ROLES OF HISTONE-DNA INTERACTIONS AND HISTONE H1 IN REGULATING CHROMATIN STRUCTURE IN PLANT AND FUNGAL MODEL SYSTEMS

by

ALEXANDRE MATTE SANTOS

(Under the Direction of XIAOYU ZHANG)

ABSTRACT

We examined the roles of histone-DNA interactions and histone H1 in regulating chromatin structure in *Arabidopsis thaliana* and *Neurospora crassa*. Using K-means clustering, we categorized Arabidopsis genes based on their nucleosome organization. Results showed that for a subset of actively transcribed genes lacked stereotypical organization with no discernible consequences on transcriptional level. This difference in organization seems to be predicted by the surrounding DNA sequence and may be related to transcriptional consistency and inducibility. We also examined the effects on chromatin structure in Neurospora following the loss of histone H1. Unexpectedly we saw little difference in nucleosome organization in the H1 deletion mutant; however these result should not discount the potential role of histone H1 in regulation chromatin structure in other organisms.

INDEX WORDS: Histone-DNA interactions, nucleosome organization, development, Histone H1, Neurospora crassa, Arabidopsis thaliana, transcriptional regulation

ROLES OF HISTONE-DNA INTERACTIONS AND HISTONE H1 IN REGULATING CHROMATIN STRUCTURE IN PLANT AND FUNGLA MODEL SYSTEMS

by

ALEXANDRE MATTE SANTOS

B.S., University of Georgia, 2011

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2015

ROLES OF HISTONE-DNA INTERACTIONS AND HISTONE H1 IN REGULATING CHROMATIN STRUCTURE IN PLANT AND FUNGLA MODEL SYSTEMS

by

ALEXANDRE MATTE SANTOS

Major Professor:	Xiaoyu Zhang
Committee:	Zachary Lewis
	Wolfgang Lukowitz

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2015

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

LITERATURE REVIEW

Chromatin is a complex of DNA and proteins that serves many structural and regulatory roles in the cell. The nucleosome, an octamer composed of two copies of histones H2A, H2B, H3 and H4, wrapped by ~146 bp of DNA, is the most basic subunit of chromatin. Chromatin was initially believed to play an exclusively structural role; specifically in the compaction of DNA. More recent results however have shown that chromatin also plays an important role in the regulation of gene transcription. There are many proposed mechanisms by which this occurs but this thesis focuses on one: the organization of nucleosomes into landscapes that can regulate transcriptional level and variability.

Epigenetics

Epigenetics is the study of heritable changes in transcription which cannot be attributed to DNA sequence("Epigenetic Mechanisms of Gene Regulation (1996, Volume 32)," n.d.). This definition allows for the inclusion of a multitude of processes that have been demonstrated to influence transcriptional regulation. There are three extensively documented epigenetic

factors that have been shown to be critical for proper transcriptional regulation: DNA cytosine methylation, histone post-translational modifications, and chromatin structure.

Cytosine DNA methylation

Cytosine methylation has been shown to be involved in transcriptional regulation for almost all major model system(Cao & Jacobsen, 2002; Cokus et al., 2008; Feng et al., 2010; Okano, Bell, Haber, & Li, 1999; Tamaru & Selker, 2001). Mutations that interfere with the establishment, maintenance, or reading/detection of methylation marks are often accompanied by severe developmental phenotypes. It is also the only epigenetic factor for which there is a mechanism for meiotic inheritance stability(Bird, 2002; Johannes et al., 2009; Schmitz et al., 2013). Interestingly, while DNA methylation is equally important in various organisms, there is considerable variation between model systems in terms of distribution and regulatory functions(Steven R. Eichten, Vaughn, Hermanson, & Springer, 2013; Feng et al., 2010; Meissner et al., 2008). In animals systems, for example, methylation is present throughout the entire genome except for CpG islands in promoter regions of actively transcribed genes(Kinney et al., 2011; Okano et al., 1999). There have been hundreds of documented tissue-specific differentially methylated regions that correlate perfectly with transcription, demonstrating the necessity of methylation for proper tissue-specific regulation(Kinney et al., 2011; Meissner et al., 2008). Plants, however, show a much sparser distribution of methylation, with no clear correlation between methylation level and transcriptional activity(Steve R Eichten et al., 2011; Steven R. Eichten et al., 2013). Results have shown that the context of gene body methylation can be correlated with certain transcriptional

behavior, but these differ significantly from animals models. For example, transposable elements and other permanently silenced regions will be methylated in the CHH context (where H is any base other than G), and never expressed; in contrast, genes that are constitutively expressed, such as housekeeping genes, will often feature extensive body methylation in the CG context(Cokus et al., 2008; Suzuki & Bird, 2008; Zilberman, Gehring, Tran, Ballinger, & Henikoff, 2007). These significant differences between models have made the study of cytosine methylation particularly difficult, nevertheless it remains an area of research for many labs.

Histone modifications

Histone modifications along with DNA methylation are some of the most extensively studied epigenetic factors. While there are a multitude of different types of modifications, generically they are small molecules covalently bonded to the long, largely unstructured tail of histones. Some of the best studied modifications include methylation, phosphorylation, and acetylation, but new modifications are still being discovered(Tan et al., 2011). Modifications vary not only based off of type but also based on the amino acid residue to which they're ligated. A small number of modifications have risen to prominence following their characterization; these include trimethylation of lysine 27 on histone H3 (H3K27me3), trimethylation of lysine 4 on histone H3 (H3K4me3), and trimethylation of lysine 27 on histone H3 (H3K36me3) (Akkers et al., 2009; Karlić, Chung, Lasserre, Vlahovicek, & Vingron, 2010; Sims & Reinberg, 2009; Wei et al., 2009). Similar to DNA methylation, histone modifications have been documented in every model system to date. The deletion of proteins responsible for ligation or binding these modifications often has severe developmental consequences(Bhaumik,

Smith, & Shilatifard, 2007; He & Amasino, 2005; Li, 2002; Lin & Dent, 2006). David Allis famously proposed the existence of a "histone code", a combination of histone modifications in specific combinations and distributions which would act as instructions for the cell in terms of transcriptional programing(Jenuwein & Allis, 2001). Some rough "rules" of this code have been proposed such as, H3K4me3/2/1 and H3K36me3 being found on actively transcribed genes, and H3K27me3/2/1 found on developmentally silenced genes(Akkers et al., 2009; Karlić et al., 2010; Roudier et al., 2011; Sims & Reinberg, 2009; Wei et al., 2009). However given the differences in histone modifications between model systems and the extensive backlog of uncharacterized and undiscovered histone modifications it is unlikely the histone code will be fully characterized in the near future(Berger, 2007).

Chromatin structure

The study of chromatin structure seeks to understand how large and small scale changes in chromatin are being used to regulate transcription. For the purposes of this paper I'll focus on small scale changes in nucleosome organization in, and bordering, genic regions. Given the early assumption of chromatin acting solely as a compacting agent, it is unsurprising that nucleosomes were mainly seen as scaffolding proteins, ubiquitously but randomly distributed throughout the genome. Most information regarding nucleosomes organization in genes came only from a few well studied genes in *Saccharomyces cerevisiae,* PHOS5, GAL10, HIS3(Cavalli & Thoma, 1993; Durrin, Mann, & Grunstein, 1992; Han & Grunstein, 1988; Paranjape, Kamakaka, & Kadonaga, 1994). With the advent of microarray technology came the field of genomics but early, low density microarrays could only show that there was a general depletion of

nucleosomes in regions upstream or transcription start sites(Bernstein, Liu, Humphrey, Perlstein, & Schreiber, 2004; C.-K. Lee, Shibata, Rao, Strahl, & Lieb, 2004). As technology improved it became obvious that nucleosomes were not randomly distributed especially bordering genic regions that were actively transcribed. This unexpected result became even more interesting when it was shown that highly transcribed genes had well-organized nucleosomes and poorly expressed genes did not(Yuan et al., 2005). One proposed mechanism was that histone-DNA interactions could help position nucleosomes by creating a barrier against which to anchor nucleosomes(Kaplan et al., 2010; Segal & Widom, 2009a; Suter, 2000). An alternative mechanism suggested that chromatin remodelers were responsible for the non-random distribution; these DNA translocases had been previously shown to be capable of directly reorganizing nucleosomes using a wide variety of mechanisms(Valouev et al., 2008; Yen, Vinayachandran, Batta, Koerber, & Pugh, 2012; Zhang et al., 2009). Unfortunately this disagreement fueled the existence of a dichotomy as groups lined up behind their mechanism of choice and preached its dominance in positioning and organizing nucleosomes(Kaplan et al., 2009; Valouev et al., 2008; Yen et al., 2012; Zhang et al., 2009).

Histone-DNA interactions

The importance of histone-DNA interactions in positioning nucleosomes was largely championed by the Widom lab, whose results showed definitively that the DNA sequences associated with nucleosomes is not random(Anderson & Widom, 2000; W. Lee et al., 2007; Segal & Widom, 2009b; Thåström et al., 1999; Widom, 2002; Yuan et al., 2005). They described two mechanisms for positioning nucleosomes: sequences with offset periodicities of GC and AT

rich dinucleotides for anchoring nucleosomes, and large poly-A tracts for creating barriers(Kaplan et al., 2009, 2010; Segal & Widom, 2009a). Results showed that the ideal nucleosomes DNA features GC dinucleotide rich sequences every 10 bp and AT rich dinucleotides also in a 10bp phase, interposed by 5bp. There is evidence that GC rich dinucleotides have an increased flexibility which could function as hinges around allowing the DNA to bend more easily when contouring the nucleosome(Kaplan et al., 2009, 2010). In contrast, AT rich dinucleotides are more rigid and allow DNA to interface more closely with the sides of the nucleosome(Segal & Widom, 2009a). A combination of both could theoretically produce an environment with much lower potential energy, discouraging movement. The poly-A sequences are a feature of many promoter regions, especially in the nucleosome free region. The rigidity of AA dinucleotides means that even short oligomers (7x-9x) could create a region incapable of bending around nucleosomes, creating a hard barrier against which nucleosomes cannot slide past(Segal & Widom, 2009a; Suter, 2000). The research culminated in the development of models which would predict nucleosome occupancy based entirely on DNA sequence(Kaplan et al., 2009; Morozov et al., 2009; Xi et al., 2010). The models were based entirely on the associated DNA of reconstituted nucleosomes, a technique which involves combining nucleosomes and fragmented DNA in vitro and allowing them to reconstitute free of any type of cell extract. There is some disagreement on the efficacy of these models particularly in predicting the positions of nucleosomes in genic regions(Struhl & Segal, 2013; Valouev et al., 2008; Zhang et al., 2009). This suggests that the more organized nucleosomes, like those found in the genic arrays of transcribed genes for example, may not entirely be the result of histone-DNA interactions. The models nevertheless are important for identifying any

regions where histone-DNA interactions could play a substantial role in positioning(Beh, Kaplan, Muller, Muir, & Landweber, 2014). Research from Pugh has shown that the addition of cell extract, or even individual nucleosomes remodelers, and ATP in in vitro reconstituting nucleosomes can reproduce native nucleosomes distributions with much higher fidelity, again questioning the dominance of histone-DNA interactions in positioning nucleosomes(Yen et al., 2012).

Chromatin remodelers

As the name implies, chromatin remodelers are a class of protein complexes known for their ability to alter, or remodel, chromatin structure by directly positioning, sliding, or ejecting nucleosomes. Remodelers have been classified into four different families based on their catalytic domains, these include: SWI/SNF, ISWI, CHD, and INO80(Clapier & Cairns, 2009; Workman, 2003). The mechanisms for each class of remodelers were mostly characterized in *Saccharomyces cerevisiae*, there is evidence however that homologs in other systems behave similarly, despite some moderate differences(Bao & Shen, 2007; Corona et al., 1999; Marfella & Imbalzano, 2007; I Whitehouse et al., 1999; Iestyn Whitehouse, Stockdale, Flaus, Szczelkun, & Owen-hughes, 2003). The Pugh lab showed how reconstituted nucleosomes do not reproduce observed distributions until chromatin remodelers, in the form of cell extract, and ATP are added to the experiment(Yen et al., 2012). This obviously shows the major role that remodelers play in creating the organized nucleosomes in actively transcribed genes. Like all other epigenetic factors, the deletion of these remodelers can cause significant developmental

aberrations(Bao & Shen, 2007; Corona et al., 1999; Marfella & Imbalzano, 2007; I Whitehouse et al., 1999; Iestyn Whitehouse et al., 2003).

Other factors in nucleosome organization

Histone variants have been of particular interest recently after their signal was shown to correlate with transcriptional activity. The effects these variant histones have, either on histone-DNA interactions or in the recruitment of remodelers, still remain to be characterized. One last factor that may play an important role in nucleosome organization is the presence of Histone H1. Also known as the linker histone, it has been found to bind to the linker regions adjacent to the nucleosome bound DNA(Allan, Hartman, Crane-Robinson, & Aviles, 1980; Noll & Kornberg, 1977; Thoma, 1979). This binding behavior and the paucity of research regarding its role in chromatin structure make it an attractive object of study.

Nucleosome organization and transcriptional regulation

Evidence for the role of nucleosome organization in transcriptional regulation extends beyond the correlation between structure and transcriptional level. Raveh-Sadka et al. showed that by altering the DNA sequence in promoter regions to position nucleosomes and by altering the distance of this sequence from the transcription start site create a molecular “dimmer” switch capable of fine tuning transcriptional level and consistency(Raveh-Sadka et al., 2012). Transcriptional consistency refers to the variability at which transcription occurs in a single cell. Other results also showed that nucleosome organization within genes was also heavily linked to transcriptional consistency, where the genes with more stereotypical nucleosome organization

where the more consistent the transcription rate was(Durrin et al., 1992). In addition, some dynamical models based off of gene regulatory networks have demonstrated that manipulating transcriptional consistency could be one of the major mechanisms for controlling cell fate during development(Alvarez-Buylla et al., 2008; Banerji et al., 2013; Eldar & Elowitz, 2010; Elowitz, Levine, Siggia, & Swain, 2002; Garg, Mohanram, Di Cara, De Micheli, & Xenarios, 2009; McAdams & Arkin, 1997; Raser & O'Shea, 2004).

Objective

The role of chromatin structure in epigenetic regulation has been well established; however it is the only epigenetic factor that has been conclusively shown to affect transcriptional consistency. Likewise the impact of histone-DNA interactions and chromatin remodelers on nucleosome organization is irrefutable. However much remains to be shown about the relative contributions for each on a more selective basis. The objective of this thesis is to help categorize genes by the relative contributions of each mechanism in organizing their nucleosomes, and perhaps gain some insight into the functional differences between these categories. Furthermore there are many factors that still remain to be characterized, with regards to their role in chromatin structure; therefore the second part of this thesis concerns the elucidation of contributions of histone H1 in organizing nucleosomes.

CHAPTER 2

CHARACTERIZING NOVEL DISTRIBUTIONS OF H3K4ME3 IN *ARABIDOPSIS THALIANA*

The purpose of these experiments, initially, was to understand the functional causes and consequences for a novel distribution of H3K4me3 found in a subset of genes. We determined that this distribution was effectively a proxy for the underlying nucleosome organization, which in turn reflected the magnitude of the contributions of Histone-DNA interactions in positioning the first nucleosome.

Materials and methods

Tissue for H3K4me3 chromatin immunoprecipitation was collected from 12-day-old seedlings and 50 base pair, single ended reads were sequenced using the Illumina platform(Shendure & Ji, 2008). Tissue for Micrococcal Nuclease digestion was collected from 12-day-old seedlings, digested with 20 units of Microccal nuclease for 60 minutes and 50 base pair single ended reads were sequenced using the Illumina platform. All reads were mapped using bowtie2 with default parameters, and clustering was performed using the hclust() function from the stats package in R version 3.2.1 (Blashfield, n.d.; Langmead & Salzberg, 2012).

GO term enrichment analysis was done using the Generic GO term finder utility from the Lewis-Sigler institute(Boyle et al., 2004). Predicted nucleosome occupancy was computed using version 3.0 of the models described in Kaplan et al, 2008. Nucleosome calling was performed using DANPOS2 with default parameters (Chen et al., 2013).

All metaplots were generated at single nucleotide resolution and read counts were divided by # of million mapped reads in order to normalize counts to library size.

Identification of new H3K4me3 distributions using K-Means clustering

We clustered genes based on their distribution of H3K4me3 Chromatin Immunoprecipitation-sequencing reads in a 500 bp region centered on the transcription start site (TSS). Reads were mapped to the 2010 TAIR assembly of the Arabidopsis genome using bowtie2 with default parameters. Clustering was performed using the hclust function in R; distance matrices were calculated using Euclidean distance and agglomerative clustering was performed using Ward's method. Using K-means clustering where k = 3 resulted in three classes of genes grouped by the pattern of their read distributions which can be seen in Figure 1. The decision to restrict the number of clusters to only three is undoubtedly under clustering, however as a first pass analysis we wanted to understand what was the next most obvious distribution beyond the two that have already been described. Two of these groups showed distributions that were already well documented: narrow and flat. The first of these contained 13547 genes and recapitulated the previously seen distribution believed to be present in all genes with H3K4me3: a "narrow" peak immediately downstream of the TSS. The second distribution contained 11309 members and corresponded to genes not marked with H3K4me3, the lack of enrichment manifests as a "flat" distribution of reads around the TSS. The third pattern contained 3597 genes and showed a much greater enrichment of reads than the "flat" group but also a much broader distribution than the "narrow" group. Interestingly we found that the genes in this group are disproportionately more likely to be upregulated during heat-stress suggesting that they may

be primed for inducibility. Moreover we found that a significant percentage of genes in this group were enriched for Gene Ontology (GO) terms related to stress response (Table 1). GO term enrichment was calculated using the Lewis-Sigler Institute GoTermFinder. Finally, results from other labs have found similar distributions of H3K4me3 in Arabidopsis and other model systems and also found that they demonstrate greater inducibility(Benayoun et al., 2014). Taken together, these results suggest that genes in the "broad" group may have some functional differences.

Potential causative factors for broad distribution of H3K4me3

We used MNase-seq to identify the nucleosome distribution for these categories of genes. Mononucleosomal DNA was selected by cutting out the ~150 bp sized band following gel electrophoresis. Reads were mapped to the TAIR10 genome using bowtie2 with standard parameters. We found that genes in the narrow group have especially well organized nucleosomes with a prominent/well positioned +1 nucleosome and a densely packed array of nucleosomes (Figure 2, blue line). In contrast the broad group showed a very poorly positioned +1 nucleosome and a distinct lack of nucleosome arrays (Figure 2, red line). We believe that this less stereotypical positioning is the causative reason for the similarly wide distribution of H3K4me3 described previously.

Role of histone-DNA interactions in positioning nucleosomes is minimal in broad group

Nucleosomes organization can be affected by many factors; we sought to characterize the prominence of one of these factors, histone-DNA interactions in each group. To compare the

role of histone-DNA interactions in organizing the nucleosomes we analyzed the predicted nucleosome occupancy using the models developed by the Widom lab. This model calculates the likelihood of a nucleosome occupying a location based purely on the underlying DNA sequence. As input we used 10kb regions DNA sequences, to avoid the boundary effects of the model, corrected for orientation and centered on the TSS for all genes. For each nucleotide the likelihood of being occupied by a nucleosome was calculated; distributions of these likelihoods over a 2kb window centered on the TSS were used for downstream analyses. Results showed a substantial peak in predicted occupancy immediately downstream of the TSS in the narrow and flat groups indicating that DNA is playing a large role in stabilizing the +1 nucleosome in these groups (Figure 4, top panel). Considering the broad group did not have a corresponding peak it suggests that histone-DNA interactions may not play as significant a role in positioning the +1 nucleosome in these genes (Figure 3, top panel). Interestingly, these results also suggest that the contributions of histone-DNA interactions seem to be minimal in organizing nucleosomes downstream of the +1.

Identification of stable +1 nucleosomes and their underlying sequences

We used the DANPOS2 program to identify stable nucleosomes using the MNase-seq datasets. Reads were not used as pairs for input; all other parameters were left as default. Plotting the centers of these called nucleosomes for each group confirms that genes in the narrow group have organized nucleosomes that aren't found in the broad group (Figure 5). To identify the underlying sequences of the +1 nucleosomes we selected called nucleosomes centers within a 400 bp window centered on the TSS. From these nucleosome centers we examined the

dinucleotide frequency in a 500 bp window centered on the called +1 nucleosome and found that GC and AT dinucleotide frequencies for the narrow and broad group differ significantly (Figure 6). The genes in the narrow and flat groups have an enrichment of GC dinucleotides particularly within 100 bp of the +1 nucleosome center, which is not found in genes from the broad group. This seems to indicate that nucleosomes in the broad group depend less on histone-DNA interactions for organization than nucleosomes in the narrow and flat group.

Figure 1. Genes grouped by distribution of H3K4me3 distribution. Distribution of H3K4me3 ChIP-seq reads in a 2000 bp window centered on the TSS. Genes in the "narrow" and "flat" classes recapitulate documented H3K4me3 distributions for transcriptionally active and silent genes, respectively. Genes in the "broad" class showed a much wider distribution of reads relative to the narrow group.

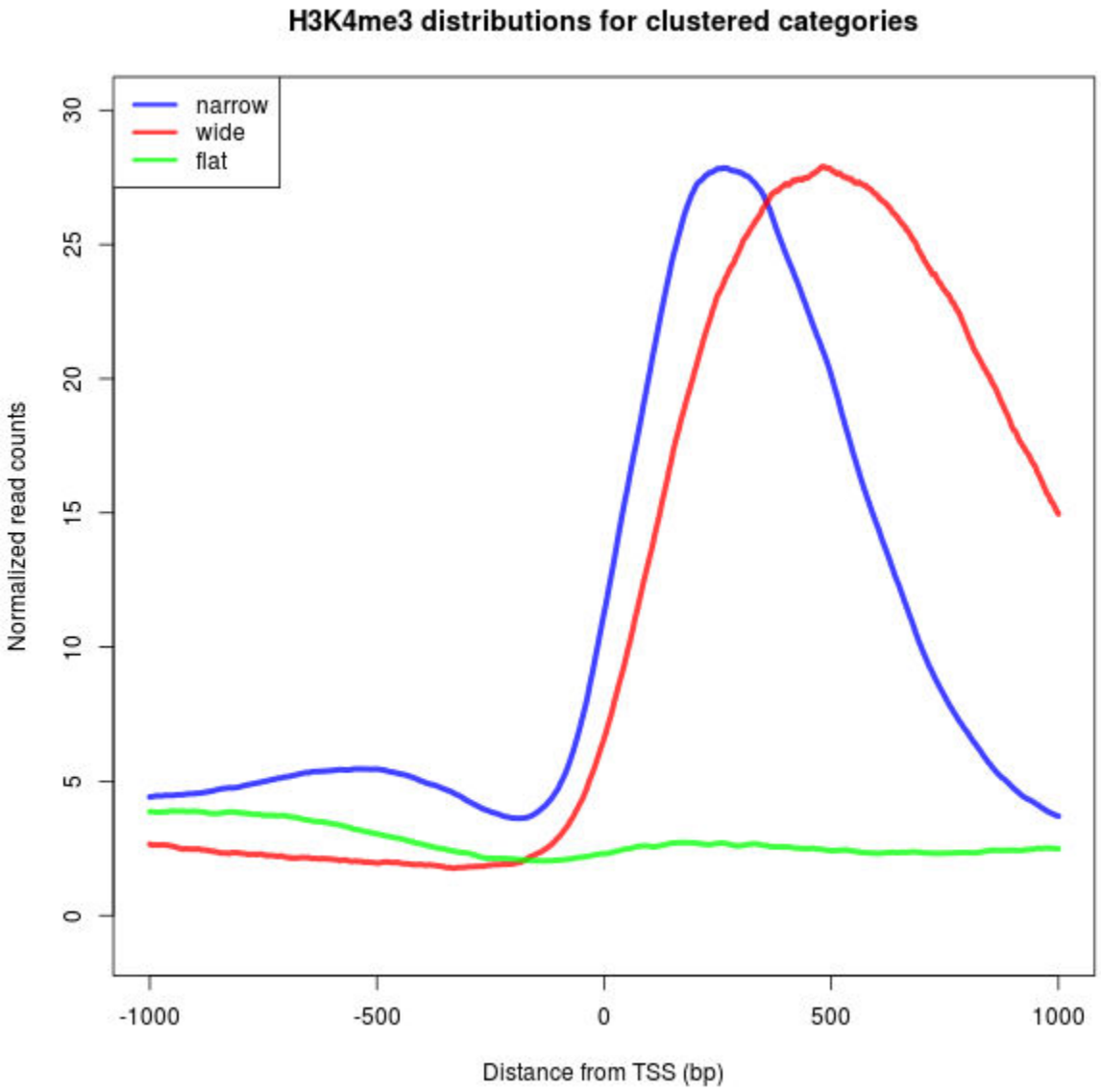


Table 1. Top 35 most significantly enriched GO terms for genes in the wide category. GO term enrichment was performed using online GOTermFinder tool from Princeton University. A significant fraction of genes are classified as being involved in the responses to factors such as stress, chemicals, or abiotic stimuli.

GO Term ID	GO Term	Frequency	log10 p-value
GO:0008152	metabolic process	70.11%	-127.6095
GO:0009987	cellular process	60.34%	-135.2861
GO:0071704	organic substance metabolic process	53.36%	-118.4773
GO:0044238	primary metabolic process	50.98%	-98.978
GO:0044237	cellular metabolic process	49.40%	-109.1738
GO:0044699	single-organism process	48.04%	-144.5691
GO:0044763	single-organism cellular process	36.04%	-133.0559
GO:0044710	single-organism metabolic process	29.63%	-84.3419
GO:0009058	biosynthetic process	28.86%	-79.0964
GO:0006807	nitrogen compound metabolic process	27.80%	-67.3772
GO:1901576	organic substance biosynthetic process	27.67%	-77.5484
GO:1901360	organic cyclic compound metabolic process	26.96%	-65.1922
GO:0006725	cellular aromatic compound metabolic process	26.29%	-63.9391
GO:0065007	biological regulation	25.92%	-144.0842
GO:0050896	response to stimulus	24.28%	-140.9146
GO:0050789	regulation of biological process	23.58%	-144.3171
GO:0050794	regulation of cellular process	20.71%	-116.8367
GO:1901362	organic cyclic compound biosynthetic process	16.58%	-70.3926
GO:0019222	regulation of metabolic process	15.91%	-70.2336
GO:0019438	aromatic compound biosynthetic process	15.78%	-66.6722
GO:0006793	phosphorus metabolic process	15.45%	-65.6112
GO:0006796	phosphate-containing compound metabolic process	15.25%	-66.051
GO:0018130	heterocycle biosynthetic process	15.24%	-59.8656
GO:0006950	response to stress	14.04%	-63.8009
GO:0080090	regulation of primary metabolic process	13.86%	-65.4899
GO:0031323	regulation of cellular metabolic process	13.83%	-63.9551
GO:0044281	small molecule metabolic process	13.04%	-66.6583
GO:0051716	cellular response to stimulus	12.89%	-71.0257
GO:0009889	regulation of biosynthetic process	12.64%	-59.0319
GO:0042221	response to chemical	12.64%	-115.1785
GO:0044711	single-organism biosynthetic process	11.29%	-72.1294
GO:0007154	cell communication	10.06%	-71.3148
GO:0044707	single-multicellular organism process	9.99%	-57.2091
GO:0007275	multicellular organismal development	9.90%	-57.2259

Figure 2. Nucleosome organization is different for each group. Distribution of MNase-seq reads in a 2000 bp window centered on the TSS. Genes in the "narrow" class show a well positioned +1 nucleosome and a densely packed nucleosome array downstream, both of which are characteristic of genes with highly organized nucleosomes. Genes in the "flat" class also featured a surprisingly well positioned +1 nucleosome, but do not have the downstream nucleosome arrays. Genes in the "broad" class do not feature an especially well positioned +1 nucleosome or nucleosome arrays.

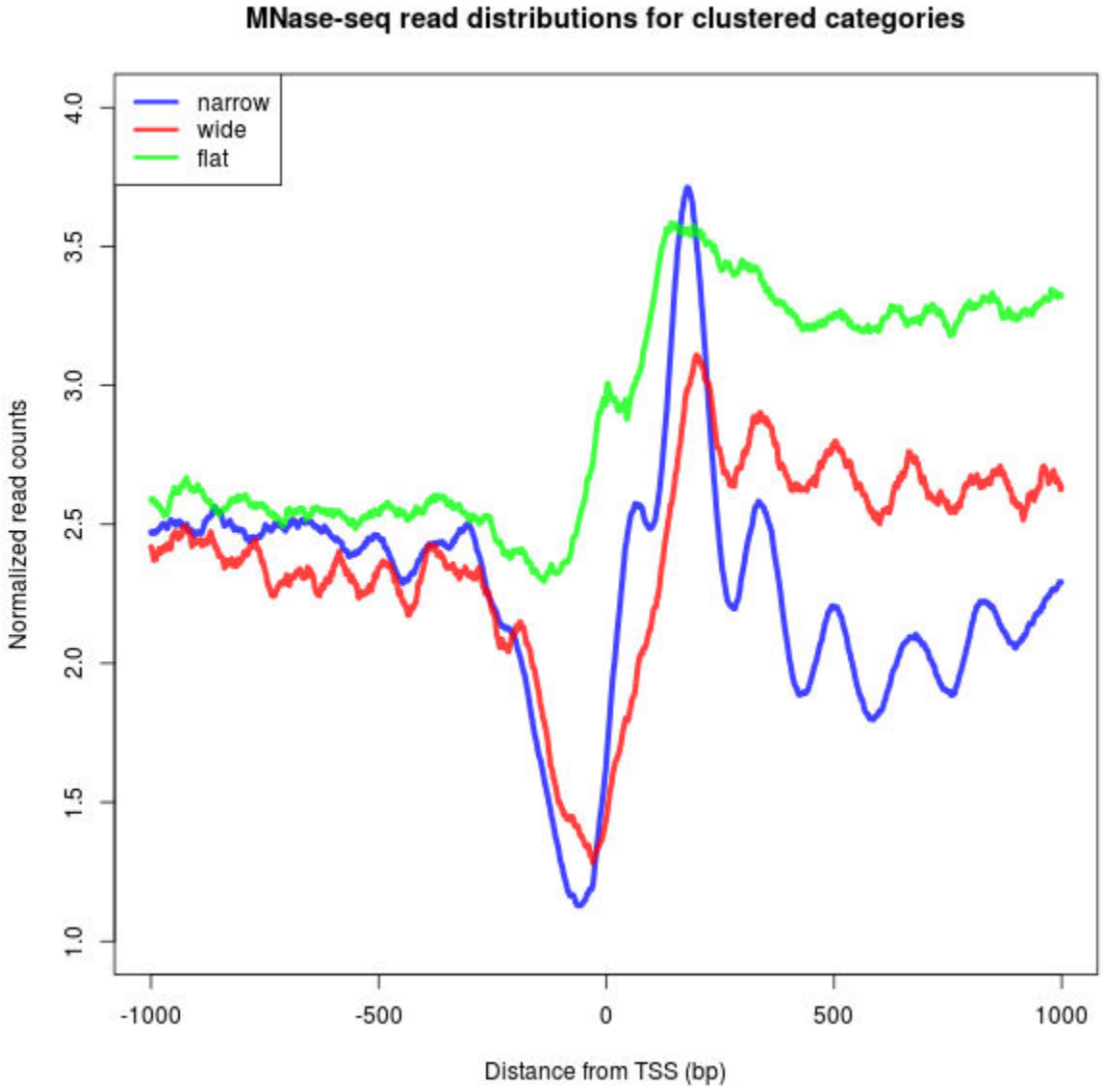
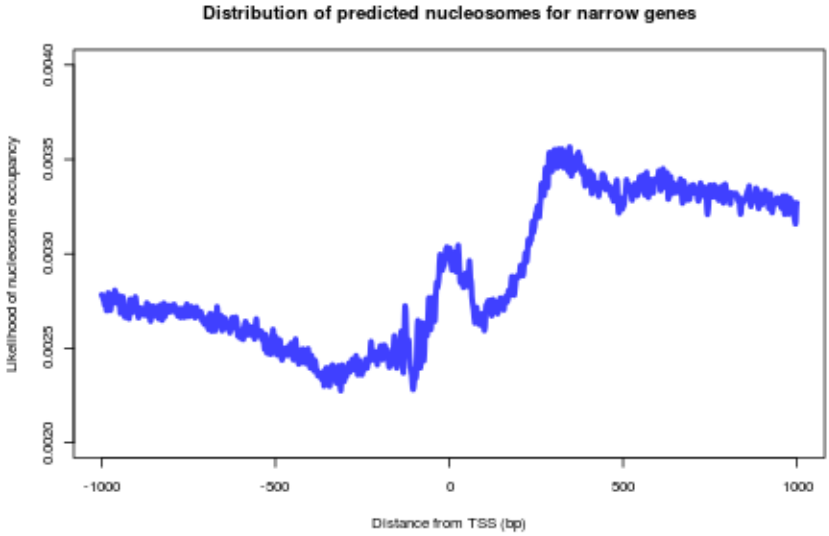
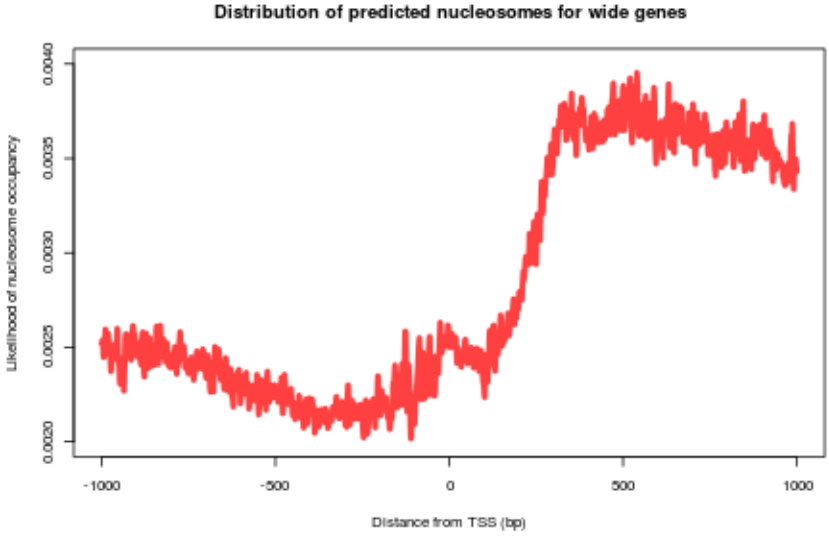

MNase-seq read distributions for clustered categories
narrow
wide
flat
Normalized read counts
4.0
3.5
3.0
2.5
2.0
1.5
1.0
-1000
-500
0
500
1000
Distance from TSS (bp)

Figure 3. Nucleosome prediction models support observed nucleosomes organization. Metaplot of predicted nucleosome occupancy for each class of genes. The "narrow" and "flat" groups both feature a peak in occupancy immedietly downstream of the TSS, largely corresponding the with the typical position of +1 nucleosomes, which is absent in the "broad" class of genes.


Distribution of predicted nucleosomes for narrow genes
Likelihood of nucleosome occupancy
0.0020
0.0025
0.0030
0.0035
0.0040
-1000
-500
0
500
1000
Distance from TSS (bp)


Distribution of predicted nucleosomes for wide genes
Likelihood of nucleosome occupancy
0.0020
0.0025
0.0030
0.0035
0.0040
-1000
-500
0
500
1000
Distance from TSS (bp)

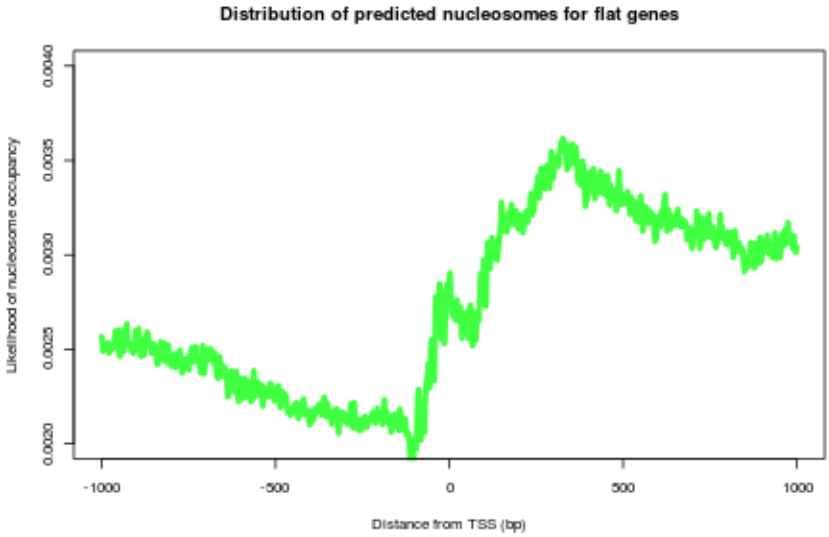
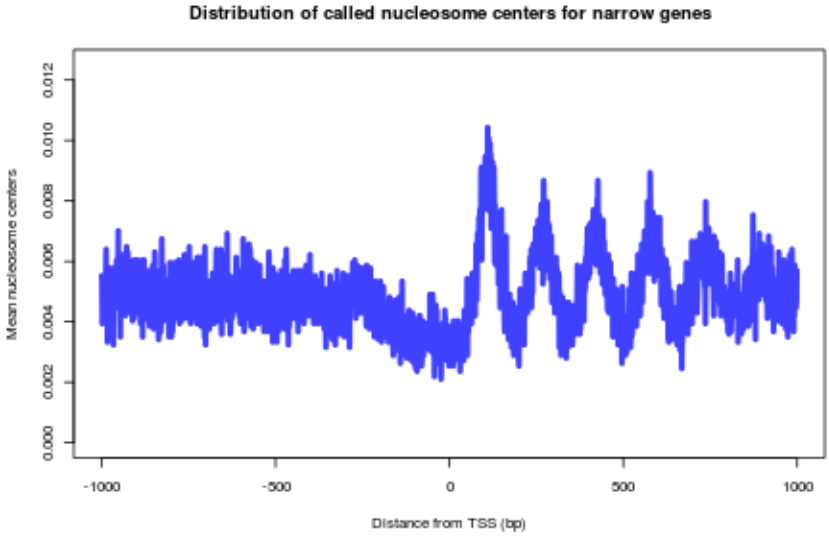
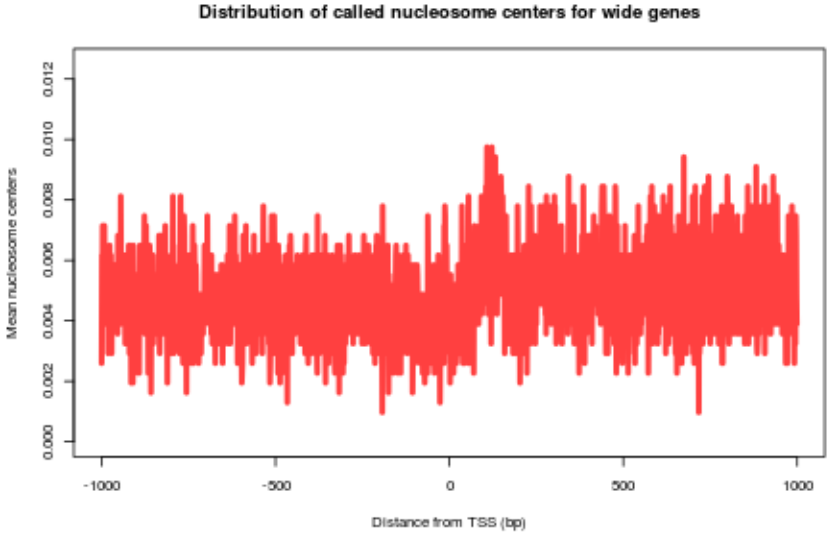
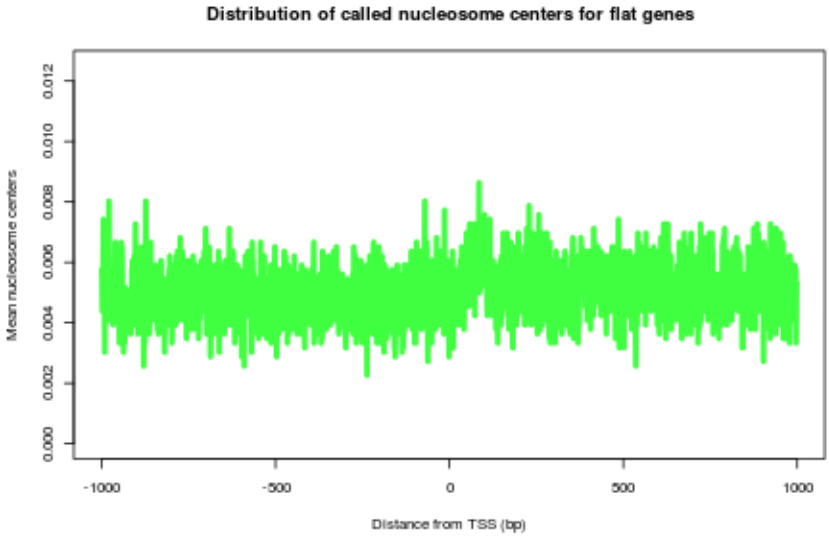

Distribution of predicted nucleosomes for flat genes
Likelihood of nucleosome occupancy
0.0020
0.0025
0.0030
0.0035
0.0040
-1000
-500
0
500
1000
Distance from TSS (bp)

Figure 4. Called nucleosomes match MNase-seq read distributions. Metaplots (individual classes first three panels, combined in last panel) showing distribution of called nucleosome centers for each class of genes. Results for the called nucleosomes recapitulate distributions seen with MNase-seq reads.


Distribution of called nucleosome centers for narrow genes
Mean nucleosome centers
0.000
0.002
0.004
0.006
0.008
0.010
0.012
-1000
-500
0
500
1000
Distance from TSS (bp)


Distribution of called nucleosome centers for wide genes
Mean nucleosome centers
0.000
0.002
0.004
0.006
0.008
0.010
0.012
-1000
-500
0
500
1000
Distance from TSS (bp)


Distribution of called nucleosome centers for flat genes
Mean nucleosome centers
0.000
0.002
0.004
0.006
0.008
0.010
0.012
-1000
-500
0
500
1000
Distance from TSS (bp)

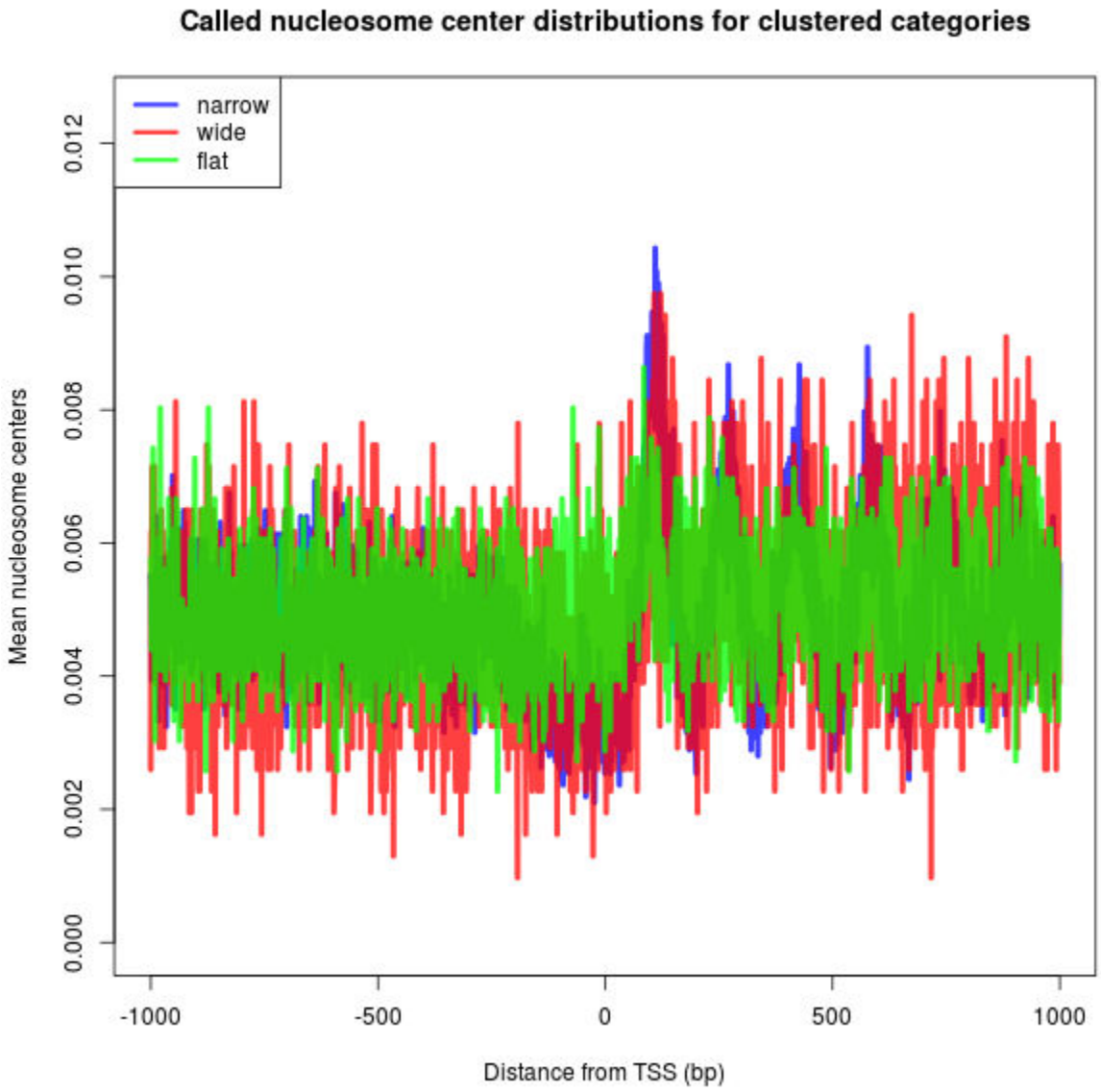
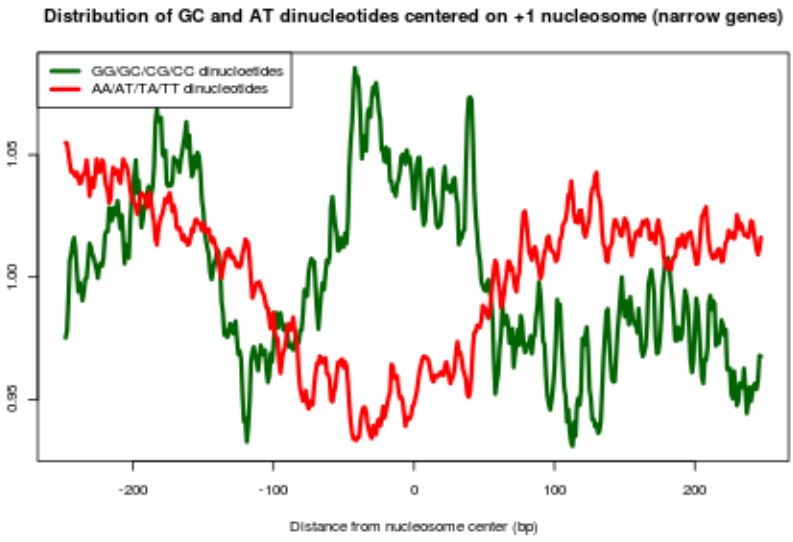
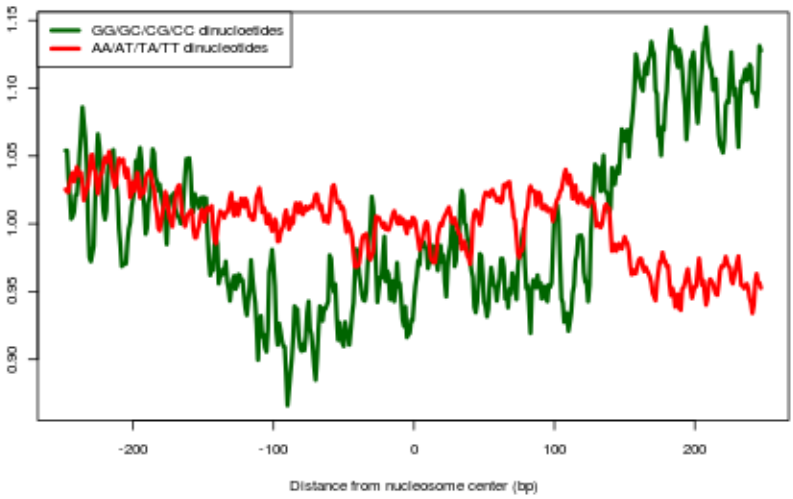
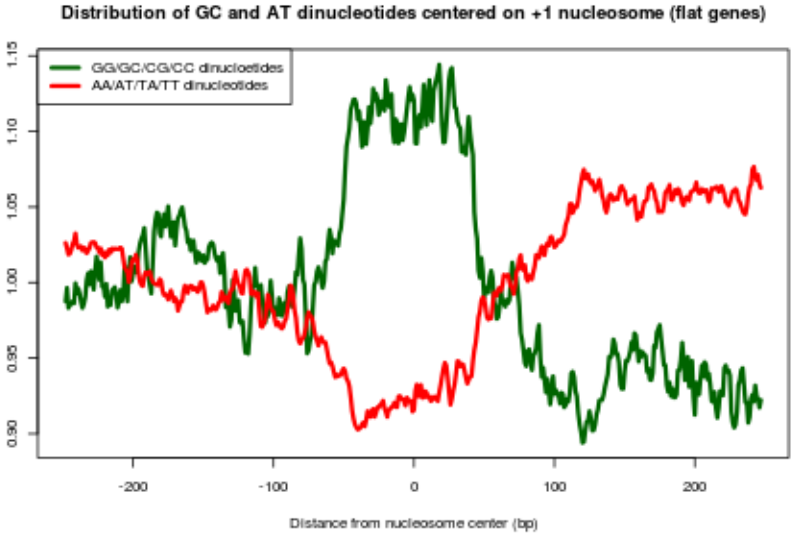

Called nucleosome center distributions for clustered categories
narrow
wide
flat
Mean nucleosome centers
0.000
0.002
0.004
0.006
0.008
0.010
0.012
-1000
-500
0
500
1000
Distance from TSS (bp)

Figure 5. +1 nucleosomes in broad group are surrounded by distinct dinucleotide frequencies. Metaplot of frequencies of GG/GC/CG/CC (green lines) and AA/AT/TA/TT (red lines) dinucleotides in a 500 bp window centered on the +1 nucleosome center for genes in each group. Genes in "narrow" and "flat" groups both feature an increase in GC rich dinucleotides and a decrease in AT rich dinucleotides localized around the nucleosome centers.


Distribution of GC and AT dinucleotides centered on +1 nucleosome (narrow genes)
GG/GC/CG/CC dinucleotides
AA/AT/TA/TT dinucleotides
0.95
1.00
1.05
-200
-100
0
100
200
Distance from nucleosome center (bp)


Distribution of GC and AT dinucleotides centered on +1 nucleosome (broad genes)
GG/GC/CG/CC dinucleotides
AA/AT/TA/TT dinucleotides
0.90
0.95
1.00
1.05
1.10
1.15
-200
-100
0
100
200
Distance from nucleosome center (bp)


Distribution of GC and AT dinucleotides centered on +1 nucleosome (flat genes)
GG/GC/CG/CC dinucleotides
AA/AT/TA/TT dinucleotides
0.90
0.95
1.00
1.05
1.10
1.15
-200
-100
0
100
200
Distance from nucleosome center (bp)

CHAPTER 3

CHARACTERIZING THE ROLE OF HISTONE H1 IN *NEUROSPORA CRASSA*

With these experiments we sought to understand the role of Histone H1 in regulating chromatin structure in the model system *Neurospora crassa*. This model system was ideal for answering these questions due to its small genome size, and its single copy of histone H1, making the creation of an H1 knockout strain very simple.

Materials and methods

Tissue for H1 ChIP-seq was harvested from 5-hour-old germinating conidia. Tissue for MNase-seq was also harvested from 5-hour-old germinating conidia and digested with 20 units of Micrococcal nuclease for 60 minutes. The digested product was size separated by gel electrophoresis and ~146 basepair bands were isolated for sequencing. All reads were mapped using bowtie2 with default parameters(Langmead & Salzberg, 2012).

Tissue for RNA-seq was collected from 5-hour-old germinating conidia and 100 basepair paired-end reads were sequenced on the Illumina platform. RNA-seq reads were mapped using tophat2 with default parameters, and differential expression analysis was performed using CuffDiff(Trapnell et al., 2012, 2013). Gene models (Version 4) created by the Broad institute were used as parameters for CuffDiff(Galagan et al., 2003).

Functional annotation was performed using the FungiFun online application(Priebe, Linde, Albrecht, Guthke, & Brakhage, 2011).

Predicted nucleosome occupancy was computed using version 3.0 of the models described in Kaplan et al, 2008.

All metaplots were generated at single nucleotide resolution and read counts were divided by # of million mapped reads in order to normalize counts to library size.

Histone H1 localization

To determine the localization of Histone H1 we performed ChIP-seq on 3XFLAG-tagged H1. Reads were mapped to the NC12.V7 genome using bowtie2 with default parameters. We found that H1 is depleted in coding regions of genes but enriched in promoter regions (Figure 6, top panel). Interestingly this enrichment is positively correlated with the transcriptional level of the adjacent genes; the most highly transcribed genes showed the greatest enrichment while the most lowly transcribed genes showed the least (Figure 2). We also found that H1 is enriched at the borders of heterochromatin domains, but not in or surrounding regions covered by H3k27me3 (Figure 6, middle and bottom panels).

Differential expression after loss of H1

We determined which genes are dependent on Histone H1 for proper transcriptional regulation by performing RNA-seq on WT and *ΔhH1* strains. Reads were mapped to the NC12.V7 genome using tophat with default parameters and differentially expressed genes were called using cuffdiff. We found that 439 genes are statistically significantly differentially expressed, of which 202 had an increased expression in the *ΔhH1* mutant and 237 had a decreased expression (Figure 8). We also found that differentially expressed genes were enriched for GO terms

related to ribosome synthesis (Figure 9). GO term enrichment analysis was performed using FungiFun2, however this enrichment only pertained to a minority of differentially expressed genes (< 20).

Modest changes in nucleosome organization in the absence of H1

We performed MNase-seq on WT and *ΔhH1* strains to determine if nucleosome organization is dependent on histone H1, and whether changes in this organization could be responsible for the differential expression of some genes in the *ΔhH1* mutant. Chromatin was isolated from X day old colonies and digested with Micrococcal nuclease at a concentration of X for X days. Mononucleosomal DNA was selected by cutting out ~146bp band after size separation using gel electrophoresis.

The MNase-seq experiment was performed using paired end sequencing, which allowed us approximate the length of DNA bound by nucleosomes by using the insert size of the reads as a proxy. We found no significant difference between the insert sizes of MNase reads from the WT and *ΔhH1* strains leading us to believe that there are no significant differences in the length of nucleosome-associated DNA in the WT and ΔhH1 strain (Figure 10).

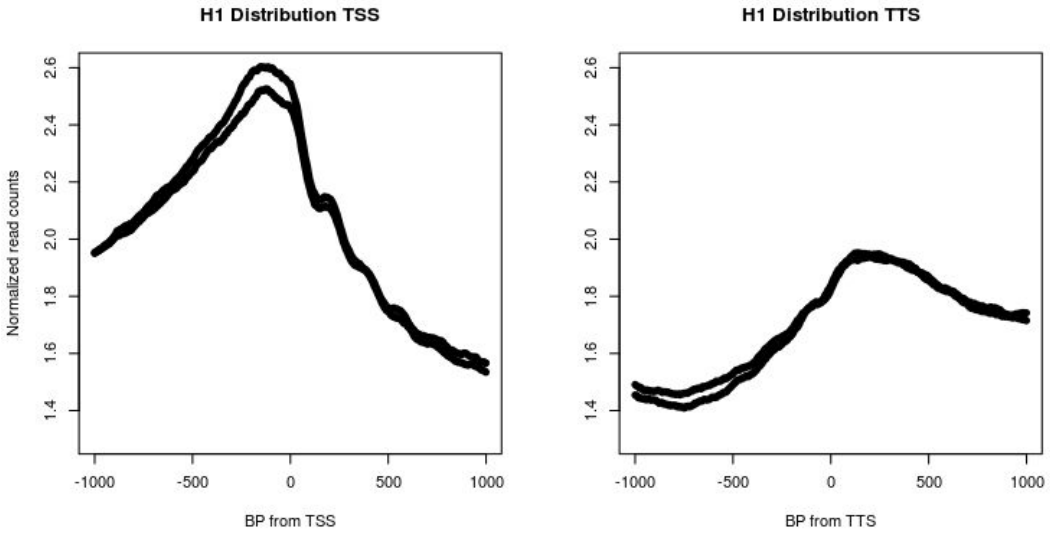
After filtering all reads with an insert size >= 90 and <= 180 we compared the nucleosome distribution surrounding the TSS of all genes, separated into quartiles by expression in WT. We found only modest changes in nucleosomes organization between WT and *ΔhH1* strains, the most significant of which were found in the middle two quartiles (Figure 11). These quartiles had fewer reads corresponding to the +1 nucleosome suggesting that the first nucleosome may be dependent on Histone H1.

In addition to the loss of reads from the +1 nucleosome we also noticed a loss of reads from the upstream edge of the nucleosome free region. This effect is apparent in all four quartiles but is more pronounced in the quartile of genes with the lowest expression in WT.

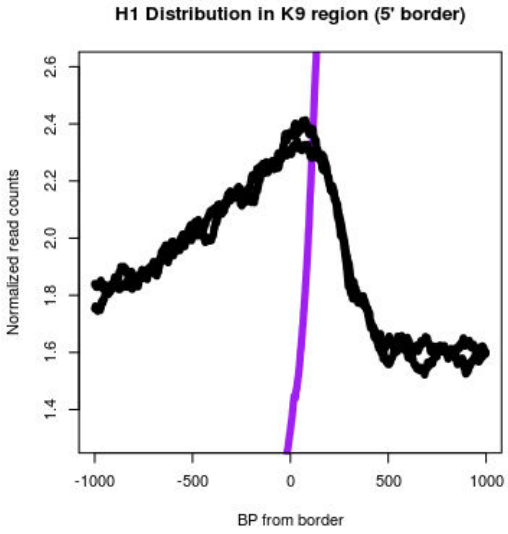
Misregulation of genes in ΔhH1 mutant not attributable to changes in nucleosome organization

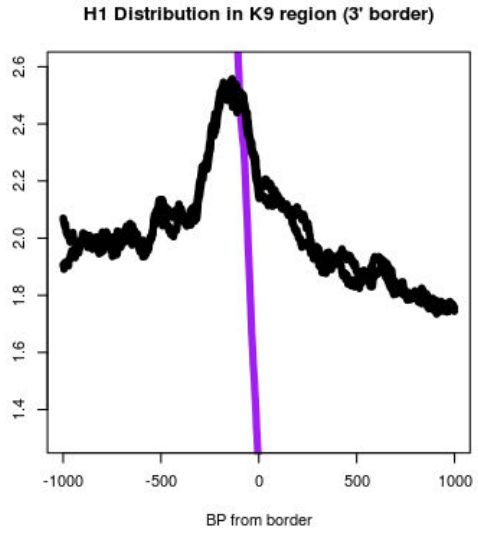
We used the models for nucleosome prediction developed by the Widom lab to determine if differentially expressed genes and their promoters had sequences which could result in delicately positioned nucleosomes; these models predict nucleosome occupancy based solely on DNA sequence, observed nucleosomes which cannot be predicted, and therefore stabilized, by DNA sequence alone may be more dependent on H1 for stability. Unfortunately our results showed that genes that are misregulated in the *ΔhH1* strain show almost no changes in nucleosome organization (Figure 12, top panels). The nucleosome predictions showed that genes which were down regulated had an unusually well predicted -2 nucleosome; however this was not seen in the MNase results in WT or *ΔhH1*. A similar peak of predicted occupancy was found in the set of genes which were not differentially expressed in ΔhH1 suggesting that this may be a common feature in Neurospora genes. This peak does roughly coincide with the reduction of MNase-reads on the upstream boundary of the NFR, but it is unsure if these two observations are related. Interestingly this predicted -2 nucleosome peak is absent in the set of genes up-regulated in *ΔhH1*.

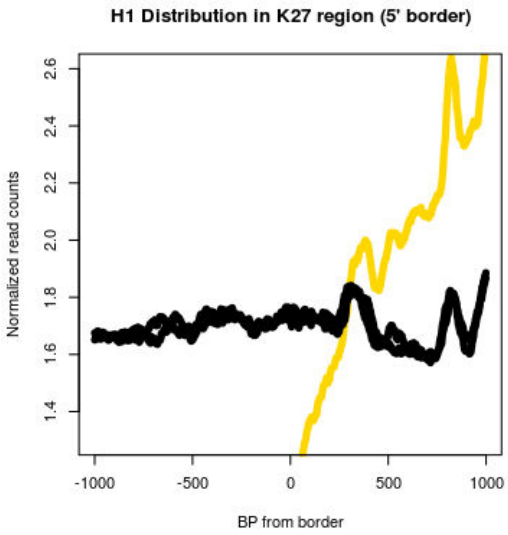
Figure 6: Histone H1 is enriched at borders of genic and heterochromatic regions, but not at H3K27me3 regions. Metaplots showing the distribution of H1 ChIP-seq reads surrounding the TSS (top left) and TSS (top right) of genes. Metaplots showing the distribution of H1 ChIP-seq reads (black line) surrounding 5' most and 3' most borders of regions bound by H3K9me3 (purple line) (middle left, middle right), and H3K27me3 (yellow line) (bottom left, bottom right). Each line represents one replicate from the ChIP-seq experiment.


H1 Distribution TSS
Normalized read counts
1.4
1.6
1.8
2.0
2.2
2.4
2.6
-1000
-500
0
500
1000
BP from TSS

H1 Distribution TTS
1.4
1.6
1.8
2.0
2.2
2.4
2.6
-1000
-500
0
500
1000
BP from TTS


H1 Distribution in K9 region (5' border)
Normalized read counts
1.4
1.6
1.8
2.0
2.2
2.4
2.6
-1000
-500
0
500
1000
BP from border


H1 Distribution in K9 region (3' border)
1.4
1.6
1.8
2.0
2.2
2.4
2.6
-1000
-500
0
500
1000
BP from border


H1 Distribution in K27 region (5' border)
Normalized read counts
1.4
1.6
1.8
2.0
2.2
2.4
2.6
-1000
-500
0
500
1000
BP from border

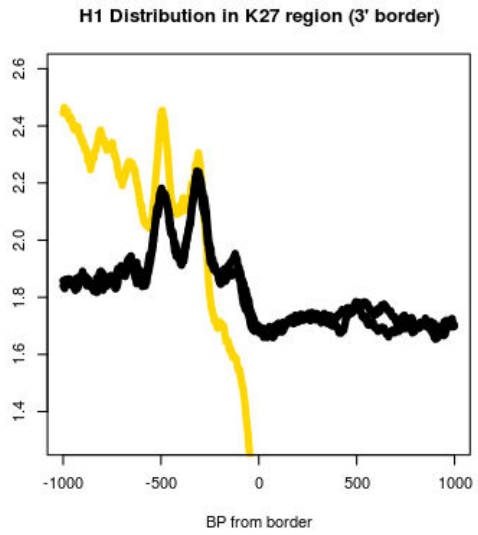

H1 Distribution in K27 region (3' border)
1.4
1.6
1.8
2.0
2.2
2.4
2.6
-1000
-500
0
500
1000
BP from border

Figure 7: Histone H1 enrichment in promoter regions is positively correlated with transcriptional level of the downstream gene. Metaplots showing the distribution of H1 ChIP-seq reads surrounding the TSS (left panel) and TTS (right panel) of four quartiles of genes separated by expression in WT. Warmer colors represent more highly transcribed genes and cooler colors represent least expressed genes. Each line represents one replicate from the ChIP-seq experiment.

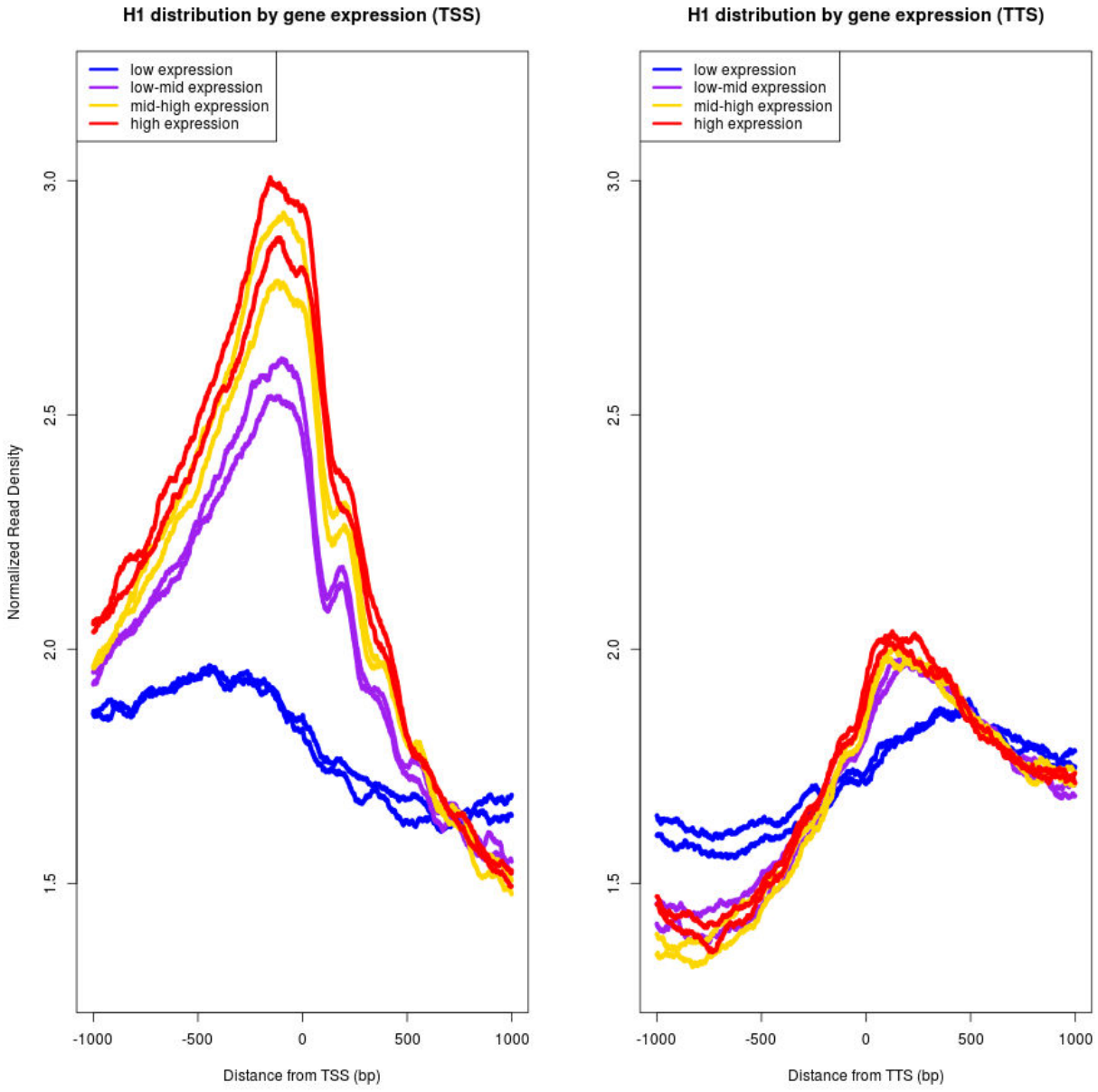

H1 distribution by gene expression (TSS)
low expression
low-mid expression
mid-high expression
high expression
Normalized Read Density
3.0
2.5
2.0
1.5
-1000
-500
0
500
1000
Distance from TSS (bp)
H1 distribution by gene expression (TTS)
low expression
low-mid expression
mid-high expression
high expression
3.0
2.5
2.0
1.5
-1000
-500
0
500
1000
Distance from TTS (bp)

Figure 8: Genes misregulated in H1 mutant not particularly highly or lowly expressed in WT. Histogram showing frequency of normalized expression values (log2(FPKM)) for genes not differentially expressed in *ΔhH1* (light gray), genes significantly down-regulated (blue), and genes significantly up-regulated (red).

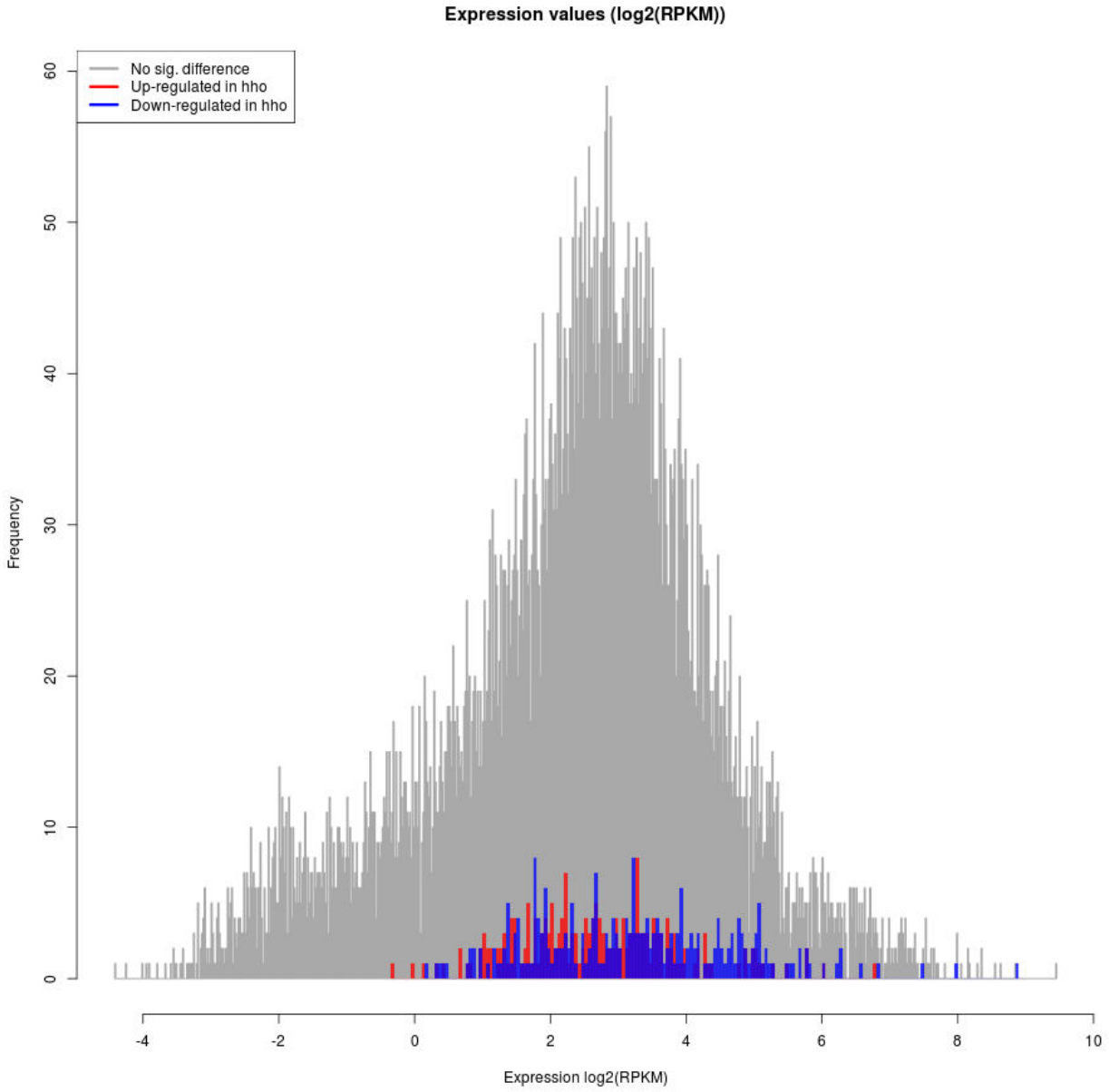


Figure 9: GO term enrichment for differentially expressed genes show enrichment in certain categories. Pie chart of the five most significantly enriched GO terms in the set of differentially expressed genes; most of the categories can be relate to ribosome synthesis.

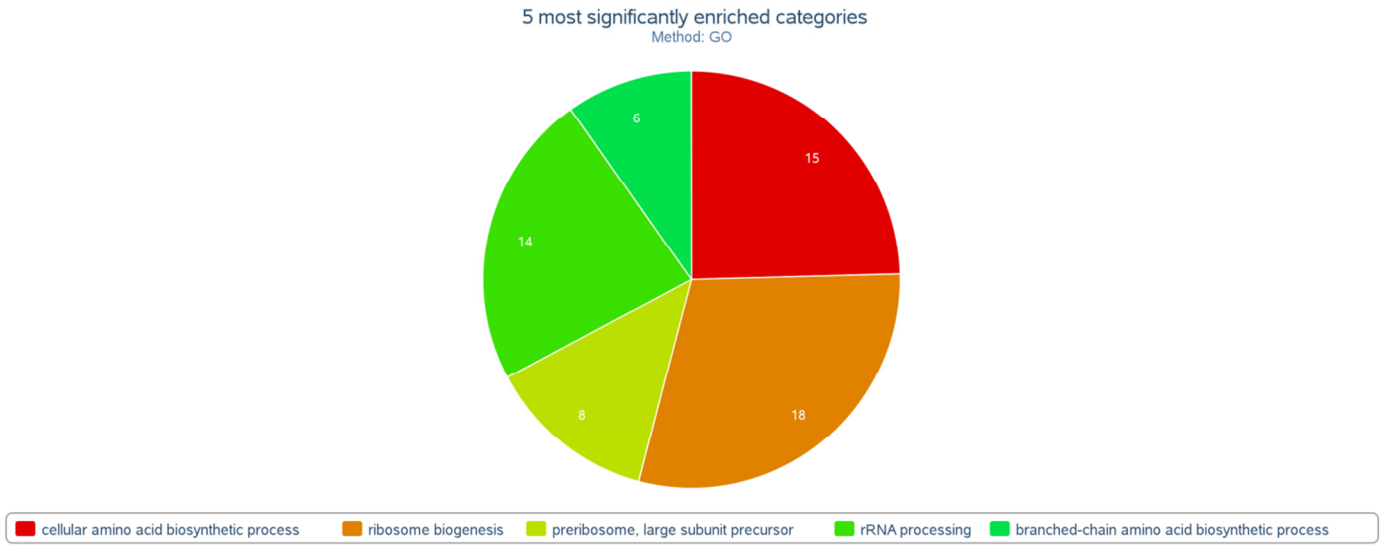


Figure 10: Paired-end read insert sizes are not statistically different in H1 mutant and WT. Boxplots showing distribution of insert sizes for MNase-seq experiments with light digestion (top panel) and heavy digestion (bottom panel). No significant differences between insert sizes between any experiments with the same MNase digest time.

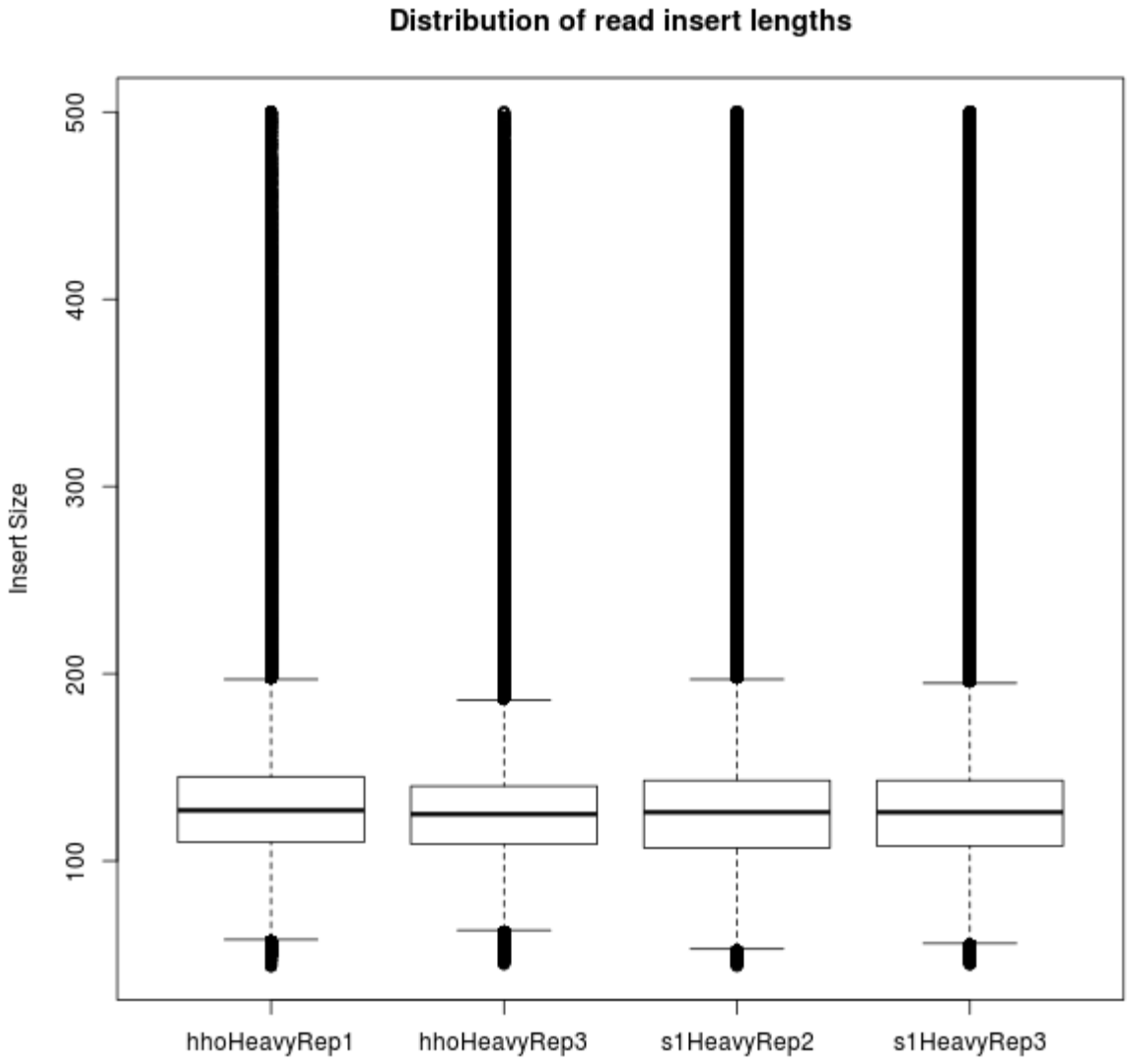


Figure 11: Minor changes in nucleosome organization observed in all genes in H1 mutant. Genes were separated into quartiles based on their expression in WT. Metaplots of MNase-seq read distribution surrounding the TSS of genes in WT (black lines) or *ΔhH1* (coloured lines). Each line represents one replicate from the MNase-seq experiment.

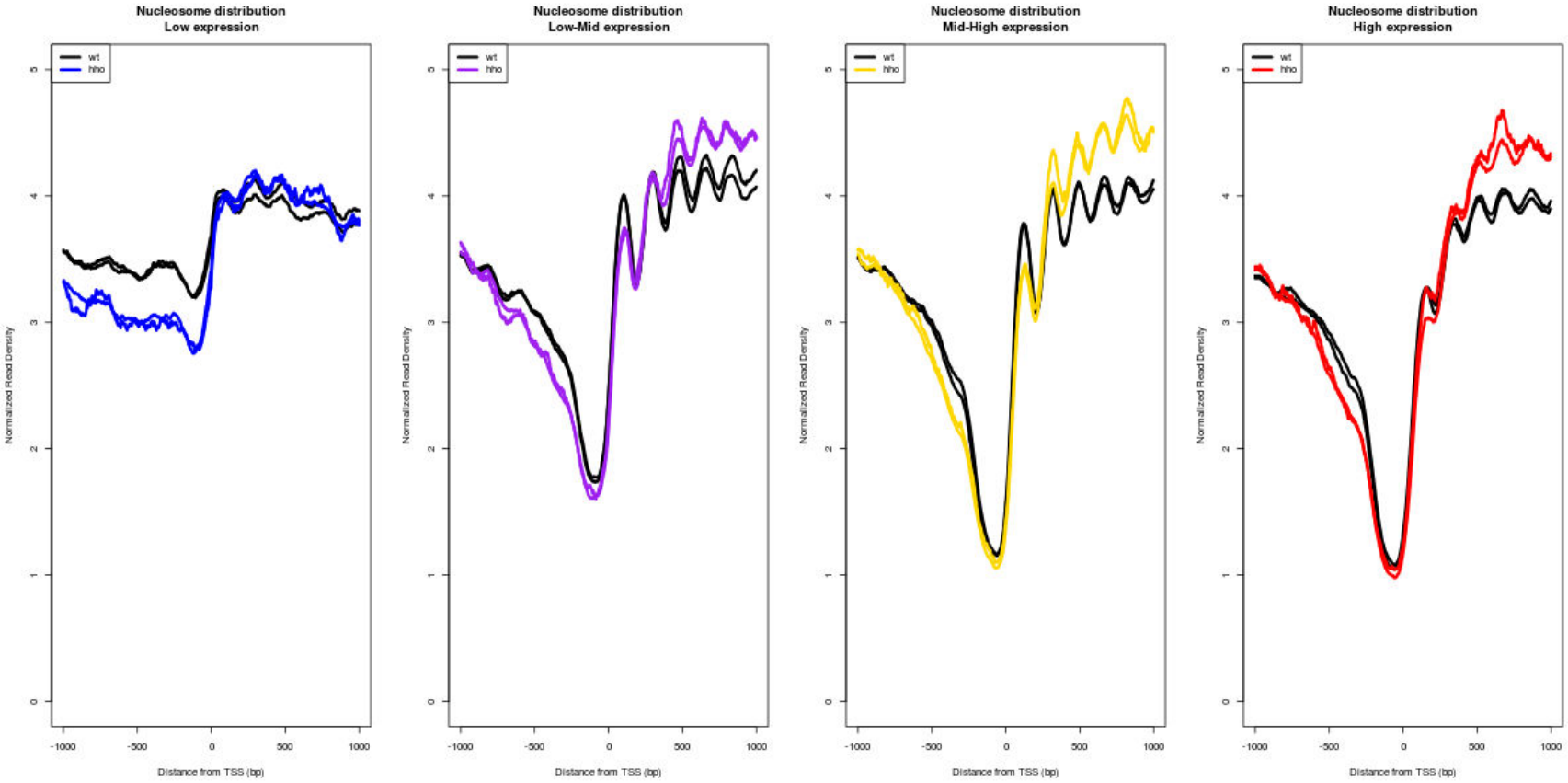
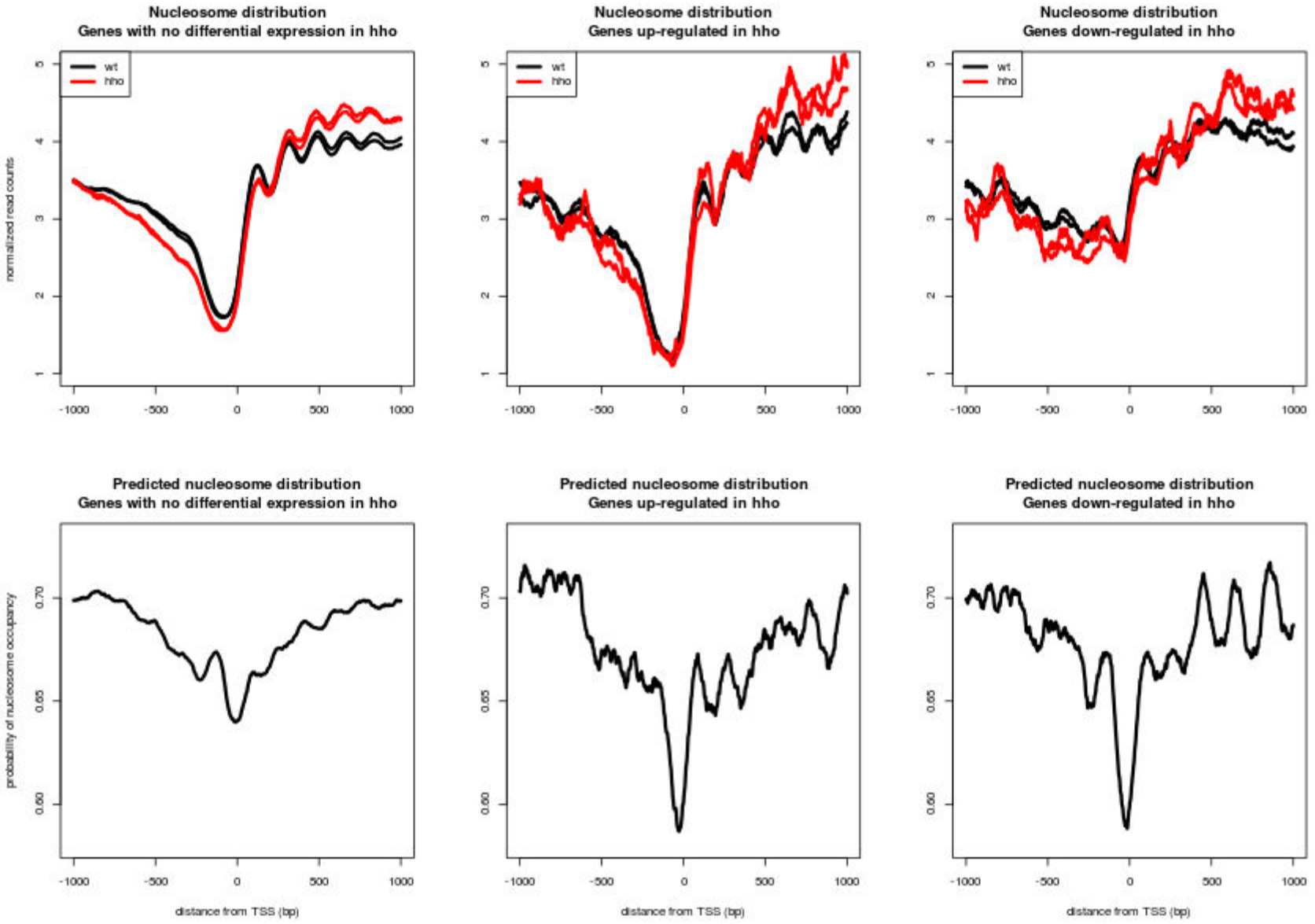

Nucleosome distribution
Low expression
wt
hho
Normalized Read Density
Distance from TSS (bp)
Nucleosome distribution
Low-Mid expression
wt
hho
Normalized Read Density
Distance from TSS (bp)
Nucleosome distribution
Mid-High expression
wt
hho
Normalized Read Density
Distance from TSS (bp)
Nucleosome distribution
High expression
wt
hho
Normalized Read Density
Distance from TSS (bp)

Figure 12: No major differences in nucleosome distribution in differentially expressed genes. Genes were grouped by their differential expression in ΔhH1. Metaplots of MNase-seq read distribution surrounding the TSS of genes in WT (black lines) or *ΔhH1* (red lines). Each line represents one replicate from the MNase-seq experiment.


Nucleosome distribution
Genes with no differential expression in hho
Nucleosome distribution
Genes up-regulated in hho
Nucleosome distribution
Genes down-regulated in hho
wt
hho
normalized read counts
Predicted nucleosome distribution
Genes with no differential expression in hho
Predicted nucleosome distribution
Genes up-regulated in hho
Predicted nucleosome distribution
Genes down-regulated in hho
probability of nucleosome occupancy
distance from TSS (bp)

CHAPTER 4

DISCUSSION

Arabidopsis thaliana experiments

These experiments began with the observation that patterns of H3K4me3 could have variable patterns of distribution in actively transcribed genes and that these were most likely mediated by nucleosome organization. The use of k-means clustering on an epigenetic feature has shown that the distributions we previously thought of as canonical may hide more complex patterns. However, the decision to use three clusters was arbitrary and further development of this technique should ensure that there are measures to determine what the most informative number of clusters should be for a given data set.

We believe that nucleosome distributions are responsible for the observed histone modification distributions and we sought to understand why some actively transcribed genes could have such disorganized nucleosomes.

We examined how the effects of histone-DNA interactions in organizing nucleosomes differed between the clusters and found that these interactions were not equal in all genes. There is some support for the functional differences we found in our categories, but further work should be done to determine the extent of this correlation. Further work is especially needed to elucidate the mechanism behind this difference, what the relationships is between inducibility and disorganized nucleosomes. Alternatively further studies could elucidate the relationship between genes with organized nucleosomes and constitutive expression.

Some labs have modeled gene regulatory networks responsible for conferring cell identity as dynamical systems. They concluded that cells transitioning from a stable cell state, such as those found in the shoot apical meristem, to their final cell states would require a degree of instability in the transcriptional regulation of some key attractors; Attractors in this context refers to genes acting as a sort of keystone in initiating and stabilizing a self-reinforcing regulatory network. One potential mechanism behind this instability may be disorganized nucleosomes. Several groups have shown that organized nucleosomes can result in lower transcriptional consistency, contributing to the aforementioned instability(Durrin et al., 1992; Han & Grunstein, 1988; Raveh-Sadka et al., 2012). In the future we would like to identify genes that are acting as attractors to determine if they generally have disorganized nucleosomes which could lend support to this theory.

Neurospora crassa experiments

Our results showed that histone H1 does not seem to contribute substantially to nucleosome positioning in the TSS. The H1 deletion strain showed some minor differences in nucleosome organization but these were relatively mild. Appropriately, defects in transcriptional regulation were similarly mild with only ~450 genes differentially expressed in the mutant. Given these results it may be that histone H1 only becomes relevant in certain developmental contexts. One of the more obvious defects is the mild hypermethylation of genomic DNA in the H1 deletion strain. This could potentially explain the differential expression described earlier, but how this could be linked to Histone H1 is still unknown.

As previous results have shown, not all epigenetics factors are equally important in all systems nor do they use the same mechanisms. It may not be appropriate to generalize these results to any other systems. These experiments should be replicated in several other systems, some that preferably have a much more complex development, in order to analyze the potential of histone H1 as an epigenetic factor.

REFERENCES

Akkers, R. C., van Heeringen, S. J., Jacobi, U. G., Janssen-Megens, E. M., Françoijs, K.-J., Stunnenberg, H. G., & Veenstra, G. J. C. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. *Developmental Cell*, *17*(3), 425–34. doi:10.1016/j.devcel.2009.08.005

Allan, J., Hartman, P. G., Crane-Robinson, C., & Aviles, F. X. (1980). The structure of histone H1 and its location in chromatin. *Nature*, *288*(5792), 675–679. doi:10.1038/288675a0

Alvarez-Buylla, E. R., Chaos, A., Aldana, M., Benítez, M., Cortes-Poza, Y., Espinosa-Soto, C., … Padilla-Longoria, P. (2008). Floral morphogenesis: stochastic explorations of a gene network epigenetic landscape. *PloS One*, *3*(11), e3626. doi:10.1371/journal.pone.0003626

Anderson, J. D., & Widom, J. (2000). Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *Journal of Molecular Biology*, *296*(4), 979–87. doi:10.1006/jmbi.2000.3531

Banerji, C. R. S., Miranda-Saavedra, D., Severini, S., Widschwendter, M., Enver, T., Zhou, J. X., & Teschendorff, A. E. (2013). Cellular network entropy as the energy potential in Waddington's differentiation landscape. *Scientific Reports*, *3*, 3039. doi:10.1038/srep03039

Bao, Y., & Shen, X. (2007). INO80 subfamily of chromatin remodeling complexes. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *618*(1–2), 18–29. doi:http://dx.doi.org/10.1016/j.mrfmmm.2006.10.006

Beh, L. Y., Kaplan, N., Muller, M. M., Muir, T. W., & Landweber, L. F. (2014). *DNA-guided establishment of canonical nucleosome patterns in a eukaryotic genome*. *bioRxiv*. Cold Spring Harbor Labs Journals. doi:10.1101/013250

Benayoun, B. A., Pollina, E. A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E. D., … Brunet, A. (2014). H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency. *Cell*, *158*(3), 673–688. doi:10.1016/j.cell.2014.06.027

Berger, S. L. (2007). The complex language of chromatin regulation during transcription. *Nature*, *447*(7143), 407–12. doi:10.1038/nature05915

Bernstein, B. E., Liu, C. L., Humphrey, E. L., Perlstein, E. O., & Schreiber, S. L. (2004). Global nucleosome occupancy in yeast. *Genome Biology*, *5*(9), R62. doi:10.1186/gb-2004-5-9-r62

Bhaumik, S. R., Smith, E., & Shilatifard, A. (2007). Covalent modifications of histones during development and disease pathogenesis. *Nature Structural & Molecular Biology*, *14*(11), 1008–16. doi:10.1038/nsmb1337

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, *16*(1), 6–21. doi:10.1101/gad.947102

Blashfield, R. K. (n.d.). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods.

Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M., & Sherlock, G. (2004). GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics (Oxford, England)*, *20*(18), 3710–5. doi:10.1093/bioinformatics/bth456

Cao, X., & Jacobsen, S. E. (2002). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Current Biology : CB*, *12*(13), 1138–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12121623

Cavalli, G., & Thoma, F. (1993). Chromatin transitions during activation and repression of galactose-regulated genes in yeast. *The EMBO Journal*, *12*(12), 4603–13. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=413896&tool=pmcentrez&rendertype=abstract

Chen, K., Xi, Y., Pan, X., Li, Z., Kaestner, K., Tyler, J., … Li, W. (2013). DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Research*, *23*(2), 341–51. doi:10.1101/gr.142067.112

Clapier, C. R., & Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annual Review of Biochemistry*, *78*, 273–304. doi:10.1146/annurev.biochem.77.062706.153223

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., … Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, *452*(7184), 215–9. doi:10.1038/nature06745

Corona, D. F., Längst, G., Clapier, C. R., Bonte, E. J., Ferrari, S., Tamkun, J. W., & Becker, P. B. (1999). ISWI is an ATP-dependent nucleosome remodeling factor. *Molecular Cell*, *3*(2), 239–45. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10078206

Durrin, L. K., Mann, R. K., & Grunstein, M. (1992). Nucleosome loss activates CUP1 and HIS3 promoters to fully induced levels in the yeast Saccharomyces cerevisiae. *Mol. Cell. Biol.*, *12*(4), 1621–1629. doi:10.1128/MCB.12.4.1621

Eichten, S. R., Swanson-Wagner, R. a, Schnable, J. C., Waters, A. J., Hermanson, P. J., Liu, S., … Springer, N. M. (2011). Heritable epigenetic variation among maize inbreds. *PLoS Genetics*, *7*(11), e1002372. doi:10.1371/journal.pgen.1002372

Eichten, S. R., Vaughn, M. W., Hermanson, P. J., & Springer, N. M. (2013). Variation in DNA Methylation Patterns is More Common among Maize Inbreds than among Tissues. *The Plant Genome*, *6*(2). doi:10.3835/plantgenome2012.06.0009

Eldar, A., & Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature*, *467*(7312), 167–73. doi:10.1038/nature09326

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science (New York, N.Y.)*, *297*(5584), 1183–6. doi:10.1126/science.1070919

Epigenetic Mechanisms of Gene Regulation (1996, Volume 32). (n.d.). Retrieved May 19, 2015, from http://cshmonographs.org/index.php/monographs/issue/view/087969490.32

Feng, S., Cokus, S. J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M. G., … Jacobsen, S. E. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8689–94. doi:10.1073/pnas.1002720107

Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., … Birren, B. (2003). The genome sequence of the filamentous fungus Neurospora crassa. *Nature*, *422*(6934), 859–68. doi:10.1038/nature01554

Garg, A., Mohanram, K., Di Cara, A., De Micheli, G., & Xenarios, I. (2009). Modeling stochasticity and robustness in gene regulatory networks. *Bioinformatics (Oxford, England)*, *25*(12), i101–9. doi:10.1093/bioinformatics/btp214

Han, M., & Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters in vivo. *Cell*, *55*(6), 1137–1145. doi:10.1016/0092-8674(88)90258-9

He, Y., & Amasino, R. M. (2005). Role of chromatin modification in flowering-time control. *Trends in Plant Science*, *10*(1), 30–5. doi:10.1016/j.tplants.2004.11.003

Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science (New York, N.Y.)*, *293*(5532), 1074–80. doi:10.1126/science.1063127

Johannes, F., Porcher, E., Teixeira, F. K., Saliba-Colombani, V., Simon, M., Agier, N., … Colot, V. (2009). Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genetics*, *5*(6), e1000530. doi:10.1371/journal.pgen.1000530

Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., … Segal, E. (2010). Nucleosome sequence preferences influence in vivo nucleosome organization. *Nature Structural & Molecular Biology*, *17*(8), 918–20; author reply 920–2. doi:10.1038/nsmb0810-918

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., … Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, *458*(7236), 362–6. doi:10.1038/nature07667

Karlić, R., Chung, H.-R., Lasserre, J., Vlahovicek, K., & Vingron, M. (2010). Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(7), 2926–31. doi:10.1073/pnas.0909344107

Kinney, S. M., Chin, H. G., Vaisvila, R., Bitinaite, J., Zheng, Y., Estève, P.-O., … Pradhan, S. (2011). Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. *The Journal of Biological Chemistry*, *286*(28), 24685–93. doi:10.1074/jbc.M110.217083

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–9. doi:10.1038/nmeth.1923

Lee, C.-K., Shibata, Y., Rao, B., Strahl, B. D., & Lieb, J. D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genetics*, *36*(8), 900–5. doi:10.1038/ng1400

Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R., & Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genetics*, *39*(10), 1235–44. doi:10.1038/ng2117

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews. Genetics*, *3*(9), 662–73. doi:10.1038/nrg887

Lin, W., & Dent, S. Y. R. (2006). Functions of histone-modifying enzymes in development. *Current Opinion in Genetics & Development*, *16*(2), 137–42. doi:10.1016/j.gde.2006.02.002

Marfella, C. G. A., & Imbalzano, A. N. (2007). The Chd family of chromatin remodelers. *Mutation Research*, *618*(1-2), 30–40. doi:10.1016/j.mrfmmm.2006.07.012

McAdams, H. H., & Arkin, a. (1997). Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(3), 814–9. Retrieved

from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=19596&tool=pmcentrez&rendertype=abstract

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., … Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766–70. doi:10.1038/nature07107

Morozov, A. V, Fortney, K., Gaykalova, D. A., Studitsky, V. M., Widom, J., & Siggia, E. D. (2009). Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Research*, *37*(14), 4707–22. doi:10.1093/nar/gkp475

Noll, M., & Kornberg, R. D. (1977). Action of micrococcal nuclease on chromatin and the location of histone H1. *Journal of Molecular Biology*, *109*(3), 393–404. doi:10.1016/S0022-2836(77)80019-3

Okano, M., Bell, D. W., Haber, D. a, & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10555141

Paranjape, S. M., Kamakaka, R. T., & Kadonaga, J. T. (1994). Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Annual Review of Biochemistry*, *63*, 265–97. doi:10.1146/annurev.bi.63.070194.001405

Priebe, S., Linde, J., Albrecht, D., Guthke, R., & Brakhage, A. A. (2011). FungiFun: a web-based application for functional categorization of fungal genes and proteins. *Fungal Genetics and Biology : FG & B*, *48*(4), 353–8. doi:10.1016/j.fgb.2010.11.001

Raser, J. M., & O'Shea, E. K. (2004). Control of stochasticity in eukaryotic gene expression. *Science (New York, N.Y.)*, *304*(5678), 1811–4. doi:10.1126/science.1098641

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., … Segal, E. (2012). Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nature Genetics*, *44*(7), 743–50. doi:10.1038/ng.2305

Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., … Colot, V. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *The EMBO Journal*, *30*(10), 1928–38. doi:10.1038/emboj.2011.103

Schmitz, R. J., He, Y., Valdés-López, O., Khan, S. M., Joshi, T., Urich, M. a, … Ecker, J. R. (2013). Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Research*, *23*(10), 1663–74. doi:10.1101/gr.152538.112

Segal, E., & Widom, J. (2009a). Poly(dA:dT) tracts: major determinants of nucleosome organization. *Current Opinion in Structural Biology*, *19*(1), 65–71. doi:10.1016/j.sbi.2009.01.004

Segal, E., & Widom, J. (2009b). What controls nucleosome positions? *Trends in Genetics : TIG*, *25*(8), 335–43. doi:10.1016/j.tig.2009.06.002

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, *26*(10), 1135–45. doi:10.1038/nbt1486

Sims, R. J., & Reinberg, D. (2009). Processing the H3K36me3 signature. *Nature Genetics*, *41*(3), 270–1. doi:10.1038/ng0309-270

Struhl, K., & Segal, E. (2013). Determinants of nucleosome positioning. *Nature Structural & Molecular Biology*, *20*(3), 267–73. doi:10.1038/nsmb.2506

Suter, B. (2000). Poly(dAmiddle dotdT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Research*, *28*(21), 4083–4089. doi:10.1093/nar/28.21.4083

Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews. Genetics*, *9*(6), 465–76. doi:10.1038/nrg2341

Tamaru, H., & Selker, E. U. (2001). A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature*, *414*(6861), 277–83. doi:10.1038/35104508

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., … Zhao, Y. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, *146*(6), 1016–28. doi:10.1016/j.cell.2011.08.008

Thåström, A., Lowary, P. T., Widlund, H. R., Cao, H., Kubista, M., & Widom, J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *Journal of Molecular Biology*, *288*(2), 213–29. doi:10.1006/jmbi.1999.2686

Thoma, F. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *The Journal of Cell Biology*, *83*(2), 403–427. doi:10.1083/jcb.83.2.403

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, *31*(1), 46–53. doi:10.1038/nbt.2450

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–78. doi:10.1038/nprot.2012.016

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., … Johnson, S. M. (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome Research*, *18*(7), 1051–63. doi:10.1101/gr.076463.108

Wei, G., Wei, L., Zhu, J., Zang, C., Hu-Li, J., Yao, Z., … Zhao, K. (2009). Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. *Immunity*, *30*(1), 155–67. doi:10.1016/j.immuni.2008.12.009

Whitehouse, I., Flaus, a, Cairns, B. R., White, M. F., Workman, J. L., & Owen-Hughes, T. (1999). Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature*, *400*(6746), 784–7. doi:10.1038/23506

Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M. D., & Owen-hughes, T. (2003). Evidence for DNA Translocation by the ISWI Chromatin-Remodeling Enzyme, *23*(6), 1935–1945. doi:10.1128/MCB.23.6.1935

Widom, J. (2002). Role of DNA sequence in nucleosome stability and dynamics. *Quarterly Reviews of Biophysics*, *34*(03), 269–324. doi:10.1017/S0033583501003699

Workman, J. L. (Ed.). (2003). *Protein Complexes that Modify Chromatin* (Vol. 274). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-55747-7

Xi, L., Fondufe-Mittendorf, Y., Xia, L., Flatow, J., Widom, J., & Wang, J.-P. (2010). Predicting nucleosome positioning using a duration Hidden Markov Model. *BMC Bioinformatics*, *11*(1), 346. doi:10.1186/1471-2105-11-346

Yen, K., Vinayachandran, V., Batta, K., Koerber, R. T., & Pugh, B. F. (2012). Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell*, *149*(7), 1461–73. doi:10.1016/j.cell.2012.04.036

Yuan, G.-C., Liu, Y.-J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J., & Rando, O. J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. *Science (New York, N.Y.)*, *309*(5734), 626–30. doi:10.1126/science.1112178

Zhang, Y., Moqtaderi, Z., Rattner, B. P., Euskirchen, G., Snyder, M., Kadonaga, J. T., … Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nature Structural & Molecular Biology*, *16*(8), 847–52. doi:10.1038/nsmb.1636

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., & Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, *39*(1), 61–9. doi:10.1038/ng1929