DIGITALCOMMONS
@WAYNESTATE

Wayne State University

Wayne State University Dissertations

1-1-2016

Whsc1l1 Regulates Estrogen Receptor Activity In Sum44 Breast Cancer Cells

Jonathan Curtis Irish
Wayne State University,

Follow this and additional works at: https://digitalcommons.wayne.edu/oa_dissertations

 Part of the Oncology Commons

Recommended Citation
Irish, Jonathan Curtis, "Whsc1l1 Regulates Estrogen Receptor Activity In Sum44 Breast Cancer Cells" (2016). *Wayne State University Dissertations*. 1641.
https://digitalcommons.wayne.edu/oa_dissertations/1641

This Open Access Dissertation is brought to you for free and open access by DigitalCommons@WayneState. It has been accepted for inclusion in Wayne State University Dissertations by an authorized administrator of DigitalCommons@WayneState.

WHSC1L1 REGULATES ESTROGEN RECEPTOR ACTIVITY IN SUM44 BREAST CANCER CELLS

by

JONATHAN CURTIS IRISH

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2016

MAJOR: Cancer Biology

Approved By:

_____

Advisor Date

_____

_____

_____

_____

© COPYRIGHT BY

JONATHAN CURTIS IRISH

2016

All Rights Reserved

DEDICATION

I dedicate this work to my loving, patient wife Cesira.

To my parents, Lori and Sean, who have been unwavering in their support of me and my family during my temporally and geographically long educational journey.

Finally, to my grandparents and aunt, Curtis, Norma, and Linda, for motivating me to study cancer.

ACKNOWLEDGEMENTS

I would like to thank my mentor, Dr. Ethier, for providing an environment where I could learn think critically and become proficient at applying the scientific method to problems in biology. I owe my future in science in no small part to the time spent with Dr. Ethier and all of the members of the Ethier lab from whom I was able to learn so much. Thank you also to Dr. Wilson and the MUSC sequencing core for their patience and expertise during my struggles with my next-gen sequencing experiments.

I also would like to thank Dr. Yang for his co-mentorship at Wayne State, and his support in helping me with my understanding of my work in the context of previous work performed by him and others in the Ethier lab, and especially with his help in organizing and structuring my writing for my dissertation.

Thank you to Dr. Sheng for giving me my first chance in the lab, and for allowing me to develop my passion for developing the technical skills required to perform challenging experiments soundly.

To my committee, particularly Dr. Matherly, and Dr. Brush, thank you so much for your guidance and patience. I could not have done this without your support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BrCa	breast cancer
ChIPseq	chromatin immunoprecipitation - next-gen DNA sequencing
CpG	a cytosine followed by a guanine on the same DNA strand
ERa	estrogen receptor alpha protein
ESR1	estrogen receptor alpha gene
FDR	false discovery rate
GSEA	Gene Set Enrichment Analysis
H1	histone subunit H1 gene
H2A	histone subunit H2A gene
H2B	histone subunit H2B gene
H3	histone subunit H3 gene
H3K4me1	histone H3 lysine 4 monomethylation
H3K4me2	histone H3 lysine 4 dimethylation
H3K4me3	histone H3 lysine 4 trimethylation
H3K9me1	histone H3 lysine 9 monomethylation
H3K9me2	histone H3 lysine 9 dimethylation
H3K9me3	histone H3 lysine 9 trimethylation
H3K36me1	histone H3 lysine 36 monomethylation
H3K36me2	histone H3 lysine 36 dimethylation
H3K36me3	histone H3 lysine 36 trimethylation

H4	histone subunit H4 gene
HAT	histone acetyltransferase
HDAC	histone deacetylase
HER2	Erb-B2 Receptor Tyrosine Kinase 2 gene
HMT	histone methyltransferase
iPS	induced pluripotent stem cell
KEGG	Kyoto Encyclopedia of Genes and Genomes
KLF4	Kruppel-like factor 4 gene
LacZ	beta-galactosidase gene
NSD1	Nuclear SET Domain-containing 1 gene
Oct4/POU5F1	POU class 5 homeobox 1 gene
PR	progesterone receptor
RNAseq	mRNA next-gen sequencing
RTK	receptor tyrosine kinase
SAM	s-adenosyl methionine
SOX2	Sex Determining Region Y box 2 gene
shRNA	short-hairpin RNA
shWHshort	shRNA targeting the WHSC1L1 short isoform
shWHlong	shRNA targeting the WHSC1L1 long isoform
shWHtotal	shRNA targeting both long and short WHSC1L1 isoforms
TCGA	The Cancer Genome Atlas
WHSC1	Wolf Hirschhorn Syndrom Candidate gene

WHSC1L1	Wolf Hirschhorn Syndrome Candidate 1-Like One gene
WH-short	WHSC1L1 short isoform (catalytically inactive)
WH-long	WHSC1L1 long isoform
WH-total	both long and short WHSC1L1 isoforms

CHAPTER 1 INTRODUCTION

1.1 General Introduction

Breast cancer, or invasive carcinoma originating from epithelial tissue of the breast, is the most common cancer site in women (excluding skin cancers), and is second only to lung cancer as a cancer-related cause of death in women as of 2011 (placing breast cancer as the third most frequent cause of death in U.S women, behind heart disease and skin cancers, respectively) (SEER 2016).

Breast cancer is estimated to account for 29% of all cancer diagnoses in U.S women as of 2015. While the breast cancer mortality rate has been decreasing at a steady but small rate over the past 25 years (approximately 1%), the rate of breast cancer incidence has remained essentially steady. In terms of odds, a woman living in the U.S. has a 1 in 8 chance of being diagnosed with breast cancer during her lifetime, and a 1 in 40 chance of dying from breast cancer (ACS 2016).

Cancer is not a single disease, but rather a family of approximately 200 illnesses sharing common genetic lesions and pathologies, classified into types primarily by the organ or tissue of origin of the primary tumor. As modern medicine has succeeded in eradicating or radically reducing infectious disease in countries such as the United States, cancer has become increasingly dominant as a cause of death and morbidity.

In summary, breast cancer is a major health issue for women in the United States, and even incremental improvements in the prevention, early diagnosis, and treatment of breast cancer could have a large positive effect on public health.

All cancers have in common the basic property of uncontrolled proliferation of malignant cells derived from the tissue of the host organism, which no longer respond to host regulation of cell morphology, growth, proliferation, or programmed cell death. This uncontrolled proliferation of cells that no longer contribute to the function of their tissue of origin results in the pathology associated with cancer, namely that the increasing physical disruption caused by the rapidly growing cancer cells displaces normal tissue, interfering with normal physiology, as well as competing for nutrients and sometimes eliciting chronic inflammation due to the immune response attempting to clear the tumor cells. While primary tumors can be fatal due to obstruction or crowding out of normal tissue (brain tumors for example), most cancers cause death by metastasizing to organs and tissues distant from the site of the primary tumor, resulting in physical disruption of multiple organs and tissues throughout the body. As eloquently stated by Harold Varmus in his Nobel Banquet speech, cancer is not a foreign pathogen invading our bodies, but is instead "a distorted version of our normal selves", making the fight to eradicate cancer a fight against our own body, a fact that has profound implications when considering effective cancer treatments. Indeed, the fundamental strategy of classic chemotherapy is

to kill enough cancer cells to cause remission before killing so much good tissue that the patient dies from the chemotherapy. This is decidedly rudimentary, and yet classical cytotoxic chemotherapies targeting dividing cells have been the most successful therapies for human cancers to date. As the fields of molecular biology and genomics have matured, the search for targeted cancer therapies, which can be defined as therapies specifically targeting the biology of cancer cells, has become a top priority of cancer research.

For targeted therapy to succeed, a basic requirement is the ability to sensitively and specifically discriminate between cancer cells and normal cells, and then develop targeted therapies that are both effective at killing or senescing cancer cells, while leaving normal cells and physiological processes intact. This basic requirement for the development of targeted cancer therapies is by itself an excellent reason for continued basic research of cancer genomics, as the defining characteristic differentiating a cancer cell from a normal cell is the presence of oncogenic and tumor suppressor inactivating mutations in cancer cells, and their absence in normal cells. As the rest of this thesis will describe, detailed characterization of specific oncogene-activating mutations that frequently occur in breast cancer is a critically important step in the understanding of the biology of breast cancer, and a basic requirement for the intelligent development of targeted therapies to treat cancer while minimizing side effects

1.2 Genomic basis of cancer

Breast cancer, like all known malignancies, is a disease caused by the mutation of specific genes causing either loss of function or loss of regulation. Genes with loss of function upon mutation are termed tumor suppressor genes, and genes which gain constitutive activation or altered function after mutation are termed activated oncogenes. Modern understanding of cancer as a disease caused by mutations in genes occurred relatively recently. The seminal work by Bishop and Varmus in the early 1970s showed conclusively that one of the tumor viruses shown to cause cancer in chickens, v-src, was a mutated remnant of a gene originating in the chicken, c-src, which caused cancer when re-introduced by viral infection into a host chicken (Varmus, Weiss et al. 1972). They also made the connection that humans could and did develop cancer through stochastic mutations in SRC, the human ortholog of c-src, marking the beginning of modern oncogene research (Stehelin, Varmus et al. 1976, Spector, Varmus et al. 1978, Parker, Varmus et al. 1981).

The human SRC proto-oncogene encodes a protein with non-receptor tyrosine kinase activity that interacts with several important mitogenic and cell cycle control signaling pathways. Oncogene activation of v-src occurs when mutations in tyrosine 527 or 416 inactivate the regulatory SH3 region of the protein, resulting in constitutive activation (Frame 2002). Human SRC has been shown to act as an oncogene in human cancers, however human SRC

oncogenic activity is usually activated by overexpression of the gene as opposed to a functional mutation in the protein (Wang, Fromowitz et al. 1991, Zheng, Wang et al. 1992). Similar to SRC, most oncogenes discovered have coded for proteins involved in some part of a cell signaling pathway promoting cell growth or division.

While oncogene activation is accepted by most of the scientific community as a required step in development of malignant disease, it is necessary but not sufficient. Full malignancy in most if not all cancers occurs only after both a mutation activating at least one oncogene, and an additional inactivating mutation in at least one tumor suppressor gene. Tumor suppressor genes are classified as genes whose normal function is usually regulation of cell growth or proliferation, response to DNA damage, or regulation of apoptosis (programmed cell death). Tumor suppressor genes are defined by the fact that loss of function mutations contribute to malignancy, in contrast to oncogene activation, where mutation constitutively activates gene function and results in deregulated increased gene activity. A classic model for the development of malignant disease through multiple successive mutations in both tumor suppressor genes and proto-oncogenes is the model proposed by Burt Vogelstein et al, which modeled the development of colon cancer through the accumulation of mutations in multiple genes (Fearon, Hamilton et al. 1987, Fearon and Vogelstein 1990, Aguirre-Ghiso 2007).

Focusing on oncogenes, different categories of oncogenes can be classified based on function. As mentioned earlier, most "classical" oncogenes are members of signaling pathways promoting cell survival, growth, metabolism, or proliferation. Organized according to the structure of a cell (from exterior to interior), oncogenes could be classified into growth factors, receptor tyrosine kinases, non-receptor tyrosine kinases (intracellular), non-receptor serine/threonine kinases and associated co-regulators (intracellular), GTPases, and transcription factors (Cannistra 1990).

It is important to the subject of this thesis to note that the canonical classification of oncogene functional groups does not include epigenetic regulators, genes that regulate chromatin structure and thus regulation of gene expression through accessibility of transcription factors to their target gene locus. Until relatively recently, chromatin modifiers were not considered to be oncogenic, as no driving oncogenes with epigenetic regulatory activity had been reported. While the mechanism of chromatin modifying enzymes and their role as oncogenes will be described and discussed elsewhere in this document, it is important to note that oncogenic chromatin modifiers, sometimes termed "oncomodifiers", represent a different class of oncogenes with biology distinct from the majority of well-characterized oncogenes.

1.3 Breast Cancer Oncogenomics

While all cancers share generally similar patterns of mutations in specific classes of proto-oncogenes, the specific genes most often mutated and the mechanism of gene mutation are often specific to cancer types and sub-types. Recent studies of larger numbers of patient tumor samples from different cancer types, such as The Cancer Genome Atlas (TCGA) study, have provided a detailed ontology of mutations in different cancer types, with breast cancer benefitting from the collection of a large number of patient samples. The number of breast cancer samples collected exceeds 1000 samples, more than almost any other cancer type, providing unprecedented views into the nature of mutations in breast cancer.

The 1105 breast cancer samples in the TCGA study portray gene amplification events as the second-most common mechanism of oncogene activation in breast cancers, second only to PIK3CA mutations at 31.3%, with amplifications in chromosomes 8 and 11 occurring most frequently. Notably, the most common point mutations found in breast cancers are loss-of-function (LOF) mutations in the TP53 gene (31.3%). A study by Ciriello et all on datasets collected by The Cancer Genome Atlas (TCGA) have shown that more than 75% of functional mutations found in breast cancer are the result of copy number alterations, and less than 15% of driving mutations are caused by point mutations (Ciriello, Miller et al. 2013).

1.4 Epigenetic Regulation

Epigenetic regulation from a molecular and cell biological perspective can be defined as the set of structural attributes of chromatin that alter gene function without altering DNA sequence. Epigenetic changes are commonly capable of persisting across mitotic and meiotic events, although in modern definitions of epigenetic regulation, heritability of an epigenetic mark is not necessarily a defining characteristic (Waddington 2012, Felsenfeld 2014, Deans and Maggert 2015). Unlike genetic changes, which are defined as changes in DNA sequence, epigenetic changes manifest as changes in the structure of chromatin in the form of covalent modifications to DNA or histones (Berger, Kouzarides et al. 2009). There are three main classes of epigenetic modification: DNA methylation, modulation of transcript availability by RNA, and histone modification. Both DNA methylation and histone modifications affect gene function by altering the physical accessibility of the DNA sequence coding for a gene, either allowing or preventing transcriptional machinery to bind and initiate transcription. RNA-induced silencing acts between transcription and translation by either promoting degradation of the nascent mRNA transcript or sequestration of the transcript in a compartment where it cannot be translated into protein. While epigenetic regulation is most commonly thought of in terms of transcriptional regulation, DNA methylation and certain histone modifications also affect the rate of somatic mutation, which sometimes leads to genetic changes in gene expression or function, including oncogene activation or tumor suppressor inactivation.

1.4.1 DNA Methylation

DNA methylation is defined as the covalent addition of a methyl group (-CH3) to a cytosine base in a molecule of DNA. Canonical DNA methylation in humans occurs at so-called CpG sites, a term for a cytosine immediately followed by a guanine in the 3' direction (CpG is used to avoid confusion with CG, which when written without the “p" refers to pairing of cytosine and guanine on complementary strands of DNA). Although most methylated cytosines in humans have a guanine base on their 3' side, there are other di-nucleotide and tri-nucleotide patterns that are subject to methylation in humans (and many other configurations in other organisms). Addition of a methyl group to cytosine converts the nucleotide to 5-methyl cytosine, methylating the DNA at that base. DNA methylation is generally a repressive mark, inhibiting transcription. CpG methylation of DNA has been shown to inhibit transcription by physically preventing binding of transcriptional enzymes (RNA polymerase and others) to the DNA, or by recruiting proteins that recognize and bind CpG-methylated DNA, recruiting other chromatin-modifying enzymes such as histone deacetylases and histone methyltransferases, resulting in heterochromatic structure at the site of CpG methylation (Bird 2002). DNA methylation is most often found in gene promoter sequences in humans, and is one of the primary mechanisms of gene silencing in cells (Jones 2012). The promoters of most cell-specific genes that are responsible for the structure and function of other cell types are found to be heavily DNA methylated in differentiated cells (Riggs 1975). In particular, the

promoters of master regulators of pluripotency, namely Oct4 (POU5F1), SOX2, and KLF4 among others, are highly methylated and repressed in differentiated cells. Yamanaka, Jaenisch, and others discovered induced pluripotency in adult cells by overexpressing exogenous Oct4 (POU5F1), Sox2, Klf4, and Myc in differentiated cells (Takahashi and Yamanaka 2006, Wernig, Meissner et al. 2007). Exogenous expression of these proteins induced a pluripotent stem cell (iPS) phenotype by removing DNA methylation at the promoters of the endogenous POU5F1 gene, resulting in the maintenance of an induced pluripotent phenotype, persisting after exogenous Oct4 expression was curtailed (and assuming no differentiation signals are present) (Maherali, Sridharan et al. 2007). This was a remarkable discovery, as induced pluripotent stem cells appear to be indistinguishable from embryonic stem cells, and iPS cells can in fact be used to create chimeric organisms that develop normally. The fact that modulation of an epigenetic mark, in this case DNA methylation, can result in a change in phenotype to pluripotent stem cells, which can then differentiate into any phenotype, shows the incredible power of epigenetic regulation of gene expression. While a cell signaling response to stimuli in a cell's environment is capable of making diverse changes to gene expression, epigenetic regulation can render entire regions of the genome more or less accessible to transcriptional regulation. To use a crude analogy: if cell signaling represents the tactical playbook of a cell to its environment, then epigenetic regulation may represent the long-term strategy of the cell, operating within the constraints of the

rulebook spelled out by the genotype. In the case of DNA methylation, this analogy appears to largely hold; DNA methylation patterns in normal adult cells remain mostly constant in terminally differentiated cells, and induced pluripotency, as well as malignancy, are exceptions that prove the rule.

1.4.2 DNA methylation in cancer

Human cancers exhibit altered patterns of DNA methylation. Globally (at genome scale), most cancer cells are hypomethylated (decreased levels of DNA methylation), while specific gene promoter regions are hypermethylated (increased DNA methylation) (Feinberg and Vogelstein 1983). In mammalian cells, CpGs often occur in clustered groups of a few hundred to a few thousand bases, and have been named CpG-islands. The incidence of CpG dinucleotides in the human genome is ~1.7% of all nucleotides (based on the UCSC hg19 genome assembly), thus potential CpG methylation sites make up a small proportion of dinucleotide sequences in the human genome.

Despite the small number of CpGs genome-wide, the distribution of CpGs is not random; CpGs are enriched proximal to promoter regions in humans. Approximately 70% of human gene promoter regions contain CpG islands, enabling the prediction of unknown gene promoters based on local CpG content (Bird 1986).

In many cancers, promoters of genes that play a role in suppressing tumor growth and progression are found to be hypermethylated, resulting in suppression of gene expression. Methylation of DNA cytosines is catalyzed by a

family of proteins called DNA methyltransferases. The three known human DNA methyltransferases are DNMT1, DNMT3A, and DNMT3B (Meilinger, Fellinger et al. 2009). Inactivating mutations in DNMT3A have been found to promote oncogenesis, characterizing DNMT3A as a tumor suppressor.

1.4.3 Histone Modification

Histones are globular protein complexes that are found in the nuclei of eukaryotic cells (Van Holde, Sahasrabuddhe et al. 1974). Histones are normally found tightly associated with DNA: the strand of DNA composing each chromosome is wound around histones in a beads-on-a-string formation, with each histone surrounded by 147 nucleotides, and 50 nucleotides connecting each histone-DNA complex, which is called a nucleosome (Schalch, Duda et al. 2005). The histone core is composed of 8 protein subunits: two each of H2A, H2B, H3, and H4 (Luger, Mader et al. 1997). The histone H1 subunit is not part of the globular core, but instead binds to the core proteins at either end, while holding the incoming and outgoing ends of the DNA strand against the globular core (Figure 0 [renumber]). Histones have a net basic charge, which aids in binding acidic DNA to the histone (Rohs, West et al. 2009). The wrapping of DNA around histones efficiently reduces the length and volume of the DNA in each eukaryotic cell, compacting the chromosomes to various degrees depending on the stage of the cell cycle. DNA that is tightly associated with histone is called heterochromatin, after the appearance of alternating dark and light bands in micrograph images when stained, while loosely associated

chromatin is called euchromatin. While the role histones play in compacting and organizing chromatin was apparent from simple observation, the complexity of the biology of histones and chromatin was not discovered until the advent of modern molecular biological methods. Since their discovery, it has been shown that histones are not simply passive scaffolding for DNA, but that the state of chromatin structure affects gene expression (Kayne, Kim et al. 1988). It was hypothesized early on that tightly packed chromatin was inaccessible to transcriptional machinery, meaning that heterochromatic DNA was not capable of being transcribed. The inactivation of one X chromosome in female cells is a prime example of this phenomenon: one of the two X chromosomes in each female cell is condensed tightly into heterochromatin, called a Barr body. The heterochromatic state of the second X chromosome prevents expression of the genes on the inactivated chromosome, which maintains the normal levels of proteins encoded by these genes, ensuring normal cell function. Barr bodies were discovered in 1949, prior to modern molecular biology techniques, yet Murray Barr correctly predicted that the function of Barr body formation was to prevent gene expression from more than one X chromosome per cell (Barr and Bertram 1949). As the field of molecular biology advanced, a virtuous cycle of structural and mechanistic discovery began, which has led to our current understanding of the process of histone modification, and the remarkable observation that covalent modification of specific amino acids just a few residues

apart on the histone core polypeptide can have opposing effects on transcription of gene loci proximal to the histone so modified.

1.4.4 Histone Variants

In addition to covalent modification of histones with different functional groups, alternate variants of the histone subunits themselves can be substituted for the canonical subunit variants. These alternative variants of the core histone subunits have slightly different amino acid content, and are substituted into specific regions of chromatin, where they alter chromatin conformation and function. The histone H3 variant H3.3 has been found to be present in chromatin containing gene loci that are actively transcribed, but more recently it was found at the pericentric and telomeric heterochromatin regions, representing silenced chromatin (Drane, Ouararhni et al. 2010). CENPA or cenH3 is a histone H3 variant found in the cetromeric regions of chromosomes, and plays a critical role in kinetochore formation, specifically the binding of spindle fibers to chromatin for chromatid migration during anaphase (Howman, Fowler et al. 2000).

1.4.5 Histone Tail Modification

Histone modification is the term given to a class of reactions that catalyze the post-translational covalent addition or removal of small molecules to one of the histone core protein subunits, H2A, H2B, H3, or H4. The majority of these covalent modifications target the amino-terminal “tail" of amino acids that protrudes from the tertiary structure of the core histone subunits, however amino acids in the globular region of the core histone subunits can also be covalently

modified. The first approximately 70 amino acids from the amino-terminal end of each histone subunit form the “tail" region, a region with little or no secondary or tertiary structure that protrudes from the globular region of the histone complex. Because the tail regions of each histone subunit extend beyond the globular histone structure, they are free to interact with neighboring nucleosomes, transcription factors, or other chromatin-associated proteins. These tail regions are rich in basic amino acids, particularly lysine and arginine (Vidali, Gershey et al. 1968).

1.4.6 Histone Acetylation

In contrast to the indirect and relatively complex action of histone methyl marks, histone acetylation alters chromatin state in a more direct way: addition of acetyl groups to histone core tail lysines neutralizes the positive charge that attracts the negatively charged DNA backbone, resulting in decreased histone/DNA affinity, and ultimately dissociation of DNA from the histone to form euchromatin. The effect of histone acetylation on gene transcription was discovered by Vincent Allfrey in 1964 (Allfrey, Faulkner et al. 1964), however the mechanism for addition or removal of acetyl groups to histones was not known, nor was it known how the process was regulated. Histone acetylation was considered to be a general feature that resulted in euchromatin, while absence of acetylated histones resulted in heterochromatin. The first indication that a class of enzymes existed that catalyzed addition of acetyl groups to histones at specific amino acid residues started to become apparent in the early 90s, when C. David

Allis et al (early pioneers of chromatin immunoprecipitation) and others successfully characterized histone acetyltransferases (HATs) that targeted acetylation of specific lysine residues on specific histone subunits, as well as lysines on some other proteins, such as transcription factors (Sobel, Cook et al. 1994, Brownell, Zhou et al. 1996, Vettese-Dadey, Grant et al. 1996). HATs catalyze the transfer of an acetyl group from Acetyl-coenzyme A (Ac-CoA) to a specific lysine on the histone subunit. Around the same time, histone deactylases (HDACs), the class of enzymes that catalyze the removal of acetyl groups from histones, were first characterized, including Rpd3 in yeast and HDAC1 in humans (Carmen, Rundlett et al. 1996, Taunton, Hassig et al. 1996). There are several HAT and HDAC classes, each of which contains several genes with relatively high sequence and structure homology, particularly in their catalytic domains.

1.4.7 Histone Acetyltransferases (HATs)

The Gcn5-related family of histone methyltransferases (GNATs) includes the human genes PCAF, HAT1, ATF-2, and KAT2A. The family is named after the first histone acetylase to be characterized, Gcn5, a yeast HAT first described by Brownell et al in 1996. Human homologues include PCAF and HAT1. Proteins in this family contain a bromodomain that recognizes other acetylated proteins, as well as a catalytic HAT domain conserved across eukaryotes from yeast to human (Brownell, Zhou et al. 1996).

The MYST family of HATs includes MOZ, MOF, MORF, HBO1, and KAT5 in humans. The name “MYST" originates from the names of four of the early genes found in this family: MOZ, Ybf2 (later known as Sas3), Sas2, and Tip60 (now known as KAT5). Structurally, MYST-family proteins usually contain zinc-finger domains and chromodomains. MYST family proteins all share a motif in their HAT domain that is common to the GNAT family. The p300/CBP family consists of just two members, p300 and CBP.

1.4.8 Histone Phosphorylation

Like the phosphorylation of other proteins such as receptors and second messengers involved in cell signaling, histones are also phosphorylated, leading to the binding and activation of downstream effectors that promote various effects on chromatin conformation. Phosphorylation of histone residues is often accompanied by crosstalk with other histone modifications, which cooperate to alter chromatin conformation. Serine, threonine, and tyrosine are all phosphorylation substrates on histone subunits.

The most well-characterized histone phosphorylation mark is phosphorylation of S139 the H2AX histone subunit in response to DNA damage, which results in recruitment of histone acetylases to promote relaxation of the chromatin surrounding the break (presumably to allow DNA repair machinery access to the DNA damage), and initiation of the DNA damage response checkpoint, which halts the cell in G1/S so that DNA repair can be completed. Phosphorylation of S139 on H2AX histones is catalyzed by the kinases ATM and

ATR, two well-known components of the DNA damage response pathway. Subsequent to phosphorylation of H2AX, which spans a multi-megabase region surrounding the DNA damage, the H4 subunits of histones in this region become acetylated, which leads to promotion of a euchromatic state in the chromatin proximal to the damage (Li, Li et al. 2011). Transition to euchromatin around the damaged DNA is probably necessary to allow DNA repair protein complexes to bind and repair the damage, and is not possible without this coordinated epigenetic transition in response to DNA damage (Goto, Tomono et al. 1999, Wei, Yu et al. 1999).

Histone phosphorylation has also been shown to regulate transcription, particularly promotion of transcription. Several serines and tyrosines on histone subunits H2B and H3 have been found to be associated with changes in transcriptional regulation of several genes, particularly proliferation-associated proto-oncogenes such as c-myc. Among other mechanisms, phosphorylation of H2B and H3 will recruit histone acetyltransferases to the histone, which promotes euchromatin formation and generally an increase in transcription of proximal loci.

Phosphorylation of H3 can also promote demethylation of repressive histone marks concomitantly with promotion of acetylation. Phosphorylation of Threonines 6 and 11 on histone H3 recruits GASC1, a histone demethylase, to the histone, where it catalyzes the removal of a transcriptionally repressive H3K9 methyl mark. Phosphorylation of H3T6 and H3T11 is triggered by androgen receptor (AR) binding to the region, and the resulting histone phosphorylation

and demethylation is part of the mechanism of AR-dependent transcription (Metzger, Yin et al. 2008).

1.4.9 Histone Ubiquitination

Histone ubiquitination was discovered by Goldknopf in 1975, well before some other histone modifications, yet it remains one of the least characterized histone modifications in terms of function and genome-wide patterns (Busch and Goldknopf 1981). Histone ubiquitination state is linked to the levels of H3K4 and H3K36 methylation on a nucleosome. Interestingly, both serial addition and serial removal of a monoubiquitin residue to or from a histone can both promote activation of transcription, unlike other histone modifications where addition or removal of a particular mark usually have opposing effects on chromatin conformation (Zhang 2003, Weake and Workman 2008).

1.5 Histone Methylation

Histone methylation refers to the class of histone modifications that describe the substitution of a methyl group ($-CH_3$) for one, two, or all three of the hydrogens in the amino ($-NH_3$) group at the end of the R-group of lysine or arginine residues. These modifications are called mono-, di-, or tri-methylation, respectively. The methyl group added to the histone residue is sourced from S-adenosyl methionine (SAM). Enzymes that catalyze the transfer of a methyl group from SAM to lysines or arginines are known as histone methyltransferases (HMT) or protein arginine methyltransferases (PRMT), respectively. Functionally, methylation of histone subunit lysines and arginines regulates chromatin

structure, transcriptional regulation, and DNA repair. While some histone methylation marks remain stable through cell divisions and even from one generation to the next, the process of histone methylation is actually quite dynamic.

Unlike other covalent modifications of histone residues, histone methylation does not alter the net charge on the histone, and so does not appear to regulate chromatin structure or gene transcription through direct changes in histone/DNA affinity. Instead, histone methylation represents what has been proposed by some to be a “histone code", a pattern of methylation marks that is read by large multi-protein complexes, which scan the methylation state of a group of histones and recruit, or block, transcriptional machinery, modifiers of chromatin structure, or other histone modifying complexes(Barski, Cuddapah et al. 2007). A model has been created to illustrate this concept, sometimes called the “read-write-erase" model. This model classifies HMTs and PRMTs as “writers", proteins which apply methyl marks to specific residues on histones. Histone modifications may then be read by histone ”readers”: proteins which are part of large multi-protein complexes that will catalyze additional histone modifications based on the existing pattern of histone modifications. HMTs and PRMTs are in fact members of such “reader" complexes, and the proteins themselves often have specific protein domains that recognize protein-DNA complexes, as well as specific covalent modifications on histones. This can result

in a positive or negative feedback loop that maintains high or low levels of histone methylation at specific regions of the genome.

There are two main classes of lysine histone methyltransferases: one class is named after the catalytic domain, named after 3 genes in model organisms where this class of HMTs was first discovered, Suppressor of Variegation, Enhancer of Zeste, Trithorax (SET). The other class encompasses all histone methyltransferases that have a non-SET catalytic domain. Methylation of specific residues only a few amino acids apart on the histone tail can have opposing effects on chromatin conformation or transcription: some promote transcription, while others inhibit.

Known substrates for methylation on histones include lysines 4, 9, 27, 36, and 79 and arginines 8 and 17 on H3; and lysines 20 and 59, and arginine 3 on H4. Hypermethylation of lysines 4, 36, and 79 on histone H3 are associated with actively transcribed chromatin in the euchromatic state, while lysines 9 and 27 on histone H3 and lysine 20 on H4 are associated with transcriptional repression.

Known arginine methylation substrates include arginines 8 and 17 on histone H3, and arginine 3 on histone H4. Histone H1 is also methylated on lysine 26. Below these histone methylation marks, the enzymes catalyzing their addition and removal, and their biological function will be described in detail.

1.5.1 Histone lysine demethylases

Histone demethylases represent of class of proteins that remove a methyl group from lysine residues, often on histones, but also from lysines of other

methylated proteins, both nuclear and cytosolic. Known histone demethylases are classified by their catalytic domains and required cofactors, and are organized into two families, the LSD family and the Jumonji family. The LSD family of lysine demethylases contains two members, LSD1 and LSD2. LSD family proteins contain an amine oxidase-like (AOL) domain, and a SWIRM domain (Hojfeldt, Agger et al. 2013). Flavin adenine dinucleotide (FAD) is reduced to $FADH_2$, removing hydrogen from the methyl group, and then the carbon leaves the nitrogen on the lysine and combines with water to form formaldehyde. Because of the mechanism of lysine demethylation exhibited by LSD family proteins, they are not capable of demethylating tri-methylated lysine, but can demethylate di- and mono-methylated lysine.

The second group of histone demethylases belongs to the Jumonji C (JMJC) family. This family of histone demethylases contain a Jumonji C catalytic domain which binds Fe^{2+} and α–ketoglutarate cofactors. The mechanism of JMJC lysine demethylases allows them to demethylate tri-methylated as well as di-methylated and mono-methylated lysine.

1.5.2 Histone H3 methylation on lysine 4

Lysine 4 is the most distal lysine on the histone H3 tail with evidence of methylation. A notation for communicating the level of methylation at a specific residue on histones: the histone subunit is noted, followed by the lysine number, and finally the level of methylation on that lysine. For example, mono-methylation of lysine 4 on the histone H3 subunit would be written as H3K4me1, and di-

methylation and tri-methylation written as H3K4me2 and H3K4me3, respectively. H3K4 methylation is most often found on or near actively transcribed promoter regions, and mostly in the tri-methylated form, however H3K4me2 and H3K4me1 marks are also commonly found proximal to active promoters. In many cell types, such as pluripotent stem cells in the process of differentiation during development, approximately 80% of actively transcribed promoters are marked by H3K4me3. H3K4me3 marks are found almost exclusively near the promoter regions of actively transcribed genes.

While there are some reports of H3K4me2 marks playing a role in regulating de-methylation of DNA CpGs, the majority of H3K4me2 appears to co-localize with H3K4me3 near promoters. H3K4me1, in contrast, appears to be a mark for distal enhancers, as it is often found marking distal enhancer regions for many genes, and is more evenly distributed genome-wide than either H3K4me2 or H3K4me3.

Histone methyltransferases that catalyze addition of methyl groups to H3K4 include several proteins. Mixed Lymphocyte Leukemia 1 (MLL1), also known as lysine methyltransferase 2A (KMT2A), is a key histone methyltransferase that specifically tri-methylates lysine 4 on histone H3 (Ruthenburg, Allis et al.).

MLL1 co-localizes with RNA polymerase II (Pol II) on the promoter regions of genes, the genomic region where almost all H3K4me3 marks are found. MLL1 is part of a multi-protein complex called the MLL complex, which contains several histone-modifying enzymes in addition to MLL1. The MLL complex is involved in

the differentiation of hematopoietic stem cells to form lymphocytes. Mutations in MLL1 through chromosomal translocations result in activation of MLL1 into an oncogene in acute lymphoid and acute myelogenous leukemias. MLL1 oncogenic activity is at least partially due to dysregulation of the HOXA family of genes, which, when prevented from tapering off their expression during hematopoesis, inhibit terminal differentiation of the lymphoid progenitor cells into mature white blood cells. MLL1 is paralogous to the Set1 gene in yeast.

ASH1L (absent-small-homeotic-like-1) is another gene shown to methylate H3K4 residues, although H3K36 has also been reported as a substrate (Zhu, Li et al. 2016). ASH1L contains a SET domain, and is part of the SET family of HMTs. ASH1L also contains a PHD domain (Plant HomeoDomain) as well as a bromodomain, two domains common in proteins that recognize and bind chromatin. ASH1L is also involved in positively regulating HOX genes during development (Schuettengruber, Martinez et al. 2011).

1.5.3 H3K9 methylation

Hypermethylation of lysine 9 on histone H3 is generally associated with repression of transcription, however various levels of H3K9 methylation are found in both repressed and actively transcribed chromatin, and mono-, di-, and tri-methylated H3K9 have been found to have diverse binding patterns and biological functions. H3K9me2 is reported to be mutually exclusive of H3K4 methylation, and several protein complexes that contain H3K9

methyltransferases have been reported to be in complex with histone demethylases that target H3K4 methylation (Shi, Sun et al. 2011).

Genome-wide, regions of H3K9 methylation are often much broader than other histone methylation marks, with some H3K9me2 regions spanning megabases of sequence in one contiguous region. H3K9me1 marks are often found localized to the promoter regions of actively transcribed genes; H3K9me2 marks are often spread across broad regions. Constitutive heterochromatin, such as the chromatin found in inactive X chromosomes in female cells, possesses H3K9me2 marks on the majority histones. It has been shown that a region of H3K9me2 marks upstream of the Xist gene, the gene responsible for initiation of X-inactivation, serves as the starting point for the spread of H3K9me2 marks across the entire X chromosome that is in the process of inactivation. The spread of H3K9me2 marks, at least in the process of X-inactivation, is thought to be due to the recognition of existing H3K9me2 marks by the G9a-GLP-Wiz complex, causing a positive feedback loop that causes H3K9me2 to spread. Such auto-recruitment of HMTs appears to be more general than just H3K9 methylation.

The primary HMT for monomethylation and dimethylation of lysine 9 on histone H3 is actually a heterodimer of the G9a and GLP proteins, two SET-domain containing proteins that can form homo-dimers, but are found in cells almost exclusively as G9a-GLP heterodimers (Tachibana, Sugimoto et al. 2001, Shankar, Bahirvani et al. 2013). Both G9a and GLP are members of the Suv39h family of proteins (Suv stands for Suppressor of Variegation, one of the three

protein families referred to in the name of the SET domain). G9a and GLP have very high homology, and have similar structure, differing only in the lengths of some domains. Mutant knockout experiments have shown that inactivation of the SET domain of G9a reduces histone H3K9me1 and H3K9me2 levels far more than mutation of the GLP SET-domain. G9a-GLP is joined by a third protein, Wiz, a zinc-finger protein that binds the SET domains of both G9a and GLP. Experiments by Ueda et all have shown that G9a-GLP-Wiz forms a more stable complex than either homodimer + Wiz complex, which may explain why the G9a-GLP-Wiz trimeric complex is the only configuration found in vivo.

As for other histone lysines, H3K9 methylation is a substrate for demethylation by histone lysine demethylases. One particular H3K9me3 demethylase known to be involved in cancer is the GASC1 gene (KDM4C), which is amplified and overexpressed in breast cancers (Liu, Bollig-Fischer et al. 2009).

1.5.4 H3K27 methylation

H3K27 hypermethylation is usually a strong mark of repressed chromatin, and is usually localized to the promoter regions of genes. In developing progenitor and stem cells, it has been found that the promoters of several genes important in differentiation are bound by histones with both H3K4me3 and H3K27me3 marks. The presence of both of these marks with opposing functions (H3K4me3 is activating, H3K27me3 is repressive) on histones near a promoter are called bivalent chromatin marks, and are found proximal to the promoters of

genes that are not expressed, but will soon be expressed once cell differentiation begins (Bernstein, Mikkelsen et al. 2006, Ku, Koche et al. 2008).

The hypothesis for bivalent chromatin marks is that, following the signal from a differentiation factor, a histone lysine demethylase will remove the H3K27me3 marks on the promoters, signaling activation of transcription due to the already present H3K4me3 marks on these promoters, allowing more rapid activation of transcription of these genes. While it is well established that these marks do occur at the promoters of several genes involved in cell differentiation, there is still some uncertainty about the role these bivalent domains play in the biology of differentiation, and whether H3K27 demethylation at bivalent chromatin marks actually result in rapid activation of these genes.

1.5.5 H3K79 methylation

Methylation of lysine 79 on histone H3 is catalyzed by a distinct class of histone methyltransferases known as the DOT1-like (DOT1L) family. The DOT1L family of HMTs do not contain a SET domain, but do share a different catalytic domain. DOT1L is the only known HMT that has lysine 79 on histone H3 as a substrate. DOT1L is capable of adding methyl groups to non-methylated, mono-methylated, and di-methylated H3K79 (Vlaming and van Leeuwen 2016). Interestingly, di- and tri-methylation of H3K79 is not processive, meaning that DOT1L can transfer between one and three methyl groups to lysine 79. As such, it does not require independent step-wise additions of single methyl groups in separate reactions. H3K79 methylation was discovered in S. cerevisiae when the

deletion of the Dot1 gene in yeast resulted in the complete loss of H3K79 methylation, which disrupted the normal silencing of telomeric genes (thus the name: Disruptor Of Telomere-1). DOT1L is the human homologue of Dot1 in yeast.

The discovery of Dot1 function is an interesting study in the complexity of histone modifications and their biology. While the function of Dot1 is to methylate lysine 79 on histone H3, which results in activation of transcription, deletion of the gene had the opposite effect one would naively expect: transcription of telomeric genes increased when Dot1 was deleted. Paradoxically, overexpression of Dot1 also increased telomeric gene expression. The hypothesis for this phenomenon is that methylation of K79 on non-telomeric histones by Dot1 maintained a small region at the telomere of the chromosome, which would be repressed by other histone modifications. High levels of Dot1 overexpression could inhibit the repression of telomeric gene expression, but total abrogation of Dot1 expression by gene deletion has a similar effect. When deleted, the lack of Dot1 activity and the resulting abrogation of all H3K79 methylation results in too many potential substrates for the telomeric repression machinery, diluting its activity, and resulting in de-repression of telomeric genes. While this phenomenon of both extremes of expression of a gene having similar biological effects may be a rare example, considering it is the sole HMT for lysine 79 on H3, Dot1/DOT1L serves as an important example of the complexity and counter-intuitive nature of the biology of histone modifications.

1.5.6 H3K36 methylation

Methylation of lysine 36 on histone H3 is less understood than many other histone methyl marks, both in terms of H3K36 methylation patterns and characterization of the identity and activity of histone methyltransferases that target H3K36 as a substrate. H3K36 methylation has been reported as a mark of both transcriptional activation and repression for chromatin regions where it is found. Recent work has shown that H3K36 methylation generally occurs in broad regions spanning the bodies of genes, and often increases near the 3' end of the transcribed region of a gene (Wagner and Carpenter 2012). H3K36 methylation has also been shown to promote elongation of the nascent transcript as the transcriptional machinery leaves the promoter region immediately after initiation of transcription, and may also act as a marker for recently transcribed genes, preventing immediate re-association of the Pol II complex and initiation of a second round of transcription before the current transcript is complete (Bell, Wirbelauer et al. 2007, Berger 2007).

1.6 NSD family of histone lysine methyltransferases

NSD1 is the prototype for the nuclear SET domain-containing (NSD) family of histone lysine methyltransferases (HKMTs) named for their localization to the nucleus and their shared catalytic SET domain (Jaju, Fidler et al. 2001). There are three members, NSD1, WHSC1, and WHSC1L1, of which NSD1 and

WHSC1L1 are both known to be oncogenic (Kim, Kee et al. 2006, Kim, Kee et al. 2007). The NSD1 gene is located on the long arm of chromosome 5, and codes for three known transcript isoforms. The longest isoform codes for a 2696 amino-acid protein containing a catalytic SET domain, a post-SET domain, and several PHD, zinc finger, and PWWP chromatin interaction domains(Vougiouklakis).

Mutations in NSD1 are associated with Sotos syndrome, a genetic overgrowth disease that causes overgrowth in children between 1 and 2 years old. The disease is caused by a decrease in the levels of NSD1 protein due to loss-of-function mutations in one copy of the NSD1 gene (Kurotaki, Imaizumi et al. 2002, Douglas, Hanks et al. 2003, Wernig, Meissner et al. 2007). Translocations with breakpoints within the NSD1 gene are found in some childhood leukemias, and NSD1 has been shown to be a driving oncogene in these cancers. NSD1 sometimes fuses with NUP98, resulting in constitutive activation of the histone methyltransferase activity of NSD1 (Jaju, Fidler et al. 2001). Lysine 36 on the histone H3 subunit has been reported as the primary substrate for NSD1 methylation. NSD1 has also been shown to interact with the androgen receptor (AR), specifically the ligand binding domain and the DNA-interaction domain of AR.

WHSC1 is the second member of the NSD family, and is located on the short arm of chromosome 4. Loss of function mutations (usually small deletions) in WHSC1 cause Wolf-Hirschhorn Syndrome, a disease characterized by microcephaly, mental deficiency, low muscle tone, and other defects in growth,

presenting at birth. Several of the symptoms of Wolf-Hirschhorn Syndrome are opposites of the symptoms observed in Sotos syndrome, a disease caused by loss of function of NSD1 (Douglas, Hanks et al. 2003, Douglas, Coleman et al. 2005). WHSC1 has high homology with WHSC1L1, with both proteins containing many of the same PWWP and PHD domains, the catalytic SET domain followed by post-SET domain. WHSC1 also contains as AWS domain, which WHSC1L1 lacks. The primary substrate of WHSC1 has been reported to be lysine 36 on histone H3, where it adds a tri-methyl mark (Nimura, Ura et al. 2009).

While there are six known isoforms coded for by WHSC1, they fall into two main classes. The long isoforms encode for a 1365 amino-acid protein that contains the catalytic SET domain, while the short isoforms encode for a 627 amino-acid protein that lacks a catalytic SET domain. WHSC1 translocations have been found to be driving oncogenes in multiple myelomas (Huang, Wu et al. 2013, Min, Ezponda et al. 2013).

The Wolf Hirschhorn Syndrome Candidate 1-Like 1 (WHSC1L1) gene is located on the short arm of chromosome 8, in the 8p11-p12 band. WHSC1L1 is the third gene in the NSD family of SET-domain containing histone-lysine methyltransferases (HKMTs), the other members consisting of Nuclear SET domain-containing protein 1 (NSD1) and Wolf Hirschhorn Syndrome Candidate 1 (WHSC1) (Stec, van Ommen et al. 2001). WHSC1L1 has the greatest protein sequence homology with WHSC1, although all three proteins contain similar domains, including a catalytic SET domain, a post-SET domain,

and one or more PWWP and PHD domains (He, Li et al. 2013). The WHSC1L1 gene codes for two known isoforms; the first isoform codes for the full-length 1437 amino-acid protein (NM_023034) containing the catalytic SET domain, and the second isoform (NM_017778) codes for a shorter protein that consists of the first 647 amino-acids of the long isoform.

The short isoform of WHSC1L1 (WHSC1L1short) contains only a single PWWP domain, and lacks the catalytic SET domain, PHD domains, and other PWWP domains (Angrand, Apiou et al. 2001). WHSC1L1 expression is localized to the nucleus (Figure 2.1). While the long isoform of WHSC1L1 is much more rich in functional motifs than the relatively simple short isoform, WHSC1L1-short is the predominantly expressed isoform in most normal breast epithelial samples as well as tumor samples from most breast cancers.

1.7 Epigenetic Dysregulation in Breast Cancer

For several decades following the initial discovery by Bishop and Varmus that the transforming oncogenes found in tumor viruses are actually remnants of human genes that also act as oncogenes when mutated in human cells, most oncogenes discovered belonged to a few functional classes. Common functions of classical oncogenes include receptor tyrosine kinases (RTKs), second-messengers downstream of RTKs, transcription factors, and a few other classes of proteins involved in cell signaling pathways. The discovery that genes encoding histone modifiers and other epigenetic regulators could act as oncogenes is fairly recent, and dramatically increases the potential complexity of

oncogenesis, both as a somatic disease and as a product of germline predisposition. It is important when discussing the oncogenic potential of histone modifiers and other epigenetic regulatory genes that the concept of epigenetics, specifically the heritability of epigenetic marks, is not conflated with the role mutated chromatin-modifying enzymes play in driving cancer through altering the marks on chromatin that confer heritable patterns of gene expression. While the effect of inherited epigenetic marks on the development of cancer has been studied, the subject of this work describes the dysregulation of a histone methyltransferase that acts as an oncogene through overexpression as a result of a somatic event that resulted in gene amplification of the histone methyltransferase WHSC1L1.

Several other genes that code for histone modifying enzymes have also been described as oncogenes when mutated, and these mutations have been somatic. It should be stressed that these are somatic events that then proceed to act as oncogenes through dysregulation of epigenetic marks, and are not inherited epigenetic marks that lead to cancer in the individual, nor are these mutations and their resulting epigenetic changes themselves heritable. For this reason, the term “epigenetic" will be used as sparingly as possible going forward, as the definition of the term epigenetics is more closely aligned with the heritable transfer of phenotypic traits through regulation of gene expression that is not encoded by DNA sequence, the key word being heritable.

1.7.1 WHSC1L1 is amplified in 12-15% of Breast Cancers

Investigation of patterns of amplification on the short arm of chromosome 8 resulted in detailed mapping of copy number changes. Using array comparative genomic hybridization (aCGH) and microarray expression analysis, Gelsi-Boyer et al conducted a detailed investigation of the 8p11-p12 amplicon on a panel of breast cancer tissue samples breast cancer cell lines, identifying four distinct regions of amplification in the 8p11-p12 amplicon (Gelsi-Boyer, Orsetti et al. 2005, Pole, Courtay-Cahen et al. 2006). These regions were labeled A1, A2, A3, and A4; regions were characterized by the common co-amplification of the “core" genes in each region in several samples in the panel. The WHSC1L1 locus was identified as one of 11 core genes in the A2 region of the 8p11-12 amplicon, together with part of ASH2L, STAR, LSM1, BAG4, DDHD2, PPAPDC1B, LETM2, FGFR1, C8orf86, and TACC1.

Bilal et al have found amplification of chromosomes 8 and 17 to be predictive of tumor recurrence and poor outcome in ER+ breast cancers with these amplifications (Bilal, Vassallo et al. 2012).

Previous work by the Ethier lab and several other groups has identified several driving oncogenes in the 8p11-p12 amplicon (Table 1.1), several of which have chromatin-modifying activity. These include, among others, MYST3 (KAT6A), ASH2L, WHSC1L1 (Forozan, Veldman et al. 1999, Forozan, Mahlamaki et al. 2000, Garcia, Pole et al. 2005, Yang, Streicher et al. 2006, Yang, Liu et al. 2010). As mentioned above, a key question this project

attempted to answer is whether there are multiple driving oncogenes within an amplicon, or conversely, one primary driving oncogene surrounded by passenger genes that are present as an artifact of the mechanism of amplification.

gene_symbol	sub-amplicon_cluster
PROSC	A1
DDHD2	A2
RAB11FIP1	A1
LSM1	A2
BRF2	A1
PPAPDC1B	-
WHSC1L1	A2
FGFR1	A2
TM2D2	A2
AP3M2	A4
ASH2L	A1;A2
BAG4	A1;A2
FUT10	A3
EIF4EBP1	A1;A2
LETM2	A2
AGPAT6	A3;A4
POLB	A4
VDAC3	A4
HOOK3	-
TACC1	A2
KAT6A (MYST3)	A4
ERLIN2	A1

Table 1.1: Genes amplified in the 8p11-12 amplicon in breast cancer.

1.7.2 Breast cancer sub-types

Classification of cancer types is fundamental to the detailed description of a specific cancer, and this identification of cancer type provides a foundation upon which almost all cancer research and treatment is built (Song, Merajver et al. 2015).

Before the development of modern molecular biology techniques, the features used to classify a cancer were the organ where the primary tumor was found, followed by the type of tissue constituting the majority of the tumor mass, and lastly the presence or absence of tumor cells in lymphatic tissue and/or metastasis.

Other than blood cancers (which are classified by cell type rather than tissue type), most if not all cancers are described by tissue type and organ or region of origin, e.g. breast carcinoma, lung carcinoma, or osteosarcoma.

While anatomical and histological features were some of the earliest methods used to classify cancers (according to the modern definition of malignant disease), these methods persist as the primary classifiers of cancer today. Despite the distortion wrought by oncogenesis, the origin of the tumor cells remains some of the most important information used to diagnose a cancer, make a prognosis, and determine the appropriate treatment. Indeed, the location of a patient's primary tumor determines which specialist they will see to be treated for their disease: oncologists and basic researchers specialize in one or a

small number of types of cancer, based on anatomical location and histological type.

1.7.3 Histological Subtyping of Breast Cancer

While classification of tumor types by organ and tissue of origin have some prognostic value, depending on the cancer type, the advent of cancer sub-typing by molecular marker detection brought forth a new era in cancer classification. Breast cancer is an excellent example of the development of classifications of distinct sub-types within the more general diagnosis of breast carcinoma. McGuire et al used dextran coated charcoal (DCC) and sucrose gradient centrifugation to quantitatively detect levels of ERa in samples from 64 primary and metastatic breast tumors, showing that expression of ERa varied widely in expression among the samples (McGuire 1973). The authors hypothesized that one of several plausible explanations for the broad range in ERa expression was that some breast cancers did not express ERa (which we now know to be the case). It was already known in sheep that oophorectomy (removal of the ovaries) could slow the growth of some breast tumors, but had little or no effect on others. The observed range in ERa expression in the breast tumor cohort suggested that classification of breast tumors by ERa status might be able to predict response to estrogen ablation therapy.

By 1974, several studies had shown that measurement of ERa in a patient tumor sample was predictive of the patient's response to anti-estrogen therapy. Later, it was determined that hormone receptor status was both predictive of

response to hormone therapies, and prognostic in that patients positive for ERa were more likely to survive in the short term (approximately 15 years) than breast cancer patients that lacked hormone receptor expression. The end result of this progress in molecular sub-typing was the establishment of two main sub-types of breast carcinoma: ER-positive and ER-negative.

Progesterone receptor testing was also developed, and added to ERa testing after it was determined that testing for the progesterone receptor could further sub-divide the ER-positive sub-type (Shyamala, Chou et al. 2002). The resolution of breast cancer sub-typing increased again with the discovery that the human epidermal growth factor receptor (HER2) gene is amplified in ~20% of human breast cancers (King, Kraus et al. 1985) by Stu Aaronson and colleagues, followed by the work by Slamon and others showing a strong correlation between levels of HER2 expression in breast cancers and the aggressiveness of the disease (Slamon, Clark et al. 1987, Ignatoski, Lapointe et al. 1999). This work led to the addition of a third molecular sub-type of breast cancer: HER2-amplified. Importantly, this work to identify the HER2-amplified breast cancer subtype was driven by the concurrent development of a targeted antibody therapy against HER2 activity, trastuzumab (brand name Herceptin), developed by Micheal Shepard and others at Genentech (Lewis, Figari et al. 1993).

The development of HER2 testing brought the number of breast cancer sub-types to three: ER/PR positive, HER2-amplified, and so-called triple-

negative, after the fact that these tumors would test negatively for all the molecular markers.

By testing patient breast tumor samples for these molecular markers, physicians were able to make much more accurate prognoses, and accurately predict response to either hormone therapy or treatment with trastuzumab. Patients with the triple-negative sub-type occupy a difficult corner of the breast cancer classification space: triple-negative breast cancers are particularly aggressive, with a mean 5-year survival rate as low as 22% for patients with metastatic disease, and the molecular profile characteristic of these tumors predicts that they will respond poorly or not at all to hormone therapy or treatment with HER2 inhibitors. However, triple-negative breast cancers are sometimes more sensitive initially to chemotherapy than hormone-receptor positive or HER2-amplified breast cancers, reinforcing the advantage of adding molecular screening to the anatomical and histological diagnostic procedures (Baselga, Norton et al. 1998, Bagaria, Ray et al. 2014).

Confirmation that testing for molecular biomarkers in patient tumors has predictive and prognostic value increased the clinical importance of more finely classifying breast cancers into sub-types, and supported the hypothesis that cancer is a highly heterogeneous collection of related diseases (Sorlie, Perou et al. 2001, Curtis, Shah et al. 2012). The implication of such heterogeneity at the level of protein expression is that increased knowledge of the specific molecular profile of a patient's tumor will increase the predictive and prognostic power of

physicians when treating breast cancer, as well as leading to more focused research that might lead to new therapies, particularly for triple-negative breast cancer patients, and increased understanding of cancer biology. This potential, together with new microarray technology that enabled the measurement of mRNA expression levels of thousands of genes from a single whole mRNA sample, led to the molecular profiling experiments that enable our current classification system of five breast cancer sub-types: luminal A, luminal B, triple-negative (sometimes called basal), HER2-amplified, and claudin-low (Perou, Sorlie et al. 2000, Bertucci, Orsetti et al. 2008, Prat and Perou 2011).

Some tests to determine breast cancer sub-type have already made a great improvement in the quality of care for a patient. The targeted treatments for patients with HER2 amplified breast cancers are a prime example. Before the test for HER2 amplification and the development of the targeted therapy Herceptin, breast cancer patients with HER2 amplifications were suffering poorer outcomes than patients with normal HER2 status due to the increased aggressiveness exhibited by HER2 amplified tumors. With the availability of trastuzumab and the routine testing of breast tumor biopsies or mastectomy tissue for HER2 amplification, these patients now have up to a 35% response rate to trastuzumab, disease-free survival (DFS) increased from 62.2% to 73.7%, and overall survival (OS) increased from 75.2% to 84% (Callahan and Hurvitz 2011, Perez, Romond et al. 2014).

For luminal breast cancer sub-types, IHC-based testing for hormone receptor status has also been an important advance, leading to targeted hormone therapies and sometimes less aggressive treatment of patients due to the knowledge that their tumors would be likely to respond completely to hormone therapy and surgery, saving them from the morbidity associated with traditional chemotherapy. However, the discovery of two major sub-types of hormone positive breast cancer, Luminal A and Luminal B, has made it clear that the probability of response to hormone therapy is lower and the risk of recurrence higher for those patients with disease classified as Luminal B (Sotiriou and Pusztai 2009). This suggests that continuing to develop classification techniques based on gene expression signatures will further advance our understanding of breast cancer, and will likely further sub-divide current breast cancer sub-types into additional meaningful classes, each with their own indications for treatment.

While we are currently in the age of classifying breast cancers based on the aforementioned staging and grading of the tumor, IHC and FISH testing for hormone receptor status and HER2 amplification, and now transcriptome signatures such as the PAM50 breast cancer panel, the next logical step in classification methodology may soon be whole-exome or whole-genome DNA sequencing of both the primary tumor and a patient blood sample to determine all somatic and germline mutations present in the tumor compared to the patient's normal tissue. Costs for next-generation sequencing such as Illumina paired-end short read sequencing or long read sequencing as performed by Pacific

Biosciences are becoming affordable enough to be used to collect genome-wide mutation, copy number alteration, variation, and transcript expression data for every patient tumor biopsied with the intention to diagnose or rule out cancer.

While gene expression signatures such as PAM50 have already increased our knowledge of breast cancer sub-types to the point where we can differentiate between luminal A and luminal B cancers, current methodologies for clinical testing of patients suspected of having breast cancer still do not attempt to detect all mutations or copy number alterations (CNAs) in a patient's tumor. If basic cancer research has shown one thing, it is that every cancer is slightly different. The discovery of oncomodifiers has shown that, together with the unique combination of driving mutations present in a given cancer, inherited or oncogene-driven epigenetic states provide an additional layer of complexity contributing to the cancer phenotype (Al-Hajj, Wicha et al. 2003, Dontu, Abdallah et al. 2003, Dontu, Jackson et al. 2004).

This dissertation will aim to expand on knowledge of transforming properties of the 8p11-12 amplicon in breast cancer, specifically the role of WHSC1L1 amplification and overexpression on the regulation of tumor cell prolifεραtion, target gene expression, and chromatin binding events, both epigenetic marks and transcription factor binding.

CHAPTER 2 KNOCKDOWN OF WHSC1L IN SUM44 CELLS ALTERS CELL PROLIFERATION AND GENE EXPRESSION

2.1 Introduction

Previous work in the Ethier lab and by others has established that WHSC1L1 is amplified and overexpressed as part of an amplified region called the 8p11-p12 amplicon, a 30-50 gene region amplified in 12-15% of breast cancers, and that WHSC1L1 acts as a driving oncogene in these cells. Additionally, other genes in the 8p11-p12 amplicon often found co-amplified with WHSC1L1 include ASH2L and MYST3 (KAT6A), both of which are chromatin modifiers. Among the candidate driving oncogenes present in the 8p11-p12 amplicon is FGFR1, which was initially predicted to be the principle driver in the amplicon, however it has been shown that in many breast cancers with the 8p11-p12 amplicon, FGFR1 is not highly expressed and FGFR1 expression does not appear to be necessary for maintenance of a tumor phenotype in these cells.

To determine which genes were the strongest candidate driver oncogenes in cells with the 8p11-p12 amplicon, Yang and others in the Ethier lab performed a functional screen of individual genes in the amplicon by overexpressing each candidate oncogene in MCF10A cells and testing for a transformed phenotype by soft-agar colony formation, 3D colony morphology, and proliferation. A complementary screen was performed by shRNA knockdown of each candidate oncogene in the SUM44 breast cancer cell line, measuring loss of cell

proliferation and loss of soft-agar colony formation (add cite)[Yang06, Yang10}. In these studies, it was found that WHSC1L1-short was sufficient to induce a transformed phenotype in MCF10A cells when overexpressed by lentiviral transduction.

Tests for induction of a transformed phenotype included colony formation of MCF10A cells in the absence of insulin on 60mm culture plates, as well as soft-agar colony formation, and finally colony morphology when cultured in 3D Matrigel. These experiments clearly indicated that MCF10A-WHSC1L1-short cells formed the largest and highest number of colonies in the absence of insulin, demonstrating growth factor independent growth, as well as forming the largest number of soft-agar colonies, and finally showing obvious disorganized colony morphology in 3D Matrigel. Additionally, in SUM44 cells, two different shRNAs knocking down total WHSC1L1 (both short and long isoforms of WHSC1L1) resulted in significantly reduced proliferation and reduced colony size and number in SUM44. Proliferation was also reduced in one SUM52-shWHSC1L1total cell line, transduced with the same shRNA that had the strongest effect in SUM44. It should be noted that these WHSC1L1 knockdown experiments in SUM44 used pGIPZ shRNA vectors against total WHSC1L1, while the experiments shown here used pLKO lentiviral shRNA vectors purchased from Sigma.

As WHSC1L1 was the strongest candidate driver oncogene of the 8p11-p12 amplified and overexpressed genes tested, it was a suitable target for further

investigation. The initial objective was to develop a model of WHSC1L1 expression loss in SUM44 cells, and measure the effect of either specific shRNA knockdown of WHSC1L1-short, or knockdown of total WHSC1L1 on cells proliferation and the expression profile of the transcriptome in the WH-knockdown cells vs. a control condition (a vector containing shRNA for LacZ, a bacterial transcript that should have no effect on any human transcript). Use of off-the-shelf vectors enabled expeditious screening for the best knockdown among several shRNAs and selection of the best vectors for knockdown of total WHSC1L1 or WHSC1L1-short.

The first set of experiments was designed to establish stable knockdown of either both the long and short isoforms of WHSC1L1, or the short isoform only, in SUM44 cells. The use of the 3'UTR-specific WHSC1L1-short shRNAs was key to achieving specific knockdown, as the coding sequence of WHSC1L1-short is identical to the long isoform until the splicing of an alternate exon 10, which contains the short isoform-specific 3' UTR region targeted by the WHSC1L1-short shRNAs. Once stable knockdown could be confirmed, the objectives were to study the effect of knockdown of WHSC1L1 on cell proliferation and on gene expression, specifically the effect of knockdown of WHSC1L1-short alone, to determine whether we would observe similar effects to previous experiments with total WHSC1L1 knockdown when only the short isoform was depleted.

2.2 Results

2.2.1 WHSC1L1-short is the major expressed isoform in breast epithelial cells

As previously mentioned, WHSC1L1 is known to have histone methyltransferase activity targeting histone H3 lysines as substrate. To support our hypothesis that overexpression of WHSC1L1-short alters histone methylation patterns in key genes that leads to promotion and maintenance of an oncogenic phenotype, we first sought to identify basic properties of WHSC1L1, such as the cellular localization of the protein isoforms, and the relative abundance of the long and short isoforms at the transcript and protein level. We would expect WHSC1L1 to be localized to the nucleus if it is in fact binding histone proteins as a substrate, and, while little can be directly interpreted from relative isoforms abundance, it is important to determine whether each isoform is actually expressed at significant levels.

To determine the subcellular localization of WHSC1L1 protein in breast epithelial cells, we cultured primary human mammary epithelial cells (HMEC) immortalized with viral E6/E7 proteins and MCF10A cells, both of which were lentivirally transduced with the pLenti6.3-WHshortV5 construct, then exposed to polyclonal WHSC1L1 antibody followed by FITC-conjugated secondary antibody. The nuclei of both HMEC-E6/E7-pL6.3-WHshort and MCF10A-pL6.3-WHshort cells exhibited strong green fluorescence in the nuclei of many cells (nuclei stained with DAPI), while cytoplasmic green fluorescence is not detected (Figure

2.1). This indicates that WHSC1L1 is expressed in the nucleus of these models of breast epithelial cells in culture.

To find the relative expression of WHSC1L1 transcript for the short and long isoforms of WHSC1L1 in breast tissue, RNAseq whole RNA expression data from The Cancer Genome Atlas Breast Cancer dataset was used to determine the expression levels of WHSC1L1-short and WHSC1L1-long. A boxplot of the distribution of expression levels for each isoform for all TCGA BrCa samples (Figure 2.2a) shows that the mean expression of the short isoform of WHSC1L1 is higher than the mean expression of the long isoform. Immunobloting for WHSC1L1 protein in SUM44 luminal B breast cancer cells and MCF10A breast epithelial cells (Figure 2.2b) indicates that WHSC1L1 protein expression in these cell lines is concordant with observed WHSC1L1 mRNA expression levels in TCGA breast cancer samples. Of note, WHSC1L1-short transcript expression was also higher than WHSC1L1-long expression in the normal breast tissue samples collected by TCGA (data not shown).

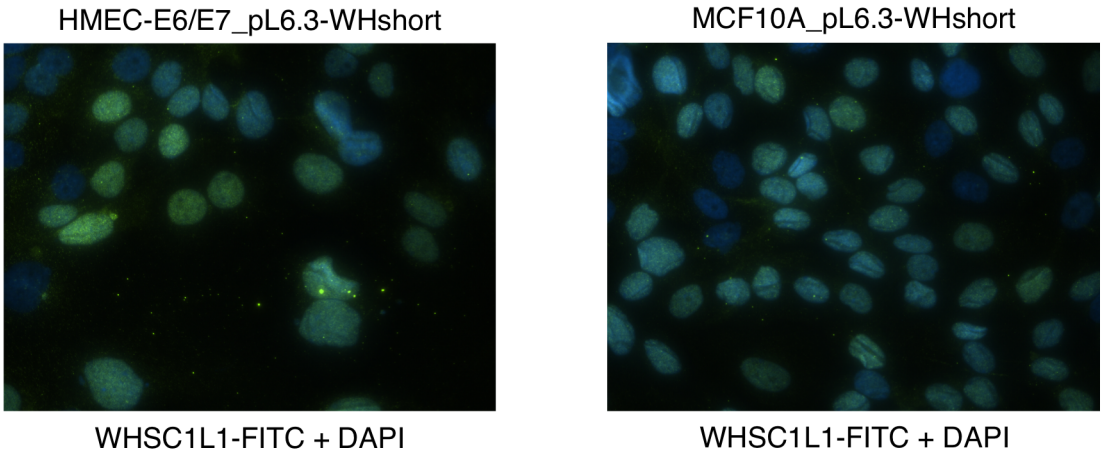


Figure 2.1. WHSC1L1 cellular localization in tumor cell lines.

Immunofluorescence micrographs of human mammary epithelial cells immortalized with viral E6/E7 and MCF10A breast epithelial cells lentivirally transduced with WHSC1L1-short. WHSC1L1 antibody visualized with anti-rabbit FITC and nuclei stained with DAPI.

2.2.2 WHSC1L1-short knockdown reduces cell growth

To confirm that knockdown of WHSC1L1 reduces SUM44 proliferation, and to determine the effect of WHSC1L1-short knockdown, SUM44 cells were lentivirally transduced with shLacZ (control), shWHtotal or shWHshort short-hairpin RNA (shRNA). Of the five WHtotal shRNAs and two WHshort shRNAs, shWHtotal_2 and shWHshort_1 achieved the greatest reduction in WHSC1L1 transcript levels when measured by real-time PCR (Figure 2.1). Knockdown of WHSC1L1 expression was validated by western blot. Both shWHtotal_1 and shWHtotal_2 reduced levels of WHSC1L1 (Figure 2.3a), with the most knockdown occurring in the short isoform. The shWHtotal_2 cells exhibited better knockdown of both isoforms of WHSC1L1; this shRNA vector was chosen for use in later experiments. Both of the shWHshort vectors knocked down WHSC1L1-short levels (Figure 2.3b), however shWHshort_1 consistently had greater knockdown of the short isoform, and was used for later experiments.

With a working WHSC1L1-knockdown model in SUM44 cells, we next investigated the effect of WHSC1L1total and WHSC1L1-short knockdown on cell proliferation. A proliferation assay was performed to measure changes in proliferation after knockdown of WHSC1L1 over a period of 10 days. SUM44 shLacZ cells exhibited an 8.8-fold increase in cell number over the 10-day culture period, while the shWHtotal cell number increased 5.6-fold, and the shWHshort

cells increased 3.8-fold (Figure 2.4a). While both the shWHtotal and the shWHshort cells proliferated more slowly than the shLacZ controls, knockdown of the short isoform of WHSC1L1 had the greatest negative effect on growth.

A boxplot of the distribution of the fold-increase in cell number for 3 repeats of the proliferation assay comparing growth of SUM44 shWHtotal and shWHshort vs shLacZ cells (Figure 2.4b) shows that the interquartile range (IQR) of the SUM44 shLacZ cell number for the three experiment replicates is higher than the IQR for SUM44 shWHshort.

Proliferation of SUM44 slowed with WHSC1L1-short knockdown, but did not stop. Observing this decrease in proliferation upon knockdown of WHSC1L1-short, we became interested in whether there was a buildup of cells in either G1 or G2 in the SUM44 shWHshort cells. The DNA content flow cytometry plot (Figure 2.5) shows DNA content in SUM44 shLacZ and shWHshort by flow-cytometry. The fraction of S-phase cells (blue shaded region) was reduced in the shWHshort cells, and there were increases in both the G1 and G2 phase populations compared to shLacZ control.

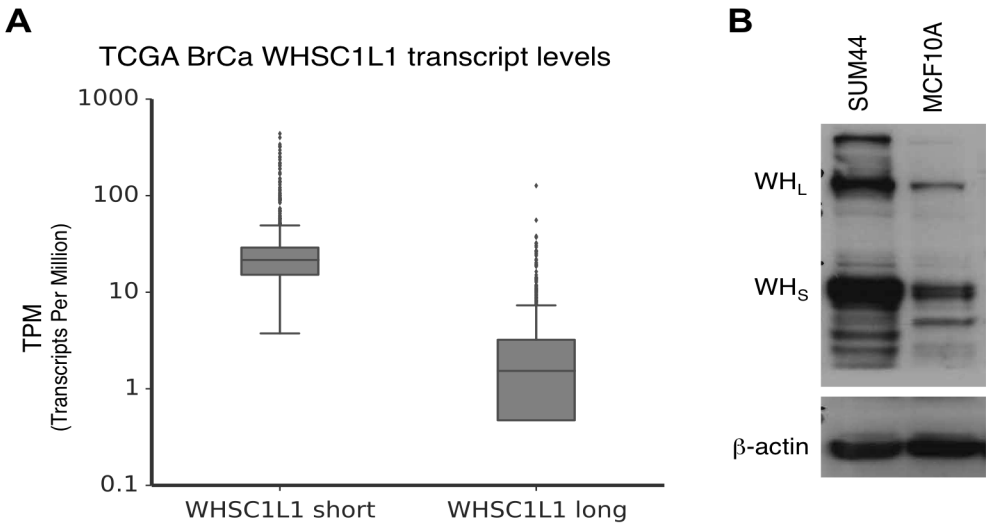


Figure 6.2 WHSC1L1 expression patterns in breast cancer samples and in SUM44 and MCF10A.

a. Transcript expression levels of WHSC11L1 short and long in TCGA breast cancer samples. **b.** Immunoblot for WHSC1L1 in SUM44 breast cancer cells and in MCF10A cells, a model of non-malignant breast epithelial tissue.

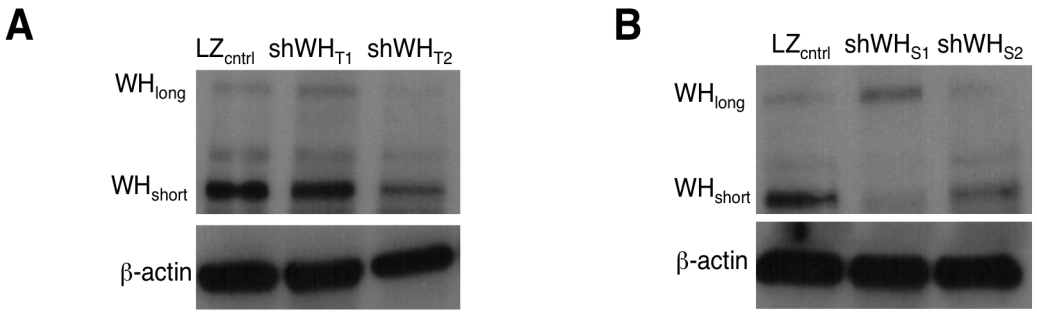


Figure 2.3 Knockdown of WHSC1L1 total and short isoform alone alters WHSC1L1 expression in SUM44 cells.

a. Immunoblot of WHSC1L1 from whole cell lysate of SUM44 shLacZ and shWHSC1L1total cells. **b.** WHSC1L1 immunoblot from whole cell lysate of SUM44 shLacZ and shWHSC1L1short cell lines.

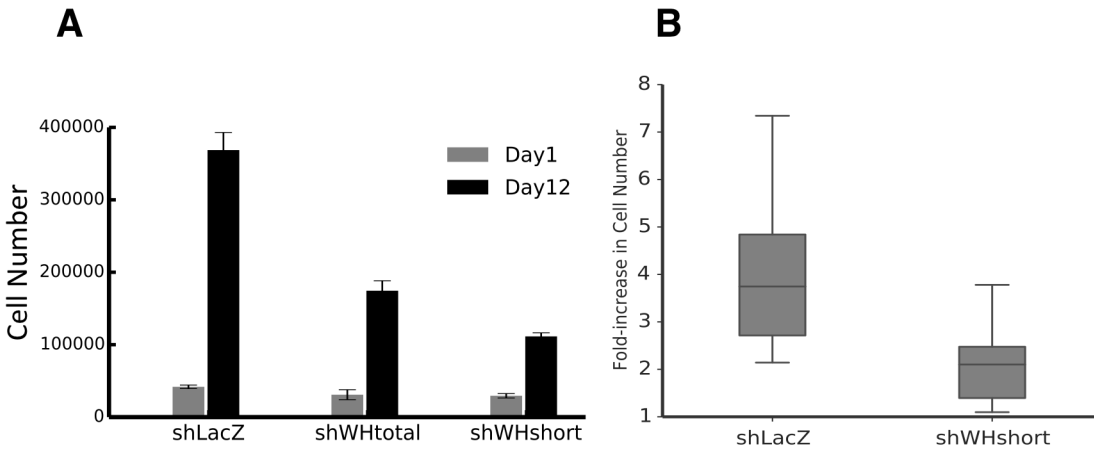


Figure 2.4 Treatment with beta-estradiol affects proliferation of SUM44 cells.

a. Proliferation of SUM44 shWHSC1L1total and shWHSC1L1short cells vs shLacZ control on days 0 and 12. Error bars represent 1.98 * standard error mean (SEM) of 3 experiments. **b.** Boxplot showing the distribution of the fold-increase in cell number in SUM44 shWHSC1L1short vs shLacZ cells after 12 days. Error bars represent 3 experiments.

2.2.3 Knockdown of WHSC1L1-short alters expression of several genes active in breast cancer

WHSC1L1-long possesses histone methyltransferase activity, promoting hypermethylation of histone H3K36. H3 lysine 36 di- and tri-methylation has been shown to promote transcriptional elongation, thus promoting expression of gene loci with these histone marks. While this mechanistic knowledge of WHSC1L1 activity is evidence that WHSC1L1 can be a transcriptional activator, little is known about the effect of WHSC1L1 expression level on the regulation of target gene expression, particularly in the setting of gene amplification. Further, it is not known how WHSC1L1-short, lacking the catalytic SET domain and thus direct methyltranferase activity, regulates gene expression, or whether both isoforms of WHSC1L1 regulate the same target genes. As the target genes regulated by WHSC1L1 expression are not well understood in normal cells, nor in cancer cells where WHSC1L1 is amplified and over expressed, we performed a series of experiments to investigate the effect of WHSC1L1-short and WHSC1L1-total knockdown on the gene expression profile of SUM44 cells.

To carry out these experiments, we first performed expression profiling of SUM44 shWHtotal cells to identify genes regulated by WHSC1L1 expression. SUM44 cells were transduced with pGIPZ shWHtotal lentivirus and selected for Puromycin resistance. Total mRNA extracted from shLacZ and shWHtotal

SUM44 cells was hybridized to an Agilent 2-color microarray and the expression profiles for the two conditions were compared. Knockdown of total WHSC1L1 resulted in significant changes in gene expression (Supplementary Table 1). Among genes with significant transcript downregulation in SUM44 shWHtotal knockdown cells, we identified several that were interesting in the context of ER-positive Luminal B breast cancers, including ESR1, ERBB4, MYCN, CD24, and CD44.

As SUM44 cells express high levels of ER-alpha, we focused on ESR1 and HER4, and validated the microarray data by RT-PCR (data not shown).

To confirm and extend the results obtained from the microarray experiments, we next performed RNAseq expression profiling analysis following knockdown of WHSC1L1-total or WHSC1L1-short. The expression profiles of SUM44 shLacZ, SUM44 shWHtotal, and SUM44 shWHshort cells were measured by Illumina 151 base-pair paired-end sequencing of whole mRNA extracts from each cell line, followed by differential expression analysis using the DESeq2 Bioconductor package to identify significantly differentially expressed genes between SUM44 shLacZ and either SUM44 WHSC1L1 knockdown sample. A plot of RNAseq results (Figure 2.6) shows log2foldchange values for ESR1 in SUM44 shWHshort vs shLacZ DESeq2 differential expression analysis of RNAseq expression data. As shown in the plot, there was a 1.06-fold decrease in ESR1 in shWHtotal cells (FDR 0.97), and a 1.78-fold decrease in ESR1 in shWHshort cells (FDR 0.13).

With confirmation by both microarray and RNAseq that WHSC1L1 knockdown significantly reduced ESR1 expression, we wanted to know whether the downregulation of ESR1 transcript observed in the expression profiling data would correspond to a reduction in levels of ER-alpha protein upon knockdown of WHSC1L1, particularly WHSC1L1-short. ER-alpha protein levels in whole cell lysates from SUM44 shLacZ, shWHtotal, and shWHshort were measured by immunoblot (Figure 2.7a). As seen in the figure, both HER4 and ER-alpha protein levels were markedly reduced in the shWHtotal SUM44 cells. ER-alpha levels were even further reduced in SUM44 shWHshort cells (Figure 2.7b). These data indicate that the amplification and overexpression of the short isoform of WHSC1L1 observed in SUM44 cells is necessary for the high levels of ERa transcript and protein expression in these cells, which is greater than ERa expression in MCF7, the prototypical ER+ breast cancer model cell line (Figure 2.7c).

RNAseq expression profiling of SUM44 shWHshort cells compared to SUM44 shLacZ control identified 1130 significantly differentially expressed genes (p-value < 0.05, FDR < 0.15). Of those 1130 genes, 622 were upregulated and 508 were downregulated in the shWHshort knockdown cells relative to shLacZ control. A total of 237 genes were differentially expressed in SUM44 shWHtotal cells compared to SUM44 shLacZ controls; 114 were upregulated, and 123 downregulated. A total of 63 genes were significantly differentially expressed in

both WHshort and WHtotal knockdown samples compared to shLacZ control. The differentially expressed genes for each condition are shown in the Appendix.

To identify patterns of biological significance in the differentially expressed genes, we ran gene set enrichment analysis (GSEA) on the expression values of all genes measured by RNAseq in SUM44 WHSC1L1 knockdown cells compared to shLacZ controls. A false discovery rate (FDR) of 0.25 was chosen as the cutoff for significant gene set enrichment, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) gene set list was selected to query for enrichment. GSEA of the SUM44 shWHSC1L1-short vs. shLacZ samples identified 10 KEGG pathway gene sets significantly enriched in SUM44 shWHSC1L1-short relative to shLacZ (Table 2.1). Top hits included gene sets involved in DNA replication and DNA repair, as well as steroid biosynthesis.

As shown in Table 2.2, only 3 KEGG pathway gene sets were significantly negatively enriched in SUM44 shWHSC1L1-short cells compared to shLacZ cells.

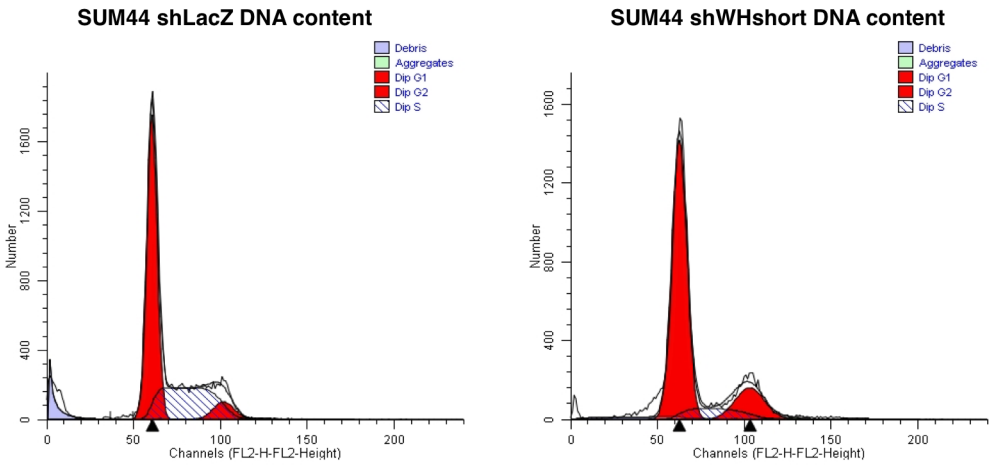


Figure 2.5 Knockdown of WHSC1L1short reduces the S-phase fraction in SUM44 cells.

DNA content assays by flow cytometry of SUM44 shLacZ (left panel) and shWHSC1L1short (right panel).

2.2.4 SUM44 cells respond negatively to estrogen

SUM44 cells express high levels of estrogen receptor alpha, in fact at significantly higher levels than in MCF7, the prototype cell line for ER-positive breast cancer. Both MCF7 and SUM44 were isolated from pleural effusion metastases in patients with late stage breast cancer. While the sensitivity of MCF7 cells in vitro to estradiol has been comprehensively described in the literature, it remains one of the few ER-positive breast cancer cell line models, and represents the Luminal-B subtype. Upon observing the effect of WH knockdown on ER-alpha protein expression levels, we became interested in the response of SUM44 cells to treatment with estrogens, and whether WH knockdown would affect the SUM44 response to estrogen.

Proliferation assays were conducted on SUM44 shLacZ, shWHtotal, and shWHshort cells. The cells were cultured in serum-free media that did not contain phenol red, and were treated with either vehicle control or varying concentrations of beta-estradiol. A plot of cell proliferation (Figure 2.8) shows SUM44 shLacZ and shWHshort cell numbers after 11 days of culture in increasing concentrations of beta-estradiol. SUM44 shLacZ cell proliferation began to decrease at 1pM beta-estradiol, and decreased in a dose-dependent manner through 1nM, indicating that picomolar concentrations of estrogen have an inhibitory effect on SUM44 proliferation. In contrast, proliferation of SUM44 shWHshort cells

increased in 1pM estrogen, and further increased at 10pM, then steadily decreased at 100pM and 1nM beta-estradiol, in parallel with the proliferation decrease seen in the shLacZ cells.

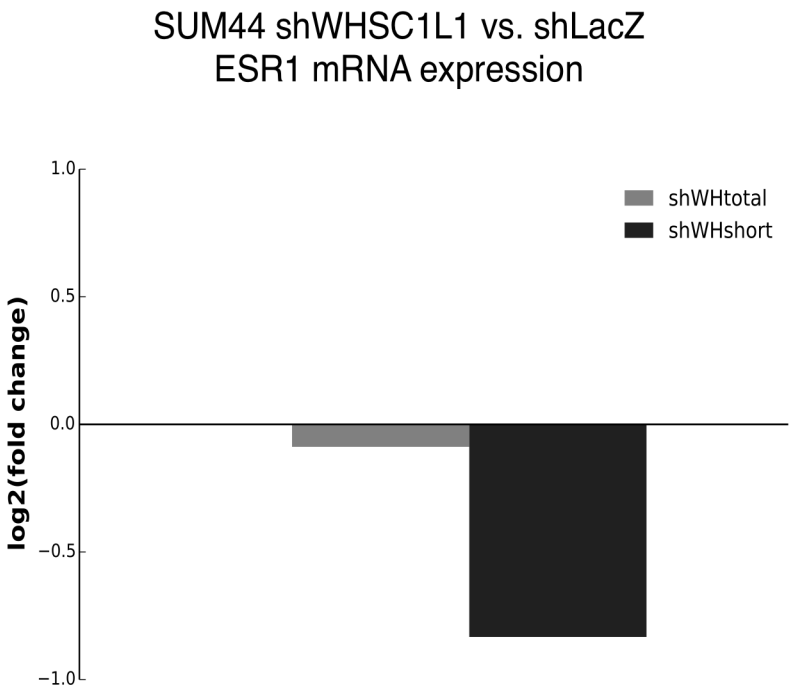


Figure 2.6 Knockdown of WHSC1L1 total and short isoform alone alters expression of ESR1 transcript in SUM44 cells.

RNAseq differential expression of ESR1 transcript levels in SUM44 shWHSC1L1short vs shLacZ control cells. Experiment performed in duplicate with differential expression performed using DESeq2.

RANK	DETAILS	SIZE	NES	FDR
1	DNA_REPLICATION	36	2.23	0.0
2	ANTIGEN_PROCESSING_AND_PRESENTATION	88	2.0	0.002
3	CELL_CYCLE	124	1.94	0.003
4	SYSTEMIC_LUPUS_ERYTHEMATOSUS	134	1.93	0.003
5	SPLICEOSOME	125	1.92	0.003
6	STEROID_BIOSYNTHESIS	16	1.84	0.008
7	HOMOLOGOUS_RECOMBINATION	28	1.82	0.01
8	MISMATCH_REPAIR	23	1.81	0.01
9	CITRATE_CYCLE_TCA_CYCLE	30	1.75	0.018
10	PYRIMIDINE_METABOLISM	96	1.75	0.017

Table 2.1. Gene set enrichment analysis (GSEA) positively-enriched processes in SUM44 shWHSC1L1short vs shLacZ cells.

RANK	DETAILS	SIZE	NES	FDR
1	DNA_RIBOSOME	86	-3.23	0
2	ARACHIDONIC_ACID_METABOLISM	58	-1.66	0.101
3	NITROGEN_METABOLISM	23	-1.58	0.134

Table 2.2 GSEA negatively-enriched processes in SUM44 shWHSC1L1short vs shLacZ cells.

RANK	DETAILS	SIZE	NES	FDR
1	CELL_CYCLE	123	2.8	0.0
2	DNA_REPLICATION	35	2.44	0.0
3	OOCYTE_MEIOSIS	102	2.12	0.001
4	MISMATCH_REPAIR	22	2.09	0.001
5	HOMOLOGOUS_RECOMBINATION	27	2.02	0.003
6	PYRIMIDINE_METABOLISM	92	1.98	0.003
7	SPLICEOSOME	125	1.96	0.004
8	NUCLEOTIDE_EXCISION_REPAIR	43	1.93	0.005
9	PROTEASOME	41	1.9	0.007
10	STEROID_BIOSYNTHESIS	16	1.9	0.006
11	FOLATE_BIOSYNTHESIS	11	1.89	0.006
12	P53_SIGNALING_PATHWAY	66	1.87	0.007
13	SYSTEMIC_LUPUS_ERYTHEMATOSUS	99	1.87	0.007
14	GALACTOSE_METABOLISM	23	1.84	0.01
15	PROGESTERONE_MEDIATED_OOCYTE_MATURATION	79	1.82	0.011
16	CITRATE_CYCLE_TCA_CYCLE	29	1.77	0.018
17	BASE_EXCISION_REPAIR	33	1.77	0.017
18	PENTOSE_PHOSPHATE_PATHWAY	25	1.7	0.03
19	RNA_DEGRADATION	56	1.69	0.032
20	BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS	21	1.64	0.045

Table 2.3 GSEA positively-enriched processes in SUM44 cells treated with beta-estradiol.

2.2.5 Gene expression profiling of SUM44 cells after treatment with estrogen

SUM44 cells represent a model of highly ER-positive luminal B breast cancer. As shown above, SUM44 cells are exquisitely sensitive to estrogen, responding to as little as 10pM beta-estradiol, and actually reducing their proliferation rate upon treatment with greater than 10pM beta-estradiol. Given this extraordinary response to estrogen, we became interested in the effect of estrogen treatment on gene expression in SUM44 cells.

To determine the effect of estrogen treatment on SUM44 gene expression, we cultured the cells in estrogen-free conditions for 72 hours, then added 10nM beta-estradiol and waited 6, 12, and 24 hours after treatment before harvesting whole mRNA and measuring transcript expression by RNAseq. We then performed GSEA on the expression profiling data to determine which KEGG pathways were positively or negatively enriched in SUM44 estrogen-treated cells compared to SUM44 cells grown in normal conditions. As shown in Table 2.3, treatment with estrogen for 6 hours resulted in positive enrichment of cell-cycle, DNA replication, meiosis, and DNA mismatch repair KEGG pathway gene sets. Interestingly, with the exception of the lupus erythematosus and antigen processing and presentation gene sets, all of the gene sets positively enriched in SUM44 shWHshort cells were also positively enriched after treating SUM44 cells

with estrogen. In addition to the common gene sets, several additional gene sets were positively enriched, including several pathways related to steroid hormone receptor activation. The p53 pathway was also positively enriched upon treatment of SUM44 with estrogen. Overall, a total of 20 gene sets were enriched with an FDR less than or equal to 0.05.

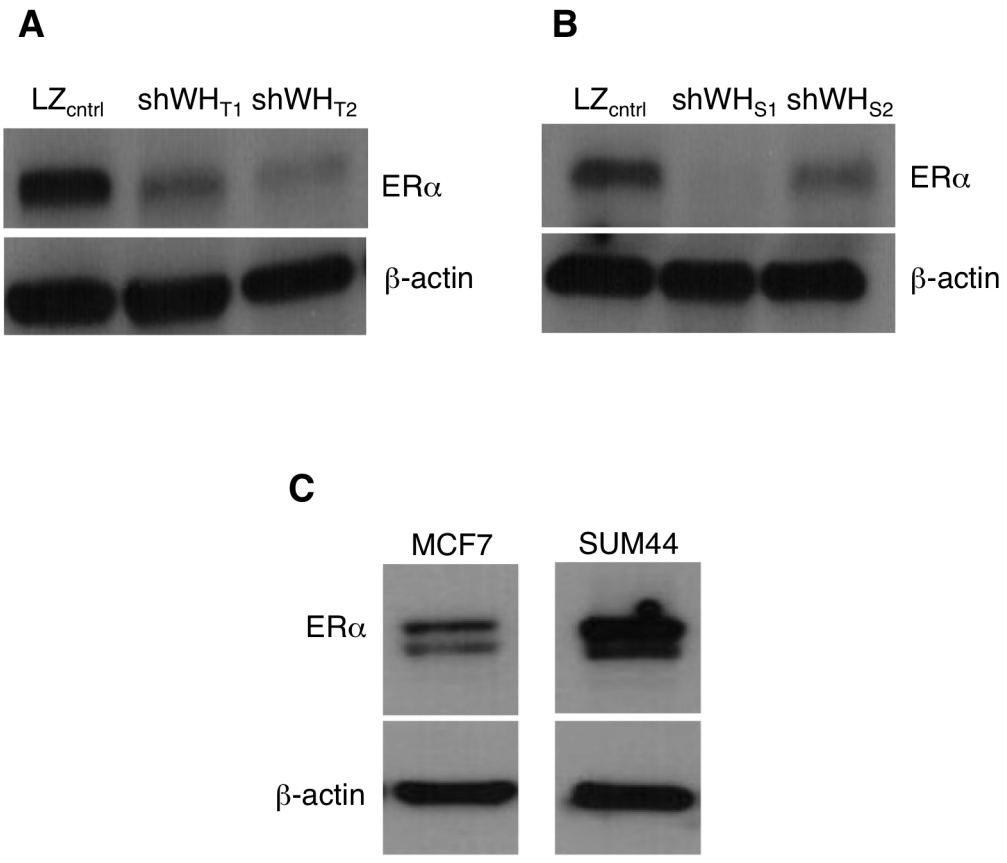


Figure 2.7 Knockdown of WHSC1L1 total and short isoform alone alters expression of ERa protein in SUM44 cells.

a. Immunoblot of ERa from whole cell lysate of SUM44 shLacZ and shWHSC1L1total cells. **b.** ERa immunoblot of whole cell lysate from SUM44 shLacZ and shWHSC1L1short cells. **c.** ERa immunoblot of MCF7 and SUM44 whole cell lysate.

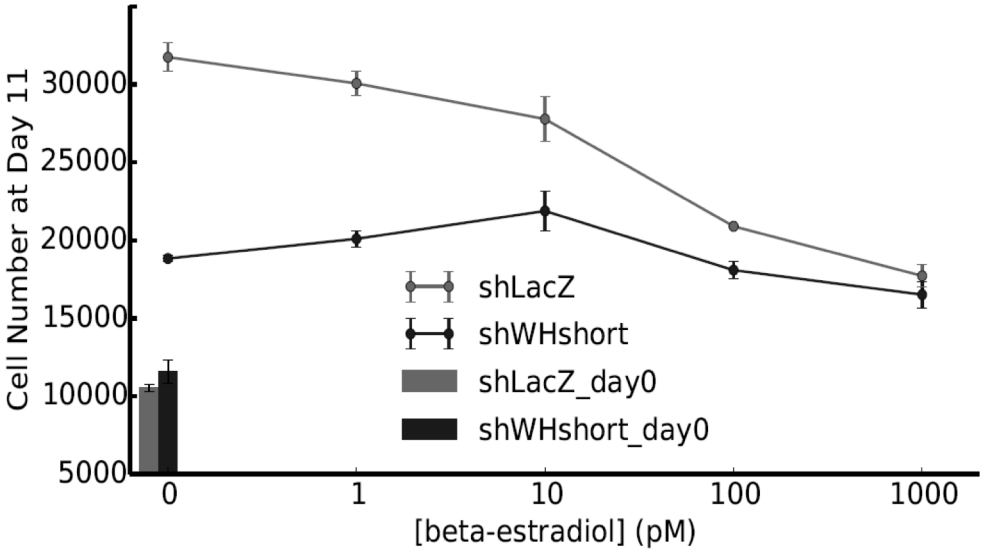


Figure 2.8 SUM44 proliferation is sensitive to treatment with beta-estradiol.

Proliferation of SUM44 shWHSC1L1short and shLacZ cells in response to increasing amounts of beta-estradiol. Error bars show 1.8 * SEM for 3 experiments.

2.3 Discussion

We were able to successfully create WHSC1L1-total and WHSC1L1-short knockdown models using a lentiviral shRNA expression vector stably transduced in SUM44 cells, achieving a reduction in both transcript and protein levels of WHSC1L1-long and WHSC1L1-short as shown by real-time PCR and western blotting, respectively. As previously shown, knockdown of total WHSC1L1 negatively affected SUM44 cell proliferation, and our results with WHSC1L1-short knockdown indicated an even greater negative effect on proliferation when only the short isoform was knocked down. The lack of a model that knocked down the long isoform of WHSC1L1 alone complicates the interpretation of the proliferation experiments, however the data with the total and short-only WHSC1L1 knockdown could support a model where the activities of the long and short isoforms of WHSC1L1 are opposed. While shRNA knockdown of total WHSC1L1 resulted in successful knockdown of both isoforms of WHSC1L1, knockdown of WHSC1L1-short with the short isoform-specific shRNA appeared to achieve a greater knockdown of WHSC1L1-short than knockdown of either isoform in the shWHtotal cell lines.

Also, interestingly, in the SUM44 shWHshort-1 cells, which achieved the greatest knockdown of WHSC1L1-short, the level of WHSC1L1-long appears to increase noticeably. This increase can also be observed at the mRNA level in the RNAseq expression profiling data (Appendix). These results suggest

autoregulation of WHSC1L1 expression by WHSC1L1-short, and the increased level of WHSC1L1-long in the SUM44 shWHshort cells, together with the much greater number of significantly differentially expressed genes upon WHSC1L1-short knockdown, suggests that SUM44 cells are much more sensitive to expression levels of the short isoform of WHSC1L1. It would be interesting to investigate whether changes in WHSC1L1-short expression cause a greater magnitude of changes compared to changes in WHSC1L1-long in other cell lines as well.

Knockdown of WHSC1L1 resulted in differential expression of many genes known to be interesting in breast cancer and notably resulted in significant downregulation of ESR1 transcript, which in turn resulted in a decrease in ERa expression. In addition to ERa, HER4 mRNA and protein levels were reduced in SUM44 shWHSC1L1-short cells relative to shLacZ control.

While we chose to focus on ERa downregulation due to the importance of the estrogen receptor in breast cancer and specifically the high ERa levels observed in SUM44 cells, several other interesting genes were differentially expressed upon WHSC1L1 knockdown, and the mechanism of regulation of these genes by WHSC1L1 should be investigated. One important general conclusion that can be made from the RNAseq expression profiling data is that knockdown of WHSC1L1-short had a positive overall effect on gene expression, increasing the expression level of over 600 genes, while downregulating approximately 500. It should also be noted that the total number of significantly

differentially regulated genes in the SUM44 shWHSC1L1total cells was much smaller than the number that were significantly altered in SUM44 shWHSC1L1short.

One possible interpretation of this finding is that the long isoform and short isoform of WHSC1L1 have opposing effects on the set of target genes shared by both isoforms. Also, the observation that WHSC1L1-long increased upon knockdown of WHSC1L1-short supports the possibility that the two isoforms of WHSC1L1 reciprocally regulate each others' expression. The net increase in expression upon WHSC1L1-short knockdown could indicate that the effect of WHSC1L1 amplification and overexpression in SUM44 cells has a net suppressive effect on gene expression, which would be counterintuitive to the effects of most oncogenes, which generally upregulate pathways involved in cell growth, proliferation, and survival. Indeed, the fact that TP53 is mildly upregulated upon knockdown of WHSC1L1-short, along with positive enrichment for DNA damage repair pathways, may indicate that WHSC1L1 amplification plays a role in suppression of tumor suppressor pathways in SUM44. The fact that the same pathways are upregulated upon treatment of SUM44 with estrogen is admittedly baffling, although extremely interesting.

There is not enough data from these experiments to propose a model for the similar response of SUM44 cells to WHSC1L1-short knockdown and treatment with estrogen, however the evidence strongly suggests that treatment of SUM44 cells with estrogen has a similar effect as downregulating ERa

expression (as has been shown to occur when WHSC1L1-short is knocked down). This is a surprising finding, and one that needs to be repeated in SUM44 and additional ER-positive breast cancer cell lines with WHSC1L1 overexpression. The possible relationship between WHSC1L1-short overexpression and the response of this highly ER-positive cell line to estrogen suggest that SUM44 may be an informative model for increasing our understanding of mechanisms for endocrine resistance in breast cancer (Gururaj, Rayala et al. 2006).

It should be noted that mRNA for the TSKU gene was the fourth most abundant transcript found in SUM44 shLacZ cells, and was also significantly upregulated in SUM44 shWHSC1L1-short cells. This is interesting considering the fact that the TSKU locus is part of an amplified region on the long arm of chromosome 11 that is also amplified in SUM44 cells. The high expression of TSKU is expected, given the gene amplification. However, the fact that TSKU expression appears to be regulated by expression of WHSC1L1-short, a transcript coded for by another amplified region on a different chromosome, suggests that there may be functional cooperation among genes in different amplified regions that co-occur in breast cancer. The question whether co-amplified genes share common regulatory pathways is an important question in cancer. It's logical that the presence of multiple amplified regions in a cancer is due to selection for the co-amplification of these regions, as they are presumably rare events that would not be expected to co-occur without selective pressure.

2.4 Materials and Methods

2.4.1 shRNA vectors

Lentiviral shRNA expression vectors were from the pLKO shRNA catalog purchased from Sigma. Vectors used include TRCN0000415241 (WHSC1L1-short), and the WHSC1L1-total vectors, TRCN0000425711, TRCN0000015615, and TRCN0000015616.

2.4.2 Lentiviral infection protocol

Lentivirus for lentiviral transduction of shRNAs into SUM44 cells was prepared using the Sigma Mission lentiviral packaging system (Sigma, shp-001) in 293FT packaging cells following the manufacturer's protocol. Briefly, each construct was co-transfected into 293FT cells with Sigma pLKO shRNA vectors and virus was harvested according to the manufacturer's instructions. Target cells were transduced with packaged virus in growth media supplemented with 5 ug/ml polybrene. Cells were cultured for three days to allow for expression of the resistance marker. Non-transduced cells were eliminated from the culture by addition of the selection agent puromycin (3ug/mL).

2.4.3 Antibodies

Antibody for WHSC1L1 was purchased from ProteinTech Group Inc. (Cat. No.11345-1-AP). Estrogen receptor alpha antibody was purchase from Bethyl Labs (Cat. No. A300-498A). ChIP-grade anti-estrogen receptor alpha antibodies

were purchased from Santa Cruz (sc-543x) and Thermo Scientific (MA5-13065). ChIP-grade histone antibodies were used for both ChIP and immunoblotting, and were purchased from Abcam and Millipore: total H3 (Abcam ab1791), H3K36me3 (Abcam ab9050), H3K36me2 (Abcam ab9049), and H3K4me3 (Millipore 17-678).

2.4.4 Cell culture

SUM44PE cells were cultured in serum-free Ham's F-12 with supplements (Ham's F-12 with 1ug/ml hydrocortisone, 1 mg/ml bovine serum albumin, 10 mM Hepes, 5 mM ethanolamine, 5ug/ml transferrin, 10 nM triiodothyronine, 50 nm sodium selenate, 25ug/ml gentamicin, 2.5ug/ml fungizone, and 5ug/ml insulin), as described previously (Ethier, Chiodino et al. 1990). For estrogen-free culture conditions, phenol red-free Ham's F-12 media was substituted for normal Ham's F-12; all supplements were as in normal media. Cells transduced with lentiviral shRNAs were selected and maintained in 3 ug/mL puromycin.

2.4.5 Proliferation assays

Proliferation assays were performed in 6-well culture plates seeded with $5x10^5$ cells per well. At 24 hours and 10 days after seeding, cells were washed three times with PBS and agitated on a rocker table with 0.5 ml of a HEPES/$MgCl_2$ buffer (0.01 M HEPES and 0.015 M $MgCl_2$) for 5 minutes. Cells were then lysed for 10 minutes with a solution of ethyl hexadecyldimethylammonium, and the nuclei were counted using a Z1 Coulter Counter (Beckman Coulter, Brea, CA, USA).

2.4.6 Western blotting

Cells were plated and grown to 90% confluency. Where indicated, the cells were treated with 500 nM gefitinib for the indicated times. Cells were then lysed in a buffer containing 20 mM Tris-HCl (pH 8.0), 137 mM NaCl, 1% NP40, 10% glycerol, 1 mM Na_3VO_4, and 1X Protease Inhibitor cocktail (Calbiochem, 539131), and protein concentrations were measured by Bradford assay (Bio-Rad). Laemmli sample buffer was added to the lysates and the samples were boiled for 5 minutes before being separated by electrophoresis on SDS-polyacrylamide gels (Bio-Rad). After transferring the proteins to polyvinylidene difluoride (PVDF) membranes, blots were probed overnight at 4° Celsius with the indicated antibodies.

2.4.7 RNA expression profiling

Total RNA was prepared using a Qiagen RNeasy Plus Mini Kit and processed by the MUSC Genomics core for 2x101 cycles, paired-end RNA sequencing on an Illumina HiScanSQ. RNA integrity was verified on an Agilent 2200 TapeStation (Agilent Technologies, Palo Alto, CA). A total of 100-200 ng of total RNA was used to prepare RNA-Seq libraries using the TruSeq RNA Sample Prep kit following the protocol described by the manufacturer (Illumina, San Diego, CA). Sequencing was performed on an Illumina HiScanSQ. Samples were demultiplexed using CASAVA (Illumina, San Diego, CA). Fastq files were used to map reads to the human genome (hg19, UCSC) utilizing Tophat2 with default settings. A total of 100-200 ng of total RNA was used to prepare RNA-Seq

libraries using the TruSeq RNA Sample Prep kit following the protocol described by the manufacturer (Illumina, San Diego, CA).

2.4.8 Bioinformatics and Statistical analysis

Paired end sequencing of 151 base-pair reads was performed on RNA samples using an Illumina HiSeq2500 with each sample sequenced to a depth of approximately 10 million reads. Data was subjected to Illumina quality control (QC) procedures (>80% of the data yielded a Phred score of 30), and preprocessing using Trimmomatic which removed adapter sequences and filtered low quality reads (Bolger, Lohse et al. 2014). Further data QC was performed using FastQC, prior to aligning the reads to the human genome HG19 using Tophat2 (Kim, Pertea et al. 2013). The resulting SAM files were inputted into the Python package HTSeq and quantitative readouts for each sample obtained in the form of count data. In order to infer differential signal within the data sets with robust statistical power, we utilize DEseq2, a method which tests for differential expression based on a model using negative binomial distribution (Love, Huber et al. 2014). Transcript count data from DESeq2 analysis was ranked according to q-value, which is the smallest false discovery rate (FDR) at which a transcript is called significant. Statistical analysis of Pathways and Gene Ontology terms was carried out using this sorted transcript list (Subramanian, Tamayo et al. 2005) which was subjected to Gene Set Enrichment Analysis (GSEA) using the ToppGene Suite (Chen, Bardes et al. 2009).

CHAPTER 3 ESTROGEN RECEPTOR ALPHA BINDS CHROMATIN IN THE ABSENCE OF ESTROGEN IN SUM44 CELLS

3.1 Introduction

Estrogen receptor signaling is important in the growth and development of tissue, and plays a key role in ER+ breast cancer. The presence or absence of ERa expression in a malignant breast tumor has a profound effect on the aggressiveness of a malignant breast tumor, affects the options for treatment, and influences the overall prognosis. ERa testing, together with tests for progesterone receptor expression and HER2 amplification/overexpression are the primary molecular markers used to classify breast cancers at diagnosis. Nearly 75% of breast cancers diagnosed are found to be positive for ERa expression, and treatments that inhibit activation of the estrogen receptor are effective in most patients with ER+ disease (Rose, Kamby et al. 1986).

Given the importance of estrogen signaling in ER+ breast cancers, there is high demand in basic cancer research for a cell line that models ER+ breast cancer in the lab. The MCF7 cell line, developed in 1973 at the Michigan Cancer Foundation (now Karmanos Cancer Institute) in Detroit, was one of the first cell line models of human breast cancer that could be cultured indefinitely, and remains essentially the sole model of ER+ breast cancer used for basic research as of 2015. One active area of basic research using MCF7 cells is the study of cellular response to inhibition of estrogen signaling, specifically inactivation of

estrogen receptor alpha. The majority of our current understanding of the cellular response of ER+ breast cancer cells to estrogen agonists and antagonists is based on research conducted with MCF7 cells. While the research conducted so far is invaluable to our understanding of ER biology in ER+ breast cancers, investigation of ERa biology in additional breast cancer cell lines will most likely strengthen some findings, while possibly contradicting others, eventually leading to a more complete understanding of the role of ERa in breast cancer biology.

3.2 Results

3.2.1 ERa binds chromatin in the absence of estrogen in SUM44 cells

The high levels of ERa protein expression in SUM44 cells and the cells' lack of response to exogenous estrogen suggest that estrogen receptor activity is an important driver of the biology of these cells in the absence of estradiol. We next sought to determine ERa binding sites in SUM44 cells, and to measure the relative binding intensity of ERa in estrogen-free and estrogen-containing conditions. To measure ERa binding to chromatin in SUM44, we cultured cells in phenol red-free SFIH media for 72hrs to maintain estrogen-free culture conditions, and treated with either vehicle or 10nM beta-estradiol for 45 minutes immediately before harvesting the cells for chromatin immunoprecipitation. We then chromatin immunoprecipitated the fragmented chromatin with anti-ERa antibody and sequenced the ChIP'd DNA on an Illumina Hi-Seq 2500 (Goren, Ozsolak et al. 2010). Three biological replicates of each sample type, vehicle and 10nM beta-estradiol, were prepared and sequenced. After alignment and peak

calling with two different peak callers, Homer's findPeaks and MACS2, we identified a consensus peakset containing the binding regions found to be present in at least 2 of the 3 replicates for each condition. We found 20820 ERa binding regions present in estrogen-free SUM44 shLacZ samples, and 33121 ERa peaks in the SUM44 plus 10nM estrogen samples.

To determine the effect of estrogen treatment on ERa binding in SUM44 cells, we next performed differential occupancy and differential binding analyses on SUM44 shLacZ 10nM estradiol vs estrogen-free cells using the DiffBind package by Anders et al. There was a large overlap of 17661 peaks shared between the estrogen-free and 10nM estrogen samples (Figure 3.1a), however the presence of a large number of peaks in the absence of estrogen is particularly interesting. The peaks in each of the three groups, estrogen-free only, estrogen-containing only, and plus/minus estrogen, were annotated to find nearby genes, and the gene annotations were grouped according to the three conditions: estrogen-free ERa binding, estrogen-rich ERa binding, and ERa binding under both conditions. The majority of the 6246 ERa-bound genes in SUM44 were bound by ERa in both estrogen-free and estrogen-rich conditions: 4761 genes with 17661 ERa binding sites (Figure 3.1b). In addition to an occupancy analysis of the ERa binding sites in 10nM estrogen vs. estrogen-free conditions, a differential binding analysis of the

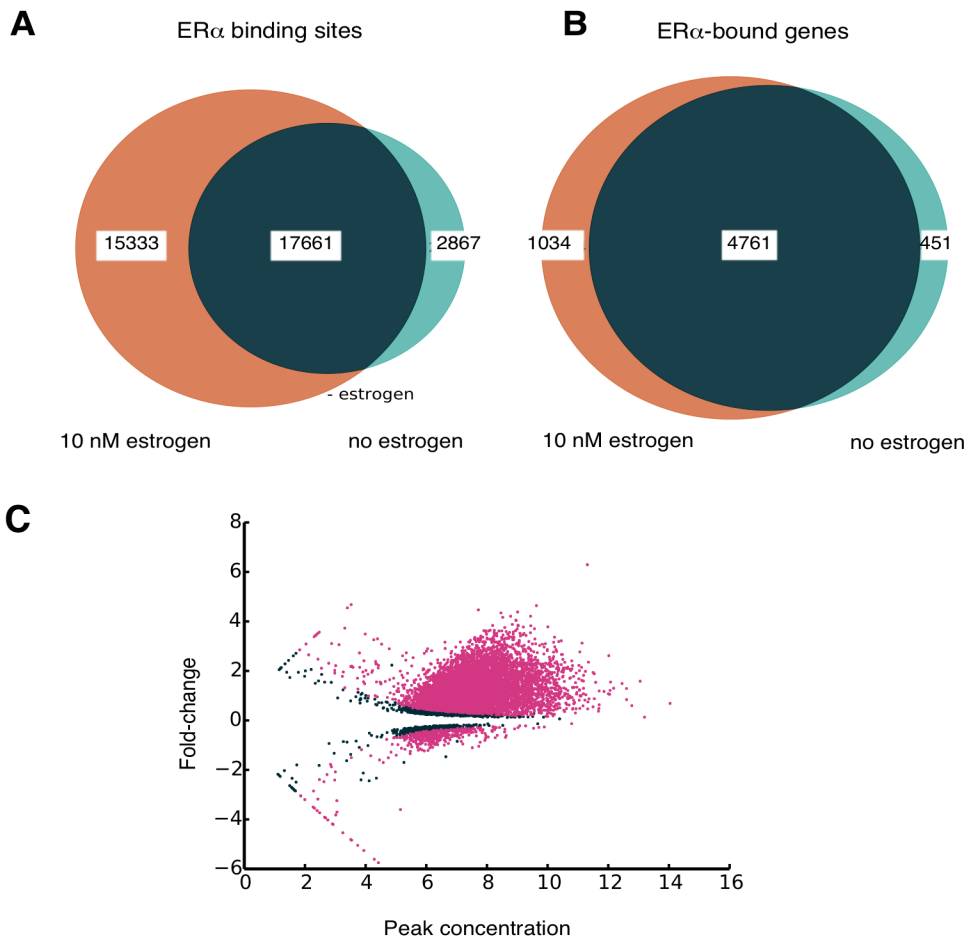


Figure 3.1 ERa binding in SUM44 cells in estrogen-free conditions compared to treatment with beta-estradiol.

a. Venn diagram of ERa binding sites in SUM44 cells treated with 10nM beta-estradiol vs. cells cultured in estrogen-free conditions, as measured by ERa ChIPseq. **b.** Venn diagram ERa-bound genes by ERa ChIPseq in estrogen-free vs 10nM beta-estradiol conditions for SUM44 cells. **c.** MA-plot showing plotting the fold change in ERa binding vs. mean peak intensity after treatment of SUM44 cells under estrogen free conditions with 10nM beta-estradiol for 45 minutes before performing ERa ChIPseq.

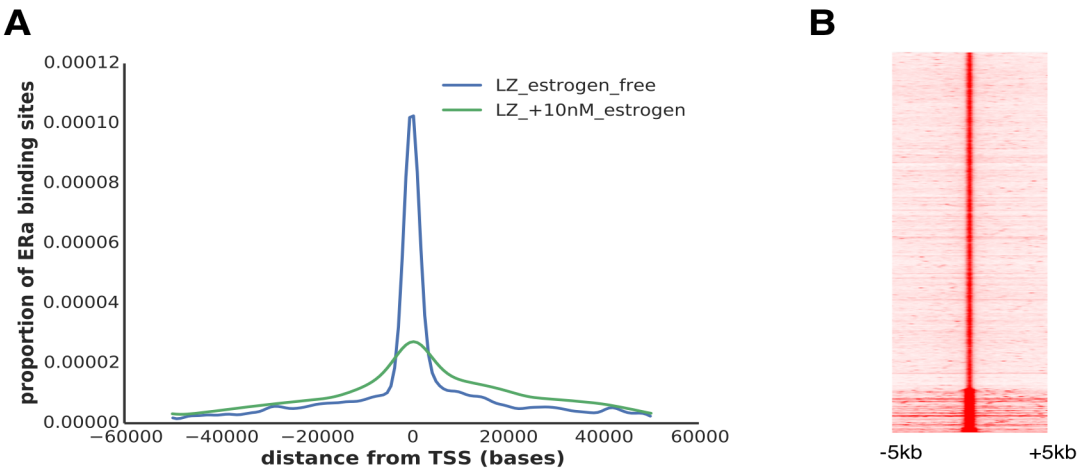


Figure 3.2 ERa binding in SUM44 cells in estrogen-free conditions and after 10nM beta-estradiol treatment.

a. Kernel density estimate plot of the distribution of ERa binding site distances from a transcription start site (TSS). **b.** Heatmap of ERa binding for all ERa binding sites in non-treated SUM44 cells over a 10kb window centered on the peak maximum.

ERa binding sites present in both conditions was performed. Figure 3.1c shows an MA plot of ERa binding in SUM44 cells treated with 10nM beta-estradiol relative to SUM44 cells in estrogen free conditions. This plot shows the fold-change in significantly differentially bound ERa binding sites (pink data points) as a function of overall peak intensity (x-axis). The plot shows that there were 7574 ERa binding sites with significant changes in ERa binding, and that most sites common to both conditions with moderate initial binding intensity showed an increase in ERa binding in the presence of estrogen. As seen in the MA-plot, the ERa binding sites that did decrease after treatment with estrogen exhibited relatively weak binding.

While the above differential ERa binding analysis showed that most changes to ERa binding upon treatment of SUM44 cells with estrogen manifested as increased binding intensity to sites already bound by ERa in estrogen-free conditions, we were also interested in whether there was a shift in peak intensity based on distance of the peak from the TSS of the most proximal gene. To investigate the distribution of peak area as a function of distance from the transcription start sites of all ER-bound genes, a kernel density estimate plot was created based on the distributions of all ERa consensus peaks in SUM44 cells grown in estrogen-free conditions and SUM44 +10nM estrogen cells. Figure 3.2a shows the distribution of ERa peaks in SUM44 estrogen-free cells (the blue line) compared to SUM44 cells treated with 10nM beta-estradiol (green line). In the estrogen-starved SUM44 cells, most ERa-bound regions are within 10kb of a

TSS. In contrast, the ERa bound to chromatin in the estrogen-treated SUM44 cells is much more spread out, with a lower proportion within 10kb of a TSS. Figure 3.2b shows that the profile of most ERa binding sites is unimodal and narrow, with the base of the peak covering less than 1kb. As seen at the bottom of the heatmap, there is a subset of ERa peaks that are wider, however these peaks are in the minority.

3.2.2 Motif enrichment at ERa binding sites

Motif enrichment analysis was run on the ERa binding sites unique to each condition to determine if there were changes in the motifs bound by ERa in the presence or absence of estrogen. As shown in Figure 3.3, the top 3 motifs enriched by ERa binding sites common to both estrogen-free and estrogen-containing SUM44 cells are ERE, FOXA1, and Foxa2. The ERa binding sites unique to plus estrogen conditions also enrich ERE motifs most strongly, followed by Esrrb, then FOXA1. Interestingly, the ERa binding regions unique to estrogen-free conditions do not enrich for ERE motifs; the top three motifs enriched in SUM44 estrogen-free ERa binding sites are Foxa2, FOXA1, and Fox:Ebox motifs.

While the strength of enrichment for a given motif, shown by p-value, changed the ranking of the top three motifs enriched for a given condition, the actual percentage of ERa binding sites bound to a given motif exhibited a different pattern. As shown in the lower panel of Figure 3.3, the percentage of ERa peaks bound to ERE motifs increased from 17.8% of the ERa peaks

common to both estrogen conditions to 27.7% of peaks unique to the 10nM estrogen condition. In contrast, the ERa peaks unique to the estrogen-free conditions contained essentially no ERE motifs. The percentage of ERa peaks bound to FOXA1 motifs decreased from 36.7% for the common peaks to 27.7% for the 10nM estrogen unique peaks, while the proportion of estrogen-free only peaks bound to FOXA1 remained near 37%. The proportion of ERa peaks bound to Foxa2 motifs also remained constant between the common peaks and the estrogen-free unique peaks, and fell to 20.1% of the peaks unique to the 10nM estrogen condition. These results indicate that ERa is not directly binding EREs in the absence of estrogen, but that response elements for pioneer factors such as Foxa2 and FOXA1 are most strongly associated with estrogen-free ERa binding, suggesting a mechanism other than direct recognition of ERE motifs by the estrogen receptor. Having analyzed motif enrichment by ERa binding in SUM44, we then determined which genomic features were enriched by ERa binding in the estrogen-free and 10nM estrogen conditions in SUM44 shLacZ cells.

ERE

ESRRB

FOXA1

Fox:Ebox

Foxa2

both +/- estrogen			10nM estrogen only			estrogen-free only		
motif	%peaks	p-value	motif	%peaks	p-value	motif	%peaks	p-value
ERE	17.8	1E-1583	ERE	27.7	1E-2832	Foxa2	27.9	1E-227
FOXA1	36.7	1E-1388	Esrrb	21.9	1E-777	FOXA1	36.2	1E-213
Foxa2	27.9	1E-1367	FOXA1	27.7	1E-646	Fox:Ebox	32.3	1E-193

Figure 7.3 ERa motif enrichment in SUM44 shWHSC1L1short and shLacZ cells.

Motifs significantly enriched on or near ERa binding sites in SUM44 cells under either estrogen-free, estrogen-treated (10nM), or both conditions.

3.2.3 ERa binding decreases upon WHSC1L1-short knockdown

Given the evidence that WHSC1L1short is necessary for high levels of ERa expression in SUM44 cells, we next asked how ERa binding to chromatin is affected by knockdown of WHSC1L1short, both in overall binding intensity and the pattern of ERa binding to specific motifs, genomic features, or gene sets. We performed ERa ChIPseq on SUM44 shWHshort cells that were cultured in estrogen-free conditions for 72 hours and then either treated with vehicle or 10nM beta-estradiol for 45 minutes before fixing for ChIP, with three biological replicates per condition. Unlike the SUM44 shLacZ cells, we were not able to detect any well-defined peaks in the estrogen-free conditions for any of the replicates. A few peaks were called, but the signal-to-noise ratio was low, eroding confidence in the results.

A total of 12254 peaks were called in SUM44 shWHshort cells treated with 10nM beta-estradiol. To determine the effect of WHSC1L1short knockdown on ERa binding in SUM44 in the presence of 10nM estradiol, we performed a differential occupancy analysis and differential binding analysis using the DiffBind package as described earlier. As shown in Figure 3.4a, 11212 ERa peaks were found to co-occupy the same regions in both shLacZ and shWHshort cells, with 21864 peaks present only in shLacZ, and 1042 peaks present only in shWHshort cells. As shown by the MA-plot in Figure 3.4b, differential binding analysis found a total of 6320 significantly differentially bound peaks in shWHshort

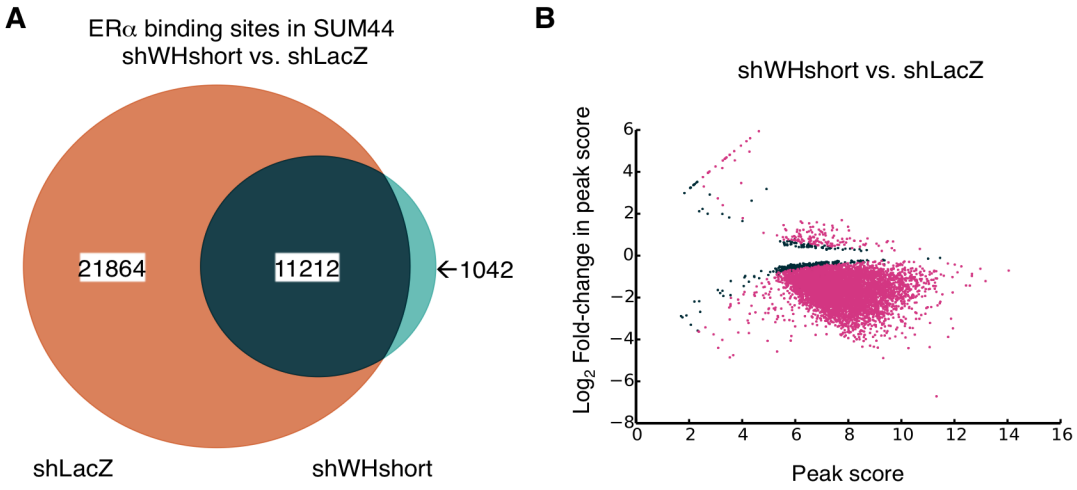


Figure 3.4 ERa binding in SUM44 shWHSC1L1short vs shLacZ cells after treatment with 10nM beta-estradiol.

a. Venn diagram of ERa binding sites in SUM44 shWHSC1L1short cells compared to shLacZ control cells. **b.** MA-plot showing the change in peak intensity (score) vs. mean peak intensity in SUM44 upon knockdown of WHSC1L1short compared to shLacZ control.

10nM estradiol vs shLacZ 10nM estradiol cells, and the majority of ERa binding sites showed a decrease in binding intensity in the shWHshort cells. Motif enrichment analysis of these peaks showed similar motifs enriched in the shLacZ and shWHshort SUM44 cells, however there were differences in the proportion of peaks binding each motif, and the third most enriched motif changed from Foxa2 in shLacZ to Esrrb in shWHshort cells, both in 10nM estrogen conditions. As shown in the figure, the proportion of peaks bound to EREs increased from 29.9% in shLacZ to 39.2% in shWHshort cells, and FOXA1 showed a similar increase from 49.7% to 57.5% of all ERa peaks in shLacZ and shWHshort, respectively. It is notable that FOXA1 was the most common motif enriched by ERa binding in SUM44 cells in all conditions. The proportion of peaks bound to promoters was also altered by WHSC1L1short knockdown. In SUM44 shWHshort cells in the presence of 10nM estradiol, promoter enrichment was reduced from 4.96-fold to 3.1-fold, indicating that WHSC1L1short knockdown resulted in a smaller, but still considerably large, proportion of ERa binding sites localized to promoter regions in SUM44.

3.3 Discussion

Testing for genome-wide ERa binding by ChIPseq in SUM44 cells resulted in detection of high levels of ERa bound to chromatin in the absence of ligand. This contrasts with the reported results of ERa ChIPseq experiments in other ER+ cell lines such as MCF7. Most ERa breast cancer biology is based on experiments conducted on the MCF7 cell line, which has been reported to have

relatively little ERa bound in the absence of estrogen. Our results in SUM44 cells indicate that ERa binding to chromatin is much higher in SUM44 cells, suggesting that ER+ breast cancers may have varying levels of ERa bound to chromatin in low-estrogen conditions. While the presence of ERa bound to chromatin in the absence of exogenous estrogen does not indicate positive regulation of gene expression by ERa, indirect evidence from the RNAseq expression profiling data show that several genes bound by ERa in the absence of estrogen were expressed at high levels in SUM44 cells. Further experiments should be done to determine whether the ERa bound near these genes is actually positively regulating expression in an estrogen-independent manner.

The observation that relatively few additional genes were bound by ERa after treatment with estrogen is unexpected, and suggests that most ERa in SUM44 cells is localized to the nucleus and bound to target regions on chromatin, and is activated by ligand binding, resulting in the recruitment of co-factor promoting transcription, or the dissociation of inhibitory factors, also resulting in promotion of transcription. Additional ChIPseq experiments for co-factors of ERa may result in elucidation of which mechanism is occuring at these sites in SUM44 cells. Overall, I conclude from these results that there is extensive binding of ERa in the absence of exogenous estrogen in SUM44 cells, and treatment of SUM44 with beta-estradiol results in increased binding intensity of the majority of the existing binding sites, with a relatively small number of additional binding sites observed after estrogen treatment.

DNA motif enrichment analysis at ERa binding sites resulted in enrichment of expected motifs, including ERE, FOXA1, ESRRB, and Foxa2. It is apparent when comparing the two treatment conditions that treatment with estrogen results in a large increase in binding intensity at ERE sites on chromatin. In the absence of estrogen, FOXA1 is the most abundant DNA sequence motif bound by the estrogen receptor.

While the difficulty in detecting a sufficient number of ERa binding sites in SUM44 shWHshort cells in the absence of estrogen leaves our investigation of ERa binding incomplete, the almost complete abrogation of ERa protein in SUM44 shWHshort cells provides allows for the reasonable assumption that there was in fact little to no estrogen receptor actually present in the cells to bind to chromatin, which would explain the negative result. While it is impossible to interpret a negative, this does seem to be a plausible explanation.

The successful measurement of ERa binding in SUM44 shWHshort cells in the presence of 10nM estrogen gives us some idea of the effect of WHSC1L1-short knockdown on ERa binding to chromatin, although interpretation of the results is difficult when taking into consideration the earlier results showing that WHSC1L1-short is necessary for ERa expression in SUM44, and that SUM44 proliferation responds negatively to treatment with even nanomolar amounts of estrogen. With that said, the DiffBind analysis of ERa binding in SUM44 shWHshort cells treated with estrogen compared to SUM44 shLacZ cells treated with 10nM estrogen suggests that WHSC1L1-short knockdown does not alter the

location of ERa binding, but does result in a genome-wide decrease in ERa binding intensity. This appears to be a logical and unsurprising result of the decrease in levels of ERa protein upon WHSC1L1-short knockdown.

Interestingly, there is evidence that an NSD-family member, WHSC1, is recruited to the ESR1 genomic locus by BRD3/4 in a breast cancer model cell line, and directly binds the ERa gene, promoting expression of ERa in these cells (Feng, Zhang et al. 2014). WHSC1L1 has high protein homology with WHSC1 (NSD2), and our data suggest that WHSC1L1-short may be affecting ERa expression in a similar way, although we lack direct evidence of WHSC1L1-short protein binding to the ESR1 locus. Also, WHSC1L1-short itself has been found to bind directly to the extra-terminal domain of BRD4 while also binding to chromatin in cancer cells (Rahman, Sowa et al. 2011).

3.4 Materials and Methods

3.4.1 Chromatin immunoprecipitation followed by Illumina sequencing

For ChIP-seq analysis of ERa binding sites, we followed procedures originally developed by Carroll, Brown and co-workers (Carroll, Liu et al. 2005, Carroll, Meyer et al. 2006) (personal communications with Tom Westerling). Cells were cultured in 150 mm plates to ~90% confluence (approximately $1x10^8$ cells). Cells were treated with 2 mM disuccinyl-glutamate (DSG) in PBS for 30 minutes at room temperature, then treated with 1% fresh formaldehyde for 2.5 minutes before quenching with 125 mM glycine for 5 minutes. Cells were washed with cold PBS, and chromatin was prepared using the Covaris High-Cell SDS

chromatin shearing kit with SDS buffer (520076) according to the manufacturer's recommended protocol. Chromatin was then sheared with a Covaris S220 sonicator using standard settings for 8 minutes. Chromatin preparation was optimized for SUM-44 cells.

ChIP was performed using the Magna ChIP A/G ChIP Kit (17-10085) with an antibody cocktail containing equal concentrations of the ERa antibodies from Thermo Scientific and Santa Cruz. Sequencing libraries from ERa ChIP DNA were prepared with Rubicon ThruPLEX librarypreparation kits, and sequenced as 35bp single-end reads on an Illumina Hi-Seq 2500. Libraries for three biological replicates were prepared and sequenced for each sample type. Read data for each sample was aligned with Bowtie2, and peakswere called with both the MACS version 2 and the homer tools FindPeaks peak callers.

To generate consensus peaksets for downstream analysis, the called peaksets from eachpeak caller for each sample were input into DiffBind, which created consensus peaksetsconsisting of peaks present in at least 2 replicates for each sample type. These consensus peaksets were used for downstream analysis including differential occupancy and differential binding.

3.4.2 RNA expression profiling

Total RNA was prepared using a Qiagen RNeasy Plus Mini Kit and processed by the MUSC Genomics core for 2x101 cycles, paired-end RNA sequencing on an Illumina HiScanSQ. RNA integrity was verified on an Agilent 2200 TapeStation (Agilent Technologies, Palo Alto, CA). A total of 100-200 ng of

total RNA was used to prepare RNA-Seq libraries using the TruSeq RNA Sample Prep kit following the protocol described by the manufacturer (Illumina, San Diego, CA). Sequencing was performed on an Illumina HiScanSQ. Samples were demultiplexed using CASAVA (Illumina, San Diego, CA). Fastq files were used to map reads to the human genome (hg19, UCSC) utilizing Tophat2 with default settings. A total of 100-200 ng of total RNA was used to prepare RNA-Seq libraries using the TruSeq RNA Sample Prep kit following the protocol described by the manufacturer (Illumina, San Diego, CA).

3.4.3 Bioinformatics and Statistical analysis

Each ChIPseq experiment was repeated three times. Fastq files were aligned to the UCSC hg19 human reference genome assembly using Bowtie2. Aligned reads were converted from SAM to BAM format, sorted and indexed. BED format files were obtained from BAM files. Quality control tools including FastQC and Homer makeTagDirectory were employed to verify read quality and the strength of the ChIP signal (Heinz, Benner et al. 2010). After QC, aligned reads for ChIP and input samples were analyzed using MACS version 2 to call peaks (Feng 2011, Feng, Liu et al. 2012). Peaks were also called with Homer's findPeaks and MACS version 1.4. Called peak sets were utilized for downstream analyses including motif enrichment analysis via Homer's findMotifsGenome, and gene annotation with Homer's annotatePeaks. Bedgraph files output by MACS v2 were converted to bigWig format for peak visualization on the UCSC Genome Browser. CEAS and Homer's annotatePeaks were used to determine which

genomic features were enriched under called peaks. Other tools were used for visualization and analysis of ChIPseq peakset data, including DiffBind to measure ERa peak occupancy and differential binding intensity (Ross-Innes, Stark et al. 2012). SeqMiner was used to perform unsupervised K-means clustering of peaks (Ye, Krebs et al. 2011). To generate a consensus peakset for a specific experimental condition, the DiffBind package was used to generate a peakset consisting only of peaks called in two or more replicates. Consensus peak sets were then used for differential occupancy and differential binding analyses using DiffBind. SeqMiner was used to perform unsupervised k-means clustering of peaks.

CHAPTER 4 HISTONE H3K4ME3 METHYLATION IS AFFECTED GLOBALLY AND NOT LOCALLY BY WHSC1L1 KNOCKDOWN

4.1 Introduction

WHSC1L1 has been reported to have histone methyltransferase activity on lysines 4, 9, and 36 on histone H3, however only the long isoform of WHSC1L1 contains a catalytic SET domain capable of catalyzing the addition of a methyl group to a lysine on a histone tail (add cite)[He13, Rahman11}. Considering our findings that both total and short isoform-specific WHSC1L1 knockdown resulted in changes in expression profiles in SUM44 cells, we considered it important to determine whether histone methylation levels are changing in response to knockdown of WHSC1L1-long or WHSC1L1-short in SUM44 shWH-total and shWH-short cells. Our initial aim was to investigate global methylation levels of lysines 4, 9, 27, and 36 on histone H3 by immunoblotting whole cell lysates with antibodies specific to mono-, di-, and tri-methylated lysines on histone H3. If we identified changes in specific methylation marks on specific lysines on histone H3 in SUM44 shWH-total or shWH-short cells compared to shLacZ control, our plan was to proceed with ChIPseq using the same ChIP-grade antibodies used for immunoblotting to determine genome-wide changes in methylation at high resolution.

4.2 Results

4.2.1 WHSC1L1-short knockdown increases global levels of histone H3

WHSC1L1 has been reported to methylate histone H3 on lysines 4, 9, and 36, although recent literature suggests that the main activity of WHSC1L1 is H3K36 dimethylation. However, understanding of WHSC1L1 substrates is still an active area of research. To investigate the mechanism of transformation by WHSC1L1, we hypothesized that knockdown of WHSC1L1 would alter histone H3 methylation levels. To test this, we looked for global changes in H3K36 and H3K4 methylation levels by western blot. Knockdown of WHSC1L1-short in SUM44 cells resulted in a large increase in total histone H3 levels, and a concomitant increase in H3K36me2 levels that was roughly proportional to the observed increase in total histone H3.

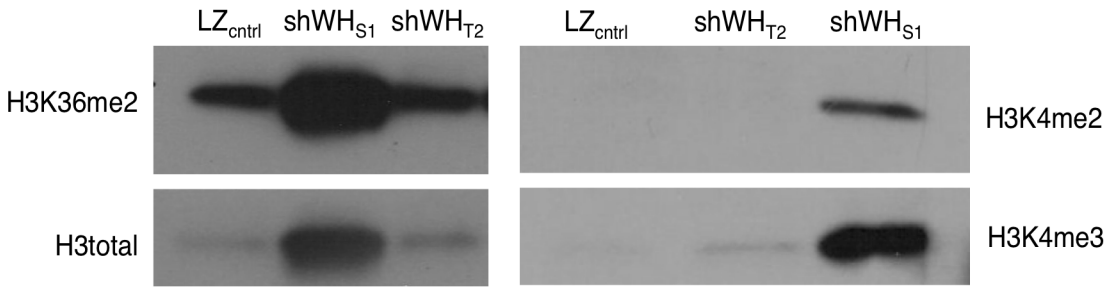


Figure 8.1 Global histone H3 methylation levels at lysines 4 and 36 in SUM44 shWHSC1L1short and shWHSC1L1total vs shLacZ control cells.

Immunoblot of histone H3 subunits H3K4me2, H3K4me3, H3K26me2, and H3K36me3 in SUM44 shWHSC1L1short and shWHSC1L1total knockdown cell lines compared to shLacZ control. A total histone H3 antibody was also used as a control for total H3 levels in each sample.

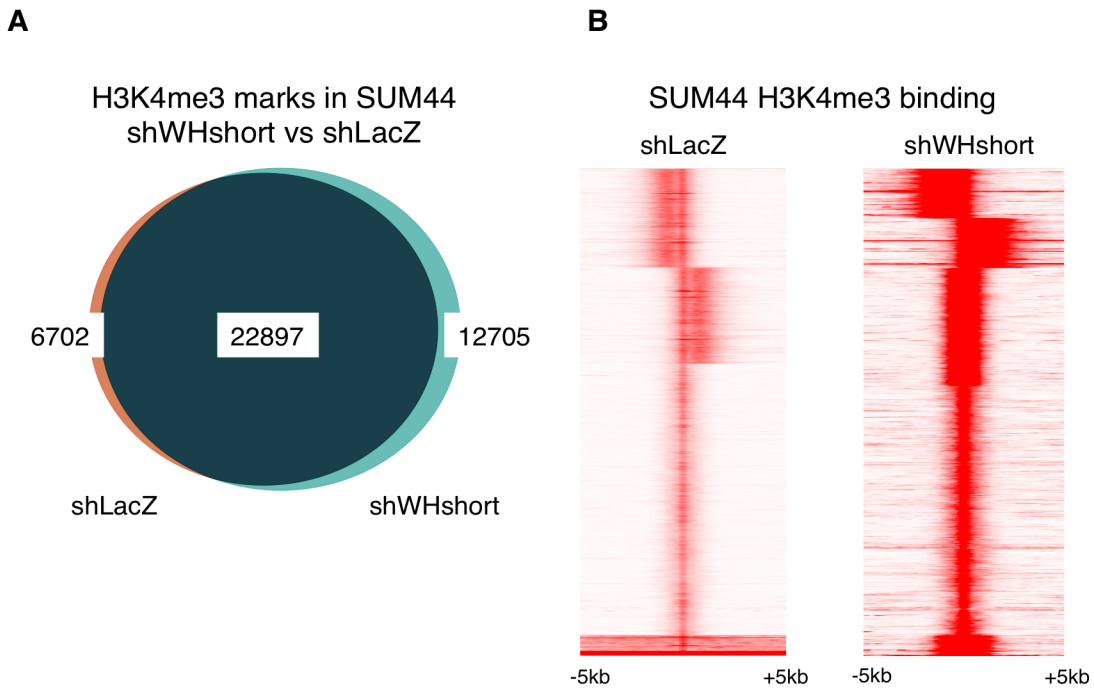


Figure 4.2 Histone H3K4me3 ChIPseq in SUM44 shWHSC1L1short vs shLacZ cells.

a. Venn diagram comparing histone H3 lysine 4 trimethyl (H3K4me3) sites in SUM44 shWHSC1L1short vs shLacZ control cells. **b.** Heatmaps of genomic regions marked by H3K4me3 in SUM44 shWHSC1L1short vs shLacZ control cells. The heatmap x-axis origin is centered on the summit of the H3K4me3 peak, and shows the relative intensities of the peaks over a 10kb genomic window.

The SUM44 shWHtotal cells did not show a consistent increase in either total histone H3 or levels of the H3K36me2 mark. Histone H3K4me2 and H3K4me3 levels also increased in the shWHshort knockdown cells, but not the shWHtotal cells (Figure 4.1, right panel). Expression profiling corroborates the western data, showing an increase in histone levels in the shWHshort cells; the transcript levels of for at least one gene for each histone subunit increased upon knockdown of WHSC1L1-short (Table 4.1).

gene symbol	$\log_2$(fold change)	FDR (B-H)
HIST1H4H1	1.538	0.052
HIST1H3H	1.524	0.075
HIST1H2BN	1.508	0.033
HIST1H3G	1.454	0.074
HIST1H1C	1.075	0.055

Table 4.1: Histone Transcripts significantly differentially expressed in SUM44 shWHSC1L1short cells vs. shLacZ control.

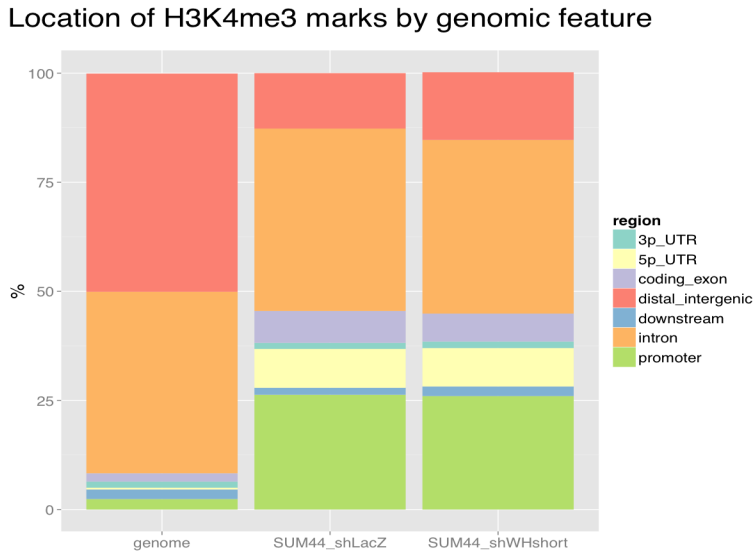
4.2.2 H3K4me3 ChIPseq

Given the observed increase in global H3K4me3 protein levels in SUM44 shWHshort cells, we wanted to investigate H3K4me3 patterns in SUM44 cells, and then look for changes in H3K4me3 patterns in SUM44 WHSC1L1-short cells vs shLacZ control. H3K4me3 was measured in SUM44 shWHshort cells and shLacZ cells by ChIPseq, with two biological replicates per condition. We observed a total of 29599 peaks in SUM44 shLacZ cells, and a total of 35602 peaks in SUM44 shWHshort. As shown in Figure 4.2a, the majority of the H3K4me3 sites detected in SUM44 cells were common to both shLacZ and shWHshort conditions. It should be noted that the mean peak intensities of the peaks unique to either the shWHshort or shLacZ cells were 4.7 and 4.4, respectively, while the overall mean peak intensity was 27.5, indicating that the unique H3K4me3 sites were in general not as strong as the sites common to both conditions. The heatmaps showing all H3K4me3 sites detected in SUM44 shLacZ and SUM44 shWHshort cell lines (Figure 4.2b) show similar but not identical patterns of H3K4me3 sites.

Having identified the sites of H3K4 trimethylation in SUM44 shLacZ and shWHshort cells, we next asked which genomic features were most enriched by H3K4me3 sites for each condition in SUM44 cells. We used the CEAS tool

released by Liu et al to annotate peaks for proximal genomic features based on peak location. The majority of H3K4me3 sites were primarily near transcription start sites (TSS) of genes (Figure 4.3a, Figure 4.3b). While intronic regions were enriched slightly more than promoter regions, Figure 4.3b shows that this is a result of the broad H3K4me3 peak extending from the promoter into the first intron. The profile plot shows clearly that the average H3K4me3 site is a bi-modal region approximately 2kb in width, with a "notch" of low intensity immediately upstream of the TSS. It should be noted that very few H3K4me3 sites were detected in distal intergenic regions, and also that the breakdown of genomic features with H3K4me3 marks is very similar in SUM44 shWHshort and shLacZ cell lines.

A



B

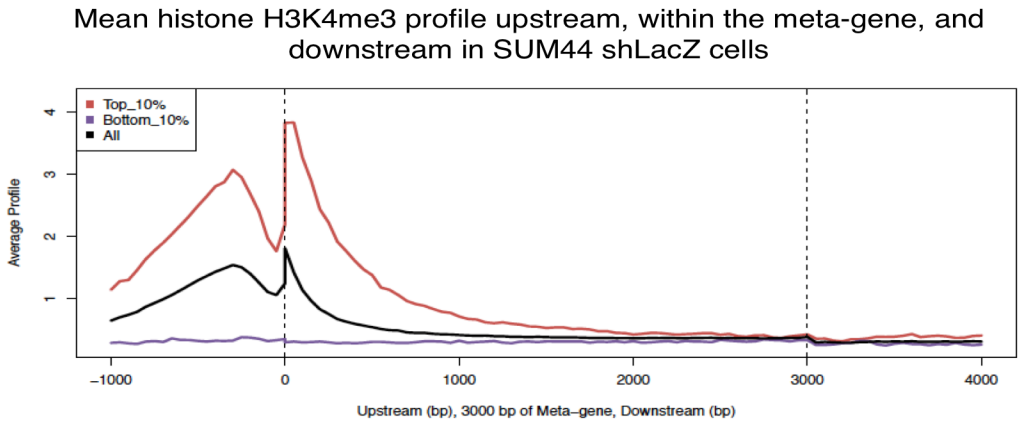


Figure 4.3 Genomic feature localization of histone H3K4me3 marks in SUM44 cells.

a. Distribution of H3K4me3 marks across different genomic features in SUM44 shWHSC1L1short and shLacZ cells, with the first column showing the proportion of the human genome each feature represents. H3K4me3 genomic locations determined by ChIPseq **b.** Histone H3K4me3 mean intensity profile of H3K4me3 sites across a map of a gene, from 1kb upstream of the TSS to 4kb downstream. The 3 profiles represent the mean intensity of all H3K4me3 sites, and the blue and red curves represent H3K4me3 binding for the bottom 10% and top 10% of expressed gene loci, as measured by ChIPseq and RNAseq in SUM44 shLacZ cells.

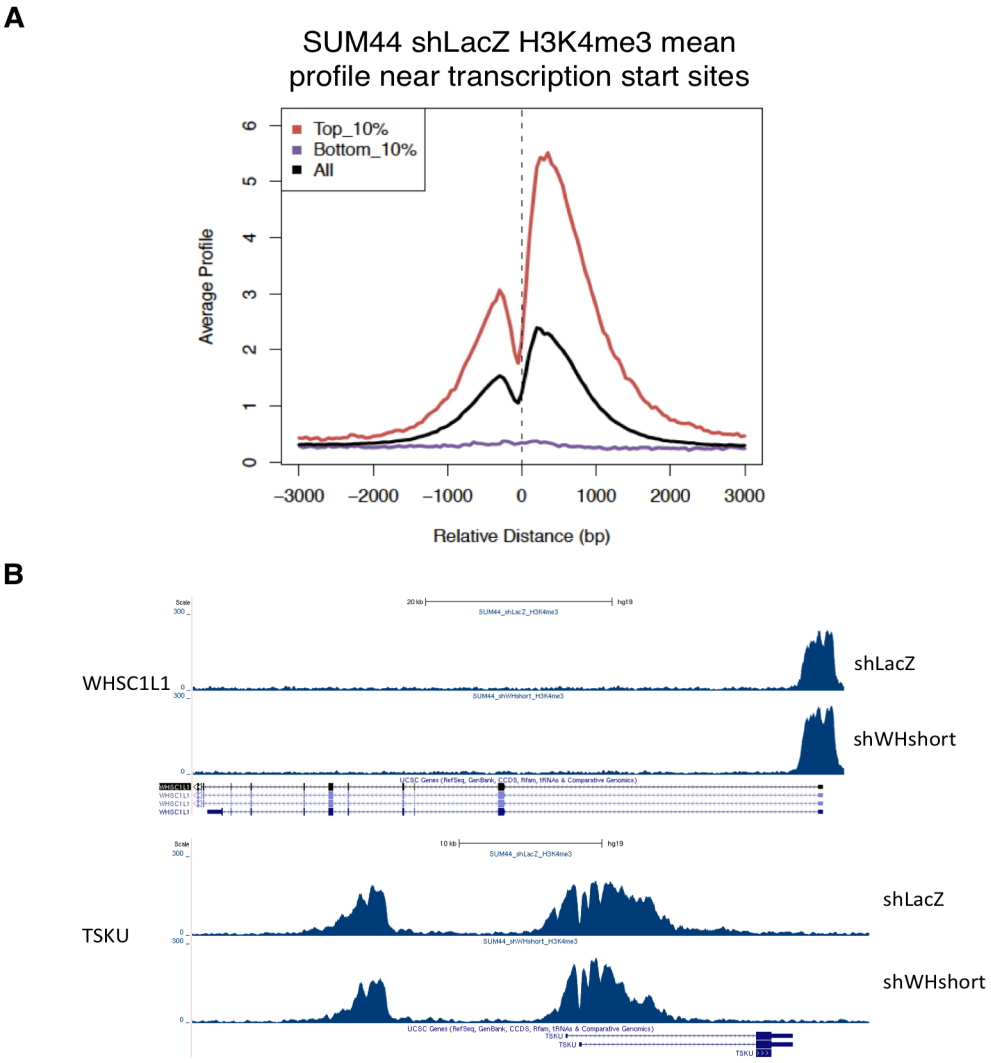


Figure 4.4 Histone H3K4me3 patterns and mean intensity profiles for SUM44 shWHSC1L1short vs shLacZ cells.

a. Histone mean occupancy profile plot for trimethylated histone H3 subunits in SUM44 shLacZ cells. The black trace shows the mean profile for all H3K4me3 marks genome-wide, and the red and blue marks represent H3K4me3 profiles for the top 10% and bottom 10% of expressed genes, respectively. H3K4me3 occupancy measured by 25bp single-end ChIPseq, and gene expression measured by 151bp paired-end RNAseq and absolute counts in SUM44 shLacZ. Profiles are representative of duplicate experiments, and plot data is from one of the experiments. **b.** Gene tracks showing H3K4me3 peaks (peaks shown using a wiggle file generated using the MACS2 peak caller). Profiles are from one experiments and are representative of duplicate experiments.

Trimethylation of histone H3 has been shown to be a mark of actively transcribed chromatin. The mean peak profile data together with the RNAseq expression profiling data we collected on the SUM44 shWHshort and shLacZ cell lines allowed us to determine the mean H3K4me3 peak profiles for groups of genes based on their expression by RNAseq (Figure 4.4a). Based on absolute expression (tags per million mapped reads) from RNAseq data, we took the top 10% of expressed genes and the bottom 10% of expressed genes, and used these two gene sets to group the gene-annotated H3K4me3 ChIPseq data. The mean H3K4me3 peak profile for the top 10% of expressed genes in SUM44 shLacZ cells (the red line) is similar in shape to the overall mean peak profile (black line), although the intensity is markedly greater. Conversely, for the bottom 10% of expressed genes, the profile is flat (blue line), indicating that there is little to no H3K4me3 at or near the promoters of the least expressed genes in SUM44 shLacZ cells. The results for SUM44 shWHshort cells are similar (data not shown).

After determining that we had successfully detected H3K4me3 sites in SUM44 shLacZ and shWHshort cells, we next asked whether there were any H3K4me3 sites in SUM44 shLacZ cells that were significantly different in intensity in the shWHshort cells. To measure differential H3K4me3 intensity, we ran DiffBind comparing the SUM44 shWHshort H3K4me3 ChIPseq data to SUM44 shLacZ, however only 2 significantly differentially intense sites were

found (no MA-plot shown). A 1.7-fold increase was identified near the TSS of DST in SUM44 shWHshort, which was upregulated 1.8-fold in shWHshort cells by RNAseq, and a 1.4-fold increase in H3K4me3 near the TSS for BCAR3, which was not significantly differentially expressed. Figure 4.4b shows the H3K4me3 profile for WHSC1L1 and TSKU, two genes significantly downregulated in SUM44 shWHshort cells. While these are just two examples out of tens of thousands, the profiles shown are representative of the majority of the observed peaks, and it can be seen qualitatively that there is little difference in H3K4me3 pattern or intensity in SUM44 shWHshort cells compared to SUM44 shLacZ controls.

4.3 Discussion

Existing literature reports several substrates of WHSC1L1 activity on histone H3, including H3K4, H3K9, and H3K36. Recent publications focus on H3K36 as the primary substrate for WHSC1L1 methylation, however we wished to also investigate H3K4 as a substrate, as relatively little is known about WHSC1L1 activity. Our initial investigation into changes in global histone methylation levels resulted in an unexpected change in not only histone methylation, but in the global levels of total histone H3, as seen by the large increase in global histone H3, H3K36me2, H3K36me3, and H3K4me3 in SUM44 shWHshort cells compared to shLacZ control. Firstly, the fact that methylation of both lysines 4 and 36 increased in response to knockdown of the short isoform of WHSC1L1 is intriguing as it may support the hypothesis that WHSC1L1-short is

acting in a dominant-negative fashion by preventing methylation of these histone lysines by WHSC1L1-long. Full interpretation of the effects of WHSC1L1 knockdown on global histone methylation is difficult with the models used, as the effect of WHSC1L1-long knockdown is confounded by the concomitant knockdown of WHSC1L1-short in the shWHtotal SUM44 cells.

While the observed increase in histone lysine methylation upon knockdown of WHSC1L1-short is intriguing, the global change in total histone H3 calls into question the interpretation of the western data as an actual relative increase in histone lysine methylation at the tested residues.

Together with the RNAseq expression data showing a significant increase in at least one transcript for every histone subunit upon knockdown of WHSC1L1-short, it appears that the overall histone levels incresed in the SUM44 shWHshort cells. Whether the relative levels of histone H3K36me2, H3K36me3, or H3K4me3 increased is inconclusive at best based on the western data alone. In our attempt to identify specific H3K4me3 binding sites, we were able to successfully detect a large number of H3K4me3 sites in both SUM44 shLacZ and SUM44 shWHshort cells by ChIPseq. This is likely due to the quality of the "ChIP-grade" antibody available for this histone methyl mark, as well as the relatively narrow pattern of H3K4 methylation, in contrast to the much wider patterns characteristic of other histone methylation marks, such as H3K36 di- or trimethylation. The genomic features found in proximity to observed H3K4me3 marks was consistent with published literature on H3K4me3: almost all marks

were found in the vicinity of gene loci, mostly within 5kb of the TSS, with a characteristic bi-modal pattern for many peaks located immediately upstream of the TSS. At the gross level, i.e. the number of H3K4me3 sites found, the mean profile of the marks, and the distribution of the genomic features where the peaks were located, there was surprisingly little difference between the SUM44 shWHshort and the SUM44 shLacZ samples. This is consistent with the interpretation of the western data as a global increase in total histone levels in the SUM44 shWHshort cells, and not an increase in histone lysine methylation at lysine 4.

In addition to the high levels of similarity in the number, peak profile, location, and enrichment of genomic features in the SUM44 shWHshort and shLacZ H3K4me3 ChIPseq data, the DiffBind differential binding intensity results indicated essentially no difference in H3K4me3 intensity between the two conditions. While 2 sites were detected as significantly increased in SUM44 shWHshort cells relative to shLacZ, only one of them, the DST gene, exhibited a change in expression: a 1.8-fold increase concomitant with the slightly incresed H3K4me3 intensity, from 25.01 to 32.5. Inspection of individual loci significantly differentially expressed in SUM44 shWHshort cells relative to shLacZ control shows qualitative similarity in the patterns and intensities of H3K4me3 peaks for these genes, including WHSC1L1, TSKU, ESR1, and others.

Based on the western data, together with the qualitative and quantitative results of the H3K4me3 ChIPseq analyses, I conclude that WHSC1L1

knockdown has no significant effect on the histone methylation patterns or intensity in SUM44 cells. While the results for these experiments looking for a change in H3K4me3 were negative, the results remain important in terms of ruling out lysine 4 on histone H3 as a substrate for WHSC1L1.

Future work could include investigation of H3K36me2 ChIPseq to determine whether and where lysine 36 on histone H3 is differentially methylated upon knockdown of WHSC1L1-short. I failed to collect useful data for H3K36 methylation by ChIPseq after several attempts with different experimental conditions. While I tried only one antibody, it was sold as ChIP-grade, and there were successful publications using the antibody for their ChIPseq work. I estimate that the two major factors that prevented a successful H3K36me2 ChIPseq dataset were the lack of proper wash conditions when performing the ChIP, and possible difficulty in tuning the peak calling software to accurately call broad peaks. It should be noted however that I did not achieve sufficient signal-to-noise ratios with my ChIP for the peak calling to be truly tested.

A final point that needs to be made about these conclusions is that, given the negative result for H3K4me3 and the lack of results for a different histone lysine, I was not able to show with any certainty that knockdown of WHSC1L1 in SUM44 cells results in net changes in the proportion of histone H3 with methylation of one or more lysines. This leaves open the possibility that WHSC1L1-short is acting in a histone-independent manner to regulate expression of ESR1 and other target genes. While the data I have collected

showing strong nuclear localization of WHSC1L1 in several breast epithelial and breast cancer cell lines is suggestive of a nuclear protein as a target, it remains unclear what the primary mechanism of transformation by WHSC1L1 overexpression might be. However, recent work in mouse leukemia cells does show that WHSC1L1-short is likely acting in a complex with BRD4 to recruit CHD8 to chromatin, resulting in chromatin conformation changes due to helicase activity and not histone methylation (Shen, Ipsaro et al. 2015).

4.4 Materials and Methods

4.4.1 Global histone methylation western blotting

Experiments to determine global histone methylation levels were performed by lentivirally transducing SUM44 cells to generate SUM44 shLacZ, shWHtotal, and shWHshort cells as described earlier (p. 36-37), then collecting whole cell lysates after 72 hours of selection with puromycin. Whole cell lysates were loaded on 5-15% polyacrylamide gradient gels in SDS-based loading buffer containing beta-mercaptoethanol (reducing conditions), and run for 35 minutes at 150V according to the manufacturer's protocol (Bio-Rad). The gels were transferred to PVDF membranes and immunoblotted with the following antibodies: H3K4me2, H3K4me3, H3K36me2, and H3K36me3.

4.4.2 Chromatin Immunoprecipitation followed by Illumina sequencing (ChIPseq)

ChIPseq was performed as described above with some differences. SUM44 shWHshort and shLacZ cells were cultured to ~90% confluence, with puromycin selection for 72 hours post-infection with lentivirus to transduce the cells with the appropriate shRNA vector as described above. On the day of the experiment, the cells were fixed with 1% fresh formaldehyde for 2.5 minutes (no disuccinyl glutamate treatment) before following the Covaris SDS-based chromatin shearing protocol as described. Sheared chromatin equivalent to approximately $1x10^6$ cells was loaded into each ChIP reaction in 1mL total volume, and 2 ug of H3K4me3 ChIP-grade antibody was used to immunoprecipitate the chromatin overnight at 4° Celsius. The wash steps also differed from the ERa ChIPseq protocol: after the magnetic Protein A/G beads were collected for each wash, the beads were washed once each with Low-Salt, High-Salt, and Lithium Chloride ChIP, and then two times with TE, all buffer solutions from the Millipore Magna-ChIP kit (Millipore).

CONCLUSIONS

The work presented here has been an attempt to satisfy four general aims. First, we sought to confirm that WHSC1L1-short is the specific WHSC1L1 isoform responsible for maintenance of a tumor phenotype in SUM44 breast cancer cells. While previous work by Yang and others in the Ethier lab had previously found that WHSC1L1-short appeared to be the WHSC1L1 isoform responsible for inducing a transformed phenotype in MCF10 cells, we have now shown that shRNA knockdown of WHSC1L1-short transcript in SUM44 cells does in fact have the greatest negative effect on SUM44 cell proliferation, suggesting the WHSC1L1-short is in fact the oncogenic isoform of WHSC1L1 in SUM44 cells (Yang, Liu et al. 2010).

Second, we sought to observe transcriptome-level changes in gene expression in SUM44 shWHSC1L1-short cells relative to SUM44 shLacZ controls to determine whether knockdown of WHSC1L1-short, which lacks a catalytic domain, would significantly alter gene expression. As shown, we observed significant changes in gene expression in several hundred genes in SUM44, with an overall trend of upregulation upon WHSC1L1-short knockdown.

Third, based on our observations that SUM44 expresses high levels of ERa, and that knockdown of WHSC1L1-short significantly downregulated expression of ERa transcript in SUM44, we sought to measure ERa binding to

chromatin in both SUM44 shLacZ (negative control) and SUM44 shWHSC1L1-short cells. Our observations that ERa is bound to chromatin at thousands of sites in SUM44 cells in the absence of estrogen was an unexpected finding, and contrasts with the findings of other groups measuring ERa.

Fourth, based the known activity of the long isoform of WHSC1L1 as a histone methyltransferase, we asked whether knockdown of WHSC1L1-short in SUM44 cells would alter histone lysine methylation patterns in SUM44 cells. Most literature on WHSC1L1 activity indicates that lysine 36 on histone H3 is the substrate for methylation by WHSC1L1 (di Luccio and Morishita 2012, Saloura, Vougiouklakis et al. 2016). However, we chose to investigate additional histone methyl marks that had been previously reported, including H3k4me3. While the results of the experiment testing whether WHSC1L1-short knockdown altered histone H3K4me3 levels in SUM44 cells did not result in any significant difference between the treatment and control samples, we did benefit by learning about the locations of histone H3K4me3 sites genome-wide in SUM44 cells. It should be noted that attempts were made to ChIP additional histone lysine methylation marks, including H3K36me2 and H3K36me3, however all experiments failed to achieve a sufficiently strong signal-to-noise ratio by ChIP, most likely due to technical errors and my failure to find the correct conditions for ChIP. While there was little change in H3K4me3 levels between the shWHshort and shLacZ cells, the occupancy analysis appears to agree with the western data in that the number of genomic loci with tri-methylated H3 increases by

approximately 6000 peaks. Figure 4.2a shows that the occupancy analysis found over 12000 areas with H3K4me3 in shWHshort cells that were not present in shLacZ. Also, analysis of the distribution of H3K4me3 levels in shWHshort vs shLacZ cells suggests that WHSC1L1short knockdown increases weak H3K4me3 levels, while initially strong levels of H3K4me3 decrease. I have not yet figured out how to show whether this apparent change is statistically significant, but it may be an interesting finding, if in fact WHshort knockdown is causing more widespread H3K4me3 marks genome-wide, but at the same time a slight decrease in H3K4me3 levels at some promoters with high levels in shLacZ cells. This is slightly at odds with our claim that there is no significant decrease in H3K4me3 levels at promoters for genes significantly downregulated by WHSC1L1short knockdown.

As mentioned in the introduction, researchers contributing to TCGA found that copy number alterations, specifically large amplifications, are the major mechanism of driving mutations in breast cancer at over 75% (Ciriello, Miller et al. 2013). This is strong evidence that changes in gene copy number are the primary mechanism of oncogene activation and tumor-suppressor inactivation in breast cancer. The Ciriello TCGA study identified Based on TCGA array CGH data, the amplifications most commonly found co-amplified with WHSC1L1 amplifications are chromosomes 1, 16, 11, 17, and 20, in descending order of coincidence, occurring in 82%, 78%, 66%, 61%, and 60% of WHSC1L1-amplified individuals. The stark structural differences in loss of genomic integrity between

function-altering point mutations common in colon and uterine cancers and copy number changes found in cancers such as breast cancer raise an important question: do cancers driven by copy number changes rely on the activation of a greater number of oncogenes to maintain an oncogenic phenotype than cancers driven by point mutations? Understanding whether there are multiple driving oncogenes in an amplified region, or rather there is a single driving oncogene selected for in a large region surrounded by non-oncogenic “passenger" genes incidental to the mechanism of mutation is an important question. Three reasons that understanding whether multiple genes in an amplicon (or across multiple amplicons) interact to drive cancer are amplicon formation, strategies for successful treatment of the primary tumor, and prevention of recurrent disease refractory to the initial treatment. With exceedingly few exceptions, such as the Gleevec for chronic myelogenous leukemias with Bcr-Abl fusions, a unifying truth of targeted cancer therapies (and cancer therapy in general) is that a drug-resistant population of tumor cells will emerge and cause cancer recurrence at a probablility approaching one, given a large enough time scale. A characteristic feature of these recurrent tumor cells is the presence of a different driving oncogene or set of oncogenes from those driving the initial primary tumor.

WHSC1L1 knockdown reduces SUM44 cell proliferation

As shown in Figure 2.8, total knockdown of WHSC1L1 in SUM44 cells decreases proliferation, which WHshort knockdown results in a still greater decrease in proliferation. Because we did not investigate knockdown of the long

isoform of WHSC1L1 alone, it is possible that the milder decrease in cell proliferation observed in SUM44 shWHtotal cells was due to the decrease of WHshort, and that the greater decrease of WHshort in the shWHshort SUM44 cells resulted in the greater decrease in proliferation observed. This cannot be determined conclusively without selective knockdown of WHlong alone, but the observed decreases in proliferation are proportional to the level of WHshort knockdown.

The effects of WHSC1L1 knockdown on transcript expression

Knockdown of WHshort in SUM44 shWHshort cells resulted in differential expression of both a greater number of different gene transcripts and larger differences in transcript levels than knockdown of WHlong and WHshort together in shWHtotal SUM44 cells. The observed 10-fold difference in the number of significantly differentially expressed genes (1130 vs 130 genes in shWHshort vs shWHtotal cells, respectively) suggests that transduction of SUM44 cells with an shRNA that knocked down both isoforms of WHSC1L1 dampened the effect of knockdown of each isoform individually, and that the effect of each WHSC1L1 isoform on transcriptional regulation may be in opposition. This provides additional indirect support for our hypothesis that the short isoform of WHSC1L1 may be acting in a dominant-negative fashion toward the long WHSC1L1 isoform. However, as seen in Figure 2a, and as seen in the TCGA RNAseqV2 BrCa data, the short isoform of WHSC1L1 is almost always expressed at higher levels than the long, not only in the cancer cells, but also in the normal matched

samples collected by TCGA to complement the tumor samples. If WHshort is the predominant isoform in normal tissue, the dominant-negative model for antagonism of WHlong activity becomes a poorer fit. Also, it should be mentioned that transcript level does not always correlate precisely with protein activity, which is not yet well characterized at a qualitative level for WHSC1L1, let alone with any quantitative accuracy in multiple tumor and normal tissue types. The fact that multiple WHshort protein molecules are likely present for every one molecule of WHlong suggests that perhaps two or more WHSC1L1short molecules complex with a single WHSC1L1long molecule, together with other chromatin modifying enzymes to form a “reader” complex that is capable of recognizing and binding a more diverse set of chromatin states than would a complex containing only a single WHlong molecule. This model would suggest that the catalytically inactive short isoform of WHSC1L1 is important for binding of WHlong to sites where it can methylate histone H3, and that WHlong binding to chromatin may be inefficient or less diverse in the absence of WHshort. These hypotheses could be tested by investigating protein-protein interactions between WHlong and WHshort, as well as measuring differential binding of WHlong to chromatin by ChIPseq in SUM44 shWHshort cells compared to shLacZ controls. While direct detection of the specific isoform bound to chromatin by ChIP would not be trivial, the existing model for knockdown of WHshort alone together with shLacZ control would serve as a good system for measuring the effect of WHshort levels on WHSC1L1 binding to chromatin. Interestingly, we have shown that WHlong

protein levels are increased in SUM44 shWHshort cells compared to shLacZ controls. It could be that the increase is an attempt at compensation for the lost chromatin binding capacity due to the absence of WHshort. Protein-protein interaction assays to detect binding of WHlong to WHshort would be essential to support this hypothesis of WHshort aiding WHlong binding to chromatin.

WHSC1L1-short expression is necessary for ERa expression in SUM44 cells

Remarkably, expression of estrogen receptor alpha in the highly ER+ SUM44 cell line is dependent on overexpression of the short isoform of WHSC1L1. As shown, shRNA knockdown of WH-short and total WHSC1L1 both resulted in reduction of ERa expression at the transcript and protein levels. This suggests that the amplification of WHSC1L1 as part of the 8p11-p12 amplicon is necessary for overexpression of ERa in breast cancers with amplified WHSC1L1.

ER binds chromatin extensively in SUM44 estrogen-free cells

A primary finding of our research was that the estrogen receptor binds chromatin extensively in SUM44 breast cancer cells in the absence of estradiol or other agonist. This observation, together with the observation that many ER target genes known to be expressed in ER+ breast cancer were actively transcribed in SUM44 in estrogen-free conditions, indicates ERa is both bound to EREs and is likely activating transcription of ER target genes in an estrogen-independent manner in SUM44 cells. ERa activity at these sites in estrogen-free conditions should be confirmed by a reporter assay consisting of a vector

containing an estrogen response element (ERE) upstream of a reporter gene to attempt to measure agonist-free ERa activity in SUM44 cells. Knockdown of ERa with shRNA or siRNA vectors should also One key question that arises from these observations is whether WHSC1L1 amplification and overexpression in SUM44 results in a genome-wide shift toward euchromatin, increasing the number of exposed EREs and increasing the space in general for binding of ERa and other associated pioneer factors such as FOXA1 that preceed ERa binding to chromatin. To answer this question, a potentially important future experiment would be investigation of the effect of WHshort knockdown on chromatin conformation genome-wide (Hi-C) and at specific loci (3C or 4C).

While I did not perform experiments which were able to measure WHSC1L1-short chromatin binding directly, recent work has shown that WHSC1L1-short links BRD4 and CHD8 by forming a complex, recruiting CHD8 to chromatin. Residues C-terminal to the single PWWP domain on WHSC1L1-short bind the ET domain of BRD4 while the N-terminal end binds CHD8 (Rahman, Sowa et al. 2011, Shen, Ipsaro et al. 2015, Zhang, Zeng et al. 2016). While their work was performed in murine leukemia cells, they did show that this mechanism was required for tumor cell proliferation.

CHD8 is a helicase, and has been shown to act as a transcriptional repressor (Clapier and Cairns 2009, Ho and Crabtree 2010). This is consistent with our observation of a net increase in gene expression upon knockdown of WHSC1L1-short in SUM44 cells, and importantly, there is now direct evidence

for a model of chromatin modification modulated by the catalytically inactive short isoform of WHSC1L1, and it does not apparently involve inhibition of WHSC1L1-long histone methylation activity. It will be very interesting to map WHSC1L1-short vs WHSC1L1-long binding sites in SUM44 and other breast cancer cell lines to determine whether the BRD4/WHSC1L1-short/CHD8 complex binds the ESR1 locus directly. One hypothesis would be that upregulation of ERa expression by WHSC1L1-short expression is indirect, as the direct activity of WHSC1L1-short in this model is transcriptional repression.

Final Conclusions

In this dissertation I have presented work that focused on the role of WHSC1L1 overexpression in maintenance of a cancer phenotype in SUM44 breast cancer cells, particularly the relationship between WHSC1L1 overexpression and expression of the estrogen receptor. Significant findings include confirmation that overexpression of the short isoform of WHSC1L1 is necessary for maintenance of a cancer phenotype in SUM44 breast cancer cells, and that WHSC1L1-short is required for estrogen receptor alpha expression at the transcript and protein level in SUM44. Others have recently shown that the short isoform of WHSC1L1 acts as an adapter protein creating a complex including BRD4 and CHD8, two enzymes involved in chromatin remodeling (Shen, Ipsaro et al. 2015). WHshort knockdown has a greater effect on SUM44 cell growth, overall differential gene expression, and ERa expression than WHtotal knockdown. The low number of significantly differentially expressed

genes in the shWHtotal cells is likely due to the fact that the shWHtotal vector reduced WHlong levels much more than WHshort; specific knockdown of WHshort resulted in an approximately 5-fold increase in the number of differentially expressed genes. This evidence that the short isoform of WHSC1L1 has a greater effect on gene expression than the long isoform may provide clues about the function of WHSC1L1-short; we hypothesize that the short isoform is more promiscuous, able to complex with many different chromatin modifying proteins to form complexes with more diverse chromatin-modifying activity than the long isoform. As WHSC1L1-short is not able to directly catalyze methylation of histone lysine residues, its function may be instead to recruit several different chromatin modifiers with its single PWWP domain, perhaps with different activities. Also, while recent evidence shows WHSC1L1-short acting as a link between BRD4 and CHD8 to promote helicase activity, it may still be possible that WHSC1L1-long and WHSC1L1-short antagonize each other, and that knockdown of both isoforms results in a regulatory state where the effect of knockdown of either isoform alone is attenuated by the concomitant knockdown of the other. Positive determination of the effect of one isoform of WHSC1L1 on the other would require further experiments, including WHSC1L1 ChIPseq to first determine whether both isoforms of WHSC1L1 the same genomic loci. The global increase in histone H3 levels observed in the shWHshort cells may be explained by the buildup of G2-arrested cells after WHSC1L1-short knockdown. This may also be consistent with the apparent contradiction of slowed

proliferation together with positive enrichment of mitotic cell cycle and DNA replication KEGG pathways in exhibited by the shWHshort cells. The observation that alternate isoforms of H3 are upregulated upon knockdown of WHshort suggests that alternate isoforms of H3, such as H3.3, may be substituted for H3.1 in chromatin. For future work, chromatin conformation could be investigated by DNase-seq hypersensitivity assay or Formaldehyde-Assisted Isolation of Regulatory Elements followed by short read sequencing (FAIRE-seq) to determine both the overall state of chromatin in the cell and specific regions of altered chromatin conformation. One hypothesis to explain the lack of expected clustering by replicate and condition of the SUM44 shLacZ and shWHshort cells based on ERa occupancy, together with the observed identical mean peak profile shape and location in both shLacZ and shWHshort, indicate that the location of ERa binding sites do not change after WHSC1L1-short knockdown, either in the absence or presence of ligand. However, the change in mean peak profile intensity combined with the expected clustering of samples by condition when clustered by differential binding intensity is consistent with the significant reduction in ERa protein levels in the shWHshort cells, and indicates there is simply less ERa available to bind chromatin.

APPENDIX SUM44 RNA-SEQ DIFFERENTIAL EXPRESSION RESULTS

Table A.1 SUM44 WHSC1L1short shRNA knockdown differential gene-level transcript expression by RNAseq

HGNC_gene_symbol	log2fold-change	p-value	FDR (q-value)
NELL2	4.21243	3.80E-18	2.74E-14
CEND1	3.49112	3.79E-11	6.07E-08
CNIH2	3.47971	1.12E-13	2.69E-10
RRM2	2.83998	5.67E-09	3.89E-06
LRRK2	2.80148	3.66E-22	5.28E-18
VEPH1	2.73812	2.44E-14	7.03E-11
DAB2	2.71662	1.75E-07	5.15E-05
STAB2	2.69528	1.17E-07	4.12E-05
RIBC2	2.63562	4.19E-10	5.49E-07
E2F7	2.59767	1.92E-07	5.15E-05
PCLO	2.56236	4.15E-08	1.92E-05
MSRB3	2.44857	9.07E-08	3.35E-05
RNF128	2.4245	2.15E-05	0.00220184
BPIFA4P	2.38396	0.000135738	0.00854558
E2F2	2.35443	1.63E-09	1.57E-06
GFRA2	2.29534	0.000222107	0.0116441
SACS	2.29352	8.53E-08	3.24E-05
FAM169A	2.24946	2.38E-06	0.000429159
STRA6	2.24448	3.34E-05	0.00305027
C10orf114	2.22678	1.67E-07	5.13E-05
CLSPN	2.21416	3.56E-08	1.77E-05
TRIP13	2.21166	1.24E-17	5.97E-14
HELLS	2.1998	8.23E-08	3.21E-05
ETV1	2.17921	3.44E-05	0.00312274
PLS3	2.17862	7.95E-05	0.00570208
CDCA7	2.15269	1.82E-07	5.15E-05
MCM10	2.15252	5.14E-06	0.000772727
COL12A1	2.14029	4.97E-16	1.79E-12
AMOTL1	2.12106	0.000218147	0.0115662
RAPGEF4	2.11885	0.00017222	0.00989204
HAS3	2.11234	0.00061495	0.0233309

ANKRD1	2.10533	0.000534165	0.0215644
TYMS	2.10505	1.36E-06	0.000280438
TK1	2.09423	3.53E-11	6.07E-08
CORO1A	2.08402	7.00E-08	2.80E-05
CHRNA5	2.06679	2.12E-06	0.000391642
CACNA2D1	2.06284	4.49E-07	0.000102795
FBXO43	2.05454	0.000128467	0.0081883
FAM64A	2.04838	7.32E-05	0.00538199
GHR	2.03573	9.15E-10	1.02E-06
ASF1B	2.03353	1.24E-07	4.26E-05
RAD51	2.02363	1.34E-05	0.00158828
DHFR	2.02285	4.26E-08	1.92E-05
FAM111B	2.02145	5.30E-06	0.000788439
BPIFB1	2.00671	0.000146576	0.00891642
B3GNT6	2.00413	0.000612379	0.0232946
E2F8	1.99529	2.47E-08	1.27E-05
ORC1	1.99393	4.27E-06	0.00069893
SPC25	1.98276	1.83E-06	0.000361763
RAD51AP1	1.96965	5.71E-08	2.42E-05
NPAS3	1.96294	0.000196925	0.0108858
ALOXE3	1.96017	4.91E-08	2.14E-05
ATP10D	1.9431	5.48E-06	0.000806596
NR4A3	1.93259	0.000492655	0.0203234
ESCO2	1.93069	1.66E-05	0.00185966
MNS1	1.93012	4.12E-05	0.00364282
KIF18B	1.92156	0.000103273	0.00695743
GTSE1	1.9193	4.18E-05	0.00367802
CDC45	1.90854	1.45E-06	0.000294892
SNX7	1.90619	0.00082274	0.0278438
C1orf95	1.90532	1.35E-07	4.53E-05
GPR3	1.89788	0.00057438	0.0222603
CLEC3B	1.89377	0.000141175	0.00872416
DMBX1	1.88504	2.09E-05	0.00215572
RSAD2	1.87337	0.000626078	0.0235237
BUB1B	1.87321	8.13E-05	0.00577361
RDM1	1.87293	0.000202727	0.0109877
RELT	1.87004	1.46E-07	4.66E-05
SYT1	1.86779	6.17E-06	0.000898407
CCKBR	1.85841	0.00288453	0.0642755

F13A1	1.85473	5.54E-05	0.00451004
KIAA0101	1.85452	4.32E-05	0.00377257
SYNE1	1.8534	4.68E-06	0.000733451
DISP2	1.84749	0.000174295	0.00997145
APOBEC3B	1.84033	3.00E-11	6.07E-08
SPC24	1.83958	6.04E-05	0.00470738
SRGAP2	1.83911	7.13E-09	4.47E-06
MYBL2	1.82795	0.000272955	0.013385
WDR17	1.82175	0.000587391	0.0224626
HEG1	1.81633	1.87E-09	1.68E-06
PCDH7	1.8129	1.42E-08	8.55E-06
ARNTL2	1.80929	0.00230648	0.0552195
UHRF1	1.80428	1.95E-05	0.00206532
LAMB1	1.80115	7.59E-07	0.000160931
TMPO-AS1	1.79864	2.34E-08	1.25E-05
PLLP	1.79727	2.49E-05	0.00245859
LOC100499484	1.79459	0.00163623	0.0435231
CDCA5	1.77845	1.84E-05	0.00200501
NCAPG	1.77712	0.000334681	0.0153665
C18orf56	1.7771	0.00139205	0.0392744
PRUNE2	1.7766	0.000880724	0.029324
MGAT5B	1.77282	0.00156216	0.0424136
TCF4	1.77213	2.22E-07	5.63E-05
LRRC4C	1.76333	0.00470508	0.0857013
LINC00284	1.7319	0.00172319	0.0449245
EPHA7	1.72725	0.000670556	0.0246617
DZIP1L	1.72517	0.00582697	0.0978301
EXO1	1.7216	0.000372007	0.016656
CDKN2C	1.7175	5.06E-06	0.000772727
HMGN5	1.70272	2.95E-05	0.00274343
AURKB	1.70203	0.000805746	0.0277091
ZWINT	1.69974	1.75E-05	0.00193001
ACBD7	1.69566	0.000573292	0.0222603
CTGF	1.69559	0.000802603	0.0276821
SKA3	1.69113	0.000280285	0.0136014
KIF11	1.6878	0.000310681	0.0145425
CCDC151	1.67986	9.46E-05	0.00649385
CDC25A	1.67494	2.70E-05	0.00259668
DTL	1.67022	0.000196123	0.0108858

BRCA2	1.66951	0.000268322	0.013385
CMPK2	1.66942	0.00396343	0.0773217
C6orf100	1.66925	0.00360706	0.0731402
ZNF367	1.66746	8.97E-06	0.00117606
LOC401242	1.66538	0.00603184	0.100186
ELOVL7	1.66315	4.05E-09	3.43E-06
GUCY1A2	1.65415	1.00E-05	0.00128818
FRMPD4	1.65232	0.00442408	0.0820724
HMGB2	1.6489	5.41E-05	0.0044586
STIL	1.64763	0.000175105	0.00997819
MLF1IP	1.64301	3.08E-05	0.00282949
KIFC1	1.64178	0.000447726	0.0191539
BLM	1.63668	0.000232559	0.0119743
PBK	1.63265	0.00145117	0.0402408
KATNAL1	1.62974	7.18E-05	0.00535713
LINC00634	1.62042	0.000878117	0.0293051
BIRC5	1.62007	0.00114451	0.0351821
ELOVL6	1.61321	0.00011242	0.00746895
MLLT11	1.61022	2.30E-08	1.25E-05
RAD54L	1.6051	2.17E-05	0.00220508
CDCA3	1.60233	0.00171634	0.0448546
FAM72A	1.60116	0.00415025	0.079653
PKD1L1	1.60049	0.000125604	0.00804815
NXPE3	1.59837	9.43E-08	3.40E-05
CPT1C	1.59824	0.00365639	0.0735205
VASH2	1.59619	0.000440074	0.0189389
SLC16A10	1.59333	4.64E-05	0.00392764
CENPM	1.58991	5.79E-05	0.00458818
HJURP	1.58918	0.00151072	0.0415033
ROR1	1.58804	0.00555885	0.0954149
ENC1	1.58503	2.05E-06	0.00038388
TCF19	1.58376	5.68E-05	0.0045735
ZNF488	1.5835	0.00140532	0.0394172
NDC80	1.58312	0.0025405	0.0586961
KIAA1524	1.57174	0.000261941	0.0131582
CHML	1.56981	4.68E-06	0.000733451
PCSK6	1.5635	0.000495931	0.0203699
MXD3	1.55198	7.79E-07	0.000162678
SLCO4A1	1.55118	1.49E-05	0.0017212

PSMB9	1.54766	1.98E-06	0.000381033
HIST1H4H	1.53799	0.00213767	0.0523436
FAM216A	1.53699	1.49E-05	0.0017212
ANLN	1.5356	0.000354617	0.0161392
PAG1	1.53523	0.000487998	0.0202466
TRIM9	1.53459	1.52E-05	0.00173722
ZNF878	1.53422	0.00656396	0.10638
FBXO5	1.53193	0.00134622	0.0386623
CENPA	1.53148	0.00138987	0.0392744
TICRR	1.52862	0.000482592	0.0201085
AP3S2	1.5267	0.010622	0.140751
HIST1H3H	1.52444	0.00379429	0.0749141
RHEBL1	1.52283	0.00120615	0.0366084
BRIP1	1.52152	0.000453045	0.0193241
POLE2	1.51632	0.00140968	0.0394629
MAP1A	1.51623	0.000322183	0.0149354
SYNM	1.51147	4.90E-07	0.000110272
SYP	1.5098	0.000528093	0.0214465
HIST1H2BN	1.50757	0.00107328	0.0334199
TTLL7	1.50539	1.04E-09	1.07E-06
LINC00669	1.50256	0.0031275	0.0670097
IL15RA	1.49733	1.93E-05	0.00206003
STMN3	1.49605	2.21E-07	5.63E-05
TPX2	1.49366	0.00130417	0.0380611
MICB	1.49203	0.000283499	0.0136696
SYCE2	1.4889	0.00264164	0.0604811
TMCC2	1.48821	0.00272606	0.0616683
SPAG5	1.48725	0.00039802	0.0174947
MMP16	1.48536	0.00759187	0.114864
PKMYT1	1.48263	0.000255629	0.012886
HMGCS1	1.48223	0.000149756	0.00903363
C1QL1	1.47951	0.00682472	0.108656
TROAP	1.47911	0.00227229	0.0549161
KCNQ5	1.47873	0.0112826	0.145948
STK10	1.47678	4.89E-09	3.71E-06
BSN	1.476	1.90E-05	0.00204815
CENPO	1.47353	2.59E-05	0.00252701
CDH7	1.46987	0.0103487	0.139151
SYT9	1.46887	0.00389381	0.0763696

VIM	1.4642	0.00363319	0.0732127
OAS3	1.45889	1.08E-05	0.00136539
CTNNAL1	1.45504	1.92E-07	5.15E-05
HIST1H3G	1.45431	0.0037091	0.0739615
FAM83D	1.45183	0.00422574	0.0804052
BICD1	1.45055	0.000154892	0.00926059
FAM81A	1.43651	0.00429035	0.0809607
FOXM1	1.43532	0.00191119	0.0483397
CCDC136	1.43476	0.000618213	0.0233319
GDAP1	1.43443	6.32E-05	0.00486918
ANP32E	1.43329	0.00126688	0.0374275
NEIL3	1.43172	0.00219017	0.0532473
PRC1	1.42942	0.000553965	0.021941
POLQ	1.4288	0.00350656	0.0720144
TACC3	1.42807	0.00115639	0.0354716
CDK1	1.42427	0.00312154	0.0670097
HPS5	1.42266	5.54E-07	0.000122867
CDT1	1.42163	4.65E-05	0.00392764
LRP8	1.42112	5.51E-05	0.00451004
DISP1	1.41651	0.00254522	0.058711
PRICKLE1	1.41359	0.00328285	0.0688921
TEAD4	1.41279	6.43E-08	2.65E-05
GOLIM4	1.41054	4.57E-10	5.49E-07
ITGAM	1.41005	0.00121219	0.0367147
CENPK	1.40904	0.0010227	0.0321225
CENPI	1.39616	0.00558886	0.0956943
DNA2	1.39102	3.01E-05	0.00278037
CDKN2B-AS1	1.38575	0.00283838	0.0636406
PTPN14	1.37918	3.57E-07	8.45E-05
LYPD5	1.37429	0.00591896	0.0987657
SHC4	1.36828	0.00120403	0.0366084
SKA1	1.36414	0.00842536	0.122572
ZNF788	1.36203	0.00246886	0.0574958
RIMKLA	1.36098	0.00366338	0.0735584
SLC36A1	1.3584	1.72E-07	5.15E-05
SLC25A19	1.35254	9.02E-05	0.00624925
SLCO2A1	1.34909	0.00217761	0.0531848
SRGAP2D	1.34897	0.00011607	0.00763404
GLYCTK	1.34777	0.00144644	0.0402408

SMC2	1.34758	0.00073352	0.0262619
CTSL2	1.33719	0.000168736	0.00984888
GPR19	1.33666	0.00433903	0.0812413
RASSF8	1.33256	5.87E-05	0.00461277
MPP1	1.33039	0.00592594	0.098768
LOXL3	1.32639	9.51E-05	0.00649965
ADAMTS12	1.3246	1.32E-05	0.00158229
CIT	1.32238	0.00575065	0.0975052
TBC1D8B	1.31873	0.000246892	0.0125776
FAM111A	1.31674	2.04E-06	0.00038388
POLR3G	1.31334	0.000313617	0.0145852
CDC7	1.30899	0.00719454	0.112866
CDC25C	1.30899	0.0046412	0.0848335
EPB41L2	1.30892	8.15E-06	0.00114077
WSCD1	1.3088	8.55E-06	0.00116287
ANO2	1.30836	0.000364762	0.0164337
ZDHHC23	1.30714	0.000272811	0.013385
RFC3	1.30705	2.22E-05	0.00220508
PART1	1.30661	0.00326768	0.0687662
FANCD2	1.30194	0.000130859	0.00827454
TUBA1B	1.30167	6.96E-05	0.00523211
PLCE1-AS1	1.29855	0.00442896	0.0820724
IL22RA1	1.29479	0.0112916	0.145948
CENPL	1.29162	0.00035991	0.0162659
SPARC	1.291	0.00174514	0.0453327
C9orf40	1.28674	0.000334024	0.0153665
CENPP	1.2857	0.00333844	0.0695525
FANCI	1.28522	0.000895829	0.02969
KIAA0319	1.28519	0.00908762	0.128953
DNAJC6	1.27826	0.00161174	0.043185
WDR76	1.27816	0.00056973	0.0222603
LMNB1	1.27652	0.00442021	0.0820724
TENM3	1.27557	2.75E-05	0.0026216
KIF23	1.27492	0.00654836	0.106315
GOT1	1.27195	3.56E-07	8.45E-05
SGOL1	1.27164	0.00981822	0.135195
MED12L	1.27128	0.00359296	0.0730603
NCAPH	1.2682	0.00156986	0.0424629
IFIH1	1.26653	7.24E-05	0.00535713

FLVCR1-AS1	1.26532	0.000488715	0.0202466
GSG2	1.2608	0.00932975	0.131226
MELK	1.2591	0.00475661	0.0861508
RGL1	1.25704	0.0020207	0.0501419
NCAPG2	1.25598	0.000199278	0.0108858
CCNE2	1.25418	0.000248571	0.0126185
SLC2A6	1.25393	0.00269182	0.0613752
FEN1	1.25186	5.89E-05	0.00461277
RASGRP3	1.2502	0.00378137	0.0748847
TTK	1.24807	0.00542533	0.0940851
SNAP25	1.24523	0.00213847	0.0523436
ARHGAP11A	1.24436	0.00667043	0.106853
MT2A	1.2407	0.000179724	0.0102011
PEG10	1.23966	0.000203521	0.0109894
FGFR2	1.23964	0.00421332	0.0804052
ADORA2B	1.23881	0.00717623	0.112824
CCDC85A	1.23181	0.00470801	0.0857013
TBC1D4	1.22891	0.00106784	0.0333227
TPTE2P1	1.22345	0.00575994	0.0975052
LYPD6	1.22298	0.000550863	0.0218782
IKBKE	1.22245	6.81E-05	0.00519156
CHEK1	1.21995	0.00354467	0.0722823
TP53I3	1.21655	1.34E-05	0.00158828
PARPBP	1.20778	0.0102512	0.138252
ATF3	1.20099	0.0092082	0.130024
APLN	1.20076	0.00943574	0.13233
ZNF460	1.20004	0.0112235	0.145594
RAI14	1.19974	0.0104243	0.139428
ENOX1	1.19774	0.0022959	0.0550749
MCM2	1.19532	0.000233878	0.0119993
H2AFX	1.19495	2.93E-06	0.000509257
SYNE2	1.19341	2.21E-05	0.00220508
PCDHGC3	1.19248	0.00195901	0.049204
CDC6	1.1918	0.00274991	0.061946
LOC148709	1.19048	0.000367006	0.0164832
TBC1D1	1.18803	0.00138987	0.0392744
ADAM22	1.18628	0.00856017	0.123908
WDR62	1.18366	0.000302027	0.0143879
SFXN5	1.18257	2.00E-07	5.25E-05

PSIP1	1.1788	6.10E-05	0.00472438
ERCC6L	1.17439	0.00718575	0.112851
ORC6	1.17059	0.00115998	0.0355062
ZNF432	1.16956	0.0112298	0.145594
NES	1.16363	0.00415475	0.079653
STMN1	1.15981	0.00352936	0.0722767
MSMO1	1.15712	0.0001416	0.00872416
GINS4	1.15463	0.00122765	0.036873
KIAA1161	1.15442	7.48E-05	0.00544352
KIF20B	1.14914	0.00472847	0.0859011
HSPA2	1.14892	0.00029214	0.0140393
STAMBPL1	1.14599	0.00440893	0.0820724
SMAD6	1.14515	0.00577055	0.0975052
ZNF695	1.14392	0.0114789	0.147496
ZAK	1.14138	0.00399503	0.0777279
IL1RAPL2	1.13827	0.00776292	0.116506
FANCA	1.13776	0.000545559	0.0217828
MYO16	1.13616	2.75E-06	0.000489791
TTC7B	1.13094	0.000522317	0.0212719
RCBTB2	1.13051	0.0034343	0.0708331
MTFR2	1.12883	0.00270036	0.0614055
XPR1	1.12682	2.07E-05	0.00215572
LMNB2	1.1253	0.00363238	0.0732127
TMEM180	1.12517	0.00501175	0.0896457
RNASEH2A	1.12331	0.00312808	0.0670097
ARHGEF39	1.1233	0.00339913	0.0704099
ARHGAP33	1.12216	0.00129731	0.0380149
ROR2	1.11949	0.00233754	0.055703
KCNK3	1.11803	0.00127222	0.0375084
C5orf34	1.11595	0.00654136	0.106315
AP4B1	1.11471	0.000413901	0.0180825
ATAD2	1.11468	0.00153405	0.0418871
TPM3P9	1.11462	0.00527955	0.0925041
OXCT1	1.11188	0.00087225	0.0291769
CEP85L	1.10679	0.000538071	0.0215644
DOCK5	1.10673	1.49E-06	0.000299161
B3GALTL	1.10549	0.000759062	0.0267565
STARD8	1.10307	0.00569285	0.0972439
ACYP1	1.10082	0.0013407	0.0386579

RRM1	1.1006	0.00127488	0.0375101
LRRCC1	1.09998	0.00154147	0.0420101
LDOC1	1.09681	0.000658698	0.0243499
NRM	1.09563	0.003198	0.0679872
AURKA	1.09561	0.0078983	0.117878
HCN3	1.09549	0.000190964	0.0107544
CENPN	1.09439	0.00305498	0.0662311
CCDC15	1.08925	0.0103564	0.139151
RASAL2	1.08451	8.55E-05	0.0059575
MCM5	1.08177	0.00122347	0.036858
LHX2	1.08097	0.00227918	0.0549161
ANKRD30B	1.08061	0.00204493	0.0504825
ZBED5-AS1	1.08044	0.000585539	0.0224514
TAP2	1.07963	2.58E-05	0.00252701
HSPA4L	1.07936	0.00237957	0.0563529
GEN1	1.07929	0.00883438	0.126594
CXCL16	1.07705	0.0042431	0.0804052
HIST1H1C	1.07451	0.0022785	0.0549161
CCNF	1.07165	0.0013581	0.0387745
PFKP	1.07161	0.000445683	0.0191233
STON2	1.07137	0.000584715	0.0224514
CENPH	1.07021	0.00320672	0.0679872
COPG1	1.06877	3.80E-07	8.83E-05
TTF2	1.06464	0.00019974	0.0108858
INSIG1	1.06341	7.95E-05	0.00570208
GMNN	1.06081	0.00090496	0.0298554
PCDHA3	1.06073	0.00125854	0.0374111
MAD2L1	1.06069	0.00475642	0.0861508
NCAPD2	1.05714	0.00198333	0.04947
LATS2	1.05455	0.000617733	0.0233319
OSBPL6	1.05191	0.0043811	0.0818166
KPNA5	1.05001	0.00296643	0.0651936
EAF2	1.04897	8.27E-05	0.00581924
PPFIBP1	1.0486	0.000811074	0.0277091
PLK4	1.04719	0.00990011	0.135667
ZWILCH	1.0419	0.00379651	0.0749141
FAM21A	1.03942	0.0111362	0.145032
DENND5B	1.03869	3.02E-06	0.000518602
BORA	1.03691	0.00370138	0.0739451

IFIT5	1.03606	0.00030994	0.0145425
CEP97	1.03549	0.00248319	0.0576491
ACAT2	1.03521	0.00424976	0.0804052
IBA57	1.03176	4.51E-05	0.00390026
FDPS	1.0311	0.000154302	0.00926059
DBF4B	1.02899	0.00517653	0.0914746
TNFAIP8L1	1.02888	0.00735494	0.113773
CLPB	1.02868	0.00158704	0.0427671
PIK3R1	1.02581	0.00136089	0.0387745
CDK6	1.02504	0.000918539	0.0301653
ZIC3	1.02038	0.0069399	0.109844
NHS	1.01472	0.00156523	0.042417
LOC643837	1.01321	0.00050201	0.0205028
MMS22L	1.00884	0.00430797	0.0809998
SYNGR3	1.00769	0.0103983	0.139428
CLCN5	1.00767	0.00247193	0.0574958
DONSON	1.00762	0.000817951	0.0278245
STIM2	1.00461	0.00477112	0.0863052
LRRC49	0.999362	0.00228166	0.0549161
PRKCA	0.997279	0.00810991	0.119673
FLVCR1	0.996571	0.00133002	0.0385037
UPF3B	0.99657	0.00124061	0.0371076
PDE4DIP	0.99464	8.79E-06	0.00117368
SCAMP5	0.989379	8.55E-06	0.00116287
ATP5D	0.989282	0.000971004	0.0311781
ATP11C	0.989165	0.000116711	0.00763404
SLC4A8	0.987678	0.000643432	0.0238467
MAPK8IP1	0.985241	0.00197495	0.0494319
PHACTR2	0.984499	0.00107646	0.0334469
NBN	0.982813	0.00779616	0.116837
SQLE	0.981725	0.00119771	0.0365061
KPNA2	0.980041	0.00038626	0.0170819
STARD9	0.979853	0.00574852	0.0975052
SKA2	0.976684	0.00123945	0.0371076
EBP	0.973094	0.000226484	0.0117454
NUP205	0.972971	0.00235896	0.0561206
FREM2	0.971049	0.00352695	0.0722767
DDX60	0.967736	0.00965747	0.134148
DBF4	0.967432	0.00302239	0.0657609

SCLT1	0.962529	0.00389873	0.0763696
LCORL	0.961218	0.000845222	0.0284709
SAP30	0.960177	0.00218022	0.0531848
FOXN2	0.959233	0.00820911	0.120887
KNTC1	0.955058	0.00446362	0.0826085
ASAP1	0.954513	0.00473091	0.0859011
FAM160A1	0.954195	0.000394948	0.0174127
BEX5	0.952618	0.00183262	0.0468926
PHTF2	0.95204	0.000221774	0.0116441
CNTRL	0.95052	0.00310037	0.0669132
RYR2	0.949209	0.0107828	0.142064
C17orf53	0.947903	0.0065971	0.106387
PSMB10	0.94751	0.00099819	0.031698
E2F1	0.942328	0.00176954	0.0458016
PANK1	0.941239	0.00967632	0.134148
LRR1	0.940095	0.00237314	0.0563529
DHCR7	0.938694	0.00201883	0.0501419
UNC13B	0.938372	0.00722613	0.113238
AIM1	0.934966	0.000683839	0.0248335
MCM3	0.933803	0.00125444	0.0373664
SEC23A	0.933332	0.000572138	0.0222603
MASTL	0.931323	0.00198031	0.04947
OSBPL11	0.930779	0.000457512	0.0194571
CEP152	0.92563	0.00164058	0.0435258
ITPR2	0.924488	0.00515802	0.0913785
SLCO3A1	0.922575	0.0112118	0.145594
HLA-H	0.92257	0.00581684	0.0978301
LCP1	0.922476	0.000225643	0.011744
SMC1A	0.922008	0.00464269	0.0848335
HIP1	0.919469	0.00505672	0.0900033
DEK	0.91851	0.0100922	0.136877
STXBP1	0.913497	0.00013753	0.00862073
TCP11L1	0.913222	0.00146313	0.0404874
MCM6	0.912942	0.00968039	0.134148
THOP1	0.911996	2.65E-05	0.00256082
ANKRD13B	0.908889	0.00482919	0.0870281
MGST3	0.908832	0.000626558	0.0235237
TMEM198	0.908174	0.00430928	0.0809998
DPP3	0.907333	2.19E-05	0.00220508

ALDH3B1	0.904533	0.00408863	0.0788046
FAM3C	0.898821	0.0017174	0.0448546
LDLR	0.898813	0.00490429	0.0880512
METTL7A	0.897796	0.00193931	0.0488793
PPIL1	0.897542	0.00734216	0.113773
LIPE	0.895762	0.00488366	0.0877901
DUT	0.894969	0.00391162	0.076518
LPIN1	0.894869	0.00126606	0.0374275
SLC41A2	0.886086	0.000210028	0.011251
CCDC77	0.884743	0.000429008	0.0186295
EZH2	0.883777	0.0104835	0.139428
MCM7	0.883153	0.00578255	0.0975052
ZNF525	0.881039	0.00662989	0.10644
FRY	0.880216	0.000694095	0.0251426
MAML2	0.879751	0.00212667	0.0523211
FAM198B	0.877871	0.00555357	0.0954149
LOC646214	0.876825	0.00834677	0.121797
APOL3	0.876128	0.000981077	0.0312924
FA2H	0.875996	0.000237048	0.0121189
DOT1L	0.87506	0.000312784	0.0145852
FANCM	0.873804	0.0106628	0.141033
HAUS2	0.87332	0.00124832	0.037261
SLX4	0.872313	0.00348205	0.0716129
E2F3	0.871111	0.000424493	0.0184892
DLAT	0.86987	0.000935702	0.0304515
PRPS1	0.866535	0.000171505	0.00989204
ALDOC	0.861118	0.011084	0.144482
POLA2	0.858399	0.00822441	0.120887
ANAPC15	0.853716	0.0104443	0.139428
GPR98	0.853153	0.00751422	0.11469
INCENP	0.852699	0.00184028	0.0469284
SFN	0.846951	0.003149	0.067258
TEAD1	0.84671	0.00744045	0.11432
ITGB8	0.846213	0.0109975	0.143745
TNS1	0.845767	4.52E-05	0.00390026
MIS18BP1	0.842141	0.00853974	0.123736
DTYMK	0.842114	0.00450241	0.083113
ALYREF	0.840344	0.00301522	0.0657609
HMGCR	0.83901	0.0116396	0.149163

MCU	0.838668	0.00393122	0.0767972
REPS2	0.836786	0.00548433	0.0945978
IMPA1	0.835749	0.00576945	0.0975052
CISD1	0.832802	0.000768806	0.0269681
FASTKD1	0.832448	0.00082131	0.0278438
PFKM	0.825855	7.42E-05	0.00543109
LIG1	0.825672	0.0031153	0.0670097
TRAF3	0.82546	0.00335643	0.0698263
PCNA	0.82383	0.0102489	0.138252
GTDC1	0.823574	0.00512039	0.0910242
TMEM38B	0.82204	0.00070606	0.0255119
POP1	0.820319	0.00515933	0.0913785
MVD	0.81786	0.00838812	0.122153
DDX12P	0.815977	0.0101705	0.137422
APOL2	0.813713	0.004373	0.0817712
AFG3L2	0.813325	3.87E-05	0.00346126
LAGE3	0.808539	0.00440211	0.0820724
CHAF1A	0.807995	0.0041809	0.0800457
PAK2	0.805508	0.00374451	0.074359
MDH1	0.805157	0.000221333	0.0116441
TYRO3	0.803125	0.00467705	0.0853532
GLMN	0.800999	0.00952083	0.133135
MARVELD2	0.799318	0.000123562	0.00795264
MYO9A	0.798361	0.0100031	0.136437
PPIC	0.796195	0.00122459	0.036858
MVK	0.7949	0.00112846	0.0347628
SNX8	0.794746	0.00544836	0.0942964
ATP2B4	0.789523	0.00837673	0.12211
ALDH9A1	0.788591	0.000738307	0.0262819
TDRKH	0.785267	0.000546949	0.0217828
POLA1	0.783385	0.010526	0.139736
ACOT7	0.781429	0.0032318	0.068318
NCAPD3	0.78138	0.0104249	0.139428
DNPH1	0.780592	0.00183446	0.0468926
LETM2	0.780285	0.000847819	0.0284919
SLC25A30	0.779687	0.00494913	0.0887458
ECSIT	0.777683	0.00246869	0.0574958
UBE2S	0.777395	0.00926788	0.130583
PPP2R1B	0.776173	0.00979057	0.135072

ACVR1C	0.776148	0.00547402	0.0945978
PARP2	0.770941	0.000795059	0.0275538
AGL	0.770928	0.00594784	0.0990185
TCTA	0.7695	0.0105521	0.139954
DNAJC9	0.76932	0.00127939	0.0375662
ATP6V0A1	0.769104	0.00238045	0.0563529
MYO1D	0.766001	0.0013333	0.0385215
APAF1	0.765506	0.0072659	0.113383
RTTN	0.762866	0.00455752	0.0838083
C12orf4	0.762633	0.001496	0.0412387
ITGA2	0.761037	0.00692178	0.109781
LINC00294	0.760652	0.0080984	0.119626
FAM155B	0.759197	0.00682823	0.108656
ZSWIM5	0.758671	0.00353752	0.0722823
EGLN3	0.757974	0.0116645	0.14922
L2HGDH	0.75742	0.00733715	0.113773
TUBB4B	0.75554	0.0114192	0.146992
PKN3	0.755475	0.0105111	0.139667
8-Mar	0.755372	0.0024726	0.0574958
CDK8	0.754501	0.00752563	0.11469
HAUS5	0.754245	0.00823408	0.120887
ARMC6	0.750857	0.00321765	0.0681188
ABHD6	0.750138	0.00406586	0.0785757
ZKSCAN2	0.749712	0.00553468	0.0953327
C4orf21	0.748688	0.00595628	0.0990446
RBM12B	0.739692	0.00995353	0.136079
BTBD1	0.738006	0.000411777	0.0180444
MYH10	0.733254	0.0033014	0.069081
ZNF860	0.729381	0.00916348	0.129759
CFL2	0.728647	0.00608608	0.10097
NFKB2	0.72576	0.00865095	0.124472
BHLHB9	0.721765	0.0108408	0.142291
PSMB8	0.720348	0.00111459	0.0345572
NDRG2	0.712113	0.00924005	0.130346
PPP1R21	0.7104	0.00385788	0.0758786
BTN2A2	0.708798	0.00782178	0.117099
RCAN1	0.702042	0.00852376	0.123736
CHCHD3	0.700728	0.00441955	0.0820724
TAP1	0.69879	0.00327208	0.0687662

DSN1	0.695578	0.00733289	0.113773
CYCS	0.693735	0.00160582	0.043112
KCNC4	0.692126	0.00227357	0.0549161
IPO9	0.690808	0.0104382	0.139428
KCTD10	0.690664	0.0106927	0.141224
GPR137C	0.69026	0.010854	0.142291
CBX6	0.68811	0.00293986	0.065106
CHTF18	0.685385	0.00678615	0.108345
GOT2	0.680253	0.00917141	0.129759
OPA3	0.67799	0.00411763	0.0791519
MYO19	0.673648	0.00239487	0.0564165
PATL1	0.665251	0.00759968	0.114864
TMEM237	0.662799	0.01079	0.142064
EPB41L4A	0.660656	0.00766258	0.115556
SUPT16H	0.655083	0.00462781	0.0848335
ASAP2	0.654086	0.0072879	0.113589
ZNFX1	0.652344	0.00967653	0.134148
DNAJC2	0.651909	0.00503274	0.0896873
TRIM14	0.649032	0.0103859	0.139417
SLC25A13	0.648802	0.00303582	0.0659147
PLCE1	0.646939	0.0117451	0.149982
EPT1	0.645487	0.00482054	0.0869808
ACSS3	0.643994	0.0112995	0.145948
IDH3A	0.640344	0.00995957	0.136079
SUCLG1	0.640029	0.00950905	0.133099
VCL	0.637432	0.0075231	0.11469
PITRM1	0.62763	0.00210566	0.0518927
ADIPOR2	0.626442	0.00624264	0.103093
LUZP1	0.626004	0.00996736	0.136079
TPI1	0.619397	0.0027967	0.0628038
WASF1	0.613017	0.00961064	0.134001
VAC14	0.610859	0.00661277	0.10644
ZNF330	0.599251	0.00877919	0.126065
CNTROB	0.599228	0.00298276	0.0653561
TDRD7	0.595862	0.00990896	0.135667
OGDH	0.590377	0.00452009	0.0833326
HCFC2	0.585451	0.0094978	0.133071
TOMM40L	0.580473	0.0100885	0.136877
PHF14	0.577677	0.00862322	0.124445

PET112	0.564802	0.0108977	0.1427
HARS	0.563603	0.00964264	0.134148
TOP3A	0.550843	0.0110756	0.144482
PPARD	0.546019	0.00463373	0.0848335
SMAD1	0.51098	0.0104543	0.139428
MRPS5	0.498317	0.0113952	0.146945
TEAD2	-0.51487	0.010084	0.136877
VMP1	-0.525157	0.00683604	0.108661
PCYOX1	-0.534758	0.0108543	0.142291
TM9SF2	-0.551405	0.00918264	0.12979
TPD52L1	-0.551863	0.0109513	0.143271
ALG13	-0.558571	0.0108566	0.142291
LSM1	-0.560658	0.00676267	0.10809
PTTG1IP	-0.562381	0.00900943	0.127969
ARIH2	-0.567511	0.00629499	0.103601
REEP3	-0.568439	0.007997	0.118634
RING1	-0.576308	0.00361211	0.0731402
PCDH1	-0.578252	0.00733503	0.113773
ANAPC16	-0.586422	0.00402794	0.0781985
FOXP1	-0.589257	0.00773092	0.116465
TRAPPC6A	-0.589967	0.0104809	0.139428
DNAL4	-0.590388	0.0033622	0.0698456
SEMA4B	-0.590463	0.00631848	0.103678
CLDN7	-0.590846	0.00966514	0.134148
GGCX	-0.591447	0.00805453	0.119344
PCMTD2	-0.593706	0.00927492	0.130583
ZNF83	-0.600686	0.0093565	0.131346
DYNLL1	-0.605153	0.00755825	0.114864
HNRNPA0	-0.606201	0.00637436	0.104194
FAU	-0.609184	0.00761534	0.114964
RWDD3	-0.61603	0.00301387	0.0657609
FIS1	-0.616842	0.00525174	0.0922221
USP28	-0.621547	0.00849442	0.123452
SEPN1	-0.621943	0.00885268	0.126594
GSKIP	-0.622023	0.00785095	0.117292
GPAA1	-0.622218	0.00796694	0.118634
CMC1	-0.625762	0.0117219	0.149818
UBA52	-0.626161	0.00985896	0.135368
COPS7A	-0.62678	0.00455563	0.0838083

TMED4	-0.628338	0.00354145	0.0722823
GPR108	-0.631619	0.00369093	0.0739057
SHARPIN	-0.63432	0.00298289	0.0653561
TCEB2	-0.634656	0.0106969	0.141224
ZNF512B	-0.636311	0.00341044	0.0705429
SKP1	-0.6383	0.00741304	0.11406
ATRAID	-0.638875	0.00739643	0.113926
SLC2A1	-0.642261	0.0108345	0.142291
STK16	-0.645447	0.0042458	0.0804052
TNFRSF25	-0.645652	0.00826455	0.121087
ALS2CL	-0.64641	0.00627646	0.103415
POLM	-0.647383	0.00972788	0.134465
MBTD1	-0.648554	0.0100628	0.136877
TTC17	-0.649491	0.00132277	0.0384483
ZSCAN18	-0.652059	0.00230959	0.0552195
ENKD1	-0.653845	0.0100339	0.136728
NUDT13	-0.65433	0.00885303	0.126594
AMZ2P1	-0.657589	0.00407569	0.0786603
TCAIM	-0.659247	0.00893095	0.127105
DDRGK1	-0.659547	0.00294859	0.065126
DUSP22	-0.662962	0.00775607	0.116506
RBKS	-0.665498	0.00583574	0.0978301
CPD	-0.667658	0.000913188	0.0300581
TMEM219	-0.668638	0.00537227	0.0934318
LUC7L	-0.669889	0.00430245	0.0809998
SERP1	-0.671544	0.00529996	0.0926177
DEF6	-0.672511	0.00724965	0.113383
ETHE1	-0.672795	0.0104708	0.139428
CBLN3	-0.67354	0.0052188	0.0917555
UBL3	-0.673737	0.0112117	0.145594
ENGASE	-0.676026	0.00798815	0.118634
LOC100129034	-0.676315	0.00274884	0.061946
EPS8L1	-0.676471	0.00426751	0.0806352
FAAH	-0.676552	0.00520053	0.0915947
TAX1BP1	-0.67729	0.00184237	0.0469284
RNASET2	-0.679953	0.000643193	0.0238467
LTBP4	-0.683268	0.00264293	0.0604811
TC2N	-0.68638	0.00776596	0.116506
RGL2	-0.687284	0.00758205	0.114864

NIT2	-0.692607	0.00135561	0.0387745
TMEM66	-0.695948	0.00320498	0.0679872
FAM110A	-0.696041	0.00706229	0.111275
NFYA	-0.696527	0.00134527	0.0386623
DNAJC4	-0.697157	0.00130998	0.0381536
CLSTN1	-0.698406	0.00168014	0.0441212
AUP1	-0.698681	0.00272809	0.0616683
MANBA	-0.700245	0.010115	0.136927
SNRNP70	-0.703713	0.00531093	0.092697
SMIM14	-0.703983	0.00877235	0.126065
SMIM22	-0.704028	0.00697881	0.110322
UBE2E1	-0.704097	0.00520331	0.0915947
C1RL	-0.707492	0.00799283	0.118634
NFATC4	-0.707754	0.00796626	0.118634
MAPK13	-0.70802	0.000187302	0.0105896
DHPS	-0.710303	0.00649032	0.10573
FAM109A	-0.710465	0.00148084	0.040899
MAT2A	-0.711014	0.000979504	0.0312924
ANKRA2	-0.713877	0.00135981	0.0387745
SIL1	-0.714805	0.00102159	0.0321225
PPP1R26	-0.715578	0.00193886	0.0488793
COMMD1	-0.715906	0.00357366	0.0727704
C12orf57	-0.716093	0.00151933	0.041564
PEA15	-0.716277	0.00859402	0.124148
TNFSF13	-0.721931	0.00739378	0.113926
C6orf120	-0.722664	0.000538364	0.0215644
SLC25A20	-0.723847	0.00572056	0.0975052
MAN2B1	-0.729352	0.0065723	0.10638
C6orf57	-0.729614	0.0028949	0.064407
SH3PXD2A	-0.730632	0.00203072	0.0503039
RPL9	-0.73244	0.00424924	0.0804052
SRD5A3	-0.733646	0.00449173	0.0830221
ENDOV	-0.733733	0.00173729	0.0452102
NTAN1	-0.733956	0.0024173	0.056852
OSTC	-0.734835	0.00648564	0.10573
PPP1R35	-0.735139	0.0058812	0.0982495
TXNIP	-0.735814	0.00213806	0.0523436
SDC3	-0.735982	0.000831742	0.0280825
CERS6	-0.737071	0.00337683	0.0700487

ATP8B1	-0.738388	0.00111902	0.03462
R3HDM4	-0.741467	0.000726288	0.026112
FAM173B	-0.743019	0.00228614	0.0549322
ICK	-0.744444	0.000904234	0.0298554
SLC3A2	-0.744648	0.00760077	0.114864
NUMB	-0.744829	0.000281141	0.0136014
BCDIN3D	-0.745141	0.011204	0.145594
TMBIM6	-0.745235	0.000269625	0.013385
GLUD2	-0.746813	0.00272903	0.0616683
CAMKK1	-0.747365	0.0110506	0.144308
C11orf54	-0.747462	0.00611799	0.101383
SKP1P2	-0.748108	0.00376127	0.0745891
SSPN	-0.748503	0.00363466	0.0732127
IDH1	-0.749834	0.00548545	0.0945978
RPL13	-0.750182	0.0106327	0.140764
BNIP3L	-0.750726	0.000225545	0.011744
IFT43	-0.750762	0.00077859	0.0271134
PPM1N	-0.750788	0.010459	0.139428
ISYNA1	-0.752398	0.0114572	0.147349
RGS14	-0.753254	0.0107313	0.141549
SNHG6	-0.753648	0.00433723	0.0812413
AP1G2	-0.756317	0.000736944	0.0262819
TNRC6C-AS1	-0.756954	0.00502262	0.0896873
APMAP	-0.757439	0.00726056	0.113383
USB1	-0.757476	0.00203739	0.0503826
RPL18A	-0.758494	0.0058599	0.0981209
SIPA1	-0.758946	0.00757836	0.114864
LOC100131564	-0.760848	0.00658383	0.106387
TAGLN	-0.761401	0.00330623	0.069081
PLA2G6	-0.761688	0.00577708	0.0975052
PRKRA	-0.764518	0.00309083	0.0668074
FAM211B	-0.764595	0.00631238	0.103678
C11orf35	-0.764705	0.00775711	0.116506
STX16	-0.76475	0.000464703	0.019647
GLUL	-0.765688	0.0101148	0.136927
ELOVL5	-0.767557	0.00295545	0.065126
HOMER3	-0.76844	0.00076162	0.0267812
BCL3	-0.770339	0.0075938	0.114864
PCSK4	-0.772646	0.00808193	0.119505

ARHGAP30	-0.77346	0.00894035	0.127113
SLC9A3R2	-0.77442	0.00302124	0.0657609
ARHGAP4	-0.774753	0.00497185	0.0890425
LYRM1	-0.774837	0.00197038	0.0494033
NPAS1	-0.775122	0.00654367	0.106315
SUMO3	-0.776232	0.00333499	0.0695525
GPR160	-0.776507	0.00674711	0.107961
AMN1	-0.778197	0.00245931	0.0574958
OSGEPL1	-0.778293	0.00565185	0.096658
LINC00116	-0.778489	0.00737229	0.113797
AGFG1	-0.779107	0.00983993	0.135332
SPAG4	-0.780924	0.00620709	0.102741
LRIG1	-0.784525	0.00528063	0.0925041
LRP3	-0.786132	0.00887968	0.126751
MFSD6	-0.786283	0.000581683	0.0224228
HOXC13	-0.786923	0.00423348	0.0804052
ZNF680	-0.791276	0.00746694	0.114388
PQLC3	-0.792632	0.00182039	0.0466986
VAT1	-0.792656	0.00659124	0.106387
KIF16B	-0.794054	0.00515734	0.0913785
CLINT1	-0.794242	0.000210253	0.011251
SSR4	-0.796862	0.00177813	0.0459415
KAZN	-0.797089	0.00636506	0.104194
LOC100507557	-0.797089	0.0098469	0.135332
TLE2	-0.806614	0.00265407	0.0606398
MFN1	-0.808489	0.000600305	0.0228958
LAMA3	-0.808917	0.00799834	0.118634
EHF	-0.809108	0.00974873	0.134624
ADAM15	-0.810991	0.000269626	0.013385
GOLGA8A	-0.814434	0.00542962	0.0940851
RPS13	-0.815761	0.00731629	0.113773
EGFL7	-0.818887	0.010085	0.136877
SURF1	-0.820833	0.000808346	0.0277091
RPS23	-0.82189	0.00784097	0.117265
RPS20	-0.821986	0.000928502	0.0302855
C19orf53	-0.823042	5.57E-05	0.00451079
PP7080	-0.823324	0.00759773	0.114864
SLC35E2B	-0.823619	5.77E-05	0.00458818
CYTH2	-0.823959	0.00307359	0.0665345

ERGIC3	-0.824316	0.00016366	0.00959143
SLC35E2	-0.825064	0.00121773	0.0368051
MPST	-0.826162	0.00403007	0.0781985
EML2	-0.826403	0.010294	0.138699
GRIN1	-0.827285	0.0117566	0.149995
C11orf1	-0.828439	0.00180149	0.0463787
FADS3	-0.830256	0.000757639	0.0267565
NCK2	-0.830456	0.00573139	0.0975052
GBAS	-0.832773	0.0116648	0.14922
ESR1	-0.833012	0.0098884	0.135643
FBXL17	-0.833737	0.00479413	0.0866128
RPS29	-0.833929	0.00295884	0.065126
FHL3	-0.834033	0.00682102	0.108656
SRPK3	-0.834408	0.00538069	0.093462
POU2F3	-0.835585	0.00555931	0.0954149
PLEKHG5	-0.836838	0.00189983	0.0482215
RPS4X	-0.83824	0.00971453	0.134409
PNPLA2	-0.838618	0.00890801	0.127029
LOC100652768	-0.839354	0.0063212	0.103678
EIF3E	-0.839801	0.00264095	0.0604811
TMEM198B	-0.841137	0.00580212	0.097721
THBS3	-0.843698	0.000677463	0.0247617
TMOD3	-0.846224	7.61E-05	0.00551427
SPSB2	-0.847041	0.0018099	0.0465122
KIFC2	-0.847671	0.000963249	0.0309981
PTPRF	-0.849735	0.00027135	0.013385
RPS7	-0.85417	0.00145143	0.0402408
GLI4	-0.856252	0.00644573	0.105241
IGSF8	-0.856391	0.000559294	0.0220913
TBC1D16	-0.858075	0.000141465	0.00872416
ZP3	-0.859352	0.00261647	0.0601621
KRCC1	-0.859706	0.000156303	0.00927333
MINPP1	-0.859797	0.000756881	0.0267565
TMEM9	-0.86155	0.000111397	0.00743527
FAM174B	-0.861821	0.00164237	0.0435258
SPPL2A	-0.861866	0.000306717	0.0144508
ARL1	-0.86219	0.00161453	0.043185
ERVK13-1	-0.864365	0.00315507	0.0672879
TMEM123	-0.866658	0.00954391	0.133199

RPL5	-0.866791	0.00269477	0.0613752
NOTCH3	-0.86744	1.25E-05	0.00152959
RAB4A	-0.871258	0.00406128	0.0785757
NANOS1	-0.876729	0.000798801	0.0276171
PABPC1L	-0.877514	0.00745373	0.11432
GPR153	-0.877674	0.00102254	0.0321225
RPL18	-0.877678	0.00577307	0.0975052
EPHB4	-0.878061	0.000635186	0.023724
ULK1	-0.878251	0.000161979	0.00953161
LGALS3	-0.879487	0.000341112	0.0156121
GLUD1	-0.880882	0.00194658	0.048977
PRSS30P	-0.882084	0.00694096	0.109844
NR4A2	-0.882205	0.000734102	0.0262619
LINGO1	-0.883165	0.00014379	0.00882136
CCBP2	-0.883665	0.000534171	0.0215644
TMCO3	-0.883675	0.00432373	0.0811656
EEF1A1	-0.886904	0.00980511	0.135144
FAM210B	-0.887675	0.00705402	0.111267
RPL38	-0.888778	0.00370316	0.0739451
NEBL	-0.889141	9.34E-06	0.00121341
RPL32	-0.890337	0.00278466	0.062631
EPB41L4A-AS1	-0.892122	0.00621765	0.102798
CELSR1	-0.892809	0.000960605	0.0309936
CD81	-0.894731	9.97E-05	0.00675046
PFKFB3	-0.896034	0.000197835	0.0108858
CAT	-0.898054	0.000210708	0.011251
TMEM179B	-0.90201	0.000303386	0.0143879
SLC11A2	-0.903617	0.00017117	0.00989204
NRCAM	-0.904292	0.000117023	0.00763404
RPS6	-0.904893	0.00291765	0.0647136
PVRL4	-0.905668	0.00295329	0.065126
THNSL2	-0.90745	3.95E-05	0.00351259
AIFM2	-0.907618	0.00285349	0.0637811
LOC642852	-0.90798	0.00157546	0.0425344
RPS27A	-0.913186	0.000579379	0.0223938
PTER	-0.913332	0.000157105	0.0092827
IRF2BP2	-0.915342	0.000979648	0.0312924
PRODH	-0.916159	0.00517745	0.0914746
STX1A	-0.919568	0.0036103	0.0731402

RPL23	-0.922254	0.00858591	0.124148
MMP17	-0.92237	0.00823164	0.120887
SUMF1	-0.922475	0.000293425	0.0140542
HSPB1	-0.924254	0.0074468	0.11432
GNPTG	-0.925549	0.00291421	0.0647136
RPS15A	-0.928909	0.00185864	0.0472594
LFNG	-0.930159	0.00176133	0.045671
MYO15B	-0.930245	0.000818311	0.0278245
PIGS	-0.931796	0.000128927	0.0081883
CSAD	-0.932528	0.00383014	0.0754359
COL1A1	-0.934014	0.00317492	0.0676112
RPL23P8	-0.935384	0.00747409	0.114388
EFEMP2	-0.937716	0.00752085	0.11469
OLFM2	-0.938714	0.00864377	0.124472
RPS18P9	-0.940563	0.00239201	0.0564165
CD44	-0.948044	0.0023223	0.0554316
EFCAB4A	-0.949582	0.00270583	0.061433
GPR155	-0.950683	0.00048027	0.0200697
RFWD2	-0.950991	0.000121636	0.0078638
RGS19	-0.952292	0.000200093	0.0108858
STARD10	-0.952548	0.00853274	0.123736
EPHX1	-0.952892	0.00423729	0.0804052
RPS18	-0.953405	0.00132913	0.0385037
MXRA7	-0.95488	0.000743098	0.0263873
GNS	-0.955849	0.000882749	0.029324
MOSPD1	-0.95857	0.00222498	0.0540026
MAP4K1	-0.958994	0.0081834	0.120634
ID2	-0.96038	0.00583324	0.0978301
RPS12	-0.962015	0.000772408	0.0270287
NAMPT	-0.96587	0.000358702	0.0162623
MYOM1	-0.967669	0.00885991	0.126594
GOLGA8B	-0.972983	7.25E-05	0.00535713
S100A1	-0.973124	0.00418634	0.0800457
C3orf35	-0.976364	0.00253433	0.0586475
CHAD	-0.980183	0.00657448	0.10638
CCDC64B	-0.981041	0.000270551	0.013385
GUCY2EP	-0.981757	0.0114161	0.146992
RPS27	-0.983833	0.000475237	0.0199171
SPDYC	-0.986111	0.000677379	0.0247617

IQCK	-0.987574	0.000218215	0.0115662
RPL34	-0.989094	0.00537247	0.0934318
MLPH	-0.990662	8.44E-05	0.00590906
RPL31	-0.994335	0.00342567	0.0707562
NAAA	-1.00096	0.00095382	0.0309712
DCST2	-1.00229	0.000566348	0.0222603
ALDH2	-1.00486	0.0016326	0.0435231
RPL26	-1.00673	0.00136784	0.0388957
DDX11L2	-1.00957	0.00520051	0.0915947
TRIL	-1.00985	5.34E-05	0.00442803
TMC4	-1.01021	2.86E-06	0.000502644
MAPK10	-1.01055	0.00829117	0.121354
KRTCAP3	-1.01401	0.000198177	0.0108858
RIN2	-1.01653	0.00140473	0.0394172
LINC00893	-1.01855	0.00238948	0.0564165
XBP1	-1.02005	0.00373145	0.0742019
STC2	-1.02426	0.00167991	0.0441212
EEF1E1	-1.02847	0.00143709	0.0400746
PGRMC1	-1.02867	0.000460188	0.0195133
LOC440311	-1.02959	0.00575095	0.0975052
FGFR3	-1.03173	0.00913771	0.129536
ARHGEF25	-1.03224	6.97E-05	0.00523211
HPN	-1.03565	0.000680143	0.0247617
CCDC40	-1.03763	0.00245145	0.0574675
LOC100131320	-1.04094	0.00102121	0.0321225
ZFAS1	-1.04399	0.000381436	0.0169205
DNAI1	-1.04411	0.00140208	0.0394172
GTF2A2	-1.04747	1.21E-05	0.00148756
CNTD1	-1.05455	0.00326283	0.0687662
GLTSCR2	-1.05611	0.000957423	0.0309936
ST6GALNAC6	-1.05756	0.00101193	0.0320636
RPL37A	-1.05787	0.000810393	0.0277091
GSDMB	-1.05808	0.000922746	0.030166
COL18A1	-1.06388	0.0016597	0.0437439
FGF11	-1.06417	0.0019053	0.0482754
SKAP1	-1.07015	0.000274104	0.0133958
LINC00087	-1.07727	0.00218941	0.0532473
ARFGAP3	-1.07924	6.72E-07	0.000146826
HGD	-1.08141	0.0032528	0.0686612

SLC1A5	-1.0819	0.00106387	0.0332708
SNHG5	-1.08403	0.00064159	0.0238467
BAIAP3	-1.08448	0.000777146	0.0271134
PRRT1	-1.08472	0.00311484	0.0670097
RPL37	-1.08698	0.000473369	0.0198967
ASAH1	-1.08783	0.00164929	0.043629
RHOU	-1.09457	0.00806803	0.119422
LINC00514	-1.0977	0.0082502	0.121
PCBP2	-1.09863	3.34E-06	0.000566984
ALG14	-1.09957	2.51E-07	6.23E-05
MXI1	-1.10286	0.00625731	0.103217
VIPR1	-1.10562	2.27E-06	0.000414299
KNDC1	-1.10712	0.00459628	0.0844134
SLC6A9	-1.10858	0.003636	0.0732127
SYTL1	-1.10956	0.000145241	0.00887265
VWCE	-1.11062	0.00261443	0.0601621
TSPAN31	-1.12185	3.74E-08	1.80E-05
FTH1	-1.12387	0.00287664	0.0641989
SLC27A1	-1.12522	1.97E-06	0.000381033
BCL6	-1.12537	0.000664745	0.0245106
AKR1C3	-1.13002	0.00934875	0.131346
LOC100133669	-1.13215	0.000471711	0.019885
VTCN1	-1.13271	0.000570443	0.0222603
NOD2	-1.13338	0.000713588	0.0257195
FLJ10038	-1.13574	0.00502692	0.0896873
RINL	-1.13591	0.00313895	0.0671428
ABCC11	-1.13606	0.00661872	0.10644
GNG7	-1.13941	0.00112258	0.0346558
TPM2	-1.14109	0.00126124	0.0374143
RPL27	-1.14336	0.000538473	0.0215644
RNF152	-1.1521	0.000228973	0.0118319
FCGRT	-1.15335	0.00142114	0.0397066
LOC100289495	-1.16127	0.000435681	0.0188625
DLG5	-1.16893	5.34E-05	0.00442803
ADCY1	-1.16941	0.00833988	0.121797
RAB12	-1.17078	4.98E-06	0.000771838
MYB	-1.17293	0.00587418	0.0982461
MMP14	-1.17492	0.00830269	0.121399
KPNA7	-1.17495	0.00702075	0.110863

DGKQ	-1.17759	8.23E-05	0.00581668
DFNB59	-1.17925	0.00330501	0.069081
SCNN1B	-1.19126	0.00686366	0.10898
DUSP5	-1.19835	1.54E-05	0.00174887
C5orf63	-1.19878	0.00953257	0.13317
CAPN9	-1.19918	0.00199056	0.0495646
C20orf26	-1.20264	0.00734726	0.113773
RPL15	-1.20638	9.21E-05	0.00635042
IL13RA1	-1.21327	7.02E-09	4.47E-06
CD24	-1.22673	0.00405396	0.0785564
IDUA	-1.22883	4.66E-05	0.00392764
LINC00340	-1.22975	0.00155119	0.0421952
FAM107B	-1.23296	2.79E-05	0.00262869
CACNA2D4	-1.23473	0.00536865	0.0934318
ANXA2P1	-1.23522	0.00284772	0.0637508
TSPAN6	-1.23784	1.27E-05	0.00153859
10-Sep	-1.23823	8.71E-06	0.00117368
SLC9A2	-1.2391	0.00024965	0.0126288
CYP4X1	-1.24223	0.000785157	0.0272762
ATP6V1C2	-1.24454	0.000921935	0.030166
TMEM240	-1.25763	0.00411635	0.0791519
VMAC	-1.25997	5.01E-05	0.00419769
GSTA4	-1.2635	8.44E-06	0.00116287
GSDMC	-1.2643	1.43E-07	4.66E-05
FAM117B	-1.26511	0.00011521	0.00761919
NAALADL2	-1.26865	0.00372262	0.0741284
TPT1	-1.26894	0.000140403	0.00872416
LGI2	-1.27317	6.85E-05	0.00519978
ANXA2P2	-1.27443	4.63E-05	0.00392764
CALML6	-1.27533	0.000305648	0.0144476
FUT1	-1.27992	0.00343984	0.070846
FLJ34503	-1.28047	0.010323	0.138961
WIPF1	-1.28489	0.00244699	0.0574564
VEGFA	-1.30779	5.76E-05	0.00458818
ZAP70	-1.31493	0.00368738	0.0739057
ANXA2	-1.3162	2.08E-05	0.00215572
SPON2	-1.32275	0.00103764	0.032521
AGAP11	-1.32404	5.15E-06	0.000772727
DEPTOR	-1.32841	0.000869314	0.0291463

LOC399715	-1.33005	0.00117599	0.0359201
TMEM52	-1.33427	6.38E-05	0.00488968
GOLM1	-1.34756	0.00163455	0.0435231
GAS5	-1.35295	1.98E-05	0.00207867
NFU1	-1.35364	0.00151136	0.0415033
BSCL2	-1.35398	0.000679912	0.0247617
MET	-1.35988	0.000375836	0.0167236
ASS1	-1.36127	8.06E-05	0.00575433
IL20RA	-1.36143	4.57E-06	0.000731719
SSPO	-1.36149	1.70E-05	0.00188096
PPM1J	-1.36201	0.000493389	0.0203234
ANKRD30A	-1.36259	4.02E-06	0.000666107
AKR1B1	-1.36303	0.000984507	0.0313325
ENG	-1.36409	0.000121361	0.0078638
PCP2	-1.36979	1.02E-05	0.00130445
SLC30A8	-1.38176	0.000570072	0.0222603
CAPN12	-1.38288	3.41E-06	0.000571444
TSKU	-1.38485	0.000108727	0.00729075
PRKAG2-AS1	-1.38838	0.00557222	0.0955229
UBXN10	-1.39664	0.00151787	0.041564
CEACAM5	-1.39667	0.00302417	0.0657609
CSF3R	-1.39992	0.00636767	0.104194
APOD	-1.40486	3.23E-07	7.90E-05
YBX2	-1.41296	0.00666073	0.106816
LINC00672	-1.41381	0.00389263	0.0763696
CPNE7	-1.42467	0.00379845	0.0749141
FLJ23867	-1.42536	9.96E-05	0.00675046
SLCO2B1	-1.44026	0.00253379	0.0586475
GSTA2	-1.4422	0.0089238	0.127105
ZNF558	-1.44307	1.86E-07	5.15E-05
RND2	-1.45169	0.00968637	0.134148
RBM20	-1.45878	0.0113078	0.145948
TLR9	-1.4597	0.0115101	0.147635
CCDC146	-1.46238	0.000155446	0.00926059
AKAP6	-1.46666	0.00138119	0.0391981
SLC5A7	-1.47417	0.00736741	0.113797
UPP2	-1.478	0.000499226	0.020447
CBFA2T3	-1.48284	0.00165862	0.0437439
CLEC1A	-1.48307	0.0115076	0.147635

AKR1C1	-1.5032	0.000630652	0.0236159
NAGS	-1.50455	0.00707616	0.111372
CECR1	-1.51333	4.33E-06	0.00070148
GPR87	-1.51427	0.00159296	0.0428466
UCN2	-1.52429	0.000960957	0.0309936
OPRL1	-1.53427	0.00179402	0.046269
GABRD	-1.54356	1.18E-05	0.00147866
TMSB15A	-1.55548	0.0112936	0.145948
COL3A1	-1.55974	0.00726684	0.113383
FBLN2	-1.56662	0.00864173	0.124472
CPAMD8	-1.57964	2.79E-05	0.00262869
BCRP3	-1.58448	0.000302937	0.0143879
SLPI	-1.6058	0.000148993	0.00902533
C20orf203	-1.6066	0.0101706	0.137422
STAT4	-1.60983	0.00885457	0.126594
CEACAM6	-1.64313	0.0016832	0.0441212
RASD2	-1.64983	6.44E-06	0.000927914
LINC00930	-1.65614	0.00130043	0.038029
LINC00304	-1.65717	6.95E-07	0.000149632
MUCL1	-1.66151	0.000280596	0.0136014
TMEM59L	-1.68435	0.00398697	0.0776758
LOC100289187	-1.68549	0.000354867	0.0161392
XYLT1	-1.68764	0.00528774	0.0925161
KRT38	-1.69434	0.0066281	0.10644
CHI3L2	-1.70273	0.000373164	0.0166561
KRT17	-1.73186	0.00485103	0.0873125
C6orf141	-1.7349	1.59E-05	0.00178818
SPINK8	-1.82077	0.000197177	0.0108858
CLEC7A	-1.82225	3.86E-05	0.00346126
LRRC32	-1.8311	1.21E-05	0.00148756
PEBP4	-1.87085	0.000439683	0.0189389
AZGP1	-1.96749	1.93E-07	5.15E-05
ATAD3C	-1.97713	1.67E-08	9.62E-06
ABCC12	-1.98199	7.85E-06	0.00111419
GJC3	-1.98811	8.91E-06	0.00117606
PDE2A	-2.00118	4.58E-09	3.67E-06
TBX4	-2.03199	0.000193358	0.0108468
RASIP1	-2.06261	1.49E-05	0.0017212
AGT	-2.07213	7.88E-06	0.00111419

LEMD1	-2.07605	0.000171779	0.00989204
RASD1	-2.09605	1.85E-05	0.00200859
KRT16	-2.11607	0.000332786	0.0153665
IZUMO1	-2.15674	2.91E-05	0.00272747
AKR1C2	-2.20714	1.67E-07	5.13E-05
SHISA9	-2.71682	3.08E-10	4.44E-07
PRSS3	-3.10627	5.24E-09	3.78E-06
LRRK2	2.76706	3.30E-12	4.45E-08
FRMPD4	2.74108	2.32E-07	0.000802072
SNPH	2.50413	2.09E-06	0.00270252
GPR158	2.378	3.92E-05	0.0154172
LOC645638	2.35482	1.99E-05	0.0112465
SYNE1	2.21137	2.21E-06	0.00270252
CYP2C8	2.1685	6.23E-05	0.0204647
STRA6	2.16048	7.40E-06	0.00532797
STAB2	2.1565	4.68E-05	0.016587
CNIH2	2.11714	0.000126763	0.0284583
ABI3BP	2.0586	0.00010475	0.0271209
RAD21L1	1.98781	5.53E-05	0.018631
RCBTB2	1.98087	1.27E-08	8.54E-05
C1orf95	1.98026	3.59E-07	0.000806297
NELL2	1.94265	0.000149532	0.032487
TRIM22	1.91057	0.000175449	0.0363583
F13A1	1.9037	2.17E-05	0.0112465
NXPE3	1.87834	1.02E-06	0.00152124
CAMK2N2	1.85558	0.000238771	0.0428833
REPS2	1.84049	0.000182849	0.0367608
PKD1L1	1.81587	4.48E-05	0.016587
PFN1P2	1.78891	0.00129276	0.112345
COLEC12	1.76561	0.00195515	0.134955
SYT1	1.76509	0.000101975	0.0269334
GPER	1.74572	2.45E-05	0.0117748
PART1	1.72574	2.82E-05	0.0126725
SLC26A3	1.71116	0.00218333	0.140784
ALDOC	1.6669	0.000193919	0.0373155
ADAMTS12	1.64802	7.52E-06	0.00532797
DHFR	1.63849	0.00110611	0.106424
GHR	1.62505	2.98E-07	0.000802072
CXCL17	1.61766	0.00118091	0.109702

MPP1	1.58976	0.000867624	0.0933047
SCGN	1.57411	0.0012729	0.111561
KIAA1324L	1.5658	0.000235159	0.0428833
CES1	1.54048	0.00126269	0.111561
IL1RAPL2	1.53265	0.000939602	0.0977896
PFKFB1	1.51988	2.77E-07	0.000802072
PCDHGC3	1.51562	0.000584842	0.0748657
HMGN5	1.51503	0.00115085	0.108459
TRO	1.50536	3.89E-05	0.0154172
FAM111B	1.50047	0.000116779	0.0271209
LOC729041	1.47244	0.000619766	0.0752095
CCDC151	1.46632	0.00127546	0.111561
ELOVL6	1.45665	0.00196953	0.134955
SLC2A6	1.45617	0.000341334	0.0553948
C15orf48	1.42316	0.00120608	0.110862
SRGAP2	1.4009	0.00115948	0.108459
CCDC136	1.37201	0.00171398	0.127298
APOL3	1.34773	0.000841647	0.0921706
TTLL7	1.34727	4.55E-07	0.000876191
VEPH1	1.33362	0.000528223	0.0712776
GATA3-AS1	1.32512	0.00168728	0.126265
OXCT1	1.32086	2.17E-05	0.0112465
MLLT11	1.30676	8.06E-05	0.0221502
ZBED5-AS1	1.29923	4.01E-05	0.0154172
ABCA5	1.29692	7.72E-05	0.0221184
WDFY3-AS2	1.28864	0.00124644	0.111561
METTL7A	1.28604	4.61E-05	0.016587
OBSCN	1.26239	0.000237519	0.0428833
TTC18	1.21929	0.0021345	0.139571
FAM111A	1.21626	0.000111736	0.0271209
PLCD4	1.20247	5.40E-05	0.018631
INSIG1	1.1982	0.000669736	0.0798349
NBPF10	1.19497	0.000438354	0.065607
FFAR2	1.18962	0.000605555	0.0748657
STK10	1.18577	0.000249119	0.0435797
KRT7	1.17234	0.000448825	0.0658022
LRP8	1.14967	0.000430357	0.065607
SORBS1	1.14157	0.00149835	0.118028
STX1B	1.12764	0.00135159	0.112906

GIPR	1.12483	0.000872783	0.0933047
MAGEH1	1.08837	0.000436995	0.065607
LOC729737	1.07951	0.00125317	0.111561
TMEM64	1.0724	0.00211917	0.139245
APOL1	1.07145	0.00115923	0.108459
KMO	1.06589	0.00259884	0.148332
CLN6	1.06275	0.00219962	0.14109
CMYA5	1.05754	0.00181007	0.12969
GOLGA2P5	1.0374	2.60E-05	0.0120785
SCAMP5	1.02841	0.00132963	0.112906
C9orf96	1.02596	0.00178834	0.129511
DOC2A	1.00935	0.000497226	0.0705013
MYO16	0.998879	3.91E-05	0.0154172
MVK	0.974622	2.30E-05	0.0114715
NOTCH2NL	0.968749	0.000916966	0.0964964
COL12A1	0.958922	0.00188419	0.133579
ENPP5	0.957253	0.00051543	0.0712776
COQ10A	0.946274	0.00104552	0.103366
PTPN14	0.945855	0.00145369	0.116555
LPIN1	0.933355	0.00126339	0.111561
AP4B1	0.93218	0.00231812	0.143265
ELOVL7	0.927248	0.00232847	0.143265
SLC36A1	0.917707	0.00136242	0.112906
COPG1	0.916752	2.54E-06	0.00285427
C1orf51	0.900304	0.000734128	0.0819686
LOC145837	0.895014	0.000600821	0.0748657
LCP1	0.887433	0.0020753	0.137787
ALDH3B2	0.881453	0.0020559	0.137776
PTPRU	0.871198	0.000578919	0.0748657
PC	0.870815	0.00224889	0.142889
GOLIM4	0.863273	0.00134631	0.112906
FKBP10	0.859984	0.000547749	0.0730513
PFKM	0.853081	3.73E-05	0.0154172
BTN2A2	0.83724	0.000454313	0.0658022
EML3	0.828823	0.00100139	0.100663
SPIN3	0.81381	0.00240469	0.144295
ALDH3B1	0.784482	0.000969513	0.098673
TOMM40L	0.782841	0.000358117	0.0574267
NUCB1	0.766174	0.000705174	0.0818853

PDE4DIP	0.754244	0.00179908	0.129592
SLC25A35	0.713699	0.00235886	0.143773
PLEKHA5	0.673017	0.00255097	0.148332
SLC26A2	0.668646	0.0016709	0.126265
KIAA1598	-0.642883	0.00250821	0.148182
CALM2	-0.664414	0.001518	0.118881
TMEM9	-0.683076	0.000606014	0.0748657
PCNP	-0.734345	0.00201943	0.136692
MISP	-0.738468	0.00256098	0.148332
SEH1L	-0.743447	0.00257151	0.148332
S100A14	-0.748624	0.00172944	0.127298
DYNLT3	-0.754159	0.00215045	0.139935
GMFB	-0.758642	0.00250101	0.148182
JDP2	-0.760815	0.000663094	0.0797489
PQLC3	-0.774298	0.00259749	0.148332
YIPF4	-0.780146	0.00105293	0.103366
RPS19	-0.781687	0.00105285	0.103366
GLUL	-0.786898	0.00157996	0.122311
CACNA1I	-0.791745	0.00229149	0.143265
CLDND1	-0.83151	0.000974277	0.098673
RPS5	-0.832131	0.00207652	0.137787
AK3	-0.83293	0.00068064	0.080177
MEX3D	-0.841342	0.0021844	0.140784
RPS11	-0.849103	0.00196027	0.134955
OSTC	-0.851914	0.00138304	0.112906
SPOPL	-0.852034	0.000340508	0.0553948
STK17A	-0.85637	0.00255334	0.148332
WHSC1L1	-0.856384	0.000948247	0.0977896
STX8	-0.861365	0.00059298	0.0748657
LGALS3	-0.862214	0.000435065	0.065607
PAFAH1B2	-0.881254	0.000174684	0.0363583
FAM173B	-0.902263	0.000376932	0.0595395
CNIH	-0.904576	0.0023178	0.143265
PKP4	-0.9048	7.61E-05	0.0221184
MGLL	-0.914809	0.000255664	0.0441511
RPS16	-0.916005	0.000736318	0.0819686
C1orf198	-0.918594	0.00186932	0.133226
SLC1A4	-0.922634	0.000286029	0.0484305
RSL24D1	-0.932577	0.0017577	0.128478

VMA21	-0.934748	6.48E-05	0.0207928
CCDC71L	-0.938266	0.00233345	0.143265
IRF2BP2	-0.962077	0.000287635	0.0484305
TNFAIP8	-0.964265	0.00131034	0.112906
C5orf15	-0.965323	0.00111409	0.106431
RAB12	-0.966198	0.0013243	0.112906
CPNE3	-0.987221	0.000684511	0.080177
SMIM20	-0.987408	0.00176454	0.128478
SOX4	-0.987627	3.91E-06	0.00344533
EIF3E	-1.01053	0.00105898	0.103366
PTAFR	-1.01554	0.00246364	0.146837
JUND	-1.01715	1.29E-05	0.00828583
SLC1A5	-1.01762	0.00165503	0.125951
MTHFD2	-1.01767	0.000451547	0.0658022
CLIC4	-1.02756	0.00258871	0.148332
FCHO1	-1.02766	0.000121693	0.0277832
TSPAN6	-1.02919	0.00222055	0.141758
CDK14	-1.03464	6.83E-06	0.00532797
MYC	-1.03621	0.000523983	0.0712776
FAM167A	-1.03802	0.00233068	0.143265
EEF2	-1.03928	0.00136674	0.112906
EEF1E1	-1.04186	0.000906826	0.0961807
TES	-1.05203	0.00120985	0.110862
CAMK2N1	-1.05481	0.00195236	0.134955
AHR	-1.05641	0.00139916	0.113435
EPB41L4A-AS1	-1.07519	0.00159936	0.123105
AZGP1	-1.07615	0.000529158	0.0712776
CHST8	-1.07682	0.00168704	0.126265
IQCK	-1.08487	0.000111643	0.0271209
CPT1A	-1.08946	2.00E-05	0.0112465
CDC42SE2	-1.09201	9.05E-05	0.0243893
SLC39A10	-1.09217	0.00148624	0.117763
CXXC5	-1.09429	0.000114125	0.0271209
HSPB1	-1.11089	0.00200732	0.136558
SIAH1	-1.11234	1.99E-05	0.0112465
GADD45B	-1.12989	0.000187237	0.0370354
PIK3CD	-1.13908	0.00237842	0.144295
IER3	-1.14482	0.000853719	0.0927387
OSGIN1	-1.1722	0.00137588	0.112906

ZFAS1	-1.17705	0.00155992	0.121458
ETNK2	-1.17949	0.00146366	0.11666
ARL4C	-1.18118	0.0022851	0.143265
GAS5	-1.18155	0.00031828	0.0529288
TAF9B	-1.2217	7.03E-05	0.0215157
FKBP4	-1.2236	0.00134798	0.112906
FHL2	-1.23093	0.00239666	0.144295
GALNT3	-1.24396	0.000520415	0.0712776
IGFBP4	-1.24483	0.0019435	0.134955
AKR1B10	-1.25354	0.00190907	0.134634
EGR1	-1.27487	0.00241027	0.144295
EIF4EBP1	-1.30086	0.00016633	0.035563
CPAMD8	-1.31017	1.23E-05	0.00825541
HR	-1.31181	0.00172091	0.127298
SNHG5	-1.33353	0.000115947	0.0271209
B3GNT3	-1.33353	0.000759262	0.0838299
MYB	-1.34323	6.85E-05	0.0214595
PDE2A	-1.34822	0.00162782	0.124583
LFNG	-1.35933	4.09E-06	0.00344533
NPNT	-1.3873	0.000496431	0.0705013
RASL11A	-1.3894	0.000565692	0.0739793
CA8	-1.3981	0.000220382	0.0418105
CEBPB	-1.44585	7.72E-05	0.0221184
VTCN1	-1.44926	0.000380134	0.0595395
MTHFD1L	-1.45764	0.000247692	0.0435797
KLF4	-1.46551	0.00197374	0.134955
GJA1	-1.47487	0.00233989	0.143265
TIAM1	-1.48433	0.000189713	0.0370354
RNF165	-1.54648	0.00258722	0.148332
TMEM240	-1.55546	0.000714406	0.0819686
INSM1	-1.56465	0.000951035	0.0977896
EGFL6	-1.56486	0.00211119	0.139245
RAET1E	-1.58735	0.00261504	0.148627
LIMCH1	-1.59506	0.000112518	0.0271209
GFRA1	-1.60909	0.000720935	0.0819686
DGAT2	-1.60923	0.00107788	0.104454
GPR87	-1.61354	0.000725571	0.0819686
C18orf32	-1.6184	0.00204502	0.137732
SLC6A9	-1.64652	6.08E-07	0.00102368

LOC100133669	-1.68275	3.85E-06	0.00344533
FLRT3	-1.71496	4.03E-06	0.00344533
TAT	-1.7276	0.000611376	0.0748657
FOXP2	-1.80867	0.00124289	0.111561
C3	-1.82133	0.00140636	0.113435
DHRS2	-1.8366	0.000178911	0.0365141
ANXA1	-1.8424	0.000231267	0.0428833
GPR124	-1.85838	0.000561842	0.0739793
TFF2	-2.25491	0.000144108	0.0318219
KRT16	-2.30423	7.97E-05	0.0221502

REFERENCES

ACS (2016). "American Cancer Society."

Aguirre-Ghiso, J. A. (2007). "Models, mechanisms and clinical evidence for cancer dormancy." Nat Rev Cancer **7**(11): 834-846.

Al-Hajj, M., M. S. Wicha, A. Benito-Hernandez, S. J. Morrison and M. F. Clarke (2003). "Prospective identification of tumorigenic breast cancer cells." Proc Natl Acad Sci U S A **100**(7): 3983-3988.

Allfrey, V. G., R. Faulkner and A. E. Mirsky (1964). "Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis." Proc Natl Acad Sci U S A **51**: 786-794.

Angrand, P. O., F. Apiou, A. F. Stewart, B. Dutrillaux, R. Losson and P. Chambon (2001). "NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines." Genomics **74**(1): 79-88.

Bagaria, S. P., P. S. Ray, M. S. Sim, X. Ye, J. M. Shamonki, X. Cui and A. E. Giuliano (2014). "Personalizing breast cancer staging by the inclusion of ER, PR, and HER2." JAMA Surg **149**(2): 125-129.

Barr, M. L. and E. G. Bertram (1949). "A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis." Nature **163**(4148): 676.

Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev and K. Zhao (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-837.

Baselga, J., L. Norton, J. Albanell, Y. M. Kim and J. Mendelsohn (1998). "Recombinant humanized anti-HER2 antibody (Herceptin) enhances the antitumor activity of paclitaxel and doxorubicin against HER2/neu overexpressing human breast cancer xenografts." Cancer Res **58**(13): 2825-2831.

Bell, O., C. Wirbelauer, M. Hild, A. N. D. Scharf, M. Schwaiger, D. M. MacAlpine, F. Zilbermann, F. van Leeuwen, S. P. Bell, A. Imhof, D. Garza, A. H. F. M. Peters and D. Schübeler (2007). "Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila." The EMBO Journal **26**(24): 4974.

Berger, S. L. (2007). "The complex language of chromatin regulation during transcription." Nature **447**(7143): 407-412.

Berger, S. L., T. Kouzarides, R. Shiekhattar and A. Shilatifard (2009). "An operational definition of epigenetics." Genes Dev **23**(7): 781-783.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber and E. S. Lander (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-326.

Bertucci, F., B. Orsetti, V. Negre, P. Finetti, C. Rouge, J. C. Ahomadegbe, F. Bibeau, M. C. Mathieu, I. Treilleux, J. Jacquemier, L. Ursule, A. Martinec, Q.

Wang, J. Benard, A. Puisieux, D. Birnbaum and C. Theillet (2008). "Lobular and ductal carcinomas of the breast have distinct genomic and expression profiles." Oncogene **27**(40): 5359-5372.

Bilal, E., K. Vassallo, D. Toppmeyer, N. Barnard, I. H. Rye, V. Almendro, H. Russnes, A. L. Borresen-Dale, A. J. Levine, G. Bhanot and S. Ganesan (2012). "Amplified loci on chromosomes 8 and 17 predict early relapse in ER-positive breast cancers." PLoS One **7**(6): e38575.

Bird, A. (2002). "DNA methylation patterns and epigenetic memory." Genes & development **16**(1): 6-21.

Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." Nature **321**(6067): 209-213.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.

Brownell, J. E., J. Zhou, T. Ranalli, R. Kobayashi, D. G. Edmondson, S. Y. Roth and C. D. Allis (1996). "Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation." Cell **84**(6): 843-851.

Busch, H. and I. L. Goldknopf (1981). "Ubiquitin - protein conjugates." Mol Cell Biochem **40**(3): 173-187.

Callahan, R. and S. Hurvitz (2011). "Human epidermal growth factor receptor-2-positive breast cancer: Current management of early, advanced, and recurrent disease." Curr Opin Obstet Gynecol **23**(1): 37-43.

Cannistra, S. A. (1990). "Chronic myelogenous leukemia as a model for the genetic basis of cancer." Hematol Oncol Clin North Am **4**(2): 337-357.

Carmen, A. A., S. E. Rundlett and M. Grunstein (1996). "HDA1 and HDA3 are components of a yeast histone deacetylase (HDA) complex." J Biol Chem **271**(26): 15837-15844.

Chen, J., E. E. Bardes, B. J. Aronow and A. G. Jegga (2009). "ToppGene Suite for gene list enrichment analysis and candidate gene prioritization." Nucleic Acids Res **37**(Web Server issue): W305-311.

Ciriello, G., M. L. Miller, B. A. Aksoy, Y. Senbabaoglu, N. Schultz and C. Sander (2013). "Emerging landscape of oncogenic signatures across human cancers." Nat Genet **45**(10): 1127-1133.

Clapier, C. R. and B. R. Cairns (2009). "The biology of chromatin remodeling complexes." Annual review of biochemistry **78**: 273-304.

Curtis, C., S. P. Shah, S. F. Chin, G. Turashvili, O. M. Rueda, M. J. Dunning, D. Speed, A. G. Lynch, S. Samarajiwa, Y. Yuan, S. Graf, G. Ha, G. Haffari, A. Bashashati, R. Russell, S. McKinney, M. Group, A. Langerod, A. Green, E. Provenzano, G. Wishart, S. Pinder, P. Watson, F. Markowetz, L. Murphy, I. Ellis, A. Purushotham, A. L. Borresen-Dale, J. D. Brenton, S. Tavare, C. Caldas and S. Aparicio (2012). "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups." Nature **486**(7403): 346-352.

Deans, C. and K. A. Maggert (2015). "What do you mean, "epigenetic"?" Genetics **199**(4): 887-896.

di Luccio, E. and M. Morishita (2012). "Structural insights into the regulation and the recognition of histone marks by the SET domain of histone lysine HMTase NSD1, NSD2/MMSET/WHSC1, and NSD3/WHSC1L1." Cancer Research **72**(8 Supplement): 1060-1060.

Dontu, G., W. M. Abdallah, J. M. Foley, K. W. Jackson, M. F. Clarke, M. J. Kawamura and M. S. Wicha (2003). "In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells." Genes Dev **17**(10): 1253-1270.

Dontu, G., K. W. Jackson, E. McNicholas, M. J. Kawamura, W. M. Abdallah and M. S. Wicha (2004). "Role of Notch signaling in cell-fate determination of human mammary stem/progenitor cells." Breast Cancer Res **6**(6): R605-615.

Douglas, J., K. Coleman, K. Tatton-Brown, H. E. Hughes, I. K. Temple, T. R. Cole, N. Rahman and C. Childhood Overgrowth (2005). "Evaluation of NSD2 and NSD3 in overgrowth syndromes." Eur J Hum Genet **13**(2): 150-153.

Douglas, J., S. Hanks, I. K. Temple, S. Davies, A. Murray, M. Upadhyaya, S. Tomkins, H. E. Hughes, T. R. Cole and N. Rahman (2003). "NSD1 mutations are the major cause of Sotos syndrome and occur in some cases of Weaver syndrome but are rare in other overgrowth phenotypes." Am J Hum Genet **72**(1): 132-143.

Drane, P., K. Ouararhni, A. Depaux, M. Shuaib and A. Hamiche (2010). "The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3." Genes Dev **24**(12): 1253-1265.

Ethier, S. P., C. Chiodino and R. F. Jones (1990). "Role of growth factor synthesis in the acquisition of insulin/insulin-like growth factor I independence in rat mammary carcinoma cells." Cancer Res **50**(17): 5351-5357.

Fearon, E. R., S. R. Hamilton and B. Vogelstein (1987). "Clonal analysis of human colorectal tumors." Science **238**(4824): 193-197.

Fearon, E. R. and B. Vogelstein (1990). "A genetic model for colorectal tumorigenesis." Cell **61**(5): 759-767.

Feinberg, A. P. and B. Vogelstein (1983). "Hypomethylation distinguishes genes of some human cancers from their normal counterparts." Nature **301**(5895): 89-92.

Felsenfeld, G. (2014). "The evolution of epigenetics." Perspect Biol Med **57**(1): 132-148.

Feng, J., T. Liu, B. Qin, Y. Zhang and X. S. Liu (2012). "Identifying ChIP-seq enrichment using MACS." Nature Protocols **7**: 1728+.

Feng, J. J. (2011). Using MACS to identify peaks from ChIP-Seq data**:** Unit-2.14.

Feng, Q., Z. Zhang, M. J. Shea, C. J. Creighton, C. Coarfa, S. G. Hilsenbeck, R. Lanz, B. He, L. Wang, X. Fu, A. Nardone, Y. Song, J. Bradner, N. Mitsiades, C. S. Mitsiades, C. K. Osborne, R. Schiff and B. W. O'Malley (2014). "An epigenomic approach to therapy for tamoxifen-resistant breast cancer." Cell Res **24**(7): 809-819.

Forozan, F., E. H. Mahlamaki, O. Monni, Y. Chen, R. Veldman, Y. Jiang, G. C. Gooden, S. P. Ethier, A. Kallioniemi and O. P. Kallioniemi (2000). "Comparative

genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data." Cancer Res **60**(16): 4519-4525.

Forozan, F., R. Veldman, C. A. Ammerman, N. Z. Parsa, A. Kallioniemi, O. P. Kallioniemi and S. P. Ethier (1999). "Molecular cytogenetic analysis of 11 new breast cancer cell lines." Br J Cancer **81**(8): 1328-1334.

Frame, M. C. (2002). "Src in cancer: deregulation and consequences for cell behaviour." Biochim Biophys Acta **1602**(2): 114-130.

Garcia, M. J., J. C. Pole, S. F. Chin, A. Teschendorff, A. Naderi, H. Ozdag, M. Vias, T. Kranjac, T. Subkhankulova, C. Paish, I. Ellis, J. D. Brenton, P. A. Edwards and C. Caldas (2005). "A 1 Mb minimal amplicon at 8p11-12 in breast cancer identifies new candidate oncogenes." Oncogene **24**(33): 5235-5245.

Gelsi-Boyer, V., B. Orsetti, N. Cervera, P. Finetti, F. Sircoulomb, C. Rouge, L. Lasorsa, A. Letessier, C. Ginestier, F. Monville, S. Esteyries, J. Adelaide, B. Esterni, C. Henry, S. P. Ethier, F. Bibeau, M. J. Mozziconacci, E. Charafe-Jauffret, J. Jacquemier, F. Bertucci, D. Birnbaum, C. Theillet and M. Chaffanet (2005). "Comprehensive profiling of 8p11-12 amplification in breast cancer." Mol Cancer Res **3**(12): 655-667.

Goren, A., F. Ozsolak, N. Shoresh, M. Ku, M. Adli, C. Hart, M. Gymrek, O. Zuk, A. Regev, P. M. Milos and B. E. Bernstein (2010). "Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA." Nat Methods **7**(1): 47-49.

Goto, H., Y. Tomono, K. Ajiro, H. Kosako, M. Fujita, M. Sakurai, K. Okawa, A. Iwamatsu, T. Okigaki and T. Takahashi (1999). "Identification of a novel phosphorylation site on histone H3 coupled with mitotic chromosome condensation." Journal of Biological Chemistry **274**(36): 25543-25549.

Gururaj, A. E., S. K. Rayala, R. K. Vadlamudi and R. Kumar (2006). "Novel mechanisms of resistance to endocrine therapy: genomic and nongenomic considerations." Clin Cancer Res **12**(3 Pt 2): 1001s-1007s.

He, C., F. Li, J. Zhang, J. Wu and Y. Shi (2013). "The methyltransferase NSD3 has chromatin-binding motifs, PHD5-C5HCH, that are distinct from other NSD (nuclear receptor SET domain) family members in their histone H3 recognition." J Biol Chem **288**(7): 4692-4703.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities." Molecular Cell **38**(4): 576-589.

Ho, L. and G. R. Crabtree (2010). "Chromatin remodelling during development." Nature **463**(7280): 474-484.

Hojfeldt, J. W., K. Agger and K. Helin (2013). "Histone lysine demethylases as targets for anticancer therapy." Nat Rev Drug Discov **12**(12): 917-930.

Howman, E. V., K. J. Fowler, A. J. Newson, S. Redward, A. C. MacDonald, P. Kalitsis and K. H. Choo (2000). "Early disruption of centromeric chromatin

organization in centromere protein A (Cenpa) null mice." Proc Natl Acad Sci U S A **97**(3): 1148-1153.

Huang, Z., H. Wu, S. Chuai, F. Xu, F. Yan, N. Englund, Z. Wang, H. Zhang, M. Fang, Y. Wang, J. Gu, M. Zhang, T. Yang, K. Zhao, Y. Yu, J. Dai, W. Yi, S. Zhou, Q. Li, J. Wu, J. Liu, X. Wu, H. Chan, C. Lu, P. Atadja, E. Li, Y. Wang and M. Hu (2013). "NSD2 is recruited through its PHD domain to oncogenic gene loci to drive multiple myeloma." Cancer Res **73**(20): 6277-6288.

Ignatoski, K. M., A. J. Lapointe, E. H. Radany and S. P. Ethier (1999). "erbB-2 overexpression in human mammary epithelial cells confers growth factor independence." Endocrinology **140**(8): 3615-3622.

Jaju, R. J., C. Fidler, O. A. Haas, A. J. Strickson, F. Watkins, K. Clark, N. C. Cross, J. F. Cheng, P. D. Aplan, L. Kearney, J. Boultwood and J. S. Wainscoat (2001). "A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia." Blood **98**(4): 1264-1267.

Jones, P. A. (2012). "Functions of DNA methylation: islands, start sites, gene bodies and beyond." Nature Reviews Genetics **13**(7): 484-492.

Kayne, P. S., U. J. Kim, M. Han, J. R. Mullen, F. Yoshizaki and M. Grunstein (1988). "Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast." Cell **55**(1): 27-39.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley and S. L. Salzberg (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." Genome Biol **14**(4): R36.

Kim, S. M., H. J. Kee, N. Choe, J. Y. Kim, H. Kook and S. B. Seo (2007). "The histone methyltransferase activity of WHISTLE is important for the induction of apoptosis and HDAC1-mediated transcriptional repression." Exp Cell Res **313**(5): 975-983.

Kim, S. M., H. J. Kee, G. H. Eom, N. W. Choe, J. Y. Kim, Y. S. Kim, S. K. Kim, H. Kook and S. B. Seo (2006). "Characterization of a novel WHSC1-associated SET domain protein with H3K4 and H3K27 methyltransferase activity." Biochem Biophys Res Commun **345**(1): 318-323.

King, C. R., M. H. Kraus and S. A. Aaronson (1985). "Amplification of a novel v-erbB-related gene in a human mammary carcinoma." Science **229**(4717): 974-976.

Ku, M., R. P. Koche, E. Rheinbay, E. M. Mendenhall, M. Endoh, T. S. Mikkelsen, A. Presser, C. Nusbaum, X. Xie, A. S. Chi, M. Adli, S. Kasif, L. M. Ptaszek, C. A. Cowan, E. S. Lander, H. Koseki and B. E. Bernstein (2008). "Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains." PLoS Genetics **4**(10): e1000242.

Kurotaki, N., K. Imaizumi, N. Harada, M. Masuno, T. Kondoh, T. Nagai, H. Ohashi, K. Naritomi, M. Tsukahara, Y. Makita, T. Sugimoto, T. Sonoda, T. Hasegawa, Y. Chinen, H. A. Tomita Ha, A. Kinoshita, T. Mizuguchi, K. Yoshiura Ki, T. Ohta, T. Kishino, Y. Fukushima, N. Niikawa and N. Matsumoto (2002). "Haploinsufficiency of NSD1 causes Sotos syndrome." Nat Genet **30**(4): 365-366.

Lewis, G. D., I. Figari, B. Fendly, W. L. Wong, P. Carter, C. Gorman and H. M. Shepard (1993). "Differential responses of human tumor cell lines to anti-p185HER2 monoclonal antibodies." Cancer Immunol Immunother **37**(4): 255-263.

Li, W., F. Li, Q. Huang, J. Shen, F. Wolf, Y. He, X. Liu, Y. A. Hu, J. S. Bedford and C.-Y. Li (2011). "Quantitative, noninvasive imaging of radiation-induced DNA double-strand breaks in vivo." Cancer research **71**(12): 4130-4137.

Liu, G., A. Bollig-Fischer, B. Kreike, M. J. van de Vijver, J. Abrams, S. P. Ethier and Z. Q. Yang (2009). "Genomic amplification and oncogenic properties of the GASC1 histone demethylase gene in breast cancer." Oncogene **28**(50): 4491-4500.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-260.

Maherali, N., R. Sridharan, W. Xie, J. Utikal, S. Eminli, K. Arnold, M. Stadtfeld, R. Yachechko, J. Tchieu, R. Jaenisch, K. Plath and K. Hochedlinger (2007). "Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution." Cell Stem Cell **1**(1): 55-70.

McGuire, W. L. (1973). "Estrogen receptors in human breast cancer." J Clin Invest **52**(1): 73-77.

Meilinger, D., K. Fellinger, S. Bultmann, U. Rothbauer, I. M. Bonapace, W. E. Klinkert, F. Spada and H. Leonhardt (2009). "Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells." EMBO Rep **10**(11): 1259-1264.

Metzger, E., N. Yin, M. Wissmann, N. Kunowska, K. Fischer, N. Friedrichs, D. Patnaik, J. M. Higgins, N. Potier, K. H. Scheidtmann, R. Buettner and R. Schule (2008). "Phosphorylation of histone H3 at threonine 11 establishes a novel chromatin mark for transcriptional regulation." Nat Cell Biol **10**(1): 53-60.

Min, D. J., T. Ezponda, M. K. Kim, C. M. Will, E. Martinez-Garcia, R. Popovic, V. Basrur, K. S. Elenitoba-Johnson and J. D. Licht (2013). "MMSET stimulates myeloma cell growth through microRNA-mediated modulation of c-MYC." Leukemia **27**(3): 686-694.

Nimura, K., K. Ura, H. Shiratori, M. Ikawa, M. Okabe, R. J. Schwartz and Y. Kaneda (2009). "A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome." Nature **460**(7252): 287-291.

Parker, R. C., H. E. Varmus and J. M. Bishop (1981). "Cellular homologue (c-src) of the transforming gene of Rous sarcoma virus: isolation, mapping, and transcriptional analysis of c-src and flanking regions." Proc Natl Acad Sci U S A **78**(9): 5842-5846.

Perez, E. A., E. H. Romond, V. J. Suman, J. H. Jeong, G. Sledge, C. E. Geyer, Jr., S. Martino, P. Rastogi, J. Gralow, S. M. Swain, E. P. Winer, G. Colon-Otero, N. E. Davidson, E. Mamounas, J. A. Zujewski and N. Wolmark (2014). "Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831." J Clin Oncol **32**(33): 3744-3752.

Perou, C. M., T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown and D. Botstein (2000). "Molecular portraits of human breast tumours." Nature **406**(6797): 747-752.

Pole, J. C., C. Courtay-Cahen, M. J. Garcia, K. A. Blood, S. L. Cooke, A. E. Alsop, D. M. Tse, C. Caldas and P. A. Edwards (2006). "High-resolution analysis of chromosome rearrangements on 8p in breast, colon and pancreatic cancer reveals a complex pattern of loss, gain and translocation." Oncogene **25**(41): 5693-5706.

Prat, A. and C. M. Perou (2011). "Deconstructing the molecular portraits of breast cancer." Mol Oncol **5**(1): 5-23.

Rahman, S., M. E. Sowa, M. Ottinger, J. A. Smith, Y. Shi, J. W. Harper and P. M. Howley (2011). "The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3." Mol Cell Biol **31**(13): 2641-2652.

Riggs, A. D. (1975). "X inactivation, differentiation, and DNA methylation." Cytogenetic and Genome Research **14**(1): 9-25.

Rohs, R., S. M. West, A. Sosinsky, P. Liu, R. S. Mann and B. Honig (2009). "The role of DNA shape in protein-DNA recognition." Nature **461**(7268): 1248-1253.

Rose, C., C. Kamby, H. T. Mouridsen, L. Bastholt, H. Brincker, H. Skovgaard-Poulsen, A. P. Andersen, H. Loft, P. Dombernowsky and K. W. Andersen (1986). "Combined endocrine treatment of postmenopausal patients with advanced breast cancer. A randomized trial of tamoxifen vs. tamoxifen plus aminoglutethimide and hydrocortisone." Breast Cancer Res Treat **7 Suppl**: S45-50.

Ross-Innes, C. S., R. Stark, A. E. Teschendorff, K. A. Holmes, H. R. Ali, M. J. Dunning, G. D. Brown, O. Gojis, I. O. Ellis, A. R. Green, S. Ali, S. F. Chin, C. Palmieri, C. Caldas and J. S. Carroll (2012). "Differential oestrogen receptor binding is associated with clinical outcome in breast cancer." Nature **481**(7381): 389-393.

Ruthenburg, A. J., C. D. Allis and J. Wysocka "Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark." Molecular Cell **25**(1): 15-30.

Saloura, V., T. Vougiouklakis, M. Zewde, K. Kiyotani, J. Park, G. Gao, T. Karrison, M. Lingen, Y. Nakamura and R. Hamamoto (2016). "WHSC1L1 drives cell cycle progression through transcriptional regulation of CDC6 and CDK2 in squamous cell carcinoma of the head and neck." Oncotarget.

Schalch, T., S. Duda, D. F. Sargent and T. J. Richmond (2005). "X-ray structure of a tetranucleosome and its implications for the chromatin fibre." Nature **436**(7047): 138-141.

Schuettengruber, B., A.-M. Martinez, N. Iovino and G. Cavalli (2011). "Trithorax group proteins: switching genes on and keeping them active." Nat Rev Mol Cell Biol **12**(12): 799-814.

SEER (2016). "SEER Cancer Statistics Factsheets: All Cancer Sites."

Shankar, S. R., A. G. Bahirvani, V. K. Rao, N. Bharathy, J. R. Ow and R. Taneja (2013). "G9a, a multipotent regulator of gene expression." Epigenetics **8**(1): 16-22.

Shen, C., Jonathan J. Ipsaro, J. Shi, Joseph P. Milazzo, E. Wang, J.-S. Roe, Y. Suzuki, Darryl J. Pappin, L. Joshua-Tor and Christopher R. Vakoc (2015). "NSD3-Short Is an Adaptor Protein that Couples BRD4 to the CHD8 Chromatin Remodeler." Molecular Cell **60**(6): 847-859.

Shen, C., J. J. Ipsaro, J. Shi, J. P. Milazzo, E. Wang, J. S. Roe, Y. Suzuki, D. J. Pappin, L. Joshua-Tor and C. R. Vakoc (2015). "NSD3-Short Is an Adaptor Protein that Couples BRD4 to the CHD8 Chromatin Remodeler." Mol Cell **60**(6): 847-859.

Shi, L., L. Sun, Q. Li, J. Liang, W. Yu, X. Yi, X. Yang, Y. Li, X. Han, Y. Zhang, C. Xuan, Z. Yao and Y. Shang (2011). "Histone demethylase JMJD2B coordinates H3K4/H3K9 methylation and promotes hormonally responsive breast

carcinogenesis." Proceedings of the National Academy of Sciences **108**(18): 7541-7546.

Shyamala, G., Y. C. Chou, S. G. Louie, R. C. Guzman, G. H. Smith and S. Nandi (2002). "Cellular expression of estrogen and progesterone receptors in mammary glands: regulation by hormones, development and aging." J Steroid Biochem Mol Biol **80**(2): 137-148.

Slamon, D. J., G. M. Clark, S. G. Wong, W. J. Levin, A. Ullrich and W. L. McGuire (1987). "Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene." Science **235**(4785): 177-182.

Sobel, R. E., R. G. Cook and C. D. Allis (1994). "Non-random acetylation of histone H4 by a cytoplasmic histone acetyltransferase as determined by novel methodology." J Biol Chem **269**(28): 18576-18582.

Song, Q., S. D. Merajver and J. Z. Li (2015). "Cancer classification in the genomic era: five contemporary problems." Hum Genomics **9**(1): 27.

Sorlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. Eystein Lonning and A. L. Borresen-Dale (2001). "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." Proc Natl Acad Sci U S A **98**(19): 10869-10874.

Sotiriou, C. and L. Pusztai (2009). "Gene-expression signatures in breast cancer." N Engl J Med **360**(8): 790-800.

Spector, D. H., H. E. Varmus and J. M. Bishop (1978). "Nucleotide sequences related to the transforming gene of avian sarcoma virus are present in DNA of uninfected vertebrates." Proc Natl Acad Sci U S A **75**(9): 4102-4106.

Stehelin, D., H. E. Varmus, J. M. Bishop and P. K. Vogt (1976). "DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA." Nature **260**(5547): 170-173.

Tachibana, M., K. Sugimoto, T. Fukushima and Y. Shinkai (2001). "Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3." J Biol Chem **276**(27): 25309-25317.

Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-676.

Taunton, J., C. A. Hassig and S. L. Schreiber (1996). "A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p." Science **272**(5260): 408-411.

Van Holde, K. E., C. G. Sahasrabuddhe and B. R. Shaw (1974). "A model for particulate structure in chromatin." Nucleic Acids Res **1**(11): 1579-1586.

Varmus, H. E., R. A. Weiss, R. R. Friis, W. Levinson and J. M. Bishop (1972). "Detection of avian tumor virus-specific nucleotide sequences in avian cell DNAs (reassociation kinetics-RNA tumor viruses-gas antigen-Rous sarcoma virus, chick cells)." Proc Natl Acad Sci U S A **69**(1): 20-24.

Vettese-Dadey, M., P. A. Grant, T. R. Hebbes, C. Crane- Robinson, C. D. Allis and J. L. Workman (1996). "Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro." EMBO J **15**(10): 2508-2518.

Vidali, G., E. L. Gershey and V. G. Allfrey (1968). "Chemical studies of histone acetylation. The distribution of epsilon-N-acetyllysine in calf thymus histones." J Biol Chem **243**(24): 6361-6366.

Vlaming, H. and F. van Leeuwen (2016). "The upstreams and downstreams of H3K79 methylation by DOT1L." Chromosoma **125**(4): 593-605.

Vougiouklakis, T. T. "The NSD family of protein methyltransferases in human cancer." Epigenomics **7**(5): 863-874.

Waddington, C. H. (2012). "The epigenotype. 1942." Int J Epidemiol **41**(1): 10-13.

Wagner, E. J. and P. B. Carpenter (2012). "Understanding the language of Lys36 methylation at histone H3." Nat Rev Mol Cell Biol **13**(2): 115-126.

Wang, P., F. Fromowitz, M. Koslow, N. Hagag, B. Johnson and M. Viola (1991). "c-src structure in human cancers with elevated pp60c-src activity." Br J Cancer **64**(3): 531-533.

Weake, V. M. and J. L. Workman (2008). "Histone ubiquitination: triggering gene activity." Molecular cell **29**(6): 653-663.

Wei, Y., L. Yu, J. Bowen, M. A. Gorovsky and C. D. Allis (1999). "Phosphorylation of histone H3 is required for proper chromosome condensation and segregation." Cell **97**(1): 99-109.

Wernig, M., A. Meissner, R. Foreman, T. Brambrink, M. Ku, K. Hochedlinger, B. E. Bernstein and R. Jaenisch (2007). "In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state." Nature **448**(7151): 318-324.

Yang, Z. Q., G. Liu, A. Bollig-Fischer, C. N. Giroux and S. P. Ethier (2010). "Transforming properties of 8p11-12 amplified genes in human breast cancer." Cancer Res **70**(21): 8487-8497.

Yang, Z. Q., K. L. Streicher, M. E. Ray, J. Abrams and S. P. Ethier (2006). "Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer." Cancer Res **66**(24): 11632-11643.

Ye, T., A. R. Krebs, M. A. Choukrallah, C. Keime, F. Plewniak, I. Davidson and L. Tora (2011). "seqMINER: an integrated ChIP-seq data interpretation platform." Nucleic Acids Res **39**(6): e35.

Zhang, Q., L. Zeng, C. Shen, Y. Ju, T. Konuma, C. Zhao, Christopher R. Vakoc and M.-M. Zhou (2016). "Structural Mechanism of Transcriptional Regulator NSD3 Recognition by the ET Domain of BRD4." Structure **24**(7): 1201-1208.

Zhang, Y. (2003). "Transcriptional regulation by histone ubiquitination and deubiquitination." Genes & development **17**(22): 2733-2740.

Zheng, X. M., Y. Wang and C. J. Pallen (1992). "Cell transformation and activation of pp60c-src by overexpression of a protein tyrosine phosphatase." Nature **359**(6393): 336-339.

Zhu, L., Q. Li, S. H. K. Wong, M. Huang, B. J. Klein, J. Shen, L. Ikenouye, M. Onishi, D. Schneidawind, C. Buechele, L. Hansen, J. Duque-Afonso, F. Zhu, G.

Mas Martin, O. Gozani, R. Majeti, T. G. Kutateladze and M. L. Cleary (2016). "ASH1L links histone H3 lysine 36 di-methylation to MLL leukemia." Cancer Discovery.

ABSTRACT

WHSC1L1 AMPLIFICATION AND OVEREXPRESSION REGULATES ESTROGEN RECEPTOR ACTIVITY IN SUM44 BREAST CANCER CELLS

by

JONATHAN CURTIS IRISH

December 2016

Advisor: Stephen P. Ethier, Ph.D.

Co-Advisor: Zeng Quan Yang, Ph.D

Major: Cancer Biology

Degree: Doctor of Philosophy

Breast cancer is the most common cancer and the second leading cause of cancer death in women. While ER-positive breast cancer subtypes are initially well-managed by targeted therapies targeting estrogen signaling, many women are suffering from recurrence of a more aggressive, hormone insensitive cancer 5 or more years after initial remission. Late recurrence of hormone resistant breast cancer in patients who were previously successfully treated with anti-estrogen therapies worsens overall long-term outcomes, and specific oncogenic

mutations may be driving late recurring aggressive disease in these patients. More complete characterization of the oncogenome of a tumor may allow for the possibility of customized therapy targeting each patient's specific oncogene-activating mutations, and in doing so increase the probability of durable remission over the long term. Basic research on cell lines modeling different breast cancer subtypes enables discovery and characterization of novel driving oncogenes and their context in a breast cancer subtype model.

The Wolf Hirschhorn Syndrome Candidate 1-Like 1 gene (WHSC1L1) is one of approximately 50 genes in the chromosome 8p11-p12 amplicon, an amplified region of the short arm of chromosome 8 found in 12-15\% of human breast cancers, as well as other cancer types such as lung. Amplification of the 8p11-p12 region is most often found in breast cancers of the luminal B subtype. WHSC1L1 is a member of the NSD family of histone lysine methyltransferases, SET domain-containing proteins which catalyze the addition of a methyl group to lysines on the amino-terminal tail of histone H3 subunits. The WHSC1L1 gene expresses two known isoforms which code for two distinct proteins, WHSC1L1-long and WHSC1L1-short. The short isoform of WHSC1L1 codes for the first 647 of the 1437 amino acids present in the long isoform, and lacks the catalytic SET domain and several PHD and PWWP chromatin interacting domains, containing a single PWWP domain and a recently characterized acidic transactivation domain.

In both normal and tumor breast tissue, WHSC1L1-short is expressed at greater levels than WHSC1L1-long. Several breast cancer cell lines established in the Ethier lab harbor WHSC1L1 amplifications and also overexpress both isoforms of WHSC1L1. The SUM44 cell line is a highly ER-positive cell line model of luminal B breast cancer isolated from a pleural effusion metastasis of a patient with aggressive disease. It is known that the short isoform of WHSC1L1, WHSC1L1-short, is a potent driving oncogene in SUM44 cells, however the specific mechanism of WHSC1L1-short as an oncomodifier is not known.

To investigate WHSC1L1-short function as an oncogene in SUM44 cells, we developed an shRNA knockdown model that specifically knocked down expression of WHSC1L1-short through its unique 3' UTR sequence (shWHSC1L1-short), and a model that knocked down both WHSC1L1 isoforms (shWHSC1L1-total). We found that knockdown of both total WHSC1L1 and WHSC1L1-short alone negatively affected SUM44 proliferation, and that WHSC1L1-short knockdown had a larger effect than knockdown of both isoforms.

After finding that WHSC1L1 expression was required for typical proliferation rates of SUM44 cells, we performed genome-wide expression profiling of SUM44 WHSC1L1-short and total WHSC1L1 knockdown cell lines relative to a control SUM44 line transduced with shRNA against lacZ. Again we found that knockdown of the WHSC1L1-short alone had a greater affect than total WHSC1L1 knockdown, this time on the number of significantly differentially

expressed genes; 1131 genes were found to be differentially expressed in the WHSC1L1-short knockdown cells relative to shLacZ control, while 238 genes were differentially expressed in the total WHSC1L1 knockdown SUM44 cells relative to shLacZ control. Interestingly, the ESR1 gene, which codes for the estrogen receptor alpha protein, was significantly downregulated by WHSC1L1-short knockdown. This was confirmed by immunoblotting with ERa antibody. While total WHSC1L1 knockdown also had a negative effect on ERa protein levels in SUM44, knockdown of WHSC1L1-short alone almost completely abrogated ERa in SUM44 as measured by western blot. We subsequently found that SUM44 cells were extremely sensitive to treatment with beta-estradiol, and that proliferation actually decreased upon as little as 100 picomolar beta-estradiol treatment, with dose-dependent decreases in proliferation as estrogen concentrations increased. We also found SUM44 cells to be relatively insensitive to Tamoxifen.

Knockdown of WHSC1L1-short reduced proliferation of SUM44 cells in estrogen-free conditions, and treatment of SUM44 shWHSC1L1-short cells with increasing concentrations of beta-estradiol resulted in a marginal increase in proliferation up to 100pM beta-estradiol, then proliferation decreased with increasing beta-estradiol concentrations similar to results seen in SUM44 shLacZ control cells.

After observing that WHSC1L1-short overexpression was required for expression of ERa in SUM44 cells, we asked whether knockdown of WHSC1L1-

short affected genome-wide binding patterns of the estrogen receptor in SUM44 cells. Interestingly we found that ERa was binding to thousands of genomic loci in the absence of exogenous estrogen. Treatment with high doses (10nM) of beta-estradiol for 45 minutes resulted in an approximately even increase in ER-alpha binding across sites already bound in the absence of estrogen, with some additional weak binding sites, but no significant changes in the pattern of ERa binding. No ERa binding sites were detected in SUM44 shWHSC1L1-short cells under estrogen-free conditions, and weak ERa binding was detected in SUM44 shWHSC1L1-short cells treated with 10nM beta-estradiol at loci where strong ERa binding was observed in control SUM44 cells, suggesting that WHSC1L1-short knockdown was reducing ERa expression levels, which made less ERa available to bind to chromatin in SUM44 shWHSC1L1-short cells.

Our investigation of WHSC1L1 oncogenic activity in SUM44 cells resulted in the interesting observation that the short isoform of WHSC1L1 is required for expression of the estrogen receptor alpha in these cells, and that ERa is bound extensively to chromatin without activation of ERa by estrogen. SUM44 is a model for luminal B breast cancer, and is highly ER-positive, and expresses little to no progesterone receptor (PR). While the implications of ERa expression dependence on WHSC1L1-short overexpression in SUM44 cells are not yet clear, the extensive binding of ERa to estrogen response elements (ERE's) in the absence of exogenous estrogen and the negative proliferative response of SUM44 to estrogen indicate that WHSC1L1 amplification and overexpression

may alter the biology of the estrogen receptor in breast cancers harboring WHSC1L1 amplification and overexpression. Additionally, the differences seen in ERa binding to chromatin and the negative response of SUM44 cells to ERa agonists illustrate the importance of researching ER-positive breast cancer using additional cell line models rather than consistently using MCF7 cells to represent ER-positive disease.

The dominant role of the catalytically-inactive short isoform of WHSC1L1 in regulating ERa expression and maintaining the proliferation rate of SUM44 cells suggests that catalytically inactive isoforms of chromatin modifying enzymes can be important regulators of gene expression. Interestingly, recent work has shown that WHSC1L1-short is likely not regulating target gene expression through histone methylation, but instead is acting as a co-factor for a different chromatin-binding complex, the BRD4-CHD8 complex, which has been shown to be recruited to superenhancer regions (marked by histone acetylation) by WHSC1L1-short, which results in activation of pTEFb through BRD4 and directly activates target gene transcription. It will be important to avoid assumptions about binding substrate identities of catalytically inactive isoforms of future chromatin modifiers of interest, as the catalytically inactive isoforms of these genes may also bind to chromatin substrates unrelated to the substrate of the catalytically active isoform.

AUTOBIOGRAPHICAL STATEMENT

EDUCATION

Wayne State University Detroit, MI 2009-2016

Doctor of Philosophy: Cancer Biology

Wayne State University Detroit, MI 2007-2009

Master of Science: Basic Medical Sciences

Michigan State University East Lansing, MI 2003-2006

Bachelor of Science: Biochemistry

and Molecular Biology (Double Major in Physiology)