Yu-Han Hsu
4/28/13
Biophysics 205 Project Proposal

Investigating the relationship between nucleosome remodeling and transcription factor binding during transcriptional perturbation

Introduction

Numerous studies have indicated that there is a close relationship between nucleosome remodeling and transcription in eukaryotic cells [1]. Genome-wide maps of nucleosome distribution in *Saccharomyces cerevisiae* have also shown that nucleosome positions in promoter regions often coincide with transcription factor (TF) binding sites [2, 3]. In particular, Shivaswamy *et al.* studied dynamic nucleosome remodeling in *S. cerevisiae* in response to heat shock and found that many nucleosome displacement events lead to increased accessibility of predicted TF binding sites upstream of genes induced in stress response [2]. Furthermore, nucleosome eviction and appearance in promoter regions are generally associated with gene activation and repression, respectively, after heat shock perturbation. While this study demonstrated that the relationship between nucleosome remodeling and TF binding events exists on a genome-wide level during transcriptional perturbation, it lacked direct experimental data to support this relationship and to establish the timing and ordering of these events.

Several mechanisms linking nucleosome displacement to TF binding have been proposed and studied. A common hypothesis is that, when a TF binding site on genomic DNA is occupied by a nucleosome, the TF would be unable to access the binding site; upon transcriptional perturbation, the nucleosome would be displaced by factors such as chromatin remodeling complexes and histone chaperones, thus allowing the TF to bind to this site and regulate transcription of a downstream gene. However, many studies have suggested that some TFs are able to bind to their sites wrapped up within nucleosomes, which may in turn destabilize the nucleosomes and allow other factors to remove them from the sites [1]. In fact, some studies have observed that multiple transcription factors may act together to bring about nucleosome displacement at a promoter. It would be informative to determine which, if any, of these mechanisms dominate at a genome-wide level. That is, does nucleosome remodeling usually precede TF binding? Or does TF binding lead to nucleosome remodeling?

To further investigate the relationship between nucleosome remodeling and TF binding during transcriptional perturbation, we propose to study these events throughout a heat shock time course in *Saccharomyces cerevisiae*. Cell samples will be collected at different time points and genomic DNA fragments that are occupied by either nucleosomes or heat shock response TFs will be determined at each time point. Genomic sites that show significant nucleosome or TF occupancy changes over time will be analyzed in order to find any correlation between nucleosome remodeling and TF binding. Clustering and gene enrichment analysis will be performed to characterize similar occupancy patterns. Multiple sites within a promoter region will also be examined together to discover large-scale

patterns related to transcriptional regulation. Finally, depending on the preliminary results and analysis, further experimentation will be performed, possibly using knockout yeast strains, in order to validate or test our hypotheses.

Specific Aims

Aim 1: Generate nucleosome and TF occupancy-time profiles for relevant genomic sites during heat shock response

S. cerevisiae cultures will be grown at 30°C to early log phase in rich medium, then subjected to 39°C heat shock condition. Cell samples will be collected at 0, 1, 2, 4, 8, 16, and 32 minutes after heat shock. These time points were chosen based on results of previous yeast heat shock response studies: widespread changes in gene expression could be observed after ~5 minutes of heat shock, while some promoters start to show nucleosome displacement and TF binding activity after 1 to 4 minutes, with most activity peaking between 8 to 16 minutes [4, 5]. Because of these rapid changes, it is important to collect denser time points towards the beginning of the time course in order to establish a possible chronological order for nucleosome remodeling and TF binding events that occur at the same site.

The nucleosome occupancy-time profiles will be generated using a recently developed method that utilizes genetically engineered histones to map nucleosome positions [3]. This method can produce higher quality nucleosome occupancy data compared to the more traditional micrococcal nuclease (MNase) digestion method used by Shivaswamy *et al.* [2] because it eliminates systematic biases that might be generated by MNase sequence preferences. Briefly, a yeast strain with a cysteine mutation in histone H4 (H4S47C mutation) will be used in the heat shock time course. After cell samples are collected, chemical reactions involving the mutated cysteines will be carried out to introduce highly specific cleavage of nucleosomal DNA at sites flanking the center of the nucleosome. DNA fragments of ~150-200bp will then be purified and sequenced using high-throughput parallel sequencing. Sequencing results will be analyzed as described by Brogaard *et al.* [3] to determine genomic sites bound by nucleosomes.

For generating occupancy-time profiles of TFs, the ChIP-exo (chromatin immunoprecipitation followed by exonuclease digestion) method developed by Rhee and Pugh [6] will be used to identify genomic sites bound by several heat shock response TFs. The ChIP-exo procedure is similar to ChIP-seq (ChIP with massively parallel sequencing) except that it involves an exonuclease digestion process that trims ChIP DNA to a small distance from the protein binding site, thus allowing binding sites to be identified with near single base-pair resolution. For this project, it is especially important to obtain high-resolution data for both nucleosome and TF occupancy in order to compare them directly across time points. In the first round of experiments, three yeast transcription factors involved in heat shock response, Hsf1, Msn2, and Msn4, will be pulled down using ChIP-exo. Hsf1 is a major heat shock response regulator that binds the HSE (heat shock element)

motif as a homotrimeric complex [7, 8]. Msn2/4 are partially redundant transcriptional activators that bind to the STRE (stress response element) motif in response to a variety of stress conditions [9, 10]. Previous experimental results have shown that Hsf1 and Msn2/4 each induce over 150 target genes during heat shock, making them good candidates for large-scale TF binding experiments as proposed in this project [4, 11]. Furthermore, antibodies for pulling down these TFs are readily available and have been tested in ChIP assays before [5].

To assess the reproducibility of the nucleosome and TF mapping results produced in this project, biological and technical replicates for a subset of the cell samples will be generated and evaluated. To assess the quality of the results, the nucleosome occupancy data will be compared to previously generated nucleosome maps [2, 3], while the TF binding data will be validated using reported binding motifs and target genes for each of the three TFs [4, 7-11]. Finally, depending on how sequencing results from each type of experiments are analyzed, it might be necessary to normalize the nucleosome and TF occupancy results to produce comparable occupancy scores. This will enable easier comparison of nucleosome and TF occupancy-time profiles in subsequent analysis steps.

Aim 2: Characterize the relationship between nucleosome and TF occupancy-time profiles

There are several ways to characterize the relationship between nucleosome remodeling and TF binding events. First, a simple analysis can be done to test if nucleosomes and TFs are more likely to occupy the same genomic sites. The number of overlapping sites that are occupied (i.e. with occupancy score above a certain threshold) by both a nucleosome and a TF at some point in the heat shock time series will be counted and its statistical significance will be evaluated by calculating the number of "random" overlaps that should occur given the genome-wide frequencies of nucleosome and TF occupancies. If the observed number of overlapping sites is significantly higher than one would expect by chance, we can conclude that nucleosomes and TFs tend to occupy the same genomic sites. A similar approach can be used to assess if nucleosome remodeling are associated with TF binding by counting the number of nucleosome eviction, nucleosome appearance, and TF binding events and determining how frequently these events occur at overlapping sites compared to random expectations. An alternative approach of examining the general association between nucleosome and TF occupancy at overlapping sites would be to generate a scatter plot of nucleosome occupancy scores vs. TF occupancy scores, where each data point corresponds to a particular site at a certain time point. Regression analysis can then be performed on this plot to determine if a general correlation trend exists. However, if there are multiple conflicting trends within the dataset, this kind of analysis might not be informative.

Next, changes in nucleosome and TF occupancy scores over time will be analyzed in more detail. For each genomic site of interest, a plot of occupancy scores vs. time can be generated to visualize how the amount of nucleosome and TF binding changes throughout the heat shock time course. To visualize patterns of the occupancy-time profiles at a genome-wide level, heat maps will be generated by clustering the nucleosome or TF

occupancy scores. Each row in the heat map would be a genomic site bound by a nucleosome or TF and each column would be a different time point during the heat shock time course. We can compare the heat map of nucleosome occupancy to the heat maps of Hsf1 and Msn2/4 occupancy and examine if any meaningful clustering patterns emerge. Enrichment analysis of downstream genes can be carried out to find any overarching themes for sites that fall into the same clustering pattern. Expression data from previously reported heat shock time series [4] can also be used to narrow down the analysis to genes that show significant heat-induced activation or repression. It will also be interesting to see if similar clusters would be formed using nucleosome occupancy scores vs. using a TF occupancy score. If the same sites tend to be clustered together using the two independent scores, the results would support the existence of specific mechanisms that are linking nucleosome remodeling and TF binding events together.

After examining the characteristics of individual sites, it will be useful to group and analyze all genomic sites within the same promoter regions together to form a more complete picture of promoter activities leading to transcription. Plenty of evidence from past studies has indicated that nucleosome and TF binding sites within a promoter region may show very different occupancy patterns upon transcriptional perturbation. For instance, nucleosome eviction at one site might be accompanied by nucleosome appearance at a nearby site [2]. Hsf1 also shows differential binding activity to different HSEs within the same promoter [8]. By comparing occupancy-time profiles of neighboring sites, we might gain a better understanding of how nucleosome and TFs interact with each other. In fact, it has been suggested that Hsf1 and Msn2/4 can cooperate with each other to influence nucleosome remodeling at the promoters of some heat shock response genes [5]. It will be interesting to see if the data generated by this project can further support this hypothesis.

Aim 3: Identify interesting results for further validation and experimentation

By analyzing the nucleosome and TF occupancy-time profiles as described above, correlation patterns between nucleosome remodeling and TF binding may be identified. Based on these patterns, we might be able to form hypotheses regarding how nucleosome remodeling and TF binding interact with each other. For instance, if we see that the majority of TF binding occurs before nucleosome displacement from genomic DNA, a logical hypothesis would be that TF binding at a site causes nucleosome displacement from that site. On the other hand, if we observe that nucleosome displacement events tend to occur at earlier time points compared to TF binding events, we may hypothesize that nucleosome remodeling is necessary before TF binding can occur. It is possible that both scenarios will be observed in numerous occupancy-time profiles, in which case it will be informative to investigate when each mechanism is utilized. Since correlation does not necessarily indicate causality, follow-up experiments will be required to test these hypotheses.

Although the specifics of the follow-up experiments will depend on the results we generate from previous aims, one possibility is to conduct heat shock time courses using

gene knockout yeast strains. A previous study has used a *SNF2* single knockout and a *MSN2/4* double knockout strain to study nucleosome remodeling and TF binding at the promoters of several heat shock proteins [5]. *SNF2* encodes a component of the chromatin remodeling SWI/SNF complex, and its deletion can eliminate nucleosome displacement at some genomic sites. In particular, the *HSP12* promoter is usually regulated by both *HSF1* and *MSN2/4* and shows nucleosome displacement during heat shock in wild-type yeast cells. However, in the *SNF2* knockout strain, almost no nucleosome displacement activity was observed at the *HSP12* promoter, and Hsf1 binding was also eliminated. Interestingly, in the *MSN2/4* knockout strain, there was also no nucleosome displacement or Hsf1 binding at the *HSP12* promoter. These findings implicate a complex relationship between Hsf1, Msn2/4, and nucleosome displacement, and the genome-wide, higher-resolution data that will be generated in this project might be able to examine this relationship in more detail. By performing heat shock time courses using *SNF2*, *MSN2/4*, and *HSF1* knockout strains, we might be able to identify and validate cases in which TF binding is required for nucleosome remodeling, and vice versa.

Finally, if the results of these experiments turn out to be informative, we might expand our project to study various TFs under different perturbation conditions. It is entirely possible that trends we observe by studying nucleosome, Hsf1, and Msn2/4 behaviors during heat shock do not encompass the full range of mechanisms that can affect the relationship between nucleosome remodeling and TF binding. Experiments that focus specifically on nucleosome remodeling mechanisms, such as the *SNF2* knockout assays described above, might also be able to provide a different perspective to the problem discussed in this project.

References

[1] Workman, J. Nucleosome displacement in transcription. *Genes Dev.* **20**, 2009-2017 (2006).

[2] Shivaswamy, S *et al.* Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.* **6**(3), e65 (2008).

[3] Brogaard, K, Xi, L, Wang, J & Widom, J. A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486**, 496-501 (2012).

[4] Gasch, A *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241-4257 (2000).

[5] Erkina, TY, Tschetter, PA & Erkine, AM. Different requirements of the SWI/SNF complex for robust nucleosome displacement at promoters of Heat Shock Factor and Msn2- and Msn4-regulated heat shock genes. *Mol. Cell. Biol.* **28**(4), 1207-1217 (2008).

[6] Rhee, HS & Pugh, BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408-1419 (2011).

[7] Jakobsen, BK & Pelham, HR. Constitutive binding of yeast heat shock factor to DNA *in vivo*. *Mol. Cell. Biol.* **8**(11), 5040-5042 (1988).

[8] Giardina, C & Lis, JT. Dynamic protein-DNA architecture of a yeast heat shock promoter. *Mol. Cell. Biol.* **15**(5), 2737-2744 (1995).

[9] Martinez-Pastor, MT *et al*. (1996). The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress-response element (STRE). *EMBO J*. **15**(9), 2227-2235 (1996).

[10] Schmitt, AP & McEntee, K. Msn2p, a zinc finger DNA-binding protein, is the transcriptional activator the multistress response in *Saccharomyces cerevisiae*. *PNAS* **93**, 5777-5782 (1996).

[11] Hahn, J *et al*. Genome-wide analysis of the biology of stress responses through heat shock transcription factor. *Mol. Cell. Biol.* **24**(12), 5249-5256 (2004).