2017

The 8p11-P12 Amplicon Oncogenes ASH2L and NSD3 Regulate Cell Cycle Progression via Epigenetic Alterations and Result in Overexpression and Estrogen-Independent Activation of ERα in Breast Cancer

Jamie Nicole Mills
Medical University of South Carolina

Follow this and additional works at: https://medica-musc.researchcommons.org/theses

JAMIE NICOLE MILLS. The 8p11-p12 Amplicon Oncogenes ASH2L and NSD3 Regulate Cell Cycle Progression via Epigenetic Alterations and Result in Overexpression and Estrogen-Independent Activation of ERα in breast cancer. (Under the direction of STEPHEN P. ETHIER)

The 8p11-p12 Amplicon Oncogenes ASH2L and NSD3 Regulate Cell Cycle Progression via Epigenetic Alterations and Result in Overexpression and Estrogen-Independent Activation of ERα in breast cancer

by

Jamie Nicole Mills
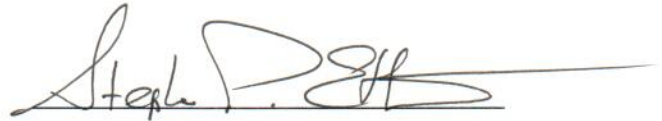
A dissertation submitted to the faculty of the Medical University of South Carolina in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Graduate Studies.

Department of Pathology and Laboratory Medicine

2017

Approved by:

Chairman, Advisory Committee

_____
Stephen P. Ethier

_____
Melissa Cunningham

_____
John Wrangle

_____
Victoria Findlay

_____
Amanda LaRue

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

Appendix A. NSD3-S and ESR1 knockdown microarray results.

Appendix B. H3K4me3 ChIP-seq quality control and enrichment reports.
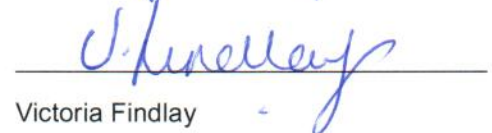
Appendix C. H3K4me3 ChIP-seq peak datasets: LacZ control and ASH2L knockdown.

Appendix D. Gene lists: overall H3K4me3 ChIP-seq peaks.

Appendix E. Gene lists and ToppFun analysis: promoter only H3K4me3 ChIP-seq peaks.

Appendix F. ASH2L knockdown RNA-seq results.

Appendix G. Gene lists and ToppFun analysis: ASH2L knockdown RNA-seq downregulated genes.

Appendix H. Gene lists and ToppFun analysis: ASH2L knockdown H3K4me3 ChIP-seq and RNA-seq.

Appendix I. Gene lists and ToppFun analysis: ASH2L, NSD3-S, and ESR1 knockdown expression analysis and ASH2L knockdown H3K4me3 ChIP-seq overlap.

LIST OF ABBREVIATIONS

4OH-tamoxifen – hydroxytamoxifen

AF – activating function (ERα domain)

AI – aromatase inhibitor

AML – acute myeloid leukemia

ASH2L – Absent, Small, or Homeotic disc 2-Like

BET – bromodomain and extraterminal domain

CCND1 – cyclin D1 gene

CDK – cyclin-dependent kinase

CHD – chromodomain

ChIP – chromatin immunoprecipitation

CNA – copy number alteration

CTD – comparative toxicogenomics database

DBD – DNA binding domain

DNMT – DNA methyltransferase

EGFR – epidermal growth factor receptor

ER/ERα – estrogen receptor (alpha)

ERE – estrogen response element

ESR1 – estrogen receptor alpha gene

FDR – false discovery rate

GISTIC – genomic identification of significant targets in cancer

GO – gene ontology

GRH – gonadotrophin releasing hormone

H – histone

HAT – histone acetyltransferase

HBSS – Hank's buffered saline solution

HDAC(i) – histone deacetylase (inhibitor)

HER2 – human epidermal growth factor receptor 2

HOX – homeobox gene family

HMT – histone methyltransferase

HMTi – histone methyltransferase inhibitors/inhibition

hPTM – histone post-translational modification

IGV – integrated genome viewer

IHC - immunohistochemistry

IP - immunoprecipitation

IPTG – isopropyl β-D-1-thiogalactopytanoside

K – lysine

LBD – ligand binding domain

LSCC – lung squamous cell carcinoma

MLL – mixed lineage leukemia

NLS – nuclear localization signal

NS – non-significant

NSD – Nuclear-receptor binding SET Domain-containing

NSD3-L – long isoform of NSD3

NSD3-S – short isoform of NSD3

NSD3-T – total NSD3 (both isoforms)

PBS – phosphate buffered saline

PCA – principal component analysis

PcG – polycomb group

PHD – plant homeodomain

PR – progesterone receptor

PRC – polycomb repressive complex

PWWP – proline-tryptophan-tryptophan-proline

R – arginine

RT – room temperature

RT-PCR – reverse transcription polymerase chain reaction

SAC – SET-associated cysteine-rich

SERD – selective estrogen receptor degrader

SERM – selective estrogen receptor modulator

SET – Suppressor of variegation 3-9, Enhancer of zeste and Trithorax

-seq – high throughput sequencing

shRNA – short hairpin RNA

TAD – transcription activating domain

TBST – tris-buffered saline + Tween 20

TCGA – the cancer genome atlas

TNBC – triple negative breast cancer

trxG – trithorax group

WAR – WDR5-ASH2L-RbBP5 sub-complex

WRAD – WDR5-RbBP5-ASH2L-DPY-30 sub-complex

WHSC1L1 – Wolf-Hirschhorn Syndrome Candidate 1-Like 1

ABSTRACT

Breast cancer is a highly heterogeneous disease classified clinically by expression of estrogen receptor alpha ERα, progesterone receptor, and human epidermal growth factor receptor. Molecular expression profiling identified a luminal breast cancer sub-type that can be sub-divided into luminal A and B. Compared to luminal A, luminal B tumors have increased proliferation, poor prognosis, endocrine therapy resistance, and complex genomes, including amplification of the 8p11-p12 genomic region. This amplicon occurs in 15% of primary breast tumors, correlates with poor prognosis and tamoxifen resistance, and harbors several oncogenes. Two of these oncogenes, ASH2L and NSD3 (WHSC1L1), promote transcription via epigenetic modification of histone proteins. NSD3 has a long isoform that is associated with di-methylation of lysine 36 on histone 3 (H3K36me2) and a short isoform that lacks a catalytic SET domain but retains the ability to interact with chromatin. ASH2L also lacks a catalytic SET domain yet is tightly and specifically linked to tri-methylation of lysine 4 on histone 3 (H3K4me3) in gene promoters. In this study, we tested the hypothesis that ASH2L and NSD3 cooperate to regulate expression of a suite of genes important in breast cancer, including ESR1, which encodes ERα. We discovered that NSD3-short is the major oncogenic isoform of NSD3 and its amplification and overexpression leads to overexpression and estrogen-independent activation of ERα. We also demonstrated that knockdown of ASH2L reduces H3K4me3 specifically in promoters of genes important to cell cycle progression. ASH2L also regulates promoter H3K4me3 at NSD3 and expression of both NSD3 and ERα. Knockdown of ASH2L reduced sensitivity to the cell cycle inhibitor palbociclib in the 8p11-p12 amplicon-bearing SUM-44 cell line. Together, the data presented here identify a role for ASH2L and NSD3 in cooperative regulation of genes important to cell

cycle regulation, including ESR1, and demonstrate that ERα is active in an estrogen-independent manner in the context of overexpression of these oncogenes. We have discovered a novel mechanism of endocrine resistance in luminal B breast cancers and provided evidence for the 8p11-p12 amplicon as a biomarker of patients who will respond to cell cycle inhibitors and epigenetic therapies against histone methyltransferase enzymes.

CHAPTER 1: Introduction and Review of Literature

1.1. Breast Cancer Classification and Treatment

a. Breast cancer history, classification, and estrogen receptor alpha (ERα)

Breast cancer is the most common cancer in women worldwide and accounts for more than 40,000 deaths in the United States annually (Morris and Carey 2007, Chang 2012). As early as 1896, oophorectomy was described by Beatson as providing benefit to patients with advanced breast cancer (Cadoo *et al.* 2013, Mancuso and Massarweh 2016, Tabarestani *et al.* 2016), although the mechanism, now known to be estrogen deprivation, would not be understood until the discovery of estrogen receptor alpha (ERα) by Elwood Jensen *et al.* in 1958 (Hartman *et al.* 2009, Pritchard 2013). With this discovery, the importance of estrogen signaling in mammary gland ductal development and breast cancer genesis and progression was identified and became the focus of targeted therapies in breast cancer treatment.

Currently, breast cancer is diagnosed and treated based on clinicopathologic analysis (Geyer *et al.* 2012, Cadoo *et al.* 2013, Netanely *et al.* 2016). Tumors are assessed by immunohistochemistry (IHC) for estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) (Cui *et al.* 2005, Reis-Filho and Pusztai 2011, Eroles *et al.* 2012). Tumors negative for all three markers are known as triple negative breast cancers (TNBC), are treated primarily with chemotherapy, and have the worst prognosis (Sorlie *et al.* 2001, Eroles *et al.* 2012, Cobain and Hayes 2015). HER2-amplified tumors are treated with chemotherapy and targeted HER2 antibody therapies (Yang *et al.* 2009, Cobain and Hayes 2015). Tumors expressing ER usually also express PR as its expression is under ER regulation, with

loss of PR acting as a marker of poorer prognosis (Margueron *et al.* 2004, Cui *et al.* 2005, Cornen *et al.* 2014). Tumors expressing both ER and PR are termed "hormone receptor-positive" and make up the majority of breast cancer cases (Scott *et al.* 2011, Geyer *et al.* 2012). Approximately 70-80% of breast cancers express ERα (ER+ breast cancer) and have the best prognosis due to the less aggressive nature of the tumors and availability of therapies targeting ERα function (Schiff *et al.* 2003, Cui *et al.* 2005, Howell 2006, Robertson 2007, Thomas and Munster 2009, Pathiraja *et al.* 2010, 2012, Chang 2012, Geyer *et al.* 2012, Cadoo *et al.* 2013, Ignatiadis and Sotiriou 2013, Kerdivel *et al.* 2013, Dabydeen and Furth 2014, Lumachi *et al.* 2015, Nagaraj and Ma 2015, Wang and Yin 2015, De Marchi *et al.* 2016, Selli *et al.* 2016).

ERα is encoded by the ESR1 gene and is a transcription factor that responds to estrogens, most predominantly 17β-estradiol (Hartman *et al.* 2009, Pathiraja *et al.* 2010, Chang 2012). ERα is comprised of two transcriptional activating domains, AF-1 and AF-2, which flank a central hinge region that includes the ligand-binding and DNA-binding domains (Figure 1.1) (Cui *et al.* 2005, Nicholson and Johnston 2005, Hartman *et al.* 2009, Kerdivel *et al.* 2013, Wang and Yin 2015, Angus *et al.* 2016). AF-1 is ligand-independent and is activated by phosphorylation and other post-translational modification events while AF-2 is ligand-dependent (Cui *et al.* 2005, Dowsett *et al.* 2005, Nicholson and Johnston 2005, Hartman *et al.* 2009, Thomas and Munster 2009, Chang 2012, Kerdivel *et al.* 2013, Zhang *et al.* 2013, Nagaraj and Ma 2015, Wang and Yin 2015, De Marchi *et al.* 2016, Steelman *et al.* 2016, Tabarestani *et al.* 2016). Binding of estradiol at the ligand-binding domain stimulates dimerization and binding to chromatin, recruitment of co-regulatory molecules, and gene transcription (Cui *et al.* 2005, Nicholson and Johnston 2005, Howell 2006, Thomas and Munster 2009, Fedele *et al.*

2012, Kerdivel *et al.* 2013, Zhang *et al.* 2013, Nagaraj and Ma 2015, Wang and Yin

2015, De Marchi *et al.* 2016, Steelman *et al.* 2016, Tabarestani *et al.* 2016). The

resulting effect on gene expression is highly cell- and tissue-type specific and is

regulated by ligand type, post-translational modifications to ERα, other transcription

factors, the balance of ERα co-activator versus co-repressor components available, the

epigenetic state of the chromatin, growth factor signaling pathways, and many other

factors (Schiff *et al.* 2003, Margueron *et al.* 2004, Cui *et al.* 2005, Kristensen *et al.* 2005,

Nicholson and Johnston 2005, Thomas and Munster 2009, Pathiraja *et al.* 2010, Chang

2012, Fedele *et al.* 2012, Hervouet *et al.* 2013, Kerdivel *et al.* 2013, Zhang *et al.* 2013,

Nagaraj and Ma 2015, Wang and Yin 2015, De Marchi *et al.* 2016). Transcript variants of

ERα with non-genomic functions, such as cross-talk with growth factor receptor signaling

pathways, have also been well-described in the literature but are beyond the scope of

this review (Howell 2006, Chang 2012, Fedele *et al.* 2012, Kerdivel *et al.* 2013, Zhang *et al.* 2013, Wang and Yin 2015, Steelman *et al.* 2016). Generally, estrogens exert

Figure 1.1. Schematic representation of ERα structure. Diagram of the domains of ERα, including AF-1 and AF-2, nuclear localization region, DNA-binding domain and ligand-binding domain (together the hinge region) and known ERα modification sites as described in the key (lower left). Diagram based on representations from Kerdivel *et al* 2013.
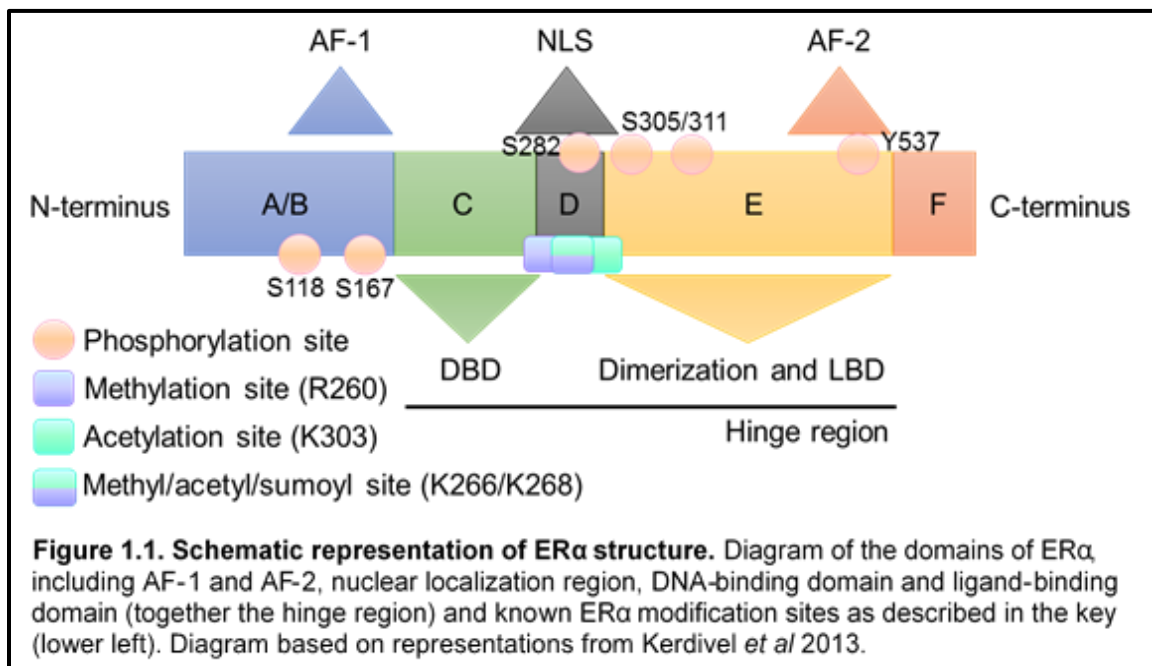
mitogenic effects via ERα-mediated transcriptional activation of genes involved in cell cycle and other cellular growth processes (Sorlie *et al.* 2001, Allred *et al.* 2004, Kerdivel *et al.* 2013, Wang and Yin 2015).

b. Treatment of ER+ breast cancer

The selective estrogen receptor modulator (SERM) tamoxifen was introduced in the 1970s and has since been the most widely used endocrine therapy in advanced breast cancer (Morris and Wakeling 2002, Margueron *et al.* 2004, McKeage *et al.* 2004, Hartman *et al.* 2009, Huang *et al.* 2011, Chang 2012, Lumachi *et al.* 2015, Nagaraj and Ma 2015, De Marchi *et al.* 2016, Selli *et al.* 2016). The introduction of tamoxifen saw a reduction in breast cancer recurrence by 39% and mortality by 30% (Nagaraj and Ma 2015). Tamoxifen binds to ERα at the ligand-binding domain and prevents signaling through the ligand-dependent AF-2 domain but is unable to block the AF-1 domain, which is activated through phosphorylation in a ligand-independent manner (McKeage *et al.* 2004, Dowsett *et al.* 2005, Howell 2006, Hartman *et al.* 2009, Chang 2012). As a result, tamoxifen exerts both antagonist and agonist effects on ERα (Morris and Wakeling 2002, Schiff *et al.* 2003, Allred *et al.* 2004, Margueron *et al.* 2004, Dowsett *et al.* 2005, Nicholson and Johnston 2005, Hartman *et al.* 2009, Chang 2012, Dabydeen and Furth 2014, Lumachi *et al.* 2015, De Marchi *et al.* 2016, Selli *et al.* 2016).

The benefit of tamoxifen use in breast cancer is centered upon its role as an ERα antagonist in breast tissue. The active metabolite of tamoxifen, 4-hydroxytamoxifen (4OH-tamoxifen), binds ERα with 25% of the affinity of 17β-estradiol and induces a conformational change to prevent recruitment of co-activators and instead favoring co-repressor recruitment, thereby blocking ERα-mediated transcription of proliferative

genes (Morris and Wakeling 2002, Margueron *et al.* 2004, Cui *et al.* 2005, Dowsett *et al.* 2005, Howell 2006, Chang 2012, De Marchi *et al.* 2016). Tamoxifen functions as an ERα agonist in bone to maintain bone density and blood to reduce serum cholesterol, positive effects for patients, but also in liver and uterus, increasing the risk of developing endometrial cancer with prolonged use (Morris and Wakeling 2002, Nicholson and Johnston 2005, Howell 2006, Hartman *et al.* 2009, De Marchi *et al.* 2016). Many adverse effects of tamoxifen are due to its agonist activity, including gastrointestinal disturbances (e.g. nausea, vomiting), hot flushes, joint pain, and headaches (McKeage *et al.* 2004, Howell 2006). The recommended duration of tamoxifen use is at least 5 years after diagnosis and initial treatment and these adverse effects are the primary cause of non-compliance during adjuvant therapy (Chanrion *et al.* 2008, Scott *et al.* 2011, Chang 2012, Ignatiadis and Sotiriou 2013, Lumachi *et al.* 2015, Nagaraj and Ma 2015).

Due in part to the agonist properties of tamoxifen, recurrence while on therapy occurs (Nicholson and Johnston 2005, Robertson 2007, Huang *et al.* 2011, Bilal *et al.* 2012, Chang 2012, Cadoo *et al.* 2013, Ignatiadis and Sotiriou 2013, Kerdivel *et al.* 2013, Nagaraj and Ma 2015, Kumler *et al.* 2016, Lim *et al.* 2016, Mancuso and Massarweh 2016, Selli *et al.* 2016, Tabarestani *et al.* 2016). As such, additional non-steroidal agents similar to tamoxifen have been developed, such as toremifene, idoxifene, and droloxifene, but none have treatment advantages over tamoxifen (Morris and Wakeling 2002). To circumvent tamoxifen resistance, another class of anti-estrogen therapies was developed to abrogate estrogen-mediated activation of transcription via ERα through a reduction in circulating estrogen levels (Morris and Wakeling 2002, Nicholson and Johnston 2005, Kerdivel *et al.* 2013, Robertson *et al.* 2014, Lumachi *et al.* 2015). Estrogens are derived from androgens by the enzyme aromatase in peripheral tissues,

especially adipose tissue in postmenopausal women, as well as within the tumor in the context of breast cancer (Morris and Wakeling 2002, Nicholson and Johnston 2005, Pritchard 2013, Dabydeen and Furth 2014, Lumachi *et al.* 2015, Nagaraj and Ma 2015, Selli *et al.* 2016). First-generation aromatase inhibitors (AIs) did not demonstrate therapeutic benefit over tamoxifen, but the third-generation non-steroidal AIs anastrozole and letrozole and steroidal exemestane surpassed tamoxifen as first-line therapy for ER+ breast cancer in postmenopausal women (Morris and Wakeling 2002, Cui *et al.* 2005, Nicholson and Johnston 2005, Ariazi *et al.* 2006, Howell 2006, Robertson 2007, Hartman *et al.* 2009, Robertson *et al.* 2014). AIs increase gonadotropin releasing hormone (GRH) in premenopausal patients, which results in increased estrogen levels, therefore these compounds are under investigation in conjunction with GRH-inhibitors to extend their use to this patient population (Tabarestani *et al.* 2016). The introduction of AIs as treatment alternatives, both as first-line therapies and for patients who failed to respond to tamoxifen, was a major step forward in breast cancer management.

Aromatase inhibitors have shown improvement in patient survival and lack the increased risk of endometrial cancer of tamoxifen (Chang 2012). AIs have similar adverse effects otherwise, including disease recurrence despite initial response (Fedele *et al.* 2012, Ignatiadis and Sotiriou 2013, Nagaraj and Ma 2015, Mancuso and Massarweh 2016). More recently, pure ERα antagonists were developed, known as selective estrogen receptor degraders (SERDs), such as fulvestrant (ICI-182,780) (Dowsett *et al.* 2005, Howell 2006, Johnston and Cheung 2010). Fulvestrant binds to ERα at the ligand-binding domain with 89% the affinity of 17β-estradiol, inducing a different conformational change than tamoxifen that results in inhibition of dimerization, blocking both the AF-1 and AF-2 domains from participating in transcription (Morris and

Wakeling 2002, McKeage *et al.* 2004, Dowsett *et al.* 2005, Nicholson and Johnston 2005, Howell 2006, Scott *et al.* 2011, Lumachi *et al.* 2015). This results in complete abrogation of transcription of ERα-target genes and destabilization of ERα, prompting degradation by the ubiquitin-proteasome complex (Morris and Wakeling 2002, Margueron *et al.* 2004, McKeage *et al.* 2004, Dowsett *et al.* 2005, Johnston and Cheung 2010, Scott *et al.* 2011). Clinical trials revealed that fulvestrant is at least as effective as anastrozole and was thus approved in 2002 for the treatment of hormone receptor-positive metastatic breast cancer that had failed other endocrine therapies (Dowsett *et al.* 2005). In 2010, fulvestrant was approved as a second-line therapy for postmenopausal patients with hormone receptor-positive breast cancer (Scott *et al.* 2011, Ignatiadis and Sotiriou 2013).

Due to the different mechanism of action of SERDs, cross-resistance to tamoxifen and anastrozole is rare (McKeage *et al.* 2004, Dowsett *et al.* 2005, Johnston *et al.* 2005, Howell 2006, Johnston and Cheung 2010, De Marchi *et al.* 2016). Adverse effects of fulvestrant are similar to the other endocrine therapies and slightly less common, with gastrointestinal disturbances and joint pain being most cited (McKeage *et al.* 2004, Howell 2006, Scott *et al.* 2011, Lumachi *et al.* 2015). Clinical trials focused on optimal dosing of fulvestrant, however, revealed the dose-dependent nature of this drug for reducing ERα levels (Robertson 2007, Scott *et al.* 2011, Robertson *et al.* 2014, Nagaraj and Ma 2015). The ability to achieve steady-state levels adequate for complete inhibition of ERα in patients, especially those with highly ER-positive tumors, without reaching the limits of adverse effects is difficult (Robertson 2007). This therapy is also given via intramuscular injection as opposed to orally, affecting patient compliance (McKeage *et al.* 2004). A second generation of SERD therapies are under development

and testing to attempt to combat these issues (Angus *et al.* 2016, Mancuso and Massarweh 2016).

Unfortunately, resistance to endocrine therapy of all types is common (Allred *et al.* 2004, Kerdivel *et al.* 2013, Pritchard 2013, Nagaraj and Ma 2015). Absence of ERα is a very good negative predictor of response to endocrine therapy and its expression is an accurate positive predictor of response in approximately 50% of ER+ breast cancers (Lonning *et al.* 2005, Thomas and Munster 2009, Shiu *et al.* 2010, Cobain and Hayes 2015, Wang and Yin 2015, Azim *et al.* 2016). Since half of ER+ breast cancer patients do not demonstrate durable response to endocrine therapy, there is a pressing need to discover novel biomarkers that will complement ERα in predicting endocrine therapy response (Chanrion *et al.* 2008, Thomas and Munster 2009, Reis-Filho and Pusztai 2011, Geyer *et al.* 2012, Habashy *et al.* 2012, Cobain and Hayes 2015, Mancuso and Massarweh 2016, Selli *et al.* 2016, Rakha and Green 2017). Similarly, many groups are investigating parallel processes, such as activation of the PI3K/AKT/mTOR pathway and the cyclin D1/CDK4/6-mediated cell cycle pathway, to determine if coupling endocrine therapy with inhibitors of these pathways will improve patient outcomes (Caldon *et al.* 2012, Ignatiadis and Sotiriou 2013, Dabydeen and Furth 2014, Lumachi *et al.* 2015, Nagaraj and Ma 2015, Azim *et al.* 2016, Knudsen and Witkiewicz 2016, Kumler *et al.* 2016, Lim *et al.* 2016, Mancuso and Massarweh 2016, Steelman *et al.* 2016). Indeed, these inhibitors have shown promising results in subsets of breast cancer patients in combination with endocrine therapy (Cornen *et al.* 2014, Nagaraj and Ma 2015, Azim *et al.* 2016, Knudsen and Witkiewicz 2016, Mancuso and Massarweh 2016, Steelman *et al.* 2016). Some of the mechanisms of endocrine therapy resistance are known, such as loss of ERα expression, ESR1 amplifications, point mutations in the ligand binding

domain of ERα, and aberrant activation of growth factor signaling pathways, but these alterations do not account for all patients with poor response to endocrine therapies (Cui *et al.* 2005, Loi *et al.* 2009, Pathiraja *et al.* 2010, Chang 2012, Fedele *et al.* 2012, Kerdivel *et al.* 2013, De Marchi *et al.* 2016, Selli *et al.* 2016, Tabarestani *et al.* 2016). More research is necessary to determine additional mechanisms behind endocrine therapy resistance in order to identify biomarkers for which patients will require combinatorial treatment strategies and to identify potential targets for novel therapies.

Breast cancer has long been described as a highly heterogenetic disease (Carey *et al.* 2006, Morris and Carey 2007, Elsheikh *et al.* 2009, Kao *et al.* 2009, Ellis *et al.* 2012, Cadoo *et al.* 2013, Dabydeen and Furth 2014, Cobain and Hayes 2015, Selli *et al.* 2016). The current clinical and histopathologic parameters used to classify breast cancers are insufficient to capture this diversity and predict treatment response (Sorlie *et al.* 2001, Sorlie *et al.* 2003, Kao *et al.* 2009, Habashy *et al.* 2012, Tang and Tse 2016). With expansion of global-scale profiling techniques such as microarrays, next generation sequencing, and many others, there has been a paradigm shift toward breast cancer classification based on molecular expression profiling.

c. Molecular profiling of breast cancers

In 2000, Perou *et al.* identified two main types of breast cancer that correlate to ERα status of the tumors: those that express genes known to be associated with basal cells and those with a more luminal cell-like expression profile (Perou *et al.* 2000). Many groups since have advocated that the delineation of these two cancer types is so robust that they should be treated as different diseases and given their own cancer classes in large databases such as the cancer genome atlas (TCGA) (Sorlie *et al.* 2001, Reis-Filho

and Pusztai 2011, Geyer *et al.* 2012, Habashy *et al.* 2012). The basal tumors then

subdivide into basal-like, normal-like, and HER2+ subtypes while the luminal group

divides into luminal A and luminal B (Figure 1.2) (Sorlie *et al.* 2001). Even within the

basal and luminal groups, a high degree of heterogeneity remains (Sorlie *et al.* 2001,

Morris and Carey 2007, Shiu *et al.* 2010, Geyer *et al.* 2012, Ignatiadis and Sotiriou 2013,

Cornen *et al.* 2014). Indeed, the luminal group, which are primarily ER+ and express ER-

associated genes such as ESR1, XBP1, CCND1, and GATA3 (Perou *et al.* 2000, Sorlie

et al. 2001, Sotiriou *et al.* 2003, 2012, Habashy *et al.* 2012, Shan *et al.* 2014, Tang and

Tse 2016), has been a topic of controversy in the field. Some advocate for the existence

of three subtypes, luminal A, B, and C, while others have focused on identifying

clinically-useful markers to separate luminal A tumors from luminal B, and others still

have argued that this tumor type actually comprises a continuum rather than individual

and separable subtypes of breast cancer (Shiu *et al.* 2010, Geyer *et al.* 2012, Netanely

et al. 2016, Tang and Tse 2016). The St Gallen International Consensus on the Primary

Therapy of Early Breast Cancer 2013 separated this group of tumors into luminal A,

luminal B HER2-, and luminal B HER2+ (Judes *et al.* 2016).

Regardless of the specific breakdown, it is generally accepted that luminal A

tumors differ from luminal B tumors in many characteristics (Figure 1.2). Luminal A

tumors are the most common type of all breast cancers (approximately 55%) and have

the highest ERα and ERα-target gene expression, including PR, low expression of genes

associated with proliferation, no HER2 amplifications, low rates of GATA3 mutations,

simple karyotypes, and higher rates of PIK3CA mutations than luminal B tumors (Raica

et al. 2009, Shiu *et al.* 2010, Geyer *et al.* 2012, Ignatiadis and Sotiriou 2013, Tang and

Tse 2016). These tumors tend to be well-differentiated and include all cases of lobular
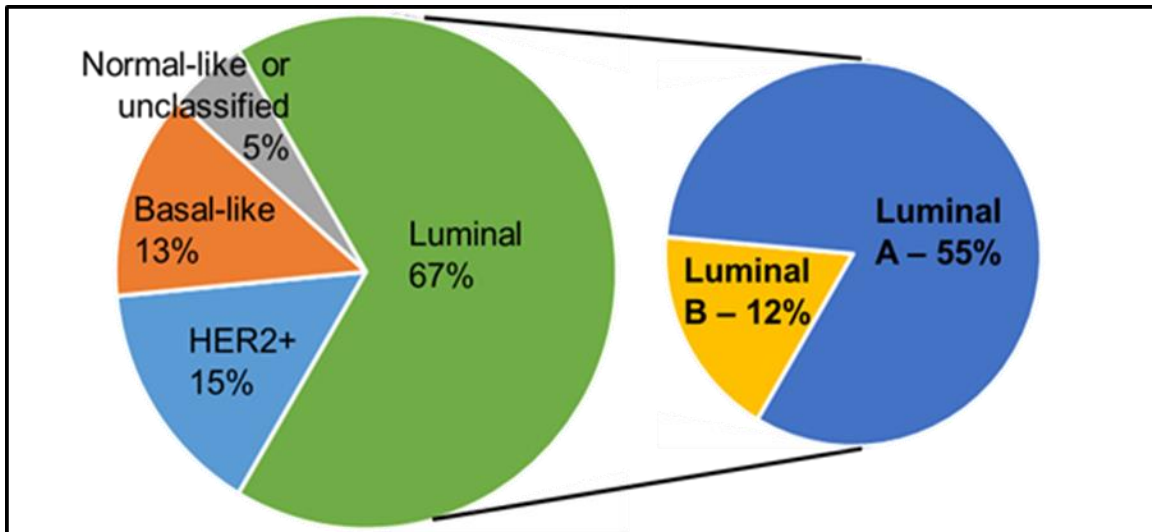
Characteristic	Luminal B	Luminal A
Incidence	12%	55%
ERα expression	lower	high
ERα negativity	up to 6%	up to 15%
ERα target genes	lower	high
PR	+/-	+
Proposed PR IHC % cutoff	<20	>20
HER2 amplification	30%	<10%
Proliferation-associated gene expression	higher	low
Proposed Ki-67 % cutoff	>14	<14
Histology	lobular/ductal and less differentiated	lobular and well-differentiated
Aggressiveness	higher	low
GATA3 mutation rate	higher	low
PIK3CA mutation rate	32%	45%
TP53 mutation rate	32%	13%
karyotypes	complex	simple
endocrine therapy response rate	lower	high
chemotherapy response rate	intermediate	poor
prognosis	intermediate	good

Figure 1.2. Breast cancer intrinsic subtype proportions and characteristics of luminal tumors. Pie chart depicting the approximate percentage of all breast cancers accounted for by each molecular subtype. Together, the luminal subtype makes up approximately 67% with luminal A and B accounting for approximately 55% and 12% of total breast cancers, respectively. Corresponding table describes selected characteristics of luminal A and luminal B tumors.

carcinoma *in situ* as well as most cases of invasive lobular carcinoma and multiple other histologic subtypes (Shiu *et al.* 2010, Eroles *et al.* 2012, Dabydeen and Furth 2014). Luminal A tumors tend to be highly responsive to endocrine therapies and have significantly better prognosis and relapse rates than luminal B tumors, which comprise approximately 12% of all breast cancers and have overall lower expression of ERα and associated genes (Raica *et al.* 2009, Turner *et al.* 2010, Tang and Tse 2016). Up to 6% of luminal B tumors are classified as ER-negative (Eroles *et al.* 2012). These tumors have higher expression of proliferation-associated genes, including MKI67, CCNB1, and MYBL2, and indeed these genes are often cited as the major distinction between luminal A and B tumors (Cheang *et al.* 2009, Sircoulomb *et al.* 2011, Eroles *et al.* 2012, Ades *et al.* 2014). The luminal B subtype also tends to be more aggressive, higher grade, less differentiated, primarily invasive ductal carcinoma of multiple histologic types, usually HER2-negative (although half of HER2+ tumors fall under this category), and less commonly PR+ than luminal A (Raica *et al.* 2009, Eroles *et al.* 2012, Ignatiadis and Sotiriou 2013, Tang and Tse 2016). Complex karyotypes with multiple copy number aberrations are common with a lower PIK3CA (32% vs 45%) but higher TP53 (32% vs 13%) mutation rate and more frequent gain of MDM2 and cyclin D1 (Geyer *et al.* 2012, Ades *et al.* 2014, Tang and Tse 2016). Luminal B tumors also tend to have poorer response to endocrine therapy than their luminal A counterpart (Turner *et al.* 2010, Tang and Tse 2016). Although luminal B tumors do respond better to chemotherapy than luminal A, response is still much poorer than the other breast cancer subtypes, leaving few good therapeutic options for these patients (Sircoulomb *et al.* 2011, Eroles *et al.* 2012, Cadoo *et al.* 2013, Ignatiadis and Sotiriou 2013, Kerdivel *et al.* 2013).

Defining molecular subtypes has major implications for identifying and developing novel therapeutic strategies for breast cancer management, therefore a means by which luminal A and luminal B tumors can be clinically differentiated, both from the other intrinsic subtypes as well as from each other, is essential for improving therapeutic interventions (Miller *et al.* 2009, Cornen *et al.* 2014, Tang and Tse 2016, Rakha and Green 2017). Several attempts have been made to identify a genetic signature based on the initial intrinsic gene sets that would function as a prognostic and predictive clinical assay (van 't Veer *et al.* 2002, Morris and Carey 2007, Chanrion *et al.* 2008, Cheang *et al.* 2009, Geyer *et al.* 2012, Ignatiadis and Sotiriou 2013, Luo *et al.* 2017). Though there is moderate clinical utility, these attempts have largely been unable to capture the biologic complexity and accurately predict tumor classification (Sorlie *et al.* 2006, Reis-Filho and Pusztai 2011, Geyer *et al.* 2012, Ignatiadis and Sotiriou 2013, Selli *et al.* 2016, Tang and Tse 2016). The exception has been Oncotype DX, an RT-PCR-based test that predicts benefit of chemotherapy and risk of distant recurrence and has demonstrated usefulness in predicting patients who will derive minimal benefit from chemotherapy (Morris and Carey 2007, Reis-Filho and Pusztai 2011, Bilal *et al.* 2012, Geyer *et al.* 2012, Cobain and Hayes 2015).

Clinicopathologic markers in addition to ER, PR, and HER2 have also been proposed to distinguish the intrinsic subtypes. The addition of IHC assessment of basal cytokeratin expression, such as CK5, has improved the identification of luminal- versus basal-type tumors (Raica *et al.* 2009, Eroles *et al.* 2012, Tang and Tse 2016). The two major proteins proposed to differentiate luminal A and B by IHC are PR and Ki-67, which is a marker of proliferation. Cutoffs of 20% for PR and 14% for Ki-67 have been endorsed by the 2011 St. Gallen International Breast Cancer Conference, with luminal A

tumors having higher PR expression and luminal B having higher Ki-67 (Eroles *et al.* 2012, Geyer *et al.* 2012, Ignatiadis and Sotiriou 2013, Ades *et al.* 2014, Tang and Tse 2016). Due to the lack of true bimodal distribution of these subtypes, these cutoffs are subjective, arbitrary, and vary widely based on tumor heterogeneity, even within one tumor biopsy (Geyer *et al.* 2012, Tang and Tse 2016). Additionally, lack of standardization of these tests poses another complicating factor in using IHC to define tumor subtypes (Eroles *et al.* 2012, Geyer *et al.* 2012). Indeed, even IHC assessment of ERα expression is highly variable and subjective. ER+ tumors are defined by at least 1% of the cells staining positive for nuclear ERα and are treated with endocrine therapy, despite studies demonstrating that tumors with 1-10% ERα staining behave much more like basal tumors (Allred *et al.* 2004, Dabydeen and Furth 2014, Lumachi *et al.* 2015, Tang and Tse 2016). Identification of reliable biomarkers that can encompass the vast heterogeneity of breast cancers is essential to moving this field forward and improving patient outcomes.

1.2. Breast Cancer Genomics

a. Genomic alterations in luminal B breast cancer

Although breast cancer classification by molecular profiling is complicated, evolving, and not yet ready for full clinical use, some of the information obtained from these studies can be immediately utilized to inform clinical trial design and spark novel scientific studies that will improve understanding of breast cancer pathogenesis and lead to identification of potential therapeutic targets. One of the major differences between luminal A and B breast cancer subtypes, other than the difference in expression of genes associated with proliferation, is the genomic instability associated with luminal B

tumors (Kristensen *et al.* 2005, Geyer *et al.* 2012, Cornen *et al.* 2014). As previously

mentioned, luminal A tumors tend to have relatively simple karyotypes (Kao *et al.* 2009,

Geyer *et al.* 2012). They are often diploid with infrequent translocations and copy

number alterations (CNAs), although low-level gains of 1q and 16p and deletions of 16q

are not uncommon, especially as the tumors become higher grade and ERα and

proliferation gene expression patterns move more toward the luminal B profile

(Bergamaschi *et al.* 2006, Shiu *et al.* 2010, Geyer *et al.* 2012, Reynisdottir *et al.* 2013,

Ades *et al.* 2014, Cornen *et al.* 2014). This tendency again highlights the continuum

pattern rather than bimodal distribution of these two luminal breast cancer subtypes.

Three major genomic phenotypes have been described in breast cancer (Shiu *et

al.* 2010). In contrast to the "simplex" phenotype of luminal A, luminal B tumors have

much greater genomic instability and frequency of high-level CNAs and aneuploidy (Shiu

et al. 2010, Cornen *et al.* 2014). These tumors fall into a "firestorm" or "amplifier" pattern

characterized by multiple recurrent amplifications (Shiu *et al.* 2010, Sircoulomb *et al.*

2011). The third breast cancer genomic phenotype is termed "sawtooth" or "complex"

pattern, which demonstrates alternating low-level gains and deletions that affect multiple

chromosomes in the same tumor type (Shiu *et al.* 2010). These three phenotypes are

highly correlated with tumor grade (Shiu *et al.* 2010, Geyer *et al.* 2012, Habashy *et al.*

2012, Cornen *et al.* 2014). Luminal B tumors often have losses of Xp, 22q, 18p, 16q,

17q, 14q, 13q, 11q, 10q, 9p, 8p, 6q, 4p, and 1p and gains of 20q, 17q, 16p, 11q, 10p,

8q, and 1q (Garcia *et al.* 2005, Reis-Filho *et al.* 2006, Bernard-Pierrot *et al.* 2008,

Haverty *et al.* 2008, Shiu *et al.* 2010, Sircoulomb *et al.* 2011, Geyer *et al.* 2012, Habashy

et al. 2012).

In addition to large areas of chromosomal gains and losses, luminal B tumors often have more focal CNAs (Yang *et al.* 2006, Sircoulomb *et al.* 2011, Geyer *et al.* 2012, Ades *et al.* 2014, Cornen *et al.* 2014). High level amplification of specific genomic regions known as amplicons is a common mechanism of overexpression of the oncogenes found in that location (Ray *et al.* 2004, Bergamaschi *et al.* 2006, Ellis *et al.* 2007, Shiu *et al.* 2010, Yang *et al.* 2010, Bilal *et al.* 2012). Since amplicons frequently harbor multiple candidate oncogenes, identifying the "drivers" from the "passengers" is key to elucidating drug targets (Garcia *et al.* 2005, Haverty *et al.* 2008, Shiu *et al.* 2010, Mahmood *et al.* 2013, Chen *et al.* 2014, Irish *et al.* 2016). For example, amplification of 17q12 is well-known to induce overexpression of the HER2 oncogene and is associated with poor endocrine response and outcome (Still *et al.* 1999, Sorlie *et al.* 2001, Yang *et al.* 2004, Bilal *et al.* 2012). As HER2 is a driving oncogene, targeted therapies such as trastuzumab, a monoclonal antibody directed toward HER2, has been a highly successful intervention (Yang *et al.* 2010). Similarly, other amplicons are known to harbor driver oncogenes. C-MYC is found on the 8q24 amplicon (Still *et al.* 1999, Ray *et al.* 2004, Yang *et al.* 2004, Yang *et al.* 2006, Bernard-Pierrot *et al.* 2008, Yang *et al.* 2010). The 11q12-q14 amplicon, which harbors CCND1, an ERα-target gene associated with cell cycle progression, is found in up to 30% of breast cancers and is commonly associated with endocrine resistance and poorer outcome in ER+ tumors (Still *et al.* 1999, Ray *et al.* 2004, Yang *et al.* 2004, Yang *et al.* 2006, Bernard-Pierrot *et al.* 2008, Kwek *et al.* 2009, Shiu *et al.* 2010, Yang *et al.* 2010, Sircoulomb *et al.* 2011, Bilal *et al.* 2012, Ignatiadis and Sotiriou 2013, Lim *et al.* 2016). Despite frequent losses of 8p in luminal B breast cancers, a small region near the centromere is commonly amplified, known as the 8p11-p12 amplicon (Figure 1.3) (Ray *et al.* 2004, Garcia *et al.* 2005, Gelsi-

Boyer *et al.* 2005, Cooke *et al.* 2008, Shiu *et al.* 2010, Turner *et al.* 2010, Sircoulomb *et al.* 2011, Geyer *et al.* 2012, Ades *et al.* 2014, Cornen *et al.* 2014). This amplicon is fairly prevalent, found in 15% of all breast cancers but up to 32% of luminal B breast cancers, and is associated with worse outcome in these cancers (Ray *et al.* 2004, Yang *et al.* 2004, Gelsi-Boyer *et al.* 2005, Shiu *et al.* 2010, Sircoulomb *et al.* 2011, Reynisdottir *et al.* 2013, Zhang *et al.* 2013).

Figure 1.3. Schematic representation of the 8p11-p12 amplicon. Top panel depicts a cartoon representation of chromosome 8 with bands corresponding to those following Giemsa staining. The 8p11-p12 region is indicated. Middle panel depicts a cartoon representation of the 8p11-p12 amplicon region with landmark genes identified and the sub-regions A1-A4 identified. Bottom panel depicts a cartoon representation of the A2 sub-region of the 8p11-p12 amplicon region with genes identified as blue segments. ASH2L and NSD3 are highlighted in red to bring attention to their locations.

b. The 8p11-p12 amplicon and candidate oncogenes

Several groups have identified and validated oncogenes from the 8p11-p12 region and linked this amplicon to luminal B, ER+ tumors with poor metastasis-free survival, high histologic grade and Ki-67 proliferation indices, and resistance to

endocrine therapy (Ray *et al.* 2004, Garcia *et al.* 2005, Streicher *et al.* 2007, Bernard-Pierrot *et al.* 2008, Elsheikh *et al.* 2009, Shiu *et al.* 2010, Yang *et al.* 2010, Sircoulomb *et al.* 2011, Bilal *et al.* 2012, Geyer *et al.* 2012, Wu *et al.* 2012, Mahmood *et al.* 2013, Cornen *et al.* 2014). The 8p11-p12 amplicon ranges in size from 1 to 11 Mb but is most commonly approximately 5 Mb in length (Garcia *et al.* 2005, Yang *et al.* 2006, Melchor *et al.* 2007). Although there is variation in the specific start and end sites among tumors and cell lines that have been mapped, the NRG1 gene is the most common telomeric breakpoint site and usually includes loss of the remaining p arm of chromosome 8 (Ray *et al.* 2004, Garcia *et al.* 2005, Gelsi-Boyer *et al.* 2005, Melchor *et al.* 2007, Cooke *et al.* 2008), which harbors important tumor suppressor genes such as DLC1, SGCZ, and TUSC3 (Reis-Filho *et al.* 2006, Cooke *et al.* 2008) . The amplicon has four sub-regions of amplification, named A1-A4 from telomere to centromere (Gelsi-Boyer *et al.* 2005, Melchor *et al.* 2007, Sircoulomb *et al.* 2011, Reynisdottir *et al.* 2013, Turner-Ivey *et al.* 2014), with up to 21 genes demonstrating coordinated amplification and overexpression in breast cancer, depending on the technique used, thereby identified as candidate oncogenes (Ray *et al.* 2004, Yang *et al.* 2004, Garcia *et al.* 2005, Yang *et al.* 2006, Bernard-Pierrot *et al.* 2008, Cooke *et al.* 2008, Kwek *et al.* 2009, Yang *et al.* 2010, Holland *et al.* 2011, Bilal *et al.* 2012, Cornen *et al.* 2014, Turner-Ivey *et al.* 2014, Irish *et al.* 2016). Additionally, this amplicon is not unique to breast cancer, but has also been implicated in lung, bladder, pancreatic, and ovarian cancers (Figure 1.4) (Bernard-Pierrot *et al.* 2008, Yang *et al.* 2010, Dutt *et al.* 2011, Mahmood *et al.* 2013, Luo *et al.* 2017).

Many amplicons are well-known to play major roles in breast tumorigenesis and amplification is a common mechanism of overexpression on oncogenes. For example, the 17q12 amplicon contains HER2 (ERBB2) and identification and characterization of

Figure 1.4. Frequencies of ASH2L and NSD3 alterations in primary tumors (TCGA).
Histogram representing mutations of WHSC1L1 (NSD3) and ASH2L in all tumor types available in TCGA as of March 2017. Mutation frequency is noted on the y-axis and cancer type and specific study listed on the x-axis with inclusion of mutation and CNA data indicated by a + sign. Amplifications of WHSC1L1 and ASH2L are represented in red. Samples with lower than 7% total mutation rates (116 studies) were eliminated from the histogram.

this oncogene has led to the development of targeted therapies, such as trastuzumab (Still *et al.* 1999, Sorlie *et al.* 2001, Yang *et al.* 2004, Yang *et al.* 2010, Bilal *et al.* 2012). This oncogene has been considered the only driving oncogene from the 17q12 locus, but more recent studies have identified other genes, once thought of as passengers, to have driving roles in a subset of breast cancers with this amplicon, including GRB7 and STARD3 (Sorlie *et al.* 2001, Ray *et al.* 2004, Bernard-Pierrot *et al.* 2008, Haverty *et al.* 2008, Mahmood *et al.* 2013). Similarly, the 8q24 amplicon is common in breast tumors and harbors C-MYC, a well-known breast cancer oncogene (Still *et al.* 1999, Ray *et al.* 2004, Yang *et al.* 2004, Yang *et al.* 2006, Bernard-Pierrot *et al.* 2008). The 11q12-q14

amplicon, which contains CCND1, the gene encoding the cell cycle protein cyclin D1, has been shown in numerous studies to enhance breast cancer progression (Still *et al.* 1999, Ray *et al.* 2004, Yang *et al.* 2004, Yang *et al.* 2006, Bernard-Pierrot *et al.* 2008, Kwek *et al.* 2009, Shiu *et al.* 2010, Yang *et al.* 2010, Sircoulomb *et al.* 2011, Bilal *et al.* 2012, Ignatiadis and Sotiriou 2013, Lim *et al.* 2016). This amplicon also harbors CTTN (EMS1), GAB2, and other suspected oncogenes. Interestingly, the 11q12-q14 amplicon is commonly associated with the 8p11-p12 amplicon and cross-talk between these two genomic regions has been suggested, discussed in more detail below (Reis-Filho *et al.* 2006, Kwek *et al.* 2009, Sircoulomb *et al.* 2011, Bilal *et al.* 2012, Cornen *et al.* 2014).

Following discovery of the 8p11-p12 amplicon in breast cancer, several groups set out to map the region and identify the driver and passenger oncogenes (Still *et al.* 1999, Ray *et al.* 2004, Yang *et al.* 2004, Garcia *et al.* 2005, Gelsi-Boyer *et al.* 2005, Yang *et al.* 2006). Profiling of 8p11-p12 amplicon-bearing tumors and cell lines revealed incredible complexity and heterogeneity in this region, with multiple amplicon peaks observed (Yang *et al.* 2004, Reis-Filho *et al.* 2006, Melchor *et al.* 2007, Kwek *et al.* 2009, Wu *et al.* 2012, Luo *et al.* 2017). Several strong candidate oncogenes were identified based on commonality of amplification and coordinated overexpression, including: FGFR1, ZNF703, BRF2, RAB11FIP1, LSM1, PPAPDC1B, ASH2L, DDHD2, EIF4EBP1, KAT6A, TC-1, WHSC1L1, TACC1, ERLIN2, and PROSC (Ray *et al.* 2004, Garcia *et al.* 2005, Gelsi-Boyer *et al.* 2005, Yang *et al.* 2006, Bernard-Pierrot *et al.* 2008, Haverty *et al.* 2008, Elsheikh *et al.* 2009, Kwek *et al.* 2009, Yang *et al.* 2009, Reynisdottir *et al.* 2013, Zhang *et al.* 2013, Cornen *et al.* 2014). Indeed, many of these genes have been investigated further and their transforming properties validated. For example, TACC1 was implicated by Still *et al.* (1999) in cell transformation, anchorage-

independent growth, and proliferation. Similarly, LSM1 has been implicated by Streicher *et al.* (2007) as involved in grow-factor independent growth and soft agar colony-forming ability, both properties of transformed cells, when overexpressed in the non-transformed MCF10A breast cancer cell line. DDHD2 has also been implicated in regulating insulin-independent growth and disorganized acini formation in MCF10A cells grown in Matrigel when its overexpression is induced (Yang *et al.* 2010). PPAPDC1B was described by Bernard-Pierrot and colleagues as displaying transforming properties when overexpressed in a fibroblast cell line and having a negative regulatory effect on ERα while also promoting survival, proliferation, anchorage-independent growth in breast, lung, and pancreatic cancers (Bernard-Pierrot *et al.* 2008, Mahmood *et al.* 2013). This group also identified WHSC1L1, a histone methyltransferase, as a significant oncogene from the 8p11-p12 genomic region, but suggested it not to be the best avenue of investigation due to other oncogenes having greater potential as druggable targets (Bernard-Pierrot *et al.* 2008). They later pursued this gene further and cited its potential and importance as a drug target for epigenetic therapies, also mentioning ASH2L, another factor involved in histone methylation from the 8p11-p12 amplicon, as potentially interesting as well (Mahmood *et al.* 2013). These two oncogenes will be discussed in more detail in subsequent sections. Similarly, our group has also established KAT6A (MYST3), a histone acetyltransferase (HAT) discovered as an important gene via RNA-interference screening, to be essential for proliferation, colony formation, and anchorage-independent growth in the amplicon-bearing SUM-52 breast cancer cell line (Turner-Ivey *et al.* 2014). This study also identified several important cancer-related genes and biological pathways under control of this chromatin-modifying oncogene.

One hypothesis regarding amplicons is that the smallest area commonly amplified will contain the driving oncogene(s). To investigate this concept, a 1 Mb region in sub-region A1 was identified as the most conserved area of amplification (Garcia *et al.* 2005). Several groups identified five candidate oncogenes from this 1 Mb region based on coordinate amplification and overexpression: PROSC, ZNF703 (FLJ14299), ERLIN2 (C8orf2, SPFH2), BRF2, and RAB11FIP1 (ROC) (Garcia *et al.* 2005, Melchor *et al.* 2007, Haverty *et al.* 2008, Holland *et al.* 2011, Reynisdottir *et al.* 2013). Subsequent studies confirmed the transforming potential of several of these genes. ERLIN2 overexpression in the non-transformed MCF10A cell line resulted in cell proliferation in the absence of insulin and insulin-like growth factors, anchorage-independent growth, and highly proliferative colony formation (Yang *et al.* 2010). This gene has been shown to act as a "non-classical" oncogene, indirectly promoting cell survival via modulation of the stress response by the endoplasmic reticulum, an important adaptation in transformed cells (Wang *et al.* 2012). Little work has been done on BRF2 in 8p11-p12 amplicon-bearing cells. Our group demonstrated its overexpression can induce insulin-independent growth in MCF10A cells (Yang *et al.* 2010) but another group cited its elimination from consideration as a potential oncogene due to high expression in MCF10A cells at baseline (Kwek *et al.* 2009). Similarly, PROSC is often identified as a strong candidate oncogene (Reis-Filho *et al.* 2006, Haverty *et al.* 2008, Kwek *et al.* 2009, Yang *et al.* 2010, Holland *et al.* 2011, Reis-Filho and Pusztai 2011, Reynisdottir *et al.* 2013, Luo *et al.* 2017), but is understudied in the context of the 8p11-p12 amplicon. RAB11FIP1 is also understudied in this context, but has been implicated in PI3K/AKT/mTOR pathway activation in luminal B breast cancer and may play a role in protein trafficking (Yang *et al.* 2010, Cornen *et al.* 2014). Sircoulomb *et al.* (2011) and

Holland *et al.* (2011) simultaneously described ZNF703 (FLJ14299) as the only candidate oncogene of these five that is commonly and coordinately amplified and overexpressed across breast tumors harboring this site of focal amplification. These groups identified ZNF703 as a co-factor for a nuclear repressor complex that may contribute to transcriptional regulation and control of cell proliferation and clonogenicity via regulation of ERα and TGF-β, respectively, and both studies linked the 8p11-p12 amplicon to endocrine resistance.

Although not contained in this 1 Mb region, one of the more well-known oncogenes from this amplicon is fibroblast growth factor receptor-1 (FGFR1), which is located more proximally in sub-region A2 and amplified in 16-27% of breast cancers (Shiu *et al.* 2010, Turner *et al.* 2010, Ades *et al.* 2014). One study linked amplification and overexpression of FGFR1 to anchorage-independent proliferation, endocrine therapy resistance, and poor prognosis (Ignatiadis and Sotiriou 2013). Interestingly, FGFR1 is not always amplified, amplification does not always correlate with overexpression, its knockdown does not always result in decreased proliferation, and treatment with ligands FGF1 and FGF2 have been shown to actually inhibit proliferation of some 8p11-p12 amplicon-bearing breast cancer cells (Still *et al.* 1999, Ray *et al.* 2004, Yang *et al.* 2004, Haverty *et al.* 2008, Elsheikh *et al.* 2009, Turner *et al.* 2010, Mahmood *et al.* 2013). Therefore, FGFR1 is not universally accepted as a driver of all cancers with 8p11-p12 amplification and these tumors are more likely driven by oncogenes in close proximity to this gene (Melchor *et al.* 2007, Elsheikh *et al.* 2009, Turner *et al.* 2010, Bilal *et al.* 2012). More recent studies investigating the role of FGFR1 as a driving oncogene from this genomic region actually identified eight other amplicon genes as better predictors of FGFR1 amplification than FGFR1 itself, including ASH2L,

BAG4, BRF2, DDHD2, LSM1, PROSC, RAB11FIP1, and WHSC1L1 (NSD3) (Luo *et al.* 2017). Two of these genes, ASH2L and NSD3, also mentioned above, are the focus of the work presented here and are discussed in detail below.

As investigation of the oncogenes from the 8p11-p12 amplicon has progressed, new insight has been gained into the function of this genetic change in breast cancer. This has led to the development of new hypotheses regarding the potential cooperation of the multiple driving oncogenes that have been identified and described. As mentioned previously, the 11q12-q14 amplicon has been hypothesized to cooperate with the 8p11-p12 amplicon due to their coordinated amplification, where amplification of 8p11-12 occurs in 40% of 11q12-14 amplified cases (Reis-Filho *et al.* 2006, Kwek *et al.* 2009, Shiu *et al.* 2010, Yang *et al.* 2010, Sircoulomb *et al.* 2011, Bilal *et al.* 2012, Ignatiadis and Sotiriou 2013, Lim *et al.* 2016). CCND1 (11q13) induces expression of ZNF703 (8p12) via the RB/E2F pathway and expression of ZNF703, BAG4, RAB11FIP1, and WHSC1L1 (NSD3), all 8p11-p12 amplicon genes, is increased when 11q12-q14 is amplified, independent of concomitant 8p11-p12 amplification (Kwek *et al.* 2009). Overexpression of CCND1 in the SUM-44 cell line, which harbors both amplicons, resulted in upregulation of these 8p11-p12 amplicon genes. Although not fully explored, this study also indicated evidence that FGFR1 and DDHD2 from the 8p11-p12 amplicon may interact with C-MYC, a commonly amplified oncogene from the 8q24 region (Kwek *et al.* 2009). Together, this study provides evidence that the complex genomic structures of luminal B tumors may be due in part to interactions between driving oncogenes from multiple amplicons.

It was suggested that oncogenes from the 8p11-p12 amplicon cooperate to promote tumorigenesis, a concept that is supported by groups studying this amplicon

(Gelsi-Boyer *et al.* 2005, Yang *et al.* 2006, Sircoulomb *et al.* 2011, Wu *et al.* 2012, Luo *et al.* 2017). This hypothesis of cooperation partially explains the complexity of this amplicon and the plethora of oncogenes in this "hot spot" of genomic activity (Ray *et al.* 2004). The candidate oncogenes from this region can often be grouped by similar function, such as regulation of RNA metabolism, vesicle trafficking, tyrosine kinase activity, chromatin maintenance, and regulation of ERα (Gelsi-Boyer *et al.* 2005, Yang *et al.* 2010, Turner-Ivey *et al.* 2014, Irish *et al.* 2016). The latter two functions are of particular interest to our group since all three of the 8p11-p12 oncogenes implicated in epigenetic regulation of chromatin structure, KAT6A (MYST3), WHSC1L1 (NSD3), and ASH2L, are also implicated in regulation of ERα and/or ESR1, the gene which encodes ERα (Ades *et al.* 2014, Qi *et al.* 2014, Turner-Ivey *et al.* 2014, Irish *et al.* 2016). Indeed, genetic abnormalities, such as amplifications, are increasingly linked to epigenetic dysregulation in cancers, leading to enhanced focus on understanding the role of chromatin modification in tumorigenesis and key players and processes that can become therapeutic targets in breast and other cancers.

1.3. Breast Cancer Epigenomics

a. Epigenetics overview

Epigenetics is emerging as an important mechanism of transcriptional dysregulation in cancer and understanding the effects of various chromatin-modifying enzymes on gene expression patterns related to cell proliferation, survival, DNA repair, and other carcinogenic processes is essential to prevention, diagnosis, and treatment of cancer. Epigenetics is defined as changes to the expression profile of a cell not accounted for by alterations to the DNA sequence (Miyamoto and Ushijima 2005,

Kristensen *et al.* 2009, Lustberg and Ramaswamy 2009, Jovanovic *et al.* 2010, Pathiraja *et al.* 2010, Grzenda *et al.* 2011, Lustberg and Ramaswamy 2011). It includes DNA methylation, histone post-translational modifications (hPTM), non-coding RNA, and nucleosome remodeling and is a multi-layer regulation system that is complex by its nature (Margueron *et al.* 2004, Jovanovic *et al.* 2010, Rijnkels *et al.* 2010, Lustberg and Ramaswamy 2011, Abdel-Hafiz and Horwitz 2015, Noberini and Bonaldi 2017).

The first and best-characterized epigenetic modification investigated is DNA methylation at CpG islands in promoter regions by DNA methyltransferase (DNMT) enzymes (Lustberg and Ramaswamy 2009). In general, DNA methylation results in gene silencing and cancers tend to have global hypermethylation (suppression) with hypomethylation (expression) of specific genes, such as those involved in cell cycle progression (Lo and Sukumar 2008, Jovanovic *et al.* 2010, Pathiraja *et al.* 2010, Huang *et al.* 2011). For this reason, inhibitors of the methyltransferase enzymes responsible for these events have proven clinically successful at re-expression of heavily methylated genes and have a positive influence on patient survival (Cai *et al.* 2011, Falahi *et al.* 2014). One such example is the re-expression of ERα and subsequent re-sensitization of ER-negative breast cancers to endocrine therapy (Kristensen *et al.* 2009, Falahi *et al.* 2014, Abdel-Hafiz and Horwitz 2015).

In addition to methylation of the DNA, modifications to the chromatin structure and regulation of access to the genetic information contained therein is emerging as an essential component of tumor initiation and progression in breast and other types of cancer (Lo and Sukumar 2008, Grzenda *et al.* 2011, Kim *et al.* 2016). Chromatin is composed of nucleosomes, units of 147 base pair segments of DNA wrapped around octamers of histone proteins, made up of two each of histone proteins H2A, H2B, H3,

and H4 (Lo and Sukumar 2008, Kristensen *et al.* 2009, Souza *et al.* 2009, Jovanovic *et al.* 2010, Grzenda *et al.* 2011, Kumar *et al.* 2016). These nucleosomes are linked via histone H1 and complexed with additional proteins into areas of highly condensed, transcriptionally inactive heterochromatin and open, transcriptionally active euchromatin (Lo and Sukumar 2008, Kristensen *et al.* 2009, Shan *et al.* 2014, Kumar *et al.* 2016). Within nucleosomes, the N-terminal tails of the histone proteins are left exposed, rendering them susceptible to post-translational modification by epigenetic enzymes (Liu *et al.* 2009, Jovanovic *et al.* 2010, Kim *et al.* 2016). Known hPTMs include acetylation, methylation, sumoylation, ubiquitination, phosphorylation, and ADP-ribosylation (Lo and Sukumar 2008, Souza *et al.* 2009, Jovanovic *et al.* 2010, Grzenda *et al.* 2011, Huang *et al.* 2011). Different modifications on different histone tail amino acid residues in different regions of the genome produce a variety of effects on the transcriptional program of the cell (Figure 1.5) (Liu *et al.* 2009, Kuo *et al.* 2011, Kim *et al.* 2014, Osmanbeyoglu *et al.* 2014, Kumar *et al.* 2016). These modifications have such a profound effect on gene expression that they account for much of the cellular programming during embryogenesis that produces a vast number of functionally differentiated cell and tissue types from a single set of genetic information, and conversely their dysregulation can aberrantly revert differentiated cells back to a pluripotent state, which has major implications in tumorigenesis (Lo and Sukumar 2008, Liu *et al.* 2009, Sircoulomb *et al.* 2011, Nagamatsu *et al.* 2012, Osmanbeyoglu *et al.* 2014).

Since epigenetic factors play a major role in altering the expression of large sets of genes simultaneously, it is no surprise that aberrant expression of these enzymes can lead to cancer (Lo and Sukumar 2008, Demircan *et al.* 2009, Liu *et al.* 2009, Yang *et al.* 2009, Kim *et al.* 2014, De Marchi *et al.* 2016, Kim *et al.* 2016). Global studies of hPTMs

Figure 1.5. Illustration of major histone modifications and effects on chromatin. Top panel depicts heterochromatin and the histone modifications that favor this configuration: H3K9me3, H3K27me3, H4K20me3, and DNA methylation. Bottom panel depicts euchromatin and the histone modifications that favor this configuration: H3K4me3, H3K9ac, H3K36me3, H4K5ac. These listed modifications are not comprehensive. A key in included for interpretation of the different representations. Figure is based on a similar representation from Wu *et al* 2010.

demonstrate that early stage tumors have a different epigenomic profile than late stage tumors and the different molecular subtypes of breast cancer can be further clarified by analyzing their epigenomic states (Garapaty *et al.* 2009, Messier *et al.* 2016). An

important aspect of epigenetics is the ability of the modifications to chromatin structure to be both stable, with heritability of the epigenetic state of cells being an essential component of development and maintenance of phenotype, and dynamic, with different patterns of gene expression required throughout the life of a cell and the mitotic process (Miyamoto and Ushijima 2005, Lustberg and Ramaswamy 2009, Pathiraja *et al.* 2010, Allali-Hassani *et al.* 2014). As such, most epigenetic modifications, unlike genetic changes, are reversible (Lo and Sukumar 2008, Simon and Lange 2008, Kristensen *et al.* 2009, Lustberg and Ramaswamy 2009, Pathiraja *et al.* 2010, Cai *et al.* 2011, Morishita and di Luccio 2011, Liu *et al.* 2014, Michalak and Visvader 2016). These features make epigenetic enzymes and the chromatin alterations they generate key drug targets in cancer and other diseases.

b. Histone acetylation and methylation

The best-studied hPTM thus far is acetylation. Histone acetylation usually occurs on lysine (K) residues in the exposed histone tails, especially K4, K9, and K27 of H3 and K5, K8, K12, and K16 of H4 (Figure 1.5) (Lo and Sukumar 2008, Kristensen *et al.* 2009, Lustberg and Ramaswamy 2009, Huang *et al.* 2011, Judes *et al.* 2016, Messier *et al.* 2016). Histone acetylation weakens the charge interaction between DNA and histones, relaxing the chromatin, and therefore is generally associated with transcriptionally active genes (Lo and Sukumar 2008, Lustberg and Ramaswamy 2009, Pathiraja *et al.* 2010, Zhang *et al.* 2013). In normal cells, histone acetylation is regulated by opposing activity of histone acetyltransferase (HAT) and histone deacetylase (HDAC) enzymes (Margueron *et al.* 2004, Lo and Sukumar 2008, Huang *et al.* 2011). Acetylated histones also interact with other epigenetic changes, such as DNA methylation, to contribute to overall transcriptional balance (Lustberg and Ramaswamy 2009, Pathiraja *et al.* 2010).

Some cancer types have high levels of global histone acetylation associated with aberrant transcription of cell cycle progression programs while others have low levels, correlating to overactivity of HDACs that shut down transcription of tumor suppressors (Lo and Sukumar 2008, Elsheikh *et al.* 2009, Judes *et al.* 2016, Messier *et al.* 2016). HDAC inhibitors have been studied for their potential to promote hyperacetylation and transcription of pro-apoptotic genes and show promising results in clinical trials, especially in combination with DNMT inhibitors and endocrine therapies (Margueron *et al.* 2004, Lo and Sukumar 2008, Kristensen *et al.* 2009, Lustberg and Ramaswamy 2009, Cai *et al.* 2011, Connolly and Stearns 2012, Abdel-Hafiz and Horwitz 2015, Borbely *et al.* 2015, Connolly *et al.* 2016, Walsh *et al.* 2016). HDAC inhibitors can also promote acetylation of proteins such as HSP90, a chaperone protein for clients such as ERα, AKT, HER2, cRAF, and others, which are then destabilized and degraded (Lo and Sukumar 2008, Lustberg and Ramaswamy 2009). Depending on the tumor type and subtype, this can be beneficial in cancer treatment. HDAC inhibitors have shown success in several hematologic malignancies but have yet to demonstrate much benefit in breast cancer (Kerdivel *et al.* 2013, Falahi *et al.* 2014).

Histone methylation is less understood and more complex than histone acetylation (Lo and Sukumar 2008, Huang *et al.* 2011). Histone methylation occurs at K and arginine (R) residues, though is almost exclusively found on K, and can be both activating and repressive depending on the specific residue, location within a gene and how many methyl groups are added (Lo and Sukumar 2008, Elsheikh *et al.* 2009, Morishita and di Luccio 2011). K residues can accept up to three methyl groups in a stepwise fashion, and different enzymes are involved in catalyzing mono-, di-, and tri-methylation as well as de-methylation events (Li *et al.* 2009, Liu *et al.* 2009, Huang *et al.*

2011, Kim *et al.* 2014, Kim *et al.* 2016). In general, the most common activating

methylation marks include H3K4me3, H3K36me3, and H3K79me3 and repressive marks

include H3K9me3, H3K27me3, and H4K20me3 (Lo and Sukumar 2008, Elsheikh *et al.*

2009, Souza *et al.* 2009, Fang *et al.* 2010, Jovanovic *et al.* 2010, Huang *et al.* 2011,

Morishita *et al.* 2014). Similarly, arginine residues can be methylated by protein-arginine

methyltransferase family proteins, and H4R3me3 is associated with activated

transcription (Elsheikh *et al.* 2009, Butler *et al.* 2011). Although the exact specificity of

different histone methyltransferase (HMT) enzymes responsible for placing these marks

is still controversial, some general patterns have emerged.

       The first HMTs, though not yet known to function as such, were identified in the

1940s in *Drosophila melanogaster* when it was observed that loss of homeobox (HOX)

gene expression in male flies produced extra sex combs, and the phenomenon was

termed "polycomb" (Grzenda *et al.* 2011). These polycomb group (PcG) enzymes that

negatively regulate HOX gene expression have since been identified as the SET

(S̲uppressor of variegation 3-9, E̲nhancer of zeste and T̲rithorax) domain-containing

HMTs EZH1 and EZH2 and their complex partners (Angrand *et al.* 2001, Lo and

Sukumar 2008, Simon and Lange 2008, Grzenda *et al.* 2011). The SET domain provides

the catalytic activity for the 4 subgroups, PRC1 (polycomb repressive complex), which is

associated with transcriptional maintenance, and PRC2, PRC3, and PRC4 that are

associated with transcriptional initiation (Grzenda *et al.* 2011, Huang *et al.* 2011, Yuan *et*

al. 2011, Garapaty-Rao *et al.* 2013). PcG factors exert their repressive function primarily

via tri-methylation of H3K27 and H3K9 and are critical for embryonic development, stem

cell maintenance, differentiation, proliferation, and invasion (Rampalli *et al.* 2007, Simon

and Lange 2008, Pathiraja *et al.* 2010, Grzenda *et al.* 2011, Yuan *et al.* 2011, Judes *et*

al. 2016). EZH2 is one of the most studied HMTs to date and EZH2 inhibitors are now in clinical trials (Garapaty-Rao *et al.* 2013, Sato *et al.* 2017). Activity of PcG enzymes is found at both active and inactive genes and transcriptional repression and chromatin compaction is thought to be the "default" setting of cells, with activation of transcription occurring through "de-repression" in a tightly controlled fashion (Rampalli *et al.* 2007, Yuan *et al.* 2011).

Until the 1980s, it was unclear how PcG enzyme activity was counterbalanced (Grzenda *et al.* 2011). Also in *Drosophila melanogaster*, the trithorax group (trxG) of chromatin modifiers was discovered (Ikegawa *et al.* 1999, Luscher-Firzlaff *et al.* 2008, Grzenda *et al.* 2011, Yuan *et al.* 2011). This family of proteins is made up of HMT enzymes that also contain SET domains, which possesses the catalytic HMT function (Ikegawa *et al.* 1999, Angrand *et al.* 2001, Cao *et al.* 2010, Judes *et al.* 2016). TrxG enzymes counteract PcG repression primarily by H3K4 tri-methylation, mediated by MLL (mixed lineage leukemia) and other ASH2L-containing complexes, leading to active transcription (Guertin *et al.* 2006, Rampalli *et al.* 2007, Grzenda *et al.* 2011). Many promoters possess both H3K27me3 and H3K4me3 at high levels and are termed "poised" chromatin regions, though the "switch" that determines which direction gene transcription ultimately favors is unclear and may depend on cross-talk with other layers of epigenetic regulation, such as DNA methylation and histone acetylation (Grzenda *et al.* 2011, Wan *et al.* 2013, Ullius *et al.* 2014, Katoh 2016). MYC may be involved in recruitment of other enzymes, such as p300, that de-methylate and acetylate H3K27 for transcriptional activation in an ASH2L-dependent manner (Hervouet *et al.* 2013, Ullius *et al.* 2014, Erfani *et al.* 2015). SET domain-containing proteins also di-methylate H3K36, another marker of active gene transcription (Li *et al.* 2009, Liu *et al.* 2009, Kuo *et al.*

2011, Yuan *et al.* 2011, French *et al.* 2014, Liu *et al.* 2014, Zhu *et al.* 2016). These

enzymes include ASH1L and the NSD family of HMTs, discussed in more detail along

with ASH2L and the MLL complexes in subsequent sections (Figure 1.6).

Although chromatin-modifying enzymes are often studied individually, the reality

is that many of these enzymes are found in complexes together, and activation of one

tends to facilitate recruitment of others (Wang *et al.* 2007, Nguyen *et al.* 2008, Simon

and Lange 2008, Li *et al.* 2009, Yates *et al.* 2010, Huang *et al.* 2011, Kumar *et al.* 2016).

This is exemplified in the SWI/SNF and similar complexes that have been well-

characterized and are known to have a vast selection of subunits that contribute to

chromatin architecture in a plethora of biological contexts (Yates *et al.* 2010, Kumar *et

al.* 2016). Similarly, transcription factors such as ERα rely on several epigenetic proteins

for local remodeling of nucleosomes to facilitate transcription, create positive feedback

loops to maintain transcription, and auto-regulate shutting down of these pathways to

prevent aberrant gene expression (Margueron *et al.* 2004, Lo and Sukumar 2008, Li *et

al.* 2009, Fang *et al.* 2010, Jacques-Fricke and Gammill 2014, Locke *et al.* 2015). Some

of the better-characterized binding partners of ERα include the HATs p300/CBP, PCAF,

and SRC1 (NCoA1, nuclear co-activator 1), which function as co-activators to destabilize

the nucleosomes and open the chromatin for transcription (Margueron *et al.* 2004,

Garapaty *et al.* 2009, Hervouet *et al.* 2013, Zhang *et al.* 2013). Various HMT,

demethylase, and DNMT enzymes, including HMTs MLL1 and MLL2, also complex with

ERα to close the chromatin once the transcriptional pattern is complete (Cui *et al.* 2005,

Huang *et al.* 2011, Hervouet *et al.* 2013, Zhang *et al.* 2013, Ades *et al.* 2014). In addition

to histones, epigenetic factors can modify non-histone proteins. For example, SMYD2, a

HMT, methylates K266 in the hinge region of the ERα protein itself in the absence of

Figure 1.6. Hierarchical representation of histone lysine methyltransferase enzyme relationships. Known lysine-specific HMT enzymes and corresponding substrates, including both validated and suggested substrates, appear on the right side with relationships illustrated by line groupings. Line lengths are arbitrary. The NSD family is bold along with the corresponding validated substrate H3K36. ASH2L does not have a SET domain but is homologous to ASH1L, listed below the NSD family, and complexes with the MLL family listed at the top of the diagram. Repressive marks (H3K27, H3K9, and H4K20) identify PcG group enzymes and activating marks (H3K4 and H3K36) identify trxG group enzymes, respectively. Diagram based on Wu *et al* 2010.

17β-estradiol and LSD1, a demethylase, must first remove this inhibitory K266me prior to acetylation at the same residue by p300/CBP, which contributes to ERα activation upon stimulation with 17β-estradiol (Zhang *et al.* 2013). Many other enzymes are also implicated in direct modification of ERα, particularly related to the hinge-region ligand-binding and ligand-independent AF-1 domains (Margueron *et al.* 2004, Kerdivel *et al.* 2013). When any of the mechanisms contributing to transcriptional access are disrupted, tumorigenesis can result.

An essential feature of epigenetics is the ability for precise interpretation of the combinatorial nature of histone modifications and translation to transcriptional signals (Grzenda *et al.* 2011, Hervouet *et al.* 2013). This is accomplished by the various "reader" domains possessed by epigenetic enzymes. Bromodomains recognize acetylated histones and recruit additional factors to the chromatin (Lo and Sukumar 2008, Grzenda *et al.* 2011, Rahman *et al.* 2011, Wagner and Carpenter 2012). The bromodomain and extraterminal domain (BET) family of proteins, BRDT, BRD2, BRD3, and BRD4, have been implicated as "linker" or "bridge" proteins in various cancer types between acetylated histones and other epigenetic factors and have been the focus of drug development leading to clinical trials in recent years (Morishita and di Luccio 2011, Rahman *et al.* 2011, Feng *et al.* 2014, Zou *et al.* 2014, Nagarajan *et al.* 2015, Shen *et al.* 2015, Crowe *et al.* 2016, Kumar *et al.* 2016, Zhang *et al.* 2016). Other common domains include WD40, ADD, and chromodomains (CHD), which recognize methylated lysine residues (Guccione *et al.* 2007, Lo and Sukumar 2008, Grzenda *et al.* 2011, Kim *et al.* 2016). Plant homeodomains (PHD) are zinc finger motifs that recognize methylated lysines and arginine residues, preferentially H3K4me3 and H3K9me3 (Huang *et al.* 1998, Grzenda *et al.* 2011, Wagner and Carpenter 2012, He *et al.* 2013, Allali-Hassani

et al. 2014, Morishita *et al.* 2014, Kim *et al.* 2016). PWWP (Pro-Trp-Trp-Pro) domains are highly conserved and preferentially recognize H3K36me1 and H3K36me2 (Yang *et al.* 2010, Grzenda *et al.* 2011, Wagner and Carpenter 2012, Allali-Hassani *et al.* 2014, Morishita *et al.* 2014, Wen *et al.* 2014, Shen *et al.* 2015, Kim *et al.* 2016). The PHD5-C5HCH domain unique to the NSD family of HMTs also plays an important role in directing these enzymes to chromatin, preferentially binding H3K4me0 and H3K9me3 (He *et al.* 2013). These reader domains generally bind multiple histone and non-histone PTMs with substrate affinities dependent upon the three-dimensional structures of the domains, presence of coregulatory factors that cause conformational changes to those binding pockets, nearby histone and DNA modifications that can have allosteric effects on HMTs, and other factors, increasing the complexity of the rich language of the histone code.

As the field of epigenetics has expanded exponentially over the past two decades (Huang *et al.* 2011), it has become clear that the complexity and interaction between all the chromatin regulatory marks, together termed the "histone code" (Grzenda *et al.* 2011, Hervouet *et al.* 2013), needs further evaluation. There is a clear link between aberrant expression of epigenetic modifiers of chromatin and pathology, especially tumorigenesis (Miyamoto and Ushijima 2005, Lo and Sukumar 2008, Demircan *et al.* 2009, Li *et al.* 2009, Kim *et al.* 2014, Paska and Hudler 2015, Kim *et al.* 2016, Zhu *et al.* 2016). Globally-targeted inhibition of some of these enzymes, such as the DNMT, HDAC, and bromodomain inhibitors, have proven that this pathway of investigation is worthwhile and has highlighted the need to fine-tune pharmacologic control of epigenetic editing in a cancer subtype-specific manner (Lo and Sukumar 2008, Falahi *et al.* 2014, French *et al.* 2014). For a more detailed discussion of the topics

briefly addressed in this section, interested readers are referred to several comprehensive review articles (Lo and Sukumar 2008, Lustberg and Ramaswamy 2009, Jovanovic *et al.* 2010, Rijnkels *et al.* 2010, Veeck and Esteller 2010, Cai *et al.* 2011, Connolly and Stearns 2012, Falahi *et al.* 2014, Lin *et al.* 2015, Paska and Hudler 2015, Michalak and Visvader 2016, Walsh *et al.* 2016, Damaskos *et al.* 2017). Only the implications of histone methylation on lysine residues 4 and 36 of histone 3 will be discussed further as these groups are regulated by the two 8p11-p12 amplicon factors involved in histone methylation, ASH2L and NSD3, respectively.

1.4. Histone Methyltransferases from the 8p11-p12 Amplicon

a. NSD3 (WHSC1L1) and H3K36me2

The Nuclear-receptor binding SET Domain-containing (NSD) family of proteins is part of the trithorax (trx) group of chromatin-modifying enzymes and has three main members: NSD1, NSD2 (WHSC1/MMSET), and NSD3 (WHSC1L1 for Wolf-Hirschhorn Syndrome Candidate 1-Like 1) (Huang *et al.* 1998, Yang *et al.* 2010). These enzymes are HMTs and share a large degree of homology with one another as well as with other SET domain-containing proteins conserved throughout evolution (Angrand *et al.* 2001, Li *et al.* 2009, Wu *et al.* 2010, Yang *et al.* 2010, Zhou *et al.* 2010, Morishita and di Luccio 2011, Katoh 2016). All three family members are dysregulated in a number of benign and malignant disease states and they have all been established as oncogenes (Douglas *et al.* 2005, Yang *et al.* 2010, Morishita and di Luccio 2011, He *et al.* 2013). Much of the work on these enzymes has been done in hematological malignancies, primarily acute myeloid leukemia (AML) for NSD1 and NSD3 and multiple myeloma for NSD2, but they have also been implicated in lung, pancreatic, colorectal, liver, bladder,

glioblastoma, neuroblastoma, head and neck carcinoma, and breast cancers (Stec *et al.* 2001, Li *et al.* 2009, Taketani *et al.* 2009, Zhou *et al.* 2010, Morishita and di Luccio 2011, Morishita and di Luccio 2011, He *et al.* 2013, Morishita *et al.* 2014, Suzuki *et al.* 2015, Saloura *et al.* 2016). Though there has been much debate over the specificity of these HMTs and some controversy remains, it has been established that they all possess H3K36 methylation ability and NSD3 is unambiguously implicated in di-methylation of this lysine residue, which is generally associated with active transcription (Li *et al.* 2009, Kuo *et al.* 2011, Morishita and di Luccio 2011, Rahman *et al.* 2011, Wagner and Carpenter 2012, He *et al.* 2013, Allali-Hassani *et al.* 2014, Feng *et al.* 2014, French *et al.* 2014, Jacques-Fricke and Gammill 2014, Morishita *et al.* 2014, Saloura *et al.* 2016, Bennett *et al.* 2017). NSD2 is implicated in the greatest number of histone methylation events (Li *et al.* 2009, Morishita and di Luccio 2011, Morishita *et al.* 2014). In addition to histone methylation, NSD1 has demonstrated methylation of NF-κb and NSD3 has been shown to methylate the epidermal growth factor receptor (EGFR), activating downstream events in the absence of EGF that lead to DNA replication in squamous cell carcinomas of the head and neck (Morishita and di Luccio 2011, Saloura *et al.* 2017). NSD2 and NSD3 share the highest degree of homology compared to NSD1, and all three have highly conserved sequences (Angrand *et al.* 2001, Yang *et al.* 2010, Zhou *et al.* 2010, Morishita and di Luccio 2011). Knockout mouse models confirm that the functions of these three enzymes are non-redundant and they likely perform unique roles in development and, then, in tumorigenesis as well (He *et al.* 2013, Allali-Hassani *et al.* 2014, Bennett *et al.* 2017).

NSD3 has three isoforms (Figure 1.7). The longest isoform, dictated NSD3-L, is 1437 amino acids from 24 exons located at 8p11.23 and contains an N-terminal PWWP

domain followed by 4 sequential PHD fingers, a second PWWP domain, a SAC (SET-associated cysteine-rich) domain, SET domain (catalytic HMT domain), and a final PHD finger linked to a C5HCH (cys-his-rich) domain that is unique to the NSD family (Angrand *et al.* 2001, Stec *et al.* 2001, Yang *et al.* 2010, Zhou *et al.* 2010, Morishita and di Luccio 2011, Morishita and di Luccio 2011, He *et al.* 2013, Saloura *et al.* 2016). WHISTLE is a shorter, C-terminal version of this protein generated by a downstream promoter that is specific to testes and bone marrow (Kim *et al.* 2007, Zhou *et al.* 2010, Bennett *et al.* 2017). It recruits HDAC1 with its PWWP domain and methylates H3K4 and H3K27 to repress transcription and may also induce apoptotic cell death through activation of caspase-3 (Kim *et al.* 2007, Zhou *et al.* 2010). This isoform has not been shown to have a role in tumorigenesis (Bennett *et al.* 2017). Alternative splicing of the NSD3 gene at exon 9 produces a truncated 645 amino-acid protein, NSD3-S, that contains only the first PWWP chromatin-binding domain and lacks the catalytic SET and other binding domains (Angrand *et al.* 2001, Stec *et al.* 2001, Yang *et al.* 2010, Zhou *et al.* 2010, Saloura *et al.* 2017). While both the long and short isoforms of NSD3 demonstrate overexpression at the transcript and protein levels in tumor versus normal tissues in multiple cancer types, NSD3-S is consistently identified as expressed to a

Figure 1.7. Representation of the three isoforms of NSD3. Top panel depicts the long, 1437 amino acid isoform of NSD3 that contains two PWWP domains, 5 PHD domains, one SAC domain, one SET domain, and a C5HCH domain. Middle panel depicts the short, 645 amino acid isoform of NSD3 that contains only a PWWP domain and the N-terminal exons of the long isoform. Bottom panel depicts WHISTLE, a testes-specific, 495 amino acid isoform of NSD3 that contains the C-terminal region of the long isoform, including one PWWP domain, the SAC, SET, C5HCH, and fifth PHD domains.

higher degree than NSD3-L (Stec *et al.* 2001, Yang *et al.* 2010, Shen *et al.* 2015, Irish *et al.* 2016).

There is mounting evidence to indicate that NSD3-S is more transforming than NSD3-L despite lacking the catalytic SET domain. The 8p11-p12 amplicon often contains breakpoints and deletions in the 3' end of the NSD3 sequence, originally leading to the conclusion that it was not a driving oncogene from this amplicon (Yang *et al.* 2010, Dutt *et al.* 2011). Several studies have alternatively suggested that the short isoform actually is the dominant oncogenic form of this protein and is under selective pressure during the amplification process to preserve the 5' end of the transcript (Nagamatsu *et al.* 2012). Shen *et al.* (2015) demonstrated NSD3-S as the essential oncogenic isoform in AML by serving as an adaptor protein to link BRD4 to other chromatin-modifying enzymes such as CHD8, and the PWWP domain was necessary for this function. This group also identified a new domain on NSD3-S, a transcription activating domain (TAD) comprised of the highly acidic first 100 amino acids, suggesting that NSD3 can exert oncogenic function and transcriptional activation independent of HMT activity. This is consistent with other studies that have demonstrated the necessity of NSD3 in regulating gene transcription but with obvious disconnect between transcriptional outcome and H3K36me2 alterations (Jacques-Fricke and Gammill 2014). Others have postulated that NSD3-S acts by binding to chromatin through its PWWP domain and preventing the methyltransferase function of NSD3-L by occupying the binding sites (Stec *et al.* 2001, Irish *et al.* 2016). It is important to note that, although NSD3-S somewhat unexpectedly has been shown to play a more transforming role than NSD3-L in some studies, there is also evidence to indicate that the enzymatic activity of NSD3-L that generates H3K36me2 is also important in some tumors (Chen *et al.* 2014,

Liu *et al.* 2014, Suzuki *et al.* 2015). While the distinction between the two main isoforms of NSD3 remains to be fully understood, it has become clear that NSD3 is an important oncogene in many cancer types.

Although one study demonstrated that knockdown of NSD3-L in non-NSD3-amplified cell lines resulted in increased cell cycle progression and proliferation, suggesting that NSD3-L may be a tumor suppressor gene (Zhou *et al.* 2010), all others have confirmed it is a bona fide oncogene (Morishita and di Luccio 2011, He *et al.* 2013, Chen *et al.* 2014). This study also demonstrated increased NSD3-L transcript levels compared to NSD3-S (Zhou *et al.* 2010), a finding that is challenged by the body of literature on this gene as well as findings from our laboratory in the same MCF7 breast cancer cell line (Stec *et al.* 2001, Yang *et al.* 2010, Shen *et al.* 2015, Irish *et al.* 2016). NSD3, like NSD1, has been shown to create an oncogenic fusion protein with NUP98 in AML and is associated with poor prognosis and increased aggressiveness, possibly by competing with EZH2-mediated repression of HOX-A gene expression (Rosati *et al.* 2002, Wang *et al.* 2007, Li *et al.* 2009, Taketani *et al.* 2009). Similarly, a fusion protein of the first 7 exons of NSD3 with the NUT gene in NUT midline carcinomas of the lung and mediastinum produced a phenotype of high proliferation and lack of differentiation (French *et al.* 2014). This cancer is highly aggressive with few treatment options and poor patient survival and would benefit from inhibition of NSD3, which induced differentiation and reduced proliferation in a NUT midline carcinoma model cell line (French *et al.* 2014, Suzuki *et al.* 2015). Increased NSD3 expression is associated with reversion of cell phenotype toward pluripotency in other models as well (Yang *et al.* 2010, Nagamatsu *et al.* 2012), consistent with a role in tumorigenesis. NSD3 has been implicated in pancreatic cancer and non-small cell lung cancer progression through its

chromatin modifying behavior (Zhou *et al.* 2010, Morishita and di Luccio 2011, Mann *et al.* 2012, Kang *et al.* 2013, Mahmood *et al.* 2013, Jacques-Fricke and Gammill 2014, Suzuki *et al.* 2015).

NSD3 has been shown to bind several other proteins. BRD4, whose bromodomain recognizes acetylated histones, also possesses an extraterminal (ET) domain that binds NSD3 and other chromatin-modifying enzymes (Rahman *et al.* 2011, Wagner and Carpenter 2012, Feng *et al.* 2014, Shen *et al.* 2015, Crowe *et al.* 2016, Zhang *et al.* 2016). The BRD4-NSD3 unit participates in activating transcription at several important target genes, including CCND1 (Rahman *et al.* 2011, Wagner and Carpenter 2012). This complex is found primarily in the gene body and is associated with H3K36me2, the histone mark for which NSD3 is responsible (Rahman *et al.* 2011). Rahman *et al.* (2011) also identified BRD2 as an NSD3-binding partner. The unique PHD5-C5HCH domain of the NSD family members has been extensively studied by He *et al.* (2013) and, though each member has different binding affinity, this motif in NSD3 was demonstrated to prefer H3K4me0 and H3K9me3, both markers of repressed chromatin. This is consistent with previous studies demonstrating that NSD3 complexes with LSD2, a H3K4 demethylase, and G9a, another SET enzyme responsible for the repressive H3K9me3 mark (Morishita and di Luccio 2011, Wagner and Carpenter 2012, He *et al.* 2013, Feng *et al.* 2014). It has been proposed that these proteins coordinate the precise dynamic balance of H3 modifications that facilitate optimal transcriptional elongation while simultaneously preventing unwanted initiation of intragenic transcription at target genes (Feng *et al.* 2014).

Methylation of H3K36 is generally associated with transcriptional activity. H3K36me2, the major histone-methyl mark of NSD3, is an important hPTM for the

recruitment of other enzymes involved in transcriptional activation (Li *et al.* 2009, Wagner and Carpenter 2012). Methylated H3K36 tends to shift from me2 at the promoter to me3 at the 3' end (Li *et al.* 2009, Rahman *et al.* 2011, Wagner and Carpenter 2012, Zhu *et al.* 2016). There is only one human enzyme known to tri-methylate H3K36: SETD2 (Wagner and Carpenter 2012, Michalak and Visvader 2016, Zhu *et al.* 2016). Knockdown of NSD3 or its binding partner BRD4 often indirectly decreases H3K36me3 levels by preventing the di-methylation step upon which SETD2 builds and these two factors have been shown to form a complex together with RNA polymerase II to keep tight regulation on chromatin structure and prevent transcriptional initiation outside the transcriptional start site (Li *et al.* 2009, Fang *et al.* 2010, Rahman *et al.* 2011, Wagner and Carpenter 2012, Jacques-Fricke and Gammill 2014, Katoh 2016). Increased levels of H3K36 methylation are associated with cancers such as AML (Morishita *et al.* 2014, Shen *et al.* 2015). H3K36me3 recruits corepressors, such as HDAC complexes and LSD2, to aid in this regulation (Fang *et al.* 2010, Zhu *et al.* 2016). H3K36 methylation is important early step in activation of transcription by antagonizing repressive marks, such as H3K27me3, and these hPTMs rarely coexist (Yuan *et al.* 2011). Additionally, H3K36 methylation is implicated in alternative splicing events, X-inactivation, and DNA repair and recombination (Wagner and Carpenter 2012). Due to the dynamic nature of chromatin regulation and complexity of the histone code, H3K36me2 has also been implicated in transcriptional repression (Morishita and di Luccio 2011, Wagner and Carpenter 2012). The PWWP domains of the NSD family members are able to read H3K36me0 and H3K36me1 to recruit these enzymes for further methylation by the catalytic SET domain, which all HMT proteins possess, including non-NSD family members such as ASH1L, which shares similarities with NSD1

(Wagner and Carpenter 2012, Liu *et al.* 2014, Katoh 2016). With the tightly coordinated balance associated with H3K36me2 in regulating transcription, it is no surprise that aberrant expression of the enzymes that alter this control have vast implications in disease state such as cancer.

In breast cancer, NSD3 has been identified as one of the top most important HMTs in tumorigenesis. One study analyzed a dataset of 958 breast tumors using TCGA copy number and expression data and GISTIC analysis and determined that, of the 51 known human HMTs, NSD3 has the highest correlation between copy number amplification and coordinate overexpression, thus emphasizing the importance of understanding this oncogene in breast cancer (Liu *et al.* 2014). NSD3 was found amplified in 19% of luminal B breast cancers and was expressed to a higher degree in luminal versus other breast cancer subtypes (Liu *et al.* 2014). Overexpression of NSD3 in the non-transformed MCF10A resulted in a transformed phenotype while knockdown in NSD3-overexpressing breast cancer cell lines reduced proliferation (Yang *et al.* 2010, Liu *et al.* 2014, Irish *et al.* 2016). Knockdown of NSD3 in the amplicon-bearing luminal B breast cancer cell line SUM-44 by our group resulted in decreased expression of ESR1 by microarray, and chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) with an ERα antibody cocktail demonstrated ERα actively bound to chromatin without estrogen in a FOXA1-dependent manner, which was then abrogated by knockdown of NSD3 (Irish *et al.* 2016). Together, these data provided the foundation for further exploration of the role of NSD3 in luminal B breast cancer as described in this work.

Although there has been much progress in elucidating the function of NSD3 and its role in tumorigenesis over the past decade, there remains much to be determined

about this driving oncogene. Due to the highly conserved SET domain and mounting evidence that inhibition of this enzyme would be beneficial in multiple cancer types, several groups have suggested NSD3 and its family members as an excellent model for drug development that would revolutionize the field of epigenetic therapies in cancer (Chen *et al.* 2014, Katoh 2016, Michalak and Visvader 2016, Saloura *et al.* 2016). In addition to the catalytic SET domain, drug targets that would inhibit the PWWP domain common to both NSD3-S and NSD3-L would be highly beneficial for blocking non-enzymatic effects as well (Shen *et al.* 2015). This study seeks to identify the role of NSD3 amplification and overexpression in 8p11-p12 amplicon-bearing breast cancer cell lines and to extend the work previously done in our laboratory linking knockdown of NSD3 to a reduction in ERα expression in this context.

b. ASH2L and MLL complexes

The human ASH2L (<u>A</u>bsent, <u>S</u>mall, or <u>H</u>omeotic disc <u>2</u>-<u>L</u>ike) gene was first discovered by Ikegawa *et al.* in 1999 by large scale genome sequencing and computer-based gene prediction methods as being homologous to the *Drosophila melanogaster* trithorax group (trxG) gene *ash2*. This group of proteins antagonizes transcriptional silencing by the polycomb group (PcG) of epigenetic factors as discussed in previous sections. ASH2L has two isoforms: a 628-amino acid ASH2L1 and a 501-amino acid ASH2L2 protein (Wang *et al.* 2001). ASH2L has a PHD finger motif and thus has been implicated in transcriptional regulation since its discovery (Ikegawa *et al.* 1999). It also has a WH (winged helix) motif for DNA binding, a SPRY domain, and a DPY-30 binding motif, but it lacks the SET domain characteristic of the trxG family and related ASH1L protein (Ikegawa *et al.* 1999, Chen *et al.* 2012). ASH2L is implicated in tight regulation of H3K4 methylation, which is associated with activated transcription (Dou *et al.* 2006,

Guccione *et al.* 2007, Luscher-Firzlaff *et al.* 2008, Wan *et al.* 2013, Judes *et al.* 2016). Specifically, ASH2L knockdown primarily affects H3K4me3 levels as opposed to mono- or di-methylation (Steward *et al.* 2006, Demers *et al.* 2007, Fossati *et al.* 2011).

H3K4 tri-methylation was originally thought to be irreversible, barring histone protein turnover, and its regulation in cells of utmost importance to transcriptional regulation in differentiation and development (Steward *et al.* 2006). While the latter is still true, H3K4me3 is now known to be de-methylated by LSD1 (Blobel *et al.* 2009, Fang *et al.* 2010, Huang *et al.* 2011, Hervouet *et al.* 2013, Kim *et al.* 2014, Abdel-Hafiz and Horwitz 2015). The primary H3K4 HMT family in mammalian cells is the SET1 family, which includes MLL proteins 1-4, SET1A, and SET1B, and is named for their highly conserved SET domain that is also found in other HMT groups, such as the NSD family as previously described (Dou *et al.* 2006, Steward *et al.* 2006, Tan *et al.* 2009, Cao *et al.* 2010, Wu *et al.* 2010, Xiao *et al.* 2011, Chen *et al.* 2012, Kim *et al.* 2014, Qi *et al.* 2014, Katoh 2016). MLL proteins actually have very weak enzymatic activity and require binding partners, one of which is ASH2L (Southall *et al.* 2009, Chen *et al.* 2012). In contrast to H3K4me1 and H3K4me2, H3K4me3 is strictly associated with active gene transcription and its distribution closely overlaps that of elongating RNA polymerase II (Demers *et al.* 2007, Guccione *et al.* 2007, Luscher-Firzlaff *et al.* 2008). H3K4me3 is often localized to the 5' (proximal promoter) end of target genes and ASH2L is also confined to this region, though MLL may spread further along the gene (Demers *et al.* 2007, Luscher-Firzlaff *et al.* 2008). In contrast, H3K4me2 has a greater distribution, suggesting a major role for ASH2L as the primary activator of H3K4 tri-methylation in promoter regions (Guccione *et al.* 2007).

Since ASH2L does not possess a SET domain like other trxG proteins, it is known to form complexes with the MLL family of SET domain proteins and is a component of all H3K4 HMTs characterized to date (Ikegawa *et al.* 1999, Butler *et al.* 2011, Ernst and Vakoc 2012, Wan *et al.* 2013, Qi *et al.* 2014, Butler *et al.* 2017). MLL proteins are preferentially H3K4 mono-methylases with weak enzymatic activity and ASH2L is critical for enhancement of HMT function and regulation of MLL-mediated catalysis of H3K4me3 (Patel *et al.* 2009, Chen *et al.* 2012). RbBP5, another MLL complex member, is implicated in activating MLL-mediated catalysis of the intermediate H3K4me2 state (Dou *et al.* 2006, Ernst and Vakoc 2012). The third critical MLL complex participant is WDR5, which is important for the stability of the protein-protein interactions and recruitment to chromatin (Dou *et al.* 2006, Steward *et al.* 2006). Menin, encoded by the MEN1 gene, is also often found together with the WDR5-ASH2L-RbBP5 (WAR) subunit and has demonstrated properties of tumor suppression as well as the ability to directly bind RNA polymerase II (Guertin *et al.* 2006, Guccione *et al.* 2007, Luscher-Firzlaff *et al.* 2008). DPY-30 is also commonly associated with WAR and ASH2L has a DPY-30 binding domain on its C-terminal end (Patel *et al.* 2011, Chen *et al.* 2012, Ali and Tyagi 2017). The sub-complex containing all 4 subunits, known as WRAD, can exist independent of MLL, with and without additional binding partners, implying that these genes may coordinate to recruit other chromatin-modifying enzymes or may have actions outside of hPTM (Patel *et al.* 2011, Ernst and Vakoc 2012, Ali *et al.* 2014, Ali and Tyagi 2017). Interestingly, although knockdown of ASH2L usually reduces global levels of H3K4me3, MLL-dependent H3K4 mono- and di-methylation at some genes is maintained, suggesting it may still interact with the chromatin and act as an HMT, but cannot perform tri-methylation without ASH2L (Guertin *et al.* 2006, Tan *et al.* 2009,

Butler *et al.* 2011, Fossati *et al.* 2011, Chen *et al.* 2012). Knockdown of MLL, however, does not alter methylated H3K4 levels, implying that ASH2L may be responsible for H3K4me3 catalysis by other HMT enzymes as well, perhaps by interaction with other SET domain-containing enzymes, and emphasizing the essential role that ASH2L plays in this histone methyl mark (Steward *et al.* 2006, Demers *et al.* 2007, Blobel *et al.* 2009, Butler *et al.* 2011, Wan *et al.* 2013).

The requirement of the non-SET binding partners ASH2L, RbBP5, and WDR5 for HMT activity at H3K4me3 is unique (Patel *et al.* 2011). As such, several studies have attempted to understand the biological role of the WRAD sub-complex without a SET domain-containing component (Steward *et al.* 2006, Patel *et al.* 2009, Ernst and Vakoc 2012, Ali and Tyagi 2017). One study demonstrated that the sub-complex interacts with nuclear hormone transcription factor coregulatory proteins NRC and NIF-1 and maintained affinity for H3 in this context (Garapaty *et al.* 2009), but the role of ASH2L in promoting HMT activity outside SET domain-containing proteins remained undefined. Patel *et al.* (2009) suggested that the sub-complex itself may possess H3K4-specific methylation ability when MLL is not present, although this activity is not through a SET domain and the mechanism is unknown. This finding was confirmed by their 2011 description of this phenomenon as well (Patel *et al.* 2011). Later, Cao *et al.* (2010) demonstrated that HMT activity is intrinsic to the ASH2L-RbBP5 heterodimer and requires the highly conserved SPRY domain of ASH2L. This group also suggested that ASH2L and the SET domain of MLL1 form a joint catalytic subunit for promoting H3K4 methylation and that overall activity of the MLL complex is dependent upon binding of the methyl-donor SAM to ASH2L (Cao *et al.* 2010). The WRAD sub-complex HMT activity has also been suggested as a second active site in the MLL complex, one

contributing factor to enhanced H3K4 methylation when the entire complex is assembled (Ernst and Vakoc 2012, Patel *et al.* 2014). These findings provided the foundation for the "two-step" hypothesis of MLL complex activation whereby a minimal association between ASH2L and RbBP5 is required to provide a second catalytic pocket in the three-dimensional structure of the complex, thus enhancing the overall methylation ability of the MLL family of proteins (Ernst and Vakoc 2012, Patel *et al.* 2014, Li *et al.* 2016, Ali and Tyagi 2017). This hypothesis explains the ability of ASH2L to act as an HMT even in the absence of a SET domain.

MLL complexes are known to interact with a plethora of proteins, including other epigenetic factors, transcription factors, and basal transcription machinery (Dou *et al.* 2006, Demers *et al.* 2007). ASH2L has similarly been shown to directly bind transcription factors such as AP2δ, as well as the basal transcriptional machinery (Tan *et al.* 2008). This study suggested that the critical role of ASH2L is as a mediator between chromatin-modifying enzymes and factors involved in transcription. Yates *et al.* (2010) demonstrated binding of the WAR sub-complex, and specifically ASH2L, to CHD8, an APT-dependent chromatin remodeler that alters H3 methylation patterns of HOX genes. Additionally, MLL complexes have been shown by several groups to interact directly with ERα, modifying the histone code near ERα-target genes to promote transcription via promotion of catalysis of H3K4me3 by ASH2L (Hervouet *et al.* 2013, Zhang *et al.* 2013, Bhan *et al.* 2014, Qi *et al.* 2014). ASH2L and the complexes in which it participates have major effects on gene expression during embryogenesis and beyond (Dou *et al.* 2006, Butler *et al.* 2011). Tight regulation of ASH2L is therefore required but poorly understood. One study identified arginine residues on ASH2L that can be di-methylated by PRMT1, but the biological effect of this modification is unknown (Butler *et al.* 2011).

Similarly, MLL1 can auto-methylate ASH2L in the MLL-WRAD complex and, although the function of this methylation is unknown, one possibility is that it acts as a mechanism of autoregulation (Patel *et al.* 2014). WDR5 binds acetylated H3 tails with high affinity, preventing the HMT activity of ASH2L-containing MLL complexes (Guccione *et al.* 2007, Avdic *et al.* 2011). Of course, the actions of PcG proteins are well-known to antagonize H3K4 methylation by these trxG enzymes as well (Grzenda *et al.* 2011). The balance of other epigenetic marks, transcription factors, and protein-protein interactions could also play a role in keeping ASH2L and associated complexes in check, but these events currently remain unclear and are understudied to date (Dou *et al.* 2006, Butler *et al.* 2011, Vedadi *et al.* 2017).

Clearly, ASH2L plays a major role in regulation of gene expression. It is especially implicated in transcription of genes associated with development, stem cell divisions, and mitotic regulation (Ikegawa *et al.* 1999, Demers *et al.* 2007, Butler *et al.* 2011, Kawabe *et al.* 2012, Ali *et al.* 2014). Knockdown of ASH2L results in decreased expression of genes associated with cell cycle, proliferation, and survival in a range of tissue types, including hematologic cells, breast tumors, skeletal muscle, osteosarcomas, and gliomas (Wang *et al.* 2001, Rampalli *et al.* 2007, Luscher-Firzlaff *et al.* 2008, Pullirsch *et al.* 2010, Schram *et al.* 2013, Ali *et al.* 2014, Erfani *et al.* 2015, Zhu *et al.* 2016). Knockdown of ASH2L in embryonic stem cells results in reduced pluripotency due to globally decreased H3K4me3 and increased H3K9me3 levels, a mark of silenced chromatin (Wan *et al.* 2013). Similarly, MLL and the WRAD sub-complex is linked to cell cycle progression through S and M phases (Ali *et al.* 2014). Given the link to stem cell potential and cell cycle regulation, dysregulation of ASH2L can be linked to tumorigenesis.

Similar to NSD3, much of the work on ASH2L in cancer has been done in hematologic models. It was first identified as an oncogene by Lüscher-Firzlaff *et al.* (2008) where they demonstrated that ASH2L cooperates with MYC and RAS, is overexpressed in human tumor samples, and results in decreased tumor growth upon knockdown. The interaction of ASH2L with MYC was later confirmed and implicated in regulation of bivalent chromatin with high levels of both activating marks H3K4me3 and H3K27ac (Ullius *et al.* 2014). ASH2L is important for differentiation of erythroid cell lineages and its aberrant expression is implicated in hematologic malignancies (Wang *et al.* 2001, Cao *et al.* 2010). MLL is often involved in translocations or is mutated or deleted in these cancers (Luscher-Firzlaff *et al.* 2008, Southall *et al.* 2009, Ali *et al.* 2014), thus raising questions about ASH2L function outside of MLL complexes in this context. Additionally, ASH2L has been described as transforming in some osteosarcomas and recently was implicated in regulation of EGFR expression via H3K4me3 in gliomas (Schram *et al.* 2013, Erfani *et al.* 2015). Interestingly, ASH2L demonstrated tumor suppressor characteristics in cell lines from several cancer types with enriched H3K4me3 at p53 pro-apoptotic gene promoters (Mungamuri *et al.* 2015). The authors implicated ASH2L in promoting stability of the initiation complex, though the specific role of ASH2L in this setting was not clearly defined. A dataset of 511 AML patient samples analyzed by reverse phase protein array (RPPA) revealed that ASH2L expression is inversely correlated with patient survival and expression of cell adhesion and cell cycle inhibition genes (Butler *et al.* 2017).

ASH2L is not well-characterized in breast cancer. Along with other 8p11-p12 amplicon genes, ASH2L was identified as one of the core luminal B genes associated with cell cycle and proliferation (Cornen *et al.* 2014). Multiple analyses attempting to

identify the driver oncogene(s) from the 8p11-p12 amplicon called out the potential of ASH2L as such (Garcia *et al.* 2005, Yang *et al.* 2006, Kwek *et al.* 2009, Yang *et al.* 2010, Cornen *et al.* 2014, Turner-Ivey *et al.* 2014), but investigation into its biological function has not yet been fully explored in a luminal B breast cancer model. One group demonstrated recruitment of ASH2L to the promoter of ESR1 by GATA3, where ASH2L potentiated the effects of the GATA3 transcription factor to induce ERα overexpression in breast cancer cell lines and primary tumors (Qi *et al.* 2014). Since ASH2L is one of the most coordinately overexpressed 8p11-p12 amplicon genes when it is amplified (Sircoulomb *et al.* 2011), and given the link between this amplicon and ERα as reported by our work on NSD3 (see Chapter 2) and its prevalence in the ER+ luminal subtype of breast cancer, understanding the link between ASH2L, ERα, and other amplicon genes such as NSD3 could provide exciting new insight into an understudied oncogene and lead to new therapeutic opportunities in this patient subset. This study seeks to explore the role of ASH2L in epigenetic regulation of gene expression in luminal B breast cancer cells with the goal of understanding another step in the mechanism of breast tumorigenesis.

1.5. Significance

ER+ breast cancer, once thought to be a relatively straightforward disease dependent upon estrogen, has since proven to be an incredibly heterogeneous set of diseases with varying degrees of ERα expression and activity that makes reliable prediction of response to endocrine therapy nearly impossible. The call to discover oncogenic signatures that will inform individualized therapy for breast cancer patients is urgent. At the cross-roads of cancer genomics and epigenomics is amplification-induced

overexpression of chromatin-remodeling enzymes. The 8p11-p12 amplicon harbors two histone methyltransferases, NSD3 and ASH2L, which are capable of exerting a vast array of aberrations in the expression profiles of the tumors in which they are overexpressed, and therefore are the focus of the remainder of this work. HMTs have been cited as promising candidates for suppression in the treatment of various cancers (Chen *et al.* 2014, Liu *et al.* 2014, Michalak and Visvader 2016), and understanding how these enzymes function in the amplified and overexpressed state often observed in breast cancer will facilitate drug development in this field. Indeed, recent studies utilizing large-scale databases have identified several HMTs as compelling targets for drug development in cancer therapeutics and suggested that the NSD family of HMTs will provide the optimal starting place for structure-based design (Chen *et al.* 2014, Bennett *et al.* 2017). These studies have called for an improved understanding of chromatin-modifying enzymes as a whole and NSD3 and ASH2L and the complexes in which they function specifically. These oncogenic enzymes have the potential to serve both as biomarkers for therapy response and as novel therapeutic targets themselves and represent an exciting new arm of investigation in breast cancer research and treatment.

CHAPTER 2: NSD3 amplification and overexpression results in overexpression and estrogen-independent activation of the estrogen receptor in human breast cancer.

2.1. NSD3 is overexpressed in 8p11-p12 amplicon-bearing cell lines and is a verified oncogene.

Nuclear Receptor Binding SET Domain protein 3 (NSD3; formerly known as WHSC1L1) was verified as an oncogene by our laboratory (Yang *et al.* 2006, Yang *et al.* 2010, Irish *et al.* 2016). We surveyed a panel of normal breast, breast cancer, and lung squamous cell carcinoma cell lines for expression of NSD3 and discovered that NSD3 is overexpressed at both the message and protein levels in amplicon-bearing cell lines compared to control and that expression is highest in the luminal B breast cancer cell line SUM-44 (Figure 2.1-A). Additionally, we noted that the short isoform of NSD3 (NSD3-S) is expressed to a higher degree than the long isoform (NSD3-L), a finding consistent with data from our lab and others that have demonstrated a role for NSD3-S as the primary transforming isoform of this oncogene (Yang *et al.* 2010, Shen *et al.* 2015). These data correlate with the Cancer Genome Atlas (TCGA), which also showed that, for a group of 964 breast cancer tumors, NSD3-S was overexpressed compared to NSD3-L in the context of NSD3 amplification (Figure 2.1-B) (Gao *et al.* 2013).

Since the effects of the two isoforms of NSD3 on cell proliferation had not been previously differentiated, we performed growth assays following knockdown of NSD3-S or total (NSD3-T) using shRNA that targeted either a unique sequence in the 3'UTR of the NSD3-S transcript or a sequence common to the coding region of both isoforms, respectively. Knockdown with either construct did not affect cell proliferation compared to LacZ control in the amplicon-null MCF7 breast cancer cell line or amplicon-bearing DMS-114 lung squamous cell cancer (LSCC) cell line (Figure 2.1-C and Figure 2.1-D). Knockdown of NSD3 with both constructs in the amplicon-bearing SUM-44 breast
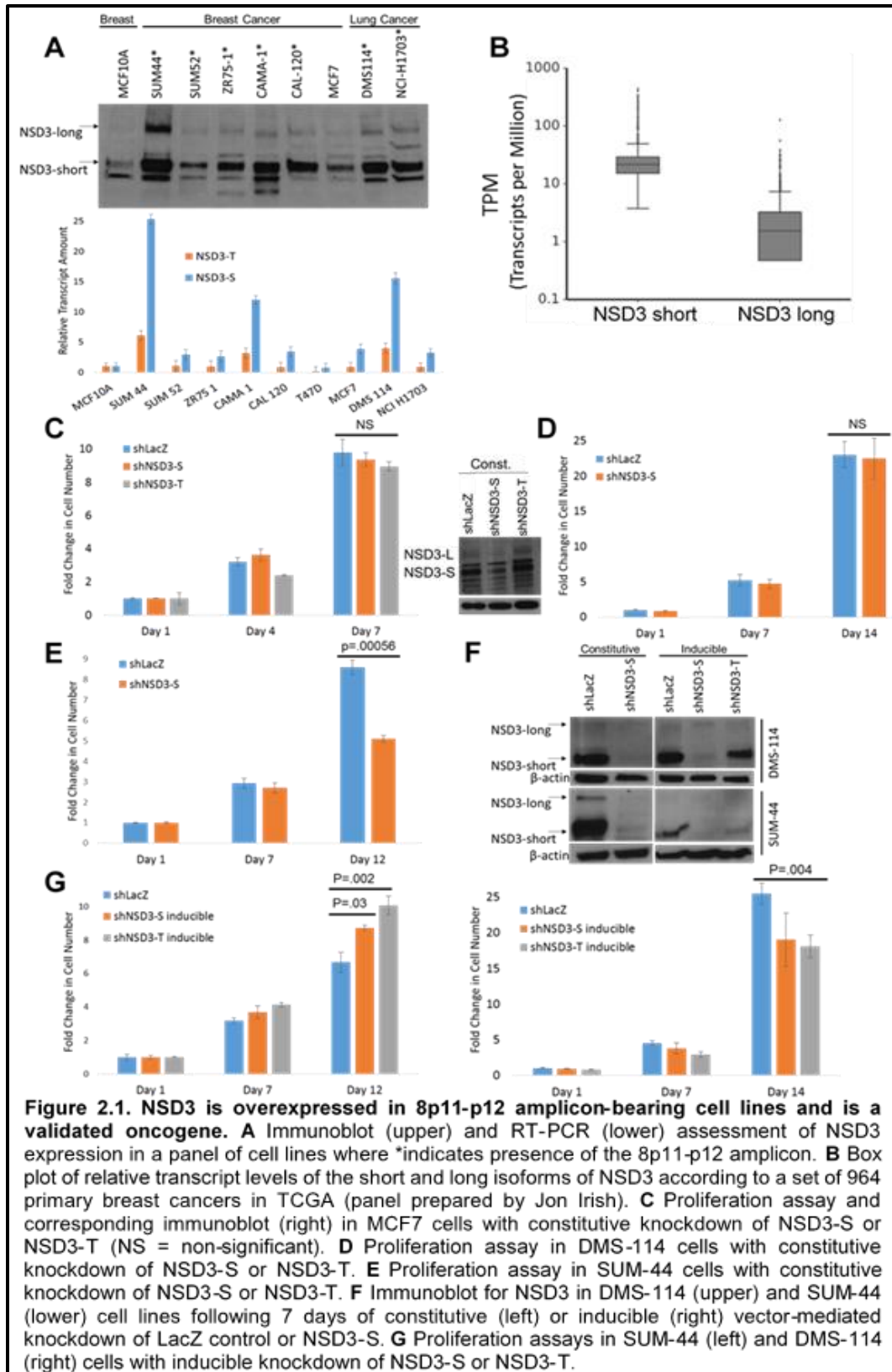
Figure 2.1. NSD3 is overexpressed in 8p11-p12 amplicon-bearing cell lines and is a validated oncogene. **A** Immunoblot (upper) and RT-PCR (lower) assessment of NSD3 expression in a panel of cell lines where *indicates presence of the 8p11-p12 amplicon. **B** Box plot of relative transcript levels of the short and long isoforms of NSD3 according to a set of 964 primary breast cancers in TCGA (panel prepared by Jon Irish). **C** Proliferation assay and corresponding immunoblot (right) in MCF7 cells with constitutive knockdown of NSD3-S or NSD3-T (NS = non-significant). **D** Proliferation assay in DMS-114 cells with constitutive knockdown of NSD3-S or NSD3-T. **E** Proliferation assay in SUM-44 cells with constitutive knockdown of NSD3-S or NSD3-T. **F** Immunoblot for NSD3 in DMS-114 (upper) and SUM-44 (lower) cell lines following 7 days of constitutive (left) or inducible (right) vector-mediated knockdown of LacZ control or NSD3-S. **G** Proliferation assays in SUM-44 (left) and DMS-114 (right) cells with inducible knockdown of NSD3-S or NSD3-T.

cancer cell line, however, did reduce cell proliferation (Figure 2.1-E). Interestingly, the inhibitory effect on growth was greater with knockdown of NSD3-S compared to NSD3-T (Irish *et al.* 2016), again underscoring the oncogenic role of the short isoform in particular. These proliferation assays were repeated using IPTG-inducible shRNA constructs and, although these constructs were able to achieve similar levels of NSD3 knockdown as the constitutive vectors in SUM-44 and DMS-114 cells (Figure 2.1-F), the effects on cell growth were not recapitulated in either cell line (Figure 2.1-G). This and other data (Figure 2.2-A and Figure 2.2-B) prompted our lab to discontinue the use of the IPTG-inducible shRNA system.

To investigate the role of this oncogene further, our lab generated a transgenic FVB mouse model with targeted overexpression of NSD3 to the mammary epithelium (Turner-Ivey *et al.* 2017). NSD3 transgenic females exhibited morphological abnormalities during mammary gland development that resulted in a lactation defect due to the inability of alveoli to undergo full functional differentiation. By 40 weeks of age, these NSD3 transgenic females also developed mammary gland dysplasias, hyperplasias, carcinomas in situ, and invasive carcinomas. Together, these data validate previous work demonstrating the transforming ability of NSD3 and confirm its role as an important driving oncogene in human breast cancer.

2.2. NSD3 knockdown results in reduced expression of estrogen receptor alpha (ERα).

To investigate further the role of NSD3 in breast cancer, we performed microarray analysis following NSD3-S knockdown in SUM-44 cells and noted decreased expression of multiple important genes, including ESR1, which encodes ERα (Table 2.1). To confirm and explore this relationship, SUM-44 cells were infected with LacZ

Table 2.1. Genes with reduced transcript expression following knockdown of NSD3-S in SUM-44 cells. Changes in gene expression by microarray analysis of select genes following knockdown of NSD3-S in SUM-44 cells. Data collected by Jon Irish.

Gene Symbol	Fold Change	P-value (corrected)
MYCN	-7.43	.03
ESR1	-7.33	.004
CXCR4	-4.88	.007
ID2	-4.30	.004
RET	-3.70	.008
MYB	-2.92	.016
CD24	-2.62	.010
ALDH2	-2.55	.002
ERBB3	-2.50	.009
ERBB4	-2.12	.002

control or IPTG-inducible NSD3-S or NSD3-T shRNA constructs and harvested at 12 hour intervals following initiation of IPTG-treatment to assess NSD3 and ERα levels by western blotting and ESR1 levels by RT-PCR. This system demonstrated very poor knockdown of either NSD3 isoform and relatively little change in ERα or ESR1 levels (Figure 2.2-A). A second attempt at this experiment demonstrated similar results (Figure 2.2-B). Coupled with the inconsistent results observed on growth patterns using this inducible shRNA system (Figure 2.1-F and 2.1-G), we concluded this system was unreliable and switched to the constitutive shRNA system exclusively. Knockdown of NSD3-S and NSD3-T reduced ERα levels compared to control beginning at 72 hours post-infection (Figure 2.2-C). ERα expression consistently reached peak knockdown at 7 days post-infection with shNSD3-S or NSD3-T virus (Figure 2.2-D). This elucidation of the timing of NSD3-induced reduction in ERα levels was essential to begin to understand the mechanism and implications of this relationship.

Figure 2.2. NSD3 knockdown results in knockdown of estrogen receptor alpha (ERα). A Immunoblot depicting NSD3 and ERα expression in SUM-44 cells with inducible knockdown of NSD3-S or NSD3-T untreated or treated with IPTG for 12, 24, 36, 48, or 72 hours. Corresponding RT-PCR for ESR1 appears in the lower left. **B** Immunoblot depicting NSD3 and ERα expression in SUM-44 cells with inducible knockdown of NSD3-S or NSD3-T treated with IPTG for 24, 48, or 72 hours. **C** Immunoblot depicting NSD3 and ERα expression in SUM-44 cells with constitutive knockdown of NSD3-S or NSD3-T 48, 72, or 96 hours after infection with lentiviral shRNA constructs. **D** Immunoblot depicting NSD3 and ERα expression in SUM-44 cells with constitutive knockdown of NSD3-S or NSD3-T 7 days after infection with lentiviral shRNA constructs.

2.3. NSD3 overexpression results in overexpression of estrogen receptor alpha (ERα).

To explore further the effect of NSD3 amplification and overexpression on ERα

expression, we assessed ERα levels by immunoblot in a panel of breast and LSCC cell

lines and determined that that SUM-44 cells express extremely high levels of ERα protein (Figure 2.3-A, upper). This finding was especially interesting considering that MCF7 cells are held as the benchmark ER+ cell line for ERα expression *in vitro* and ZR-75-1, CAMA-1, and T47D are also described as ER+ and widely used to study ERα biology in breast cancer cell lines. SUM-44 cells have markedly increased ERα expression compared to all of these lines. RT-PCR for ESR1 expression in this panel of cell lines revealed that message level expression followed the same pattern as protein expression (Figure 2.3-A, lower).

To assess the biological relevance of the dramatic overexpression of ERα in SUM-44 cells in relation to ERα levels observed in patients, we performed RT-PCR to determine message level expression of NSD3 and ESR1 in a group of patient tumor specimens compared to the panel of cell lines (Figure 2.3-B). While we noted 7 cases of high ESR1 expression without corresponding overexpression of NSD3, indicating that ESR1 can be overexpressed without NSD3, we discovered that every case of NSD3 overexpression also displayed high levels of ESR1. TCGA also demonstrates this link between NSD3 and ESR1 overexpression (Figure 2.3-C). We obtained a pleural effusion sample from a breast cancer patient at MUSC with metastatic ER+ breast cancer resistant to endocrine therapy and performed immunoblots for NSD3 and ERα (Figure 2.3-D). While we were unable to isolate a homogeneous population of cells from this specimen, even this heterogeneous sample demonstrated NSD3 expression greater than MCF7 and CAMA-1 cells and ERα expression comparable to the ER+ CAMA-1 line. From these data we concluded that ESR1 can be overexpressed without NSD3 overexpression, suggesting that there are multiple mechanisms for ERα upregulation in

human breast cancer, but not vice versa, demonstrating an important link between the NSD3 oncogene and the transcription factor ERα. Importantly, we found that only the SUM-44 cell line recapitulated the expression levels of these two genes in patient tumors (Figure 2.3-B). SUM-44 is the only *in vitro* model that represents this important subgroup of breast cancer patients and serves as a means by which to study the biology of tumors with NSD3-induced overexpression of ERα.

Since we previously demonstrated that knockdown of NSD3-S reduced ERα expression in SUM-44 cells (Figure 2.2), we next tested whether overexpression of NSD3-S in amplicon-null MCF7 (ER+) cells or amplicon-bearing CAMA-1 (ER+) and SUM-52 (ER-negative) cells would induce overexpression of ERα as detected by immunoblot. NSD3-S overexpression did not alter ERα levels in MCF7 cells but did increase ERα protein levels in both of the amplicon-bearing cell lines (Figure 2.3-E, upper). Message level expression of ERα downstream target genes in CAMA-1 cells also increased, suggesting that the upregulated ERα by NSD3-S overexpression is transcriptionally active (Figure 2.3-E, lower). This suggests that NSD3 is necessary but not sufficient to overexpress ERα and requires one or more additional factors present on the 8p11-p12 amplicon. Since NSD3 is an epigenetic modifier of chromatin, we assessed a panel of cell lines for expression of the two other chromatin modifiers present in the 8p11-p12 genomic region: KAT6A and ASH2L. RT-PCR revealed that, of the breast cancer cell lines, CAMA-1, SUM-52, and SUM-44 have the highest levels of KAT6A (Figure 2.3-F) and SUM-44 expresses the highest ASH2L protein levels, followed by SUM-52, CAMA-1, and MCF7 (Figure 3.1-A). Together, these data led to our hypothesis that these three 8p11-p12 amplicon oncogenes, ASH2L, KAT6A, and NSD3,

Figure 2.3. NSD3 overexpression results in overexpression of estrogen receptor alpha (ERα). A Immunoblot for ERα (upper) and RT-PCR for ESR1 (lower) in a panel of cell lines. B RT-PCR for NSD3 and ESR1 transcript expression in a selected group of patient tumor samples and cell lines. C Box plot demonstrating ESR1 mRNA expression as a function of NSD3 mRNA expression in a set of 964 primary breast cancers in TCGA (prepared by Jon Irish). D Immunoblot for NSD3 and ERα expression in cell lines and cells isolated from pleural effusion fluid from an MUSC breast cancer patient with ER+ disease resistant to endocrine therapy. E Immunoblot (upper panel) showing NSD3 and ERα expression in amplicon-null MCF7 cells and amplicon-bearing CAMA-1 and SUM-52 cells following infection with an NSD3-S or LacZ overexpression vector (data collected by Brittany Ivey) and RT-PCR for ERα-target genes following NSD3-S overexpression compared to LacZ control in CAMA-1 cells (lower panel) where relative transcript level expression is plotted for the ERα-target genes listed on the x-axis. F RT-PCR for KAT6A transcript expression in a panel of cell lines.
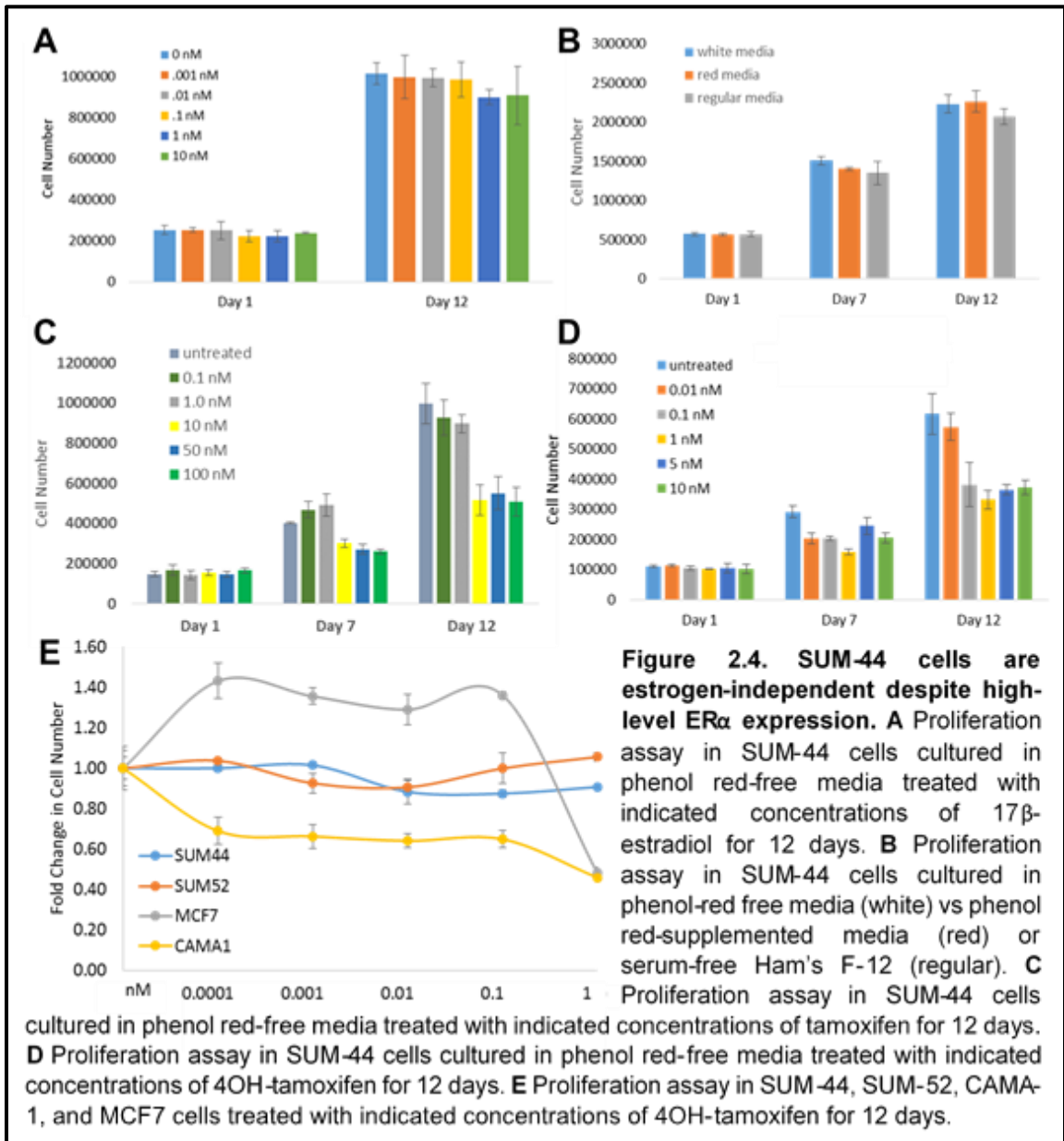
work cooperatively to promote tumorigenesis in luminal B breast cancer. This work discovered a clear link between NSD3 amplification and overexpression and overexpression of ERα in human breast cancer and identified SUM-44 as a unique model by which this patient group can be studied further.

2.4. SUM-44 cells are estrogen-independent despite high-level ERα expression.

Since SUM-44 cells have extremely high levels of ERα expression, we hypothesized that this line would be highly responsive to its mitogenic ligand, 17β-estradiol, and highly sensitive to the SERM tamoxifen as this drug is used clinically to treat ER+ breast cancers. To test this hypothesis, we measured the proliferation of SUM-44 cells grown in phenol red-free media and treated with increasing doses of estrogen over 12 days and observed no difference in growth across these concentrations (Figure 2.4-A). Since SUM-44 cells are routinely grown in serum free, low phenol red media, we tested the effect of phenol red, often cited as having a stimulatory effect on ERα (Wesierska-Gadek *et al.* 2006), in this cell line using "white" phenol red-free, "red" phenol red-supplemented, and "regular" low phenol red Ham's F-12 media. We observed no difference in cell growth due to the phenol red content of the media (Figure 2.4-B).

We next tested the effect of the SERM Tamoxifen on SUM-44 cells and, while higher concentrations of the drug did inhibit the growth of these cells, they were still able to proliferate and reached a plateau where increasing the amount of drug did not have a corresponding increase in response (Figure 2.4-C). We repeated this assay, performed in phenol red-free media, using 4OH-tamoxifen, the active metabolite of tamoxifen, and again observed a plateau in the proliferation response of this cell line (Figure 2.4-D). To

assess how this observation compared to other cell lines, we treated the 8p11-p12

amplicon-bearing lines SUM-52, SUM-44, and CAMA-1 as well as the amplicon-null

MCF7 line with 4OH-tamoxifen and observed that SUM-44 cells had little response up to

the 1 nM treatment dose, mirroring the ER-negative SUM-52 line rather than the

tamoxifen-sensitive ER+ CAMA-1 line (Figure 2.4-E). Interestingly, 4OH-tamoxifen

Figure 2.4. SUM-44 cells are estrogen-independent despite high-level ERα expression. A Proliferation assay in SUM-44 cells cultured in phenol red-free media treated with indicated concentrations of 17β-estradiol for 12 days. B Proliferation assay in SUM-44 cells cultured in phenol-red free media (white) vs phenol red-supplemented media (red) or serum-free Ham's F-12 (regular). C Proliferation assay in SUM-44 cells cultured in phenol red-free media treated with indicated concentrations of tamoxifen for 12 days. D Proliferation assay in SUM-44 cells cultured in phenol red-free media treated with indicated concentrations of 4OH-tamoxifen for 12 days. E Proliferation assay in SUM-44, SUM-52, CAMA-1, and MCF7 cells treated with indicated concentrations of 4OH-tamoxifen for 12 days.

elicited an agonistic effect on the ER+ MCF7 cell line before achieving antagonistic

concentrations (Figure 2.4-E). These results are an important step toward understanding

patients with highly ER+ tumors that do not respond to traditional endocrine therapies.

In addition to these findings, ERα chromatin immunoprecipitation and high-

throughput sequencing (ChIP-seq) was performed in our laboratory (Irish *et al.* 2016).

These data demonstrated that ERα maintained the ability to bind to chromatin without

estrogen in SUM-44 cells. This work also showed that treatment with estrogen increased

the number of peaks per gene and size of those peaks, indicating that these cells are

capable of responding to estrogen, but no significant alterations of the interaction of ERα

with the chromatin were observed in this context. Knockdown of NSD3-S abrogated the

ability of ERα to bind to chromatin without estrogen, suggesting that the estrogen-

independent activity of ERα is dependent upon NSD3. Based on these observations, we

concluded that, despite high-level ERα expression, the SUM-44 cell line is independent

of estrogen signaling for growth and survival.

2.5. SUM-44 cells are ERα-dependent.

One explanation for the estrogen-independent state of ERα in this model is that

the receptor is no longer functional. Indeed, loss of ERα or its function is a known

mechanism of endocrine resistance in ER+ human breast tumors (Kerdivel *et al.* 2013).

Since we clearly demonstrate here that SUM-44 cells maintain ERα expression and

previous characterization of this cell line detected wild-type ERα (Ethier 1996), we next

sought to determine whether the receptor was functional in SUM-44 cells. Previously

performed ChIP-seq experiments demonstrated that ERα is indeed capable of binding to

chromatin without estrogen and ChIP-PCR validated enrichment of several important

ERα-target genes (Figure 2.5-A). Enrichment was enhanced by treatment with estrogen, again underscoring that this receptor maintains its normal structure and function. In MCF7 cells, there was no enrichment of target genes over negative control until estrogen was added. Interestingly, enrichment of the gene encoding the progesterone receptor (PR), an ERα pioneer factor associated with good prognosis genes that is lost in SUM-44 cells, was dramatically increased with estrogen treatment in MCF7 cells but not in SUM-44, further confirming that ERα is active in these cells but is behaving differently than in MCF7 cells (Mohammed *et al.* 2015). Indeed, ERα ChIP-seq in SUM-44 cells identified FOXA1 binding motifs in addition to estrogen response elements (ERE) (Irish *et al.* 2016), consistent with reports that alternative pioneer factors are involved in ERα-mediated gene transcription upon loss of PR (Ross-Innes *et al.* 2012).

To test the ability of ERα to alter the expression of target genes, we performed RT-PCR following siRNA-mediated knockdown of ESR1 or FOXA1. ERα target gene expression decreased with knockdown of both ESR1 and FOXA1 (Figure 2.5-C). When ESR1 was knocked down using three different shRNA constructs, MCF7 cells demonstrated reduced proliferation while SUM-44 cells were unable to survive or proliferate (Figure 2.5-D), suggesting that SUM-44 cells are highly dependent on ERα for growth and survival. Array analysis following knockdown of either NSD3-S or ESR1 revealed that the top-scoring biological processes of the genes affected by knockdown of both ESR1 and NSD3-S are related to cell cycle regulation and DNA replication (Table 2.2; Appendix A). Together, these data demonstrate that SUM-44 cells are highly dependent on ERα signaling for growth, survival, and gene transcription. We conclude

that ERα is active in an estrogen-independent manner in the 8p11-p12-amplified SUM-44 cell line.

Since the SUM-44 cell line represents the group of patients who respond poorly to AIs and SERMs yet maintain high-level ERα expression, we tested the sensitivity of this cell line to the SERD Fulvestrant. At increasing doses of drug, we observed corresponding decreasing levels of ERα expression and proliferation of the cells (Figure 2.5-E). When this response was compared to other cell lines, we noted that SUM-44 cells are not as sensitive as the other ER+ MCF7 and CAMA-1 cell lines (Figure 2.5-F). Since the mechanism of action of the SERD compounds is dose-dependent and directly proportional to the ability of the drug to degrade the receptor, it is expected that a cell line such as SUM-44, which has the highest ERα expression of the cell lines tested (Figure 2.3-A), would appear to be resistant to fulvestrant due to the increased requirement to achieve effective dose.

Table 2.2. Biological processes to which genes with altered expression following knockdown of NSD3-S and ESR1 were annotated. Top scoring biological processes and corresponding p-values in common between NSD3-S and ESR1 knockdown microarrays in SUM-44 cells.

| Biological Process	P-value
Mitotic cell cycle	1.00E-24
Cell Cycle	1.00E-24
Mitotic cell cycle process	1.00E-24
DNA metabolic process	1.00E-24
Cell cycle process	1.00E-24
Mitotic nuclear division	1.60E-22
Organelle fission	9.30E-22
Nuclear Division	1.00E-21
Organelle organization	1.20E-21
DNA replication	4.40E-21
Cell division	4.70E-21
Mitotic cell cycle phase transition	1.50E-20
Cell cycle phase transition	2.80E-20
DNA strand elongation	8.80E-19
DNA strand elongation involved in DNA replication	8.90E-19

This is an important finding for breast cancer patients who fall into this subtype of cancer and provides foundational evidence for the 8p11-p12 amplicon as a marker for those patients who would benefit from addition of a SERD therapy to their AI/SERM regimen. These findings also underscore the need for development of additional SERD compounds that can achieve adequate degradation of ERα without subjecting patients to intolerable adverse effects.

2.6. Summary

Together, these data demonstrate that amplification and subsequent overexpression of the NSD3 oncogene results in overexpression of the transcription factor ERα. Under these circumstances, the receptor is activated in an estrogen-independent manner, rendering traditionally-used endocrine therapies, such as Tamoxifen, ineffective. These cells do, however, respond to the SERD Fulvestrant, but require high-dose treatment for adequate degradation of ERα. These data provide rationale for continued SERD development and offer a foundation for implementation of these types of therapies in patients using the 8p11-p12 amplicon and NSD3 overexpression as a biomarker. Further investigation is required to elucidate the mechanism of NSD3-induced overexpression and activation of ERα, which could provide additional therapeutic strategies to combat endocrine-resistant breast cancer.

Figure 2.5. SUM-44 cells are ERα-dependent. A ERα ChIP-PCR in SUM-44 cells after treatment with 17β-estradiol for 6 hours. **B** ERα ChIP-PCR in MCF7 cells after treatment with 17β-estradiol for 6 hours. **C** Immunoblot (inset) for protein expression and RT-PCR for transcript expression of selected ERα-target genes following siRNA-mediated knockdown of ESR1 (left) or FOXA1 (right) in SUM-44 cells. **D** Immunoblot for ERα expression (upper) and corresponding cell photographs (lower) following shRNA-mediated knockdown of LacZ control or ESR1 by three different shRNA constructs in MCF7 (left) and SUM-44 (right) cells. **E** Proliferation assay in SUM-44 cells treated with indicated concentrations of fulvestrant for 12 days with corresponding immunoblot demonstrating dose-dependent degradation of ERα (inset). **F** Proliferation assay in SUM-44, SUM-52, CAMA-1, and MCF7 cells treated with indicated concentrations of fulvestrant for 12 days. (Figure on page 69).
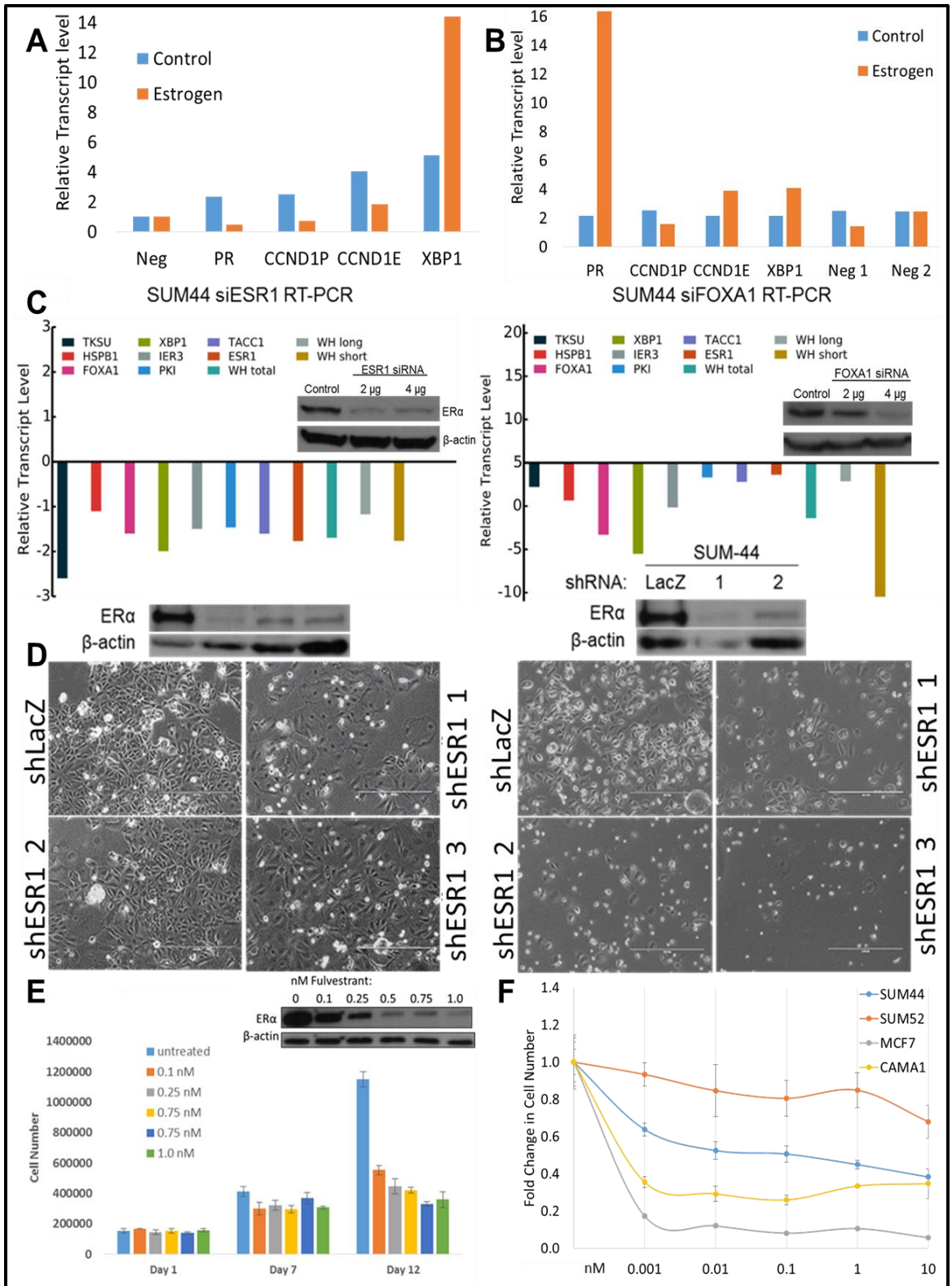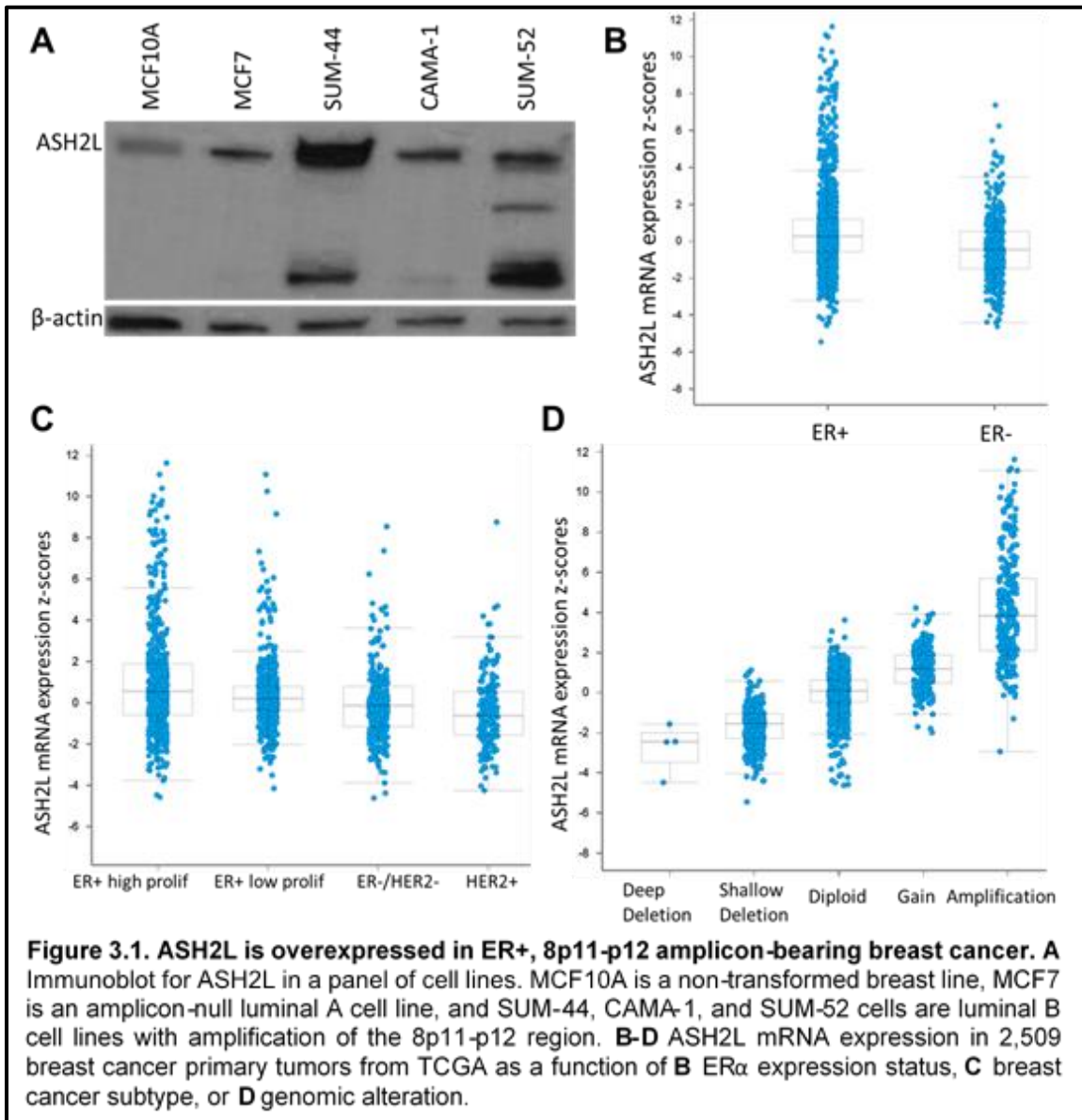
CHAPTER 3: ASH2L regulates gene expression via H3K4me3 in promoters and knockdown of ASH2L reduces expression of NSD3, ERα, and other genes important in cell proliferation processes.

3.1. ASH2L is overexpressed in ER+, 8p11-p12 amplicon-bearing breast cancer.

ASH2L (Absent, Small, or Homeotic disc 2-Like) is a chromatin-modifying factor located in the 8p11-p12 genomic region and is amplified in approximately 12% of breast cancers (TCGA). In a panel of breast cancer cell lines, ASH2L is overexpressed to varying degrees compared to the non-transformed MCF10A breast cell line (Figure 3.1-A). Of note, ASH2L is known to have two isoforms expressed in human tissues and multiple bands observed in ASH2L immunoblots suggest the presence of at least one other ASH2L isoform expressed in the amplicon-bearing SUM-44, CAMA-1, and SUM-52 cell lines (Figure 3.1-A). This possibility was not explored further and only the highest molecular weight ASH2L isoform was pursued in this study as this is the major isoform.

According to TCGA, ASH2L mRNA expression is higher in clinically-defined ER+ breast tumors compared to ER-negative tumors (Figure 3.1-B). Similarly, mRNA expression is greatest in the TCGA-defined highly proliferative subgroup of ER+ breast tumors as compared to low proliferation ER+, ER-negative, and HER2-amplified tumors (Figure 3.1-C). Amplification of ASH2L is highly correlated with overexpression at the message level (Figure 3.1-D), consistent with our observations at the protein level by immunoblot in cell lines with ASH2L amplification (Figure 3.1-A), as well as several large-scale studies of the 8p11-p12 amplicon which identified ASH2L as one of the candidate oncogenes where amplification is well-correlated with overexpression. These studies suggest ASH2L as a potential driving oncogene from this genomic region, but it remains understudied in breast cancer.

ASH2L is implicated in the tumorigenesis of osteosarcoma, glioma, and leukemia

and is known to be primarily involved in tri-methylation of lysine (K) 4 on histone H3

(H3K4me3) by interacting with binding partners, such as WDR5, RbBP5, DPY-30, and

MLL proteins, which are SET domain-containing histone methyltransferases (HMTs)

(Dou *et al.* 2006). ASH2L has demonstrated HMT activity on H3K4 as a heterodimer with

RbBP5 even without a SET domain-containing collaborator and is part of all known

Figure 3.1. ASH2L is overexpressed in ER+, 8p11-p12 amplicon-bearing breast cancer. A Immunoblot for ASH2L in a panel of cell lines. MCF10A is a non-transformed breast line, MCF7 is an amplicon-null luminal A cell line, and SUM-44, CAMA-1, and SUM-52 cells are luminal B cell lines with amplification of the 8p11-p12 region. **B-D** ASH2L mRNA expression in 2,509 breast cancer primary tumors from TCGA as a function of **B** ERα expression status, **C** breast cancer subtype, or **D** genomic alteration.
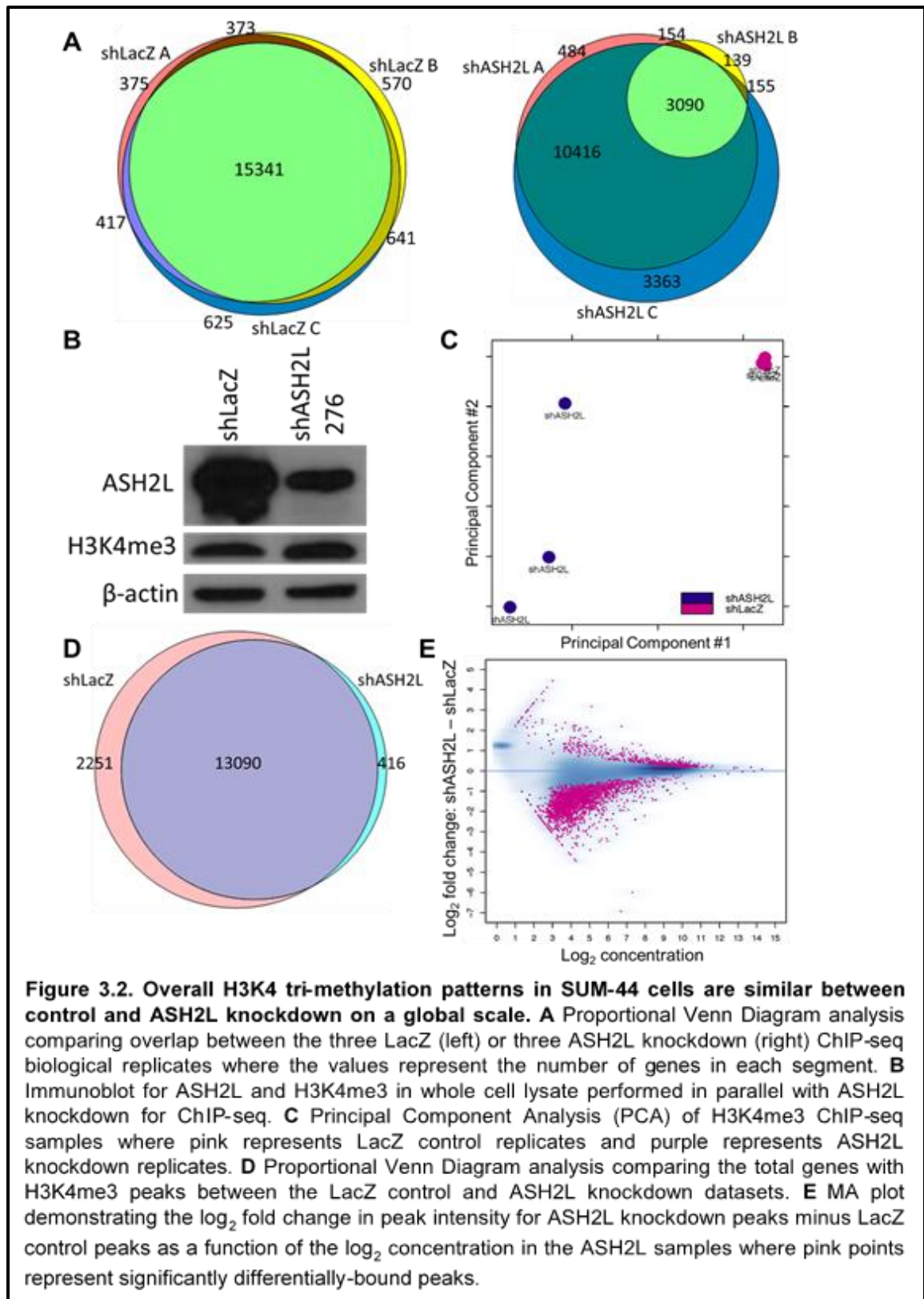
H3K4 HMT complexes characterized to date (Butler *et al.* 2017). As such, we sought to determine the influence of ASH2L on the H3K4me3 status in a luminal B breast cancer model. We selected the SUM-44 cell line, in which ASH2L is amplified and overexpressed to the highest degree of the cell lines tested (Figure 3.1-A).

3.2. Overall H3K4 tri-methylation patterns in SUM-44 cells are similar between control and ASH2L knockdown on a global scale.

To assess the influence of ASH2L overexpression on H3K4 tri-methylation, we performed chromatin immunoprecipitation and high throughput sequencing (ChIP-seq) utilizing an H3K4me3 antibody following shRNA-mediated knockdown of ASH2L or LacZ control in SUM-44 cells. Three biological replicates for each of shLacZ control and shASH2L were performed, designated A, B, and C. Quality control measures can be found in Appendix B and complete gene lists and peak data can be found in Appendix C. Sequences were aligned and annotated and the overlap between these annotated gene names corresponding to peaks called for the replicates was analyzed by Venn Diagram comparison (Figure 3.2-A). Similar numbers of peaks were called across the three LacZ control replicates with near complete overlap, corresponding to 15,341 genes with H3K4me3 enrichment in common. Overlap between ASH2L knockdown replicates A and C was also nearly identical with peaks called in a similar number of genes, both to each other as well as the shLacZ samples. Replicate B of the ASH2L knockdown group demonstrated overall poor enrichment for H3K4me3 in the paired ChIP versus input samples (Appendix B), and thus a much smaller number of genes were called in this sample. Nevertheless, the degree of overlap of this gene set with the other two ASH2L knockdown replicates was high. Due to the discrepancy in enrichment in this ChIP compared to the other two, accepting only the genes enriched in all three replicates would erroneously suggest that overall H3K4me3 levels were reduced by ASH2L

knockdown. However, western blot analysis of whole cell lysates isolated in parallel with these ChIP samples indeed showed that global levels of H3K4me3 were not reduced by knockdown of ASH2L (Figure 3.2-B). Principal component analysis also demonstrated the very close similarity between the LacZ control replicates and relative variability in the ASH2L knockdown replicates, confirming the observations by the Venn Diagram-style comparisons of annotated gene names (Figure 3.2-C). Based on these data, the genes were summed across the three replicates within each shRNA type for further analysis.

To determine the set of genes with reduced H3K4me3 enrichment due to knockdown of ASH2L, the LacZ control gene set was compared to that of the ASH2L knockdown group (Figure 3.2-D; Appendix D). We had hypothesized that there would be a significant proportion of genes with decreased enrichment of H3K4me3 upon knockdown of ASH2L, corresponding to the far left segment of the Venn Diagram in Figure 3.2-D, which represents the genes with peaks called in the LacZ control but not ASH2L knockdown group. While 2,251 genes do indeed fall into this category, the vast majority (13,090) of genes demonstrated enrichment in common between LacZ control and ASH2L knockdown. Since this Venn Diagram strategy of comparison employed a binary, presence or absence assessment of the two gene sets, differences in peak concentration between the two conditions were not taken into account. Analysis of the $\log_2$ fold change between the ASH2L knockdown and LacZ groups as a function of the $\log_2$ concentration of each sequence represented did indeed show that overall H3K4me3 enrichment was reduced following knockdown of ASH2L (Figure 3.2-E). These data demonstrate that knockdown of ASH2L does reduce the robustness of H3K4me3 peaks, but does not cause them to completely disappear. This is likely due to residual

Figure 3.2. Overall H3K4 tri-methylation patterns in SUM-44 cells are similar between control and ASH2L knockdown on a global scale. A Proportional Venn Diagram analysis comparing overlap between the three LacZ (left) or three ASH2L knockdown (right) ChIP-seq biological replicates where the values represent the number of genes in each segment. **B** Immunoblot for ASH2L and H3K4me3 in whole cell lysate performed in parallel with ASH2L knockdown for ChIP-seq. **C** Principal Component Analysis (PCA) of H3K4me3 ChIP-seq samples where pink represents LacZ control replicates and purple represents ASH2L knockdown replicates. **D** Proportional Venn Diagram analysis comparing the total genes with H3K4me3 peaks between the LacZ control and ASH2L knockdown datasets. **E** MA plot demonstrating the log$_2$ fold change in peak intensity for ASH2L knockdown peaks minus LacZ control peaks as a function of the log$_2$ concentration in the ASH2L samples where pink points represent significantly differentially-bound peaks.

expression of ASH2L in the knockdown samples (Figure 3.2-B). The implications of this
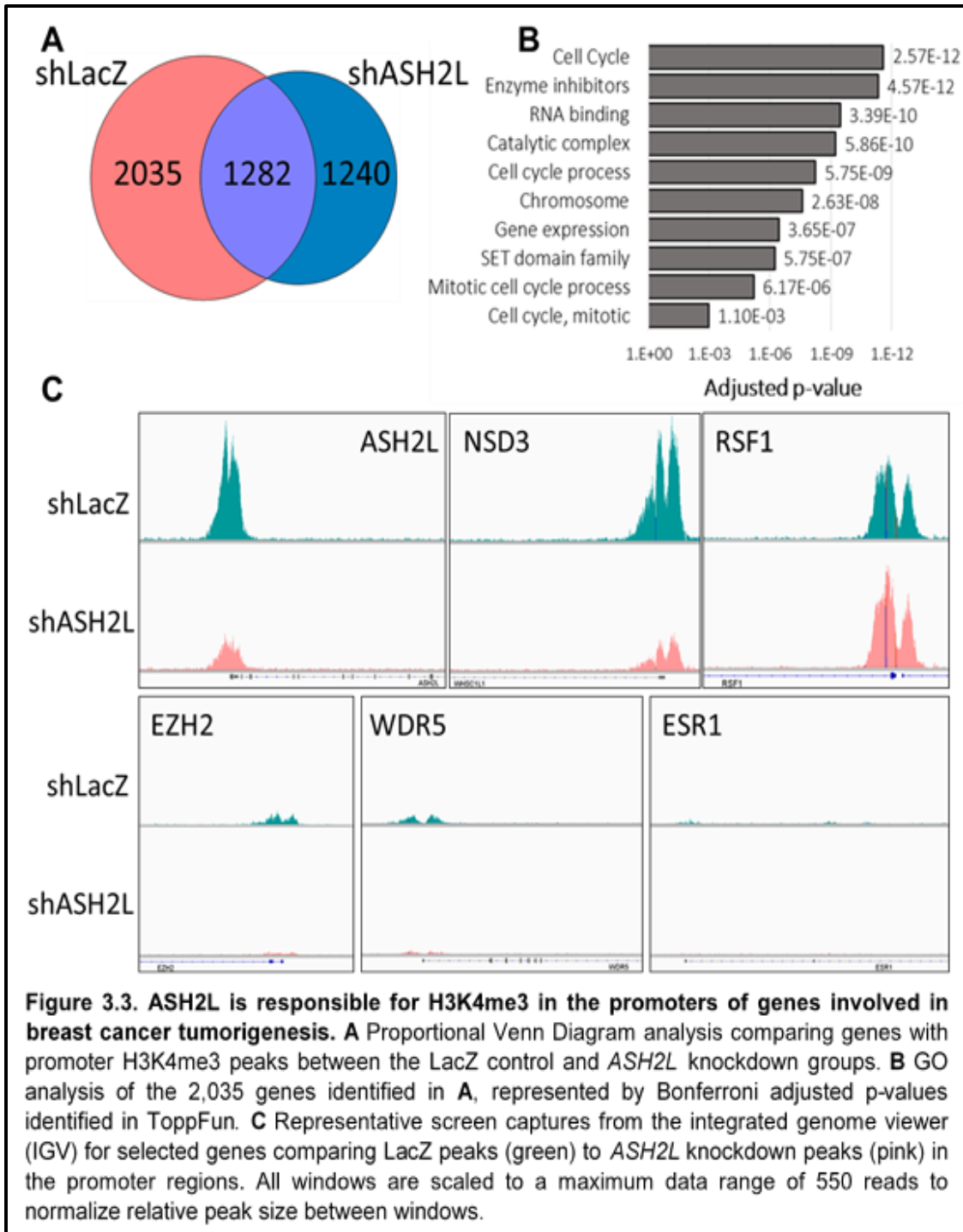
scenario on gene transcription are unknown.

Together, the results of the ChIP analysis demonstrated that knockdown of

ASH2L reduced the size of H3K4me3 peaks in SUM-44 cells, but the number of peaks

and the overall distribution of H3K4 tri-methylation between control and ASH2L

knockdown groups was largely unchanged. This observation was in contrast to our

original hypothesis that knockdown of ASH2L would reduce global patterns of H3K4me3

in SUM-44 cells. Instead, we discovered that H3K4me3 peaks are similar between

control and ASH2L knockdown, which then led us to the hypothesis that ASH2L may

play a role in H3K4 tri-methylation in a more specific set of genes.

3.3. ASH2L is responsible for H3K4me3 in the promoters of genes involved in breast cancer tumorigenesis.

H3K4me3 is known to be most abundant in the 5' proximal promoter regions of

actively transcribed genes (Demers *et al.* 2007, Luscher-Firzlaff *et al.* 2008). Similarly,

ASH2L has been shown to be confined to these regions as well (Demers *et al.* 2007). In

order to identify the specific gene set with ASH2L-regulated H3K4me3 peaks, we sorted

the overall peak data sets by detailed annotation and isolated the peaks with promoter

annotations only. Promoter H3K4me3 peaks were found in 3,317 genes from the LacZ

group and 2,522 genes from the ASH2L knockdown group. Comparison between the

two gene sets identified 2,035 genes that lose H3K4 tri-methylation following knockdown

of ASH2L and are therefore found uniquely in the LacZ control gene set (Figure 3.3-A).

These genes annotate primarily to cell cycle processes by ToppFun analysis (Figure 3.3-

B and Appendix E). Somewhat unexpectedly, there were also a significant number of

genes (1,240) with promotor H3K4 tri-methylation that demonstrated enrichment upon

ASH2L knockdown only. This is especially interesting since ASH2L is tightly and specifically linked to H3K4me3 and may indicate that another HMT gains H3K4 tri-methylation ability in the context of 8p11-p12 amplification in breast cancer.

Several interesting patterns emerged upon closer inspection of the 2,035 genes that lost H3K4me3 enrichment following ASH2L knockdown. Twelve of the 21 candidate oncogenes from the 8p11-p12 amplicon region are contained in this 2,035 gene set, including ADAM9, ASH2L, DDHD2, DUSP4, FGFR1, IKBKB, LSM1, PLEKHA2, POMK, SMIM19, UBXN8, and WHSC1L1 (NSD3), and thus we predict have increased H3K4me3 in their promotor regions due to ASH2L overexpression when the amplicon is present. These genes have some of the strongest peak scores and most have multiple peaks called in their promoter region in the shLacZ control group but do not have any peaks called in the ASH2L knockdown group (Table 3.1). Examination of the peaks in the integrated genome viewer (IGV) demonstrated the robust nature of the promoter H3K4me3 peaks in these genes, including WHSC1L1 (NSD3) and ASH2L itself, as opposed to other significant genes identified in this dataset, such as EZH2 and WDR5 (Figure 3.3-C). RSF1 was included here to illustrate the dramatic decrease in peak size in the promoter peaks of NSD3 and ASH2L upon ASH2L knockdown since RSF1 was not identified as differentially regulated between control and ASH2L knockdown datasets. ESR1 was included here to demonstrate negative ChIP data as neither the control nor ASH2L knockdown samples identified H3K4me3 peaks at the ESR1 gene promoter. These data highlight the cooperative nature of this amplicon and suggest that ASH2L may be involved in generating a positive feedback loop of active transcription of amplicon genes.

Figure 3.3. ASH2L is responsible for H3K4me3 in the promoters of genes involved in breast cancer tumorigenesis. A Proportional Venn Diagram analysis comparing genes with promoter H3K4me3 peaks between the LacZ control and *ASH2L* knockdown groups. **B** GO analysis of the 2,035 genes identified in **A**, represented by Bonferroni adjusted p-values identified in ToppFun. **C** Representative screen captures from the integrated genome viewer (IGV) for selected genes comparing LacZ peaks (green) to *ASH2L* knockdown peaks (pink) in the promoter regions. All windows are scaled to a maximum data range of 550 reads to normalize relative peak size between windows.

In addition to the 8p11-p12 amplicon genes listed above, several important

epigenetic regulators were identified in this 2,035 gene set (Table 3.1). As mentioned,

the other oncogenic 8p11-p12 amplicon HMT NSD3 was found in this set, as was EZH2, the major HMT of the polycomb group complexes that are antagonized by trithorax group proteins such as ASH2L. The ASH2L sub-complex member WDR5 appeared here, as did KMT2B (MLL4), one of the SET domain-containing enzyme family that is activated by ASH2L to generate H3K4me3. These data again suggest that knockdown of ASH2L breaks several positive feedback loops in which it participates to maintain active transcription via chromatin remodeling. While the loss of H3K4me3 peaks upon ASH2L knockdown provides some insight into the genes that may be under epigenetic control of this oncogene, this dataset does not provide information regarding the transcriptional effect that loss of this histone modification induces. Therefore, additional exploration of the effects of ASH2L knockdown on the transcriptome was carried out.

3.4 ASH2L knockdown reduces expression of genes with decreased promoter H3K4me3 levels.

To determine the influence of ASH2L knockdown on gene expression, total RNA was harvested and sequenced in triplicate following ASH2L knockdown using two different shRNA constructs in SUM-44 cells. The resulting datasets (Appendix F) were compared to RNA-seq from shLacZ control triplicate samples run in parallel. We identified 6,473 genes downregulated in common between control and ASH2L knockdown with the ASH2L 275. We similarly identified 5,766 genes downregulated in common between control and knockdown with the ASH2L 276 construct. We used a false discovery rate (FDR) cutoff of 0.4 for this analysis, a measure of the probability of type I error, in this case the assumption that these genes were significantly differently expressed from control when they actually were not. To increase statistical stringency, we next compared the differentially expressed genes from the two ASH2L shRNA constructs compared to control and identified 3,278 genes that were downregulated in

Table 3.3: Selected genes with promoter H3K4me3 peaks unique to shLacZ ChIP-seq samples. Known processes or associations of the gene are described and the average peak score of promoter region peaks, $\log_{10}$ p-value, and $\log_{10}$ q-value are reported for each gene.

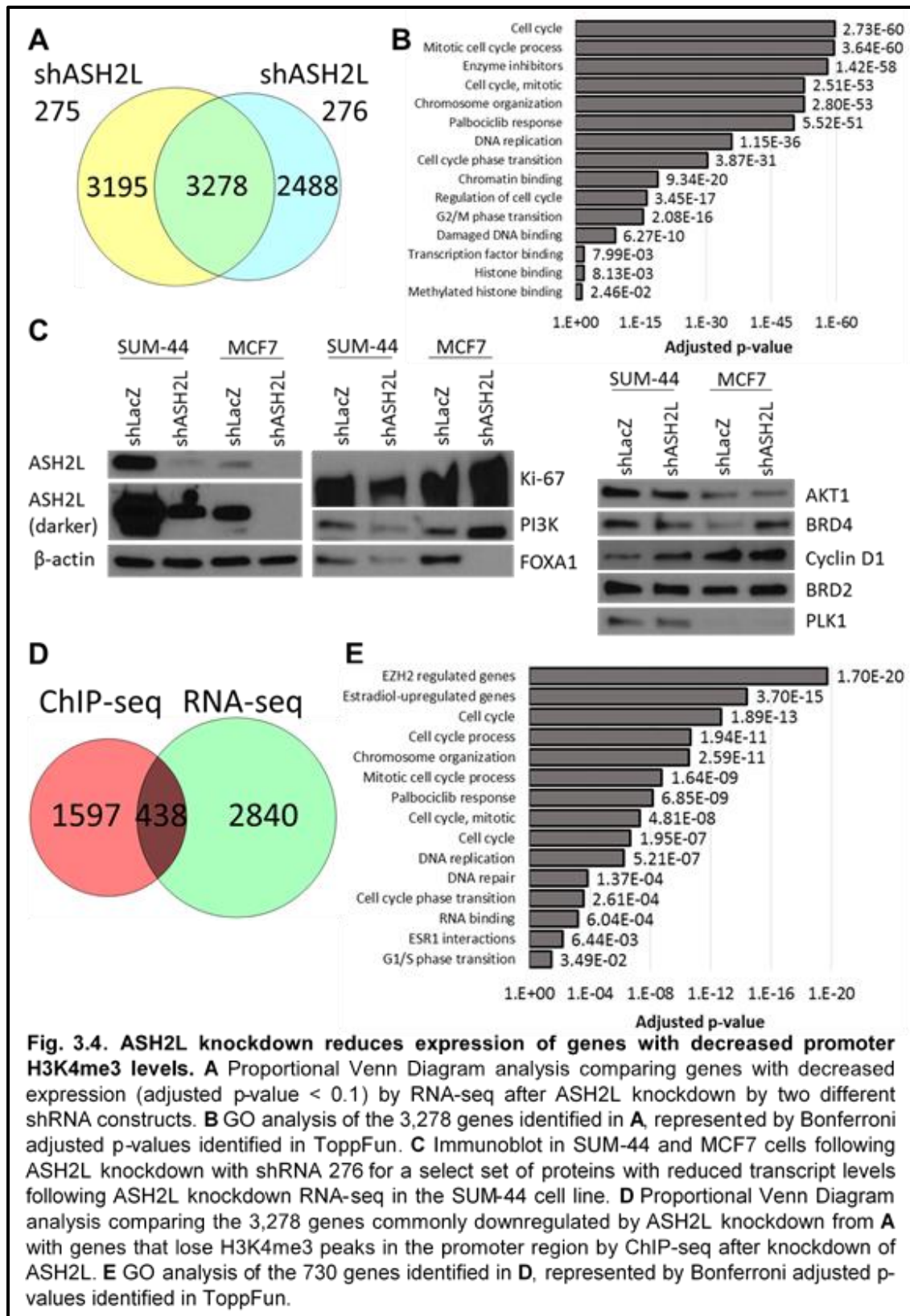| Gene Name	Description	Peak Score	log10 p-value	log10 q-value
DDHD2	8p11-p12 amplicon gene	1548	165.33	161.63
LSM1	8p11-p12 amplicon gene	1483	163.67	160.04
ASH2L	8p11-p12 amplicon gene	1284	151.53	148.00
FGFR1	8p11-p12 amplicon gene	1420	145.81	142.05
WHSC1L1 (NSD3)	8p11-p12 amplicon gene	1405	144.25	140.51
ADAM9	8p11-p12 amplicon gene	1024	109.46	106.39
CCNE2	cell cycle regulation	889	99.71	96.84
PLEKHA2	8p11-p12 amplicon gene	748	86.65	83.86
CDK6	cell cycle regulation	649	76.49	73.85
NCOA7	transcription-associated factor	574	73.66	71.05
BRD2	chromatin-associated factor	657	70.71	68.13
IKBKB	8p11-p12 amplicon gene	579	68.71	66.14
WDR5	ASH2L binding partner	476	43.78	41.35
IRX2	embryogenesis	482	54.34	51.91
FGFR4	cell signaling pathway	441	51.58	49.16
UBE2C	associated with tamoxifen resistance	485	51.04	48.52
EZH2	polycomb group HMT	482	50.68	48.28
POMK	8p11-p12 amplicon gene	465	49.00	46.50
ASF1B	associated with tamoxifen resistance	369	46.63	42.27
RECQL4	transcription-associated factor	440	46.54	44.09
SMIM19	8p11-p12 amplicon gene	360	41.93	39.63
AURKA	cell cycle regulation	392	41.67	39.26
PIK3CA	cell signaling pathway	345	41.51	39.19
SETD2	HMT (SET family)	388	41.35	39.03
KMT2B (MLL4)	ASH2L binding partner, HMT (SET family)	347	38.48	36.18
UBXN8	8p11-p12 amplicon gene	318	38.43	36.13
DUSP4	8p11-p12 amplicon gene	353	37.77	35.40

common between them (Figure 3.4-A; Table 3.2; Appendix G). In order to assess the biological processes affected by this suite of genes, we performed ToppFun analysis on the 3,278 genes and discovered that they are involved in cell cycle regulation, DNA replication, and chromatin organization (Figure 3.4-B). These results are consistent with previously implicated functions of ASH2L. We validated knockdown of several of these genes by immunoblot, including Ki-67, PI3K, and FOXA1 (Figure 3.4-C). Expression of

Table 3.4: Selected genes commonly downregulated between shASH2L 275 and shASH2L 276 in SUM-44 cells. Known processes or associations of the gene are described and log$_2$ fold change from shLacZ control and associated p-value and adjusted p-value are reported for each of the two shRNA constructs.

Description	shASH2L 275			shASH2L 276		
	log2 fold change	p-value	adj p-value	log2 fold change	p-value	adj p-value
H3K4me3 (activating)	-4.33	<1.05E-305	<4.06E-303	-2.65	<2.92E-74	<2.25E-72
Luminal B marker	-1.74	<1.05E-305	<4.06E-303	-0.88	2.92E-74	2.25E-72
Transcription-associated, tamoxifen resistance	-1.92	<1.05E-305	<4.06E-303	-0.55	1.31E-45	4.34E-44
Transcription factor	-1.78	1.05E-305	4.06E-303	-0.49	4.62E-41	1.24E-39
Luminal B marker, transcription factor, ERα-associated	-1.50	6.65E-198	9.96E-196	-0.51	3.61E-09	1.74E-08
Cell cycle regulation	-1.46	3.50E-137	2.69E-135	-0.75	4.35E-36	9.59E-35
Cell cycle regulation, tamoxifen resistance	-1.09	2.83E-112	1.43E-110	-0.47	3.93E-27	5.89E-26
Cell cycle regulation	-0.57	4.71E-80	1.43E-78	-0.12	1.09E-04	3.15E-04
Luminal B marker, transcription factor, ERα-associated	-0.73	1.37E-73	3.64E-72	-0.27	8.91E-16	7.08E-15
H3K37me3 (repressive)	-0.87	9.83E-71	2.52E-69	-0.54	4.95E-24	6.34E-23
Transcription factor	-0.68	8.73E-54	1.55E-52	-0.73	4.65E-10	2.43E-09
Transcription factor	-0.74	1.45E-50	2.35E-49	-0.07	6.86E-02	1.14E-01
Cell cycle regulation, tamoxifen resistance	-0.62	1.50E-32	1.44E-31	-0.59	2.95E-07	1.18E-06
Transcription factor	-0.47	2.94E-26	2.30E-25	-0.32	6.74E-15	5.04E-14
Luminal B marker, proliferation	-1.01	8.32E-20	5.04E-19	-0.71	5.30E-06	1.81E-05
Cell cycle regulation, ERα-associated	-0.47	9.47E-19	5.47E-18	-0.37	1.24E-02	2.47E-02
8p11-p12 amplicon gene	-0.25	1.39E-14	6.64E-14	-0.50	6.47E-58	3.17E-56
Cell cycle regulation	-0.28	1.80E-12	7.68E-12	-0.14	1.75E-05	5.61E-05
8p11-p12 amplicon gene; Histone acetyltransferase	-0.25	1.94E-08	6.42E-08	-0.12	5.86E-03	1.25E-02
ASH2L binding partner	-0.31	3.97E-07	1.18E-06	-0.33	1.52E-09	7.55E-09
8p11-p12 amplicon gene; H3K36me2 (activating)	-0.18	4.69E-06	1.29E-05	-0.22	6.97E-09	3.26E-08
H3K36me3 (activating)	-0.19	2.35E-04	5.50E-04	-0.17	8.09E-04	2.03E-03
Bromodomain protein	-0.10	4.15E-03	8.31E-03	-0.42	5.65E-42	1.57E-40

these genes did decrease upon knockdown of ASH2L in SUM-44 cells but had varied changes in expression in MCF7 cells, in which ASH2L is not amplified and its expression is much lower than in SUM-44 cells. These data validate the RNA-seq results and confirm that knockdown of ASH2L reduces gene expression. Interestingly, several of the genes with reduced transcript expression following ASH2L knockdown RNA-seq analysis did not demonstrate reduced protein levels (Figure 3.4-C), suggesting a disconnect between transcript- and protein-level expression for a subset of ASH2L-regulated genes.

In order to determine which of the genes with decreased H3K4me3 promoter peaks in ASH2L knockdown cells also demonstrated decreased expression upon knockdown of ASH2L, the 3,278 genes downregulated in common between both ASH2L knockdown constructs compared to control were then cross-referenced with the 2,035 genes that lost H3K4me3 peaks after knockdown of ASH2L. This strategy identified 438 genes with coordinated decreased expression and H3K4me3 peak loss (Figure 3.4-D). ToppFun analysis of this set of genes assigned many of the same processes and pathways as the RNA-seq gene set alone, including palbociclib response, cell cycle processes, and chromosome organization (Figure 3.4-E and Appendix H). This gene set, however, was also assigned to several additional processes and pathways, such as genes that are upregulated in MCF7 cells upon treatment with estradiol, proteins that interact with ESR1/ERα, DNA and RNA binding, and histone methyltransferase activity. Together, these results demonstrate that ASH2L regulates H3K4me3 in the promoter regions of important genes related to breast cancer, primarily related to cell cycle and chromatin modification, and some of these genes are downregulated at the transcript and protein level as a result.

Fig. 3.4. ASH2L knockdown reduces expression of genes with decreased promoter H3K4me3 levels. A Proportional Venn Diagram analysis comparing genes with decreased expression (adjusted p-value < 0.1) by RNA-seq after ASH2L knockdown by two different shRNA constructs. B GO analysis of the 3,278 genes identified in A, represented by Bonferroni adjusted p-values identified in ToppFun. C Immunoblot in SUM-44 and MCF7 cells following ASH2L knockdown with shRNA 276 for a select set of proteins with reduced transcript levels following ASH2L knockdown RNA-seq in the SUM-44 cell line. D Proportional Venn Diagram analysis comparing the 3,278 genes commonly downregulated by ASH2L knockdown from A with genes that lose H3K4me3 peaks in the promoter region by ChIP-seq after knockdown of ASH2L. E GO analysis of the 730 genes identified in D, represented by Bonferroni adjusted p-values identified in ToppFun.

3.5 ASH2L, NSD3, and ERα influence the expression of similar suites of genes related to cell cycle progression.

The 438 genes identified in the previous section as downregulated upon loss of promoter H3K4me3 due to ASH2L knockdown were also implicated in ESR1 interactions and response to estradiol (Figures 3.4-D and 3.4-E; Appendix H). Given these results, the inclusion of NSD3 in this group of 438 ASH2L-regulated genes, and the link between NSD3 overexpression and estrogen-independent activation of ERα (see Chapter 2), we hypothesized that ASH2L is responsible for promoter H3K4me3-mediated NSD3 overexpression in the context of amplification of these two oncogenes, thus influencing ERα expression and activity and perpetuating the alterations to the transcriptome in the setting of 8p11-p12 amplified breast cancer. Knockdown of ASH2L resulted in reduced protein expression of NSD3 and ERα in SUM-44 cells (Figure 3.5-A), supporting the hypothesis that these three factors are cooperating to alter transcription in luminal B breast cancer models. Additionally, preliminary evidence in support of this hypothesis from TCGA demonstrates that ASH2L and NSD3 overexpression at the transcript level is highly correlated in a cohort of 2,509 primary breast tumors, with a Pearson correlation score of 0.81 and Spearman rank of 0.65 (Figure 3.5-B).

To test this hypothesis further, we compared the 438 genes with reduced H3K4me3 and expression upon ASH2L knockdown with genes that were downregulated upon ESR1 or NSD3-S knockdown using a p-value cutoff of 0.1 (Figure 3.5-C). In this analysis, 320 genes were unique to the ASH2L knockdown H3K4me4 ChIP-seq/RNA-seq downregulated group. We ran ToppFun analysis on this set of 320 genes and discovered they align with processes primarily involved in DNA repair and chromosome organization. (Figure 3.5-D and Appendix I). Interestingly, "ESR1 interactions" was

Figure 3.5. ASH2L, NSD3, and ERα influence the expression of similar suites of genes related to cell cycle progression. A Immunoblot for ERα and NSD3 following knockdown of ASH2L in SUM-44 cells. **B** NSD3 transcript expression as a function of ASH2L transcript expression in 2,509 primary breast cancer samples from TCGA. Pearson=0.81 and Spearman=0.65. **C** Proportional Venn Diagram analysis comparing genes with decreased expression and reduced promoter H3K4me3 after ASH2L knockdown (upper left) with genes demonstrating downregulation (adjp<0.1) by microarray analysis following knockdown of NSD3-S (upper right) or ESR1 (lower). **D** GO analysis of the 320 genes identified in **C**, represented by Bonferroni adjusted p-values identified in ToppFun. **E** Immunoblot for total ERα and ERα-pS167 in SUM-44 (left) and MCF7 (right) cells following knockdown of ASH2L and treatment with the mTOR inhibitor KU-0063796 at 1 μM for 24 hours. **F** Immunoprecipitation of ASH2L and immunoblot for ASH2L and ERα

identified as a significant process ascribed to the 320 genes regulated by ASH2L, indicating that the group of genes under ASH2L transcriptional regulation are ERα - interacting proteins. As expected, these genes include chromatin remodeling enzymes and factors associated with transcription, such as TOPBP1, ZNF131, SMARCE1, TADA3, and several ribosomal proteins. PIK3CA also appears in this set of 320 genes, and the PI3K/AKT/mTOR pathway is known to play a role in activation of ERα via phosphorylation of S167 of the ERα protein. Treatment of LacZ control or ASH2L knockdown SUM-44 cells with 1 µM of the mTOR inhibitor KU-0063794 for 24 hours did not reduce ERα-pS167 levels in SUM-44 or MCF7 cells, however (Figure 3.5-E). ASH2L and its sub-complex binding partner WDR5 are included in the list of genes that annotate in ToppFun to ESR1 interactions, suggesting that these proteins themselves may cooperate with ERα. ASH2L immunoprecipitation (IP) followed by ERα blotting identified a possible interaction between these two proteins (Figure 3.5-F). Although the MLL proteins are known to bind to ERα, this is the first study to provide evidence that ASH2L and ERα may directly interact and this possibility should be explored further.

From the overlapping gene sets in Figure 3.5-C, the 58 genes commonly downregulated by ASH2L and ESR1 knockdown were primarily involved in the cell cycle and estradiol-regulated processes. The 16 genes commonly downregulated by ASH2L and NSD3 knockdown were involved in processes having to do with transcription and chromatin/DNA interactions (data not shown). We have previously analyzed the overlap between ESR1 and NSD3 function (Chapter 2; (Irish *et al.* 2016)). Since the set of genes here is specifically those which were found to be downregulated by microarray analysis (p<0.1), we analyzed these genes by ToppFun and found that they annotate to

processes similar to what we previously reported, including cell cycle and estradiol-mediated processes (data not shown).

3.6 Knockdown of ASH2L results in decreased proliferation and clonogenicity in luminal B breast cancer and reduces sensitivity to palbociclib.

Since ASH2L knockdown reduces expression of NSD3 and ERα, we hypothesized that ASH2L indirectly mediates the expression of the genes regulated by these two factors. The H3K4me3 ChIP-seq data indicated that ASH2L directly affects epigenetic regulation of NSD3 but not ESR1 (Figure 3.3-C) and we already validated the causal relationship between knockdown of NSD3 and decreased ERα expression (Chapter 2), so it is possible that the downregulation of these two factors occurs sequentially following knockdown of ASH2L. To test this hypothesis, whole cell lysate was harvested from SUM-44 cells at 5, 7, 11, and 14 days following infection with ASH2L-specific or LacZ control shRNA vectors and probed for ASH2L, NSD3, and ERα protein expression (Figure 3.6-A). Although previous immunoblots for these proteins performed as knockdown validation steps for other experiments indicated that the reduction in NSD3 expression may precede that of ERα, and that shRNA-mediated NSD3 knockdown-induced reduction in ERα expression takes approximately seven days, this more controlled test did not reproduce the temporal aspect of ASH2L knockdown-mediated reduction of NSD3 and ERα expression.

Some of the 44 genes shown in Figure 3.5-C derived by comparison of the ASH2L knockdown H3K4me3 ChIP-seq, ASH2L knockdown RNA-seq, and ESR1 and NSD3-S knockdown arrays gene sets appear in Table 3.3. This gene set is associated with several interesting processes by ToppFun analysis (Figure 3.6-B and Appendix I). Many of these processes are associated with the cell cycle. Since this is a common

theme associated with ASH2L-mediated transcriptomic alterations and high proliferation rates are associated with luminal B breast cancer and poor prognosis, we measured cell growth following shRNA-mediated knockdown of ASH2L in several cell lines (Figure 3.6-C). In both amplicon-bearing (SUM-44, CAMA-1, and SUM-52) and amplicon-null (MCF7) cell lines, knockdown of ASH2L by two different shRNA constructs reduced cell proliferation compared to shLacZ control. Knockdown levels were reported by immunoblot analysis for ASH2L expression in the four cell lines (Figure 3.6-C). Similarly,

Table 3.3. Summary of genes identified by ASH2L knockdown H3K4me3 ChIP-seq, ASH2L knockdown RNA-seq, NSD3-S knockdown array, or ESR1 knockdown array. Selected genes identified by RNA-seq as commonly downregulated between both shRNA constructs against ASH2L are reported in this table. Known processes and associations are described for each gene. Check marks indicate whether the gene was identified as having lost promoter H3K4me3 upon ASH2L knockdown (ChIP-seq), downregulated by both shRNA constructs against ASH2L (RNA-seq), downregulated by NSD3-S knockdown (NSD3 array), or downregulated by ESR1 knockdown (ESR1 array). Blank cells indicate the gene was not identified in that sample set.

Gene Name	Description	ChIP-seq	RNA-seq	NSD3 array	ESR1 array
EZH2	polycomb group HMT:H3K27 (repressive)	✓	✓	✓	✓
RECQL4	associated with tamoxifen resistance	✓	✓	✓	✓
UBE2C	associated with tamoxifen resistance	✓	✓	✓	✓
CCNE2	cyclin E2, cell cycle regulator	✓	✓	✓	✓
CKD2	cyclin dependent kinase (cyclin E2), associated with tamoxifen resistance	✓	✓	✓	✓
BRD2	bromodomain protein, chromatin-associated factor	✓	✓	✓	✓
ASF1B	associated with tamoxifen resistance	✓	✓	✓	✓
NSD3	HMT: H3K36me2, 8p11 oncogene	✓	✓	✓	
AURKA	aurora A kinase, cell proliferation	✓	✓		✓
ASH2L	H3K4me3, 8p11 oncogene	✓	✓		
PIK3CA	PI3K, cell signaling oncogene/survival pathway	✓	✓		
WDR5	ASH2L binding partner	✓	✓		
SETD2	HMT: H3K36me3	✓	✓		
ESR1	estrogen receptor alpha, transcription factor		✓	✓	✓
CCNB1	cyclin B1, cell cycle regulator		✓	✓	✓
ATAD2	bromodomain protein, ER co-activator		✓	✓	✓
CCNB2	associated with tamoxifen resistance		✓	✓	✓
TK1	associated with tamoxifen resistance		✓	✓	
AURKB	aurora B kinase, cell proliferation		✓		✓
MYBL2	proliferation marker of luminal B		✓		✓
CCNE1	cyclin E1, cell cycle regulator		✓		✓
CCND1	cyclin D1, cell cycle regulator, 11q14 amplicon		✓		✓
PLK1	Kinase involved in cell proliferation		✓		
MLL5	Mixed lineage leukemia family HMT		✓		
FOXA1	ER pioneer factor		✓		
FLT3LG	ligand of FLT3 (cell proliferation and survival)		✓		
GATA3	TF, ER partner, luminal BC marker		✓		
SMYD3	HMT, transcription complex member		✓		
AKT1	cell survival pathway signaling		✓		

knockdown of ASH2L was performed in SUM-44 and MCF7 cells and clonogenicity

measured by colony forming assay (Figure 3.6-D). Knockdown of ASH2L reduced

clonogenicity in both cell lines to approximately the same degree, and the 276 ASH2L

shRNA construct demonstrated a much greater effect than the 275 construct in both cell

lines, although both achieved statistical significance (Figure 3.6-D, upper). Together,

these data demonstrate that knockdown of ASH2L decreases cell proliferation and

clonogenic potential in luminal breast cancer cell lines regardless of 8p11-p12

amplification status.

In addition to cell cycle processes, analysis of ASH2L-regulated genes

consistently identified palbociclib response as a significant process to which they are

annotated (Figure 3.4-B, Figure 3.4-E, Figure 3.6-B). Palbociclib selectively inhibits

cyclin dependent kinases (CDK) 4 and CDK6, which bind cyclin D1 (CCND1) to advance

cell cycle progression from G_1 to S phase (Nagaraj and Ma 2015). CCND1 is found on

the 11q14 amplicon, which SUM-44 cells harbor in addition to the 8p11-p12 amplicon

(Kwek *et al.* 2009). CDK6 is also amplified in this cell line and was identified by siRNA-

based screening technique performed in our lab as essential for growth and survival of

these cells. Palbociclib is approved for use in postmenopausal women with ER+/HER2-

negative breast cancer and clinical trials are in progress for several similar compounds

(Knudsen and Witkiewicz 2016). To determine the sensitivity of ASH2L-overexpressing

cell lines to palbociclib, we treated a panel of cell lines with increasing doses of

palbociclib and assessed cell number on day seven (Figure 3.7-A). We observed that

the ASH2L-low MCF7 and CAMA-1 cell lines were more sensitive to palbociclib than the

SUM-44 and SUM-52 cells. To validate the bioinformatic assignment of ASH2L-

regulated genes to palbociclib response, we knocked down ASH2L in SUM-44 and
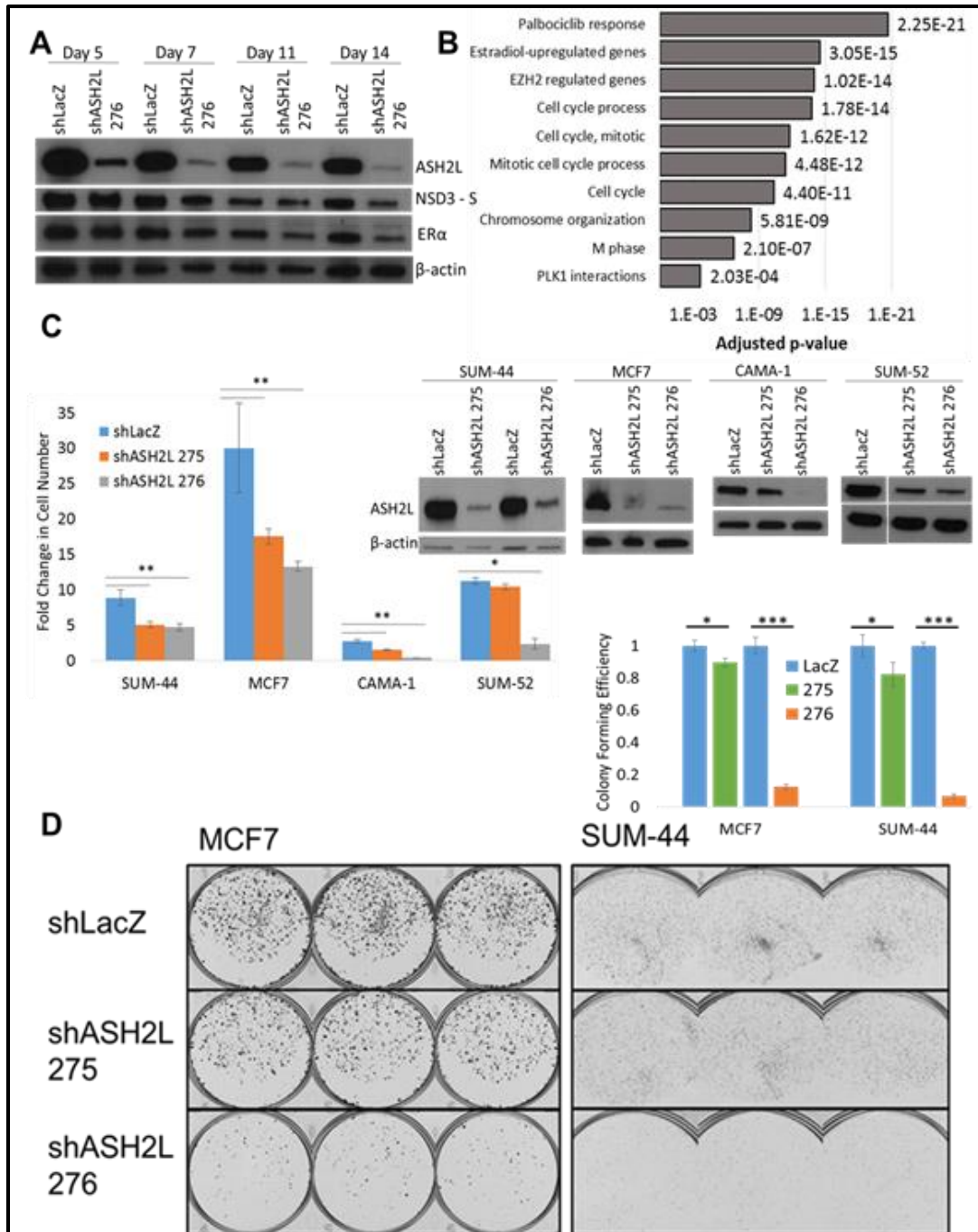
Figure 3.6. Knockdown of ASH2L results in decreased proliferation and clonogenicity in luminal B breast cancer. A Immunoblot for NSD3-S and ERα in SUM-44 cells 5, 7, 11, and 14 days after ASH2L knockdown. **B** ToppFun analysis on the 44 genes identified in **Figure 3.5-C**, represented by Bonferroni adjusted p-values identified in ToppFun. **C** Proliferation assay in SUM-44, MCF7, CAMA-1, and SUM-52 cells following ASH2L knockdown. Corresponding immunoblot for ASH2L appears in the upper right. **D** Colony forming assay plate photographs (stained with crystal violet) of MCF7 (left) or SUM-44 (right) cells following ASH2L knockdown and corresponding quantification (upper right). *indicates p<0.02 and ***indicates p<6$_E$-6.

MCF7 cells with increasing doses of palbociclib and measured their growth (Figure 3.7-B). Knockdown of ASH2L reduced sensitivity of SUM-44 cells to palbociclib but had no effect on MCF7 sensitivity to this compound. Of note, MCF7 had greater overall sensitivity to palbociclib than SUM-44 cells. MCF7 cells do not have amplification of 8p11-p12 or 11q14.

To further examine the relationship between genes regulated in expression by ASH2L, NSD3, and ESR1, and genes associated with response to palbociclib, a sub-set of genes was identified and expression level assessed by qRT-PCR. The set of 44 genes regulated by ASH2L, NSD3, and ESR1 included 10% (17 of 172) of the genes annotated as involved in response to palbociclib in ToppFun according to the comparative toxicogenomics database (CTD) (Table 3.4). From this set of 17 genes that lost promoter H3K4me3 upon ASH2L knockdown and were downregulated in response to knockdown of ASH2L, NSD3, and ESR1, we selected a sub-set of five genes that were downregulated in response to treatment with palbociclib according to CTD data, including FBXO5, EZH2, TTK, CCNE2, and BUB1 (Table 3.4). To determine if treatment with palbociclib reduced expression of these genes in the context of ASH2L overexpression, we treated SUM-44 and MCF7 cells with palbociclib and harvested RNA 24 hours after treatment with 100 nM palbociclib for analysis of relative transcript expression by RT-PCR (Figure 3.7-C). Expression of *FBXO5, EZH2, TTK, CCNE2,* and *BUB1* was reduced in the palbociclib-sensitive MCF7 cell line, as predicted by the CTD data, however SUM-44 cells did not demonstrated a reduction in transcript level for these genes. ASH2L and ESR1 transcript levels were not significantly affected by palbociclib treatment (Figure 3.7-C). RNA was also harvested 2 and 96 hours following treatment with 100 nM palbociclib, demonstrating a recovery in gene expression at 96

Figure 3.7. Knockdown of ASH2L alters response to palbociclib. A Proliferation assay in SUM-44, MCF7, CAMA-1, and SUM-52 cells following treatment with palbociclib at indicated doses. **B** Proliferation assay in SUM-44 and MCF7 cells treated with increasing doses of palbociclib following knockdown of ASH2L. *indicates p<0.003 and **indicates p<0.0004. **C** qRT-PCR for palbociclib-responsive and ASH2L-regulated genes as indicated in Table 3.4 following 24 hours of 100 nM palbociclib treatment in MCF7 and SUM-44 cells. **D** qRT-PCR for palbociclib-responsive and ASH2L-regulated genes as indicated in Table 3.4 harvested 2, 24, or 96 hours after treatment with 100 nM palbociclib in SUM-44 and MCF7 cells.

hours post-palbociclib in the MCF7 cell line and an increase in transcript expression in

the SUM-44 line for the five genes assessed (Figure 3.7-D). These data indicate that

palbociclib resistance in the SUM-44 cells may be due to ASH2L-mediated transcriptional upregulation of genes that must be downregulated for response to palbociclib, thus preventing the cytostatic effects of this compound. This hypothesized mechanism should be investigated further.

The smallest set of genes analyzed in ToppFun, the 44 in common to ASH2L knockdown H3K4me3 ChIP-seq, ASH2L knockdown RNA-seq, NSD3-S and ESR1 knockdown microarrays (Figure 3.5-C), annotated to palbociclib response with the strongest p-value (Figure 3.6-B; Table 3.4), indicating that ASH2L, NSD3, and ERα all have a role in mediating expression of the suite of genes responsible for sensitivity to palbociclib and highlighting again the cooperative potential of these oncogenes in 8p11-p12-amplified breast cancer. This set of 44 overlapping genes (Appendix H) likely regulated by ASH2L, NSD3, and ESR1 requires further validation of their expression and investigation of the biological processes to which they connect that is beyond the scope of this study.

3.7 Summary

Little was known about ASH2L in breast cancer prior to these studies. Here, we demonstrated that ASH2L is involved in H3K4 tri-methylation specifically at the promoters of target genes without altering global H3K4me3 levels. Knockdown of ASH2L results in decreased expression of a subset of these genes with ASH2L-mediated promoter H3K4me3 marks, and these genes are implicated in cell cycle regulation and chromosome organization. ASH2L potentially exerts downstream effects by modulating several proteins important to cell signaling pathways, such as the PI3K/AKT1/mTOR pathway, and through expression of cell cycle proteins such as cyclin D1. Additionally,

ASH2L may be responsible for H3K4me3-mediated transcriptional regulation of itself as well as the co-amplified HMT NSD3, which has been shown to overexpress and activate ERα in an estrogen-independent manner (Chapter 2). Together, these data implicate ASH2L in the regulation of a large suite of genes, both directly and indirectly, when amplified and overexpressed in luminal B breast cancers. These studies identify ASH2L as an upstream regulator of many targetable pathways in breast cancer, and therefore this oncogene has the potential to inform drug discovery. Further investigation into the ASH2L complex composition in this setting and confirmation of the factors and pathways under control of this oncogene will be essential to extend these studies, ultimately leading to therapies that will benefit the breast cancer patient population with ASH2L amplification and overexpression.

Table 3.4. Genes annotated to palbociclib response in ToppFun. Genes from the input set that aligned to the ToppFun annotation set for Palbociclib Response are listed in this table. The input set consisted of 44 genes that were downregulated upon knockdown of ASH2L, NSD3-S, and ESR1 and lost promoter H3K4me3 peaks after ASH2L knockdown. Genes selected for validation by qRT-PCR are indicated.

Gene Symbol	Gene Name	PCR
KIF20A	Kinesin family member 20A	
HMGB2	High mobility group box 2	
MCM7	Minichromosome maintenance complex component 7	
HMMR	Hyaluronan mediated motility receptor	
FBXO5	F-box protein 5	✓
KIF15	Kinesin family member 15	
EZH2	Enhancer of zeste 2 polycomb repressive complex 2 subunit	✓
CDC25C	Cell division cycle 25C	
CENPE	Centromere protein E	
TTK	TTK protein kinase	✓
CENPM	Centromere protein M	
ASF1B	Anti-silencing function 1B histone chaperone	
NCAPG2	Non-SMC condensing II complex subunit G2	
GMNN	Geminin, DNA replication inhibitor	
CCNE2	Cyclin E2	✓
UBE2C	Ubiquitin conjugating enzyme E2 C	
BUB1	BUB1 mitotic checkpoint serine/threonine kinase	✓

CHAPTER 4: Discussion and Future Directions

4.1. 8p11-p12 Amplified Breast Cancer

a. Breast cancer *in vitro* models

Due to the incredible heterogeneity of breast cancer, a major challenge in research is acquiring laboratory models able to encompass the multiple patient subpopulations observed clinically (Benjamini and Hochberg 1995, Kao *et al.* 2009, Dabydeen and Furth 2014). The MCF7 cell line is considered the gold standard by which ER+ breast cancers are studied *in vitro* and *in vivo*, however this luminal A cell line does not sufficiently capture the complexity of the luminal type tumors as a whole (Comsa *et al.* 2015). This cell line has a simplex genomic pattern and is sensitive to endocrine therapies such as tamoxifen (Shiu *et al.* 2010, Comsa *et al.* 2015). While attempts have been made to induce tamoxifen resistance in MCF7 cells and these altered cell lines have been thoroughly investigated, this artificial representation of endocrine resistance does not translate to patient tumors and the genomic profiles of these modified MCF7 cell lines do not resemble patient-derived luminal B cell lines (Ross-Innes *et al.* 2012).

Model cell lines with intrinsic tamoxifen resistance have been reported but generally have lost ERα expression or dependence in culture, such as the SUM-52 cell line. While it is well-known that a significant number of patient tumors retain ERα expression even in the context of resistance to endocrine therapy, an *in vitro* model accurately representing this patient population has not previously been available (Cui *et al.* 2005). Similarly, although cell lines with amplification of the 8p11-p12 amplicon have been reported, these models have varying degrees of expression of the oncogenes contained in this region (Ray *et al.* 2004, Garcia *et al.* 2005, Gelsi-Boyer *et al.* 2005).

Approximately 15% of primary breast tumors have 8p11-p12 amplification and this number increases in metastatic samples, correlating with poor prognosis and ERα expression (Bernard-Pierrot *et al.* 2008, Tabarestani *et al.* 2016). Several groups have linked the 8p11-p12 amplicon with endocrine resistance (Shiu *et al.* 2010, Luo *et al.* 2017). Amplification is a major mechanism of oncogene overexpression and activation (Ray *et al.* 2004, Yang *et al.* 2006, Streicher *et al.* 2007, Bernard-Pierrot *et al.* 2008), and thus there is a need for more model cell lines by which to study the highly heterogenic 8p11-p12 amplicon.

The SUM-44 cell line, originating from metastatic cells in the pleural effusion fluid from a breast cancer patient with ER+ disease (Ethier *et al.* 1993), reflects the characteristics of a subset of breast cancer patients not previously represented *in vitro*, including intrinsic tamoxifen resistance and amplification of the 8p11-p12 genomic region (Ray *et al.* 2004, Irish *et al.* 2016). This cell line expresses ESR1/ERα to a higher degree than other ER+ cell lines, even MCF7 cells, a biologically relevant observation when compared to patient tumor sample ESR1 expression (Figure 2.3-B). SUM-44 cells have amplification of 8p11-p12 with abundant overexpression of the amplicon genes, even compared to other amplicon-bearing cell lines such as CAMA-1 and SUM-52 (Figure 2.3-A). Interestingly, the SUM-52 cell line, originally ER+, has lost ERα expression in culture and although the CAMA-1 cell line retains ERα expression, the level is very low compared to MCF7 and SUM-44. Together, the data presented in this work establish a model by which endocrine resistant, highly ER+ breast tumors can be investigated. Due to the heterogeneity of breast cancer patients, establishment of additional models is necessary, but here we offer an important advancement in the ability to understand the

8p11-p12 amplicon in luminal B breast cancer and the effects of the amplicon genes on constitutive activation of ERα as well as the transcriptome as it relates to tumorigenesis.

b. NSD3

NSD3 (WHSC1L1) was identified and validated as an 8p11-p12 amplicon oncogene in breast and other cancers (Yang *et al.* 2010, Bilal *et al.* 2012, He *et al.* 2013, Mahmood *et al.* 2013, Chen *et al.* 2014, French *et al.* 2014). We and others have demonstrated that the catalytic, SET domain-containing NSD3-L isoform is not the major oncogenic isoform (Shen *et al.* 2015, Irish *et al.* 2016). Instead, NSD3-S, produced by alternative splicing at exon 10 of the NSD3 (WHSC1L1) gene, is more highly expressed than NSD3-L (Figure 2.1-A) and has greater transforming ability when overexpressed in the MCF10A breast cell line (Yang *et al.* 2010). Knockdown of this isoform reduced cell proliferation and expression of ESR1, the gene that encodes ERα (Figure 2.2). Further investigation of this relationship identified that NSD3-mediated overexpression of ERα resulted in estrogen-independent activity of this receptor (Figure 2.5) and tamoxifen resistance (Figure 2.4). In the context of the 8p11-p12 amplicon, ERα is targeted to FOXA1 binding sites in addition to EREs (Irish *et al.* 2016), resulting in downstream transcription of a suite of genes correlating with poor prognosis in breast cancer (Figure 2.5-C). Knockdown of NSD3 and ESR1 identified altered expression of genes associated with cell cycle processes (Table 2.2), suggesting that one of the roles of NSD3-S is to mediate ERα activity and estrogen dependence.

The mechanism of NSD3-induced overexpression and estrogen-independent activation of ERα is poorly understood. Since the effect is mediated primarily by NSD3-S, investigation of the role of NSD3-S in general will elucidate possible effects on ERα.

Shen *et al.* (2015) identified NSD3-S as an adaptor protein between BRD4, a member of the bromodomain and extraterminal domain (BET) family of chromatin readers, and CHD8, a chromodomain helicase DNA binding protein involved in transcriptional regulation via nucleosome remodeling. BRD4 is tethered to chromatin by acetylated histones and is essential for maintenance of leukemia through interactions with other proteins. Identification of NSD3-S as an adaptor protein that aides BRD4 interactions with these other factors is a novel finding that has implications for breast cancer treatment. Preliminary studies in our laboratory indicate that SUM-44 cells may be sensitive to the BRD4 inhibitor JQ1 and that NSD3 binds BRD4 by co-IP (data not shown). BET inhibitors are currently under investigation in several cancer types (French *et al.* 2014), and inclusion of NSD-S as a potential biomarker for patient response would be an exciting arm of investigation.

In addition to its role as a linker protein, NSD3-S has been hypothesized to inhibit histone and non-histone protein methylation. NSD3-S retains the N-terminal PWWP domain and maintains the ability to read methylated H3K36 residues, thereby occupying potential binding sites of the NSD3-L isoform (Stec *et al.* 2001). In addition to the implications of such a role for NSD3-S on gene transcription, this hypothesis is applicable to ERα activation. In estradiol-mediated activation of ERα, the protein is de-methylated and acetylated at the K266 residue (Zhang *et al.* 2013). In the context of NSD3-S amplification and overexpression, NSD3-S may bind ERα-K266 and prevent methylation, thereby leaving the protein susceptible to activation by acetylation, perhaps by another amplified and overexpressed 8p11-p12 amplicon oncogene, KAT6A, a HAT enzyme. The potential influence of NSD3-S on ERα PTM could be investigated by ERα proteomic analysis following knockdown of NSD3-S. Similarly, ERα is activated by

several phosphorylation events and any potential effects of NSD3-S on ERα

phosphorylation could be identified by the same technique.

Since ERα is a receptor that is activated by ligand-stimulated dimerization, it is

possible that NSD3-S does not affect the ERα protein itself and, instead, ERα is

activated in an estrogen-independent manner by NSD3-induced overexpression leading

to spontaneous dimerization. ERα is known to have a ligand-independent AF-1 domain

(Cui *et al.* 2005), therefore abundant expression of the receptor as is observed with

NSD3 amplification and overexpression may be sufficient to result in estrogen-

independent ERα activity. AF-1 is known to be involved in transcription of a suite of

genes different from those mediated by the ligand-dependent AF-2 domain (Wang and

Yin 2015), consistent with ERα ChIP-seq and microarray results that demonstrated

differential binding and expression profiles in 8p11-p12 amplicon-bearing SUM-44 cells

compared to amplicon-null MCF7 cells (Irish *et al.* 2016). Testing the hypothesis that

abundant overexpression of ERα is sufficient to induce estrogen-independent activation

would require overexpression of the receptor in MCF7 cells and assessing ERα activity

by reporter assay and RT-PCR for ERα target genes in the absence of estrogen. If these

experiments indeed demonstrate estrogen-independent ERα activity, we would then

predict that these MCF7 cells induced to recapitulate the ERα expression levels

observed in SUM-44 cells would show tamoxifen resistance *in vitro* but would retain their

sensitivity to fulvestrant.

If the estrogen-independent activation of ERα is due to its overexpression rather

than PTM, then we would hypothesize that the effect of NSD3 overexpression to induce

ERα overexpression reported here is due to epigenetic modifications by NSD3 that lead

to increased transcription of ESR1. Since NSD3-S is implicated in chromatin modification as an adaptor protein and NSD3-L is implicated in H3K36me3 in gene bodies of actively transcribed genes, analysis of the epigenetic landscape at the ESR1 gene with knockdown of NSD3-S and NSD3-T would identify the relationship between these two isoforms of NSD3 and their transcriptional effects on ESR1. Indeed we have shown that knockdown of NSD3 reduced ESR1 transcript and ERα protein levels (Figure 2.2), but the precise temporal relationship between this events has yet to be determined, therefore it remains unclear whether NSD3 knockdown first affects ESR1 transcription or ERα protein stability/activity/PTM that then results in a positive feedback loop whereby activated ERα is able to increase its own expression by its actions as a transcription factor. Other ERα co-factors and modifying proteins that may be involved in this process also remain yet unknown. Clearly, there is still a great deal of investigation necessary to elucidate the mechanism of ERα overexpression in the context of NSD3 amplification and overexpression.

The discovery of the relationship between overexpression of the 8p11-p12 amplicon oncogene NSD3 and overexpression and estrogen-independent activation of ERα has exciting implications for the treatment of amplicon-bearing luminal B breast cancers. We have demonstrated that these cells, though tamoxifen resistant (Figure 2.4-E), respond to the SERD fulvestrant (Figure 2.5-E and Figure 2.5-F) and treatment with fulvestrant recapitulates the cytotoxic effects of ESR1 knockdown in this cell line. The link between NSD3 and ERα described here is a novel mechanism of endocrine resistance in ER+ breast cancer.

In addition to breast cancer, the 8p11-p12 amplicon is present in several other cancer types (Gao *et al.* 2013, Mahmood *et al.* 2013). NSD3 is a validated oncogene in lung squamous cell carcinoma (LSCC), bladder carcinoma, and AML (Mahmood *et al.* 2013, Shen *et al.* 2015). These cancers are not ERα-dependent as is the case in breast cancer, and we demonstrated that two LSCC cell lines with high expression of NSD3 have undetectable levels of ERα expression (Figure 2.1-A and Figure 2.3-A). As such, we hypothesize that NSD3 has an important role in tumorigenesis outside of its effects on ERα. NSD3 is the least studied member of the NSD family of SET domain HMT proteins and the effects of NSD3 and its substrate, H3K36me2, remain understudied in cancer. In order to investigate the genes under epigenetic regulation by NSD3, we attempted to perform ChIP-seq analysis in SUM-44 cells utilizing an NSD3 antibody developed by the Vakoc group at the Cold Spring Harbor Laboratory (Shen *et al.* 2015), but were unsuccessful. The H3K4me3 ChIP-seq performed in this study in the context of ASH2L knockdown demonstrated the potential efficacy of employing a similar strategy to investigate the effects of NSD3 on H3K36me2 in the SUM-44 cell line as a future direction of this project.

Figure 4.1 illustrates the three proposed mechanisms of NSD3 oncogenesis. Panel A depicts the epigenetic role of NSD3-L, which catalyzes H3K4me2 in gene bodies to promote active transcription. We hypothesize that, in addition to a suite of genes involved in regulation of the cell cycle, NSD3 promotes transcription of ESR1, thereby inducing ERα overexpression. Since NSD3-S lacks the catalytic SET domain, its mechanism for promoting tumorigenesis remains largely unknown. Figure 4.1-B depicts a possible role for NSD3-S based on the model proposed by Shen *et al.* (2015). Here, NSD3-S acts as an adaptor protein for recruitment of additional chromatin-modifying

enzymes and transcription factors, thus affecting the transcriptome. This mechanism relies on chromatin binding by BRD protein family, which recognize acetylated histones (Lo and Sukumar 2008). We have previously hypothesized that the 8p11-p12 amplicon oncogene KAT6A is responsible for histone acetylation in this context. Alternatively, anchoring of NSD3-S to the chromatin by the BRD family proteins may act to prevent binding by NSD3-L, thus preventing H3K36me2 in those gene bodies. The exact mechanism and downstream implications of possible NSD3-S inhibition of NSD3-L remain to be elucidated. Finally, we have also hypothesized that NSD3-S binds K residues on the ERα protein itself, preventing its methylation and rendering it susceptible to activation by acetylation (Figure 4.1-C). These different mechanisms of action are not mutually exclusive and all potentially influence cell cycle progression, tumorigenesis, and endocrine resistance and therefore extension of the work reported here should explore each of these hypotheses.

c. ASH2L

ASH2L is understudied in breast cancer despite having been identified by several groups as a potential driving oncogene from the 8p11-p12 amplicon region (Garcia *et al.* 2005, Kwek *et al.* 2009, Cornen *et al.* 2014). Here, we identified a role for ASH2L in the regulation of H3K4me3 in promoter regions and expression of genes associated with cell cycle (Chapter 3). Knockdown of ASH2L resulted in decreased expression of an important set of genes, reduced proliferation and clonogenicity, and reduced sensitivity to the cell cycle inhibitor palbociclib. Together, the results of this study identify ASH2L as a driving oncogene in ER+ breast cancer and as a potential biomarker for palbociclib sensitivity in endocrine resistant tumors, a therapy currently in clinical trials for the treatment of ER+ breast cancer. Further investigation of ASH2L as a biomarker for drug

response and as a potential drug target itself could identify new therapeutic strategies to

increase survival in patients with the 8p11-p12 amplicon.

ASH2L lacks the catalytic SET domain characteristic of HMT enzymes (Ikegawa

et al. 1999). While this oncogene does possess methyltransferase activity when coupled

Figure 4.1. Summary model of NSD3 functions. A Representation of the role of NSD3-L in placing H3K36me2 in gene bodies in a BRD2- or BRD4-dependent manner. ESR1 and cell cycle genes are included in the regulated gene set based on data reported in chapter 2. **B** Depiction of the hypothesized role for NSD3-S in interference with H3K36me2 and/or recruitment of additional chromatin-modifying proteins/transcription factors in a BRD2- or BRD4-dependent manner. **C** Representation of ERα activated by acetylation and phosphorylation. The hypothesized functions of NSD3 depicted by panels **A-C** all lead to the transcriptional upregulation, tumor progression, and endocrine resistance reported and described in chapter 2.

with its binding partner RbBP5 (Tan *et al.* 2009), ASH2L commonly relies on the

formation of complexes with SET domain-containing proteins, such as the MLL family of

HMTs, where the role of ASH2L lies primarily in promoting catalysis of H3K4me3

(Steward *et al.* 2006). ASH2L is present in all known complexes responsible for

H3K4me3 and is tightly and specifically linked to this histone methyl mark (Butler *et al.*

2017). Here, we demonstrated that knockdown of ASH2L reduced the robustness of

global H3K4me3 peaks across the genome in SUM-44 cells (Figure 3.2-E). Remaining

H3K4me3 is likely a result of incomplete knockdown of ASH2L, which was not achieved

due to the extreme level of ASH2L overexpression in the context of its amplification

(Figure 3.2-B). Despite the reduction in robustness, ASH2L knockdown had little effect

on the overall distribution of H3K4me3 peaks (Figure 3.2-D). We were, however, able to

identify a major role for ASH2L in the elimination of H3K4me3 peaks specifically in the

promotor regions of genes (Figure 3.3-A), a function in agreement with the previous

studies on ASH2L-directed H3K4 tri-methylation (Demers *et al.* 2007).

The complex in which ASH2L functions in the context of the 8p11-p12 amplicon

is yet unknown and investigation of the protein(s) assisting ASH2L would be an

interesting future direction of this project. It is possible that ASH2L is forming a complex

with the MLL family as is well-documented, but with the extremely high level of

amplification-induced overexpression of ASH2L it is also possible that novel binding

partners are favorable in this context. Since ASH2L lacks a catalytic SET domain, the

direct versus indirect role of ASH2L on histone methylation is unclear. The effects of

ASH2L on H3K4me3 elucidated in this study may be a reflection of indirect effects via

the influence of ASH2L on binding partners. Indeed, we demonstrated here that the

ASH2L binding partner WDR5 has reduced promoter H3K4me3 peaks and transcript

expression upon ASH2L knockdown (Tables 3.1 and 3.2). WDR5 is responsible for chromatin binding, recruitment of ASH2L and other MLL complex proteins, and stability of the ASH2L-MLL complex (Steward *et al.* 2006, Guccione *et al.* 2007, Ali and Tyagi 2017). Identification of regulation of WDR5 expression by ASH2L reveals a feedback loop by which ASH2L may increase its chromatin-binding ability. Lack of corresponding evidence of ASH2L-mediated regulation of MLL family members suggests that ASH2L is relying on alternative binding partners for its function in the context of amplification-induced overexpression in breast cancer. Suspected ASH2L binding partners, such as other SET domain-containing proteins, could be identified by ASH2L IP followed by immunoblotting for such factors, although novel ASH2L binding proteins would not be identified by this method and would require additional techniques such as proteomic analysis.

According to the ChIP-seq and RNA-seq analyses presented in Chapter 3, ASH2L mediates H3K4me3 at its own promoter as well as several other chromatin-modifying enzymes such as NSD3 and EZH2, suggesting that ASH2L amplification establishes a positive feedback loop whereby amplified ASH2L drives expression of itself, leading to increased transcription of cell cycle genes and chromatin-modifying enzymes that also alter downstream gene expression. Similarly, ASH2L knockdown abolished H3K4me3 in the promoter regions of over half of the candidate oncogenes from the 8p11-p12 amplicon region (Table 3.1), suggesting that ASH2L may have a role in modulating the expression of these amplified genes. Of these amplicon genes, however, RNA-seq analysis identified only POMK and NSD3 as transcriptionally downregulated upon knockdown of ASH2L but also identified the 8p11-p12 amplicon gene KAT6A, another epigenetic factor, as being under ASH2L transcriptional regulation

without corresponding modification of the H3K4me3 status of its promoter region (Table 3.2). Further investigation of the relationship between promoter H3K4me3 and transcription is required. Epigenetic regulation is complex and other chromatin-modifying or epigenetic reader enzymes may be altered in the context of ASH2L knockdown to modify the histone code, accounting for the discrepancy between H3K4me3 and transcription.

The ChIP-seq and RNA-seq datasets provide a plethora of avenues for further investigation of the role of this oncogene in cell and tumor biology. Importantly, the analysis presented in Chapter 3 focused solely on genes which lost H3K4me3 promoter tri-methylation with a corresponding reduction in gene expression upon ASH2L knockdown. Genes with increased expression following ASH2L knockdown may represent tumor suppressors that are reduced when ASH2L is amplified and overexpressed. Genes that gained H3K4me3 following knockdown of ASH2L may represent an interesting group of factors, especially considering other groups have reported ASH2L knockdown abolishes H3K4 tri-methylation as ASH2L is required to promote catalysis of this reaction. Of note, we did not recapitulate the global reduction in H3K4me3 upon ASH2L knockdown that other studies have demonstrated, indicating that either ASH2L is overexpressed to such a degree in SUM-44 cells that residual levels are sufficient to maintain H3K4me3 methyltransferase activity or that another factor is compensating for ASH2L reduction in the context of 8p11-p12 amplification. Identification of an HMT with H3K4 tri-methylation capabilities independent of ASH2L would be a novel finding in the field.

Finally, immunoblots designed to validate the ASH2L RNA-seq knockdown results identified several proteins with unchanged expression despite message-level

downregulation (Figure 3.4-C), suggesting a disconnect between transcriptional

regulation and ultimate expression of several key factors. The complex nature of

epigenetic alterations, transcriptional effects, and protein expression and stability leave

several questions unanswered from these studies. The data indicate ASH2L influences

H3K4me3 in the promoter of NSD3 and knockdown of ASH2L reduces NSD3 expression

(Figure 3.3-C, Figure 3.5-A, and Table 3.2), therefore we hypothesize that amplification

of ASH2L leads to overexpression of co-amplified NSD3, which then influences ESR1

and ERα expression and activity (Chapter 2), as well as cell cycle genes and ASH2L

itself, establishing a positive feedback loop. This hypothesis is illustrated by Figure 4.2-

A. More investigation into the complex biology of 8p11-p12 amplified breast tumors is

required to fully understand the implications of ASH2L and its direct and indirect impact

on breast cancer progression and therapy resistance.

d. Cooperating oncogenes and ERα

One hypothesis generated from this study is that ASH2L and NSD3 cooperate to

alter epigenetic regulation of gene transcription and expression and activity of ERα in the

context of 8p11-p12 amplified breast cancer. We demonstrated that knockdown of

ASH2L abolished H3K4me3 in the promoter of NSD3 (Figure 3.3-C and Table 3.3),

reduced NSD3 and ESR1 transcript levels (Table 3.2), and decreased NSD3 and ERα

protein expression (Figure 3.5-A and Figure 3.6-A) in SUM-44 cells. Knockdown of

NSD3 reduced ESR1 and ERα expression (Figure 2.2). From these data we conclude

that ASH2L regulates NSD3 expression via promoter H3K4me3 and NSD3 then

regulates expression and estrogen-independent activation of ERα. Further study is

necessary to determine if ASH2L is directly involved in overexpression and/or activation

of ERα and whether these two oncogenes, ASH2L and NSD3, cooperate to modify the chromatin state in breast tumors in which they are amplified and overexpressed.

As discussed above, the complex in which ASH2L is found is essential to its function. Since ASH2L lacks the catalytic SET domain of other HMT enzymes, recruitment of SET proteins is essential to ASH2L function (Dou *et al.* 2006). NSD3-L possesses a SET domain that primarily catalyzes H3K36me2 in gene bodies and is associated with active transcription (Li *et al.* 2009). NSD3-L has also been implicated in H3K4 methylation, but this histone mark has not been well-characterized for this enzyme (Jacques-Fricke and Gammill 2014). IP of ASH2L followed by immunoblot for NSD3 would determine whether ASH2L and NSD3 form a complex and, if so, which isoform of NSD3 is involved. These proteins do not necessarily need to bind one another, however, in order to coordinately regulate gene expression (Figure 4.2-C). ASH2L NSD3-S lacks a catalytic SET domain like ASH2L yet displays oncogenic characteristics when overexpressed in non-transformed breast cells (Yang *et al.* 2010). These two oncogenes may therefore provide similar functions in breast cancer cells outside of histone methylation. Since the stability and activity of several important proteins, ERα included, depends on PTM such as methylation (Zhang *et al.* 2013), global proteomic analysis of the methylome in breast cancer cell lines following knockdown of either ASH2L or NSD3-S would be an interesting future direction of this project.

In order to interact with chromatin, NSD3 requires the assistance of chromatin reader proteins such as those of the BET family (Wagner and Carpenter 2012). Other studies have identified BRD4 as an NSD3 binding partner, but other BRD family proteins may also be involved (Shen *et al.* 2015). BRD2 was identified here by H3K4me3 ChIP-seq and RNA-seq with ASH2L knockdown, NSD3-S knockdown array, and ESR1

knockdown array (Table 3.3). Immunoblot analysis following knockdown of ASH2L in SUM-44 and MCF7 cells did not reveal a change in BRD2 expression, however (Figure 3.4-C). Whole cell lysate was harvested for the blots shown in Figure 3.4-C seven days post infection with ASH2L shRNA constructs, therefore longer incubation may be required to detect changes in protein expression. Alternatively, a disconnect between promoter H3K4me3, transcription, and protein expression may be present in these cells and investigation of factors involved in translational regulation would be an interesting future direction of this project. One such factor, EIF4EBP1, is found on the 8p11-p12 amplicon, is amplified and overexpressed in SUM-44 cells, and is implicated in translational dysregulation in cancer (Karlsson *et al.* 2011). This factor may mediate the proteomic effects of transcriptional dysregulation by ASH2L, NSD3, and ERα.

While we discovered a direct link between knockdown of NSD3-S and reduced expression of ESR1 and ERα (Figure 2.2), we were not able to induce overexpression of ERα in the amplicon-null MCF7 cell line simply by overexpression if NSD3-S (Figure 2.3-E). Successful induction of ERα overexpression in the amplicon-positive cell lines CAMA-1 and SUM-52 suggests that other amplicon components are required for overexpression and activity of ERα in this setting. IP followed by immunoblot for ERα identified a potential interaction between these two proteins (Figure 3.5-F). The MLL complex, which includes ASH2L, is known to interact directly with ERα and ERα relies on histone- and chromatin-modifying enzymes to promote transcription at ERα-target genes (Cui *et al.* 2005, Turner-Ivey *et al.* 2014). Therefore, we hypothesize that ASH2L and ERα cooperate to alter gene transcription in the context of 8p11-p12 amplification, including increased transcription of ESR1 (Figure 4.2-B). Additional co-IP assays are

required to validate the interaction between ASH2L and ERα and close study of the

transcriptional effects of these two factors would be necessary to explore this

hypothesis.

The complex nature of amplification-induced overexpression and subsequent

activation of oncogenes, epigenetic modifications, transcriptional regulation, and ERα-

Figure 4.2. Summary model of ASH2L functions. A Representation of the role of ASH2L described in chapter 3 whereby ASH2L is responsible for placing transcriptionally-activating H3K4me3 in promoter regions of genes, resulting in transcription of 8p11-p12 amplicon genes, establishing a positive feedback loop of ASH2L auto-upregulation, and leading to the overexpression of NSD3, which then leads to overexpression and estrogen-independent activation of ERα as described in chapter 2. **B** Depiction of the hypothesized role for ASH2L in cooperating with ERα to regulate transcription. **C** Diagram of the hypothesized role of NSD3 in cooperation with ASH2L to affect transcription via modification of chromatin structure either at the same genes under ASH2L promoter regulation or a complementary gene set unique to NSD3. The hypothesized functions of ASH2L and NSD3 depicted by panels **A-C** all lead to the transcriptional upregulation, tumor progression, and endocrine resistance reported and described in chapter 2 and 3.

signaling renders study of individual genomic alterations difficult. This study attempted to elucidate the roles of two important oncogenes, ASH2L and NSD3, and characterize the transcriptional effects of these two epigenetic factors. There are a multitude of pathways to investigate stemming from these studies, some of which are discussed here. Ultimately, the goal of these *in vitro* studies is to identify novel mechanisms of drug treatments for patients with 8p11-p12 amplicon-bearing ER+ breast cancer. The implications of these studies on breast cancer management are discussed below.

4.2. ER+ Breast Cancer Treatment

a. Endocrine therapies

While tamoxifen and the aromatase inhibitors have made great strides in the treatment of ER+ breast cancer, not all patients with ER+ disease respond to these drugs. Mechanisms of endocrine resistance have long been the subject of investigation, and several have been identified. First, loss of ERα expression can occur (Cui *et al.* 2005). These cells are no longer dependent upon the receptor for growth and survival, therefore endocrine therapies are irrelevant to these processes, much like tumors that are ER-negative upon diagnosis. Second, point mutations in the ligand-binding domain of ERα can abrogate the ability of tamoxifen to bind and activate the ligand-independent AF-1 domain of ERα (Chang 2012, Angus *et al.* 2016). This mechanism renders AIs ineffective as well since ERα is no longer dependent upon estrogen for activation and is capable of signaling through the ligand-independent AF-1 domain. Third, amplification of ESR1, though rare, does occur (Tabarestani *et al.* 2016). This mechanism is not yet well-characterized, but does correlate with endocrine resistance. Finally, we demonstrate here a novel mechanism of endocrine resistance: overexpression and

estrogen-independent activation of ERα by the 8p11-p12 amplicon oncogene NSD3, perhaps in cooperation with other amplicon-oncogenes (Chapter 2). This mechanism also requires further investigation, as discussed above, but highlights again the diversity of breast tumors and the genomic alterations that can have vast effects on drug sensitivities, underscoring the need to better characterize luminal B breast cancers to identify novel treatment strategies to overcome endocrine resistance.

Several drugs have undergone clinical trials for the treatment of endocrine-resistant, ER+ breast cancer. Fulvestrant was developed in an attempt to expand the SERM repertoire to pure ERα antagonists (Howell 2006). Fulvestrant (ICI 182,780; Faslodex) binds ERα and induces a conformational change different than that of tamoxifen, preventing dimerization and targeting the receptor for degradation by the proteasome system (Dowsett *et al.* 2005). Fulvestrant binding also inactivates both the AF-1 and AF-2 domains of ERα, unlike tamoxifen, completely abolishing ERα-mediated gene transcription (McKeage *et al.* 2004). Since the mechanism of action of Fulvestrant is primarily due to degradation of ERα, cross-resistance with tamoxifen and anastrozole is minimal (De Marchi *et al.* 2016). Several clinical trials assessing efficacy and dosing of Fulvestrant led to its approval in 2010 for the treatment of post-menopausal women with ER+ breast cancer nonresponsive to other endocrine therapy (Ignatiadis and Sotiriou 2013).

While Fulvestrant has been a major step forward in breast cancer treatment, the ability of this drug to degrade ERα is dose-dependent and achieving steady-state doses adequate for tumors with high-level ERα expression is a major clinical challenge (Robertson 2007). Several other SERDs are under development and testing, but interest

in this drug class has diminished due to apparent resistance in patients. This so-called

resistance, however, may be due the dose-dependent nature of Fulvestrant, reflecting

inadequate bioavailability. The work presented here is an important step toward shifting

the paradigm regarding ERα expression and endocrine therapy resistance. Most luminal

B tumors have relatively low ERα expression and it has long been thought that these

tumors are not dependent upon the receptor for growth and survival, thereby accounting

for the overall poor response to endocrine therapies (Ades *et al.* 2014). Here, we

demonstrate that high-level ERα expression does occur in patient tumors with endocrine

resistance, this receptor is a driving oncogene in this setting, and abolishing ERα

expression is detrimental to these cancer cells (Chapter 2). The data presented in this

work support continued development of the SERD drug class and underscore the need

for such a drug that can effectively eliminate high ERα overexpression in patients.

b. Signaling pathway therapies

In addition to therapies targeting ERα signaling, several other pathways have

been implicated in ER+ breast tumors. The "non-genomic" effects of ERα have been

described in detail in several review articles, all concluding that cytoplasmic and/or

membrane-bound ERα cross-talks with growth factor receptor pathways such as EGFR

(McKeage *et al.* 2004, Dowsett *et al.* 2005, Johnston *et al.* 2005, Kristensen *et al.* 2005,

Nicholson and Johnston 2005, Ariazi *et al.* 2006, Howell 2006, Chang 2012, Fedele *et al.* 2012, Ignatiadis and Sotiriou 2013, Kerdivel *et al.* 2013, Nagaraj and Ma 2015, Wang

and Yin 2015, Tabarestani *et al.* 2016). In breast cancer, ERα expression (ERα-66) is

essentially exclusively nuclear, as demonstrated by IHC staining of patient biopsies and

ERα ChIP experiments identifying chromatin-bound ERα prior to treatment with 17β-

estradiol (Cui *et al.* 2005, Howell 2006, Angus *et al.* 2016, Irish *et al.* 2016). Smaller ERα isoforms, most commonly ERα-36 and ERα-46, have been described as mediating these non-genomic ERα activities and the sensitivity of these isoforms to endocrine therapy is unknown, though both retain the DNA-binding and ligand-binding domains (Wang and Yin 2015). Additionally, the role of ERβ in breast cancer is not fully characterized (Hartman *et al.* 2009). Despite the questions that still remain surrounding non-genomic ERα signaling, it is known that cross-talk with cell signaling pathways does occur in the form of post-translational modifications.

The PI3K/AKT/mTOR pathway is known to activate ERα by phosphorylation of several serine residues, including S118, S167, and S305 (Kerdivel *et al.* 2013, Zhang *et al.* 2013, Steelman *et al.* 2016). The EGFR pathway can phosphorylate ERα tyrosine residue Y537 and p38/MAPK is implicated in threonine phosphorylation on ERα-T311 (Cui *et al.* 2005, Kerdivel *et al.* 2013). Several drugs targeting these pathways have been tested in ER+ breast cancer, but success is limited and biomarkers of response are unknown (Morris and Wakeling 2002, Fedele *et al.* 2012). Since the PI3K/AKT/mTOR pathway was identified as possibly regulated by ASH2L (Chapter 3), we hypothesized that inhibition of mTOR would alter ERα phosphorylation, possibly identifying the mechanism behind estrogen-independent activation of ERα in the SUM-44 cell line. Reduced PI3K expression by ASH2L knockdown was confirmed by immunoblot but AKT1 was unchanged (Figure 3.4-C) and immunoblot for ERα-pS167 following treatment with the mTOR inhibitor KU-00063794 did not identify reduction of phosphorylation at this site (Figure 3.5-E), therefore the mechanism of estrogen-independent ERα activation remains unknown.

Another strategy for overcoming endocrine resistance is to target genes downstream of ERα (Mancuso and Massarweh 2016). ERα drives cell cycle progression from G_1 to S phase through activation of its transcriptional program (Wang and Yin 2015). CCND1 (cyclin D1) is an ERα-target gene that mediates G_1 to S phase transition by binding to CDK4 or CDK6 and stimulating transcription (Dabydeen and Furth 2014). Cyclin D1 is associated with several types of cancer, including breast, and is correlated with endocrine resistance (Dabydeen and Furth 2014, Nagaraj and Ma 2015). As discussed in section 1.2b, amplification of CCND1, located at 11q13, occurs in a subset of breast cancer patients and the 11q12-q14 amplicon often co-occurs with the 8p11-p12 amplicon (Still *et al.* 1999, Reis-Filho *et al.* 2006, Kwek *et al.* 2009, Karlsson *et al.* 2011, Sircoulomb *et al.* 2011, Bilal *et al.* 2012, Cornen *et al.* 2014). Cooperation of 11q14 oncogenes with oncogenes from the 8p11-p12 amplicon has been suggested (Kwek *et al.* 2009, Karlsson *et al.* 2011) and the data presented here supports this hypothesis. The CDK4/6 inhibitor palbociclib is in clinical trials in ER+ breast cancer and is demonstrating increased progression free survival in conjunction with endocrine therapy in luminal breast cancers (Nagaraj and Ma 2015, Turner *et al.* 2015, Angus *et al.* 2016, Knudsen and Witkiewicz 2016, Mancuso and Massarweh 2016). Genes with reduced H3K4me3 promoter peaks and transcript expression following ASH2L knockdown in SUM-44 cells annotated to the "palbociclib response" with a p-value of 5.7_E-7, suggesting that knockdown of ASH2L reduces the expression of a suite of genes that sensitize cells to palbociclib (Figure 3.4-E). We tested response to palbociclib in the SUM-44 and MCF7 cell lines following knockdown of ASH2L and validated that ASH2L knockdown reduces palbociclib response in SUM-44 cells (Figure 3.6-E). MCF7 cells were more sensitive to palbociclib than SUM-44 cells and ASH2L knockdown did not

alter sensitivity. Interestingly, CCND1 was identified as downregulated by knockdown of ASH2L (Table 3.2), yet immunoblot for cyclin D1 expression demonstrated increased cyclin D1 protein levels in SUM-44 cells and no change in MCF7 following knockdown of ASH2L (Figure 3.4-C). MCF7 cells do not have either the 8p11-p12 or 11q12-q14 amplicons present in SUM-44 cells, underscoring the complexity of luminal B breast cancer genomic alterations and the potential interaction between multiple oncogenes affecting the biology and drug sensitivities of these tumors.

Although therapies targeting signaling pathways that complement ERα activity have demonstrated some improvement in the treatment of ER+ breast cancer, resistance and lack of biomarkers are still common clinical challenges. The work presented here identified NSD3 as a potential biomarker for SERD response and ASH2L as a potential biomarker for palbociclib response in endocrine-resistant tumors. The 8p11-p12 amplicon represents not only a source of therapy response markers, but understanding the mechanism of action of the oncogenes present in this genomic region provides the potential for development of novel therapies that will affect the genomic and epigenomic events responsible for tumorigenesis. Indeed, epigenetic therapies are under clinical investigation and have demonstrated some success in the treatment of breast and other cancer types (Kerdivel *et al.* 2013, Ades *et al.* 2014). This work on NSD3 and ASH2L provides additional evidence that can be utilized for the development and novel therapies and improvement of those already under clinical investigation.

c. Epigenetic therapies for ER+ breast cancer

Unlike genetic alterations, epigenetic changes leading to cancer are reversible (Lo and Sukumar 2008, Simon and Lange 2008, Kristensen *et al.* 2009). Coupled with

increased understanding of the histone modifying enzymes dysregulated in cancer and their downstream effects, the potential therapeutic benefits of targeting these factors is becoming apparent (Morishita and di Luccio 2011). Of the epigenetic drugs, DNMT and HDAC inhibitors are the best-studied. Use of these drugs has identified the need to better understand optimal timing of epigenetic therapy, appropriate biomarkers, and specificity of chromatin-modifying enzymes for tailored therapeutic benefit (Lustberg and Ramaswamy 2009, Pathiraja *et al.* 2010, Cai *et al.* 2011).

HDAC inhibitors (HDACi) have been explored in several disease states, including ER+ breast cancer (Thomas and Munster 2009). HDACi are implicated in cell cycle arrest by nuclear export of cyclin D1, which interacts with histone modifying enzymes to execute its transcriptional program, and by favoring hyperacetylation of genes, including pro-apoptotic genes, leading to their expression (Alao *et al.* 2004, Inoue and Fry 2015). HDACi themselves do not discriminate transcription of pro- versus anti-apoptotic genes, leaving certain tumors exposed to the possibility of HDACi-induced accelerated growth until biomarkers can be established to predict this differential response (Lo and Sukumar 2008, Lustberg and Ramaswamy 2009). HDAC inhibition also re-sensitizes cells to endocrine therapies by interrupting the interaction of HDACs with DNMTs, reversing the DNA methylation-induced silencing of ESR1 (Allred *et al.* 2004, Kristensen *et al.* 2009, Thomas and Munster 2009, Abdel-Hafiz and Horwitz 2015). This mechanism, however, only affects tumors where the mechanism of endocrine resistance is loss of ERα expression. Alternatively, HDACi have been implicated in hyperacetylation of the chaperone protein HSP90 (Lo and Sukumar 2008, Lustberg and Ramaswamy 2009). ERα is one of the many client proteins of HSP90, so hyperacetylation that destabilizes HSP90 results in degradation of ERα. HDACi, however, can also alter the chromatin

structure, allowing histone acetylation-mediated opening of genetic material for increased transcription by ERα (Margueron *et al.* 2004).The multiple and potentially conflicting mechanisms by which HDACi function highlight the need to better understand epigenetic factors and tailor inhibition toward remedy of specific dysregulated processes.

Clinical investigation of HDACi has provided an important step forward in the use of epigenetic therapy to treat cancer (Lustberg and Ramaswamy 2011, Connolly and Stearns 2012, Walsh *et al.* 2016). Since histone methylation is not as well-understood as DNA methylation and histone acetylation, HMT inhibition (HMTi) is not as developed as inhibition of other epigenetic enzymes (Morishita and di Luccio 2011). As the field of epigenomics progresses, however, the importance of this mechanism of gene regulation has become clear and HMTi is advancing (Morishita and di Luccio 2011, French *et al.* 2014, Liu *et al.* 2014). Several groups have used large databases such as GISTIC and TCGA to determine optimal factors for drug targeting and EZH2, MLL, and the NSD family of HMTs have been identified as some of the most ideal candidates (Liu *et al.* 2014, Michalak and Visvader 2016). EZH2 inhibitors have progressed furthest of the HMTi class, but even these drugs have only reached early stage clinical trials (Garapaty-Rao *et al.* 2013, Curry *et al.* 2015, Michalak and Visvader 2016, Sato *et al.* 2017). Other HMTi remain under preclinical investigation and require additional *in vitro* studies of HMT activity and mechanisms of action, such as those described in this work, for continued advancement of this field (Liu *et al.* 2014, Katoh 2016, Michalak and Visvader 2016).

EZH2 is a PcG protein implicated in H3K27 methylation-mediated transcriptional repression and is one of the best-studied HMT enzymes to date (Lo and Sukumar 2008, Pathiraja *et al.* 2010, Grzenda *et al.* 2011). It functions in polycomb repressive complexes (PRC) and is antagonized by the Trx group proteins such as ASH2L and the

MLL complexes (Simon and Lange 2008, Grzenda *et al.* 2011, Huang *et al.* 2011). HMT inhibitors (HMTi) have been primarily targeted to EZH2 and have been designed based on the paradigm that histone methylation is a negative regulator of tumor suppressors, and therefore inhibition of EZH2 and other PcG factors will alleviate this repression and halt cell cycle progression (Curry *et al.* 2015, Sato *et al.* 2017). There is some evidence to indicate that this strategy may be efficacious as EZH2 inhibitors indeed demonstrate reduced growth in tumor cell lines *in vitro* (Garapaty-Rao et al. 2013). Unfortunately, these therapies are highly dose-dependent and require further investigation to determine the precise timing for optimal treatment (Lustberg and Ramaswamy 2009, Garapaty-Rao *et al.* 2013). Several combination therapies have been tested in an attempt to circumvent these issues and have displayed more success than individual compounds, however these studies have also revealed the need to better understand the histone modifications at particular loci to optimize HMTi therapies as global inhibition is not necessarily the best strategy in all cases (Lo and Sukumar 2008, Curry *et al.* 2015, Sato *et al.* 2017). Several companies are developing additional HMT inhibitors currently, including specific EZH2 inhibitors as well as generalized SET domain inhibitors, which would not only target repressive histone methyltransferases but also transcriptional activators such as NSD3 and the MLL family proteins (Michalak and Visvader 2016).

Another strategy by which HMT activity can be targeted is via chromatin-interacting proteins such as those of the chromatin-reader bromodomain and chromodomain families. Many HMT enzymes, NSD3 and ASH2L included, rely on these reader proteins for recruitment to directed sites on the chromatin. Bromodomain inhibitors have been under development and testing and have shown promising results in the clinic (French *et al.* 2014, Borbely *et al.* 2015). These chromatin readers, however,

are promiscuous and broad inhibition of their activity has unknown implications. One advantage to this strategy of epigenetic inhibition is that non-SET domain-containing proteins, such as ASH2L and NSD3-S, are also inhibited in the manner. NSD3-S is known to rely on the BRD family of bromodomain proteins for chromatin tethering, and inhibition of BRD4 has demonstrated abrogated NSD3-S interaction with chromatin (Shen *et al.* 2015). NSD3-S bound to BRD4 also binds other factors, such as the chromodomain protein CHD8, a potential novel drug target in cancer (Shen *et al.* 2015). Another bromodomain protein, ANCCA, is implicated in recruitment of MLL complexes to histones (Zou *et al.* 2014). These chromatin-reading protein families are candidates for drug development and understanding their mechanism of action and effects on the epigenomic state of cancer cells is essential to optimizing their therapeutic benefits.

Epigenetic therapies represent a novel arm of therapeutic investigation and, though still in the early stages of development compared to many other cancer treatment strategies, comprise a promising source of compounds with the potential to revolutionize cancer treatment. The work presented here on NSD3 and ASH2L, both associated with transcriptional activation by histone methylation activity, challenges the current paradigm that epigenetic factors function as oncogenes by transcriptional repression of tumor suppressors. The antagonistic relationship of these trithorax group proteins on the polycomb group counterparts such as EZH2 is an interesting and extremely complicated topic of future investigation. Overall, it is clear that further investigation of the specific actions of epigenetic factors dysregulated in cancer is necessary in order to advance the field of epigenomics and identify efficacious strategies for targeting the biological effects of these alterations in patients.

4.3. Concluding Remarks

Breast cancer is the most common cancer in women and second leading cause of cancer death in women in the United States, resulting in the loss of 40,000 lives each year (Morris and Carey 2007, Chang 2012, 2014, Lumachi *et al.* 2015, Selli *et al.* 2016). Great strides have been made over the last four decades in the diagnosis and treatment of breast cancer patients, however there remains much progress to be made. Resistance to endocrine therapies is still a major clinical challenge in ER+ breast cancer and the need to understand mechanisms of resistance and develop novel therapeutics is urgent. Identification of intrinsic subtypes of breast cancer, classification of tumors based on expression profiles, and technological advancements coupled with the development of large databases tracking genomic alterations in primary breast tumors has allowed novel biological mechanisms of tumorigenesis to be elucidated and investigated. Similarly, the field of epigenomics has opened a novel avenue of investigation into tumor biology and identified a suite of potential therapeutic candidates and biomarkers in histone methyltransferase enzymes.

At the cross-road between genomic and epigenomic alterations in cancer lies amplification-induced overexpression of oncogenes that modify the histone code. The 8p11-p12 amplicon is found in 15% of primary breast tumors and harbors three epigenetic modifiers of chromatin, two of which, NSD3 and ASH2L, are implicated in regulation of histone methylation and transcriptional activation. NSD3 has been previously validated as an oncogene (Yang *et al.* 2010, Mahmood *et al.* 2013), however little was known about the biological role of the two isoforms of NSD3. This work established a link between amplification and overexpression NSD3 and overexpression and estrogen-independent activation of ERα (Figure 4.3). We have identified and

described here a novel mechanism of endocrine resistance in luminal B breast cancer and demonstrated a need for improved therapies that reduce receptor activity irrespective of estrogen status. Additional work is required to fully understand the implications of the non-catalytic short isoform of NSD3 and the mechanism behind NSD3-induced overexpression and activation of ERα.

ASH2L has been identified by several groups as a candidate oncogene from the 8p11-p12 region (Ray *et al.* 2004, Garcia *et al.* 2005, Gelsi-Boyer *et al.* 2005, Cornen *et al.* 2014), however this epigenetic factor has been understudied in cancer to date. The data presented here identify a role for ASH2L in regulation of H3K4me3 specifically in the promoter regions of genes associated with cell cycle progression, epigenetic regulation of chromatin, and palbociclib response. We also demonstrated that knockdown of ASH2L reduced expression of NSD3 and ERα, extending our understanding of the relationship between these oncogenes (Figure 4.3). It is clear from the data presented in this work and the unanswered questions that remain that the underlying biology of 8p11-p12 amplicon oncogene-mediated epigenetic dysregulation is complex. Further investigation of the transcriptional regulation of gene expression by NSD3 and ASH2L, cooperation between these oncogenes, effects on the proteome, and the mechanism of estrogen-independent ERα activation will provide insight into the essential biology of a significant subset of breast cancer patients at risk of succumbing to their disease due to endocrine therapy resistance.

Histone methyltransferases are increasingly recognized not just for their potential as drug targets but also as biomarkers for therapeutic response (Connolly and Stearns 2012, Chen *et al.* 2014, Michalak and Visvader 2016). NSD3 and ASH2L both have the

potential to be direct drug targets due to their oncogenic effects on transcription and cell

behavior, and these oncogenes and the 8p11-p12 amplicon also have the potential to be

utilized as biomarkers to predict treatment response to SERD and cell cycle therapies.

The work presented here provides a rationale to include survey of NSD3 and ASH2L

expression in clinical trials investigating these types of compounds. The downstream

Figure 4.3. Summary model. Schematic representation of the model whereby overexpressed ASH2L is responsible for upregulation of other 8p11-p12 amplicon oncogenes as well as genes involved in cell cycle progression and palbociclib response via placement of transcriptionally-activating H3K4me3 in promoter regions. ASH2L-mediated overexpression of NSD3 then leads to overexpression and estrogen-independent activation of ERα and the hypothesized roles for ASH2L in cooperating with NSD3 and/or ERα to regulate transcription are also depicted here, culminating in transcription of a suite of genes that promotes breast tumor progression. Red symbols indicate opportunities for pharmacological intervention in the processes presented in the model. Blue arrows indicate data presented in this manuscript, orange indicates relationships previously published, and green indicates unknown relationships. Dotted lines represent hypothesized relationships.

processes and pathways under control of these two epigenetic oncogenes can be targeted therapeutically, extending the future implications of this study. The work presented here provides the foundation for a multitude of additional lines of investigation and has the potential to be translated into novel therapeutic strategies to improve survival for an important group of patients with ER+ breast cancer.

CHAPTER 5: Materials and Methods

5.1. Cell culture

All cells were cultured at 37°C in 10% CO_2.

5.1a. SUM-44. Cells were cultured in serum-free Ham's F-12 medium supplemented with 1 μg/ml hydrocortisone, 1 mg/ml bovine serum albumin, 10 mM HEPES, 5mM ethanolamine, 5 μg/ml transferrin, 10 nM tiiodothyronine, 50 nm sodium selenite, 25 μg/ml gentamicin, 2.5 μg/ml fungizone, and 5 μg/ml insulin. For estrogen-free culture conditions, phenol red-free Ham's F-12 media was substituted for normal Ham's F-12 with all the above supplements. 3 μg/ml puromycin was used for selection and maintenance of shRNA-transduced SUM-44 cells.

5.1b. SUM-52. Cells were cultured in Ham's F-12 medium supplemented with 5% fetal bovine serum, 1 μg/ml hydrocortisone, 25 μg/ml gentamicin, 2.5 μg/ml fungizone, and 5 μg/ml insulin. 2 μg/ml puromycin was used for selection and maintenance of sRNA-transduced SUM-52 cells.

5.1c. MCF-7 and CAMA-1. Cells were cultured in DMEM supplemented with 10% fetal bovine serum, 25 μg/ml gentamicin, and 2.5 μg/ml fungizone. 2 μg/ml puromycin was used for selection and maintenance of sRNA-transduced MCF-7 and CAMA-1 cells.

5.1d. MCF10A. Cells were cultured in serum-free Ham's F-12 medium supplemented with 1 μg/ml hydrocortisone, 1 mg/ml bovine serum albumin, 10 mM HEPES, 5mM ethanolamine, 5 μg/ml transferrin, 10 nM tiiodothyronine, 50 nm sodium selenite, 25 μg/ml gentamicin, 2.5 μg/ml fungizone, 10 ng/ml EGF and 5 μg/ml insulin. 1 μg/ml puromycin was used for selection and maintenance of sRNA-transduced MCF10A cells.

5.1e. ZR-75-1. Cells were cultured in RPMI supplemented with 10% fetal bovine serum, 25 μg/ml gentamicin, and 2.5 μg/ml fungizone.

5.2 Antibodies

Antibodies used in these studies were diluted in 5% milk and used at a dilution of 1:1000 for immunoblot unless otherwise stated. Antibodies include:

NSD3 (WHSC1L1): Proteintech 11345-1-AP; Vakoc Lab antibody (Shen *et al.* 2015)

ERα: Cell Signaling 8644S, Santa Cruz HC-20 (1:50 dilution for ERα ChIP-PCR cocktail); NeoMarkers ER Ab-10 (TE111.5D11) (1:50 dilution for ERα ChIP-PCR cocktail)

pS167 ERα: Cell Signaling 5587S

ASH2L: Bethyl A300-489A; 1:1000 for immunoblot and 1:50 for IP

H3K4me3: Abcam ab8580; 1:1000 for immunoblot and 1:50 for ChIP-seq

FOXA1: Abcam ab109760

AKT1: Cell Signaling 2938S

BRD2: Cell Signaling 5848S

BRD4: Cell Signaling 13440S

Cyclin D1: Cell Signaling 2926P (courtesy Dr. Carroll Lab)

Ki-67: Dako A0047 (courtesy Dr. Carroll Lab)

PI3 Kinase: Abcam ab137815

PLK1: Abcam ab47867

IgG: Cell Signaling 3900S

β-actin: Cell Signaling 3700S

5.3. shRNA constructs

Lentiviral shRNA expression vectors from the pLKO shRNA catalog were purchased from Sigma. The vectors used were the NSD3-short vector TRCN0000415241, NSD3-total vectors TRCN000425711, TRCN0000015615, and TRCN0000015616, ASH2L

vectors TRC0000019275 and TRC0000019276, ESR1 vector TRCN0000003301 and LacZ control vector. Plasmid DNA for all constructs was obtained using the Promega PureYield plasmid midiprep system (A2495) according to manufacturer's instructions.

5.4. Lentiviral production and transformation

Lentivirus was prepared using Sigma Mission lentiviral packaging system (Sigma, shp-001) in 293FT packaging cells following manufacturer's instructions. Briefly, each construct was co-transfected into 293FT cells with Sigma pLKO shRNA vectors and Lipofectamine 2000 in antibiotic-free media. Virus was harvested 48 hours after transfection, filtered, aliquoted, and stored at -80°C until use. Target cells were transduced with virus in growth media supplemented with 5 µg/ml polybrene using 0.5-1.0 ml virus per 1 million cells. Cells were selected for resistance marker expression and maintained in puromycin-containing growth media (1-3 µg/ml as listed above) beginning 24-48 hours after infection. Cells were cultured at least 5 days to allow expression of the construct.

5.5. Immunoblotting

Cells were plated and growth to 75-90% confluency. Cells were treated as indicated, either with drug compounds or shRNA infection, for the specified timepoints. Cells were lysed in buffer containing 20mM Tris-HCl (pH 8.0), 137 mM NaCl, 1% NP40, 10% glucerol, 1 mM Na_3VO_4, and 1x protease inhibitor cocktail (Calbiochem 539131). Protein concentrations were measured by Bradford assay (Bio-Rad). Laemmli sample buffer was added to the lysates and the samples were boiled for 5 min 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°C with the indicated antibodies. Blots were washed 3x10 min in 1x tris-buffered saline with

tween 20 (TBST), probed 1 hour with the appropriate secondary antibody, incubated 1 min in ECL and developed on a Konica SRX-101A medical film processor. After developing, membranes were re-probed with a new primary antibody or discarded.

5.6. Proliferation assays

Proliferation assays were performed in 6-well culture plates seeded in triplicate with 200,000 cells per well. At each timepoint, cells were washed once with PBS and agitated on a rocker table with 0.5 ml HEPES/MgCl$_2$ buffer (0.01 M HEPES and 0.015 M MgCl$_2$) for 5 min. Cells were then lysed for 10 min with addition of 50 µl of an ethyl hexadecyldimethylammonium solution and nuclei were counted using a Z1 Coulter Counter (Beckman Coulter, Brea, CA, USA) according to manufacturer's instructions. Data are reported as averages where error bars represent +/- one standard deviation.

5.7. Drug treatment proliferation assays

Drug treatment assays were performed in 6-well culture plates seeded in triplicate with 250,000 - 500,000 cells per well. Cells were treated every 24 or 48 hours (2-5 treatments total) with the indicated drug compound and concentration. When control wells reached 90% confluence, cells were washed once with PBS and agitated on a rocker table with 0.5 ml HEPES/MgCl$_2$ buffer (0.01 M HEPES and 0.015 M MgCl$_2$) for 5 min. Cells were then lysed for 10 min with addition of 50 µl of an ethyl hexadecyldimethylammonium solution and nuclei were counted using a Z1 Coulter Counter (Beckman Coulter, Brea, CA, USA) according to manufacturer's instructions. Counts were normalized to the average of the untreated samples and are reported as averages where error bars represent +/- one standard deviation.

5.8. Colony forming assay

Cells were plated in triplicate in 6-well plates at varying clonal cell densities. MCF7 cells were plated at 5,000 cells per well and SUM-44 cells at 15,000 cells per well. MCF7 cells were cultured as normal. SUM-44 cells required 1:1 ratio of regular media to conditioned media from SUM-44 cells growing at high density. Puromycin-supplemented media was used in these assays to maintain selection of infected cells. Media was changed 24 hours after plating and then 7 days later. Colonies formed for two weeks and then were fixed with 1 ml/well 3.7% paraformaldehyde for 20 minutes at RT. Colonies were stained with 1 ml/well 0.2% crystal violet for 15 minutes at RT and de-stained with dH$_2$O. Colonies were allowed to dry prior to imaging, counting, and quantification on the GelCount System from Oxford Optronix.

5.9. qRT-PCR

Total RNA was harvested using a Qiagen RNeasy Plus Mini Kit (74136) according to manufacturer's instructions. RNA was quantified on the NanoDrop 2000c spectrophotometer and 2 ug RNA utilized for each cDNA reaction using the iScript Adv cDNA Kit for RT-qPCR (172-5038) according to manufacturer's instructions. Quantitative RT-PCR was then performed using 5 ng of cDNA per reaction in triplicate for each primer and sample pair in Roche FastStart Universal SYBR Green Master (Rox) mix (04913914001). Cycling conditions were 95°C for 10 min, then 40 cycles of 95°C 10 sec, 51°C 10 sec, 68°C 40 sec. GAPDH primers (IDT PrimeTime qPCR Primer set Hs.PT.39a.22214836) were used as control for each sample, analyzed by the ΔΔCt method (Livak and Schmittgen 2001). Values are reported as averages of triplicate wells where error bars represent +/- one standard deviation. Primer sets were designed using IDT custom parameters and conditions for each primer pair are as reported below.

Name	Sequence 5'-3'	Tm (°C)	%GC
ASH2L – F	CTCCGTGGATGAAGAGAA	51.1	50.0
ASH2L – R	GGTAGACAGGATGAGGTATC	51.2	50.0
ESR1 – F	GCTTCGATGATGGGCTTACT	54.9	50.0
ESR1 – R	CCTGATCATGGAGGGTCAAATC	55.3	50.0
FBXO5 – F	GCCTCAAAGCCTGTATTC	50.8	50.0
FBXO5 – R	CAAATCCACAGCCTTCTC	50.7	50.0
EZH2 – F	TGACTGCTTCCTACATCC	51.1	50.0
EZH2 – R	CTTTGCTCCCTCCAAATG	51.0	50.0
TTK – F	GTTGTGCCTGGATCTAAAC	51.0	47.4
TTK – R	CCAGAGGTTCCTTGAAATG	50.7	47.4
BUB1 – F	GGTTAATCCAGCACGTATG	50.8	47.4
BUB1 – R	ACTGGTGTCTGCTGATAG	51.1	50.0

5.10. Immunoprecipitation

Whole cell lysate was harvested using the buffer and protocol described under section

5.6 or in FAK buffer supplemented with 1 mM Na_3VO_4, and 1x protease inhibitor cocktail.

Samples were measured by Bradford assay (Bio-Rad) and diluted with lysis buffer to

1000 µg protein per reaction. Antibodies were added at 1:50 dilution and samples

incubated overnight at 4°C with rotation. Protein A/G beads were incubated with

lysate/antibody samples for 1 hour at RT with rotation prior to washing 3x in TBST with

1x protease inhibitor cocktail. Protein was eluted in 50 µl 1x laemmli sample buffer

(diluted to 1x with lysis buffer), boiled 5 min, and 25 µl loaded per lane on SDS-

polyacrylamide gels (Bio-Rad). Whole cell lysate samples were created in parallel from

the same samples prior to incubation with antibody and flow through was collected prior

to wash steps. IgG as well as targeted antibodies were used. Gels were transferred,

probed, and developed as described in section 5.5.

5.11. Microarray

RNA integrity was verified on an Agilent 2200 TapeStation (Agilent Technologies, Palo

Alto, CA). 100-200 ng of total RNA will be used to prepare cRNA utilizing Ambion

Illumina TotalPrep RNA Amplification Kit (Thermo Fisher Scientific, MA). 0.75 μg of

cRNA per sample was hybridized to Illumina's (HumanHT-12 v4.0 Expression BeadChip

following the manufacturer's WGGEX Direct Hybridization protocol (Illumina, CA).

Processed array were scanned on an Illumina HiScan. Raw data was produced using

Illumina Genome Studio followed by differential gene expression analysis utilizing Partek

Genomics Suite (St. Louis, MO).

5.12. RNA-seq

Total RNA was prepared using a Qiagen RNeasy Plus Mini Kit and processed by the

MUSC Genomics core for 2 x 125 cycles, paired-end RNA sequencing on an Illumina

HiSeq 2500. 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 manufacturer's

instructions (Illumina, San Diego, CA). Sequencing was performed on an Illumina HiSeq

2500. 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. Samples were prepared in triplicate as biological replicates and

sequenced simultaneously.

5.13. Chromatin immunoprecipitation (ChIP)

5.13a. Chromatin isolation and shearing. For each ChIP, 8-15 million cells (1-2 15 cm

dishes) per treatment condition were used as starting material and processed with the

Covaris truChIP chromatin shearing reagent kit (Covaris, 520154) according to

manufacturer's instructions. Briefly, cells were first washed with RT PBS then

crosslinked using 2 mM disuccinyl-glutamate in PBS at RT for 30 min followed by

fixation in 1x Covaris fixing buffer with 1% formaldehyde for 5 min (MCF-7) or 2.5 min

(SUM-44). After quenching 5 min in Covaris quenching buffer, cells were washed and scraped in 5 ml/plate cold 1x PBS, spun at 7,500xg for 5 min at 4°C, washed in PBS, then resuspended in 1x lysis buffer and transferred to a microfuge tube. The samples were incubated 10 min with rotation at 4°C, spun at 1,700xg for 5 min at 4°C, resuspended in 1x wash buffer, incubated 10 min with rotation at 4°C, and spun at 1,700xg 5 min at 4°C. The pellets were then washed twice in 1x shearing buffer without resuspension using the same pelleting conditions as above. Upon final wash, pellets were resuspended in 1 ml shearing buffer and transferred to a Covaris glass sonication tube. Pellets of duplicate samples were combined at this step to 1 ml total volume for shearing. Each sample was sheared on the Covaris S220 series sonicator under high cell ChIP standard conditions for 10 min (MCF-7) or 8 min (SUM-44). Sheared chromatin was then transferred to a DNA LoBind tube, centrifuged at 10,000xg for 10 min at 4°C, and transferred to a clean DNA LoBind tube for storage at 4°C (up to 72 hours).

5.13b. ChIP reaction setup. ChIP reactions were set up in duplicate (IgG control, H3K4me3), triplicate (ERα), or quadruplicate (NSD3) according to antibody efficiency. Reactions included 900 µl ChIP dilution buffer (0.01% SDS, 1.10% Triton X-100, 1x TE buffer pH 8.0, 167 mM NaCl), 100 µl sheared chromatin, 10 µl (1x) protease inhibitor cocktail (Thermo, 78442), and 4 µg antibody as indicated. Reactions were incubated at 4°C with end-to-end rotation overnight.

5.13c. Epitope verification. Upon first use of a new ChIP antibody, a 10 ul aliquot of each sheared chromatin sample was mixed with 10 ul of laemmli sample buffer, and run as a normal western blot (see section 5.5) to verify that the shearing process did not disrupt the epitope to be used in the ChIP process.

5.13d. Size fragment verification. Upon first use of a new cell line, including

knockdown cells of a previously used cell line, a 25 ul aliquot of each sheared chromatin

sample was analyzed by Agilent 2200 TapeStation (Agilent Technologies, Palo Alto, CA)

according to manufacturer's instructions.

5.13e. Chromatin washing and DNA elution. Protein A/G beads were pre-blocked with

0.5% BSA in PBS and 20 µl beads with 10 µl protease inhibitor cocktail added to each

ChIP reaction at 4°C for 1 hour with end-to-end rotation. Supernatant was removed and

beads were washed 4x in wash buffer (1x TE buffer pH 8.0, 0,5 M LiCl, 100 mM NaCl,

0.10% sodium deoxycholate, 1% Triton X-100; recipe courtesy of the Brown Lab) and 2x

in 1x TE pH 8.0. Beads were resuspended in 50 µl 1x TE buffer pH 8.0 with 2 µl RNase

and incubated with input samples (10 µl non-ChIP'd sheared chromatin, 40 µl 1x TE, 2 µl

RNase) for 30 min at 37°C. 50 µl elution buffer (0.2 M sodium bicarbonate, 2% SDS)

and 2.5 µl proteinase K were added to each sample and incubated at 65°C overnight

with shaking (1400 rpm). DNA was then purified using AMPure XP beads (Beckman

Coulter, A6381) according to manufacturer's instructions and quantified by PicoGreen

DNA kit (Thermo, 7605).

5.13f. Preparation for sequencing. Sequencing libraries were prepared from ChIP

DNA by the MUSC Genomics Core Facility with Rubicon ThruPLEX library preparation

kits and sequenced as 35 bp single-end reads on an Illumina HiSeq 2500. Libraries for

three biological replicates were prepared and sequenced for each sample type.

5.14 Bioinformatics and Statistical Analysis

5.14a. Microarray. Data were normalized using the linear and LOWESS (LOcally

WEighted Scatterplot Smoothing) method (Yang *et al.* 2002). Normalized data were

imported into GeneSpring for analysis. Statistical significance of differentially expressed

genes was determined using four replicate measurements for each probe (gene). A t-test was performed with the null hypothesis that the normalized log fold changes reflecting the change in gene expression (44/NS) equal zero. The Benjamini and Hochberg multiple test correction was used to determine the false discovery rate (FDR) (Benjamini and Hochberg 1995).

5.14b. RNA-seq. Paired end sequencing was performed on RNA samples using an Illumina HiSeq 2500 with each sample sequenced to a depth of ~50 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 *et al.* 2014). Further data QC was performed using FastQC prior to aligning the reads to the human genome hg19 using Tophat2 (Kim *et al.* 2013). The resulting SAM files were sorted and inputted into the Python package HTSeq to generate count data for gene-level differential expression analyses. In order to infer differential signal within the data sets with robust statistical power, we utilized DEseq2 (Love *et al.* 2014). Transcript count data from DESeq2 analysis of the samples were sorted according to their adjusted p-value or q-value, which is the smallest false discovery rate (FDR) at which a transcript is called significant. FDR adjustment is needed with large data sets such as RNAseq. FDR is the expected fraction of false positive tests among significant tests and was calculated using the Benjamini-Hochberg multiple testing adjustment procedure (Benjamini and Hochberg 1995). Using the transcript count data from DESeq2 analysis, we selected transcripts with negative log fold change values, indicating decreased expression in test versus control samples, and $p<0.1$. This consensus gene set was then used for downstream

analysis, including comparison between different shRNA constructs to generate a more statistically stringent list of genes.

5.14c. ChIP-seq. Single end sequencing was performed on ChIP DNA using an Illumina HiSeq 2500 with each sample sequenced to a depth of ~50 million reads. Data was subjected to Illumina quality control (QC) procedures (data yielded Phred scores >30). Further data QC was performed using FastQC prior to aligning the reads to the human genome hg19 using Bowtie2. Aligned reads were converted from SAM to BAM format, sorted, and indexed. BED format files were obtained from BAM files. Peaks were called with the MACS version 2 peak caller (Zhang *et al.* 2008). ChIP samples were compared to matched input samples to generate peak sets for each replicate, which were then compared to matched condition biological replicates as described in Chapter 3, section 2 using an FDR cutoff of 0.4. Consensus peak sets generated by replicate overlap were used for downstream analysis, including comparison between treatment conditions.

5.14d. Gene Ontology and Pathway Analysis. Various data comparisons were performed between ChIP-seq, RNA-seq, and microarray gene lists as described in Chapters 2 and 3. Statistical analysis of Gene Ontology and other processes, pathways, and interactions was performed on these gene lists by gene list enrichment analysis using the ToppGene suite (Chen *et al.* 2009). P-values and percentage of genes in input compared to annotated gene lists in ToppFun are reported in results sections with Bonferroni corrected p-values and absolute number of genes aligning to each process reported in Appendices according to gene set (See List of Appendices on page v.).

5.14e. General Statistics. Statistical analysis was performed in Microsoft excel. Student's t-test was used and two-tail p-values reported. Graphs with error bars represent +/- one standard deviation unless otherwise reported.

5.15. Patient tumor cell culture

Samples were obtained according to IRB protocol (Pro00048028). Using sterile technique, samples were aliquoted to 50 ml falcon tubes, spun at 1,000 rpm for 5 min at RT, and the supernatant removed and properly discarded. Cells were resuspended in 1x Hank's Buffered Saline Solution (HBSS), consolidated, and spun again at 1,000 rpm for 5 min at RT. Supernatant was removed and resuspension and spin step was repeated. To lyse red blood cells, 9 ml sterile water was added to the tubes followed immediately by 1 ml 10x HBSS. Tubes were spun again at 1,000 rpm for 5 min at RT. Supernatant was removed and cells were resuspended in Ham's F-12 media supplemented with insulin, hydrocortisone, and 5% FBS (see section 5.1b). Cells were counted and plated in 6-well plates at a density of 1 million cells/well. Cells were cultured in various media types and supplements of estradiol but could not be cultured more than a few weeks. One well was lysed for western blotting after 2.5 weeks in culture and blotted for NSD3 and ERα expression according to section 5.5.

REFERENCES CITED

"<a systematic evaluation of miRNA mRNA interactions involved in the migration and invasion of breast cancer cells.pdf>."

(2012). "Comprehensive molecular portraits of human breast tumours." Nature **490**(7418): 61-70.

(2014). "SEER Cancer Statistics Factsheets: Lung Cancer."   Retrieved February 13, 2014, from http://seer.cancer.gov/statfacts/html/lungb.html.

Abdel-Hafiz, H. A. and K. B. Horwitz (2015). "Role of epigenetic modifications in luminal breast cancer." Epigenomics **7**(5): 847-862.

Ades, F., D. Zardavas, I. Bozovic-Spasojevic, L. Pugliano, D. Fumagalli, E. de Azambuja, G. Viale, C. Sotiriou and M. Piccart (2014). "Luminal B breast cancer: molecular characterization, clinical management, and future perspectives." J Clin Oncol **32**(25): 2794-2803.

Alao, J. P., E. W. Lam, S. Ali, L. Buluwela, W. Bordogna, P. Lockey, R. Varshochi, A. V. Stavropoulou, R. C. Coombes and D. M. Vigushin (2004). "Histone deacetylase inhibitor trichostatin A represses estrogen receptor alpha-dependent transcription and promotes proteasomal degradation of cyclin D1 in human breast carcinoma cell lines." Clin Cancer Res **10**(23): 8094-8104.

Ali, A. and S. Tyagi (2017). "Diverse roles of WDR5-RbBP5-ASH2L-DPY30 (WRAD) complex in the functions of the SET1 histone methyltransferase family." J Biosci **42**(1): 155-159.

Ali, A., S. N. Veeranki and S. Tyagi (2014). "A SET-domain-independent role of WRAD complex in cell-cycle regulatory function of mixed lineage leukemia." Nucleic Acids Res **42**(12): 7611-7624.

Allali-Hassani, A., E. Kuznetsova, T. Hajian, H. Wu, L. Dombrovski, Y. Li, S. Graslund, C. H. Arrowsmith, M. Schapira and M. Vedadi (2014). "A Basic Post-SET Extension of NSDs Is Essential for Nucleosome Binding In Vitro." J Biomol Screen **19**(6): 928-935.

Allred, D. C., P. Brown and D. Medina (2004). "The origins of estrogen receptor alpha-positive and estrogen receptor alpha-negative human breast cancer." Breast Cancer Res **6**(6): 240-245.

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.

Angus, L., N. Beije, A. Jager, J. W. Martens and S. Sleijfer (2016). "ESR1 mutations: Moving towards guiding treatment decision-making in metastatic breast cancer patients." Cancer Treat Rev **52**: 33-40.

Ariazi, E. A., J. L. Ariazi, F. Cordera and V. C. Jordan (2006). "Estrogen receptors as therapeutic targets in breast cancer." Curr Top Med Chem **6**(3): 181-202.

Atlas, T. C. G. (2015). "cBioPortal." Retrieved Feb 2, 2015, 2015, from http://www.cbioportal.org/index.do.

Avdic, V., P. Zhang, S. Lanouette, A. Voronova, I. Skerjanc and J. F. Couture (2011). "Fine-tuning the stimulation of MLL1 methyltransferase activity by a histone H3-based peptide mimetic." FASEB J **25**(3): 960-967.

Azim, H. A., L. Kassem, I. Treilleux, Q. Wang, M. A. El Enein, S. E. Anis and T. Bachelot (2016). "Analysis of PI3K/mTOR Pathway Biomarkers and Their Prognostic Value in Women with Hormone Receptor-Positive, HER2-Negative Early Breast Cancer." Transl Oncol **9**(2): 114-123.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methological) **57**(1): 289-300.

Bennett, R. L., A. Swaroop, C. Troche and J. D. Licht (2017). "The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in Cancer." Cold Spring Harb Perspect Med.

Bergamaschi, A., Y. H. Kim, P. Wang, T. Sorlie, T. Hernandez-Boussard, P. E. Lonning, R. Tibshirani, A. L. Borresen-Dale and J. R. Pollack (2006). "Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer." Genes Chromosomes Cancer **45**(11): 1033-1040.

Bernard-Pierrot, I., N. Gruel, N. Stransky, A. Vincent-Salomon, F. Reyal, V. Raynal, C. Vallot, G. Pierron, F. Radvanyi and O. Delattre (2008). "Characterization of the recurrent

8p11-12 amplicon identifies PPAPDC1B, a phosphatase protein, as a new therapeutic target in breast cancer." Cancer Res **68**(17): 7165-7175.

Bhan, A., I. Hussain, K. I. Ansari, S. A. Bobzean, L. I. Perrotti and S. S. Mandal (2014). "Histone methyltransferase EZH2 is transcriptionally induced by estradiol as well as estrogenic endocrine disruptors bisphenol-A and diethylstilbestrol." J Mol Biol **426**(20): 3426-3441.

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.

Blobel, G. A., S. Kadauke, E. Wang, A. W. Lau, J. Zuber, M. M. Chou and C. R. Vakoc (2009). "A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit." Mol Cell **36**(6): 970-983.

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

Borbely, G., L. A. Haldosen, K. Dahlman-Wright and C. Zhao (2015). "Induction of USP17 by combining BET and HDAC inhibitors in breast cancer cells." Oncotarget **6**(32): 33623-33635.

Butler, J. S., Y. H. Qiu, N. Zhang, S. Y. Yoo, K. R. Coombes, S. Y. Dent and S. M. Kornblau (2017). "Low expression of ASH2L protein correlates with a favorable outcome in acute myeloid leukemia." Leuk Lymphoma **58**(5): 1207-1218.

Butler, J. S., C. I. Zurita-Lopez, S. G. Clarke, M. T. Bedford and S. Y. Dent (2011). "Protein-arginine methyltransferase 1 (PRMT1) methylates Ash2L, a shared component of mammalian histone H3K4 methyltransferase complexes." J Biol Chem **286**(14): 12234-12244.

Cadoo, K. A., M. N. Fornier and P. G. Morris (2013). "Biological subtypes of breast cancer: current concepts and implications for recurrence patterns." Q J Nucl Med Mol Imaging **57**(4): 312-321.

Cai, F. F., C. Kohler, B. Zhang, M. H. Wang, W. J. Chen and X. Y. Zhong (2011). "Epigenetic therapy for breast cancer." Int J Mol Sci **12**(7): 4465-4487.

Caldon, C. E., C. M. Sergio, J. Kang, A. Muthukaruppan, M. N. Boersma, A. Stone, J. Barraclough, C. S. Lee, M. A. Black, L. D. Miller, J. M. Gee, R. I. Nicholson, R. L. Sutherland, C. G. Print and E. A. Musgrove (2012). "Cyclin E2 overexpression is associated with endocrine resistance but not insensitivity to CDK2 inhibition in human breast cancer cells." Mol Cancer Ther **11**(7): 1488-1499.

Cao, F., Y. Chen, T. Cierpicki, Y. Liu, V. Basrur, M. Lei and Y. Dou (2010). "An Ash2L/RbBP5 heterodimer stimulates the MLL1 methyltransferase activity through coordinated substrate interactions with the MLL1 SET domain." PLoS One **5**(11): e14102.

Carey, L. A., C. M. Perou, C. A. Livasy, L. G. Dressler, D. Cowan, K. Conway, G. Karaca, M. A. Troester, C. K. Tse, S. Edmiston, S. L. Deming, J. Geradts, M. C. Cheang, T. O. Nielsen, P. G. Moorman, H. S. Earp and R. C. Millikan (2006). "Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study." JAMA **295**(21): 2492-2502.

Chang, M. (2012). "Tamoxifen resistance in breast cancer." Biomol Ther (Seoul) **20**(3): 256-267.

Chanrion, M., V. Negre, H. Fontaine, N. Salvetat, F. Bibeau, G. Mac Grogan, L. Mauriac, D. Katsaros, F. Molina, C. Theillet and J. M. Darbon (2008). "A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer." Clin Cancer Res **14**(6): 1744-1752.

Cheang, M. C., S. K. Chia, D. Voduc, D. Gao, S. Leung, J. Snider, M. Watson, S. Davies, P. S. Bernard, J. S. Parker, C. M. Perou, M. J. Ellis and T. O. Nielsen (2009). "Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer." J Natl Cancer Inst **101**(10): 736-750.

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.

Chen, Y., F. Cao, B. Wan, Y. Dou and M. Lei (2012). "Structure of the SPRY domain of human Ash2L and its interactions with RbBP5 and DPY30." Cell Res **22**(3): 598-602.

Chen, Y., J. McGee, X. Chen, T. N. Doman, X. Gong, Y. Zhang, N. Hamm, X. Ma, R. E. Higgs, S. V. Bhagwat, S. Buchanan, S. B. Peng, K. A. Staschke, V. Yadav, Y. Yue and H. Kouros-Mehr (2014). "Identification of druggable cancer driver genes amplified across TCGA datasets." PLoS One **9**(5): e98293.

Cobain, E. F. and D. F. Hayes (2015). "Indications for prognostic gene expression profiling in early breast cancer." Curr Treat Options Oncol **16**(5): 23.

Comsa, S., A. M. Cimpean and M. Raica (2015). "The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research." Anticancer Res **35**(6): 3147-3154.

Connolly, R., H. Li, R. C. Jankowitz, Z. Zhang, M. A. Rudek, S. C. Jeter, S. Slater, P. Powers, A. C. Wolff, J. H. Fetting, A. M. Brufsky, R. Piekarz, N. Ahuja, P. W. Laird, H. Shen, D. J. Weisenberger, L. Cope, J. G. Herman, G. Somlo, A. Garcia, P. A. Jones, S. B. Baylin, N. E. Davidson, C. Zahnow and V. Stearns (2016). "Combination Epigenetic Therapy in Advanced Breast Cancer with 5-Azacitidine and Entinostat: A Phase II National Cancer Institute/Stand Up to Cancer Study." Clin Cancer Res.

Connolly, R. and V. Stearns (2012). "Epigenetics as a therapeutic target in breast cancer." J Mammary Gland Biol Neoplasia **17**(3-4): 191-204.

Cooke, S. L., J. C. Pole, S. F. Chin, I. O. Ellis, C. Caldas and P. A. Edwards (2008). "High-resolution array CGH clarifies events occurring on 8p in carcinogenesis." BMC Cancer **8**: 288.

Cornen, S., A. Guille, J. Adelaide, L. Addou-Klouche, P. Finetti, M. R. Saade, M. Manai, N. Carbuccia, I. Bekhouche, A. Letessier, S. Raynaud, E. Charafe-Jauffret, J. Jacquemier, S. Spicuglia, H. de The, P. Viens, F. Bertucci, D. Birnbaum and M. Chaffanet (2014). "Candidate luminal B breast cancer genes identified by genome, gene expression and DNA methylation profiling." PLoS One **9**(1): e81843.

Crowe, B. L., R. C. Larue, C. Yuan, S. Hess, M. Kvaratskhelia and M. P. Foster (2016). "Structure of the Brd4 ET domain bound to a C-terminal motif from gamma-retroviral integrases reveals a conserved mechanism of interaction." Proc Natl Acad Sci U S A **113**(8): 2086-2091.

Cui, X., R. Schiff, G. Arpino, C. K. Osborne and A. V. Lee (2005). "Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy." J Clin Oncol **23**(30): 7721-7735.

Curry, E., I. Green, N. Chapman-Rothe, E. Shamsaei, S. Kandil, F. L. Cherblanc, L. Payne, E. Bell, T. Ganesh, N. Srimongkolpithak, J. Caron, F. Li, A. G. Uren, J. P. Snyder, M. Vedadi, M. J. Fuchter and R. Brown (2015). "Dual EZH2 and EHMT2 histone methyltransferase inhibition increases biological efficacy in breast cancer cells." Clin Epigenetics **7**: 84.

Dabydeen, S. A. and P. A. Furth (2014). "Genetically engineered ERalpha-positive breast cancer mouse models." Endocr Relat Cancer **21**(3): R195-208.

Damaskos, C., S. Valsami, M. Kontos, E. Spartalis, T. Kalampokas, E. Kalampokas, A. Athanasiou, D. Moris, A. Daskalopoulou, S. Davakis, G. Tsourouflis, K. Kontzoglou, D. Perrea, N. Nikiteas and D. Dimitroulis (2017). "Histone Deacetylase Inhibitors: An Attractive Therapeutic Strategy Against Breast Cancer." Anticancer Res **37**(1): 35-46.

De Marchi, T., J. A. Foekens, A. Umar and J. W. Martens (2016). "Endocrine therapy resistance in estrogen receptor (ER)-positive breast cancer." Drug Discov Today **21**(7): 1181-1188.

Demers, C., C. P. Chaturvedi, J. A. Ranish, G. Juban, P. Lai, F. Morle, R. Aebersold, F. J. Dilworth, M. Groudine and M. Brand (2007). "Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus." Mol Cell **27**(4): 573-584.

Demircan, B., L. M. Dyer, M. Gerace, E. K. Lobenhofer, K. D. Robertson and K. D. Brown (2009). "Comparative epigenomics of human and mouse mammary tumors." Genes Chromosomes Cancer **48**(1): 83-97.

Dou, Y., T. A. Milne, A. J. Ruthenburg, S. Lee, J. W. Lee, G. L. Verdine, C. D. Allis and R. G. Roeder (2006). "Regulation of MLL1 H3K4 methyltransferase activity by its core components." Nat Struct Mol Biol **13**(8): 713-719.

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

Dowsett, M., R. I. Nicholson and R. J. Pietras (2005). "Biological characteristics of the pure antiestrogen fulvestrant: overcoming endocrine resistance." Breast Cancer Res Treat **93 Suppl 1**: S11-18.

Dutt, A., A. H. Ramos, P. S. Hammerman, C. Mermel, J. Cho, T. Sharifnia, A. Chande, K. E. Tanaka, N. Stransky, H. Greulich, N. S. Gray and M. Meyerson (2011). "Inhibitor-sensitive FGFR1 amplification in human non-small cell lung cancer." PLoS One **6**(6): e20351.

Ellis, M. J., L. Ding, D. Shen, J. Luo, V. J. Suman, J. W. Wallis, B. A. Van Tine, J. Hoog, R. J. Goiffon, T. C. Goldstein, S. Ng, L. Lin, R. Crowder, J. Snider, K. Ballman, J. Weber, K. Chen, D. C. Koboldt, C. Kandoth, W. S. Schierding, J. F. McMichael, C. A. Miller, C.

Lu, C. C. Harris, M. D. McLellan, M. C. Wendl, K. DeSchryver, D. C. Allred, L. Esserman, G. Unzeitig, J. Margenthaler, G. V. Babiera, P. K. Marcom, J. M. Guenther, M. Leitch, K. Hunt, J. Olson, Y. Tao, C. A. Maher, L. L. Fulton, R. S. Fulton, M. Harrison, B. Oberkfell, F. Du, R. Demeter, T. L. Vickery, A. Elhammali, H. Piwnica-Worms, S. McDonald, M. Watson, D. J. Dooling, D. Ota, L. W. Chang, R. Bose, T. J. Ley, D. Piwnica-Worms, J. M. Stuart, R. K. Wilson and E. R. Mardis (2012). "Whole-genome analysis informs breast cancer response to aromatase inhibition." <u>Nature</u> **486**(7403): 353-360.

Ellis, M. J., M. Dixon, M. Dowsett, R. Nagarajan and E. Mardis (2007). "A luminal breast cancer genome atlas: progress and barriers." <u>J Steroid Biochem Mol Biol</u> **106**(1-5): 125-129.

Elsheikh, S. E., A. R. Green, E. A. Rakha, D. G. Powe, R. A. Ahmed, H. M. Collins, D. Soria, J. M. Garibaldi, C. E. Paish, A. A. Ammar, M. J. Grainge, G. R. Ball, M. K. Abdelghany, L. Martinez-Pomares, D. M. Heery and I. O. Ellis (2009). "Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome." <u>Cancer Res</u> **69**(9): 3802-3809.

Erfani, P., J. Tome-Garcia, P. Canoll, F. Doetsch and N. M. Tsankova (2015). "EGFR promoter exhibits dynamic histone modifications and binding of ASH2L and P300 in human germinal matrix and gliomas." <u>Epigenetics</u> **10**(6): 496-507.

Ernst, P. and C. R. Vakoc (2012). "WRAD: enabler of the SET1-family of H3K4 methyltransferases." <u>Brief Funct Genomics</u> **11**(3): 217-226.

Eroles, P., A. Bosch, J. A. Perez-Fidalgo and A. Lluch (2012). "Molecular biology in breast cancer: intrinsic subtypes and signaling pathways." <u>Cancer Treat Rev</u> **38**(6): 698-707.

Ethier, S. P. (1996). "Human breast cancer cell lines as models of growth regulation and disease progression." <u>J Mammary Gland Biol Neoplasia</u> **1**(1): 111-121.

Ethier, S. P., M. L. Mahacek, W. J. Gullick, T. S. Frank and B. L. Weber (1993). "Differential isolation of normal luminal mammary epithelial cells and breast cancer cells from primary and metastatic sites using selective media." <u>Cancer Res</u> **53**(3): 627-635.

Falahi, F., M. van Kruchten, N. Martinet, G. A. Hospers and M. G. Rots (2014). "Current and upcoming approaches to exploit the reversibility of epigenetic mutations in breast cancer." <u>Breast Cancer Res</u> **16**(4): 412.

Fang, R., A. J. Barbera, Y. Xu, M. Rutenberg, T. Leonor, Q. Bi, F. Lan, P. Mei, G. C. Yuan, C. Lian, J. Peng, D. Cheng, G. Sui, U. B. Kaiser, Y. Shi and Y. G. Shi (2010). "Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation." Mol Cell **39**(2): 222-233.

Fedele, P., N. Calvani, A. Marino, L. Orlando, P. Schiavone, A. Quaranta and S. Cinieri (2012). "Targeted agents to reverse resistance to endocrine therapy in metastatic breast cancer: where are we now and where are we going?" Crit Rev Oncol Hematol **84**(2): 243-251.

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.

Fossati, A., D. Dolfini, G. Donati and R. Mantovani (2011). "NF-Y recruits Ash2L to impart H3K4 trimethylation on CCAAT promoters." PLoS One **6**(3): e17220.

French, C. A., S. Rahman, E. M. Walsh, S. Kuhnle, A. R. Grayson, M. E. Lemieux, N. Grunfeld, B. P. Rubin, C. R. Antonescu, S. Zhang, R. Venkatramani, P. Dal Cin and P. M. Howley (2014). "NSD3-NUT fusion oncoprotein in NUT midline carcinoma: implications for a novel oncogenic mechanism." Cancer Discov **4**(8): 928-941.

Gao, J., B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander and N. Schultz (2013). "Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal." Sci Signal **6**(269): pl1.

Garapaty-Rao, S., C. Nasveschuk, A. Gagnon, E. Y. Chan, P. Sandy, J. Busby, S. Balasubramanian, R. Campbell, F. Zhao, L. Bergeron, J. E. Audia, B. K. Albrecht, J. C. Harmange, R. Cummings and P. Trojer (2013). "Identification of EZH2 and EZH1 small molecule inhibitors with selective impact on diffuse large B cell lymphoma cell growth." Chem Biol **20**(11): 1329-1339.

Garapaty, S., C. F. Xu, P. Trojer, M. A. Mahajan, T. A. Neubert and H. H. Samuels (2009). "Identification and characterization of a novel nuclear protein complex involved in nuclear hormone receptor-mediated gene regulation." J Biol Chem **284**(12): 7542-7552.

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.

Geyer, F. C., D. N. Rodrigues, B. Weigelt and J. S. Reis-Filho (2012). "Molecular classification of estrogen receptor-positive/luminal breast cancers." Adv Anat Pathol **19**(1): 39-53.

Grzenda, A., T. Ordog and R. Urrutia (2011). "Polycomb and the emerging epigenetics of pancreatic cancer." J Gastrointest Cancer **42**(2): 100-111.

Guccione, E., C. Bassi, F. Casadio, F. Martinato, M. Cesaroni, H. Schuchlautz, B. Luscher and B. Amati (2007). "Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive." Nature **449**(7164): 933-937.

Guertin, D. A., K. V. Guntur, G. W. Bell, C. C. Thoreen and D. M. Sabatini (2006). "Functional genomics identifies TOR-regulated genes that control growth and division." Curr Biol **16**(10): 958-970.

Habashy, H. O., D. G. Powe, T. M. Abdel-Fatah, J. M. Gee, R. I. Nicholson, A. R. Green, E. A. Rakha and I. O. Ellis (2012). "A review of the biological and clinical characteristics of luminal-like oestrogen receptor-positive breast cancer." Histopathology **60**(6): 854-863.

Hartman, J., A. Strom and J. A. Gustafsson (2009). "Estrogen receptor beta in breast cancer--diagnostic and therapeutic implications." Steroids **74**(8): 635-641.

Haverty, P. M., J. Fridlyand, L. Li, G. Getz, R. Beroukhim, S. Lohr, T. D. Wu, G. Cavet, Z. Zhang and J. Chant (2008). "High-resolution genomic and expression analyses of copy number alterations in breast tumors." Genes Chromosomes Cancer **47**(6): 530-542.

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.

Hervouet, E., P. F. Cartron, M. Jouvenot and R. Delage-Mourroux (2013). "Epigenetic regulation of estrogen signaling in breast cancer." Epigenetics **8**(3): 237-245.

Holland, D. G., A. Burleigh, A. Git, M. A. Goldgraben, P. A. Perez-Mancera, S. F. Chin, A. Hurtado, A. Bruna, H. R. Ali, W. Greenwood, M. J. Dunning, S. Samarajiwa, S. Menon, O. M. Rueda, A. G. Lynch, S. McKinney, I. O. Ellis, C. J. Eaves, J. S. Carroll, C. Curtis, S. Aparicio and C. Caldas (2011). "ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium." EMBO Mol Med **3**(3): 167-180.

Howell, A. (2006). "Pure oestrogen antagonists for the treatment of advanced breast cancer." Endocr Relat Cancer **13**(3): 689-706.

Huang, L., S. Zhao, J. M. Frasor and Y. Dai (2011). "An integrated bioinformatics approach identifies elevated cyclin E2 expression and E2F activity as distinct features of tamoxifen resistant breast tumors." PLoS One **6**(7): e22274.

Huang, N., E. vom Baur, J. M. Garnier, T. Lerouge, J. L. Vonesch, Y. Lutz, P. Chambon and R. Losson (1998). "Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators." EMBO J **17**(12): 3398-3412.

Huang, Y., S. Nayak, R. Jankowitz, N. E. Davidson and S. Oesterreich (2011). "Epigenetics in breast cancer: what's new?" Breast Cancer Res **13**(6): 225.

Ignatiadis, M. and C. Sotiriou (2013). "Luminal breast cancer: from biology to treatment." Nat Rev Clin Oncol **10**(9): 494-506.

Ikegawa, S., M. Isomura, Y. Koshizuka and Y. Nakamura (1999). "Cloning and characterization of ASH2L and Ash2l, human and mouse homologs of the Drosophila ash2 gene." Cytogenet Cell Genet **84**(3-4): 167-172.

Inoue, K. and E. A. Fry (2015). "Aberrant expression of cyclin D1 in cancer." Sign Transduct Insights **4**: 1-13.

Irish, J. C., J. N. Mills, B. Turner-Ivey, R. C. Wilson, S. T. Guest, A. Rutkovsky, A. Dombkowski, C. S. Kappler, G. Hardiman and S. P. Ethier (2016). "Amplification of WHSC1L1 regulates expression and estrogen-independent activation of ERalpha in SUM-44 breast cancer cells and is associated with ERalpha over-expression in breast cancer." Mol Oncol.

Jacques-Fricke, B. T. and L. S. Gammill (2014). "Neural crest specification and migration independently require NSD3-related lysine methyltransferase activity." Mol Biol Cell **25**(25): 4174-4186.

Johnston, S. J. and K. L. Cheung (2010). "Fulvestrant - a novel endocrine therapy for breast cancer." Curr Med Chem **17**(10): 902-914.

Johnston, S. R., L. A. Martin, J. Head, I. Smith and M. Dowsett (2005). "Aromatase inhibitors: combinations with fulvestrant or signal transduction inhibitors as a strategy to overcome endocrine resistance." J Steroid Biochem Mol Biol **95**(1-5): 173-181.

Jovanovic, J., J. A. Ronneberg, J. Tost and V. Kristensen (2010). "The epigenetics of breast cancer." Mol Oncol **4**(3): 242-254.

Judes, G., A. Dagdemir, S. Karsli-Ceppioglu, A. Lebert, M. Echegut, M. Ngollo, Y. J. Bignon, F. Penault-Llorca and D. Bernard-Gallon (2016). "H3K4 acetylation, H3K9 acetylation and H3K27 methylation in breast tumor molecular subtypes." Epigenomics **8**(7): 909-924.

Kang, D., H. S. Cho, G. Toyokawa, M. Kogure, Y. Yamane, Y. Iwai, S. Hayami, T. Tsunoda, H. I. Field, K. Matsuda, D. E. Neal, B. A. Ponder, Y. Maehara, Y. Nakamura and R. Hamamoto (2013). "The histone methyltransferase Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1) is involved in human carcinogenesis." Genes Chromosomes Cancer **52**(2): 126-139.

Kao, J., K. Salari, M. Bocanegra, Y. L. Choi, L. Girard, J. Gandhi, K. A. Kwei, T. Hernandez-Boussard, P. Wang, A. F. Gazdar, J. D. Minna and J. R. Pollack (2009). "Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery." PLoS One **4**(7): e6146.

Karlsson, E., M. A. Waltersson, J. Bostner, G. Perez-Tenorio, B. Olsson, A. L. Hallbeck and O. Stal (2011). "High-resolution genomic analysis of the 11q13 amplicon in breast cancers identifies synergy with 8p12 amplification, involving the mTOR targets S6K2 and 4EBP1." Genes Chromosomes Cancer **50**(10): 775-787.

Katoh, M. (2016). "Mutation spectra of histone methyltransferases with canonical SET domains and EZH2-targeted therapy." Epigenomics **8**(2): 285-305.

Kawabe, Y., Y. X. Wang, I. W. McKinnell, M. T. Bedford and M. A. Rudnicki (2012). "Carm1 regulates Pax7 transcriptional activity through MLL1/2 recruitment during asymmetric satellite stem cell divisions." Cell Stem Cell **11**(3): 333-345.

Kerdivel, G., G. Flouriot and F. Pakdel (2013). "Modulation of estrogen receptor alpha activity and expression during breast cancer progression." Vitam Horm **93**: 135-160.

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, J. H., A. Sharma, S. S. Dhar, S. H. Lee, B. Gu, C. H. Chan, H. K. Lin and M. G. Lee (2014). "UTX and MLL4 coordinately regulate transcriptional programs for cell proliferation and invasiveness in breast cancer cells." Cancer Res **74**(6): 1705-1717.

Kim, S., S. Natesan, G. Cornilescu, S. Carlson, M. Tonelli, U. L. McClurg, O. Binda, C. N. Robson, J. L. Markley, S. Balaz and K. C. Glass (2016). "Mechanism of Histone H3K4me3 Recognition by the Plant Homeodomain of Inhibitor of Growth 3." J Biol Chem **291**(35): 18326-18341.

Kim, S. M., H. J. Kee, N. Choe, J. Y. Kim, H. Kook, 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.

Knudsen, E. S. and A. K. Witkiewicz (2016). "Defining the transcriptional and biological response to CDK4/6 inhibition in relation to ER+/HER2- breast cancer." Oncotarget **7**(43): 69111-69123.

Kristensen, L. S., H. M. Nielsen and L. L. Hansen (2009). "Epigenetics and cancer treatment." Eur J Pharmacol **625**(1-3): 131-142.

Kristensen, V. N., T. Sorlie, J. Geisler, A. Langerod, N. Yoshimura, R. Karesen, N. Harada, P. E. Lonning and A. L. Borresen-Dale (2005). "Gene expression profiling of breast cancer in relation to estrogen receptor status and estrogen-metabolizing enzymes: clinical implications." Clin Cancer Res **11**(2 Pt 2): 878s-883s.

Kumar, R., D. Q. Li, S. Muller and S. Knapp (2016). "Epigenomic regulation of oncogenesis by chromatin remodeling." Oncogene **35**(34): 4423-4436.

Kumler, I., A. S. Knoop, C. A. Jessing, B. Ejlertsen and D. L. Nielsen (2016). "Review of hormone-based treatments in postmenopausal patients with advanced breast cancer focusing on aromatase inhibitors and fulvestrant." ESMO Open **1**(4): e000062.

Kuo, A. J., P. Cheung, K. Chen, B. M. Zee, M. Kioi, J. Lauring, Y. Xi, B. H. Park, X. Shi, B. A. Garcia, W. Li and O. Gozani (2011). "NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming." Mol Cell **44**(4): 609-620.

Kwek, S. S., R. Roy, H. Zhou, J. Climent, J. A. Martinez-Climent, J. Fridlyand and D. G. Albertson (2009). "Co-amplified genes at 8p12 and 11q13 in breast tumors cooperate with two major pathways in oncogenesis." Oncogene **28**(17): 1892-1903.

Li, Y., J. Han, Y. Zhang, F. Cao, Z. Liu, S. Li, J. Wu, C. Hu, Y. Wang, J. Shuai, J. Chen, L. Cao, D. Li, P. Shi, C. Tian, J. Zhang, Y. Dou, G. Li, Y. Chen and M. Lei (2016). "Structural basis for activity regulation of MLL family methyltransferases." Nature **530**(7591): 447-452.

Li, Y., P. Trojer, C. F. Xu, P. Cheung, A. Kuo, W. J. Drury, 3rd, Q. Qiao, T. A. Neubert, R. M. Xu, O. Gozani and D. Reinberg (2009). "The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate." J Biol Chem **284**(49): 34283-34295.

Lim, E., G. Tarulli, N. Portman, T. E. Hickey, W. D. Tilley and C. Palmieri (2016). "Pushing estrogen receptor around in breast cancer." Endocr Relat Cancer **23**(12): T227-T241.

Lin, I. H., D. T. Chen, Y. F. Chang, Y. L. Lee, C. H. Su, C. Cheng, Y. C. Tsai, S. C. Ng, H. T. Chen, M. C. Lee, H. W. Chen, S. H. Suen, Y. C. Chen, T. T. Liu, C. H. Chang and M. T. Hsu (2015). "Hierarchical clustering of breast cancer methylomes revealed differentially methylated and expressed breast cancer genes." PLoS One **10**(2): e0118453.

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.

Liu, L., S. Kimball, H. Liu, A. Holowatyj and Z. Q. Yang (2014). "Genetic alterations of histone lysine methyltransferases and their significance in breast cancer." Oncotarget.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-408.

Lo, P. K. and S. Sukumar (2008). "Epigenomics and breast cancer." Pharmacogenomics **9**(12): 1879-1902.

Locke, W. J., E. Zotenko, C. Stirzaker, M. D. Robinson, R. A. Hinshelwood, A. Stone, R. R. Reddel, L. I. Huschtscha and S. J. Clark (2015). "Coordinated epigenetic remodelling of transcriptional networks occurs during early breast carcinogenesis." Clin Epigenetics **7**: 52.

Loi, S., C. Sotiriou, B. Haibe-Kains, F. Lallemand, N. M. Conus, M. J. Piccart, T. P. Speed and G. A. McArthur (2009). "Gene expression profiling identifies activated growth factor signaling in poor prognosis (Luminal-B) estrogen receptor positive breast cancer." BMC Med Genomics **2**: 37.

Lonning, P. E., T. Sorlie and A. L. Borresen-Dale (2005). "Genomics in breast cancer-therapeutic implications." Nat Clin Pract Oncol **2**(1): 26-33.

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.

Lumachi, F., D. A. Santeufemia and S. M. Basso (2015). "Current medical treatment of estrogen receptor-positive breast cancer." World J Biol Chem **6**(3): 231-239.

Luo, J., S. Liu, S. Leung, A. A. Gru, Y. Tao, J. Hoog, J. Ho, S. R. Davies, D. C. Allred, A. L. Salavaggione, J. Snider, E. R. Mardis, T. O. Nielsen and M. J. Ellis (2017). "An mRNA Gene Expression-Based Signature to Identify FGFR1-Amplified Estrogen Receptor-Positive Breast Tumors." J Mol Diagn **19**(1): 147-161.

Luscher-Firzlaff, J., I. Gawlista, J. Vervoorts, K. Kapelle, T. Braunschweig, G. Walsemann, C. Rodgarkia-Schamberger, H. Schuchlautz, S. Dreschers, E. Kremmer, R. Lilischkis, C. Cerni, A. Wellmann and B. Luscher (2008). "The human trithorax protein hASH2 functions as an oncoprotein." Cancer Res **68**(3): 749-758.

Lustberg, M. B. and B. Ramaswamy (2009). "Epigenetic targeting in breast cancer: therapeutic impact and future direction." Drug News Perspect **22**(7): 369-381.

Lustberg, M. B. and B. Ramaswamy (2011). "Epigenetic Therapy in Breast Cancer." Curr Breast Cancer Rep **3**(1): 34-43.

Mahmood, S. F., N. Gruel, R. Nicolle, E. Chapeaublanc, O. Delattre, F. Radvanyi and I. Bernard-Pierrot (2013). "PPAPDC1B and WHSC1L1 are common drivers of the 8p11-12 amplicon, not only in breast tumors but also in pancreatic adenocarcinomas and lung tumors." Am J Pathol **183**(5): 1634-1644.

Mancuso, M. R. and S. A. Massarweh (2016). "Endocrine therapy and strategies to overcome therapeutic resistance in breast cancer." Curr Probl Cancer **40**(2-4): 95-105.

Mann, K. M., J. M. Ward, C. C. Yew, A. Kovochich, D. W. Dawson, M. A. Black, B. T. Brett, T. E. Sheetz, A. J. Dupuy, D. K. Chang, A. V. Biankin, N. Waddell, K. S. Kassahn, S. M. Grimmond, A. G. Rust, D. J. Adams, N. A. Jenkins and N. G. Copeland (2012). "Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma." Proc Natl Acad Sci U S A **109**(16): 5934-5941.

Margueron, R., V. Duong, A. Castet and V. Cavailles (2004). "Histone deacetylase inhibition and estrogen signalling in human breast cancer cells." Biochem Pharmacol **68**(6): 1239-1246.

McKeage, K., M. P. Curran and G. L. Plosker (2004). "Fulvestrant: a review of its use in hormone receptor-positive metastatic breast cancer in postmenopausal women with disease progression following antiestrogen therapy." Drugs **64**(6): 633-648.

Melchor, L., M. J. Garcia, E. Honrado, J. C. Pole, S. Alvarez, P. A. Edwards, C. Caldas, J. D. Brenton and J. Benitez (2007). "Genomic analysis of the 8p11-12 amplicon in familial breast cancer." Int J Cancer **120**(3): 714-717.

Messier, T. L., J. A. Gordon, J. R. Boyd, C. E. Tye, G. Browne, J. L. Stein, J. B. Lian and G. S. Stein (2016). "Histone H3 lysine 4 acetylation and methylation dynamics define breast cancer subtypes." Oncotarget **7**(5): 5094-5109.

Michalak, E. M. and J. E. Visvader (2016). "Dysregulation of histone methyltransferases in breast cancer - Opportunities for new targeted therapies?" Mol Oncol **10**(10): 1497-1515.

Miller, W. R., A. Larionov, L. Renshaw, T. J. Anderson, J. R. Walker, A. Krause, T. Sing, D. B. Evans and J. M. Dixon (2009). "Gene expression profiles differentiating between breast cancers clinically responsive or resistant to letrozole." J Clin Oncol **27**(9): 1382-1387.

Miyamoto, K. and T. Ushijima (2005). "Diagnostic and therapeutic applications of epigenetics." Jpn J Clin Oncol **35**(6): 293-301.

Mohammed, H., I. A. Russell, R. Stark, O. M. Rueda, T. E. Hickey, G. A. Tarulli, A. A. Serandour, S. N. Birrell, A. Bruna, A. Saadi, S. Menon, J. Hadfield, M. Pugh, G. V. Raj, G. D. Brown, C. D'Santos, J. L. Robinson, G. Silva, R. Launchbury, C. M. Perou, J. Stingl, C. Caldas, W. D. Tilley and J. S. Carroll (2015). "Progesterone receptor modulates ERalpha action in breast cancer." Nature **523**(7560): 313-317.

Morishita, M. and E. di Luccio (2011). "Cancers and the NSD family of histone lysine methyltransferases." Biochim Biophys Acta **1816**(2): 158-163.

Morishita, M. and E. di Luccio (2011). "Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1." Biochem Biophys Res Commun **412**(2): 214-219.

Morishita, M., D. Mevius and E. di Luccio (2014). "In vitro histone lysine methylation by NSD1, NSD2/MMSET/WHSC1, and NSD3/WHSC1L." BMC Struct Biol **14**(1): 25.

Morris, C. and A. Wakeling (2002). "Fulvestrant ('Faslodex')--a new treatment option for patients progressing on prior endocrine therapy." Endocr Relat Cancer **9**(4): 267-276.

Morris, S. R. and L. A. Carey (2007). "Molecular profiling in breast cancer." Rev Endocr Metab Disord **8**(3): 185-198.

Mungamuri, S. K., S. Wang, J. J. Manfredi, W. Gu and S. A. Aaronson (2015). "Ash2L enables P53-dependent apoptosis by favoring stable transcription pre-initiation complex formation on its pro-apoptotic target promoters." Oncogene **34**(19): 2461-2470.

Nagamatsu, G., S. Saito, T. Kosaka, K. Takubo, T. Kinoshita, M. Oya, K. Horimoto and T. Suda (2012). "Optimal ratio of transcription factors for somatic cell reprogramming." J Biol Chem **287**(43): 36273-36282.

Nagaraj, G. and C. Ma (2015). "Revisiting the estrogen receptor pathway and its role in endocrine therapy for postmenopausal women with estrogen receptor-positive metastatic breast cancer." Breast Cancer Res Treat **150**(2): 231-242.

Nagarajan, S., E. Benito, A. Fischer and S. A. Johnsen (2015). "H4K12ac is regulated by estrogen receptor-alpha and is associated with BRD4 function and inducible transcription." Oncotarget **6**(9): 7305-7317.

Netanely, D., A. Avraham, A. Ben-Baruch, E. Evron and R. Shamir (2016). "Expression and methylation patterns partition luminal-A breast tumors into distinct prognostic subgroups." Breast Cancer Res **18**(1): 74.

Nguyen, P., G. Bar-Sela, L. Sun, K. S. Bisht, H. Cui, E. Kohn, A. P. Feinberg and D. Gius (2008). "BAT3 and SET1A form a complex with CTCFL/BORIS to modulate H3K4 histone dimethylation and gene expression." Mol Cell Biol **28**(21): 6720-6729.

Nicholson, R. I. and S. R. Johnston (2005). "Endocrine therapy--current benefits and limitations." Breast Cancer Res Treat **93 Suppl 1**: S3-10.

Noberini, R. and T. Bonaldi (2017). "A Super-SILAC Strategy for the Accurate and Multiplexed Profiling of Histone Posttranslational Modifications." Methods Enzymol **586**: 311-332.

Osmanbeyoglu, H. U., R. Pelossof, J. F. Bromberg and C. S. Leslie (2014). "Linking signaling pathways to transcriptional programs in breast cancer." Genome Res **24**(11): 1869-1880.

Paska, A. V. and P. Hudler (2015). "Aberrant methylation patterns in cancer: a clinical view." Biochem Med (Zagreb) **25**(2): 161-176.

Patel, A., V. Dharmarajan, V. E. Vought and M. S. Cosgrove (2009). "On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex." J Biol Chem **284**(36): 24242-24256.

Patel, A., V. E. Vought, V. Dharmarajan and M. S. Cosgrove (2011). "A novel non-SET domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex." J Biol Chem **286**(5): 3359-3369.

Patel, A., V. E. Vought, S. Swatkoski, S. Viggiano, B. Howard, V. Dharmarajan, K. E. Monteith, G. Kupakuwana, K. E. Namitz, S. A. Shinsky, R. J. Cotter and M. S. Cosgrove (2014). "Automethylation activities within the mixed lineage leukemia-1 (MLL1) core complex reveal evidence supporting a "two-active site" model for multiple histone H3 lysine 4 methylation." J Biol Chem **289**(2): 868-884.

Pathiraja, T. N., V. Stearns and S. Oesterreich (2010). "Epigenetic regulation in estrogen receptor positive breast cancer--role in treatment response." J Mammary Gland Biol Neoplasia **15**(1): 35-47.

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.

Pritchard, K. I. (2013). "Endocrine therapy: is the first generation of targeted drugs the last?" J Intern Med **274**(2): 144-152.

Pullirsch, D., R. Hartel, H. Kishimoto, M. Leeb, G. Steiner and A. Wutz (2010). "The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation." Development **137**(6): 935-943.

Qi, J., L. Huo, Y. T. Zhu and Y. J. Zhu (2014). "Absent, small or homeotic 2-like protein (ASH2L) enhances the transcription of the estrogen receptor alpha gene through GATA-binding protein 3 (GATA3)." J Biol Chem **289**(45): 31373-31381.

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.

Raica, M., I. Jung, A. M. Cimpean, C. Suciu and A. M. Muresan (2009). "From conventional pathologic diagnosis to the molecular classification of breast carcinoma: are we ready for the change?" Rom J Morphol Embryol **50**(1): 5-13.

Rakha, E. A. and A. R. Green (2017). "Molecular classification of breast cancer: what the pathologist needs to know." Pathology **49**(2): 111-119.

Rampalli, S., L. Li, E. Mak, K. Ge, M. Brand, S. J. Tapscott and F. J. Dilworth (2007). "p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase

complexes to specific genes during differentiation." Nat Struct Mol Biol **14**(12): 1150-1156.

Ray, M. E., Z. Q. Yang, D. Albertson, C. G. Kleer, J. G. Washburn, J. A. Macoska and S. P. Ethier (2004). "Genomic and expression analysis of the 8p11-12 amplicon in human breast cancer cell lines." Cancer Res **64**(1): 40-47.

Reis-Filho, J. S. and L. Pusztai (2011). "Gene expression profiling in breast cancer: classification, prognostication, and prediction." Lancet **378**(9805): 1812-1823.

Reis-Filho, J. S., P. T. Simpson, N. C. Turner, M. B. Lambros, C. Jones, A. Mackay, A. Grigoriadis, D. Sarrio, K. Savage, T. Dexter, M. Iravani, K. Fenwick, B. Weber, D. Hardisson, F. C. Schmitt, J. Palacios, S. R. Lakhani and A. Ashworth (2006). "FGFR1 emerges as a potential therapeutic target for lobular breast carcinomas." Clin Cancer Res **12**(22): 6652-6662.

Reynisdottir, I., A. Arason, B. O. Einarsdottir, H. Gunnarsson, J. Staaf, J. Vallon-Christersson, G. Jonsson, M. Ringner, B. A. Agnarsson, K. Olafsdottir, R. Fagerholm, T. Einarsdottir, G. Johannesdottir, O. T. Johannsson, H. Nevanlinna, A. Borg and R. B. Barkardottir (2013). "High expression of ZNF703 independent of amplification indicates worse prognosis in patients with luminal B breast cancer." Cancer Med **2**(4): 437-446.

Rijnkels, M., E. Kabotyanski, M. B. Montazer-Torbati, C. Hue Beauvais, Y. Vassetzky, J. M. Rosen and E. Devinoy (2010). "The epigenetic landscape of mammary gland development and functional differentiation." J Mammary Gland Biol Neoplasia **15**(1): 85-100.

Robertson, J. F. (2007). "Fulvestrant (Faslodex) -- how to make a good drug better." Oncologist **12**(7): 774-784.

Robertson, J. F., J. Lindemann, S. Garnett, E. Anderson, R. I. Nicholson, I. Kuter and J. M. Gee (2014). "A good drug made better: the fulvestrant dose-response story." Clin Breast Cancer **14**(6): 381-389.

Rosati, R., R. La Starza, A. Veronese, A. Aventin, C. Schwienbacher, T. Vallespi, M. Negrini, M. F. Martelli and C. Mecucci (2002). "NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15)." Blood **99**(10): 3857-3860.

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.

Saloura, V., T. Vougiouklakis, M. Zewde, X. Deng, K. Kiyotani, J. H. Park, Y. Matsuo, M. Lingen, T. Suzuki, N. Dohmae, R. Hamamoto and Y. Nakamura (2017). "WHSC1L1-mediated EGFR mono-methylation enhances the cytoplasmic and nuclear oncogenic activity of EGFR in head and neck cancer." Sci Rep **7**: 40664.

Saloura, V., T. Vougiouklakis, M. Zewde, K. Kiyotani, J. H. 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 **7**(27): 42527-42538.

Sato, T., M. Cesaroni, W. Chung, S. Panjarian, A. Tran, J. Madzo, Y. Okamoto, H. Zhang, X. Chen, J. Jelinek and J. J. Issa (2017). "Transcriptional Selectivity of Epigenetic Therapy in Cancer." Cancer Res **77**(2): 470-481.

Schiff, R., S. Massarweh, J. Shou and C. K. Osborne (2003). "Breast cancer endocrine resistance: how growth factor signaling and estrogen receptor coregulators modulate response." Clin Cancer Res **9**(1 Pt 2): 447S-454S.

Schram, A. W., R. Baas, P. W. Jansen, A. Riss, L. Tora, M. Vermeulen and H. T. Timmers (2013). "A dual role for SAGA-associated factor 29 (SGF29) in ER stress survival by coordination of both histone H3 acetylation and histone H3 lysine-4 trimethylation." PLoS One **8**(7): e70035.

Scott, S. M., M. Brown and S. E. Come (2011). "Emerging data on the efficacy and safety of fulvestrant, a unique antiestrogen therapy for advanced breast cancer." Expert Opin Drug Saf **10**(5): 819-826.

Selli, C., J. M. Dixon and A. H. Sims (2016). "Accurate prediction of response to endocrine therapy in breast cancer patients: current and future biomarkers." Breast Cancer Res **18**(1): 118.

Shan, L., X. Li, L. Liu, X. Ding, Q. Wang, Y. Zheng, Y. Duan, C. Xuan, Y. Wang, F. Yang, Y. Shang and L. Shi (2014). "GATA3 cooperates with PARP1 to regulate CCND1 transcription through modulating histone H1 incorporation." Oncogene **33**(24): 3205-3216.

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.

Shiu, K. K., R. Natrajan, F. C. Geyer, A. Ashworth and J. S. Reis-Filho (2010). "DNA amplifications in breast cancer: genotypic-phenotypic correlations." Future Oncol **6**(6): 967-984.

Simon, J. A. and C. A. Lange (2008). "Roles of the EZH2 histone methyltransferase in cancer epigenetics." Mutat Res **647**(1-2): 21-29.

Sircoulomb, F., N. Nicolas, A. Ferrari, P. Finetti, I. Bekhouche, E. Rousselet, A. Lonigro, J. Adelaide, E. Baudelet, S. Esteyries, J. Wicinski, S. Audebert, E. Charafe-Jauffret, J. Jacquemier, M. Lopez, J. P. Borg, C. Sotiriou, C. Popovici, F. Bertucci, D. Birnbaum, M. Chaffanet and C. Ginestier (2011). "ZNF703 gene amplification at 8p12 specifies luminal B breast cancer." EMBO Mol Med **3**(3): 153-166.

Sorlie, T., C. M. Perou, C. Fan, S. Geisler, T. Aas, A. Nobel, G. Anker, L. A. Akslen, D. Botstein, A. L. Borresen-Dale and P. E. Lonning (2006). "Gene expression profiles do not consistently predict the clinical treatment response in locally advanced breast cancer." Mol Cancer Ther **5**(11): 2914-2918.

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. E. 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.

Sorlie, T., R. Tibshirani, J. Parker, T. Hastie, J. S. Marron, A. Nobel, S. Deng, H. Johnsen, R. Pesich, S. Geisler, J. Demeter, C. M. Perou, P. E. Lonning, P. O. Brown, A. L. Borresen-Dale and D. Botstein (2003). "Repeated observation of breast tumor subtypes in independent gene expression data sets." Proc Natl Acad Sci U S A **100**(14): 8418-8423.

Sotiriou, C., S. Y. Neo, L. M. McShane, E. L. Korn, P. M. Long, A. Jazaeri, P. Martiat, S. B. Fox, A. L. Harris and E. T. Liu (2003). "Breast cancer classification and prognosis based on gene expression profiles from a population-based study." Proc Natl Acad Sci U S A **100**(18): 10393-10398.

Southall, S. M., P. S. Wong, Z. Odho, S. M. Roe and J. R. Wilson (2009). "Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks." Mol Cell **33**(2): 181-191.

Souza, P. P., P. Volkel, D. Trinel, J. Vandamme, C. Rosnoblet, L. Heliot and P. O. Angrand (2009). "The histone methyltransferase SUV420H2 and Heterochromatin Proteins HP1 interact but show different dynamic behaviours." BMC Cell Biol **10**: 41.

Stec, I., G. J. van Ommen and J. T. den Dunnen (2001). "WHSC1L1, on human chromosome 8p11.2, closely resembles WHSC1 and maps to a duplicated region shared with 4p16.3." Genomics **76**(1-3): 5-8.

Steelman, L. S., A. M. Martelli, L. Cocco, M. Libra, F. Nicoletti, S. L. Abrams and J. A. McCubrey (2016). "The therapeutic potential of mTOR inhibitors in breast cancer." Br J Clin Pharmacol **82**(5): 1189-1212.

Steward, M. M., J. S. Lee, A. O'Donovan, M. Wyatt, B. E. Bernstein and A. Shilatifard (2006). "Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes." Nat Struct Mol Biol **13**(9): 852-854.

Still, I. H., M. Hamilton, P. Vince, A. Wolfman and J. K. Cowell (1999). "Cloning of TACC1, an embryonically expressed, potentially transforming coiled coil containing gene, from the 8p11 breast cancer amplicon." Oncogene **18**(27): 4032-4038.

Streicher, K. L., Z. Q. Yang, S. Draghici and S. P. Ethier (2007). "Transforming function of the LSM1 oncogene in human breast cancers with the 8p11-12 amplicon." Oncogene **26**(14): 2104-2114.

Suzuki, S., N. Kurabe, I. Ohnishi, K. Yasuda, Y. Aoshima, M. Naito, F. Tanioka and H. Sugimura (2015). "NSD3-NUT-expressing midline carcinoma of the lung: first characterization of primary cancer tissue." Pathol Res Pract **211**(5): 404-408.

Tabarestani, S., M. Motallebi and M. E. Akbari (2016). "Are Estrogen Receptor Genomic Aberrations Predictive of Hormone Therapy Response in Breast Cancer?" Iran J Cancer Prev **9**(4): e6565.

Taketani, T., T. Taki, H. Nakamura, M. Taniwaki, J. Masuda and Y. Hayashi (2009). "NUP98-NSD3 fusion gene in radiation-associated myelodysplastic syndrome with t(8;11)(p11;p15) and expression pattern of NSD family genes." Cancer Genet Cytogenet **190**(2): 108-112.

Tan, C. C., K. V. Sindhu, S. Li, H. Nishio, J. Z. Stoller, K. Oishi, S. Puttreddy, T. J. Lee, J. A. Epstein, M. J. Walsh and B. D. Gelb (2008). "Transcription factor Ap2delta associates with Ash2l and ALR, a trithorax family histone methyltransferase, to activate Hoxc8 transcription." Proc Natl Acad Sci U S A **105**(21): 7472-7477.

Tan, C. C., M. J. Walsh and B. D. Gelb (2009). "Fgfr3 is a transcriptional target of Ap2delta and Ash2l-containing histone methyltransferase complexes." PLoS One **4**(12): e8535.

Tang, P. and G. M. Tse (2016). "Immunohistochemical Surrogates for Molecular Classification of Breast Carcinoma: A 2015 Update." Arch Pathol Lab Med **140**(8): 806-814.

Thomas, S. and P. N. Munster (2009). "Histone deacetylase inhibitor induced modulation of anti-estrogen therapy." Cancer Lett **280**(2): 184-191.

Turner-Ivey, B., S. T. Guest, J. C. Irish, C. S. Kappler, E. Garrett-Mayer, R. C. Wilson and S. P. Ethier (2014). "KAT6A, a chromatin modifier from the 8p11-p12 amplicon is a candidate oncogene in luminal breast cancer." Neoplasia **16**(8): 644-655.

Turner-Ivey, B., E. L. Smith, A. C. Rutkovsky, L. S. Spruill, J. N. Mills and S. P. Ethier (2017). "Development of mammary hyperplasia, dysplasia, and invasive ductal carcinoma in transgenic mice expressing the 8p11 amplicon oncogene NSD3." Breast Cancer Res Treat.

Turner, N., A. Pearson, R. Sharpe, M. Lambros, F. Geyer, M. A. Lopez-Garcia, R. Natrajan, C. Marchio, E. Iorns, A. Mackay, C. Gillett, A. Grigoriadis, A. Tutt, J. S. Reis-Filho and A. Ashworth (2010). "FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer." Cancer Res **70**(5): 2085-2094.

Turner, N. C., J. Ro, F. Andre, S. Loi, S. Verma, H. Iwata, N. Harbeck, S. Loibl, C. Huang Bartlett, K. Zhang, C. Giorgetti, S. Randolph, M. Koehler and M. Cristofanilli (2015). "Palbociclib in Hormone-Receptor-Positive Advanced Breast Cancer." N Engl J Med **373**(3): 209-219.

Ullius, A., J. Luscher-Firzlaff, I. G. Costa, G. Walsemann, A. H. Forst, E. G. Gusmao, K. Kapelle, H. Kleine, E. Kremmer, J. Vervoorts and B. Luscher (2014). "The interaction of MYC with the trithorax protein ASH2L promotes gene transcription by regulating H3K27 modification." Nucleic Acids Res **42**(11): 6901-6920.

van 't Veer, L. J., H. Dai, M. J. van de Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards and S. H. Friend (2002). "Gene expression profiling predicts clinical outcome of breast cancer." Nature **415**(6871): 530-536.

Vedadi, M., L. Blazer, M. S. Eram, D. Barsyte-Lovejoy, C. H. Arrowsmith and T. Hajian (2017). "Targeting human SET1/MLL family of proteins." Protein Sci.

Veeck, J. and M. Esteller (2010). "Breast cancer epigenetics: from DNA methylation to microRNAs." J Mammary Gland Biol Neoplasia **15**(1): 5-17.

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.

Walsh, L., W. M. Gallagher, D. P. O'Connor and T. Ni Chonghaile (2016). "Diagnostic and Therapeutic Implications of Histone Epigenetic Modulators in Breast Cancer." Expert Rev Mol Diagn **16**(5): 541-551.

Wan, M., J. Liang, Y. Xiong, F. Shi, Y. Zhang, W. Lu, Q. He, D. Yang, R. Chen, D. Liu, M. Barton and Z. Songyang (2013). "The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells." J Biol Chem **288**(7): 5039-5048.

Wang, G., G. Liu, X. Wang, S. Sethi, R. Ali-Fehmi, J. Abrams, Z. Zheng, K. Zhang, S. Ethier and Z. Q. Yang (2012). "ERLIN2 promotes breast cancer cell survival by modulating endoplasmic reticulum stress pathways." BMC Cancer **12**: 225.

Wang, G. G., L. Cai, M. P. Pasillas and M. P. Kamps (2007). "NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis." Nat Cell Biol **9**(7): 804-812.

Wang, J., Y. Zhou, B. Yin, G. Du, X. Huang, G. Li, Y. Shen, J. Yuan and B. Qiang (2001). "ASH2L: alternative splicing and downregulation during induced megakaryocytic differentiation of multipotential leukemia cell lines." J Mol Med (Berl) **79**(7): 399-405.

Wang, Z. Y. and L. Yin (2015). "Estrogen receptor alpha-36 (ER-alpha36): A new player in human breast cancer." Mol Cell Endocrinol **418 Pt 3**: 193-206.

Wen, H., Y. Li, Y. Xi, S. Jiang, S. Stratton, D. Peng, K. Tanaka, Y. Ren, Z. Xia, J. Wu, B. Li, M. C. Barton, W. Li, H. Li and X. Shi (2014). "ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression." Nature **508**(7495): 263-268.

Wesierska-Gadek, J., T. Schreiner, M. Gueorguieva and C. Ranftler (2006). "Phenol red reduces ROSC mediated cell cycle arrest and apoptosis in human MCF-7 cells." J Cell Biochem **98**(6): 1367-1379.

Wu, H., J. Min, V. V. Lunin, T. Antoshenko, L. Dombrovski, H. Zeng, A. Allali-Hassani, V. Campagna-Slater, M. Vedadi, C. H. Arrowsmith, A. N. Plotnikov and M. Schapira (2010). "Structural biology of human H3K9 methyltransferases." PLoS One **5**(1): e8570.

Wu, J., S. Liu, G. Liu, A. Dombkowski, J. Abrams, R. Martin-Trevino, M. S. Wicha, S. P. Ethier and Z. Q. Yang (2012). "Identification and functional analysis of 9p24 amplified genes in human breast cancer." Oncogene **31**(3): 333-341.

Xiao, Y., C. Bedet, V. J. Robert, T. Simonet, S. Dunkelbarger, C. Rakotomalala, G. Soete, H. C. Korswagen, S. Strome and F. Palladino (2011). "Caenorhabditis elegans chromatin-associated proteins SET-2 and ASH-2 are differentially required for histone H3 Lys 4 methylation in embryos and adult germ cells." Proc Natl Acad Sci U S A **108**(20): 8305-8310.

Yang, Y. H., S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai and T. P. Speed (2002). "Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation." Nucleic Acids Res **30**(4): e15.

Yang, Z. Q., D. Albertson and S. P. Ethier (2004). "Genomic organization of the 8p11-p12 amplicon in three breast cancer cell lines." Cancer Genet Cytogenet **155**(1): 57-62.

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., G. Liu, A. Bollig-Fischer, R. Haddad, A. L. Tarca and S. P. Ethier (2009). "Methylation-associated silencing of SFRP1 with an 8p11-12 amplification inhibits canonical and non-canonical WNT pathways in breast cancers." Int J Cancer **125**(7): 1613-1621.

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.

Yates, J. A., T. Menon, B. A. Thompson and D. A. Bochar (2010). "Regulation of HOXA2 gene expression by the ATP-dependent chromatin remodeling enzyme CHD8." FEBS Lett **584**(4): 689-693.

Yuan, W., M. Xu, C. Huang, N. Liu, S. Chen and B. Zhu (2011). "H3K36 methylation antagonizes PRC2-mediated H3K27 methylation." J Biol Chem **286**(10): 7983-7989.

Zhang, Q., L. Zeng, C. Shen, Y. Ju, T. Konuma, C. Zhao, C. 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, X., X. Mu, O. Huang, Z. Xie, M. Jiang, M. Geng and K. Shen (2013). "Luminal breast cancer cell lines overexpressing ZNF703 are resistant to tamoxifen through activation of Akt/mTOR signaling." PLoS One **8**(8): e72053.

Zhang, X., K. Tanaka, J. Yan, J. Li, D. Peng, Y. Jiang, Z. Yang, M. C. Barton, H. Wen and X. Shi (2013). "Regulation of estrogen receptor alpha by histone methyltransferase SMYD2-mediated protein methylation." Proc Natl Acad Sci U S A **110**(43): 17284-17289.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li and X. S. Liu (2008). "Model-based analysis of ChIP-Seq (MACS)." Genome Biol **9**(9): R137.

Zhou, Z., R. Thomsen, S. Kahns and A. L. Nielsen (2010). "The NSD3L histone methyltransferase regulates cell cycle and cell invasion in breast cancer cells." Biochem Biophys Res Commun **398**(3): 565-570.

Zhu, L., Q. Li, S. H. Wong, M. Huang, B. J. Klein, J. Shen, L. Ikenouye, M. Onishi, D. Schneidawind, C. Buechele, L. Hansen, J. Duque-Afonso, F. Zhu, G. M. Martin, O. Gozani, R. Majeti, T. G. Kutateladze and M. L. Cleary (2016). "ASH1L Links Histone H3 Lysine 36 Dimethylation to MLL Leukemia." Cancer Discov **6**(7): 770-783.

Zou, J. X., Z. Duan, J. Wang, A. Sokolov, J. Xu, C. Z. Chen, J. J. Li and H. W. Chen (2014). "Kinesin family deregulation coordinated by bromodomain protein ANCCA and histone methyltransferase MLL for breast cancer cell growth, survival, and tamoxifen resistance." Mol Cancer Res **12**(4): 539-549.

APPENDIX A. NSD3-S and ESR1 knockdown microarray results.

Immunoblots (lower right) for ERα and NSD3 (WH) following knockdown of ESR1 or NSD3, respectively, corresponding to microarray samples appear below. Venn Diagram represents genes in the ESR1 knockdown compared to LacZ control (left) or NSD3 knockdown compared to LacZ control (right). Overlapping (left) or NSD3-unique (right) gene sets were analyzed in ToppFun to produce the tables found below the diagram. Full gene lists and associated statistical measures can be located at https://www.researchgate.net/profile/Jamie_Mills2/publications. File NSD3-short knockdown microarray DOI: 10.13140/RG.2.2.13586.25286 and file ESR1 knockdown microarray DOI: 10.13140/RG.2.2.11908.53126.

Biological Process	P-value
Vesicle-mediated transport	3.30E-04
ER to Golgi vesicle-mediated transport	3.70E-04
Pigment metabolic process	7.00E-04
Pigment biosynthetic process	8.20E-04
Intracellular transport	8.20E-04

Biological Process	P-value
Mitotic cell cycle	1.00E-24
Cell Cycle	1.00E-24
Mitotic cell cycle process	1.00E-24
DNA metabolic process	1.00E-24
Cell cycle process	1.00E-24
Mitotic nuclear division	1.60E-22
Organelle fission	9.30E-22
Nuclear Division	1.00E-21
Organelle organization	1.20E-21
DNA replication	4.40E-21
Cell division	4.70E-21
Mitotic cell cycle phase transition	1.50E-20
Cell cycle phase transition	2.80E-20
DNA strand elongation	8.80E-19
DNA strand elongation involved in DNA replication	8.90E-19

APPENDIX B. H3K4me3 ChIP-seq quality control and enrichment reports.

Output from sequencing quality control measures appears in the table below. The figure represents Phred scores for the ChIP-seq sequencing run. All samples are nearly identical, therefore only one is presented (shASH2L-A-H3K4me3). Cross-correlation plots appear on page 208 and enrichment plots on page 209.

Sample-Name	Total Sequences	Type	# reads	%GC
2247_shASH2L-A-H3K4me3	28430917	ChIP	28387407	45
2248_shASH2L-A-input	28847916	Input	28819172	41
2249_shASH2L-B-H3K4me3	28142484	ChIP	28057625	43
2250_shASH2L-B-input	35706228	Input	35668119	43
2251_shASH2L-C-H3K4me3	35126397	ChIP	34986667	51
2252_shASH2L-C-input	32529399	Input	32449969	43
2253_shLacZ18-A-H3K4me3	32483014	ChIP	32393402	48
2254_shLacZ18-A-input	33296775	Input	33219481	44
2255_shLacZ19-B-H3K4me3	35198935	ChIP	35069303	49
2256_shLacZ19-B-input	35411340	Input	35328228	43
2257_shLacZ20-C-H3K4me3	34957285	ChIP	34835920	49
2258_shLacZ20-C-input	46720830	Input	46614287	41

Figure 4. (Legend on next page)

$$NSC = \frac{cc(fragment\ length) - min(cc)}{min(cc)}$$

$$RSC = \frac{cc(fragment\ length) - min(cc)}{cc(read\ length) - min(cc)}$$
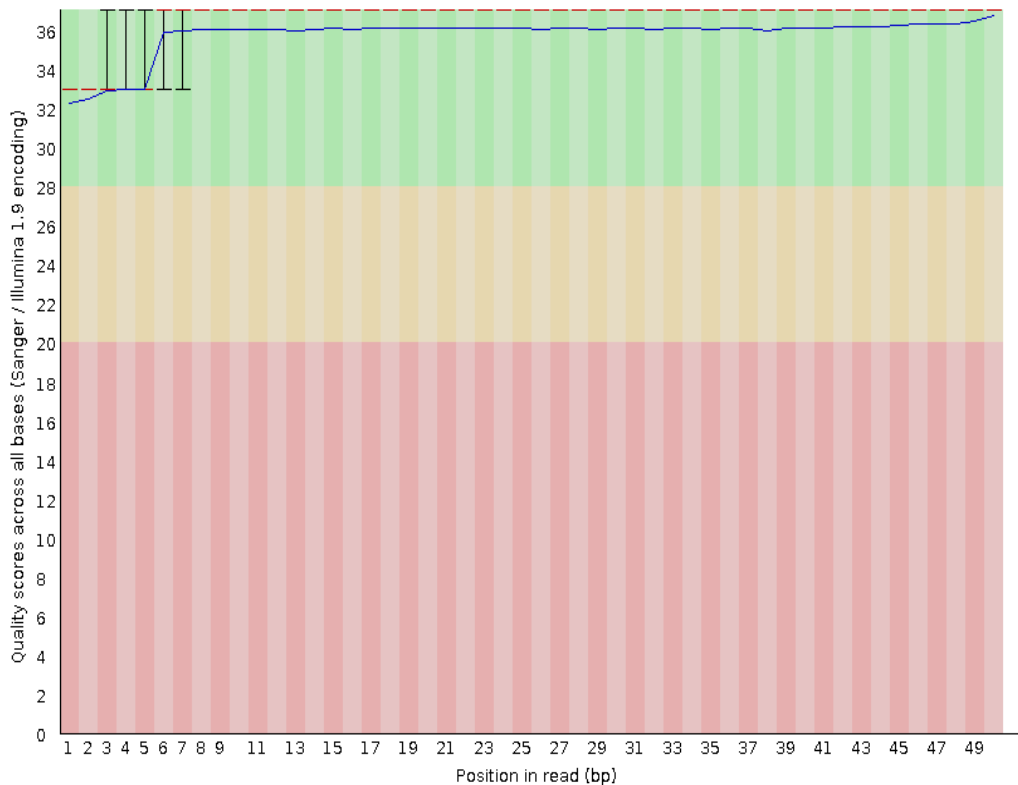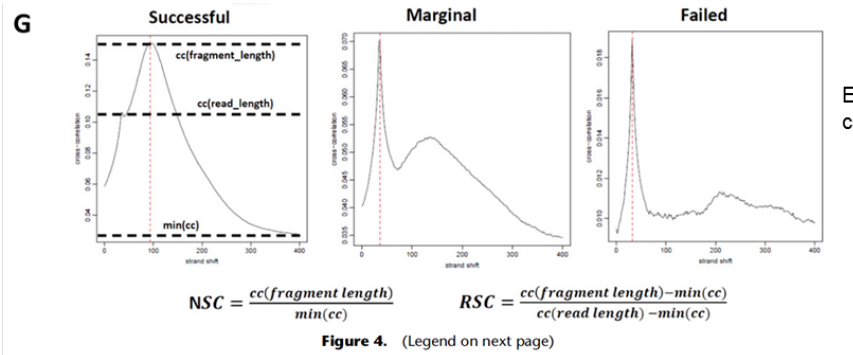
Example of successful cross-correlation plots

LacZ H3K4me3 ChIP samples
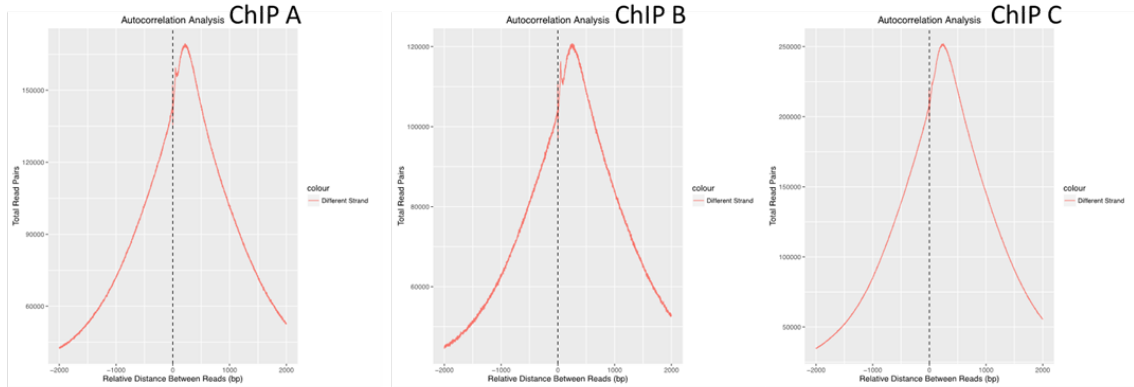
ASH2L H3K4me3 ChIP samples
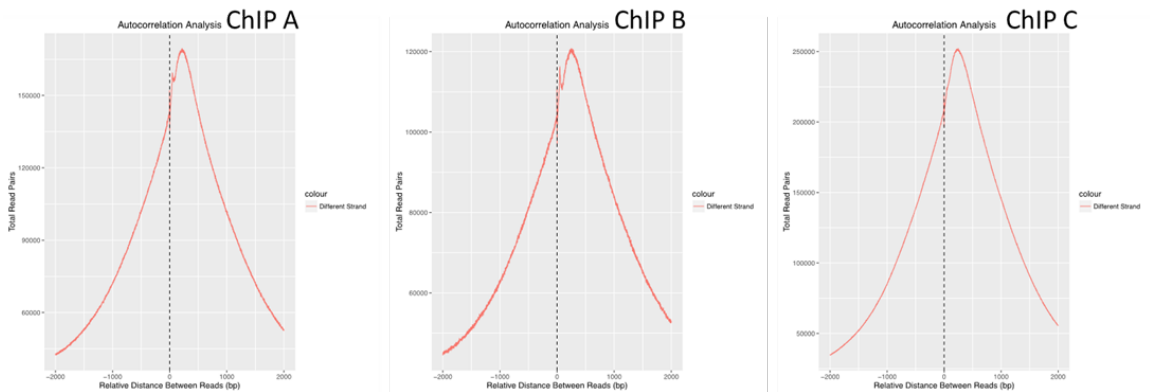
Example: Hoping for a peak here
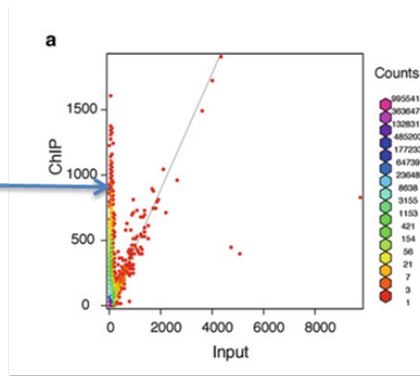
Example of successful enrichment plots

LacZ H3K4me3 ChIP samples
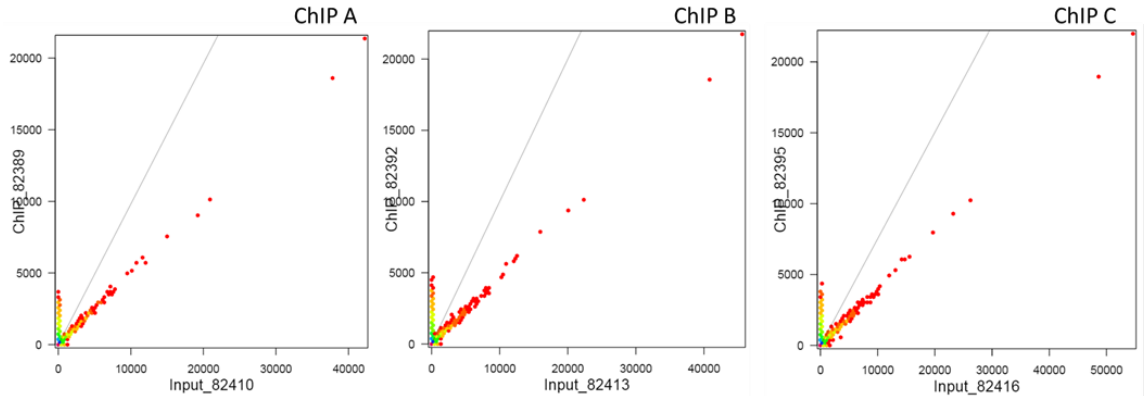
ASH2L H3K4me3 ChIP samples
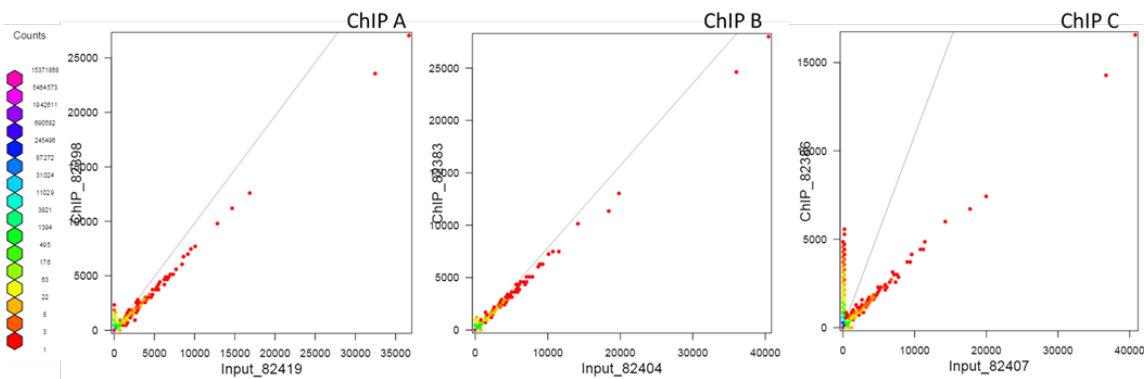
APPENDIX C. H3K4me3 ChIP-seq peak datasets: LacZ control and ASH2L knockdown.

Output lists from MACS2 ChIP-seq analysis can be found at https://www.researchgate.net/profile/Jamie_Mills2/publications. File H3K4me3 ChIP-seq MACS2 shLacZ DOI: 10.13140/RG.2.2.21345.71526 and file H3K4me3 ChIP-seq MACS2 shASH2L DOI: 10.13140/RG.2.2.28056.60162.

Output lists from MACS2 ChIP-seq analysis Venn Diagrams comparing LacZ and ASH2L overall gene lists appear in this appendix. These lists correspond to **Figure 3.2-D** but do not report the 13,090 genes in common.

LacZ only (2,251)	L1TD1	KCNH6	MIR6124	C1orf131	CLEC14A
LOC100133669	PKIG	KCNH7	LOC101927854	COL14A1	LOC101060091
HTR6	GSTM5	FBXO15	SNORA59A	CPZ	DCN
HTR7	KCNC1	PKP1	SNORA59B	APOA4	MIR941-1
LURAP1L	KCNC2	LOC100505716	ADAM18	CEBPE	MIR1289-1
FAM19A2	KCNC3	LOC100289495	BRMS1	TULP4	RPS6KB1
BBX	LOC283683	KCNF1	LOC105376331	IL21R	SNORA31
FAM19A1	LOC79999	HTR1F	LOC101927865	LOC401312	TYW5
FAM19A5	LOC100288181	TNFSF13B	PLD5	LOC100505625	LOC101927694
FAM19A3	BBC3	MBL2	KCNV1	EVI2A	SNORA24
EXOC3-AS1	NEFH	C1R	RPS6KA2-AS1	LOC100288254	CYP2E1
GSTK1	MYBPC3	LOC284950	SPINK14	MAB21L2	NGEF
WDR83OS	MYBPC1	KCNJ1	LOC101927881	C1orf116	ACTL7B
DHDH	MYBPC2	COL26A1	MIR6165	LOC100506869	FPR3
BCR	KCND1	KCNJ8	TBX22	SPRR2F	FPR2
ORAI3	KCND2	ANKRD62P1-PARP4P3	TBX18	SLC35D3	C2
ORAI2	KCND3	CDRT15	CD6	UBA6-AS1	GLYATL1
AGTR1	GSTO1	MIR4776-1	CD5	LOC101927907	GRIN2C
VOPP1	GTF2H2B	LGALS9	MNX1-AS1	XCR1	GRIN3B
ORAI1	PKN3	INHBA-AS1	MIR6132	LOC101927914	SNORA84
AGTR2	SYNPO2	NKAIN3	CFH	LOC440895	DIO3OS
COL13A1	GSTP1	SLC10A6	ELL2	SLC35F4	CCDC162P
TARSL2	PTAFR	LOC101805491	MOCS2	SIPA1L1	F8
SHMT1	KCNA2	FAM193A	MOCS3	TCEAL2	F9
B3GLCT	KCNA3	SSNA1	PLIN1	EMC8	SNORA68
PLBD1-AS1T	LOC642423	LOC101927822	PLIN5	NFYA	DPY19L1P1
RDH11	KIAA2022	FAM35A	DIO2-AS1	C6orf58	FGF14-AS1
RDH16	KCNA5	FAM196B	CKM	C3orf79	SNORA98
RDH14	KCNA6	LOC101927839	HAO2-IT1	CLEC19A	GC
RDH13	ADAM28	LOC101927847	DNM3OS	NPC1L1	SNORA90
GSTM3	ADAM32	ENPP2	DIO2	LDLRAD4-AS1	GGT3P
PKIB	ADAM30	MIR6125	MIRLET7I	DAO	SNORA87
	POFUT2	LOC101927849	LOC439933	DNAJC3-AS1	UBE2D4
	MYBPHL		C1orf158	C3orf22	LOC102724784

TNFRSF11B	CHGA	TDRD12	RGS20	ACY3	RAB32
SOCS2	LSP1P3	LOC339059	NRIP3	ZNF587B	RAB36
COL28A1	NCRUPAR	KCNAB3	RGS22	FSHB	LOC102723344
BDNF	PNMT	RBBP4	CLASP1	PRSS3P2	CSPG4
TNFRSF10C	MIR6086	MROH8	LOC101927798	FSHR	HPN-AS1
SNORA80B	RPL32	RPL13AP3	DLK1	SNORA17A	SLC13A1
SH2D4B	MIR6083	SLC46A2	NOBOX	GCK	SLC13A2
CECR1	ABCA13	MIR320B1	RABL6	EPGN	HERC2P9
LOC102723427	MIR6089	IQCF2	CDCA5	CFHR2	FBXL22
CECR3	GPR179	TSHZ1	CDCA8	CFHR1	CALML3
PTPRZ1	MIR6071	TSHZ2	TPBGL	CFHR5	MIR7978
SLITRK1	DKK3	FAM180A	LOC400553	SLX4IP	TKTL2
SLITRK6	CPED1	SRGAP2-AS1	F13B	SNORA11B	XXYLT1-AS1
ENTPD1	SNORA74B	RAB40AL	LOC150935	SNORA111	GPR37L1
AARD	MIR6078	CX3CL1	LOC101243545	FTO-IT1	CYCSP52
MST1L	MEG9	ANKRD7	SLC7A3	SYT2	EFCAB6-AS1
LOC102724710	OCSTAMP	B3GNT5	PTF1A	HIGD2B	TWIST2
SWI5	VTRNA2-1	FAM181B	FRY	SYT8	USP32P1
TG	CPEB1	FAM181A	LOC100132735	SYT6	HBM
TH	GPR182	AKAP6	CIR1	CTD-3080P12.3	LOC100288748
SNAR-C3	FRMPD2B	AKAP4	MFNG	MS4A7	RAB6C
SLC7A13	ELN	LOC100506682	NDC80	ANKS3	HDC
SNAR-C5	GPR132	LOC100506688	IGSF11	AGPAT4	GABRP
SNAR-C4	SNORA70B	OLAH	RLBP1	SPOCK3	FAIM2
HYAL1	RNF165	LOC102724804	LOC389602	SPOCK1	NANOS2
FAM170A	RNF139	PPY2P	LOC102724511	ACKR2	NANOS3
SNAR-B1	LIMD1-AS1	CIB4	TTC39C	CISTR	LOC55338
CRIP3	CHML	CIB3	MCM3AP	ACMSD	HRASLS2
SNAR-A7	GPR156	FAM186A	PCBP2-OT1	ZFPM2-AS1	SCUBE1
CRIPT	MYBPH	ELMO1	PRKCQ-AS1	C10orf55	LOC100144595
LOC101927623	GPR142	DLG1	NFIA-AS2	H19	CETN4P
ECD	KRT8P41	TGFBI	MIR6516	LOC102723362	LOC642943
PHOX2A	COL16A1	ANKRD44-IT1	MIR7846	TIMP3	OACYLP
LOC101928977	CHRM3-AS2	HSPB9	MIR6513	TIMP1	USP46-AS1
CHD5	CHPF	LRRC74A	EIF1B	LOC388242	EPB41L4A-AS2
LOC101927640	ERG	BBOX1-AS1	MIR6504	TP53TG5	CCBE1
RGPD4-AS1	IQCA1	C5orf66-AS1	MIR6502	C10orf91	CNBD1
FAM178B	PAUPAR	KCNA10	LOC101926940	LOC102724652	SPARCL1
LOC101928988	PPIAL4A	ECT2L	SMIM10L2A	LOC102724651	MIR5095

NYAP2	TM4SF4	SERPINE1	NTSR1	CCDC18-AS1	ACOXL
SNORD2	TM4SF1	EARS2	TVP23C-CDRT4	AGR2	LOC100996624
MGC34796	CCAT2	ACOT8	NTSR2	JAM3	LOC100996634
LOC339568	STX1B	LOC400661	NMBR	LINC00989	CNR2
MIR6715A	LOC103312105	SERPINA3	VCX2	CYSRT1	LOC100996635
PTH2R	PPP1R3A	C17orf50	MIR7705	EFEMP1	LINC00911
PGAM4	TSPAN8	SERPINB5	ESM1	UGT2B10	LINC00901
ERVFRD-1	TSPAN1	HSPA1L	CEACAM3	ATP11C	NALCN-AS1
LOC100652999	BPIFB2	LOC102724484	NRN1L	LGSN	MRGPRE
BEND3P3	PPP1R1C	C17orf67	ADRB3	SRSF5	WDR72
GPR65	MIR550A1	TUSC8	CEACAM6	LVCAT5	WDR77
LRRC70	ANKRD18B	ATP8A2	ESRG	LOXHD1	LINC00882
LRRC71	MIR550A3	TUSC3	MIR217HG	TMEM26-AS1	SYNJ2-IT1
SLC9C1	AKR1B1	C17orf99	C10orf107	OVOL1-AS1	LOC649133
LOC100129316	KIF23	TECTA	C10orf113	BFSP2	MIR8067
SH2D6	PDZRN3	RNF216-IT1	C10orf126	ZBTB20-AS4	MIR8065
SLC9A3	TMPRSS9	SLX1B-SULT1A4	BDKRB1	UGT2B28	LINC00842
SLC9A4	STX19	C17orf80	MIR5197	ARL13A	LINC00845
LOC100505920	RXRB	ITPR1	MIR5189	GZMB	MIR8059
GPR87	PRLH	ATP8B4	C14orf119	SRD5A3-AS1	LINC00877
GPR88	RXRG	TECRL	RWDD3	ANPEP	IGLL5
GPR85	RFTN1	PSCA	LOC100192426	WWC2-AS1	IGLL1
GPR84	MIAT	SYCP1	KIAA1024L	FAM9B	LINC00865
LOC100287072	PRND	PSD3	C14orf105	MTRNR2L6	LINC00867
LRRC52	DOK5	SLC24A2	LOC400997	MTRNR2L3	LINC00861
LOC100130331	FUT9	SEMA3D	ZC3H12D	ANKFN1	RERGL
SLC38A4	MIR5008	SEMA3B	LOC158434	CACNA2D3-AS1	SSUH2
LOC100287010	EGFR-AS1	ANKRD30B	NPBWR2	LOC100287792	MIR8082
IQCH-AS1	LOC102724297	CMA1	RPS14P3	C6orf141	MAP3K19
LOC100129345	AFF2	MIR1972-2	FAM111A	LOC102724096	DRC3
L3MBTL4	B3GALT5-AS1	SYCP3	MRVI1	ETS1	MIR8079
PRCD	ELMO1-AS1	LOC101926908	HLA-DRB1	FAM3B	PARD3-AS1
IFI16	SSTR1	LMAN1L	PGCP1	PHGR1	CSTL1
SLC12A2	CRMP1	SLC6A12	LINC00977	ACOT12	PLA2G12B
FCGR2B	CLMP	PLPP1	LINC00970	LINC00922	DNAJC7
FCGR2C	PDLIM4	SLC6A20	LINC00971	UBA7	HCG17
ZNF790-AS1	PLA2G2F	PTDSS1	CHST9	CNN1	MIR135A2
ANTXRLP1	S100G	DPH5	CHST7	LINC00954	GJA10
HNF1B	LL22NC01-81G9.3	LOC100505795	LINC00967	LOC102724053	OR7E156P

PWRN4	LYN	THEMIS	PATL2	SLC9A3R1	PEBP4
FEZF2	NOX3	RPGRIP1L	SCGB2A2	FRG2DP	NRTN
PWRN2	DLEU1-AS1	TAC4	VSTM2A	TBXAS1	FAM150B
CTD-2151A2.1	UCN3	TAC1	KCNJ13	SP140L	WBP11P1
AMIGO2	FLVCR2	SLC52A1	KCNIP4	LINC00581	RARRES3
ZDHHC22	VWC2L-IT1	BMP5	KCNJ15	LINC00578	VIL1
GDPD4	OR14A16	BMP2	PARTICL	LINC00570	OR51A2
PPP2CA	LOC338694	CPS1	HABP2	LINC00571	MUC21
EGFLAM	LOC103344931	PNLIPRP2	VCX3B	LINC00572	C1orf94
ANKRD34C-AS1	LYRM2	USP12-AS1	ARHGAP26-IT1	MIR486-2	ADGRE5
LINC00838	LOC653786	CDHR3	OTOG	SMIM18	ADGRF1
CDH26	CCDC144CP	RAB40A	RAB39A	MUC5AC	ADGRF4
LINC00824	MYLK2	SRD5A1P1	CXCR6	LINC00563	PKD1L1
MIR3622B	NRROS	LOC653712	DISC2	SPACA7	ASB15
MRS2P2	CXCL17	NPVF	AAMDC	LINC00592	EYA1
GAGE10	CXCL14	LAMB3	CTAGE10P	LINC00593	LOC105375650
MFSD4	CPB2	SCIMP	SNAR-A12	TBX5	GABRG2
COL6A3	DSEL	BCL6B	CPLX3	PINK1	FAM169B
AKNAD1	CPA4	FAM231D	KLRC3	LINC00589	HYALP1
CDH17	C7orf65	FAM231B	CXCL8	CPNE8	LINC00400
CDH19	FOXN1	SKP1P2	SPO11	CPNE7	MIR7151
LTF	SMARCA5-AS1	NBAT1	DOCK9	CYP3A7-CYP3A51	MIR7150
CST9L	C7orf69	NKX2-3	LOC100288911	ZNF890P	LINC00434
LTK	SCHIP1	PCSK6-AS1	LINC00538	NUDCD1	LINC00424
ATP1A4	CELF3	CDK15	LINC00523	OR51Q1	CEP83
ATP1A2	DSG1	FAM101A	LINC00524	FMR1-AS1	PAM
OPCML	SLC25A11	COL21A1	LINC00525	SFTPB	MESP2
NOS2	MFF	GCNT4	LINC00520	SFTPC	AIFM2
SAMD14	SLC25A31	LOC145474	CRABP1	SLC25A25-AS1	HADHB
LIPC	NSMCE2	GUCA2A	SHANK2-AS3	DEPDC4	CEP63
LIPM	SFRP1	TRMT5	LINC00558	HSD17B3	MIR4472-2
GABARAPL3	CTXN3	BEST2	LINC00559	OSM	ALOX5
CSNK2A3	COL9A1	BEST3	CRB2	MYO9B	GPC5-AS1
FLJ31356	RERG-AS1	BEST1	LINC00511	ACTL8	GABRA5
HS3ST3A1	MLN	NEUROD1	IL12RB2	SMAD1-AS1	DHRS1
HACD2	DIRC3	LOC100130691	LINC00507	FAM133CP	DHRS9
LY9	PLVAP	WEE2-AS1	CENPO	LOXL3	AGBL1-AS1
MIR1973	ACRBP	NEB	ADGRL3	NR3C1	LINC00499
GPIHBP1	MIR1910	BNC1	CDKL4	LOXL2	LINC00494

FRMD4A	FAM163A	ANP32A-IT1	C11orf86	IGSF1	PRPF38A
VASH1	TRPV5	DIAPH3-AS1	LINC00609	PYHIN1	MIR548D2
GOLGA8G	NR2E1	SPATA12	LINC00606	KLLN	LARP7
PIP	AMTN	MIR3156-2	LINC00603	SERP2	L1CAM
LINC00460	MIR7110	MGC16025	PRSS45	TFEC	LGALS17A
LINC00462	MIR7111	LOC100996455	LINC00639	LOC102546299	MIR130A
CHMP1B2P	MIR7109	EDA2R	MIR548AE1	TFF1	MIR3973
HMGCS2	CLDN2	ANO4	CCL17	MRRF	EMX2OS
MIR1273E	CLDN1	ANO5	BRD4	ETNPPL	NBPF25P
MIR1273H	PSMD5	ANO1	PRSS35	MGC32805	SCTR
LOC100130880	FAM133A	RAMP3	LINC00626	GALNT12	BBOX1
LINC00485	IDO1	SLC30A8	C11orf53	EFTUD1	BMP8B
LMF2	CLCNKB	NPSR1-AS1	NKX6-3	SCGN	MIR3938
THSD4	AGAP2	LINC00709	LHCGR	COL4A3	LINC01467
CCDC70	MRPL14	CTB-12O2.1	NKX6-2	SCG2	LINC01468
LINC00476	AGAP4	COL5A2	ITGA11	MGAT4EP	MIR1343
LINC00477	COL24A1	LINC00701	VAC14-AS1	RALYL	LINC00160
LINC00469	TDP2	LINC00702	NACAD	DKFZp434J0226	BRE-AS1
CCDC60	WFIKKN1	SCARF2	LOC145845	AGAP6	LINC01483
KLHDC7B	SNAR-H	GALNT8	LOC102546228	HKDC1	LINC01485
OR52N2	EFCAB9	TEX13A	LINC00692	CAMK1G	LINC01482
OR52M1	MYT1L-AS1	HSP90AB4P	DUSP26	GMPR2	RFXANK
CTAGE15	MIR3689C	FAM132A	POM121L2	TFPI	G6PC2
C2CD4A	LEFTY1	FAM132B	SLC16A12-AS1	B3GAT1	SP5
TRPM1	MRPL47	GALNT4	RP1	PGLYRP2	SP9
TRPM5	SMTNL2	BPESC1	C4orf45	MIR3929	LINC01479
LOC105747689	COL11A1	FAM131C	LINC00687	MIR3935	MIR548A1
HEPACAM2	TMEM75	RPLP0P2	MRLN	EPSTI1	MIR548A3
PRL	PSMB3	TMEM92-AS1	SALRNA3	MIR548F5	MSLN
CLDN14	CEPT1	LINC00648	KLF3	MIR548F4	UNC80
HNRNPH2	TMCO3	LINC00644	C4orf26	UNC5C	LINC01426
CLDN17	WARS2-IT1	LINC00683	PCED1B-AS1	MIR3923	LINC01422
CST6	PTOV1-AS2	GALR3	MYOZ2	MIR3925	LINC01423
CST5	RBPMS	LINC00675	MUCL1	MIR3924	MSMB
IFNB1	DCDC5	LINC00667	FANK1-AS1	SCO1	CERS3
KBTBD8	PTPRC	PRSS23	SMIM2	MIR3920	MIR548AQ
CHRNB2	MIR8071-1	MYOM2	IGSF5	FAXDC2	MIR548AV
LMO3	FASLG	LINC00615	SMIM5	ADAMTS12	MIR548AD
PYY	RDH10-AS1	LINC00613	IGSF6	ADAMTS14	MIR548AC

LINC01416	FREM3	MTTP	LINC00313	BRINP2	LSM3
MIR548AL	ZSCAN10	C12orf80	LINC00304	BRINP1	MIR1206
MIR548AM	FAM120AOS	C12orf71	TEDDM1	C9orf129	MIR1207
MTUS2-AS1	ADAMTSL1	C12orf74	DSCAML1	FAM181A-AS1	MIR1208
NOP56	GAST	LINC01270	MIR3714	GABRG3-AS1	MIR1202
PDCD1	ASAP1-IT2	LINC01272	LINC01609	TAF11	MIR1200
LINC01432	ANGPTL2	ZNF341-AS1	TUBA3E	ASB5	LINC00269
STH	ANGPTL4	LINC01268	LINC01608	LINC01507	LSP1
MIR1293	LNX1-AS1	LINC01266	VENTXP7	OR1I1	SGCD
TNFSF18	OLFML3	LINC01256	NRG3-AS1	LINC01525	LINC01549
MSR1	LINC01354	LINC01258	ASPRV1	PRCAT47	LINC01543
DDIAS	LINC01350	LINC01298	LINC00383	LINC01526	LINC01544
TNFSF12	TNF	TM4SF20	ADCYAP1	LINC01523	PROSER2-AS1
APOBEC3A	TNR	LINC01287	OVCH1-AS1	GSDMC	ASPG
MIR1284	LINC01344	MIR26A1	MIR374B	RNASE10	NLGN3
LDHAL6B	TOX	LINC01280	CPAMD8	TMEM196	LINC01571
LINC00052	LINC01333	LOC100128993	LINC00399	ZFAT-AS1	LINC00240
SYP	LINC01335	ZMAT4	BCL2L14	LINC01510	LINC00243
C14orf1	LINC01375	CPSF2	CNTNAP2	LINC01512	LINC01574
MRPS34	LINC01370	SELV	LINC00330	TMEM182	LINC01570
PRICKLE2	TPR	UCN	MIR378G	MIR5689HG	MIR1256
GAD2	ESPNP	HEPH	MIR378I	SLC2A14	MIR1252
APTR	LINC01361	C15orf43	LINC00359	CACNA1C-IT2	LINC00238
LINC01399	LINC00028	C15orf53	LINC00348	OR1M1	LINC00239
CLYBL-AS1	LINC00029	C15orf54	LINC00343	TMEM159	LINC01564
TGM1	SYNE3	GBP2	MMADHC	LINC01594	CDRT7
TGM6	LINC01304	NXF5	ART4	LINC01591	LINC01559
RAG2	PROX1-AS1	MUC6	ARSI	DOCK4-AS1	LINC00222
TGM4	LINC01307	CAPN14	TAS2R38	LINC01581	LINC01553
CARTPT	DKFZP586I1420	FSIP2	CCNJL	DEFB112	LINC01554
FGF11	LINC01331	LINC01621	TAS2R60	SSBP3-AS1	CBLN1
STEAP3-AS1	TTN	LINC01622	BANK1	CALCRL	SCARNA3
PRICKLE2-AS3	RNU11	ZFYVE19	DEFB132	LINC00282	CD101
IL1RL2	LINC01322	LINC01612	DEFB130	ASMT	KPRP
OXCT1	MIR147A	LINC01614	BANF1	ANKRD29	PDE1C
TGS1	HDAC11	LINC00320	BANF2	LINC00276	ATP13A5-AS1
ZPLD1	AQP7	KCND3-AS1	PDGFRA	ANKRD24	UBAC2
MSANTD1	AQP5	LINC00310	VCX	ANKRD22	GDF7
MBNL1	TXK	LINC00311	HS1BP3-IT1	LINC00271	SCARNA5

SCARNA6	SYT16	MMS19	MYLK	TMEM198B	PMCHL1
CLIC6	SUGCT	LOC101928008	FLJ16171	LINC01140	PMCHL2
CLIC5	TSPAN18	LOC101929331	JPH2	LINC01141	OR5T2
STAB2	LOC154449	CDC14C	JPH3	NBPF7	AACSP1
PRMT5	HHIP	ARMC2-AS1	MYOG	LINC01137	BZW1
LINC01499	LOC63930	SNRPN	MYOC	LINC01176	LINC01106
RAPSN	RIIAD1	BAIAP2-AS1	MIR1537	LINC01170	LINC01107
TLDC2	LOC101929199	LOC101929384	SH3BP5-AS1	C5orf64	LINC01102
THRB-AS1	STAT5A	LOC101928052	LINC01227	C5orf67	LINC01103
PPP1R26-AS1	REG4	EIF2B1	TCP11	C5orf66	LINC01105
MANEA	CTB-178M22.2	LOC101929395	S100A2	LINC01164	ADAM8
PDE8B	CD200	CATSPER3	FRMD1	LINC01162	LINC01133
AMY1B	IL20	ANGPT2	FRMD3	LINC01163	LINC01126
ASIC4	TBX2-AS1	SNX29P1	FRMD7	CRYBB2	MSC-AS1
ASIC3	WFDC3	KIAA0586	LINC01213	RGS5	SERINC4
DNAJC16	IL18	P2RX2	S100A9	ARID2	TUBA8
FRMD6-AS2	EHD4-AS1	LOC101928093	S100A8	RGS8	GOLGA8CP
TMEM100	IL19	TRY2P	GIMD1	GGT1	DMRTA2
OR2F1	PABPC3	KRT3	LINC01247	RGS7	TEX36-AS1
ARID3C	MGAT2	TLN2	LINC01248	MAT1A	TNNI1
MGC15885	IL34	TLN1	MIR4444-1	CALCB	LINC01070
NDST4	NUDT7	PGAM1P5	B3GALT6	NXPH2	LOC101929064
GATA4		2-Sep OR4E2	CLLU1	LOC286437	C6orf222
PYDC2	OR5K1	ZNRF2P2	LINC01235	KTN1-AS1	LINC01095
HOXC12	EPHA5-AS1	SOSTDC1	NALT1	CHURC1	LINC01096
LOC100507065	FBN3	TLR9	EFCAB11	CNTFR	LINC01091
LOC100268168	C8orf49	TLR7	ATP7B	ARIH2OS	LINC01093
LOC101929207	SNTG1	TLR4	UPP2	UNC13D	GOLGA8IP
FAM201A	TFAP2D	FLJ42627	LOC101929034	HNRNPKP3	OR6B3
LIMS3	TMEM233	ANXA10	SHCBP1L	HOXB1	LINC01085
GMDS-AS1	FBN1	PCDHA10	LINC01206	LOC442497	LOC101929099
ADIPOQ	GRM7-AS2	PRPF4	LINC01204	HIF1A	CIDEB
LINC01010	RER1	EFNA2	CTCFL	KIAA0319L	LINC01082
LOC101929241	MAN1B1	PI4KA	ATP5S	H3F3C	CCT6A
MFSD6L	LOC100507006	MIR8069-1	XRCC4	FOXG1-AS1	LINC01029
CTAG2	NYNRIN	MIR1825	XRCC3	H2AFY2	TNNT2
LOC101929282	ZC2HC1B	COLEC11	INTS12	HIF3A	LINC01016
TSPAN32	NAIF1	MKNK1-AS1	LINC01151	PDCD4-AS1	LINC01017
KCTD21	TPRG1	OPA1-AS1	C5orf46	NBPF13P	TOR4A

RANBP1	MIR4424	FBXW12	LOC101927123	MIR4499	PADI4
STEAP4	MIR4427	SPINK8	LOC101927131	MIR4494	PADI6
ARHGAP44	MIR4422	GRIP2	ADCY4	MIR3165	MIR569
BASP1	TRIM63	PCP4	LOC101927142	ACTRT1	MIR577
PDIA2	MIR4418	OR9Q1	ARFRP1	MIR4481	MIR570
ARHGAP19	TRIM54	LOC284080	FGD6	LOC102467226	LOC101929413
A2ML1	TRIM52	LOC100506127	TPMT	PDPN	MIR587
NMRAL1	MIR1302-1	CNTN5	NDUFC2-KCTD14	DRT15L2	TRAM2-AS1
ARHGAP23	MIR1302-4	CNTN6	LOC101927166	MIR3153	LOC285629
NPY6R	MIR1302-5	CNTN1	FGF7	MIR4470	LOC101929420
MIR2909	C16orf86	CNTN2	CSGALNACT1	LOC102467212	RAB11B-AS1
OPRM1	MIR1302-6	TNFAIP3	TMEM178B	LOC102467213	LOC101928105
CYP1B1-AS1	TP63	CYTIP	EDAR	LOC102467214	ELOVL3
LOC101929625	GIN1	LOC100507477	ZNF793	LOC102467217	LOC101929439
MIR30B	SLPI	SHROOM3	DDX11L5	TNKS2-AS1	POU5F2
MIR30A	TRIM15	TCAF2P1	LOC101927189	MIR4478	LOC101929448
TEX40	RIN3	SHROOM4	LOC101927190	MIR3146	LOC101929452
TIPARP-AS1	NACAP1	LOC100507468	SVILP1	MIR3140	LOC101929450
MOV10L1	MIR4325	SUMO1P3	CTDSPL	MIR4474	LOC101928135
LMCD1-AS1	MIR4328	EMC3-AS1	MIR4532	TCTEX1D4	KPNA7
BTBD19	MIR4320	SLC3A1	MIR3201	LOC100507334	LOC101928140
LOC101928322	FER1L4	LOC101929608	DDC-AS1	MIR3181	UNQ6494
MIR340	SPHKAP	TTC13	MFAP3	MIR184	BEAN1
NANOG	MIR4311	PRSS1	MIR3139	LOC100507351	LOC101929488
LOC101928333	MIR4310	WIPF1	MIR4465	LRRTM3	SCGB2B3P
LOC101927023	MIR4312	NPIPB11	MIR4460	HLX-AS1	SRD5A2
MIR30C2	OR9G1	SLC4A5	MIR3123	LOC101929709	LOC101928162
RIC3	ECM1	EVA1A	MIR3122	LOC101929717	LOC102467655
SNX17	GRIK5	GJA4	MIR4450	IQCG	ZNF658
DPP10-AS1	NMRK1	RSPH6A	ZNF829	IQCK	ZNF652
LY6K	MIR4300HG	LOC101928435	MIR4442	ALDH1A2	POMT2
LOC101927053	MIR4290HG	MIR205	MIR4439	DAW1	WISP1
LOC101928386	MIR5691	FAM171A1	MIR4431	MIR139	WISP2
LOC101927055	SCARNA26A	GJC2	MIR4430	MIR145	C2orf81
LY6H	CYP8B1	MIR211	ZCCHC13	ERVH48-1	EVADR
PARD6G-AS1	LOC100507406	PLAC1	C2CD3	MIR554	MIR6829
IP6K3	MIR5682	LOC101928449	MIR3173	MIR556	MIR6811
RHPN1-AS1	GRIP1	LOC101929771	VWA2	SLC5A7	ZNF721
LOC101927087	SPINK4	LOC101928443	MIR3169	PADI3	MIR6761

MIR6757	LOC100507175	EDNRA	ANAPC1P1	LOC100506526	MIR3653
LOC441601	MIR455	MIR6888	LOC101927450	MIR4798	CXorf65
MIR6741	CRACR2A	MIR6889	LOC101928782	NBPF20	MAGEL2
DCTN1-AS1	LOC101928201	CA3-AS1	RPL13AP20	MAP1LC3C	LOC284632
LOC101241902	SLC6A5	MIR6884	LOC101927460	DDR2	MIR605
CHRDL2	LOC101928211	JMJD6	AFMID	SNORD17	TMEM132B
FER1L6-AS2	SLC6A4	WFDC11	LOC101927472	TBC1D30	MIR618
ACADL	LOC101929541	MIR6854	LOC101927481	LOC101928708	LOC644762
TMEM150C	GLI2	MIR5590	MIR4802	FAM71D	REEP2
ACTN1-AS1	LOC100507195	MIR4269	SPP2	ZNF33BP1	REEP6
MIR544B	RIPPLY2	MIR4268	ZNF469	FAM71C	SNORD137
4-Mar	KRT7A2P1	MIR4263	POM121L4P	CAV3	MIR3692
FRMD8P1	OCA2	CLLU1OS	MIR708	CAV1	CEBPA-AS1
MIR6772	LOC101929563	MPPED1	CASC6	IRAIN	GRTP1-AS1
MIR6773	LOC101928233	MIR4254	OLFML2A	LOC101927406	LURAP1L-AS1
VIPR1-AS1	LOC101928241	MIR5692B	CASC1	LOC101928738	PECAM1
DUOXA2	SULF1	CACNA1C-AS2	RBM5-AS1	NAT8B	MIR3686
CAPN3	LOC101929577	GLYCAM1	SPATA8	LOC101927412	MIR2355
SATL1	POU4F1	S100A7L2	BRIX1	LOC101927416	MIR3681
MIR548Q	FIG4	KRT78	ODF4	LOC283177	RPEL1
CCDC148	MIR421	MIR4297	POM121L12	TTI1	NLRC3
MIR548X	SARM1	MIR4282	ISY1	NAV3	NLRC4
CCDC140	FILIP1	GPR158-AS1	ELFN1	SNORD46	MIR3678
LOC100507205	SLC28A2	COBLL1	ZNF408	LOC101927421	MIR3672
TREH	CSF3R	MIR4289	PGM3	LOC101928896	MIR365B
ZNF385D	ZNF559	MIR4288	UGT2A3	LOC101927560	MIR4999
LOC285593	LOC101928273	MIR4284	UGT2A1	DDX11	MIR3664
NHEG1	SKINTL	MIR4287	SPTB	LOC101927571	MIR3660
LOC440311	SLC15A3	MIR4270	ITGAD	LOC101927592	CBLC
DBX1	SLC15A5	MIR4272	SPATC1	DEAR	MAGEA9
MIR492	STX18-IT1	MIR4276	ITGA8	LYPLAL1-AS1	METTL1
CCDC185	HOPX	MIR4275	UGT2B7	MIR629	MIR2278
CCDC182	ZNF536	MIR34A	CFAP58-AS1	ABCC12	LOC100506403
CCDC179	EFHB	KRT82	LOC728730	MIR3616	HNCAT21
CREB3L3	KIAA1671	KRT85	MIR4771-2	KIAA1024	SLC22A16
MS4A15	LOC101928298	CMTM3	CYP4B1	MIR635	SLC22A11
LOC100507144	MIR4302	LOC153910	CAHM	MIR634	DCST1
PIH1D3	BACH1-IT2	MIR383	UNC93A	MIR645	GPBP1L1
LOC101929529	SKOR1	LOC101928778	NAIP	ABI3BP	UBL4B

NBR2	LOC101927285	GPC6	MIR4743	PTCHD3
HILS1	F13A1	LOC102723831	ARMCX4	PTCHD4
IGSF11-AS1	MIR944	CD69	MIR4735	LOC100506207
LOC101928812	TPRG1-AS2	GSN-AS1	JMJD1C-AS1	GDF10
H1FOO	C9orf92	EP300	MIR4725	POU6F2-AS1
CHMP4A	C9orf62	LOC100132057	UGT3A2	AXL
LOC101928834	NCK1	OSCAR	MIR4720	MIR2053
NLRP8	ATP11A-AS1	SH3PXD2A-AS1	DSCR8	LOC102467080
NLRP1	SLCO1B7	GUCY2EP	LOC729987	NUP50-AS1
GAS2L2	MIR4655	RASGRP2	SSX8	OR5AC2
LOC100133050	LOC645752	SIRPB2	MIR4708	APLNR
LOC101927523	MIR4651	A4GALT	MIR4715	HDGFL1
LOC101927526	MIR4654	DDX53	MIR4713	ABCG2
YBX2	NDUFB9	TMIGD1	ALK	PPP4R1L
MIR592	MIR4644	NDUFS1	C16orf47	PLCXD3
CASP4	MIR4643	DDX25	MIR4789	MIR95
VWC2L	MIR938	LOC101928516	MIR4786	FAM60A
LOC440446	MIR933	ATP6V1B1-AS1	TRAF2	RMND5A
LOC283299	C9orf57	GNAT1	TRAF1	ZNF189
LOC100506497	SNORD45A	DAPK1	LOC729970	THSD4-AS2
SND1-IT1	C9orf24	LOC101927359	MIR4777	THSD4-AS1
LOC101928539	IL12A	LOC101927354	MIR2117	LOC101927311
CCDC148-AS1	ZNF436-AS1	GPN3	MIR4774	
RBFOX3	LOC102723854	LOC101928694	LUZP6	
EID3	LRRN4	MIR874	CLEC3A	
ZNF277	CD28	MIR4703	CLEC3B	
ABCB4	CD22	LOC101927379	MIR4766	
ABCB5	DIRAS3	5-Mar	MIR4767	
LOC100132111	CCR8	SCARNA18		4-Sep
LOC101927244	CCR7	LOC101927391	MIR4762	
LOC283440	CD36	RPS3	MGAT4C	
LOC101927243	CCR2	GPR1	CLEC2B	
CLUL1	CD53	FMO1	MIR4755	
LOC101927257	RTCA-AS1	FMO4	ABCC5-AS1	
LRRD1	GLCCI1	SCARNA27	FAM65B	
FOXCUT	MKRN9P	LHX4-AS1	STAM	
FABP3	KIF25-AS1	FMOD	MIR4676	
KCCAT333	GPC3	CDH6	FAM69C	
NCF4	LOC102723828	CDH5	VLDLR	

ASH2L only (416)

BLM	PARP3	KPTN	SNORD104	KRT16P33	SNAR-A6
DRAP1	NOL8	DNAJC27	KRTCAP2	KRT16P34	FAM171B
TAX1BP3	TMEM9B-AS1	FRMD6-AS1	ABCA2	KRT16P35	LRWD1
VPS35	AKT1S1	SGSH	PIGG	KRT16P36	MIR6087
C1orf109	WBP11	SH3YL1	C9orf85	KRT16P37	PKI55
C6orf52	RBM18	SMPD4	ZNF174	KRT16P38	GPR150
GBAT2	PREPL	NUDT2	ZNF143	KRT16P39	ABHD8
TYW3	UGDH	KCTD10	TIGD7	KRT16P40	CHRD
SNORA57	FAM161B	UBE2M	GPATCH11	KRT16P41	PPIAL4F
SNORA7A	HAGLR	YTHDF3	PHOSPHO2	KRT16P42	FAM27L
PSMD5-AS1	HIF1A-AS1	OTUD6B-AS1	KRT16P4	KRT16P43	RNF32
PCSK7	PSMA7	FTSJ2	KRT16P5	KRT16P44	OVOL1
MRPS7	AASDHPPT	RASAL2-AS1	KRT16P6	KRT16P45	FRG2
CHFR	ATG9A	RQCD1	KRT16P7	KRT16P46	RELL2
STRN3	TEN1	DPY19L3	KRT16P8	KRT16P47	NFKBIB
LOC101927795	RIC8B	PRR19	KRT16P9	KRT16P48	HSPB3
ACP2	MORN2	SKA2	KRT16P10	KRT16P49	POLN
CDC37	KLC4	HOXC5	KRT16P11	KRT16P50	SYS1
ERMARD	NACC1	METTL18	KRT16P12	FBXO18	SNORA11E
TIMM9	USP46	TGIF2-C20orf24	KRT16P13	IGF1R	LDB2
NIFK-AS1	SCO2	ALKBH6	KRT16P14	DIAPH2-AS1	TIMP4
RPRD1B	DZIP3	GID8	KRT16P15	GRM3	SOCS2-AS1
ILK	KRIT1	NCAPD2	KRT16P16	REREP3	GPNMB
GTF2H1	SDF4	ZNF891	KRT16P17	SSTR5-AS1	RPL10A
GTF2F1	SDF2	SMC5	KRT16P18	HECW1	MIR663B
C17orf89	DKFZP434I0714	LOC101928438	KRT16P19	ST3GAL3	ESCO2
RAD51	RNASEH1	PAAF1	KRT16P20	DGUOK-AS1	LEF1
CMC2	CAPN15	PPHLN1	KRT16P21	LOC401324	CKS2
PSMB8-AS1	PRDM10	RIPK2	KRT16P22	LOC102723769	SH2D7
IFT122	GATC	ELOVL2	KRT16P23	FAM47E	HIST2H4B
SRSF9	HIST1H2BG	MIR6733	KRT16P24	FPGT	LOC388436
MIR1184-2	HELQ	CCDC122	KRT16P25	SPECC1L	C14orf180
MKKS	ST7-OT4	CCDC189	KRT16P26	DBI	TTTY23B
YY1AP1	CAPN10	ZNF271P	KRT16P27	PRKAR1B	MIR3914-1
CIRBP-AS1	RAB11FIP2	NUDT16	KRT16P28	DPY19L1P2	ESX1
WDR54	ZBED5	HPS4	KRT16P29	PCSK2	MAMDC2
COG8	THYN1	MIPEP	KRT16P30	SNAR-C2	LINC00960
	ERP44	ITGA3	KRT16P31	MRPS9	LINC00999
	STAG2	QRSL1	KRT16P32	SNAR-A3	LOC100129520

TMEM63B	EEF1E1	ARL2	TRIM48	RNMT
MTRNR2L9	RPL36A-HNRNPH2	WTR1	LOC101927100	SERF1A
HAVCR1P1	ALPI	DHX29	STT3A-AS1	MIR607
PRDM5	CEP76	LINC00293	LOC101928495	MIR3621
ZNF670-ZNF695	GNPTG	DEFB115	DAB1	MIR3680-2
FAM122B	GABRA2	LINC00207	MIR4528	LHFPL3
BLOC1S5	CCDC47	TRMT44	MFAP1	BMPR1B
FAM126B	NPIPA7	CORO7	STXBP4	LOC101928851
PHACTR3	CCDC58	TMEM138	PDCD11	THBS2
DNAJC4	NPIPB3	ATP6V1D	ZNF816	PID1
PARVB	ISPD-AS1	GDNF	ZCCHC17	TAPBPL
ATP5J2-PTCD1	PTN	POLDIP2	MIR198	LOC101927237
GDPD3	KLHDC3	SGTA	LGALS8-AS1	DFFB
CD40LG	ALDH16A1	TMEM234	MKRN2OS	PIGB
STARD4	LOC650226	XPC	MIR3191	ZNF214
NELFE	MRPL32	PCDHA2	LOC285696	ZBTB25
MYLK3	RBPJL	LL22NC03-75H12.2	MIR2588	DDX54
SLC25A40	KIF20A	TXNL4B	IFNG-AS1	NDUFS8
RBM14-RBM4	DCLRE1B	SCRN3	SLC16A1	NDUFS6
NPC2	OSTF1	FOXD4L1	FNDC5	LOC100133331
NEGR1-IT1	BUD31	ANGPT1	ZNF716	LCE3A
FKBP1A	TCF24	SHISA8	LOC440300	CDK4
LOXL1-AS1	ANKRD26P1	H2AFJ	CCDC121	AMN
NPR1	RHO	LINC01232	CCDC137	SPAG5
RAB4B-EGLN2	CACNG1	LINC01203	KIAA0391	TP53TG3B
MUT	HRAT92	LYSMD1	FIBP	MIR4283-2
WNT10A	HTR7P1	TTTY3B	MIR423	MIR4674
PRDX6	MCPH1	PAX1	MIR4300	C7orf55-LUC7L2
GCOM1	USP20	FZD10-AS1	JMJD4	
PMF1-BGLAP	DDIT3	PPAN-P2RY11	JMJD8	
TIPARP	MSH3	SLX1A-SULT1A3	MIR4265	
ARHGAP1	TGM2	GOLT1B	PACRG	
KHDRBS2	LINC01346	MOSPD2	PSMA3-AS1	
C14orf39	LINC01340	DUXAP10	NUDT21	
LRTOMT	LINC01356	TEX41	MAP1LC3B	
LINC00564	TENM4	FGFR1OP2	TBC1D3H	
LOC102606465	HIST1H2BB	LOC101929646	LOC100506585	
ALG5	RBMXL1	PABPC5-AS1	RSPH9	
ALG2	PFKFB2	LOC101929681	MIR663AHG	

APPENDIX E. Gene Lists and ToppFun Analysis: promoter only H3K4me3 ChIP-seq peaks.

Output lists from MACS2 ChIP-seq analysis Venn Diagrams comparing LacZ and ASH2L promoter peaks only gene lists appear in this appendix. These lists correspond to **Figure 3.3-A.** ToppFun analysis for the 2,035 LacZ only genes can be found at this link: https://toppgene.cchmc.org/output.jsp?userdata_id=dade9860-36b3-4a1b-b16f-e3507b08ff76.

LacZ only (2,035)					
	ITGB8	TYMP	ATG14	SNORA71D	DXO
	SAP30L	LOC654841	CDK19	SDHA	BZW2
MIR760	RBL2	BBS9	FTSJ2	DNAH7	AZIN1
DDHD2	EXOC3	MIR378A	TADA3	POLD2	METTL21A
LSM1	CCNE2	SMG7	TMEM127	STK16	PLCG1-AS1
ASH2L	RELL2	CCT7	RABL2A	ZHX2	GNAI1
ITGB1BP1	HADHA	NEURL2	GLI3	NUDT1	MORN2
FGFR1	MIR4757	UPP1	NFS1	TRAF3IP2-AS1	RABL2B
WHSC1L1	KHK	EFNA5	DESI1	MIR4651	ANKS1A
TMEM56-RWDD3	EPB41L4A	SNX3	SBDSP1	SRD5A1	ZNF655
BRE	LINC00472	ARPC1A	ADGRF3	TRIAP1	TBC1D22A
SUPT7L	ERO1B	CBY1	LOC100506127	DOCK4	NSL1
ZNF79	LOC653602	ZDHHC14	STRADB	ST5	TIGD6
GEN1	DYNLRB1	RPS21	HIVEP2	PARD3B	RB1CC1
ALG8	PGM3	ABHD1	MBOAT2	LRRC73	UNC50
DNAJC27	LRWD1	ZFAND2A	Sep7	TAP1	LRRC16A
AGBL5	ATP6V0E2-AS1	ZNF496	TTC31	FAF2	FAHD2A
LOC728024	SSH3	SLX1A-SULT1A3	DKFZP434I0714	AK9	EIF1B-AS1
AGBL5-AS1	CENPO	SHARPIN	LOC100996437	ARFRP1	PLPP1
CCDC121	TNRC18	MAF1	BTBD9	SATB2	TBCC
LOC102723729	SF3B6	PRKAG2-AS1	ZNF786	SERAC1	ZFP36L2
FAM228B	NIFK-AS1	DNTTIP1	COA4	DHX57	TSGA10
MYCNOS	MTRR	LOC154761	PCIF1	CRNDE	ZNF142
CPSF3	IRX5	MPLKIP	AP5Z1	MDH2	EPB41L1
ADAM9	TTC21A	AMMECR1L	RING1	ERCC3	PPP1R21
GPN1	SCAMP1-AS1	ZDHHC3	ZNF346	PCOLCE-AS1	ZNF277
SNX17	LOC101928222	ZNF860	POLR3F	TCAF2	PHF3
DNAJC27-AS1	SFT2D1	FOXO3	PAX8-AS1	C6ORF120	TMEM237
WDR35	GRM8	NLN	ASB3	ACVR2B-AS1	ZFP2
PRR3	TPST1	LZTS3	ZNF398	SMPD4	UXS1
KLHL7	CDPF1	RNASET2	CPSF2	MAVS	C22ORF23
PITPNM1	NANP	PLEKHA2	CYP51A1	HCG17	POP1
	RBM33	DEPDC1B	SGMS1	SMPD2	LOC100505771

PLOD3	AUP1	E2F3	AP4M1	NSMAF	EFHC1
PSMA2	MCM7	LOC100289230	SNORA71E	DZANK1	MNX1
FAHD2B	EAF1	FAM134A	IFT43	PHF14	MRPL36
HEY2	SNORA3A	RAB23	PLAG1	COL4A3BP	TUBA4A
ZNF250	COL4A4	NSMCE2	RPL15	AHI1	COPS5
KIAA0196	SUCLG1	MFSD14A	ATG9A	MTBP	TOMM6
OSTM1	DNAJC30	MOGS	MLLT4	RAB22A	KIAA0141
PSEN2	CHKB-CPT1B	TIGD1	PGAP1	C5ORF15	SNORA9
STRIP2	TMEM60	NCOA7	BRD9	ZBTB49	LINC01024
RSBN1L	LMBRD2	MRPL32	MOSPD3	C2ORF68	TUBB2A
SPATA2	PNKD	PIGF	RIPK2	TTI1	POLR1C
STK36	CHRAC1	BUD31	SGK1	ARPC2	CIR1
WDPCP	AZI2	TYW5	FZD5	LYPD6	FOXP4-AS1
TMEM43	CEBPB-AS1	RPRD1B	URGCP	TP53I3	ST3GAL5-AS1
ZNF664-FAM101A	RTKN	TMEM170B	ALKBH4	PDE10A	SLC35B4
ZNF292	TTC30B	ST3GAL1	RBSN	BRD2	PPIC
CPSF4	FBXO32	ATXN1	FIGN	C14ORF178	C2ORF76
FBXL6	PRR15	LHX4	LOC100288181	SUCLG2	TMEM198
TMEM243	CEP68	ZNF141	VPS50	EIF6	ROPN1L-AS1
SLF1	MTHFD1L	C20ORF27	ALG12	PDAP1	SLC25A46
PRKCE	C8ORF33	GMDS-AS1	XRCC6	COPS6	LOC150776
ING2	GLI4	MFSD9	DDX11	FAM120B	EIF3B
SLC12A9	NOL7	NCAPG2	CNOT8	AP3B1	FOXN2
KCTD13	FLJ38576	PLEKHA8	CHCHD7	GFOD1	IRAK1BP1
NOD1	EFCAB6	MBLAC1	IFT22	BYSL	FRMD6
UGP2	FAHD2CP	TMEM129	PCED1A	EIF5B	STK4
DARS-AS1	TBPL1	PPP3CA	SNHG15	HIP1	LIPT1
SNX13	SAP130	PFDN6	ABLIM2	KIZ	MIA3
LINC01124	HSF1	SCO2	APLF	UGDH-AS1	IKBKB
ICA1L	MKLN1	NDUFAF2	ZNF622	METTL1	ERBB2IP
RNF4	EGR1	ZNF7	PSMG3	ZBTB9	FGFR1OP
ANAPC1	FBXW7	RMND5B	STAG3L1	CDK11B	CACNA2D1
CRNKL1	PSMB8	OSBPL8	QRSL1	AP4S1	NUP50-AS1
CDK6	CLCN7	WDYHV1	FKBP14	ZRANB3	RPIA
GDF9	ZNF623	LOC100288152	NIPBL	GABBR1	MRPS27
FER	SP4	RAE1	RASA1	CCNG2	THOC7
TRIM36	SPAG1	CCDC146	GAK	ATRAID	RAET1K
ZNF252P	ETF1	RGS2	RMDN2	PRIM2	ANKRA2
C2	FAM114A2	THEM6	PSMB8-AS1	STAG3L2	SBDS
MBLAC2	TMPPE	ARHGEF5	NDUFS6	ABRACL	HSPA4
	COPG2	SOX12	PEX1	TRIM52	DBNDD2

FASTK	CIDECP	ARL6IP6	ARL15	CUX1	MAT2A
EPHA4	LIFR-AS1	CCDC115	KIAA1841	MAFF	SMIM15
SMAD1	EDEM2	TMEM192	PCYOX1L	MTX3	KLF10
BLCAP	HDAC3	OTULIN	BRIX1	BARD1	COX5B
SMAP1	MAPK3	NOP14	MIR7111	SUPT3H	MRPL33
OGFRP1	HEATR5B	LTV1	ZNF775	GPANK1	LRRC14
ANKIB1	TAX1BP1	IFRD1	PPP1R12A	CARF	HARS
CUL9	IRS1	UFSP1	DPM1	PANK2	ARHGEF35
WDFY3	PRKAG1	PDCD2	CTDSP1	CASP8AP2	SLC12A2
ARMC10	HIST2H2AA3	NEMP2	TLK1	ST3GAL5	CHPF
CLASP1	ASCC3	RPS6KA2	GID8	CEP55	ZBED5-AS1
NXT1	SPRY1	ZNRD1	NFYA	BTBD3	CLHC1
GNPDA1	LOC100129518	RFC2	AFAP1	ACAA1	LINC01117
DLL1	PNO1	COBLL1	FBXO48	GNMT	PPIL3
ACTA2	LTBP1	MOB4	CDK5RAP1	INHBB	YWHAZ
RETSAT	IL15	ABHD18	SEMA4F	FAM161A	TTL
ALDH6A1	LNPEP	PLEC	R3HDM1	PUS1	ANKHD1
CSE1L	ZSWIM6	LRRC61	HTT-AS	ESRP1	STK32B
C5ORF22	ZNF343	RPS27A	C2ORF81	ETV1	ZNF721
EIF2AK3	CCT6P1	SEC61G	KIAA0825	RALY	FAM184A
AKAP9	RPP40	NOP16	HIST1H4A	MITF	WDR75
HOMER1	RPL26L1	JADE2	ZNF75A	ASAP1	TMEM230
ATG10	PDSS2	FBXO4	LRRC6	POMZP3	MRPS30
NIF3L1	RDH10	FOXP4	USP46-AS1	HOXC13-AS	CFAP36
CDC40	MIR6850	PDGFA	TAF2	SNRPE	SREK1IP1
GVQW2	LOC100129917	FAM126B	WDR54	LMBR1	NCAPH2
KRIT1	ITPA	EIF3L	OR2A1-AS1	ERCC8	ZC3H3
ZNF76	MRPL14	BAG6	ATF6B	LINC01184	NSA2
C4ORF46	GTSE1	DDR1	CCDC109B	PAIP2	C8ORF37
TSEN2	LYAR	SELO	WDR46	TSTA3	PPID
PUF60	NRF1	NUDT6	MZT2B	ZRSR2	RAB24
EPC2	LOC101927795	ICA1	NDUFA6	SYNCRIP	CCT6A
TMEM177	STXBP5	ZNF706	TPD52L2	NTPCR	LOC101928530
CENPM	CEBPZ	DHX35	GGCX	EMC3-AS1	TECPR1
ZSCAN32	RQCD1	UGDH	KIAA1715	CITED2	NUP50
TMEM217	RNF181	MIR933	KCND3	NR3C2	PSMB1
CNPPD1	VPS13B	DHX29	SLC38A9	TRERF1	SKIV2L
FGD5-AS1	TMEM9B	CWC27	BRD1	PAXIP1-AS1	PDE5A
PGM2	SIRT5	MED7	PCNA-AS1	PTDSS1	PLCB4
HSPA4L	ZNF696	PAK1IP1	ANXA4	DIMT1	THUMPD3
VPS13C	TCP1	NDUFA2	GDE1	AFTPH	GALK2

C5ORF45	SNW1	LOC728743	GPATCH11	PHOSPHO2	HIST1H2BN
SHPRH	KLHL8	TTK	LOC101927157	ZSCAN31	SEMA4C
RBPJ	FAM229B	C1ORF112	DCAF17	PPT2-EGFL8	LINC01012
GINM1	MRPL51	CLPTM1L	DDX41	TDG	HMGCR
XRCC4	PRPF6	DIDO1	WASF1	MCFD2	SERPINB6
STMND1	NR4A2	CCT5	UMAD1	PLCL1	RAP1GDS1
CKMT2-AS1	SPAG16	SLC25A27	IRF5	LOXL3	HNRNPDL
PIGG	BCAP29	MRPS7	LOC100294362	ORC4	TRIQK
LOC100132356	KLHDC3	NDUFB3	LOC344967	BLOC1S1	DYNC1LI1
LINC01003	RAD52	RAD17	APBB3	SLX4IP	GGT7
AUTS2	EIF4E	TDP2	INTS1	LZTFL1	ST13
CHAC2	TRIP6	SLC35B3	LRRC1	CHSY3	DUS4L
TNPO1	CENPH	ZNF619	SLC16A1	EP300-AS1	NDUFB9
DENND3	HELZ	HRSP12	MROH1	C9ORF78	WTAP
WDR60	CCDC88A	PPP4R3B	EMC3	HIF1A	CROT
BRI3	MCM8	GTF2IRD1P1	TTC23L	MTMR12	ITPR1
MUT	ATL2	ATP5I	FOXQ1	VWA3B	USP39
RNF103-CHMP3	TRAPPC13	KIF20A	LOC100133669	MEA1	CDKN2AIP
CDS2	PYGB	LOC101929710	CNOT6L	PDP1	CAMKMT
RARS2	AIG1	AVL9	FAM151B	PRELID2	WBSCR27
PREPL	CLOCK	THADA	CXCR4	EPHB6	OTX1
ORC2	RRM2B	ZSCAN16-AS1	RTN1	B3GALT6	NUDCD1
CTC-338M12.4	SLC18B1	LSM3	ZNF815P	INSIG1	UBE2B
OXNAD1	CTSO	CASP3	RWDD4	PREX1	TFAP2A
JRK	LOC100287944	ARFIP1	ZNF789	HELQ	BMP2K
RTN4	RBM48	PYGO2	LYRM4	WDSUB1	TMED2
GSS	TAB2	PLCD1	CARD10	RTN4IP1	BOLA3
NAPRT	RPL23AP7	MEIS1	ADSL	ZNF584	LINC00899
C6ORF52	TNPO3	FBXO5	RNF139-AS1	TGFA	NOA1
FAM160A1	NDUFS1	FAM86EP	PRMT3	HLA-H	PTPRA
AKR7A2	SOWAHC	TP53TG1	CPSF3L	SESN1	FAM47E-STBD1
NRN1	STK19	LRP12	C6ORF226	SLC4A11	KRBA1
TOP1MT	NCAPH	LOC100505921	TPMT	RNF123	DOPEY1
ZNF131	ZNF2	LETM1	SKP2	ATP13A2	MRPL2
CNPY4	GZF1	CYP4V2	FANCD2	PTPN18	SAMD10
MIR4469	ABALON	RGMB-AS1	TRIM41	FARS2	DCK
DPY19L4	YIPF3	TRAM2	ADA	KMT2E-AS1	PLK4
HNRNPA2B1	GPR155	RNGTT	WAPL	HCFC1	PRKRA
KLC4	KIF15	MDH1B	AK6	GATA2-AS1	PLEKHN1
NDRG1	ERMARD	DENND6B	C9ORF9	TMEM161B-AS1	JADE1
FAM86DP	POLH	RPS23	LOC221946		PRICKLE4

SMARCAL1	NAA15	SLC25A40	ICE1	CFAP221	SLC1A5
HIST1H2BD	TTC37	SCOC	USP37	CLIC1	FGFR4
ST7-OT4	ZC3H14	ACVR1C	SLBP	RANBP3	SYS1-DBNDD2
TMEM161B	BTD	JARID2-AS1	MDGA1	TBC1D14	LIPA
LOC101929709	TRAPPC9	MED25	DLX2-AS1	COPG1	METAP2
SLC10A7	LCA5	SNORA13	HDAC11-AS1	USP49	DAP3
MCEE	SLC17A5	TMEM30A	FAXC	FAM86HP	TGFBRAP1
TST	ENPP4	MKL1	NPHP1	PSRC1	LSM5
TCTE3	STX11	YTHDF3-AS1	PTGES2	ZNF10	CTDSPL
PMS2P1	RPF2	MRPL9	EHBP1	MMAA	LOC153684
LOC101926913	RBM24	C1ORF159	EMC6	FLOT1	STX18
STAG3L4	FAM135A	PPWD1	NR1D2	ZNF382	MRPS2
KCNK12	IQGAP2	COQ6	ZKSCAN8	AIMP2	MOCS2
LYRM2	SIX2	TMEM67	MIR3661	EXD3	C20ORF194
UBE2D3	LRRFIP2	HGS	HMGB2	KCTD7	CDC25C
FIS1	GSTCD	IRAIN	ACVR1B	MIR6791	ADD1
FBXO30	ACOT8	TFAP2A-AS1	COMMD1	PHF10	KIAA1429
PSMA7	CEP250	HM13-AS1	BAG2	NDUFA12	FUBP3
PCBP1-AS1	RASA4	NPY5R	DBN1	PSMG2	YIPF5
LSM14B	OTUD6B-AS1	ZNF709	BIN1	HSPD1	LYRM4-AS1
MFAP3L	TMEM128	MROH8	GPSM3	TRAPPC11	SLC25A32
TMCO6	ACTR3	SCRIB	RHPN1-AS1	PLOD1	CCDC74A
PSMG3-AS1	NEMP1	WWC2	FAM86B1	C19ORF52	RBM42
NDUFAF5	ANKZF1	GTF2I	AKAP7	RPS12	BRD8
CCDC136	CRLS1	SAFB	TFB1M	CFLAR	TMEM203
TAF6	PXT1	CEP44	CDC37	NAGA	YTHDF3
PSIP1	TSNARE1	GGA3	RGS19	DAXX	KIAA0319
MACROD2	UFL1	SNORA17A	PPIP5K2	PAICS	FBXL20
ZNF596	DMC1	LOC553103	PMS1	METAP1D	C12ORF60
GALNT10	MOB1B	CD3EAP	ITGA9-AS1	MIR4792	UBE2C
STAM	LOC101929231	DBF4	MIR6821	LOC731157	KLHL18
ARHGEF18	ITPR1-AS1	GHR	MIR6875	PARK2	C5ORF56
TRMT11	CTNND2	RAVER1	HIST1H2BK	NEU1	FGD6
TBC1D32	LOC100652758	PASK	POLR1A	SLC30A6	LOC100287015
SNORD54	IRX2	DTX3	ADAT2	ATXN7L1	TAF9
RRP8	SLCO4A1	BFSP1	PDCD6	PMEPA1	CCNT2
RPL7L1	HEXA	PRRT3-AS1	SDCCAG3	LRPAP1	NELFB
CHMP4C	MTA3	ARF4	STRN	USP20	REPIN1
POLR2H	MPST	MAU2	STK25	NUP188	KCNIP3
PLEKHA3	MIR6892	GGA1	LOC101926935	FAM53A	NR2C2
NCK2	GTF2F1	C17ORF80	THAP5	ACBD4	TXLNG

TEX10	JAKMIP1	ZNF561-AS1	SMARCE1	DNAJC28	KCNK15-AS1
PRPF31	TSG101	PAQR8	CELSR1	RFX3	RECQL4
EZH2	NABP1	NUS1	CEP76	ZNF230	MRPL12
SLC30A4	NBEAL1	NUDT4P1	ORMDL1	C17ORF89	PTMA
ZNF8	HIST2H4B	LOC102723824	THCAT158	ERF	C18ORF54
RASA4B	H1FX-AS1	STARD7-AS1	GLT8D1	SMARCD3	NEK1
CDK16	GNPDA2	TUBB	PRPF4	ZNF567	GLYCTK
FAAP100	CLDN4	SMC1A	LRCH4	COPS7B	UBE2G2
ST7-AS1	TATDN1	FAM86B3P	FST	SNORD52	ANO10
SIGMAR1	WNT7B	RPL7A	ZNF470	ELOVL2-AS1	ZNF239
SERINC1	SCARNA13	NAT14	MBTD1	ABT1	SEC14L1
SEC22C	HIST1H2AG	ARHGEF16	FAM134C	CISD2	LOC284930
PCDH7	SGCE	AGPAT1	ZNF790	COQ4	HIST1H2BG
SGOL1	FAM86B2	SDCBP2	FAM86JP	POLRMT	ZSCAN26
MAP3K14-AS1	AP5S1	NUFIP1	ASPH	MKKS	ZBTB12
SNORD84	SUGP2	MRGBP	KIAA1958	RAI14	ZNF717
CAPN7	LOC100507053	GPAT3	RBBP8	SUCO	PNPLA6
STIM2	TGS1	GMPPA	BOD1L1	NDUFA8	TPRKB
INSIG2	TTLL12	PEX26	BZW1	CSTF1	XRCC2
PCBP1	NOXA1	FAM110A	MYADM	SNX5	FAM210A
MAP4K4	MFSD10	MAD2L1	SNTA1	HBP1	TM7SF2
FBXO38	LINC00602	ACAD8	MMP16	PPM1L	CERK
SIPA1L3	EXTL3	SPPL2B	ANKRD31	HMGN4	ZNF594
ZYX	RRM1	PNISR	SYS1	TAF4B	PLIN2
GABPB1	APBB2	COQ5	SRRD	PEX13	TEF
TMEM167A	MANEA-AS1	MTIF2	TFEB	KHDC1	LRRC37A3
RASGRP3	PRR34	NME2	SUZ12P1	CHCHD5	SMIM7
OXLD1	AMZ2	CBX7	SKIDA1	RIBC2	CAPNS1
CHERP	SNORD12B	CEP72	USP5	ATP2C1	TYK2
ANAPC10	C1GALT1C1L	GMNN	HAUS8	HSD11B1L	AKAP1
TBC1D16	DCTD	CETN3	ERP44	USP34	NDUFV3
C3ORF58	RAB3A	EWSR1	TNFRSF10A	ASF1B	OSMR-AS1
HOMER3	MAFG	ZDHHC9	PEX3	BUB1	LINC00652
ZBTB21	UBE2V1	CDADC1	C2ORF27B	MICB	SLC25A1
MIR615	TAF8	BEND3	SDHAP3	RRP1B	PTCH1
DNPEP	POMK	CCDC157	HIST1H2AJ	CENPP	USP40
TMEM198B	STK11IP	PPP1R18	HIST1H3B	GPR108	UBE2E1-AS1
MCPH1	PITRM1	KRT18	ARID3A	PTCD3	NIPAL2
SQSTM1	SNAI1	USP6NL	OSMR	NNT-AS1	ELFN2
CALM2	NUDCD2	GATA6	C17ORF67	DNAJB2	XAB2
SKIV2L2	RPS18	MET	NCBP2	RFXANK	P3H4

SCLY	LENG8	RNMT	UTP18	RNASEH1	SMARCAD1
MIF4GD	PHF1	ZIC3	S1PR5	LOC100506990	NOL4L
MIB1	ZFYVE28	BCAS3	PTRH1	MIR375	TSEN34
LRRC8A	LIMCH1	IFT20	HIST1H2AE	PLEKHH3	MIR425
SLC38A7	GM2A	PARP2	TIAM1	FRG1HP	WDR91
PTPN23	ODF2	MORN4	AFF1	NDUFA11	R3HDM4
LMTK3	ISOC2	CDKN2B-AS1	LOC644656	ARSG	L3MBTL1
PGLS	RAET1E-AS1	RRAGD	RIOK1	ZC2HC1C	TRIM35
RIOK2	STXBP4	FNDC3B	GNAS	ENTHD2	PQLC1
SACM1L	LGALSL	IKBKAP	WDR5	PRR18	SYNE4
HN1	MIR4754	LOC389765	ID1	RPS6KB1	KIAA1468
SMKR1	KIAA1328	GSTA4	PAPD7	TEN1-CDK3	AP1M1
THOC1	THAP7-AS1	WBP2	DTD2	CEP126	PEG10
KDSR	UBA5	ZNF529	CBR3-AS1	HPCAL1	NADK2
GMFB	FBXL12	PPAN	OTUD6B	HSPBP1	MYRIP
SGSH	MTRF1L	ANKRD17	TBCB	SLC25A14	BLOC1S5
HIST1H2AM	KANSL1L	ACAA2	ANKRD37	CCDC171	CPLX1
RNU86	POLR3D	BBS12	HGH1	POLR2D	PPP1R11
PCBP4	RPL12	C3ORF14	GAD1	RPP25L	PDK2
TPGS2	MRPL30	SLC44A4	IKZF2	CLTCL1	C18ORF8
NUTF2	ZSCAN16	LINC01336	RSPH1	TNPO2	TRAF2
STMN3	DLG1-AS1	PHF5A	BPGM	LOC101928438	COPE
PLRG1	TOPORS-AS1	FAM73B	OSGIN2	PSMC5	ABCG1
TLR5	SGPP2	ORAI2	ZNF829	ZNF580	LPP
ACVR1	CERCAM	SSNA1	ABTB1	LOC100272217	BCL11A
VPS13A-AS1	DGCR6L	SOX9	LPIN3	NCALD	C6ORF62
C8ORF59	C5ORF66	RPL9	SNORA26	PECR	AMZ2P1
RUFY1	MRPS23	INAFM1	NDUFB2-AS1	DLX4	UBE2O
MAN1B1-AS1	TIPARP-AS1	LOC389641	TCF19	NPAS2	RRAGA
ZNF197	NDUFA13	MIR6790	REEP6	POP4	RPS28
PTPDC1	TTC3	DUSP18	GTPBP2	LINC01607	TUSC2
TCTEX1D2	ELOVL2	KLHL2	MED12L	PITPNB	FANCC
CEP78	OXSM	CCDC137	ATG12	SH3BGR	ENPP5
DTWD2	DPY19L3	RASL11B	SLC25A23	HIST1H2BH	TTC28-AS1
PRCC	PDE6D	LRRC45	PDE11A	ZNF570	CGGBP1
SNRNP200	ANKRD33B	SYNJ1	CMYA5	SNX30	PLCD3
C2ORF73	ZBTB2	HERC2P2	SEPSECS-AS1	MIR330	PIGX
ACSL6	NFKBIL1	ZNF830	PAQR3	UBL5	PAF1
MIR34AHG	BECN1	CENPQ	HIST1H2AK	JMJD6	ARHGEF3
SIRT2	TBCK	C19ORF33	ZDHHC4	MIR4665	VAV2
LOC145694	MIR1204	NREP	TEFM	PSMG1	TMEM135

NDUFA5	ZNF793	HMBOX1	2-Sep	ZNF573	CCM2
SUN2	FBXO10	GOLGA2	FBXO7	PNPLA7	HNRNPUL1
CDC14B	AURKA	NKILA	ZNF566	SLC26A11	JUNB
SWI5	DST	HMMR	CABYR	ANO2	MSH3
MSL1	ACP1	WSB1	ASAH1	C19ORF66	THAP4
ARMC9	LOC100507564	SRP72	RPL10	GRB2	PINX1
TRIP12	PIK3CA	GPR107	TRIM46	ZNF571	ITGA3
COX11	FZD3	THAP9	CWC25	ADAMTS19-AS1	SSH2
SEPSECS	DCLRE1A	DGCR6	SH3YL1	SPIRE1	PPP1R10
HPS4	NCK1	SNAP25	ZNF569	MIR4530	NEK11
VEGFC	KIAA1524	AKT1S1	UVSSA	LINC00667	RAB3C
SYBU	EBNA1BP2	ZNF565	RAP2B	B3GALNT1	TBC1D5
CLPTM1	NF1	ZNF700	TXNRD2	GADD45G	TMX3
PNPO	SETD2	DIRC2	BLOC1S3	NARF	ADAM17
MCTP1	ACOX3	USP43	SVOPL	AP2A1	TXNIP
DIP2A	RANGAP1	SPRYD3	TXNDC5	C11ORF68	NME6
EDN1	CAPN10-AS1	SLC16A2	TNRC6B	MIR4664	LOC100128398
NOTCH1	KLHL26	GYS1	MRPL41	ADGRL1	POLR2I
GOLGA1	TESK1	MYO19	SYCE3	VAMP2	MIR8069-2
HIST1H3H	TOPBP1	CENPE	CAB39	RIC1	TBC1D2
ZNF322	SNRPD2	CCDC39	SORBS3	LAMB1	NUP62
EPN1	INSR	SEC24B	SSR4	KIF9-AS1	ASTN2
AIMP1	DCTN4	NUDT18	CCDC43	SUPT5H	KHDRBS3
SMIM19	CDKN2A	C9ORF40	RND1	ZNF473	RALBP1
GSDMD	DNAJC7	MAFA	NRM	GPRIN1	UBXN7
FLAD1	TSPAN33	NCL	IL1R1	MCCC1	MAPRE2
STRA13	LOC101929762	MAPT-IT1	AXIN1	C19ORF44	HNRNPUL2-BSCL2
CANX	HIST1H3G	MAPK11	MRPL27	SCP2	ZNF311
CYTH1	CDC20B	RCHY1	TTC39C-AS1	CEP112	MED16
PAQR5	TMEM147	CNBP	ANKRD6	CEBPA	ARVCF
ARHGDIA	KCTD2	ZNF446	MED22	DENND4A	PI4KA
TMCO3	ATP13A1	EXOC7	MSX2	NXF1	ADRB2
ZNF791	FKRP	DALRD3	LARP7	MRPL1	DDX5
ZNF845	SNHG16	SLCO5A1	RAD18	MYPOP	CCDC74B
TP53I13	SS18L1	ADNP-AS1	RFX1	LDOC1L	DTYMK
CPQ	ZNF436-AS1	RHOQ	LOC100419583	HCG25	ARIH2OS
MTERF4	KIF4A	DNASE1L1	RNF152	TAOK1	EPM2AIP1
INTU	SEC61B	SLC25A33	RBM12B-AS1	CTDP1	ZNF561
POLE4	MPND	ABCB8	FAM162A	THAP8	NDUFA3
RBM18	HIRA	PEX12	CDKN1C	SLC2A11	SNORA16A
JAK2	COMMD10	SLC16A3	FAM207A	GRHL2	

TRIM37
SNORD58B
RUNDC1
ZNF354C
AGPAT5
ATP5G1
BAG1
PPP1R26
SDSL
PLEKHH2
SH3GL1
SH3BP2
NKX3-2
EBAG9
HMG20B
PSMD5-AS1
CASKIN2
IRGQ
ZNF607
IDUA
ZNF846
KMT2B
ARHGEF6
FAIM
UBXN8
MSL2
DCAF16
EIF4ENIF1
MRPL34
PRR7
LSS
PIGP
JAK1
DEDD2
MAP1LC3A
PAXBP1
SNORD50B
ZNF879
PSMD7
NBR2
MYO1B

UBA6
CELSR3-AS1
ZMYM3
HNRNPL
ARMC1
WDR89
ESCO2
SLC1A1
SEC24B-AS1
POLQ
PARP3
CECR5
DLX1
GUCD1
ALG3
LIN54
SNORD61
ME2
MIR4674
SAMD4B
ADCY3
MRPS18C
SIRT6
LINC01431
ABHD17C
SEC16A
FLT3
C9ORF69
POT1-AS1
SLC35G1
TECR
ASXL3
RUBCN
INVS
DUSP4
MIS18A
SNHG5
CC2D1A
ADARB1
ALG2
MRPS12

MGC57346
EEF1D
NUDT16
SCARB2
XRCC1
SS18L2
BCL2L12
LOC728323
ALYREF
C21ORF58
AASDH
HPS5
NOL8
LINC01534
WDR62
SUPT6H
NNT
CAPN10
RRP9
RNF214
MIR4523
TEN1
CCDC58
LACC1
TRIM32
HIST1H2BM
DIS3
SPATA13
PSMD5
VRK3
RNASEH1-AS1
LOC339874
CCDC25
MIR4449
NFKBIB
C9ORF163
NIP7
ZGRF1
FAM47E

SNRNP40
POLA2
MIR5700
HIST2H2BF
AGO3
RCOR2
MAP4K2
LMO7DN
GDF11
LMO7
MEGF6
ELP4
MYCL
TMEM138
ATP8B2
PATL1
FZD10
CLYBL
PITHD1
PDE3A
RILPL1
GON4L
ATP11A
LINC01167
CHST11
KMO
CHID1
MIR9-1
GMEB1
ABHD13
DPYSL4
NELL2
ZFP91
LOC101928414
TMCC2
FAAP20
ZNF385A
POU6F1
ZNF326
DEF8
PSMB2

USPL1	DAGLA	ANAPC15	SSBP3	PKN2	FNDC5
CACNB3	STARD10	GALNT4	ARHGEF2	KCNH3	ADAMTSL3
FANK1	TMEM39B	ANKRD63	DNAJC3	RAB3IL1	RNVU1-20
PPFIA2	RGL1	ZMYM2	HOXC-AS3	ZCCHC11	PCDH17
METTL21B	LATS2	PFKFB2	PGBD5	POLR3GL	SDHAF2
SNORD15A	DNALI1	PLEKHB1	BLACAT1	PPP1R1A	ATP2C2
ZNF664	HS6ST3	EIF4G2	RASSF7	BORA	HIST2H4A
LOC101927204	PADI2	ZBTB41	KCNC1	HMGA2	SWT1
PDE3B	MSRB3	YBX1	TMCC3	SLC22A18	C1ORF145
LOC101927583	LRTOMT	SAMD11	C1ORF53	P4HA3	SYT17
USP12	TMEM53	PDE1B	GUCY1A2	GOLGA8A	DIXDC1
SETDB2	AHCTF1	COL16A1	PPP2R5B	LOC101927318	SCFD1
TUB	LOC101929234	NES	TESC	TSC22D1	TTC39A
SPATA1	NDFIP2-AS1	LRRC63	GTF2H3	INPP5B	PAH
TMEM200B	TNS2	CDH15	PHLDA2	PLBD1-AS1	BCL2L14
GSE1	PRICKLE1	C11ORF91	RFX5	CACNA1C	B3GALNT2
RAD54L	SNAI3	KRT8	ERMAP	KIF2C	CAND1.11
ZBTB7B	DDX55	ARL14EP	SF1	SPATA6	S100A16
PXMP2	EVA1B	ARV1	RCAN3	DYNLRB2	SHF
SPRYD4	GTF2B	LOC101928443	MANEAL	FAM78B	SMPD3
RAB30	CLP1	KCNT2	DAPK2	ESRRA	GOLGA8F
DNAJA3	PNRC2	BEND5	PLCG2	MAF	GOLPH3L
PHC2	SCUBE2	VANGL2	SERPING1	KCNK2	HMCN1
ALKBH3	TAF6L	GAN	CYB5R2	TPBGL	HELB
PUSL1	TRMT112	FAM86C1	TCTN1	IGFALS	SLC4A8
LOC101928737	TUFT1	KCNJ11	WDR11-AS1	MZT1	SIVA1
LOC101928043	SLC3A2	TMEM266	CFAP45	LOC100287036	TRIM34
LOC101929657	KCTD14	ZNF687	LOC100132057	USP35	CUEDC2
EIF3J-AS1	CELF3	THBS1	CNIH3	NADK	ETV6
MCOLN2	BPNT1	SMIM22	CHRM3	KCNQ4	CELF1
FMNL3	HMGB1	NUDT4	MSI1	RAP2A	AMDHD1
COLCA1	SMYD2	BCAR1	WDR26	LRRK2	UBE2L6
EIF4G3	JDP2	SPG20-AS1	GJB2	ATF7IP	MIR5087
SHISA2	ZBTB8OS	LHX9	SPDYC	DISP2	VPS36
LIX1L	RWDD3	FNDC3A	INPP5A	ROM1	CSNK2A2
QSOX1	DR1	FAM103A1	LOC284648	DOC2A	ZCCHC14
ZC3H12C	CACNA1C-AS1	SP1	MIR7155	TMC7	SCN8A
FAM124A	RBM26	MIR190B	TRIM67	FAM183A	ST3GAL4
LOC100288798	ZMYM5	PRRG4	MCF2L-AS1	MIR5691	PLA2G4B
GPBP1L1	KCTD10	SNHG1	INTS6	ZNF778	LINC01144
HNRNPA1L2	CCDC24	ALG14	MIPEP	IFNLR1	PPFIA4

KMT5A	ZNF644	ZNF597	NQO1	ALG5	COX15
IRX6	CCDC83	C11ORF1	GLUD1P3	SOX5	C14ORF37
HYPK	C1ORF27	ADAMTS7	GLB1L2	HS2ST1	IRAK4
PLPP4	EFEMP2	PTPN6	PLK3	MKNK1	GAS6
SYT7	STYK1	ARPC5	DYDC2	EXOSC8	LOC101928979
ARHGEF25	MKI67	MPPED2	MTMR10	YOD1	SHOC2
TMEM87A	LINC00939	TMX2	SYNC	CALCOCO1	ANP32AP1
NUP93	PIGBOS1	TAT-AS1	CPEB3	EXD2	LOC283922
PRC1	TMEM126B	RIMKLA	BCL9L	LINC00346	ANKRD2
C10ORF54	AP5B1	SYT13	USP3	KCNK13	CENPJ
PHF11	NPAS3	HSPA7	DHRS12	ITPRIP	PRDM10
KIAA1804	MAB21L3	FALEC	FGF9	RAB30-AS1	C16ORF71
FUCA1	PLEKHG4	ARID5B	ONECUT1	EHF	MRPS31P5
PLEKHD1	TPT1	MAP1LC3B	SMIM2-AS1	AIM1L	NSUN4
RGMA	POMT2	CYB5B	ATMIN	ALG6	TIMM10
TK2	KLHDC4	LBHD1	FGGY	TMEM126A	CACHD1
TTBK2	CALB2	TDRD5	MMP25	KIAA1217	C11ORF74
ZNF140	MRPS35	RBPMS2	PTPN20	RPS27L	TBX6
LINGO1	SLC2A1	MIR4692	AGAP11	ARNT2	SPNS2
MSANTD2	FLJ37453	POU3F1	FMN1	FOXD3-AS1	XRCC6BP1
EEF1DP3	DYNC1LI2	CIB2	LOXL4	FOXA1	IQSEC3
RNF6	ITFG1-AS1	PRSS27	CORO1A	MEIOB	C16ORF74
MMACHC	RRP15	SNX33	RTN4RL2	FREM2	USP54
PLLP	UGGT2	SHCBP1	XYLT1	C11ORF45	FAM25A
MIR4500HG	SLC16A7	MSANTD4	TRIM13	CHRM1	COQ9
CATSPER2	DNAJC3-AS1	FTH1	HSPA12A	ITGB1	CYP2J2
CHD4	NPAT	GPR161	MRE11A	TMEM256	LOC101929574
GUCY2EP	TPTE2	ST3GAL2	LINC00434	TMEM167B	MAD2L2
LOC101929340	TENM4	SETD6	RHOBTB1	IFIT3	KIF22
SNHG21	STOX1	SFR1	ZNF248	SUPV3L1	GPS2
MIR3124	DMAP1	SLC43A2	UTP20	ARHGEF12	MIR4513
EEA1	KCNJ5	LIPT2	GRID1	KBTBD6	NRG3
RAB39A	LINC01572	CTSC	C11ORF70	ESRRG	DHX38
HNRNPA1P10	NTHL1	ABCG4	TDRD3	ATM	TMOD2
UCHL3	PLXNC1	LIG4	LYRM1	PKD1L2	GLIS2-AS1
WDR81	ACBD7	NEGR1	BEAN1	ARNT	LOC105376671
ESYT1	SEMA6D	RAP1A	LOC644919	ZNF263	MAEL
INO80	FAM81A	NCDN	BEST3	PKMYT1	PRIMA1
BARX2	PLEKHO1	GJC2	CBFA2T3	TEKT1	ACTR1A
ZFPL1	SVIL	CDT1	ADD3-AS1	IGSF9B	RUSC1-AS1
BMF	NPR1	FAM216A	NME3	KCNQ1OT1	CCDC64B

B4GALNT1	LINC00871	SLC25A21	SPA17	LOC81691	ACSM3
LOC283140	USP2	USP3-AS1	MIR5093	FLT1	HOXC9
LOC105447648	RPS17	RIMS3	FAM63A	AK4	FAM90A1
RXFP4	PRUNE	A2M-AS1	NMNAT1	FBXO33	MAFTRR
SPRED1	ANO9	PAN3-AS1	LNX2	ZZEF1	ANAPC16
FZD10-AS1	SPTB	BEGAIN	KCNIP2	SORT1	ADPRM
BAIAP3	PKNOX2	LGR5	DNAJC16	CIDEB	MYO1C
LARP6	CELF6	VWA2	WNT4	SNORD49B	RAVER2
ZNF436	PC	CHD8	C11ORF97	ADAMTSL4	CSTF2T
SAXO2	APOLD1	HTR7	ADPRHL1	C10ORF10	PLXDC2
TMEM255B	ZDHHC24	NLK	FHOD1	TP73	MRPL55
FAM181A-AS1	AARS	BBS10	LRRC38	MIR497HG	RFX7
CXCL16	LINC01461	NDUFB1	SNORD45C	MED31	ACIN1
ATP1A1	PFKFB3	SEMA7A	DHDDS	SLX1B	CPEB1
ARMC4	RNASEK-C17ORF49	NOP9	KIAA1462	CHMP4A	CTNS
FSIP1		MIR193BHG	TLCD2	ENOX1	ZNF25
NPIPA7	SAT2	AGBL4	ARPIN	SMIM1	ADPGK
GRK5	PVRL1	ANKFY1	SPAG5	SAMD4A	DBNDD1
TSHR	AURKB	RPS6KA4	BIRC3	WSCD1	GRIK4
PKM	C11ORF80	NFYC-AS1	LOC101929089	MIR7846	SNORD42B
FPGT	ARHGAP22	CFAP70	MXI1	MIR9-3HG	KCTD11
SLC11A2	SRSF11	GRHL3	SNRNP25	FAM57A	ALOX15P1
SHISA6	CTC1	HHEX	ABCC4	PALB2	CISTR
NFIA	SPAG7	PPP1R36	LOC283575	SIAE	HIC1
NRDC	CCDC6	ACOT7	FAM25C	ZDHHC1	LOC102723809
HMOX2	TMEM57	PROSER1	C11ORF71	APITD1	PRPSAP2
LRRC8D	ZFP69B	BDKRB2	CDKL1	PIK3AP1	ARL6IP1
MTHFR	MIR365A	PTPRB	PLD2	LINC00437	ERCC4
TRABD2B	BIRC2	TMC5	NIPAL3	PARS2	DBT
DNAJC14	DEXI	STAM-AS1	RPAIN	MIR132	NAA60
SMCO4	ZNF684	ANKDD1A	APLP2	C1ORF168	HS3ST3A1
FAM13C	TLL2	TSPAN2	MED11	UBTD1	RHBDL1
ZRANB2	RPL13	PDCD6IPP2	SALL2	EDC3	FBXO2
TAF1D	LINC00466	SC5D	PITPNM3	SV2B	POLR2A
SLC45A1	DCUN1D2	TRIM33	SELL	LOC101928453	LINC01515
TMEM254-AS1	GPRC5B	LRRC40	SULT1A2	NXN	GCOM1
PCAT29	EFCAB7	SCN4B	PGM1	DGKA	TTC36
NUDT7	SLC25A30-AS1	ADRB1	ADRA2A	CCSER2	RBM4B
MAN1C1	GFOD2	HOXC5	DLG4	MEIS3P1	FLJ36000
TUBGCP2	PGR	ANKRD52	STIP1	NPRL3	NCOR1
NRGN	DDX19A	FMO9P	NECAP1	RGS6	HBA2

TSC22D1-AS1	NEUROG3	RBM7	MIR6859-2	ASPG	FAM111A
SULT1A1	HBA1	BMP8B	ZNF219	WBP1L	DDX11L9
MTOR	WDR63	C1ORF233	MIR6859-3	SREBF1	NBEAP1
RNF167	C14ORF93	USP2-AS1	BAMBI	EFCAB14	MIR4708
CELSR2	LOC100190986	MORN1	PPP4R4	RPS13	PTEN
FAM64A	NIPA1	ABR	GLRX5	LTB4R	KBTBD3
COX10	FAM155A	C16ORF52	MIR4307HG	WHAMMP3	EML5
DPAGT1	MIR3175	PRTFDC1	PRCP	RPP30	SUGT1
STXBP6	MTIF3	SESN3	USP28	RERE	SERINC4
CD164L2	DEPDC1-AS1	MFSD6L	MBTPS1	RNF187	USP31
ALDH3B2	TMEM120B	DENND2D	PHF23	ASAH2B	DDX23
ANKRD35	SLC2A4	ROR1	MIR6732	TCTEX1D4	LEPROT
TINF2	C15ORF59	EPHA10	IGF2	LINC01481	GNRHR2
DACT1	DIO3	CYP26A1	CYB5D1	HSD11B2	ACADVL
TYW3	ZSWIM7	ARHGAP32	TANGO6	ALDOC	BBOX1
TAF13	MIR4492	SERPINA1	TNFSF12	FLCN	CLCF1
HAX1	POLR1D	LOC440300	AUNIP	CRISPLD2	LOC101929613
LOC101928162	LOC101926933	TMEM107	SLC2A1-AS1	CFAP58	MIR3176
PDSS1	IREB2	TVP23B	SCMH1	TRNP1	RASAL2
LRIG3	HPCAL4	SIRT1	GLI1	PPP2R1B	LOC101054525
CASKIN1	RPL26	MPDU1	FBXO44	CHAC1	TTLL5
LDB1	HYDIN	PTCHD2	TACC2	SEPN1	ZNF592
AK5	ALDH1A3	ARRB2	USP21	BMP8A	MYBBP1A
ZDHHC16	MT1A	RAB1B	KIF11	SHBG	LINC00359
MAP3K9	EML1	ARMC3	ETFA	CREB3L1	MTERF2
NATD1	LINC00842	RAB11FIP2	CYB5D2	C1R	CEP83-AS1
PRKCQ	C15ORF61	C10ORF11	BRSK2	PMP22	C11ORF88
UBXN10-AS1	DYNC2H1	MSS51	CRB1	LAMB3	JMJD7-PLA2G4B
MIR4691	MFSD13A	TOM1L2	DERL2	PDE2A	DISC1
WBP4	GRTP1	PTGER2	MCOLN3	KIAA0754	PPP1R14D
EIF1AD	SPNS1	CYB5RL	HSPA2	KIF1BP	ESR2
MKX-AS1	OAF	SNORA48	RAB40C	MIR4487	ADIPOR2
NEDD8-MDP1	LOC100288846	SOS2	ADAMTS17	SMC3	SRPR
TNRC6A	IL10RA	GPR3	IFI27L2	LOC284023	C1QL4
BCL6B	SLC5A11	SLX1A	GYLTL1B	ALDH4A1	SPIRE2
SPAG5-AS1	FAM107B	EGR2	PMVK	DDX11L10	CTH
LOC100506022	LOC102724933	MIR6775	SMG5	RBBP4	GNPAT
WRAP53	TRIM3	LOC728392	FAF1	IPO13	HP1BP3
C17ORF107	CT62	TMX2-CTNND1	HOXC12	KMT2A	DPY19L2
POMGNT1	MRPS31	SARM1	ZZZ3	C15ORF48	SNORD55
GNAO1	RNF207	TPM3	ETNK2	FAM111B	PALMD

LOC102724467	SLC25A45	SDHD	LAMC1	KIN
DIAPH3	LTBP2	SERINC2	P2RY6	ZCCHC24
RTCA	SFXN4	MRGPRX3	TMTC2	ACY3
CDC42BPA	NHLRC4	EXOC3L4	PTGER3	BACE1
FOSL1	ORMDL2	GCHFR	GBP4	ADAM8
POLR2L	ZG16B	NR0B2	LOC101929718	KIAA0895L
POTEB3	RAB3B	SSX2IP	GPATCH4	NDUFS8
PRKCQ-AS1	AGO4	MKL2	DLG2	RAB2B
NLRP14	NEURL1	PIGK	FAR2	PAQR7
SCNN1B	B4GALNT3	SH2D5	TMEM69	CKS1B
TRIM6	LRRC8C	ARID1A	LINC00866	LMO3
ZNF214	COMMD6	STIL	E2F8	UBAC2-AS1
FAM160A2	NME4	KAZN	GORAB	NUBP2
HYI	SPRY2	LRRC52	HERC1	ATP5S
PSKH1	UPF2	TM9SF2	OTOG	
SSTR5-AS1	CIT	ST3GAL4-AS1	CAMK2G	
GOLGA6L7P	SPPL2A	VAMP1	LGALS8	
DDX11L1	DDN	ABCC11	ATG101	
POU2F3	LINC00638	LOC101928322	TUBB3	
HIST2H2BA	EP400NL	DNAJA4	CPB2-AS1	
ARHGAP1	LOC101927045	FAM83G	TTF2	
FANCA	ACTR10	C1ORF220	ALG1	
FOXO6	PACSIN3	ABCD2	COA7	
SHC1	HINFP	C10ORF2	USP15	
ELOVL1	LINC00938	HIPK3	RORA	
GOT1	TRNAU1AP	CFAP54	LOC101928069	
CCDC144CP	SDC3	TLN2	FOXM1	
TLX1NB	YRDC	SSU72	CCKBR	
UFM1	NUMA1	C15ORF57	ZMYM6NB	
ZC3H13	SAMD13	SYNM	LOC103171574	
TGM1	CDYL2	RPS14P3	CAPN1	
CDC7	SLC25A35	ALDH1A2	EFNB2	
LRRC32	ANAPC5	MIR4537	PCBD1	
JMJD4	MED18	PGLYRP4	ZNF281	
KIF18A	TIGD3	MTMR11	LOC100996455	
CAP1	HTR6	GLYCAM1	RMI2	
RSRC2	SYT8	SNX22	INTS7	
MIR4710	MUC1	LINC01488	ASB7	
SBF2-AS1	HNRNPUL2	MIR484	FGF11	
NBEA	REEP3	CCDC85B	MDK	
HAS3	F2	CD59	KCNQ1	

Common (1,282)

RSBN1	ADAM15	DYNLL1	ATHL1	SLC12A6	PGAP2
MAGI3	TFB2M	UBQLN4	PIGV	PRKAB2	CD81-AS1
SLC16A1-AS1	OSCP1	TSACC	LMNA	CD63	ARL1
WDR77	MIR7641-2	TPRG1L	TRAPPC3	CCT2	AXDND1
HIPK1-AS1	NT5DC3	LINC01389	APIP	LOC148413	TMEM5
GSTM4	TDRKH	TMEM59	LINC01128	KLHL17	LOC101928580
BAG5	MRPL20	LOC101927415	SMIM12	SMAP2	TMEM52
LRIG2	FBXO3	AASDHPPT	FRMD5	ELF3	DPH6-AS1
RPLP0	RNF121	TAGLN2	SLCO3A1	TMEM179B	LYPLAL1
METTL10	E2F7	ANKLE2	CPTP	RARG	HIST2H2AA4
CNN3	EEF1G	PIBF1	FOXJ3	RBM15	PMM2
PEX10	C11ORF57	C11ORF73	RASSF3	TROVE2	IPO9-AS1
EIF5	METTL20	GGPS1	UBAC2	KBTBD4	PPCS
XRCC3	EIF2S1	ARHGEF7	COX4I1	MIR203A	RBM34
PPP1R13B	LINC01351	ATF1	ASIC1	LOC101927495	RPAP3
CCDC64	TOLLIP	UBR4	SH3BP5L	QSER1	GALE
RAB35	PEF1	NAV2	COX20	TMEM9B-AS1	SNX1
RNF141	TMPO-AS1	PDIK1L	C1ORF131	SLC50A1	BCAR3
ZFYVE21	CAPZB	MIR2276	STAT6	DEDD	ZNF143
B3GNT4	CSAD	EIF3J	DIABLO	TUBGCP4	FAM213B
LINC00959	ATP6V1D	EXOC8	ANGEL2	TSPAN31	TATDN3
PLEK2	NCSTN	CAPS2	MTF1	RAD51AP1	PACS2
LOC101928068	ITGA5	SPPL3	LYRM5	RIT1	LPCAT4
TKFC	EIF2B1	CCND1	S100A13	ZCCHC17	FURIN
CAPRIN2	SYDE2	PLEKHA8P1	MVK	RBBP5	WSB2
PYM1	CCDC59	EFNA1	LOC101927587	AKT1	RNU6-2
PSMD9	UBE3B	MAGOH	SYF2	ZMYND11	MIB2
AMPD2	USF1	LOC101929224	PPHLN1	TMEM80	VEZT
RSF1	VPS33A	PSMD13	PARPBP	DDX51	HRAS
NUSAP1	SOCS4	RHOG	ANP32E	SRRM1	TP53BP1
CAMKK2	TAF1A-AS1	MIR6733	SPRYD7	BROX	SLC38A1
LOC100505666	DCLRE1B	KCTD21-AS1	TSSC4	ZNF605	DTWD1
TPCN1	C1ORF109	ENO1-AS1	LRRC41	CORO1C	CD82
ACTR6	KIAA1033	HIST2H2BE	MFSD5	TSPAN4	RRAS2
BLM	PHRF1	GDPD5	GATC	EMSY	AP3B2
NEDD1	TMEM56	P3H1	IBA57	DDIAS	NOC2L
PBXIP1	SNORA57	ART5	SRSF9	AQP11	TAF5L
BTBD10	POLR3B	PKP2	BRAP	NDUFAF1	SRP14
CAPN5	MIR3652	HEATR1	FKBP11	AP4E1	POLE
	ZC3H10	FADS2	ILK	PCCA	ZNF891
	TMEM79	RAB3IP	NAA40	TPR	TARBP2

MIR5187	PKP3	KIF26A	RNF34	CHP1	TMEM54
SCAF11	PGM2L1	SHMT2	SPCS2	HMGCL	SMAGP
CEP170B	UBN1	TTC13	CES2	LOC148709	TCF12
NADSYN1	SLC43A1	ZNF821	BTBD6	ITPKA	FLYWCH2
PAK6	GPT2	VPS33B	NHLRC2	MCRS1	TRMT1L
C12ORF49	CRTC3-AS1	MIR616	P2RY2	ADCK3	MVD
RNPS1	ARNTL	LOC100996255	VPS45	RIC8B	GBAT2
RAB13	SLC25A22	TERF2IP	TTLL7	ZNF768	MLLT10
TINAGL1	MRTO4	SH3D21	LOC100288069	MTA1	CFDP1
C16ORF59	PRDM11	PLBD2	ADAL	HSDL1	SLC30A7
STARD9	TSC2	SPATA33	EPS8L2	GNS	ZNF500
ACD	TMEM234	NOP10	KSR2	DNAJC17	NUDT21
IST1	PQLC2	HEBP1	SYT1	LOC643339	RBM8A
SNRNP35	PRPF38A	BICD1	MMS19	CCDC189	HHAT
FZD4	EFTUD1	LRRC4C	ABCA3	NDUFA4L2	THAP2
YY1AP1	GPR89A	USB1	ADAR	NAGPA-AS1	MIR1915
ISG20L2	ABCB10	TOLLIP-AS1	ILF2	VAMP4	MYCBP
SNAP47	FOXN3	CCNF	ARHGAP5	C16ORF95	ADCY9
ING1	TNFRSF19	INAFM2	LINC00936	ATG16L2	UBALD1
PAAF1	COX14	ZNF785	GAPDH	MLEC	TMEM72-AS1
SOCS2	CRYL1	ISCU	F3	KIAA0430	INPPL1
TMED5	GPR137	AEBP2	KMT5B	ZNF84	PPFIBP2
SHISA9	PTBP2	SDR39U1	BMS1P5	ZNF503-AS2	SMAD6
CDCA5	SMARCC2	MAGOHB	TPP2	ALG11	MAP2K5
GNAI3	KCTD21	DENND5B	PDIA3	DDX11-AS1	PARD3
PIP5K1A	SMARCD1	WARS2	IDH3A	PGAM5	C14ORF119
IKBIP	ZFP69	HCFC2	RCBTB1	RSL1D1	LOC643770
SERF2	MRPL48	TP53BP2	ZBTB18	MIR762	MDM2
FGF14-IT1	C1ORF101	PFKP	MRPL49	MIR4512	SNORA14B
KRTCAP2	TXNDC12	DVL1	USP24	DMXL2	FAM86FP
SUOX	RNF41	DGAT2	5-Mar	KATNB1	GOLT1B
NUDT22	TBC1D24	MTFR1L	RANBP10	NUP133	RAB6A
NUF2	UBAP2L	RPL4	EEF2KMT	ANKRD13C	CHTF8
POLR3C	DNAJC24	AAGAB	NIT1	NFAT5	ENSA
DUS2	XPO4	ZNF668	RASSF8	LLPH-AS1	NR1H3
TRAPPC2L	GALNT6	GNPTG	FLJ10038	PEX11B	RNF111
UVRAG	COG8	TSPAN9	C14ORF79	C15ORF41	C14ORF80
PTDSS2	CD2BP2	SDF4	ZBED5	FOS	GPR19
TMEM50A	PCED1B	TJP1	APH1B	LOC100147773	FEN1
TMTC3	PAOX	SCYL2	TCP11L1	ARMC5	SLC2A13
MAPKAPK5	EXO1	CMC2	UBE2Q2	RCC2	NUTM2A-AS1

KIF14	RCE1	FAHD1	SLX4	MAN2C1	IL15RA
TMEM183B	IGSF9	INTS4	POLL	TRIP4	RSU1
NUTM2B-AS1	VPS29	SLC7A6OS	DDIT3	CASC2	KCNAB2
LRRC57	MIR200C	PARGP1	PHKB	PIF1	VPS35
TXNL4B	LAMTOR5-AS1	CHEK1	BORCS5	AHSA1	TXNDC17
COQ7	CDK2	SLIRP	SDF2	KLHL35	HENMT1
CDIPT	TMEM51	CIART	PHYH	NCAPD3	KIAA0586
TSR3	FLVCR1	DOCK9-AS2	PYGL	DCAF11	PCDH9
DHCR7	PLEKHG6	DICER1	OAZ3	VIM-AS1	KCNMB4
NAB2	TARS2	HNRNPF	TAF12	EMC8	TCP11L2
DLL4	HTATIP2	PIK3CD-AS2	SNORD74	CUL4A	LIN7C
CHTF18	FAM213A	CIAPIN1	STARD5	CAMTA1	NCKAP5L
RAB27A	LRP1	CIPC	GFRA1	ENKD1	L3HYPDH
AFG3L1P	ICE2	TSR1	ADK	GTF2IP20	HECTD2
LOC100506844	TMEM208	HMG20A	GPR158	PTPN14	NOP2
NCAPD2	SLC37A4	TGFB2-AS1	RELA	GPALPP1	MEIS2
TIGAR	PLEKHG5	LRRC20	ARHGAP11B	FBXL19-AS1	CKAP5
FRA10AC1	MIR6741	NFATC4	HIRIP3	M6PR	MIR210HG
MTMR2	GPN3	LYSMD1	ZNF488	ATP1A1-AS1	GSTO2
ARHGEF17	CCNY	CEP170	MANSC1	SUV39H2	DLG5
ATP5C1	TMEM254	GMPR2	15-Sep	SNORD68	YAP1
GDAP2	SKA3	CCPG1	TCEB2	C10ORF111	MADD
ANKRD42	AP3M1	CREBL2	LOC102724571	RCOR3	CREBBP
DUT	CCDC122	SMG6	LRRC28	GALNS	CLSTN3
FANCM	HIF1A-AS1	MIR22HG	CASP9	BTG2	NMRAL1
MRPS11	GABARAPL2	CFL2	DCAF6	MIR3656	HSPA6
CDKN2C	HOXC-AS2	TEAD4	RNPC3	LDHA	OPTN
PDDC1	DPF3	PRKRIR	HIST2H2AC	PIAS3	IGHMBP2
FAM168A	C10ORF25	RPAP2	CASC4	ACP2	HSPA14
RPLP2	SLC39A13	MTX1	THYN1	SEC14L5	NARFL
RPUSD4	TCHP	AKAP11	ZNF213-AS1	CCNL2	PLEKHO2
SVIL-AS1	NEBL-AS1	ICMT	LINC00115	EMC4	DHRS4L2
FBXL8	DUSP10	ZC3H7A	NEK9	TM9SF1	MRPL43
ACBD5	CTAGE5	MTHFSD	SBF2	ADM	CNST
AIP	LRP6	SOCS1	RPPH1	PPFIA1	SNORD60
CARHSP1	SFMBT2	C14ORF142	TRAPPC4	NDUFA9	ZWINT
ATAD3A	TAF3	RABGAP1L	LRRC16B	BMS1P6	GSTZ1
HYKK	GPR176	FAM212B	MIR4515	ERLIN1	ORC6
IFT140	CCDC82	DRAP1	MLST8	CCDC88C	ZNF174
ASH1L	ATE1	SUSD4	CENPL	RAPGEF3	ZNF689
UBC	RNLS	FAM24B	ATF6	TMEM219	RIC8A

BLOC1S6	LRFN4	SH2B1	TROAP	AP1G2	ACADSB
ZBTB1	ENO4	FRMD6-AS1	PRDM2	POU2F1	NME7
ASUN	ST8SIA1	SLC39A9	IDH2	MYO9A	ARID4A
RNASE4	B9D1	MSTO2P	POLDIP2	CYP1A1	ZNF641
DYRK3	SEZ6L2	CHST15	CDC123	GLUD1P7	FTO
NRP1	VAV3	SNHG10	CFAP43	MIEF2	CNTROB
AP5M1	TIMM8B	EXOC5	ZFP3	RNF166	SCO1
CCDC84	METTL15	NUP107	RSPRY1	LCOR	DDX12P
PANK1	CORO1B	NDUFV1	APEX1	OGFOD1	CDC42BPB
TNKS2	LINC01134	PEX5	CTDNEP1	IQCH-AS1	SP7
PAPSS2	HEXA-AS1	SLC16A13	ST20	CYLD	MFGE8
SPATA7	NFKB2	METTL18	TLE3	KATNAL1	CFL1
ZNF697	MAP4K5	ZNF487	MASTL	L2HGDH	ANKRD13D
GNB5	RGS10	ARHGAP42	SLC38A6	CCDC77	SRM
SF3B3	BAHD1	MUTYH	SEMA4B	SEMA6C	POLG
RRN3	KIAA0391	MAPK8IP3	C1ORF174	COA6	CCDC184
TIMM23	SLC25A11	FAM21C	HAUS2	SNHG12	GLOD4
FAM173A	ATP2A1-AS1	FAM149B1	SPG7	EIF5A	ALDH9A1
CWC15	PDE8A	WDR3	DPP3	SLC44A3	WDR73
HPS1	TMCO4	STRN3	VWA9	IL4R	PFKM
MPG	VTI1A	TIMM9	PPP3CB-AS1	ATF3	LRRC8B
RECQL	ATG13	STX8	FBRS	ALDH3B1	ZCRB1
ZEB1-AS1	BMS1P4	CENPBD1	TMEM116	LCMT1	LOC102606465
CIB1	FUT8	LOC100506023	SPSB3	INPP5K	TLCD1
ZNF438	ZNF32	PEX11A	RAB26	INSM2	GEMIN4
EMC7	CRIP2	RPL36AL	DDB2	WARS	FAM86C2P
UBL7-AS1	THNSL1	RNF31	PRKG1	TP73-AS1	SOCS2-AS1
LOC100289511	MIER1	KIF1C	ENTPD7	ARL3	FUK
ASCC1	ATG2B	SCNM1	HES4	MRPS34	SEPHS1
GPATCH2	RASAL2-AS1	CNOT2	PRPF40B	TXNDC16	TIGD7
FAM118B	BRMS1L	DHRS4	MRPS16	TNFAIP1	NSMCE1
LOC101559451	ANXA11	GTF3C1	CDC42SE1	AK7	ZKSCAN2
DGKH	SLC22A17	IKZF5	CINP	DNAH2	DPH5
JRKL	SNX29	FZD8	PEMT	ERAL1	NARS2
NUDT16L1	VPS26B	DCAF4	LOC100130950	INPP5F	UBFD1
EHD4	USMG5	CAMTA2	CEP57	HTR7P1	CDIP1
RPA1	DHRS4L1	SLC19A2	MESP1	POMP	PSMA3-AS1
DPP8	BMPR1A	CCDC7	KTN1	SPG21	ATXN7L2
CHFR	FAM65A	METTL9	LOC100130987	PRMT5	ARHGAP11A
RAD9A	BBS4	PLA2G15	EPC1	PDCD11	SETD3
PLEKHH1	RBM23	STAT2	LOC101929099	PCSK7	WDR20

SRCAP	LOC399815	RITA1	C16ORF58	GPAM	DCAF5
WEE1	DDHD1	STT3A-AS1	ECD	CDCA3	KRTAP5-AS1
TMED8	NEURL4	PHB2	IQCC	WBP11	ACTN1
CDC14A	ARF1	RANGRF	VPS11	ATN1	ENTPD5
ZSCAN2	SHPK	FAM21A	KIAA0319L	C10ORF131	AP4B1
PBLD	MAP3K11	FERMT2	LINC00612	ZNF511	KDM4A-AS1
HEATR5A	CTNNBIP1	FLII	C15ORF39	SSH1	NPC2
PRSS22	BHLHE41	C16ORF62	BANF1	PPP2R3C	FAM69A
TBL3	FBXO34	PER3	RAB5B	RHBG	ERC1
HELLS	MYH10	GTPBP4	C14ORF132	TIMM10B	DCTN5
FHAD1	COPS3	FBXO18	PAPOLA	COPS7A	PWP1
NBL1	SLC36A4	DDX24	CEP104	APAF1	SFXN3
CREM	CKAP2	NR4A1	SFXN2	GLMN	FGFR1OP2
TUFM	C16ORF13	POLR3K	LINC01465	EDRF1	ELMSAN1
C17ORF100	GSG2	ZNF839	ERI2	UEVLD	LINC00933
SRP54	MIR762HG	CHCHD1	AGL	C15ORF40	RELT
MYL6B	COG7	LZIC	AMN1	DFFB	DTL
CAPN15	PML	FKBP2	SPEN	APBB1	PAN2
GID4	RRN3P3	WDR37	FAM161B	SZT2	FCF1
BTBD7	VPS53	COQ10A	TRIM45	SCARB1	MTHFD1
ARID3B	SLC35A3	DRAM2	RPS2	SPATA41	CHD9
DECR2	ZNF286B	YIF1A	TMEM199	GAS5	LRRN2
THAP10	C11ORF24	LINC00235	HERC2P3	ZNF276	ELMOD1
BCAS2	POP5	PHF13	PTGR2	DAD1	S100PBP
VANGL1	ANKS3	KAT6B	LIN52	RHOD	SMPD1
RABGGTA	PHC1	CLCN6	MEX3A	OGFOD2	FRRS1
HOMEZ	ARG2	PELI2	PARP6	FAN1	STT3A
CYP46A1	VWA8	MOAP1	MIR4687	DDX54	ALG9
LYSMD4	CNNM2	ANKRD16	LOC101928118	SARNP	PIGB
TRAP1	ATAD1	TRIM8	TMEM251	SERGEF	EIF3I
DHRS1	MAX	RB1	CAPZA1	TRMT61A	PRH1-PRR4
TM9SF3	RER1	LOC100134368	MGAT2	TAF10	MIR3178
LRR1	LINS1	ECE1	COPS2	ZNF485	INF2
COX10-AS1	NEO1	TIMM23B	MLF2	AMIGO1	RASSF10
WAC	TM2D3	RTCA-AS1	SLC25A28	DNAL1	ST3GAL3
TDP1	ZBTB25	TOX4	SLC41A1	LIMA1	ZMPSTE24
HERC2	PARK7	APITD1-CORT	SASS6	RPS6KL1	SMG7-AS1
CSGALNACT2	AKIP1	ACYP1	ERCC6-PGBD3	NFYC	MIR135B
MIS12	ZNF18	CISD1	PPME1	AREL1	CEP83
MCU	CCS	NUDT8	PPP2R5C	RNF40	MED8
KDM8	TAX1BP3	SEC61A2	CEPT1	ARHGAP12	BLOC1S2

PRDX6   LOC101927765
ARID2   OSBPL9
PLBD1   ACOT2
C12ORF65   MPHOSPH9
MRPL37   SNORD46
TPI1   PPOX
CD27-AS1   AKAP5
C1ORF43   ACAP3
EXT2   STX4
ADCY6   RAD51-AS1
BCCIP   TMEM8A
RHNO1   TMC3-AS1
IMMP1L   C14ORF1
TAPBPL
ZFYVE27
GTF2H1
ISCA2
YIPF1
RAB15
HYOU1
KLHDC9
SH2B3
ZNF408
POC1B
CDK2AP1
DENND2C
B4GALT2
RPUSD1
SEC23A
EDARADD
RAD51
KIF20B
LAMTOR2
WDR90
RUSC1
MIR1282
ZPR1
SNORD59A
IPMK
ARAP1
ANG

APPENDIX F. ASH2L knockdown RNA-seq results.

RNA-seq results following knockdown of ASH2L with two shRNA constructs (275 and 276). Sequencing read count data for each knockdown sample was compared to read count data for LacZ control. Genes were ranked by negative $\log_2$ Fold Change from control and can be found at https://www.researchgate.net/profile/Jamie_Mills2/publications. File ASH2L 275 knockdown RNA-seq DOI: 10.13140/RG.2.2.36445.20967 and file ASH2L 276 knockdown RNA-seq DOI: 10.13140/RG2.2.23652.58249.

APPENDIX G. Gene lists and ToppFun analysis: ASH2L knockdown RNA-seq downregulated genes.

The 3,278 genes in common between the two shRNA constructs, corresponding to **Figure 3.4-A**, are reported below and ToppFun analysis on these genes can be found at this link: https://toppgene.cchmc.org/output.jsp?userdata_id=4af8986f-347e-45e2-81e8-d235a39427ed.

PGLYRP2	PAXBP1-AS1	RASL10B	PEG10	C14orf1
HTR6	STX2	FBXO38	LOC102724933	FAM193A
CAPNS1	MBD3	FAM90A1	FBXO3	FAM35A
PRKACA	GATSL3	DRAP1	FBXO5	MRPS30
ZNF19	MCRS1	SP2	BSN	SH3BGRL
DHDH	BMP8A	SP4	ATXN7L3	MRE11A
ZNF14	ZNF76	SP3	MT1A	PEX10
ZNF16	SKI	KCNK5	CRISPLD2	PDCL3
SIGIRR	KCNG1	APIP	SUB1	FAM195B
VOPP1	PDCD5	SPDL1	FLYWCH2	KCNS3
SF1	MSI1	NEO1	CERKL	TACC3
SHMT1	MSI2	CERS6	DHX9	HDGF
NPRL3	FBXO24	DHPS	ZBTB2	RAD1
ADAMTS15	FADS2	VPS13B	NEU4	INPPL1
ADAMTS13	NEK2	CCDC183-AS1	RACGAP1	IGF1R
PRPF38A	MAP3K6	UHRF2	FAM192A	TGM1
TMEM254-AS1	KCNH1	UHRF1	NTNG1	PPIP5K2
ZNF34	FBXW4	SRF	BBS1	FAM198B
FRG1B	PIAS3	LRTM2	XBP1	IGFLR1
DHFR	MSH6	MLLT1	SPDYC	ENPP1
LAS1L	MSH2	U2AF1	LOC100506990	TGM3
STRN	PKP4	SRR	MC1R	CARD10
SFXN2	MIR210HG	KCNJ3	FBXL3	ZGRF1
SFXN5	BLM	NOP58	LOC100505666	ADAM11
KCNC1	IREB2	GSTZ1	APOBEC3B	SDHC
LOC100288181	FBXO45	NECAP1	MRPS11	DENND6B
KCND2	FBXO46	BRE	DMBX1	GRK1
KCND3	C6orf25	ST7	RAC3	GYLTL1B
BGN	FBXO43	KCNK1	LINC00052	GRK6
MMP16	NPAS1	SLC23A2	PRPF40B	PIBF1
RAPGEF5	FBXO41	SLC23A1	CACYBP	SDK1
GTF2H2B	NEK8	CCP110	PITPNM3	IFT22
PKN3	NEK4	LGALS4	FAM193B	TAT
CCM2L	API5	LGALS1	MICU2	IFT52

GRM2	MEX3A	C1orf189	GLOD5	COPG2
GRM8	CDKN1C	FKBPL	BCOR	NUPL1
MYBL2	SETD4	SKIDA1	TCERG1	LOC100294145
CA8	SETD5	LEMD3	ZMYND8	SPPL2B
MYBL1	SETD2	FKBP7	MKI67	ROGDI
EEF1G	CDKN1B	FKBP3	KIAA0930	C3orf67
ELF2	SETD6	FKBP5	NFU1	EME1
ELF4	ASH2L	CNP	TANGO6	EME2
ZMYND10	FGD5-AS1	C1orf167	CPT1C	GBAP1
ZMYND11	MTG2	COPS6	NIF3L1	IL11RA
SHANK3	TK1	MTL5	UHRF1BP1	ACYP1
IFT74	RRP15	C1orf159	STRIP2	MBNL1-AS1
MTA1	DENND2D	MCM7	GSE1	TIMELESS
OXCT1	ZSCAN5A	ZC3H4	ARGLU1	PMEL
MVB12B	ASH1L	C1orf131	C6orf99	ISYNA1
MVB12A	COPZ1	MCM8	GBAS	GSTCD
TDH	CIT	COL14A1	C12orf75	ALOX15
RRM1	MTF1	MCM9	CAB39L	PMF1
RRM2	RARA	ZC3H6	SLC35F4	CAMSAP1
WARS2	ANGPTL2	SYNE4	SIPA1L2	EMG1
CCS	RCCD1	TRO	C6orf48	UBB
CD9	GAS6	CEP170B	SIPA1L3	ZMAT5
SLIRP	EXOC2	PRKG1	GSG2	LONRF1
FRMPD4	PLIN5	ESPL1	ZYG11A	LONRF2
ELK1	NFIX	MCM3	LOC101927934	CPSF7
RRP9	C1orf192	MCM4	HEG1	CPSF6
MTBP	METTL21A	MCM5	SCAND2P	DAK
ALCAM	PAPSS1	MCM6	CALML3-AS1	DIAPH3
PRKCI	MINOS1	MCM2	SDCCAG8	PXMP2
DNAAF3	CDKN2C	DKFZP586I1420	DTYMK	PXMP4
SDSL	MOCOS	HDAC4	TYMS	ZCCHC4
NUP85	ELP5	HDAC2	EMC1	FGFR3
VPS16	ELP6	HDAC3	EMC6	ZCCHC3
ST3GAL3	NFIA	DGCR11	NOP14-AS1	ZCCHC2
PLK4	NFIB	HDAC1	NFYA	THRA
BTF3	C1orf174	TULP3	FAM46B	DBI
C3orf18	TNC	TTK	MTX2	RBM3
PLK3	OIP5	TENC1	NFYB	HEPH
PLK1	GTF2I	HDAC6	C6orf57	DCK
MEX3D	KCMF1	VPS52	RSG1	ZAP70
C3orf14	MAP2K4	C1orf106	PRKRA	RBL1

DCX	SIRT5	CNTNAP2	BAIAP2L2	BCCIP
RBKS	EMP2	NXT1	EPHA7	CCND1
NRXN1	EMP3	NXT2	LOC100188947	C11orf1
DDO	SIRT3	ATRAID	MXRA8	C9orf106
NXF1	F7	NGRN	ULK2	CCNE2
POC1A	F8	MB	DXO	CCNE1
BUB1	GPRC5A	LINC00337	MXRA7	C9orf116
HES2	CDC42EP4	MRPL9	FKBP1AP1	VCL
HES4	ADSSL1	LINC00339	PAXIP1	RTN3
LOC91450	CDC42EP2	DGCR5	RTF1	RTN1
DEK	ARF5	DGCR2	AARD	KLHL11
PDE10A	TOB2P1	EMX1	ART4	ENHO
ATP13A2	STAU2	VPS37C	DHX34	PTPRG-AS1
RBMX	UNG	RAD23A	PCBP2	CCNB2
MDC1	GLRA2	PIK3CA	SWI5	CCNB1
C22orf23	UNK	DTL	ARSG	KLHL26
SCAMP5	FGF14-AS2	LRP1	N4BP3	KLHL22
LOC101927691	PMS1	LRP5	ZBED4	STT3A
EML1	LRG1	PRIMPOL	ARSB	STT3B
LOC493754	SERBP1	CECR1	RAB11FIP4	ACTL10
EML6	RIMBP3	CECR2	TH	YWHAE
HGC6.3	HEPHL1	CECR5	DHX29	YWHAB
NGEF	PITX1	OR2W3	AARS	CHD3
SHOX2	LAG3	CECR6	CETN2	MEA1
SDR39U1	PPP1R16A	KATNA1	CETN3	PIDD1
C5	UBE2E1	TIFA	PCSK4	FAM178A
PRKAR2B	RBX1	SLITRK6	ENC1	C9orf142
PRKAR2A	PFKFB1	LOC100506844	DHX15	SFR1
RSU1	PFKFB3	C1orf229	NCOA5	DARS2
NETO2	ARL3	CNNM3	IMMP1L	CCNA2
LINC00304	NTPCR	C1orf226	GREB1L	RCN3
RPLP1	BCL2L12	LINGO1	FAM172A	SGK223
RPLP0	SOCS3	DUT	GCH1	SFPQ
ECSIT	CPAMD8	BAIAP2	DEFB132	PSRC1
TUBA1C	SOCS7	LRR1	NEDD1	TBC1D5
TUBA1B	BRPF3	EIF4A3	GLB1L	CHDH
PP7080	LOC100132529	TONSL	YIPF6	TBC1D7
HEY2	MDM2	UBE2L6	AUNIP	LOC150776
RPLP2	FGFR1OP	NUBP1	TARDBP	PNKP
IKBIP	TUBGCP3	UBE2L3	MAD2L1	TAF15
ACTL6A	TUBGCP4	TXNDC17	KLHL36	TAF10

C9orf173	SLCO4A1	ANKRD24	CBLN3	KPTN
ASB8	RNF166	ESRRG	SLC46A1	TLDC1
ACIN1	PCYOX1	LSM5	GEMIN8P4	STAC3
QRICH1	RNF165	LSM3	C21orf2	DLC1
ASB1	RPL12	LSM2	LOC102724814	TOMM7
RPL30	RPL11	CNOT1	MCCC2	CIB3
RPL32	GPR153	ZFP62	MTUS2	CIB1
RPL34	RNF138	TADA3	CARHSP1	RDM1
THYN1	CHN2	CNOT3	ASTL	FBL
GCSH	RPL13	SPERT	RRAGB	PTTG1
RPL35	ANKRD54	RNF14	VDAC1	NUSAP1
ADNP2	RPL18	ADRA1A	ZFPM1	TMEM14A
RPL38	RNF130	PTBP1	CBLL1	TMEM14B
RTTN	ABI2	RPL7A	TSHZ1	PAM16
RPL37	ABI3	DNMT3B	TRMT2B	ARID5B
ANKS1B	MSRB2	PDDC1	GMCL1	TMEM134
MED1	KRT8P41	WASH1	NUAK1	LMX1B
RPL23	ANKRD39	IQCB1	NUAK2	PDE9A
RPL22	LINC00263	EP400NL	TNRC6C-AS1	H6PD
IFT80	HJURP	HIAT1	UBAC2	DNAJC21
LOC100131138	DACH1	DNMT3A	CCDC88B	ZC3H18
IFT88	GRB14	ADRA2B	CCDC88A	SUDS3
IFT81	GRB10	PTBP2	CCDC88C	ZC3H13
RPL27	ANKRD31	ADRA2A	CLIC3	ASIC3
RPL26	ANKRD32	PTBP3	GDI1	ZC3H14
ALG10B	RNF44	ASPM	OLA1	DNAJC19
NEDD4L	RNF40	SNRPA1	C21orf58	DNAJC11
PNPO	BSCL2	EIF4E2	PRMT6	HSPA2
RPL41	CLIP3	SSX2IP	LINC00174	DNAJC16
PRRC2B	MELK	GATA3-AS1	MAP4K1	FES
SLC16A14	ZFP41	SMARCD1	PRMT5	DLG4
PRRC2A	NHP2	NLGN2	LINC00176	CLASRP
TMEM198	ERP27	RBBP4	PRMT1	HSPB6
SPEF1	ABHD2	PUF60	LINC00173	FRMD6-AS1
TMEM180	ABHD1	RBBP8	PRMT3	ELAVL1
PCOLCE-AS1	OARD1	RBBP7	POC5	DHTKD1
GPR137	ABL1	MEST	B3GNT6	TMEM104
EMD	ANKRD10	LINC00235	RPUSD3	TMEM107
ZC3H10	ABL2	RNF20	B3GNT4	TMEM109
TMEM173	ESRRA	SMARCE1	C21orf67	PDE4A
RNF168	ANKRD26	ILF3-AS1	SGK3	POLI

GUSB	PDGFC	NDC80	CD207	CACNA1F
POLE	LOC642852	MCMDC2	SLC2A4RG	TMEM240
POLH	MAGOHB	FAM208A	SUGP2	CACNA1H
PDE4C	FASN	TSPAN12	DHRS13	TMEM241
CTPS2	SLC25A21-AS1	TSPAN11	WFDC6	C8orf48
RNF208	CATSPER2P1	CEP290	IL20	SYCE3
CTPS1	CHAF1B	APAF1	DHRS11	SYCE2
ARID1A	CHAF1A	PRPS1	ZNF587B	TMEM249
UPK3B	ATIC	CDC45	PABPC4	TFAP2B
PALLD	SMPD4	ATXN1L	PABPC5	UBE2C
RGS19	LIN52	EIF4H	DMC1	GGH
POLDIP3	SPHK2	EIF4E	PABPC3	UBE2A
PON2	POT1	EIF4B	PABPC1	SYT7
DBF4B	ACSBG1	CDC5L	LOC100288637	HIGD2A
STRN4	XRCC6BP1	FAM209B	GBA	TMEM231
HSPD1	ATL3	TSPAN13	PCDP1	TMEM237
POLN	WRB	SYT12	MGAT3	UBE2S
TSPEAR	ATP5A1	EIF3L	RELT	UBE2T
HSPE1	LOC100652758	EIF3I	GCA	LOC730183
MEIS3	WTIP	EIF3F	EPGN	UBE2N
DLK2	TTC30B	ATXN10	NUDT7	MXD3
RGS10	TTC30A	NUDT1	MIS18A	FBN1
POMK	SLC27A1	CDC20	NUDT8	MS4A7
RGS12	GMDS-AS1	EIF2D	HNRNPA2B1	DMKN
POP1	ATN1	EPC2	OTULIN	RERG
TESK2	PLOD1	CIRBP	ZFAND3	SGOL2
F11R	CEP250	MFSD10	RPP30	SGOL1
LOC642846	CDC25C	CSNK2B	GEN1	MAN1A2
CAPN10-AS1	CDC25A	KCTD13	SLX4IP	LOC100130075
HDGFRP2	ACLY	KCTD18	POU3F2	SPIDR
FRS3	SLC27A2	EIF1B	TMEM253	EPN2
RABL3	DLX1	KCTD15	SPOCD1	TMEM210
CDCA2	CORO1A	FBF1	SCNN1G	C8orf59
CDCA3	CORO1C	OMA1	TMEM256	RER1
CDCA4	IGSF21	KALRN	SHQ1	TMEM219
CDCA5	DHRS4-AS1	MBTD1	SCNN1D	AGPAT5
CDCA7	EIF4EBP2	XRRA1	TMEM258	CASC10
CDCA8	SLC14A1	CEP192	EHBP1L1	RAP1GDS1
GATA4	FUS	IFFO2	PPIH	C8orf82
GATA3	FUZ	ALDH6A1	FBN2	AGPAT1
GATA2	DDX11-AS1	CDC16	CACNA1D	C8orf86

TMEM201	TMSB15B	TTC21B	CISD3	KANSL1L
PABPN1	CACTIN	CEP112	RPL18A	MYEF2
TMEM209	LOC100131347	LPP-AS2	PPP2R1A	VAX2
FBRS	FAM213B	SEC16A	KAT8	RAB11B
SSRP1	TAF9B	CSPG4	KAT7	FAIM3
MAP10	DMTN	WDR1	SEPHS1	TATDN1
ANKLE2	C10orf91	CEP164	KRT2	COTL1
EPS8	YTHDF2	LOC101928053	KRT8	TRAPPC13
DLEU1	CSNK1A1	SNRNP70	FLT3LG	NANOS1
CDK5RAP3	C10orf95	WDR5	ESCO2	HRASLS2
CDK5RAP2	SCAF4	WDR6	KRT5	SHISA2
REV1	LOC100129046	TTC23L	MAPK8IP1	SHISA9
GTPBP1	ADM5	RFC5	SIK3	ALG10
GTPBP3	TMEM39B	RFC3	TLN2	FANCD2
ZC2HC1C	RAB3A	RFC4	DNA2	PRKAR2A-AS1
PLEKHS1	CEP120	PLEKHG2	SPTBN5	PRPF8
MAP1A	CEP128	RFC1	TWIST1	SCUBE2
PPIEL	SLC8A2	RFC2	HTATSF1	SCUBE3
ANAPC13	BZRAP1	TFDP1	LOC100126784	RPRD1B
SEH1L	SALL4	TFDP2	TP53BP2	RPRD1A
CTBP2	PHYHD1	MTFR2	TNRC18	POU2F1
ACMSD	SLC39A6	PLEKHH2	TP53BP1	EIF2S1
DICER1-AS1	SLC39A4	CATSPER3	CCR10	POU2F3
BARD1	PLEKHJ1	PLEKHH3	SETDB1	TLX1
TPRG1	RARS2	MTFP1	LOC81691	PYCRL
SETD1A	TLE3	RAB26	FUBP3	SOWAHD
IRF2BP1	SUPT16H	RAB23	FUBP1	ATP5G2
GP2	CEP131	RAB28	TOP2A	SIX4
IRF2BP2	SPICE1	RNGTT	COL18A1	H2AFJ
ADI1	VRK1	KRR1	ZNRF2P1	SIX5
C10orf82	VRK3	FBXL19	VAV3	RHBDF1
GMEB1	ILF3	CEP152	PSENEN	H2AFY
GMEB2	RAB36	TPRKB	RAB6C	H2AFZ
YJEFN3	RAB37	FBXL14	TOP3A	H2AFX
KRI1	EIF4E1B	SNRNP40	GFRA1	RHBDF2
TEAD2	SNRPG	P2RX2	LIME1	PPP2R5D
TEAD3	SNRPF	VAPB	MIR17HG	PPP2R5C
LOC100133985	SNRPC	RAB13	RFNG	TRAPPC6A
LOC148709	SNRPA	PANK3	ZNF8	RFT1
TAF6L	CSNK1G2	TRAF3IP3	ZNF7	SKIV2L2
NCOR1	SNRPB	PPP2R2A	PSMC3IP	COLEC12

STOML2	GGA1	PPM1L	XRCC5	AP2B1
MAPK9	TCOF1	PPM1N	PPP1R21	PLEKHA4
MAPK7	SRSF10	PPM1J	XRCC2	DPY19L4
TMEM30B	EXTL3	PPM1K	XRCC3	ERCC6L
CKLF	SRSF11	PPM1D	EAF1	MIDN
PGAP1	LRRC27	BARX2	ERAL1	CLGN
PGAP2	LRRC20	SH3BP1	RPL27A	PALM
RFX2	UCKL1-AS1	LINC01208	CALM3	ARID2
PDZK1	GPR64	ATP9B	CALM2	GGT1
RFX1	RUVBL2	SH2B2	PPP1R11	GPANK1
LOC102724386	RUVBL1	SLC12A2	GPAA1	DOK3
UQCRC2	ZFP69B	CYB5B	PPP1R10	MZF1
NCAPG	MYT1	NICN1	KIF14	PSMD10
NCAPH	SLC9B1	SLC12A1	HNF4G	NXPH4
GIGYF2	SAP30	MBOAT1	AP2A1	DBIL5P
HSD11B2	OLFM2	RPL23A	BRCA1	FUT8
FIGNL1	SLC9A5	ATOX1	KIF11	KIF3A
MYO6	FRMD8	PRC1	BRCA2	NPHP1
RHBDL1	SAP25	ZAK	KIF15	ISOC1
PRRT3	SLC9A8	PLEKHB1	QSOX2	CAMTA1
PRRT2	MRTO4	IFI35	GRIPAP1	NPHP4
IL22RA1	LRRC56	CCAR1	LYNX1	PHF10
PRRT4	TACSTD2	TMPO	TMEM198B	PHF12
GPR19	GPR83	TMEM194B	ZFR	PAN2
EFCAB4A	CRNDE	TMEM194A	KIF25	KIF4A
HSPB11	MMP24-AS1	PAQR4	KIF24	PHF13
JPH3	GPR97	PAQR5	KIF23	PHF19
UPF1	EFCAB12	HNF1A	KIF22	LOC115110
LRRC46	EFCAB11	TM4SF4	NBPF1	MAST4
PGAM1	ATP7B	STX1B	KIF27	AFF3
PGAM2	ERC1	PPP1R3C	KIF2C	TMEM191A
CCSAP	DNM1L	CD2BP2	LINC01136	AFF1
TMC4	CTNNBL1	L3MBTL2	STX12	TIMM17B
PTGR2	PELP1	PPP1R35	DERL3	NCBP1
LRRC41	SIGMAR1	ANKRD19P	TMPRSS2	AZGP1P1
CKS2	VEZF1	TSPAN9	RXRA	PLA2G4B
KANK2	PPM1A	ERI1	EXOSC9	MASTL
LRRC14	NT5C3B	ERI2	EXOSC8	ZP1
LRRC15	GIGYF1	ERI3	EXOSC3	SEPT5-GP1BB
PELI3	PPM1B	XRCC6	EXOSC2	SAMHD1
GGA2	GTF3C4	XRCC4	MICB	CLN6

AURKB	LOC730101	GTF2F1	EBF4	PSPH
PARL	TRIP6	LINC01132	LOC101929080	LOC101926888
AURKA	UNC119	METTL2B	BOLA3	NPY1R
PDLIM1	RAD17	DNMT1	GRWD1	KCTD3
TAS1R1	RAD18	LINC01134	ALDH7A1	PDIA5
PDLIM4	CDC42SE2	RHOBTB2	1-Mar	ADRB2
PDLIM7	SERPINA3	LINC01124	C6orf203	SEZ6L2
U2SURP	SERPINA1	SCCPDH	SRPK3	AGER
MAPK10	SPIN4	SERINC2	SRPK1	DPP3
MAPK11	SPIN3	SYCP2	HBS1L	PDZD4
R3HDM1	SERPINA6	PSD4	FEN1	ARHGAP20
MBTPS2	ING1	HYDIN	DAZAP1	MYH11
CNTFR	FCRLB	SLC37A4	RPPH1	ALYREF
IPP	ING5	NIPSNAP1	ZMYM3	IFT122
IPW	TRIM9	PAPD7	LINC01023	ESR1
FDXR	ING2	RPL37A	RHOH	TDRD3
EARS2	TRIM6	SLC24A3	NME4	NMRAL1
ACOT7	CSRP2	SEMA3C	AP3B2	SMAD7
TRIQK	C17orf53	SFTPA2	IL17RB	ARHGAP33
SERPINF1	AP3S1	UNC93B1	RHOB	ARHGAP30
PASK	CNTD1	DPCD	RAD51C	DESI1
SYTL4	NUDT16L1	AP3M2	DNASE1L2	DESI2
NISCH	C14orf169	CTNNAL1	LINC01016	URB2
HOXB9	C14orf164	ANKRD30B	RAD54B	DLEU2L
ACOT2	IVD	SEMA4D	PBLD	NT5DC1
ACOT1	ITPKA	ANKRD30A	KIAA0754	ERBB4
PRR12	HSPA14	RPL35A	KIAA0753	NT5DC2
PRR11	BLCAP	CDC42BPB	RAD54L	NCAM2
PRR14	MZT1	TPCN1	LCMT2	REPS1
ACOT4	TM9SF1	RAD52	ANKMY1	LGI1
SKA3	C17orf62	RAD51	RANBP1	KDF1
SKA1	ATP8A2	SOGA1	FLJ30403	NAP1L1
SKA2	CLSTN2	DPF1	PRSS30P	SHCBP1
FAM160B2	RTKN2	DPF2	DNM1P35	NAP1L4
RAD21	TUSC1	TNS4	CYP2D7P	FMNL3
HOXC4	BZW2	SKP1	CAMK1	NUP107
PRR22	TUBB3	TNS1	LY6G5B	ALKBH7
PRR24	GYG2	TIPIN	RAB3GAP1	ALKBH6
SIVA1	WHSC1L1	SLC6A16	CMIP	ALKBH5
LOC284009	IGFBP5	SACM1L	ARHGAP40	CACNA2D2
GLUD1	IGFBP4	CMC2	KNTC1	CACNA2D4

LOC100630918	TP53I11	TRIM46	PTMS	TTC36
MTHFD1	LINC00993	E2F2	PPP3CB	SHROOM4
MTHFD2	ATAD1	ZNF865	PCM1	PTGES3L
SLBP	CASP8AP2	E2F7	PPP3CC	POLD3
URI1	ATAD5	E2F8	GIT1	CYTH2
MEPCE	ATAD2	TRIM45	SLTM	POLD1
LYPD5	SNX22	CNBP	ARHGEF2	POLD2
LYPD6	APEX1	CABLES2	PTMA	LINC00910
LOH12CR1	NCAPD2	TRIM36	PCNA	TTC12
ATRIP	RIF1	TRIM32	AHCY	TTC18
JAG2	TMEM167A	TNFRSF6B	COL6A4P1	TTC17
TOE1	LOC101928378	C16orf80	NOXO1	POLE4
HEXIM2	LAPTM4B	ECI2	FITM2	POLE2
BTBD19	THAP9-AS1	TRIM28	SLX4	COL7A1
TEX30	EFEMP1	TRIM24	GRIN1	ERCC1
NCAPH2	UGT2B15	GIN1	NEK11	CTBP1-AS2
PBX3	UGT2B17	TMEM170B	NUP188	PRSS1
TEX38	SRSF2	SNHG15	CELSR1	SMC5
PBX1	SRSF3	KEL	CELSR2	SMC6
FAM111B	SRSF6	SNHG10	CELSR3	SMC3
FAM111A	SRSF7	TRIM13	LINC00938	FAM86FP
AGO1	SRSF9	C16orf62	FAM3B	SMC4
NCAPG2	PKMYT1	LOC153684	FAM3D	SMC2
KDM8	LOC100129534	ACTR3B	SPINK8	FAM122B
TFCP2L1	KIAA1524	WWC2-AS2	FBXW11	GJA1
FAM115A	RHPN1-AS1	APELA	LOC400927	WDR90
IL18BP	UFD1L	HLTF	CCNB1IP1	CIRBP-AS1
CASP14	ECE1	PCK1	LINC00925	KIAA1468
MAMDC4	AP4M1	B4GALNT4	NUP160	KIAA1467
DUSP5	GRID1	PCK2	LINC00957	LUC7L
RING1	TNFRSF18	C16orf59	LINC00958	LINC00898
IFT140	TNFRSF19	BRI3BP	CNTLN	LINC00899
TEX14	ZNF878	POLA1	NUP155	PRSS3
AGPS	TRIM58	POLA2	PCIF1	LINC00893
CHST1	TNFRSF25	TP73	LINC00940	LINC00894
TP53I13	DCTPP1	GRIK5	TTC25	NUP210
URM1	TRIM55	GRIK3	TNFAIP8	KNSTRN
LOC101927021	SRSF1	GRIK4	LMNB2	WDR76
AGR2	MTHFSD	LOC100129518	LTB4R	HCFC1
AGR3	ARL13B	MTMR9LP	LMNB1	ADCY9
LINC00997	E2F1	LOC100996693	MKL2	ARMC3

NUP205	MAMSTR	ZNF764	MLEC	CCDC28B
WDR62	TPM1	PAAF1	TMEM184B	CNDP2
WDR65	THOC1	CKAP2L	PUM1	LIG1
LOC101927100	THOC3	CTDSPL	TRIM52-AS1	TMEM9B-AS1
GJC3	CHD1L	SMC1A	SMU1	LIG3
NCSTN	THOC2	C1QL4	TPX2	MATR3
BLOC1S1	THOC6	ZNF75A	RPS25	TCTEX1D2
BLOC1S3	ZNF799	SMC1B	RPS28	LOC440028
PLAC1	C20orf96	DNAJC1	ADCK2	CDH24
FDFT1	BZRAP1-AS1	NPEPL1	RPS29	TCTEX1D4
WDR52	GLT8D1	CTDSP1	ZNF813	NONO
WDR54	HMMR	DNAJC9	RPS20	ZSCAN2
WDR60	LOC101593348	FGL1	RPS23	GPSM2
MKS1	ORC5	PDHB	DCP1B	LOC100506022
SMG6	ORC4	ZNF749	C2CD3	TOPBP1
LINC00853	ORC6	LOC100130417	ZCCHC17	IQCC
HELLS	FGGY	ZNF747	ZCCHC14	RPS3A
WDR37	ORC1	HN1L	LRRCC1	IQCD
WDR33	UBR7	KIAA0101	ZCCHC24	IQCG
KIFC2	ORC3	MFAP3	PARP1	IQCH
KIFC1	UBR5	UMODL1	PARP2	IQCK
ZNF724P	ZNF789	PARS2	ZDHHC12	RABEP1
FAM129A	FANCI	RPS15	GP1BA	FCHSD2
ADCY3	CHTOP	DYNLL1-AS1	WIBG	DDAH1
ADCY1	OSBPL6	RPS14	FYTTD1	OAS3
FGD1	COA1	RPS16	AMIGO3	AKT1S1
SLC17A9	FANCM	RPS19	NOL9	CDH13
PPP1R7	FANCL	RPS18	EPPIN-WFDC6	THOP1
USP1	FANCA	PDK3	DRP2	ATP1A4
PCGF1	FANCC	RPS11	PURA	ATP1A3
RSPH4A	FANCB	BRICD5	PDPR	LIN9
WDR18	FANCG	RPS13	GINS1	ATP1A1
LOC101928491	FANCF	RPS12	GINS2	LOC100507373
VMA21	LHX2	CRNKL1	GINS3	OGFRL1
PCYOX1L	LHX1	SRRM1	GINS4	CHSY1
PARK7	ZNF775	NEIL3	MLH1	SUV420H2
PPP1CC	DDX11L2	CSTF3	AIG1	NSDHL
IPO9	UFSP1	RPAP3	GRHL2	SAMD14
RIMS3	UBTF	FXR2	MPHOSPH9	SAMD15
HMCES	PHACTR1	CHTF8	ZWILCH	LIPJ
SLC26A10	ZNF765	WBSCR27	LGALS8-AS1	DAXX

STARD7	KPNB1	SLC25A14	COMMD3	SNORD88B
AKIP1	ZNF672	C2orf81	COMMD6	ARHGEF39
PADI2	ZNF670	ZNF624	COMMD4	HMGA1
PARPBP	RNASEH2B	SLC25A22	CSDE1	BMP7
OAZ3	CPB2	MMAB	C2orf16	TAB2
HMGXB4	RNASEH2A	FOXJ3	THRIL	TAB1
INCA1	PEF1	SLC25A39	USP27X-AS1	RBM25
ERLIN1	TROAP	ZNF619	RABGAP1	NAA40
ERLIN2	MAZ	ZNF618	CCDC104	RBM26
SASS6	FOXO6	SLC25A33	DOT1L	HMGB2
TRAM2-AS1	THADA	PVRL1	PGRMC1	PSIP1
FGFBP3	GLIS2	SLC25A35	HEATR2	ATP2A1
ELOVL5	STK31	LPGAT1	HEATR4	HMGB1
ELOVL2	NPAT	FOXI3	NPM1	SEC61A2
LOC101928103	STK35	ZDHHC6	MOK	BCL7C
HIST2H2BE	ZNF668	ZDHHC4	LOC100129722	BCL7B
SUMF1	ZNF665	FOXH1	ALDH4A1	ZNF385C
SLC29A4	MCC	THAP3	FKBP1A	ZNF385B
SLC29A2	STK24	ZNF718	PER3	RBM33
TFAP2A-AS1	STK25	ZNF717	ESYT1	HS3ST3B1
PROCA1	SCGB3A1	THAP7	FOXA1	ARHGEF10
FSTL3	ARHGAP11A	ZNF714	CCDC167	RBM17
RECQL4	CELF1	TMEM87A	MPG	RBM14
C7orf34	MCM10	DSN1	SNIP1	DYNLT1
YLPM1	NRARP	ZNF701	OSR2	ARHGEF19
STXBP5L	SQRDL	ZNF700	CCDC159	PEX1
MPLKIP	FOXM1	RIMKLB	CCDC157	PEX7
ZNF696	CAPZB	RBM10	PCOLCE	LPHN1
PECR	MDK	INTS2	RPGRIP1L	SPOPL
NRROS	SUMO3	INTS7	UTP11L	NAA35
CSE1L	ZNF649	RCOR2	NPNT	RBM23
FAM228B	MYADML2	RCOR1	WBP11	RBM24
SLC16A8	SUMO2	SUV39H2	CCDC151	RBM22
SLC25A42	ZNF646	DOC2A	CCDC150	RBM4B
ZNF687	ZNF644	SUV39H1	CCDC142	BSN-AS2
KPNA2	SLC25A15	CCDC125	RBM8A	CCDC183
PRPF4B	DBF4	BMI1	ARHGEF26	CCDC170
FAM227A	SLC25A17	PEPD	CCDC148	FAM98C
ACAP3	LAMTOR5-AS1	UTP14C	FAM95A	TAF6
MAK	ZNF639	FOXD2	ZFHX2	THRSP
KIAA1731	SLC25A10	CCDC114	CCDC138	NAA16

TAF4	DISP2	ZNF608	CDK5R1	LRRC14B
CPVL	NCL	RUSC1	NRK	SPIB
CCDC173	ZNF557	ZNF606	NRM	KIAA1107
CCDC171	NUTM2D	CHEK2	NRL	C14orf79
MYB	ZNF551	CHEK1	NUCKS1	SYNGAP1
MS4A14	GLO1	ARL6IP6	TOMM20	RAVER1
LOC100507144	GLB1L2	RECQL	CHCHD7	AKT1
SARNP	PHF21B	FLJ37453	DLG5-AS1	LINC00536
PIH1D2	STIP1	PFN4	TOMM34	PKNOX1
NHLRC1	ZNF876P	PFN1	PREPL	GTSE1
CDKN2B-AS1	ZNF549	C19orf40	CTTNBP2	MNX1
PSMG3-AS1	CYP4F12	C19orf43	LTBP4	TSTD1
NR2C2AP	CHERP	C19orf48	CPLX3	SKAP1
NSUN5P1	LOC101928292	C19orf44	ENKD1	CRABP2
F8A1	ZNF534	PATZ1	TBCB	DOCK9-AS2
FZD8	PCNXL3	FIS1	TBCA	RNF219
PA2G4	NES	C19orf54	TBCK	PLA2G6
PPP5C	PCNXL4	SS18L1	MRGBP	RHOT2
PLCH1	ZNF530	C19orf55	CXCL9	RNF214
CDK12	C19orf66	SNHG5	DOCK7	C14orf93
GLI1	ASB16-AS1	ZWINT	GAS6-AS2	LINC00552
PFAS	C19orf68	C19orf57	GAS6-AS1	NR2F2
GLI3	ZNF525	MYL6B	EGFL6	CYP4F8
PITPNA-AS1	C19orf71	PRPF19	NS3BP	AIF1L
PLCB3	LOC646719	MTERFD1	SENP1	ZNF496
PLCB1	LCA5L	CCHCR1	HSH2D	DDX12P
PLCB2	SMDT1	MIR4263	TRNP1	ARHGAP1
ASAH1	ZNF519	KREMEN1	LUC7L3	MED19
LOC101929567	SCGB2A1	KREMEN2	MNS1	MED16
ZNF592	GLS2	KRT32	DOCK2	MED18
CARD9	LOC100128361	PACS2	DOCK1	MECP2
SAAL1	PAXIP1-AS1	KRT38	LUC7L2	MECOM
LRPPRC	NHS	KRT37	TSEN34	DLGAP5
RFWD3	UGGT2	MIR5587	NXN	ZNF486
ALMS1	UGGT1	ZNF137P	TUFT1	RMI2
TRMT6	PLCL2	DCPS	LTC4S	RMI1
ZNF574	MND1	AAMDC	KLRG2	JOSD2
PHLDB1	METAP1	CA12	CMTM4	BAZ1B
NBN	ARL6IP1	KRT19	CMTM6	MED22
ZNF566	C19orf25	KRT18	TICRR	MED25
PHLDA1	SOX2	TRIP13	CMTM1	MED20

L2HGDH	NUDT12	NAF1	NUDCD1	GPRIN1
ZNF473	ZCRB1	LINC00565	MIS18BP1	EEF1A2
ODC1	KHDRBS1	ITGA9	CHAMP1	TBC1D24
CENPA	LOC285847	ZNRF1	LOC100506548	WDFY1
CENPC	SMYD5	ZNRF2	SREBF1	C1orf74
NEURL1B	AGAP11	ZNRF3	DUS2	UBIAD1
EXOSC10	GMPS	HNRNPA1	KIF9	PKDCC
PRR14L	TNRC18P1	HNRNPA0	DEPDC1	TCF7
ZNF467	SLC9A3R1	HNRNPA3	DEPDC5	PKD1P1
ZSWIM5	HNRNPDL	CYP4B1	GNAI3	TBC1D13
CENPU	ZNF428	TBX6	GNAI1	TBC1D17
CENPW	BRIX1	HNRNPAB	GNAI2	VIL1
WNT7B	SMYD4	SUCLA2	RANBP3L	NAT8
KIAA1161	SMYD3	NAGS	PNMA1	NASP
CENPE	NAE1	MYO1G	TCEB3	UBXN11
ABCB10	NADK	SMARCB1	TCEB1	HOOK2
CENPH	ODF2	CRKL	RIBC1	HOOK1
ETNK2	RABL2A	OPHN1	RIBC2	ALPL
CENPI	RABL2B	SNRPD1	CAP2	TCF4
CENPJ	ARPC4	CPNE3	NOL4L	TCF3
CENPK	NACA	SNRPD3	TCEA3	AMOTL1
TAOK2	ZNF415	SPA17	TTC3	DNAL4
TSR2	CNTROB	SCAPER	RTN4IP1	DNAL1
CENPL	MIPEP	CCDC78	CAPS	GALK1
CENPM	ITGB1	ALG6	CRTC1	HP1BP3
CENPN	ISY1	ACTN1	C1orf56	FBXL19-AS1
CENPO	RMND1	KCNK12	LOXL3	WHSC1
CENPQ	ITGB2	ACTN4	BUB1B	NAV2
CRAT	ITGAL	UNC93A	TTF2	TTI2
CDKN3	NRGN	FAM76B	ACTG2	TTI1
TAF1A-AS1	LOC728752	CCDC87	LOXL1	ACTA1
EXOC3L1	EXO1	LOC100129931	ACTG1	VIMP
EXOC3L4	PPP6R3	SEC14L2	NR3C2	FAM72B
GMNN	LOC100130992	BRIP1	EEF1B2	OTUB1
CASC5	TMEM255B	EGR3	AMMECR1L	RNF8
CASC2	ACTR5	RANBP17	TBC1D31	ASAP2
CASC3	SPTB	WDYHV1	NRTN	GNG3
ZNF446	MMS22L	TSPEAR-AS1	TEKT5	GNG4
NUDT15	LRTOMT	NOL11	MORC4	TRA2B
NUDT13	ITGA7	THTPA	CASK	GNG7
ZNF443	UGT2B4	NANP	EEF1A1	DNAI1

EYA3	LOC283335	CCDC12	ZGLP1	CHMP4A
MYO15B	ABCA17P	CCDC15	POLR3H	MFI2-AS1
OTUD3	ELMOD3	LOC54944	LINC00467	PXN
GPATCH2	GNL1	TMEM132A	ISCA2	HRH3
XYLT2	COL2A1	PHKB	PNN	CLSPN
HNRNPU	TCP1	MIR614	DENR	FAM83D
LOC101927560	C18orf54	REEP5	METTL1	FAM83A
ADARB1	ADRBK2	REEP2	HSD11B1L	CBX7
ADARB2	NSD1	CNIH2	CIRH1A	CBX6
FGFRL1	ABI3BP	DIDO1	C2CD4C	CBX5
DHRS4L2	ZNF347	AMD1	MAATS1	CBX4
DDX11	FARP2	ADAMTS4	ZRANB3	CBX3
FNDC3A	FARP1	PSTPIP2	LOC728819	TRPV1
RHNO1	NR1H3	EMILIN1	RAET1E	EIF3IP1
MCMBP	DHRS2	CEBPA-AS1	RAET1K	CAPRIN1
CEP78	ZNF33A	CCDC34	RNPS1	TRPV4
PRELID2	CEP41	HAUS4	HEPACAM2	CHMP2A
AFAP1L2	ZNF331	HAUS3	XPO1	SAE1
ZNF384	CHKB-AS1	HAUS6	XPO6	KLHDC9
NFATC3	TM7SF2	HAUS5	CHMP1A	CBY1
LOC155060	CYSLTR1	HAUS1	TUBB	8-Sep
NFATC2	ZNF326	CCDC74B	NHP2L1	SPRED2
HNRNPL	HMGN2	CCDC74A	HNRNPH1	TRPS1
EHD1	ZNF324	NOS1AP	HNRNPH3	CASP2
HNRNPK	C18orf25	PWARSN	CSTA	1-Sep
HNRNPD	HMGN1	THSD4	DCST1	SAFB
HNRNPC	MAGI2	NSL1	DCST2	POMGNT2
FARSB	CXorf65	TGFBR3L	SLC22A31	2-Sep
CEP89	KAT6A	FIP1L1	MPP2	IGDCC3
CEP57	PGP	HAUS8	PTGES3	AMTN
PHF2	ZNF317	POLR2A	XPO7	PFDN6
NRSN2	TSSK6	ASTE1	LHFPL1	LOC100506499
PHF7	STAMBP	HAUS7	CT62	FAM84B
ABCC11	ZNF311	POLR2B	KDM1A	KDM3B
PHF6	CCDC24	POLR2F	RBMX2	PERM1
PHF8	CCDC30	POLR2G	BACE1-AS	THBS3
TCTN2	DVL2	CCDC77	LOC100506469	FAM168B
TCTN1	ZNF304	POLR2I	NRCAM	UGP2
PBK	CCDC17	ASCC1	HADH	PCED1A
ZNF367	CCDC18	CYP51A1	PHYH	GNA12
ELMOD2	CCDC19	CCDC59	HMHA1	KDM4A

KDM4B	ZNF250	UBALD1	SFMBT1	DCLRE1C
KDM4D	PIGO	HAT1	C9orf37	CD81
LOC101927550	PIGP	LOC145783	SPATA17	GEMIN2
FAM167B	TMEM52	ISG20L2	KIF20A	GXYLT1
CDK2AP1	MRPL37	ZBTB17	KIF20B	BAHCC1
EHMT2	CTCF	RBPMS	WNK4	PPFIA3
LOC101928530	EFCAB6	ZBTB14	SLAMF9	SLC30A8
KRTCAP2	MRPL35	GOT2	IFITM10	SLC30A1
PPP4R2	PIGX	ZBTB12	LRRN4	ATAD3C
PPP4R4	MRPL42	MLXIPL	HSPA12A	HNRNPUL1
PSMD3	EDRF1	C9orf91	PAICS	HNRNPUL2
PSMD1	EFCAB2	PTPRA	HSPA12B	SLC1A4
ZNF287	PSMA6	PIN1	ANLN	PDS5B
WDHD1	MRPL50	VPS9D1-AS1	PTPRS	SLC1A5
RBFOX2	PSMA4	SLC25A5-AS1	FAM57B	H1FX
ZNF282	PSMA2	ZNF512B	DIRAS1	PACSIN1
CLDN5	LRRC9	FUNDC1	FAM57A	SUPT5H
DDX39A	TMEM106C	HMG20B	TECR	SSH1
PSME1	PIGK	HMG20A	PTPRK	H1F0
ZNF275	ZNF239	MIR4664	BABAM1	LOC100996447
TTYH3	PIGM	HSP90AA1	SPTLC1	DDX46
ZNF273	SMTNL2	HAX1	RSRC1	DDX47
TTYH1	PET112	SYNPO2L	CCR8	DDX42
IL20RA	SRRT	ITIH4	CCR7	DDX50
IQGAP3	ZNF518A	MAK16	CCR6	NDUFS8
MRPL11	PSMB4	PARD6G	CCT3	NDUFS2
AMZ2	TMEM67	NDUFB1	LOC100128885	NUP54
ZBTB7B	HAS3	ZBTB45	KIF25-AS1	DDX23
LOC100133445	FLNA	NTN1	HSF2	KDELC2
ABCA2	FLNB	NTN3	CAMK2N2	FAM132B
CADM4	TMCO3	PARD6B	CCT7	GALNT2
CADM1	APITD1	GANAB	CCT4	SPATA33
CADM2	PSMC6	KIAA1328	TMUB1	BUD13
ABCA4	PSMC3	FCHO1	RPA2	CLUAP1
ABCA7	ZNF214	NDUFA8	KIF26A	NUP37
PSMG1	SAYSD1	DONSON	RPA3	RPL5
ZNF256	KIAA0895L	PTPN9	SPC24	RPL3
EZH2	ZBTB22	CCNF	SPC25	GABPB2
ZNF253	LOC728554	FAM53A	LOC643072	TRRAP
PIGU	ZNF207	SCN1A	DCLRE1B	STMN2
PIGT	ZBTB32	ZFP1	DCLRE1A	STMN3

RPL8	C11orf82	CUL9	DCAF16	STAM
RPL9	C11orf80	NOTCH3	HIRIP3	HTR7P1
RPL6	SF3B4	NOTCH1	MAL2	SPRTN
STMN1	SF3B2	AK3	DCAF10	ANP32B
DAPK3	GPS2	AK7	PTPN18	TMCC1
CLPSL1	AFP	AK8	SPAG7	CENPBD1
ASF1B	BCLAF1	AK9	SPAG5	BAP1
TSPYL5	C11orf73	DSCR3	SPAG6	KLLN
NOLC1	RPS3	C11orf30	NFASC	TRAM2
LOC101928674	LINC00634	CLEC7A	IMPDH1	PPCDC
ZNF143	BRD9	NACC1	MGAT4A	HINT2
ACD	RPS2	MPDU1	MGAT4B	LOC102723927
ZNF141	AGA	MAFA	C4orf27	TSC22D3
RDX	SF3A3	ABHD15	ALG1L9P	LOC102546294
CDC7	SF3A1	AQPEP	CYP4Z1	CCKBR
UPF3A	SF3A2	SNAI3-AS1	MYOZ1	RBM12B
CDC6	BRF1	CDK4	MUCL1	LOC100507501
GPN1	PRSS36	CDK2	BROX	PRIM2
GPN3	C11orf58	ALB	ARC	PRIM1
TADA2A	C11orf57	CDK1	CEP57L1	RBM15B
LINC00669	AGBL5	BRK1	CCDC42B	PTCHD2
ZNF136	MORN1	C16orf45	XRN2	DZANK1
RET	MORN3	DUSP16	TFAM	CILP2
MYOM1	HYLS1	PAGR1	FAM64A	LOC100506207
ZNF131	DGKD	SIN3A	TELO2	GDF11
EP300-AS1	CTNND2	BAI2	IGSF8	DMRT2
TCF20	WASH2P	EFHD2	SNORD35B	SEMA3B-AS1
NUF2	ZUFSP	EFHD1	ANP32A	FLJ37035
LINC00618	CDH1	BAI1	TEX264	GALNT14
GARNL3	RPS9	C4orf36	TMTC2	GALNT16
TCF19	RPS7	AMH	TMTC1	DSCC1
NDC1	RPS8	MSH5-SAPCD1	SMIM7	SCGN
TIGD3	GPT2	LOC145837	IGSF9	APLP1
TAPT1-AS1	RPS6	LOC729970	TMTC4	COL4A2
BRD2	RPSA	TRAF3	INPP5B	STK4
BRD1	C11orf35	BHLHE40-AS1	CYP4X1	DMAP1
AEN	TRAIP	EFHC1	STC2	STIL
TIMM22	ARMCX4	C4orf46	SCAI	KIF18A
TIMM21	TEX9	LOC284837	BANP	KIF18B
C11orf84	MADD	MRM1	CDT1	ATG16L2
KIAA1257	TMPO-AS1	DCAF15	CDSN	LOC100507564

B4GALT2
USP37
ABCD3
USP31
CHTF18
USP39
PKD1
SLC2A8
CKS1B
HEBP1
GMPPA
AOX1
USP30
USP46
ABCC8
CKAP5
KLF16
MST1P2
USP40
SSR4P1
MCPH1
USP13
HCN3
USP10
USP19
ELOVL2-AS1
STOM
MUTYH
TCFL5
USP21
USP22
ZNF182
ZNF180

APPENDIX H. Gene lists and ToppFun analysis: ASH2L knockdown H3K4me3 ChIP-seq and RNA-seq.

The 438 genes in common between the ASH2L knockdown RNA-seq and H3K4me3 ChIP-seq, corresponding to **Figure 3.4-D**, are reported below and ToppFun analysis on these genes can be found at this
link: https://toppgene.cchmc.org/output.jsp?userdata_id=aab4d5af-6c57-42af-8c9e-6eb4b19d807d.

ASH2L	BZW2	AP4M1	ALDH6A1	GPANK1
WHSC1L1	METTL21A	PGAP1	CSE1L	CASP8AP2
BRE	GNAI1	BRD9	NIF3L1	NTPCR
GEN1	RABL2B	LOC100288181	ZNF76	NR3C2
AGBL5	NSL1	XRCC6	PUF60	PAXIP1-AS1
FAM228B	PPP1R21	DDX11	EPC2	LRRC14
GPN1	TMEM237	CHCHD7	CENPM	SLC12A2
CCNE2	POP1	IFT22	FGD5-AS1	MRPS30
CENPO	PSMA2	PCED1A	HDAC3	NCAPH2
TNRC18	HEY2	SNHG15	LOC100129518	LOC101928530
GRM8	ZNF250	PEX1	GTSE1	XRCC4
NANP	STRIP2	DZANK1	VPS13B	WDR60
RBM33	ING2	MTBP	SIRT5	RARS2
CCT7	KCTD13	TTI1	ZNF696	PREPL
CBY1	UGP2	LYPD6	TCP1	ZNF131
ABHD1	LINC01124	PDE10A	ARL6IP6	DPY19L4
ZNF496	CRNKL1	BRD2	OTULIN	HNRNPA2B1
MPLKIP	TRIM36	COPS6	UFSP1	RAD52
AMMECR1L	MCM7	METTL1	RFC2	EIF4E
TADA3	EAF1	ZRANB3	EIF3L	TRIP6
RABL2A	TTC30B	ATRAID	DHX29	CENPH
GLI3	EFCAB6	PRIM2	PCYOX1L	CCDC88A
DESI1	SP4	EFHC1	BRIX1	MCM8
PCIF1	COPG2	MNX1	ZNF775	TRAPPC13
RING1	RAB23	TMEM198	CTDSP1	AIG1
CYP51A1	RPRD1B	LOC150776	NFYA	TAB2
POLD2	TMEM170B	STK4	R3HDM1	NCAPH
NUDT1	ZNF141	FGFR1OP	ZNF75A	KIF15
AK9	GMDS-AS1	RAET1K	WDR54	POLH
CRNDE	NCAPG2	BLCAP	KCND3	TTK
PCOLCE-AS1	PFDN6	CUL9	BRD1	DIDO1
SMPD4	ZNF7	NXT1	BARD1	RAD17
DXO	WDYHV1			

ZNF619
KIF20A
THADA
LSM3
FBXO5
TRAM2
RNGTT
DENND6B
RPS23
TTC23L
RTN1
ZNF789
CARD10
PRMT3
FANCD2
LOXL3
ORC4
BLOC1S1
SLX4IP
EP300-AS1
MEA1
PRELID2
RTN4IP1
ATP13A2
PTPN18
HCFC1
RAP1GDS1
HNRNPDL
TRIQK
USP39
WBSCR27
NUDCD1
BOLA3
LINC00899
PTPRA
DCK
PLK4
PRKRA
KCNK12
FIS1
PSMG3-AS1

TAF6
PSIP1
STAM
ZC3H14
RBM24
GSTCD
CEP250
DMC1
CTNND2
LOC100652758
SLCO4A1
GTF2F1
MED25
MRPL9
TMEM67
TFAP2A-AS1
GTF2I
SAFB
DBF4
RAVER1
PASK
GGA1
USP37
SLBP
NPHP1
EMC6
HMGB2
RHPN1-AS1
RGS19
PPIP5K2
PMS1
STRN
STK25
PSRC1
PHF10
HSPD1
PLOD1
RPS12
DAXX
PAICS

NUP188
FAM53A
SLC1A5
LSM5
CTDSPL
LOC153684
CDC25C
FUBP3
CCDC74A
UBE2C
EZH2
ZNF8
SIGMAR1
SGOL1
FBXO38
SIPA1L3
TMEM167A
CHERP
TMEM198B
MCPH1
CALM2
SKIV2L2
TATDN1
WNT7B
SUGP2
MFSD10
EXTL3
RRM1
AMZ2
RAB3A
POMK
RPS18
TUBB
SMC1A
RPL7A
AGPAT1
MRGBP
GMPPA
MAD2L1
SPPL2B
CBX7

GMNN
CETN3
CCDC157
KRT18
SMARCE1
CELSR1
GLT8D1
MBTD1
RBBP8
MMP16
ANKRD31
SKIDA1
HAUS8
ELOVL2-AS1
NDUFA8
PPM1L
RIBC2
HSD11B1L
ASF1B
BUB1
MICB
RECQL4
PTMA
ZNF239
ZBTB12
ZNF717
TPRKB
XRCC2
TM7SF2
SMIM7
CAPNS1
USP40
SACM1L
THOC1
STMN3
TCTEX1D2
CEP78
ODF2
KIAA1328
KANSL1L
RPL12

TTC3
ELOVL2
ZBTB2
TBCK
PARP2
CDKN2B-AS1
RPL9
CENPQ
AFF1
WDR5
PAPD7
TBCB
TCF19
ZDHHC4
LOC100506990
PLEKHH3
ARSG
ZC2HC1C
SLC25A14
CCDC171
PECR
PSMG1
NOL4L
TSEN34
SYNE4
KIAA1468
PEG10
PPP1R11
RPS28
FANCC
PIGX
SWI5
PNPO
NOTCH1
PAQR5
TMCO3
TP53I13
POLE4
AURKA

LOC100507564
PIK3CA
DCLRE1A
KIAA1524
SETD2
CAPN10-AS1
KLHL26
TOPBP1
SS18L1
KIF4A
HMMR
AKT1S1
ZNF700
CENPE
MAFA
NCL
MAPK11
CNBP
ZNF446
SLC25A33
SEP2
ZNF566
ASAH1
TRIM46
BLOC1S3
SYCE3
NRM
MED22
RAD18
RFX1
CDKN1C
AP2A1
MIR4664
SUPT5H
ZNF473
GPRIN1
CEP112
NXF1
GRHL2
HNRNPUL1

PPP1R10
NEK11
TBC1D5
POLR2I
ZNF311
MED16
ADRB2
CCDC74B
DTYMK
AGPAT5
SDSL
PLEKHH2
HMG20B
DCAF16
PIGP
ZMYM3
HNRNPL
ESCO2
CECR5
DLX1
ADCY3
SEC16A
TECR
MIS18A
SNHG5
ADARB1
BCL2L12
ALYREF
WDR62
RRP9
RNF214
TRIM32
VRK3
ZGRF1

APPENDIX I. Gene lists and ToppFun analysis: ASH2L, NSD3-S, and ESR1 knockdown expression analysis and ASH2L knockdown H3K4me3 ChIP-seq overlap.

The 44 genes in common between the ASH2L knockdown RNA-seq and H3K4me3 ChIP-seq and NSD3-S and ESR1 knockdown microarrays, corresponding to **Figure 3.5-C**, are reported below and ToppFun analysis on these genes can be found at this link: https://toppgene.cchmc.org/output.jsp?userdata_id=be0407ac-a6be-4e45-b138-28ef7cb260d7.

The 320 genes that remain unique to the ASH2L knockdown RNA-seq and H3K4me3 ChIP-seq group following the comparison in **Figure 3.5-C** also appear below and ToppFun analysis on these genes can be found at this link: https://toppgene.cchmc.org/output.jsp?userdata_id=69413b86-3639-4bd6-8210-dad543db7860.

44 Genes:

CCNE2
RING1
MCM7
NCAPG2
CSE1L
CENPM
TCP1
RFC2
DHX29
CENPH
KIF15
TTK
KIF20A
LSM3
FBXO5
RTN1
MEA1
WBSCR27
PLK4
PRKRA
STAM
USP37
SLBP
HMGB2
SLC1A5
LSM5
CDC25C

UBE2C
EZH2
SGOL1
AMZ2
MAD2L1
GMNN
HAUS8
ASF1B
BUB1
RECQL4
ELOVL2
CENPQ
DCLRE1A
KIAA1524
HMMR
CENPE
ADCY3

320 Genes:

ASH2L
BRE
GEN1
FAM228B
GPN1
TNRC18
GRM8
NANP
RBM33
CBY1
ABHD1
ZNF496
MPLKIP
TADA3
RABL2A
GLI3
DESI1
PCIF1
CYP51A1
NUDT1
AK9
CRNDE
PCOLCE-AS1
SMPD4
DXO
BZW2
METTL21A

GNAI1
RABL2B
NSL1
PPP1R21
TMEM237
POP1
ZNF250
STRIP2
UGP2
LINC01124
CRNKL1
TRIM36
EAF1
TTC30B
EFCAB6
SP4
COPG2
RPRD1B
TMEM170B
ZNF141
GMDS-AS1
AP4M1
PGAP1
BRD9
LOC100288181
XRCC6
DDX11

CHCHD7
IFT22
PCED1A
SNHG15
PEX1
DZANK1
TTI1
ZRANB3
ATRAID
PRIM2
EFHC1
MNX1
TMEM198
LOC150776
STK4
RAET1K
BLCAP
CUL9
NXT1
NIF3L1
ZNF76
PUF60
EPC2
FGD5-AS1
GTSE1
VPS13B
SIRT5
ZNF696

ARL6IP6	DENND6B	RAVER1	CCDC157	CCDC171
OTULIN	RPS23	PASK	KRT18	PECR
UFSP1	TTC23L	GGA1	SMARCE1	NOL4L
EIF3L	ZNF789	NPHP1	CELSR1	TSEN34
PCYOX1L	PRMT3	EMC6	GLT8D1	SYNE4
ZNF775	LOXL3	RHPN1-AS1	MBTD1	KIAA1468
CTDSP1	ORC4	PPIP5K2	RBBP8	PEG10
NFYA	BLOC1S1	PMS1	MMP16	PPP1R11
R3HDM1	SLX4IP	STRN	SKIDA1	RPS28
ZNF75A	EP300-AS1	STK25	ELOVL2-AS1	PIGX
WDR54	PRELID2	PSRC1	NDUFA8	SWI5
KCND3	RTN4IP1	PHF10	PPM1L	PNPO
BRD1	ATP13A2	PLOD1	RIBC2	NOTCH1
GPANK1	PTPN18	RPS12	PTMA	PAQR5
CASP8AP2	HCFC1	DAXX	ZBTB12	TMCO3
NTPCR	RAP1GDS1	NUP188	ZNF717	TP53I13
PAXIP1-AS1	HNRNPDL	FAM53A	XRCC2	POLE4
LRRC14	TRIQK	CTDSPL	TM7SF2	LOC10050756
SLC12A2	NUDCD1	CCDC74A	SMIM7	PIK3CA
LOC10192853	BOLA3	ZNF8	CAPNS1	SETD2
XRCC4	LINC00899	FBXO38	SACM1L	CAPN10-AS1
WDR60	PTPRA	SIPA1L3	THOC1	KLHL26
ZNF131	DCK	CHERP	STMN3	TOPBP1
DPY19L4	KCNK12	TMEM198B	TCTEX1D2	SS18L1
HNRNPA2B1	FIS1	MCPH1	KIAA1328	ZNF700
RAD52	PSMG3-AS1	CALM2	KANSL1L	MAFA
EIF4E	TAF6	SKIV2L2	RPL12	MAPK11
TRIP6	PSIP1	TATDN1	TTC3	CNBP
CCDC88A	ZC3H14	WNT7B	ZBTB2	ZNF446
MCM8	RBM24	SUGP2	TBCK	SLC25A33
TRAPPC13	GSTCD	MFSD10	CDKN2B-AS1	SEP2
AIG1	CEP250	EXTL3	AFF1	ZNF566
TAB2	DMC1	RAB3A	WDR5	ASAH1
NCAPH	CTNND2	POMK	PAPD7	TRIM46
POLH	LOC10065275	TUBB	TBCB	BLOC1S3
DIDO1	SLCO4A1	SMC1A	ZDHHC4	SYCE3
RAD17	GTF2F1	AGPAT1	LOC10050699	NRM
ZNF619	MED25	MRGBP	PLEKHH3	MED22
THADA	TFAP2A-AS1	GMPPA	ARSG	RAD18
TRAM2	GTF2I	SPPL2B	ZC2HC1C	RFX1
RNGTT	DBF4	CBX7	SLC25A14	CDKN1C

AP2A1
MIR4664
SUPT5H
ZNF473
GPRIN1
CEP112
NXF1
HNRNPUL1
PPP1R10
NEK11
TBC1D5
ZNF311
MED16
ADRB2
CCDC74B
DTYMK
PLEKHH2
HMG20B
DCAF16
ZMYM3
HNRNPL
ESCO2
CECR5
DLX1
SEC16A
TECR
MIS18A
ADARB1
ALYREF
RRP9
TRIM32
VRK3
ZGRF1