GENOME-WIDE ANALYSIS OF TRANSCRIPTIONAL REPRESSION MECHANISMS IN THE DROSOPHILA EMBRYO

By

Kurtulus Kok

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Genetics – Doctor of Philosophy

2016

ABSTRACT

GENOME-WIDE ANALYSIS OF TRANSCRIPTIONAL REPRESSION MECHANISMS IN THE DROSOPHILA EMBRYO

By

Kurtulus Kok

Transcriptional repressors control temporal and spatial patterns of gene expression in animal development through co-repressors, which interact with histone modifying enzymes and chromatin remodelers. In this work, I used genome-wide approaches to understand the mechanisms of repressors and corepressors by identifying biochemical changes on chromatin using the Drosophila Hairy long-range transcriptional repressor protein as a paradigm. I found that Hairy induces wide-spread and diverse changes in histone modifications. Intriguingly, many sites are targeted errantly by Hairy to modify chromatin landscape, even though gene expression is unaffected. I propose that many eukaryotic transcription factors may induce similar dynamic modifications on off-target sites, and speculate that this errant activity may provide a path for creation of new regulatory elements, facilitating the evolution of novel transcriptional circuits. In addition to these evolutionary insights, my studies on roles of corepressors showed that CtBP corepressor – also associated with short-range repressors - contributes to Hairy mediated repression in a quantitative and gene-specific fashion, suggesting that different classes of transcriptional repressors can utilize common corepressors to effect distinct histone modification patterns across the genome.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation and gratitude to my mentor, Dr. David Arnosti, for his enthusiasm, patience, encouraging my research and for allowing me to grow as a better scientist. His advice on both research as well as on my career have been priceless. He has been a tremendous mentor for me. Without his supervision and constant help this dissertation would not have been possible. I would like to acknowledge all my lab members for their support, insight, discussions, and being great friends, with special thanks to Dr. Sandhya Payankaulam for being a excellent colleague. I would also like to thank Dr. R. William Henry, Dr. Lee Kroos, Dr. Shin-Han Shiu, and Dr. Michael Thomashow for serving as my committee members. My research project was substantially enriched by their brilliant comments, suggestions and thoughtful contributions. I also want to thank to the Genetics Program for providing an excellent environment for graduate students and their financial support. A special thanks to my family in Turkey. I am grateful to my parents, brothers and sisters, whose supports and encouragements sustained me thus far. I would like to acknowledge all of my friends who helped, and supported me during my graduate life. At the end I would like express appreciation to my beloved wife, Khalida Gasimova, who was always my support.

TABLE OF CONTENTS

LIST OF FIGURES

KEY TO ABBREVIATIONS

CtBP- C-terminal binding protein

Gro- Groucho

Sir2- Silent Information Regulator 2

dTopors- *Drosophila* homolog of Topoisomerase I (TopI) binding protein (Topors)

HDAC- Histone Deacetylase

HAT- Histone Acetyltransferase

HMT- Histone Methyltransferase

KDM- Lysine Demethylase

E(Z)- Enhancer of Zaste

HES- Hairy/Enhancer of Split/Deadpan

PcG- Polycomb Group

TF- Transcription Factor

TFBS- Transcription Factor Binding Site

GRN- Gene Regulatory Network

bHLH- Basic helix-loop-helix

ChIP- chromatin immunoprecipitation

FAIRE- Formaldehyde Assisted Isolation of Regulatory Elements

MNase- Micrococcal nuclease

STARR-seq- self- transcribing active regulatory region sequencing

modENCODE- Model Organism Encyclopedia Of DNA Elements

eve- even-skipped

GTFs- General Transcription Factors

PIC- Preinitiation Complex

TSS- transcription start site

UAS- upstream activation sequences

ES- embryonic stem cells

PSM- presomitic mesoderm

NPC- neuronal progenitor cells

HD- Huntington's disease

GR- glucocorticoid receptor

ER- human estrogen receptor

CHAPTER I

Cis-regulatory regions in gene regulation and the specificity of transcription factor interactions: a "shotgun" model for promiscuous interactions of metazoan transcription factors[1]

Keywords: Hairy, evolution, gene regulatory network, chromatin, transcription factor

[1]Chapter I is presented in the form of a manuscript for publication as: Kurtulus Kok, and David N. Arnosti. Duck hunting and specificity of transcription factor interactions: a "shotgun" model for promiscuous interactions of metazoan transcription factors.

1.Introduction

Regulation of genetic information in a precise temporal and spatial manner is essential for proper cell behavior and development. The initial level of gene expression, transcription, is controlled by the interaction of soluble *trans* regulatory factors and the DNA switch elements, or *cis*-regulatory elements (CREs), with which they interact (Lee and Young 2013). In bacteria, recognition sites for regulatory proteins often consist of motifs 12-30 bp in size; there is a good correlation between the presence of a site in the genome and occupancy by cognate transcription factors (TF), with few exceptions (Rodionov 2007). In metazoans, CREs contain shorter, typically 6–10 bp binding sites that interact with TF required for transcriptional regulation (Arnosti and Kulkarni 2005). Occupancy of these sites is more context-dependent, being influenced by

nucleosome occupancy, histone modifications, and presence of other proteins (Voss and Hager 2014; Lickwar et al. 2012). The challenge to specificity of binding to functional elements is further influenced by the generally larger size of the genome, the presence of repeat elements and the overall higher percentage of non-coding sequences. These factors mean that the challenge to achieve specificity and precision of TF binding to drive regulatory programs is very great in higher eukaryotes.

Considering metazoan TF-DNA interactions, sites of high affinity can be found throughout the genome; a TF that recognizes a 6-bp motif would be expected to bind every ~4 kb. For example, the MyoD TF typically binds to CANNTG motifs; there are more than fourteen million such consensus “E-boxes” in the human genome (Cao et al. 2010). However, only a small fraction of motifs are occupied in a given cellular or developmental context, due to the influence of chromatin packaging and other factors (Voss and Hager 2014). Indeed, metazoan TFs are typically found to bind ~10^3-10^4 regions in a given developmental stage or cell type (Biggin 2011).

Even though the relationship between *in vivo* binding and intrinsic DNA recognition properties is poorly understood, the selection of optimal sets of targets is influenced by DNA sequence, as well as the chromatin state and protein-protein interactions (Hager et al. 2009; Slattery et al. 2011). Specificity of binding preferences can be conferred by not only the consensus motif but also

flanking nucleotides, which may influence the physical structure of a region of DNA or allow binding of a cooperatively acting factor (Villar et al. 2014). CREs often contain motif clusters recognized by different TFs, permitting binding in a cooperative manner by homotypic or heterotypic combinatorial interactions (Villar et al. 2014). Furthermore, nucleosomes often compete with TFs to contact DNA, leading to ordered assembly of TF complexes that depends on the recruitment of histone remodeling factors (Hager et al. 2009). Post-translational modifications of histone proteins may further influence specific binding by attracting protein complexes such as Polycomb-group proteins that impact TF binding and activity (Rando 2012).

Thermodynamic and kinetic properties of TFs influence DNA binding levels – in general, higher concentrations of a TF in the nucleus lead to greater levels of binding (Hager et al. 2009). In addition, the dynamics of protein interactions with the DNA impacts the sensitivity and robustness of regulatory responses. Changing the transcription factor concentration of a nucleus is central to cellular response and differentiation. Indeed, it is not coincidental that the canonical Yamanaka factors used to induce cellular reprogramming are all TFs (Takahashi and Yamanaka 2006).

In addition to the complexity of factors that dictate specific TF binding, there is an additional, fundamental issue in understanding these protein-DNA regulatory systems: only a small subset of TF-DNA interactions appear to be functionally

important in transcriptional regulation. The majority of the thousands of interactions between TF and DNA appear to be nonfunctional and thus inconsequential in regulating transcription. What distinguishes functional from nonfunctional binding is still unclear. The implications of TF activity on these “off-target” sites are addressed in this review.

2. What is a CRE?

To establish proper transcriptional outputs across the genome, short collections of DNA sequences termed CREs are required for regulation of gene expression. Bacterial CREs often have only a few binding sites, in contrast to eukaryotic CREs that may contain dozens of sites (Payankaulam et al. 2010). In higher eukaryotes, three general classes of CREs are found: core promoters, enhancers (both promoter-proximal and distal) and insulators (Maston et al. 2006). The core promoter consists of a region of about 100 bp that overlaps the transcription start site (TSS) and recognized by general transcription factors to assist formation and activation of the RNA polymerase complex. In the well-characterized system of Drosophila, core promoters have been classified into focused or dispersed types (Juven-Gershon et al. 2008). Focused promoters have either a single TSS or few tightly clustered start sites. In contrast, start sites are spread over 50–100 nucleotides in dispersed promoters. Although textbook illustrations often feature focused promoters, in many eukaryotes, they represent a minority of actual promoters. In vertebrates, for instance, a majority of promoters are do not drive single or even clustered initiation events (Juven-Gershon et al. 2008). Drosophila

core promoters may contain sequences such as the TFIID-binding TATA box, initiator (INR), downstream core promoter element (DPE), and Motif Ten Element (MTE), typically in focused promoters; mammalian promoters similarly contain elements such as TATA and B recognition element (BRE) which bind basal machinery to position the start of transcription, or they may be characterized by CpG-rich regions associated with highly dispersed initiation events. TATA box and BRE are the most ancient motifs, and conserved from Archaea to humans (Lenhard et al. 2012).

The specific combinations of these motifs in core promoters may represent different promoter types associated with distinct classes of genes such as developmental regulation, housekeeping, and tissue-specific differentiation (Ohler 2006; Engström et al. 2007; Zabidi et al. 2015; Wei and Arnosti 2015). The basal promoter can influence which CREs can communicate with a gene, as well as controlling RNA polymerase II stalling (Zeitlinger et al. 2007; Kadonaga 2012).

Basal promoter sequences are required for transcription, but the levels and timing of expression are generally controlled by additional regulatory elements contained within enhancers, which are *cis* elements typically ranging in size from 100-1000 bp, featuring collections of sequence-specific motifs that load TF proteins, and their associated co-activators and co-repressors (Spitz and Furlong 2012). TF binding at CREs often alter chromatin architecture through chromatin-

modifying factors such as histone modifiers or nucleosome remodelers (Fuda et al. 2009). Some TFs defined as pioneer factors function to exclude nucleosomes and recruit other factors by creating accessible DNA regions (Zhang et al. 2012).

Combinatorial interactions of TF and co-factors fine-tune the activity of the enhancers to produce a regulatory output that can act in a modular fashion, independent of the distance and orientation to the TSS (Arnosti and Kulkarni 2005). Although bacterial regulatory motifs are generally located within a few hundred bp of the TSS, some bacterial CREs possess enhancer-like activities and can be located over one kilobase from the TSS (Xu and Hoover 2001). Yeast regulatory sequences, termed upstream activation sequences (UASs), usually contain multiple TF binding motifs (Dobi and Winston 2007). Similar to the situation in bacteria, these regulatory elements act locally, and are usually located within a few hundred bp of the core promoter. Changing their position to a more distal region inactivates the regulatory elements, which is a useful property, considering the dense packing genes in *S. cerevisiae* (Dobi and Winston 2007).

The DNA sequence information present in enhancers enables TFs to direct exact temporal and spatial activity of a gene in development, and during changes in cell physiology in response to signaling. The regulatory code of the animal genome is highly complex and context-specific, and depends on cell-type (Lee and Young 2013). The general outlines of eukaryotic transcriptional control

systems have yielded to dedicated research of the past three decades, however the detailed ‘grammar’ pertaining to the specific ways that DNA sequence influences complex patterns of gene expression remains a formidable challenge. There are many factors that influence how sets of binding sites may function together to influence transcription of genes. Three general models proposed for enhancer activity include the enhanceosome, the billboard and the TF collective. In the enhanceosome model, highly scaffolded binding of all TFs to the enhancer is essential for very specific cooperative interactions to occur. Thus, the position and stoichiometry of binding motifs are constrained; small changes in DNA sequence can disrupt cooperative interactions and significantly impact the output (Arnosti and Kulkarni 2005). In contrast, the billboard model suggests that enhancers serve as less-constrained ‘information display’ elements where TFs still can act cooperatively, but with flexible positioning and composition of their binding sites (Arnosti and Kulkarni 2005). The TF collective model was proposed based on the observations that TFs bind to enhancers even in the absence of recognizable cognate motifs. Therefore, indirect DNA recognition through protein-protein interactions may be another layer to the DNA code to activate enhancers (Spitz and Furlong 2012).

The long-range interactions of enhancers with promoters can be restricted and blocked by insulator CREs, which are bound by sequence-specific proteins such as CTCF. Similar to enhancers, they can be located anywhere, but unlike enhancers, they possess no intrinsic activation or repressing function. Another

important role of insulators is to prevent spreading of repressive heterochromatin (Valenzuela and Kamakaka 2006).

3. How do we find CREs?

The complex regulatory logic underlying specific gene networks has remained incomplete in part because we lack a comprehensive “parts list” of regulatory elements. Three different approaches have been useful in bridging this gap. First, bioinformatic approaches use expression data with DNA sequence information about possible TF binding sites present near genes to infer regulatory networks. Second, chromatin accessibility and protein binding information derived the genome-wide chromatin immunoprecipitation (ChIP) experiments has provided useful landmarks within the genome that correlate with transcriptionally active regions. Finally, functional assays of candidate elements, a traditionally slow process, has recently become amenable to high-throughput methods, providing significant additional orthogonal perspectives on regulatory elements (Rando and Chang 2009; Schones and Zhao 2008; Hawkins et al. 2010; Hardison and Taylor 2012; Shlyueva et al. 2014).

3.1. Functional assays

The earliest characterization of CREs involved the use of functional assays, in which a candidate sequence is linked to a basal promoter to drive a reporter gene, *in vitro*, in transient assays, or more usefully, when integrated into a genomic locus. To obtain information about spatial and temporal activity of enhancers in whole organisms or cells, commonly used approaches include *in*

situ mRNA hybridization, fluorescent protein readout, or enzyme activity assays. Especially in cell culture, these assays can be quantitative, but are most commonly used qualitatively when applied to whole tissues. Using such approaches, the first description of transcriptional enhancers was achieved; these studies reported the activities of viral and subsequently metazoan sequences that potently regulate transcription independent of the location and orientation with respect to the TSS. Cell culture or transgenic embryo assays are still valuable tests to individual candidates and verify enhancers identified by genome-wide studies. However, traditionally, these approaches were used to assay only small numbers of enhancers. Recently, high-throughput approaches have been developed to discover enhancers on a genome-wide scale. In some cases, traditional reporter assays have been scaled up to survey thousands of different transgenic lines; these studies generally have not completely mapped the entire genome. Yeast one-hybrid assays coupled with a robotic mating platform and automated readout quantification allowed identification of CREs in *Caenorhabditis elegans* (Reece-Hoyes et al. 2011). Unbiased genome-wide search for enhancers in the ascidian *Ciona intestinalis* using a *lacZ* reporter gene driven by 138 random genomic DNA fragments with an average size of 1.7 kb provided tissue-specific identification of CREs (Harafuji et al. 2002). Fragments representing 13% of the non-coding non-repetitive genome in Drosophila were characterized for activity in the embryo, resulting in the identification of thousands of potential enhancers (Kvon et al. 2014). Similarly, hundreds of genomic regions were tested in mouse embryo using transposon-associated

regulatory sensors (Chen et al. 2013). An alternative approach uses self-transcribing active regulatory region sequencing (STARR-seq), in which randomly fragmented libraries of candidate sequences are placed downstream of a minimal promoter to drive their own transcripts. The activity can be measured quantitatively by high-throughput sequencing of mRNA (Arnold et al. 2013). This approach only surveys activity in a given cell type, but the coverage encompasses most of the genome.

3.2. TF motifs and sequence conservation

CREs represent collections of motifs for sequence-specific transcriptional regulators, thus one straightforward approach to identification of CREs is identification of conserved sequences outside of protein coding regions. Such studies of sequence conservation have unearthed hundreds of human ultraconserved non-coding DNA sequences at least 200 bp long (~1.3 kbp average length); some have been tested for regulatory activity in transgenic mouse embryos. 45% of these sequences drove expression in embryos as tissue-specific enhancers, indicating that the conservation of sequence may reflect their roles in regulation (Pennacchio et al. 2006). In many cases, the level of sequence conservation of *bona fide* enhancers is not nearly as absolute as those found in ultraconserved sequence, but there is information available to indicate likely regulatory regions (Frazer et al. 2004; Johnson et al. 2005). However, many regulatory regions lack this level of conservation. For instance, functionally conserved enhancers from *even-skipped* genes from different species of Drosophila drive identical expression patterns in *D. melanogaster*

embryos, but the enhancer sequences do not show significant similarity, indicating a substantial change with time (Hare et al. 2008). Of course, not all of the sequence of an enhancer is comprised of binding motifs; spaces between motifs may freely vary without an impact on function. Even so, functional sites may vary extensively. A comparison of confirmed binding sites within regulatory elements in *D. melanogaster* and presumed homologous elements in *D. pseudoobscura* showed that overall conservation of binding sites was only slightly greater than overall conservation between these two species (Sinha and Siggia 2005). In some cases, such divergence at the sequence level may represent changes in function, although the *cis* element is still active. A comparison of mouse and human DNaseI hypersensitive patterns, which are a measure of TF-DNA interactions *in vivo*, revealed extensive turnover of TF motifs in shared *cis*-regulatory regions (Vierstra et al. 2014). Therefore, predictions based solely on sequence conservation can result in low accuracy; better results are likely obtained if sequence-based analysis is combined with genome-wide *in vivo* datasets discussed below.

Attempts to map regulatory regions based on the presence of a motif for a particular TF have been stymied by additional factors that impact DNA interactions. *In vivo*, most canonical TF motifs are not occupied, due to nucleosome occupancy, lack of cooperating neighboring sites, or other context-specific effects. Because single TF sites may not be functional, and in light of the flexibility of arrangement of regulatory sequences, an alternative bioinformatic

approach involves the detection of clusters of particular motifs, either corresponding to known regulatory factors or simply overrepresented in the region of interest (Kazemian et al. 2014). Not all enhancers contain homotypic repeats for bindings sites, however. Furthermore, this approach does not capture the interactions of TFs with non-canonical 'cryptic' motifs (Narasimhan et al. 2015). In Drosophila, where evolution of binding events by embryonic TF has been extensively studied, a common finding is that many regulatory regions maintain activity through evolutionary time, even as the underlying sequences or TF binding show considerable variation. Thus, the regulatory "logic" of these elements may change slower than the actual DNA sequence (He et al. 2011; Paris et al. 2013). Comparative analysis of genome-wide binding profiles of three TFs in liver among six rodents, including five closely related mouse species, showed that the differences in TF binding site contents among these species emerge frequently, even more quickly than in the cases studied in Drosophila species. In this case, the differences in occupancies observed were not explained by gain or loss of the cognate TF motifs, suggesting that other factors were affected, and binding alterations were a reflection of combinatorial binding (Stefflova et al. 2013). In contrast, the analysis of TF binding in livers of five highly divergent vertebrates showed that loss of TF binding mostly coincided with changes in the directly bound motifs (Schmidt et al. 2010a). In summary, while there is clearly important information in DNA sequences that influences TF binding with enhancers, in many systems, consideration of DNA motifs alone,

even measured in clusters, may give erroneous predictions of likely regulatory sequences.

3.3. Open chromatin structure

The interaction of regulatory factors with CREs leads to stereotypical changes in chromatin structure such as nucleosome movement or depletion, with concomitant increases in DNA accessibility to nuclease treatment. Changes in chromatin structure may reflect the competitive binding of TFs with nucleosomes to open a region of DNA, or introduction of bends in the DNA. Additionally, TF can recruit remodeling complexes that move and displace nucleosomes in an ATP-dependent manner. Some regulatory factors, such as the mammalian FOXA1 and Drosophila Zelda proteins, are defined as pioneer factors. These proteins may access regulatory regions early in development to deplete nucleosomes, providing accessible regions for binding of other TFs later in development. Such structural changes in chromatin over CREs are identified by digestion of accessible DNA using DNase I or micrococcal nuclease (MNase). These approaches can be coupled to deep sequencing to extend the analysis to the whole genome. An alternative technology, FAIRE (formaldehyde-assisted identification of regulatory elements), identifies genomic regions with reduced nucleosome and protein content by light chemical crosslinking with formaldehyde and deep sequencing. The simplicity of the ATAC-seq approach, which measures chromatin accessibility by virtue of insertion of easily identified transposon tagging elements, has also provided valuable information about structural properties of chromatin (Bao et al. 2015). Different techniques used to

measure chromatin structure have distinct advantages and disadvantages. DNase- and MNase-seq have higher sensitivity compared to FAIRE-seq, but their inherent cleavage patterns show more variation due to the sequence-specificity of the enzymes. FAIRE-seq, on the other hand, requires only physical shearing to identify regions of interest. At deeper sequencing levels, DNase-seq can provide TF footprints with high resolution (Simon et al. 2012). It is important to note that even though CREs can be discovered using these techniques, not all regions presenting these features necessarily regulate gene expression; open regions may also represent features such as insulator elements and regions primed by pioneer factors to be activated at some later juncture.

Evidence that DNaseI hypersensitive regions do indeed correspond to functional elements is supported by genetics studies. Genome-wide mapping and correlation of expression quantitative trait loci (eQTLs) and DNase I hypersensitive sites (DHS) were used to characterize regulatory regions in 70 human lymphoblastoid cell lines. Thousands of DHS sites were significantly associated with genotype of a nearby genetic variation, which are enriched within TF binding sites. 16% of these loci were linked to nearby genes with changed gene expression (Degner et al. 2012). Not all DHS may represent *bona fide* regulatory regions, but the correlation indicates that these approaches can identify regulatory variations linked to phenotypes, complex traits and diseases (Albert and Kruglyak 2015).

3.4. Chromatin marks

In eukaryotes, chromatin is dynamically regulated through posttranslational modifications of histone tails or by introduction of specific histone variants (Jiang and Pugh 2009). Specific patterns of histone marks are associated with nucleosomes present on or near CREs (Rando 2012); these include enrichment of H3K27Ac and H3K4me1 at enhancers, enrichment of H3K4me3, but lower levels of H3K4me1 at promoters and the presence of H3K9me3 and H3K27me3 at silent regions (Zhou et al. 2011). Some enhancers carry marks associated with both activation and repression. Such 'poised' enhancers exhibit bivalent features e.g. H3K4me1 and H3K27me3. Other enhancers (termed "latent") lack active marks, but may gain H3K4me1 and H3K27ac upon signaling stimulation. Several chromatin features are widely used to predict tissue-specific enhancer activity. For instance, in one study, twelve of eighteen regions characterized by tissue-specific H3K27Ac showed activity as embryonic forebrain enhancers in the mouse (Nord et al. 2013). Combinations of histone modifications can also be informative; in human embryonic stem (ES) cells, the presence of both H3K27Ac and H3K4me1 is correlated with enhancers near active genes, while H3K4me1 alone without H3K27Ac marks inactive enhancers. The differentiation of ES cells into a neuronal pathway was associated with the loss of the H3K27me3 mark and gain of H3K27Ac at many poised enhancers (Rada-Iglesias et al. 2011). Combinations of marks can indicate active elements, such as elevated H3K4me1 and H3K27ac, with low H3K4me3 and absence of H3K27me3 or H3K4me1 (Heintzman et al. 2007). Another study correlated H3K27ac and H3K79me3 to

increase the prediction accuracy (Bonn et al. 2012), but there is still no agreement for the types of marks most effective in identifying enhancers, since additional contextual factors may affect activity of the enhancer. For instance, in considering p300/CBP histone acetyltransferases in mouse fibroblasts, a significant correlation between histone acetylation and target gene activity was not found (Bedford and Brindle 2012). In addition, many active enhancers (~40%) do not exhibit H3K27Ac at active mesodermal enhancers in Drosophila embryos (Bonn et al. 2012). Similarly, H3K27me3 is enriched at only 35% of silenced promoters, and only a fraction of repressed genes ever acquired this repressive mark in T cell differentiation at inactivated regions (Zhang et al. 2012). Therefore, while broad correlations do indicate that such marks are associated with CRE function, the causative roles of such modifications and context-specificity for the function of particular CREs are unclear and require further research.

3.5. Enhancer transcription as a mark of enhancer activity

Genome-wide GRO-seq, PRO-seq and CAP-seq studies map the activity of eukaryotic RNA polymerase II, regardless of whether the enzyme is producing a stable transcript or not (Guertin et al. 2012; Kwak et al. 2013; Gu et al. 2012). In these studies, Adelman, Lis, the FANTOM consortium and others have found that regulatory regions are frequently bidirectionally transcribed; equivalent levels of divergent transcription leading to unstable “eRNAs” have been suggested to mark active enhancers, as compared to the histone modifications discussed above (De Santa et al. 2010; Hah et al. 2011; Shlyueva et al. 2014).

3.6. Enhancer-promoter interactions

A widely accepted model for enhancer function is “looping”, the direct interaction of enhancers and promoters. These interactions are supported by physical *in vivo* measurements. In fact, the association of enhancers and target promoters has been used to detect the location of enhancers. One approach has focused on profiling the binding of proteins that play important roles in bridging enhancers and promoters. For example, ChIP-seq analysis of cohesin and Mediator revealed information about enhancers in murine embryonic stem cells (Kagey et al. 2010). A more commonly used strategy, chromatin conformation capture or 3C and its high-throughput variations (e.g. Hi-C), involves fixing spatially proximal chromosomal regions with formaldehyde treatment. The subsequent ligation of sheared genomic DNA reveals long-range contacts (Lieberman-Aiden et al. 2009). The capture of such conformations shows that spatial organization of chromosomes in the nucleus is not random, and is similar enough across cells to provide reproducible, broad patterns, however, such chromosome “packing” has not in general been proven to be entirely or even largely a function of enhancer-promoter connections (Ghavi-Helm et al. 2014). A modified version of this method, chromatin interaction analysis with paired-end tag sequencing (ChIA–PET), uses probes against a protein of interest to detect chromatin interactions. Since enhancers stimulate RNAP II complexes, use of antibodies to RNAP II in this method provides insights into enhancer and target gene interaction (Li et al. 2012). This method reveals direct targets of enhancers, which may interact with distal TSS that are not the most proximal to the CRE. Such empirically measured

interactions are valuable, as computational analysis of genome-wide expression and protein-binding data usually involves linking the enhancers to the most proximal genes. The resolution of such chromatin conformation approaches can be limited, e.g. 0.1-1 Mb, which can be improved with greater sequencing depth to distances of ~20 kbp, often sufficient to determine likely target TSS of enhancers (van Steensel and Dekker 2010).

3.7. TF binding

Tissue-specific occupancy of TFs is a strong predictor of CREs, especially if the combinatorial patterns of multiple TFs commonly associated with active regulatory elements are taken into account. In a pioneering study, the temporal and combinatorial occupancy of five TFs during different stages of Drosophila mesoderm development were used to train machine learning models to predict high accuracy enhancer predictions (Zinzen et al. 2009). As part of their biochemical functions, TFs recruit co-factors that alter the chromatin environment and interact with the basal machinery. The chromatin association of co-factors such as the p300 histone acetyltransferase is therefore also used to identify CREs; for example in the mouse embryonic forebrain, midbrain and limb tissues p300 occupancy was used to predict enhancer activities. Most of the predicted regions (75 out of 86) showed consistent expression in tissues in transgenic mouse assays where the candidates were predicted (Visel et al. 2009). The binding of transcriptional co-repressors is also informative; in early embryonic stages of *Xenopus tropicalis*, occupancy of the TLE/Groucho factor provides

even better information than p300 binding to predict tissue-specific CREs (Yasuoka et al. 2014).

4. Off-target activities of TFs - nonfunctional vs. functional regulation

To understand the regulation of gene expression by TFs in molecular terms, it is essential to determine both physical interactions between regulatory factors and genes, as well as potential functional consequences. Physical localization of regulatory proteins by ChIP and DNase I digestion on the genome can be carried out with high precision, however, assigning functional roles to specific binding interactions is challenging. A correlation of changes in gene expression with the depletion or overexpression of TFs that bind near specific loci is often the primary information used to infer activity of presumed CREs; this approach has been employed from yeast to mammals. In many metazoan studies, a common finding is that even though TFs bind to thousands of regions in genome, only small subsets of these interactions are associated with changes in gene expression. In general, the overall view is that the majority of the interactions between TFs and genome may be non-functional, and are not important for the activity of GRNs.

In a comprehensive analysis of yeast GRNs, targets of 263 TFs were identified by measuring transcriptional profiles in strains that contained deletions in genes for specific TF (Hu et al. 2007). Considering the DNA binding profile for 188 of these 263 TFs (Harbison et al. 2004) and correlation with these transcriptionally altered targets, a small fraction of directly bound genes were observed to show

transcriptional effects upon deletion of the gene for specific TFs. The many apparently non-functional TF binding interactions might be partially explained by backup systems in GRNs, whereby more than one input stabilizes a particular regulatory link, so that ablation of one factor has only a minor impact (Gitter et al. 2009). Indeed, in yeast mutations in TFs with paralogs were found to be less likely to produce a transcriptional response than similar disruption of TFs without paralogs, indicating overlapping functions and redundancy (Gitter et al. 2009). A different approach to understand the functional consequences of TF binding employed measurement of high-resolution temporal dynamic binding of the yeast Rap1 factor using genome-wide competition ChIP, in which Rap1-Myc and Rap1-Flag were induced by different promoters, allowing detection of dynamic Rap1 binding. By analyzing RNAP II recruitment and transcript levels, it appeared that there is a stronger correlation between function and Rap1 binding turnover than steady-state occupancy (Lickwar et al. 2012).

One of the best described systems for metazoan TF interactions comes from genome-wide mapping of TF occupancy in the Drosophila embryo; for a number of developmental factors, there were thousands of bound regions with varying levels of DNA occupancy (Li et al. 2008; MacArthur et al. 2009). To determine whether magnitudes of the DNA binding features reflected function, e.g. higher-affinity binding at regulated genes and lower-affinity binding at non-regulated genes (Biggin 2011) a number of medium or low occupancy sites were tested in transgenic reporter assays, with the finding that lowly bound regions tended not

to drive gene expression, in contrast to highly bound regions that drive patterns of expression in the embryo (Fisher et al. 2012). This study added support to the model that much low-affinity or off-target binding may be of no functional significance.

Studies of mammalian systems similarly show a discordance between TF binding and transcriptional impact. The expression profile of mouse liver genes in the presence or absence of exogenous glucocorticoid was integrated with genome-wide location of glucocorticoid receptor (GR) binding. About 20% of GR bound genes were transcriptionally responsive to GR (Phuc Le et al. 2005). In revealing human estrogen receptor (ER) regulatory circuitry in breast cancer cells, coupled analysis of gene expression and ER binding showed that a small fraction (13%) of direct ER target genes were highly responsive to estrogen (Kwon et al. 2007). Similar conclusions were reached from analysis of p63, a homolog of the p53 tumor suppressor. Depletion of p63 protein in cervical carcinoma cells showed that 10–20% of the p63-bound sites were associated with changes in neighboring gene expression (Yang et al. 2006).

A special feature of some *in vivo* occupied sites that sets functional sites apart from nonfunctional sites is the presence or absence of additional TF. Evidence for such combinatorial functional interactions of multiple TFs at high occupancy sites was found for the transcriptional repressor REST and five of its corepressors in mouse embryonic stem cells. A heterogeneous pattern of REST

interactions was suggested; weaker binding of REST with low recruitment of corepressors at about half of its non- functional binding sites, and stronger measured interactions of REST with higher, though variable, corepressor presence at almost all functional sites (Yu et al. 2011).

Recently, the intersection of gene expression data with TF binding data from systematic knockdown of 59 TFs in lymphoblastoid cell lines led to a similar conclusion that most TF-DNA interactions are non-functional (Cusanovich et al. 2014). An enrichment of higher binding affinity (inferred from peak heights) and higher number of TF peaks were found at presumably functional TF binding sites (i.e. present near functionally impacted transcription units), indicating the importance of combinatorial interactions for transcriptional regulation. A different result was obtained from genome-wide analysis of MyoD binding in differentiating muscle cells: the inferred binding affinity for MyoD (based on peak heights) at thousands of locations across the genome in myoblasts and myotubes did not discriminate inactive sites from “active sites”, which were defined as those at which MyoD binding is associated with upregulation of nearby gene upon differentiation (Cao et al. 2010).

A major limitation of these studies is that the identification of *cis*-regulatory elements as “functional” is guided solely by the correlation of levels of nearby transcripts and protein occupancy of a DNA element. It cannot be ruled out that some of these binding interactions are a result of, rather than cause of,

transcriptional regulation of the gene of interest. An alternative approach to identification of active vs. inactive regions is CRISPRi, in which targeted genomic loci are silenced in cells by the expression of a nuclease-dead Cas9 protein fused to a KRAB transcriptional repressor domain, guided by sRNA to specific regions of interest (Larson et al. 2013). The interpretation of such experiments is complicated by the possible redundancy in gene regulatory regions, which may lead to false negative results.

5. Evolution of CREs

In addition to changes through time in protein function through mutations in coding sequences, an additional source of evolutionary diversity has been attributed to divergence in regulation of gene expression (King and Wilson 1975). Jacob and Monod have proposed roles of *cis*-regulatory mutations in evolution even before this study (Monod and Jacob 1961). After decades of research on evolutionary gene control, many types of data sets support the idea that changes in regulatory sequences are major source of variation driving evolution, as these elements are less constrained than coding sequences (Wray 2007). As described below, several mechanisms increasing diversity of gene regulation have been discovered (Carroll 2008; Wittkopp and Kalay 2011; Villar et al. 2014).

Substitutions, insertions, and deletions of individual nucleotides may alter *cis*-regulatory activity by loss or gain of protein binding motifs, or their position within an element. For example, the loss of pigmentation in Drosophila species has

been directly attributed to multiple substitutions that inactivate a key enhancer (Jeong et al. 2008). Similarly, multiple single-nucleotide substitutions in one of five enhancers of transcription factor encoding genes controlling Drosophila hair-like trichome morphology change the level and timing of transcription and ultimately generate diverse trichome appearances (Frankel et al. 2011). In addition to mutation of specific binding sites, changes in relative spacing can similarly alter transcriptional read-out. Some male-specific abdominal pigmentation in different Drosophila species varies according to insertions in a key CRE changing binding site spacing (Williams et al. 2008). Such alterations are found in higher eukaryotes as well; sixteen substitutions clustered in a short regulatory element, human-accelerated conserved noncoding sequence 1, are sufficient to lead to gain of function in a developmental enhancer driving human-specific limb expression (Prabhakar et al. 2008).

An additional source of phenotypic novelty is generated by entirely new expression pattern of genes, which may be caused by *de novo* generation CREs from non-regulatory/non-mobile sequences or recruitment of CREs active in other contexts. Conversion of coding sequences to regulatory elements following gene duplication has also been observed (Eichenlaub and Ettwiller 2011). Such larger-scale evolutionary innovations in gene regulatory networks are widespread after individual gene or whole genome duplications. In both bacteria and yeast, systems that lend themselves to comprehensive genomic analysis, only a small fraction of regulatory protein-target gene interactions have evolved by entirely

new interconnection of transcription factors with target genes without homologs. Instead, almost 90% of the interactions have evolved by duplication of either a transcription factor or a target gene, followed by diversification; about one-half of these interactions involved the gain of new interactions (Teichmann and Babu 2004). Although alterations in trans-acting factors would be expected to generate more pleiotropic effects, in bacteria, duplication and changes in trans factors are observed much more frequently than duplication and changes in CREs (Madan Babu and Teichmann 2003). In yeast, analysis of genome sequences, microarray data, and transcriptional regulatory networks showed that expression and regulatory network interactions change very rapidly after gene duplication, much faster than the rate of change of protein sequences (Gu et al. 2005). Therefore, early rapid evolution after gene and genome duplication is crucial for continuously increasing the complexity of the yeast regulatory network.

In addition to the creation of novel regulatory links, evolutionary changes also frequently involve loss of molecular function, such as the loss of CREs. Several examples indicate that this process is widespread in biology. A “weak spot” for regulatory changes was found in environmentally-selected body characteristics of the three-spine stickleback. In this fish, recurrent regulatory mutations in natural populations caused the deletion of a tissue-specific enhancer of a homeobox transcription factor gene. The fish exhibit a loss of a pelvic structure associated with predation (Chan et al. 2010). The emergence of rapid anaerobic growth in *S. cerevisiae* paralleled the uncoupling of expression of cytoplasmic and

mitochondrial ribosomal protein genes, which appears to be caused by loss of a specific regulatory motif from dozens of these genes' promoters (Ihmels et al. 2005). In humans, the specific loss of hundreds of non-coding sequences that are highly conserved in chimpanzees and other mammals indicates that there may be human-specific changes in regulation of genes involved in steroid hormone signaling and neural function (McLean et al. 2011).

Movement and multiplication of transposable elements can extensively reshape the genome, and along with it, the spectrum of CREs from bacteria to animals. Insertion sequence 5 (IS5), the most prevalent IS element in E. coli K12, can activate or inactivate genes to provide potential evolutionary benefits (Zhang and Saier 2011; Saier and Zhang 2014; Wang and Wood 2011). In Drosophila, the Accord LTR retrotransposon insertion brings tissue-specific regulatory sequences, resulting in up-regulation of an insecticide resistance gene (Chung et al. 2007; Schmidt et al. 2010b). Human fetal γ- and adult β-globin genes are regulated by a distally located endogenous retrovirus ERV-9 retrotransposon, whose LTR carries multiple TF binding sites to stimulate RNA polymerase II activity at globin TSS through long-range interactions (Pi et al. 2010).

In addition, retroelements might shape genome organization through expansion of genome. Comparative analysis of six mammals revealed species-specific divergence of motifs for the CTCF boundary element binding protein in rodents, dogs, and opossum caused by activation of retroelements (Schmidt et al. 2012).

Overall, the abundance of these sequences in higher vertebrate genomes is paralleled by their possible roles in acting as CREs. Genome-wide binding analysis showed that a large fraction of *bona fide* binding sites for five of seven mammalian TFs were within distinctive families of transposable elements (Bourque et al. 2008). A larger study of 26 pairs of orthologous TFs in human and mouse cell lines showed that 20% of binding sites were embedded within transposable elements, and a majority of these binding events were cell type-specific, with corresponding cell type-specific histone modifications suggestive of CRE function. In addition, most of the binding sites contributed by transposable elements were species-specific, suggesting that transposable elements are an important driving force for regulatory innovation (Sundaram et al. 2014). Indeed, transposable elements represent about one-half of open chromatin sites as revealed by DNase-seq analysis of normal, embryonic, and cancerous human cells. Especially in embryonic and cancer cells, active cell type-specific sequences are derived from endogenous retrovirus sequences (Jacques et al. 2013). An estimated 20% of gene regulatory sequence in the human genome showing cross-species conservation was co-opted from mobile elements such as SINE, LINE, LTR and DNA transposons, including more than 1,000 primate- and human-accelerated elements (Lindblad-Toh et al. 2011; Lowe and Haussler 2012).

Rewiring of entire gene regulatory networks is a less-common but widespread phenomenon in biology. Here, transposons may again provide the raw material

for such innovations, such as in the evolution of the mammalian placenta. Genomic studies indicate that nearly 1500 genes were recruited into endometrial expression; about 13% of these genes are within 200 kb of a eutherian-specific transposable element. TFs essential for pregnancy bind to these transposons, which carry the epigenetic signatures of CREs (Lynch et al. 2011). Hormone-responsive regulatory elements distributed throughout the mammalian genome are similarly suggested to be carried by transposable elements (Lynch et al. 2015). Transposable elements may contribute to other species-specific developmental networks, as in regulation of genes by OCT4 and NANOG in pluripotent stem cells. Only ~5% of the regions bound by these factors are the same in human and mouse cells; transposable elements contribute up to a quarter of the bound sites, so they provide an important part of novel connections made by these embryonic stem cell determinants (Kunarso et al. 2010).

The global importance of mobile elements on the evolution of gene regulation is still not settled (de Souza et al. 2013). A recent study suggested that ancestral DNA exaptation, not lineage-specific expansions of repeat elements, is the major source for the most of the recently evolved enhancers in liver where active CREs are identified by H3K27Ac and H3K4me3 ChIP-seq in twenty mammalian species (Villar et al. 2015).

6. Dynamics of histone modifications – cause or effect?

During the process of development, there are many genomic regions impacted as cells transition through different chromatin “states”. For instance, as human embryonic stem cells differentiate into neural progenitor cells, ~3400 genes are differentially expressed, concomitant with wide-spread changes in various histone marks at ~10,000-50,000 regions. More abundant changes were observed at enhancer marks H3K27Ac and H3K4me1 compared to the promoter mark H3K4me3 and the repressive mark H3K27me3 (Ziller et al. 2015). Such complex dynamics associated with progressive activation of a GRN pose a real challenge to mechanistic understanding what causes certain changes, and whether specific modifications are functional or are merely “going along for the ride”. Therefore, mutagenic approaches have been used to focus on the contributions of individual chromatin modifying enzymes to gene regulation.

Judging from dramatic experimental interventions, such as the non-lethal effect of deletion of entire histone tails in yeast, it is probable that histone modifications play important modulatory roles, rather than representing the central elements of transcription itself. A central unanswered question is what fraction of modifications to histones and nucleosomes are functionally relevant in any particular context. Frequently, the deletion of genes for global chromatin regulators affects expression of only a limited subset of genes bound to or modified by the regulator in question. Of 165 chromatin machinery components studied by mutation in yeast, 80% were associated with some alteration of gene

expression patterns. However, pleiotropic changes in histone marks were always found to result in very specific, not broad, effects on expression of a subset of genes. For example, removal of regulators of H3K4 (Set1), H3K36 (Set2), and H3K79 (Dot1) methylation causes differential expression of 55, 72 and 2 genes respectively (Lenstra et al. 2011). As with the association of TF with different genomic loci, the disruption of specific interactions may not produce a measureable effect because the element/interaction is truly nonfunctional, or redundant aspects to the control of gene expression masks the perturbation, or our assays lack the sensitivity needed to detect moderate effects in a laboratory setting.

General correlations between active marks and transcription are under debate in higher eukaryotes as well. Genomic binding of the histone acetylase coactivator CBP/p300 is taken as a proxy for active transcriptional enhancers, but the induction of histone acetylation by recruitment of this complex does not always result in gene activation, however (Bedford and Brindle 2012). With respect to transcriptional corepressors, the depletion of the NAD^+ dependent HDAC SIRT1 during the transition from quiescence to proliferation of skeletal muscle stem cells results in increased H4K16 acetylation and the transcriptional activation of muscle-specific genes. However, H4K16ac status at other several thousand loci was also altered without any impact on the expression of nearby genes (Ryall et al. 2015). One possibility is that interactions by these coactivators and corepressors with common regions of chromatin may represent a means by

which dynamic cycles of histone modification induce nucleosome turnover, maintaining accessibility at CREs (Henikoff and Shilatifard 2011).

In analyzing a role for MyoD in reprogramming the myoblast epigenetic state, genome-wide binding of this transcription factor, and identification of its possible functional targets in skeletal muscle cell differentiation were analyzed by measuring MyoD binding, gene expression and epigenome profiles in myoblasts and differentiated cells. From the standpoint of genes showing upregulation, MyoD binding was significantly associated with these loci (63%). However, MyoD was also found to bind to thousands of other sites. Activities of MyoD on these inactive sites at chromatin level were assessed by profiling histone marks H3K4me3 and H4Ac. As for the MyoD-bound sites associated with gene activation, a substantial increase in H4Ac was found the inactive sites, indicating that dynamic chromatin marks may by themselves be insufficient to predict transcriptional function of CREs (Cao et al. 2010).

Another example for lack of association of broad chromatin alterations with genes expression comes from analysis of the core circadian clock in mammals. Oscillatory transcriptional regulation in circadian clock is based on transcriptional activation of genes under control of the CLOCK and BMAL1 activators, which are feedback inhibited by PER and CRY repressors (Kok and Arnosti 2015). Even though associated rhythmic changes of histone marks corresponding to oscillatory expression of thousands of genes through a complex regulatory

network have been described, pervasive circadian rhythms in H3K4me3, H3K9ac, and H3K27ac occur at other thousands of expressed genes whether or not gene expression cycling was detectable (Koike et al. 2012).

Altered transcriptome and chromatin modification profiles are also commonly seen in diseases. Recent genome-wide studies addressed possible correlations between changes in histone modifications with transcriptional abnormalities across the entire transcriptome. Based on small-scale studies, a potential mechanism for changes in transcription has been thought to be alterations in histone modifications, such as decreases in H3K9ac and H3K14ac. Transcriptional dysregulation is a key pathogenic feature in Huntington's disease, and the analysis of H3K9/K14Ac levels and transcriptome patterns in a mouse model of Huntington's disease showed no clear pattern between loss of histone H3K9/K14Ac, predominately located within the coding region, with changes in gene expression (McFarland et al. 2012). Similarly, using a mouse disease for amyloid-forming polygluatamine repeat proteins, hundreds of hypoacetylated loci for H3K9/14 and H4K12 were identified. Most of these loci were not associated with transcriptional dysregulation, except for a subset of genes that showed both reduced transcript levels and deficits in H3K9/14 acetylation at the TSS (Valor et al. 2013). The global lack of correlation between differential histone acetylation and gene expression may indicate that other mechanisms might account for expression changes associated with disease (Valor 2014).

7. Futile cycling - implications for evolution of GRNs

Our analysis of the Hairy transcriptional repressor, the eponymous member of the conserved Hairy Enhancer of Split (HES) transcription factors, has provided important perspective on the topic of functional and non-functional interactions with respect to chromatin dynamics. Our studies in the Drosophila embryo showed that Hairy interacts and changes the chromatin environment dynamically by removing active histone marks such as H4Ac, H3K27Ac and H3K4me1 at many parts of the genome; the expression of some genes is impacted but most are not (Kok et al. 2015). This finding let us to propose the so-called “shotgun model” for this apparent off-target activity of TFs on chromatin modifications; the analogy stems from the experience of waterfowl hunters using a shotgun: many pellets are fired, but few are expected to hit the duck flying overhead. Similarly, a cellular regulatory mechanism may accommodate having many off-target binding events, as long as important sites are regulated by Hairy. Yet the Hairy molecules that don’t “hit the target” – functional targets - still appear to be quite active, biochemically, inducing chromatin modifications that are similar to those seen on transcriptionally controlled loci. Our studies indicate that Hairy may be relatively nonselective about where it can attract chromatin modifying agents across the genome. This indicates that Hairy is not an inert actor on many non-functional sites. As described in Chapter 2, this biochemical activity may provide a unique molecular exaptation to generate novel edges between nodes of a standing GRN on an evolutionary timescale. Since the induced chromatin changes are inconsequential, these effects will not be selected against during

genomic evolution. However, simple acquisition of an activator site may generate an on/off switch, which is a feature of higher eukaryotic enhancers containing combined action of activators and repressors. This feature is likely to be even more pronounced in mammals, since accumulation of these novel TF binding sites over time is much faster in mammals due to their larger genomes and smaller breeding populations (Stefflova et al. 2013). Hairy activity on chromatin may provide a pervasive and accessible entry point for evolution of novel gene regulatory switches. The existence of partially active gene switch events through non-target site binding by Hairy and other transcriptional regulators indicates that there might be a lower than expected threshold for evolution of novel CREs and GRN links, as few new genetic changes would be required to “capture” a Hairy regulated site by activators to make a complete on/off switch. Further studies are required to determine how many other TF may evoke similar chromatin regulating events on inactive elements, and how much selection there is at a biological level against these biochemical processes.

APPENDIX

APPENDIX

Dynamic reprogramming of chromatin: paradigmatic palimpsests and HES factors[2]

Abstract

Temporal and spatial control of transcription in development is dictated to a great extent by transcriptional repressors. Some repressor complexes, such as Polycomp-group proteins, induce relatively long-term non-permissive states, whereas others such as Hairy/Enhancer of Split (HES) family repressors are linked to dynamically modulated chromatin states associated with cycling expression of target genes. The mode of action and specificity of repressors involved in mediating this latter form of epigenetic control are unknown. Oscillating expression of HES repressors controlled by signaling pathways such as Notch suggests that the entire ensemble of HES–associated co-repressors and histone modifying complexes readily cycle on and off genes. Dynamic interactions between these factors and chromatin seem to be crucial in maintaining multipotency of progenitor cells, but the significance of such interactions in more differentiated cells is less well understood. We discuss here how genome-wide analyses and real-time gene expression measurements of HES regulated genes can help decipher the detailed mechanisms and biological importance of highly dynamic transcriptional switching mediated by epigenetic changes.

Keywords: HES, Hairy, oscillatory gene expression, repression, chromatin

[2]This works was published as the following manuscript: Kok K, Arnosti DN. 2015. Dynamic reprogramming of chromatin: paradigmatic palimpsests and HES factors. *Front Genet* **6**: 29.

1. Introduction

Dynamic cellular processes in biological systems require modulated and adaptable responses at the level of gene expression. Variations in the internal and external environment provoke short- and long-term changes in gene expression, which help maintain cellular physiology; these controls are also a fundamental point of evolutionary changes (López-Maury et al., 2008). Some variability in output of gene regulatory networks (GRN) is an inescapable consequence of molecular noise, including stochastic switching of promoter activity or “bursts”. Such random fluctuations can be easily propagated to downstream genes or buffered out, and may play significant roles in physiological regulation, differentiation, adaptation and evolution (Eldar and Elowitz, 2010). In addition to the impact of stochastic molecular processes on gene expression, organisms from bacteria to animals have evolved a wide variety of specialized oscillatory gene expression mechanisms to respond to predictable and unpredictable environmental fluctuations and effect developmental programs (Paszek et al., 2010; Young and Kay, 2001). The levels of mechanistic complexity vary among oscillatory systems, but they share common regulatory principles, including negative feedback loops (Figure 1-A-1A). These core features were successfully used to design simple synthetic oscillatory networks that accurately predict the dynamic behavior of biological systems, which are generally more complex and feature robustness to genetic and environmental influences (Elowitz and Leibler, 2000; Tigges et al., 2009; Cookson et al., 2009).

2. Design and function of oscillating gene networks

A classic example of oscillatory transcriptional regulation is the ability of the circadian clock to adjust output of many genes in preparation for predictable daily changes in light, food, and temperature (Bell-Pedersen et al., 2005). Although regulation is highly complex, the core of the vertebrate molecular clock is based on transcriptional activation of genes under control of the CLOCK and BMAL1 activators. These factors drive expression of many genes during the day, including the PER and CRY repressors, which feedback inhibit and block CLOCK/BMAL1 action during the nighttime (Figure 1-A-1B) (Ko and Takahashi, 2006; Baggs and Hogenesch, 2010). Repression is relieved by phosphorylation, ubiquitination, and degradation of PER and CRY, leading to a feedback loop with a period of ~24 hours (Busino et al., 2007).

Genome-wide studies have revealed associated rhythmic changes of histone marks corresponding to oscillatory expression of thousands of genes coordinating biological cycles through a complex regulatory network (Koike et al., 2012; Feng et al., 2011). A recent study from the Takahashi laboratory provided a comprehensive overview of chromatin-associated dynamics of circadian cycling in the murine liver. Using time-dependent ChIP-seq analysis of transcription factors (BMAL1, CLOCK, NPAS2, PER1, PER2, CRY1, CRY2, p300 and CBP), RNA Pol II, and histone marks (H3K9Ac, H3K27Ac, H3K4me1, H3K4me3, H3K36me3 and H3K79me2), the authors identified three phases in the circadian

clock corresponding to genes in a transcriptionally poised, activated, and repressed states (Koike et al., 2012).

In addition to predictable daily cycles, cells need to respond to rapid changes and variations during development and growth. Ultradian oscillations often feature a time period of minutes to hours and are triggered by intrinsic and environmental signals. One of the best described such instances is represented by the p53 pathway; this transcription factor can display dynamic behavior in response to DNA damage and other cellular stress to protect cells against malignant transformation (Batchelor et al., 2011). p53 expression is regulated by a negative feedback loop. The MDM2 regulator normally keeps p53 activity at low levels by binding to the factor's DNA binding domain, inducing a change in subcellular localization from the nucleus to the cytoplasm, and inducing ubiquitylation for eventual degradation of p53 (Wu et al., 1993; Haupt et al., 1997). After DNA damage, the p53 protein is phosphorylated, preventing the interaction of MDM2 with p53 and resulting in activation of p53 (Kruse and Gu, 2009). p53 transcriptionally activates expression of many genes including *MDM2*, resulting in a time-delay feedback inhibition that can exhibit oscillations of both p53 and Mdm2 (Figure 1-A-1C) (Lahav et al., 2004; Lev Bar-Or et al., 2000; Bose and Ghosh, 2007). Depending on the dynamic control of p53, different cellular responses can be elicited. Cells can undergo a transient cell cycle arrest and recover from the DNA damage (Purvis et al., 2012). In addition to transient responses, the p53 pathway also triggers terminal fates such as apoptosis and

senescence. In contrast to oscillatory output, sustained p53 expression affects the expression of a different set of genes, leading to senescence (Purvis et al., 2012). Therefore, depending on the dynamics of the input, distinct chromatin and regulatory changes can be imparted on a gene network to transmit information and alter cellular fate.

Oscillations are also seen in differentiation and embryonic development. One of the best-studied examples involves the transcriptional repressor Hes1 that controls the differentiation of neurons and formation of somite segments in the vertebrate hindbrain (Figure 1-A-1D) (Koike et al., 2012; Kageyama et al., 2007). Hes1 belongs to the conserved family of Hairy Enhancer of Split (HES) transcriptional repressors that recruit common co-repressors of the Groucho/TLE family (Aloia et al., 2013; Davis and Turner, 2001). The eponymous Drosophila Hairy repressor functions as a so-called long-range repressor that remodels large blocks of chromatin upon transcriptional repression. Hairy mediates wide-spread and coupled loss of active histone marks H4Ac, H3K27Ac, H3K4me1 and H3K4me3 on many embryonic genes (Li and Arnosti, 2011; Kok et al., submitted). Furthermore, Hairy represses its own transcription by removing these active marks, consistent with the previously observed autoregulatory mechanism of related mammalian HES proteins (Kageyama et al., 2007).

A conserved feature of regulatory pathways involving HES proteins is the role of Notch signaling. Upon ligand binding, Notch is cleaved and released from the

plasma membrane to translocate to the nucleus, where it associates with and activates the *Hes1* promoter. Hes1 protein negatively regulates its own promoter, establishing a feedback loop (Fischer and Gessler, 2007). This feedback loop can induce oscillations in Hes1 protein levels (Kageyama et al., 2007). Periodic temporal expression of Hes1 plays a crucial role in formation of somites, which give rise to the vertebrae, ribs, skeletal muscles and dermis (Aulehla and Herrmann, 2004). These segments are formed from the anterior region of the presomitic mesoderm (PSM) by periodic Notch signals. Notch coordinates Hes1 oscillations, which progress from the posterior to anterior region of the PSM. One wave of expression of this so-called segmentation clock lasts two hours, marking the boundary for a new somite that forms at the end of the embryo (Pourquié, 2003). In this setting, temporal oscillations are converted into a spatial pattern of somite boundaries. A large number of genes involved in cell signaling are periodically expressed during this segmentation process in mouse (Dequéant et al., 2006). Comparison of the mouse, chicken and zebrafish PSM oscillatory transcriptomes revealed networks of 40 to 100 conserved cycling genes that are activated downstream of the Notch, Fibroblast Growth Factor and Wnt pathways (Krol et al., 2011). Thus, the segmentation clock is controlled by conserved multiple signaling pathways. The common oscillatory genes in all vertebrates include at least one member of the Hes/Her family. However, the identity of cyclic genes varies from species to species as well, indicating evolutionary plasticity of the segmentation networks (Krol et al., 2011).

In contrast to the fate-determining effects of Hes1 oscillations in the PSM, cyclic behavior of Hes1 in neuronal progenitor cells (NPC) is associated with stabilization of the undifferentiated phenotype. In these cells, Hes1 mRNA, protein, and activity oscillate with a two hour period (Hirata et al., 2002). Hes1 represses transcription of proneural transcription factors such as *Ascl1*, inducing oscillations in levels of that factor. Interestingly, self-renewal of NPCs and their eventual proper differentiation is achieved only when Hes1 and downstream genes are periodically expressed (Imayoshi and Kageyama, 2014). Sustained expression of Hes1 constitutively in NPCs represses proneural genes, blocking proliferation and inducing quiescence (Baek et al., 2006). This observation indicates that active division of NPCs is dependent on the oscillatory expression of fate determination factors. Neuronal fate choice is determined by sustained expression of Ascl1 after cell division. During differentiation, Hes1 oscillations cease as Notch inputs diminish, leading to upregulation of Ascl1 (Imayoshi et al., 2013). Using a light-activatable system, the impact of oscillating and sustained expression of Ascl1 on proliferation and differentiation of NPCs was tested. A three hour periodic expression of Ascl1 supported proliferation of NPCs, whereas sustained expression resulted in differentiation (Imayoshi et al., 2013). Similar roles for Hes1 oscillation has been observed in embryonic stem cells (Kobayashi et al., 2009).

The types of chromatin dynamics occurring on genes entrained under the circadian clock system have not been well documented for oscillations involving

ultradian factors such as HES proteins and other bHLH transcription factors. However, a recent study suggested that the Ascl1 bHLH factor, which shows oscillatory expression complementary to that of Hes1 in neuronal progenitors, is critical for formation of open chromatin during reprogramming through its activities as a pioneer factor on enhancers (Wapinski et al., 2013). Less is known about the chromatin modifying properties of Hes1 itself, however, the homologous Drosophila protein Hairy has a direct role in chromatin modification, and this protein impacts the chromatin state of hundreds of loci on a genome-wide scale (Li and Arnosti, 2011; Kok et al., submitted). As HES transcription factors share common structural features, including DNA binding and effector domains, as well as conserved developmental roles, the biochemical properties are likely to be similar.

How general are the dynamic chromatin responses associated with activation and repression of genes such as those targeted by HES factors? The time-delays associated with activating or repressing promoters are a function of dynamics of protein complexes. Even in steady-state situations, transcription factors are observed to continuously associate and dissociate with target loci, a feature not revealed by ChIP experiments but that is demonstrated by direct imaging as well as in vitro approaches (Voss and Hager, 2014). However, as observed for the prolactin promoter, stochastic chromatin processes can render promoters refractory to stimulation. Such refractory periods would block transmission of dynamic signals (Harper et al., 2011). Indeed, high-resolution

temporal measurement of mRNA of many mammalian genes from single cells reveals that distinct regulatory regions confer gene-specific switching rates with different refractory periods (Suter et al., 2011). Such differences may cause differential oscillation of genes in response to stimuli. Fine time-scale analysis of global gene expression triggered by the inflammatory cytokine TNF showed oscillations in >5000 genes that are involved in multiple pathways, with different genes oscillating either very rapidly or after a lag phase (Sun et al., 2008). Cyclic interaction of transcription factors with promoters can extend from seconds for bursting promoters to minutes for developmental oscillators to hours for circadian clocks. A single promoter may experience both fast (2 min) and slow (40 min) periodic binding of a single transcription factor, as with Ace1 occupancy of the yeast *CUP1* promoter (Karpova et al., 2008). The authors suggest that fast cycling is responsible for the initial period of gene expression, while slow cycling represents the fine-tuning of expression levels associated with slow-period oscillating nucleosome occupancy. A short-period ultradian cycling has also been described for the estrogen receptor, involving periodic binding and assembly of chromatin complexes in mammalian cells, however recent high-resolution studies of RNA polymerase activity have not supported this picture (Hah et al., 2011; Voss and Hager, 2014).

In development, oscillatory circuits affect not only specific networks of genes relating to patterning, as described for Hes1, but also can include many synchronized genes not linked to circadian control. Large-scale transcriptome

analysis in C. elegans larvae revealed robust ~8 hour cycling of thousands of genes, which may be related to developmental processes such as molting (Hendriks et al., 2014). In contrast to the simple synthetic biology circuits tested in bacteria, such large-scale oscillatory behavior likely involves more components than a single negative feedback loop (Sun et al. 2008). The coordinated expression of many genes in these systems indicates that persistent chromatin changes are not likely to prevent genome-wide oscillatory coordination, thus the dynamic chromatin changes found for HES factors are likely to be representative of many regulatory mechanisms.

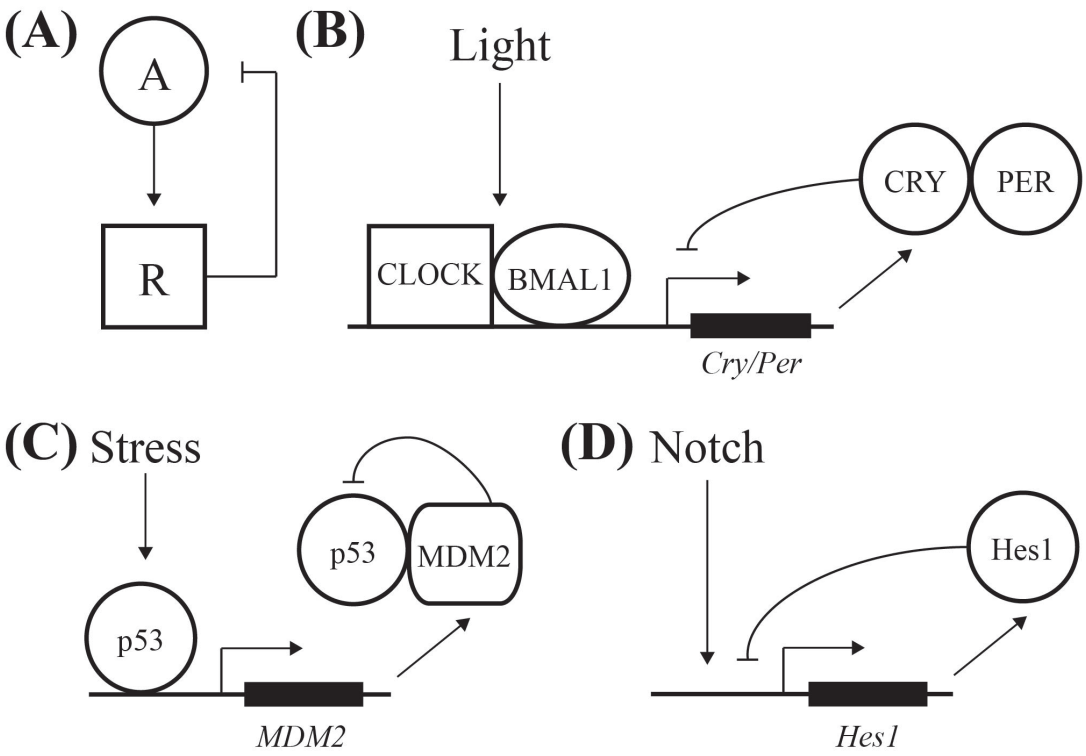


Figure 1-A-1 Negative feedback loops at the core of transcriptional oscillators (A) Diagram of a simple negative feedback loop for oscillatory behavior. An activator "A" increases activity of a repressor "R", which in turn decreases activity of the activator. **(B)** Major factors driving daily oscillations of the circadian clock, whereby CLOCK/BMAL1 drive expression of the inhibitory factors CRY/PER. **(C)** Stress and DNA damage activation of the p53 pathway, whereby 5-9 hr. ultradian oscillations in p53 activity drive expression of p53 inhibitor MDM2. **(D)** *Hes1* expression is driven by Notch signaling and feedback inhibited by Hes1, with an oscillation of ~2-3 h.

3. Oscillatory behavior and chromatin dynamics

The biochemical mechanisms by which transcriptional oscillations can be induced are in many cases better understood than the physiological significance of such dynamics. In the case of circadian regulated genes, adaptation to predictable environmental changes, such as food availability, temperature or light, is a clear driver of such dynamics. In development, the dynamic readout of HES activity represents a morphological pattern generator. In other cases, it is not clear whether the cycling is a necessary feature of the system, or tolerated as an also-acceptable form of control that may or may not have superior regulatory properties. Arguing against a view that cycling occurs by chance is the likelihood that randomly propagated oscillations though a multi-level network should eventually cancel out, thus it is likely that there is selection for coordinated responses at some level. Depending on the nature of downstream targets, cycles of transcriptional output may be “integrated” to a steady-state approximation of the average level of signaling, or it may be “propagated”, if dynamics of the downstream gene expression is as fast as the cycling signal (Hoffmann, 2002) (Figure 1-A-2A).

Oscillatory behavior may be eventually damped by several layers of a gene regulatory cascade. For example, in the case of cyclical expression of Hes1, expression of several downstream targets also alternates, but the overall undifferentiated state of the cell - represented by the global activity or inactivity of many genes - stays constant, indicating that at least at a larger scale, such

oscillatory behavior is subsumed into a stable phenotype. Alternatively, the oscillatory action at one level of a gene regulatory network may better ensure that a particular level of expression within a critical range is maintained, rather like a singer who uses vibrato to hold a particularly difficult note (Imayoshi et al., 2013). At the same time, the interlocking feedback loops that permit oscillation also provide the control points that can be shifted to move a cell into a different gene regulatory, and eventually differentiated state. These arguments are attractive in pointing out possible adaptive features of oscillatory regulation, however, testing the null hypothesis is difficult. It may be that just as transcriptional “bursting” is an inevitable consequence of micro-scale chromatin movements, longer period, regular transcriptional oscillations may be system properties that arise as a secondary consequence of core properties of the system, such as robustness. Alternatively, or in addition, many oscillations that are observed are consequences of a few key dynamic drivers that must show periodic changes; the ancillary downstream changes may not important for natural selection acting on gene expression (Paszek et al., 2010; Cheong and Levchenko, 2010).

What is known about the required chromatin dynamics that are associated with oscillatory gene regulation? Circadian regulated genes exhibit cyclical chromatin responses that reset every day (Koike et al., 2012). In the developmental settings for Hairy and HES protein activity, the targets of these proteins are often active only transiently, implying very dynamic chromatin responses. For instance, the

activators of *ftz*, a gene that is repressed by Hairy, are present on the genome for only minutes during early embryogenesis, and repressive countermeasures would be required only for a similarly brief time. Indeed, we find that in cases of artificial induction of Hairy, dramatic chromatin deacetylations are quickly reversed as soon as Hairy levels drop, indicating that the repressor is working against a background of cellular chromatin modifying activities that quickly restore a landscape to the status quo ante (K. Kok, unpublished observations). Hes1 action, although not studied at the chromatin level, must similarly be transient in terms of perdurance, as downstream transcriptional targets quickly follow changes in the levels of Hes1 over a period of hours. Thus, in general, HES protein directed alterations to genome-wide chromatin states may be very transient (Figure 1-A-2B). In some regulatory circuits, we do know that chromatin states are locked in, preserving a particular epigenetic mark through multiple mitoses – these markers involve Polycomb complexes in Drosophila and higher metazoans, as well as DNA methylation signals in vertebrates. Significantly, both of these systems can be deployed in alternate modes, so that in some instances DNA methylation and Polycomb-regulated effects are transient (Aloia et al., 2013). Are global chromatin modifications just reflections of gene regulatory effects rather than drivers of the system? To what extent are these chromatin changes important for setting the boundary conditions for oscillatory gene responses? Systems and synthetic biology approaches will converge with developmental gene regulation to deliver answers to these intriguing questions.

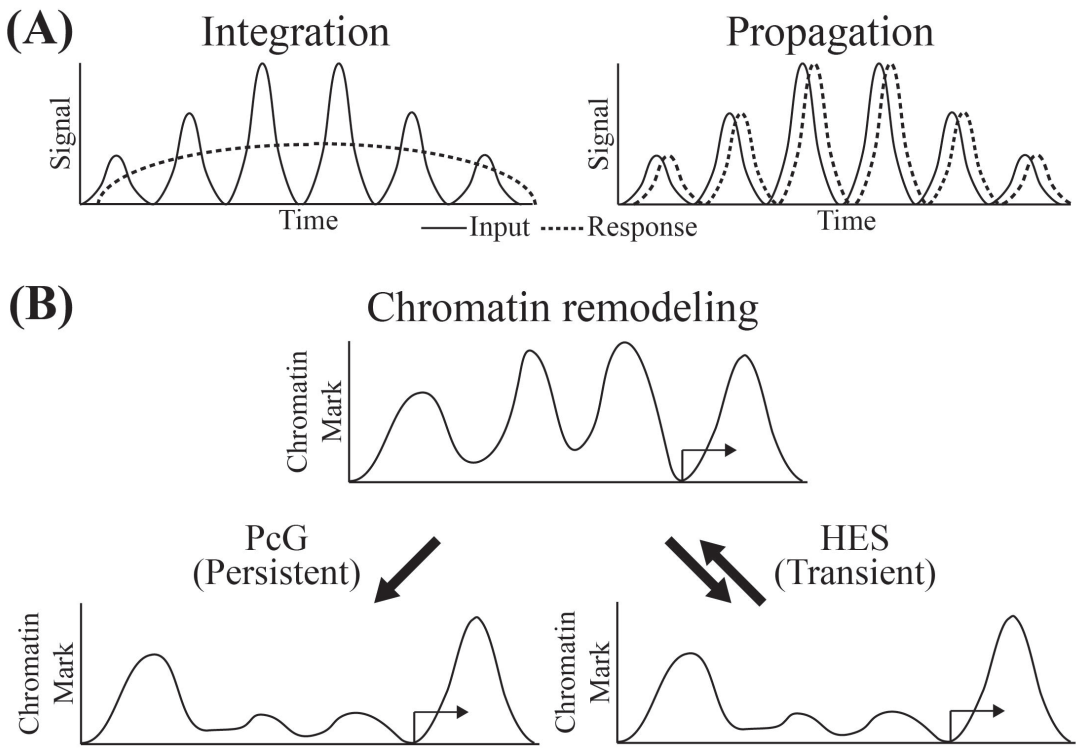


Figure 1-A-2 Oscillator inputs and possible outcomes (A) The oscillating input signal can be integrated or propagated to generate a sustained or dynamic response, respectively. Time on the x-axis might be minutes to hours. **(B)** Outcome at the chromatin level might be persistent and long-term in integrated response by PcG regulation or transient and reversible in propagated response by HES regulation.

4. Acknowledgments

This work was supported by NIH GM056976 to D.N.A. and a fellowship from the MSU Genetics Program to K.K.

REFERENCES

REFERENCES

Aloia, L., Di Stefano, B., and Di Croce, L. (2013). Polycomb complexes in stem cells and embryonic development. *Development* 140, 2525–2534. doi:10.1242/dev.091553.

Albert FW, Kruglyak L. 2015. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* **16**: 197–212.

Aulehla, A., and Herrmann, B. G. (2004). Segmentation in vertebrates: clock and gradient finally joined. *Genes & Development* 18, 2060–2067. doi:10.1101/gad.1217404.

Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. 2013. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science* **339**: 1074–1077.

Arnosti DN, Kulkarni MM. 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* **94**: 890–898.

Baek, J. H., Hatakeyama, J., Sakamoto, S., Ohtsuka, T., and Kageyama, R. (2006). Persistent and high levels of Hes1 expression regulate boundary formation in the developing central nervous system. *Development* 133, 2467–2476. doi:10.1242/dev.02403.

Baggs, J. E., and Hogenesch, J. B. (2010). Genomics and systems approaches in the mammalian circadian clock. *Current Opinion in Genetics & Development* 20, 581–587. doi:10.1016/j.gde.2010.08.009.

Bao X, Rubin AJ, Qu K, Zhang J, Giresi PG, Chang HY, Khavari PA. 2015. art%3A10.1186%2Fs13059-015-0840-9. *Genome Biol* 1–17.

Batchelor, E., Loewer, A., Mock, C., and Lahav, G. (2011). Stimulus-dependent dynamics of p53 in single cells. *Molecular Systems Biology* 7, 488. doi:10.1038/msb.2011.20.

Bedford DC, Brindle PK. 2012. Is histone acetylation the most important physiological function for CBP and p300? *Aging (Albany NY)* **4**: 247–255.

Bell-Pedersen, D., Cassone, V. M., Earnest, D. J., Golden, S. S., Hardin, P. E., Thomas, T. L., and Zoran, M. J. (2005). Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat. Rev. Genet.* 6, 544–556. doi:10.1038/nrg1633.

Biggin MD. 2011. Animal transcription networks as highly connected, quantitative

continua. *Developmental Cell* **21**: 611–626.

Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczynski B, Riddell A, Furlong EEM. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* **44**: 148–156.

Bose, I., and Ghosh, B. (2007). The p53-MDM2 network: from oscillations to apoptosis. *J. Biosci.* 32, 991–997.

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew J-L, Ruan Y, Wei C-L, Ng H-H, et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research* **18**: 1752–1762.

Busino, L., Bassermann, F., Maiolica, A., Lee, C., Nolan, P. M., Godinho, S. I. H., Draetta, G. F., and Pagano, M. (2007). SCFFbxl3 controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins. *Science* 316, 900–904. doi:10.1126/science.1141194.

Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL, et al. 2010. Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Developmental Cell* **18**: 662–674.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**: 25–36.

Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**: 302–305.

Chen C-K, Symmons O, Uslu VV, Tsujimura T, Ruf S, Smedley D, Spitz F. 2013. TRACER: a resource to study the regulatory architecture of the mouse genome. *BMC Genomics* **14**: 215.

Cheong, R., and Levchenko, A. (2010). Oscillatory signaling processes: the how, the why and the where. *Current Opinion in Genetics & Development* 20, 665–669. doi:10.1016/j.gde.2010.08.007.

Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, Daborn PJ. 2007. Cis-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the Drosophila melanogaster insecticide resistance gene Cyp6g1. *Genetics* **175**: 1071–1077.

Cookson, N. A., Tsimring, L. S., and Hasty, J. (2009). The pedestrian

watchmaker: genetic clocks from engineered oscillators. *FEBS Letters* 583, 3931–3937. doi:10.1016/j.febslet.2009.10.089.

Cusanovich DA, Pavlovic B, Pritchard JK, Gilad Y. 2014. The Functional Consequences of Variation in Transcription Factor Binding ed. Y. Pilpel. *PLoS Genet* **10**: e1004226.

Davis, R. L., and Turner, D. L. (2001). Vertebrate hairy and Enhancer of split related proteins: transcriptional repressors regulating cellular differentiation and embryonic patterning. *Oncogene* 20, 8342–8357. doi:10.1038/sj.onc.1205094.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei C-L, Natoli G. 2010. A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers ed. J.S. Mattick. *Plos Biol* **8**: e1000384.

de Souza FSJ, Franchini LF, Rubinstein M. 2013. Exaptation of Transposable Elements into Novel Cis-Regulatory Elements: Is the Evidence Always Strong? *Molecular Biology and Evolution* **30**: 1239–1251.

Degner JF, Pai AA, Pique-Regi R, Veyrieras J-B, Gaffney DJ, Pickrell JK, De Leon S, Michelini K, Lewellen N, Crawford GE, et al. 2012. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**: 390–394.

Dequéant, M.-L., Glynn, E., Gaudenz, K., Wahl, M., Chen, J., Mushegian, A., and Pourquié, O. (2006). A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science* 314, 1595–1598. doi:10.1126/science.1133141.

Dobi KC, Winston F. 2007. Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **27**: 5575–5586.

Eichenlaub MP, Ettwiller L. 2011. De novo genesis of enhancers in vertebrates. *Plos Biol* **9**: e1001188.

Eldar, A., and Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature* 467, 167–173. doi:10.1038/nature09326.

Elowitz, M. B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. doi:10.1038/35002125.

Engström PG, Ho Sui SJ, Drivenes O, Becker TS, Lenhard B. 2007. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Research* **17**: 1898–1908.

Feng, D., Liu, T., Sun, Z., Bugge, A., Mullican, S. E., Alenghat, T., Liu, X. S., and

Lazar, M. A. (2011). A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science* 331, 1315–1319. doi:10.1126/science.1198125.

Fischer, A., and Gessler, M. (2007). Delta Notch and then? Protein interactions and proposed modes of repression by Hes and Hey bHLH factors. *Nucleic Acids Research* 35, 4583–4596. doi:10.1093/nar/gkm477.

Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmann R, MacArthur S, Thomas S, Stamatoyannopoulos JA, Eisen MB, et al. 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. *Proceedings of the National Academy of Sciences* **109**: 21330–21335.

Frankel N, Erezyilmaz DF, McGregor AP, Wang S, Payre F, Stern DL. 2011. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* **474**: 598–603.

Frazer KA, Tao H, Osoegawa K, de Jong PJ, Chen X, Doherty MF, Cox DR. 2004. Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional. *Genome Research* **14**: 367–372.

Fuda NJ, Ardehali MB, Lis JT. 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**: 186–192.

Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EEM. 2014. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**: 96–100.

Gitter A, Siegfried Z, Klutstein M, Fornes O, Oliva B, Simon I, Bar-Joseph Z. 2009. Backup in gene regulatory networks explains differences between binding and knockout results. *Molecular Systems Biology* **5**: 276.

Gu W, Lee H-C, Chaves D, Youngman EM, Pazour GJ, Conte D Jr, Mello CC. 2012. CapSeq and CIP-TAP Identify Pol II Start Sites and Reveal Capped Small RNAs as C. elegans piRNA Precursors. *Cell* **151**: 1488–1500.

Gu X, Zhang Z, Huang W. 2005. Rapid evolution of expression and regulatory divergences after yeast gene duplication. *Proc Natl Acad Sci USA* **102**: 707–712.

Guertin MJ, Martins AL, Siepel A, Lis JT. 2012. Accurate prediction of inducible transcription factor binding intensities in vivo. *PLoS Genet* **8**: e1002610.

Hager GL, McNally JG, Misteli T. 2009. Transcription dynamics. *Molecular Cell* **35**: 741–753.

Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL. 2011. A

rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145**: 622–634.

Harafuji N, Keys DN, Levine M. 2002. Genome-wide identification of tissue-specific enhancers in the Ciona tadpole. *Proc Natl Acad Sci USA* **99**: 6802–6805.

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne J-B, Reynolds DB, Yoo J, et al. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**: 99–104.

Hardison RC, Taylor J. 2012. Genomic approaches towards finding cis-regulatory modules in animals. *Nat Rev Genet* **13**: 469–483.

Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB. 2008. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. *PLoS Genet* **4**: e1000106.

Harper, C. V., Finkenstädt, B., Woodcock, D. J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D. G., Mullins, J. J., Rand, D. A., Davis, J. R. E., et al. (2011). Dynamic analysis of stochastic transcription cycles. *Plos Biol* 9, e1000607. doi:10.1371/journal.pbio.1000607.

Haupt, Y., Maya, R., Kazaz, A., and Oren, M. (1997). Mdm2 promotes the rapid degradation of p53. *Nature* 387, 296–299. doi:10.1038/387296a0.

Hawkins RD, Hon GC, Ren B. 2010. Next-generation genomics: an integrative approach. *Nat Rev Genet* 1–11.

He Q, Bardet AF, Patton B, Purvis J, Johnston J, Paulson A, Gogol M, Stark A, Zeitlinger J. 2011. High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. *Nat Genet* **43**: 414–420.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–318.

Hendriks, G.-J., Gaidatzis, D., Aeschimann, F., and Großhans, H. (2014). Extensive oscillatory gene expression during C. elegans larval development. *Molecular Cell* 53, 380–392. doi:10.1016/j.molcel.2013.12.013.

Henikoff S, Shilatifard A. 2011. Histone modification: cause or cog? *Trends in Genetics* **27**: 389–396.

Hirata, H., Yoshiura, S., Ohtsuka, T., Bessho, Y., Harada, T., Yoshikawa, K., and Kageyama, R. (2002). Oscillatory expression of the bHLH factor Hes1

regulated by a negative feedback loop. *Science* 298, 840–843. doi:10.1126/science.1074560.

Hoffmann, A. (2002). The Ikappa B-NF-kappa B Signaling Module: Temporal Control and Selective Gene Activation. *Science* 298, 1241–1245. doi:10.1126/science.1071914.

Hu Z, Killion PJ, Iyer VR. 2007. Genetic reconstruction of a functional transcriptional regulatory network. *Nat Genet* **39**: 683–687.

Ihmels J, Bergmann S, Gerami-Nejad M, Yanai I, McClellan M, Berman J, Barkai N. 2005. Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science* **309**: 938–940.

Imayoshi, I., and Kageyama, R. (2014). bHLH factors in self-renewal, multipotency, and fate choice of neural progenitor cells. *Neuron* 82, 9–23. doi:10.1016/j.neuron.2014.03.018.

Imayoshi, I., Isomura, A., Harima, Y., Kawaguchi, K., Kori, H., Miyachi, H., Fujiwara, T., Ishidate, F., and Kageyama, R. (2013). Oscillatory control of factors determining multipotency and fate in mouse neural progenitors. *Science* 342, 1203–1208. doi:10.1126/science.1242366.

Jacques P-É, Jeyakani J, Bourque G. 2013. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* **9**: e1003504.

Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. 2008. The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. *Cell* **132**: 783–793.

Jiang C, Pugh BF. 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* **10**: 161–172.

Johnson DS, Zhou Q, Yagi K, Satoh N, Wong W, Sidow A. 2005. De novo discovery of a tissue-specific gene regulatory module in a chordate. *Genome Research* **15**: 1315–1324.

Juven-Gershon T, Hsu J-Y, Theisen JW, Kadonaga JT. 2008. The RNA polymerase II core promoter - the gateway to transcription. *Current Opinion in Cell Biology* **20**: 253–259.

Kadonaga JT. 2012. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol* **1**: 40–51.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**:

430–435.

Kageyama, R., Ohtsuka, T., and Kobayashi, T. (2007). The Hes gene family: repressors and oscillators that orchestrate embryogenesis. *Development* 134, 1243–1251. doi:10.1242/dev.000786.

Karpova, T. S., Kim, M. J., Spriet, C., Nalley, K., Stasevich, T. J., Kherrouche, Z., Heliot, L., and McNally, J. G. (2008). Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science* 319, 466–469. doi:10.1126/science.1150559.

Kazemian M, Suryamohan K, Chen J-Y, Zhang Y, Samee MAH, Halfon MS, Sinha S. 2014. Evidence for deep regulatory similarities in early developmental programs across highly diverged insects. *Genome Biol Evol* **6**: 2301–2320.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116.

Ko, C. H., and Takahashi, J. S. (2006). Molecular components of the mammalian circadian clock. *Human Molecular Genetics* 15 Spec No 2, R271–7. doi:10.1093/hmg/ddl207.

Kobayashi, T., Mizuno, H., Imayoshi, I., Furusawa, C., Shirahige, K., and Kageyama, R. (2009). The cyclic gene Hes1 contributes to diverse differentiation responses of embryonic stem cells. *Genes Dev.* 23, 1870–1875. doi:10.1101/gad.1823109.

Koike N, Yoo S-H, Huang H-C, Kumar V, Lee C, Kim T-K, Takahashi JS. 2012. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* **338**: 349–354.

Kok K, Arnosti DN. 2015. Dynamic reprogramming of chromatin: paradigmatic palimpsests and HES factors. *Front Genet* **6**: 29.

Kok K, Ay A, Li LM, Arnosti DN. 2015. Genome-wide errant targeting by Hairy. *eLife* **4**.

Krol, A. J., Roellig, D., Dequéant, M.-L., Tassy, O., Glynn, E., Hattem, G., Mushegian, A., Oates, A. C., and Pourquié, O. (2011). Evolutionary plasticity of segmentation clock networks. *Development* 138, 2783–2792. doi:10.1242/dev.063834.

Kruse, J.-P., and Gu, W. (2009). Modes of p53 regulation. *Cell* 137, 609–622. doi:10.1016/j.cell.2009.04.050.

Kunarso G, Chia N-Y, Jeyakani J, Hwang C, Lu X, Chan Y-S, Ng H-H, Bourque G. 2010. Transposable elements have rewired the core regulatory network of

human embryonic stem cells. *Nat Genet* **42**: 631–634.

Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, Dickson BJ, Stark A. 2014. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**: 91–95.

Kwak H, Fuda NJ, Core LJ, Lis JT. 2013. Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science* **339**: 950–953.

Kwon Y-S, Garcia-Bassets I, Hutt KR, Cheng CS, Jin M, Liu D, Benner C, Wang D, Ye Z, Bibikova M, et al. 2007. Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters. *Proc Natl Acad Sci USA* **104**: 4852–4857.

Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A. J., Elowitz, M. B., and Alon, U. (2004). Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat Genet* 36, 147–150. doi:10.1038/ng1293.

Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. 2013. CRISPR_interference_CRISPRi_for_sequence-specific_. *Nature Protocols* **8**: 2180–2196.

Lee TI, Young RA. 2013. Transcriptional regulation and its misregulation in disease. *Cell* **152**: 1237–1251.

Lenhard B, Sandelin A, Carninci P. 2012. Regulatory elements: Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Publishing Group* **13**: 233–245.

Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NACH, Margaritis T, van de Pasch LAL, van Heesch SAAC, Brok MO, Groot Koerkamp MJA, et al. 2011. The specificity and topology of chromatin interaction pathways in yeast. *Molecular Cell* **42**: 536–549.

Lev Bar-Or, R., Maya, R., Segel, L. A., Alon, U., Levine, A. J., and Oren, M. (2000). Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11250–11255. doi:10.1073/pnas.210171597.

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, et al. 2012. Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell* **148**: 84–98.

Li, L. M., and Arnosti, D. N. (2011). Long- and Short-Range Transcriptional Repressors Induce Distinct Chromatin States on Repressed Genes. *Curr Biol* 21, 406–412. doi:10.1016/j.cub.2011.01.054.

Li X-Y, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Hendriks CLL, et al. 2008. Transcription Factors Bind Thousands of Active and Inactive Regions in the Drosophila Blastoderm. *Plos Biol* **6**: e27.

Lickwar CR, Mueller F, Hanlon SE, McNally JG, Lieb JD. 2012. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**: 251–255.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289–293.

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**: 476–482.

López-Maury, L., Marguerat, S., and Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet.* 9, 583–593. doi:10.1038/nrg2398.

Lowe CB, Haussler D. 2012. 29 mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome. *PLoS ONE* **7**: e43128.

Lynch VJ, Leclerc RD, May G, Wagner GP. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* **43**: 1154–1159.

Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza SL, Mazur EC, Emera D, Sheikh SZ, Grützner F, Bauersachs S, et al. 2015. Ancient Transposable Elements Transformed the Uterine Regulatory Landscape and Transcriptome during the Evolution of Mammalian Pregnancy. *Cell Reports*.

MacArthur S, Li X-Y, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keränen SVE, et al. 2009. Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol* **10**: R80.

Madan Babu M, Teichmann SA. 2003. Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic Acids Research* **31**: 1234–1244.

Maston GA, Evans SK, Green MR. 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genom Human Genet* **7**: 29–59.

McFarland KN, Das S, Sun TT, Leyfer D, Xia E, Sangrey GR, Kuhn A, Luthi-Carter R, Clark TW, Sadri-Vakili G, et al. 2012. Genome-wide histone acetylation is altered in a transgenic mouse model of Huntington's disease. *PLoS ONE* **7**: e41423.

McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT, et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**: 216–219.

Monod J, Jacob F. 1961. General conclusions — teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol* **26**: 389–401.

Narasimhan K, Pillay S, Huang YH, Jayabal S, Udayasuryan B, Veerapandian V, Kolatkar P, Cojocaru V, Pervushin K, Jauch R. 2015. DNA-mediated cooperativity facilitates the co-selection of cryptic enhancer sequences by SOX2 and PAX6 transcription factors. *Nucleic Acids Research* **43**: 1513–1528.

Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, et al. 2013. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**: 1521–1531.

Ohler U. 2006. Identification of core promoter modules in Drosophila and their application in accurate transcription start site prediction. *Nucleic Acids Research* **34**: 5943–5950.

Paris M, Kaplan T, Li X-Y, Villalta JE, Lott SE, Eisen MB. 2013. Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression. *PLoS Genet* **9**: e1003748.

Paszek, P., Jackson, D. A., and White, M. R. (2010). Oscillatory control of signalling molecules. *Current Opinion in Genetics & Development* 20, 670–676. doi:10.1016/j.gde.2010.08.004.

Pourquié, O. (2003). The segmentation clock: converting embryonic time into spatial pattern. *Science* 301, 328–330. doi:10.1126/science.1085887.

Payankaulam S, Li LM, Arnosti DN. 2010. Transcriptional Repression: Conserved and Evolved Features. *Curr Biol* **20**: R764–R771.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**: 499–502.

Phuc Le P, Friedman JR, Schug J, Brestelli JE, Parker JB, Bochkis IM, Kaestner

KH. 2005. Glucocorticoid receptor-dependent gene regulatory networks. *PLoS Genet* **1**: e16.

Pi W, Zhu X, Wu M, Wang Y, Fulzele S, Eroglu A, Ling J, Tuan D. 2010. Long-range function of an intergenic retrotransposon. *Proceedings of the National Academy of Sciences* **107**: 12992–12997.

Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Morrison H, Fitzpatrick DR, Afzal V, et al. 2008. Human-specific gain of function in a developmental enhancer. *Science* **321**: 1346–1350.

Purvis, J. E., Karhohs, K. W., Mock, C., Batchelor, E., Loewer, A., and Lahav, G. (2012). p53 dynamics control cell fate. *Science* 336, 1440–1444. doi:10.1126/science.1218351.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–283.

Rando OJ. 2012. Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Current Opinion in Genetics & Development* **22**: 148–155.

Rando OJ, Chang HY. 2009. Genome-wide views of chromatin structure. *Annu Rev Biochem* **78**: 245–271.

Reece-Hoyes JS, Diallo A, Lajoie B, Kent A, Shrestha S, Kadreppa S, Pesyna C, Dekker J, Myers CL, Walhout AJM. 2011. Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nat Methods* **8**: 1059–1064.

Rodionov DA. 2007. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev* **107**: 3467–3497.

Ryall JG, Dell'Orso S, Derfoul A, Juan A, Zare H, Feng X, Clermont D, Koulnis M, Gutierrez-Cruz G, Fulco M, et al. 2015. The NAD(+)-dependent SIRT1 deacetylase translates a metabolic switch into regulatory epigenetics in skeletal muscle stem cells. *Cell Stem Cell* **16**: 171–183.

Saier MH, Zhang Z. 2014. Transposon-mediated directed mutation controlled by DNA binding proteins in Escherichia coli. *Front Microbiol* **5**: 390.

Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**: 335–348.

Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter

C, Watt S, Martinez-Jimenez CP, Mackay S, et al. 2010a. Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding. *Science* **328**: 1036–1040.

Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, Bogwitz MR, Martin J, Daborn PJ, Goddard ME, Batterham P, et al. 2010b. Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. *PLoS Genet* **6**: e1000998.

Schones DE, Zhao K. 2008. Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet* **9**: 179–191.

Shlyueva D, Stampfel G, Stark A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**: 272–286.

Simon JM, Giresi PG, Davis IJ, Lieb JD. 2012. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nature Protocols* **7**: 256–267.

Sinha S, Siggia ED. 2005. Sequence turnover and tandem repeats in cis-regulatory modules in drosophila. *Molecular Biology and Evolution* **22**: 874–885.

Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ, et al. 2011. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**: 1270–1282.

Spitz F, Furlong EEM. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**: 613–626.

Stefflova K, Thybert D, Wilson MD, Streeter I, Aleksic J, Karagianni P, Brazma A, Adams DJ, Talianidis I, Marioni JC, et al. 2013. Cooperativity and Rapid Evolution of Cobound Transcription Factors in Closely Related Mammals. *Cell* **154**: 530–540.

Sun, L., Yang, G., Zaidi, M., and Iqbal, J. (2008). TNF-induced gene expression oscillates in time. *Biochemical and Biophysical Research Communications* 371, 900–905. doi:10.1016/j.bbrc.2008.03.114.

Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Research* **24**: 1963–1976.

Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332, 472–474. doi:10.1126/science.1198817.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse

embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663–676.

Teichmann SA, Babu MM. 2004. Gene regulatory network growth by duplication. *Nat Genet* **36**: 492–496.

Tigges, M., Marquez-Lago, T. T., Stelling, J., and Fussenegger, M. (2009). A tunable synthetic mammalian oscillator. *Nature* 457, 309–312. doi:10.1038/nature07616.

Valenzuela L, Kamakaka RT. 2006. Chromatin insulators. *Annu Rev Genet* **40**: 107–138.

Valor LM. 2014. Transcription, Epigenetics and Ameliorative Strategies in Huntington's Disease: a Genome-Wide Perspective. *Mol Neurobiol*.

Valor LM, Guiretti D, Lopez-Atalaya JP, Barco A. 2013. Genomic landscape of transcriptional and epigenetic dysregulation in early onset polyglutamine disease. *J Neurosci* **33**: 10471–10482.

van Steensel B, Dekker J. 2010. Genomics tools for unraveling chromosome architecture. *Nat Biotechnol* **28**: 1089–1095.

Vierstra J, Rynes E, Sandstrom R, Zhang M, Canfield T, Hansen RS, Stehling-Sun S, Sabo PJ, Byron R, Humbert R, et al. 2014. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**: 1007–1012.

Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, et al. 2015. Enhancer Evolution across 20 Mammalian Species. *Cell* **160**: 554–566.

Villar D, Flicek P, Odom DT. 2014. Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nat Rev Genet* **15**: 221–233.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854–858.

Voss TC, Hager GL. 2014. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet* **15**: 69–81.

Wang X, Wood TK. 2011. IS5 inserts upstream of the master motility operon flhDC in a quasi-Lamarckian way. *ISME J* **5**: 1517–1525.

Wapinski, O. L., Vierbuchen, T., Qu, K., Lee, Q. Y., Chanda, S., Fuentes, D. R., Giresi, P. G., Ng, Y. H., Marro, S., Neff, N. F., et al. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* 155,

621–635. doi:10.1016/j.cell.2013.09.028.

Wei Y, Arnosti DN. 2015. Enhancer trafficking: free throws and three-pointers. *Developmental Cell* **32**: 135–137.

Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB. 2008. The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. *Cell* **134**: 610–623.

Wittkopp PJ, Kalay G. 2011. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Publishing Group* **13**: 59–69.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* **8**: 206–216.

Wu, X., Bayle, J. H., Olson, D., and Levine, A. J. (1993). The p53-mdm-2 autoregulatory feedback loop. *Genes & Development* 7, 1126–1132.

Xu H, Hoover TR. 2001. Transcriptional regulation at a distance in bacteria. *Curr Opin Microbiol* **4**: 138–144.

Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, Struhl K. 2006. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Molecular Cell* **24**: 593–602.

Yasuoka Y, Suzuki Y, Takahashi S, Someya H, Sudou N, Haramoto Y, Cho KW, Asashima M, Sugano S, Taira M. 2014. Occupancy of tissue-specific cis-regulatory modules by Otx2 and TLE/Groucho for embryonic head specification. *Nature Communications* **5**: –.

Young, M. W., and Kay, S. A. (2001). Time zones: a comparative genetics of circadian clocks. *Nat. Rev. Genet.* 2, 702–715. doi:10.1038/35088576.

Yu HB, Johnson R, Kunarso G, Stanton LW. 2011. Coassembly of REST and its cofactors at sites of gene repression in embryonic stem cells. *Genome Research* **21**: 1284–1293.

Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A. 2015. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**: 556–559.

Zeitlinger J, Stark A, Kellis M, Hong J-W, Nechaev S, Adelman K, Levine M, Young RA. 2007. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nature Publishing Group* **39**: 1512–1516.

Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. 2012. Dynamic

Transformations of Genome-wide Epigenetic Marking and Transcriptional Control Establish T Cell Identity. *Cell* **149**: 467–482.

Zhang Z, Saier MH. 2011. Transposon-mediated adaptive and directed mutations and their potential evolutionary benefits. *J Mol Microbiol Biotechnol* **21**: 59–70.

Zhou VW, Goren A, Bernstein BE. 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12**: 7–18.

Ziller MJ, Edri R, Yaffe Y, Donaghey J, Pop R, Mallard W, Issner R, Gifford CA, Goren A, Xing J, et al. 2015. Ziller 15:neuronal diffn epigenetics. *Nature* **518**: 355–359.

Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EEM. 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **461**: 65–70.

CHAPTER II

Genome-wide errant targeting by Hairy[3]

1. Abstract

Metazoan transcriptional repressors regulate chromatin through diverse histone modifications. Contributions of individual factors to the chromatin landscape in development is difficult to establish, as global surveys reflect multiple changes in regulators. Therefore, we studied the conserved HES family repressor Hairy, analyzing histone marks and gene expression in Drosophila embryos. This long-range repressor mediates histone acetylation and methylation in large blocks, with highly context-specific effects on target genes. Most strikingly, Hairy exhibits biochemical activity on many loci that are uncoupled to changes in gene expression. Rather than representing inert binding sites, as suggested for many eukaryotic factors, many regions are targeted errantly by Hairy to modify the chromatin landscape. Our findings emphasize that identification of active *cis*-regulatory elements must extend beyond the survey of prototypical chromatin marks. We speculate that this errant activity may provide a path for creation of new regulatory elements, facilitating the evolution of novel transcriptional circuits.

Key words

transcription, repression, chromatin, gene regulatory network

[3]Chapter II is presented in the form of a manuscript published as: Kok K, Ay A, Li Li.M, and Arnosti DN. Genome-wide errant targeting by Hairy.

2. Introduction

Metazoan transcriptional circuitry features activation and repression signals that constitute robust regulatory networks important for the unfolding of developmental programs. In the Drosophila embryo, localized transcriptional repressors provide essential patterning information that establishes the primary anterior-posterior and dorsal-ventral axes of the organism. The action of transcriptional repressors is heterogeneous and can exhibit context effects; one of the most striking aspects involves the different classes of repressors that mediate distinct chromatin changes on target genes. Short-range acting proteins Snail and Knirps interfere with transcription only when their cognate binding sites are located within close range of the activator binding sites (Gray and Levine 1996). These proteins interact with evolutionarily conserved corepressors that possess chromatin modifying activities (Nibu et al. 1998; Payankaulam and Arnosti 2009). Paradoxically, these same cofactors are also recruited by another class of repressors, the long-range transcriptional repressors, exemplified by the Hairy factor (Paroush et al. 1994; Barolo and Levine 1997; Poortinga et al. 1998). This protein is a founding member of the Hairy/Enhancer of Split (HES) transcription factors, which play essential roles in animal development, including segmental gene patterning in the early embryo and specification of neuronal differentiation in response to Notch signaling (Kageyama et al. 2007). Thus, elucidation of molecular mechanisms of Hairy activity will shed light on a number of important gene circuits that are prominently represented in key developmental

pathways. The biochemical function of Hairy is associated with long-range chromatin modifications, which endow this factor with the ability to interfere with multiple *cis*-regulatory regions, including activators bound over 1 kb distal to the Hairy binding sites. The long-range effect has been proposed to be due to the recruitment of the corepressor Groucho (Gro), that can oligomerize to spread over large areas of the genome, and colocalization of HDAC to the target genes resulting in deacetylation of specific lysine residues in histones H3 and H4 (Courey and Jia 2001; Martinez and Arnosti 2008). In our previous studies, we showed that Hairy induced extensive tracts of deacetylation on *ftz*, a segmental patterning gene expressed early in embryogenesis (Li and Arnosti 2011).

While potent in repression potential, Hairy and other long-range repressors are apparently restricted in their ability to exercise transcriptional effects by the local *cis*-regulatory context in which binding sites are located. Hairy was demonstrated to lack long-range effects on a distal RACE enhancer in the embryonic dorsal ectoderm, when Hairy binding motifs were situated in an element with activators that are restricted to mesoderm/neurectoderm regions. Furthermore, the Dorsal protein, when itself acting as a long-range repressor, is dependent on neighboring Cut and Dri transcription factor motifs to function, indicating that long-range repression complexes may require specific *cis*-regulatory grammar (Nibu et al. 2001; Cai et al. 1996).

The action of eukaryotic transcriptional repressors involves a number of biochemical activities, including direct antagonism of transcriptional activators and assembly of chromatin-associated factors that are correlated with gene silencing (Perissi et al. 2010). Specific types of covalent histone modifications, such as H3 and H4 deacetylation, H3K9 trimethylation and H3K27 trimethylation are correlated with repressed genes, but there is still no general understanding of how important in a quantitative sense such modifications are for inhibition of transcription at specific genes. Context effects for a particular transcriptional repressor can influence what sort and how much of a response will be generated. At a genome-wide level, specific chromatin features correlate with transcriptionally repressed genes (e.g. H3K9 and 27 methylation, reduced levels of H3 and H4 acetylation, binding of HP1), however these marks are also found within highly active loci (modENCODE Consortium et al. 2010). The epigenetic signature of transcriptional repression is thus context-dependent, consistent with a revised picture of the simple “histone code” hypothesis. In the context of specific transcriptional repressors, we know little about how the context of distinct factors present at *cis*-regulatory elements shapes their action. Genome-wide information obtained from chromatin immunoprecipitation experiments should provide information about molecular targets and action of transcription factors, however, in addition to *bona fide* regulatory targets, metazoan transcription factors typically associate with a large number of in vivo binding sites of unknown significance. Recent studies have suggested that these interactions represent off-target genomic interactions, driven by low binding specificity of transcription

factors and a general affinity for open chromatin of active enhancers (MacArthur et al. 2009). A survey of possible “off target” binding elements suggested that these tend to be of lower affinity and are transcriptionally inert (Fisher et al. 2012). As noted above, previous studies of Hairy suggested that the protein is unable to mediate transcriptional repression in the absence of other factors co-occupying regulatory elements (Nibu et al. 2001).

Identification of functional properties of Hairy transcends the simple biochemical elucidation of repression; this protein is representative of the regulatory factors comprising conserved gene regulatory networks (GRN) that constitute the basis of animal development. Molecular studies have demonstrated that the acquisition or loss of binding sites or entire regulatory modules appears to drive significant changes in gene expression that initiate critical evolutionary transitions, such as elaboration of novel limb structures (Tanaka et al. 2011; Pavlopoulos et al. 2009; Khila et al. 2009). Significantly, although relatively subtle changes have been linked to such important evolutionary innovations, it appears that functional conservation of gene expression is also compatible with major changes in the structure of transcription control regions (Hare et al. 2008). The constraints for reorganization of existing *cis*-regulatory elements, or appearance of such elements *de novo*, are poorly understood; in some cases, the exact placement of multiple transcription factor motifs is essential for transcriptional function, while the composition of other genetic switches appears to be very loosely organized (Arnosti and Kulkarni 2005). The existence of a large fraction of “off-target”

binding sites both complicates the analysis of important functional links, and the interpretation of potential evolutionary changes. Thus, elucidation of the functional targets and chromatin effects of Hairy can provide important insights on the basic substance of evolutionary variation. In this study, we use genetic tools to mediate induction of Hairy on a short time scale, permitting us to identify direct regulatory targets and chromatin effects on a genome-wide level. In addition to identifying common features of Hairy repression mechanisms across many targets, we also show that this protein exerts pervasive biochemical activity to change chromatin states at many loci unlinked to gene expression, revealing a possible pathway to evolution of novel gene regulatory connections.

3. Results

3.1. Genome-wide transcriptional regulation by Hairy

To study transcriptional repression at the genome-wide level at this important developmental stage, we profiled changes in transcriptome, epigenome and RNA polymerase II binding regulated by Hairy in the blastoderm embryo using an inducible system as described previously to capture direct effects with high temporal resolution (Li and Arnosti 2011) (Figure 2-1A). Hairy is first expressed in the Drosophila blastoderm embryo in a seven stripe pattern, which is important in controlling downstream pair rule genes that direct segmentation (Ish-Horowicz and Pinchin 1987). Here, we express Hairy with a brief heatshock, throughout the embryo, which is sufficient to completely repress target genes such as *ftz* (Figure 2-1A,B). We treated the control embryos identically to embryos carrying the

inducible Hairy transgene to test for possible nonspecific effects of heat shock on gene expression and chromatin marks. In this system, heat shock alone has no effect on the expression patterns of the pair rule and other genes analyzed, and the chromatin marks in heat shocked control embryos were indistinguishable from chromatin patterns previously reported for untreated embryos (Li and Arnosti 2011 and K. Kok, data not shown). In total, we identified 241 down-regulated and 146 up-regulated transcripts in response to induction of Hairy (Figure 2-1C). Our microarray analysis captured previously identified targets of Hairy, showing downregulation of *en, edl, Impl2,* and *prd,* as well as *ftz,* all of which were previously found to be derepressed in *h* embryos (Ish-Horowicz and Pinchin 1987; Bianchi-Frias et al. 2004).

Differentially regulated genes were compared to those physically bound by Hairy (MacArthur et al. 2009); 70% of down-regulated genes are bound by Hairy, suggesting that most of these are likely to be direct targets (Figure 2-1C). In contrast, only 30% of up-regulated genes are bound by Hairy, indicating that majority of these genes may be indirect targets. *In situ* hybridization and RT-qPCR confirmed the repression of a number of target genes we identified (Figure 2-1B,D). Many of these genes, including *odd*, *comm*, *comm2*, *edl*, *en*, *Impl2, prd*, and *18w*, have striped expression patterns complementary to that of Hairy, supporting direct regulation by the repressor. Furthermore, consistent with known biological functions of Hairy, gene ontology analysis showed that categories for down-regulated genes are significantly enriched in transcriptional regulation, cell

fate commitment and neurogenesis ($p<3.7e-18$). GO categories for the set of upregulated genes were of lower statistical significance, and included reproductive processes ($p<0.03$) (Supplementary file 1,2).

Expression of the majority of genes bound by Hairy did not change (Figure 2-1C), consistent with previous observations that metazoan transcription factors have apparently many “nonfunctional” interaction sites in the genome (Cusanovich et al. 2014).

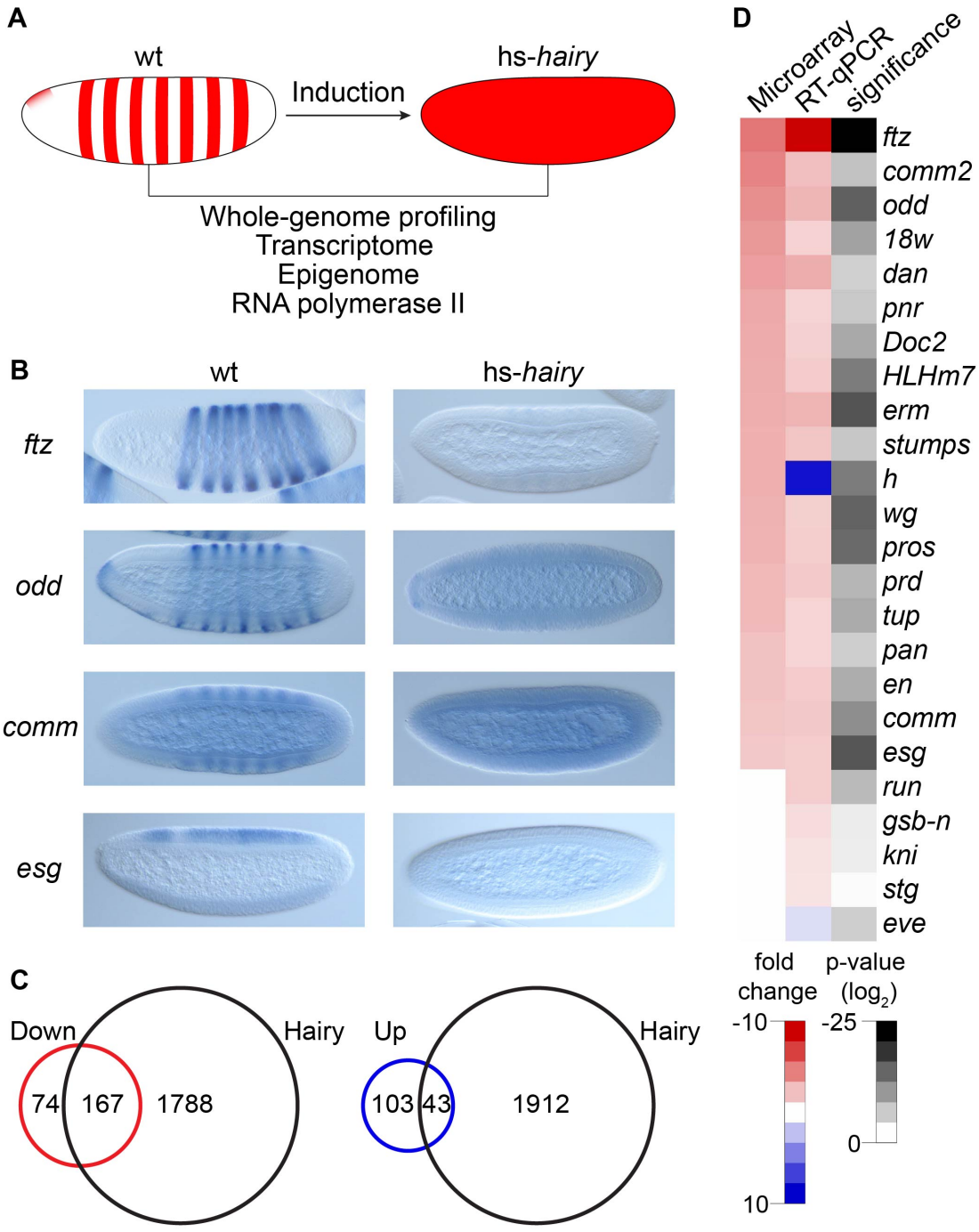


Figure 2-1 Global analysis of Hairy regulation. (A) Schematic expression of Drosophila embryo system used for Hairy repression, with outline of the genome-wide analysis of transcription, chromatin, and RNA polymerase II. **(B)** Repression of *ftz*, *odd*, *comm* and *esg* revealed by *in situ* hybridization in wild-type (wt) and Hairy transgenic embryos (hs-*hairy*) after 20 min induction. Similar repression of *18w, HLHm7* and *erm* was also observed (not shown). **(C)** Transcriptionally regulated (red, down; blue, up) and Hairy bound genes identified by microarray and ChIP-chip (MacArthur et al. 2009). A larger fraction of down-regulated genes were physical targets of Hairy than for up-regulated genes (significance: p=3.8e-

95 and p=3.5e-08 respectively, hyper-geometric test). Differentially expressed genes are selected based on $p<0.05$ and fold change > 2. **(D)** Validation of microarray data by RT-qPCR, showing concordance between these methods. Genes are ranked by the fold change from the microarray measurements. Significance was tested by Student's T test. y-axis values were normalized as described in Materials and Methods.

3.2. Coordinate chromatin transitions mediated by Hairy on diverse genes

Identification of functional and physical Hairy targets allowed us to study gene-specific chromatin changes associated with repression. We performed epigenomic profiling via chromatin immunoprecipitation-high throughput sequencing (ChIP-seq) of chromatin marks that are often correlated with specific features of *cis*-regulation; H4Ac, H3K27Ac, and H3K4me1 at promoters and enhancers; H3K4me3 at transcription start sites (TSS); H3K36me3 at gene body regions; and H3K9me3 at repressed regions of chromatin (Zhou et al. 2011). The measured signals for specific marks were highly reproducible in separate biological replicates, and Hairy-induced changes in histone marks were consistently observed at specific loci, such as the widespread loss of the H4Ac signal on the *ftz* locus, with little change to the overall global chromatin landscape (Figure 2-2-1A). As was apparent from comparison of control chromatin profiles, the induction of Hairy did not cause a global impact on histone marks. In the presence or absence of induced Hairy, the genome features for multiple chromatin marks are virtually identical, except in very discrete regions where there are significant changes (Figure 2-2-1A-C).

Using single gene techniques, we previously found that Hairy induces a widespread histone H4 deacetylation throughout the entire *ftz* locus (Li and Arnosti 2011). To determine if these are general properties of Hairy, we compared all affected loci genome-wide. We observed that on a number of transcriptionally repressed target genes, H4 deacetylation is coupled with loss of

the active marks H3K27Ac and H3K4me1. Widespread reduction of these active marks affecting > 1 kb blocks was observed on many genes repressed by Hairy, including *ftz* and other segmentally expressed genes such as *h* and *18w* (Figure 2-2). Notably, Hairy regulates its own transcription by chromatin alteration, consistent with autoregulatory mechanism of related mammalian HES proteins (Kageyama et al. 2007). In addition to removal of enhancer marks, repression on *h* and *18w* resulted in demethylation of the promoter mark H3K4me3. Furthermore, action of Hairy on another pair rule gene, *odd,* was limited to removal of acetyl marks on H4 and H3K27; methylation marks on H3K4 are untouched (Figure 2-2D). These results suggest that Hairy mediates coordinated sets of chromatin transitions. The chromatin changes did however exhibit heterogeneous characteristics; the sizes of altered chromatin domains varied on different repressed genes. For example, changes in levels of H4Ac involved blocks with a range of sizes; generally larger than 1 kb, with the average ~2.5 kb. Somewhat smaller chromatin blocks were associated with repression of the *HLHm7*, *gogo, pros* and *tup* genes, which showed just as robust regulation of transcription as those genes with large tracts of chromatin modification (Figure 2-2E, F, G, H).

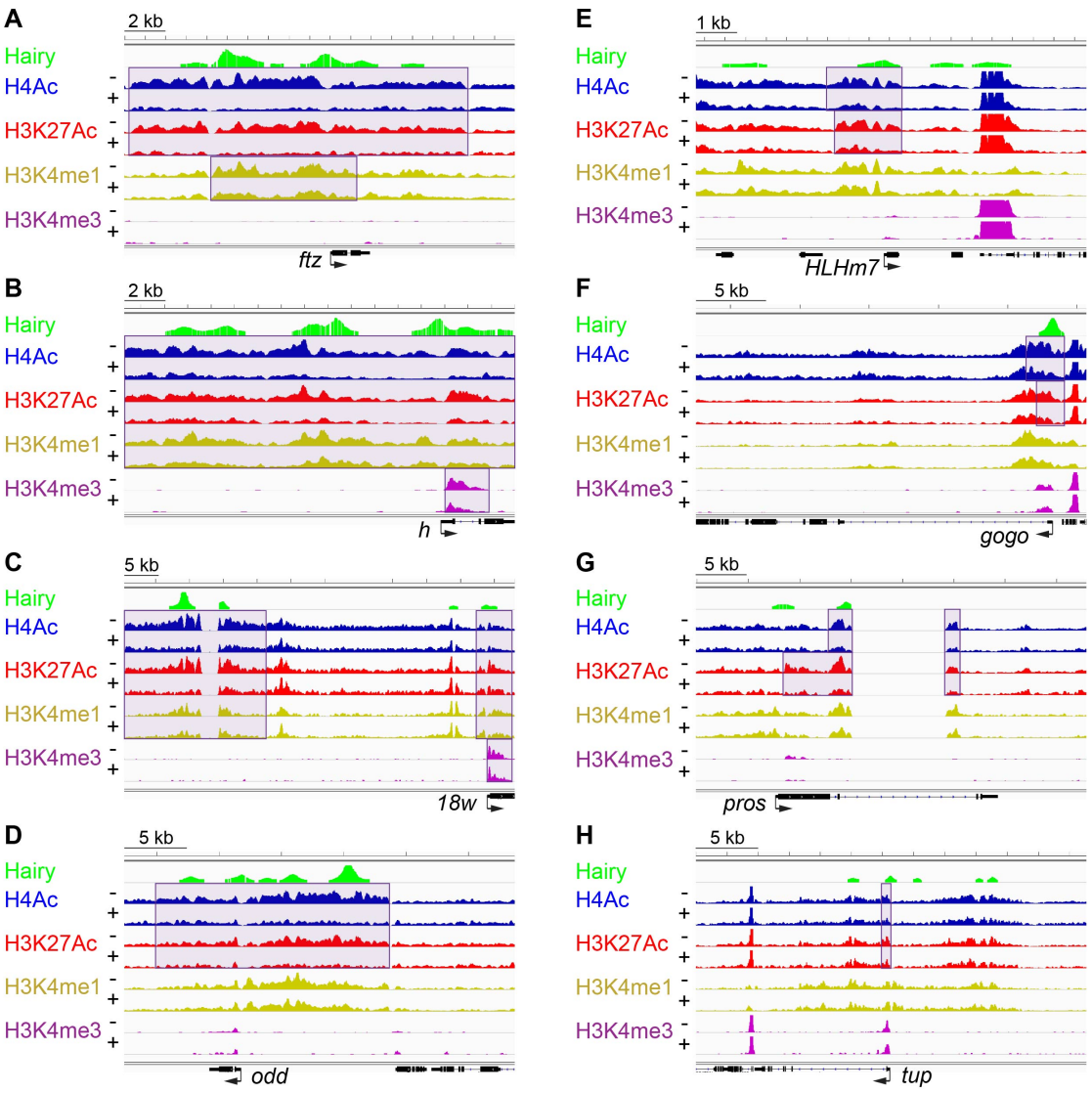


Figure 2-2 Examples of coupled, large-scale chromatin changes mediated by Hairy. ChIP-seq tracks for H4Ac, H3K27Ac, H3K4me1 and H3K4me3 are shown at repressed genes before (-) and after Hairy (+) induction, with gene models below. **(A-D)** Coupled reduction of active histone marks was observed in a wide-spread fashion on *ftz, h, 18w* and *odd* genes (scale at top left). **(E-H)** Relatively smaller blocks of chromatin changes were detected *on HLHm7, gogo, pros* and *tup* genes. Significantly changed regions (shaded boxes) were identified by the diffReps program. Hairy binding (top track) from MacArthur et al. 2009.

The largest ranges of size in chromatin domains were observed for H4Ac, but similar, although smaller ranges were also seen for H3K27Ac and H3K4me1 marks (Figure 2-3, Figure 2-3-1A,B and Supplementary file 3). We found strong correlations between the sizes of the domains of chromatin modification and the direct action of Hairy. Hairy-bound blocks of deacetylation were significantly larger than those not bound by Hairy, and smaller correlations were noted for other modifications, indicating that deacetylation is especially likely to show “spreading” characteristics (Figure 2-3, Figure 2-3-1A,B and Supplementary file 3).

These results suggest that widespread effects found at H4Ac, H3K27Ac and H3K4me1 marks are dependent on presence of Hairy and are consistent with a long-range “spreading” repression mechanism. We saw no correlation between the height or extent of Hairy binding sites and the range of chromatin alteration, suggesting that the effectiveness of this protein is not merely a function of number of binding sites (Figure 2-3-2A,B). Other local factors may dictate how extensively modifications are propagated on individual genes. Therefore, Hairy induces diverse chromatin transitions associated with gene silencing, indicating that there are gene-specific features dictating how repression is mediated at individual genes.

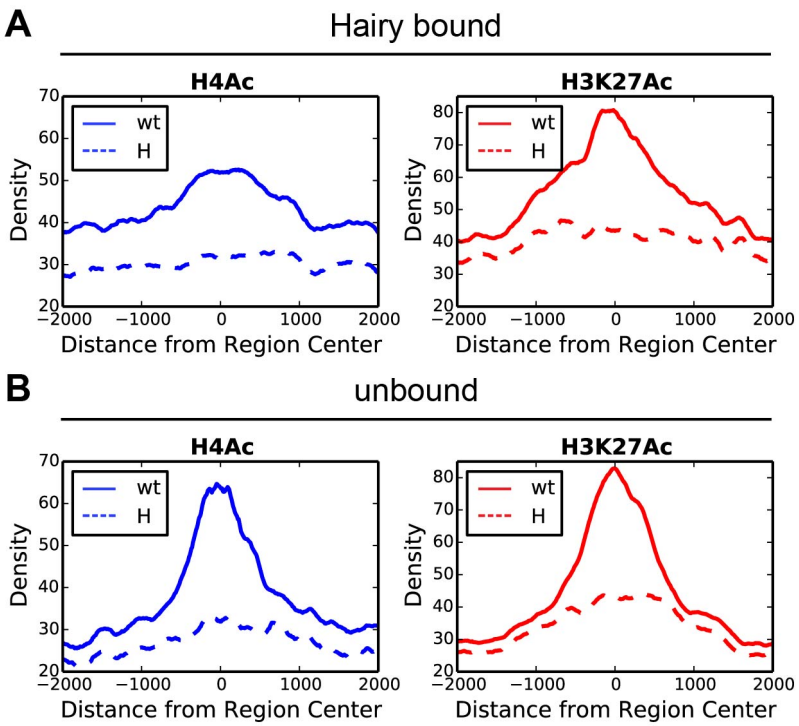


Figure 2-3 Direct Hairy target genes exhibit broad domains of chromatin effects. Distribution of genome-averaged ChIP-seq signals before (straight line) and after (dashed line) Hairy induction, showing 4 kb window around affected regions. **(A)** Distributions of histone H4Ac and H3K27Ac marks of direct Hairy targets were significantly broader than for regions **(B)** not bound by Hairy (p=2.55e-92 and p=5.63e-70 respectively; KM test).

3.3. Global set of chromatin modifications mediated genome-wide

These observations suggest there are context-specific aspects to chromatin modifications directed by Hairy. To determine the nature of changing chromatin states at different genomic loci, we compared the complete set of significant alterations in all measured chromatin marks observed after Hairy induction, regardless of transcriptional effects on the neighboring genes. We observed both loss and gain of these marks on hundreds of regions. Most frequently observed were changes in H4Ac, H3K27Ac, H3K4me1 and H3K36me3; changes in some chromatin marks were much more frequent than in others, indicating that there is some heterogeneity in the impact of Hairy on different regions (Figure 2-4A). The changes in levels of these marks is not simply due to increased or decreased histone density, as histone H3 levels generally were unchanged (Figure 2-4A). The roughly equal abundance of regions showing loss or gain of acetylation and methylation would indicate that either secondary effects are common, or that Hairy may exert distinct biochemical activities on different loci. The correlation of Hairy-bound regions with repressed transcripts, as well as the association of Hairy binding with longer-range deacetylations, but not with increased acetylation, supports the idea that indirect effects are common. Indeed, focusing specifically on genes targeted by Hairy, we found that H4 histone deactylation was strongly enriched compared to acetylation gains, suggesting that deacetylations are direct effects (Figure 2-4B and Figure 2-4-1A). Further support comes from consideration of the actual Hairy occupancy of the chromatin blocks in question; there was significant correlation between Hairy binding and

chromatin blocks exhibiting decreased, but not increased acetylation (Figure 2-4-2).

With respect to another chromatin mark, changes in histone methylation revealed an unexpected and interesting trend. Both decreases and increases in H3K4me1 signals were significantly associated with Hairy-bound genes; decreases were especially found in those regions directly bound by Hairy (Figure 2-4-1A, Figure 2-4-2 and Figure 2-4B). At the same time, about one-quarter of the genes that were transcriptionally silenced by Hairy showed increases in H3K4me1, although these regions of increase did not overlap with Hairy binding. The increase in this mark may represent a reaction of proximal promoter chromatin to distal enhancer silenced by Hairy.

H3K36me3 modification is often associated with active transcription. We found a small fraction of transcriptionally regulated genes that exhibited changes in the mark upon transcriptional repression (Figure 2-4C and Figure 2-4-1B). These findings indicate that Hairy repression does not require H3K36me3 changes. Indeed, the direct effect of H3K36me3 on transcription is complex, as has been found for many other histone marks. For example, upregulation of KDM4A histone demethylase target genes in Drosophila occurs without increases in H3K36me3 (Crona et al. 2013). Similar studies with elongation factor Spt6 in Drosophila further indicate that *Hsp70* gene expression is not correlated to H3K36me3 levels (Ardehali et al. 2009). In fact, H3K36me3 may in some

contexts contribute to gene silencing, due to its presence in heterochromatic domains (Chantalat et al. 2011) and in other cases, removal of H3K36me3 is required to promote transcriptional elongation (Kim and Buratowski 2007).

A smaller number of H3K9me3 regions were observed to change globally, or on genes that were associated with Hairy (Figure 2-4A,B). Very few repressed genes showed any alteration in this mark, thus it appears that repression mediated by Hairy does not require changes in such repressive histone modifications (Figure 2-4C), consistent with our previous report that repression on *ftz* did not change H3K27me3 levels (Li and Arnosti 2011). Indeed, other studies have found that these marks are not always simply coupled to repression. For example, only a modest correlation between H3K9me3 and H3K27me3 levels and gene silencing was observed in human cells (Barski et al. 2007; Zhang et al. 2012). In the differentiation of T and B cells, only a small fraction of repressed genes ever acquire H3K27me3 (Zhang et al. 2012; McManus et al. 2011). Interestingly, H3K9me3 was found to be enriched in many active promoters and associated with transcriptional elongation in vertebrates (Squazzo et al. 2006; Vakoc et al. 2005).

Consequently, of the assessed modifications, it appears that Hairy predominantly works to modify acetyl and methyl marks of H4, H3K27 and H3K4 and represses gene expression primarily by eliminating active marks.

Of all chromatin regions impacted by Hairy, only a small number are associated with genes demonstrating measurable transcriptional changes (Figure 2-4C). Thus, it is striking that the majority of chromatin changes are decoupled from any detectable effect on gene expression (Figure 2-4-1B). For the many cases where chromatin effect was unlinked to changed gene expression, we observed extensive chromatin alterations associated with both silent and active genes. For example, chromatin transitions occur on transcribed genes not functionally repressed by Hairy, as seen on the *pyr* gene (Figure 2-5A). In this case, the gene may remain active because the necessary *cis*-regulatory elements are located distally and are still able to interact with the promoter and activate it. In other cases, chromatin changes flank silent loci; *nht* undergoes widespread deacetylation and demethylation even though it is silent during this developmental stage of embryos (Figure 2-5B). In some cases, binding and changing chromatin near inactive genes by Hairy in the blastoderm embryo may involve the interaction of Hairy with DNA elements that will become active at a later developmental stage, however, this seems unlikely in the case of *nht*, a testes-specific gene. Here, the physical binding by Hairy and subsequent impact on chromatin may represent “errant targeting”. Overall, chromatin changes were observed to correlate with over half of the regions bound by Hairy, suggesting that in most cases, this protein is biochemically active on chromatin, whether or not the changes lead directly to gene repression (Figure 2-5-1).

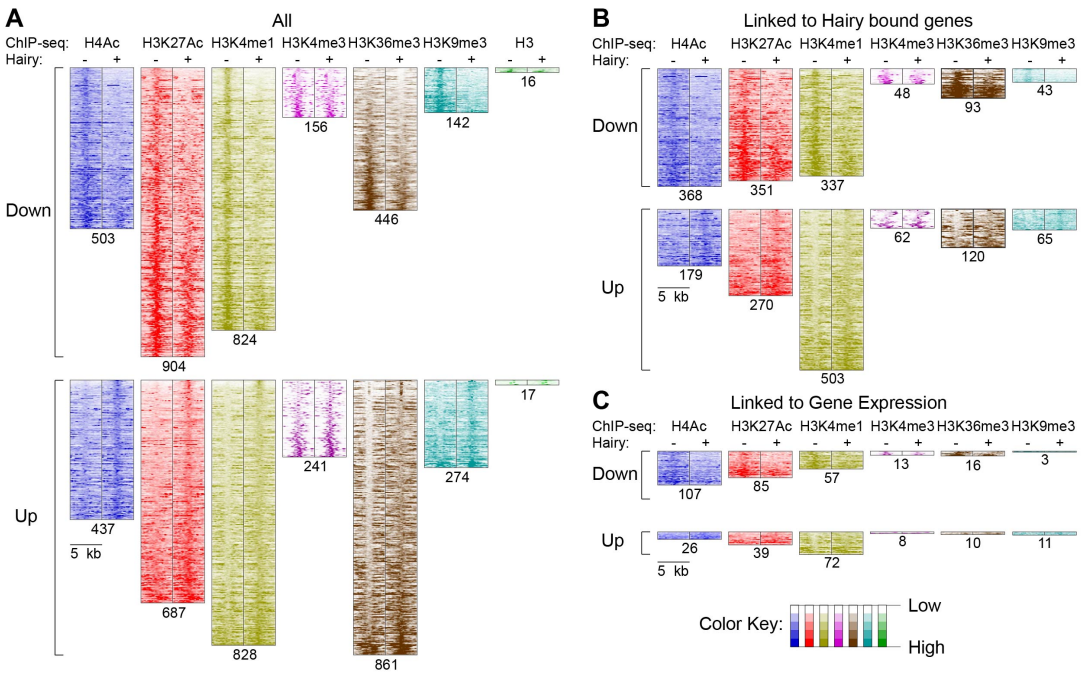


Figure 2-4 Pervasive genome-wide chromatin effects of Hairy. (A) All reduced (top) and increased (bottom) chromatin marks in the genome for H4Ac, H3K27Ac, H3K4me1, H3K4me3, H3K36me3, H3K9me3 and H3 shown as heatmaps for 5 kb windows from the center of significantly affected regions before (-) and after (+) Hairy induction. The number of affected regions indicated below each mark. **(B)** Affected chromatin regions associated with Hairy-bound genes show preferential enrichments for H4Ac, H3K27Ac, and H3K4me1. All affected regions were assigned to closest genes, and those in the vicinity of Hairy-bound genes are shown. **(C)** Subset of modified regions from **(B)** that were linked to genes transcriptionally regulated by Hairy. Significance of enrichment for chromatin modifications shown in Figure 2-4-1A,B.

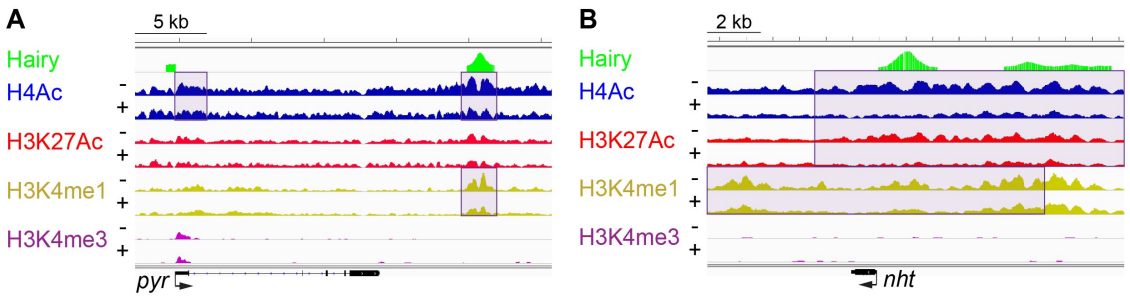


Figure 2-5 Examples of chromatin-modified loci unlinked to changes in gene expression. (A) *pyr* is actively transcribed, and not significantly repressed by Hairy, **(B)** while *nht* is not expressed at this stage. ChIP-seq tracks for H4Ac, H3K27Ac, H3K4me1 and H3K4m3 are shown before (-) and after Hairy (+) induction.

3.4. Hairy coordinates sets of modifications on preferred gene regions

The individual cases described in Figure 2-2 suggest that Hairy organizes a coordinated set of chromatin changes involving both deacetylation and demethylation of multiple histone residues. To determine if such alterations are a general property of the repressor, we assessed the extent of coordination of modifications on all individual blocks of affected chromatin. Changes in H4Ac, H3K27Ac and H3K4me1 marks were significantly correlated at many loci (Figure 2-6A). Deacetylation events were also strongly correlated with loss of both H3K4me1 and H3K4me3, indicating that Hairy may form complexes containing both deacetylase and demethylase activities. Indeed, the CtBP cofactor is known to bind both of these classes of enzymes. However, Hairy is not mediating only one average type of transformation; removal of methyl groups from H3K4me1 and H3K4me3 is catalyzed by distinct classes of enzymes; Hairy is likely to interact with both, allowing for removal of H3K4me1 marks on distal sites and H3K4me3 at TSS (Figure 2-6B,C). A very similar pattern of correlations between acetylation marks, and between acetylation and methylation marks was observed for regions with increased acetylation and methylation. These elements may represent to a large extent indirect targets of Hairy, as no significant overlap between Hairy binding and these modified regions was found (Figure 2-4-2).

Where does Hairy most commonly mediate significant chromatin modifications? We compared the location of individual histone marks genome-wide to those altered by Hairy expression. Although a third of Hairy binding sites are promoter-

proximal, where the majority of H4 and H3K27 actetylation occurs, the large majority of affected chromatin sites were found on intergenic and intronic regions, suggesting that successful alterations are targeted to distal sites that may represent transcriptional enhancers (Figure 2-6B). By contrast, changes in the methylation marks H3K4me1, H3K4me3, H3K36me3, and H3K9me3 are found in the genomic regions where they are naturally enriched (Figure 2-6B). For instance, H3K4me3 marks are enriched at TSS, as are the bulk of the altered chromatin sites. Hairy may thus have privileged sites on which it is more likely to induce chromatin changes; promoter regions may be in general more resistant to acetylation changes if strong activators are replenishing acetylation marks at these loci. In addition, transcriptional targets of Hairy are enriched in developmentally regulated genes, which typically possess larger *cis*-regulatory regions with multiple distal enhancers (Supplementary file 1) (Nelson et al. 2004).

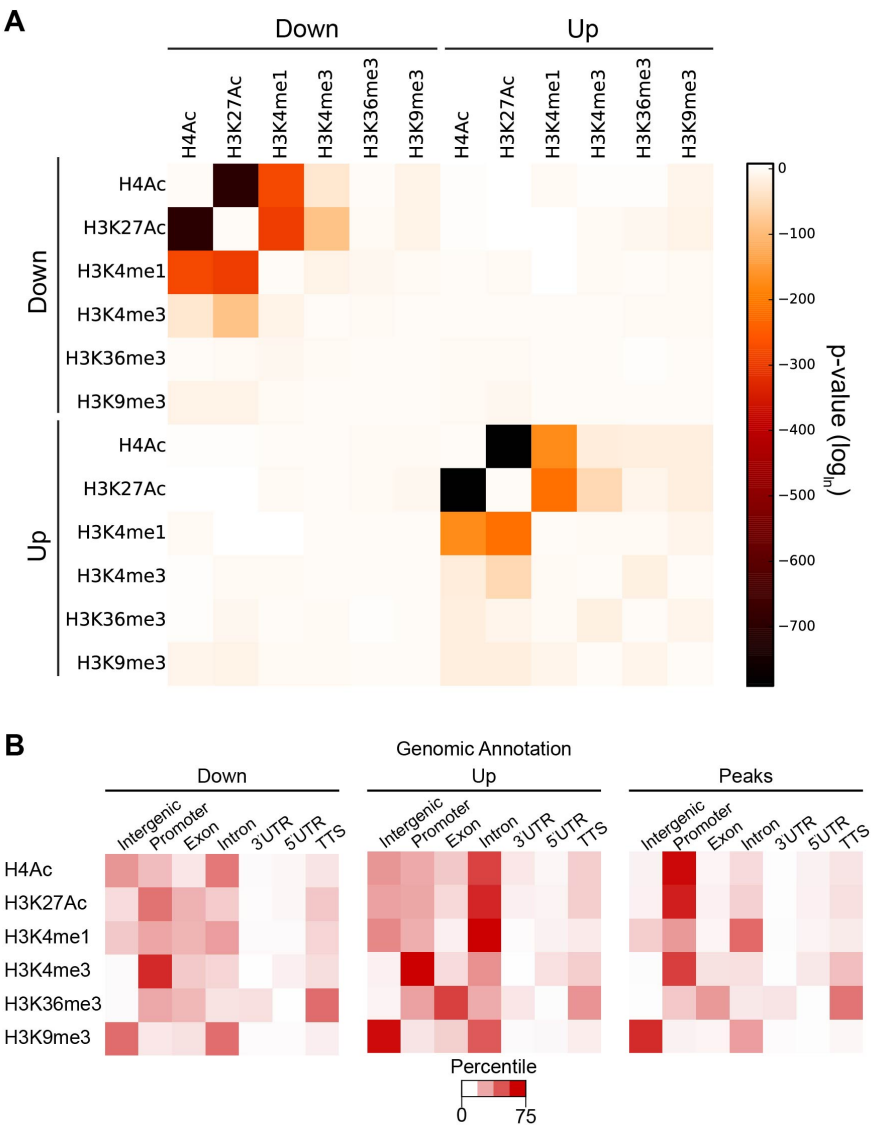


Figure 2-6 Coordination in changes of specific chromatin modifications by Hairy. (A) Very strong overlap between decreases in regions of H4Ac, H3K27Ac, H3K4me1 (heat map, upper left quadrant). Similar coordination between increases of H4Ac, H3K27Ac, H3K4me1 was noted (lower right quadrant). Combined increases and decreases of different marks were rarely observed. **(B)** Distribution of modified blocks by genomic regions show preferential action of Hairy at a distance from TSS. Affected regions were mapped to intergenic regions, promoter, exon etc. Overall distribution of genomic peaks for measured marks shown at right; the distributions for affected H4Ac and H3K27Ac regions deviated from the genomic averages (left, decreased, and center, increased levels).

3.5. RNA polymerase II and silencing by Hairy

To directly assess the influence of Hairy on transcriptional machinery, we compared the genome-wide occupancy of RNA polymerase II (Pol II) before and after Hairy induction. 75 of 241 repressed genes exhibited changes in Pol II occupancy (Figure 2-7A). Only three of those are not directly bound by Hairy, indicating a direct regulation by Hairy in loss of Pol II signal. A marked decrease of Pol II occupancy was observed at the *ftz* promoter, gene body and distal downstream region (Figure 2-7B). Loss of binding at the promoter, or the body of the gene, or both was detected on other loci (Figure 2-7C-I). Thus, the loss of Pol II on the promoter and gene body of *ftz* is not universally associated with transcriptional repression; on other genes, silencing of a distal enhancer may interfere with promoter release without blocking polymerase recruitment to the promoter, consistent with recent studies implicating transcriptional signaling in promoter escape, rather than promoter recruiting (Lagha et al. 2013). As expected, genes with associated chromatin changes without any impact on transcription did not show any change on Pol II occupancy (Figure 2-7J,K).

95 repressed genes bound by Hairy did not show any change in Pol II occupancy (Figure 2-7A). It is possible that Hairy induces a slower transit rate of Pol II without any detectable change in Pol II binding. It has been suggested that repression through elongation control may cause no change in Pol II binding on *slp1* and *Hsp70* (Wang et al. 2007; Adelman et al. 2006; Ardehali et al. 2009). Our previous analysis of *eve* repression by short-range repressor Knirps showed

similar effects (Li and Arnosti 2011). Therefore, Hairy may interfere with gene expression at different steps of the transcription cycle, as also suggested for repression by GR, indicating gene specific repression mechanisms (Gupte et al. 2013). An additional consideration is that genes featuring poised polymerase at the promoter in many or most nuclei, but are only expressed in a few nuclei, will have weak signals at the body of the gene. Therefore, the lack of change in Pol II levels on the gene body would reflect the inherently low signal, rather than a distinct biochemical mechanism. This explanation may account for a considerable number of affected genes where no changes in Pol II levels are observed after repression.

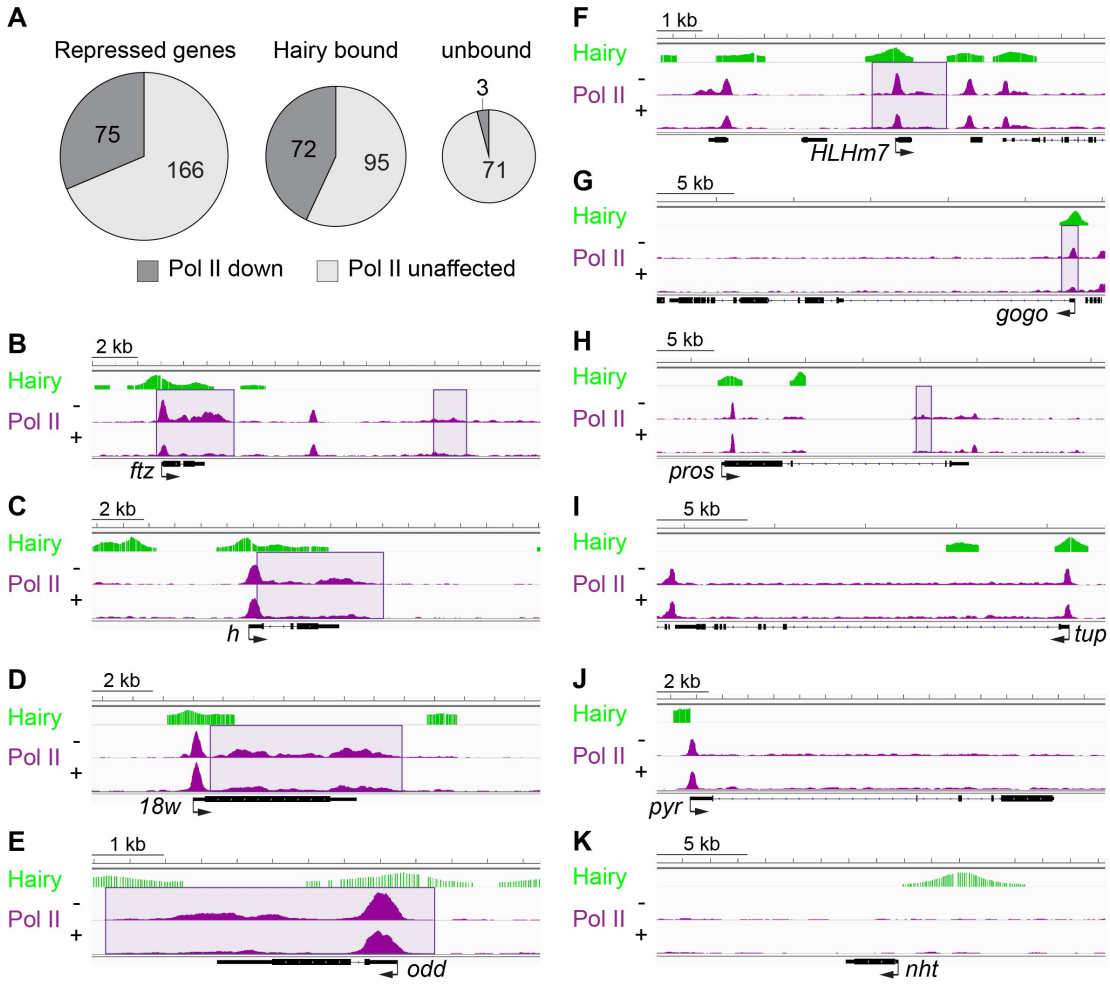


Figure 2-7 Diverse impact on RNA polymerase II occupancy by Hairy. (A) A minority of genes show significant changes in Pol II occupancy after Hairy repression, although a larger proportion of the directly targeted genes have measureable decreases in Pol II. "Repressed genes" shows entire set of transcriptionally downregulated genes, with reduced Pol II occupancy shown in dark gray. Subsets of genes directly bound or not bound by Hairy shown in center and at right. **(B-I)** Pol II occupancy on transcriptionally regulated genes before (-) and after (+) Hairy induction. Pol II occupancy decreases in the promoter and gene body of *ftz* and *odd,* only on the gene body of *h, 18w* and *pros,* and only at the promoter of *HLHm7* and *gogo*. Pol II signal was not changed significantly on *tup.* **(J, K)** Consistent with lack of transcriptional effects on other genes with associated chromatin modifications, Pol II occupancy on *pyr* is not changed, and absent on *nht*.

3.6. Predicting a “successful” repression context

The complexity of chromatin transitions observed genome-wide in the wake of Hairy expression prompted us to ask which features best predict successful repression of a target gene, versus those genes with no chromatin responses or exhibiting errant targeting by Hairy. Here, we alter the expression of only one regulatory factor, rather than the many changes in regulatory factors observed over a developmental time course, therefore our data sets are enriched for direct action of Hairy, potentially simplifying the search space. We sought out correlations between dynamic histone marks, Pol II, Hairy, CtBP and Gro and the repression of targeted genes. Direct inspection reveals that occupancy by Hairy, Gro, and decreases in Pol II are strong predictors of repression, as are several histone marks, compared to genes not repressed unaffected or those activated (Figure 2-8A). However, there are numerous loci that do not fit these simple generalizations. To more systematically assess the connections between these different observed states and transcriptional repression, we applied machine learning to analyze features that may be implicated in the activity of Hairy. We tested 41 features, including the number of observed peaks for Hairy, CtBP, and Gro; the number, width, and magnitude of altered chromatin blocks, and distance to TSS for 583 genes (241 repressed, 146 activated and 196 unaffected genes; activated and unaffected genes were grouped as nonrepressed genes). To identify the most informative features, four different feature selection algorithms were used to rank the information content of the 41 measured properties associated with the genes; the top twenty of these features were then used for

predictions (Supplementary file 4). We then tested four classifiers, using 90% of the data for training and 10% for predictions, with 10-fold cross-validation. Overall, each of the classifiers performed better than background, with Random Forests showing superior performance of ~75% accuracy for repressed and nonrepressed genes (Figure 2-8B). Three of the feature selection algorithms used with this classifier employed very similar features to achieve this high level of accuracy (Supplementary file 4), indicating that certain features are most informative. The presence and properties of Hairy and Gro peaks are good indicators, although not sufficient information by themselves. RNA polymerase II properties, transcript levels, and chromatin modifications, especially H3K4me1 and H4Ac, whether causal or not, are also a close reporter of gene activity. The overall performance differences in these methods are frequently observed in machine learning studies, and likely reflect the underlying data structure and types of features available for analysis. Genes that were correctly predicted as repression targets generally had the most differential features, including binding by Hairy and Gro, and changes in histone modifications. The genes that were least successfully called had one or no differential features, and may represent genes that are expressed in fewer cells and at lower levels where measurement of chromatin changes in a global population is difficult (Figure 2-8C). The unrepressednonrepressed gene *pyr* was consistently called as “repressed” by the machine learning algorithms, as it exhibited chromatin signatures similar to those found on genes that were actually repressed (Figure 2-8C). In this case, we propose that the relevant enhancers lie outside of the chromatin regions

affected by Hairy. Such genes may represent loci that are poised for capture in the Hairy regulatory network through stepwise acquisition of activator binding sites. Overall, this analysis indicates that from the perspective of Hairy biochemistry, there are intuitive and some non-intuitive combinations of chromatin dynamics that typify this protein's action in the context of transcriptional repression, rather than a "practice" site, but other factors predominate in many instances. The missing information likely relates to the activity of bona fide *cis*-regulatory elements that are acting on genes in the vicinity of Hairy, which is partially but incompletely known from genome-wide studies (Kvon et al. 2014).

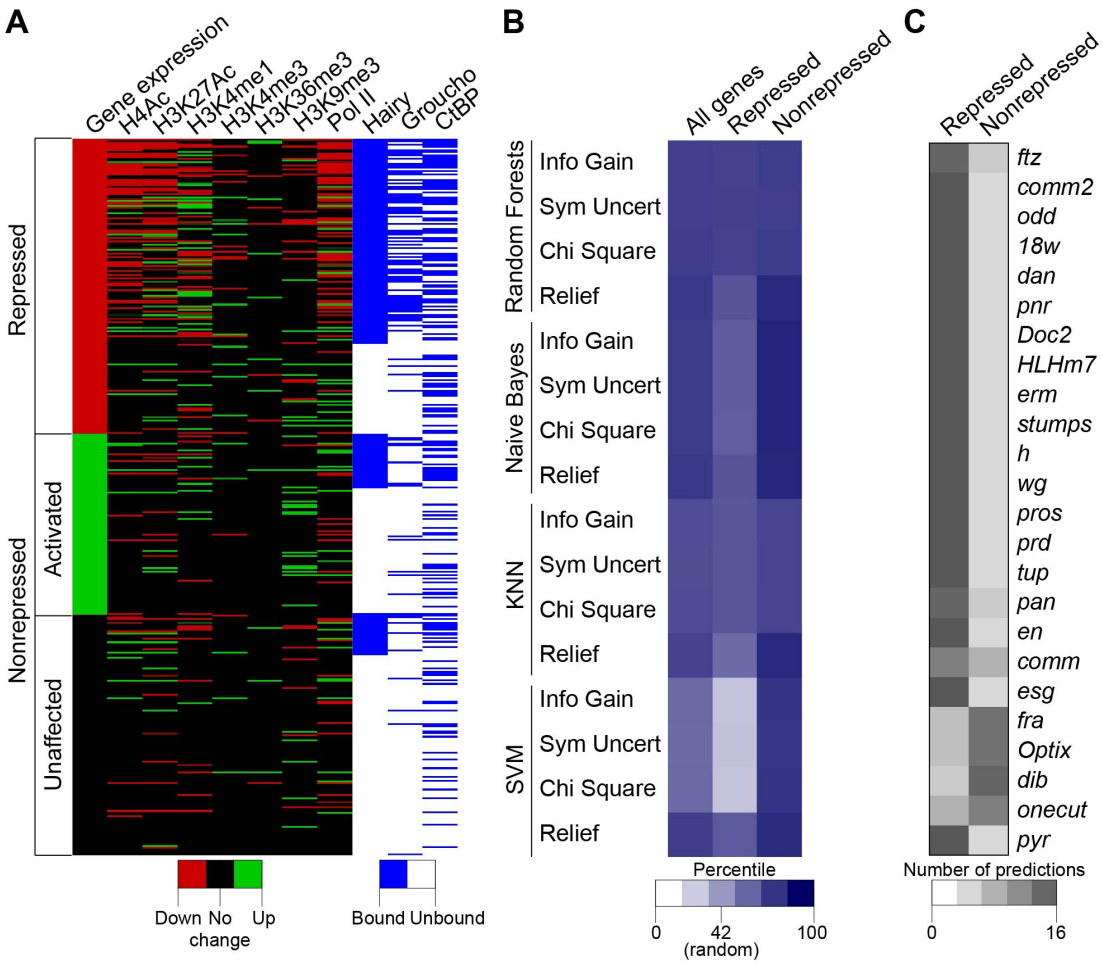


Figure 2-8 Machine learning reveals complex chromatin code for repression of Hairy target genes. (A) Changes in histone marks, Pol II occupancy and Hairy, Gro and CtBP binding on repressed (red), activated (green) and unaffected (black) genes upon Hairy induction. Genes were grouped by change in expression, then subgrouped into Hairy bound or unbound, and finally ranked by fold change in gene expression. Activated and unaffected genes were grouped as nonrepressed genes. **(B)** Relative success rate at calling repressed and nonrepressed genes for four different machine-learning models. Background prediction for this entire set is expected to be 58%; Random Forests, Naive Bayes, KNN classifiers had an average success of 75% overall, while the SVM classifier was not better than background. Classifiers were used in conjunction with Information Gain, Symmetrical Uncertainty, Chi Square and Relief feature selection algorithms. The average prediction accuracies of each method are shown in the first column. Expected random success (42%) for repressed genes (middle column) shown on heat map scale bar. **(C)** Model predictions for subset of repressed genes including those identified in Figure 2-1; top 19 were successfully predicted by almost all methods. *fra*, *Optix*, *dib*, and *onecut* were genes with disparate predictions that had few measureable chromatin features. At bottom, uniform false “repressed” calls for *pyr*, which was not transcriptionally repressed.

4. Discussion

By testing direct effects of the Hairy repressor in the embryo, we conclude that this protein coordinates a stereotypical set of chromatin modifications, modulated by local context, that underlie its function as a long-range repressor. Most remarkably, these changes on chromatin impact large segments of the genome that are not directly relevant to gene expression in this developmental context. We speculate that these off-target activities may provide an easy entry point for evolution of novel regulatory switches (Figure 2-9). Our mechanistic analysis of Hairy provides insights into likely mechanisms of related HES factors, as well as other transcriptional repressors that serve as scaffolds for chromatin modifying complexes. Hairy interacts with the widely utilized cofactors Gro, CtBP, and the Sir2 HDAC, and here we provide for the first time a genome-wide picture of the biochemical activities of this archetypal repressor.

How is transcription actually controlled by Hairy? The associated chromatin modifications may be effects, rather than direct causes of gene silencing. Our previous studies indicated that Hairy modulated transcription independent of activator occupancy or SAGA co-activator occupancy (Martinez and Arnosti 2008). These previous observations raised the possibility that Hairy acts through entirely independent pathways from that employed by activators to block transcription. Our work here indicates that Hairy does indeed directly reverse chromatin marks associated with activators, and may therefore work through a

dynamic competition with these activators, undoing their positive influence on the chromatin environment that would be necessary for RNA polymerase initiation and/or elongation (Figure 2-9A). Indeed, Hairy repression is readily reversible, with genes showing reversion to an active state minutes after depletion of the overexpressed repressor (K. Kok, unpublished results).

The genome-wide analysis of repression by Hairy revealed an unexpected facet of chromatin activity and highlights the need to consider the activity of “off target” sites in generating novel elements, particularly because for Hairy at least (and likely other factors that employ the same cellular machinery) they are “shovel ready” and not constrained by complex *cis*-regulatory grammar. Metazoan transcription factors typically interact with thousands of discrete sites in the genome, but only a small subset of these interactions correlate with observable effects on gene expression. In this study, we combined analysis of gene expression and chromatin dynamics in a way that allowed us to attribute effects directly to the induction of Hairy, inferences that would be difficult with a loss-of-function assay due to kinetics of depletion and secondary effects. In contrast, many other genome-wide data sets provide a static snapshot of the extant chromatin landscape or track complex changes through development, which represents the complex contributions of many activators and repressors. Previous studies have noted the presence of detectable but lowly-occupied sites, which have been suggested to reflect non-specific, non-functional interactions that are unavoidable by-products of proteins binding to large genomes (Fisher et

al. 2012). Other studies have emphasized that transcription factors may have general nonspecific interaction with HOT sites that represent preferences for open chromatin (Gerstein et al. 2010; modENCODE Consortium et al. 2010). In general, the overall view is that whether or not these interactions are conserved, they may be of little functional consequence, and are not important for activity of GRNs (Cusanovich et al. 2014). Importantly, considering our finding that “off-target” Hairy sites still appear to regulate chromatin structure, we should fundamentally reconsider how we interpret genome-wide data sets. Frequently, an increase in H3K27 acetylation is taken as an indication that the element is an active enhancer, without further functional tests (e.g. Villar et al. 2015). Of course, correlated gene expression measurements indicate that such elements are likely to be enhancers in many cases, but genomic consideration of chromatin marking must not automatically equate changes in certain active marks with enhancers.

Our study provides a new perspective on these previous observations, in that essentially trivial biological interactions may have consequences in evolutionary time. We show that Hairy is engaged apparently in errant targeting of chromatin on many loci during the period when it is expressed, and demonstrate that in many cases, little distinguishes the types of chromatin effects observed on functionally repressed genes targets compared to “non-functional” interactions on other loci (Figure 2-9A-B). Thus, unlike an earlier model for Hairy action, in which the protein is active only when embedded in a previously active enhancer (Nibu

et al. 2001), our work demonstrates that Hairy is able to mediate biochemical activities in most bound regions, indicating that there is little context necessary for the protein to function. Therefore, Hairy may be relatively nonselective about where it can attract chromatin-modifying agents across the genome. Much molecular biology research has emphasized the high degree of cooperativity necessary for metazoan transcription factors to work well. Enhanceosomes, patterning elements and other enhancers give aberrant readouts if correct stoichiometries and spacings are not respected. These findings suggest that random individual sites are less likely to generate a suitable transcriptional readout. At least for repressors such as Hairy, the demands for generating biochemical activity are lower than anticipated, indicating that enhancers may have a lower threshold for formation that we might have expected. Although some of the targeted genes that are not transcriptionally affected may represent ectopic binding events of the induced Hairy protein, most sites are found in ChIP analysis of endogenous Hairy. The unresponsive genes may in some cases represent later targets of Hairy, may be already repressed by endogenous Hairy, or might may have responses too small to measure in this system, however it is likely that there are hundreds of changed chromatin regions that not formally part of the functional Hairy GRN. Thus, a large fraction of the genomic interactions are likely to be with regions that are not strongly selected on an evolutionary timescale. As long as the induced chromatin changes are inconsequential, these effects will not be selected against during genomic evolution. This biochemical activity, however, may provide a unique molecular exaptation to generate novel

edges between nodes of a standing GRN (Figure 2-9B). Most enhancers involve the combined action of transcriptional activators and repressors, thus errant targeting may facilitate formation of new modules with gain of a few activator binding sites (Gould and Vrba 1982) (Figure 2-9C).

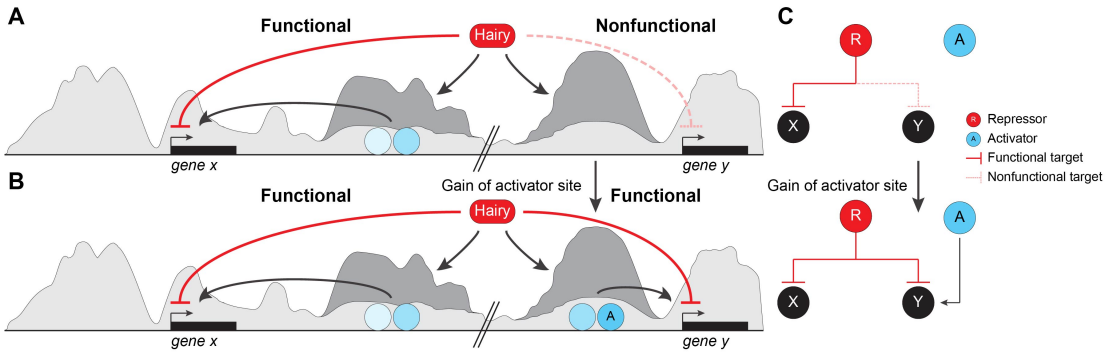
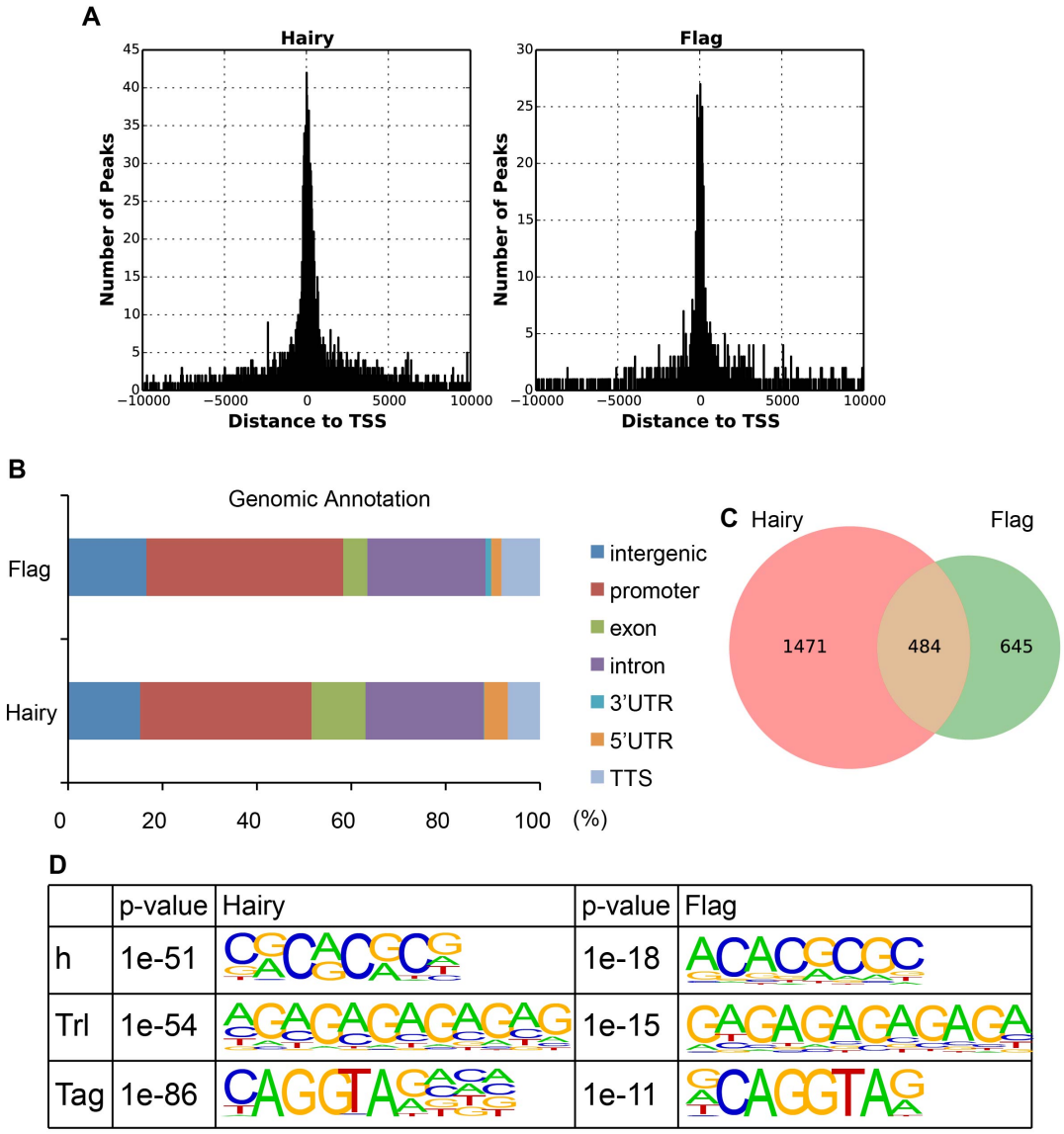


Figure 2-9 Pervasive biochemical activities on "off-target" loci may represent molecular exaptations that generate novel edges between nodes of a standing GRN. Functional and nonfunctional interactions of Hairy with chromatin are depicted. **(A)** Hairy repression of target genes results in loss of active histone marks such as H4Ac, H3K27Ac, and H3K4me1 (dark gray peaks; *gene x*). Hairy interacts with many other nonfunctional targets where it carries out biochemical activities similar to those seen on transcriptionally controlled loci (*gene y*). The latter chromatin changes are inconsequential and unlikely to be evolutionarily selected. **(B)** Gain of activator sites in a region of Hairy-modified chromatin may generate an on/off switch and result in functional targeting. **(C)** Schematic representation of cooption of Hairy physical interaction into modified GRN.



D

	p-value	Hairy	p-value	Flag
h	1e-51		1e-18	
Trl	1e-54		1e-15	
Tag	1e-86		1e-11	

Figure 2-1-1 Similarity between binding of endogenous Hairy and overexpressed Hairy protein. **(A)** Similarity of promoter proximities. Histogram shows the global distribution of Hairy peaks around TSS identified from ChIP-seq of induced Flag tagged Hairy protein (right panel) and by previous ChIP-chip detection of endogenous Hairy binding (left panel) (MacArthur et al. 2009). **(B)** Genomic annotation of peaks show similar binding distributions on genic and intergenic regions. **(C)** Area-proportional Venn diagram showing significant overlap between endogenous and induced Hairy binding (p=2.15e-159). **(D)** *De novo* motif analysis reveals similar motifs enriched under peaks of both data, including canonical Hairy binding site (CACGCG). We used the ChIP-chip data from MacArthur et al. 2009 for our analysis because the Flag epitope gave low signals overall, although high-confidence functional targets such as *ftz, Impl2, odd, h, 18w, wg, tup, pros, nht,* and *en* were found.

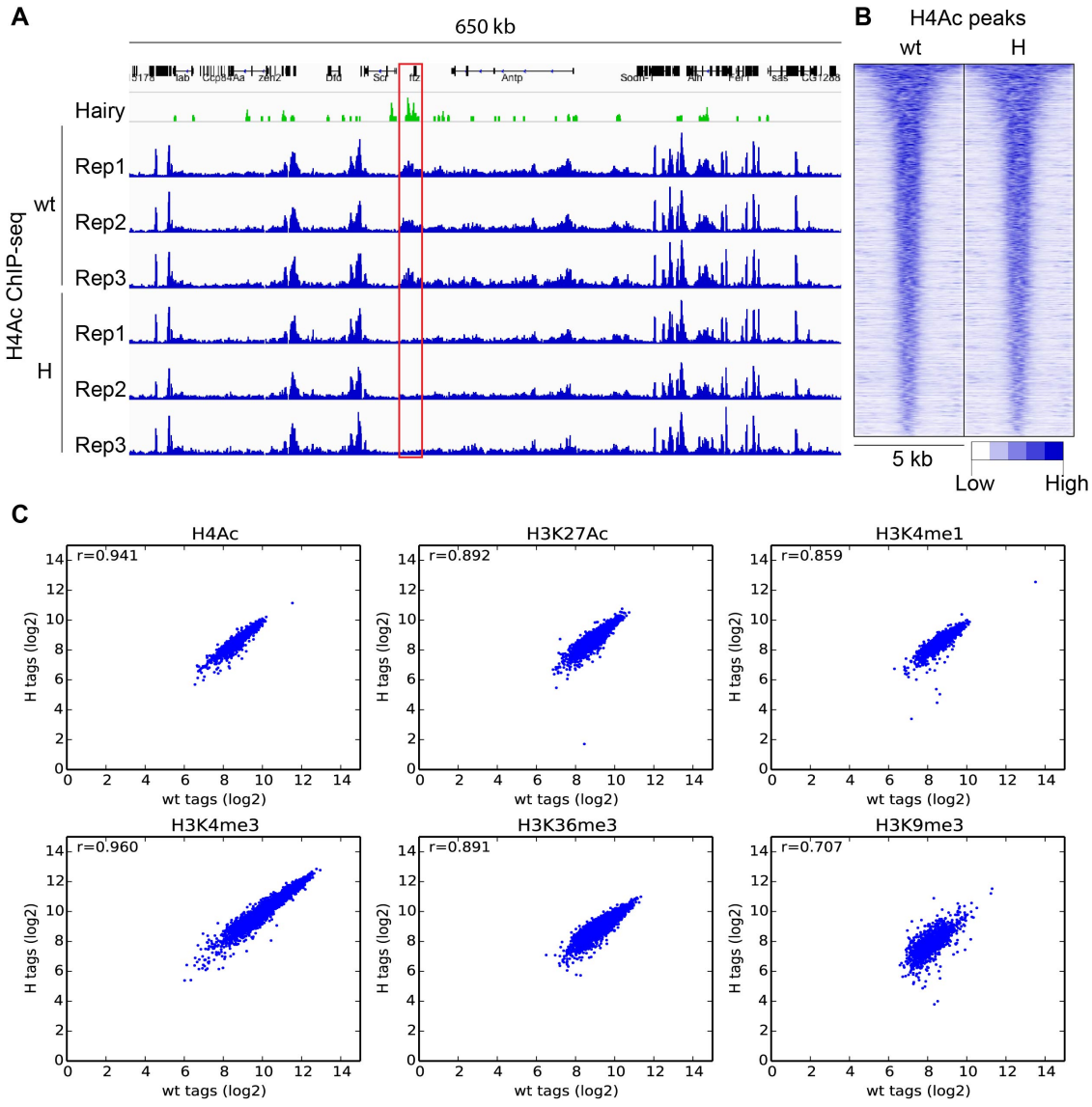


Figure 2-2-1 ChIP-seq reproducibility of biological replicates and variation between wild-type (wt) and transgenic embryos (H). (A) Specific reduction of H4Ac signal at *ftz* locus (red box) in three biological replicates after induction of Hairy (H). **(B)** H4Ac peaks were not altered globally in genome by Hairy expression. Heatmap show 5 kb window centered on called H4Ac peaks, ranked by peak height. **(C)** Measured global chromatin features were similar in wt and H samples, indicating that Hairy does not affect the majority of chromatin features throughout the genome. Scatter plots indicate the correlation (r=Pearson's correlation coefficient) between wt and H embryos for H4Ac, H3K27Ac, H3K4me1, H3K4me3, H3K36me3 and H3K9me3 marks. Each dot represents a peak. ChIP-seq read counts on the axis are transformed to log2 base.

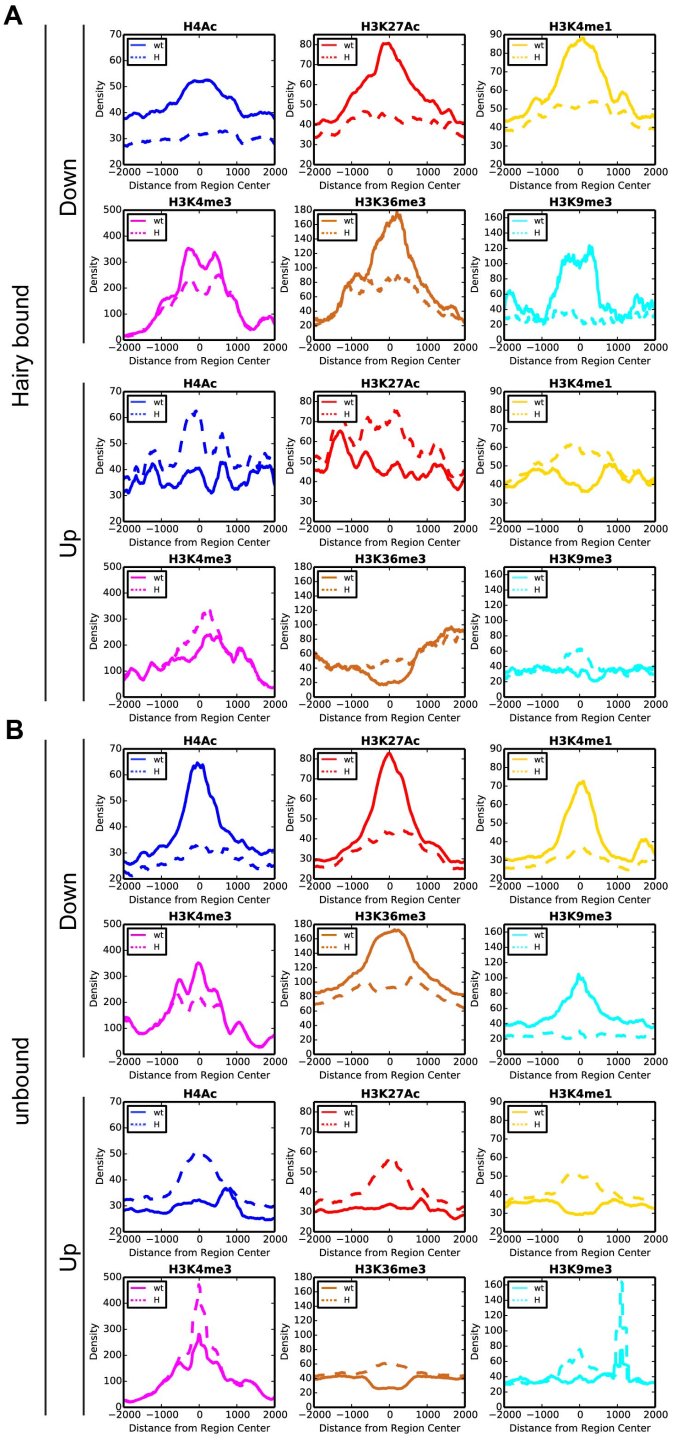


Figure 2-3-1 Distinct chromatin profiles associated with direct and indirect Hairy targeted loci. Histograms show the distribution of averaged ChIP-seq signals in a window of 4 kb centered on differentially changed regions associated with Hairy bound **(A)** and unbound **(B)** genes in the wild-type (wt, solid lines) and Hairy induced (H, dashed lines) embryos for H4Ac, H3K27Ac, H3K4me1, H3K4me3, H3K36me3 and H3K9me3.

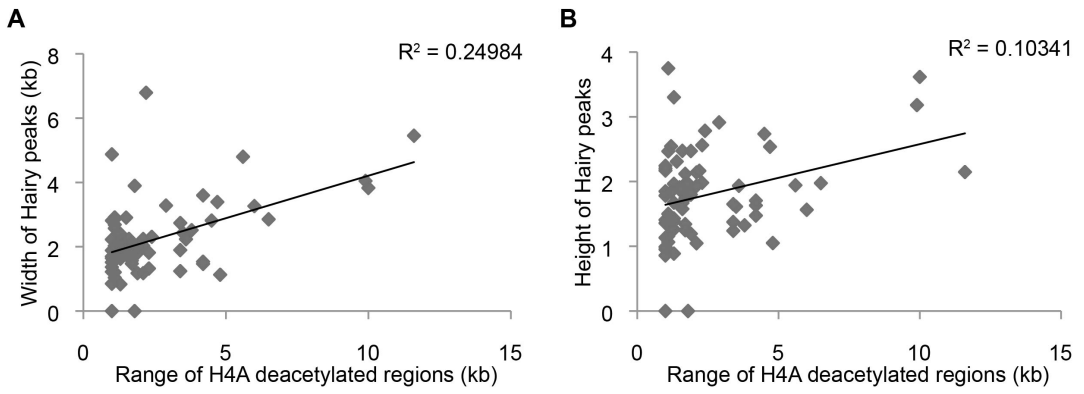


Figure 2-3-2 Little correlation between height of Hairy peaks or width of Hairy-bound region and extent of H4 deacetylation blocks and width (A) or height (B) of Hairy peaks. Other marks also exhibited little correlation between Hairy peak width and height and range of chromatin alterations (not shown).

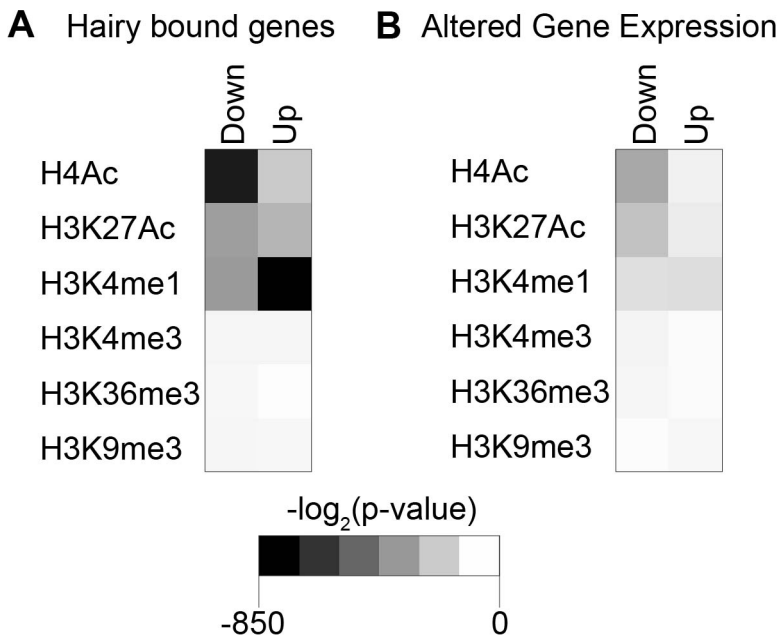


Figure 2-4-1 Significance of individual histone modifications associated with Hairy bound genes and transcriptionally regulated genes. (A) Strongest link between loss of H4Ac, gain of H3K4me1, and presence of Hairy on genes. **(B)** Transcriptionally repressed genes associated with loss of H4Ac, H3K27Ac, and gain or loss of H3K4me1.

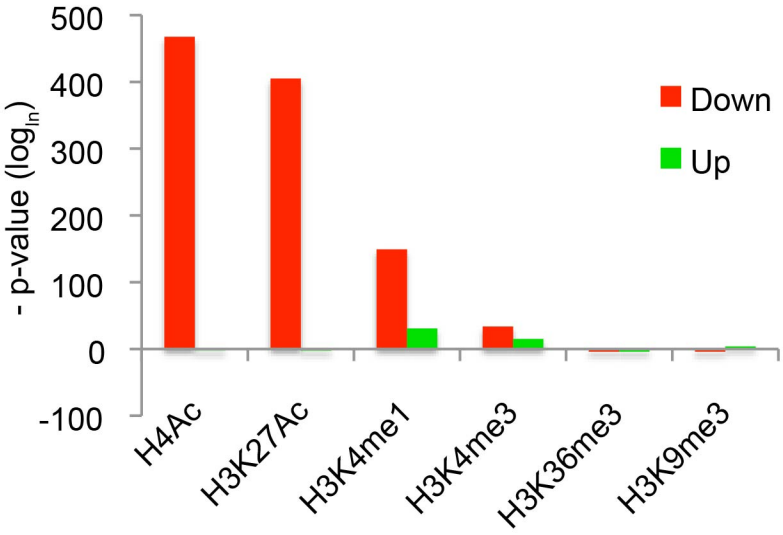


Figure 2-4-2 Strong correlation between the presence of Hairy binding and chromatin alterations on specific chromatin blocks. (A) Reduced H4Ac, H3K27Ac, and H3K4me1 significantly associated with Hairy binding. Hairy bound regions overlapped with chromatin blocks by at least 1 bp. y-axis indicates –p-value (log_{ln}).

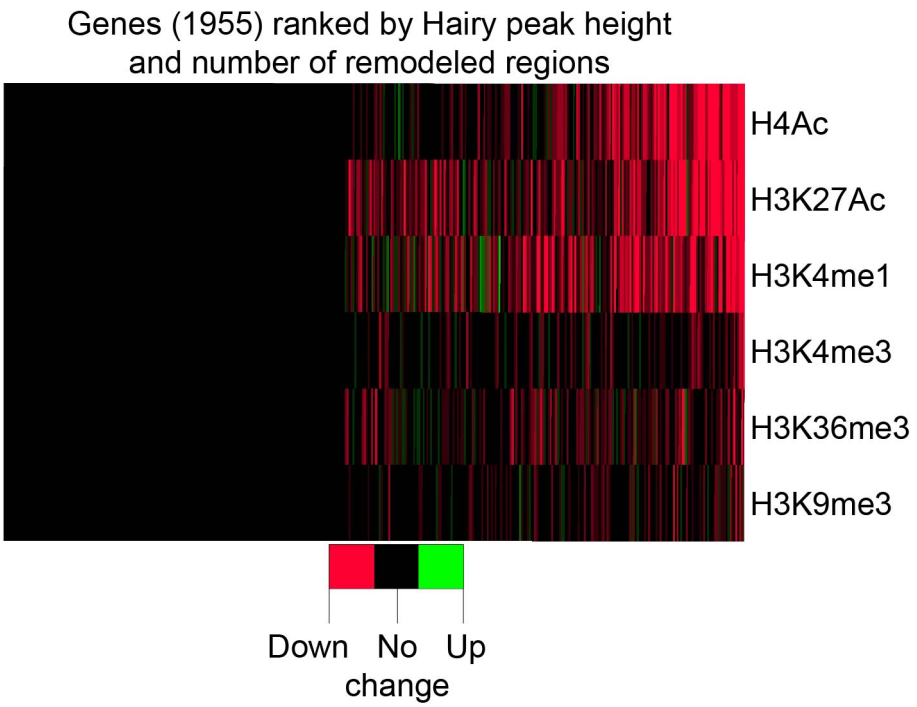


Figure 2-5-1 Global association of Hairy binding with histone mark alterations. Changes in histone marks, predominantly reductions, were detected for more than half of the Hairy bound genes. Genes were divided into two groups; no detectable histone mark changes vs. at least one change, and then ranked (right to left) by height of Hairy peaks and total number of changes in histone marks. Differential changed regions of histone marks and Hairy peaks were assigned to genes with the closest TSS.

5. Materials and methods

Plasmid construction: The heat-inducible *hairy* gene was created by introducing a multiple cloning site (MCS) containing Kozak sequence, initiator ATG and HindIII/BglII sites into the 5' portion of the *hairy* ORF in the pCaSpeR-hsh using EcoRI/BstEII sites as described previously (Li and Arnosti 2011). 400 bp of upstream promoter, 5' UTR, Kozak sequence, initiator ATG, HindIII/BglII sites, coding sequence and entire *hsp70* 3' UTR from the modified pCaSpeR-hsh were amplified using 5' and 3' primers with AgeI/KpnI sites and subcloned to the modified pattB vector (Sayal et al. 2011). Oligonucleotides with sequence encoding the double Flag epitope, as described in (Zhang and Arnosti 2011), was inserted 5' of the coding sequence after the ATG using HindII/BglII sites, so that Hairy protein was expressed with the double Flag tag at the N terminus.

Embryo collection, in situ hybridization and antibody staining of Drosophila embryos: For chromatin analysis 2-3.5 hr. embryos were collected and 20 min heat-shock treated for induction of transgenes as described previously (Li and Arnosti 2011). We treated the wild-type embryos similar to embryos carrying inducible transgene to control for possible nonspecific effects of heat shock. Heat shock alone has no effect on the expression or chromatin patterns (data not shown). For analysis of gene expression by *in situ* hybridization, embryos were fixed and stained using anti-digoxigenin-UTP-labeled RNA probe for *ftz* as described previously (Struffi 2004).

Quantitative reverse transcriptase PCR analysis: Total RNA from embryos was purified using RNeasy columns (Qiagen), and reversed transcribed using a High Capacity cDNA Reverse Transcription Kit from Invitrogen/Applied Biosystems. The cDNA was then analyzed by real-time PCR using the primer pairs located at transcription units. Data was normalized to *act5c.* Values for wild-type embryos were set to 1; results represent the average of 2-8 biological replicates. Statistical significance was tested using Student's T test and $p<0.05$. Amplicons were designed using Primer Express and Primer-BLAST.

Expression profiling analysis: Total RNA from 2-3 hr embryos was purified using RNeasy columns (Qiagen). Samples were amplified and labeled using the Quick AMP Labeling kit (Agilent) and hybridized to 8x15K Customized Drosophila Genome Oligo Microarrays (Agilent) according to the manufacturer's instructions. Slide image data was quantified using Agilent's Feature Extraction software. Four biological replicates were performed for each sample. Differential gene expression analysis was performed with the GeneSpring program (Agilent). Functional annotation of down- and up-regulated genes was done using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Dennis et al. 2003). Differentially regulated gene symbols and their fold changes are listed in Supplementary file 5.

Chromatin Immunoprecipitation: Heat shocks and ChIPs were performed as described previously (Li and Arnosti 2011), with the exceptions that embryos were sonicated for a total of 20 times using a Branson sonicator in 1 ml of sonication buffer. After precipitation of chromatin-antibody complexes, protein A beads were washed twice with low-salt buffer, once with high-salt buffer, once with LiCl buffer and twice with Tris-EDTA. We used the following antibodies: rabbit IgG (5 µl, Santa Cruz Biotechnology), rabbit anti-H3 (1 µl, Abcam), rabbit anti-acetyl H4 (1 µl, Upstate), rabbit anti-acetyl H3K27 (1 µl, Abcam), rabbit anti-monomethyl H3K4 (1 µl, Abcam), rabbit anti-trimethyl H3K4 (1 µl, Abcam), rabbit anti-trimethyl H3K36 (2 µl, Abcam), rabbit anti-trimethyl H3K9 (3 µl, Abcam), rabbit anti-Flag (5 µl, Sigma), rabbit anti-Rpb3 (5 µl, gift from Carla Margulies, LMU University of Munich).

ChIP-seq: Libraries. DNA from chromatin immunoprecipitation (10 ng) was adapter-ligated and PCR amplified (18 cycles) as described in (Ford et al. 2014). DNA ligated to the adapter was size selected for 300-500 bp. Illumina HiSeq single-end reads were checked using FastQC and HOMER for sequence quality, base sequence and GC content, sequence duplication, sequence bias, overrepresented sequences and Kmer content. Reads were aligned to genome (BDGP 5.70) with Bowtie version 1.0.0 using -m 1 --best parameters. Tags that only mapped uniquely to the genome were considered for further analysis. Summary of tags generated is shown on Supplementary file 6. ChIP-Seq experiments were visualized as custom tracks using Integrative Genomics

Viewer (Broad Institute). Total uniquely mapped tags were normalized to 10 million reads to generate tracks. y-axis values shown in all figures use the same scale for an individual measurement of each histone modification in the individual panels. For reasons of clarity, scales can vary between different panels.

Mapping differential regions. We detected the regions where chromatin states are changed upon induction of Hairy by comparing the level of histone marks at particular genomic locations. Differentially changed genomic regions were identified using the diffReps program (Shen et al. 2013), which uses a sliding window approach to scan the genome and find regions showing read count differences. Default window size with --nsd broad --meth nb parameters was used for the analysis. For downstream analysis, we used regions with $p<0.05$ and fold change (log2)>0.4 or fold change (log2)<-0.4. Input was sequenced from nontransgenic (wt) and Hairy overexpressing embryos and used as background.

HOMER (Hypergeometric Optimization of Motif EnRichment) was used for peak finding and downstream data analysis (Heinz et al. 2010).

Identification of ChIP-seq peaks: Using HOMER with default settings, peaks for histone marks and Flag tagged Hairy protein were identified using signals from H3 ChIP and input respectively as background.

Annotation of significantly affected regions: Regions detected by diffReps or peaks called by HOMER were associated with genes by identifying the nearest RefSeq TSS and annotated to a genomic feature such as intergenic, intron, exon etc.

Normalization of ChIP-seq tags for histograms, heatmaps, and scatter plots: We

normalized the total number of mapped tags to 10 million for each sample using HOMER so that the read densities were comparable.

Comparison of ChIP-seq experiments using histograms: ChIP-seq densities of a 4 kb window centered at affected regions detected by diffReps were determined using HOMER. The program normalizes the output histogram such that the resulting units are per bp per region with bin size of 10 bp. Plots were generated using matplotlib (Hunter 2007).

Comparison of ChIP-seq experiments using heatmaps: Data matrices were generated using HOMER by counting total tags in a 5 kb window around affected regions or peaks and normalizing to 10 million reads with bin size of 25 bp. Data was visualized using Java Tree View (Saldanha 2004).

Comparison of ChIP-seq experiments using scatter plots: Tag densities were calculated by counting the tags at regions defined by peak coordinates of the first experiment (x axis) and compared to the second experiment (y axis). Data was log2 transformed and plotted using matplotlib. Pearson's Correlation Coeffcients were calculated to determine the extent of similarity between samples.

Analysis of co-occurrence of differentially changed regions: mergePeaks program of HOMER was used to find overlapping sites between differentially changed regions of different histone marks upon Hairy induction. These regions were considered as overlapped if changed regions from each experiment share at least 1 bp. Significance of co-occurance of regions was indicated by natural log p-values using the hypergeometric distribution. Positive values signify divergence.

Linking affected regions to Hairy binding: Affected regions for chromatin marks were considered as Hairy bound if the nearest gene has at least one Hairy peak. The occupancy of the induced Hairy protein was compared to that of endogenous Hairy binding by conducting ChIP-Seq analysis using the Flag epitope on the inducible protein; a large fraction (40%) of these binding sites were also found in the ChIP-chip study (p=2.15e-159). Similar Hairy binding motifs were enriched in both data sets, indicating that the induced Hairy protein has similar targeting specificity to the endogenous protein (Figure 2-1-1).

Machine learning analysis: We used the differential changes of H4Ac, H3K27Ac, H3K4me1, H3K4me3, H3K36me3, H3K9me3 and Pol II in response to Hairy as features for our analysis here. The genomic blocks detected as significantly altered by diffReps are annotated to closest TSS. We considered four features for each of the ChIP-seq data; number of blocks linked to the same gene, range of blocks, fold change of ChIP-seq signal at blocks, and distance of blocks to closest TSS. Four features from ChIP-chip data sets of Hairy (MacArthur et al. 2009), CtBP, and Gro (Nègre et al. 2011) were used; number of peaks linked to the same gene, width of peaks, peak signal, and distance of peaks to closest TSS. In addition, expression of transcripts in wild-type embryos was included as a feature. In total, these 41 features were collected for 583 genes (241 repressed, 146 activated and 196 unaffected genes; activated and unaffected genes were grouped as nonrepressed genes) in this study. Differentially regulated genes and their fold changes are listed in Supplementary file 5 and

randomly selected unaffected genes are listed in Supplementary file 7. Important features were first identified with four feature selection algorithms (Information Gain, Symmetrical Uncertainty, Chi Square and Relief). Then, to predict genes in the repressed and not-nonrepressed categories, four classifiers (Support Vector Machine (SVM), *k*-Nearest Neighbors (KNN), Naive Bayes and Random Forests) were employed. To perform this analysis, we wrote Python and Java codes to partition our dataset into 10 parts to perform feature selection and 10-fold cross validation classification utilizing the Weka machine learning software (http://www.cs.waikato.ac.nz/ml/weka/). To increase the robustness of our results we performed 50 iterations of the above procedure and combined the predicted classes for each gene to create a new aggregate predicted class for that gene. Here we took the class that has been predicted more than 50% of the 50 iterations as the predicted class of the gene. We have applied every combination of the four feature selection algorithms and four classification algorithms to the data to obtain the optimal classification methodology for our dataset. The results of our analysis are summarized in the main text.

6. Acknowledgments

We thank Carla Margulies for the Rpb3 antisera, Shin-Han Shiu, Monique Floer, and George I. Mias for critical reading of the manuscript, John Johnston and the MSU Institute for Cyber-Enabled Research (iCER) for help with the High Performance Computing Center. We thank Samuel Daulton (Colgate University) for his help in the Machine Learning analysis of the data. We also thank the Arnosti laboratory members for useful discussions, and Sandhya Payankaulam for assistance with establishing library preparation protocols.

Competing interests: The authors declare that no competing interests exist.

APPENDIX

APPENDIX

Analysis of chromatin effects induced by Hairy

Small-scale analysis of chromatin changes

The genome-wide studies described in Chapter II were initiated by small-scale analysis on a number of genes (Figure 2-A-1). The antibodies against several histone marks were verified by ChIP-qPCR on several genes identified as Hairy targets. Hairy induction resulted in deacetylation at H4 and H3K27 on *ftz, comm, comm2, 18w* and *odd*. Only a modest change was observed on another H3 acetylation mark (H3K9,14). Histone acetylation levels did not change on on *esg, erm,* and *hlhm7* genes. Methylation marks associated with enhancers (H3K4me1), promoters (H3K4me3) and gene bodies (H3K36me3) were investigated to get insights into the possible roles of KDMs in repression. A modest decrease at H3K4me1 was found in *ftz* and *18w*. Promoter mark H3K4me3 was not affected significantly. A significant reduction in H3K36me3 was detected in transcribed regions of *18w* and *odd*. H3K27me3 levels were explored to determine whether Hairy modulates repression marks. Robust levels of H3K27 trimethylation were detected on *ftz*, *odd* and *erm*. Only a slight reduction on *ftz* locus was detected in response to Hairy. To understand the effects of repression by Hairy on chromatin structure, H3, H2A.Z and H1 levels were measured. In overall, chromatin structure was not extensively influenced by repression in terms of histone density, presence of a histone variant H2A.Z

(associated with active regions) and histone H1 (associated with inactive genes). Finally, loss of Pol II occupancy was observed only on *ftz*, *18w* and *hlhm7*.

Dynamic recovery of chromatin changes induced by Hairy

Chromatin marks play key roles in establishing epigenetic memories. To understand if chromatin alterations regulated by Hairy persist long-term in embryo development, levels of *hairy* mRNA were measured 5-20 minutes after 20 minutes overexpression. *hairy* mRNA appears to be degraded within 20 minutes (Figure 2-A-2). Considering half-life (~30 min) of Hairy protein, Hairy expression is expected to drop to endogenous level in less than an hour. To detect the dynamics of histone deacetylation mediated by Hairy, acetylation marks were measured right after Hairy induction (2-3:30) and 90 minutes recovery time (3:30-5) (Figure 2-A-3). As expected, H4Ac and H3K27Ac were reduced at enhancers of *ftz*, *18w* and *odd.* However, these marks came back to endogenous levels after the recovery time. Therefore, Hairy induces dynamic transient changes in histone acetylation.

Heat shock alone does not affect chromatin

Studies in this thesis are based on inducible transgenic genes, which are driven by a heat shock promoter. In order to test if heat shock alone affects the chromatin marks, I measured H4Ac, H3K27Ac, H3K36me3 and H3 levels from wild-type embryos with and without heat shock (Figure 2-A-4). It appears that chromatin structure is not altered by heat shock alone, at least for these marks at enhancer regions of *ftz*, *18w*, and *comm*, and gene body region of *odd*, where

Hairy decrease active marks significantly. These results show that chromatin effects described in this study are specific to Hairy overexpression.

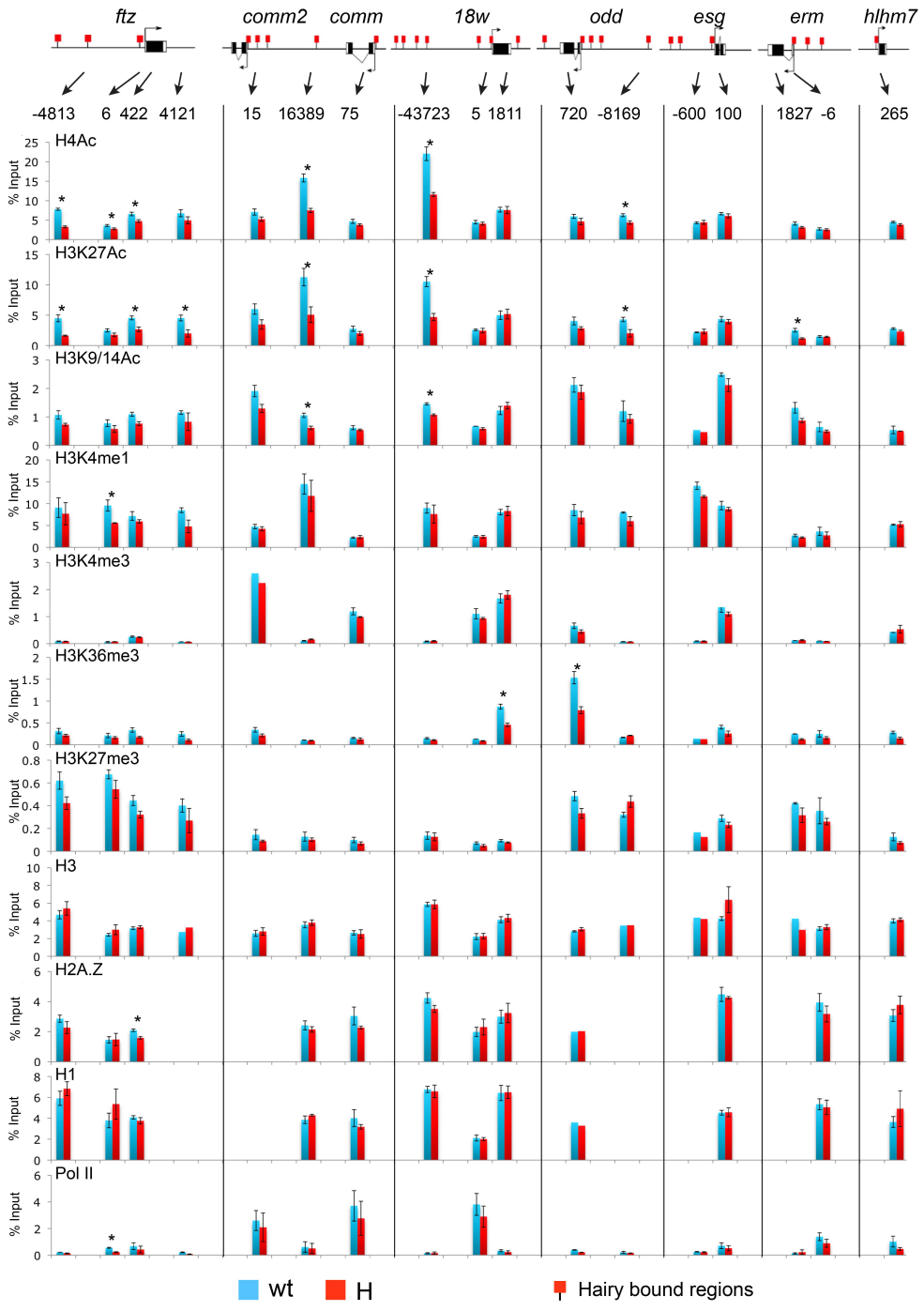


Figure 2-A-1 Small-scale survey of various chromatin signatures on transcriptionally repressed genes by Hairy induction. Changes in chromatin marks associated with active loci (H4Ac), promoters (H3K27ac, H3K9/14Ac, H3K4me3), gene bodies (H3K36me3), enhancers (H3K4me1 and H3K27ac), repressed genes (H3K27me3) in Hairy mediated repression, occupancy of histone H3, H1 and H2A.Z and total Pol II were measured by chromatin immunoprecipitation (ChIP) before (blue bar) and after (red bar) Hairy induction.

X-axis indicates the loci detected by specific primers relative to TSS, shown above the bar graphs. Y-axis shows ChIP signals as percentages of input. *p < 0.05 by Student's t test is used for results represent at least three biological replicates; error bars show standard errors. Gene structures are shown at the top. Red boxes denote Hairy bound regions based on ChIP-chip data.

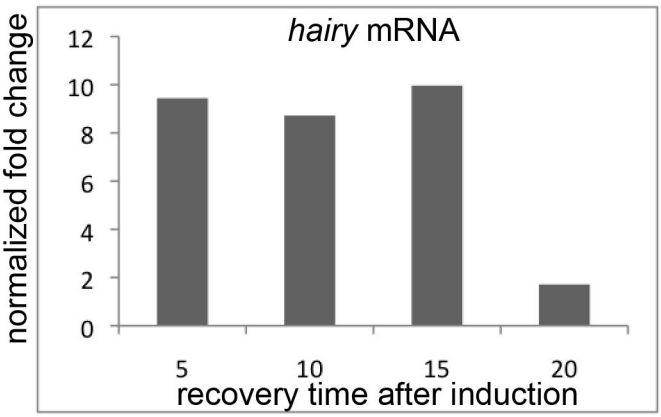


Figure 2-A-2 Decay of *hairy* mRNA. RT-qPCR analysis shows ~10 fold increase in mRNA levels after 5 minutes induction. mRNA level drops to endogenous level in ~20 minutes.

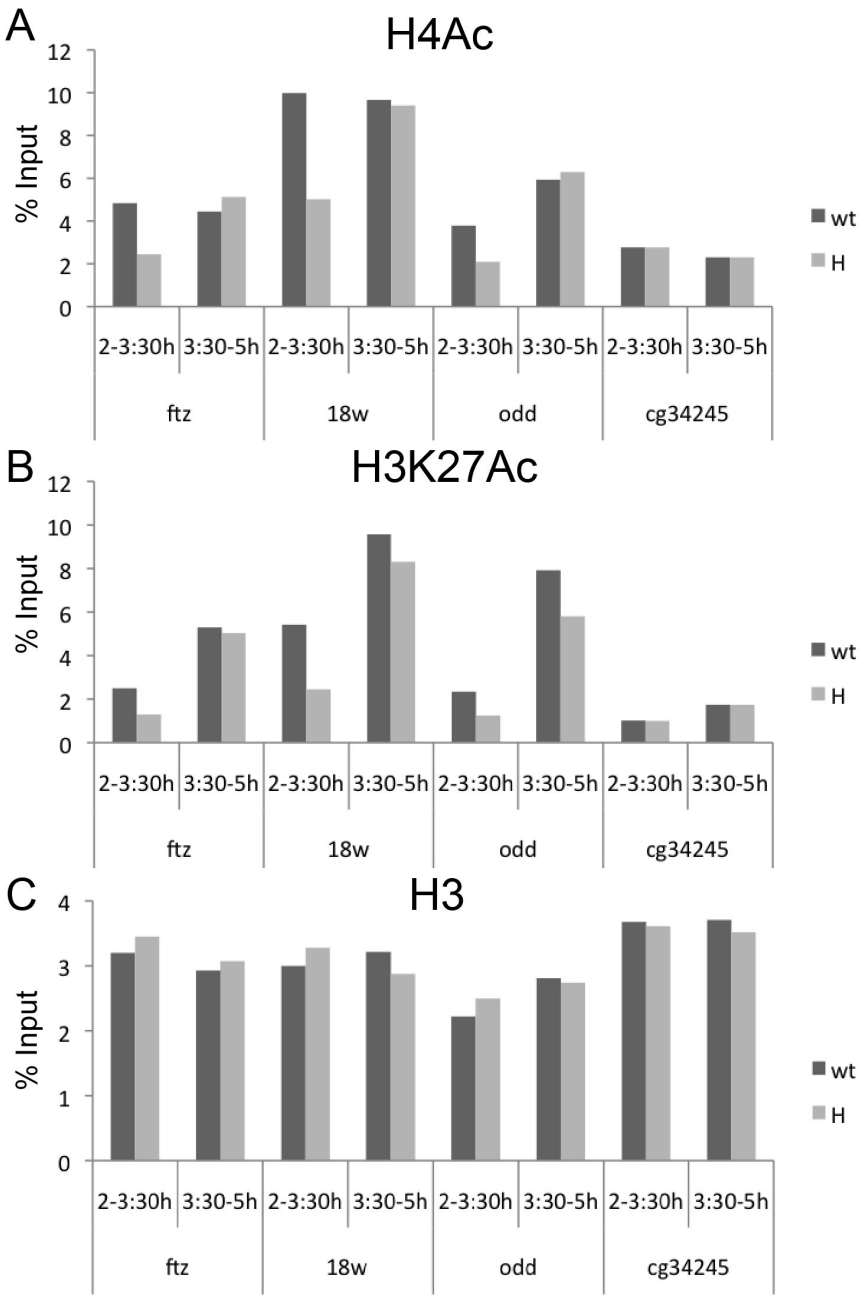


Figure 2-A-3 Fast recovery of chromatin changes induced by Hairy. Active marks **(A)** H3K27Ac and **(B)** H4Ac were reduced at enhancer regions of *ftz*, *18w* and *odd* immediately after Hairy induction (2-3:30h). These marks quickly came back in less than 90 min (3:30-5h). **(C)** No change at histone density (H3) was detected. An inactive gene, *cg34245,* was used as a negative control.

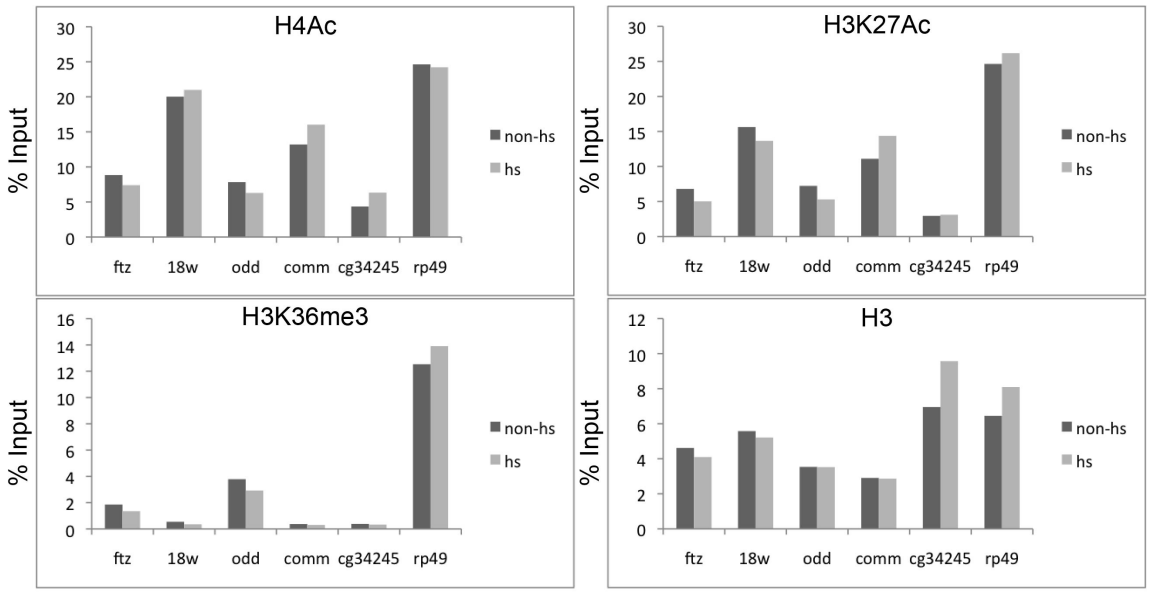


Figure 2-A-4 Heat shock did not affect chromatin signatures. Wild-type embryos were exposed to heat shock for 20 minutes. No effect was detected at H4Ac, H3K27Ac, H3K36me3 and H3 on a selected number of loci, including a ribosomal gene *rp49* and a silent gene *cg34245* in embryos.

List of Files

Supplementary File 1 GO analysis of down-regulated genes.

Supplementary File 2 GO analysis of up-regulated genes.

Supplementary File 3 Comparison of ChIP-seq signal around differentially changed histone marks using Kolmogorov Smirnov test.

Supplementary File 4 Feature ranking in predicting gene expression.

Supplementary File 5 Diffentially regulated genes identified by microarray analysis.

Supplementary File 6 Summary of sequencing reads.

Supplementary File 7 Randomly selected unaffected genes for machine learning analysis.

REFERENCES

REFERENCES

Adelman K, Wei W, Ardehali MB, Werner J, Zhu B, Reinberg D, Lis JT. 2006. Drosophila Paf1 modulates chromatin structure at actively transcribed genes. *Molecular and Cellular Biology* **26**: 250–260.

Ardehali MB, Yao J, Adelman K, Fuda NJ, Petesch SJ, Webb WW, Lis JT. 2009. Spt6 enhances the elongation rate of RNA polymerase II in vivo. *EMBO J* **28**: 1067–1077.

Arnosti DN, Kulkarni MM. 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* **94**: 890–898.

Barolo S, Levine M. 1997. hairy mediates dominant repression in the Drosophila embryo. *EMBO J* **16**: 2883–2891.

Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823–837.

Bianchi-Frias D, Orian A, Delrow JJ, Vazquez J, Rosales-Nieves AE, Parkhurst SM. 2004. Hairy Transcriptional Repression Targets and Cofactor Recruitment in Drosophila. *Plos Biol* **2**: e178.

Cai HN, Arnosti DN, Levine M. 1996. Long-range repression in the Drosophila embryo. *Proc Natl Acad Sci USA* **93**: 9309–9314.

Chantalat S, Depaux A, Hery P, Barral S, Thuret JY, Dimitrov S, Gerard M. 2011. Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome Research* **21**: 1426–1437.

Courey AJ, Jia S. 2001. Transcriptional repression: the long and the short of it. *Genes & Development* **15**: 2786–2796.

Crona F, Dahlberg O, Lundberg LE, Larsson J, Mannervik M. 2013. Gene regulation by the lysine demethylase KDM4A in Drosophila. *Developmental Biology* **373**: 453–463.

Cusanovich DA, Pavlovic B, Pritchard JK, Gilad Y. 2014. The Functional Consequences of Variation in Transcription Factor Binding ed. Y. Pilpel. *PLoS Genet* **10**: e1004226.

Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated

Discovery. *Genome Biol* **4**: P3.

Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmann R, MacArthur S, Thomas S, Stamatoyannopoulos JA, Eisen MB, et al. 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. *Proceedings of the National Academy of Sciences* **109**: 21330–21335.

Ford E, Nikopoulou C, Kokkalis A, Thanos D. 2014. A method for generating highly multiplexed ChIP-seq libraries. *BMC Res Notes* **7**: 312.

Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al. 2010. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* **330**: 1775–1787.

Gould SJ, Vrba ES. 1982. Exaptation-A Missing Term in the Science of Form. *Paleobiology* **8**: 4–15.

Gray S, Levine M. 1996. Short-range transcriptional repressors mediate both quenching and direct repression within complex loci in Drosophila. *Genes & Development* **10**: 700–710.

Gupte R, Muse GW, Chinenov Y, Adelman K, Rogatsky I. 2013. Glucocorticoid receptor represses proinflammatory genes at distinct steps of the transcription cycle. *Proceedings of the National Academy of Sciences* **110**: 14616–14621.

Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB. 2008. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. *PLoS Genet* **4**: e1000106.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**: 576–589.

Hunter JD. 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9**: 0090–0095.

Ish-Horowicz D, Pinchin SM. 1987. Pattern abnormalities induced by ectopic expression of the Drosophila gene hairy are associated with repression of ftz transcription. *Cell* **51**: 405–415.

Kageyama R, Ohtsuka T, Kobayashi T. 2007. The Hes gene family: repressors and oscillators that orchestrate embryogenesis. *Development* **134**: 1243–1251.

Khila A, Abouheif E, Rowe L. 2009. Evolution of a novel appendage ground plan in water striders is driven by changes in the Hox gene Ultrabithorax. *PLoS Genet* **5**: e1000583.

Kim T, Buratowski S. 2007. Two Saccharomyces cerevisiae JmjC domain proteins demethylate histone H3 Lys36 in transcribed regions to promote elongation. *J Biol Chem* **282**: 20827–20835.

Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, Dickson BJ, Stark A. 2014. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**: 91–95.

Lagha M, Bothma JP, Esposito E, Ng S, Stefanik L, Tsui C, Johnston J, Chen K, Gilmour DS, Zeitlinger J, et al. 2013. Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. *Cell* **153**: 976–987.

Li LM, Arnosti DN. 2011. Long- and Short-Range Transcriptional Repressors Induce Distinct Chromatin States on Repressed Genes. *Curr Biol* **21**: 406–412.

MacArthur S, Li X-Y, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keränen SVE, et al. 2009. Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol* **10**: R80.

Martinez CA, Arnosti DN. 2008. Spreading of a Corepressor Linked to Action of Long-Range Repressor Hairy. *Molecular and Cellular Biology* **28**: 2792–2802.

McManus S, Ebert A, Salvagiotto G, Medvedovic J, Sun Q, Tamir I, Jaritz M, Tagoh H, Busslinger M. 2011. The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. *EMBO J* **30**: 2388–2404.

modENCODE Consortium, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Nègre N, Eaton ML, Landolin JM, Bristow CA, Ma L, et al. 2010. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**: 1787–1797.

Nelson CE, Hersh BM, Carroll SB. 2004. The regulatory content of intergenic DNA shapes genome architecture. *Genome Biol* **5**: R25.

Nègre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R, et al. 2011. A cis-regulatory map of the Drosophila genome. *Nature* **471**: 527–531.

Nibu Y, Zhang H, Bajor E, Barolo S, Small S, Levine M. 1998. dCtBP mediates transcriptional repression by Knirps, Krüppel and Snail in the Drosophila

embryo. *EMBO J* **17**: 7009–7020.

Nibu Y, Zhang H, Levine M. 2001. Local action of long-range repressors in the Drosophila embryo. *EMBO J* **20**: 2246–2253.

Paroush Z, Finley RL, Kidd T, Wainwright SM, Ingham PW, Brent R, Ish-Horowicz D. 1994. Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79**: 805–815.

Pavlopoulos A, Kontarakis Z, Liubicich DM, Serano JM, Akam M, Patel NH, Averof M. 2009. Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. *Proceedings of the National Academy of Sciences* **106**: 13897–13902.

Payankaulam S, Arnosti DN. 2009. Groucho corepressor functions as a cofactor for the Knirps short-range transcriptional repressor. *Proceedings of the National Academy of Sciences* **106**: 17314–17319.

Perissi V, Jepsen K, Glass CK, Rosenfeld MG. 2010. Deconstructing repression: evolving models of co-repressor action. *Nat Rev Genet* **11**: 109–123.

Poortinga G, Watanabe M, Parkhurst SM. 1998. Drosophila CtBP: a Hairy-interacting protein required for embryonic segmentation and hairy-mediated transcriptional repression. *EMBO J* **17**: 2067–2078.

Saldanha AJ. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**: 3246–3248.

Sayal R, Ryu S-M, Arnosti DN. 2011. Optimization of reporter gene architecture for quantitative measurements of gene expression in the Drosophila embryo. *Fly (Austin)* **5**: 47–52.

Shen L, Shao N-Y, Liu X, Maze I, Feng J, Nestler EJ. 2013. diffReps: Detecting Differential Chromatin Modification Sites from ChIP-seq Data with Biological Replicates ed. R. Mantovani. *PLoS ONE* **8**: e65598.

Squazzo SL, O'Geen H, Komashko VM, Krig SR, Jin VX, Jang S-W, Margueron R, Reinberg D, Green R, Farnham PJ. 2006. Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome Research* **16**: 890–900.

Struffi P. 2004. Quantitative contributions of CtBP-dependent and -independent repression activities of Knirps. *Development* **131**: 2419–2429.

Tanaka K, Barmina O, Sanders LE, Arbeitman MN, Kopp A. 2011. Evolution of sex-specific traits through changes in HOX-dependent doublesex expression. *Plos Biol* **9**: e1001131.

Vakoc CR, Mandat SA, Olenchock BA, Blobel GA. 2005. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Molecular Cell* **19**: 381–391.

Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, et al. 2015. Enhancer Evolution across 20 Mammalian Species. *Cell* **160**: 554–566.

Wang X, Lee C, Gilmour DS, Gergen JP. 2007. Transcription elongation controls cell fate specification in the Drosophila embryo. *Genes & Development* **21**: 1031–1036.

Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. 2012. Dynamic Transformations of Genome-wide Epigenetic Marking and Transcriptional Control Establish T Cell Identity. *Cell* **149**: 467–482.

Zhang YW, Arnosti DN. 2011. Conserved catalytic and C-terminal regulatory domains of the C-terminal binding protein corepressor fine-tune the transcriptional response in development. *Molecular and Cellular Biology* **31**: 375–384.

Zhou VW, Goren A, Bernstein BE. 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12**: 7–18.

CHAPTER III

Contributions of CtBP and Groucho corepressors to Hairy-mediated repression in the Drosophila embryo[4]

1. Abstract

Context- specific switching between transcriptional activation and repression is a central process in development, and is orchestrated by repressors that effect changes in chromatin structure. The nonpermissive chromatin environment associated with gene silencing is created by multi-subunit protein complexes. DNA binding transcription factors associate with scaffolding proteins, histone modifying enzymes, and chromatin remodeling complexes. The non-DNA-binding components of such complexes, collectively termed corepressors, interact with transcription factors by binding to specific surface motifs. The combination of multiple enzymatic activities presumably provides desirable quantitative or qualitative effects on transcriptional targets, although this topic is relatively unexplored. One such transcriptional repressor with the capacity to recruit multiple corepressors is the Drosophila Hairy protein, a member of the Hairy/Enhancer of Split (HES) family with the ability to silence regulatory sites over 1 kb from its binding elements through a long-range mechanism. Although the functional relevance of these biochemical interactions remains controversial, Hairy can bind the Groucho, CtBP and Sir2 co-repressors, all of which are associated with histone deacetylase activity. In particular, the importance of interaction between Hairy and the CtBP corepressor, a factor that has been

clearly linked to short-range repression mechanisms, has been called into question. These conclusions were based on analysis on a few gene targets, however. In order to identify specific functions of these conserved corepressors in Hairy mediated repression, we assessed chromatin changes on a genome-wide scale by testing wild-type and a mutant version of Hairy that lacks a specific motif required for interaction with CtBP. The mutation in the CtBP interaction domain partially impairs repression ability of Hairy on some genes and completely abolishes its activity on others. These results support the model that Hairy acts in a context-dependent manner utilizing different repression mechanisms, and CtBP contributes to repression in a quantitative and gene-specific fashion.

Key words

transcription, repression, chromatin, corepressor

[4]Chapter III is presented in the form of a manuscript to be submitted for publication as: Kok K, and Arnosti DN. Contributions of CtBP and Groucho corepressors to Hairy-mediated repression in the Drosophila embryo.

2. Introduction

Transcriptional repressors including Hairy interact with multiple cofactors with distinct enzymatic functions (Payankaulam et al. 2010). The global importance of different cofactors in repression and chromatin changes is poorly understood. Hairy interacts with three corepressors: the Sir2 deacetylase binds the Hairy DNA binding domain, Groucho (Gro) binds the C-terminal WRPW motif, and C-terminal binding protein (CtBP) binds a PLSLV motif near the C-terminus (Rosenberg and Parkhurst 2002; Paroush et al. 1994; Poortinga et al. 1998).

The identification of Gro as a regulator of transcriptional repression marked an important step for the understanding of corepressor functions in eukaryotes, since maternal and zygotic functions of Gro are essential in a wide variety of developmental processes, and Gro homologs are present in yeasts, plants and most metazoans (Fisher et al. 1996; Grbavec and Stifani 1996; Paroush et al. 1994). However, identification of the transcriptional repression patterns mediated by Gro has remained incomplete despite its widespread importance in development. A variety of mechanisms have been proposed based on genetic, molecular, and biochemical data; these include histone deacetylation, direct contacts with histone proteins, and oligomerization and spreading of the cofactor on chromatin. Gro interacts with the class I histone deacetylase Rpd3/HDAC1 (Chen et al. 1999). Colocalization of Gro and Rpd3 to target genes results in deacetylation of specific lysine residues in histones H3 and H4 (Winkler et al.

2010). In light of the protein's ability to oligomerize, several groups have suggested that formation of a larger region of silent chromatin structure is possible in the wake of oligomerization and spreading over large areas of the genome (Song et al. 2004; Martinez and Arnosti 2008). Such an activity has been similarly suggested for the yeast Tup1 corepressor, with which Gro shares a limited level of structural similarity through its conserved WD40 domains (Flores-Saaib and Courey 2000). The ability of Gro to directly bind to all core histones may further contribute to Gro-mediated repression (Flores-Saaib and Courey 2000), perhaps enabling stable interaction of oligomers of the corepressor with chromatin. The interaction of Gro with H3 is significantly increased when N-terminal tail of H3 is deacetylated, suggesting that there may be positive feedback between deacetylation and physical binding (Chen et al. 1999; Winkler et al. 2010). Gro repression can alter the chromatin structure so that a closed chromatin state is formed with an increase in nucleosomal density, resulting exclusion of activators and the transcriptional machinery from the target region (Sekiya and Zaret 2007; Winkler et al. 2010). Gro may employ these chromatin-mediated mechanisms in a step-wise or mutually exclusive manner depending on gene-specific context. Chromatin-independent mechanisms may involve direct interaction of Gro with the transcriptional machinery to block preinitiation complex assembly, promoter release, or elongation (Turki-Judeh and Courey 2012).

A large number of metazoan repressors recruit C-terminal binding protein (CtBP) via PxDLS motifs to mediate their activity. Several short-range repressors such as Knirps, Kruppel and Snail rely heavily or exclusively on CtBP to meditate short-range repression, defined as repression that interferes with activators bound within approximately 100 bp (Aihara et al. 2007; Gray and Levine 1996; Turner and Crossley 2001). Although CtBP proteins are structurally related to NAD-dependent dehydrogenases, and exhibit a weak dehydrogenase activity in vitro, the significance of such activity to transcription is unclear (Kumar et al. 2002). Firm evidence exists however for the interaction between CtBP and histone modifying enzymes, in particular histone deacetylases and demethylases (Shi et al. 2003). Knirps interaction with with Rpd3 (HDAC1) is dependent on its CtBP-interaction ability (Struffi and Arnosti 2005). CtBP has also been implicated in the activity of the long-range repressor, Hairy; the corepressor interacts directly with a C-terminal motif in Hairy, adjacent to the Gro binding sequence. Mapping of genome-wide CtBP interaction sites using a chimeric CtBP-DNA methyltransferase enzyme suggested that Hairy binding overlaps to a greater extent with CtBP than with Gro, and CtBP and Hairy show synthetic genetic interactions (Bianchi-Frias et al. 2004). However, for specific target genes in the embryo, repression was found to be dependent on a Gro binding motif in Hairy, and mutation of CtBP-binding motif was not inhibitory to repression. Thus, the.proximity of the CtBP interaction motif to the Gro-binding WRPW motif was suggested to permit antagonism to occur between CtBP and Gro (Zhang and

Levine 1999). Thus, the role of CtBP in Hairy's repression activity remains to be elucidated.

CtBP also plays a role in establishing Polycomb-group (PcG) protein mediated repression of *HOX* genes during development. In the heterologous setting of the Drosophila embryo, mammalian YY1, a homolog of Drosophila Pleiohomeotic, requires CtBP to mediate repression by a PcG-dependent fashion (Atchison et al. 2003). Thus CtBP may not function only in short-range and long-range repression, but also in the unique perdurant repression effects of PcG. In addition to reversing the acetylation effects of HAT enzymes, the physical association between human CtBP1 and the HAT coactivators p300/CBP suggests that CtBP1 might also work by directly repressing their transcriptional activation function. Inhibition of the HAT activities of the coactivators might be a mechanism shared by other CtBP homologs (Kim et al. 2005).

Currently we lack information about how distinct corepressors contribute to repression, in either quantitative or qualitative fashions. To what end different transcription factors recruit distinct corepressors, and how are these cofactors work together to produce distinct patterns of chromatin modification is poorly understood. In this study, we used genome-wide methods to develop a picture of the functional roles of two highly conserved corepressors CtBP and Gro in the context of the Hairy transcriptional repressor.

3. Results

To determine the significance of CtBP binding for Hairy repression activity on a global scale, we disrupted CtBP interactions by mutation of the cognate binding site on Hairy (Figure 3-1A). This mutation has been demonstrated to abolish interaction of the two proteins (Zhang and Levine 1999). Hairy constructs were expressed in Drosophila embryos using a brief heatshock induction, as described earlier (Li and Arnosti 2011). Comparable expression of the wild-type and CtBP mutant forms of Hairy was confirmed by measuring mRNA and protein (Figure 3-1B-D).

We analyzed the repression ability of the CtBP mutant form on genes identified by microarray analysis (Kok et al. 2015). *In situ* hybridization (Figure 3-1C, shown for *ftz,* and others-not shown) and RT-qPCR (Figure 3-1D) showed that compared to wild-type Hairy, the mutant form failed to repress some genes (*odd, en, stumps, wg*) indicating Hairy acts on those genes in a CtBP-dependent manner. However, the mutant form was able to exert some repression activity on the other target genes (*ftz, pros, esg*). A modest increase in the expression of *HLHm7* may reflect antagonism of the endogenous wild-type Hairy protein by the mutant on some target genes, leading to derepression. These results clearly indicate that CtBP plays a critical role in Hairy-mediated repression, affecting gene expression in a quantitative and gene-specific fashion.

To understand the role of the CtBP interaction with Hairy in chromatin regulation, we carried out chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) analysis. The CtBP mutant form of Hairy significantly impacted H4Ac, H3K27Ac and H3K4me1 active histone marks on genes such as *ftz*, however, compared to the effects of the wild-type protein, the amount of change and range of deacetylation and demethylation were substantially decreased (Figure 3-2A). On the *gogo* gene, where wild-type Hairy induced local reduction of H4Ac and H3K27Ac, the mutant form was not able to regulate these marks (Figure 3-2B; (Kok et al. 2015)).

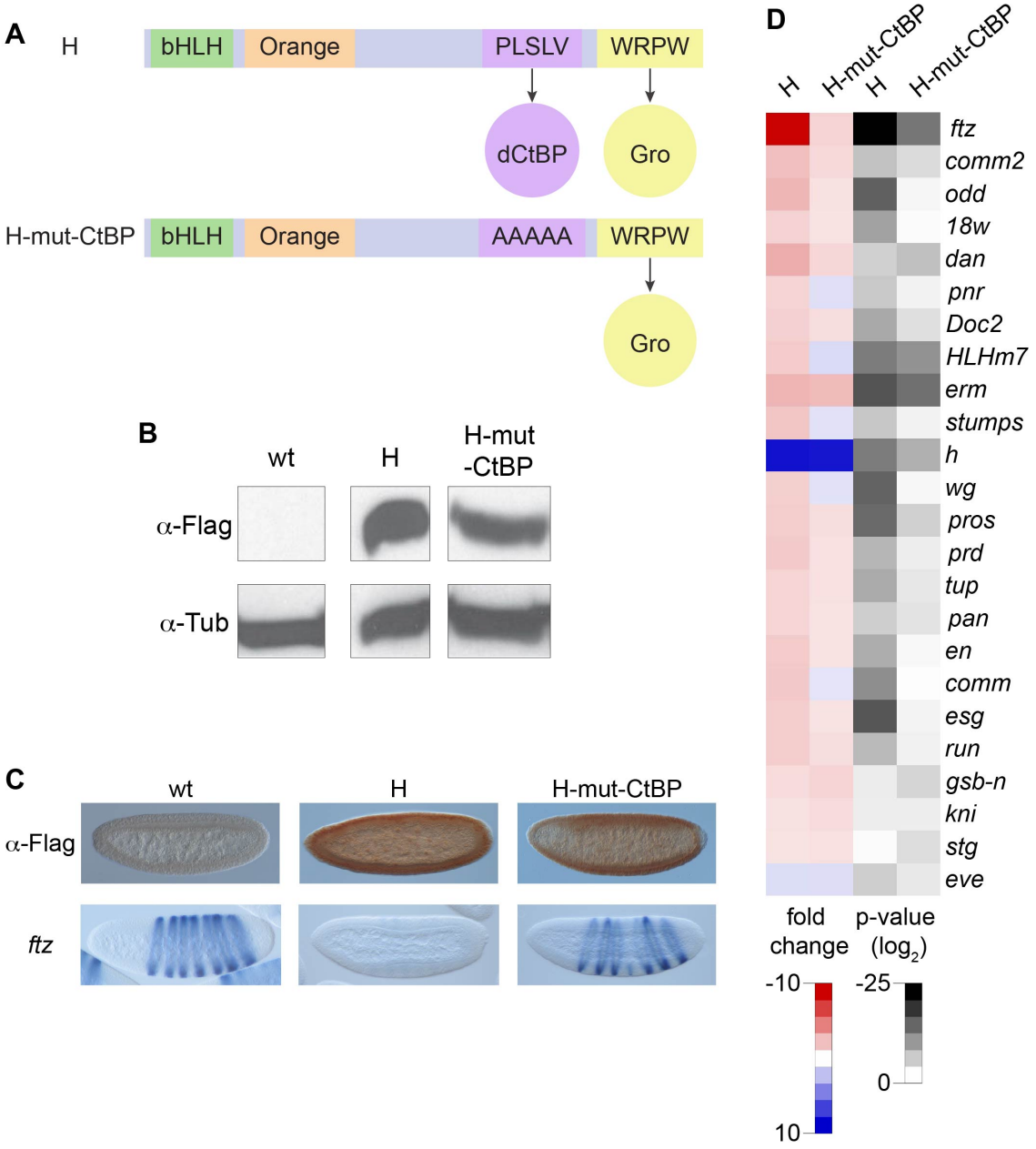


Figure 3-1 Contribution of the corepressor CtBP to Hairy mediated repression. (A) Protein domains of wild-type (H) and mutant form (H-mut-CtBP) of Hairy, and interaction with CtBP and Gro corepressors. PLSLV/AAAAA mutation prevents interaction with CtBP (Zhang and Levine 1999). **(B, C)** Western blot and immunohistochemistry analysis show similar expression for the induced proteins. Repression of *ftz* was measured by *in situ* hybridization. **(D)** RT-qPCR analysis of Hairy target genes demonstrates that H-mut-CtBP exhibits weaker repression activity on a spectrum of genes.

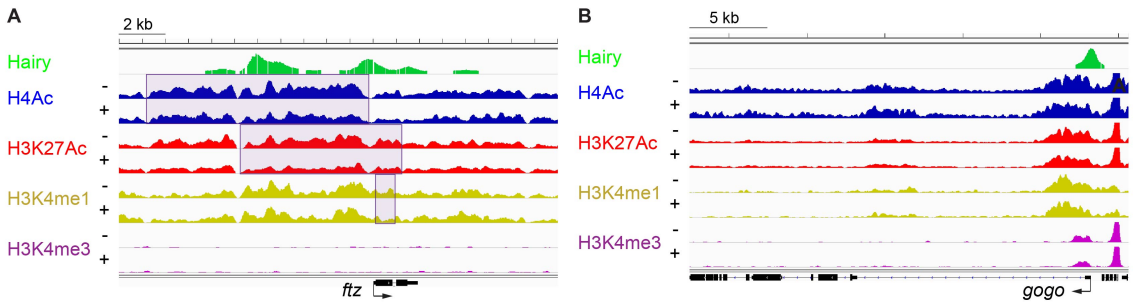


Figure 3-2 Examples of chromatin changes mediated by CtBP. ChIP-seq tracks for H4Ac, H3K27Ac, H3K4me1 and H3K4me3 are shown at Hairy regulated genes before (-) and after (+) H-mut-CtBP induction, with gene models below. The extent of change in histone modification, as well as the size of the impacted regions associated with the *ftz* gene, are less than that observed for the wild-type Hairy protein (Chapter 2). For *gogo*, histone acetylation and methylation signals were not significantly altered, unlike the local changes seen for expression of wild-type Hairy (Chapter 2).

In general, the average size of the altered chromatin regions for H4Ac and H3K27Ac marks was smaller (Figure 3-3A and Supplementary file 1). On direct targets of Hairy, the average range of deacetylation for H4 was ~700 bp and for H3K27 ~300 bp smaller than that observed for the wild-type protein, indicating that CtBP may influence how effectively Hairy-mediated events can “spread” on a locus (Figure 3-3A and (Kok et al. 2015)).

Even though mutation of the CtBP interaction motif altered Hairy’s activity on many genes, a significant overlap was found between the altered chromatin regions by the wild-type and mutant protein (Figure 3-4A-C). Where the mutant form of Hairy was found to induce changes in chromatin states of novel sites, these tended not to be near regions bound by endogenous Hairy, either because the mutant protein exhibits novel targeting, or because this protein induces unique indirect changes in chromatin on some loci (Figure 3-4-1). Overall, the wild-type Hairy repressor affected chromatin marks near hundreds of genes that were not affected at chromatin by the mutant form (Figure 3-4B,C and 3-5). This result indicates that Hairy is dependent on the CtBP interaction to change chromatin near these genes. CtBP may thus play a significant role in specificity of gene targeting and influence Hairy repression in a quantitative manner.

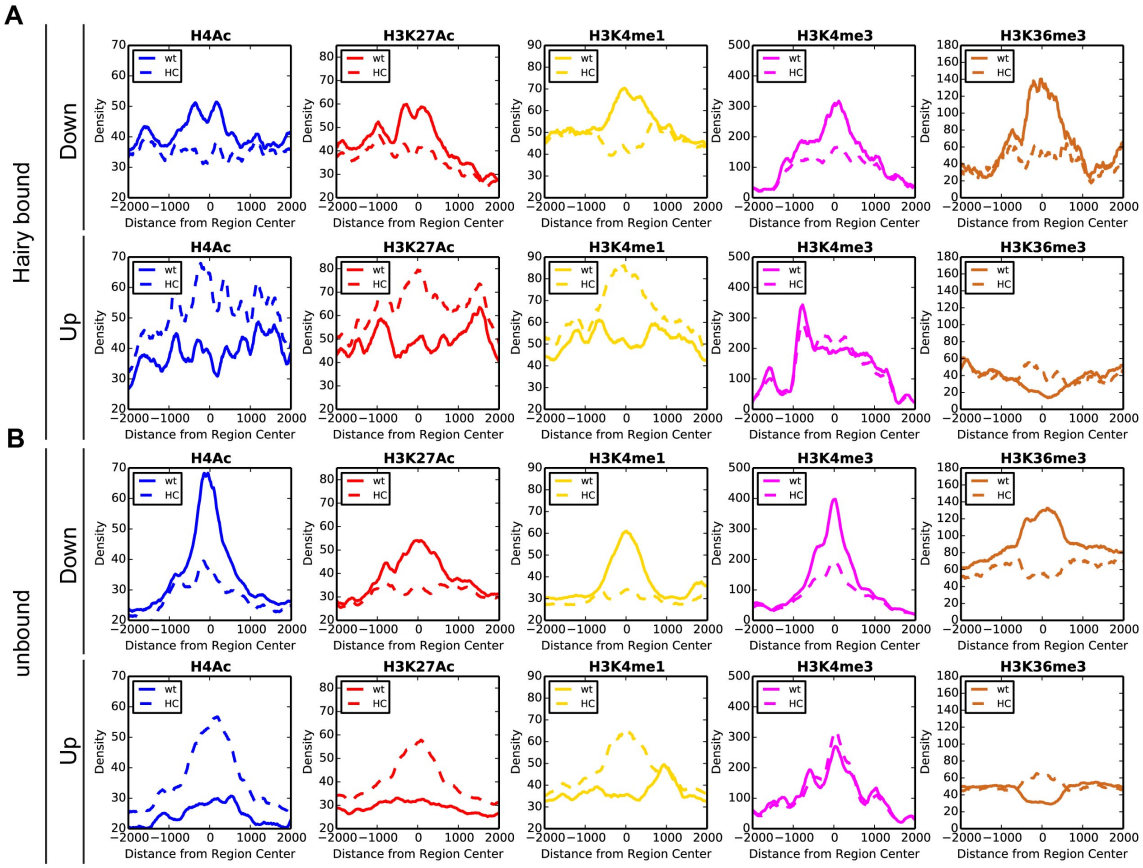


Figure 3-3 H-mut-CtBP induces smaller average tracts of modified chromatin. Histograms show the distribution of averaged ChIP-seq signal for a window of 4 kb centered on differentially changed regions associated with Hairy bound **(A)** and unbound **(B)** genes in the uninduced (wt, solid line) and Hairy mutant form induced (HC, dashed line) embryos for H4Ac, H3K27Ac, H3K4me1, H3K4me3, and H3K36me3.

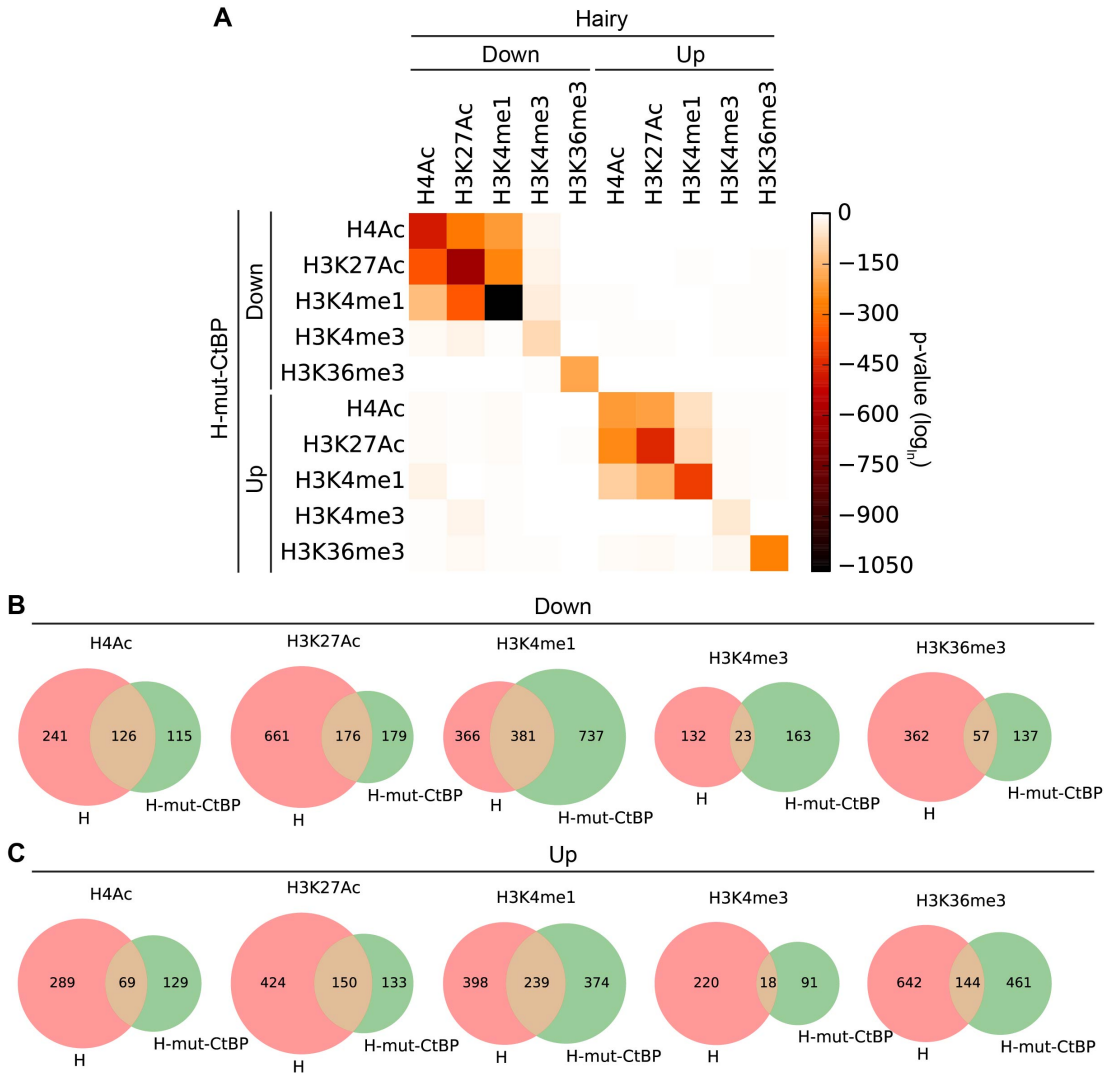


Figure 3-4 H-mut-CtBP and Hairy induce similar patterns of histone modifications. (A) Heatmap showing the overlap between significantly decreased (down) and increased (up) regions by induction of the wild-type Hairy (H, columns) and Hairy mutant form (H-mut-CtBP, row) on H4Ac, H3K27Ac, H3K4me1, H3K4me3 and H3K36me3. Bar scale is indicated as natural log of p-value. **(B, C)** Area-proportional Venn-diagrams showing the number of overlapping genes associated with loss (B) and gain (C) of H4Ac, H3K27Ac, H3K4me1, H3K4me3 and H3K36me3 as a result of induction of the wild-type (H) and mutant form (H-mut-CtBP).

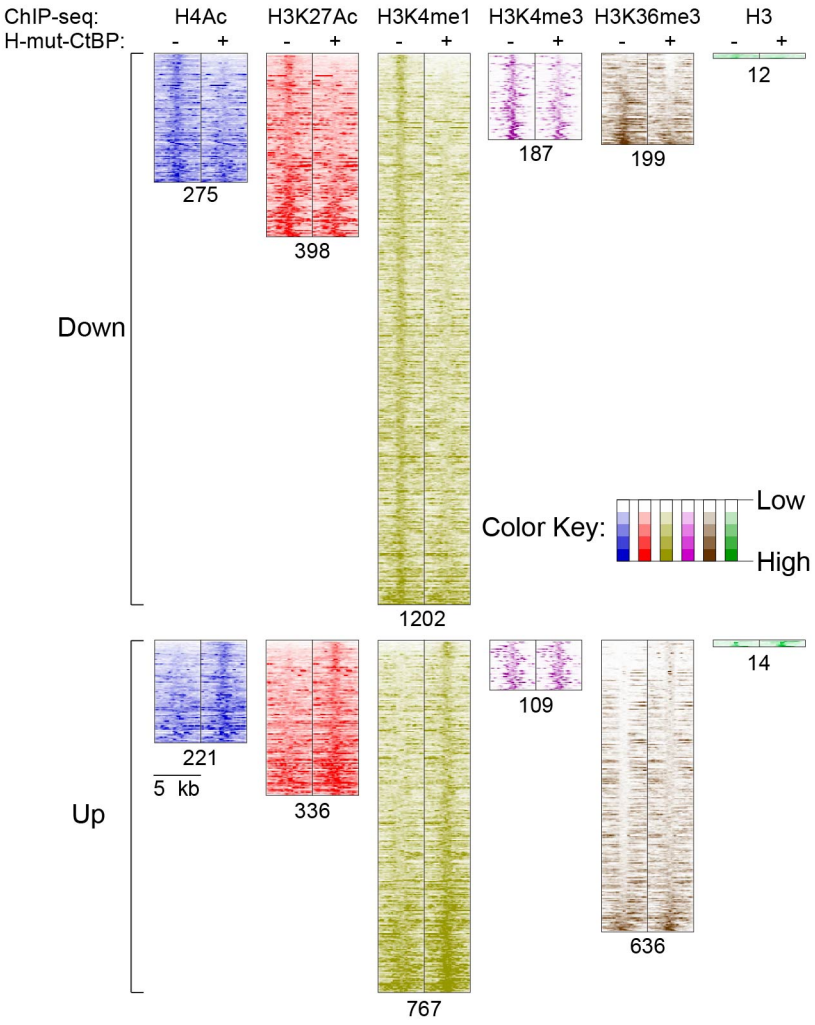


Figure 3-5 Pervasive genome-wide chromatin effects of H-mut-CtBP. All reduced (top) and increased (bottom) chromatin marks in the genome for H4Ac, H3K27Ac, H3K4me1, H3K4me3, H3K36me3, and H3 shown as heatmaps for 5 kb windows from the center of significantly affected regions before (-) and after (+) H-mut-CtBP induction. The number of affected regions indicated below each mark.

4. Discussion

Clarifying a controversial point in the field, our findings strongly implicate CtBP in contributing to Hairy repression activity, in addition to the well-characterized role for Gro in this protein's function. By studying the effect of expression of wild-type and mutant forms of the protein, we find that many genes repressed by Hairy were not effectively regulated by the CtBP mutant form. Based on the more limited tracts of histone deactylation induced by the mutant, we suggest that CtBP may stabilize Gro interactions in a repression complex, facilitating Gro spreading, either by direct interactions with Gro, or by stimulating local deacetylation to permit more extensive Gro-histone contacts. In the absence of CtBP binding, the Hairy-Gro complex appears to acquire somewhat different targeting specificity, thus the CtBP cofactor may also contribute to promoter-specificity by contacting other transcription factors.

The deep conservation of the C-terminal WRPW Gro interaction motif motif on HES family members across metazoans indicates that this corepressor is an essential part of HES protein activity. The interactions of CtBP and Gro with Hairy appear to be evolutionarily conserved; Hairy homologs throughout arthropods have recognizable C-terminal binding motifs for CtBP and Gro, and the more distantly related *HLHmdelta* gene encodes a PVNLA motif for CtBP interaction similar to the Hairy PLSLV motif (Davis and Turner 2001). In vertebrates, the HES family has undergone extensive radiation, and there is

diversification in interaction motifs for corepressors. Most HES family members possess the hydrophobic C-terminal sequence for recruitment of TLE, the Gro homolog. Similar to Hairy, mammalian HES proteins interact with Sir2-class HDACs, while CtBP interactions are less well characterized. The similarities of these repressor-corepressor interactions indicate that the biochemical properties of Hairy will be informative for studies of the wider class of HES proteins. A recent study noted that a C-terminal fragment of Hairy tethered to a heterologous DNA binding domain mediated short-range, not long-range, effects when targeted to the stripe 2 enhancer of the *even-skipped (eve)* locus, contrary to previous reports that endogenous Hairy mediates long-range repression (Crocker and Stern 2013). The fragment of Hairy tested lacked its endogenous DNA binding domain as well as another conserved functional domain, thus this protein may not fully recapitulate the activity of the wild-type protein, which appears to be mediating chromatin changes consistent with widespread effects on a genomic level. Alternatively, context-specific factors may block Hairy's range of action, and the *eve* locus may represent just such a region in which long-range effects are prohibited. Indeed, a subset of endogenous Hairy targets exhibit relatively limited sizes of chromatin modifications (Kok et al. 2015).

Other transcriptional repressors have also been found to interact with both CtBP and Gro in Drosophila, including the short-range Knirps repressor, as well as Brinker, Hairless and Runt, which may act through long-range effects (Nibu et al. 1998; Payankaulam and Arnosti 2009; Hasson et al. 2001; Barolo et al. 2002;

Wheeler et al. 2002). A combination of both Gro and CtBP activity is required for repression mediated by Knirps and Hairless (Payankaulam and Arnosti 2009; Nagel et al. 2005). On the other hand, multiple mechanisms were suggested in repression by Brinker; it employs distinct corepressors in different developmental stages (Upadhyai and Campbell 2013). Furthermore, these corepressors contribute to gene-specific regulation for Runt; Gro is required for establishment and maintenance of *en* repression but not for *Slp1* repression (Walrad et al. 2010). These metazoan transcriptional repressors thus interact with multiple corepressors to effect repression in a context-dependent manner, and the genome-wide differential dependence on CtBP seen for Hairy may be representative of the entire family of HES proteins.

Loss of the C-terminal Gro WRPW motif destabilizes the Hairy protein in embryos (Figure 3-1-1) (interestingly, the Gro binding mutant was apparently stable when expressed in cultured Drosophila S2 cells), therefore we generated forms of Hairy in which CtBP or Gro is directly fused to the body of the transcription factor (Figure 3-6A). Preliminary analysis of mRNA and protein expression of these fusion transgenes in embryos showed stabilization of Gro mutant form, which should allow us to test forms that only have CtBP or Gro (Figure 3-6B,C). Initial functional analysis suggested that Gro interaction with Hairy is important for repression of *ftz* (Figure 3-6D). How the activity of Gro is linked to the unique effects mediated by Hairy can be further investigated quantitatively by genome-wide approaches.

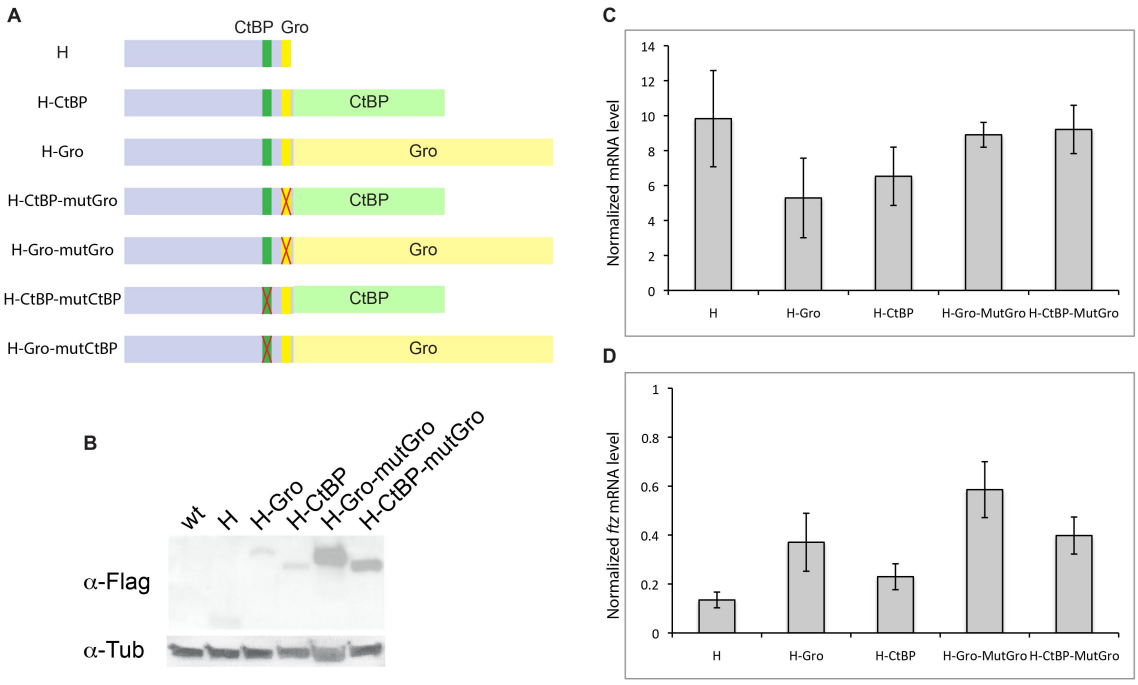


Figure 3-6 Transgenic Drosophila lines carrying Hairy fusion proteins permit the testing of the roles of distinct corepressors. (A) Wild-type or mutant forms of Hairy are fused to either CtBP or Gro for induction in embryos. **(B)** Fusion of CtBP or Gro to Hairy protein lacking the C-terminal WRPW Gro-binding motif resulted in detectable protein expression measured by Western blot. **(C)** Ectopic mRNA expression and **(D)** repression of *ftz* were tested by RT-qPCR analysis.

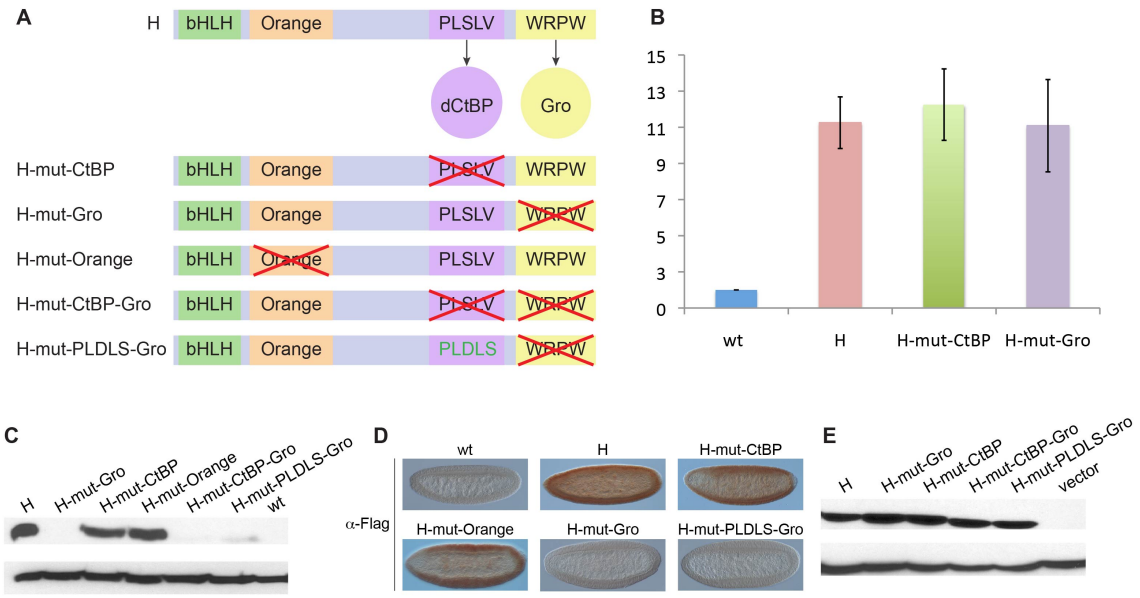


Figure 3-1-1 Destabilization of Hairy by mutation in Gro interaction domain. **(A)** Protein domains of wild-type (H) and mutant forms of Hairy, and interaction with CtBP and Gro corepressors. PLSLV/AAAAA and WRPW/AAAAA mutations prevent interaction with CtBP and Gro respectively (Zhang and Levine 1999). **(B)** RT-qPCR analysis shows similar mRNA expression for CtBP and Gro mutant proteins **(C, D)** Western blot and immunohistochemistry analysis show undetectable expression for the Gro mutant proteins, preventing us from assaying its activity. **(E)** Western blot analysis shows similar protein expression in S2 cells, indicating a role of Gro in Hairy stability in development.

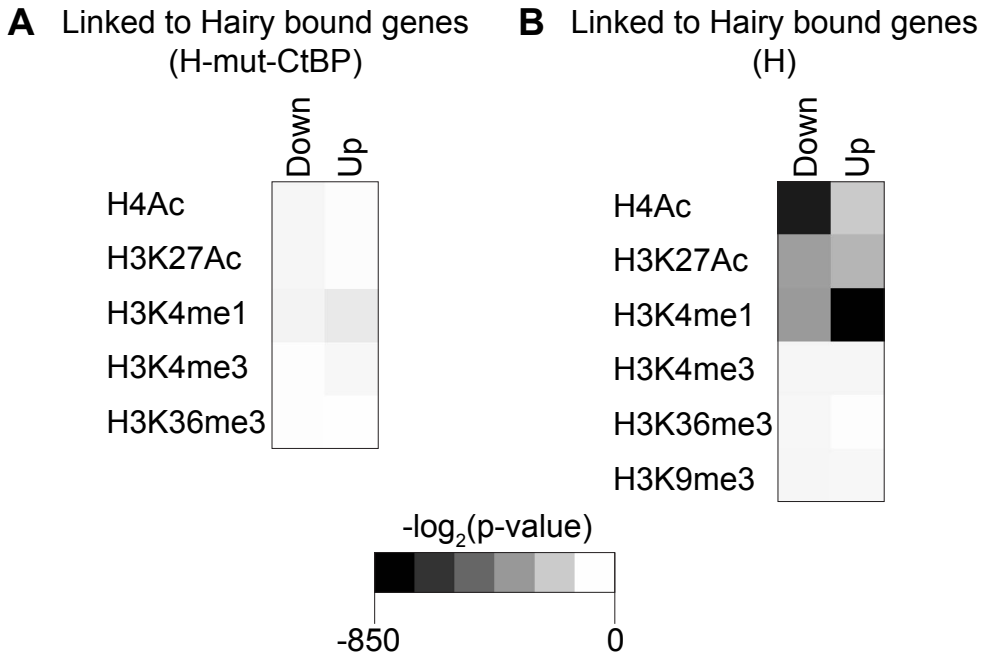


Figure 3-4-1 Significance of individual histone modifications associated with Hairy bound genes. (A) Weaker correlations between changes in chromatin marks near Hairy-bound genes for induction of H-mut-CtBP, **(B)** compared to strong link between loss of H4Ac, gain of H3K4me1, and presence of Hairy on genes altered by the wild-type protein (H).

5. Materials and methods

Plasmid construction: The wild-type heat-inducible *hairy* gene was created by introducing a multiple cloning site (MCS) containing Kozak sequence, initiator ATG and HindIII/BglII sites into the 5' portion of the *hairy* ORF in the pCaSpeR-hsh using EcoRI/BstEII sites (described previously, (Li and Arnosti 2011)). 400 bp of upstream promoter, 5' UTR, Kozak sequence, initiator ATG, HindIII/BglII sites, coding sequence and entire *hsp70* 3' UTR from the modified pCaSpeR-hsh were amplified using 5' and 3' primers with AgeI/KpnI sites and subcloned to the modified pattB vector (Sayal et al. 2011). Oligonucleotides with sequence encoding the double Flag epitope, as described in (Zhang and Arnosti 2011), was inserted 5' of the coding sequence after the ATG using HindII/BglII sites, so that Hairy protein was expressed with the double Flag tag at the N terminus. The mutant form (PLSLV/AAAAA) of Hairy was created by site-directed mutagenesis. All transgenes were integrated by Rainbow Transgenics (Camarillo, CA) into the same position (51D) in the genome.

Embryo collection, in situ hybridization and antibody staining of Drosophila embryos: For chromatin analysis 2-3.5 hr. embryos were collected and heat-shock treated for induction of transgenes as described previously (Li and Arnosti 2011). For analysis of gene expression by *in situ* hybridization, embryos were fixed and stained using anti-digoxigenin-UTP-labeled RNA probe for *ftz* as described previously (Struffi 2004).

Western blot analysis: Crude embryo lysate preparation and immunoblotting were done as described previously (Struffi 2004). 50 µg of embryo lysates were run on 10% SDS-PAGE gels, transferred to PVDF membranes, and probed with mouse monoclonal M2 anti-Flag antibody (Sigma) at 1:10,000 dilution and mouse monoclonal anti-tubulin (Iowa Hybridoma Bank) at 1:10,000 dilution. Antibody incubation was performed overnight at 4 °C for the primary antibody and 2 h at room temperature for the secondary antibody in TBST (20 mm Tris-Cl, pH 7.5, 120 mm NaCl, 0.1% Tween 20) with 5% nonfat dry milk, washed three times for 5 min each, after primary and secondary antibody incubation. Blots were developed using HRP-conjugated secondary antibodies (Pierce) and SuperSignal West Pico chemiluminescent substrate (Pierce).

Quantitative reverse transcriptase PCR analysis: Total RNA from embryos was purified using RNeasy columns (Qiagen), and reversed transcribed using a High Capacity cDNA Reverse Transcription Kit from Invitrogen/Applied Biosystems. The cDNA was then analyzed by real-time PCR using the primer pairs located at transcription units. Data was normalized to *act5c.* Values for wild-type embryos were set to 1; results represent the average of 2-8 biological replicates. Statistical significance was tested using Student's T test and $p<0.05$. Amplicons were designed using Primer Express and Primer-BLAST.

Chromatin Immunoprecipitation: Heat shocks and ChIPs were performed as described previously (Li and Arnosti 2011), with the exceptions that embryos were sonicated for a total of 20 times using a Branson sonicator in 1 ml of sonication buffer. After precipitation of chromatin-antibody complexes, protein A beads were washed twice with low-salt buffer, once with high-salt buffer, once with LiCl buffer and twice with Tris-EDTA. We used the following antibodies: rabbit IgG (5 µl, Santa Cruz Biotechnology), rabbit anti-H3 (1 µl, Abcam), rabbit anti-acetyl H4 (1 µl, Upstate), rabbit anti-acetyl H3K27 (1 µl, Abcam), rabbit anti-monomethyl H3K4 (1 µl, Abcam), rabbit anti-trimethyl H3K4 (1 µl, Abcam), and rabbit anti-trimethyl H3K36 (2 µl, Abcam).

ChIP-seq: Libraries. DNA from chromatin immunoprecipitation (10 ng) was adapter-ligated and PCR amplified (18 cycles) as described in (Ford et al. 2014). DNA ligated to the adapter was size selected for 300-500 bp. Illumina HiSeq single-end reads were checked using FastQC and HOMER for sequence quality, base sequence and GC content, sequence duplication, sequence bias, overrepresented sequences and Kmer content. Reads were aligned to genome (BDGP 5.70) with Bowtie version 1.0.0 using -m 1 --best parameters. Tags that only mapped uniquely to the genome were considered for further analysis. Summary of tags generated is shown on Supplementary file 6. ChIP-Seq experiments were visualized as custom tracks using Integrative Genomics Viewer (Broad Institute). Total uniquely mapped tags were normalized to 10 million reads to generate tracks. y-axis values shown in all figures use the same

scale for an individual measurement of each histone modification in the individual panels. For reasons of clarity, scales can vary between different panels.

Mapping differential regions. We detected the regions where chromatin states are changed upon induction of Hairy by comparing the level of histone marks at particular genomic locations. Differentially changed genomic regions were identified using the diffReps program (Shen et al. 2013), which uses a sliding window approach to scan the genome and find regions showing read count differences. Default window size with --nsd broad --meth nb parameters was used for the analysis. For downstream analysis, we used regions with $p<0.05$ and fold change (log2)>0.4 or fold change (log2)<-0.4. Input was sequenced from nontransgenic (wt) and Hairy overexpressing embryos and used as background.

HOMER (Hypergeometric Optimization of Motif EnRichment) was used for peak finding and downstream data analysis (Heinz et al. 2010).

Identification of ChIP-seq peaks: Using HOMER with default settings, peaks for histone marks and Flag tagged Hairy protein were identified using signals from H3 ChIP and input respectively as background.

Annotation of significantly affected regions: Regions detected by diffReps or peaks called by HOMER were associated with genes by identifying the nearest RefSeq TSS and annotated to a genomic feature such as intergenic, intron, exon etc.

Normalization of ChIP-seq tags for histograms, heatmaps, and scatter plots: We normalized the total number of mapped tags to 10 million for each sample using HOMER so that the read densities were comparable.

Comparison of ChIP-seq experiments using histograms: ChIP-seq densities of a 4 kb window centered at affected regions detected by diffReps were determined using HOMER. The program normalizes the output histogram such that the resulting units are per bp per region with bin size of 10 bp. Plots were generated using matplotlib (Hunter 2007).

Comparison of ChIP-seq experiments using heatmaps: Data matrices were generated using HOMER by counting total tags in a 5 kb window around affected regions or peaks and normalizing to 10 million reads with bin size of 25 bp. Data was visualized using Java Tree View (Saldanha 2004).

Comparison of ChIP-seq experiments using scatter plots: Tag densities were calculated by counting the tags at regions defined by peak coordinates of the first experiment (x axis) and compared to the second experiment (y axis). Data was log2 transformed and plotted using matplotlib. Pearson's Correlation Coeffcients were calculated to determine the extent of similarity between samples.

Analysis of co-occurrence of differentially changed regions: mergePeaks program of HOMER was used to find overlapping sites between differentially changed regions of different histone marks upon Hairy induction. These regions were considered as overlapped if changed regions from each experiment share at least 1 bp. Significance of co-occurence of regions was indicated by natural log p-values using the hypergeometric distribution. Positive values signify divergence.

Linking affected regions to Hairy binding: Affected regions for chromatin marks were considered as Hairy bound if the nearest gene has at least one Hairy peak based on a previous ChIP-chip data (MacArthur et al. 2009).

REFERENCES

REFERENCES

Aihara H, Perrone L, Nibu Y. 2007. Transcriptional repression by the CtBP corepressor in Drosophila. *CtBP Family Proteins* 18–27.

Atchison L, Ghias A, Wilkinson F, Bonini N, Atchison ML. 2003. Transcription factor YY1 functions as a PcG protein in vivo. *EMBO J* **22**: 1347–1358.

Barolo S, Stone T, Bang AG, Posakony JW. 2002. Default repression and Notch signaling: Hairless acts as an adaptor to recruit the corepressors Groucho and dCtBP to Suppressor of Hairless. *Genes Dev* **16**: 1964–1976.

Bianchi-Frias D, Orian A, Delrow JJ, Vazquez J, Rosales-Nieves AE, Parkhurst SM. 2004. Hairy Transcriptional Repression Targets and Cofactor Recruitment in Drosophila. *Plos Biol* **2**: e178.

Chen G, Fernandez J, Mische S, Courey AJ. 1999. A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in Drosophila development. *Genes Dev* **13**: 2218–2230.

Crocker J, Stern DL. 2013. TALE-mediated modulation of transcriptional enhancers in vivo. *Nat Methods* **10**: 762–767.

Davis RL, Turner DL. 2001. Vertebrate hairy and Enhancer of split related proteins: transcriptional repressors regulating cellular differentiation and embryonic patterning. *Oncogene* **20**: 8342–8357.

Fisher AL, Ohsako S, Caudy M. 1996. The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain. *Molecular and Cellular Biology* **16**: 2670–2677.

Flores-Saaib RD, Courey AJ. 2000. Analysis of Groucho-histone interactions suggests mechanistic similarities between Groucho- and Tup1-mediated repression. *Nucleic Acids Research* **28**: 4189–4196.

Ford E, Nikopoulou C, Kokkalis A, Thanos D. 2014. A method for generating highly multiplexed ChIP-seq libraries. *BMC Res Notes* **7**: 312.

Gray S, Levine M. 1996. Short-range transcriptional repressors mediate both quenching and direct repression within complex loci in Drosophila. *Genes Dev* **10**: 700–710.

Grbavec D, Stifani S. 1996. Molecular interaction between TLE1 and the carboxyl-terminal domain of HES-1 containing the WRPW motif. *Biochemical*

and Biophysical Research Communications **223**: 701–705.

Hasson P, Müller B, Basler K, Paroush Z. 2001. Brinker requires two corepressors for maximal and versatile repression in Dpp signalling. *EMBO J* **20**: 5725–5736.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**: 576–589.

Hunter JD. 2007. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **9**: 0090–0095.

Kim J-H, Cho E-J, Kim S-T, Youn H-D. 2005. CtBP represses p300-mediated transcriptional activation by direct association with its bromodomain. *Nature Structural & Molecular Biology* **12**: 423–428.

Kok K, Ay A, Li LM, Arnosti DN. 2015. Genome-wide errant targeting by Hairy. *eLife* **4**.

Kumar V, Carlson JE, Ohgi KA, Edwards TA, Rose DW, Escalante CR, Rosenfeld MG, Aggarwal AK. 2002. Transcription corepressor CtBP is an NAD(+)-regulated dehydrogenase. *Molecular Cell* **10**: 857–869.

Li LM, Arnosti DN. 2011. Long- and Short-Range Transcriptional Repressors Induce Distinct Chromatin States on Repressed Genes. *Curr Biol* **21**: 406–412.

MacArthur S, Li X-Y, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keränen SVE, et al. 2009. Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol* **10**: R80.

Martinez CA, Arnosti DN. 2008. Spreading of a Corepressor Linked to Action of Long-Range Repressor Hairy. *Molecular and Cellular Biology* **28**: 2792–2802.

Nagel AC, Krejci A, Tenin G, Bravo-Patiño A, Bray S, Maier D, Preiss A. 2005. Hairless-mediated repression of notch target genes requires the combined activity of Groucho and CtBP corepressors. *Molecular and Cellular Biology* **25**: 10433–10441.

Nibu Y, Zhang H, Bajor E, Barolo S, Small S, Levine M. 1998. dCtBP mediates transcriptional repression by Knirps, Krüppel and Snail in the Drosophila embryo. *EMBO J* **17**: 7009–7020.

Paroush Z, Finley RL, Kidd T, Wainwright SM, Ingham PW, Brent R, Ish-

Horowicz D. 1994. Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. *Cell* **79**: 805–815.

Payankaulam S, Arnosti DN. 2009. Groucho corepressor functions as a cofactor for the Knirps short-range transcriptional repressor. *Proceedings of the National Academy of Sciences* **106**: 17314–17319.

Payankaulam S, Li LM, Arnosti DN. 2010. Transcriptional Repression: Conserved and Evolved Features. *Curr Biol* **20**: R764–R771.

Poortinga G, Watanabe M, Parkhurst SM. 1998. Drosophila CtBP: a Hairy-interacting protein required for embryonic segmentation and hairy-mediated transcriptional repression. *EMBO J* **17**: 2067–2078.

Rosenberg MI, Parkhurst SM. 2002. Drosophila Sir2 is required for heterochromatic silencing and by euchromatic Hairy/E(Spl) bHLH repressors in segmentation and sex determination. *Cell* **109**: 447–458.

Saldanha AJ. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**: 3246–3248.

Sayal R, Ryu S-M, Arnosti DN. 2011. Optimization of reporter gene architecture for quantitative measurements of gene expression in the Drosophila embryo. *Fly (Austin)* **5**: 47–52.

Sekiya T, Zaret KS. 2007. Repression by Groucho/TLE/Grg Proteins: Genomic Site Recruitment Generates Compacted Chromatin In Vitro and Impairs Activator Binding In Vivo. *Molecular Cell* **28**: 291–303.

Shen L, Shao N-Y, Liu X, Maze I, Feng J, Nestler EJ. 2013. diffReps: Detecting Differential Chromatin Modification Sites from ChIP-seq Data with Biological Replicates ed. R. Mantovani. *PLoS ONE* **8**: e65598.

Shi Y, Sawada J-I, Sui G, Affar EB, Whetstine JR, Lan F, Ogawa H, Luke MP-S, Nakatani Y, Shi Y. 2003. Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422**: 735–738.

Song H, Hasson P, Paroush Z, Courey AJ. 2004. Groucho Oligomerization Is Required for Repression In Vivo. *Molecular and Cellular Biology* **24**: 4341–4350.

Struffi P. 2004. Quantitative contributions of CtBP-dependent and -independent repression activities of Knirps. *Development* **131**: 2419–2429.

Struffi P, Arnosti DN. 2005. Functional interaction between the Drosophila knirps short range transcriptional repressor and RPD3 histone deacetylase. *J Biol Chem* **280**: 40757–40765.

Turki-Judeh W, Courey AJ. 2012. *Chapter 3 - Groucho: A Corepressor with Instructive Roles in Development*. 1st ed. Elsevier Inc.

Turner J, Crossley M. 2001. The CtBP family: enigmatic and enzymatic transcriptional co-repressors. *Bioessays* **23**: 683–690.

Upadhyai P, Campbell G. 2013. Brinker possesses multiple mechanisms for repression because its primary co-repressor, Groucho, may be unavailable in some cell types. *Development* **140**: 4256–4265.

Walrad PB, Hang S, Joseph GS, Salas J, Gergen JP. 2010. Distinct contributions of conserved modules to Runt transcription factor activity. *Mol Biol Cell* **21**: 2315–2326.

Wheeler JC, VanderZwan C, Xu X, Swantek D, Tracey WD, Gergen JP. 2002. Distinct in vivo requirements for establishment versus maintenance of transcriptional repression. *Nat Genet* **32**: 206–210.

Winkler CJ, Ponce A, Courey AJ. 2010. Groucho-mediated repression may result from a histone deacetylase-dependent increase in nucleosome density. *PLoS ONE* **5**: e10166.

Zhang H, Levine M. 1999. Groucho and dCtBP mediate separate pathways of transcriptional repression in the Drosophila embryo. *Proc Natl Acad Sci USA* **96**: 535–540.

Zhang YW, Arnosti DN. 2011. Conserved catalytic and C-terminal regulatory domains of the C-terminal binding protein corepressor fine-tune the transcriptional response in development. *Molecular and Cellular Biology* **31**: 375–384.

CHAPTER IV

Conclusions and Future Directions

1. Conclusions

My doctoral research on molecular mechanisms of Drosophila gene regulation provides evidence that mechanisms of short- and long-range repressors differ fundamentally. Despite the common role of corepressors Groucho and CtBP, short- and long-range repressors impart very different alterations to chromatin structure (Figure 4-1A). Based on our laboratory's research, short-range repression mediated by Knirps is associated with local chromatin compaction, an increase in histone density, local deactylation, and activator ejection, without blocking access of RNA polymerase to the promoter, consistent with a local "quenching" model of repression (Li and Arnosti 2011). Furthermore, Knirps engages in a completely distinct set of activities, including increasing association of H1, raising levels of H3K9me3 (S. Payankaulam unpublished) (Figure 4-1B). However, long-range repression mediated by the Hairy protein involves extensive spreading of the Groucho corepressor, correlated with widespread deacetylation and elimination of the transcriptional machinery from the transcription unit, with little change in nuclease accessibility (Martinez and Arnosti 2008; Li and Arnosti 2011). My genome-wide work shows that Hairy-mediated effects have common aspects that constrast sharply with what we know about short-range action of Knirps. Hairy invokes a characteristic set and scope of histone modifications at

targeted loci, although there are gene-specific differences. Most strikingly, we noted that Hairy actively remodels many regions that are not linked to changes in gene expression. This so-called “futile cycling” indicates that Hairy is not an inert factor on non-functional sites (Kok et al., 2015). Previous studies had suggested that many “off target” sites occupied by many metazoan transcription factors are quiescent, and essentially a background removed from evolutionarily selective processes (Fisher et al. 2012). The action of Hairy at many transcriptionally nonfunctional sites indicates that there is a lower than expected threshold for evolution of novel *cis* elements and gene regulatory network links, because simple acquisition of an activator site will generate a complete on/off switch, characteristic of most higher eukaryotic enhancers.

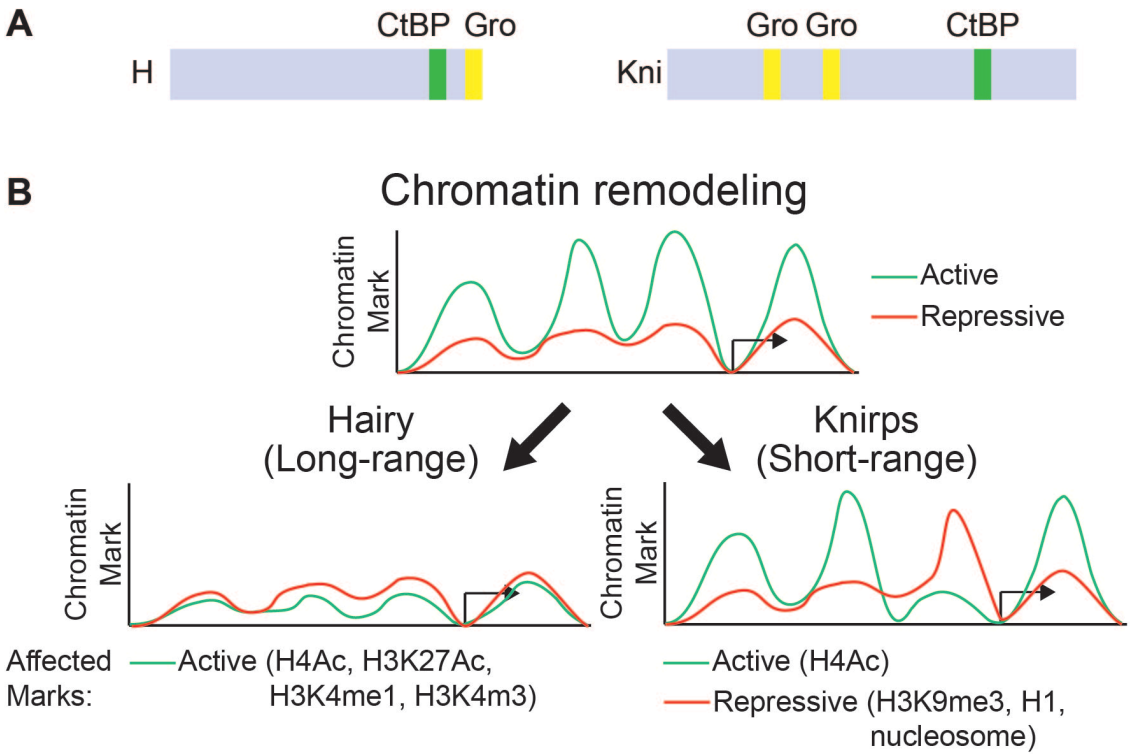


Figure 4-1 Mechanistic differences between long-range and short-range repression. (A) Hairy and Knirps interact with common conserved corepressors, Gro and CtBP. **(B)** At the top are schematic chromatin marks on a gene; in green, activating marks such as H4Ac; in red, repressive marks such as H3K9me3. Upon conversion from active to repressed state, Hairy induces long-range (multi-kbp) changes in active acetylation and methylation marks. Knirps by contrast induces very localized loss of these active marks, but also causes increase in H3K9me3, H1, and total nucleosome levels. A local decrease in MNase accessibility is also detected.

2. Future directions

There are several possible future directions that will increase our knowledge of pathways of transcriptional repression and their significance, including Hairy-mediated pervasive chromatin dynamics and the roles of corepressors CtBP and Gro in GRNs and evolution.

Whether the “futile cycling” we describe for Hairy is a common property of many repressors, and hence a basic architectural feature of gene regulatory circuits, is unknown. We should seek to determine general properties of a wide spectrum of Drosophila repressors. To determine whether features found for Hairy and Knirps are widely shared with other transcriptional repressors, I have prepared Drosophila lines containing inducible forms of four other embryonically-active transcriptional repressors that recruit CtBP and/or Gro, as well as Hunchback, which interacts with dMi-2 ATPase remodeling complex (Figure 4-2A). Characterization of these lines is incomplete; but I find that the overexpression of *odd-skipped* and *brinker* mRNA is detected in embryos (Figure 4-2B). RNA-Seq of Brk line revealed 43 down- and 52 up-regulated genes (data not shown). Conducting similar studies for the other repressors and surveying histone acetylation and methylation signals will provide evidence to determine whether each of these factors induces a unique set of changes across the genome. It is possible that there are essentially just two general forms of repression, short-range and long-range, represented by Knirps and Hairy. Molecular studies in

which *cis*-regulatory elements were modified to test distance specific effects certainly point to strong similarities in the effects of Giant, Knirps, and Snail (Hewitt et al. 1999; Gray and Levine 1996; Arnosti et al. 1996). The number of long-range repressors tested is less, so Hairy serves as our best-studied example (Dorsal in particular *cis*-regulatory contexts can be a long-range repressor, and other proteins such as Runt and Eve have not been explicitly tested for range in the same fashion, but may represent long-range actors). Alternatively, more detailed investigations may reveal that there is a continuum of effects mediated by different proteins that interact with CtBP and Groucho; genome-wide efforts will be helpful in this regard.

I predict that active chromatin remodeling at sites both functional and nonfunctional may be a global property of many transcriptional regulators, a picture that challenges the assumption that “off-target sites” are essentially silent. Identification of genome-wide targets of these repressors in the embryo will allow us to understand whether, like Hairy, different classes of repressors have unique modes of chromatin-based effects and whether CtBP and Gro induce different effects when recruited by different repressors. This will provide us with a unified picture of how repressor-, corepressor- and gene-specific effects combine to generate the dynamic patterns of transcriptional regulation in the embryo. Importantly, where a common theme of “off-target sites regulate chromatin structure” found to be a frequent feature of interactions by transcription factors, we should fundamentally reconsider how we interpret genome-wide data sets.

Too often, an increase in H3K27 acetylation is taken as proof that the element is an active enhancer, without any further functional test (Villar et al. 2015). My doctoral research provides evidence that many of these chromatin regions with active marks may represent a platform for nonfunctional biochemical activities. These regions may show higher levels of substitutions, deletions or insertions of TF binding sites in time than functional elements. Pervasive biochemical activities of TF with flexibility of binding site composition may facilitate creation of right context of enhancer grammar in evolution of gene expression.

To test the hypothesis that “off target” activities of transcriptional repressors are truly unlinked to gene expression, and represent the “background activity” proposed in our model, CRISPR should be used to delete genomic regions that are modified but not associated with expression of linked genes. Whether loss of these regions indeed has minimal impact on gene expression would be tested by *in situ* hybridization and RT-qPCR analysis. In some cases, the elements may represent enhancers that are active in other contexts, but the main idea remains the same, that for any given developmental stage, the pervasive interactions of transcriptional regulators may be quietly modifying chromatin marks that are not functional at that point in time, but may evolve to become new *cis*-regulatory modules.

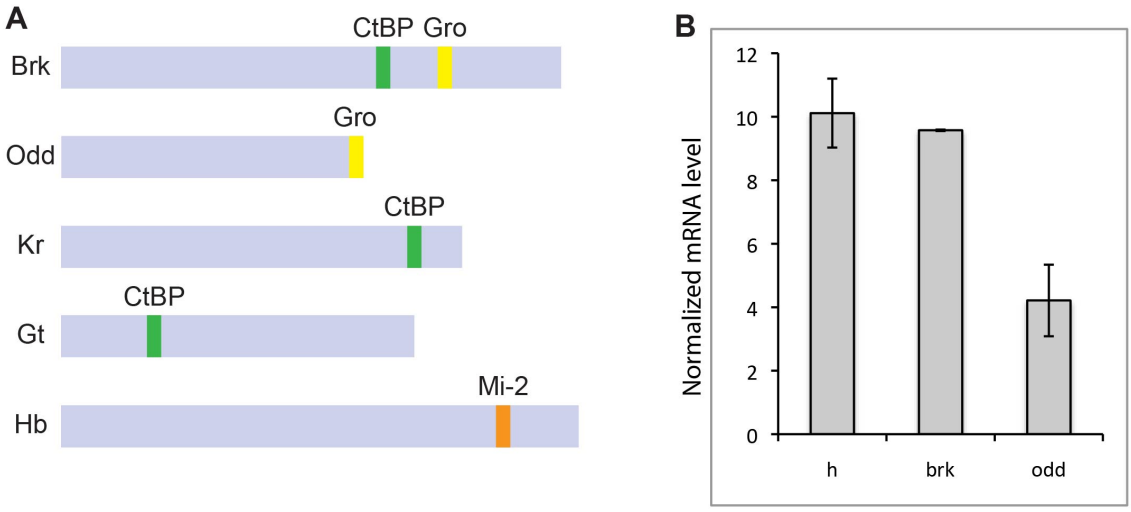


Figure 4-2 Gro and CtBP-dependent embryonic repressors. (A) Transgenic Drosophila lines express additional Gro and/or CtBP interacting repressors in embryos. Hb is unusual in that it appears to function both as repressor and activator, and is included as a repressor likely to invoke additional effects in a gene-specific manner. **(B)** Overexpression of *brk* and *odd* was tested by RT-qPCR analysis.

3. Methods Materials

Plasmid construction: The heat-inducible *brk*, *odd*, *kr gt*, and *hb* genes were created by subcloning of coding sequences to pattB vector containing 400 bp of upstream promoter, 5' UTR, Kozak sequence, initiator ATG, HindIII/BglII sites, coding sequence and entire *hsp70* 3' UTR so that the proteins were expressed with the double Flag tag at the N terminus as described in Chapter II.

Embryo collection of Drosophila embryos: For chromatin analysis 2-3.5 hr. embryos were collected and 20 min heat-shock treated for induction of transgenes as described previously as described in Chapter II.

Quantitative reverse transcriptase PCR analysis: Total RNA from embryos was purified using RNeasy columns (Qiagen), and reversed transcribed using a High Capacity cDNA Reverse Transcription Kit from Invitrogen/Applied Biosystems. The cDNA was then analyzed by real-time PCR using the primer pairs located at transcription units. Data was normalized to *act5c.* Values for wild-type embryos were set to 1; results represent the average of 2 biological replicates.

APPENDIX

APPENDIX

Evolving ways to switch genes on and off[5]

[5]This work was published as a blog at the Michigan State University NSF Center for Evolution in Action (BEACON): Kok K "Evolving ways to switch genes on and off", http://beacon-center.org/blog/2015/02/02/beacon-researchers-at-work-evolving-ways-to-switch-genes-on-and-off/

"In considering the Origin of Species, it is quite conceivable that a naturalist…might come to the conclusion that each species…had descended, like varieties, from other species. Nevertheless, such a conclusion, even if well founded, would be unsatisfactory, until it could be shown how the innumerable species inhabiting this world have been modified, so as to acquire that perfection of structure and coadaptation which most justly excites our admiration." – Charles Darwin, *Origin of Species*, 1859

At least since the appearance of Darwin's seminal work, biologists have speculated on the sources of biological variation, and many current studies have pointed to the importance of variation in gene expression as a foundational principle. Exactly what changes at a molecular level is a topic of lively interest, with important ramifications for human health. My studies of the Hairy protein, a transcriptional repressor from the fruit fly Drosophila, have revealed new

concepts on the consequence of “random” events affecting genomic interactions by transcription factors. These insights prompt us to reconsider mechanisms for the evolution of gene regulatory networks (GRNs).

One of the most important processes in biology is regulation of precise temporal and spatial use of genetic information to establish the physiological state of multicellular organisms. Proteins called transcription factors (TFs) bind to the genome and regulate the use of genetic information for embryonic development, cellular differentiation and cell fate in response to endogenous and exogenous signals. In other words, what cells are doing, how tissues work, and how organisms survive are dependent on transcriptional regulation. Therefore, understanding the mechanisms in transcription can inform and teach us about what happens when something goes wrong, which may result in diseases. TFs have to regulate gene expression at the right place at the right time (Figure 4-A-1A). In eukaryotes, this task is achieved by networks of very complex and combinatorial interactions between DNA binding proteins, co-regulators, and the matrix of DNA and histone proteins termed chromatin. Transcriptional networks represent an important evolutionary target for the development of morphological innovations. Molecular studies have demonstrated that the acquisition or loss of binding sites on DNA drive significant changes in gene expression that initiate critical evolutionary transitions (Figure 4-A-1B). Significantly, although relatively subtle changes have been linked to such important evolutionary innovations, it appears that sometimes gene expression is functionally conserved, even as

there are major changes in the structure of transcription control regions. Thus, only some rearrangements of gene control elements alter output enough to meaningfully affect biological processes.

I am using an excellent model system, the fruit fly *Drosophila melanogaster,* for the study of transcriptional networks. Since it is subject to easy manipulations, a wide range of genetic and molecular approaches have been applied to characterize regulatory interactions for several decades. Understanding the fly regulatory circuitry will help reveal similar phenomena in other animal systems, since they use closely related genes in conserved genetic pathways. In the Drosophila embryo, localized transcriptional repressors provide essential patterning information that establishes the primary anterior-posterior and dorsal-ventral axes of the organism (Figure 4-A-2). The Hairy repressor, a founding member of the Hairy/Enhancer of Split (HES) transcription factors, plays essential and conserved roles in animal development, including segmental gene patterning in the early embryo and specification of neuronal differentiation. Disruption of HES signaling is a prominent aspect of leukemia, lung and prostate cancers. Thus, elucidation of molecular mechanisms of Hairy activity could shed light on a number of important gene circuits that are prominently represented in key developmental pathways. I carried out genome-wide analysis of dynamic transformations in gene expression, chromatin modifications and transcriptional machinery to get insight into direct molecular interactions of Hairy on genome systematically.

My work revealed that Hairy removes chromatin marks associated with activators in large blocks of chromatin, at hundreds of loci throughout the genome (Figure 4-A-3). Hairy may therefore work through a dynamic competition with activators, undoing their positive effects on the chromatin states that would be necessary for RNA polymerase to engage genes to transcribe them. At the genome-wide level, an unexpected aspect of Hairy activity was observed on chromatin that may provide a pervasive and accessible entry point for evolution of novel gene regulatory switches. Metazoan TFs usually interact with thousands of regions in the genome, but only small subsets of these interactions are associated with changes in gene expression. In general, the overall view from other studies is that the majority of the interactions between TFs and genome may be non-functional, and are not important for activity of GRNs. My work demonstrates that Hairy interacts dynamically with many parts of the genome; some genes are impacted but most are not. This finding let us to propose the so-called “shotgun model” for this apparent off-target activity of TFs on chromatin modifications; many pellets are fired, but few are expected to reach the duck flying overhead. Yet the Hairy molecules that don’t “hit the target” still appear to be quite active, biochemically, inducing chromatin modifications that are similar to those seen on transcriptionally controlled loci. Hairy may be relatively nonselective about where it can attract chromatin modifying agents across the genome.

What is the significance of this chromatin modification associated with non-functional binding? For the organism, it is another instance of the extravagance of Nature –all of that chromatin modification for naught! As long as it is not particularly onerous metabolically or genetically, however, it may be the price paid for hitting the duck. “Futile cycling” by Hairy may however provide a unique mechanism for creation of new genetic switch elements; most DNA regulatory modules involve the combined action of transcriptional activators and repressors, thus these off-target sites may provide a path for evolution of novel transcriptional connections through addition of new TF binding sites. Where Hairy is busy acting as if it were shutting down a regulatory circuit by chromatin remodeling, small changes in DNA sequence that draw in existing activators may be sufficient to create a novel genetic switch, and a new connection between nodes in a genetic circuit. Thus modification of core elements of gene expression machinery may be an important answer to the question Darwin raised 150 years ago. How influential this particular mechanism may be will be the focus of future molecular work.

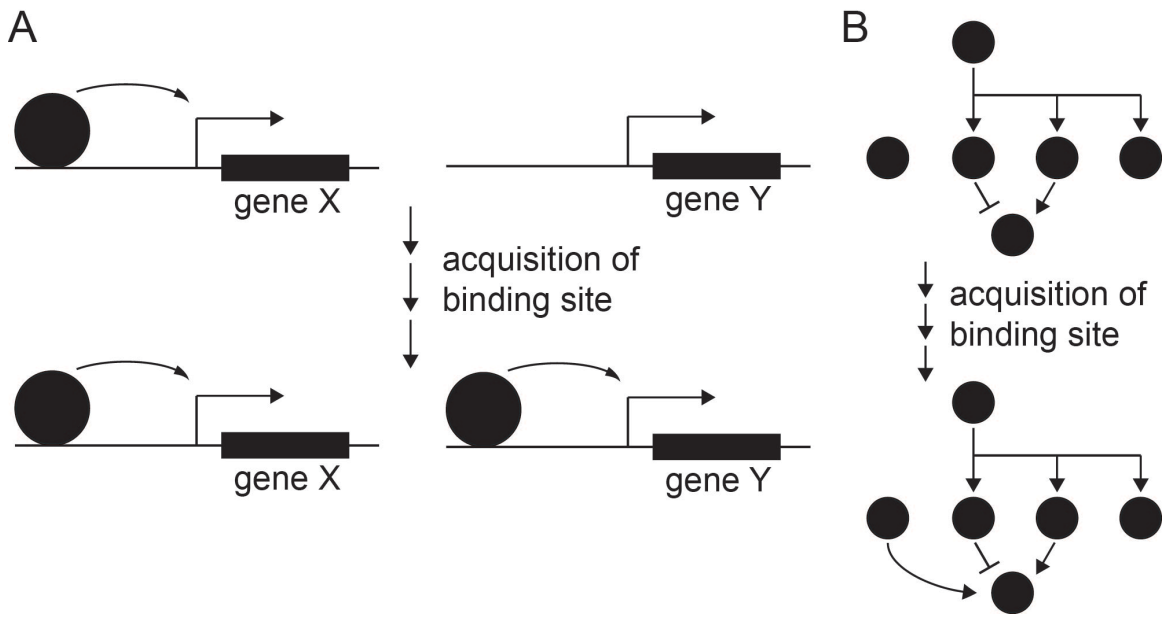


Figure 4-A-1 Evolution of *cis*-regulatory interactions through changes in DNA elements. (A) A transcriptional regulator normally regulating gene X may be recruited to an additional target gene Y by acquisition of a new binding site. (B) Interposing a new genetic link by such a modification can reconfigure a gene regulatory network (GRN).

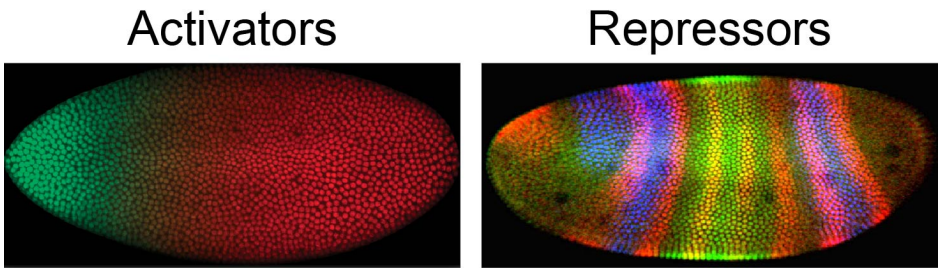


Figure 4-A-2 Patterning of the early Drosophila embryo. Distinct expression patterns of transcriptional activators and repressors, including Hairy, which is expressed in transverse stripes, spatially drive embryo patterning.

Embryo images are taken from the following sources:
https://www.microscopyu.com/featuredmicroscopist/paddock/paddockgallery.html
http://www.mun.ca/biology/desmid/brian/BIOL3530/DEVO_02/devo_02.html

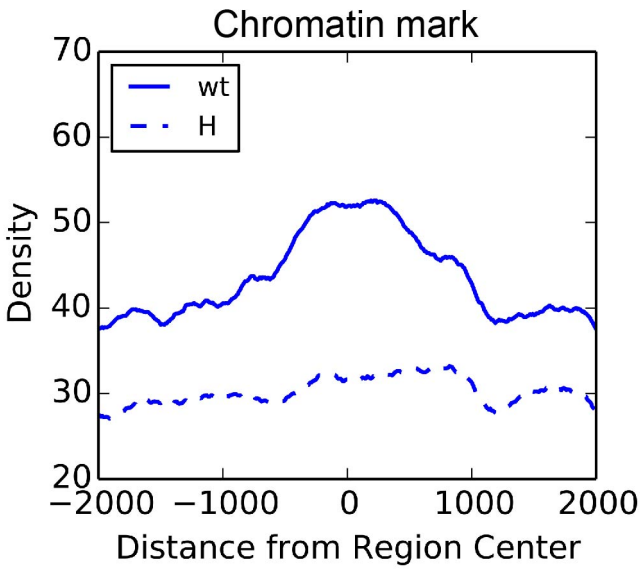


Figure 4-A-3 Chromatin marks for an "active" histone modification are lost in large blocks after expression of the Hairy transcriptional repressor. Only a small fraction of these Hairy-mediated events are associated with transcriptional regulation, however.

REFERENCES

REFERENCES

Arnosti DN, Gray S, Barolo S, Zhou J, Levine M. 1996. The gap protein knirps mediates both quenching and direct repression in the Drosophila embryo. *EMBO J* **15**: 3659–3666.

Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmann R, MacArthur S, Thomas S, Stamatoyannopoulos JA, Eisen MB, et al. 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. *Proceedings of the National Academy of Sciences* **109**: 21330–21335.

Gray S, Levine M. 1996. Short-range transcriptional repressors mediate both quenching and direct repression within complex loci in Drosophila. *Genes & Development* **10**: 700–710.

Hewitt GF, Strunk BS, Margulies C, Priputin T, Wang XD, Amey R, Pabst BA, Kosman D, Reinitz J, Arnosti DN. 1999. Transcriptional repression by the Drosophila giant protein: cis element positioning provides an alternative means of interpreting an effector gradient. *Development* **126**: 1201–1210.

Li LM, Arnosti DN. 2011. Long- and Short-Range Transcriptional Repressors Induce Distinct Chromatin States on Repressed Genes. *Curr Biol* **21**: 406–412.

Martinez CA, Arnosti DN. 2008. Spreading of a Corepressor Linked to Action of Long-Range Repressor Hairy. *Molecular and Cellular Biology* **28**: 2792–2802.

Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, et al. 2015. Enhancer Evolution across 20 Mammalian Species. *Cell* **160**: 554–566.