Alexander Garruss
Biophysics 205 Final

Several histone modifications have been identified and correlated to nearby gene expression during differentiation. One such example is the transition from a poised enhancer bearing H3K4me1 to an active state bearing H3K4me1 and H3K27ac (Rada-Iglesias et al. 2011). Although such state transitions have been documented in the literature, it remains to be seen if site-specific recruitment of chromatin modifying enzymes is sufficient to drive associated gene expression. In this proposal I outline a strategy to design and construct a protein that can has both the ability to be directed to regions of chromosomes at will, and has the enzymatic potential to acetylate H3K27. With such a dual-feature protein in hand, I propose to investigate the causal role of a H3K27 acetylation in the differential regulation of nearby gene expression in undifferentiated human embryonic stem cells (hESCs).

The field of transcriptional regulation in the context of distal enhancer regulatory elements has recently undergone a revolution. As researchers in the past normally worked on a gene-by-gene basis with a handful of regulatory regions of interest, the completion of the human genome sequence brought forth a genomic era. Today, researchers have the tools and techniques to measure the expression of tens of thousands of genes at once, and probe the individual roles of hundreds of thousands of putative regulatory elements. Connecting the spatial and temporal relevance of regulatory elements to the expression of their associated genes is now a grand challenge of the genomics era. Since greater than 97% of the genome does not code for protein, the vast landscape of regulatory potential must be explored in terms of its importance to development and disease (Herz et al. 2014).

One feature of the genomic environment in a cell that has been effective in linking distal enhancer regions to nearby gene expression is chromatin modification state. In the last few years, several groups have simultaneously profiled many chromatin modifications and identified correlations to nearby gene expression. In addition to dynamic chromatin modifications that correlate to changes in active transcription, some modifications pinpoint actively modulated regulatory regions. For instance, in hESCs the chromatin mark H3K27me3 has been shown to correlate with inactive expression of the nearest gene, while H3K27ac has been shown to correlate with relatively high expression of the nearest gene (Creygton et al. 2010). Furthermore, after differentiation of the embryonic stem cell into neuroepithelium, many activated genes appear to have enhancer regions transition from H3K27me3 to H3K27ac, suggesting a causal relationship between the chromatin state of the enhancer and the nearest gene's activity level (Rada-Iglesias et al. 2011).

While many distinct chromatin modifications have been correlated to distinct genomic properties, it remains to be clearly seen if the chromatin modifications at enhancer regions truly drive and cause the associated transcriptional changes, or if the modifications are instead a correlated consequence (Henikoff et al. 2011). The remainder of this proposal will detail one possible line of inquiry that hopes to

address the cause or consequence of chromatin modifications in enhancer-target gene regulation.

Specific Aim 1: In order to direct enhancer region transitions to an active state bearing H3K27ac, develop a Cas9-nuclease null fusion construct to the human p300 histone acetyltransferase.

The prokaryotic CRISPR (clustered regularly interspace short palindromic repeats)/Cas9 system can provide precise RNA-guided DNA binding. The system utilizes an RNA guide strand and normally a Cas9 protein with DNA nuclease activity found at two positions in in the protein (Qi et al. 2013). A critical first step in developing a fusion contruct of Cas9 to p300 for the purposes of this proposal is to remove the nuclease activity of Cas9, creating a Cas9-nuclease null (dCas9) protein. Modifying Cas9 will be done with lambda Red recombineering using the MAGE protocol in E. coli (Wang et al. 2009), whereby specific oligonucleotide sequences sharing homology to the Cas9 coding sequence introduce nucleotide changes to cause a single amino acid mutation at the two critical codons needed for active nuclease activity. Colonies of E. coli will be screened for defective Cas9 nuclease activity by expression of a guide RNA to a drug resistance gene on a plasmid. Select colonies that are still resistant to an antibacterial drug, possibly because they have a nuclease-null Cas9, will be further screened by allele-specific PCR reactions with primers specific to the codon changes. As a final screen, colonies containing plasmids with the desired dCas9 will be sequenced to ensure proper nuclease null mutations. Fusion of human p300 to C-terminus of dCas9 will be done similarly as described for a VP64:dCas9 fusion(Mali et al. 2013). The p300:dCas9 fusion will be checked by PCR for the correct size and by sequencing. Further validation of the fusion and functioning p300 will be done in vitro with an H3 histone substrate and by western blotting for acetylation.

Realization of a validated and functioning fusion construct enables expression of p300:dCas9 on a plasmid under a high-expressing promoter. The second component to testing the construct is to simultaneously express the fusion protein along with short guide RNAs (sgRNAs) to ~20 target regions in the human genome and to probe for proper recruitment and histone acetylation. The regions chosen for in vivo validation will be uniquely targetable by the sgRNA and tested one at a time. Expression of the sgRNA and fusion contruct on a plasmid will be transfected into hESC's using the lipofectamine reagent and protocol (Invitrogen). Existing data from the Gene Expression Omnibus (GEO) for chromatin profiles of H3K4me1, H3K27ac, and H3K27me3 in hESCs will be used to select candidate regions lacking both K27 modifications having the highest levels of K4me1. ChIP-PCR will first be used to detect occupancy of p300:dCas9 by using antibodies to either p300 or Cas9, and both will be tested independently. Primers specific to the regions targeted for recruitment will probe if recruitment occurred. Second, ChIP-PCR will be used to detect H3K27ac at the test regions of interest. Lastly, all ~20 sgRNA and p300:dCas9 plasmids targeting different regions will be cotransfected in a culture of human cells and be subjected to ChIP-seq using antibodies to p300,

Cas9, and H3K27ac in order to determine any off-target recruitment or aberrant acetylation, as compared to no transfection.

Several important considerations for this specific aim are as follows. The use of p300 is based on its implicated role in some of the most studied enhancer regions bearing H3K27ac to date. Furthermore, additional histone acetyltransferases, such as GCN5, are implicated in either promoter proximal acetylation or H3K9ac (Jin et al. 2011). For the purposes of attempting this novel strategy of site-directed chromatin modification, I will limit effort to a single enzyme, p300. The choice of hESC cells is based primarily on prior work describing chromatin modulation at enhancers in development, as well as the deep repository of chromatin profiles and gene expression datasets for this cell type. Guide RNAs that do not have matching homology to human sequences will be used as a negative control for changes in fusion protein recruitment and changes in acetylation as a separate ChIP-seq experiment, which will also serve as the baseline to determine fold changes of acetylation in the other, directed cases. Significant changes in ChIP-seq profiles will be assessed using common peak detection software using Poisson distribution at a Bonferroni adjusted p-value less than 0.05, using the non-targeting guide RNA as the control. The sgRNAs will be designed by finding regions of uniqueness at the target loci of interest and synthesized using standard microarray oligo-synthesis techniques, attempting 3-5 sgRNAs per site. In order to determine if global cellular pathways are disrupted by the presence of the plasmid containing a non-targeting sgRNA and fusion construct, microarray gene expression analysis will be performed to determine gene expression changes compared to normal hESCs.

Specific Aim 2: To recruit the functional fusion protein to a large number (~5,000) of loci that show enhancer state transition upon directed differentiation, bioinformatically design and synthesize many sgRNAs (and assemble a plasmid library) in order to localize the fusion product to many poised enhancers.

Because a large repository of chromatin profiles exist for H3K27ac (>20 cell type datasets at GEO for hESCs), first identify enhancer regions that lack both H3K27me3 and H3K27ac in hESCs, but have H3K27ac in any other differentiated human cell type. First, I will download and align unique reads from the ChIP-seq datasets for H3K27ac in several cell types early in differentiation of human embryos. For each dataset aligned, I will determine areas of enrichment for H3K27ac using peak detection at FDR < 0.05 with the associated whole extract controls, and determine peak regions. For each downloaded dataset, I will compare the areas of enrichment to the hESC peak regions in order to determine newly acquired H3K27ac regions in the differentiated cell type data. By determining all pair-wise differences between each differentiated cell type to hESCs, I can compile a superset of all regions that become H3K27ac, which are not acetylated in the hESC condition. I will exclude from the superset regions that have active repression of H3K27me3 in hESCs. Lastly, I will exclude further regions that lack H3K4me1 in hESCs. The final set of regions corresponds to putative enhancers in human embryonic stem cells that are poised for activity, lacking active repression

(H3K27me3), lacking the active enhancer signature (H3K27ac), but are bookmarked for activation (H3K4me1). From this set of candidate regions, I will select those that are uniquely targetable by sgRNAs, and prioritize based on the level of H3K4me1 occupancy, which reflects a proportion of cells in the population that are poised.

Utilizing the large set of candidate regions and corresponding unique sgRNA designs, I will synthesize the sgRNAs, creating a library of those oligos. This library will be transformed into a plasmid under control of an active promoter, along with the p300:dCas9 construct, as previously described. The library of plasmids corresponding to differing sgRNAs and p300:dCas9 will be transfected into the human cell culture all together.

In order to determine global recruitment of the fusion constructs, several ChIP-seq experiments will be performed on the pool of cells, using antibodies against p300 and Cas9, and whole cell extract controls. Peak regions corresponding to areas of recruitment of the fusion construct will be determined. Furthermore, a ChIP-seq experiment will be performed using an antibody for H3K27ac in order to determine sites that accepted the chromatin modification. Regions harboring both a recruited fusion protein and modified chromatin, as compared to normal hESCs, will be considered actively transformed enhancers. To address the question as to the role transformed enhancers play in driving nearest gene expression, the nearest protein-coding gene will be determined. Additionally, I will perform RNA-seq to determine gene expression for every polyadenylated transcript in normal hESCs and the pool containing transformed enhancers. After aligning the data keeping only unique reads, I will assess differential expression between all genes using a standard negative-binomial distribution, calling significantly changed genes at Bonferroni adjusted p-value less than 0.05. Two paths of analysis are possible with the combined transformed enhancer's nearest genes and the gene expression data. First, I will determine the percentage of transformed enhancers that show significantly increased gene expression. Second, I will determine a correlation coefficient between levels of H3K27ac change at enhancer regions to the level of gene expression change for the nearest protein-coding gene.

Taking a candidate approach as described using thousands of putative poised regions should allow for interpretation of the data even in the face of failed recruitment or acetylation at some loci. Since the downstream analysis will consider only those enhancers deemed to be transformed, one can limit their interpretations to those enhancer/nearest gene pairs that perform as expected. However, the choice of candidate regions may be error prone since I will use differentiated cell type chromatin profiles in order to select candidate regions. There will certainly be some cellular contexts where higher-order regulation may prevent direct activation of nearest genes. Also, the nearest gene assumption may not hold for a good number of enhancer-gene pairs, further complicating interpretation. For those transformed enhancer-nearest gene pairs that do show correlated changes, ChIP-PCR validation will be done to the enhancer regions, and PCR validation of the gene expression changes will be done expressing the

corresponding sgRNA independently. Because of the –omics approaches suggested, there will be many regions not directed by sgRNAs that will serve as negative controls for recruitment and chromatin modification. Finally, as a negative control in determining whether H3K27ac site-specific recruitment can drive nearest gene expression, candidate pairs of transformed enhancers and nearest genes that show correlative changes will be tested using a catalytically-null p300 version of the fusion construct, assembled similarly as in aim one. For several of the interesting pairs found, I want to determine if the presence of p300 alone, with acetylation potential, is sufficient to drive nearest gene expression. If the nearest gene expression, as measured by PCR, does not increase with the null p300, then one would be more confident that H3K27ac is indeed a driving force in regulating associated gene expression.

Here, I present a proposal to assess the cause of chromatin modifications on nearest gene expression. First, I will develop a fusion construct containing an active p300 fused to the C-terminus of a nuclease-null Cas9 protein. The use of Cas9 is driven by the ability to direct the protein to any unique site of DNA of interest, using only a short guide RNA. By designing the guide RNAs to target putative poised enhancers, p300 activity may be site-directed as well. Thus, a tremendous feature of the proposed work is the ability to modulate chromatin modifications in a highly specific manner. Since I only utilize a single protein fusion, p300, extensions to this work may involve other families of histone acetyltransferases. Further extending the proposed idea further to other chromatin modifying fusion constructs may open the door to discovering much more about the role of gene regulation and chromatin state, and may offer an exciting new therapeutic avenue for those suffering from "enhanceropothies" (Smith et al. 2014).

References:

Mali et al. Nature Biotechnology 2013
CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.

Creyghton et al. PNAS. 2010
Histone H3K27ac separates active from poised enhancers and predicts developmental state

Rada-Iglesias et al. Nature. 2011
A unique chromatin signature uncovers early developmental enhancers in humans.

Qi et al. Cell 2013
Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression

Jin et al. EMBO J 2011
Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation.

Wang et al. Nature 2009
Programming cells by multiplex genome engineering and accelerated evolution

Henikoff et al. Trends Genet. 2011
Histone modification: cause or cog?

Smith et al. Nat Struct Mol Biol. 2014
Enhancer biology and enhanceropathies

Herz et al. Mol Cell. 2104
Enhancer Malfunction in Cancer.