Epigenetic and Transcriptional Response to HDAC Inhibition in Triple Negative Breast Cancer Cells

By

Danielle Marie Walheim

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Molecular Physiology and Biophysics

May, 2016

Nashville, Tennessee

Approved:

Bryan Venters, Ph.D.

Peter Anthony Weil, Ph.D.

Gregor Neuert, Ph.D.

“Whether you think you can or you think you can’t, you’re right”—Henry Ford

To my wonderful parents, whom have always supported me—without them I would not be where I am today

and

To my heroes: Diane Walheim, Norma Lynn Layne and Kimberlee Layne, who have battled triple negative breast cancer and whose strength astounds me. They will forever be my inspiration.

ACKNOWLEDGEMENTS

This work is a compilation of my research since arriving at Vanderbilt. Throughout my time here I have learned more than I could have expected. Each mentor I have come across was willing to guide me in matters of both science and life, and for that I am very grateful. First, I would like to thank my advisor, Dr. Bryan Venters, who has provided me with support and guidance throughout my graduate career. Dr. Venters was very encouraging and always pushed me to pursue my dreams. I am so appreciative of his help in the development of this thesis. I would also like to thank the members of my thesis committee, Dr. Tony Weil and Dr. Gregor Neuert, who have guided me throughout my time at Vanderbilt and encouraged me to think critically for both science matters and matters of life. I appreciate my lab mate, Dr. Chirie Sumanasekera, for her valuable advice in both my scientific career and my personal life as well.

I would also like to thank my family and friends. I thank my parents who have always supported my dreams. They have both molded me into the strong and driven person I am today, and without their guidance I would never have achieved all that I have. I am forever grateful for their love and encouragement. I would like to thank my siblings, and best friends: Lauren, Matthew, Nicole and Brian, who have encouraged me along the way. My friends that I have made at Vanderbilt have also supported me immensely. I would especially like to thank Kristin Peterson and Angela Howard who graciously edited for me. Finally, I would like to thank my love, Chris Sprunger, for his endless support. Thank you all for always being my cheerleaders!

This thesis research was supported in part by the Molecular Endocrinology Training Grant, (DK07563) and VICTR resource vouchers (VR16652 and VR10785) as part of the CTSA award No. UL1TR000445.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

bp	base pairs
BWA	Burrows-Wheeler Aligner
ChAsE	Chromatin Analysis and Exploration
ChIP	Chromatin-Immunoprecipitation
ChIP-exo	ChIP with exonuclease, coupled to high-throughput sequencing
ChIP-seq	ChIP, coupled to high-throughput sequencing
DAVID	Database for Annotation, Visualization and Integrated Discovery
DNA	Deoxyribonucleic Acid
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic Acid
EMT	Epithelial to Mesenchymal Transition
ENCODE	Encyclopedia of DNA Elements
ER	Estrogen Receptor
EtOH	Ethanol
FDR	False Discovery Rate
HAT	Histone Acetyltransferase
HDAC	Histone Deacetylase
HDACi	HDAC inhibitors
Her2	Human Epidermal growth factor Receptor
HMM	Hidden Markov Modeling

HOMER	Hypergeometric Optimization of Motif EnRichment
IC_{50}	Inhibitory concentration that results in 50% cell viability
IGV	Integrative Genomics Viewer
kb	kilobase
LAD	Lamin B1 associated domains
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NCBI	National Center for Biotechnology Information
NRO	Nuclear Run-On
Pan	Panobinostat
PCR	Polymerase Chain Reaction
PNK	PolyNucleotide Kinase
Pol II	RNA-polymerase II
PR	Progesterone Receptor
PrG	Protein G
PRO-seq	Precision nuclear Run-On and sequencing assay
RefSeq	NCBI Reference Sequence Database
RNA	Ribonucleic Acid
RPKM	Reads Per Kilobase per Million of total mapped reads
RPMI	Rosewell Park Memorial Institute (media)
rpm	rotations per minute
TAP	Tobacco Acid Pyrophosphatase
TNBC	Triple Negative Breast Cancer
TSG	Tumor Suppressor Gene

TSS	Transcription Start Site
TTS	Transcription Termination Site
VANTAGE	Vanderbilt Technologies for Advanced Genomic

CHAPTER I

INTRODUCTION

Triple Negative Breast Cancer

According to the National Cancer Institute, 1 in 8 women will develop invasive breast cancer during her lifetime, with over 200,000 women in the US diagnosed each year[1]. Approximately 15% of these cancer diagnoses are classified as Triple Negative Breast Cancer (TNBC)[2], meaning that these cancers lack the estrogen receptor (ER-), progesterone receptor (PR-) and lack of amplification of human epidermal growth factor receptor (Her2-) unlike other breast cancers. TNBC is a more aggressive form of breast cancer than ER+ or Her2 breast cancers. Patients presenting with TNBC typically have tumors of higher histological grade which are more likely to recur earlier at metastatic sites[2]. Like most cancers, breast cancer is a heterogeneous disease characterized by different morphologies, biology, and clinical evolution. This is due, in part, to the presence or absence of the different hormone receptors. These hormone receptors (ER, PR, Her2) are used for targeted therapies for breast cancer patients[3,4]. However, because TNBC cancers lack these receptors, they do not respond to these common targeted therapies. Therefore, in addition to radiation or surgery, the standard course of treatment is chemotherapy—which indiscriminately kills rapidly dividing cells, not just the cancer cells. Many patients do not respond well to chemotherapies, and their cancers tend to recur with early metastasis and a poor prognosis due to developed drug resistance[5]. Due to these issues with

TNBC treatment, there is a major need to understand the molecular basis of TNBC and to develop more effective treatments with this knowledge.

TNBC subtypes

Not only is breast cancer a heterogeneous disease, but the classification of TNBC is heterogeneous itself. Expression profiling done by the Pietenpol lab has established seven subtypes within TNBC (as seen in Table 1)[6]. Of these subtypes, the basal-like (BL) subtype represents the most prominent TNBC subtype. Therefore, the work presented in this thesis focuses on a BL2 model cell line, HCC1806. The heterogeneity of the disease and the absence of well-defined molecular targets together make TNBC difficult to treat. My hope is to shed light on some up-and-coming treatments and to help gain insight into how to enhance these treatments for women with TNBC.

Epigenetics in Cancer

The traits that make up the morphology of a cell are encoded in the cell's genome. The cells within an organism all have the same genetic background, with the exception of some somatic mutations. So, how do cells with the same genetic makeup become different tissues? While mutations may be the drivers for some cancers, including some subtypes of TNBC, not all TNBCs have common somatic mutations, which makes it difficult to identify driver mutations for targeted therapies[7]. How do some of these cells become cancerous while others remain healthy? One theory is that cancer is driven by abnormalities at both the epigenetic level as well as in the genomic DNA[8].

Table 1: Preclinical Models for TNBC

Cell Line	Cancer Type	Subtype	Known Mutations
HCC38	TNBC	Basal-like 1	*CDKN2A, TP53*
MDA-MB-468	TNBC	Basal-like 1	*PTEN, RB1, SMAD4, TP53*
HCC1806	TNBC	Basal-like 2	*CDKN2A, KDM6A, STK11, TP53, UTX*
SUM149	TNBC	Basal-like 2	*BRCA1*
MDA-MB-453	TNBC	Luminal	*CDH1, PIK3CA, PTEN*
CAL51	TNBC	Mesenchymal	*PI3KCA*
Controls			
MCF10A	non-cancerous breast tissue		
MCF7	ER+ breast cancer		
HeLa	Ovarian		
HEK293	Kidney		

*****Mutations taken from COSMIC database (sanger.ac.uk/genetics/CGP/cosmic/)

Epigenetics is the study of heritable changes in gene expression that do not involve changes to the DNA sequence itself. The epigenome is made up of DNA methylation patterns and posttranslational histone modifications[9], which help determine when a gene is turned on or off. These changes can affect gene expression patterns, and therefore physiology throughout normal-development and in disease states[8]. Posttranslational modifications on histone tails make up the "histone code". It is thought that much like the genetic code, the histone code can be decoded to reveal another layer of information on the genome[10]. The modification of histone tails regulates how compact or accessible the genome is to factors that control rates of transcription. Therefore, dysregulation of epigenetic control and disruption of the chromatin are hallmarks of tumorigenesis[8], due to the cell's ability to disrupt normal transcription. While there are many known epigenetic marks, this thesis focuses on acetylation of histone tails and a drug implicated with this modification in TNBC.

The histone code can be read to reveal important biological regions of the genome, such as enhancers, promoters, heterochromatin, transcription termination sites (TTS), and repressed chromatin, to name a few. Promoters are *cis*-acting regions upstream of transcription start sites (TSS) of a gene where initiation of transcription occurs[11]. Enhancers are also *cis*-acting elements that activate transcription in an orientation and distance independent manner[12]. Transcription factors interact with promoters and enhancers to activate (or repress) transcription of genes. By defining active enhancers and promoters, we are able to gauge the epigenetic landscape and the potential effects of the epigenome on cancer cells.

Histone Deacetylases as Therapeutic Targets

As the role of epigenetic dysregulation is becoming more apparent in cancer, inhibitors targeting proteins involved in the regulation of epigenetic marks are being developed by pharmaceutical companies. Histone acetylation was the first histone modification discovered 50 years ago and the acetylation of histones has since been associated with gene activation[9]. Importantly, histone deactylase inhibitors (HDACi) have recently emerged as promising anti-cancer drugs[13].

Many biological roles have been attributed to the 18 HDACs characterized in humans. These roles include the regulation of transcription, cell cycle progression, and DNA damage and replication[9], emphasizing the complex role HDACs play in the cell. HDACs are further classified into four subcategories based on their homology to *S. cerevisiae*. One of the primary functions of HDACs is to oppose the actions of histone acetyltransferases (HATs). HATs acetylate the ε-amino groups of lysine residues of histone tails while HDACs remove the acetylation from the histone tails, which restores the positive charge of the lysine residue[9]. The change in the charge is what creates the "opening" and "closing" of the chromatin structure due to the change of affinity between the histone and DNA. When there is no acetylation on the histone, the positive charge of the lysines creates a more compact, or closed, chromatin structure. The removal of this positive charge through acetylation decreases the histone's affinity to the negatively charged DNA, which allows unwinding and opening of the chromatin[9]. This is why acetylation is associated with active transcription.

Because of the reversible nature of epigenetic alterations, epigenetic therapy has become a promising candidate for cancer research[8]. HDACi especially are gaining traction as viable targeted therapies in certain cancers. As of November 2015, three HDACi, Vorinostat, Depsipeptide and Panobinostat, have been FDA approved. The first two were approved to treat

cutaneous T-cell lymphoma while the third just received approval to treat multiple myeloma (clinicaltrials.gov). At least 12 different HDAC inhibitors are in clinical trials for both hematologic and solid tumors. Indeed, class I HDACs are aberrantly overexpressed in several cancers, including breast and lung[14-16]. Although there are no FDA approved epigenetic drugs currently for the treatment of breast cancer[17], Panobinostat is currently in clinical trials for use in breast cancer[18]. Taken together, this led us to investigate the ability of Panobinostat to halt cell growth specifically in TNBC, since the physiological impact and mechanism of action in TNBC remains largely unknown.

Tumor Suppressor Genes

Tumor Suppressor Genes (TSGs) are often mutated in cancers. These genes get their name because they regulate cell growth through preventing uncontrolled growth. In normal functioning cells, TSGs work in balance with proto-oncogenes to keep the cell under regulated control. Cancer causing TSG mutations typically inactivate the TSG, which frees the cells from the growth suppression[19]. It is thought that most malignancies occur following inactivating mutations in TSGs and activating mutations in oncogenes. TSGs fall under many categories, but some common ones are DNA damage and repair genes (*BRCA1/2, ATM*), genes involved in signaling (*PTEN*, *EGR1*) and transcription factors (*TP53*, *FOXA2*)[20]. An imbalance of TSGs and oncogenes may lead to an imbalance of proliferation and homeostasis. As proliferation begins to dominate, uncontrolled growth occurs resulting in tumorigenesis.

Specific Aims and Summary of Results

I hypothesize that transcription of TSGs is repressed in TNBC cells through aberrant epigenetic silencing via histone lysine deacetylation. My thesis project aims to test that increased chromatin acetylation associated with Panobinostat treatment leads to increased TSG expression and growth inhibition. This hypothesis is summarized in the model seen in Figure 1.

To test this hypothesis, I proposed multiple methods to determine the effects of Panobinostat. I first set out to determine the physiological impact that HDAC inhibition has on TNBC cell growth. Using a spectrum of several preclinical models representing different TNBC subtypes, I found that nearly all TNBC subtypes tested are sensitive to the HDACi, Panobinostat. Since one of the primary actions of HDACs is deacetylating histones, I predicted that Panobinostat was causing increased acetylation of histones. It was confirmed through western blot analysis that specific histone marks were becoming hyperacetylated following Panobinostat treatment.

In Aim 2, I set out to determine what the epigenetic changes were. I sought to answer the questions “where is the increase in acetylation located throughout the genome?” and “is the increase in acetylation correlated with an increase in transcription at these loci?”. As my model shows, I hypothesized that the increased acetylation of histones was occurring at the promoters and enhancers of TSGs. I also predicted that these TSGs would have upregulated expression following Panobinostat treatment. This upregulation would then be responsible for the growth hindrance of the TNBC cells. To identify the genes most sensitive to HDAC inhibition, I conducted a genome-wide transcription profiling by PRO-seq and a genome-wide epigenetic profiling by ChIP-exo at several time points of HDACi treatment.

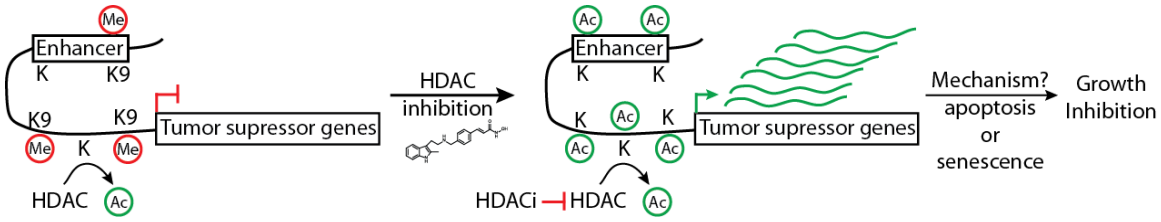


Figure 1: Model for HDACi mediated impairment of TNBC cell growth. The epigenetic state of cancerous cells is aberrant, through deacetylation of lysine (K) on histone tails and other repressive marks (i.e.) H3K9me) leading to repression of TSGs. Panobinostat leads to the hyperacetylation of these repressed regulatory regions, enabling the transcription of TSGs and therefore inhibiting cell growth.

CHAPTER II

PHYSIOLOGICAL EFFECTS OF HDAC INHIBITION ON TNBC CELLS

Introduction

While Hanahan and Weinberg established that the ability of cells to evade apoptosis is a hallmark of most cancers, each cancer may be unique in its mechanism for this evasion[21]. It is well documented that the epigenome of cancer cells is often dysregulated, and this dysregulation is one mechanism that may enable cells to bypass surveillance pathways that would otherwise cause the cells to undergo apoptosis or senescence.

Due to the overwhelming evidence that the epigenome is dysregulated in cancers, investigators have begun evaluating the effectiveness of epigenomic therapies in clinical trials. One of these therapies is the HDACi, Panobinostat (LBH589). In other preclinical studies of Panobinostat, investigators have confirmed that Panobinostat is a potent pan-HDACi that has an inhibitory effect on a wide range of hematologic malignancies at low nanomolar concentrations[22]. These affects are hypothesized to be an effect of inhibiting multiple cancer related pathways and reversing the aberrant epigenetics implicated in tumorogenesis[23]. There are currently a few preclinical studies occurring in solid tumors as well. Based on the preclinical studies, Panobinostat has undergone rapid clinical development, where it has been approved for use in some hematological malignancies and is currently in clinical trials for some solid tumors. In both hematological malignancies and solid tumors, Panobinostat has shown promising clinical responses with low levels of toxicity[2]. However, whether Panobinostat treatment results in the associated epigenetic changes in TNBC remains unclear. Establishing this link will enable

investigators to develop appropriate combinatorial therapies that may result in more effective TNBC cancer treatments.

To test our hypothesis (Figure 1) that Panobinostat inhibits a broad panel of TNBC subtypes, I assessed Panobinostat's effects on the cell growth of a representative panel of TNBC subtypes, characterized by the Pietenpol lab[6]. I further sought to characterize the effects Panobinostat had on histone acetylation and gene expression in the BL2 model, HCC1806.

Since Panobinostat is an HDAC inhibitor, our model (Figure 1) predicts that Panobinostat treatment would cause an increase of histone acetylation in the cells. The Collins-Burow's lab also showed that increased acetylation occurred following Panobinostat treatment[2], but did not analyze this increase genome-wide. Again, I wanted to both verify and expand on their findings. I set out to show that in our model cell line, HCC1806, histone acetylation increased at multiple histone marks in the 16-hour time frame of our study. These preliminary studies will allow me to continue on to my in-depth epigenetic studies.

Materials and Methods

Cell lines

Human TNBC (HCC-1806, HCC-38, MDA-MB-468, MDA-MB-453), human mammary epithelial cells (MCF-10A), cells characterized as ER-positive/PR-positive breast cancer (MCF-7), human embryonic kidney cells (HEK293T), and human cervical adenocarcinoma cells (HeLa) were all obtained from ATCC. SUM149 (TNBC cells) was provided by Asterand. CAL-51 (TNBC) cells were obtained through the Leibniz-Institut DSMZ. Each of the TNBC cell lines requires a different growth media, whose recipes are detailed in Table 2. The cells were all

Table 2 **Cell culture conditions**	
Cell Line	**Media Recipe**
HCC-1806	RPMI 1640 (Gibco 11875-093), 10% bovine serum (Gibco 16170-078), 1% penicillin/streptomycin (Gibco-15146-122), and 1% L. glutamine (Gibco 25030-081)
HCC-38	RPMI 1640, 10% bovine serum, 1% penicillin/streptomycin and 1% L. glutamine
SUM149	Ham's F12 (1:1) nutrient mix (Gibco 11765-054), 5% bovine serum, 0.2 units of insulin/mL (Novolin), and 1μg/mL hydrocortisone (Sigma H0888)
MDA-MB-468	DMEM (Gibco 11995-065), 10% bovine serum and 1% penicillin/streptomycin
CAL51	DMEM, 10% bovine serum and 1% penicillin/streptomycin
HEK293TA	DMEM, 10% bovine serum and 1% penicillin/streptomycin
HeLa	DMEM, 10% bovine serum and 1% penicillin/streptomycin
MDA-MB-453	DMEM, 10% bovine serum, and 1μg/100mL EGF (Life Technologies PHG0311)
MCF7	Advanced DMEM (Gibco 12491-015), 10% bovine serum, 1% penicillin/streptomycin and 0.2 units of insulin/mL
MCF10A	DMEM/F-12 with HEPES (Gibco 11330-032), 5% bovine serum, 1% penicillin/streptomycin, 2μg/100mL EGF, .2 units of insulin/mL, 0.1 mg/mL cholera toxin(Sigma C8052), and 0.5μg/mL hydrocortisone

maintained at 37°C in 5% CO_2.

Drug Screen

Each adherent cell line was plated in triplicate in 96-well plates, at ~20% confluency in 100µl of media (Table 2) and allowed to attach to the microtiter plate overnight. To determine the IC_{50}, the cells were then treated with a range of Panobinostat (Selleck Chemicals #S1030) concentrations (10nM, 30nM, 100nM, 300nM, 1µM, 3µM, and 10µM) or an equivalent concentration of DMSO for 72 hours.

MTT cell proliferation assay

The MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay is a colorimetric assay, in which cell proliferation is measured by the color change of the yellow MTT dye to the reduced formazan, which is purple in color. Each assay was performed according to manufacturer's instructions (Sigma-Aldrich) in triplicate. Briefly, after 72 hours of drug treatment, media was replaced with fresh media and 10µl of MTT reagent was added to each well. Cells were incubated for 1-6 hours (until color change apparent) followed by the addition of 100µl of SDS solubilizing reagent to each well. After incubating plates overnight at 37°C, the absorbance was measured at 560nm. The percent viability was determined by the ratio of drug treated cells to DMSO treated cells.

Immunoblot analysis

Cell pellets were resuspended in 1x protein sample buffer (10% glycerol, 2.5% SDS, 0.7135M β-mercaptoethanol, 0.002% bromophenol blue, 62.5mM Tris-HCl pH 6.8), and heated at 95°C for

5 minutes. SDS-PAGE was performed on 15% acrylamide gels, loading $1x10^5$ cell equivalents of whole cell extract/well. The separated proteins were then transferred to 0.2μm nitrocellulose membranes at 30V for 70 min in western transfer buffer (25mM KH_2PO_4, 25mM K_2HPO_4, 0.5mM EDTA pH 7.2) with 20% methanol. After transfer, the membranes were blocked with 5% milk in TBST (10mM Tris-Hcl, 5mM HCl, 150mM NaCl, 0.1% Tween 20) and then probed with Ac-Histone H3 (Lys9/14) (sc-8655, 1:500 dilution) from Santa Cruz Biotechnologies, H3K27Ac (ab4729, 1:2000 dilution) or H3K9Ac (ab4441, 1:2000 dilution) from Abcam. The membranes were then stripped with two washes in stripping buffer (1.5% glycine, 0.1% SDS, 10% Tween 20, pH 2.2), as described by Abcam, and probed for Histone H3 (Santa Cruz; sc-10809, 1:500 dilution) as a loading control. Each blot was probed with the appropriate anti-rabbit or anti-goat secondary antibody from Santa Cruz (sc-2004 or sc-2020). Proteins were detected with BioRad Clarity ECL substrate following manufacturer's instructions and imaged using BioRad ImageQuant apparatus.

Results and Discussion

Cytotoxicity in response to Panobinostat

Although a previous report showed that at least one TNBC cell line was sensitive to Panobinostat[2], it was not known whether the full panel of TNBC subtypes were also sensitive to Panobinostat. Thus, I wanted to test Panobinostat's effectiveness in a range of TNBC subtypes as defined by the Pietenpol lab (Table 1). After Panobinostat treatment for 72 hours, IC_{50}s were determined for 10 cell lines. The IC_{50}s were determined after graphing the MTT normalized absorbance vs. the Panobinostat treatment concentration and measured as the concentration of

Panobinostat in which 50% of the cells were viable as compared to DMSO control (set at 100% cell concentration). An example of this can be seen in Figure 2.

Panobinostat appears to have potent anti-proliferative activity against TNBC cells. The non-TNBC breast cells are less sensitive to the Panobinostat treatment than the TNBC cells, but are more sensitive than the other tissue controls. As shown in Figure 2, after 72 hours, the non-breast tissue cells remained insensitive to higher concentrations of Panobinostat. As seen in Figures 2 and 3, the IC_{50} of HCC1806 was measured at 35nM. There is a plateau at higher concentration of the drug that does not dip down to 0% viability. This may be due to maximum inhibitory effect of the drug. The IC_{50}s were determined in a similar way for each of the 10 cell lines tested which are summarized in Figure 3 and can be seen in Supplemental Figure 1 in Appendix C. The highest IC_{50} for the TNBC cell lines was MDA-MB-468 at 87nM, consistent with a previous report[2]. All of the other TNBC cell lines had IC_{50}s between 20 and 40nM. The IC_{50}s for the HeLa and HEK293 control cells were 313 and 436 nM, respectively. Meanwhile, the IC_{50}s for the breast tissue control lines were 187 and 163nM for MCF7 and MCF10A, respectively. Due to these IC_{50}s it is evident that TNBC is more sensitive to Panobinostat than other cell lines.

These findings expand upon previous reports[2], and confirm some of their initial findings. The Collins-Burow's lab also found that MDA-MB-468 was the least sensitive TNBC cell line to Panobinostat treatment. The relatively higher resistance of MDA-MB-468 to Panobinostat does not appear to be subtype specific because HCC-38 is also basal-like (BL1) TNBC. It would be interesting to determine what features of MDA-MB-468 make it more resistant to Panobinostat, but currently this remains unknown. There are a few limitations to the previous report on Panobinostat sensitivity in TNBC. Only a few subtypes of TNBC were tested with Panobinostat treatment instead of a full panel, there were few controls, and their sensitivity assays were not

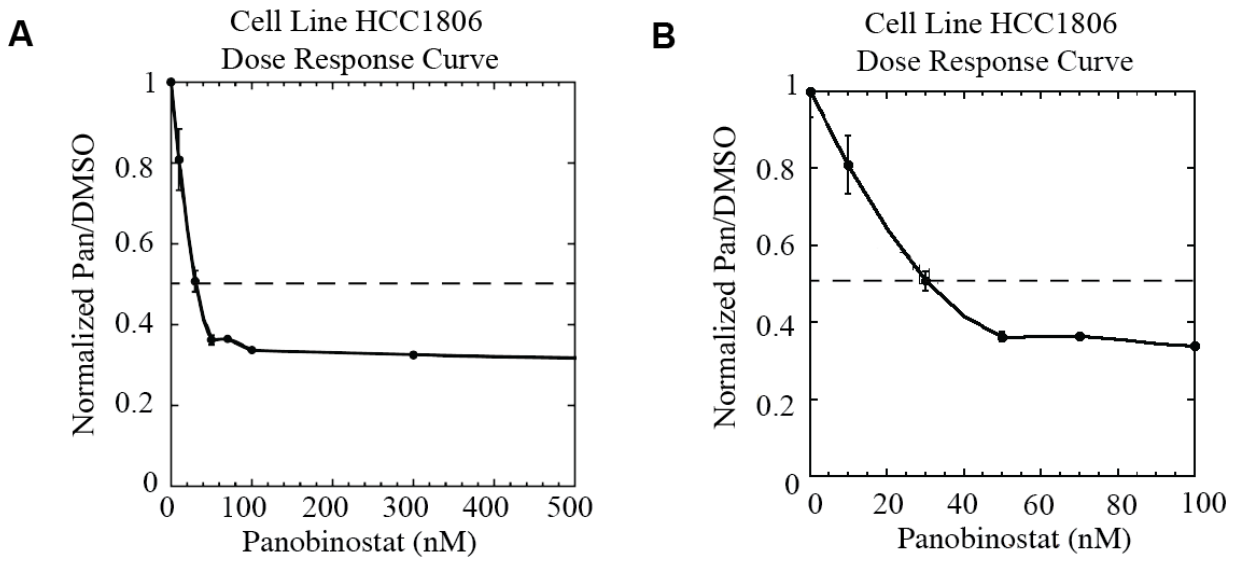


Figure 2. HCC1806 sensitivity to Panobinostat. After 72 hours of treatment with varying ranges of Panobinostat concentrations, the HCC1806 cells were subjected to the MTT assay. The MTT absorbance for each Panobinostat treatment was normalized to the absorbance for cells treated with DMSO. This ratio was then plotted against Panobinostat concentration. The dashed line represents cell viability of 50%. The corresponding x-axis value (Panobinostat concentration) represents the IC_{50}. This assay was done in triplicate, and error bars represent the standard deviation. A) The dose response curve for 0-500nM Panobinostat treatment. B) Zoomed-in dose response to 0-100nM Panobinostat treatment.

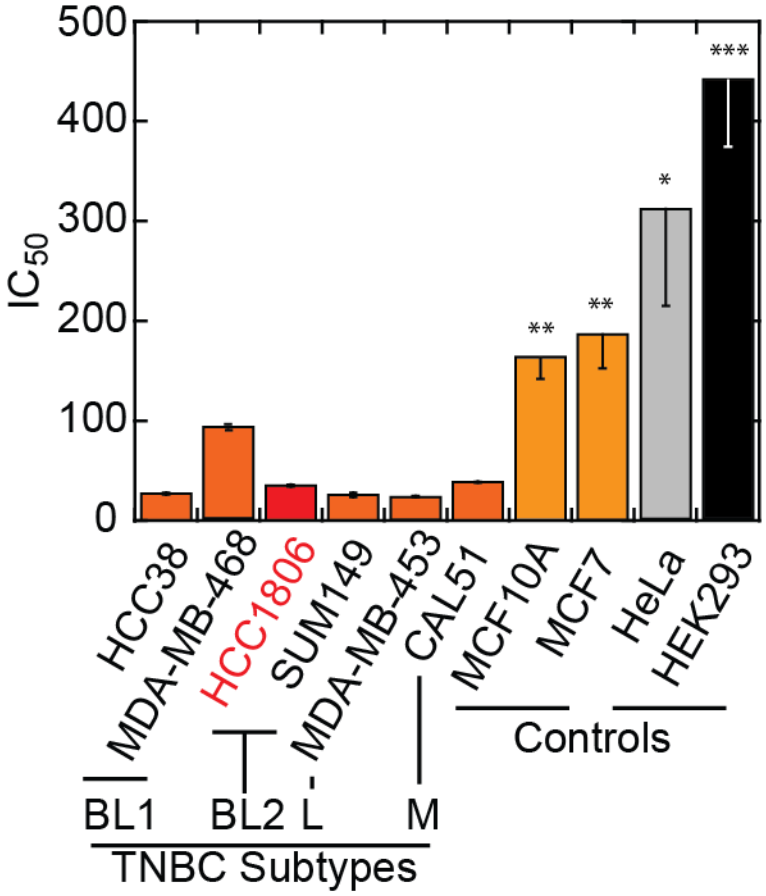


Figure 3. TNBC cells are more sensitive to Panobinostat than control.
The IC_{50} for each cell line was determined by the MTT assay and dose response curves as described in Figure 3. The HCC1806 bar in red is the cell line focused on through the remainder of this thesis. Orange bars represent the other TNBC cell lines tested. The yellow bars are other breast tissue derived cell lines and the grey and black bars represent other control lines. Each cell line was tested in triplicate, and the error bars represent the standard deviation. P-values were determined through unpaired student's t-test with unequal variance (*p-value<0.05, ** p-value <0.03, *** p-value <0.01, as compared to all TNBC cell lines), with respect to the IC_{50} of HCC1806.

dose dependent, so they were not able to quantitate IC_{50}s. In their publication, the Collins-Burows' lab only tested ER+ breast cancer cells as a control. I expanded on this by using a non-invasive breast-tissue derived cell line, MCF10A, and non-mammary tissue controls (HeLa and HEK293). The Collins-Burow group also only tested their control cells in 200nM of Panobinostat for 24 hours, and claimed that these cells were insensitive to the Panobinostat treatment. At these ranges, IC_{50}s could not be determined for control cells. I showed that the cells used for controls remained insensitive at longer hours and higher concentrations of Panobinostat, and that a wider variety of TNBC cells were sensitive under the same conditions.

Histone acetylation in response to HDACi

At the onset of this project, Panobinostat was being used to treat TNBC in Phase 1 and Phase 2 clinical trials, however the mechanism of action remained unestablished. After verifying that TNBC cells were sensitive to HDACi, I sought to determine if the bulk histone acetylation was increasing. However, since the discovery of HDAC proteins, it has also been found that HDACs can acetylate non-histone proteins, such tubulin and p53[24,25]. Mechanisms for HDACi induction of cell death have been proposed, such as interference with chaperone proteins, free radical generation and induction of DNA damage, and upregulation of some cell cycle regulators (such as p21)[26]. Thus, while my model focuses on the epigenetic basis for HDACi sensitivity, we cannot rule out other contributing factors. With the next aspect of my hypothesis, I tested whether HDACi treatment leads to an increase in histone acetylation. Immunoblot analysis (Figure 4) showed that histone acetylation did increase following Panobinostat treatment in TNBC cells. In each of the 3 replicates, the acetylation of histones increases over the course of 16 hours. Also, note that each lysine I tested increased in acetylation: "bulk acetylation" at

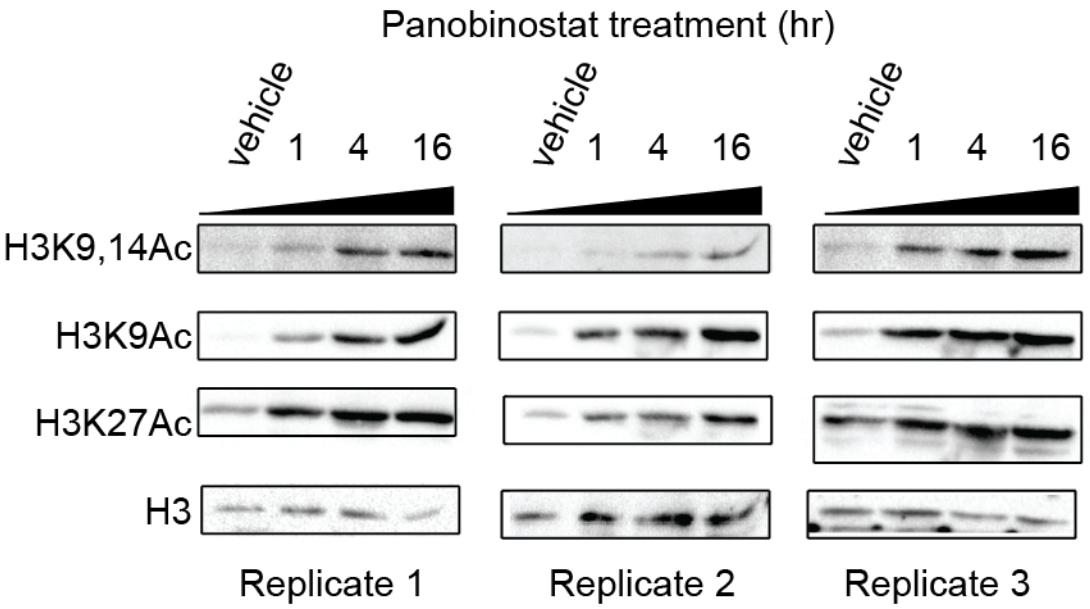


Figure 4. Histone acetylation following Panobinostat Treatment. HCC1806 cells were treated with 150nM Panobinostat and harvested at 1, 4, or 16 hours. The vehicle control is DMSO treated HCC1806 cells. The blots were probed with the antibodies indicated on the left, with H3 representing the loading control.

H3K9,14Ac, and lysine specific acetylation at H3K9 and H3K27. From this analysis, we can also see that acetylation begins to increase within 1 hour of treatment and continues to increase throughout the 16 hour time frame. This data allowed me to hone in on experimental conditions and continue on to the epigenetic and transcriptional profiling. Thus, to minimize the contribution of indirect effects, for the work in Chapter III I focused our epigenetic and transcriptional studies within the early 1-4 hour time frame.

In summary, I found that treating a diverse panel of TNBC cell lines with Panobinostat impaired their growth. A caveat to my viability assay is that it does not distinguish between cytostatic and cytotoxic growth defects. However, upon visual inspection of the treated cells, I found the confluency dramatically reduced below starting confluency (data not shown), which is consistent with at least some cytotoxic effect. Importantly, I showed that TNBC cell growth was selectively sensitive to HDAC inhibition, compared to other unrelated cell lines. Then, I established that histone acetylation increases in HCC1806 cells. Each of these observations is consistent with our hypothesis that HDACi inhibits cell growth that is associated with changes in histone acetylation, which will be tested in more detail in Chapter III.

CHAPTER III

EPIGENETIC AND TRANSCRIPTIONAL CHANGES ASSOCIATED WITH HDAC INIHIBITION

Introduction

Although there are many competing theories as to the mechanism behind cell death caused via HDAC inhibition, our model predicts that increased acetylation of enhancers and/or promoters of some TSGs leads to increased TSG expression, causing cell death.

To test our model, I first established the baseline chromatin (epigenetic) state of the HCC1806 cells. Chromatin states are defined by biologically meaningful combinations of chromatin marks linked both spatially and temporally[27]. These combinations of histone marks define certain regions of the genome with annotations such as enhancers, promoters and heterochromatin (Figure 5). To discover these states in the HCC1806 TNBC cell line I sought first to determine where each of the marks is found in the genome. I did this through ChIP-exo, a technique developed by the Pugh lab, focusing on the histone marks (Table 3) that the ENCODE project had previously published to discover chromatin states[28]. After extensive epigenetic studies, these marks are known to be associated with specific regulatory regions of the genome; I have indicated the relevant marks, their location in the genome, and the proposed effects on gene repression in Figure 5 and Table 3. For example, Pol II, H3K4me3, H2A.Z, H3K27me3, and H3K27Ac can all be found at promoter regions. Some of these marks (H3K27me3) are associated with repressed promoters while others (H3K27Ac) are associated with activated promoters. By examining where in the genome these marks are found, we can establish which regions of the genome are active and which are likely being actively repressed. The same

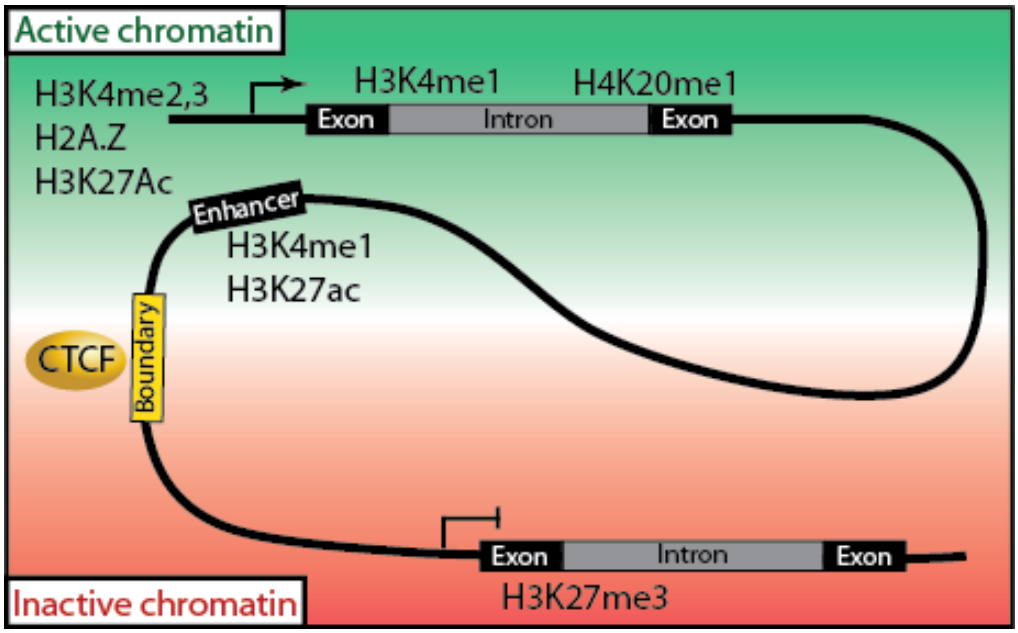


Figure 5. Overview of Histone marks. This cschematic indicates which regions of the genome where histone marks and other proteins which we tested are typically located: promoters, enhancers and transcription termination sequences. Active chromatin is shaded green and repressive chromatin is shaded red[27-29].

Genomic Location	Antibody Target	Associated with
Enhancer/Gene body	H3K4me1	Activation
Enhancer/Promoter	H3K27Ac	Activation
Promoter	H3K27me3	Repression
Promoter	H2A.Z	Activation
Promoter	H3K4me3	Activation
Gene Body	H4K20me1	Activation/Repression
Gene Body	H3K4me2	Activation
Insulator	CTCF	N/A
Promoter/Gene Body	Pol II	Activation

Table 3. Histone mark description. This table summarizes the information found in Figure 5. It describes each of the antibodies used for ChIP-exo, where in the genome these histone modifications are typically found, and whether these histone marks are known to be associated with gene activation or repression[27-29].

analysis can be performed on enhancer regions as well, using a different pattern of histone marks.

By combining chromatin immunoprecipitation assays with high throughput sequencing, ChIP-seq is a powerful tool for mapping the genome-wide locations of proteins. While ChIP-seq has been used to examine the DNA sequences for which transcription factors bind, it is also a powerful tool for examining the locations of different histone modifications throughout the genome[30]. The development of ChIP-exo has allowed improvement to resolution over the popular ChIP-seq technique. To improve resolution compared to ChIP-seq, ChIP-exo applies lambda exonuclease to immunoprecipitated chromatin, thereby identifying cross-linked points[31]. The added advantage over ChIP-seq is that ChIP-exo improves the factor mapping resolution from a few hundred base pairs to only a couple of base pairs while simultaneously increasing signal to noise[31]. I applied ChIP-exo to each of the nine targets listed in Table 3. As each cell type has a unique epigenetic landscape, these libraries will allow me to assess the epigenome of unperturbed HCC1806 cells and will also provide a reference map of chromatin states. Identifying chromatin states in unperturbed HCC1806 cells will allow me to better interpret the changes in H3K27Ac occupancy after HDACi treatment.

After examining where in the genome these marks are located, I wanted to evaluate regulatory regions in HCC1806 by analyzing the chromatin states. To learn chromatin states and the regulatory regions they define in HCC1806 cells, I used the computational program ChromHMM. The computational challenge for learning chromatin states from large data sets in multiple cell types is a bioinformatics' burden. ChromHMM was developed to automate the process of learning chromatin states and characterizing a proposed biological function for each state[27]. ChromHMM uses Hidden Markov Modeling (HMM) to analyze each individual mark by

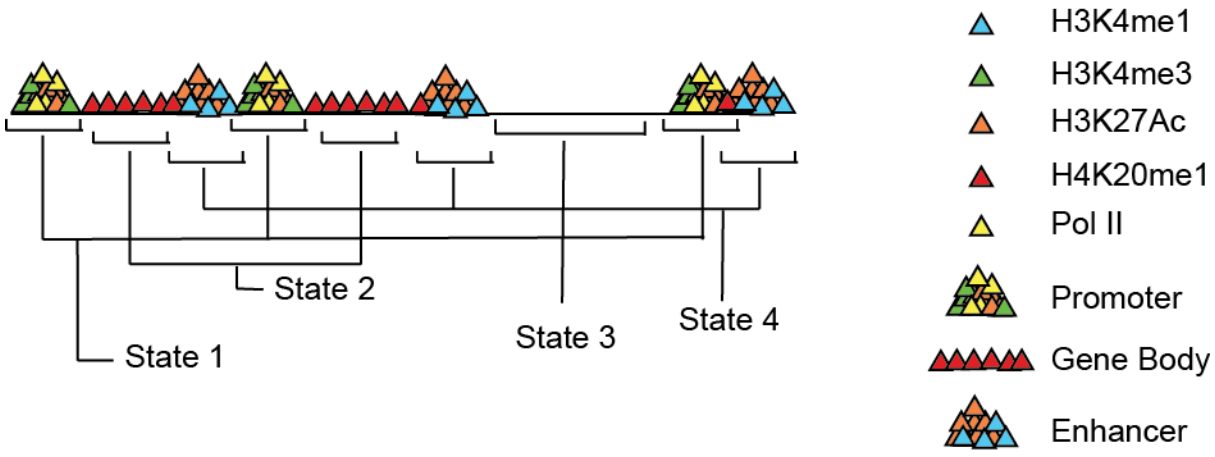


Figure 6. Model of ChromHMM pattern recognition. ChromHMM recognizes spatial patterns of chromatin marks across the genomes and classifies distinct patterns of histone marks and other factors. These states can then further be classified as different genomic regions passed on the absence or presence of particular marks. Each triangle is colored to represent a certain mark or factor as per the legend. The groupings of the states as promoters (Pol II, H3K4me3, and H3K27Ac), gene body (H4K20me1) and enhancer (H3K4me1 and H3K27Ac) based on the pattern of the particular histone marks.

examining the patterns that each mark forms and the relationship to other marks analyzed as seen in Figure 6. HMM is a common statistical tool used for temporal pattern recognition. The ChromHMM program utilizes HMM to process data on chromatin mark locations and outputs a number of states based on the probability that certain marks are found together. Each of these patterns can then be analyzed based on the presence or absence of certain marks to infer biological function, such promoters, enhancers, etc. The program uses the presence of different domains, such as proximity to TSS, exons, CpG islands, and Lamin B1 associated domains (LADs), to further annotate the different chromatin states[27]. ChromHMM learns patterns of recognition of different chromatin marks across the genome, and outputs them in such a way that you can annotate the regions as different chromatin states to learn the epigenetic profile of the cell-line.

After learning the baseline for each of the states, I sought to determine the effect of Panobinostat on a representative epigenetic mark (H3K27Ac) after 1 or 4 hours of treatment. I focused my efforts on changes to H3K27Ac because this mark is found at both promoters and enhancers, and is associated with gene activation[32,33]. Therefore, this mark is a prime example for our hypothesis that increased acetylation at both promoters and enhancers leads to increased transcription of tumor suppressor genes, leading to cell death. Following our ChromHMM analysis on the HCC1806 baseline chromatin state, we can examine the learned regions (promoters and enhancers) and how H3K27Ac increases or decreases at each of these regions following Panobinostat treatment. Since H3K27Ac is found in both promoters and enhancers and activates the corresponding genes, it is an ideal target from which to start our analysis.

To study the change in acetylation marks, we examined differential occupancy of H3K27Ac using the program HOMER. HOMER (Hypergeometric Optimization of Motif EnRichment) is a

suite of command line tools for next-gen sequencing analysis, including ChIP-seq (or –exo) and RNA-seq (or PRO-seq) analysis. For ChIP data, the software finds, annotates, quantifies, and applies statistical metrics to each peak[34]. The quantification of peaks from HOMER can then be used in a differential expression package, edgeR, to determine statistically significant (false discovery rate) changes in peak occupancy or expression. Using HOMER, we sought to uncover which genes and associated regions of the genome H2K27Ac occupancy is altered upon HDACi treatment.

The final aspect of our model I wanted to test was the connection between histone modifications and gene expression. To do this, I prepared PRO-seq libraries to collect data on *de novo* transcription before and after HDACi treatment. PRO-seq is a technique developed by the Lis lab, which modifies the classic nuclear run-on assay by adding a biotin-labeled nucleotide, allowing nascent transcripts to be isolated with streptavidin resin and mapped with next-generation sequencing[35]. From these PRO-seq libraries, I was able to determine which genes were being upregulated after HDACi treatment, with a focus on TSGs (genes that encode a protein that protects the cell from dysregulated growth). Next, I applied k-means clustering to identify groups of genes with similar patterns of expression in response to HDACi. These gene groups were further analyzed with Gene Ontology to reveal how biological pathways were affected. Then I aligned the transcriptionally upregulated genes with the differential H3K27Ac occupancy data to assess the link between promoter/enhancer acetylation and transcription in Panobinostat treated cells. Each of these steps and the results are detailed below in this final aim.

Materials and Methods

In accordance with ENCODE guidelines, we prepared and sequenced at least two biological replicates for each sample library[36].

ChIP-exo

In vivo Crosslinking, Cell Lysis and Chromatin Sonication

In vivo crosslinking, cell lysis and chromatin sonication were performed as described previously by the Odom laboratory[37]. Cells were cross-linked with 1% formaldehyde solution for 10 minutes at room temperature, and the reaction was quenched by adding 1/20 the volume of 2.5M glycine. Cells were then scraped from their plates and pelleted at 3000g for 4 min at 4°C. To obtain a nuclear lysate for ChIP that is free of most contaminating cytosolic proteins, the cell outer membrane is first lysed and solubilized, and then the nuclei are spun down. To lyse the cells, the pellet was resuspended in LB1 (50mM Hepes-KOH, pH 7.5, 140 mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) and rocked at 4°C for 10 minutes. The nuclei were then pelleted at 4°C, 2000g, for 4 minutes. The pellet was washed in 10mL of LB2 (10mM Tris-HCl, pH 8.0, 200mM NaCl, 1 mM EDTA, 0.5 mM EGTA) and rocked at 4°C for 5 minutes. The suspension was pelleted at 2000g, 4°C for 5 minutes. The pellet was then resuspended in 1.5 mL LB3 (10mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine) to lyse the nuclear membrane. This suspension was then transferred to 15mL polystyrene tubes and sonicated using the Diagenode Biorupter, on medium power, for 2 cycles lasting 10 minutes of alternating 30 seconds on and 30 seconds off (for 20 minutes total). After sonication, insoluble chromatin and nuclear debris were removed by centrifugation at 14,000rpm for 10 minutes at 4°C. The sonication conditions were

optimized to obtain DNA fragment lengths between 200-500bp (Appendix C-Supplemental Figure 2). This sonicated chromatin lysate is used in the ChIP assays detailed below.

Antibodies

The antibodies used for ChIP-exo and immunoblots include: Pol II (sc-899) from Santa Cruz Biotechnologies; H2A.Z (07-594), CTCF (07-729) and H3K27me3 (07-449) from Millipore; H3K27Ac (ab4729), H3K4me1 (ab8895), H3K4me2 (ab7766), H3K4me3 (ab8580), H3K9Ac (ab4441), and H4K20me1 (ab9051) from Abcam.

Chromatin ImmunoPrecipitation-exonuclease

A schematic of the ChIP-exo protocol described below can be seen in Appendix C, Supplemental Figure 3. Each ChIP assay was performed on 20-25 million cells using 5μg of antibody pre-coupled to 2.5μl of PrG MagSepharose Xtra beads (GE Healthcare) for 4 hours at 4°C. After an overnight chromatin immunoprecipitation on a rotator at 4°C, the beads were washed 7 times in 4°C lithium wash buffer (50mM HEPES, pH 7.6, 1mm EDTA, 0.7% Na-Deoxycholate, 1% NP-40, 0.5 M LiCl) and then twice in 4°C Tris-HCl, pH 7.5. ChIP-exo proceeded as described previously[38], with slight modifications noted below. The following incubations were performed on the PrG MagSepharose Xtra beads at 1400rpm in the thermomixer:

1. End polishing: 1mM ATP, 100μM dNTPs, 1x NEB buffer 2 (50mM NaCl, 10mM Tris-HCl, 10mM $MgCl_2$, 1mM DTT, pH 7.9), 3U T4 DNA polymerase, 5U Klenow DNA polymerase, 10U T4 PNK at 30°C for 30 min.
2. A-tailing: 100μM dATP, 1x NEB buffer 2, 5U Klenow 3'-5' exo-minus at 37°C for 30 min.
3. Ligation of ExA2-adapter: 1mM ATP, 1x NEB buffer-2, 30pmol ExA2 adapter (with index for each sample), 1000U T4 DNA ligase at 25°C for 2 hours.

4. Nick Repair: 150μM dNTP, 1x Phi29 reaction buffer (50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 10mM $(NH_4)_2SO_4$, 1mM DTT, pH7.5), 15U phi29 DNA polymerase at 30°C for 20 minutes.
5. Kinase reaction: 1mM ATP, 1x NEB buffer-2, 10U T4 PNK, for 20 minutes at 37°C.
6. *Lambda* exonuclease: 1x NEB *Lambda* exonuclease buffer (67 mM Glycine-HOH, 2.5mM $MgCl_2$, 50μg/mL BSA, pH 9.4) for 30 minutes at 37°C.
7. RecJf nuclease: 1x NEB buffer-2, 30U RecJf exonuclease for 30 minutes at 37°C.
8. Each previous step (1-7) was followed with 4 wash steps (2x RIPA-Lithium buffer washes, and 2x Tris-HCl pH7.5 washes).
9. Elution and cross-link reversal: beads incubated overnight at 65°C with 20μg Proteinase K in 200μl ChIP elution buffer (50mM Tris-HCl pH 8.0, 10mM EDTA, 1% SDS). After overnight incubation, the supernatant was transferred to a new tube with 200μl TE buffer (10mM Tris, 1mM EDTA pH 7.4).
10. The DNA was purified using phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation. The DNA pellet was resuspended in 20μl ddH_2O.
11. P7 primer extension: The DNA is denatured at 95°C for 5 minutes and then added to 5 pmol P7 primer and 1x Phi29 reaction buffer. This mixture was incubated for 5 minutes at 65°C and 2 minutes at 30°C. 10U of Phi29 polymerase was added to the reaction. This reaction was incubated in thermocycler for 20 minutes at 30°C and then 10 minutes at 65°C.
12. A-tailing: 0.1mM dATP, 1x NEB buffer-2, and 5U Klenow 3'-5' exo minus were added to the previous reaction, which was then incubated at 37°C for 30 minutes and then 75°C for 20 minutes.

13. P5 adapter ligation: 15pmol ExA1-58/1, 1x NEB ligase buffer (50mM Tris-HCl, 10mM $MgCl_2$, 1mM ATP, 10mM DTT at pH 7.5), and 1000U of T4 DNA ligase added to the previous reaction mix and incubated at 25°C for 2 hours.
14. DNA was then purified using AMPure XP beads (1.8 times volume) from Agencourt for 3 minutes. The beads were placed on a magnetic rack and the supernatant was discarded. The beads were then washed three times in 70% ethanol. The DNA was then eluted in 10mM Tris-HCl pH 7.5.
15. PCR amplification: The eluted DNA was PCR amplified using 0.5μM primers P1.3 and P2.1 in 1xQ5 reaction buffer (25mM TAPS-HCl pH 9.3, 50mM KCl, 2mM $MgCl_2$, 1mM β-mercaptoethanol), 200μM of each dNTPs, and 1U of Q5 Hot Start DNA polymerase. PCR amplification was then run in the thermocycler with the following PCR program (Table 4).
16. Gel-size purification: Excise PCR product between 200-500bp for each library from a 1.5% agarose gel (Appendix C-Supplemental Figure 4). The product was then eluted from the gel using QIAquick Gel Extraction Kit (Qiagen) and eluted in 50μl of elution buffer (EB).
17. Each library was quantified using a Qiagen Qbit fluorescence-based nucleic acid quantification system. Prior to sequencing, the Vanderbilt VANTAGE Sequencing Core conducted two quality control assays to verify DNA size (Agilent Bioanalyzer) and to precisely measure DNA concentration (qPCR). After libraries passed quality control, library DNA samples were sequenced on Illumina HiSeq 2500 or HiSeq3000/HiSeq4000 systems following manufacturer's protocol to obtain ~20M or more reads per library.

Table 4 **PCR program**		
Temp (°C)	**Time**	**Cycles**
98	30 sec	1
98	10 sec	21
52	30 sec	
72	20 sec	
72	2 min	1
4	pause	hold

Oligonucleotides

Oligonucleotides used in the ChIP-exo protocol are described in Table 5.

Oligonucleotide notes

ExA2 has a 5' phosphate required for lambda exonuclease digestion and a 3' phosphothiorate (designated by the "^{S}") to prevent contaminating nuclease digestion of the sample DNA 3'-end. ExA1-58 also has a phosphothiorate before the terminal T, so as to prevent nuclease digestion. The ExA2 oligo needs to be annealed to the ExA2-33, prior to the ExA2 ligation in ChIP-exo. The ExA1-58 oligonucleotide also needs to be annealed to ExA1-13 oligonucleotide to create the double-stranded ExA1-58/13 adapter. The annealing is done by combining each oligonucleotide at a final 15μM concentration in 100mM Tris-HCl pH 7.5 and 50mM NaCl. The slow adapter annealing then occurs during the following thermocycler program: 95°C for 5 minutes, 72°C for 5 minutes, 65°C for 2.5 minutes, 60°C for 2.5 minutes, 55°C for 1.5 minutes, 50°C for 1.5 minutes, 45°C for 1.5 minutes, 40°C for 1.5 minutes, 30°C for 3 minutes, 20°C for 3 minutes, 10°C for 3 minutes, and held at 4°C. The oligonucleotides were then aliquoted and stored in the -80°C, until use.

The barcode indexes used for Chip-exo (ExA2) and PRO-seq (RPI-index) are listed in the Table 6.

PRO-seq

Nuclei Isolation

Nuclear isolation was performed similarly to the method described previously,[39] with the following modifications. First, cells were treated for 0, 1 and 4 hours with 150nM Panobinostat

Table 5 ChIP-exo oligonucleotide sequences and purpose		
Name	**Sequence**	Purpose
ExA2	5'/Phos/CAAGCAGAAGACGGCATACGAGAT**INDEX**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC^{S}T3'	Ligates to ChIP enriched DNA to add barcode sequence to each sample, which allows sample pooling and multiplexed sequencing
ExA2-33	5'GATCGGAAGAGCACACGTCTGACTCCAGTCAC-3'	Required for Illumina sequencing chemistry
ExA1-58	5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC^{S}T-3'	Required for Illumina sequencing chemistry
ExA1-13	5' GATCGGAAGAGCG -3'	Required for Illumina sequencing chemistry
P7 primer	5' CTGGAGTTCAGACGT -3'	Used to fill-in ExA2 adapter complement
P2.1	5' AATGATACGGCGACCACC-3'	PCR amplification primer-anneals to ExA2 sequence
P1.3	5' CAAGCAGAAGACGGCATACGAG-3'	PCR amplification primer-anneals to ExA1 sequence

TABLE 6: ChIP-exo and PRO-seq Barcode Indexes

	1	2	3	4	5	6
A	I-01 ACATCGCGTGAT	I-09 AAGCTACTGATC	I-17 TGCTTTATGTTT	I-25 GTTTGTCCTATT	I-33 CTATCTTCTTCT	I-41 GCTCATATAGAT
B	I-02 GCCTAAACATCG	I-10 GTAGCCAAGCTA	I-18 GCATTTTGCTTT	I-26 AGATGTGTTTGT	I-34 GATGCTCTATCT	I-42 AGGAATGCTCAT
C	I-03 TGGTCAGCCTAA	I-11 TACAAGGTAGCC	I-19 TTTGTTGCATTT	I-27 CTGGGTAGATGT	I-35 AGCGCTGATGCT	I-43 CTTTTGAGGAAT
D	I-04 CACTGTTGGTCA	I-12 AGTCTTTACAAG	I-20 GGGGTTTTTGTT	I-28 GCCGGTCTGGGT	I-36 CGGCCTAGCGCT	I-44 TAGTTGCTTTTG
E	I-05 ATTGGCCACTGT	I-13 GCGTGGAGTCTT	I-21 CAAGTTGGGGTT	I-29 TATCGTGCCGGT	I-37 AATTATCGGCCT	I-45 CCGGTGTAGTTG
F	I-06 GATCTGATTGGC	I-14 TATATCGCGTGG	I-22 TCGCTTCAAGTT	I-30 GAGAGTTATCGT	I-38 CCGTATAATTAT	I-46 ATCGTGCCGGTG
G	I-07 TCAAGTGATCTG	I-15 GCTGTATATATC	I-23 GTCCTTTCGCTT	I-31 ATCAGTGAGAGT	I-39 TAGGATCCGTAT	I-47 TGAGTGATCGTG
H	I-08 CTGATCTCAAGT	I-16 ATGTTTGCTGTA	I-24 CCTATTGTCCTT	I-32 TCTTCTATCAGT	I-40 ATAGATTAGGAT	I-48 CGCCTGTGAGTG

or DMSO and then nuclei were isolated from $3x10^7$ cells. Four $15cm^2$ plates of 85% confluent HCC1806 cells were washed twice with cold PBS. Then, cold swelling buffer (10mM HEPES, pH 7.9, 340 mM Sucrose, 10nM KCl, 1.5mM $MgCl_2$, 5mM DTT, Roche protease inhibitors cocktail) was added to each plate and incubated for 10 minutes. Cells were collected from the plate with rubber cell scrapers, pooled in to a 50mL tubes and cells were pelleted in a swinging bucket centrifuge at 1000g, 4°C for 5 minutes. Cells were then resuspended in lysis buffer (swelling buffer with 10%glycerol and 0.2% Triton X-100) and incubated on ice for 5 minutes. Then, the cells were dounce-homogenized 50 times to facilitate cell lysis. Lysis was verified by use of the vital stain, trypan blue. The nuclei were transferred to 15 mL tubes and then pelleted in swinging bucket rotor at 1200g, 4°C for 5 minutes. Nuclei were washed and pelleted in lysis buffer and then washed and pelleted once in glycerol storage buffer (50mM Tris-HCl pH 8.3, 40% glycerol, 0.1 mM EDTA, 5 mM MgCl2, 5mM DTT, protease inhibitors cocktail (Roche)). Then, the nuclei were transferred to a 1.5 mL tube, resuspended in glycerol storage buffer at a concentration of 2.5-3.0 107/100µl of storage buffer and stored in -80°C until ready for use.

Nuclear run-on reaction

This procedure is summarized in Supplemental Figure 5, Appendix C. It is important to note that the Lis lab showed previously that no significant transcription occurs during the isolation protocol[39]. To begin the nuclear run-on reaction (NRO), $3x10^7$ nuclei were added to 2x NRO mixture (10 mM Tris-HCl pH 8.0, 300mM KCl, 1% Sarkosyl, 5 mM $MgCl_2$, 1mM DTT, 0.8 u/µl RNase inhibitor from Ambion) with 37.5µM ATP, UTP, and GTP, 500nM biotin-11-CTP (Roche), and the run-on proceded for 3 min at 30°C. Sarkosyl prevents transcription initiation by preventing reassociation of RNA polymerase to additional genomic locations (after nuclei

isolation)[40]. Therefore only genes that have engaged polymerase at the time of nuclear isolation will produce labeled transcripts. Importantly, transcripts produced prior to the addition of biotin-nucleotides will not be included in the final library pool. Finally the sub-μM concentration of biotin-CTP limits elongation to only a few base pairs. This is how this procedure is able to measure nascent transcription. Total RNA was extracted with Trizol LS (Invitrogen) for 20 minutes and then ethanol precipitated with 3x volume of 100% EtOH. The pellet was then washed in 75% EtOH, and resuspended in DEPC-treated H_2O.

Base Hydrolysis

The extracted RNA was subjected to base hydrolysis for 10 min on ice. This was done by adding NaOH to the RNA for a final concentration of 0.2 N for 10 minutes. The reaction was quenched by adding 25μl of 1M Tris-HCl pH 6.8. The reaction was then run through a P-30 column (Bio-Rad) according to the manufacturer's instructions to remove salts, nucleotides and small molecules from the reaction.

Purification of biotin-CTP incorporated nascent RNA

Homo-tetramers of Straptavidin have an extremely high affinity for biotin with a K_d on the order of $\sim 10^{-14}$ mol/L[41]. We exploited this to pull out the nascent transcripts, which are tagged with biotin, from the reaction mixture. To begin, 30μl of Streptavadin M280 (Invitrogen) beads were washed once in a solution containing 0.1 N NaOH and 50mM NaCl. Then the beads were washed twice in 100mM NaCl and resuspended in binding buffer (10mM Tris-HCl pH 7.4, 300mM NaCl, 0.1% Triton X-100). Total purified nuclear RNA was incubated with the beads for 20 minutes at room temperature. The beads were then washed once in high salt wash (50mM Tris-HCl pH 7.4, 2M NaCl, 0.5% Triton X-100), once in binding buffer, and once in no salt

wash (5mM Tris-HCl pH 7.4, 0.1% Triton X-100). The bound RNA was then extracted from the beads twice with Trizol for 10 minutes. The extraction was pooled and then RNA was EtOH precipitated as described previously.

3' RNA adapter ligation

The purpose of this step is to attach the 3' adapter to the nascent RNA. The nascent RNA was resuspended in DEPC-H_2O with 20 pmol of the Rev3 adapter. To remove any secondary RNA structures, the mixture was heated to 65°C for 1 min. The reaction was then diluted to 10pmol Rev3 with 1x T4 RNA ligase buffer (50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 1mM DTT), 2U/μL RNase inhibitor, 1mM ATP and 10U T4 RNA ligase I. This reaction proceeded for 6 hours at 20°C and was then kept at 4°C until the following morning.

2nd Purification of biotin-CTP incorporated nascent RNA

This Streptavidin bead binding and RNA extraction proceeded as described in the "Purification of biotin-CTP incorporated nascent RNA" step above.

5' cap and triphosphate repair

In order to prepare the 5' end for subsequent ligations, the 5' cap must be removed. Tobacco Acid Pyrophosphatase (TAP, Epicenter) removes the 5' cap from the RNA transcripts. The RNA was incubated in 1x TAP buffer (50mM sodium acetate pH 6.0, 1mM EDTA, 0.1% β-mercaptoethanol, and 0.01% Triton X-100), 2.5U TAP, and 1U/μl RNase inhibitor for 2 hours at 37°C. The RNA was then extracted once with Trizol and ethanol precipitated as described previously.

Then, Polynucleotide Kinase (PNK, NEB) was used to ensure a 5'-phosphate is present for subsequent ligation. The precipitated RNA is resuspended in DEPC-H_2O and then added to PNK mix with a final concentration of 1x PNK buffer (70mM Tris-HCl pH 7.6, 10mM $MgCl_2$, 5mM DTT), 1mM ATP, 1U/μL RNase inhibitor, and 5U T4 PNK. The reaction proceeded at 37°C for 1 hour. Then, RNA was extracted again by Trizol and EtOH precipitated.

5' RNA adapter ligation

After the 5' end of the RNA is repaired, the 5' adapter can be ligated to the RNA. This was done similar to the 3' adapter ligation (but with a different adapter) as follows: The nascent RNA was again resuspended in DEPC-H_2O with 20 pmol of VRA5 adapter. This suspension was heated at 65°C for 1 minute to remove any secondary RNA structures and then placed on ice. Then, the reaction was diluted to 10 pmol VRA5 adapter with 1x T4 RNA ligase buffer, 2U/μL RNase inhibitor, 1mM ATP and 10U T4 RNA ligase I. This reaction proceeded for 6 hours at 20°C and was then kept at 4°C overnight.

3rd Purification of biotin-CTP incorporated nascent RNA

This Streptavidin bead binding and RNA extraction proceeded as described in the "Purification of biotin-CTP incorporated nascent RNA" step above.

Reverse transcription

The RNA must be converted into a cDNA library prior to Illumina sequencing. This was done by reverse transcription of purified nascent RNA. The RP1 primer was used for reverse transcription. The RNA was first resuspended in 50pmol of RP1 primer and heated to 65°C for 5 minutes. Then, the reaction buffer was added to the mixture for a final concentration of 1x first

strand synthesis buffer (50mM Tris-HCl pH 8.3, 75mM KCl, 3mM $MgCl_2$), 10 mM DTT, 625pmol each dNTP and 1U/µl RNase inhibitor. After incubating for 3 minutes at 48°C, 200U of Superscript III Reverse Transcriptase Invitrogen) was added to the reaction. This reaction was then incubated for 20 minutes at 44°C and then 52°C for 45 minutes, and then held at 4°C.

Test amplification

This step is used to optimize the PCR cycle number for each library. First, 8µl of H_2O was added to the 10µl of RNA from the reverse transcriptase reaction. Then, 1µL (or 1/18 of the total RNA) was diluted 8 fold. Then, 7 more four-fold serial dilutions were made by taking 2µL of the previous dilution and adding it to 6µL of H_2O. PCR was done in the following reaction mix: 1x HF buffer (25 mM TAPS-HCl pH 9.3, 50mM KCl, 1.5mM MgCl, 1mM β-mercaptoethanol), 1M Betaine, 0.25µM primer Gx2short, 0.25µM RP1 short primer, 0.25mM each dNTP, and 0.4U of Phusion High-Fidelity DNA polymerase. The reaction ran for 35 cycles of 95°C denaturation, 56°C annealing, and 72°C extension for 30 seconds at each step. The results were analyzed by electrophoresis on a 2% agarose gel. The library should run as a DNA smear between 60bp-300bp (Appendix C, Supplemental Figure 6). The dilution that has the smear closest to this range was used to determine the cycle numbers for the full scale amplification.

Full scale amplification PCR

The remainder of the reverse transcriptase product was used for the full scale amplification. The same reaction mixture was used as in the test amplification, however, 12.5 pmol of RP1 and 12.5 pmol of RPI-index primers were used in lieu of Gx2 short and RP1 short. The libraries were then amplified for 25-29 cycles depending on the results from the test amplification. The PCR product

was run on a 2% agarose gel and size excised for 200-400bp smear (Appendix C, Supplemental Figure 7). The DNA was then extracted from the gel using QIAquick Gel Extraction Kit (Qiagen) and eluted in 50μl of elution buffer (EB).

Each library was quantified using a Qbit and sequenced on Illumina HiSeq 2500 following manufacturer's protocol.

Oligonucleotides

The oligonucleotides used in the PRO-seq protocol are described in Table 7. The indexes are described in Table 6.

Data analysis

ChIP-exo sequence read mapping and data pre-processing

All sample libraries were sequenced at the Vanderbilt VANTAGE core as single-read and 50 nucleotides in length. According to ENCODE guidelines[36], we targeted 20 million or more reads per data set. Reads that passed the base-calling quality filter were mapped to the hg19 human reference genome with BWA[42] using default parameters. We filtered out reads mapping to multiple locations, and retained only uniquely mapped reads with SAMtools v0.1.18[43]. This step provides an initial metric of library quality by looking at uniquely mapped read ratios (Appendix C-Supplemental Table 1). SAMtools was also used to sort the BAM files and create index files to view the aligned reads in the Interactive Genomics Viewer (IGV)[44]. We converted the BAM files to BED files using BEDtools v2.12.0[45]. At this stage, the chromosomes were listed as numerical values. The tools used later in the pipeline do not recognize chromosomes in this format, so we reformatted the file to change "1" to "chr1", "2" to "chr2" and so forth using

Table 7 PRO-seq Oligonucleotide Sequences and Purpose		
Name	**Sequence**	**Purpose**
Rev3	5'phosphate-GAUCGUCGGACUGUAGAACUCUGAAC-/3InvdT/	3’ RNA Adapter, template for Reverse transcription primer.
VRA5	5'-CCUUGGCACCCGAGAAUUCCA	5’ RNA Adapter, recognized by PCR primers
RP1	5’-AATGATACGGCGACCACCGAGATCTACAC GTTCAGAGTTCTACAGTCCGA	Required for Illumina sequencing chemistry
Gx2 short	5’-CAGAGTTCTACAGTCCGA	Test PCR primer
RP1 short	5’-CCTTGGCACCCGAGAATTCCA	Test PCR primer
RPI-index	5'-CAAGCAGAAGACGGCATACGAG AT **INDEX** GTGACTGGAGTT CCTTGGCACCCGAGAATTCCA	Adds barcode sequence to each sample, which allows for sample pooling when sequencing

UNIX command line:

sed '/^[^#]/s/^/chr/' inputfile.bed > inputfile_chr.bed

These files also contained short chromosome contigs that have little annotation and disrupt downstream analysis, such as chrGL000191.1, that needed to be removed. These chromosome contigs were removed with the UNIX command line:

sed '/chrGL*/d' inputfile_chr.bed > inputfile_chr_sed.bed

These large BED file sizes were then converted to a more concise Genetrack Index (*.idx) file format using the tabs2genetrack.py python script. Each individual biological replicate was then subjected to peak calling using GeneTrack[46] with a sigma of 20 (smoothes reads to call peaks) and an exclusion zone of 40 around each peak (prevents other peaks from being called in this region). Peaks were then paired if they existed on opposite strands and were offset by <80 bp (offsets in the downstream or 3' direction) and were binned with a bin size of 3. Peaks in the ENCODE backlist region, which represents non-specific sequencing and/or mapping artifacts found in nearly all data sets, were removed[36]. Biological replicate reproducibility was assessed by first matching peak-pairs across replicates and then fitting a Pearson-correlation to the resulting scatter plot, with each axis representing a biological replicate. Biological replicates that are reproducible have a strong positive correlation (R-value > 0.5) (see Appendix C-Supplemental Figure 8). In nearly all instances, since our replicates met this criteria, we merged the reads from each individual biological replicate data set, and ran the merged BAM files through the same pipeline described above to call peaks.

PRO-seq data processing

The PRO-seq data processing was performed in the same manner as ChIP-exo up to and including the peak-calling step. The ENCODE blacklist peaks[36] were removed from the peak calling outputs. Replicate correlation graphs can be seen in Appendix C Supplemental Figure 9.

ChIP-exo data analysis

Hidden Markov Modeling of ChIP-exo data

To determine the chromatin states and their locations across the HCC1806 genome, the merged files for each of the histone marks, CTCF, and Pol II were subjected to multivariate Hidden Markov Modeling with ChromHMM[27]. ChromHMM is an automated program that outputs chromatin state annotation using combinations of chromatin modification patterns. It models the chromatin states through observing combinations of chromatin marks using product independent Bernoulli random variables. This program creates robust learning of complete patterns of many chromatin modificaitons[27].

To conduct the ChromHMM analysis, first, BED files for all 6 histone modifications (H3K27me3, H4K20me1, H3K27Ac, H3K4me1, H3K4me2, H3K4me3) one histone variant (H2A.Z), one chromatin insulator (CTCF) and Pol II (with the "chr" added and the non-normal chromosomes removed) were converted to binary data using the BinarizeBed function, meaning the data is converted to the absence or presence of the mark across the genome. Then, the binary data can be piped into the LearnModel function, which will learn a certain number of chromatin states, based on the input. To identify the minimum number of non-redundant chromatin states, we processed the learn model with 4, 8, 12, 16, 20, and 24 number states. The LearnModel will

produce segmentation of chromatin states based on the input "numstates" option. The program generates both data files and graphical reports on: 1) emission states, 2) emission order/ transition states, 3) emission states' biological category, and 4) emission states' distances to TSS or TTS. The model with 12 chromatin states was chosen to proceed with further analysis, because it was the minimum state number that stratified the data without redundant states.

Analyzing differential H3K27Ac ChIP-exo occupancy with HOMER

HOMER is an all inclusive software suite that can analyze ChIP-seq and ChIP-exo peaks. It is able to find and annotate peaks based on read densities. These peak quantifications can then be used for RPKM analysis or differential occupancy analysis[34]. Here, we use HOMER to find the RPKM for our Pol II libraries and to quantify peaks in our H3K27Ac libraries to prepare them for differential occupancy analysis with edgeR. The pre-processed replicates for PRO-seq were not merged, as HOMER accepts replicates in the input. HOMER was run on the Pol II, PRO-seq, and H3K27Ac data sets. To generate correctly formatted input files for HOMER, I converted our BED files to BAM files using BEDtools bedtoBam function. This eliminated the possibility that our BED files would not be read correctly. The BAM files were then used to create Tag Directories with the makeTagDirectory function, which transforms the data into a platform independent format and performs quality control checks (such as checking GC content). We then used the findPeaks function to find enriched peaks within our data. For the Pol II data set, we used the –groseq function within the findPeaks program, because this would take into account variable size peaks, unlike the –factor function created for transcription factor peaks. The RPKM was determined with the analyzeRepeats.pl script, with the default normalization and the –rpkm function. The default normalizes each of the total reads to 10 million reads. The –rpkm function reports the normalized values as reads per kilobase per million mapped reads.

For the H2K27Ac data analysis, we used the –histone option, which calls peaks for broad regions of enrichment and variable width peaks. Then, we used the annotatePeaks.pl program with no normalization (-noadj option) to print the reads from each time point (DMSO, 1hr Pan and 4hr Pan) for each gene. A directory with the merged DMSO H3K27Ac data was used to create a tag directory and this tag directory was run through findPeaks as a control. This findPeaks output file was used as the control for annotatePeaks.pl, so only peaks that were called in the DMSO file were used to compare peaks in each of the Panobinostat time points.

edgeR

HOMER does not include its own package to find statistically significant differentially expressed/occupied genes. However, it does have a package that pipes the data obtained from HOMER's analyzeRepeats package out to the R Bioconductor package edgeR[47] for differential expression analysis. The edgeR package is able to compute the differential expression through normalization of the total reads per transcript using an overdispersed Poisson statistical model. The statistical significance is then computed using the Fisher's exact test[47]. To find differential expression from our data, I started with the output .txt file from annotatePeaks.pl and ran it through the getDiffExpression.pl, which used edgeR to calculate the differential expression of genes. I filtered out the genes that had an FDR greater than 5%.

ChIP-exo heatmaps using ChAsE

ChAsE is a Java based computational tool that provides a graphical interface for analysis of our epigenetic data using interactive heatmaps and plots[48]. Since ChAsE only accepts sequence data files in the bigWig format, we converted our BAM files to bigWig format. The other input file required by ChAsE is a GFF format reference file containing annotation intervals ±2 kilobases

from TSSs or TTSs) to display. ChAsE then aligns the sequencing reads in bigWig files to the reference intervals. We then sorted the output heatmaps based on descending maximum peak and maximum peak position. These were created for each of our ChIP libraries.

PRO-seq data analysis

HOMER

HOMER is also an all inclusive software suite for analyzing transcript abundance[34]. Here, I used HOMER to quantify my PRO-seq libraries, and to prepare the data for further differential expression. First, I converted each PRO-seq BED file replicate to a BAM file using BEDtools bedtoBam function. Tag directories were created from the BAM files with the makeTagDirectory function. At this stage, I used the option –flip to flip the strands in the PRO-seq data. The PRO-seq data is strand specific, but in the BWA mapping phase the reads were mapped to the opposite strand. Using –flip corrected this problem. After the Tag Directories were made, I used the findPeaks program to find enriched transcripts. This program can be used for both transcripts and finding peaks in ChIP data. To specify transcripts, the –groseq function was used. The –groseq function attempts *de novo* transcript identification from strand specific nascent RNA sequencing reads. HOMER has built in functions to quantify RNA expression in genes. I used HOMER's analyzeRepeats.pl program to create a gene expression matrix from each of our Tag Directories and then quantify the PRO-seq data. I used the "rna" option to specify which transcripts to analyze. This option forces the quantification from the RefSeq annotation for the specified genome—in this case hg19. I also used the option "–count genes" to have the program count tags throughout the whole gene body and not just exons. This is because

PRO-seq produces nascent transcripts, so reads can map to any part of the gene, not just exons (like post-processing RNA-seq libraries). I also wanted to quantify transcription in the positive direction of gene orientation, so I used the function "-strand +", which will measure reads on the positive strand (relative to the gene orientation). The analyzeRepeats.pl program allows us to calculate the RPKM through its normalization functions. So, I used the default normalization function which normalizes all of the total reads of each library to 10 million reads and the -rpkm function. The RPKM for each individual biological replicate was plotted against its corresponding replicate to determine positive correlation (see Supplemental Figure 9 in Appendix C).

edgeR

First, I ran analyzeRepeats.pl with all the same options as above, however instead of normalizing and finding the RPKM, I forced no normalization with -noadj and used -condenseGenes to report only one isoform of a gene per locus. I used the output .txt file from running analyzeRepeats.pl and ran it through the getDiffExpression.pl, which used edgeR to calculate the differential expression of genes. I filtered out the genes that had a FDR greater than 5%, to keep only statistically significant changes in gene expression.

Cluster analysis of differentially expressed genes

I used the open source Cluster v3.0 software[49] to cluster the gene expression data by K-means clustering on genes using the Euclidean distance similarity metric. A variety of cluster numbers were run (from k=4-12) and the number of clusters was chosen based on the minimum number with the greatest separation of non-redundant clusters. Clustering was done on 1h Panobinostat over DMSO and 4h Panobinostat over DMSO as one data set, and 1h Panobinostat over DMSO

with 4h Panobinostat over 1h Panobinostat as another data set. This second data set was included to show if genes continued to increase in expression as time went on or whether most genes increased in expression drastically at the onset of Panobinostat treatment and plateaued in expression by 4 hours. Six clusters was deemed the best k value for both of the samples, and the 6 clusters were used for all further analysis. Java TreeView[50] was used to view the clusters and create heatmap plots.

Gene ontology and pathway analysis of cluster groups

Gene ontology is the bioinformatics study of the relationships of a group of gene products enriched within a dataset. The list of genes gathered from each cluster above was piped into DAVID 6.7[51]. This program reported clusters of gene ontologies as well as KEGG pathway analyses, based on DAVID defaults. We then collected the gene ontologies that had an enrichment factor greater than 2 and a p-value less than $1.0x10^{-4}$. The top six factors (or the gene ontologies with the smallest p-values) were reported.

Results and Discussion

In this aim, I sought to define the chromatin state of HCC1806 cells. I also sought to examine how a specific chromatin state was altered following Panobinostat treatment by looking at the changes to H3K27Ac occupancy as a proxy for promoter/enhancer chromatin states. Finally, I compared these changes to the changes in transcription following HDACi treatment.

HCC1806 Chromatin State

Determining the chromatin states of different cell types probes the understanding of the mechanisms of human development and the etiology of many human diseases. By examining and predicting regulatory regions, we can develop a better understanding of the disease state.

In order to look at the chromatin state of HCC1806, I performed ChIP-exo on multiple histone marks, CTCF, H2A.Z and Pol II (Figure 5 and Table 3). Some of these histone marks have been well characterized and correlate with specific regions of the genome, such as TSS or enhancers[29]. I established HCC1806's baseline distribution of chromatin states when treated with DMSO vehicle using HMM so that I could compare drug-induced changes to the baseline state. The results of these ChIP experiments are summarized in Figures 6-8.

ChAsE heatmaps were generated for each histone mark ChIP-exo library in DMSO HCC1806 cells (Figure 6). The heat maps represent read density around the promoter (TSS) and TTS. As seen in the model of chromatin marks (Figure 5), H3K4me3, H3K4me2, H2A.Z and H3K27Ac are typically found at the promoters[28,29,36,52,53]. This was verified by our ChAsE heatmaps. In Figure 6A, the columns are centered around the TSS. Pol II, H3K4me3, H3K4me2, H2A.Z and H3K27Ac all have a high density of reads around the TSS (as seen by red color in the heatmaps). The rows are sorted based on Pol II density, so it is interesting to see that the genes with high densities of Pol II at the TSS also have corresponding high densities at the promoter in all of the promoter marking histone marks (H3K4me3, H3K4me2, H2A.Z and H3K27Ac). Meanwhile, H3K27me3, is also commonly associated with promoters, but it is associated with repressed regions[28,36]. So, it makes sense that H3K27me3 has very low density of reads around TSS that have high Pol II density with higher H3K27me3 density occurring at genes with low or no Pol II.

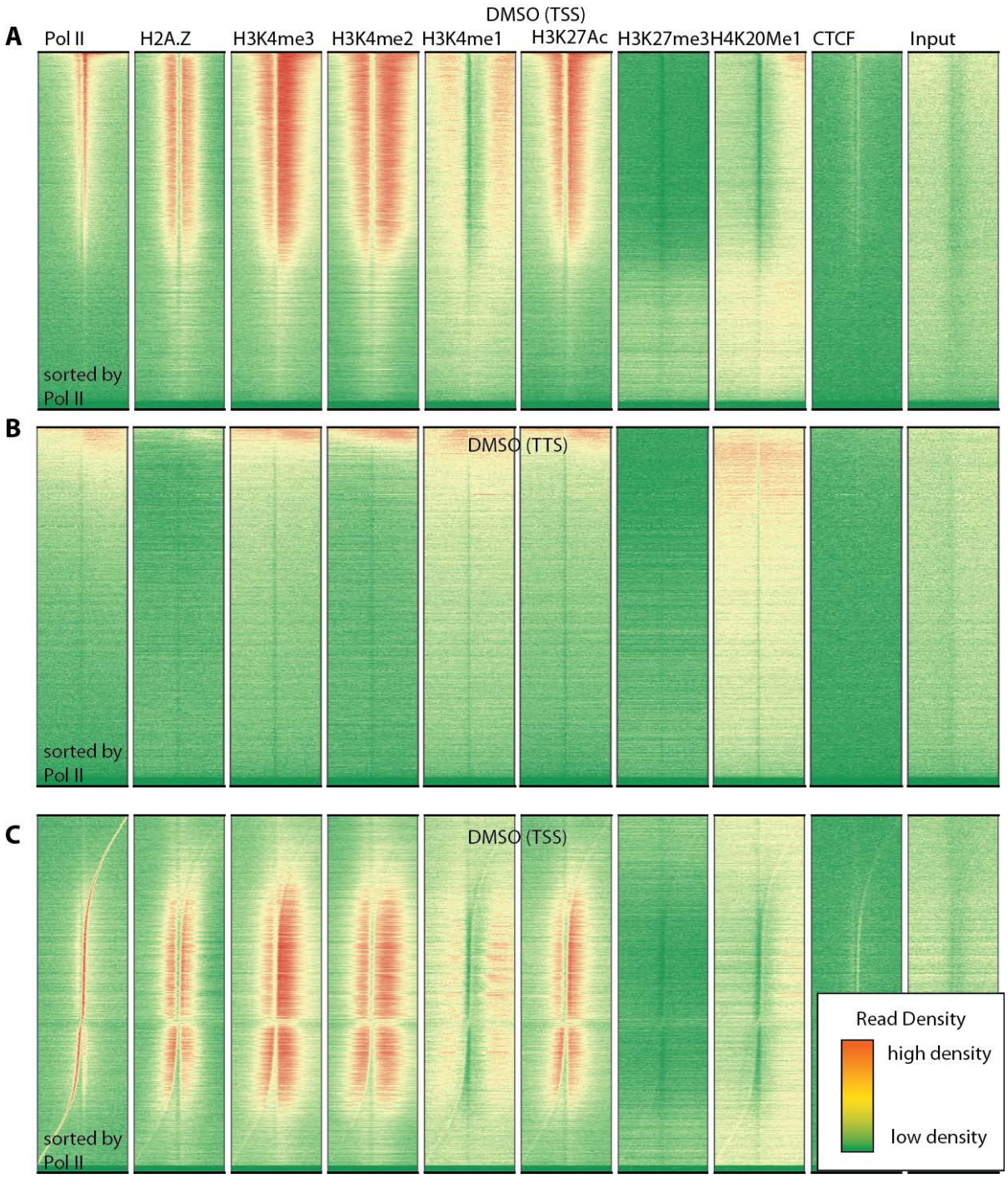


Figure 7. ChAsE heat maps of histone marks at TSS and TTS. ChIP-exo libraries were prepared for Pol II, H2A.Z, H3K4me3, H3K4me2, H3K4me1, H3K27Ac, H3K27me3, H4K20me1, and CTCF in DMSO treated HCC1806 cells. Heatmaps were produced by ChAsE and densities were sorted by Pol II densities or location. Each row represents one gene, with 18,793 lines/genes present. Red color indicates high densities and green indicates low density. A) densities +/- 2000 bp from the TSS sorted by Pol II densities. B) densities +/- 2000 bp from the TTS sorted by Pol II densities C) densities +/- 2000 bp from the TSS sorted by Pol II location. Input is included to show background densities.

It appears that H3K4me1 does not occupy the promoter, but rather several nucleosomes down from the promoter. This is consistent with previous knowledge (Figure 5) that H3K4me1 is present in intragenic activating regions of the genome, with enrichment just downstream of TSS[36]. It should be noted that H3K4me1 is also a mark of distal regulatory elements, such as enhancers[28,36]. CTCF is an insulator protein, meaning that it resides at a genomic boundary that blocks interactions between enhancers and promoters[54]. CTCF is thought to help regulate the 3D structure of chromatin. So, consistent with previous reports, CTCF has low density at both the TSS and TTS. So far, our findings of HCC1806 histone marks at promoters agree with the literature.

In contrast, relative to the termination site (TTS), there were no uniformly enriched patterns for the previously mentioned marks (Pol II, H3K4me3, H3K4me2, H2A.Z and H3K27Ac). There seems to be a few residual higher density genes, and they all correlate with Pol II, as discussed previously. This could be due, in part, to an overlap of the TTS with another gene's TSS, but these instances are uncommon in the human genome. Notably, the only mark that differs in regard to TTS peaks is H4K20me1, which our model predicts would be found at the TTS[36]. There is a higher density of H4K20me1 reads around the TTS than the other marks (as indicated by more red and yellow color in the heatmap). This correlation may not look as striking, but this is simply because the genes are still sorted by Pol II density, which would have a different pattern than H4K20me1. The input shows that there is no unusual spike in read densities and is included as a control for background, however there is a modest but uniform depletion of reads near the promoter of most genes.

Another interesting feature of the ChAsE heatmaps is the apparent trend of mutual exclusivity of Pol II and the certain histone marks. You can see that where Pol II has the highest density peaks

(TSS), the histone densities are located just to the left or the right with low density of reads at the location where Pol II density is highest. In vitro, when Pol II encounters nucleosomes, it pauses transiently, which has been well characterized as nucleosome-induced Pol II arrest[55]. Figure 6C is a different way of viewing the data from Figure 6A. In Figure 6C, the Pol II density is sorted by location relative to the TSS. This still shows that H3K4me3, H3K4me2, H2A.Z and H3K27Ac are found at active promoters (TSS where Pol II resides). This is another nice way to visualize the mutual exclusivity of the histone marks and Pol II.

The ChAsE heatmaps are useful tools for examining the positions of histone marks genome wide, identifying trends, and assessing the overall relationships across the data. However, it is not ideal for focusing in on a single gene. In genome-wide analyses, it is valuable to balance global trends with anecdotal examples that illustrate those trends in more detail at a single gene level. For this we used the Integrative Genomics Viewer (IGV), developed by the Broad Institute, to examine what each mark looked like at the TSS and enhancers of genes of interest (Figure 7). Shown in Figure 7 is the *FOXA2* gene locus, which is a tumor suppressor gene that is of interest given our model (Figure 1). The *FOXA2* gene is a hepatocyte transcription factor that is typically a transcriptional activator for liver specific genes like albumin. It is characterized as a tumor suppressor gene because the protein suppresses metastasis and the epithelial-to-mesenchymal transition (EMT) in human lung cancers[56]. The dysregulation of *FOXA2* has been noted in other cancers, including breast cancer[57].

The first thing to note in the browser view is that RNA Pol II does not have a peak within the *FOXA2* gene, suggesting that this gene may not be transcribed in the untreated HCC1806 cells. There are, however, strong H3K4me2 and H3K4me3 peaks at the promoter of *FOXA2*. There is no H3K27Ac peak within the gene body, but there is a dispersion of H3K27me3 (the repressive

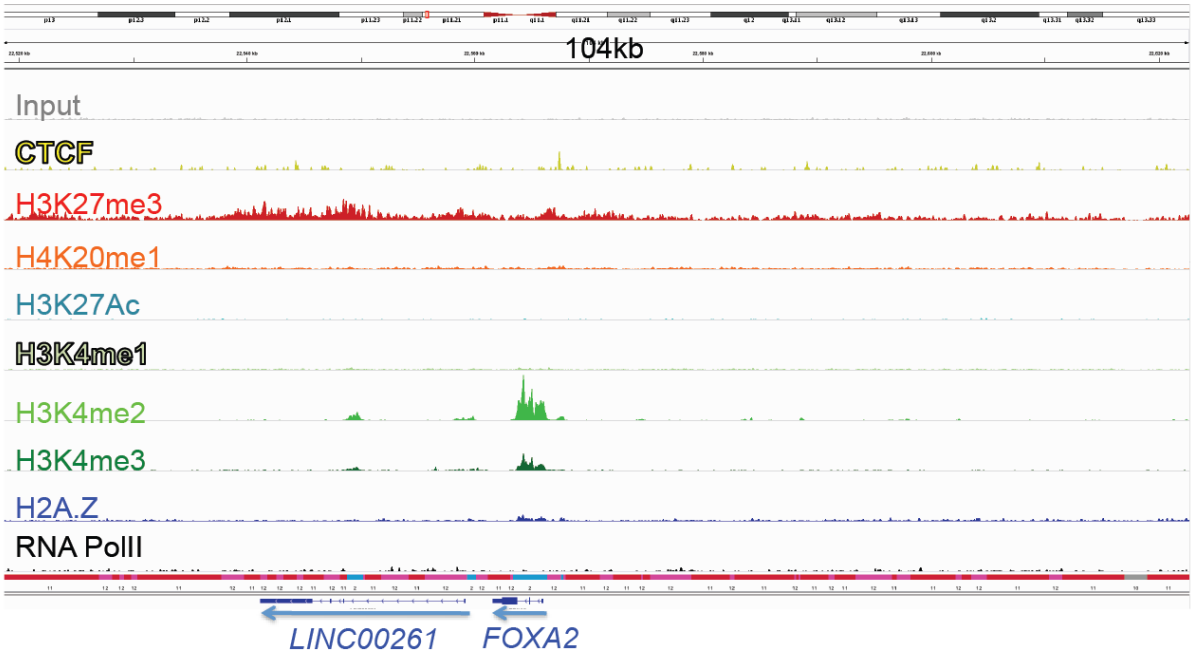


Figure 8: Browser view of ChIP signal at *FOXA2*. An IGV browser screen shot encompassing the *FOXA2* gene. The track is shown for each of the marks (labeled to the left) I performed ChIP-exo on for HCC1806 cells (DMSO control).

mark) throughout the gene body, consistent with this gene being in an inactive or repressed state. H3K4me1, CTCF and H2A.Z have slight occupancy, but no major peaks. The genome browser allows us determine which genes are associated with specific marks, and whether transcription is occurring. This adds value to global heatmap analysis patterns by illustrating the genome-wide trends in greater detail.

De novo chromatin state discovery

After analyzing these histone marks in respect to known TSS and TTS, I wanted to determine other regions of the genome where these marks may localize. Cellular epigenomes for cells in different tissues or diseases are fundamentally distinct. The epigenome plays a major role in conferring cellular identity and contributes to the stable expression profile for each cell type. In diseases, such as cancer, which are heterogeneous, the mutational and molecular profiles (such as the epigenetic landscape) of cells differ. While the mutational profile of many given cancers has been studied previously, the epigenomic landscapes of many cancer types remains unknown[58]. In order to examine the epigenetic landscape (or chromatin state) of HCC1806 TNBC cells, I used ChromHMM to run multivariate Hidden Markov Modeling (HMM) on the before-mentioned histone marks, CTCF and Pol II. HMM is a statistical model used for temporal pattern recognition, which has previously been manipulated to discover *de novo* regulatory regions in the genome. When used with chromatin mark data, the HMM model can learn patterns of chromatin mark locations and establish what regions of the genome each pattern is typically associated with. The discrete chromatin states learned from HMM can be used to infer biological function for each of the states, allowing us to make testable predictions for chromatin state functions. The ChromHMM program takes the aligned reads for each mark tested and outputs learned chromatin state model parameters and the state assignments for each genomic category based on

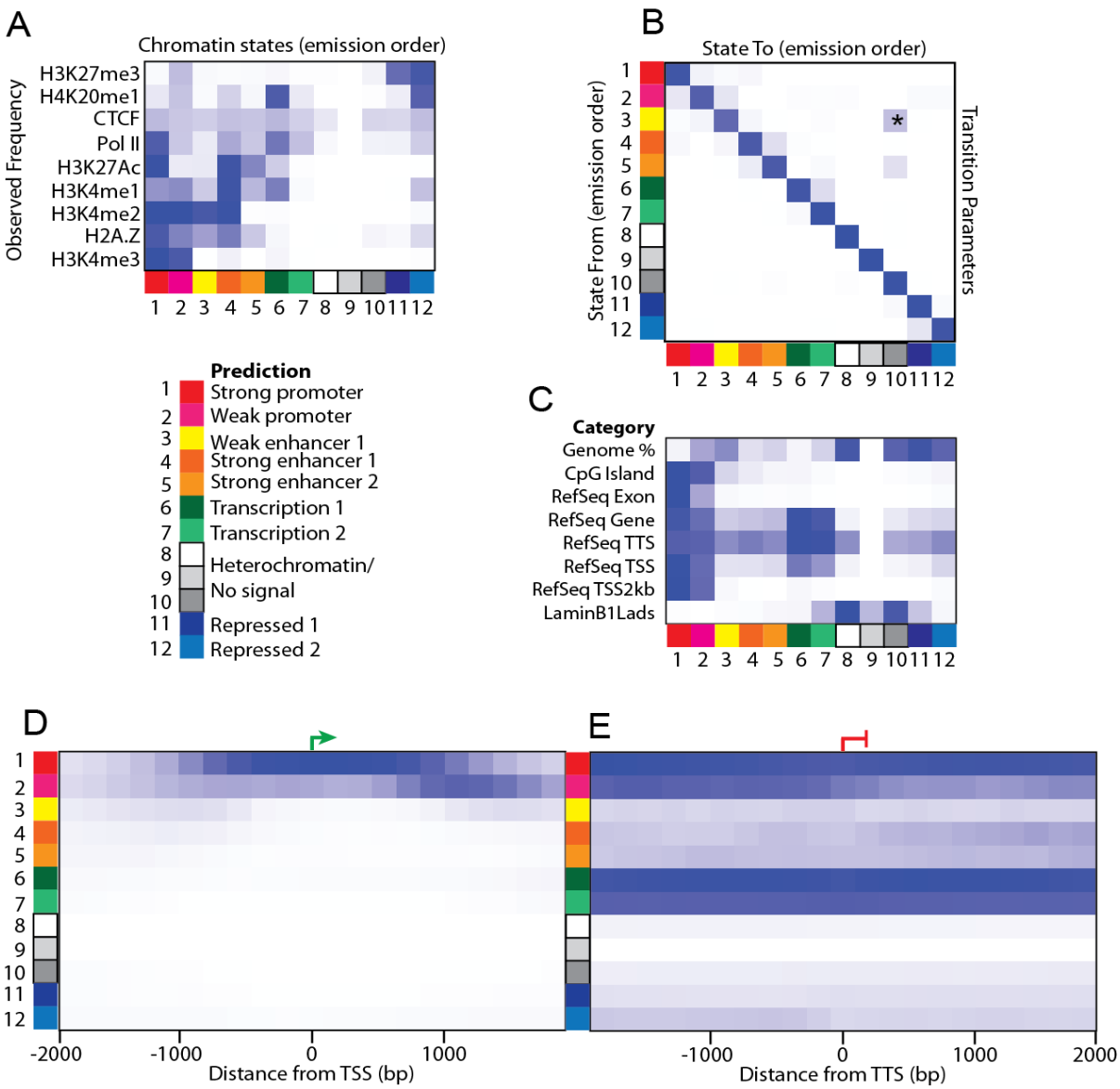


Figure 9. Predicting genomic regions with HMM. ChromHMM was run on the following HCC1806 ChIP-libraries for the DMSO control (Pol II, H2A.Z, H3K4me3, H3K4me2, H3K4me1, H3K27Ac, H3K27me3, H4K20me1, and CTCF). Twelve chromatin states were learned, because twelve chromatin states produced the greatest stratification without duplication. A) Heatmap representation of chromatin marks/proteins associated with each learned state (1-12). Blue represents the presence of that particular mark in the associated state, while white represents the absence of a mark. Functional annotation inferences for each state are listed below the heatmap and color-coded. B) The transition matrix shows the probability that the state in the row will transition to the state in the column. (*denotes the most likely state transition in the matrix, which is ~20% probable). C) The first row indicates the relative percentage of the genome represented by each state. The following rows show the relative fold enrichments of each state for the annotations listed (eg: CpG Island, RefSeq Exon, etc.). D) The enrichment of each state within -/+2000bp from the TSSs. E) The enrichment of each state within -/+2000bp of the TTS.

the temporal and spatial patterns. Figure 8A summarizes the 12 chromatin states resolved by ChromHMM from my nine ChIP-exo libraries. I chose to use 12 states because this learned model produced the best stratification of states without repeating any states. Figure 8A also summarizes our inferences (or predictions) for the biological function of each state, based on the combination of marks that are known to associate with specific regions (i.e. H3K4me3 with promoters) and the proximity of each state to a given gene annotation (Figure 8D-E), and other properties (Figure 8C).

The first output from the ChromHMM program is the "emission order", which are the learned chromatin states for our given model (Figure 8A). ChromHMM captures the emission orders as the frequency in which different histone mark combinations are found nearby one another. We then assigned a predicted biological role to each of the learned states.

Promoters are defined as the region of DNA located near the TSS where transcription of the associated gene is initiated[59]. Epigenetically, promoters are marked by the presence of H3K4me3[60]. In fact, in metazoan H3K4me3 is thought to play a role in active transcription, possibly through interactions with the TAF3 subunit of TFIID[61]. Since chromatin states 1 and 2 are the only states enriched for H3K4me3, we predicted they would be the promoter regions. ENCODE found that after H3K9Ac, H3K4me3 is the most important mark for determining regulatory region associations[36]. H3K9Ac marks active regulatory elements with a preference for promoters[36]. This data would have added interesting insights, however I had difficulty producing high quality ChIP libraries with our H3K9Ac antibodies, thus we do not have data for this mark to help corroborate putative promoters regions marked by H3K4me3. The reason for this is still unknown, but may be in part because of the low occupancy of H3K9Ac in the DMSO samples

(Figure 4). However, we were able to define promoters based on the H3K4me3 mark and H2A.Z, which is also known to be enriched at promoters[53].

Both predicted promoter states are also enriched for H3K4me1, H3K4me2, and H2A.Z. Because H2A.Z is a histone variant that is essential for chromatin architecture, it is known to flank promoters of actively transcribed genes as well as other regions of the genome, such as chromatin boundary elements, centromeres and replication origins[53]. H2A.Z is thought to contribute to higher order chromatin structure by altering interactions between tails of nearby histones[53]. Aside from the promoter regions, we also see H2A.Z enriched in our predicted enhancer regions. H3K4me1 and H3K4me2 are also found in both promoter and enhancer regions. According to the ENCODE database, H3K4me2 does mark both regulatory elements[36]. Meanwhile, H3K4me1 is associated primarily with enhancer elements[62], which I will discuss in more detail below. The enrichment of H3K4me1 in our predicted promoter regions may be due to the H3K4me1 peaks that are found just distal of the TSS[36]. Our promoter predictions may have lumped these more distal peaks in with marks that are more proximal to the TSS and are what we associate with the promoter region.

Based on the distribution and co-occurrence of these 4 marks, we inferred 2 promoter regions. We split these two promoter classifications into a strong promoter and a weak promoter classification based on the amount of Pol II occupying these regions. In gene regulation, a promoter is classified as strong or weak according to its affinity for RNA polymerase (Pol II)[52]. State 1, which we have titled the strong promoter, has higher enrichment of Pol II than state 2, the weak promoter. Along with the differential Pol II enrichment in promoters, H3K27Ac is known to specifically mark active regulatory regions at both promoters and enhancers[33,63]. The Zhao lab found that histone acetylation positively correlated with increased gene expression, but

they found that different marks associated with different regions of the genome[33]. They particularly found H3K27Ac at promoters that typically correlated with increased transcription[33]. This is another reason for our differentiation between the weak and strong promoters. Thus, our HMM results predict that the inferred strong promoter would be more active, and therefore associated with higher levels of H3K27Ac and Pol II at the promoter.

While H3K4me3 marks promoters, enhancers are marked with H3K4me1 and the absence of H3K4me3[64]. Enhancers are *cis*-active elements that activate transcription in a distance and orientation independent fashion[12]. We classified one weak enhancer (State 3) and two strong enhancerss (States 4 and 5). Each one of these states is indeed marked by the presence of H3K4me1 and the lack of H3K4me3. State 6 also meets these qualifications, but the reasoning behind categorizing it as a transcription state will be discussed in more detail below. We used H3K27Ac enrichment to help classify our enhancers. While H3K4me1 is found at all predicted enhancers, H3K27Ac is known to differentiate active and poised enhancers much like promoters[32]. The Jaenisch lab found that enhancers with the H3K27Ac mark correlate well with enhanced proximal gene activity[32], and they therefore determined that H3K27Ac is a good candidate mark for active enhancers, while H3K4me1 marks both active and inactive enhancers[32,62]. As seen in Figure 8A, the chromatin state we predicted for a weak enhancer has H3K4me1 enrichment but very little to no H3K27Ac or Pol II enrichment. Meanwhile, the two states predicted to be strong enhancers are enriched for all three. Pol II is also known to be enriched at active enhancers. Although the reasoning remains elusive, this is consistent with reports of transcription over enhancer regions[65]. One explanation is that enhancers loop Pol II to the corresponding promoter[62], and that this incidental transcription is much like divergent

transcription. Nonetheless, our predictions for enhancers are consistent with their known features based on current literature[12,32,62,65].

The next states we predicted were the transcription states. One state (State 6) was deemed a transcription state because of the enrichment of H4K20me1. H4K20me1 has been found to be enriched in active chromatin downstream of the transcription start site[66]. Because H4K20me1 is typically found at the 5' end of genes[36], State 6 was predicted to be active transcription instead of an enhancer. However, it should be noted that State 6 can not be ruled out as an enhancer, since enhancers can be found intragenically as well. The high enrichment of Pol II in State 6 and 7 also coincides with active transcription. Again, it does not rule out the possibility of other biological states; however, the appearance of Pol II and lack of H2A.Z combined in these two states is a good indicator of transcription. As previously mentioned, H2A.Z typically flanks regulatory regions (such as enhancers and promoters). These observations combined are why we predict State 6 and 7 as transcription states.

The next three learned states were classified as heterochromatin. Heterochromatin is the tightly packed form of DNA typically made up of repetitive DNA and a low density of genes[9]. While little transcription typically occurs in heterochromatin, it is a fluid structure meaning heterochromatin is capable of transitioning into active states. These three states were categorized as heterochromatin and/or "no signal" due to their lack of histone marks with ChIP-seq reads in these regions. A caveat to note is that heterochromatin harbors certain repression associated histone modifications (H3K9me3) that would better identify heterochromatin[36,67], but I did not examine these marks. The only factor that had ChIP read enrichment found throughout the entire genome was the insulator protein, CTCF, which is likely to be found on the outside of the tightly-packed, dense chromatin because a primary role of CTCF is mediating intra- and

interchromosomal contacts and establishing higher order chromatin structure[68]. Other factors or marks may be "hidden" or unreachable by ChIP antibodies in the dense heterochromatin.

The final prediction we made was for repressed states. We defined "repressed states" as genes or other dynamic regions of the genome that were actively repressed at the time of harvest in our HCC1806 cell line. This repressed state is most likely euchromatin, meaning that it is not densely packed like the predicted heterochromatin states. These states were predicted as repressed states because of the enrichment of H3K27me3. H3K27me3 enrichment is typically found at the promoter and throughout the gene body of repressed genes[69]. One interesting thing to note is that H3K27me3 is also found in State 2, which is a weak promoter state. The Majewski lab has shown that there is a subset of H3K27me3 found at TSSs associated with bivalent genes. Bivalent genes are defined here as genes that are marked with both H3K4me3 (promoter mark) and H3K27me3 (repressive mark). One caveat is that these genes may not be marked with both histone modifications in the same cell, but instead different cells within the population may have the different marks at the same location. However, if in the same cell, it is thought that the combination for these two marks keeps the gene poised to respond to environmental cues, such as developmental or growth signals[69]. This fits with our prediction that State 2 is a "weak" promoter. So, by reviewing literature and carefully analyzing which marks were found in which state, we were able to infer a biological role for each of the 12 states the ChromHMM model learned. More importantly, we now experimentally measured where these states are located throughout the HCC1806 genome. These chromatin state locations are likely unique to the HCC1806 TNBC BL2 preclinical cell model.

After ChromHMM learns the different emission states, it outputs transition states for each learned state. This is used to show how robust the learned states are. Transition probability is

based on the probability that within 200bp of the interval on the chromosome another state will be found. Each state has very low probabilities of transitioning to another state (Figure 8B), which is seen from the strong blue diagonal line. This diagonal line is the desired outcome for this analysis. At the extreme, only one state has > 0.25 probability of transitioning to another state. State 3, the predicted weak enhancer, has 0.25 probability of transitioning to state 10, one of the heterochromatin states. This means that 25 out of 100 times the state 10 is found within 200bp of state 3. This heat map illustrates the stable separation of our states. So, we believe that this transition matrix solidifies the separation of the 12 states we have chosen.

Next, the ChromHMM program overlaps each of the 12 states with known annotated features of the genome (Figure 8C). The heat map indicates the relative fold enrichments of each chromatin states for each of the annotations. The first row shows the relative frequency of each of the states throughout the entire genome. This indicates that the repressed states and heterochromatin make up the majority of the genome. Meanwhile, the transcription states, strong promoters, and strong enhancers do not represent a large proportion of the genome. However, we would not expect these regions to take up the majority of the genome, because the ENCODE project found that less than a fifth of the genome regulates the 2% of the genome that encodes proteins[36].

The remainder of the heat map shows the distribution of each state for different annotations (Figure 8C). The first of which is CpG islands, which represent promoter regions[62]. States 1 and 2 have the strongest enrichment of CpG islands, which correlates with our prediction that these states are promoters. The next five annotations are from RefSeq, which is the NCBI reference sequence database. The first of these is RefSeq Exon, which is found mainly at the two predicted promoter states. The RefSeq exon annotation is a list of exons that are found in the reference cDNA libraries, from 5' to 3'. The states defined by ChromHMM are not meant to match the

classical definitions of regions, like promoters. Instead, ChromHMM identifies regions of the genome that have histone modifications typical of promoter regions. This is one reason why the ChromHMM promoter region may match RefSeq Exon regions. The promoter states may extend into the body of the gene because one of the defining marks (H3K4me3) is known to spread downstream of the TSS. In fact the spreading of H3K4me3 indicates genes that are essential for the identity and function of the given cell type[70]. Therefore, we are still confident that States 1&2 are promoter regions. The next annotation is RefSeq genes, which is a list primarily of protein coding regions of the genome and encompasses the body of the gene. It makes sense then that all of the states that Pol II is found in (states 1-7 and 12) are RefSeq gene annotated regions. This includes both the ChromHMM promoter and transcription states. The enhancer states are also weakly enriched in this annotation. This could be because enhancers are often found intragenically within introns[71]. Lastly, the repressed states are also slightly represented in the RefSeq gene annotations. All of the states should include some genes, so this annotation fits with our predictions.

The next three states represent specific regions of the gene itself (Figure 8C). The first is the TTS. Every state seems to have some regions that overlap with TTS annotated regions. State 6 and 7 have a higher proportion of TTS than the other states, which makes sense because these are our transcription states. State 6 specifically was enriched for H4K20me1, which is a mark of TTS. However, TTS are found all throughout the genome so the fact that all states have some TTS annotated regions included is not surprising. The TSS heatmap densities are not quite as dispersed as the TTS, but there are still many states that have TSSs. The two states with the highest enrichment are our defined promoter regions (States 1&2). The other more enriched sites are the transcription states (States 6&7). Next, the ChromHMM program zooms in on the TSS

region, with regions only 2kb from the TSS. With this narrowed down region, only the two promoter regions were enriched for this annotation. Again, these annotations fit well with our predictions.

The final annotation included is for repressive chromatin environments (or heterochromatin) in the genome. They use Lamin B1 associated domains (LAD) annotation to represent these regions of the genome. LADs are regions of the genome that are located in close proximity—possibly anchored to—the nuclear lamin. The borders of these domains are typically surrounded by CTCF or by CpG islands that are located away from the LAD[72]. In agreement with our prediction of this chromatin state as heterochromatic/no signal, our heterochromatin states (States 8-10) are enriched for this annotation. The surrounding regions are transcription and a repressed state, which may indicate that these regions may move in and out of attachment to the nuclear lamin. Taken together, the predictions we made on our states based on the histone marks are in general agreement with the annotated regions found in each of the states.

In the final two images that ChromHMM generates, the program zooms in on the TSS and TTS to see which states are found within 2kb of each. In Figure 8D, we look at +/- 2000bp from the TSS. Again, we see that only State 1 and State 2 are found within this range of annotated TSSs. It seems that the strong promoter (State 1), is centered right around the TSSs while the weak promoter (State 2) is enriched slightly more downstream from the TSS. I think that these weak promoters may represent poised promoters for inactive genes, so it is interesting that the marks associated with weak promoters center ~1000bp from the TSS. The second heat map (Figure 8E) is similar to Figure 8D, but this time it zooms in to the TTS. Again, there is not too much interesting here as TTSs are found throughout the entire genome. This image does not show

much more information than Figure 8C already conveyed, but rather portrays a distance relationship for each chromatin state with prominent genomic features.

Through using ChromHMM we were able to generate a lot of predictions about certain regions in the HCC1806 genome, and reveal their corresponding locations. We found regions that are likely to be promoters and enhancers as well as regions that are repressed by H3K27me3 or are repressed through compaction in heterochromatin. While this information is available for a variety of cell lines, until now, this information was unknown for the HCC1806 cell line—which is significant since no cell line, tissue type or tumor type is identical in their epigenetic landscape. Importantly, this information can be used to provide an epigenetic frame-of-reference for how the different marks observed alter in response to different developmental signals or drug treatments. You can examine which regions change from repressed to active transcription, or poised to active promoters. Thus, ChromHMM is a very powerful tool to examine *de novo* chromatin states within a genome.

Genome-wide H3K27Ac changes in response to Panobinostat treatment

After examining the untreated epigenetic landscape of HCC1806, I was interested in changes brought on by the Panobinostat treatments. For the remaining studies, I chose to focus on H3K27Ac because of the nature of the drug treatment. With Panobinostat being an HDACi, I expected H3K27Ac to increase and I wanted to examine whether this increase was genome wide or in specific areas of the genome.

To examine this question I used the program HOMER, which is an all-inclusive suite of analysis tools for ChIP and expression data. The program can find peaks, annotate peaks (or transcripts)

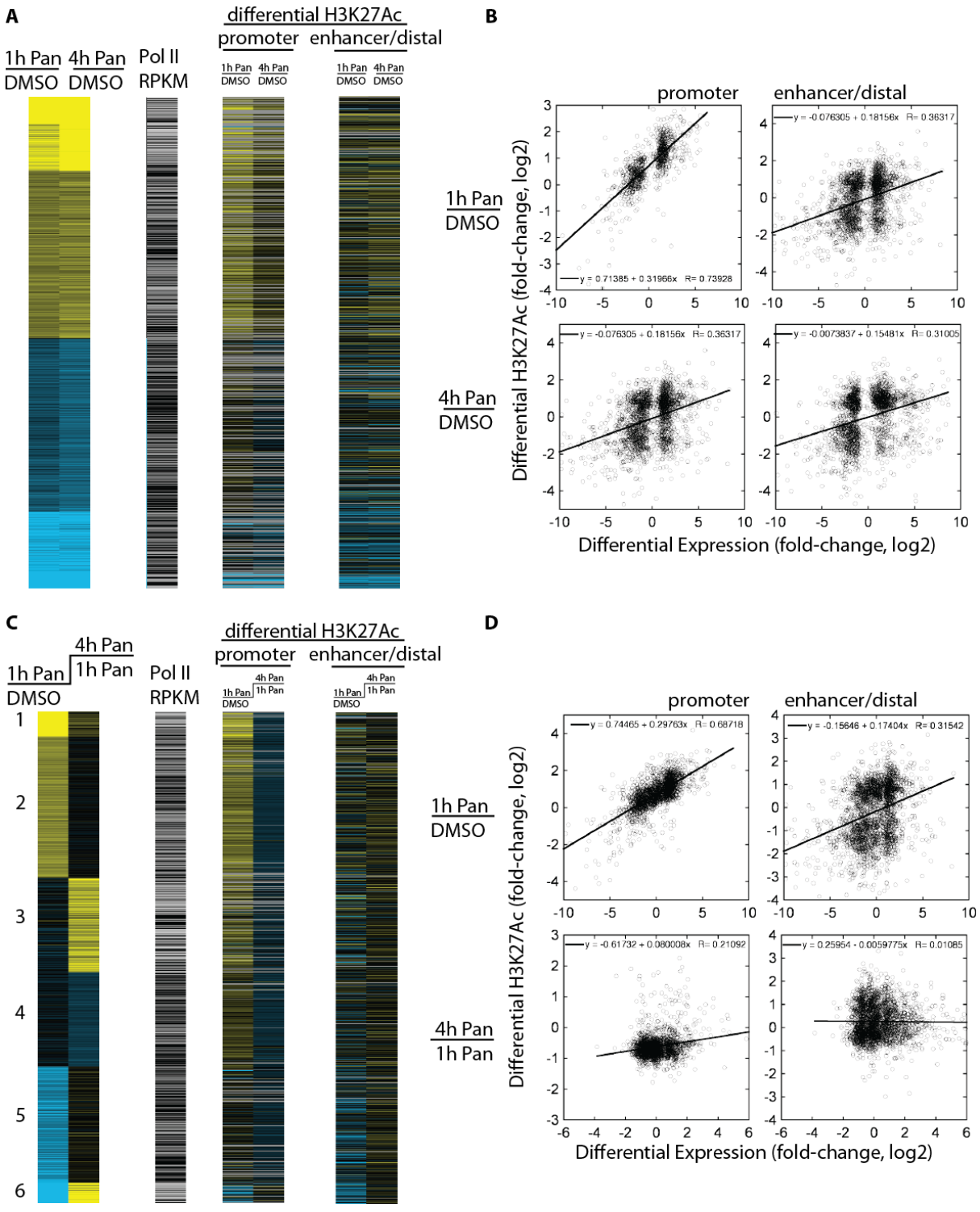

A
differential H3K27Ac
promoter
enhancer/distal
1h Pan
DMSO
4h Pan
DMSO
Pol II
RPKM
B
promoter
enhancer/distal
1h Pan
DMSO
4h Pan
DMSO
Differential H3K27Ac (fold-change, log2)
Differential Expression (fold-change, log2)
y = 0.71385 + 0.31966x R= 0.73928
y = -0.076305 + 0.18156x R= 0.36317
y = -0.0073837 + 0.15481x R= 0.31005
C
4h Pan
1h Pan
D
y = 0.74465 + 0.29763x R= 0.68718
y = -0.15648 + 0.17404x R= 0.31542
y = -0.61732 + 0.080008x R= 0.21092
y = 0.25954 - 0.0059775x R= 0.01085

Figure 10: Differential expression and differential H3K27Ac occupancy reveal correlation between H3K27Ac and change in expression. Heat maps show the differential expression and H3K27Ac differential occupancy between DMSO, 1 and 4 hour Panobinostat treatment. Yellow=increased expression or occupancy and Blue=decreased. A) Incremental differential expression was compared for 1 hour Panobinostat over DMSO and 4 hour Panobinostat over DMSO. Each column is a comparison of the 2 indicated experiments and each row represents a gene. A total of 4344 genes/rows are present. Differential H3K27Ac occupancy was aligned with the statistically significant differentially expressed genes. Promoters were any reads within 2000bp of the TSS and enhancer/distal regions were those more than 2000bp outside of the TSS. Also Pol II RPKM was used as a control that the clusters did not have aberrant expression patterns prior to treatment. B) Log fold change of gene expression vs log fold change of H3K27Ac occupancy at both promoters and enhancers for those shown in A. C) Differential expression was compared for 1 hour Panobinostat over DMSO and 4 hour Panobinostat over 1 hour Panobinostat. Each column is a comparison of the 2 indicated experiments and each row represents a gene. A total of 4004 genes/rows are present. Differential H3K27Ac occupancy was aligned with the statistically significant differentially expressed genes. Again, promoters were any reads within 2000bp of the TSS, enhancer regions were any region further than 2000bp from a TSS. The Pol II RPKM control was included as well. D) Log fold change of gene expression vs log fold change of H3K27Ac occupancy at both promoters and enhancers for those shown in C.

and prepare quantification of the peaks to be used for differential expression or occupancy analysis. I used the differential analysis to examine areas in the genome where H3K27Ac levels were altered the greatest (Figure 9).

At this point, I will only discuss the differential occupancy of H3K27Ac, and I will look more in-depth at the connection to gene expression shortly. In Figure 9A and 9C, for H3K27Ac at both promoters and enhancers/distal regions there are some genes in which H3K27Ac occupancy increases (yellow) while there are some genes where H3K27Ac occupancy decreases (blue). It is interesting to note that in the promoter regions the majority of the H3K27Ac increase occurs within the first hour, while some of the enhancer regions do not increase until the 4 hour time point (Figure 8A and 8C). This could indicate that acetylation at promoters is necessary for transcription, whereas an increase of acetylation at enhancers is only needed to increase the current level of transcription, and is therefore a secondary location of acetylation.

Finally, to independently corroborate the HOMER-based assignment of differential H3K27Ac as presumably being enhancers, I wanted to determine if these regions actually mapped to the regions we defined in the HCC1806 genome from our ChromHMM analysis. I used the intersectBed function to see which of the distal/enhancer regions where H3K27Ac occupancy changed significantly (FDR<5%) were defined as any of the enhancer states (states 3-5) by the ChromHMM analysis (Figure 10). Remarkably, we found that 75% of the distal regions with differential H3K27Ac occupancy indeed resided within ChromHMM annotated enhancers. This may be skewed slightly as the presence of H3K27Ac helped to define some of the ChromHMM states to begin with. However, this is still important to note that regulatory regions are specifically increasing in H3K27Ac occupancy, and not just random distal regions of the genome.

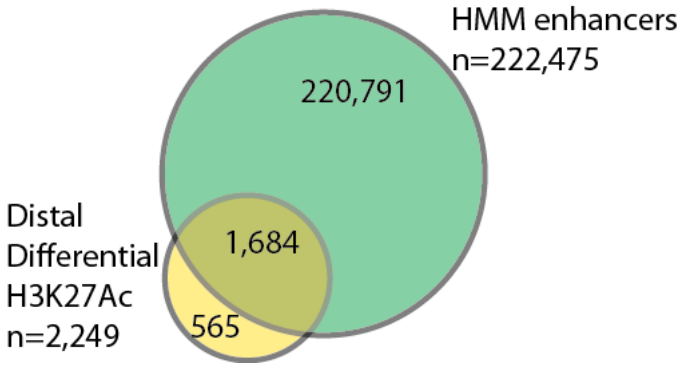


Figure 11. Majority of distal H3K27Ac differentially occupied regions are HMM enhancers. The total number of differentially expressed H3K27Ac occupied distal (>2000bp from TSS) regions aligned to the regions defined as enhancers by ChromHMM (States 3-5). P-value $1.5x10^{-125}$. P-value was calculated using the chi-squared test.

Gene expression changes post Panobinostat treatment

After looking at the changes in histone acetylation, our model predicts a corresponding increase in gene expression following Panobinostat treatment. An increase in histone acetylation following Panobinostat treatment will coincide with an increase in the expression of the genes associated with modified acetylation. To test this prediction, I first examined the changes of gene expression by itself. The first panels in figure 9A and 9C show the statistically significant (FDR <0.05) differentially expressed genes for Panobinostat 1 hour and Panobinostat 4 hour treatments. There were many genes with increased expression, but there were almost as many genes with decreased expression. While we do not know the mechanism behind decreased expression of the latter set of genes, such changes in gene expression could be due to a secondary events following HDAC inhibition. There is precedence for such indirect effects in the literature[73]. To provide more biological insight into the groups of genes that displayed increased or decreased expression in response to Panobinostat, I conducted gene ontology and pathway analysis on the six clusters of genes presented in Figure 9C (incremental differential expression). This could provide insight into which biological pathways were being altered the most following Panobinostat treatment.

Because the incremental Panobinostat 1hr treatment over DMSO clustered with Panobinostat 4hr treatment over Panobinostat 1hr treatment (Figure 9C) more distinctly resolved each cluster, I chose to examine these clusters in more depth (as opposed to Figure 9A); this analysis is summarized in Table 8. However, it should be noted that both clusters had very similar gene ontology annotations. The first 3 clusters include the genes that have increased expression levels. According to the gene ontology, some transcription factor genes, such as *FOXA2*, increase in expression. Cell cycle proteins also increase in expression, which may include regulators of the

Table 8: Gene Ontology and Pathway Analysis			
Cluster	**# of Genes**	**Gene Ontology (**p-value**)**	**Pathway** (p-value)
1	221	Regulator of Transcription (n=52, 10^{-7}): ie) *Myb, FoxA2, Pou51*	TGF-β (10^{-2})* Wnt (10^{-2})*
2	1249	Cell cycle (n=69, 10^{-4}) Drug response (n=25, 10^{-3})	p53 (n=7, 10^{-4}) MAPK (n= 25, 10^{-2})*
3	837	Chromatin assembly/disassembly (n=17, 10^{-6}) Regulation of cell death (n=45, 10^{-3})	Hedgehog (n=6, 10^{-1})*
4	832	Cell proliferation (n=38, 10^{-7})	Cancer signaling (n=28, 10^{-5}) *
5	1027	Immune response (n=43, 10^{-4})	NF-KB (n=4, 10^{-2})
6	178	Inflammatory response (n=11, 10^{-5})	

*FDA-approved inhibitor on the market targeting this pathway

cell cycle that halt unfit cells at certain checkpoints. This does, in fact, correlate with our model. *FOXA2* is a known tumor suppressor in lung cancer, so the increase of *FOXA2* following Panobinostat treatment should inhibit tumor growth. Cell cycle regulators and regulators of cell death increasing could also have a similar effect as tumor suppressor genes.

It is also interesting to look at the gene ontology of genes whose expression decreases. In the three clusters with decreased expression, cell proliferation and immune/inflammatory response are enriched, meaning that expression of proteins related to these pathways is decreasing. According to Hanahan and Weinberg, these are all pathways that cancer cells use to generate uncontrolled growth[21]. This shows that our prediction is correct—pathways that help shut down growth, such as TSGs, are upregulated. While our model did not predict that oncogenes would be shut down—it is interesting to see that these pathways are indeed downregulated.

However, an interesting response is that most of the pathways that are upregulated involve both oncogenes and tumor suppressor genes, such as *Wnt*, *TGF-β, MAPK* and *TP53*[74-77]. The dual roles of these pathways can make them difficult to treat, as they need to be maintained at a healthy balance. These pathways may get out of control and could be a good target for cells that exhibit Panobinostat resistance. It is important to note that some of the pathways that are upregulated, such as the TGF-β, Wnt, MAPK, and Hedgehog pathways have FDA approved drugs already on the market[74,77-79]. These drugs can be tested in combination with Panobinostat to see if cell death is expedited with these combinatorial therapies.

I wanted to test the extent to which the genes whose expression changes observed in the clustering analysis were TSGs, as I previously hypothesized. To do this, I did a similar analysis as H3K27Ac mapped to ChromHMM enhancer regions. In doing so, I found that 93 of the 716 TSGs[80] were upregulated (13%) following Panobinostat treatment (Figure 11).

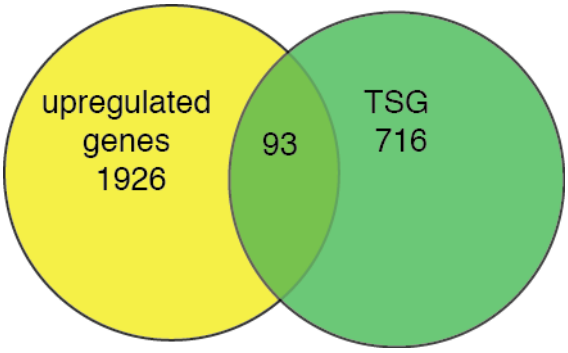


Figure 12. Venn Diagram comparing TSG to genes upregulated after Panobinostat treatment. I used IntersectBed to compare the lists on 716 TSGs to our list of upregulated genes following 1 hour of Panobinostat treatment. P-value=$5x10^{-7}$, as calculated by chi-squared tests.

I found that the upregulated TSGs had an average of 1.92 fold expression change and that ~5% of the upregulated genes were tumor suppressor genes, supporting my hypothesis. However, it should be noted that 91 down-regulated genes were also on the list of 716 TSGs, suggesting that not all TSGs respond in the same way to Panobinostat treatment. Presumably, the downregulated TSGs, although expressed at some detectable level in the HCC1806 tumor cells, were not able to subvert the oncogenic signaling pathway to drive the cell toward apoptosis.

Correlation between expression and H3K27Ac

While we learned a great deal from the expression analysis, our hypothesis was that increases in H3K27Ac occupancy are linked to increases in expression of TSGs in response to HDACi treatment. Although this general trend between H3K27Ac and expression has been observed previously in literature[32,62], it is not known how this occurs over time in response to Panobinostat in a TNBC preclinical model. I found that, especially at promoters, the H3K27Ac change in occupancy correlated with changes in expression (Figure 9B and 9D). In both examples, the correlation of expression and H3K27Ac at one hour Panobinostat treatment has an R-value of ~0.75 at the promoters. Interestingly, this correlation goes down greatly by 4 hours, with an R-value between 0.2 and 0.35. The shape of the distribution also flattens after one hour. At one hour Panobinostat treatment, the correlation between expression and occupancy at promoters has a slope of 0.32, meaning that for every 32% H3K27Ac relative fold occupancy increase, expression would increase by 100%. By four hours of treatment the slope decreased to 0.08. Meaning that expression may still be increasing, but H3K27Ac occupancy has ceased increasing. This makes sense with the treatment of Panobinostat to the cells. By four hours the occupancy of H3K27Ac may have maxed out at these locations, and because HDACs are inhibited, they are

not able to remove the acetylation. Therefore, by this time the acetylation of histones at these locations may be static-or unchanging. However, the high levels of histone acetylation at these locations are still signaling increased transcription. Therefore, while acetylation is maxed out, transcription levels can continue to increase.

We see a similar trend with the enhancers, although it is less drastic than the promoter regions. The correlation R-values at enhancers are all under 0.4. The slope for the enhancers is only 0.17 after one hour of treatment as well. This is about half of what we observed at the promoter. This means that for an increase of 17% H3K27Ac occupancy we see an increase of expression by 100%. This also tapers off by 4 hours as well, though, with a slope below 0.01—meaning that expression may still be increasing but H3K27Ac occupancy has stopped increasing. This tapering off at both locations is most likely due to the fact that H3K27Ac occupancy has reached its maximum occupancy at these locations and has not been able to be removed, because of the inhibition of HDACs. Again, you get a build-up of the H3K27Ac early on, but this continues to signal transcription, creating a constant increase of transcription at these sites. This analysis provides a genome-wide view of the correlation between acetylation and expression. Again, since they were positively correlated, this result supports our hypothesis that increased acetylation causes an increase of transcription, and that while H3K27Ac maxes out at earlier time points, an increase of expression at these genes continues to occur after the H3K27Ac tapering.

I then wanted to zoom into a gene-by-gene view of some of the TSGs that were upregulated upon Panobinostat treatment. To do this we established a list of possible candidates by filtering genes for the highest changes in gene expression and then filtered this list to look at TSG. We chose to examine *FOXA2* more in depth because it was one of the only TSGs in the 50 genes

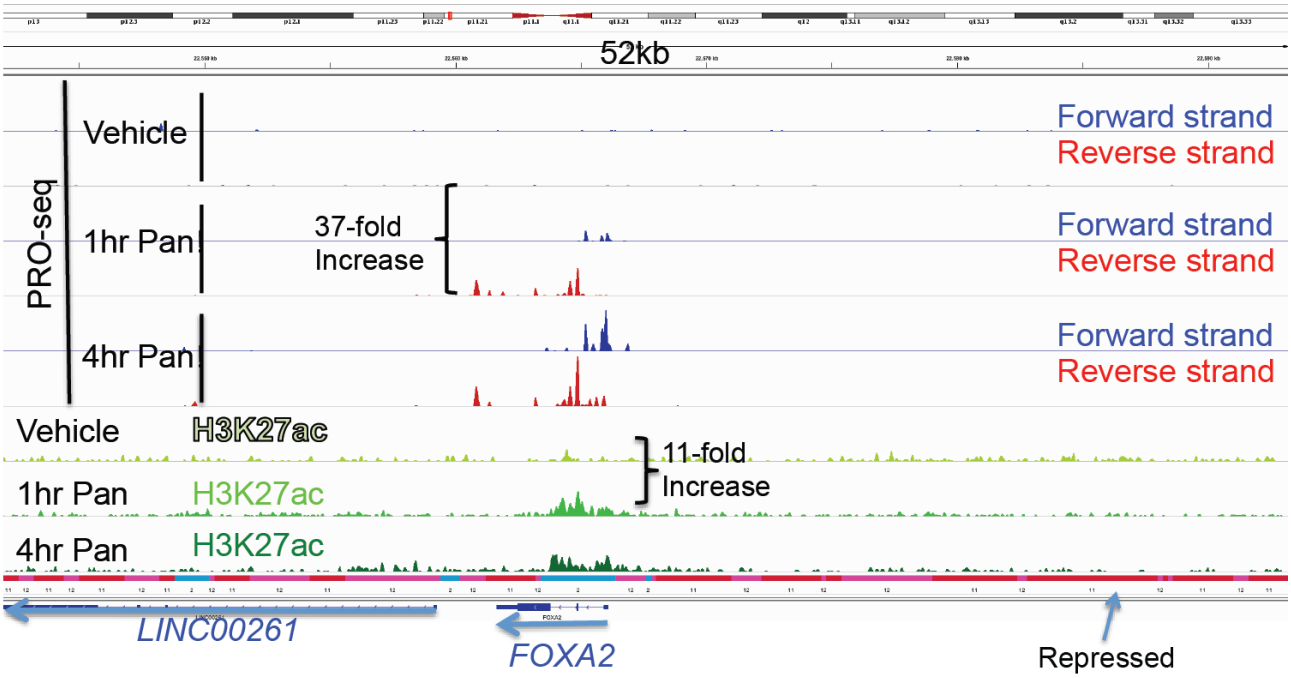


Figure 13. IGV Browser view of H3K27Ac increasing with Panobinostat treatment at TSG *FOXA2*. A browser view showing 52kb of the *FOXA2* gene, with genome tracks for H3K27Ac treated with DMSO (vehicle) and 1 and 4 hour Panobinostat as well as PRO-seq tracks for the same treatments.

with highest expression changes, so examining this gene is more relevant to my hypothesis than a non-tumor suppressor gene with slightly higher expression changes. We previously looked at *FOXA2* prior to Panobinostat treatment, but now I wanted to look at changes in H3K27Ac peaks in response to Panobinostat (Figure 12). This browser view shows us that *FOXA2* does indeed increase in both expression and H3K27Ac occupancy following Panobinostat treatment. You can see that there is a marked increase (11-fold) of transcription for *FOXA2* between DMSO and 1 hour Panobinostat libraries. The reads for PRO-seq also increase drastically (37-fold) in the same time period. So, it appears that H3K27Ac occupancy and transcription increase together. You can also see that transcription continues to increase into the 4 hour time period; however, when you look at the occupancy of H3K27Ac at 4 hours there is a decrease from the 1 hour—but still a net increase from the DMSO treated libraries. This is similar to what we saw globally in Figure 9. Again, my interpretation is that the transcription levels appear to continue to increase even though H3K27Ac occupancy is no longer changing—most likely because there is already a maximum of H3K27Ac at the promoter, allowing transcription to proceed.

Another anecdotal example of a TSG that increased in transcription following Panobinostat treatment in HCC1806 cells was *EGR1*. *EGR1* is a transcription factor, so it falls into the gene ontology corresponding to cluster 1 (Table 8), similar to *FOXA2*. *EGR1* is an immediate early gene implicated in differentiation. The genes that *EGR1* targets are required for differentiation and mitogenesis, which is why *EGR1* is classified as a tumor suppressor gene[81]. Much like *FOXA2* regulation, we see a coincident increase in both H3K27Ac occupancy and gene expression following Panobinostat treatment (Figure 13) at the *ERG1* gene.

Accordingly, much like *FOXA2*, we see an increase of H3K27Ac (1.7 fold) between the DMSO treated sample and the Panobinostat 1 hour sample at the *ERG1* promoter. We chose to examine

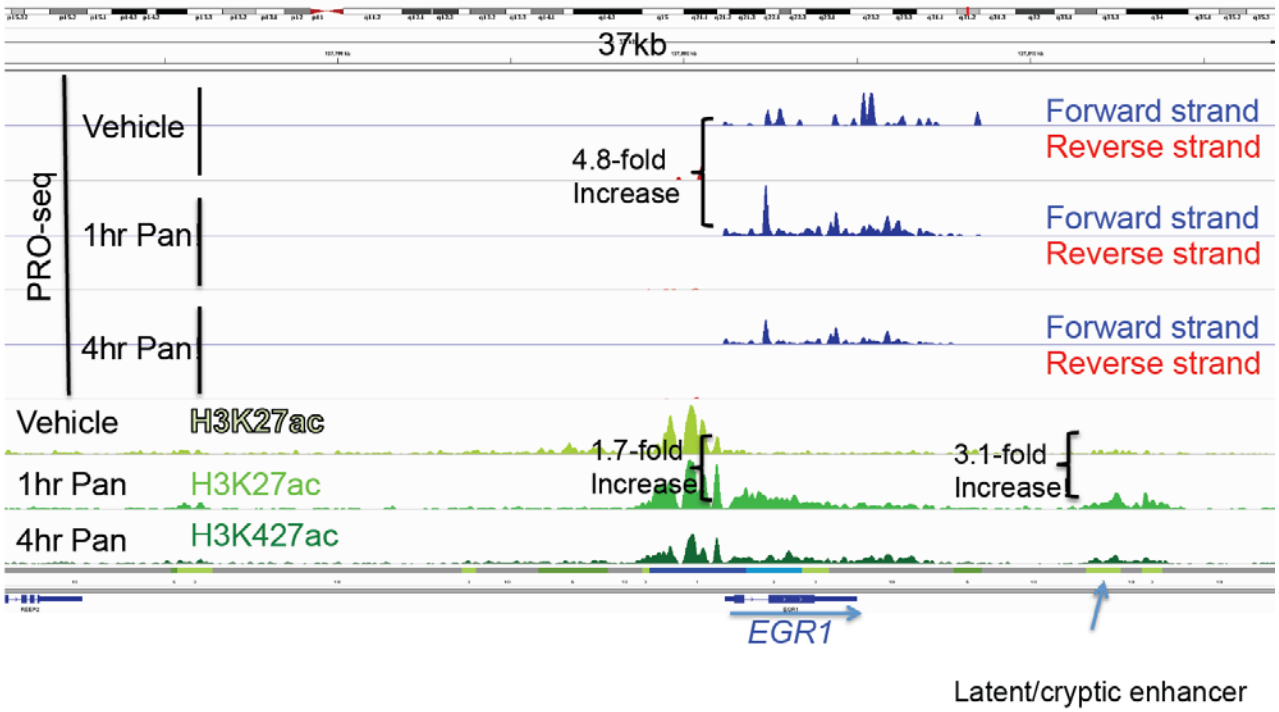


Figure 14. IGV browser view of H3K27Ac increasing with Panobinostat treatment at TSG EGR1. A browser view showing 37kb of the *EGR1* gene, with genome tracks for H3K27Ac treated with DMSO (vehicle) and 1 and 4 hour Panobinostat as well as PRO-seq tracks for the same treatments. The tracks are all normalized to show the same number of reads as equivalent density.

the *EGR1* gene, because it too was a TSG that fell in the top 50 genes with significant gene expression increases, and unlike *FOXA2*, *EGR1* had a large fold change in H3K27Ac occupancy at an enhancer region. The acetylation also goes down slightly in the 4 hour Panobinostat treated sample much like we saw in *FOXA2*, but still remains higher than the DMSO treated base-line. Again, we see an increase of transcription (4.8 fold) that corresponds with the H3K27Ac occupancy increase between DMSO and 1 hour Panobinostat. In this case, the transcription appears to go down again by the 4 hour time point. The difference between the continued increase of *FOXA2* transcription but the decrease in *EGR1* transcription at 4 hours may be linked to histone acetylation levels—as *FOXA2* has a much greater fold increase than *EGR1* so more H3K27Ac may continue to activate transcription at extended periods of time. Another difference between the two browser views of the TSGs is that you can see an *EGR1* enhancer in the genome browser, but no enhancer is shown in the *FOXA2* browser view. The enhancer is downstream from the *EGR1* TSS, but it also experiences an increase of histone acetylation between the DMSO and 1hr Panobinostat treatments (3.1 fold). This enhancer acetylation may play a role in increased expression levels as well. However, our experiments are not able to tease apart which acetylation plays a greater role on expression levels. Both of these browser views are representative illustrations of gene level examples of the genome-wide analysis we saw in Figure 9.

Concluding Remarks

In conclusion, our results support the model I originally hypothesized. First, I showed that a wide range of TNBC cells were sensitive to Panobinostat, as compared to other cancer and non-

cancerous cell lines. Next, I presented that Panobinostat treatment caused an increase of acetylation levels in HCC1806 cells and that acetylation continues to increase globally up to (and possibly beyond) 16 hours of Panobinostat treatment. These observations had been published, with limitations, by the Collins-Burow's lab previously[2]. I show that Panobinostat is effective in suppressing growth in many subtypes of TNBC and then verified that the cell lines I would be working with for the remainder of the project also exhibited sensitivity and increased histone acetylation.

After examining the phenotype of Panobinostat treated TNBC cells, I wanted to uncover the link between epigenetics, transcription and cell death. I started by looking at epigenetic and transcriptional changes that occur following the drug treatment. To uncover epigenetic changes, I had to first look at the epigenetic landscape of untreated HCC1806 cells. I used ChromHMM to identify 12 *de novo* chromatin states and their locations within the genome including promoters, enhancers, transcription, heterochromatin, and repressed states.

Once the identities and locations for the chromatin states in vehicle treated HCC1806 cells was discovered, I could start analyzing changes in acetylation at different regions of the genome. For this thesis, I focused solely on H3K27 acetylation because it is found at both promoters and enhancers, and is linked to increased expression of the genes where the modification is found. I observed that H3K27Ac increased in both promoter and enhancer regions in the genome. I utilized the learned states from ChromHMM to verify that these regions were most likely enhancers, and found that 75% of the distal (non promoter) regions where we saw an increase of H3K27Ac were indeed classified as enhancers.

Along with H3K27Ac occupancy increases, I also saw an increase of expression in a subset of genes. Specifically, our model predicted that some tumor suppressor genes would increase in H3K27Ac occupancy due to HDAC inhibition and subsequently would increase in expression to drive the cell toward programmed cell death. I examined the proportions of TSGs in the up-regulated genes, and found that 93 out of 716 TSGs (13%) were statistically significantly upregulated following Panobinostat treatment of one hour. This agrees with our model that TSGs will be upregulated following Panobinostat treatment. However, it should be noted that almost as many TSGs were down regulated following Panobinostat treatment. The reasoning behind this requires a more in-depth analysis and follow-up experiments that are beyond the scope of this thesis. This would be interesting to examine in the future directions of this project.

The final aspect of this project was examining the link between histone acetylation and increased expression. I found that at promoters, increased expression strongly correlated with increased transcription following one hour of Panobinostat treatment. I also found that there was a positive correlation between H3K27Ac occupancy at enhancers and gene expression; however, the correlation was weaker at enhancers than at promoters. These correlations also seemed to decrease drastically by 4 hours, indicating that, in fact, the maximum acetylation may have been reached at earlier time points, while expression continued to increase.

Importantly, my work on this project uncovered pathways involved in the Panobinostat response. Identification of these pathways have implications in the rational design of combinatorial therapies. Thus, experiments that follow up on this work could have a lasting impact on the lives of patients with TNBC.

In conclusion, my results and analyses support the model proposed in Figure 1 for how growth of HCC1806 TNBC cells are inhibited by blocking HDAC activity pharmacologically with Panobinostat. The results of this study are novel because it combines different genomics approaches to understand why TNBC cell are sensitive to HDACi from an epigenetic point of view. Although HDACi induced epigenetic changes are the focus of this thesis as one model for how growth of TNBC cells are impaired by HDAC inhibition, there are competing models that have been proposed. For example, one model posits that there could be other proteins that are acetylated, such as p53, α-tubulin or HIF-1α[23], that cause the change in expression. The histone acetylation at these genes may just be a side effect, and not a contributing factor—although this is less likely. So, while our model has held up through our experiments, I have not eliminated all other possibilities for mechanisms behind Panobinostat induced cell death. I would have to perform site-specific acetylation experiments to verify causal links between histone acetylation and increases of expression. To do this you could use CRISPR-Cas9 linked to p300 to direct the p300 mediated enhancer acetylation of H3K27 at specific enhancer loci. Then one could measure the expression levels of the adjacent gene or cell growth rates. This would help verify the causal link of H3K27 acetylation to both increased transcription rates and cell death, but this is not a trivial experiment. We would also have to do overexpression assays to probe which genes are mainly responsible for the cell death. For example, in future experiments, one could overexpress *ERG1* or *FOXA2* in our HCC1806 cell line and measure cell growth rates, to determine if the increase of expression of these TSGs is responsible for the cell death of HCC1806 cells following Panobinostat treatment. More likely than not, there are multiple genes whose expression changes are responsible for Panobinostat induced cell death. Therefore, my experiments are consistent with our model. Our analysis reports correlations with HDACi

treatment epigenetic changes and cell death. Because these experiments only indicate correlation, additional experiments to test the predictions our analysis makes are needed to corroborate the hypothesis.

APPENDIX A

LIST OF FILE FORMATS

BAM	Compressed binary version of the data from SAM files
BED	This format provides a flexible way to display data in an annotation track. BED files differ between programs, but have three required fields (chr=name of chromosome, chrStart and chrEnd positions).
BigWig	This is a format used to display dense and continuous data in a genome browser. BigWig will “bin” the data to show a continuous “bar-graph”-like figure across the genome.
FastQ	Text-based format for storing a DNA sequence with its corresponding quality scores
Genetrack index	(*.idx files) A tab-separated file that lists chromosome number (chr), genomic coordinates and the values on the forward or reverse strand. These files are used as inputs for peak calling programs.
GFF	A tab separated file with nine required fields: sequence name (chr), source (program generates this feature), features (name of this type of feature), start position, end position, score (between 0-100), strand (+/-), frame (reading frame, 0-2, for coding exon), group (all lines with same group linked).
SAM	Sequence Alignment/Map file. A compact representation of nucleotide sequence alignments which is index-able.

APPENDIX B

LIST OF ANALYSIS TOOLS

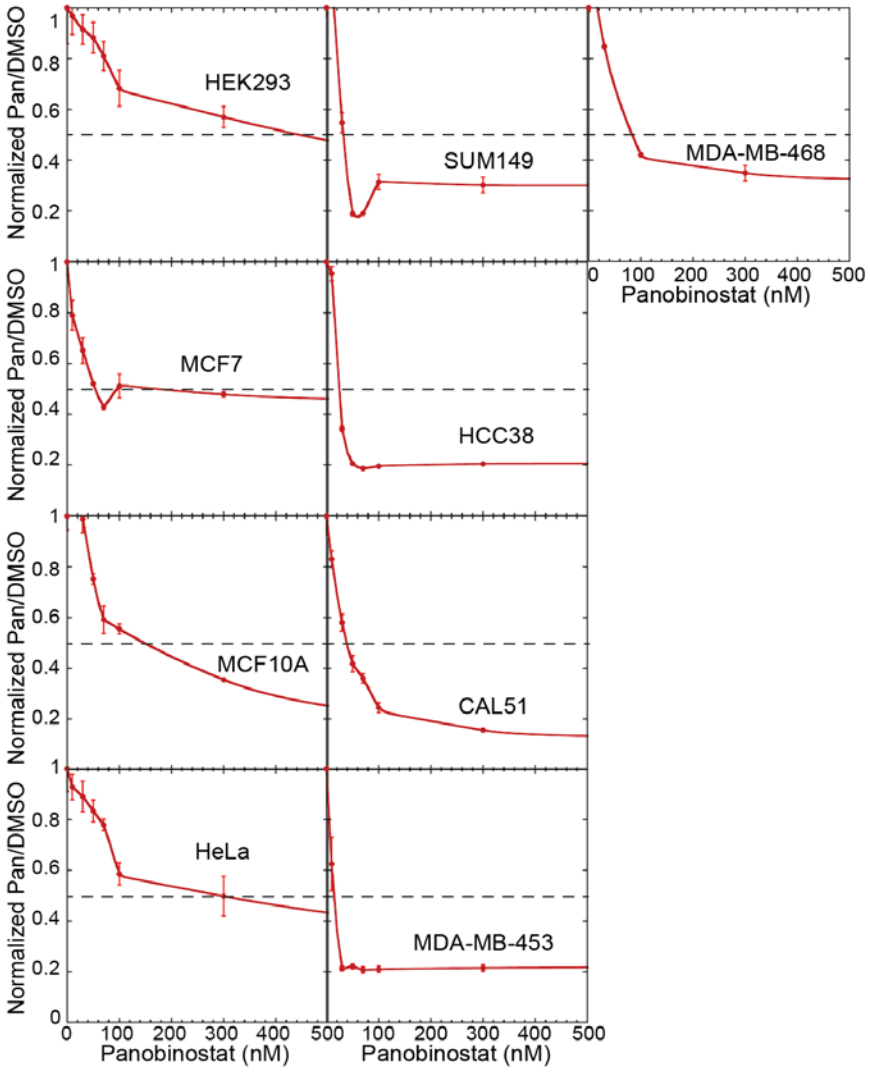
BedTools	This software suite covers a wide range of genomic analysis tools that allows you to merge, intersect, count, etc. genomic reads from a variety of file formats[45]. BedTools was used to merge our BAM files and to convert BAM files to BED files.
BWA	Burrows-Wheeler Aligner: software that maps low-divergent sequences to a large reference genome[42]. BWA was used to map libraries to the hg19 genome.
ChAsE	Software with interactive graphical interface to develop heat maps and clustering of genomic data[48]. ChAsE was used on our epigenetic data to provide heatmaps centered at TSS and TTS.
ChromHMM	An automated software program for learning chromatin states by observing combinations of chromatin modification patterns[27]. We used ChromHMM to uncover regulatory regions of the HCC1806 genome.
Cluster	This program performs cluster analysis on transcriptional data. We used k-means clustering on our data-sets[49]. Together, with TreeView, the programs are used for analyzing and visualizing the results of our PRO-seq and ChIP-exo experiments.
DAVID	Database for Annotation, Visualization and Integrated Discovery. It is a database with comprehensive set of functional annotation tools to examine

	biological meaning behind large datasets[51]. We used DAVID for gene ontology classifications.
edgeR	A bioconductor package used for differential expression analysis[47]. Used to find the differential expression of DMSO, 1hr and 4hr Panobinostat PRO-seq and H3K27Ac ChIP-exo libraries.
GeneTrack	Python software for data storage and analysis. GeneTrack fits the data (read positions and density) and generates peak predictions[46]. We used GeneTrack to call peaks for both ChIP-exo and PRO-seq libraries.
HOMER	Hypergenometric Optimization of Motif Enrichment. A suite of motif discovery software tools. HOMER encompasses programs to find peaks or transcript densities and then allows you to quantify these peaks in RPKM. While it does not encompass its own differential expression program, you can pipe the quantification from HOMER into edgeR for differential expression analysis[34]. We used HOMER to find RPKM for PRO-seq and ChIP-exo and to prepare the data for differential expression analysis of PRO-seq libraries and differential H3K27Ac occupancy.
IGV	Integrative Genomics Viewer. A visualization tool for exploring large genomic datasets[44]. We used IGV to create localized screen-shots at genes of interest.
Java TreeView	A program to allow interactive visualization of the Cluster results[50]. Used to visualize clusters from PRO-seq data, Pol II RPKMs and H3K27Ac peak data.

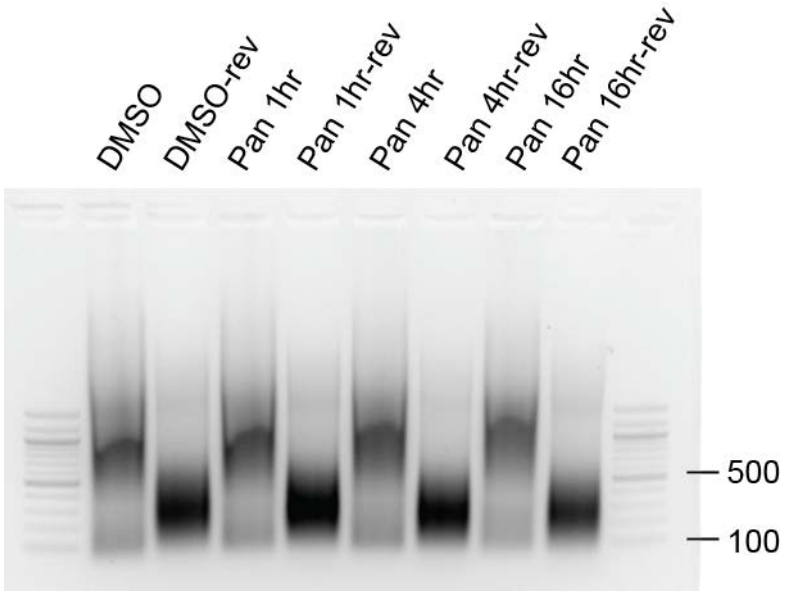
SAMtools	A software suite that provides tools to manipulate SAM and BAM files, such as sorting, indexing and merging[43]. We used SAMtools to sort and index BAM files.
Tabs2genetrack	A custom python script used to create the *.idx file from BED files. We used in the pre-processing stage prior to peak calling.
UNIX	Multitasking computer operating system which executes programs from multiple languages. We used UNIX to run all of the above programs, and we used command line to perform some extraneous modifications (see Methods).

APPENDIX C

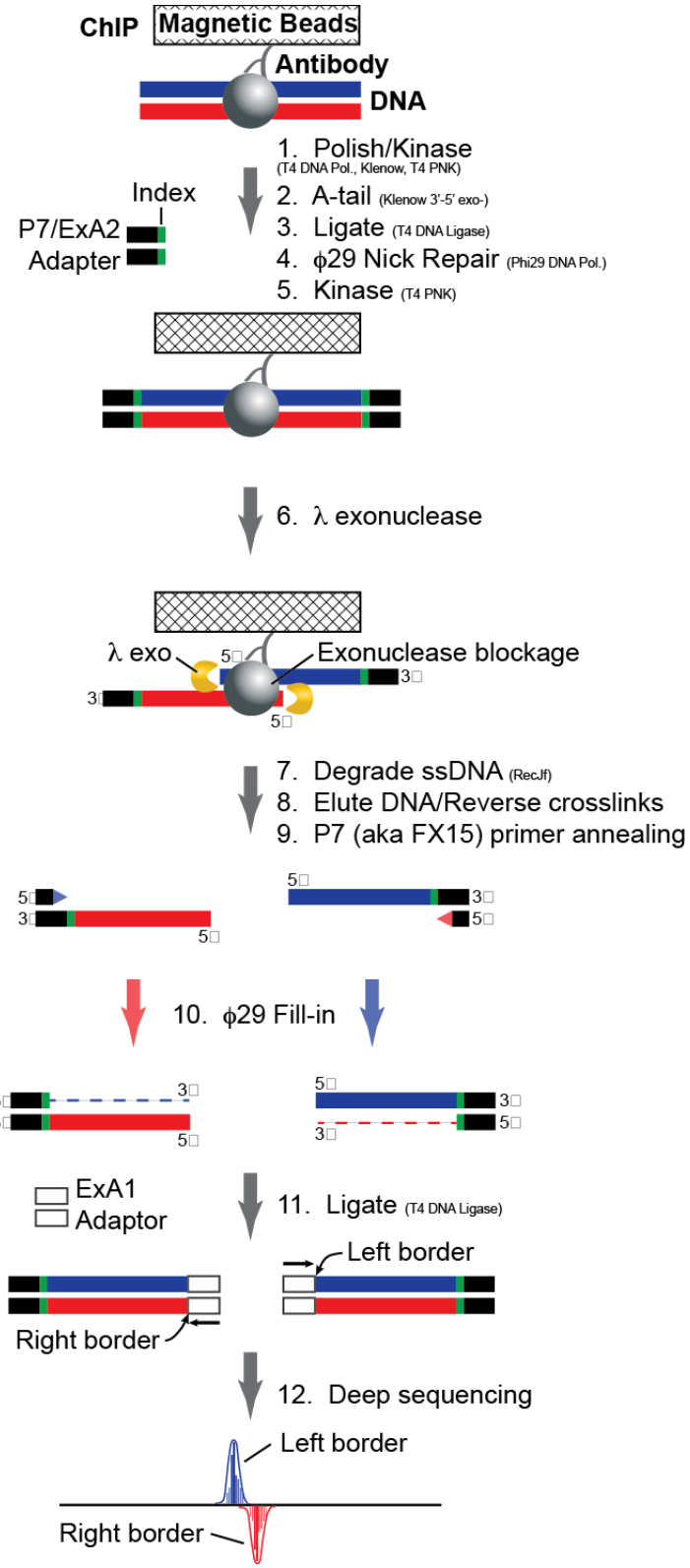
SUPPLEMENTAL FIGURES



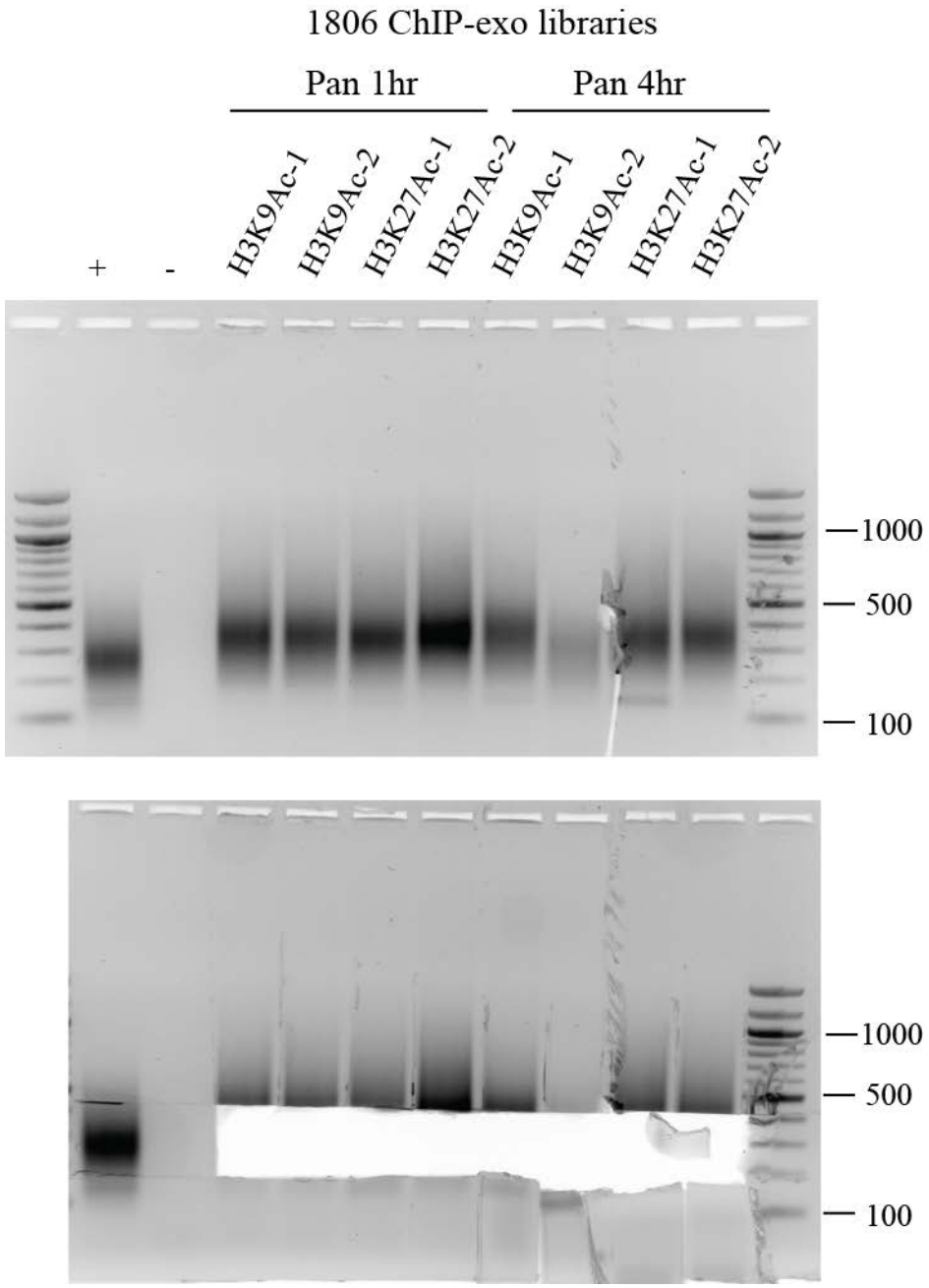
Supplemental Figure 1. Panobinostat Dose Response Curves. After 72 hours of treatment with varying ranges of Panobinostat concentrations, the cells were subjected to the MTT assay. The MTT absorbance for each Panobinostat treatment was normalized to the absorbance for cells treated with DMSO. This ratio was then plotted against Panobinostat concentration. The dashed line represents cell viability of 50%. The corresponding x-axis value (Panobinostat concentration) represents the IC_{50}. This assay was done in triplicate, and error bars represent the standard deviation.



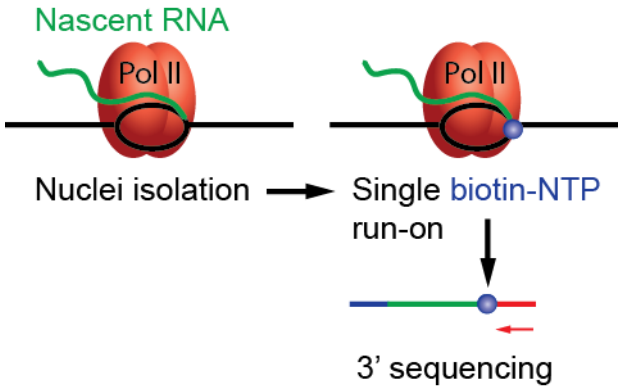
Supplemental Figure 2. Sonication Check for HCC1806. The HCC1806 cells were processed as described in the methods of Chapter III. After sonication, 10µl of the sample was removed and kept on ice while another 10µl was added to 1x the volume of TE-RNAse with Proteinase K. This later sample was heated at 37°C for 30 minutes to reverse the crosslinking (Rev samples). All samples were run on a 1.5% agarose gel at 200V for 45 minutes. The DMSO, Pan 1hr, Pan 4 hr, and Pan 16 hr samples refer to the treatment period of the cells prior to lysis and sonication. These cells were sonicated but did not undergo reversal of crosslinking. The DNA should run as a wide smear >500bp, because proteins are still cross-linked to the DNA. The DMSO rev, Pan 1hr rev, Pan 4 hr rev, and Pan 16 hr rev samples were the same samples as above, but were subjected to reverse crosslinking. These samples run in a smear between 100-500bp confirming that 20 min of sonication is appropriate to prepare DNA for ChIP-exo.



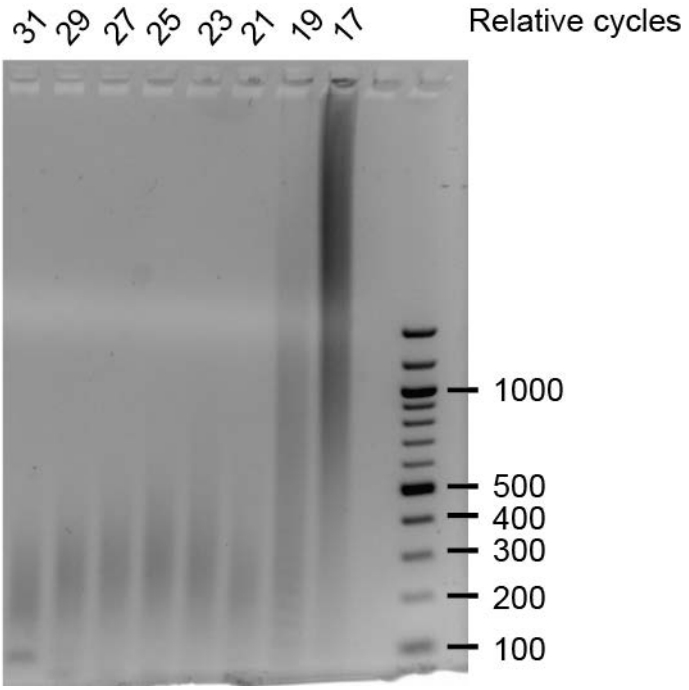
Supplemental Figure 3. Schematic of ChIP-exo. This schematic outlines the steps in the ChIP-exo library preparation.



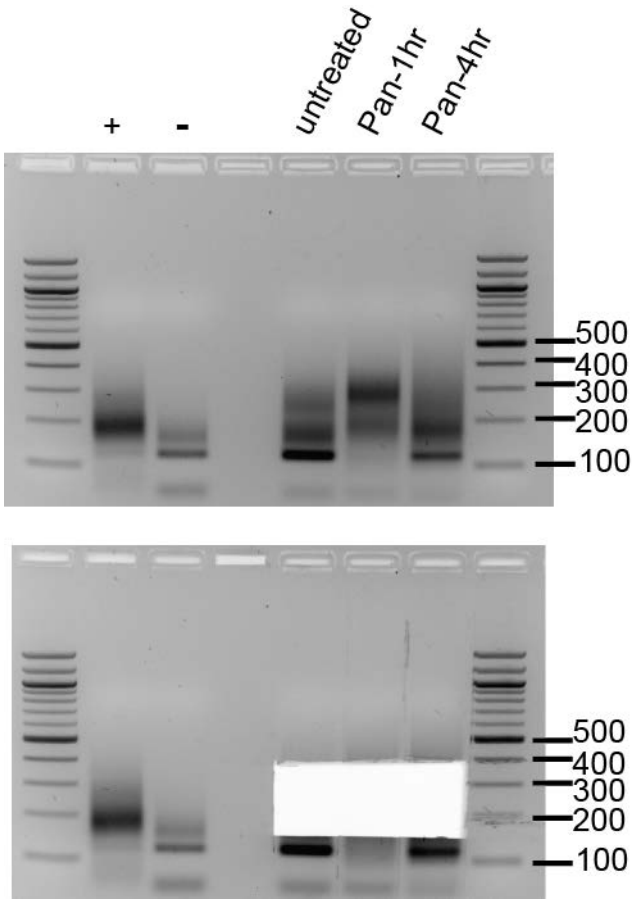
Supplemental Figure 4. ChIP-exo library excision for HCC1806. This is an example of a library test gel run on all of the ChIP-exo libraries to perform size exclusion for DNA fragments outside the 200-500bp. Agarose (1.5%) gels are run on libraries just after the PCR amplification. The smear between 200-500bp is physically excised from the gel, to remove larger DNA fragments and the adapter dimers (at 125bp). These smears then undergo gel extractions and are then quantified prior to submission to sequencing. The (+) control is a previous ChIP library and the (-) control is a no template PCR reaction. Each of the libraries (Pan 1hr H3K9Ac and H3K27Ac, and Pan 4hr H3K9Ac and H3K27Ac) were prepared in biological duplicates (1 and 2). Pan 4hr H3K9Ac-2 is an example of a library that has too low a smear density, so the quantification of the library is low and the quality of the library is also low.



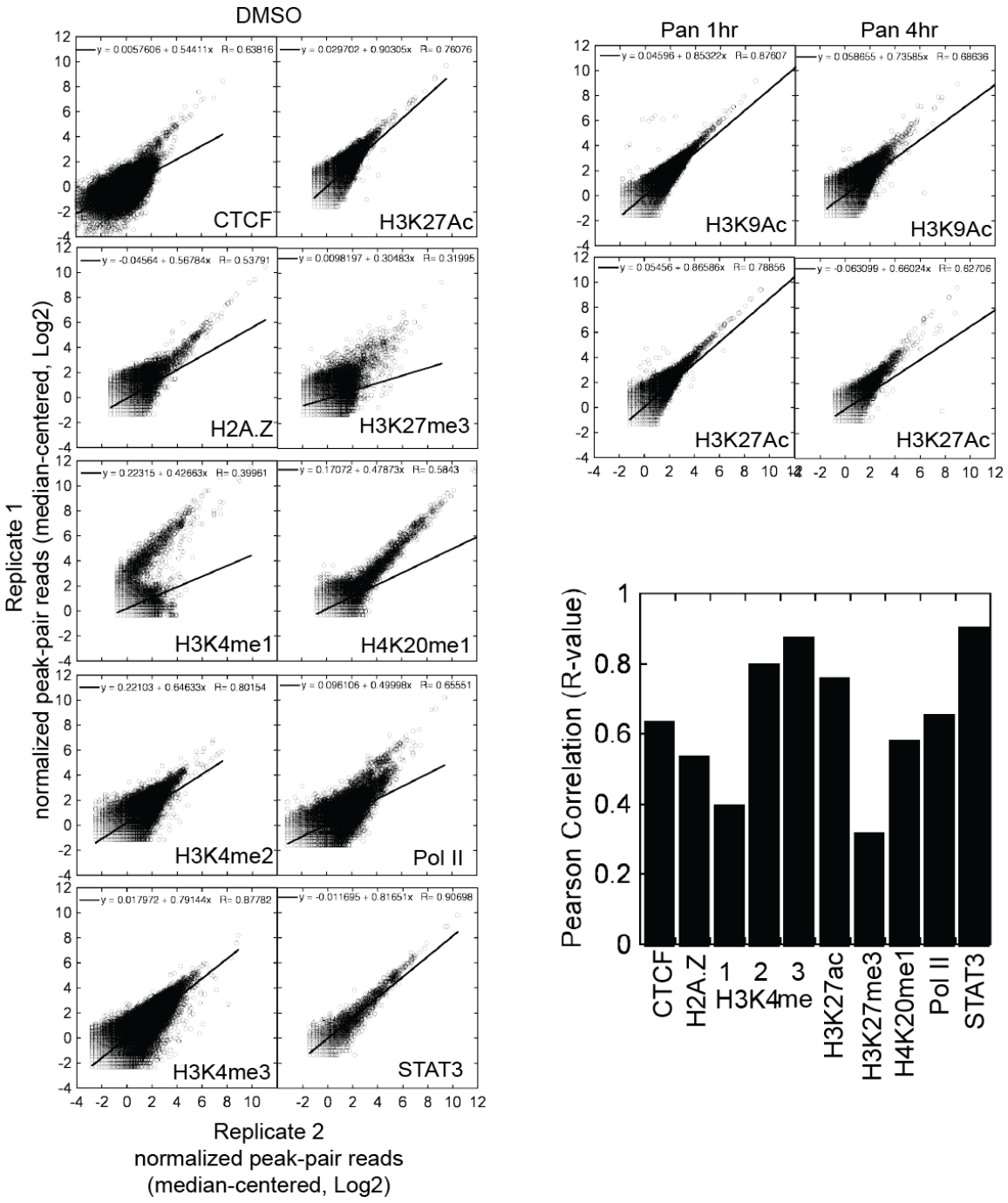
Supplemental Figure 5. PRO-seq summary. This schematic summarizes the PRO-seq protocol. Nuclei are isolated and biotin-NTP is added to the nuclei. The biotin-NTP is then incorporated into the nascent transcripts, and the biotin-RNA is affinity purified with streptavidin to prepare a library for deep sequencing.



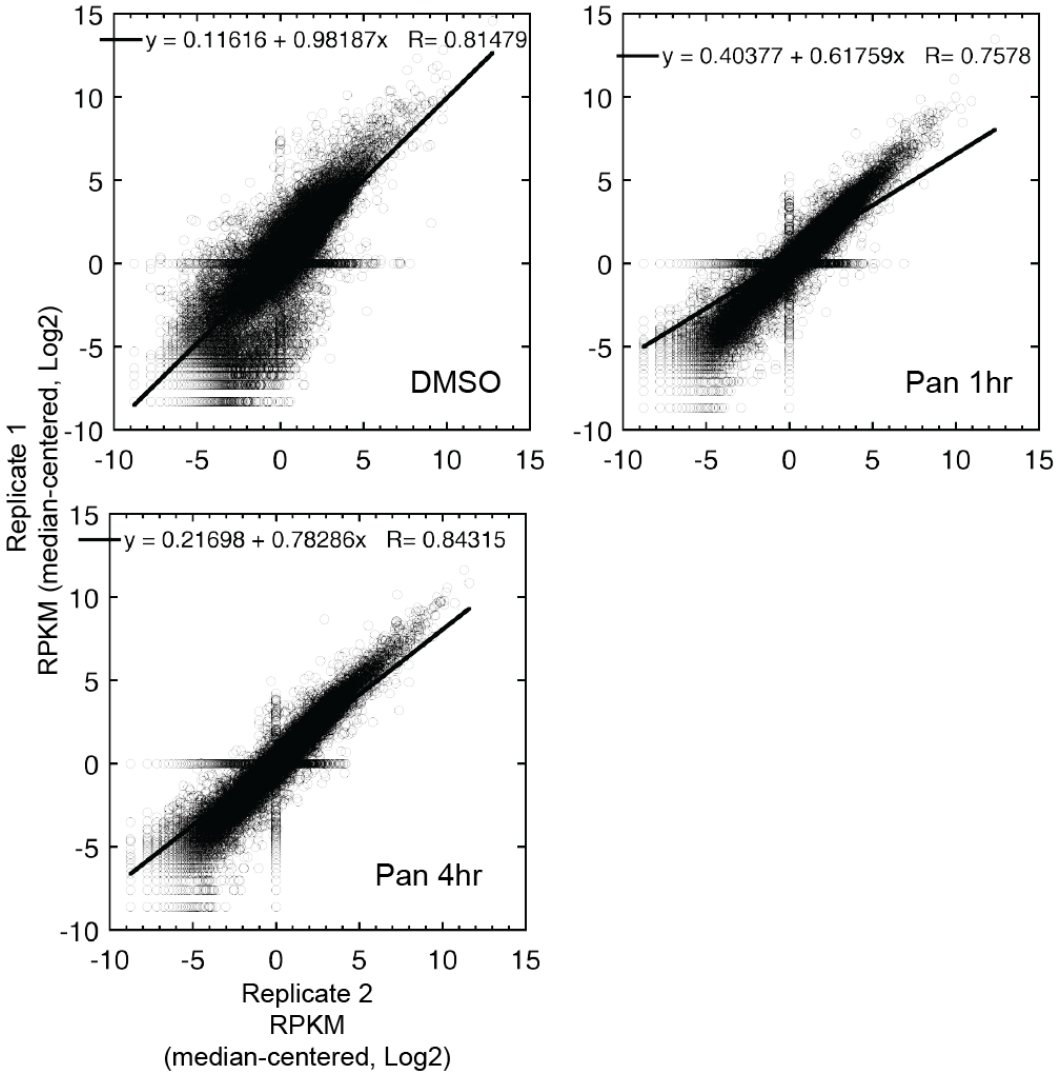
Supplemental Figure 6. HCC1806 PRO-seq Test Amplification. This is an example of the test amplification run on each PRO-seq library made. The purpose of the test amplification is to determine what PCR cycle is appropriate for full-scale amplification of each library (i.e. relative cycles), because over amplification can cause inaccurate read counts when sequenced. The samples are diluted in water and run for 35 cycles. Each dilution represents a step-wise cycle number for full scale amplification. The first dilution (or the most concentrated sample) is equivalent to 31 PCR cycles (Relative cycles), with each serial dilution representing 2 less cycles. The relative cycle that produces a smear between 200-500bp is the ideal cycle number for full scale. In each of the test amplifications, the lower dilutions showed smears the entire way up the gel, for unknown reasons. All libraries we sequenced for PRO-seq had ideal smears at 25-29 relative cycles. This sample was HCC1806 cells treated with Panobinostat for 16 hours.



Supplemental Figure 7. PRO-seq library check example for HCC1806. This gel was run on fully amplified PRO-seq libraries. In this gel the prepared libraries were HCC1806 untreated (not included in this thesis), Panobinostat 1hr treatment and Panobinostat 4 hr treatment. Library smears should run between 150-400 bp. The adapter dimer runs at 120bp. The 150-400bp smear was excised from the gel and subjected to gel extraction to prepare libraries for sequencing. The lower panel shows the gel post excision. The positive control is PCR that was run on a previous PRO-seq library. The negative control was a no template PCR.



Supplemental Figure 8. ChIP-exo Replicate Matching. Replicate matching was performed on each biological replicate as described in the ChIP-exo data processing Methods Section. Each replicate tallied the normalized peak paired reads and then plotted this log2 normalization against its corresponding biological replicate. A linear model was fit to the data, and the Pearson Correlation (R-value) for each was plotted in the bar graph above. A positive correlation was confirmed for each data set, so the data was merged and the merged data was used for all further analysis.



Supplemental Figure 9. PRO-seq Replicate Matching. Replicate matching was done on RPKM values for each of the PRO-seq replicates at each time point. A linear model was fit to the data and a Pearson-correlation (R-value) was found for each. Each of the R-values were greater than 0.75, so we believe the replicates were reproducible.

IP	Antibody	Method	Treatment	Rep1 Uniquely Mapped Reads	Rep1 Unique Mapping Rate	Rep2 Uniquely Mapped Reads	Rep2 Unique Mapping Rate
CTCF	Millipore 07-729	ChIP-exo	Vehicle (DMSO)	50,338,238	0.81	45,941,823	0.70
H2AZ	Millipore 07-594	ChIP-exo	Vehicle (DMSO)	30,152,515	0.72	18,871,682	0.76
H3K27ac	ab4729	ChIP-exo	Vehicle (DMSO)	16,499,902	0.51	29,929,229	0.68
H3K27me3	Millipore 07-449	ChIP-exo	Vehicle (DMSO)	28,166,034	0.76	31,048,885	0.83
H3K4me1	ab8895	ChIP-exo	Vehicle (DMSO)	30,710,984	0.70	26,865,944	0.72
H3K4me2	ab7766	ChIP-exo	Vehicle (DMSO)	21,002,075	0.75	56,272,491	0.76
H3K4me3	ab8580	ChIP-exo	Vehicle (DMSO)	29,367,448	0.79	57,633,130	0.75
H4K20me1	ab9051	ChIP-exo	Vehicle (DMSO)	58,682,522	0.74	34,349,834	0.77
Pol2	sc-899	ChIP-exo	Vehicle (DMSO)	59,331,303	0.72	23,255,073	0.73
Stat3	sc-482	ChIP-exo	Vehicle (DMSO)	37,062,439	0.76	11,765,498	0.73
--	--	PRO-seq	Vehicle (DMSO)	32,425,568	0.70	40,395,247	0.46
H3K9ac	ab4441	ChIP-exo	1hr Panobinostat	27,980,881	0.93	24,738,573	0.93
H3K27ac	ab4729	ChIP-exo	1hr Panobinostat	26,058,943	0.92	30,995,107	0.91
--	--	PRO-seq	1hr Panobinostat	36,292,690	0.68	54,559,196	0.64
H3K27ac	ab4729	ChIP-exo	4hr Panobinostat	24,341,541	0.93	26,516,559	0.92
H3K9ac	ab4441	ChIP-exo	4hr Panobinostat	23,767,030	0.92	22,673,957	0.91
--	--	PRO-seq	4hr Panobinostat	33,831,021	0.67	45,749,131	0.56
--	--	Input	DMSO/Pan	79,302,832	0.65	19,939,197	0.71

Supplemental Table 1. Summary of sequencing reads. The table gives an overview of the quality of the reads for each IP. The antibody, and experiment (both method and treatment are included). Then I include the number of uniquely mapped reads for each of the biological replicates, because we aimed for each library to have 20 Million reads. This helps us determine the quality of our library. I also displayed the uniquely mapped read rate, which is the number of uniquely mapped reads to the number of total reads. Mapping rates are an initial indicator of library quality.

REFERENCES

1 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2015. *CA Cancer J Clin* **65**, 5-29, doi:10.3322/caac.21254 (2015).

2 Tate, C. R. *et al.* Targeting triple-negative breast cancer cells with the histone deacetylase inhibitor panobinostat. *Breast Cancer Res* **14**, R79, doi:10.1186/bcr3192 (2012).

3 Anders, C. K. & Carey, L. A. Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer. *Clin Breast Cancer* **9 Suppl 2**, S73-81, doi:10.3816/CBC.2009.s.008 (2009).

4 Haffty, B. G. *et al.* Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. *J Clin Oncol* **24**, 5652-5657, doi:10.1200/JCO.2006.06.5664 (2006).

5 Chuang, H. C., Kapuriya, N., Kulp, S. K., Chen, C. S. & Shapiro, C. L. Differential anti-proliferative activities of poly(ADP-ribose) polymerase (PARP) inhibitors in triple-negative breast cancer cells. *Breast Cancer Res Treat* **134**, 649-659, doi:10.1007/s10549-012-2106-5 (2012).

6 Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* **121**, 2750-2767, doi:10.1172/JCI45014 (2011).

7 Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395-399, doi:10.1038/nature10933 (2012).

8 Sharma, S., Kelly, T. K. & Jones, P. A. Epigenetics in cancer. *Carcinogenesis* **31**, 27-36, doi:10.1093/carcin/bgp220 (2010).

9 Thiagalingam, S. *et al.* Histone deacetylases: unique players in shaping the epigenetic histone code. *Ann N Y Acad Sci* **983**, 84-100 (2003).

10 Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41-45, doi:10.1038/47412 (2000).

11 Gagniuc, P. & Ionescu-Tirgoviste, C. Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. *BMC Genomics* **13**, 512, doi:10.1186/1471-2164-13-512 (2012).

12 Blackwood, E. M. & Kadonaga, J. T. Going the distance: a current view of enhancer action. *Science* **281**, 60-63 (1998).

13 Liu, T., Kuljaca, S., Tee, A. & Marshall, G. M. Histone deacetylase inhibitors: multifunctional anticancer agents. *Cancer Treat Rev* **32**, 157-165, doi:10.1016/j.ctrv.2005.12.006 (2006).

14 Bartling, B. *et al.* Comparative application of antibody and gene array for expression profiling in human squamous cell lung carcinoma. *Lung Cancer* **49**, 145-154, doi:10.1016/j.lungcan.2005.02.006 (2005).

15 Minamiya, Y. *et al.* Expression of histone deacetylase 1 correlates with a poor prognosis in patients with adenocarcinoma of the lung. *Lung Cancer* **74**, 300-304, doi:10.1016/j.lungcan.2011.02.019 (2011).

16 Zhang, Z. *et al.* Quantitation of HDAC1 mRNA expression in invasive carcinoma of the breast*. *Breast Cancer Res Treat* **94**, 11-16, doi:10.1007/s10549-005-6001-1 (2005).

17 Huang, Y., Nayak, S., Jankowitz, R., Davidson, N. E. & Oesterreich, S. Epigenetics in breast cancer: what's new? *Breast Cancer Res* **13**, 225, doi:10.1186/bcr2925 (2011).

18 Anne, M., Sammartino, D., Barginear, M. F. & Budman, D. Profile of panobinostat and its potential for treatment in solid tumors: an update. *Onco Targets Ther* **6**, 1613-1624, doi:10.2147/OTT.S30773 (2013).

19 Weinberg, R. A. Tumor suppressor genes. *Science* **254**, 1138-1146 (1991).

20 Maziveyi, M. & Alahari, S. K. Breast Cancer Tumor Suppressors: A Special Emphasis on Novel Protein Nischarin. *Cancer Res* **75**, 4252-4259, doi:10.1158/0008-5472.CAN-15-1395 (2015).

21 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

22 Prince, H. M., Bishton, M. J. & Johnstone, R. W. Panobinostat (LBH589): a potent pan-deacetylase inhibitor with promising activity against hematologic and solid tumors. *Future Oncol* **5**, 601-612, doi:10.2217/fon.09.36 (2009).

23 Atadja, P. Development of the pan-DAC inhibitor panobinostat (LBH589): successes and challenges. *Cancer Lett* **280**, 233-241, doi:10.1016/j.canlet.2009.02.019 (2009).

24 Dokmanovic, M., Clarke, C. & Marks, P. A. Histone deacetylase inhibitors: overview and perspectives. *Mol Cancer Res* **5**, 981-989, doi:10.1158/1541-7786.MCR-07-0324 (2007).

25 Witt, O., Deubzer, H. E., Milde, T. & Oehme, I. HDAC family: What are the cancer relevant targets? *Cancer Lett* **277**, 8-21, doi:10.1016/j.canlet.2008.08.016 (2009).

26 Bose, P., Dai, Y. & Grant, S. Histone deacetylase inhibitor (HDACI) mechanisms of action: emerging insights. *Pharmacol Ther* **143**, 323-336, doi:10.1016/j.pharmthera.2014.04.004 (2014).

27 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).

28 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49, doi:10.1038/nature09906 (2011).

29 Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**, 817-825, doi:10.1038/nbt.1662 (2010).

30 Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813-1831, doi:10.1101/gr.136184.111 (2012).

31 Rhee, H. S. & Pugh, B. F. ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr Protoc Mol Biol* **Chapter 21**, Unit 21 24, doi:10.1002/0471142727.mb2124s100 (2012).

32 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

33 Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897-903, doi:10.1038/ng.154 (2008).

34 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

35 Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-1848, doi:10.1126/science.1162228 (2008).

36 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

37 Schmidt, D. *et al.* ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240-248, doi:10.1016/j.ymeth.2009.03.001 (2009).

38 Serandour, A. A., Brown, G. D., Cohen, J. D. & Carroll, J. S. Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties. *Genome Biol* **14**, R147, doi:10.1186/gb-2013-14-12-r147 (2013).

39 Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950-953, doi:10.1126/science.1229386 (2013).

40 Dvir, A. Promoter escape by RNA polymerase II. *Biochim Biophys Acta* **1577**, 208-223 (2002).

41 Green, N. M. Avidin. *Adv Protein Chem* **29**, 85-133 (1975).

42 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

43 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

44 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

45 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**, 11 12 11-11 12 34, doi:10.1002/0471250953.bi1112s47 (2014).

46 Albert, I., Wachi, S., Jiang, C. & Pugh, B. F. GeneTrack--a genomic data processing and visualization framework. *Bioinformatics* **24**, 1305-1306, doi:10.1093/bioinformatics/btn119 (2008).

47 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

48 Younesy, H. N., C.B.: Moller, T.; Alder, O.; Cullum, R.; Lorincz, M.C.; Karimi, M.M; Jones, S.J.M. An Interactive Analysis and Exploration Tool for Epigenomic Data. *Computer Graphics Forum* **32**, 91-100, doi:10.1111/cgf.12096 (2013).

49 de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453-1454, doi:10.1093/bioinformatics/bth078 (2004).

50 Saldanha, A. J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246-3248, doi:10.1093/bioinformatics/bth349 (2004).

51 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

52 Hon, G. C., Hawkins, R. D. & Ren, B. Predictive chromatin signatures in the mammalian genome. *Hum Mol Genet* **18**, R195-201, doi:10.1093/hmg/ddp409 (2009).

53 Fan, J. Y., Gordon, F., Luger, K., Hansen, J. C. & Tremethick, D. J. The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. *Nat Struct Biol* **9**, 172-176, doi:10.1038/nsb767 (2002).

54 Burgess-Beusse, B. *et al.* The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci U S A* **99 Suppl 4**, 16433-16437, doi:10.1073/pnas.162342499 (2002).

55 Kireeva, M. L. *et al.* Nature of the nucleosomal barrier to RNA polymerase II. *Mol Cell* **18**, 97-108, doi:10.1016/j.molcel.2005.02.027 (2005).

56 Tang, Y., Shu, G., Yuan, X., Jing, N. & Song, J. FOXA2 functions as a suppressor of tumor metastasis by inhibition of epithelial-to-mesenchymal transition in human lung cancers. *Cell Res* **21**, 316-326, doi:10.1038/cr.2010.126 (2011).

57 Miyamoto, K. *et al.* Identification of 20 genes aberrantly methylated in human breast cancers. *Int J Cancer* **116**, 407-414, doi:10.1002/ijc.21054 (2005).

58 Rivera, C. M. & Ren, B. Mapping human epigenomes. *Cell* **155**, 39-55, doi:10.1016/j.cell.2013.09.011 (2013).

59 Sandelin, A. *et al.* Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet* **8**, 424-436, doi:10.1038/nrg2026 (2007).

60 Lauberth, S. M. *et al.* H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021-1036, doi:10.1016/j.cell.2013.01.052 (2013).

61 Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58-69, doi:10.1016/j.cell.2007.08.016 (2007).

62 Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**, 825-837, doi:10.1016/j.molcel.2013.01.038 (2013).

63 Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112, doi:10.1038/nature07829 (2009).

64 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318, doi:10.1038/ng1966 (2007).

65 Ren, B. Transcription: Enhancers make non-coding RNA. *Nature* **465**, 173-174, doi:10.1038/465173a (2010).

66 Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837, doi:10.1016/j.cell.2007.05.009 (2007).

67 Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110-113, doi:10.1126/science.1060118 (2001).

68 Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194-1211, doi:10.1016/j.cell.2009.06.001 (2009).

69 Young, M. D. *et al.* ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res* **39**, 7415-7427, doi:10.1093/nar/gkr416 (2011).

70 Benayoun, B. A. *et al.* H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* **158**, 673-688, doi:10.1016/j.cell.2014.06.027 (2014).

71 Rubinstein, M. & de Souza, F. S. Evolution of transcriptional enhancers and animal diversity. *Philos Trans R Soc Lond B Biol Sci* **368**, 20130017, doi:10.1098/rstb.2013.0017 (2013).

72 Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948-951, doi:10.1038/nature06947 (2008).

73 Kim, Y. J. *et al.* HDAC inhibitors induce transcriptional repression of high copy number genes in breast cancer through elongation blockade. *Oncogene* **32**, 2828-2835, doi:10.1038/onc.2013.32 (2013).

74 Ouelaa-Benslama, R. & Emami, S. Pinworm and TNKS inhibitors, an eccentric duo to derail the oncogenic WNT pathway. *Clin Res Hepatol Gastroenterol* **35**, 534-538, doi:10.1016/j.clinre.2011.03.015 (2011).

75 Polakis, P. Wnt signaling in cancer. *Cold Spring Harb Perspect Biol* **4**, doi:10.1101/cshperspect.a008052 (2012).

76 Dhillon, A. S., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26**, 3279-3290, doi:10.1038/sj.onc.1210421 (2007).

77 Smith, M. L. & Kumar, M. A. The "Two faces" of Tumor Suppressor p53-revisited. *Mol Cell Pharmacol* **2**, 117-119 (2010).

78 Cannaerts, E., van de Beek, G., Verstraeten, A., Van Laer, L. & Loeys, B. TGF-beta signalopathies as a paradigm for translational medicine. *Eur J Med Genet*, doi:10.1016/j.ejmg.2015.10.010 (2015).

79 Shea, M., Costa, D. B. & Rangachari, D. Management of advanced non-small cell lung cancers with known mutations or rearrangements: latest evidence and treatment approaches. *Ther Adv Respir Dis*, doi:10.1177/1753465815617871 (2015).

80 Zhao, M., Sun, J. & Zhao, Z. TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res* **41**, D970-976, doi:10.1093/nar/gks937 (2013).

81 Kubosaki, A. *et al.* Genome-wide investigation of in vivo EGR-1 binding sites in monocytic differentiation. *Genome Biol* **10**, R41, doi:10.1186/gb-2009-10-4-r41 (2009).