Mitochondrial haplotype affects histone modification profiles in MNX mice

by

John T. McGuire

Submitted to the graduate degree program in Cancer Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science in Cancer Biology.

_____

Chair: Danny R. Welch, PhD

_____

Animesh Dhar, PhD

_____

Luciano DiTacchio, PhD

Date of Defense: April 26th, 2018

The thesis committee for John T. McGuire certifies that this is the approved version of the following thesis:

Mitochondrial haplotype affects histone modification profiles in MNX mice

Chair: Danny R. Welch, Ph.D.

Date Approved: May 7th, 2018

Abstract

Cancer is a growing problem. In the US, cancer is roughly tied with cardiac disease as the leading cause of death. Intriguingly, compared to the proportion of deaths attributable to primary tumors, roughly 90% of all cancer patient morbidity and mortality are a direct result of metastatic disease perturbing normal organ and system function. Research has shown that the propensity to form both spontaneous tumors and metastatic growths is a function of both the nuclear and mitochondrial genomes present within an organism. Specifically, studies have also shown that, when the nuclear background is held constant and the mitochondrial haplotype is varied, there are marked discrepancies in both tumor latency and the average number of metastatic growths observed. To further study the effects of varying the mitochondrial haplotype on cancer progression and metastasis, a unique in vivo model—known as Mitochondrial-Nuclear Exchange (MNX) mice—has been developed. Experiments using this model system further support the notion that mitochondrial haplotype can contribute significantly to tumor latency and the overall of metastatic growths. As an initial line of inquiry, the question was asked as to whether or not differences in gene expression could be present upon changing the mitochondrial haplotype. Given that clear differences in genome-wide expression profiles between wildtype and MNX mice were indeed observed, the next line of inquiry sought to determine whether or not DNA methylation profiles could be regulating the differences in expression. Likewise, significant differences in DNA methylation patterns were established. As a corollary to these findings, the subsequent line of inquiry sought to elucidate whether or not mitochondrial haplotype could influence other potential epigenetic mechanisms

responsible for the differences in gene expression profiles. Specifically, given the known link between metabolism and epigenetics, the logical hypothesis of the experiment was simply that the mitochondrial haplotype could potentially able to directly influence the epigenetic process of post-translationally modifying the N-terminal tails of histones in order to affect chromatin state dynamics in the cell. To address this hypothesis, following rigorous antibody validation, a ChIP-Seq experiment comparing wildtype and MNX mice using antibodies against four known histone modifications (H3K4me1, H3K4me3, H3K27me3, and H3K27ac) was conducted. Post-sequencing analysis indicated a selective difference in the overall profile of differentially-bound, post-translationally modified histones, thus supporting the hypothesis that mitochondrial haplotype does directly contribute to overall histone modification profiles. As a next logical step, future work will involve integrating previously generated RNA-Seq, Methyl-Seq, and ChIP-Seq experiments (in combination with extensive metabolome data) to develop hypotheses as to possible causal mechanisms involved in driving the observed differences in tumor latency and metastatic growth. Furthermore, another direction could involve expanding the ChIP-Seq study to include newly characterized acylation (propionylation, butyrylation, succinylation, malonylation, etc.) post-translational modifications which functionally are thought to overlap with acetylation and may more closely link metabolism to overall chromatin state. The results of this research not only contribute to our understanding of how mitochondria participate in influencing the DNA packaging mechanisms involved in regulating chromatin dynamics, but, as our research is not limited to any single form of cancer, has the potential to benefit to all cancer patients.

Table of Contents

Introduction

On the microscopic, molecular level, cancer evolution is a beautifully Darwinian process [1]. Cells that sustain growth and survival advantages are able to generate more progeny with the same characteristics (e.g., deregulated cell cycle control or resistance to cell death and senescence mechanisms. The result is ultimately a clonal expansion with either a few or—often—many heterogeneous sub-clonal populations comprising a tumor and simultaneously co-evolving [2]. On the macroscopic level, the uncontrolled expansion of cancer cells can overtake and replace normal, healthy cells and tissues with non-functioning, cancerous equivalents. The ensuing tissue damage can eventually instigate organ malfunction and endanger the patient if left untreated. If the cancer spreads to other distant organs or anatomical locations, the implications for the outcome of the patient are usually severe [3].

In the US alone, cancer is roughly tied with cardiac disease as the leading cause of death [4]. About 40% (or 2 in 5 individuals) of the total population will develop some form of cancer of the course of their lives [5]. While death rates due to cancer have been consistently decreasing over the past few decades for most forms of cancer, the lifetime risk of dying to cancer presently stands at 20%, (or 1 in 5 individuals) [5]. Intriguingly, relatively few cancer patients succumb to their primary tumors. Many neoplasms, if detected early in their benign and in situ forms, are usually able to be readily excised or otherwise treated to produce an essentially disease-free state. As such, only around one in ten patient deaths are directly attributable to primary neoplasia-induced complications. Unfortunately, many types of cancer are asymptomatic or have vague symptoms and are not detected until the cancer has

matured to a more advanced, malignant state capable of invading other tissues and metastasizing to distant sites within the body. Metastatic disease is extremely difficult to treat and cure. Due to increasing genomic instability and selective pressures exerted by prophylaxis (such as chemotherapeutic regimens), cell populations within secondary tumors are able to evolve rapidly and are likely exhibit resistance to treatments (due to surviving exposure to initial chemotherapeutics or possibly due to sustaining new mutations effectively conferring drug resistance). Compared to the proportion of deaths attributable to primary tumors, roughly 90% of all cancer patient deaths are directly a result of metastatic disease perturbing normal organ and system function [6]. Not all primary tumors have the same likelihood to metastasize, however [7].

Indeed, some forms of primary cancer rarely metastasize. Basal cell carcinomas, for example, generally do not spread beyond the patient's skin [8] and gliomas usually remain contained within the brain and CNS (although aberrant neoplasia in the brain—even without distant spreading—overgrows regular, healthy neighboring tissue, interferes with normal function, and can still threaten the patient's life) [9]. On the other hand, while most cancers can metastasize, certain types of cancer are known to metastasize at very high rates (such as melanomas) [7]. The tissue of origin of the primary tumor can also have a significant impact on the location(s) where metastatic disease may likely form in a patient [10]. Some high-grade primary cancers (such as breast cancer, lung cancer, and melanomas) are able to establish distant colonies in a very promiscuous manner (i.e., disease can spread to many host tissues and organs). Other certain types of advanced primary neoplasia appear to exhibit "metastatic tropism", whereby metastatic colonies form in organs and locations in a somewhat

predictable manner [11]. For example, metastatic disease originating from prostate cancer often colonizes bones, primary colorectal cancer tends to selectively form secondary growths in the liver, and in the event primary lymphoma in the lymph system and blood manages to form metastatic growths in the bones, liver, or brain [11]. The precise mechanisms and events required for cells to be able to undergo the process of metastasis are still not well understood. For instance: why do macroscopic primary tumors, which can shed millions of transformed cells into the blood stream each day [12], often form only relatively few metastatic growths [13] ?

To address such a question, several general processes involved in enabling a benign neoplastic growth to transform into distant, malignant disease have been described. As over 80% of malignant cancer involves carcinomas of epithelial origin [14], the processes describing the complex invasion-metastasis cascade will largely be in reference to this class of tumors.

In brief, resident cancer cells comprising an *in situ* primary tumor must locally invade by degrading and ultimately penetrating the basal lamina, a specialized proteinaceous extracellular matrix physically separating the epithelium from the stroma [15]. Upon invading the stroma itself, cancer cells may have the opportunity to enter blood or lymph system vasculature via the process of intravasation (possibly through interactions with macrophages in order to pass by endothelial cells of vessel walls) [15]. Once in circulation, cancer cells must overcome significant challenges, including evading anchorage-dependent programmed cell death (anoikis), being able to grow and proliferate without supportive signals from stromal constituents, and simply not being killed by the immune system or torn to shreds by hydrodynamic forces experienced

while circulating [15]. To increase their odds of survival, circulating cancer cells can associate with a sheath of blood platelets [16]. Although, despite the presence of CTCs in the blood, very few of these blood-borne pioneers manage to successfully establish secondary growths at distant sites. One significant method (apart from interactions with cell surface adherence receptors) through which cells can begin to extravasate (exit) the bloodstream is through being trapped within capillaries of organs [15]. For example, many capillaries range between 3 to 8μm wide, whereas circulating cancer cells in the blood are often 20μm or more [17]. Once trapped in the vessel, the circulating cancer cell must penetrate through the endothelial wall. Similar to how the cancer cells initially intravasated, cancer cells can extravasate through association with a macrophages or they can even begin to proliferate within the lumen of the vessel itself in order simply break through the endothelial wall [15]. The final step of the metastatic cascade is the successful colonization of distant tissues and organs [15]. Even though cancer cells may survive the journey to the foreign site, the new local environment will be very different from the environment of origin [18]. Many of the supporting and growth factors present within the primary tumor will likely be absent at the new location. As such, most micrometastatic growth are able to maintain only a state of stasis and never grow to become readily-observable neoplasms [19]. Thankfully, the overall rate of metastatic efficiency is very low and few micrometastases are able to exit their essentially proliferatively-senescent states in order to form macrometastases [15].

As one might imagine, there are certain genes which can act to promote specific steps of the metastatic cascade. For instance, downregulation of E-Cadherin contributes to a shift from an epithelial (strong Upregulation of ECM-modifying proteins

(such as MMPs, collagenases, and other proteases) can actively promote degradation of the basal lamina and autocrine expression of TFG-β enables cancer cells to proliferate in a stroma-independent manner. Given that there are a set of genes that can promote metastasis, there have likewise been a set of genes discovered that have been shown to actively suppress metastatic spread [20]. In a manner similar to that of tumor suppressor genes, metastasis suppressor genes act to prevent the occurrence of metastatic disease, albeit without having any significant effect on the growth of the primary tumor itself. Many such genes were discovered by observing which genes were upregulated/actively expressed in primary tumors and downregulated/repressed in secondary growth. Some metastasis suppressor genes have clear ties to specific steps in the metastatic cascade, whereas the mechanisms of other metastasis suppressor genes are decidedly less clear [21]. Several putative metastasis suppressor genes such as KISS1, BRMS1, and ITIH5 (among several others) are being evaluated through careful investigation [22-25].

A keystone paper from Kent Hunter in 2006 revealed that the host genetic contribution to metastatic cancer formation may be significantly more complex than first meets the eye [26]. In his paper, Hunter showed that the genetic background of the host can dramatically affect the capacity to form metastatic disease. In the experiment, Hunter bred male FVB/NJ-TgN(MMTV-PyMT)^{634Mul} mice (transgenic mice expressing the polyoma middle-T antigen), which present with spontaneous tumors within 60 days and roughly 90% of the mice form secondary growth in their lungs within 100 days, with female mice of several different inbred stains. The result was that mice formed pulmonary metastases to significantly varying degrees. The straight-forward explanation

for this observation would be that the differences were simply due to the nature of the nuclear genome (strain-specific mutations; epigenetic dissimilarities) and this hypothesis was partly proven due to the discovery of distinct metastasis suppressor genes. What Hunter did not evaluate was the possibility that the observed differences could have been due to differences in the maternally-inherited mitochondrial haplotypes of the mice (although, in his 2018 paper, Hunter does recognize recent findings in the Welch lab which show the contribution of mitochondria and metabolism to the formation of metastatic disease) [27].

Mitochondria are widely known as the “powerhouses” of the cell, as they generate most of the intracellular ATP (in addition to having roles in cell death). Embedded within the inner membrane of the mitochondrion are members of the electron transport chain family, which accept electrons from donor molecules (such as succinate, NADH, and $FADH_2$) and ultimately reduce oxygen to water. Critical in the process is that the energy released through the incremental passage of electrons is used to produce a proton gradient in the intermembrane space (essentially pumping protons again their concentration gradient). ATP Synthase (another protein in the inner mitochondrial membrane) takes advantage of the proton gradient and energetically favorable tendency to move from high to low proton concentrations in order to catalyze the synthesis of ATP, one of the most important chemical energy sources in the cell. In order to generate the electron donors necessary for electron transport and efficient aerobic respiration, however, the citric acid cycle (also known as the tricarboxylic acid cycle or the Krebs cycle) is critical. The TCA cycle occurs within the mitochondrial matrix (innermost compartment) and oxidizes acetyl-CoA through a chain of reactions to

form electron donors (such as NADH and $FADH_2$) and metabolite precursors for amino acids and other biomolecules. Acetyl-CoA can be generated primary through breakdown of carbohydrates (via glycolysis), lipids (via β-oxidation), and ketone bodies. In the event there is insufficient oxygen available to receive electrons from the electron transport chain, anaerobic respiration increasingly attempts to use pyruvate as an electron acceptor in lieu of oxygen (to regenerate NAD^+ from NADH) and produces lactic acid, ethanol, or other acids/alcohols. Fermentation is, however, a very inefficient process for producing NAD^+ and ATP. Aerobic metabolism in the mitochondrion is roughly 15 times more efficient than fermentation in terms of overall ATP generation. As is evident, mitochondria are referred to as “powerhouses” for good reason.

The presence of mitochondria likely arose in the eukaryote cell roughly 1.5 billion years ago through an endosymbiotic relationship in which a prokaryote was engulfed by a eukaryotic cell and managed to survive inside the eukaryote’s cytoplasm [28]. Since then, mitochondria have closely coevolved with eukaryotes, wherein many mitochondrial genes responsible for the electron transport complex have migrated to the nuclear genome [29]. However, critical components of the electron transport chain have remained under the control of the mitochondrion (within the mitochondrial genome) and mitochondria replicate through binary fission (a method very similar to prokaryotic division) [30]. Instead of dividing in sync with the cell cycle, mitochondria divide as per the situational metabolic demands of the cell. As one might imagine, the metabolic demand of a cell depends on the function of the specific tissue in which the cell finds itself. For example, heart, muscles, brain, kidneys, and liver cells have high metabolic/energy demand and, accordingly, high cell proportions of mitochondria [31].

The number of distinct mitochondria (although complex to count due to the nature of fused and non-fused mitochondria [32]) can range anywhere from the tens to several thousand [33].

The mitochondrial genome, by itself, is rather small and unassuming [34]. In human mitochondria, a circular dsDNA molecule comprises the mitochondrial DNA (mtDNA) and is roughly 16,000bp in length. The mtDNA codes for 37 unique genes, specifically 13 respiratory complex subunit genes (I,III, IV, V), 22 mitochondrial tRNA, and two rRNA. The haploid human genome, as a comparison, is over 3 billion bp in length. Given that each individual mitochondrion contains between two and ten copies of its genome and that there are many mitochondria per cell, the “effective” size of the entire collection of mtDNA can surprising (sometimes approaching 4% of the genetic material in the cell) [35]. However, unlike the nuclear genome, mitochondria (and hence the mtDNA) are entirely inherited from the oocyte of the mother [36, 37], while the paternal mitochondria present in the sperm are hypothesized to be degraded via ubiquitination after fertilization [38]. This process allows maternal ancestry to be inferred, somewhat analogous to how the Y chromosome can be used to trace the paternal lineage.

As one might imagine, several human diseases result from damaged or dysfunctional mitochondria. Of the diseases that do arise, roughly 15% are a result of mutations in the mtDNA itself (whereas the other 85% of mitochondrial diseases stem from mutations to metabolic genes in the nuclear genome) [39]. Symptoms of malfunctioning mitochondria essentially run a very wide gamut. Everything from muscle, vision, auditory, nerve, and internal organ function (such as the liver/kidneys/heart) can

all be perturbed. Specific examples of mitochondrial diseases include Leigh disease (neurologic and system-wide failure due to metabolic stress) [40], Pyruvate Dehydrogenase Complex Deficiency (lethargy, respiratory failure, coma, ataxia, neuropathy) [41], and Pearson Syndrome (anemia/neutropenia/thrombocytopenia due to affected hematopoietic stem cells, insufficient pancreatic function, general systemic failure) [42]. Clearly, having enough available energy and functioning metabolic pathways is imperative to good health. If parts of mitochondria-mediated metabolism are not working normally, sever systemic disease can result.

Cancer cells exhibit their own unique form of energy metabolism. As reported by the famous Otto Warburg in 1924, even when cancer cells have access to sufficient oxygen, often the cells do not readily undergo oxidative phosphorylation—instead preferring to generate lactic acid via strongly upregulated glycolysis (often despite having normally functioning respiration components) [43]. Energetically, the process of aerobic glycolysis is a poor trade off in efficiency when comparing ATP generated by the two processes (2 ATP vs 30-36 ATP). In order to feed their hunger for glycolysis, cancer cells import very high quantities of glucose, which is often made possible through upregulated glucose transporters in the cell plasma membrane. A possible explanation for such a dramatic shift in metabolism could be that the intermediate metabolites generated through glycolysis could be used to fuel anabolic reactions and drive proliferation. Another possible explanation stems from possible competition between tumor cells and T immune cells. As a result of tumor cells consuming/outcompeting T cells for glucose, T cells are effectively weakened and exhibit decreased mTOR signaling, IFN production, and inhibits the production of anti-

tumor cytokines involved in tumor clearance [44]. That being said, Warburg's observations were fairly large generalizations. Not all cancer types exhibit Warburg metabolism (due to possible metabolic heterogeneity) [45] and research suggests that cancer cells retain their ability to both Warburg metabolism and oxidative phosphorylation in order to increase adaptability [46]. Unfortunately, the role of mitochondria in the contribution to cancer is complex and often complicated by environmental dynamics, tissue-specific factors, and genetics.

One well-studied method through which mitochondrial energy metabolism can influence cancer is the generation of reactive oxygen species (ROS) [47]. During the process of oxidative phosphorylation, oxygen receives electrons in a controlled manner in order to be reduced to H_2O. The OxPhos system, however, is by no means perfect and normally exhibits a low rate of leak, wherein a small percentage (< 2%) of electrons incompletely reduce oxygen and form $^{\bullet}O_2^-$, superoxide, a radical which can deleteriously react with other biomolecules in the cell. Fortunately, cells have enzymes which act as an antioxidant defense system and catalyze the conversion of ROS radicals to less reactive molecules. For example, the enzyme Superoxide Dismutase can convert superoxide into hydrogen peroxide and oxygen. Hydrogen peroxide itself is also a member of the ROS family and the Catalase enzyme (among a few others) can actively convert hydrogen peroxide into water and oxygen. Outside of normal homeostasis, high levels of ROS and the associated oxidative stress can be detrimental to the integrity of the cell and ultimately result in cell death [48]. Low levels of ROS, on the other hand, can be involved in potentiating proliferation and survival [48]. Cancer cells often exhibit elevated ROS levels and are reliant on NAPDH-dependent antioxidant activity via

reduced glutathione and thioredoxin. Interestingly, many chemo- and radiotherapeutic regimens actually function by further increasing intracellular ROS levels (rather than by decreasing ROS) [49]. Normal cells have lower basal ROS levels and can survive increased oxidative stress, whereas cancer cells often cannot [48, 50, 51].

In the context of metastasis, the role is ROS is less clear and is dependent on the stage of disease development [52]. At low levels, mitochondrial ROS act as an intracellular, retrograde signaling mechanism to promote cancer survival and proliferation through activation of ROS-dependent growth factors [53]. ROS have also been linked to chronic inflammation by contributing to the microenvironment necessary to support neoplasia. On the other hand, high levels of ROS can contribute to initiation of cancer and disease progression [53]. For example, interactions of misregulated ROS with biomolecules (such as DNA) can result in DNA lesions and double-strand breaks, thus potentially contributing to inactivation of tumor suppressors or oncogene activation. ROS also play key roles at each step of the metastatic cascade, which has been described in the literature [54].

As one might imagine, mitochondrial haplotype (and thus overall energy metabolism) can significantly affect the epigenetic profile within a cell [55, 56]. The term “epigenetics” is defined as any heritable method for regulation of gene expression apart from manipulation of the DNA sequence itself (“epi” = above, on, over) and epigenetics is the reason for why, although essentially all cells in the body have the same genetic complement, there exists a wide variety of cell types, each with unique expression programs, phenotypes, and physiological roles. The field of epigenetics is fairly young. While Jean-Baptiste Lamarck (~1800 A.D.) and Charles Darwin (1868) both

hypothesized that environmental factors could influence heritable traits, though evidence for the presence of epigenetic machineries did not start to significantly accumulate until the twentieth century [57-62]. Examples of epigenetic mechanisms include DNA methylation, microRNAs, and post-translational modification of chromatin packaging proteins (i.e., histones) [63-66].

The direct, covalent modification of DNA via the addition of methyl groups plays a very important role in regulation of gene expression within a cell [67]. Of the four canonical DNA bases—adenine, thymine, cytosine, and guanine—two can be methylated: adenine and cytosine (of which only cytosine has been well studied) [68]. The addition of a methyl group to DNA (primarily to the C5 locus on cytosine's pyrimidine ring) is catalyzed via DNMT family enzymes in mammals at the many (~29 million) CpG dinucleotide motifs [69]. De novo DNA methyltransferases are involved in early establishment of the genomic DNA methylation pattern, whereas maintenance DNA methyltransferase enzymes are primarily responsible for re-establishing DNA methylation following DNA replication via recognition of hemi-methylated sites [70]. Reversal of DNA methylation occurs via the TET (ten eleven translocation) family of enzymes [71]. While methylated CpG moieties do not individually block transcription, DNA methylation is known for suppressing transcription activity when methylation is sufficiently dense (islands) within a genetic locus [72]. For example, high levels of methylation localized within promoter regions have been shown to inhibit gene transcription, DNA methylation also plays a critical role in suppressing the mobility of transposon elements, and normal embryonic development likewise relies on methylation in order to inactivate/silence genes involved in certain stages of embryogenesis and to

imprint maternal and paternal genomes with sex-specific methylation profiles [72-74]. Alternatively, methylation of the gene body (i.e., not within the promoter region) has actually been shown to be strongly associated with highly transcribed genes, possibly due to either inhibiting secondary promoters or to potentiating the rate of elongation or transcript splicing [75].

Perhaps unsurprisingly, aberrant DNA methylation is often an important factor in cancer initiation and progression [64]. Cancer is frequently described as a disease of the genome, characterized by a high degree of genetic heterogeneity. Given that the DNA methylation program can be drastically altered through sustained mutations to metabolic pathways such as the TCA cycle and epigenome modifying enzymes, epigenetic heterogeneity may also be playing a key role in the evolution of tumors [76-79]. Tumor initiation can be promoted via hypermethylation of tumor suppressors (such as DNA repair genes) and oncogenic hypomethylation of the genome [80, 81]. In the context of tumor progression and metastasis, the specific genes affected by changes in DNA methylation vary depending on cell type and primary and secondary site microenvironments [82]. For instance, hypermethylation of tumor suppressors often occurs very early in cancer and hypermethylation of metastasis suppressor genes would likely only provide a growth/survival advantage at late stages of cancer progression [83]. Global, heterogeneous hypomethylation would allow for disease progression via phenotypic plasticity, allowing for a population of cancer cells to evolve and adapt to the differing stresses and demands of the metastatic cascade [84, 85].

A second significant epigenetic mechanism through which cells can regulate gene expression takes the form of non-coding RNAs. This group of RNAs includes long

non-coding RNA (lncRNA) and small non-coding RNA, such as siRNA (small interfering), snRNA (small nuclear), snoRNA (small nucleolar), and miRNA (microRNA), among several others. While many non-coding RNA transcripts appear to be the result of non-selective, cryptic transcription (essentially intergenic RNA), other non-coding RNAs have been shown to have biological functions [86]. One fairly well-characterized class of non-coding RNA are microRNAs, which range between 20 and 25 nucleotides long after maturation and function by binding to mRNA transcripts and modulating subsequent translation (either via inhibition of translation or promoting degradative cleavage of the transcript itself) [87]. As such, miRNAs represent an important method of epigenetic regulation of protein translation. In the context of cancer, just like regular peptide-encoding genes, genetic loci encoding miRNA transcripts are subject to mutations and epigenetic regulation via methylation and chromatin dynamics [88]. Much as regular genes, miRNAs can also act in either an oncogenic or tumor suppressive manner (and the miRNAs are known as oncomiRs and tumor suppressor miRNAs, respectively) [89]. OncomiRs typically suppress tumor suppressor peptide translation (e.g., miR-21 modulates PTEN) and tumor suppressor miRNA downregulate oncoproteins (let-7 miRNA regulates RAS and Myc oncoproteins; let-7 expression can be throttled in cancer). Perturbation of normal miRNA expression can also contribute to all steps of the metastatic cascade. Such miRNAs involved in regulating the various steps of the metastatic cascade are known as metastamirs, which can have either pro- or anti-metastatic properties [90].

Histones and their post-translational modifications comprise yet another important form of epigenetic regulation in the cell. Chromatin structure and the

accessibility of genetic loci are critical to determining the level of transcriptional permissiveness [91]. As one can imagine, the dynamic conformation of chromatin can drastically affect the ability of transcriptional machinery and transcription factors to bind to target loci and potentiate expression. The presence and/or combination of certain histone post translational modifications are associated with specific gene regions (such as binding to promoters or enhancers) and with transcriptional (such as active or repressed transcription) and chromatin states [92]. Heterochromatin is associated with strongly repressed gene expression and is visible as dense, dark-staining regions in the nucleus. Conversely, euchromatin is associated with activate transcription and is open and does not readily stain [93, 94]. In tumorigenesis, chromatin remodeling and histone modifying proteins can become mutated and misregulated, ultimately resulting in both aberrant epigenetic and expression programs [95].

In order to play their critical role in the dynamic packaging of the genome, histone proteins are organized into octamers of histone subunits. Eukaryotic cells produce five canonical histone subunits, H1, H2A, H2B, H3, and H4, where H2A binds H2B and H3 binds H4 as heterodimers and H1 is considered to be a linker protein (contributing to high-order chromatin status) [96]. Each of the canonical histone subunit's expression levels are closely tied with S phase of the cell cycle and the coding sequences are extremely conserved throughout eukaryotes, as any perturbation in DNA packaging and chromatin conformation is likely to be highly detrimental to an organism [97, 98]. Two of each heterodimer (H2A-H2B, H3-H4) bind in order to form a functional DNA-binding unit and complete a nucleosome is formed when a histone octamer is wrapped 1.67 times by a 147bp strand of DNA [96].

Apart from the canonical histones, there exist many other histone subvariants which can be deposited into nucleosomes and which are associated with different chromatin conformations or specific chromosomal loci [99]. For example, histone H3 has at least two known subvariants, including H3.3 and cenH3. H3.3 deposition takes place specifically at highly-transcribed genes, whereas cenH3 localized to centromeres and are involved in kinetochore assembly. The other well-studied histone is histone H2A, which has at least four additional subvariants (H2A.Z, H2A.X, H2A.B, macroH2A) [100]. Histone H2A.Z is unique in that the protein has evolutionarily diverged and is roughly 60% conserved compared to canonical H2A. H2A.Z has been associated with chromatin remodeling and, in eukaryotes, tends to correlate with actively expressed genes. The role of H2A.X is somewhat less clear-cut [101]. H2A.X_{null} mice are infertile, as their spermatocytes are unable to properly condense and likewise cannot inactivate X and Y chromosome genes during meiotic division. In the context of DNA repair, upon induction of DNA damage (specifically, double-strand breaks), phosphorylation of H2A.X immediately occurs within the region of the strand break. The initial phosphorylation event propagates throughout a large portion of the break region and persists until the lesion is repairs. Intriguingly, knockout of H2A.X does not abolish lesion repair, although the mutants do exhibit a reduction in repair efficiency and potentiated sensitivity to genotoxic insults. Lastly, H2A.B and macroH2A appear to play mutually exclusive roles in the process of X chromosome inactivation and dosage compensation. MacroH2A is strongly associated with inactive heterochromatin and can be readily found in inactivated X chromosomes (Barr bodies) [102, 103]. H2A.B, on the

other hand, is essentially undetectable in Barr bodies, but is widely distributed throughout the genome at actively expressed genetic loci [104, 105].

Several diverse mechanisms are responsible for either writing, erasing, or reading post-translationally modified histone tails and thus leading to appropriate shifts in the expression program [65, 106, 107]. Writers [108] are enzymes which can catalyze the addition of PTMs to histones and are divided into families which add specific types of modifications. For example, proteins which catalyze the addition of acetyl groups are known as histone acetyltransferases (HATs), such as p300/CBP domains, and those which catalyze the addition of methyl transferases are known as histone methyltransferases (HMTs), such as SET domains. Erasers [108] are enzymes which catalyze the removal of modifications from histone tails and are likewise divided into families which can eliminate specific modifications from histone amino acids. Families of erasers include histone deacetylases (HDACs), such as sirtuins, and histone (lysine) demethylases (HDM/KDMs), such as JARIDs and Jumonji domains. Reader domains are peptides which have the ability to recognize post-translational modifications of histones [109-112]. Although much still remains unknown about readers of several modifications, there have been several discoveries and characterizations of histone readers in recent years. Examples include bromo domains (recognize acetylated histones; first discovered in HATs), chromo domains (recognize methylated histones), and Tudor domains (recognize methylation on lysine and arginine residues). Repositioning nucleosomes is also critical to maintaining proper chromatin conformation dynamics and is largely mediated through chromatin remodeling proteins in an ATP-dependent manner [113]. Chromatin remodelers (such as INO80, ISWI, or SWI/SNF)

are composed of combinations of effector domains (such as for ATPase or helicase activity) with one or more reader domains (such as bromo domains to bind histone acetylation or HSA domains to bind actin/actin-related proteins) [114, 115]. Through recognizing and binding to modified histones, chromatin remodelers are able to use their effector domains to reposition, evict, or otherwise affect nucleosome assembly and, thus, effectively regulate the replication, transcription and repair of DNA. Overall, the chromatin state of the cell is strongly influenced by the combination of writers/erasers and ATP-dependent chromatin remodelers.

While aberrant behavior of epigenome modifying proteins often does ultimately involve a genetic perturbation, subsequent shifts in the epigenetic program (and thus transcription) of a cell can provide the evolutionary motive force necessary for neoplastic growth and to cultivate the epigenetic plasticity required for the many steps of the highly inefficient metastatic cascade [116, 117]. Many histone modifying enzymes have been associated with cancer [66]. Some of the well-studied histone modifiers include the HDM (KDMs), HMT (SUV39Hs, SETs), HDAC (sirtuins), and HAT (p300/CBP) families. The aberrant regulation of such proteins has the potential to play a role in the development of essentially every known type of cancer, although the specific evolutionary contribution to cancer pathogenesis depends largely on the overall progression of the tumor [66]. In principle, abnormal histone modifying proteins can drive cancer simply through either inactivation of tumor suppressor genes or activation of oncogenes. For example, SIRT6 is a histone deacetylase and is considered to be a tumor suppressor due to the protein's direct interactions with HIF1α and MYC. SIRT6 typically co-represses HIF1α (hypoxia-inducible factor-1α), a gene which induces

glycolytic enzyme expression and a Warburg-like metabolic phenotype [118, 119], and MYC, an oncogene involved in ribosome biogenesis, glutaminolysis, and driving cell cycle progression [120]. SIRT6 has also been suggested to suppress stem-like phenotypes in cancer cells [121]. As one might imagine, in cancer, SIRT6 can suffer fatal indels, point mutations, or expression become otherwise lost, thus fueling deregulated metabolism and a stem-like state in the cell.

In eukaryotes, the N-terminal tail of histone subunits extend into the surrounding nuclear space and post-translational modifications to these tails contribute to overall chromatin conformation and DNA accessibility [122-124]. A small plethora of possible PTMs exist and include lysine-specific sumoylation, ribosylation, and ubiquitination, and arginine/lysine-specific acetylation and methylation, and serine/threonine-specific phosphorylation [125]. Recent research has shown the presence of several new lysine acylations, specifically crotonylation, hydroxyisobutyrylation, propionylation, malonylation, and succinylation [126, 127]. Many of these new acylation modification sites are suggested to overlap known acetylation and are thought to be tied closely to metabolic activity within a cell. For example, proprionyl-CoA is produced via oxidation of odd-chain lipids and catabolic breakdown of branched-chain amino acid [128] and butyryl-CoA is produced as an intermediate metabolite of lipid β-oxidation [126]. As described earlier, the overall epigenetic program is closely tied to the activity epigenome modifying enzymes, which generally rely on mitochondria-derived metabolic substrates and intermediates as metabolites [129]. This most recent research on histone lysine acylation further builds support for the concept of linking metabolism and metabolic state with chromatin conformation [126].

Intriguingly, there exists essentially a cyclic crosstalk between metabolism and epigenetics [130]. The level and subsequent catalytic activity of enzymes responsible for cellular metabolism directly control production of downstream metabolites such as acetyl-CoA, SAM, and NAD^+. Epigenome modifying enzymes (such as HATS, HMTs, and DNMTs) utilize the metabolic substrates in order to write the epigenetic program of a cell. Epigenetic modifications (such as histone PTMs and DNA methylation), in turn, can directly control the expression of the metabolic enzymes responsible for producing the required substrates in the first place [130]. Apart from a marked shift towards Warburg-like glycolysis, metabolic pathways can be disturbed as a result of sustaining mutations in proteins involved in the metabolic pathways themselves (such as converting an intermediate metabolite into a substrate usable by epignome modifying proteins) [131]. Several examples of such mutated metabolic enzymes include isocitrate dehydrogenase 1 and 2 (IDH1/2), fumarate hydratase (FH), succinate dehydrogenase (SDH), and 5-methylthioadenosine phosphorylase (MTAP). One of the first discoveries made (linking epigenetics and metabolism in cancer) was through the observation that IDH1/2 can sustain gain-of-function mutations in cancer, which results the generation of (R)-2-hydroxygluatrate via reduction of 2-oxoglutarate, an important substrate of the Jmjd-type histone demethylases (such as KDM4) and TET 5-methylsytosine hydroxylases [132]. The result of mutated IDH1/2 is ultimately histone and DNA hypermethylation associated with cancer pathogenesis [133]. In a similar manner, loss-of-function mutations to FH and SDH result in high levels of fumarate and succinate, which is able to inhibit α-kg-dependent histone demethylases (again, such as KDM4) and DNA demethylases [134]. The result of FH and SDH loss is a gain of histone and

DNA hypermethylation [135]. Furthermore, one study found FH dysfunction and the subsequent accumulation of fumarate are able to repress TET DNA demethylase activity and induce EMT (thus contributing to cancer progression). Intriguingly, another study even suggested that most of the observed variation in DNA methylation programs could ultimately be due to differences in expression of proteins involved in one-carbon-metabolism (SAH-SAM) and methionine SAM salvage [134]. While mutations in metabolic enzymes themselves are fairly rare in cancer, the ability to generate aberrant epigenetic programs independently of metabolism is a cancer cell's dream. Epigenetic plasticity is a critical evolutionary catalyst which enables cells to overcome challenges faced over the course of disease progression. Fortunately, our ever-improving knowledge of epigenome-metabolome dynamics has resulted in significant progression in efficiently targeting treatments [134]. One study observed that nutrient heterogeneity within the interior/microenvironments of melanomas resulted in decreased α-kg levels and, accordingly, in increased hypermethylation due to diminished activity of Jmjd-type histone demethylases such as KDM4 family members. The result of this hypermethylation was resistance in the melanoma to BRAF inhibitors. Once the BRAF inhibitors were paired with HMT inhibitors, BRAF resistance was able to be overcome. A final interesting therapy indirectly-involving KDM4 family members involves the gain-of-function mutations in IDH1/2. A small molecule inhibitor (AG-221) of IDH1-/2 mutant was able to effectively diminish intracellular concentrations of 2-hydroxyglutarate and, thus, reverse the previously observed histone hypermethylation through the action of Jmjd-type HDMs, including KDM4 family members [134]. As is clear, there is a strong link between metabolism and the epigenetic program within a cell. However, despite

recent advances, there is still very much to learn about characterizing the dynamics between metabolism and chromatin and epigenetics and how to apply that knowledge to best treat disease.

Developing methods through which to reliably and accurately study mitochondria-mediated effects in cancer, however, is a challenge. Over the years, several different models have been developed in order to study the contribution of mitochondria to tumorigenesis. In vitro models include using fibroblast lines, p^0 cells or cybrids, while in vivo models mostly utilize mice. Patient-derived fibroblasts have been isolated and cultured from patients with mitochondrial diseases or mtDNA mutations [136]. Unfortunately, while fibroblasts allow for potential therapeutics to be evaluated in vitro, the overall consequence of mitochondrial defects in fibroblasts can be difficult to ascertain due to metabolic variation (OxPhos vs glycolysis) observed using different culture conditions [137]. Further, such fibroblasts are—alone— generally non-tumorigenic and thus not the best model to use. An alternative to using patient-derived fibroblasts is the p^0 model for studying the effects of depleted mitochondria, which utilizes mammalian cells treated with ethidium bromide in order to perturb the structure and function of mitochondria (and thus deplete mtDNA) [138]. Unfortunately, EtBr can result in hindered cell proliferation, is genotoxic, and can lead to under desired off-target perturbations of the nuclear genome [135]. As with patient-derived fibroblasts, the p^0 model is likewise not an ideal model for studying the role of mtDNA in cancer. Another possible method consists of developing a cybrid cell line, wherein a cell with the nuclear genome of a cancer line and with the mitochondrial genome of another cell line are created. In this situation, the cybrid models can be used to address whether or not

certain phenotypes or other traits result from changing the nuclear or mitochondrial genomes [139, 140]. This model is more useful as the cybrid cells are indeed tumorigenic and can thus reliably be used in vivo [135]. A downside to the model is the dependency on mtDNA variants present in patients.

In regard to animal models available for studying mitochondrial contribution to cancer, various different models are in use [135]. Small animal models using zebra fish, nematodes, or fruit flies are also viable methods options for studying the effects of mitochondria in disease pathology, although such organisms do not, of course, contain the same organs as humans [141, 142]. Murine models are very useful for investigating the role of mtDNA in cancer as many variables that could not readily be controlled in patients can be directly controlled by a researcher using mice as subjects. Three useful murine models for studying the relationship between mitochondrial haplotype and cancer have been developed, including conplastic mice, mito-mice, and MNX mice. Conplastic mice are bred through many backcrosses in order to have the nuclear genome from one strain and the mitochondrial haplotype of another strain [143]. Using this model, one can observe the contribution of mtDNA to complex diseases, although the process of conducting enough backcrosses takes several years and the nuclear genome may not be 100% the same as the target strain. Another in vivo mouse model is the mito-mice (transmitochondrial) model, whereby various levels of mitochondrial heteroplasmy are assessed for their contribution to disease against a single nuclear genome. In the mito-mice model, a mitochondria donor cell with a specific mtDNA mutation is selected and the nucleus is removed in order to generate cytoplasts, which are fused with a rho-0 cell in order to generate a cybrids [144, 145]. After the genome if

the cybrid is removed, the cybrid can either be fused whole with an embryo in the pro-nuclear stage, or the cybrid can be treated with rhodamine-6G in order to deplete the mtDNA, whereupon the cybrid is fused with an embryonic stem cell and subsequently injected into an 8-cell mouse embryo before implantation. Due to off-target genotoxic effects of mitochondria-depleting compounds, however, the mito-mouse model is not optimal for studying the contribution of mtDNA to tumorigenesis and metastasis [135].

A suitable alternative to the mito-mouse model is the Mitochondrial Nuclear Exchange (MNX) mouse model [146]. MNX mice are generated from fertilized oocytes with a certain mitochondrial haplotype, which have their nuclear genome removed and, in place, are microinjected with the nuclear genome from another mouse strain before being implanted into a female mouse. As previously discussed above, in 2006, Kent Hunter revealed that the host genetic contribution to metastatic cancer formation may be significantly more complex than originally anticipated [26]. In his experiment, Hunter bred male FVB/NJ-TgN(MMTV-PyMT)^{634Mul} mice (transgenic mice expressing the polyoma middle-T antigen), which present with spontaneous tumors within 60 days and roughly 90% of the mice form secondary growth in their lungs within 100 days, with female mice of several different inbred stains. The result was that mice formed pulmonary metastases to significantly varying degrees. The intuitive explanation for this observation would be that the differences are likely due to the nature of the nuclear genome and this hypothesis supported due to the discovery of several metastasis suppressor genes. Critically, however, what Hunter did not evaluate was the possibility that the observed differences in disease could have been due to maternally-inherited mitochondrial haplotypes of the mice.

Using the unique Mitochondrial-Nuclear Exchange (MNX) mouse model to study the effects of varying the type of mitochondria in regard to cancer progression and metastasis—in addition to studying the interactions of mitochondria with the DNA present in the nucleus of the cell—previous research in the lab of Dr. Danny R. Welch showed that mitochondrial haplotype does have a significant influence on the rate of both tumor formation and metastasis [147]. Previous experiments in the lab have also shown that mice with different mitochondrial backgrounds have distinct patterns of gene expression and DNA methylation [148]. As a corollary to these findings, the research project for this Master's thesis was interested in investigating other critical epigenetic mechanisms through which mitochondrial haplotype is able to influence which genes are turned on or off in the cell. Given the tight link between mitochondrial metabolism and epigenetics, the logical hypothesis was simply that the mitochondrial haplotype is likewise able to influence the epigenetic process of post-translationally modifying the N-terminal tails of histones (in order to affect chromatin state dynamics in the cell. Thus, this thesis project sought to observe whether or not histone modification profiles are different depending on the mitochondrial haplotype present of each strain of MNX mouse examined. The results of this research will not only contribute to our understanding of how mitochondria participate in influencing the DNA packaging mechanisms involved in turning genes on or off, but, as our research is not limited to any single form of cancer, has the potential to benefit to all cancer patients.

Methods

Mouse Strains

For the ChIP-Seq experiment, cerebella from four different strains of mouse (*mus musculus*) were harvested, snap frozen in liquid nitrogen and stored at -80°C until use. Specific strains included: C57BL/6J$_{wt}$, C3H/HeN$_{wt}$, C57BL/6J$_{(n)}$:C3H/HeJ$_{(mt)}$, and C3H/HeJ$_{(n)}$: C57BL/6J$_{(mt)}$. All brain samples originated from 8 week old, male mice.

Histone Acid Extraction

Cells at ~75% confluence were harvested and washed twice with ice-cold PBS (or one could use roughly 25 mg of tissue). Cells or Dounce-homogenized tissue samples were suspended in Triton Extraction Buffer (TEB: PBS containing 0.5% Triton X 100 (v/v), 2 mM phenylmethylsulfonyl fluoride (PMSF), 0.02% (w/v) NaN_3) at a cell density of 10^7 cells per mL Cells were lysed on ice for 10 min with gentle stirring and then centrifuged at 6,500 x g for 10 min at 4°C to spin down nuclei. Supernatant was removed and discarded. Nuclei were washed in half the volume of TEB and centrifuged as before. Pellet was resuspended in 0.2 N HCl at a density of $4x10^7$ nuclei per mL. Histones were acid extracted over night at 4°C. Samples then centrifuged at 6,500 x g for 10 min at 4°C to pellet any debris. Supernatant was then saved, which contained the histone-enriched fraction, and protein content then determined using BCA assay. Aliquots stored at -20°C. (Protocol source: Abcam)

Western Blot Analysis

For *whole cell lysate*, cells were washed in 5mL PBS twice, whereafter 300 µL of RIPA with 1% (3 µL) HALT 100X protease inhibitor cocktail (EDTA-free) was added to the plate and incubated on ice for 15 minutes. Cells were scraped and transferred to microcentrifuge tubes, which were then centrifuged (15 min, 13,000 RPM, 4°C). After spinning, supernatant was transferred into fresh microcentrifuge tube and stored at -80°C.

For *cytoplasmic* and *nuclear protein fraction* isolation, a NE-PER Nuclear and Cytoplasmic Extraction Kit protocol was followed. CER I, CER II, and NER I solutions were supplemented with 1% HALT (EDTA-free).

To quantify the concentration of the protein fractions, a Pierce BCA (Bicinchoninic Acid) Protein Assay Kit was used. No deviations from the standard protocol (microplate procedure) were made. After 30 minutes incubation, absorbance was read at 562nm and the protein concentrations were thus determined.

Bis-Tris SDS gels were made in lab according to the following protocol:

4% stacking gel: 6.8 mL dH_2O, 1.25 mL Tris pH 6.8, 1.7 mL 30% bis-acrylamide, 100 µL 10% SDS

> *To initiate polymerization*, add 100 µL 10% ammonium persulfate (APS) and 10 µL TEMED (to 10 mL gel solution)

12% resolving gel: 3.3 mL dH_2O, 2.50 mL Tris pH 8.8, 4.0 mL 30% bis-acrylamide, and 100 µL 10% SDS

> *To initiate polymerization*, add 100 µL 10% ammonium persulfate (APS) and 4 µL TEMED (to 10 mL gel solution).

Prior to casting Bis-Tris SDS gel, protein samples of each fraction (20-50 µg; WCL/Cyto/Nuc) were diluted with 4X Lithium Dodecyl Sulfate (LDS) sample buffer supplemented with ß-Mercaptoethanol (10% v/v) and heated to 95°C (> 5 min), before being brought back to room temperature.

Once fully polymerized, the gel was placed in an appropriately sized electrophoresis chamber with/without a buffer dam. The well comb was removed carefully from gel and the chamber then filled with desired running buffer (1X Tris/Glycine/SDS; 1X MES SDS; 1X MOPS; etc). Samples were transferred to individual wells, in addition to desired protein ladder /marker. The gel was run at 100V until dye front reached bottom of gel.

After completion of electrophoresis, the gel was transferred to a PVDF/nitrocellulose membrane using either a commercial transfer pack (Bio-Rad Turbo) or a homemade transfer stack. For the homemade transfer stack using PVDF, 1X transfer buffer (with 10% v/v methanol) of the desired composition was made. Two extra-thick transfer pads were equilibrated in the activated transfer buffer. The PVDF membrane was activated by soaking in 100% methanol for ~5 sec and then moved to 1X activated transfer buffer and soaked. Once ready, PVDF membrane was placed on one of the thick transfer pads in the transfer tray; the membrane was pre-cut with an orientation mark (removed top left corner, etc). The gel was then carefully removed from its cassette and placed atop the PVDF, paying close attention to sample order orientation. A roller removed air bubbles between the gel and the membrane. The second thick transfer pad was placed atop the gel and again the roller was used to remove air bubbles. The transfer tray was

locked, placed in the transfer unit, and the desired transfer protocol (high/low MW, standard, etc) selected. After complete transfer, the PVDF membrane was removed and placed immediately in TBST+5% skim milk. Membrane was blocked for at least one hour at RT, or overnight at 4°C.

Once the membrane had blocked, primary antibody was added and incubated at room temperature for one hour or overnight at 4°C. After incubating, antibodies were removed and the membrane was washed three times for 5 minutes each in TBST. Secondary antibody was added and incubated again at room temperature for one hour or overnight at 4°C. After incubating, secondary antibodies were removed and the membrane was washed three times for 5 minutes each in TBST.

The membrane was imaged using Thermo Fisher West Dura or Femto (high sensitivity) Kit. Imaging reagent solutions were mixed at 1:1 ratio, added to the membrane, and incubated for 5-8 minutes in a dark place (or cover the membrane case with foil).

To strip PVDF after imaging, the membrane was washed three times with TBST and soak membrane in Thermo Scientific Restore PLUS stripping buffer for 5 to 15 minutes at room temperature. Stripping buffer was removed and the membrane washed with TBST. The membrane was re-blocked for 1 hour at RT as previously described. Repeated primary and secondary antibody incubation procedure for desired controls.

Modified Histone Peptide Array

To validate antibody specificity prior to ChIP-Seq, an EMD Millipore AbSurance H3 Peptide Array (#16-667) was used. Manufacturer instructions were followed. Primary antibody dilutions ranged from 1:8,000 to 1:10,000. Secondary antibodies were diluted 1:500.

Antibodies

anti-H3K4me1 – Abcam – ab8895, polyclonal rabbit IgG (lot: GR3186715-1)
anti-H3K4me3 – Abcam – ab8580, polyclonal rabbit IgG (lot: GR3190162-1)
anti-H3K27ac – Abcam – ab4729, polyclonal rabbit IgG (lot: GR3187599-1)
anti-H327me3 – Cell Signaling – 9733, monoclonal rabbit IgG
anti-normal IgG – Cell Signaling – 2729, rabbit IgG

ChIP

Flash-frozen, mouse brains (cerebella) were weighed and transferred to a mortar and pestle pre-chilled in liquid nitrogen. Brain tissue was pulverized to a fine, homogenous powder and subsequently added to 1% formaldehyde/1X PBS solution for protein crosslinking. After 10 minutes of fixing at room temperature, the samples were spun briefly and the supernatant discarded. Tissue was resuspended in 1 M Tris 8.0 pH and placed on a rocker for 10 minutes at room temperature to quench the crosslinking reaction. Again, samples were spin briefly and the supernatant discarded.

Samples were resuspended in ice-cold B1 hypotonic lysis buffer supplemented with HALT 100X EDTA-free protease inhibitor cocktail (PIC) and transferred to a 7 mL Dounce homogenizer with “tight” pestle for mechanical disruption of the outer plasma membrane. Samples were dounced between 50-60 strokes to ensure thorough membrane disruption. After douncing, sample was sieved through a 100 µm nylon DNase/RNase-free cell strainer into a new 50mL conical tube and centrifuged at 800g, 4°C, for 5 minutes. Supernatant (cytoplasmic fraction) was discarded and the nuclei resuspended in ice-cold S1supplemented with PIC. The resuspended S1 solution was transferred on top of the ice-cold S2 solution (with PIC) to create a sucrose gradient. The sucrose solutions were spun at 3000RPM, 4°C. for 10 minutes. The supernatant was carefully removed with a 10 mL pipette and discarded. The isolated nuclei pellet was resuspended in sonication/lysis buffer supplemented with PIC and allowed to incubate on ice for at least 10 minutes to ensure lysis of nuclear membranes.

For sonicating chromatin, samples were aliquoted to 0.6 mL tubes (200 µL/tube) to ensure even and consistent shearing. Using an ActiveMotif temperature-controlled water bath (indirect) sonicator, chromatin samples were sheared for 27 minutes total, using a cycle of 10 sec ON, 5 sec OFF, and 75% amplitude. The sonicator was briefly turned off every 5-7 minutes to allow the water bath to return to 2°C in order to ensure the samples did not overheat. After sonication, sample aliquots were returned to 1.7mL low-bind Eppendorf tubes.

To validate shearing of samples (fragment bands between 100-600bp), an agarose DNA gel was run. 25 µL of sheared chromatin was transferred to a 1.7 mL microcentrifuge tube, along with 250 µL elution buffer (supplemented with 250 mM NaCl). Crosslinks were reversed by incubating at 85°C for 2 hours. After reversal of crosslinks, 1-2 µL RNase/T1 were added to each sample and incubated for 30-60 minutes at 37°C. After RNase digestion, 1-1.5 µL Proteinase K was added to each sample and incubated to 30-60 minutes at 48°C. 500 µL phenol-chloroform was then added to the samples, vortexed briefly, and centrifuged at max speed, 4°C, for 20 minutes. The top layer was transferred to a new tube containing 900 µL 100% EtOH and glycogen (1000 µL:1 µL ratio). DNA was precipitated from 1hr to overnight before being centrifuged at max speed, 4°C, for 20 minutes. After spinning, the supernatant was discarded and 500 µL 80% EtOH added. Samples were vortexed briefly to resuspend the pellet and centrifuged at max speed, 4°C, for 20 minutes. After spinning, the supernatant was discarded and DNA pellet allowed to air dry for 15-20 minutes. DNA was resuspended in 20-30 µL H_2O and mixed with 6X loading dye. Samples and 1 kb marker ladders were loaded on a 1.5% agarose gel in 1X TAE and run at 100V until sufficiently resolved.

Once proper fragmentation of chromatin (100-500bp, minimal higher molecular weight bands) had been assured, protein concentrations were estimated using a Pierce BCA (bicinchoninic acid) Protein Assay Kit was used. No deviations from the standard protocol (microplate procedure) were made. After 30 minutes incubation, the

absorbance was read at 562 nm on a plate reader and the protein concentration thus determined.

After determination of protein concentrations, Dynabeads were prepared. Dynabeads (Protein A) were resuspended in their vial (vortexed briefly) and an appropriate volume of Dynabeads added to 1.7 mL low-bind microcentrifuge tubes. The volume of beads was adjusted based on required binding capacity (i.e., amount of antibody required). Bead tubes were placed on magnetic rack to separate beads from the stock solution and supernatant was discarded. Beads were washed with 200-400 µL 1X PBS+0.1% Tween-20. To pre-bind desired antibody to the beads, antibody was diluted in 100 µL 1X PBS+0.1% Tween-20 to the Dynabeads. The antibodies and beads were incubated with rotation 2 hours at room temperature. After pre-binding, tubes were placed on magnetic rack and supernatant discarded. The antibody-bead complexes were gently washed in 400-800 µL 1X PBS+0.1% Tween-20.

To set up the immunoprecipitation reactions, 15-75 µg sonicated chromatin was added per test sample tube and matching IgG tube (equal volume of antibody). Bead-chromatin samples were diluted at least 1:10 using IP buffer and incubated overnight with rotation. IP volumes ranged from 100 µL to 500 µL (depending on antibody used). Inputs for each individual mouse were generated by transferring 470 µL elution buffer to 1.7 mL low-bind microcentrifuge tubes and adding 20 µL sonicated chromatin to the tubes (one input per unique individual/mouse). After completion of IP incubation, samples were washed six times to reduce background binding. First wash: 1000 µL

wash buffer #1 for 5 minutes with rotation at 4°C. Second wash: 1000 µL wash buffer #2 for 5 minutes with rotation at 4°C. Third wash: 1000 µL wash buffer #3 for 5 minutes with rotation at 4°C. Fourth wash: 1000 µL of wash buffer #4* for 5 minutes with rotation at 4°C (*made fresh 1X wash buffer #4 with 1% sodium deoxycholate + H_2O). Fifth and Sixth washes: 1000 µL TE for 3 minutes with rotation at 4°C. Each sample was eluted in 500 µL ChIP elution buffer at 65°C for >= 30 minutes with occasional vortexing/hand-turning. After incubation, non-input samples were transferred to clean 1.7mL low-bind microcentrifuge tubes. To reverse crosslinks, both input and non-input sample tubes were supplement with 20 µL 5M NaCl and incubated for at least 15 hours at 65°C. RNAse A/T1 was then added to each sample and incubated at 37°C for two hours. To digest remaining proteins, 1-1.5 µL Proteinase K was added to each tube and incubated at 48°C for three hours. Lastly, DNA was purified using a column-based QIAquick PCR Purification Kit (eluted in 35 µL molecular biology-grade H_2O) and quantified using a Qubit fluorometer with high-sensitivity reagents. Input samples were quantified using Nanodrop.

Library Preparation

Libraries for sequencing were generated according to the Illumina TruSeq ChIP-Seq Sample Prep Protocol, available from the Illumina website (Illumina catalog: IP-202-9001DOC).

Sequencing

ENCODE guidelines and recommendations for sequencing were followed. ChIP DNA samples and input were sequenced using an Illumina HiSeq2500 TruSeq.v3 High Output platform (100bp single reads). “Broad” histone modifications (H3K4me1 and H3K27me3) were sequenced to a target of 45 million mappable reads. “Narrow” histone modification (H3K4me3 and H3K27Ac) were sequenced to a target of 20 million mappable reads. Inputs for each individual sample were sequenced to a target of 60 million mappable reads.

Results

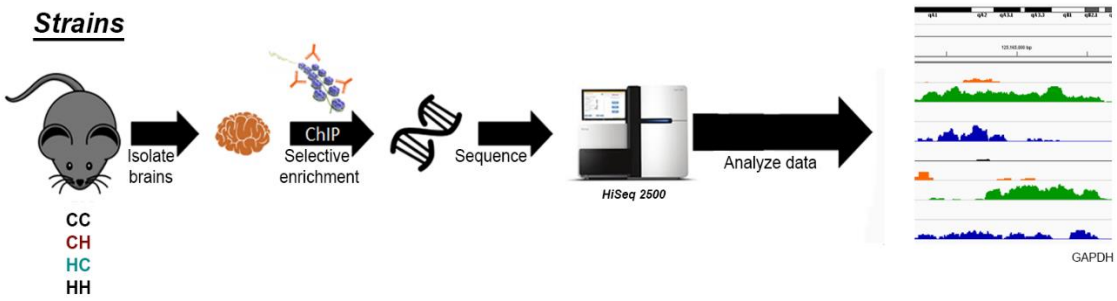


Figure 1 Experimental Outline. For this ChIP-Seq experiment, four lines of Mitochondrial Nuclear Exchange (MNX) mice were used, including two isogeneic, biological replicates per strain. Cerebellum samples were harvested from eight week old male mice and flash frozen with liquid nitrogen. Brains were initially homogenized via cryopulverization and then fixed for 10 minutes in 1% formaldehyde. After quenching the fixation reaction, brain homogenate was dounced 50-60 times, run through a 100 µm filter, and the nuclei isolated using a sucrose gradient. After adding lysis buffer, nuclei were sonicated for 27 minutes total (10 sec ON, 5 sec OFF, 75% amplitude) in a temperature-controlled water bath. Sonicated chromatin fragmentation was visualized on an agarose gel and the protein content quantified via BCA assay. For setting up the immunoprecipitation reaction, antibodies were pre-bound with Dynabeads for two hours before starting the IP. With an appropriate amount of chromatin, IPs ran for 15 hours before being washed and eluted. Post-ChIP DNA was purified using a QIAquick PCR purification kit. Upon submission to the sequencing core, DNA was run on an Agilent Bioanalyzer to further verify chromatin fragment length. After index/sample libraries were built, sequencing was conducted on an Illumina HiSeq2500. Broad histone modifications (H3K4me1/H3K27me3) were sequenced to a target of 45 million reads, whereas narrow histone modifications (H3K4me3/H3K27ac) were sequenced to a target of 20 million reads. Post-sequencing analysis was conducted in collaboration with Dr. Devin Koestler and Dr. Dong Pei in the KUMC Biostatistics department. Tools used to analyze sequencing data include FastQC, IGV, DiffBind, PAVIS, and Gene Ontology. Mouse strain key: CC = C57BL/6J$_{wt}$, CH = C57$_{nuclear}$, C3H $_{mtDNA,}$ HC = C3H$_{nuclear}$, C57 $_{mtDNA}$, HH = C3H/HeN$_{wt}$

Primary Antibody Validation

Before beginning the ChIP-Seq experiment itself, evaluating both the specificity and the selectivity of the chosen antibodies was necessary. Following published ENCODE project guidelines for primary and secondary antibody validation methodology, the first means of antibody evaluation consisted of immunoblotting using histone-enriched fractions obtained through acid extraction of B16F10 mouse melanoma whole cell lysate. For the immunoblots, concentration gradients of histone-enriched fractions were used in order to evaluate optimal, specific signal response and to monitor for potential off-binding events at different protein concentrations. As can be observed in the blots of figure 2, there are clear signals at 17 kDa for each antibody evaluated. No other bands are readily visible, except for some faint events around 75 kDa at higher loading concentrations for H3K4me1. For H3K27me3, there appear to be very faint background bands at roughly 50 kDa, though there is also background staining visible on the blot.

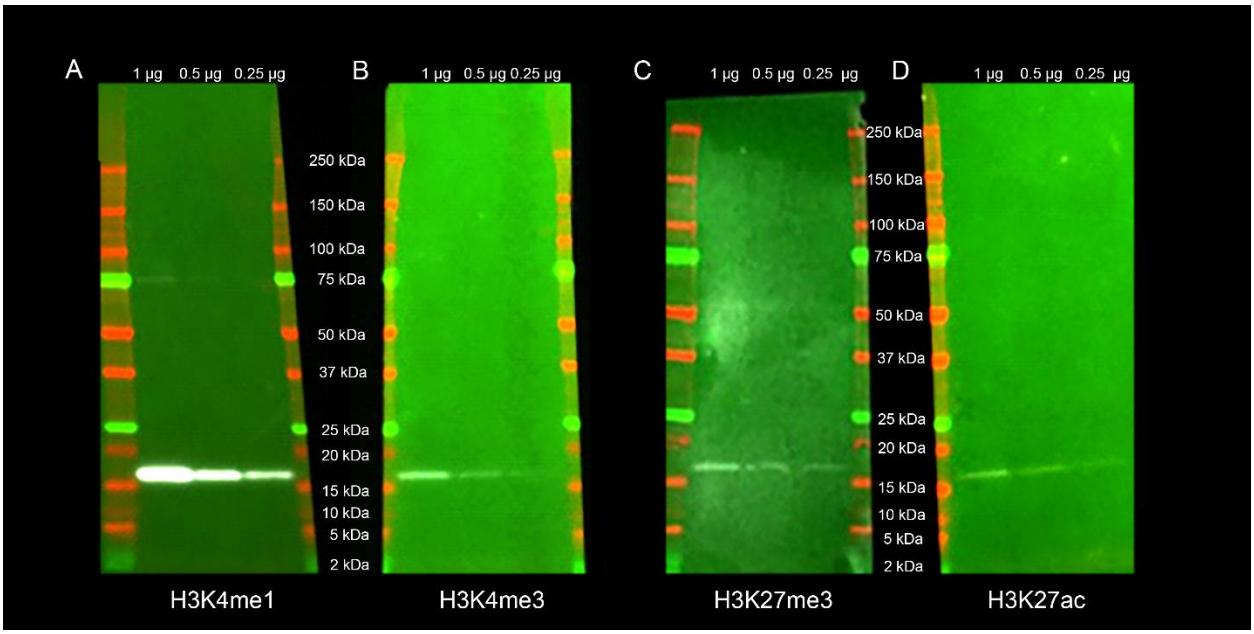


Figure 2 Primary antibody validation (immunobloting). Concentration gradients of histone-enriched fractions were used in order to evaluate optimal, specific signal response and to monitor for potential off-binding events at different protein concentrations. All antibodies were diluted 1:1000 (primary), 1:5000 (secondary). Expected molecular weight of histone H3: 17 kDa.

Secondary Antibody Validation

Following primary evaluation of the candidate antibodies using immunoblots, the antibodies were further tested using an ENCODE-recommended secondary characterization method. In order to determine the selectivity of the antibodies, a commercial peptide array from EMD-Millipore was used. The peptide array features 48 different recombinant post-translationally modified histone epitopes at two different blot concentrations. Four species-specific isotype controls (rat, sheep, mouse, rabbit) are also included. As seen in figure 3, all antibodies displayed very strong binding at a primary target and signals from the rabbit IgG controls were visible. Off-target events were observed in the H3K4me1, H3K4me3, and H3K27ac blots.

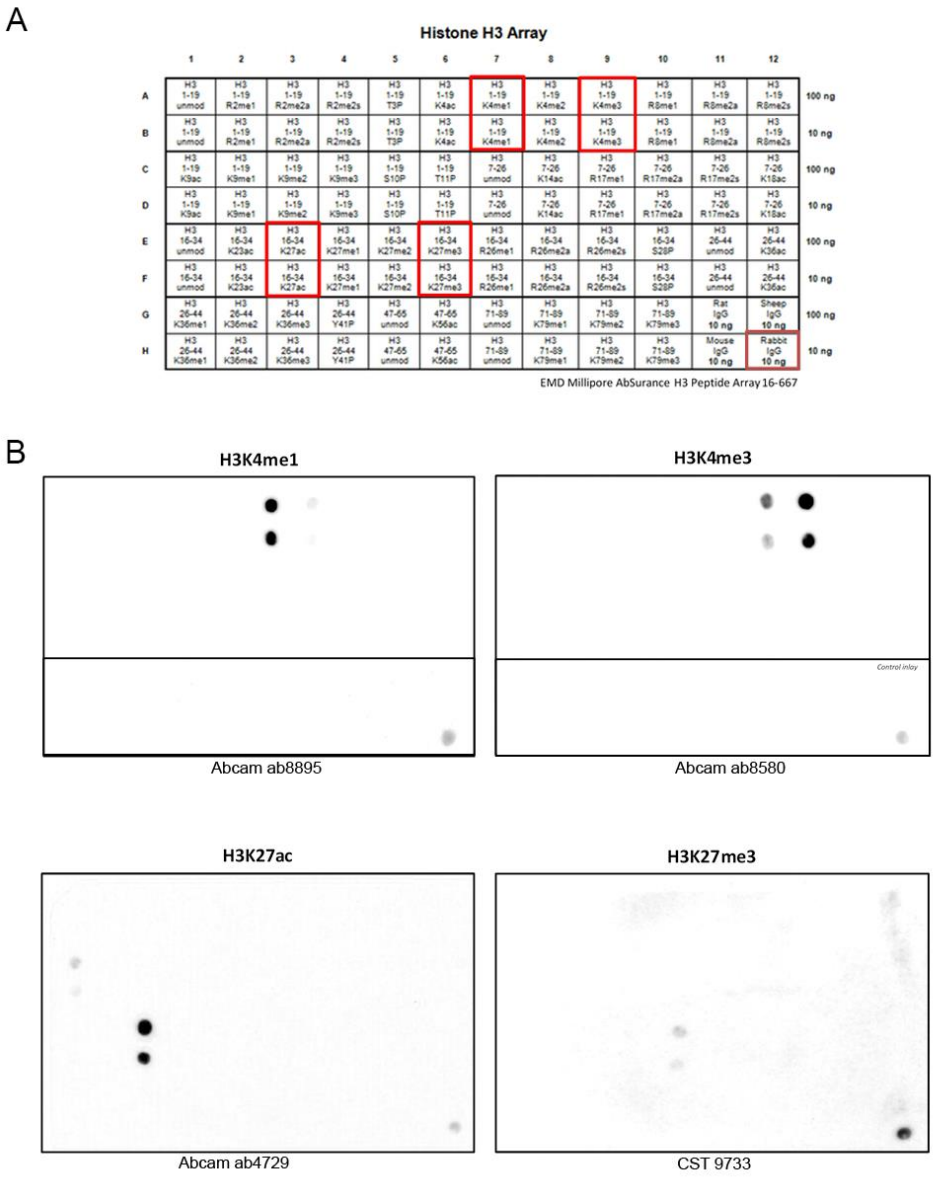


Figure 3 Secondary antibody validation (peptide array). Commercial peptide arrays were obtained from EMD-Millipore. The peptide array features 48 different recombinant post-translationally modified histone epitopes at two different blot concentrations. Four species-specific isotype controls (rat, sheep, mouse, rabbit) are also included. Primary antibody dilutions were as follows: H3K4me1 (1:9000), H3K4me3 (1:9000), H3K27me3 (1:10000), H3K27ac (1:8000). All secondary antibodies were diluted 1:5000.

Post-Sequencing Analysis

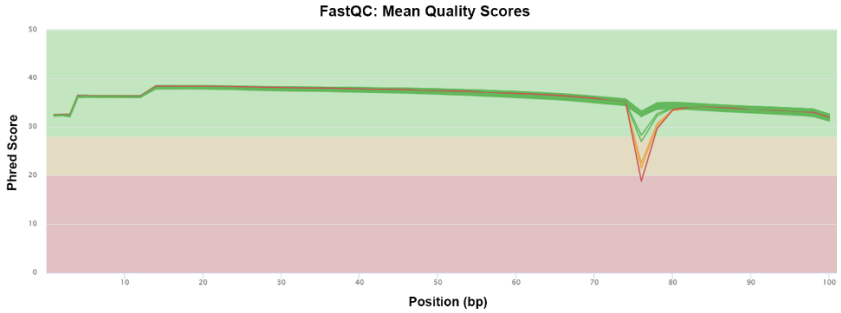
After successful completion of antibody characterization, the ChIP-Seq experiment itself was begun. A minimum of 20 ng of post-ChIP DNA was collected for each antibody per mouse and submitted to the sequencing core facility at KUMC. Upon receipt of sequencing data, all fastq files were initially evaluated for overall quality using the FastQC program. In figure 4, one representative sample was chosen. As is visible, the average percentage of duplicate reads ranged anywhere from around 10% up to roughly 20%. The mean quality score (i.e., the Phred score) for reads generally remained around 30. Interestingly, one consistent drop in read quality was visible in one specific input subgroup in many specimens (as seen by the drop in the average read quality Phred score to below 30 at roughly 80 bp). Furthermore, when observing the quality per tile in the flow cell, a large stretch of red and warm color (low Phred score) reads is visible. In order to correct this dip in quality, the specific input files containing the errors were trimmed to include only the reads prior to the 70 bp read length. As is visible, the read quality chart no longer features a decrease in read quality following truncation of the erroneous read set.

A

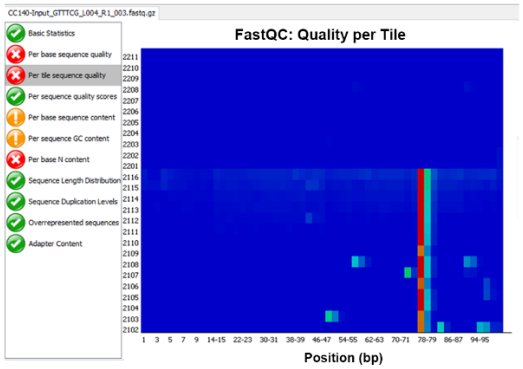
General Statistics

Sample Name	% Dups	% GC	M Seqs
CC140-1_TGACCA_L001_R1_001_fastq	11.8%	42%	4.0
CC140-1_TGACCA_L001_R1_002_fastq	11.7%	42%	4.0
CC140-1_TGACCA_L001_R1_003_fastq	11.2%	42%	3.5
CC140-1_TGACCA_L002_R1_001_fastq	11.8%	42%	4.0
CC140-1_TGACCA_L002_R1_002_fastq	11.7%	42%	4.0
CC140-1_TGACCA_L002_R1_003_fastq	10.9%	42%	3.5
CC140-1_TGACCA_L003_R1_001_fastq	12.0%	42%	4.0
CC140-1_TGACCA_L003_R1_002_fastq	11.7%	42%	4.0
CC140-1_TGACCA_L003_R1_003_fastq	10.9%	42%	3.3
CC140-1_TGACCA_L004_R1_001_fastq	11.8%	42%	4.0
CC140-1_TGACCA_L004_R1_002_fastq	11.8%	42%	4.0
CC140-1_TGACCA_L004_R1_003_fastq	10.9%	42%	3.4
CC140-2_ACTTGA_L001_R1_001_fastq	20.5%	42%	4.0
CC140-2_ACTTGA_L001_R1_002_fastq	16.4%	42%	1.7
CC140-2_ACTTGA_L002_R1_001_fastq	20.4%	42%	4.0

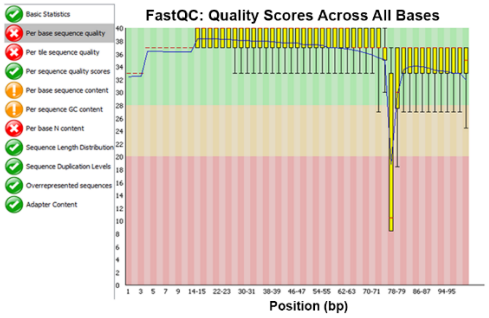
B



C



D



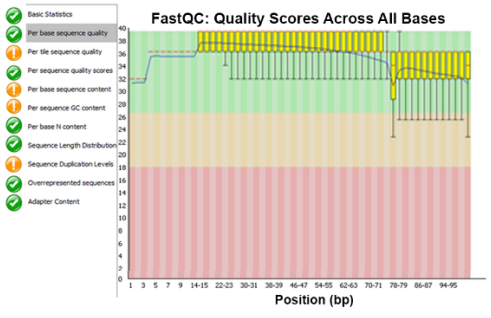
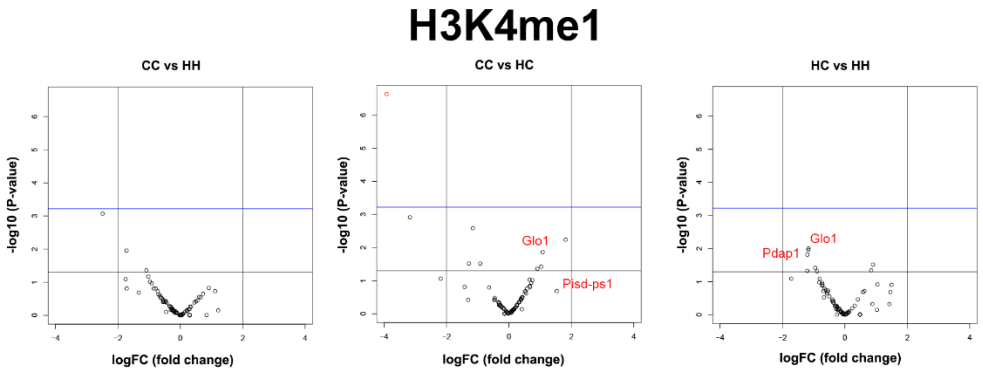


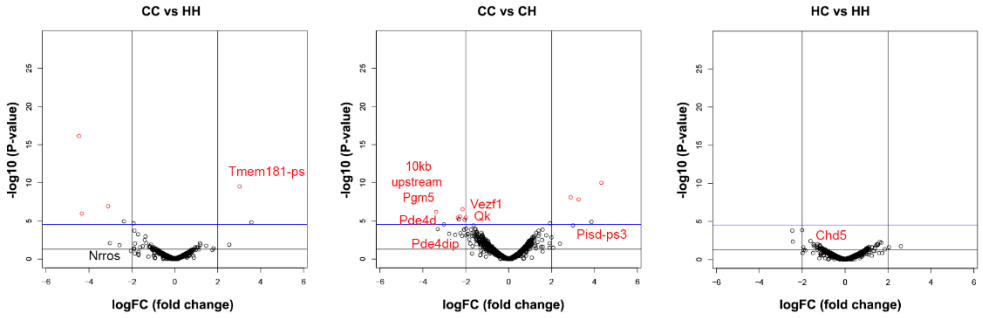
Figure 4 Post-sequencing read quality evaluation using FastQC. A) General statistics for read subgroups, including the percentage of duplicate reads, percent GC content, and number of reads per subgroup. B) Representative mean quality scores (Phred scores) for the experiment. Higher Phred scores indicate less probability of erroneous base calls being made. C) Overall quality per tile is a visual representation of the read locations within the sequencing flow cell. Solid blue indicates high Phred scores, whereas warm colors indicate lower Phred scores and lower read quality. D) (left) Pre-trimming quality scores across all bases. (right) Post-trimming quality scores across all bases. Higher Phred scores (green region of chart) indicate less probability of erroneous base calls being made and higher quality reads.

As all data sets passed initial quality control checks (or were otherwise corrected), sequence data were used to generate a series of volcano plots in order to investigate the fold changes and relative P values of ChIP signal when comparing two mouse strains. Given the lab is concerned with addressing the extent to which mitochondrial background influences the histone modification profiles, three main comparisons were utilized: CC vs. HH, CC vs. CH, and HC vs HH. Intriguingly, in reference to figure 5, while several “hits” were below the P value significance of 0.05, very few data points (only in H3K4me3 sets) were above the adjusted significance value as determined using the Bonferroni correction. Gene names in red indicate matches with previously identified RNA-Seq genes (does not indicate significance), whereas black genes have not been matched to previous RNA-Seq gene lists. For the H3K4me1, H3K4me3, and H3K27me3 shifts in the “hit” profiles between different strain comparisons can be readily observed.

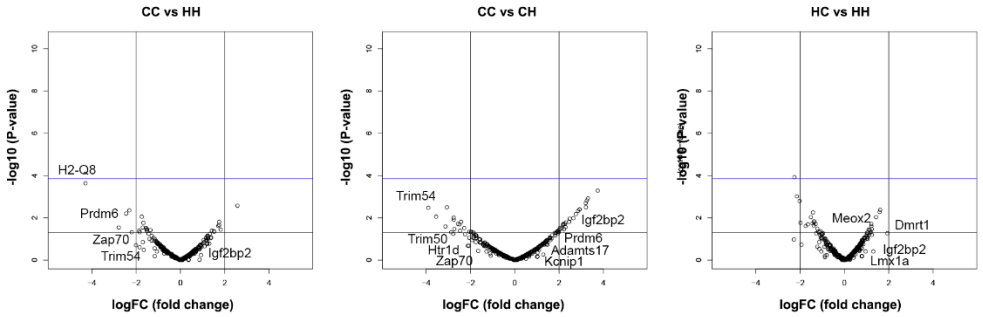
A


H3K4me1
CC vs HH
CC vs HC
HC vs HH
-log10 (P-value)
logFC (fold change)
Glo1
Pisd-ps1
Glo1
Pdap1

B


H3K4me3
CC vs HH
CC vs CH
HC vs HH
-log10 (P-value)
logFC (fold change)
Tmem181-ps
Nrros
10kb upstream Pgm5
Vezf1
Qk
Pde4d
Pde4dip
Pisd-ps3
Chd5

C


H3K27me3
CC vs HH
CC vs CH
HC vs HH
-log10 (P-value)
logFC (fold change)
H2-Q8
Prdm6
Zap70
Trim54
Igf2bp2
Trim54
Trim50
Htr1d
Zap70
Igf2bp2
Prdm6
Adamts17
Kcnip1
Meox2
Dmrt1
Igf2bp2
Lmx1a

D

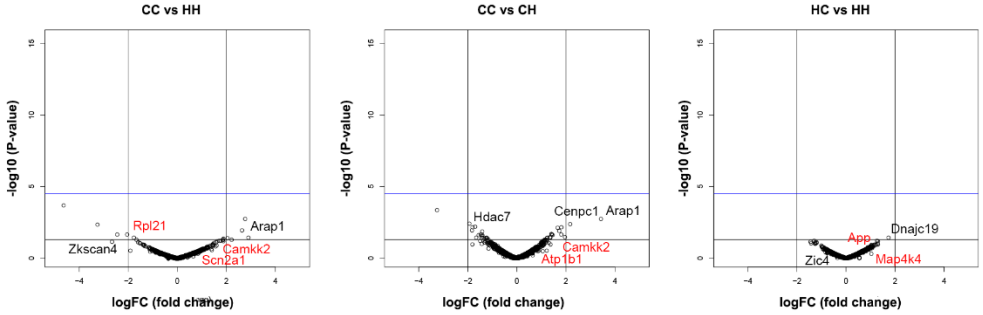
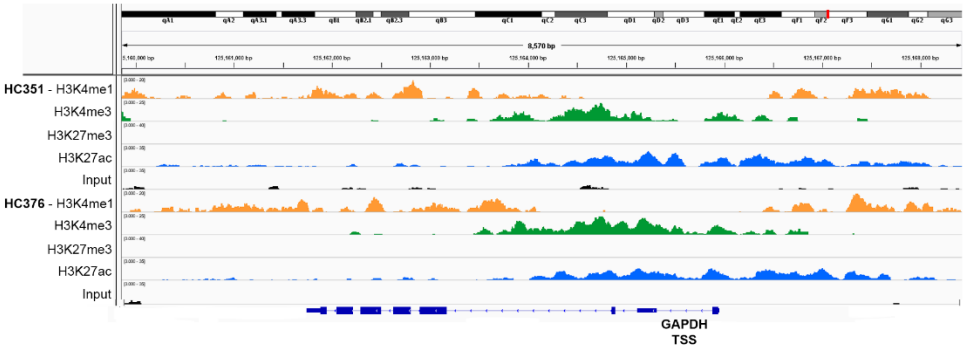

H3K27ac
CC vs HH
CC vs CH
HC vs HH
-log10 (P-value)
logFC (fold change)
Rpl21
Arap1
Zkscan4
Camkk2
Scn2a1
Hdac7
Cenpc1
Arap1
Camkk2
Atp1b1
App
Dnajc19
Zic4
Map4k4

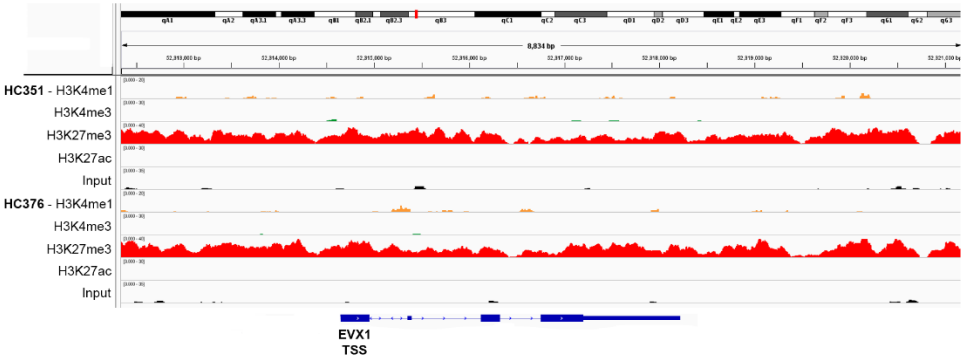
Figure 5 Volcano plots of post-differential binding analysis. A) Differential H3K4me1 (enhancer/gene body mark; active transcription) loci. B) Differential H3K4me3 (promoter region mark; active transcription) loci. C) Differential H3K27me3 (repressed transcription). D) Differential H3K27ac loci (enhancer/gene body mark; active transcription). Black bar indicates a P-value threshold of 0.05. Blue bar indicates an adjusted P-value via Bonferroni correction. Red circles indicate hits above the adjusted threshold. Red gene names indicate direct matches with previous RNA-Seq data. Log scale (base 2) used on abscissa, where log of 1 = two-fold change.

Lastly, in order to visually verify changes in histone modification profiles between strains, IGV (Integrative Genomics Viewer) was used. Data tracks were loaded into IGV (including input tracks) and, initially, well-known active or suppressed gene loci were viewed. As can be seen in figure 6, H3K4me1, H3K4me3, and H3K27ac marks were present within the GAPDH gene. Specifically, H3K4me1 can be seen to localize upstream of the TSS (in addition to within the gene itself), H3K4me3 is generally located close to the TSS and initial exons, and H3K27ac closely matches the pattern exhibited by H3K4me3. Importantly, H3K27me3 is completely absent at this locus. Upon viewing a well-known repressed gene, such as EVX1, the only strong signal is H3K27me3. After determining that the observed signal patterns at control loci closely matched those of published ENCODE ChIP-Seq data sets, other loci were examined to determine whether or not the reported signals mapped back to known genes. One of the hits from the RNA-Seq data set (Glo1) was reported to have a differential H3K4me1 profile in the ChIP-Seq dataset as well. As can be seen in figure 5, when comparing CC140 and CC142 with HC351 and HC376, there is more H3K4me1 signal throughout the Glo1 locus. The top gene hits for the various comparisons have been provided in figure 7.

A


HC351 - H3K4me1
H3K4me3
H3K27me3
H3K27ac
Input
HC376 - H3K4me1
H3K4me3
H3K27me3
H3K27ac
Input
GAPDH
TSS

B


HC351 - H3K4me1
H3K4me3
H3K27me3
H3K27ac
Input
HC376 - H3K4me1
H3K4me3
H3K27me3
H3K27ac
Input
EVX1
TSS

C

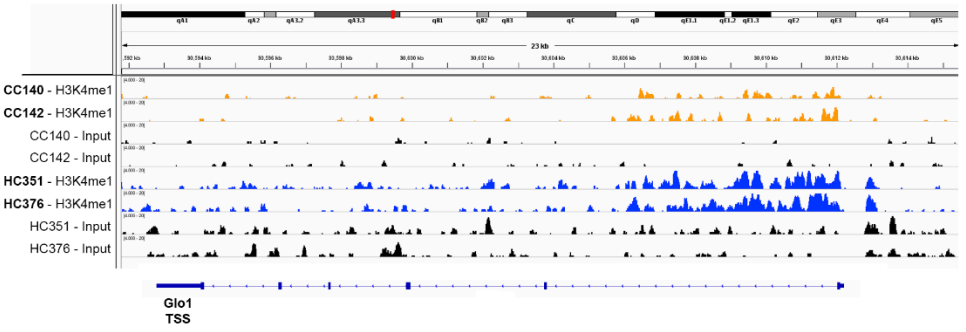

CC140 - H3K4me1
CC142 - H3K4me1
CC140 - Input
CC142 - Input
HC351 - H3K4me1
HC376 - H3K4me1
HC351 - Input
HC376 - Input
Glo1
TSS

Figure 6 Visualization of ChIP-Seq data using IGV (integrative genomic viewer). A) Verification of peak pattern at a well-known, actively-expressed control (GAPDH). Two replicate mice of one strain (HC) shown. All four antibodies loaded into tracks. H3K4me1 (orange); H3K4me3 (green), H3K27me3 (red), H3K27ac (blue), Input tracks for each strain (black). B) Verification of peak pattern at a well-known, repressed control (Evx1). Two replicate mice of one strain (HC) shown. All four antibodies loaded into tracks. H3K4me1 (orange); H3K4me3 (green), H3K27me3 (red), H3K27ac (blue), Input tracks for each strain (black). C) Verification of differential binding results at the Glo1 locus. Two replicates from each strain CC and HH shown. Only one antibody shown. H3K4me1 for CC (orange) and for HC (blue). Input tracks for each strain (black).

H3K4me1

CC vs HH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
-1.7492	1.7492	0.0805	1	chr17	13550794	13551294	NR_045437	2700054A10Rik	3050	Intron
1.2137	1.2137	0.7108	1	chr11	3126968	3127468	NR_003517	Plsd-ps1	3198	Intron
-1.0933	1.0933	0.0450	1	chr12	113100207	113100707	NM_054081	Mta1	2180	Intron
-1.0241	1.0241	0.0692	1	chr7	79484733	79465233	NM_017462	Polg	1290	Exon/CD
-0.9709	0.9709	0.0981	1	chr11	3964869	3965369	NM_022889	Pes1	1145	Intron

CC vs HC

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
-3.1691	3.1691	0.0013	0.0521	chr17	13550794	13551294	NR_045437	2700054A10Rik	3050	Intron
1.5365	1.5365	0.2087	0.8592	chr11	3126968	3127468	NR_003517	Plsd-ps1	3198	Intron
1.0858	1.0858	0.0141	0.2348	chr17	30609511	30610011	NM_025374	Glo1	2898	Intron
-0.9266	0.9266	0.0310	0.3673	chr12	113100207	113100707	NM_054081	Mta1	2180	Intron
0.9085	0.9085	0.0439	0.4053	chr4	151639053	151639553	NM_001081557	Camta1	222465	Intron

HC vs HH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
1.422761953	1.4228	0.478100733	1	chr17	13550794	13551294	NR_045437	2700054A10Rik	3050	Intron
-1.2013069	1.2013	0.015491143	0.4285883	chr7	130545894	130546394	NM_001162855	Nsmce4a	1237	Intron
-1.172828116	1.1728	0.011197542	0.4285883	chr5	145138544	145139044	NM_001033313	Pdap1	1295	Intron
-1.166821847	1.1668	0.010186784	0.4285883	chr17	30609511	30610011	NM_025374	Glo1	2898	Intron
-0.951668066	0.9517	0.038341755	0.505126161	chr16	56036516	56037016	NM_029092	Trmt10c	1008	Intron

H3K4me3

CC vs HH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
-4.471582878	4.4716	7.62E-17	1.32E-13	chr13	119488600	119489100	NM_001039244	Gm7120	812	Intron
-3.03025783	3.0303	0.008789627	1	chr17	13552629	13553129	NR_045437	2700054A10Rik	1215	Intron
-2.579878872	2.5799	0.016535572	1	chr16	32149526	32150026	NM_146069	Lrrc33	15700	Intron
-1.924711222	1.9247	2.25E-05	0.005545929	chr2	121506072	121506572	NM_026220	Mfap1a	334	Intron
-1.88267344	1.8827	0.000185253	0.039991477	chr18	40307932	40308452	NM_026135	Kctd16	49832	Intron

CC vs CH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
-2.774453196	2.7745	0.000504553	0.024204537	chr3	97887268	97887768	NM_001039376	Pde4dip	1189	Intron
-2.489370445	2.4894	0.000730316	0.030762348	chr13	94175888	94176388	NM_172589	Lhfpl2	118343	Intron
-2.370121583	2.3701	3.91E-06	0.000844661	chr10	80141743	80142243	NM_025313	Atp5d	-321	Upstream
-2.321351548	2.3214	6.09E-06	0.001168715	chr18	40307932	40308452	NM_026135	Kctd16	49832	Intron
-2.251979665	2.2520	0.000401399	0.023671134	chrx	70365093	70365593	NM_010498	Ids	-258	Upstream

HC vs HH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
2.049335614	2.0493	0.025592018	1	chr8	126593596	126594096	NM_001164598	Irf2bp2	-410	Upstream
1.724984319	1.7250	0.136621211	1	chr13	119598811	119599312	NM_001039244	Gm7120	111023	Intron
1.6893034	1.6893	0.005888583	1	chr13	119428343	119428843	NM_001079849	Paip1	-6	Upstream
1.628897548	1.6269	0.005505436	1	chr4	152338755	152339255	NM_001081376	Chd5	355	Exon/CD
1.533703108	1.5337	0.011737799	1	chr3	95356428	95356928	NM_018877	Setdb1	524	Intron

H3K27me3

CC vs HH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
-4.289420545	4.2894	0.000235762	0.083695597	chr17	35424479	35424979	NM_023124	H2-Q8	-117	Upstream
-2.782475674	2.7825	0.02942536	0.760630186	chr18	53505258	53505758	NM_001033281	Prdm6	40963	Intron
-2.304463531	2.3045	0.004464192	0.528262673	chr1	36771119	36771619	NM_009539	Zap70	9509	Intron
-1.827413074	1.8274	0.049298197	0.773894886	chr5	31116690	31117190	NM_021447	Trim54	329	Exon/CD
1.822158517	1.8222	0.037034204	0.760630186	chr14	61043671	61044171	NM_013869	Tnfrsf19	2934	Intron

CC vc CH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
-3.894568578	3.8946	0.003310246	0.162177937	chr5	31116690	31117190	NM_021447	Trim54	329	Exon/CD
3.249608592	3.2496	0.001506183	0.162177937	chr7	70377021	70377521	NR_045850	B130024G19Rik	11889	Intron
2.934865156	2.9349	0.004052924	0.162177937	chr18	53477652	53478152	NM_001033281	Prdm6	13357	Intron
-2.773970248	2.7740	0.059423352	0.427824418	chr14	104715622	104716122	NR_015593	D130009I18Rik	76726	Intron
-2.766007134	2.7660	0.010029807	0.260176347	chr5	135353609	135354109	NM_178240	Trim50	564	Intron

HC vs HH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
-2.150214671	2.1502	0.000985664	0.174955273	chr1	36771119	36771619	NM_009539	Zap70	9509	Intron
-2.021405125	2.0214	0.001603289	0.189722518	chr11	33754693	33755193	NM_001190886	Kcnip1	238250	Intron
-1.982361064	1.9824	0.017244259	0.479399794	chr5	31116690	31117190	NM_021447	Trim54	329	Exon/CD
1.593379344	1.5934	0.005312743	0.334681054	chr2	105112768	105113268	NR_015462	AI314831	13492	Intron
-1.519296755	1.5193	0.009623994	0.427064735	chr11	24075800	24076300	NM_001159289	Bcl11a	-2005	Upstream

H3K27ac

CC vs HH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
-3.251838804	3.2518	-0.996379759	0.004516293	chr17	6979068	6979568	NM_026611	Rnaset2b	459	Intron
2.901360379	2.9014	-1.231867964	0.037947512	chr7	101373802	101374302	NM_001040111	Arap1	25984	Intron
-2.664519412	2.6645	-1.729606696	0.07403928	chr11	106031555	106032056	NM_001037712	Kcnh6	23603	Intron
2.641225626	2.6412	-0.540306382	0.01184837	chr17	17830253	17830753	NR_029730	Mirlet7e	152	Downstream
-2.453380018	2.4534	-0.789896481	0.022649015	chr13	21479361	21479861	NM_001039115	Zkscan4	763	Intron

CC vs CH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
3.435470026	3.4355	-0.959664746	0.00180383	chr7	101373802	101374302	NM_001040111	Arap1	25984	Intron
-3.259313996	3.2593	-0.800055775	0.000440542	chr1	135549126	135549626	NM_173437	Nav1	35979	Intron
2.175381892	2.1754	-0.569348264	0.004384946	chr5	86066040	86066540	NM_007683	Cenpc1	-707	Upstream
1.962071532	1.9621	-1.061230512	0.035784937	chr17	17830253	17830753	NR_029730	Mirlet7e	152	Downstream
-1.9314073	1.9314	-0.018843738	0.004052886	chr15	97829073	97829573	NM_001204275	Hdac7	15179	Intron

HC vs HH

logFC	Absolute logFC	Pvalue	FDR	Chromosome	Loci Start	Loci End	Gene ID	Gene Symbol	Distance to TSS	Category
1.728575449	1.7286	0.038884257	1	chr5	39632946	39633446	NM_010474	Hs3st1	11435	Intron
-1.425526966	1.4255	0.095682247	1	chr11	81095756	81096256	NM_001034013	Asic2	872390	Intron
1.303612322	1.3036	0.067717721	1	chr6	28542562	28543062	NM_019776	Snd1	62465	Intron
-1.295658503	1.2957	0.061474933	1	chr9	91378627	91379127	NM_009576	Zic4	9906	Exon/CD
1.275914039	1.2759	0.095163657	1	chr3	95251663	95252163	NM_001168356	Bnipl	-720	Upstream

Figure 7 Top gene hits from annotation of differential binding analysis data. For each antibody, the top five genes (based on absolute log fold-change) were determined for three sets comparisons. logFC = log fold-change (base 2), absolute log fold-change, FDR = false discovery rate, distance to TSS = distance to transcription start site. Data annotated using mm10 and PAVIS. For H3K4me1, nearest gene was defined by 5 kb upstream and 1 kb downstream. For H3K4me1, nearest gene was defined by 5 kb upstream and 1 kb downstream. For H3K4me3, H3K27me3, and H3K27ac, nearest gene was defined by 2.5 bp upstream and 1 kb downstream.

Discussion

Overall, the objectives of the experiment were able to be successfully met. After conducting ChIP-Seq using brain tissue (cerebellum) from different strains of MNX mice, clear differences in histone modification profiles were observed, thus suggesting that mitochondrial haplotype does play a significant role in contributing to the dynamics of histone PTMs.

From the outset, significant emphasis was placed on quality and reproducibility. As recent publications have shown that a sizable proportion of commercial antibodies do not perform as promised, all antibodies used for ChIP-Seq were subject to initial characterization and validation. For histone modification antibody validation, ENCODE project guidelines were closely followed. Upon selection of the antibodies to be used, immunoblot assays using histone-enriched fraction from acid-extracted B16F10 whole cell lysate were run. The results were very encouraging. For all four antibodies tested (H3K4me1, H3K4me3, H3K27me3, and H3K27ac), almost no non-specific bands were observed. Though banding was present at around 50 kDa (whereas the expected molecular weight of histone H3 is around 17 kDa), the bands were very faint and well within ENCODE's suggested criteria of the target band exhibiting (at least) ten-fold signal strength compared to any off-target bands. In addition, the histone-specific band should comprise at least 50% of the total protein signal. As can be readily seen in figure 2, all four antibodies passed this initial characterization step.

Once candidate antibodies were shown to be sufficiently specific via immunoblot analysis, secondary characterization of the antibodies was conducted using a

commercial (EMD-Millipore) peptide array. On this array, 46 common post-translationally modified recombinant histone peptides (at two dot concentrations of 10 and 100 ng) were present, in addition to species-specific IgG positive controls. As depicted in figure 3, all four antibodies exhibited strong binding at the anticipated recombinant histone peptide loci. Unfortunately, determining the optimum concentration for the peptide array proved difficult. Mild chemically-similar off-target binding events were visible for H3K4me1, H3K4me3, and H3K27ac. For H3K4me1 and H3K4me3, the signal strength was so strong (even at low dilutions) that the positive control could only be imaged when the signal from the recombinant peptides was concealed with aluminum foil and the blot reimaged. As such, one can conclude that the off-target events at H3K4me2 for these two antibodies was likely a result of too-high primary concentration. Similarly, H3K27ac exhibited mild off-binding at H3K9ac. Thankfully, all of the off-target binding events observed have the same association with transcription status (active transcription) and, thus, would be unlikely to significantly throw off interpretation of ChIP-Seq results downstream. Furthermore, the primary signals are easily 10-fold stronger compared to the weaker off-binding targets and falls well within ENCODE quality control parameters.

Upon successful primary and secondary validation of the antibodies, ChIP-Seq experiments were conducted and enough DNA for sequencing was collected. Although a minimum of 10 ng per sample was necessary as a bare minimum for sequencing, actual amount of collected ChIP DNA was 20-30 ng (to hedge against the possibility of a failed sequence run and/or inaccurate/low DNA quantitation). After submitting the ChIP DNA to the KUMC sequencing core facility and receiving the fastq data files, initial

quality control using the FastQC program was conducted. Overall the quality parameters looked great for most samples—most samples exhibited low percentages of duplicate reads and high Phred scores (low probability of an inaccurately called base). One set of samples exhibited fairly high duplicate reads (CH374) for the H3K4me1 antibody and thus the CH strain H3K4me1 comparisons were dropped. Furthermore, stochastic errors during the sequencing run itself were observed. As visible in figure 4, one out of seventeen total read batches per sample experienced a sudden drop in Phred score, which was reflected as red and warm-color streaks on the flow cell quality per tile chart. At roughly 80bp in read length, reads in this tile region displayed sudden high probabilities of inaccurate read calls before self-correcting and returning to the baseline level of read quality for the end of the read. To address this, the batch of affected reads (one out of seventeen) were truncated at the 70bp position and retained in the dataset. Upon re-evaluation of read quality with FastQC, the post-correction average read quality returned to levels comparable with other read batches. Following successful quality control measures, all samples passed visual read quality inspection.

As a next step, the sequencing data were used to generate a series of volcano plots in order to examine whether or not the mitochondrial haplotype of the MNX mice had any effects on the overall histone modification profiles observed. Since the lab is interested in the mitochondrial-induced shifts, the three main comparisons between CC-HH, CC-CH, and HH-HC were utilized. In figure 5, clear shifts in the patterns of H3K4me1, H3K4me3, H3K27me3, and H3K27ac were all able to be seen. Unfortunately, as the fold changes are log scale, the more subtle shifts are likely hidden due to most genes densely clustering within the sub-logFC 2 range. Indeed, when using

the Bonferroni correction to adjust the P value in order to minimize the likelihood of selecting false-positive hits (type I errors) when evaluating multiple pair-wise tests on a single dataset, very few statistically significant hits were registered. Only hits for H3K4me3 were above the corrected P value threshold.

Lastly, in order to visually verify that the ChIP-Seq data compared to control loci of published ENCODE datasets and in order to visually determine whether or not the shifts reported by the differential binding analysis are genuine, bigwig files for samples were loaded into IGV and first examined at genetic loci known to be either strongly transcribed or strongly repressed. Visible in figure 6, samples were initially compared to the GAPDH locus, which is known to be an actively-expressed housekeeping gene. Within GAPDH, H3K4me1 was observed in expected regions (throughout the gene body and within enhancer regions), H3K4me3 was localized close to the TSS and initial exons, and H3K27ac followed a pattern closely matching that of H3K4me3. Importantly, H3K27me3 signal is absent at this locus (as would be expected, given that H3K27me3 is associated with repressed transcription and heterochromatin). Upon viewing the EVX1 locus, however, H3K27me3 signal became very evident and the active transcription histone modifications all dramatically decreased. Thus, given that the antibody profiles for the control loci were as had been anticipated, chromosome loci stemming from the DiffBind analysis were visualized. One good example of differential signal strength can be seen in the H3K4me1 signal throughout the Glo1 locus when comparing CC and HC. Clearly, there is significantly stronger signal of H3K4me1 binding in HC (blue) compared to the CC signal (orange). In the RNA-Seq dataset, Glo1 was upregulated 1.98-fold in HC relative to CC. Figure 7 provides a brief overview of

the top hits between each strain comparison. Functions of the various genes range between DNA methylation (Trmt10c), ROS regulation (Nrros), calmodulin-binding transcription activation (Camta1), ATP generation (Atp5d), mitochondrial DNA replication/repair (polg), and histone lysine methylation (Setdb1). At first glance, many of the genes that exhibit differential histone modification profiles are either directly or indirectly tied to metabolism. Upon conducting preliminary pathway analysis (not shown) using all or directional fold-change DiffBind calls within comparisons, processes as diverse as transcript splicing, G-coupled protein receptor and kinase signaling, ATP synthesis, and, reassuringly, chromatin remodeling/histone post-translational modification have been implicated.

As should now be clear, high-quality data was able to be generated using the given antibodies, validation procedures, and ChIP-Seq protocol. Most importantly, the question of whether or not mitochondrial haplotype plays a role in determining the relative histone modification profiles between different MNX mouse strains was affirmatively answered. Given this new information, several new questions emerge: what specifically about the mitochondria is responsible for these changes? Of the genes that were differentially identified, which biological pathways could be responsible for these changes? How do changes in histone modification profile relate to tumor latency and metastasis? Likely candidates for future investigation would be to determine precise changes in overall metabolite levels between the MNX mice (which is already underway in the lab), conducting integrative epigenomic analysis between the RNA-, Methyl-, and ChIP-Seq data sets, or investigating whether or not histone and/or chromatin modifying enzymes could be playing a role in generating these differences. Indeed, previous RNA-

Seq data showed distinct differences in histone acetyltransferase, histone deacetylase, and histone demethylase expression. Either of these factors or the combination of the two could likely be responsible for the observed differences. Another interesting path for future investigation would be to broaden both the total number of histone modifications examined via ChIP-Seq and to even investigate the role of more recently characterized acyl histone modifications and their link to overall metabolism within an organism. While significant, exciting in-roads have been made in regards to the contribution of mitochondrial haplotype to both cancer progression and metastasis, many questions still remain unanswered. Future research will not only further contribute to our understanding of how mitochondria participate in influencing the cellular epigenetic mechanisms involved in fine-tuning gene expression, but has the potential to benefit to all cancer patients.

References

1. Greaves, M., *Nothing in cancer makes sense except.* BMC Biol, 2018. **16**(1): p. 22.
2. McGranahan, N. and C. Swanton, *Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future.* Cell, 2017. **168**(4): p. 613-628.
3. Chen, M.T., et al., *Comparison of patterns and prognosis among distant metastatic breast cancer patients by age groups: a SEER population-based analysis.* Sci Rep, 2017. **7**(1): p. 9254.
4. Statistics, C.-N.C.f.H., *Number of deaths for leading causes of deaths (USA).* 2017.
5. Society, A.C., *Lifetime Risk of Developing or Dying From Cancer.* 2012-2014.
6. Seyfried, T.N. and L.C. Huysentruyt, *On the origin of cancer metastasis.* Crit Rev Oncog, 2013. **18**(1-2): p. 43-73.
7. Budczies, J., et al., *The landscape of metastatic progression patterns across major human cancers.* Oncotarget, 2015. **6**(1): p. 570-83.
8. Piva de Freitas, P., et al., *Metastatic Basal Cell Carcinoma: A Rare Manifestation of a Common Disease.* Case Rep Med, 2017. **2017**: p. 8929745.
9. Beauchesne, P., *Extra-neural metastases of malignant gliomas: myth or reality?* Cancers (Basel), 2011. **3**(1): p. 461-77.
10. Institute, A.C.S.-N.C., *Metastatic Cancer: Where Cancer Spreads.* 2017.
11. Obenauf, A.C. and J. Massague, *Surviving at a distance: organ specific metastasis.* Trends Cancer, 2015. **1**(1): p. 76-91.
12. Butler, T.P. and P.M. Gullino, *Quantitation of cell shedding into efferent blood of mammary adenocarcinoma.* Cancer Res, 1975. **35**(3): p. 512-6.
13. Swartz, M.A., et al., *Cells shed from tumours show reduced clonogenicity, resistance to apoptosis, and in vivo tumorigenicity.* Br J Cancer, 1999. **81**(5): p. 756-9.
14. Institute-SEER, N.C., *Cancer Classification.*
15. Lambert, A.W., D.R. Pattabiraman, and R.A. Weinberg, *Emerging Biological Principles of Metastasis.* Cell, 2017. **168**(4): p. 670-691.
16. Honn, K.V., D.G. Tang, and J.D. Crissman, *Platelets and cancer metastasis: a causal relationship?* Cancer Metastasis Rev, 1992. **11**(3-4): p. 325-51.
17. Kojic, N., et al., *A computational study of circulating large tumor cells traversing microvessels.* Comput Biol Med, 2015. **63**: p. 187-95.
18. Aiello, N.M., et al., *Metastatic progression is associated with dynamic changes in the local microenvironment.* Nat Commun, 2016. **7**: p. 12819.
19. Hurst, R.E., et al., *Targeting dormant micrometastases: rationale, evidence to date and clinical implications.* Ther Adv Med Oncol, 2016. **8**(2): p. 126-37.
20. Yan, J., Q. Yang, and Q. Huang, *Metastasis suppressor genes.* Histol Histopathol, 2013. **28**(3): p. 285-92.
21. Khan, I. and P.S. Steeg, *Metastasis suppressors: functional pathways.* Pathobiology in Focus (Nature), 2018.
22. Manley, S.J., W. Liu, and D.R. Welch, *The KISS1 metastasis suppressor appears to reverse the Warburg effect by shifting from glycolysis to mitochondrial beta-oxidation.* J Mol Med (Berl), 2017. **95**(9): p. 951-963.
23. Lin, L.Z., et al., *BRMS1 gene expression may be associated with clinico-pathological features of breast cancer.* Biosci Rep, 2017. **37**(4).
24. Koyama, R., et al., *Identification and characterization of a metastatic suppressor BRMS1L as a target gene of p53.* Cancer Sci, 2017. **108**(12): p. 2413-2421.
25. Sasaki, K., et al., *Genome-wide in vivo RNAi screen identifies ITIH5 as a metastasis suppressor in pancreatic cancer.* Clin Exp Metastasis, 2017. **34**(3-4): p. 229-239.

26. Hunter, K., *Host genetics influence tumour metastasis.* Nat Rev Cancer, 2006. **6**(2): p. 141-6.
27. Hunter, K.W., et al., *Genetic insights into the morass of metastatic heterogeneity.* Nat Rev Cancer, 2018. **18**(4): p. 211-223.
28. Ettema, T.J., *Evolution: Mitochondria in the second act.* Nature, 2016. **531**(7592): p. 39-40.
29. Berg, O.G. and C.G. Kurland, *Why mitochondrial genes are most often found in nuclei.* Mol Biol Evol, 2000. **17**(6): p. 951-61.
30. Battersby, B.J., J.C. Loredo-Osti, and E.A. Shoubridge, *Nuclear genetic control of mitochondrial DNA segregation.* Nat Genet, 2003. **33**(2): p. 183-6.
31. Fernandez-Vizarra, E., et al., *Tissue-specific differences in mitochondrial activity and biogenesis.* Mitochondrion, 2011. **11**(1): p. 207-13.
32. Youle, R.J. and A.M. van der Bliek, *Mitochondrial fission, fusion, and stress.* Science, 2012. **337**(6098): p. 1062-5.
33. Cole, L.W., *The Evolution of Per-cell Organelle Number.* Front Cell Dev Biol, 2016. **4**: p. 85.
34. Chen, X.J. and R.A. Butow, *The organization and inheritance of the mitochondrial genome.* Nat Rev Genet, 2005. **6**(11): p. 815-25.
35. Miller, F.J., et al., *Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age.* Nucleic Acids Res, 2003. **31**(11): p. e61.
36. Hutchison, C.A., 3rd, et al., *Maternal inheritance of mammalian mitochondrial DNA.* Nature, 1974. **251**(5475): p. 536-8.
37. Radzvilavicius, A.L., N. Lane, and A. Pomiankowski, *Sexual conflict explains the extraordinary diversity of mechanisms regulating mitochondrial inheritance.* BMC Biol, 2017. **15**(1): p. 94.
38. Rojansky, R., M.Y. Cha, and D.C. Chan, *Elimination of paternal mitochondria in mouse embryos occurs through autophagic degradation dependent on PARKIN and MUL1.* Elife, 2016. **5**.
39. Alston, C.L., et al., *The genetics and pathology of mitochondrial disease.* J Pathol, 2017. **241**(2): p. 236-250.
40. Ruhoy, I.S. and R.P. Saneto, *The genetics of Leigh syndrome and its implications for clinical practice and risk management.* Appl Clin Genet, 2014. **7**: p. 221-34.
41. Center, N.N.f.A.T.S.-G.a.R.D.I., *Pyruvate dehydrogenase complex deficiency.* 2017.
42. Center, N.-N.C.f.A.T.S.-G.a.R.D.I., *Pearson syndrome.* 2016.
43. Warburg, O., *"The Metabolism of Carcinoma Cells".* 1924.
44. Kouidhi, S., A.B. Elgaaied, and S. Chouaib, *Impact of Metabolism on T-Cell Differentiation and Function and Cross Talk with Tumor Microenvironment.* Front Immunol, 2017. **8**: p. 270.
45. Strickaert, A., et al., *Cancer heterogeneity is not compatible with one unique cancer cell metabolic map.* Oncogene, 2017. **36**(19): p. 2637-2642.
46. Epstein, T., R.A. Gatenby, and J.S. Brown, *The Warburg effect as an adaptation of cancer cells to rapid fluctuations in energy demand.* PLoS One, 2017. **12**(9): p. e0185085.
47. de Sa Junior, P.L., et al., *The Roles of ROS in Cancer Heterogeneity and Therapy.* Oxid Med Cell Longev, 2017. **2017**: p. 2467940.
48. Galadari, S., et al., *Reactive oxygen species and cancer paradox: To promote or to suppress?* Free Radic Biol Med, 2017. **104**: p. 144-164.
49. Zou, Z., et al., *Induction of reactive oxygen species: an emerging approach for cancer therapy.* Apoptosis, 2017. **22**(11): p. 1321-1335.

50. Yilmazer, A., *Cancer cell lines involving cancer stem cell populations respond to oxidative stress.* Biotechnol Rep (Amst), 2018. **17**: p. 24-30.
51. Zhou, D., L. Shao, and D.R. Spitz, *Reactive oxygen species in normal and tumor stem cells.* Adv Cancer Res, 2014. **122**: p. 1-67.
52. Assi, M., *The differential role of reactive oxygen species in early and late stages of cancer.* Am J Physiol Regul Integr Comp Physiol, 2017. **313**(6): p. R646-R653.
53. Prasad, S., S.C. Gupta, and A.K. Tyagi, *Reactive oxygen species (ROS) and cancer: Role of antioxidative nutraceuticals.* Cancer Lett, 2017. **387**: p. 95-105.
54. Kumari, S., et al., *Reactive Oxygen Species: A Key Constituent in Cancer Survival.* Biomark Insights, 2018. **13**: p. 1177271918755391.
55. Lee, W.T., et al., *Mitochondrial DNA haplotypes induce differential patterns of DNA methylation that result in differential chromosomal gene expression patterns.* Cell Death Discov, 2017. **3**: p. 17062.
56. Alexander N Patananan, A.J.S., Michael A Teitell, *More than a powerplant: the influence of mitochondrial transfer on the epigenome.* Current Opinion in Physiology, 2018.
57. Morgan, T.H., *Sex Limited Inheritance in Drosophila.* Science, 1910. **32**(812): p. 120-2.
58. Sturtevant, A.H., *Genetic Studies on DROSOPHILA SIMULANS. II. Sex-Linked Group of Genes.* Genetics, 1921. **6**(1): p. 43-64.
59. Muller, H.J. and E. Altenburg, *The Frequency of Translocations Produced by X-Rays in Drosophila.* Genetics, 1930. **15**(4): p. 283-311.
60. McClintock, B., *Induction of Instability at Selected Loci in Maize.* Genetics, 1953. **38**(6): p. 579-99.
61. Waddington, C.H. and E.M. Pantelouris, *Transplantation of nuclei in newt's eggs.* Nature, 1953. **172**(4388): p. 1050-1.
62. Gurdon, J.B., *Nuclear transplantation and the control of gene activity in animal development.* Proc R Soc Lond B Biol Sci, 1970. **176**(1044): p. 303-14.
63. Peng, Y. and C.M. Croce, *The role of MicroRNAs in human cancer.* Signal Transduct Target Ther, 2016. **1**: p. 15004.
64. Kim, M. and J. Costello, *DNA methylation: an epigenetic mark of cellular memory.* Exp Mol Med, 2017. **49**(4): p. e322.
65. Hyun, K., et al., *Writing, erasing and reading histone lysine methylations.* Exp Mol Med, 2017. **49**(4): p. e324.
66. Chrun, E.S., F. Modolo, and F.I. Daniel, *Histone modifications: A review about the presence of this epigenetic phenomenon in carcinogenesis.* Pathol Res Pract, 2017. **213**(11): p. 1329-1339.
67. Razin, A. and H. Cedar, *DNA methylation and gene expression.* Microbiol Rev, 1991. **55**(3): p. 451-8.
68. Jin, J., et al., *The effects of cytosine methylation on general transcription factors.* Sci Rep, 2016. **6**: p. 29119.
69. Jang, H.S., et al., *CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function.* Genes (Basel), 2017. **8**(6).
70. Lyko, F., *The DNA methyltransferase family: a versatile toolkit for epigenetic regulation.* Nat Rev Genet, 2018. **19**(2): p. 81-92.
71. Rasmussen, K.D. and K. Helin, *Role of TET enzymes in DNA methylation, development, and cancer.* Genes Dev, 2016. **30**(7): p. 733-50.
72. Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function.* Neuropsychopharmacology, 2013. **38**(1): p. 23-38.
73. Hollister, J.D. and B.S. Gaut, *Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression.* Genome Res, 2009. **19**(8): p. 1419-28.

74. Elhamamsy, A.R., *Role of DNA methylation in imprinting disorders: an updated review.* J Assist Reprod Genet, 2017. **34**(5): p. 549-562.
75. Yang, X., et al., *Gene body methylation can alter gene expression and is a therapeutic target in cancer.* Cancer Cell, 2014. **26**(4): p. 577-90.
76. Zhang, W. and J. Xu, *DNA methyltransferases and their roles in tumorigenesis.* Biomark Res, 2017. **5**: p. 1.
77. Ulrey, C.L., et al., *The impact of metabolism on DNA methylation.* Hum Mol Genet, 2005. **14 Spec No 1**: p. R139-47.
78. An, J., A. Rao, and M. Ko, *TET family dioxygenases and DNA demethylation in stem cells and cancers.* Exp Mol Med, 2017. **49**(4): p. e323.
79. Jones, P.A., J.P. Issa, and S. Baylin, *Targeting the cancer epigenome for therapy.* Nat Rev Genet, 2016. **17**(10): p. 630-41.
80. Curry, E., et al., *Genes Predisposed to DNA Hypermethylation during Acquired Resistance to Chemotherapy Are Identified in Ovarian Tumors by Bivalent Chromatin Domains at Initial Diagnosis.* Cancer Res, 2018. **78**(6): p. 1383-1391.
81. Ehrlich, M., *DNA hypomethylation in cancer cells.* Epigenomics, 2009. **1**(2): p. 239-59.
82. Cock-Rada, A. and J.B. Weitzman, *The methylation landscape of tumour metastasis.* Biol Cell, 2013. **105**(2): p. 73-90.
83. Jones, P.A. and M.L. Gonzalgo, *Altered DNA methylation and genome instability: a new pathway to cancer?* Proc Natl Acad Sci U S A, 1997. **94**(6): p. 2103-5.
84. Flavahan, W.A., E. Gaskell, and B.E. Bernstein, *Epigenetic plasticity and the hallmarks of cancer.* Science, 2017. **357**(6348).
85. Vizoso, M. and M. Esteller, *DNA methylation plasticity contributes to the natural history of metastasis.* Cell Cycle, 2015. **14**(18): p. 2863-4.
86. Palazzo, A.F. and E.S. Lee, *Non-coding RNA: what is functional and what is junk?* Front Genet, 2015. **6**: p. 2.
87. Gu, S. and M.A. Kay, *How do miRNAs mediate translational repression?* Silence, 2010. **1**(1): p. 11.
88. Oliveto, S., et al., *Role of microRNAs in translation regulation and cancer.* World J Biol Chem, 2017. **8**(1): p. 45-56.
89. Svoronos, A.A., D.M. Engelman, and F.J. Slack, *OncomiR or Tumor Suppressor? The Duplicity of MicroRNAs in Cancer.* Cancer Res, 2016. **76**(13): p. 3666-70.
90. Hurst, D.R., M.D. Edmonds, and D.R. Welch, *Metastamir: the field of metastasis-regulatory microRNA is spreading.* Cancer Res, 2009. **69**(19): p. 7495-8.
91. Li, G. and D. Reinberg, *Chromatin higher-order structures and gene regulation.* Curr Opin Genet Dev, 2011. **21**(2): p. 175-86.
92. Dong, X. and Z. Weng, *The correlation between histone modifications and gene expression.* Epigenomics, 2013. **5**(2): p. 113-6.
93. Goyanes, V.J., *Sequential staining of euchromatic and heterochromatic regions of the human Y chromosome.* J Med Genet, 1980. **17**(6): p. 468-71.
94. Almassalha, L.M., et al., *The Global Relationship between Chromatin Physical Topology, Fractal Structure, and Gene Expression.* Sci Rep, 2017. **7**: p. 41061.
95. Gibbons, R.J., *Histone modifying and chromatin remodelling enzymes in cancer and dysplastic syndromes.* Hum Mol Genet, 2005. **14 Spec No 1**: p. R85-92.
96. Ramakrishnan, V., *Histone structure and the organization of the nucleosome.* Annu Rev Biophys Biomol Struct, 1997. **26**: p. 83-112.
97. Marino-Ramirez, L., I.K. Jordan, and D. Landsman, *Multiple independent evolutionary solutions to core histone gene regulation.* Genome Biol, 2006. **7**(12): p. R122.
98. Nelson, D.M., et al., *Coupling of DNA synthesis and histone synthesis in S phase independent of cyclin/cdk2 activity.* Mol Cell Biol, 2002. **22**(21): p. 7459-72.

99. Talbert, P.B. and S. Henikoff, *Histone variants on the move: substrates for chromatin dynamics.* Nat Rev Mol Cell Biol, 2017. **18**(2): p. 115-126.
100. Henikoff, S. and M.M. Smith, *Histone variants and epigenetics.* Cold Spring Harb Perspect Biol, 2015. **7**(1): p. a019364.
101. Celeste, A., et al., *Genomic instability in mice lacking histone H2AX.* Science, 2002. **296**(5569): p. 922-7.
102. Douet, J., et al., *MacroH2A histone variants maintain nuclear organization and heterochromatin architecture.* J Cell Sci, 2017. **130**(9): p. 1570-1582.
103. Perche, P.Y., et al., *Higher concentrations of histone macroH2A in the Barr body are correlated with higher nucleosome density.* Curr Biol, 2000. **10**(23): p. 1531-4.
104. Chen, Y., et al., *H2A.B facilitates transcription elongation at methylated CpG loci.* Genome Res, 2014. **24**(4): p. 570-9.
105. Chadwick, B.P. and H.F. Willard, *A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome.* J Cell Biol, 2001. **152**(2): p. 375-84.
106. Benton, C.B., W. Fiskus, and K.N. Bhalla, *Targeting Histone Acetylation: Readers and Writers in Leukemia and Cancer.* Cancer J, 2017. **23**(5): p. 286-291.
107. Zhang, T., S. Cooper, and N. Brockdorff, *The interplay of histone modifications - writers that read.* EMBO Rep, 2015. **16**(11): p. 1467-81.
108. Gillette, T.G. and J.A. Hill, *Readers, writers, and erasers: chromatin as the whiteboard of heart disease.* Circ Res, 2015. **116**(7): p. 1245-53.
109. Gong, F., L.Y. Chiu, and K.M. Miller, *Acetylation Reader Proteins: Linking Acetylation Signaling to Genome Maintenance and Cancer.* PLoS Genet, 2016. **12**(9): p. e1006272.
110. Lu, R. and G.G. Wang, *Tudor: a versatile family of histone methylation 'readers'.* Trends Biochem Sci, 2013. **38**(11): p. 546-55.
111. Sanchez, R. and M.M. Zhou, *The PHD finger: a versatile epigenome reader.* Trends Biochem Sci, 2011. **36**(7): p. 364-72.
112. Zhao, S., et al., *Systematic Profiling of Histone Readers in Arabidopsis thaliana.* Cell Rep, 2018. **22**(4): p. 1090-1102.
113. Langst, G. and L. Manelyte, *Chromatin Remodelers: From Function to Dysfunction.* Genes (Basel), 2015. **6**(2): p. 299-324.
114. Kadoch, C. and G.R. Crabtree, *Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics.* Sci Adv, 2015. **1**(5): p. e1500447.
115. Gerhold, C.B. and S.M. Gasser, *INO80 and SWR complexes: relating structure to function in chromatin remodeling.* Trends Cell Biol, 2014. **24**(11): p. 619-31.
116. Kim, J.A. and Y.I. Yeom, *Metabolic Signaling to Epigenetic Alterations in Cancer.* Biomol Ther (Seoul), 2018. **26**(1): p. 69-80.
117. Chatterjee, A., E.J. Rodger, and M.R. Eccles, *Epigenetic drivers of tumourigenesis and cancer metastasis.* Semin Cancer Biol, 2017.
118. Kugel, S., et al., *Identification of and Molecular Basis for SIRT6 Loss-of-Function Point Mutations in Cancer.* Cell Rep, 2015. **13**(3): p. 479-488.
119. Zhong, L., et al., *The histone deacetylase Sirt6 regulates glucose homeostasis via Hif1alpha.* Cell, 2010. **140**(2): p. 280-93.
120. Sebastian, C., et al., *The histone deacetylase SIRT6 is a tumor suppressor that controls cancer metabolism.* Cell, 2012. **151**(6): p. 1185-99.
121. Ioris, R.M., et al., *SIRT6 Suppresses Cancer Stem-like Capacity in Tumors with PI3K Activation Independently of Its Deacetylase Activity.* Cell Rep, 2017. **18**(8): p. 1858-1868.

122. Shaytan, A.K., et al., *Coupling between Histone Conformations and DNA Geometry in Nucleosomes on a Microsecond Timescale: Atomistic Insights into Nucleosome Functions.* J Mol Biol, 2016. **428**(1): p. 221-237.
123. Iwasaki, W., et al., *Contribution of histone N-terminal tails to the structure and stability of nucleosomes.* FEBS Open Bio, 2013. **3**: p. 363-9.
124. Bartova, E., et al., *Histone modifications and nuclear architecture: a review.* J Histochem Cytochem, 2008. **56**(8): p. 711-21.
125. Lawrence, M., S. Daujat, and R. Schneider, *Lateral Thinking: How Histone Modifications Regulate Gene Expression.* Trends Genet, 2016. **32**(1): p. 42-56.
126. Kebede, A.F., et al., *Histone propionylation is a mark of active chromatin.* Nat Struct Mol Biol, 2017. **24**(12): p. 1048-1056.
127. Shanmugam, M.K., et al., *Role of novel histone modifications in cancer.* Oncotarget, 2018. **9**(13): p. 11414-11426.
128. Wongkittichote, P., N. Ah Mew, and K.A. Chapman, *Propionyl-CoA carboxylase - A review.* Mol Genet Metab, 2017. **122**(4): p. 145-152.
129. McDonnell, E., et al., *Lipids Reprogram Metabolism to Become a Major Carbon Source for Histone Acetylation.* Cell Rep, 2016. **17**(6): p. 1463-1472.
130. Wong, C.C., Y. Qian, and J. Yu, *Interplay between epigenetics and metabolism in oncogenesis: mechanisms and therapeutic approaches.* Oncogene, 2017. **36**(24): p. 3359-3374.
131. Zhou, Z., E. Ibekwe, and Y. Chornenkyy, *Metabolic Alterations in Cancer Cells and the Emerging Role of Oncometabolites as Drivers of Neoplastic Change.* Antioxidants (Basel), 2018. **7**(1).
132. Yang, H., et al., *IDH1 and IDH2 mutations in tumorigenesis: mechanistic insights and clinical perspectives.* Clin Cancer Res, 2012. **18**(20): p. 5562-71.
133. Kernytsky, A., et al., *IDH2 mutation-induced histone and DNA hypermethylation is progressively reversed by small-molecule inhibition.* Blood, 2015. **125**(2): p. 296-303.
134. Reid, M.A., Z. Dai, and J.W. Locasale, *The impact of cellular metabolism on chromatin dynamics and epigenetics.* Nat Cell Biol, 2017. **19**(11): p. 1298-1306.
135. van Gisbergen, M.W., et al., *How do changes in the mtDNA and mitochondrial dysfunction influence cancer and cancer therapy? Challenges, opportunities and models.* Mutat Res Rev Mutat Res, 2015. **764**: p. 16-30.
136. Murata, T., E. Mekada, and R.M. Hoffman, *Reconstitution of a metastatic-resistant tumor microenvironment with cancer-associated fibroblasts enables metastasis.* Cell Cycle, 2017. **16**(6): p. 533-535.
137. Moran, M., et al., *Mitochondrial bioenergetics and dynamics interplay in complex I-deficient fibroblasts.* Biochim Biophys Acta, 2010. **1802**(5): p. 443-53.
138. Yu, M., et al., *Depletion of mitochondrial DNA by ethidium bromide treatment inhibits the proliferation and tumorigenesis of T47D human breast cancer cells.* Toxicol Lett, 2007. **170**(1): p. 83-93.
139. Tu, Y.F., et al., *Mitochondria of highly metastatic breast cancer cell line MDA-MB-231 exhibits increased autophagic properties.* Biochim Biophys Acta, 2011. **1807**(9): p. 1125-32.
140. Kaipparettu, B.A., et al., *Crosstalk from non-cancerous mitochondria can inhibit tumor properties of metastatic cells by suppressing oncogenic pathways.* PLoS One, 2013. **8**(5): p. e61747.
141. Bussard, K.M. and L.D. Siracusa, *Understanding Mitochondrial Polymorphisms in Cancer.* Cancer Res, 2017. **77**(22): p. 6051-6059.
142. Yu, X., et al., *Dissecting the effects of mtDNA variations on complex traits using mouse conplastic strains.* Genome Res, 2009. **19**(1): p. 159-65.

143. Kretzschmar, C., et al., *Polymorphisms of the murine mitochondrial ND4, CYTB and COX3 genes impact hematopoiesis during aging.* Oncotarget, 2016. **7**(46): p. 74460-74472.
144. Yokota, M., et al., *Generation of trans-mitochondrial mito-mice by the introduction of a pathogenic G13997A mtDNA from highly metastatic lung carcinoma cells.* FEBS Lett, 2010. **584**(18): p. 3943-8.
145. Hashizume, O., et al., *Specific mitochondrial DNA mutation in mice regulates diabetes and lymphoma development.* Proc Natl Acad Sci U S A, 2012. **109**(26): p. 10528-33.
146. Feeley, K.P., et al., *Mitochondrial Genetics Regulate Breast Cancer Tumorigenicity and Metastatic Potential.* Cancer Res, 2015. **75**(20): p. 4429-36.
147. Brinker, A.E., et al., *Mitochondrial Haplotype Alters Mammary Cancer Tumorigenicity and Metastasis in an Oncogenic Driver-Dependent Manner.* Cancer Res, 2017. **77**(24): p. 6941-6949.
148. Vivian, C.J., et al., *Mitochondrial Genomic Backgrounds Affect Nuclear DNA Methylation and Gene Expression.* Cancer Res, 2017. **77**(22): p. 6202-6214.