGAV, ITG6</td>
</tr>
<tr>
<td>Cell Death</td>
<td>cell death</td>
<td>cell death of cell lines</td>
<td>0.0201</td>
<td>ATP2B1, DUSP6, F3, FST, GALNT5, GLIPR1, IL1B, INHBA, ITGA2, ITGAV, MET, NOTCH1, OBFC2A, PTGS2, S100A9, SEMA3A, SERPINB2, SGK1, SMAD2, TFRC, TOP2B, UBE2S, WEE1, XDH</td>
</tr>
<tr>
<td>Cellular Movement</td>
<td>invasion</td>
<td>invasion of normal cells</td>
<td>0.0201</td>
<td>FST, IL1B, INHBA, MET, XDH</td>
</tr>
<tr>
<td>Cell Death</td>
<td>apoptosis</td>
<td>apoptosis of endothelial cell lines</td>
<td>0.0219</td>
<td>DUSP6, IL1B, S100A9, SEMA3A, TFRC</td>
</tr>
<tr>
<td>Cell Morphology</td>
<td>permeabilization</td>
<td>permeabilization of tight junctions</td>
<td>0.0248</td>
<td>CLDN1, IL1B</td>
</tr>
<tr>
<td>Cell-To-Cell Signaling and Interaction Cellular Development</td>
<td>permeabilization</td>
<td>permeabilization of tight junctions</td>
<td>0.0248</td>
<td>CLDN1, IL1B</td>
</tr>
<tr>
<td>Cell-To-Cell Signaling and Interaction Cellular Development</td>
<td>angiogenesis</td>
<td>angiogenesis of endothelial cells</td>
<td>0.0248</td>
<td>ITGA2, MET, SEMA3A</td>
</tr>
<tr>
<td>Cell-To-Cell Signaling and Interaction Cellular Development</td>
<td>differentiation</td>
<td>differentiation of granulosa cells</td>
<td>0.0248</td>
<td>IL1B, INHBA</td>
</tr>
<tr>
<td>Category</td>
<td>Function</td>
<td>Functional Annotation</td>
<td>q-value</td>
<td>Molecules</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>---------------</td>
<td>---------------------------------------</td>
<td>---------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Cardiovascular System Development and Function</td>
<td>attachment</td>
<td>attachment of endothelial cell lines</td>
<td>0.015</td>
<td>IL1B, ITGA2, ITGAV</td>
</tr>
<tr>
<td>Hair and Skin Development and Function</td>
<td>morphogenesis</td>
<td>morphogenesis of hair follicle</td>
<td>0.015</td>
<td>FST, INHBA, NOTCH1, PTGS2</td>
</tr>
<tr>
<td>Organ Development</td>
<td>morphogenesis</td>
<td>morphogenesis of hair follicle</td>
<td>0.015</td>
<td>FST, INHBA, NOTCH1, PTGS2</td>
</tr>
<tr>
<td>Organ Morphology</td>
<td>morphogenesis</td>
<td>morphogenesis of hair follicle</td>
<td>0.015</td>
<td>FST, INHBA, NOTCH1, PTGS2</td>
</tr>
<tr>
<td>Connective Tissue Development and Function</td>
<td>activation</td>
<td>activation of fibroblasts</td>
<td>0.0199</td>
<td>IL1B, ITGA2, ITGAV</td>
</tr>
<tr>
<td>Embryonic Development</td>
<td>invasion</td>
<td>invasion of cytotrophoblastic cells</td>
<td>0.0199</td>
<td>FST, INHBA</td>
</tr>
<tr>
<td>Hair and Skin Development and Function</td>
<td>morphogenesis</td>
<td>delay in morphogenesis of hair follicle</td>
<td>0.0199</td>
<td>FST, INHBA</td>
</tr>
<tr>
<td>Hematological System Development and Function</td>
<td>differentiation</td>
<td>differentiation of Langerhans cells</td>
<td>0.0199</td>
<td>ITGA2, ITGB6</td>
</tr>
<tr>
<td>Hematopoiesis</td>
<td>differentiation</td>
<td>differentiation of Langerhans cells</td>
<td>0.0199</td>
<td>ITGA2, ITGB6</td>
</tr>
<tr>
<td>Organ Development</td>
<td>morphogenesis</td>
<td>delay in morphogenesis of hair follicle</td>
<td>0.0199</td>
<td>FST, INHBA</td>
</tr>
<tr>
<td>Organ Morphology</td>
<td>morphogenesis</td>
<td>delay in morphogenesis of hair follicle</td>
<td>0.0199</td>
<td>FST, INHBA</td>
</tr>
<tr>
<td>Reproductive System Development and Function</td>
<td>size</td>
<td>size of preantral follicle</td>
<td>0.0199</td>
<td>FST, INHBA</td>
</tr>
<tr>
<td>Reproductive System Development and Function</td>
<td>migration</td>
<td>migration of prostate cell lines</td>
<td>0.0199</td>
<td>ITGA2, ITGB6</td>
</tr>
<tr>
<td>Skeletal and Muscular System Development and Function</td>
<td>size</td>
<td>size of muscle</td>
<td>0.0199</td>
<td>H19, MET, PTGS2, F3, IL1B, ITGA2, ITGAV</td>
</tr>
<tr>
<td>Tissue Development</td>
<td>adhesion</td>
<td>adhesion of tumor cells</td>
<td>0.0199</td>
<td>H19, MET, PTGS2, F3, IL1B, ITGA2, ITGAV</td>
</tr>
<tr>
<td>Tissue Morphology</td>
<td>size</td>
<td>size of muscle</td>
<td>0.0199</td>
<td>H19, MET, PTGS2, F3, IL1B, ITGA2, ITGAV</td>
</tr>
<tr>
<td>Tumor Morphology</td>
<td>adhesion</td>
<td>adhesion of tumor cells</td>
<td>0.0199</td>
<td>H19, MET, PTGS2, F3, IL1B, ITGA2, ITGAV</td>
</tr>
<tr>
<td>Cardiovascular System Development and Function</td>
<td>angiogenesis</td>
<td>angiogenesis of endothelial cells</td>
<td>0.0248</td>
<td>IL1B, ITGA2, ITGAV, ITGA2, MET, SEMA3A</td>
</tr>
<tr>
<td>Organismal Development</td>
<td>angiogenesis</td>
<td>angiogenesis of endothelial cells</td>
<td>0.0248</td>
<td>IL1B, INHBA</td>
</tr>
<tr>
<td>Reproductive System Development and Function</td>
<td>differentiation</td>
<td>differentiation of granulosa cells</td>
<td>0.0248</td>
<td>IL1B, INHBA</td>
</tr>
</tbody>
</table>
HYPERMETHYLATION AND ONCOGENESIS

by

JOSEPH TYSON DEANGELIS, JOEL B. BERLETCH, LUCY G. ANDREWS, AND TRYGVE O. TOLLEFSBOL

CANCER EPIGENETICS

Copyright
2009
by
CRC PRESS
Used by permission
Format adapted for dissertation

90
Abstract

Many biological processes are known to lead to oncogenesis although epigenetics plays a key role in the transition to and the maintenance of a neoplastic phenotype. DNA methylation is an epigenetic process, whose mechanisms comprise an integral part of gene regulation in development, aging, memory formation, and oncogenesis. The methylation of CpG dinucleotides is carried out by a family of enzymes termed DNA methyltransferases (DNMTs) and results in a methyl group attached to the 5-position of cytosine. DNA hypermethylation can lead to gene silencing and, in rare cases, gene activation and both of these effects of hypermethylation can have a carcinogenic effect. This chapter summarizes some of the mechanisms by which DNA hypermethylation can lead to oncogenesis and how hypermethylation of the regulatory regions of specific genes contributes to the hallmarks of cancer.

Keywords: DNMTs, CpG Islands, Hypermethylation, MBDPs, HDACs, HMTs.
Introduction

DNA methylation comprises an integral part of epigenetics and its gene silencing effects have been implicated in numerous biological processes such as aging, development, memory formation and oncogenesis. DNA methylation is carried out by the DNA methyltransferase (DMNT) family of enzymes, which transfer methyl groups from S-adenosylmethionine (SAM) to the 5-position of cytosine in CpG dinucleotides [1]. Hypermethylation during oncogenesis usually occurs at CpGs located in CpG islands, which are C-G rich regions of 300-3000 bp in or near 40% of promoters in mammalian genes [1]. About 70% of the promoters of human genes contain CpG islands within close proximity. The product of DNA methylation, 5-methylcytosine, is highly susceptible to mutation due to spontaneous deamination which results in a cytosine to thymine mutation. CpG conversion to TpG over most of the genome has led to highly conserved regions of DNA in or near promoter regions that contain an unusually high percentage of CpGs. These regions are normally unmethylated and the lack of methylation has prevented accumulation of C→T mutations and has promoted the formation of CpG islands [2]. DNA methylation allows an additional tier of hereditable genetic information to be stored upon the genetic code of DNA without changes in the DNA sequence itself. The accepted paradigm of DNA methylation is that hypermethylation usually leads to silencing of gene activity, while the expression of a few genes may be activated by DNA methylation.

The gene silencing effects of hypermethylation have been linked not only to oncogenesis but also to the ability of tumors to sustain a neoplastic phenotype [2]. Methylation profiling of specific CpG islands in various tumor types has begun to reveal
the widespread significance of hypermethylation [3]. Genomic methylation profiling can assist in diagnosis as well as predict prognosis and the elucidation of the importance of DNA hypermethylation in oncogenesis has led to creation of treatments aimed at preventing as well as reversing DNA hypermethylation.

**DNA Methyltransferases**

In human cells three major DNA methyltransferases (DNMTs) are responsible for DNA methylation. Methylation patterns are maintained by DNMT1, which preferentially methylates hemimethylated DNA following DNA replication [4]. DMNT3A and DNMT3B are known as the *de novo* methyltransferases [5]. Although DNMT3A and DNMT3B show no preference for unmethylated DNA over hemimethylated DNA, the low level of *de novo* methylation carried out by DNMT1 relative to DNMT3a and DNMT3b have led to the latter two enzymes being designated as the *de novo* methyltransferases [5,6].

Changes in the expression of the DNMTs are a contributing factor leading to changes in methylation patterns. Correlation of modulations in DNMT expression with changes in methylation patterns was shown by Casillas et al. who observed a decrease in expression of DMNT1 and DNMT3a combined with increased expression of DNMT3b in aging fetal lung fibroblasts [7]. These changes in DNMT expression were shown to correlate with changes in maintenance and *de novo* methylation activity [8]. DNA methylation has also been implicated in memory formation and the age-related changes in DNMT expression and genomic methylation may also provide insight into age-related memory loss [9].
Fetal lung fibroblasts, neoplastically transformed through the addition of defined genetic elements, show a significant increase in expression of all three DNMTs [7]. A recent report has correlated DNMT up-regulation seen \textit{in vitro} with DNMT overexpression in tumors taken from lung cancer patients [10]. Although all DNMTs were up-regulated, the extent to which DNMT1 was overexpressed directly correlated with a poor prognosis, illustrating that sustained DNA methylation may not only contribute to oncogenesis, but may also help tumors maintain a malignant phenotype [10].

\textbf{Methyl-CpG Binding Proteins and Recruitment of Chromatin Remodeling Complexes}

One mechanism by which DNA hypermethylation can regulate gene expression is through the recruitment of methyl-binding proteins containing methyl-CpG-binding domains (MBD). These regions contained in proteins such as MeCP1 and MeCP2 include amino acid residues capable of binding to methylated DNA [11]. In particular, Nan and colleagues found that a short stretch of 85 amino acids in the N-terminal region of the MeCP2 protein retained the ability to bind methylated DNA following deletion analysis [12]. There are currently several proteins that have been identified and are referred to as MBD proteins (MBDPs) which have the capacity to silence transcription by binding to both hemimethylated and fully methylated DNA.

DNA methylation is not the only epigenetic modification known to influence gene expression. Chromatin structure is an important aspect of transcriptional regulation and methyl-CpG-binding proteins provide the link between DNA methylation and chromatin
remodeling-mediated gene silencing. The silencing effect is a result of tight associations between DNA and histone proteins brought on by modifications of histone tails. MeCP2 has been found to be associated with Sin3a, a histone deacetylase (HDAC) co-repressor, whereas MeCP1 complexes with HDAC1 and HDAC2 [13,14]. Histone methyltransferases (HMTs) are another family of histone modifying enzymes that repress transcription in part through methylation of lysine 9 on histone H3. HMTs have also been shown to associate with MeCP2, once again linking DNA methylation to chromatin remodeling [15].

By binding to hypermethylated DNA, the involvement of MeCPs in chromatin remodeling complexes leads to altered chromatin structure and long-term gene silencing; however without hypermethylation the MeCPs do not normally bind to DNA. The fact that most of methylation association silencing requires DNA methylation to be present leads into perhaps one of the most significant questions in the field of epigenetics: How and why are CpG islands that have been evolutionarily conserved due to a lack of methylation become suddenly hypermethylated? Studies over the last decade have finally begun to unravel a mystery that involves the interactions of a very unique and diverse group of players.

Initiation and Preservation of Aberrant Methylation Patterns

Once hypermethylation of DNA occurs, the MBDPs bind to methylated DNA, thereby preventing the binding of transcriptional activators, and facilitate recruitment of chromatin remodeling complexes that allow genes to be maintained in a silenced state. Despite genome-wide hypomethylation during oncogenesis, CpG islands located in the 5’
control region of certain tumor suppressor genes are specifically targeted and more frequently methylated than other CpG islands across the genome [3]. The process by which tumor suppressors are targeted is not yet fully understood and elucidation of exactly how this targeting takes place would be a landmark discovery.

*De novo* methylation by DNMT3a and DNMT3b appears to depend on their association with other proteins, such as transcriptional regulators and chromatin remodeling proteins. c-Myc is a well-known gene regulator that primarily functions as a transcriptional activator. c-Myc also can repress gene transcription and this ability has been shown to be an effect asserted in part by its association with DNMT3a. Association with other transcriptional regulating proteins allows c-Myc to be recruited to the promoter region of certain tumor suppressors, bringing with it DNMT3a, which in turn will methylate unmethylated CpGs [16].

Another example of how transcription factors can recruit DNMTs to the promoter region of a gene involves the promyelocytic-retinoic acid receptor (PML-RAR) fusion protein. The oncogenetic transcription factor PML-RAR has been shown to interact with DNMT1 and DNMT3a. A gene known to be regulated by PML-RAR, is the retinoic acid receptor RARβ2 [17]. The interaction of PML-RAR with DNMTs leads to DMNT localization at the RARβ2 promoter and contributes to methylation induced repression [17].

Recruitment of DNMTs by oncoproteins is only a piece of the puzzle. Another important factor is the association of DNMTs with chromatin modifying complexes, a relationship that seems to affect nearly every area of epigenetics. Histone methylation and subsequent chromatin inactivation has been shown to precede *de novo* DNA
methylations, possibly by HMT interaction with DNMT3a [18]. The cyclical association of HMTs with DNMT3a and of MeCP2 with HMTs may also help to explain why de novo-methylated CpG islands emerge in close proximity to one another in the genome [19]. Lymphoid-specific helicase (Lsh), a member of the SNF2 subfamily of helicase enzymes, has also been shown to have an effect on DNA methylation in mouse embryonic fibroblasts by interaction with DNMT3a and 3b but not DNMT1 [20]. No known report has yet to corroborate this study in humans or to illustrate its effect on oncogenesis.

Another molecule has been recently shown to assist in targeting of aberrant methylation patterns is RNA. A study by Castanotto has shown that short hairpin RNAs (shRNAs), containing sequences homologous to the promoter of the RASSF1A, mediated a DNA methylation-induced gene silencing effect [21]. The ability of RNA to target genes for methylation opens up an entirely new avenue of epigenetic research. siRNAs might allow researchers, for the first time, to target specific genes for hypermethylation. This would allow a control over gene expression never before seen.

Preservation of methylation patterns is a much simpler process but still involves most of the same players that regulate initiation of aberrant methylation patterns. Recruitment of DNMT1, HDACs, and HMTs by MBDPs allows the chromatin to remain in a condensed state and the methylation patterns to be maintained with a high degree of fidelity [15, 22, 23].

Although multiple proteins and even RNA cooperate to target and effectively silence specific tumor suppressor genes, hypermethylation is the factor that allows genes targeted for silencing to be maintained in a transcriptionally silenced state. Treatment
with the demethylating agent, 5-azacytidine (5-aza), can reactivate almost every known methylation-silenced gene, further supporting the vital role of DNA methylation in gene silencing. The mechanism by which 5-aza reverses methylation will be discussed later in this chapter.

Figure 1 is a schematic representation of the complex interactions between the various proteins that modulate initiation and preservation of hypermethylation. One possible reason why it has been so difficult to elucidate the exact cause of the appearance of aberrant methylation patterns may be that any and all of the players involved may be capable of initiating hypermethylation of a target gene. It also may be possible that one player may initiate DNA methylation in one specific gene but not in another. If this is the
case, it would further need to be resolved as to what circumstances are required for a specific methylation modulator to initiate DNA methylation.

**The Impact of Hypermethylation on Oncogenesis**

Due to the large number of genes known to be involved in oncogenesis and regulated via methylation, the following sections will illustrate a few specific examples of genes across a range of tumor types that are silenced via hypermethylation and contribute to five of the six hallmarks of cancer according to Hanahan et al, which are evading apoptosis, insensitivity to anti-growth signals, sustained angiogenesis, limitless replicative potential, and tissue evasion and metastasis [24]. The sixth hallmark, self-sufficiency in growth signals, has not been shown to be regulated by DNA methylation.

**Evading apoptosis**

The internal and external cellular environment is carefully monitored and any abnormalities or changes can trigger a cascade of events leading to apoptosis or programmed cell death. Apoptosis can be triggered by changes in growth rates, contact inhibition, radiation, and even signals from nearby cells and tissues. Multiple proteins are involved in apoptotic pathways, and although their pathways differ in their method of destruction, all lead to cell death.

The *p53* gene is one of the best known apoptosis modulators and it is inactivated by mutations in greater then 50% of cancers [25]. Loss of *p53* expression triggers a vital apoptotic pathway, capable of initiating cell death in response to a number of abnormal cellular conditions [25]. The coding region of the *p53* gene is hypermethylated in normal cells and mutations resulting from deamination of 5-methylcytosine have been shown to
lead to p53 inactivation [26]. Mouse Double Minute 2 (MDM2) is a nuclear protein that binds and represses the function of the p53 protein. The cyclin-dependant kinase inhibitor 4a gene (INK4a) encodes two proteins: p16\textsuperscript{INK4a} and p14\textsuperscript{ARF}, the latter is generated from an alternate reading frame. p14\textsuperscript{ARF} is a nuclear protein that binds and represses the function of MDM2, thus allowing p53 to carry out its apoptotic role. Hypermethylation of the 5’ control region of the \textit{INK4a/ARF} gene results in transcriptional repression of p14\textsuperscript{ARF} and allows MDM2 to bind and repress p53 [27]. Hypermethylation plays an extremely important role in preventing p53 from functioning thereby allowing oncogenesis to occur.

There are two major pathways that induce apoptosis-triggered cell death both of which can function independently but also interact with one and other; mitochondrial-mediated and caspase-mediated pathways [24]. Activation of the mitochondrial pathway results in various outcomes, one of which is hypoxia-induced cell death. BCL2/adenovirus E1B 19kDa interacting protein 3 (BNIP3) is a member of the B-cell lymphoma-2 (BCL-2) family of apoptotic-linked proteins, and when activated, localizes to the mitochondria. Localization of BNIP3 to the mitochondria results in hypoxia-induced cell death [28]. The promoter of BNIP3 contains a CpG island and is a target of hypermethylation-induced silencing [29]. Silencing of BNIP3 allows cancer cells to avoid cell death induced by hypoxia [28, 29].

There are multiple caspase pathways, all of which result in a dismantling of cellular and organelle membranes. Caspase-8, called an initiator caspase, is one of the first caspases activated, and its activation triggers the other caspases to begin the dismantling process [30]. Inactivation of caspase-8 has been shown to be a result of
hypermethylation in its 5’ control region and results in inhibition of apoptosis [31].

Apoptosis is a complex process that can be initiated by a multitude of abnormal cellular conditions and induces cell death through numerous pathways. Inactivation of apoptotic pathways is a common trend across a majority of cancers [32]. Epigenetic alterations contribute to the dysregulation of apoptotic pathways. Another consequence of blocking apoptotic pathways is a resistance to anti-cancer drugs, many of which function by inducing cell death [33].

**Insensitivity to anti-growth signals**

The propensity of cancer cells to proliferate at rates much greater than normal cells lies in their ability to bypass cell cycle checkpoints. Multiple pathways affect cellular proliferation and one pathway in particular is dysregulated in a majority of cancers. The retinoblastoma (Rb) pathway regulates the G1 to S checkpoint and is one of the effectors of anti-growth signals. Both transition to S phase and anti-growth signaling pathways involve signal transduction via phosphorylation to induce or inhibit cellular division. Hypermethylation reduces transcription of inhibitors of these pathways allowing the cell to continuously divide despite the lack of sufficient levels of growth signals or the presence of anti-growth signals. A high rate of proliferation is a major contributor to oncogenesis.

Signaling through the TGFβ pathway can lead to a block in cellular division. The Rb pathway is one of the targets of TGFβ signaling and the anti-growth signals function to prevent Rb phosphorylation. p16$$^{\text{INK4a}}$$, the other protein encoded from the $$\text{INK4a/ARF}$$ locus, is a well known tumor suppressor that asserts its effect by binding to and inhibiting cyclin dependent kinases (CDK) 4 and 6. Binding of p16$$^{\text{INK4a}}$$ to CDK4/6 prevents the
CDK complex from binding to cyclin D [34]. When the CDK4/6 complex is bound to cyclin D, the CDK4/6-cyclin D complex will phosphorylate Rb. When Rb is unphosphorylated it is bound to E2F1, a transcription factor that promotes G1 to S transition [35]. When Rb is phosphorylated, E2F1 is free to induce the G1 to S transition. In normal cells the activity of p16^{INK4a} allows the p16^{INK4a}/Rb pathway to act as a cell cycle checkpoint and in combination with TGFβ signaling provides an excellent means to prevent unnecessary proliferation [34]. This inactivation of p16^{INK4a} in cancer cells results from hypermethylation of its promoter and subsequent gene silencing [36]. Inactivation of p16 causes Rb to be constantly phosphorylated despite anti-growth signaling from TGFβ. This is yet another example of how the silencing effects brought on by hypermethylation can allow systems to be bypassed that normally serve to prevent oncogenesis.

A recent study has opened up a new area of oncogenic effects exerted by hypermethylation. The great majority of cases of hypermethylation result in silencing of tumor suppressors, however through silencing microRNA (mirRNA) genes via hypermethylation, oncogenes can become up-regulated. mirRNAs are endogenous noncoding RNAs that use RNAi pathways to post-transcriptionally silence expression of a target mRNA [37]. Lujambio et al. have shown that miR-124a, a gene that encodes a mirRNA which contains sequences complementary to CDK6 mRNA, is downregulated in tumors via hypermethylation of its promoter. Down-regulation of miR-124a causes an upregulation of CDK6 [38]. Upregulation of CDK6 leads to Rb phosphorylation, and subsequent increase in cellular proliferation [38]. The functions of mirRNAs are not fully understood and as their role in cellular biology is revealed, DNA methylation will
surely be shown to play a major role in their regulation.

**Sustained angiogenesis**

Cancer cells, like normal cells, need a constant supply of oxygen and nutrients to function properly. The process of angiogenesis creates the vessels required to sustain an abnormally high growth rate. In order for neovascularization to occur, the extracellular matrix (ECM) must be degraded and subsequently restructured. The matrix metalloproteinase (MMP) family of proteases are secreted into the ECM, facilitating its degradation [39]. Another initiator of angiogenesis is the vascular endothelial growth factor (VEGF), which promotes vessel formation through signal transduction-mediated activation of downstream elements [40]. One of the receptors of VEGF, VEGFR2, has been shown to induce angiogenesis through integrin activation [41]. Tissue inhibitor of metalloproteinases-3 (TIMP3) inhibits MMPs preventing them from degrading the ECM [42]. TIMP3 also prevents VEGF-induced angiogenesis by binding to VEGFR2 [43]. Down-regulation of TIMP3 is due in part to hypermethylation of its promoter region, and leads to increased angiogenic capabilities [44].

Various signaling systems initiate the formation of the required vessels and here also hypermethylation leads to angiogenesis by silencing genes that serve to counteract the angiogenic process. Another anti-angiogenic protein is thrombospondin-1 (TSP1). Its expression has been shown to inhibit oncogenesis and the down-regulation of this protein is a result of methylation of CpG islands in the promoter of the *TSP1* gene [45]. TSP1 down-regulation allows the ECM to be restructured during oncogenesis [46].

**Limitless replicative potential**

The ability of cancer cells to divide indefinitely is due in part to the multimeric
enzyme telomerase. Normal cells display low to undetectable levels of telomerase activity; however, immortal cells such as stem cells and cancer cells display high levels of telomerase activity [47,48]. Telomerase adds repeats of DNA hexamers to the ends of chromosomes and helps to circumvent the end-replication problem. The enzyme telomerase is composed of key subunits. One component is hTR, the RNA template that is used to synthesize the hexameric DNA. Another component is human telomerase reverse transcriptase (hTERT). The expression of hTERT has been shown to directly correlate with telomerase activity [49]. Due to the role of hTERT in extending the lifespan of cancer cells, hTERT gene regulation has been one of the prime targets for anticancer therapies. By down-regulating telomerase activity, the chromosomes of cancer cells will eventually become so short that the cells can no longer function properly and will enter into senescence.

The hTERT gene has for a long time been one of the few examples of DNA methylation leading to gene activation. Multiple reports have shown that hypermethylation of key CpG dinucleotides in the 5’ control region of the hTERT gene leads to its activation in cancer cells. However, recent studies, using more sensitive assays, have shown that an area of the hTERT promoter within close proximity of the transcriptional start site must remain unmethylated for transcription to occur [50,51]. These new reports for the first time provide insight into how hTERT transcription can occur despite hypermethylation, and illustrate that specific patterns of methylation involving key sites are more likely to control gene expression than total hypermethylation or hypomethylation. Though an area of the hTERT regulatory region must remain hypomethylated, it has also been shown that hypermethylation is still necessary for
hTERT up-regulation. CCCTC-binding factor (CTCF), a transcriptional repressor with a zinc finger domain, whose binding is methylation sensitive, has been shown to bind to an unmethylated area within the first exon of hTERT [52]. Hypermethylation of hTERT, as a result of oncogenesis, prevents the binding of CTCF. Despite this new finding, hTERT maintains its distinction as one of the few genes that requires hypermethylation for activation.

**Tissue invasion and metastasis**

Cancer metastasis has been associated with highly invasive tumors and usually results in a poor prognosis. One of the major hallmarks of malignant tumors is the loss of cellular adhesiveness allowing the cells to separate and colonize other tissues. Dysregulation of cytoskeletal signaling proteins may also play a role in promoting cancer metastasis.

Important proteins that work to maintain cell-to-cell junctions are the cadherins and the catenins, which include α-catenin, E-cadherin and β-catenin. α-catenin is known to connect E-cadherin and β-catenin to actin filaments which confer stability to the cell-cell junctions [53]. Through these observations it is suggested that suppression of E-cadherin can cause the release of cancer cells from primary tumors [54,55]. Suppression of E-cadherin expression through DNA hypermethylation was illustrated in a study by Yoshiura et al. where hypermethylation around the promoter was present and treatment with 5-azacytidine induced E-cadherin transcription [56]. The E-cadherin gene was also examined in vivo and also showed hypermethylation of its promoter region in hepatocellular carcinomas [56].

Focal adhesion proteins localized at ECM contact sites are important in
establishing proper cytoskeletal signaling. Recently, integrin linked kinase (ILK) has emerged as a crucial protein in focal adhesion and has been shown to interact with LIM- and senescent-cell-antigen-like-domains-1 (LIMS1) and parvin to form complexes in ECM sites [57-59]. A protein highly similar to LIMS1, termed LIMS2, has been illustrated to also form complexes with ILK and parvin; however, the LIMS1 and LIMS2 complexes are mutually exclusive suggesting a regulatory role of LIMS2 on the LIMS1-

<table>
<thead>
<tr>
<th>Hallmarks of Cancer</th>
<th>Gene</th>
<th>Effect Exerted by Methylation</th>
<th>Contribution to Oncogenesis</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evasion of Apoptosis</td>
<td>p53</td>
<td>Mutation</td>
<td>Mutation-linked inactivation prevents p53 from initiating apoptotic pathways.</td>
<td>[26]</td>
</tr>
<tr>
<td></td>
<td>p14ARF</td>
<td>Silencing</td>
<td>Silencing of p14ARF allows MDM2 to suppress the function of p53.</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>BNIP3</td>
<td>Silencing</td>
<td>Silencing of BNIP3 prevents hypoxia-linked apoptosis.</td>
<td>[29]</td>
</tr>
<tr>
<td>Caspase-8</td>
<td>Silencing</td>
<td>Silencing of caspase-8 prevents activation of caspase-associated apoptosis</td>
<td>[31]</td>
<td></td>
</tr>
<tr>
<td>Insensitivity to Anti-growth signals</td>
<td>p16INK4a</td>
<td>Silencing</td>
<td>Silencing of p16INK4a prevents Rb from carrying out its role in cell cycle regulation.</td>
<td>[36]</td>
</tr>
<tr>
<td></td>
<td>miR-124a</td>
<td>Silencing</td>
<td>Silencing of miR-124a causes an increase in CDK6 expression, which leads to Rb phosphorylation.</td>
<td>[38]</td>
</tr>
<tr>
<td></td>
<td>TIMP3</td>
<td>Silencing</td>
<td>Silencing of TIMP3 leads to: 1) MMPs to degrade ECM. 2) VEGF-mediated growth signals to be efficiently transduced.</td>
<td>[44]</td>
</tr>
<tr>
<td>Sustained Angiogenesis</td>
<td>TSP1</td>
<td>Silencing</td>
<td>Silencing of TSP1 promotes vessel formation.</td>
<td>[45]</td>
</tr>
<tr>
<td>Limitless Replicative Potential</td>
<td>hTERT</td>
<td>Activation</td>
<td>Activation of hTERT allows telomerase to maintain telomeres by addition of hexameric repeats.</td>
<td>[52]</td>
</tr>
<tr>
<td>Tissue Invasion and Metastasis</td>
<td>E-Cadherin</td>
<td>Silencing</td>
<td>Silencing of E-Cadherin leads to degradation of cell-to-cell junctions.</td>
<td>[56]</td>
</tr>
<tr>
<td></td>
<td>LIMS2</td>
<td>Silencing</td>
<td>Silencing of LIMS2 leads to a loss of focal adhesion at ECM contact sites.</td>
<td>[60]</td>
</tr>
</tbody>
</table>
ILK-parvin complex. Recently gastric cancer cell lines have shown increased methylation and associated silencing of LIMS2. siRNA reduction of LIMS2 expression significantly increased cell migration in SNU-484 and SNU-668 gastric cancer cell lines suggesting a regulatory role in cell migration for LIMS2. Thus, silencing of LIMS2 transcription by hypermethylation in gastric cancer may play a key role in invasiveness and metastasis [60].

Table 1 summarizes the diverse mechanisms by which hypermethylation promotes oncogenesis. The genes reviewed in this chapter are just a few of the genes regulated by DNA methylation and each year more and more genes are being reported to be modulated by hypermethylation. With the vast range of effects exerted by DNA methylation it comes as no surprise that much focus has been placed on the discovery of cancer therapies aimed at preventing, reversing, and in the case of oncogenes, inducing, DNA methylation. DNA methylation plays such a vital role in oncogenesis, before long it will be considered a hallmark of cancer itself.

**Cancer Therapeutics and Hypermethylation**

Multiple therapies have been created over the past decade to treat neoplasia by demethylation of DNA. Most of these therapies reduce methylation through inhibition of DNMTs. One class of compounds used as demethylation agents are nucleoside analogues. They have a structure similar to that of cytosine, except they have a modification at their 5 carbon and when incorporated into DNA they will covalently bind DNMTs to the DNA [61]. The DNMTs will remain bound to the DNA preventing them from carrying out methylation elsewhere. Azacytidine, zebularine, and decitabine have
all shown an ability to decrease methylation through DNMT inhibition [62]. Although these compounds are very unstable, superior delivery and handling methods have been designed to increase their efficiency as demethylating agents [63].

The major polyphenol in green tea is (−)-epigallocatechin-3-gallate (EGCG) and it has been shown to reduce hypermethylation through inhibition of DNMTs [64]. EGCG is readily available, making it an ideal compound for cancer prevention and therapy. Other molecules such as genistein, procainamide, antisense oligonucleotides, and siRNAs have all been shown to reduce hypermethylation and will be discussed in greater detail later on in this book [65-68].

**Conclusion**

The rest of this book will cover in detail the epigenetics of cancer and in almost every instance DNA methylation will be involved in some capacity. Hypermethylation, in the promoters of tumor suppressor genes, leads to silencing of gene expression and is a primary effect exerted by DNA methylation. Although the exact mechanisms by which these genes are targeted for silencing is yet to be revealed, great progress has been made over the last decade. Contributing to nearly every aspect of carcinogenesis, hypermethylation may provide one of the most promising targets for cancer therapeutics, and may someday be a mainstay of cancer therapy.
References:


110


AN OVERVIEW OF EPIGENETIC ASSAYS

by

JOSEPH TYSON DEANGELIS*, WOODROW J. FARRINGTON*, and TRYGVE O. TOLLEFSBOL

*These Authors Contributed equally to this work

Molecular Biotechnology

Copyright 2008 by Humana Press
Used by permission

Format adapted for dissertation
Abstract

A significant portion of ongoing epigenetic research involves the investigation of DNA methylation and chromatin modification patterns seen throughout many biological processes. Over the last few years, epigenetic research has undergone a gradual shift and recent studies have been directed towards a genome wide assessment. DNA methylation and chromatin modifications are an essential component of the regulation of gene activity. DNA methylation effectively down regulates gene activity by addition of a methyl group to the 5’ region of a cytosine base. Less specifically, modification of the chromatin structure can be carried out by multiple mechanisms leading to either the upregulation or down-regulation of the associated gene. Of the many assays used to assess the effects of epigenetic modifications, chromatin immunoprecipitation (ChIP), which serves to monitor changes in chromatin structure, and bisulfite modification, which tracks changes in DNA methylation, are the two most commonly used techniques.
Introduction

The field of epigenetics has rapidly developed into one of the most influential areas of scientific research and has been shown to regulate essential biological processes such as aging, development, and memory formation (1-3). The number of pathologies linked to the dysregulation of epigenetic systems continues to grow and with it the list of potential targets for epigenetic-based therapeutics. In recent years, epigenetic research has seen a shift from site-specific studies aimed at determining how epigenetic processes, such as DNA methylation and histone acetylation, regulate specific genes to a genome-wide assessment of when and why epigenetic alterations occur. The assays which are used to monitor changes in the epigenetic environment have also undergone a similar progression, marked by both the appearance of new and innovative assays as well as the modification and coupling of more traditional epigenetic assays such as bisulfite modification of DNA or chromatin immunoprecipitation (ChIP).

Epigenetics is widely defined as any heritable change among DNA and its surrounding chromatin that does not alter the sequence of the DNA. Epigenetic regulation of gene expression has been linked to discrete mechanisms that affect the stability, folding, positioning, and organization of DNA (1). The most studied of these mechanisms includes DNA methylation and chromatin remodeling, which work synergistically to organize the genome into transcriptionally active and inactive zones. Patterns of DNA methylation are established during development, and once established remain fairly stable though adult life (1). On the other hand, the modification of histones is more of a reversible process, allowing genes to be switched between transcriptionally active and inactive states. Events such as tumorigenesis can lead to a drastic alteration in
established methylation patterns or to dysregulation of chromatin remodeling processes, both of which can result in significant and consequential changes in gene expression (4).

In order to better understand these gene regulatory mechanisms, several techniques have evolved that enable investigators to physically map chromatin modifications and changes in DNA methylation patterns. Of these techniques, chromatin immunoprecipitation and bisulfite modification of DNA form the foundation for tracking chromatin changes as well as changes in DNA methylation, respectively. These techniques, coupled with an ever-growing interest in the field of epigenetics, have proven to be the fundamental driving forces responsible for the proliferation of epigenetic research.

**DNA Methylation**

Methylation at the 5-carbon of cytosine is a reversible process that can directly influence gene activity. Essential for normal embryonic development, DNA methylation plays an important role in the regulation of gene expression and affects a wide range of essential biological processes. Patterns of hemimethylation are established throughout the stages of embryogenesis, and alterations among established methylation patterns have a wide range of effects. Aberrant DNA methylation has received the most attention for its role in tumorigenesis but has also been shown to play a role in a number of other genetic disorders.

DNA methylation is a result of the enzymatic activity of the DNA methyltransferase (DNMT) family of enzymes. DNMTs catalyze the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to cytosines in CpG dinucleotides (5). The
vast majority DNA methylation is carried out by DNMT1, DNMT3a, and DNMT3b. DNMT1 was the initial methyltransferase to be identified and is largely responsible for maintaining pre-existing methylation patterns. DNMT1 is known as the maintenance methyltransferase, due to a preference for hemimethylated DNA over unmethylated DNA. The lack of de novo methylation by DNMT1 has led to the designation of DNMT3a and DNMT3b as the de novo methyltransferases, even though they show no preference for nonmethylated over hemimethylated DNA. Inactivation of genes coding for the DNMT3a and DNMT3b family of methyltransferases via gene targeting disrupts de novo methylation in developing stem cells, while displaying no significant loss of maintenance methylation in imprinted methylation patterns (6).

Evolutionarily conserved nonmethylated regions of DNA abundant in cytosine and guanine bases known as CpG islands are the most common sites of DNA hypermethylation. Promoter regions of nearly 40% of mammalian genes have been shown to either contain or to be in close proximity to one or more CpG islands (5). Hypermethylation of CpG islands usually leads to a loss in gene expression or in a small number of cases to upregulation of gene expression (7-8). Methylation profiling of specific CpG islands before, during, and after events such as the formation of a tumor may one day provide physicians with the ability to accurately diagnose diseases based on the epigenetic state of the patient. The onset of CpG island profiling was one of the first steps in the shift to a global study of epigenetic modifications and has led to the discovery of novel tumor suppressor genes regulated via DNA methylation.

The generally accepted principal surrounding DNA methylation is that hypermethylation leads to a loss of gene expression. The physical extension of the
methylated cytosine into the major groove of the DNA can result in a steric hindrance thus rendering the DNA inaccessible to the active site of a transcription factor. Binding of methyl-cytosine binding domain proteins (MBDs) to methylated DNA has also been shown to prevent transcription. In addition to the direct silencing effect of MBDs, it has now become clear that they provide the link that allows DNA methylation and chromatin remodeling to cooperatively prevent gene transcription. Association of MBDs with histone modifying enzymes, such as histone deacetylases (HDACs) and certain histone methyltransferases (HMTs) allows a gene silenced via DNA methylation to remain in a transcriptionally inactive state (9-10).

**Chromatin Remodeling and the Connection to DNA-methylation**

Basic chromatin structure within the nucleus consists of 146 kb sections of DNA wound tightly around a series of histone proteins forming the nucleosome. Histones can be modified by a number of chemical processes including acetylation, methylation, and phosphorylation, all of which can display profound effects on the regulation of gene expression (11). Levels of histone acetylation are a function of the delicate balance that exists between acetyltransferases (HATs) and HDACs. In general, histone acetylation leads to an increase in gene activity, while histone deacetylation leads to transcriptional repression. Acetylation of the histone tails by HATs exposes the underlying DNA to a variety of cellular factors that directly affect gene expression. HDACs catalyze the removal of acetyl groups from histone tails, which leads to the local condensation of chromatin and the rendering of associated DNA inaccessible (12, 13). Histone acetylation/deacetylation has been the most studied of histone modifications; however,
recently it has been shown that methylation of specific histone residues by HMTs can also play an important role in gene regulation. HMTs substitute an acetyl group with a methyl group at specific arginine and lysine residues on histone tails. Methylation of lysine residue K9 on histone H3 results in decreased transcription of the gene (14).

It is now believed that most genes silenced as a result of epigenetic changes undergo alterations in both chromatin structure and methylation status. Many of the epigenetic modifying enzymes have been shown to interact with and recruit one another. DNA methylation has been shown to influence chromatin structure due to the association of a type of MBD, methyl CpG binding proteins (MeCPs), with HDACs and HMTs. Localization of histone modifying enzymes to a gene targeted for silencing helps to maintain the gene in a transcriptionally inactive state (15). MeCPs have also been shown to interact with the DNMTs, leading to a greater likelihood that methylation patterns will be efficiently maintained (16). The complex and cyclic interactions of the various epigenetic modulators make it very difficult to attribute a change in gene expression to one single factor. As a result, investigators are now trying to understand in what order epigenetic changes occur. Does DNA methylation lead to histone deacetylation or vice versa?

In addition to attempting to understand the timing of epigenetic modifications, researchers are now trying to determine exactly how the cellular environment contributes to and regulates DNA methylation and chromatin structure. One of the most important mysteries facing epigenetists today is the elucidation of the chain of events that leads to the appearance of aberrant methylation patterns during tumorigenesis. A much more detailed picture of the shifts in methylation patterns and alterations in chromatin structure
occurring across the entire genome during abnormal processes is direly needed. The answer may lie in methylation profiling of CpG islands and other similar “epigenomic” studies that will provide insight into the timing of epigenetic alterations.

**Analysis of Changes in DNA Methylation**

There are many techniques to analyze changes in DNA methylation, with the best method depending on factors including but not limited to the availability of the DNA, total number of targets being analyzed or the desired specificity. Bisulfite modification of DNA is the foundation for the majority of assays geared toward analyzing changes in methylation patterns. The differences in bisulfite-based methylation assays arise from the manner in which bisulfite-modified DNA is analyzed. Bisulfite modification converts nonmethylated cytosines to uracils, which are then converted to thymines, whereas methylated cytosines are protected from bisulfite modification.

DNA sequencing and the use of methylation-sensitive primers (MSPs) are the two most commonly used techniques to analyze bisulfite-treated DNA. Sequencing analysis of bisulfite-modified DNA can be used to reveal the methylation status of specific cytosines, whereas MSPs can be used to quickly assess a larger number of CpG islands. Recent studies have shown that a combination of bisulfite techniques and chromatin immunoprecipitation assays can allow assessment of methylation status and chromatin structure from one sample. Zinn and colleagues have shown that DNA collected from ChIP can be analyzed for methylation status using MSPs (17).

Single nucleotide primer extension (SNuPE) is another means to analyze bisulfite-modified DNA. The extension of an oligonucleotide to the 5’ end of a CpG site using
dideoxycytidines (ddCTP) or dideoxythymidine (ddTTP) followed by real-time PCR, allows for a quantitative assessment of methylation patterns and can be applied to multiple sites simultaneously (18). A semi-quantitative method known as methylation sensitive-single strand conformation analysis (MS-SSCA) can be used to obtain an overall picture of DNA methylation. MS-SSCA can be applied across a broader range of samples and can be used to assess the ratio of methylated to nonmethylated DNA (19).

One of the oldest methods of mapping methylation alterations on a genomic scale does not use bisulfite treatment of DNA. Digestion of genomic DNA with endonucleases that differ in their methylation sensitivities is still an ideal method for obtaining a rough estimate of the totality of methylation. One of the best assays to assess a large number of CpG islands is restriction landmark genomic scanning (RLGS). This method involves the radioactive labeling of nonmethylated sequences that are targets of methylation sensitive restriction enzymes (20).

Methods of Assessing Chromatin Modifications

Assays aimed at assessing epigenetic changes have evolved remarkably over the last half decade. The chromatin immunoprecipitation (ChIP) assay, which assesses changes in chromatin structure, comprises one of the most utilized assays in epigenetic research. ChIP assays monitor DNA-protein interactions and allow the chromatin structure surrounding a specific DNA sequence to be analyzed. A conventional ChIP (xChIP) uses formaldehyde to crosslink DNA and protein, followed by immunoprecipitation of DNA-protein complexes. Once the crosslinks are reversed, recovered DNA can then be analyzed using PCR. Another commonly used form of the
ChIP assay is the native ChIP (nChIP). nChIP uses micrococcal nuclease digestion to prepare the chromatin for analysis. nChIP allows for modifications of histones, such as methylation or acetylation, to be assessed more accurately than with formaldehyde fixation; however, nChIP does not usually allow for assessment of proteins with a weak binding affinity for DNA (21). Most ChIP assays are semi-quantitative although combining either ChIP assay with real-time PCR (Q-ChIP) can achieve a quantitative measurement of the amount of DNA bound to a specific protein.

ChIP assays can also be combined with other epigenetic assays such as DNA bisulfite modification. DNA harvested from a ChIP assay can be treated with bisulfite and methylation specific primers (MSP) can be used to assess changes in DNA methylation in a ChIP-MSP (17). New uses for the ChIP assay are continuously appearing as epigenetic research comes of age. With the recent surge in epigenetic research, multiple modifications of ChIP assays have begun to appear, which allow assessment of global changes in chromatin structure. One of the most useful techniques to assess genome-wide epigenetic changes is the ChIP on Chip assay that utilizes traditional ChIP protocols combined with microarray analysis (22).

In addition to ChIP, many other assays exist that can be used to assess chromatin structure. For example, DNaseI hypersensitivity assays can be used if a more general determination of the changes chromatin has undergone is desired. DNaseI hypersensitivity sites are usually located in or around promoter regions thereby allowing for mapping of transcriptionally active vs. inactive chromatin (23). One of the most useful techniques to assess changes in chromatin structure is the use of the deacetylating agent trichostatin A (TSA), which has also been shown to be capable of treating diseases
caused by dysregulation of histone acetylation (24). At low concentrations, TSA inhibits HDACs and allows determination of exactly what role acetylation/deacetylation plays in regulation of a specific gene.

**Analysis of Epigenetic Modulating Enzymes**

Comparison of results obtained from methylation or chromatin structure analysis with results obtained from quantification of either the expression or analysis of epigenetic players like DNMTs, HDACs, or MeCPs can provide valuable insight into the causes of epigenetic alterations. Analysis of the differences in mRNA and protein levels can be quantitatively preformed using Real-time PCR and Western blot protocols, respectively. Immunoprecipitation assays such as ChIP or co-immunoprecipitation (co-IP) can be highly effective means of assessing the functions of epigenetic modulators. For example, ChIP analysis using antibodies specific to MeCPs has been used to identify new sites of DNA hypermethylation during oncogenesis (25). If an analysis of the interactions between epigenetic modulators is needed, a co-IP can be performed.

Immunoprecipitation using antibodies specific to one enzyme, followed by detection with antibodies specific to others, can help to determine recruitment capabilities. Most of the assays that are used to assess enzymatic activity involve addition of synthetic substrates specific to the enzyme being analyzed (i.e., DNA rich in CpG dinucleotides for DNMT activity assays) with cellular protein extracts. The differences in the various activity assays stem from the type of substrate used as well as from the method used to quantify enzymatic activity.
Conclusion

Over the last decade it has become very clear that epigenetic mechanisms play important roles in gene regulation. In epigenetic research two major questions have still yet to be fully answered: 1) What events lead to the dysregulation of epigenetic systems?, and 2) How do certain epigenetic changes regulate the appearance of other aberrant epigenetic alterations? To answer these questions, epigenetists now have begun a genome-wide assessment of epigenetic alterations during certain biological processes. Elucidation of both the timing and the interactions of the epigenetic modulating enzymes may hold the key to unlocking what could be one of the 21st century’s greatest scientific discoveries. As epigenetics truly comes of age, so too do the assays that have led to the realization of just how important of a role epigenetic mechanisms play in gene regulation.
References


THE EXPRESSION PROFILES OF EPIGENETIC MODULATING GENES UNDERGO DRASTIC ALTERATIONS DURING BREAST TUMORIGENESIS

by

JOSEPH TYSON DEANGELIS, TRYGVE TOLLEFSBOL

In preparation for Science

Format adapted for dissertation

129
ABSTRACT

As important to the epigenetic code as the code itself are the enzymes that are responsible for maintaining it. Epigenetic modulating enzymes establish and maintain both the histone code as well as the patterns of DNA methylation. This study illustrates that alterations in the gene expression of epigenetic modulating enzymes may be one of the initial events that leads to drastic alterations in the epigenetic code that occurs during tumorigenesis. Using previously published microarray data in combination with clustering analysis we identified two distinct gene clusters in a cell culture model of breast tumorigenesis based on the expression profiles of 63 epigenetic modulating genes. One cluster of epigenetic modulators was down regulated following introduction of hRASV12 to HMECs stably expressing SV40 early region and hTERT and one cluster of epigenetic modulating enzymes that was upregulated. Nearly every known family of epigenetic modulating enzymes was represented, yet no one family was specific to one cluster or another. Future studies will be aimed at understanding what is driving the switch from one expression profile to another and to understand what areas of the genome are most affected by this global change in epigenetic modulator expression.
All epigenetic modulating enzymes can be grouped into one of the three following categories: 1) Writers or enzymes that write the epigenetic code by the addition of functional groups to either DNA or histone proteins 2) Erasers or enzymes that erase the epigenetic code by removing functional groups from histone proteins and 3) Readers or enzymes that recognize a specific epigenetic mark and bind to DNA or histone proteins in that genomic region. Examples of writers include the DNA methyltransferases (DNMTs) such as DNMT3b, histone methyltransferases (HMTs) such as the SET family of enzymes, and histone acetyltransferases (HATs) such as the MYST family of enzymes. Examples of erasers include histone demethylases such as the Jumanji family of enzymes and histone deacetylases (HDACs) such as the sirtuin family of enzymes. Examples of readers include methyl binding domain (MBD) proteins, polycomb proteins, chromobox proteins and DNMT3L.

A vast majority of epigenetic-based research has focused on the investigation of the epigenetic code itself, and how it affects transcription of a particular gene or genes. As important to the etiology of both normal and diseased states as the epigenetic code, is an understanding of the role in which alterations in the expression of epigenetic modulating enzymes contribute to the state of a cell. Studies often focus on one particular epigenetic modification, such as histone acetylation, or one particular family of epigenetic enzymes such as HDACs. However, transcription of a downstream gene is the result of the summation of all epigenetic marks in the vicinity which are maintained by the collective expression and subsequent localization of all the epigenetic modulating enzymes to that vicinity. Epigenetic marks are a highly dynamic phenomenon, asserted by the delicate balance between functional groups being removed and those that are being
added. Disruptions in the presence of a particular epigenetic modulating enzyme would undoubtedly disrupt this balance and possibly lead to an alteration in gene expression.

Three major factors affect the availability of a particular epigenetic enzyme to a specific genomic region. The first factor is interaction with and availability of co-factors and transcription factors. None of the epigenetic modulating enzymes display independent sequence specificity, but instead rely on interaction with enzymes that are capable of binding to specific sequences of DNA, such as transcription factors. The second factor is the presence of specific epigenetic marks. For example, hypermethylation of a CpG island can lead to the binding of MBDs that in turn recruit other epigenetic modulating enzymes such as HDACs. The third and final factor affecting the availability of a particular epigenetic modulating enzyme to a specific genomic region is the gene expression of the epigenetic modulating enzymes themselves. For example we have previously shown that in fetal lung fibroblasts, during the transition to a neoplastic phenotype, DNMT1, 3a, and 3b all display a progressive increase in gene expression. Of these three factors, global changes in the gene expression of the epigenetic modulating enzymes is one of the least investigated. Alterations in the gene expression profiles of epigenetic modulating enzymes that occur during the transition from a normal to a diseased state would have profound effects on the epigenome and may be one of the first steps in the onset of complex pathologies such as cancer.

Using gene expression data from a previous unpublished microarray study, we set out to understand how the expression of genes coding for epigenetic modulating enzymes is altered during breast tumorigenesis. In 2001, Elenbaas et al. illustrated that normal human mammary epithelial (HMEC) cells can be induced to undergo a neoplastic
transformation by the serial retroviral-mediated addition of genetic elements coding for
SV40 early region (SV40ER), hTERT (the catalytic subunit of telomerase), and an
oncogenic form of hRAS. More recently, we used these same procedures to study
proteomic and transcriptomic alterations that occur during breast tumorigenesis.
Utilizing Affymetrix Human GeneST 1.0 arrays, that assesses the expression of 27,000
genes with an average of 26 probes per gene, we illustrated that following addition of the
third and final genetic element, neoplastically transformed HMECs (THMECs)
progressively begin to acquire the gene expression profile associated with a metastatic
state. We monitored three conditions or time points in the neoplastic progression of
breast cancer. The first condition was HMECs stably expressing SV40ER and hTERT
(SHMEC). The second two conditions were THMECs at 40 days and 80 days following
the introduction of hRAS to SHMECs. Using this data set we were interested in
determining which epigenetic modulating enzymes underwent alterations in their
respective gene expression. We were also interested in determining if the expression
profiles of these genes clustered together, possibly due to similar regulatory mechanisms.
First, we generated a gene set comprised of 63 epigenetic modulating genes that
displayed differential gene expression between the three conditions listed above. We
then performed a Welch’s ANOVA with Benjamini-Hochberg multiple test correction
on this gene set to ensure that the differential expression was significant ($q < .05$). The q
values for each gene can be found in Tables 1 and 2. We next used unsupervised
hierarchical clustering to determine if the any particular genes displayed similar
expression profiles. Hierarchical clustering analysis was performed using Genesprings
GX 11, with Pearson’s Centered as the distance metric and complete as the linkage rule.
The results of the unsupervised hierarchical clustering analysis can be viewed in Figure 1. Interestingly, all 63 genes were grouped into one of two well-defined clusters. One cluster contained those genes that were down-regulated following the addition of hRASV12 and one cluster contained those genes that were upregulated following the addition of hRASV12. Once we had established that there were only two primary clusters, we used a k-mean clustering algorithm to generate a mean expression profile for each cluster. The expression profiles of all the epigenetic modulating genes in each cluster across all 12 samples (3 conditions x 4 biological replicates for each condition) can be found in Figure 2. The thick black line on each graph in Figure 2 represents the mean expression profile for that cluster. Each colored line (red or green) corresponds to the expression profile of an individual epigenetic modulator. Figure 2A shows the profile of genes that were down-regulated during breast tumorigenesis. These genes are listed in Table 1. Figure 2B shows the profile of genes that were upregulated during breast tumorigenesis and a list of these genes can be found in Table 2.

Our findings for the first time illustrate that the expression profile of an entire set of genes corresponding to epigenetic modulating enzymes is drastically altered during oncogenesis. A shift as we have found in a large number of epigenetic modulating proteins would be expected to have a drastic effect on gene expression across the entire genome and lead to a severe disruption of the cellular state. What is most interesting is that members of individual families of enzymes such as the HDAC family or the Jumanji family were found in both clusters, as were members of all three groups: Readers, writers, and erasers. Not a single group or family was confined to one cluster, a very interesting finding as nearly every known family of epigenetic modulating enzymes is
represented. This suggests that some epigenetic modulating enzymes may carry out the exact same function as others, but are only utilized in a particular cellular state and thus would affect a different area of the genome.

It is widely accepted that the epigenome undergoes significant modifications that contribute to the onset of an oncogenic state. What is not understand is the factors that are driving the changes that the epigenome undergoes. Our results indicate that alterations in the expression of epigenetic modulating proteins may be this driving force. A complete switch in the relative expression levels of the epigenetic modulating enzymes listed in Tables 1 and 2 would result in the creation of an entirely new epigenetic code. Although our findings are intriguing, they elicit more questions than provide answers. For example, what is driving these observed changes in gene expression? Are there unknown feedback mechanisms? Why are the expression profiles of so many epigenetic modulating genes so similar? It is these questions that our future studies will be aimed at answering. Although a daunting task, chromatin immunoprecipitation (ChIP) on chip analysis performed for each enzyme listed in Tables 1 and 2 would provide a detailed understanding of what regions of the genome were most affected by the change in gene expression of the epigenetic modulators. This type of study would provide an unparalleled understanding of the role of epigenetic mechanisms in oncogenesis.
References


10. Casillas, M. A., Jr.; Lopatina, N.; Andrews, L. G.; Tollefsbol, T. O.,
Transcriptional control of the DNA methyltransferases is altered in aging and

11. DeAngelis, J.; Tollefsbol, T., Identification of a potential role of HNF4a in breast

generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev

difference gel electrophoresis analysis of different time points during the course of
neoplastic transformation of human mammary epithelial cells. In revision for Journal of
Proteome Research. 2010.
Figure 1. Unsupervised hierarchical clustering of differentially expressed epigenetic modulating genes in neoplastically transforming human mammary epithelial cells. Clustering analysis was performed using Pearson’s centered distance metric and complete linkage rule. Each column corresponds to one of the following conditions: SHMEC (HMEC-SV40ER-hTERT), THMEC 40d (HMEC-SV40ER-hTERT-hRASV12, 40d post-hRAS), THMEC 80d (HMEC-SV40ER-hTERT-hRASV12, 80d post-hRAS). Rows correspond to the gene names listed to the right of the heatmap. Each box represents the average normalized intensity value from four biological replicates for that condition for that particular gene. Clustering analysis revealed two well defined clusters. One cluster representing epigenetic modulating genes that were down-regulated (red -> green) following introduction of hRASV12 and one cluster representing genes that were upregulated (green ->red) following introduction of hRASV12.
Figure 2. Profile plot of differentially expressed epigenetic modulating genes. A. Cluster 1 - Genes that were downregulated following introduction of hRASV12. B. Cluster 2 - Genes that were upregulated following introduction of hRAS. Each line on the profile plot corresponds to the intensity value profile for an individual gene in that cluster. The thick black line is the mean for all the genes in that particular cluster. Each of the 12 samples (3 conditions; 4 biological replicates per condition) is represented on the x-axis and normalized intensity values are represented on the y-axis.
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Description</th>
<th>Type</th>
<th>q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDAC1</td>
<td>histone deacetylase 1</td>
<td>E</td>
<td>1.10E-04</td>
</tr>
<tr>
<td>JMJD2A</td>
<td>jumonji domain containing 2A</td>
<td>E</td>
<td>7.62E-05</td>
</tr>
<tr>
<td>PRMT6</td>
<td>protein arginine methyltransferase 6</td>
<td>W</td>
<td>7.10E-05</td>
</tr>
<tr>
<td>SMYD2</td>
<td>SET and MYND domain containing 2</td>
<td>W</td>
<td>8.40E-04</td>
</tr>
<tr>
<td>SUV39H2</td>
<td>suppressor of variegation 3-9 homolog 2 (Drosophila)</td>
<td>W</td>
<td>3.27E-04</td>
</tr>
<tr>
<td>BMI1</td>
<td>BMI1 polycomb ring finger oncogene</td>
<td>R</td>
<td>2.45E-04</td>
</tr>
<tr>
<td>SIRT1</td>
<td>sirtuin 1</td>
<td>E</td>
<td>8.89E-04</td>
</tr>
<tr>
<td>MYST4</td>
<td>MYST histone acetyltransferase (monocytic leukemia) 4</td>
<td>W</td>
<td>.0056</td>
</tr>
<tr>
<td>PCGF5</td>
<td>polycomb group ring finger 5</td>
<td>R</td>
<td>.0037</td>
</tr>
<tr>
<td>JMJD1C</td>
<td>jumonji domain containing 1C</td>
<td>E</td>
<td>.0019</td>
</tr>
<tr>
<td>JMJD2D</td>
<td>jumonji domain containing 2D</td>
<td>E</td>
<td>2.45E-04</td>
</tr>
<tr>
<td>SETD8</td>
<td>SET domain containing (lysine methyltransferase) 8</td>
<td>W</td>
<td>2.45E-04</td>
</tr>
<tr>
<td>JARID1A</td>
<td>jumonji, AT rich interactive domain 1A</td>
<td>E</td>
<td>8.47E-04</td>
</tr>
<tr>
<td>CBX5</td>
<td>chromobox homolog 5 (HP1 alpha homolog, Drosophila)</td>
<td>R</td>
<td>5.97E-04</td>
</tr>
<tr>
<td>SETDB2</td>
<td>SET domain, bifurcated 2</td>
<td>W</td>
<td>3.37E-04</td>
</tr>
<tr>
<td>PRMT5</td>
<td>protein arginine methyltransferase 5</td>
<td>W</td>
<td>.0014</td>
</tr>
<tr>
<td>MYST2</td>
<td>MYST histone acetyltransferase 2</td>
<td>W</td>
<td>4.70E-04</td>
</tr>
<tr>
<td>MBD1</td>
<td>methyl-CpG binding domain protein 1</td>
<td>R</td>
<td>1.21E-04</td>
</tr>
<tr>
<td>MBD2</td>
<td>methyl-CpG binding domain protein 2</td>
<td>R</td>
<td>4.85E-04</td>
</tr>
<tr>
<td>UHRF1</td>
<td>ubiquitin-like, containing PHD and RING finger domains, 1</td>
<td>R</td>
<td>.0026</td>
</tr>
<tr>
<td>JMJD1B</td>
<td>jumonji domain containing 1B</td>
<td>E</td>
<td>.0013</td>
</tr>
<tr>
<td>HAT1</td>
<td>histone acetyltransferase 1</td>
<td>W</td>
<td>2.91E-04</td>
</tr>
<tr>
<td>SETD8</td>
<td>SET domain containing (lysine methyltransferase) 8</td>
<td>W</td>
<td>3.47E-04</td>
</tr>
<tr>
<td>MBD4</td>
<td>methyl-CpG binding domain protein 4</td>
<td>R</td>
<td>2.36E-04</td>
</tr>
<tr>
<td>CLOCK</td>
<td>clock homolog (mouse)</td>
<td>W</td>
<td>2.36E-04</td>
</tr>
<tr>
<td>SETD7</td>
<td>SET domain containing (lysine methyltransferase) 7</td>
<td>W</td>
<td>.0015</td>
</tr>
<tr>
<td>JMJD1B</td>
<td>jumonji domain containing 1B</td>
<td>E</td>
<td>3.27E-04</td>
</tr>
<tr>
<td>HDAC3</td>
<td>histone deacetylase 3</td>
<td>E</td>
<td>.0037</td>
</tr>
<tr>
<td>SIRT5</td>
<td>sirtuin 5</td>
<td>E</td>
<td>3.42E-04</td>
</tr>
<tr>
<td>JARID2</td>
<td>jumonji, AT rich interactive domain 2</td>
<td>E</td>
<td>.0015</td>
</tr>
<tr>
<td>HDAC2</td>
<td>histone deacetylase 2</td>
<td>E</td>
<td>3.05E-04</td>
</tr>
<tr>
<td>HDAC9</td>
<td>histone deacetylase 9</td>
<td>E</td>
<td>.0201</td>
</tr>
<tr>
<td>CBX3</td>
<td>chromobox homolog 3</td>
<td>R</td>
<td>2.45E-04</td>
</tr>
<tr>
<td>EZH2</td>
<td>enhancer of zeste homolog 2 (Drosophila)</td>
<td>W</td>
<td>5.04E-04</td>
</tr>
<tr>
<td>JMJD2C</td>
<td>jumonji domain containing 2C</td>
<td>E</td>
<td>2.36E-04</td>
</tr>
<tr>
<td>UTX</td>
<td>ubiquitously transcribed tetratricopeptide repeat, X chrom</td>
<td>E</td>
<td>8.47E-04</td>
</tr>
<tr>
<td>SUV39H1</td>
<td>suppressor of variegation 3-9 homolog 1 (Drosophila)</td>
<td>W</td>
<td>2.45E-04</td>
</tr>
<tr>
<td>HDAC8</td>
<td>histone deacetylase 8</td>
<td>E</td>
<td>4.50E-04</td>
</tr>
</tbody>
</table>
### Table 2. Gene Cluster 2

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Description</th>
<th>Type</th>
<th>q value</th>
</tr>
</thead>
<tbody>
<tr>
<td>JMJD4</td>
<td>jumonji domain containing 4</td>
<td>E</td>
<td>.0111</td>
</tr>
<tr>
<td>MBD6</td>
<td>methyl-CpG binding domain protein 6</td>
<td>R</td>
<td>1.21E-04</td>
</tr>
<tr>
<td>SIRT4</td>
<td>sirtuin 4</td>
<td>E</td>
<td>3.47E-04</td>
</tr>
<tr>
<td>SETD1B</td>
<td>SET domain containing 1B</td>
<td>W</td>
<td>8.47E-04</td>
</tr>
<tr>
<td>HDAC7</td>
<td>histone deacetylase 7</td>
<td>E</td>
<td>9.06E-05</td>
</tr>
<tr>
<td>SETD1A</td>
<td>SET domain containing 1A</td>
<td>W</td>
<td>2.45E-04</td>
</tr>
<tr>
<td>MYST1</td>
<td>MYST histone acetyltransferase 1</td>
<td>W</td>
<td>8.47E-04</td>
</tr>
<tr>
<td>CREBBP</td>
<td>CREB binding protein (Rubinstein-Taybi syndrome)</td>
<td>R</td>
<td>8.47E-04</td>
</tr>
<tr>
<td>JMJD3</td>
<td>jumonji domain containing 3, histone lysine demethylase</td>
<td>E</td>
<td>9.06E-05</td>
</tr>
<tr>
<td>CBX2</td>
<td>chromobox homolog 2 (Pc class homolog, Drosophila)</td>
<td>R</td>
<td>.0037</td>
</tr>
<tr>
<td>PCGF2</td>
<td>polycomb group ring finger 2</td>
<td>R</td>
<td>7.62E-05</td>
</tr>
<tr>
<td>HDAC5</td>
<td>histone deacetylase 5</td>
<td>E</td>
<td>2.45E-04</td>
</tr>
<tr>
<td>CBX1</td>
<td>chromobox homolog 1 (HP1 beta homolog Drosophila)</td>
<td>R</td>
<td>.0023</td>
</tr>
<tr>
<td>JMJD6</td>
<td>jumonji domain containing 6</td>
<td>E</td>
<td>8.47E-04</td>
</tr>
<tr>
<td>CBX4</td>
<td>chromobox homolog 4 (Pc class homolog, Drosophila)</td>
<td>R</td>
<td>.0084</td>
</tr>
<tr>
<td>DOT1L</td>
<td>DOT1-like, histone H3 methyltransferase (S. cerevisiae)</td>
<td>E</td>
<td>.0010</td>
</tr>
<tr>
<td>MLL4</td>
<td>myeloid/lymphoid or mixed-lineage leukemia 4</td>
<td>W</td>
<td>3.47E-04</td>
</tr>
<tr>
<td>MBD3</td>
<td>methyl-CpG binding domain protein 3</td>
<td>R</td>
<td>3.47E-04</td>
</tr>
<tr>
<td>DNMT3A</td>
<td>DNA (cytosine-5-)-methyltransferase 3 alpha</td>
<td>W</td>
<td>.026</td>
</tr>
<tr>
<td>PCGF1</td>
<td>polycomb group ring finger 1</td>
<td>R</td>
<td>.0040</td>
</tr>
<tr>
<td>HDAC4</td>
<td>histone deacetylase 4</td>
<td>E</td>
<td>.0043</td>
</tr>
<tr>
<td>DNMT3L</td>
<td>DNA (cytosine-5-)-methyltransferase 3-like</td>
<td>R</td>
<td>.0056</td>
</tr>
<tr>
<td>HDAC10</td>
<td>histone deacetylase 10</td>
<td>E</td>
<td>.0098</td>
</tr>
<tr>
<td>PCGF3</td>
<td>polycomb group ring finger 3</td>
<td>R</td>
<td>.0091</td>
</tr>
</tbody>
</table>

### Tables 1 and 2 Legend

E: Eraser

W: Writer

R: Reader


10. Casillas, M. A., Jr.; Lopatina, N.; Andrews, L. G.; Tollefsbol, T. O.,
Transcriptional control of the DNA methyltransferases is altered in aging and

11. DeAngelis, J.; Tollefsbol, T., Identification of a potential role of HNF4a in breast
cancer brain metastasis. *In preparation for submission to PLoS One,* **2010.**

generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev*

difference gel electrophoresis analysis of different time points during the course of
neoplastic transformation of human mammary epithelial cells. *In revision for Journal of*
*Proteome Research.* **2010.**
GENERAL DISCUSSION

Cancer is the perfect evil. A cancer cell defies the limits of what normal, somatic cells are capable of and it does this with perfection time and time again, in patient after patient, as if it had a purpose. In fact, cancer is too perfect, too perfect to be the result of a random event. Take for instance the recent discovery that certain cancer cells display similar cell surface markers and gene expression profiles to stem cells. The concept of stemness is an evolutionary masterpiece that is the basis for nearly all life. Yet cancer cells reacquire certain stem-like features for no apparent biological purpose. How could something so perfect and so capable be derived outside the bounds of evolution only to be pathological in nature? To this point we have been under the assumption that cancer cells originate from normal cells, although no one has ever actually observed a normal cell transitioning to cancer cell in vivo. It is possible to induce a normal cell to an oncogenic state in a laboratory setting, however, this does not necessarily imply that is the way that oncogenesis occurs in vivo. The Thomson lab has shown that normal human somatic cells can be reprogrammed to a pluripotent state; however there is no recorded example of this occurring in the human body.

So if a cancer cell does not originate from a normal cell, where would it originate from? The answer to that question can be found in a recent discovery by Ratajczak and colleagues. Their group identified a population of stem cells in adult bone marrow called Very Small Embryonic-Like Stem Cells (VSELs), which are the first known
example of a cell possessing pluripotent lineage in the adult body.\textsuperscript{39} VSELs are mobilized into the human blood in response to organ damage and utilize some of the very same signaling pathways that metastatic cells use.\textsuperscript{40} These migratory stem cells have been shown to serve as a global tissue repair mechanism.\textsuperscript{39,41-46}

The notion that cancers are derived from germ-line cells and not from normal somatic cells was proposed nearly a century ago.\textsuperscript{47,48} Ratajczak recently revisited this theory with regard to VSELs.\textsuperscript{49} One of the key tenants to this theory is that the migratory ability of a VSEL provides for the presence of a pluripotent stem cell in all tissues of the adult body. Additionally, the notion that the same cell type, VSEL, can be found in all tissue types infers the possibility of a singular cell type origin to many cancers. This is in stark contrast to the current theories on the origins of cancer that carcinomas can originate from nearly any cell type.

Cancer, in most instances, is associated with or preceded by some type of tissue or cellular damage. This can be an event that may elicit an activation of VSELs and lead to the presence, at the site of the damage, of a cell that already possesses the ability to migrate throughout the human body and has stem-like characteristics. The presence of these cells in the microenvironment of a particular organ would most likely elicit the necessary differentiation pathways for that particular VSEL to become a cell of that particular organ. This would be a cell that, depending on when in the differentiation process it was observed, appear to possess both the features of the surrounding cells, and also those of a migratory stem cells. A dysregulation of VSEL differentiation as an origin to cancer is very similar to the concept that the type of leukemic cell that is
produced is in part due to the point at which a hematopoietic stem cell departs for normal differentiation pathways.

The ability of a cell to migrate throughout the body would require a very specific subset of genes to be activated, and upon arrival at the target tissue, these processes and the corresponding genes would need to be inactivated. A subset of metastasis-related genes known as metastasis suppressor genes (MSG), such as maspin, have already been shown to suppress the metastatic features that certain cancer cells display. However, their function with respect to non-pathological biological processes is for the most part unknown. MSGs could easily serve to inactivate the migratory and morphogenic features of a VSEL upon arrival at the target tissue and subsequent morphogenesis.

It has recently been shown that neoplastic transformation of human mammary epithelial cells produces breast cancer cells that display characteristics of stem cells.\textsuperscript{50} In particular these stem-like cancer cells possess a much higher metastatic potential than other breast cancer cells. \emph{It is theorized that the process of neoplastic reprogramming, which is very similar to pluripotent reprogramming, produces cells that are characteristic of VSELs.} It may actually be possible, by modifying both culture conditions and altering the genes used in transformation, to generate VSELs in a laboratory setting. The ability to synthetically generate VSELs would revolutionize modern medicine. The first step in this process would be to characterize VSELs, at the proteomic, transcriptomic, and methylomic levels, and compare them to the –omic profiles of THMECs produced by this study. This would allow the identification of what these cell types have in common and more importantly what makes them different.
Protocols and/or cell culture conditions could then be altered to accommodate these differences and potentially produce a migratory stem cell from normal, somatic cells.

The research conducted in this dissertation contains two main limitations. The first limitation is that the observed changes in expression could be the result of changes in the ratios of cell types within the cell population. To overcome this limitation it would first be necessary to identify the different cell types within the cell population and then using florescence activated cell sorting, quantify the percentage of each cell type. The second limitation is the transformation model itself. Although an ideal method for studying the initial stages of breast tumorigenesis, no cancer cell is ever generated in vivo by sequential addition/disruption of SV40ER, hTERT, and hRAS. To overcome this limitation, we would need to investigate this in mouse models, and then perform clinical studies.

In summary, this dissertation utilized a neoplastic transformation model of breast oncogenesis to investigate the molecular mechanisms involved in the generation of breast cancers. We utilized transcriptomic and proteomic analyses to identify molecules and pathways that contribute to an onset of an oncogenic phenotype. Future directions of this project include investigation of the causes of a down-regulation of maspin expression and elucidation of the precise role of HNF4a in breast cancer and breast cancer metastasis. Additionally, using DNA harvested during extraction of total cellular protein and total RNA, methylomic experiments are currently underway to understand the alterations the methylome undergoes over the same time period of breast oncogenesis investigated in this dissertation.
GENERAL LIST OF REFERENCES


THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

Institutional Animal Care and Use Committee (IACUC)

NOTICE OF APPROVAL

DATE: March 22, 2010

TO: Tollefsbol, Trygve O.
    CH-175.1170
    934-4573

FROM: Judith A. Kapp, Ph.D., Chair
      Institutional Animal Care and Use Committee

SUBJECT: Title: Epigenetics of Tea Polyphenols in Cancer Prevention
         Sponsor: NIH
         Animal Project Number: 100203406

On March 22, 2010, the University of Alabama at Birmingham Institutional Animal Care and Use Committee (IACUC) reviewed the animal use proposed in the above referenced application. It approved the use of the following species and numbers of animals.

<table>
<thead>
<tr>
<th>Species</th>
<th>Use Category</th>
<th>Number in Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mice</td>
<td>A</td>
<td>120</td>
</tr>
</tbody>
</table>

Animal use is scheduled for review one year from February 2010. Approval from the IACUC must be obtained before implementing any changes or modifications in the approved animal use.

Please keep this record for your files, and forward the attached letter to the appropriate granting agency.

Refer to Animal Protocol Number (APN) 100203406 when ordering animals or in any correspondence with the IACUC or Animal Resources Program (ARP) offices regarding this study. If you have concerns or questions regarding this notice, please call the IACUC office at 934-7852.