Balancing transcriptional activity in Drosophila through protein-protein interactions on chromatin

Balancing transcriptional activity in Drosophila through protein-protein interactions on chromatin

A dissertation presented

by

Kyle McElroy

to

The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biology

Harvard University

Cambridge, Massachusetts

April 2016

Dissertation Advisor: Professor Mitzi I Kuroda Kyle McElroy

Balancing transcriptional activity in Drosophila through protein-protein interactions on chromatin

Abstract

Chromatin plays a vital role in the implementation of gene expression programs. Several disparate groups of regulatory proteins alter chromatin state through post-translational modification of histone proteins, nucleosome remodeling, and higher order chromatin structure in order to affect gene expression. Several of these key groups, such as the Male-Specific Lethal complex and Polycomb Group have been well characterized in *Drosophila*. Yet aspects of their biology at the molecular level, such as the means by which they are faithfully targeted to regulated loci throughout the genome and the molecular mechanisms they employ to alter transcriptional state, still remain unexplained. In this dissertation I explore how identifying protein-protein interactions on chromatin reveals insights into these unanswered questions critical to chromatin biology. My results highlight the importance of balancing active and repressive chromatin states for the proper maintenance of gene expression.

The Male-Specific Lethal complex is the dosage compensation complex in *Drosophila*, which upregulates gene expression on the male X chromosome approximately two-fold. The MSL complex catalyzes an acetyl mark which may create a uniquely permissive chromatin state to promote transcriptional elongation. A proteomic screen for MSL-interacting proteins identified UpSET, the *Drosophila* homolog of yeast SET3 and mammalian MLL5. Interestingly, SET3 and UpSET have been characterized to assemble into histone deacetylase complexes. I employed genetic, genomic, and proteomic techniques to assess whether UpSET plays a role in dosage compensation. UpSET appears to play a role in limiting the level of activation of the MSL complex. Surprisingly, UpSET appears to play a more important role in the maintenance of heterochromatin.

The Polycomb Group is comprised of a well characterized set of developmental repressors. The PcG assembles into several multiprotein complexes to maintain the repressed state. The PcG is opposed by a group of activators known as the Trithorax group. Although the PcG and TrxG often appear to be recruited to the same genomic elements in different tissues, whether they might interact directly was not known. In a collaboration with Dr. Hyuckjoon Kang, I characterized the TrxG protein Female sterile (1) homeotic and found that it interacts specifically with PRC1. The data support a model that bivalency, a poised state observed in mammalian stem cells, may be critical, perhaps transiently, in the developing *Drosophila* embryo. The mechanism of coordination amongst the various PcG complexes on chromatin is not well understood. We also identified the Sex comb on midleg protein, a known member of the PcG, as a potential physical bridge between PRC1 and PRC2.

In these sets of experiments, I have characterized instances of crosstalk between activating and repressing regulators which are critical for the proper maintenance of chromatin state. Perturbations of these interactions may lead to an imbalance of regulators on chromatin and aberrant transcriptional activity. These findings highlight the need for tuning gene expression state and suggest chromatin-based mechanisms by which this can be accomplished.

Table of Contents

List of Figures

List of Tables

Acknowledgements

There are many, many people over the course of the last many, many years who have made this document possible, and it would take many, many pages to name them all and really do justice to what their contribution has meant to my life, both scientifically and personally. I will attempt to keep this succinct, perhaps most of all because it could get a little bit dusty in here.

First and foremost, I need to thank my mentor, Dr. Mitzi Kuroda. Her indispensable advice and constructive criticisms over the years have helped to shape me scientifically. She has created and nurtured a lab environment where everyone is on equal footing and there is free exchange of ideas, which certainly has helped me to grow as a scientist, and I do hope that others have found our lab interactions to be useful in their evolution as researchers as well. Within the lab, I really must thank my baymate and collaborator Dr. Hyuckjoon Kang. Without Joon, this work would not have been possible. Literally it just would not have happened. Joon is fantastically motivated and dedicated; his example has been one to strive for, and I have yet to reach it. He is also exceedingly generous with his time, having contributed beyond the call of duty for others in the lab. Dr. Artyom Alekseyenko was one of the first people I met in the lab and immediately struck me with his larger than life persona. With his expertise and direction I became acquainted with BioTAP-XL and have learned much from him. Dr. Barry Zee has been a valued resource as a technical expert for all things mass spec, able to explain its complexities with accessible ideas, and always a quick wit to bring some levity to those days when nothing seems to go right. Kuroda lab members past have also had a hand in shaping my progress toward this work, and I must thank Jumana al-Haj Abed, with whom I worked closely on revamping the nascent RNA protocol, Charlotte Wang, who initiated the UpSET project, Annette Plachetcka, who laid the ground work for my experiments on Jil-1, and Leon Lim, who helped introduce me to Drosophila, for their fundamental contributions to the direction of my research.

I would be remiss if I didn't thank the many others in the scientific community who have been a positive influence and have lent their expertise and materials to my research over the years. Thanks to the Perrimon (especially Ben Housden, Richelle Sopko, and Rich Binari), Elledge (especially Eric Wooten and Kristen Mengwasser), Winston, Bender, Wu (especially Sonny Nguyen and Brian Beliveau), Dymecki, Cepko, and Tabin laboratories. I am also grateful for the effort and patience of the HMS&BWH support staff, most especially Jeselle Gierbolini, Vonda Shannon, Kenji Kono, Tracy Dunn, the Taplin Mass Spec facility, and the SysBio FACS core. The HMS Genetics and greater Longwood community has been a warm, welcoming adoptive community from day one.

I'd also like to thank members of my home department and grad program. On the administrative end, Mike Lawrence was a tireless supporter of the MCO graduate program. Patty Perez and Fanuel Muindi have been helpful these last few months as I've tried to wrangle together this work and all the associated paperwork in time to meet submission deadlines. I am indebted to those who have served on my DAC and quals committees for their countless hours of advice and feedback: Dr. Susan Mango, Dr. Robert Kingston, Dr. Rachelle Gaudet, Dr. Andres Leschziner, Dr. Andrew Murray, and the late Dr. William Gelbart. I'd also like to thank Dr. Nicole Francis, formerly of MCB, with whom I started my graduate career. Her commitment to excellence in biochemistry is astounding, and her perseverance in the face of hardship is second to none. Former members of her lab, especially Nicole Follmer, Egbert Madamba, Leslie Beh, and Ajaz Wani, have been good friends over the years.

In addition, there are many people to thank for keeping my grounded throughout the ups and downs of grad school. My friends and classmates from the MCO program and associates, with whom I've shared in the sweet cup of victory and the bitter taste of defeat (often in the realm of table top games, but also volleyball…and I guess, science): Jimmy Z, Naked John, Srouji, Sarah, Paige, John M, Julio, Jamie, Amanda, Aaron, Jon, Alan, Nikhil, Jiang, Scott, Diego, et al. I must especially thank my long-time roommate and friend, Bumjin Namkoong, and I promise after I finish typing this I'll go do those dishes I've been neglecting. Outside of the immediate science community, I need to thank CMSC (especially Abe and Alana), Tri club (especially Trevor, Michael, and Joon), HUCA (especially Portia, George, Drew, Colin, and David Z), and the Dudley House IM community, all of whom, collectively, have made getting out of the lab and into the pool, out on the path, or onto the bike something I can enjoy.

I've been lucky enough to have a fantastic network of friends and confidants who have stayed in touch over the years. Each, in their own way, has had an influence on who I am today. Thanks to my longtime friends Xena, Laura, Pfeif, Fingerhut, and other FLOW area friends; Yalies Woo, Annie, Liam, Atisha, Adam, Alison, Sly, Brent, and SM'09 & MSW/WSW alums; Sara (who basically is my sister); Ziggy (who actually read all of this!) and Tori; Amelia, Ken, and NJWers past. These friendships have been a rock in times of hardship for which I truly am grateful.

I must thank my wonderful family, most especially, my parents, Frank and Linda, and my brother, Greg, who have been steadfast in their support throughout this long, sometimes daunting, process called grad school. Greg, it was great overlap with you here in Boston and be able to chat science and I look forward to hearing about where your studies take you. Mom and Dad, thank you for pushing me to embrace challenges and reach for a potential that sometimes I don't think I really have.

Lastly, thank you, dear reader, for taking the time to read this!

Chapter 1
Introduction

This chapter contains excerpts from McElroy, et al, 2014.

On the importance of chromatin

The central dogma of molecular biology is often summarized simply as "DNA makes RNA makes protein." Mechanistic insights at the molecular level were critically important to deciphering the process by which DNA could create RNA (and itself be copied) and by which RNA could direct protein expression. And while much of biology does indeed follow this generalized rule, it has become apparent that proteins provide feedback to the genome to influence *which* DNA gets to make RNA at any particular time. Additionally, numerous examples now exist of non-protein-coding RNAs being essential for proper gene expression and genome organization.

The level at which such proteins and non-coding RNAs influence gene expression is typically not at the level of altering DNA primary sequence. Such modification of the sequence of DNA would be untenable for the integrity of individual genes and insufficient for the rapid responses needed in biological systems. Instead, protein and RNA factors control the access of transcriptional complexes to DNA, thereby controlling gene expression. In the eukaryotic system, DNA is wrapped around an octameric core of histone proteins to assemble a unit called the nucleosome. Additional proteins and RNAs interact with nucleosomes and the associated DNA in the nucleus to form what is called chromatin.

Given the complexity of nuclear content in eukaryotes, it is lucky, then, that the original studies of the structure of DNA, which were so vitally important to working out the central dogma of molecular biology, made use of systems that generally lacked the higher order packaging of DNA of eukaryotes. Certainly the beautiful X-ray crystallography data generated by Wilkins, Franklin, and their colleagues would not have been interpretable had the DNA been chromatinized, and the seminal synthesis of these data into their model of the DNA double helix by Crick and Watson would not have been possible.

From this technical standpoint to the reality of molecular biology and genetics, it is abundantly clear that chromatin plays a key role in the cell. In addition to packaging DNA into the nucleus, the

milieu of chromatin associated proteins affects the stability of the genome, the organization of the chromosomes, timing of replication, transcriptional activity, and many other processes. Given the diversity of proteins and biochemical properties that associate with chromatin, a truly systematic study of each gene is impractical for any one researcher. Instead a common approach is to focus on groups of genes critical for specific biological processes.

Chromatin-based modulation of gene expression

One critical process for an organism is the maintenance of the transcriptional activity state of developmentally important genes. These genes have their spatial pattern of expression set during the course of development by certain transient transcription factors. Following the disappearance of these transcription factors, whether the gene is on or off needs to be preserved in the cell and any of its daughters. Epigenetic memory, that is, heritable changes in gene expression not dependent on changes to DNA primary sequence, is executed by two groups of proteins: the Polycomb group (PcG), responsible for maintenance of repression, and the Trithorax Group (TrxG), responsible for maintenance of the active state (Schuettengruber et al., 2011, Simon and Kingston, 2013, Steffen and Ringrose, 2014).

Another critical process for organisms is that of dosage compensation. Dosage compensation is the process by which the expression levels from sexually dimorphic chromosomes are tuned to balance that of autosomes. The loss of dosage compensation is highly detrimental to one or the other sex, as the proper ratios of expression of sex-linked factors become skewed. While there are several strategies to achieve dosage compensation, all strategies rely on a chromatin-based mechanism. In *Drosophila melanogaster*, a core complex of five proteins and two non-coding RNAs localize specifically to the male X chromosome and boost the expression of active genes by approximately 2-fold via the establishment of a unique chromatin state that is permissive for efficient transcriptional elongation (the Male-Specific Lethal complex, [MSL])(Chen et al., 2014, Lucchesi and Kuroda, 2015).

The overarching theme relevant to the above examples is the proper balance between activation and repression through chromatin state. In eukaryotes, the nucleus is generally organized into two types of chromatin: euchromatin and heterochromatin (Figure 1-1). Active genes are found in the euchromatic regions of the genome, and silenced genes are found in heterochromatin. Furthermore, heterochromatin can be further subdivided into constitutive and facultative heterochromatin. The differences between these flavors of heterochromatin lie in the proteins that mediate their formation and the general genomic location. The Polycomb group is the main player in facultative heterochromatin, which intermixes with euchromatin on the long chromosome arms, whereas the Heterochromatin protein 1 (HP1) family forms constitutive heterochromatin, which serves a highly structural role primarily in centromeric and pericentromic regions. The work presented herein documents several interactions between activating and repressive proteins, and characterizes the effect these interactions have on the balance between euchromatin and heterochromatin.

Transcription and Chromatin

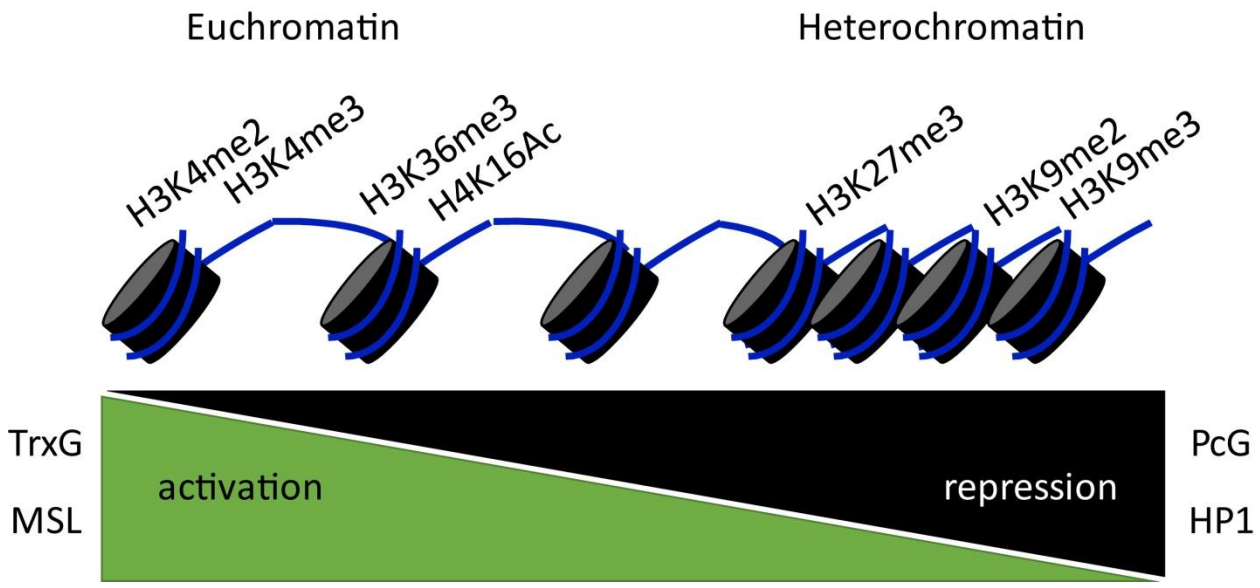
Transcription is the process by which a messenger molecular, RNA, is made from DNA, the ultimate repository of genetic information. Transcription is, by itself, a rather complicated process, without even considering the context of chromatin in which is must take place. The major agent of gene expression in the RNA Polymerase II holoenzyme. This multisubunit machine has modules dedicated to each step of transcription: from the melting of DNA strands to yield a template for RNA synthesis to proofreading to ensure fidelity of nucleotide incorporation to processivity of the synthesis reaction (Vannini and Cramer, 2012). Furthermore, each step is heavily regulated and integrates feedback from the rest of the cell to determine when and how much transcription occurs.

Transcriptional regulation can be split into three main phases: initiation, promoter clearance, and elongation. Initiation often occurs as a cascade of recruitment steps. Initially, sequence-specific

Figure 1-1: The classical view of euchromatin vs heterochromatin

Euchromatin and heterochromatin differ in the density of nucleosome packaging and in their complements of non-histone proteins and enrichments of histone post-translational modifications. Euchromatin is thought of as a much more open structure, whereas heterochromatin is more densely packed, with less DNA accessibility. Euchromatin is enriched for proteins such as the Trithorax Group (TrxG) and, in males, the Male-Specific Lethal complex (MSL). These proteins reinforce active genes and deposit histone post-translational modifications associated with active loci such as those indicated. Heterochromatin covers repressed genes and is enriched for members of the Polycomb group (PcG) and Heterochromatin Protein 1 (HP1). These two groups of repressive factors have their own associated histone marks such as Histone H3 lysine 27 trimethylation for the PcG and Histone H3 lysine 9 di- and tri-methylation for HP1. The typical dogma is that these proteins and modifications are antithetically opposed and do not associate with each other.

Figure 1-1 (Continued)



transcription factors (TFs) bind to DNA in response to stimuli or other cellular conditions. These sequence-specific factors then recruit general transcription factors (GTFs). GTFs are critical for the loading of RNA polymerase II (Pol II) and the initiation of RNA synthesis (Vannini and Cramer, 2012). The next phase of transcriptional regulation occurs when Pol II become paused, usually within a few hundred nucleotides from the transcriptional start site (TSS). Originally thought to be a rare event, new experimental technologies have revealed that promoter-proximal pausing is a very common occurrence throughout the genome (Price, 2008). Pausing is induced by negative elongation factor (NELF) and DRB sensitivity inducing factor (DSIF), which stalls engaged Pol II (Yamaguchi et al., 2013). Pausing is relieved by the cyclin-dependent kinase P-TEFb, which phosphorylates NELF, causing its dissociation, and DSIF, causing it to become a pro-elongation factor (Guo and Price, 2013, Kwak and Lis, 2013). At this point, Pol II enters the phase of productive transcriptional elongation. Upon reaching the end of the transcript, Pol II encounters a polyA signal, which causes the association of termination factors. Proper termination again requires P-TEFb activity (Laitem et al., 2015). The different stages of transcription can be identified using the modification state of the C-terminal domain (CTD) of the Pol II enzyme. The CTD of Pol II consists of a variable number (by organism) of heptad repeats, with many of the residues being post-translationally modified at different times throughout the transcriptional cycle (Egloff et al., 2012, Phatnani and Greenleaf, 2006).

All steps of transcription can be sensitive to the chromatin environment. One classic paradigm for this is the regulation of genes by nucleosome occlusion of the promoter. Nucleosomes assembled over the promoter sequence present a physical barrier to the association of sequence-specific TFs and/or GTFs with DNA (Gilchrist and Adelman, 2012). ATP-dependent chromatin remodelers can resolve this and promote transcription by sliding or evicting nucleosomes to allow DNA access. There are also chromatin looping events between the promoter and enhancer elements distal to the gene loci that greatly affect the efficacy of transcription. An important player in these sort of long-range

interactions is the Mediator complex. The Mediator complex is made up of ~30 proteins and facilitates the crosstalk between enhancer and promoter by physically bridging the looped structure, though it may also modulate Pol II function more directly (Ansari and Morse, 2013). Mediator complex can be broken down into several different modules, based on functionality, copurification, and reconsititutional structural studies (Lariviere et al., 2012). Recent work in the field of Mediator structure has synthesized these various analyses to reveal a detailed physical map for the complex and its association with Pol II (Plaschka et al., 2016).

Once engaged, nucleosomes may present physical barriers to the progression of Pol II. It has been observed using a variety of nucleosome mapping techniques that the first few downstream nucleosomes are typically well phased (Zhang et al., 2011). Whether the +1 nucleosome actually represents a barrier is somewhat debated in the field, since, while nucleosome mapping techniques clearly show strongly positioned nucleosomes, other techniques suggest these nucleosomes are in a state of high turnover (Dion et al., 2007). Some of this turnover may be mediated by the exchange of canonical histone subunits for variant histones, which may loosen the wrapping of DNA around the nucleosome (Ranjan et al., 2015, Weber et al., 2014). The presence of Pol II also has an effect on the spacing and phasing of nucleosomes. It has been proposed that the promoter-proximally paused Pol II itself promotes a phased nucleosome architecture. The mechanism for this is the fact that the paused Pol II complex is highly stable on DNA, and therefore functions as a barrier for the thermodynamic repositioning of nucleosomes back over favorable DNA sequences (Gilchrist et al., 2010, Gilchrist et al., 2008). This poises the gene for expression, provided that the next Pol II holoenzyme is recruited quickly, before nucleosomes can relax back to positions which prevent the next round of transcription.

Another consideration for the status of transcription is the chromatin state beyond simple nucleosome positioning. Chromatin packages meters of DNA into a sphere that is only microns in diameter, requiring a high degree of compaction. Nucleosome density and higher order folding of

chromatin can have both positive and negative effects. As mentioned above, higher order folding interactions bring distal enhancers adjacent to promoters which modify, typically in a positive fashion, transcription. Similarly, co-regulated genes may physically associate in subnuclear locations. This sort of folding does not always activate transcription, however, as repressive complexes also may loop to promoters and silence transcription (Kassis, 2002).

From the physical standpoint, too, more densely packed nucleosomes restrict the access of transcription factors to DNA. Perhaps the clearest example of this is during mitosis. Chromosome condensation into metaphase chromosomes poised for proper segregation generally inhibits transcription as most chromatin associated factors are evicted (Delcuve et al., 2008). A good deal of this is undoubtedly due to the extreme compaction that is required as chromatin folds from nucleosomes to the higher order 30nm fiber and beyond. While this is an extreme example, clearly the state of chromatin has profound consequences on transcription.

Chromatin state and the histone code

A key aspect in our understanding of chromatin state is the complement of histone post-translational modifications which exist in the cell. The histone proteins which make up the core octomer around which DNA is wound also extend their N-terminal tails out from the structure. These tails can be highly modified by numerous enzymes and can serve as a scaffold for protein-protein interactions on chromatin, or can change the biophysical properties of nucleosomes themselves (Zhao and Garcia, 2015). These modifications reveal insights as to the state of chromatin domains and the underlying transcriptional state of the genes contained within them. The idea of the histone code was that if one could accurately assess, predict, and/or alter the locations of histone post-translational modifications (PTMs), it would be possible to infer the state of genes (Jenuwein and Allis, 2001). Unfortunately, the idea of the histone code in a binary sense (ie, *if* this modification, *then* this *must* happen/be the case) is

not quite true. Even combinatorially, the exact input and output of histone modifications remains elusive, yet the overall correlations, as described below, have greatly advanced the field of molecular biology and molecular genetics.

Several key advances in characterizing the state of chromatin have employed large scale ChIP experiments followed either by microarrays or sequencing. By surveying a wide variety of histone PTMs, it has been possible to identify coincident marks, genome feature associations, and co-enriched chromatin proteins. In this way, chromatin state can be broken down into a number of classes based on the relative enrichment of a panel of histone modifications. For example, using 18 histone marks, Kharchenko and colleagues identified 9 different types of chromatin state (Kharchenko et al., 2011). High co-enrichment of H3K4me2/me3, and H3K9ac was the observed signature for the TSS of expressed genes. The signature for transcriptional elongation was evident by H3K36me3 across gene bodies. The chromatin signature often found over introns was enrichment of lower methylation states of H3K36 and enrichment of H3K27ac. H3K27ac is most strongly associated in the literature with enhancers; that some enhancers lie within introns may account for this enrichment. A unique form of active chromatin carries enrichment for H4K16ac, which, in Drosophila, is strongly enriched on the dosage compensated male X chromosome due to the activity of the MSL complex. Three different heterochromatin states were classified. One is enriched for H3K27me3, corresponding to PcG facultative heterochromatin, while the other two are enriched to different degrees with H3K9me2 and H3K9me3, corresponding to two types of HP1 constitutive heterochromatin. Ho, Jung, Liu, and colleagues performed similar analyses using 8 histone marks across humans, flies, and worms, and identified 16 states (Ho et al., 2014). Many of these are similar to the classifications discussed above, with the caveat that they often parse the above into more finely separated states (ie, transcription at the 5’ end of a gene vs 3’ end).

Beyond the ability to scaffold chromatin proteins, some histone PTMs are known to change the properties of nucleosomes. One mechanism for the compaction of chromatin into the 30nm fiber

discussed in the previous section is packing interactions between adjacent nucleosomes. This is achieved by the interaction between a negatively charged "acidic patch" of residues of the H2A and H2B histone subunits with the N-terminal tail of histone H4 (Kalashnikova et al., 2013). Like other histone N-terminal tails, H4 is quite basic (numerous lysines provide a positive charge). Packing between adjacent nucleosomes requires the H4 tail-H2A/H2B acidic patch interaction to neutralize charges and allow close proximity of nucleosomes. The H4K16ac PTM mark introduces an acetyl group onto the H4 tail. In addition to the physical space of the acetyl molecule, it also introduces a negative charge. Crystal structure analysis shows the H4K16 residue positioned centrally over the acidic patch during adjacent nucleosome interactions. The introduction of the negative charge from the acetyl group and the central location of the H4K16 residue in the acidic patch groove leads to a complete loss of the H4 tail/acidic patch interaction (Shogren-Knaak et al., 2006). Without this interaction, nucleosomes are unable to pack closely, and chromatin remains more extended and open. Intriguingly, this same modification is facilitated by the Male-Specific Lethal complex in Drosophila, which creates a unique, highly active chromatin state (Gelbart et al., 2009).

The Male-Specific Lethal Complex in Drosophila

The MSL complex regulates dosage compensation in Drosophila. Dosage compensation is the means by which X chromosome gene expression is adjusted to balance gene expression from the autosomes. In flies, this is achieved by hypertranscription of the active genes on the single male X chromosome. The MSL complex is made up of 5 core proteins (Figure 1-2A) and two redundant non-coding RNAs (*roX1* and *roX2*). The loss of any protein component or both RNAs leads to male lethality. Downstream of the sex determination cascade, the MSL complex is assembled in males only and targeted exclusively to active genes on the X chromosome. Recruitment to the male X appears to occur in at least three steps (Figure 1-2B). Complex assembly is initiated on the nascent *roX* RNAs at their sites

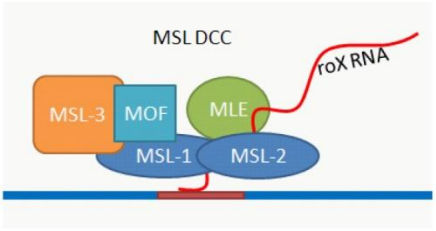
Figure 1-2: The Male-Specific Lethal Complex is the Dosage Compensation Complex for Drosophila

A) The Male-Specific Lethal complex is made up of 5 proteins (MSL-1, -2, -3, MOF, and MLE) and 2 non-coding *roX* RNAs (*roX1* and *roX2*).

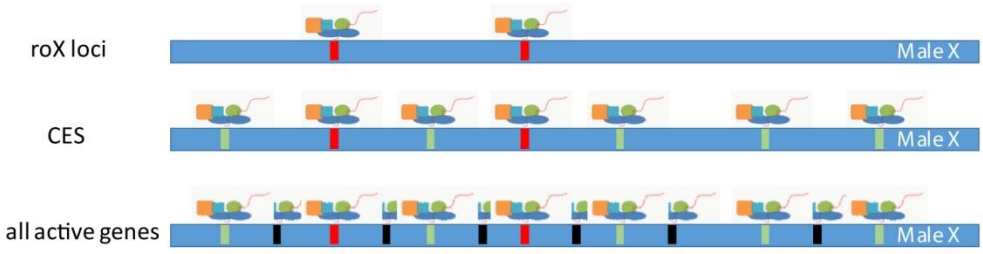
B) The MSL complex is assembled at *roX* loci and recruited to high affinity chromatin entry site (CES) on the male X chromosome. The MSL complex then spreads to all active genes on the X chromosome where it facilitates the hypertranscription of genes to achieve dosage compensation. Figure adapted from McElroy et al, 2014.

Figure 1-2 (Continued)

A



B



Adapted from McElroy et al 2014

of transcription. The second step involves targeting to several hundred chromatin entry sites (CES, also called high affinity sites, HAS) carrying a degenerate sequence motif. The third targeting step involves sequence-independent spreading *in cis* to most active genes. It is for this third step that *roX* RNAs seem most critical (reviewed in (Gelbart and Kuroda, 2009)). Evidence for this "nucleate and spread" model comes from experiments in which the MSL complex becomes targeted to active genes flanking the ectopic insertion of *roX* RNA transgenes on autosomes (Kelley et al., 1999, Larschan et al., 2007). Therefore, the initial targeting of the complex to X chromosome entry sites is critical for its specificity. However, understanding the mechanism for the selection of the initial chromatin entry sites has been challenging, since the associated MSL-response element (MRE) sequence motif is enriched <2 fold on the X chromosome versus the autosomes.

There are two main enzymatic properties known to be required for MSL function. The first, RNA helicase activity, is conferred by Maleless (MLE) (Lee et al., 1997), which recent work suggests may be critical for *roX* RNA remodeling and complex assembly (Maenner et al., 2013). The second, histone acetyltransferase activity directed towards H4K16, is catalyzed by Males absent on the first (MOF) (Akhtar and Becker, 2000, Hilfiker et al., 1997), which is likely to be key to the increase in transcriptional activity of male X-linked genes. While these activities are both essential for MSL function, they appear to be dispensable for the initial targeting of the complex. More recently, the groups of Dou and Becker showed that MSL2 has E3 ubiquitin ligase activity in vitro. Wu et al (2011) provided evidence that mammalian and fly MSL2 had activity in association with MSL1, leading to ubiquitination of H2B K34 (H2B K31 in flies)(Wu et al., 2011). In contrast, Villa et al (2012) found that Drosophila MSL2 ubiquitinates other MSL components, likely serving a stoichiometry-balancing role (Villa et al., 2012). Whether either of these functions is essential for dosage compensation has not yet been reported.

Non-enzymatic domains that have chromatin interaction capabilities are present as well. MSL3 and MOF both contain chromodomains, which are found in various chromatin-modifying proteins and

often interact with methylated histones. The MSL3 chromodomain is characterized to have H3K36me3 binding (Larschan et al., 2007), H4K20me1 binding (Kim et al., 2010), and H4K20me2 binding (Moore et al., 2010). The MSL3 chromodomain plays a role in the spread of MSL to all active genes, but, like the enzymatic activities above, is dispensable for initial targeting (Sural et al., 2008). Unsurprisingly, MLE contains several RNA-interacting motifs, including a double-stranded RNA binding domain, a DExH helicase domain, and a C terminal glycine rich region, which could be used for engaging chromatin via RNA (Izzo et al., 2008, Morra et al., 2008).

The CXC domain of the MSL2 protein has been shown to have affinity for DNA in vitro (Fauth et al., 2010). Structural analysis of this domain bound to non-specific or MRE DNA suggests that this domain may mediate recognition of the motif. Indeed, there appeared to be two modes of DNA interaction for the CXC domain: a lower affinity "scanning" mode and a higher affinity mode engaged on the MRE (Zheng et al., 2014). However, other experiments revealed another factor to be critically important for MSL complex recruitment. In an RNAi screen for factors influencing MSL complex recruitment, the gene *CG1832* was identified (Larschan et al., 2012). It was later shown that the CG1832 protein, renamed CLAMP (chromatin-linked adaptor for MSL proteins), facilitated recruitment of the MSL core complex and directly recognized the MRE motif through an array of zinc finger domains (Soruco et al., 2013). From CES, the complex binds at all active genes on the X via a spreading mechanism that includes recognition of the H3K36me3 histone mark for active transcription by MSL3 (Larschan et al., 2007).

Once bound at active genes on the X chromosome, the MSL complex achieves dosage compensation by increasing the transcriptional output of these genes. The molecular mechanism by which this occurs could, hypothetically, proceed by changes in initiation of transcription or by changes in the elongation rate or processsivity. Initial genomic analyses disagreed on the key steps in transcription that were most affected (Conrad et al., 2012, Larschan et al., 2011). Subsequently, a computational

error in Conrad et al (2012) that greatly overstated the differential in Pol II recruitment to promoters in males, was noted in the literature (Ferrari et al., 2013a, Straub and Becker, 2013, Vaquerizas et al., 2013). Along with additional data, enhanced transcriptional elongation, through the creation of a uniquely permissive chromatin environment by the MSL complex, is the model currently favored in our laboratory (Ferrari et al., 2014, Ferrari et al., 2013b).

One hurdle in the study of the MSL complex is that traditional biochemical methods fail to purify intact complexes. The reasoning for this is that members of the MSL complex are extremely sensitive to the conditions used to solubilize the complex off of chromatin. Members of the Kuroda lab developed a technique to affinity purify a bait protein and all of its interactors from crosslinked chromatin and identify them using mass spectrometry (Wang et al., 2013). From these same chromatin samples, the genomic localization of the bait protein can also be assessed (Alekseyenko et al., 2014a, Alekseyenko et al., 2014b, Alekseyenko et al., 2015). When applied to the MSL complex, all genetically known members were recovered among the top hits (Wang et al., 2013). Of the novel factors identified, I further explored how one, UpSET (CG9007), might contribute to dosage compensation, and also found an unexpected connection to heterochromatin (see Chapter 2).

The Polycomb Group

The Polycomb Group (PcG) is a set of proteins critical for the maintenance of repression of developmentally-silenced genes. The members of the PcG were originally characterized in the fruit fly *Drosophila melanogaster* as being necessary for the maintenance of the parasegment specific pattern of *Hox* gene expression (Lewis, 1978). Mutations in PcG genes lead to homeotic transformation in the fly. Homeotic transformations occur when key developmental *Hox* genes, which specify the body plan, are transcribed outside of their proper domain of expression (Jurgens, 1985, Struhl, 1981). For example, many PcG mutants give rise to ectopic sex combs, a specialized hair-like structure, typically only found

on the first legs of males. These ectopic sex combs are usually observed on the second leg, but also sometimes on the third; they may also ectopically appear on tarsal segments, indicating defects in the proximal to distal axis as well. The incidence of these ectopic sex comb structures are due to the improper activation of the homeotic gene *Sex combs reduced* (*Scr*) in regions where it normally is silent (Lewis et al., 1980, Pattatucci and Kaufman, 1991, Struhl, 1982).

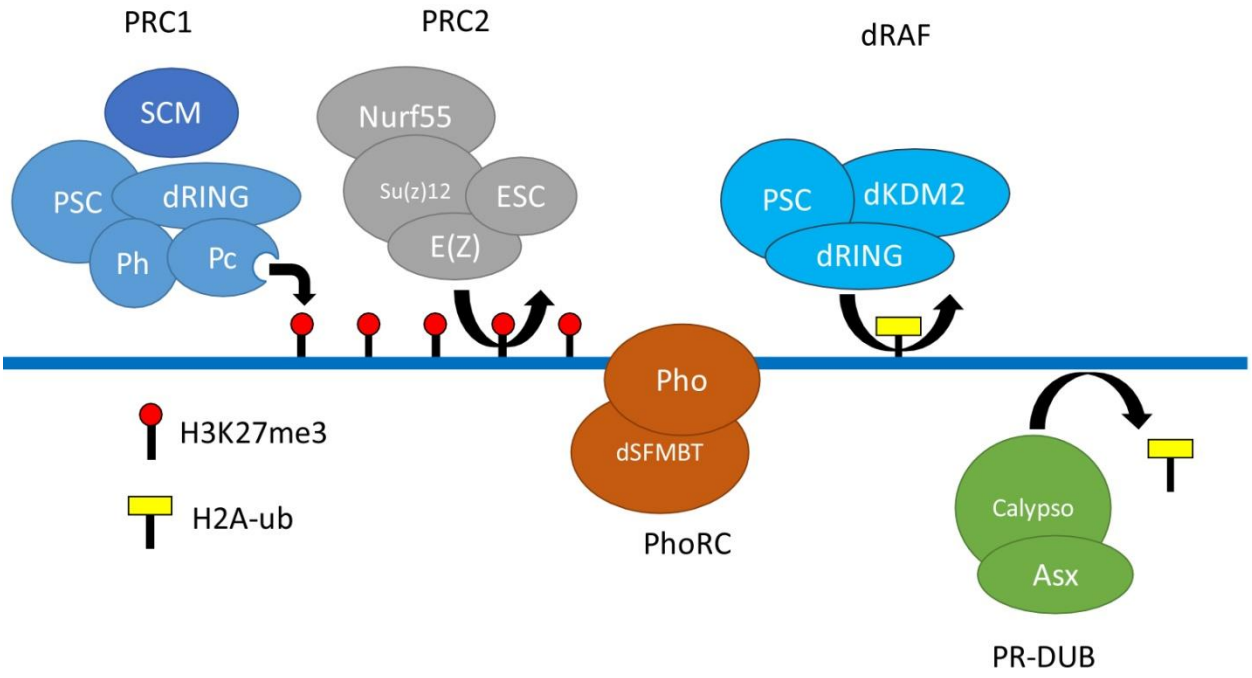
Further investigation of these proteins has identified hundreds of additional PcG-bound genes in Drosophila, suggesting that the PcG functions as a general repressor (Negre et al., 2006, Schwartz et al., 2006, Tolhuis et al., 2006). In the mammalian system, the PcG has been implicated in processes ranging from the cell cycle to cancer. At individual target genes in flies, Polycomb response elements (PREs) have been identified that can function in ectopic chromatin contexts, but these lack a strong consensus motif (Kassis and Brown, 2013). A classical model for the targeting of PcG complexes is that they recognize PREs in silenced domains that were previously established by repressive, spatially restricted transcription factors. Once PcG complexes are initially targeted to PREs, they can be stably maintained at these loci even after the original silencing factors are no longer expressed (Follmer et al., 2012, Francis et al., 2009, Lengsfeld et al., 2012). Like the MSL complex, the PcG may also have a spreading mechanism, as silenced regions can form large PcG-associated domains (Tolhuis et al., 2006). The creation of these domains, which can differ from cell type to cell type, is not fully understood, but recent work by members of our lab suggests that these domains spread to fill topologically-associated domains, which correspond to insulated neighborhoods created by the three-dimensional folding of the genome (Jung et al., 2016).

A hallmark of the PcG is that they form several distinct multiprotein complexes, with an array of different biochemical and enzymatic activities, in order to properly maintain silencing (Figure 1-3). A substantial number of the known Polycomb group proteins (PcG) have been identified as subunits of two main PcG complexes, Polycomb repressive complex 1 and 2 (PRC1 and PRC2). PRC1 complex is

Figure 1-3: The PcG forms multiprotein complexes that maintain gene repression

The Polycomb Group (PcG) forms several biochemically distinct multiprotein complexes. These complexes have an array of biochemical and enzymatic properties. PRC1 binds to H3K27me3 via the chromodomain of Pc, while the SET domain of E(z) in PRC2 catalyzes that methyl mark. PhoRC recognizes DNA in a sequence specific manner. The dRAF complex catalyzes ubiquitylation of histone H2A, whereas the PR-DUB complex catalyzes its removal. The diversity of PcG complexes work in synergistic ways to reinforce a repressed transcriptional state. Figure adapted from McElroy et al, 2014.

Figure 1-3 (Continued)



Adapted from McElroy et al 2014

thought to act as a direct executor of target gene silencing through inhibition of chromatin remodeling and chromatin compaction (Shao et al., 1999). Polycomb (Pc), Polyhomeotic (Ph), Posterior sex combs (Psc) and Sex comb extra (Sce, aka dRing) are the core subunits of PRC1 (Francis et al., 2001). Suppressor of zeste 2 (Su(z)2), which is functionally redundant with Psc (Lo et al., 2009), and Sex comb on midleg (Scm) co-purify with PRC1 at sub-stoichiometric levels and are also categorized as PRC1 subunits. RNA interference (RNAi) knockdown experiments in tissue culture cells have suggested that Scm may be particularly important for PcG complexes at PREs (Wang et al., 2010).

The PRC2 complex is responsible for trimethylation of Histone H3 lysine27 (H3K27me3) (Cao et al., 2002, Czermin et al., 2002, Kuzmichev et al., 2002, Muller et al., 2002), a mark of transcriptional repression which is known to be recognized by the chromodomain of Pc. PRC2 consists of Enhancer of zeste (E(z)) as its catalytic subunit and Suppressor of zeste 12 (Su(z)12), Extra sex combs (Esc), and Nurf55 (aka Caf1) as noncatalytic subunits. Several other PcG complexes have been identified in Drosophila. dRAF (dRING-associated factors) complex, which shares Psc and Sce/dRing with PRC1, contains the demethylase dKDM2 and is involved in H3K36me2 demethylation and H2A ubiquitylation (Lagarou et al., 2008). Polycomb repressive deubiquitinase (PR-DUB), another PcG complex, consists of Additional sex combs (Asx) and the ubiquitin carboxy-terminal hydrolase Calypso, which specifically removes monoubiquitin from histone H2A (Scheuermann et al., 2010). Classical genetic experiments imply that the cycling of ubiquitylation of H2A is important for PcG repression.

The Pho-repressive complex (PhoRC), consisting of the DNA binding proteins Pleiohomeotic (Pho) or Pleiohometic-like (Phol) together with Sfmbt (Scm-related gene containing four mbt domains), is the only PcG complex shown to have sequence specific DNA binding activity (Alfieri et al., 2013, Klymenko et al., 2006). Evidence for an interaction between PRC2 components and the DNA binding proteins Pho and Phol, as well as the requirement of Pho in PRE binding of E(z), led to a model of hierarchical recruitment of PcG complexes (Wang et al., 2004). In this model, Pho and Phol bind to PREs

and recruit PRC2 complex to PREs through their interaction. Subsequently, E(z) methylates H3K27, which results in the recruitment of PRC1 by the recognition of the histone mark by the Pc chromodomain. However, this simple model is not sufficient to explain PcG silencing. PRC1 and PRC2 components are still visible by immunostaining at many sites on polytene chromosomes in *pho* and *pho-like* double mutants (Brown et al., 2003), suggesting that additional DNA binding factors are likely to be involved in PcG recruitment and silencing.

In a collaboration with Hyuckjoon Kang, a postdoc in the lab, we sought to address unknown aspects of PcG biology. While biochemically distinct PcG complexes have been identified, we were interested in whether these complexes were similarly separable when found in the chromatin context. In the course of this exploration, we identified an already-known PcG factor, Scm, that physically interacted with both PRC1 and PRC2, which we suggest allows for coordination of these complexes for the proper maintenance of silencing (see Chapter 3). We also were interested in whether there was interplay between the PcG and the Trithorax Group (TrxG, see below), which have opposing functions, yet are found at overlapping sites throughout the genome. Our data support an interaction between PRC1 and the TrxG protein Fs(1)h, and suggests an interrelatedness of these two groups during the course of development (see Chapter 3).

The Trithorax Group

The Trithorax Group serve a very similar role in development as the PcG, except with the opposite effect. Whereas the PcG maintains the repression of *Hox* (and many other) genes, the TrxG maintains the expression of these genes. The TrxG was originally defined as those genes whose function was to antagonize PcG-mediated silencing (Kingston and Tamkun, 2014). Due to their role maintaining the expression pattern of *Hox* genes, *TrxG* genes often have homeotic phenotypes of their own, typically the reverse of PcG in that posterior to anterior transformations are observed. One example of this is the

transformation of the haltere (a rudimentary wing structure on the third thoracic segment) to wing (second thoracic segment). This striking phenotype is due to the failure to maintain expression of the *Ubx* gene (which normally functions in the haltere to suppress the wing fate) (Lewis, 1978). The model for TrxG recruitment mirrors the model for PcG recruitment, where TrxG proteins are brought to domains initially activated by developmental factors. Some of the same recruitment machinery, but in different chromatin context, may function to recruit both the PcG and TrxG (Kassis and Brown, 2013).

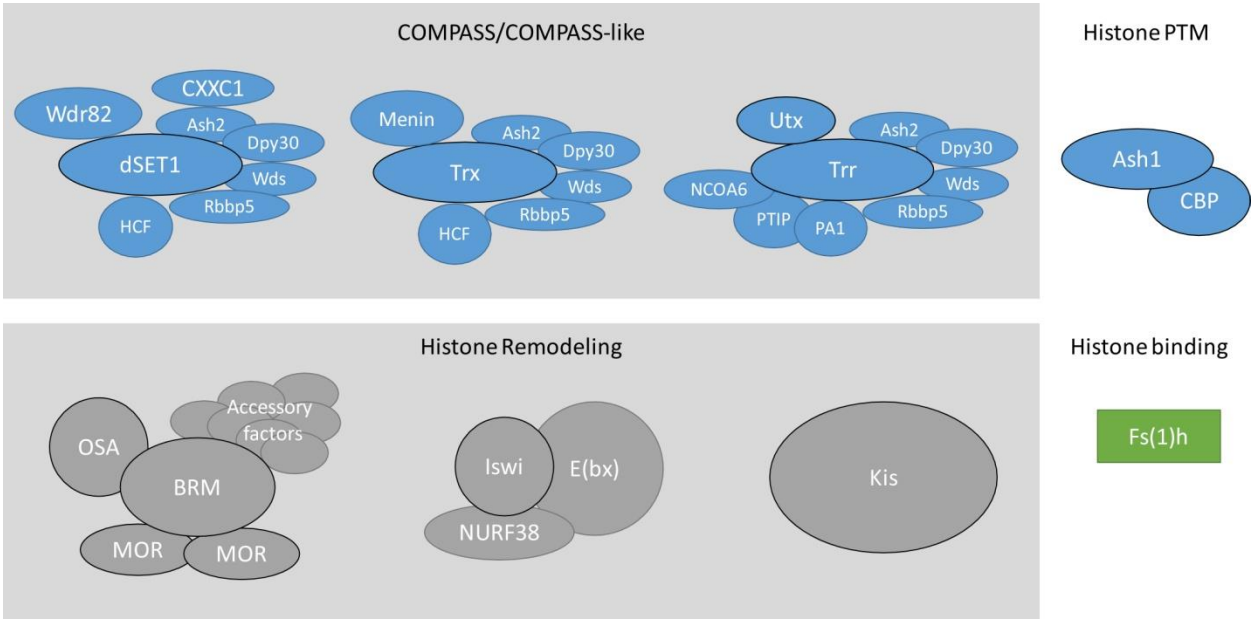
Once again similar to the PcG, the TrxG assembles into several multiprotein complexes which achieve maintenance of activation via chromatin remodeling and nucleosome post-translational modification (Schuettengruber et al., 2011) (Figure 1-4). The histone modifying complexes affect two types of marks: methylation and acetylation. Methylation marks are deposited by three complexes whose core catalytic subunits contain SET domains. These complexes are known as the COMPASS or COMPASS-like complexes and they share several accessory factors such as Ash2, DPY30, RBBP5, and WD5 (Shilatifard, 2012). Each complex is differentiated by its own unique catalytic subunit and additional specific accessory factors. The catalytic proteins are Set1, Trx, and Trr, and catalyze the methylation of histone H3 on lysine 4. Histone H3 lysine 4 trimethylation (H3K4me3) is known to be enriched at promoters and the transcriptional start site, with regions of H3K4me2 flanking it. While all three of these SET domain proteins are competent for the deposition of H3K4 methylation, Set1 is responsible for the majority of it throughout the genome, with Trx and Trr catalyzing the mark at more specific subsets of sites.

Interestingly, the Trr/COMPASS-like complex has an additional catalytic factor: the demethylase UTX (Schuettengruber et al., 2011, Shilatifard, 2012). UTX has demethylase activity directed toward H3K27, effectively erasing the PcG-catalyzed mark. The Trr complex, unlike Set1 and Trx, appears to be responsive to several signaling pathways, and thus encountering target sites in a different chromatin context than Set1 and Trx, but it remains unclear why the Trr complex would carry along this function

Figure 1-4: The TrxG forms multiprotein complexes that maintain gene activation

The Trithorax Group (TrxG) forms many multiprotein complexes to reinforce the active state of gene expression. In general these complexes either post-translationally modify histones, or remodel chromatin. COMPASS/COMPASS-like complexes are responsible for H3K4me3 via the catalytic subunits dSet1, Trx, and Trr. The Trr complex also carries Utx, an H3K27-directed demethylase. Ash1 methylates H3K36 and its partner CBP is an acetyl transferase. Chromatin remodelers from the Swi/Snf (BRM, OSA, and MOR), Iswi (Iswi), and Chd (Kis) families also facilitate active transcription by remodeling chromatin around the promoters of active genes. The BET protein Fs(1)h recognizes active chromatin via its double bromodomains, but has not be found to be a stoichiometric member of any multiprotein complexes.

Figure 1-4 (Continued)



while Set1 and Trx complexes do not. Another activity that is thought to oppose the PcG is the methylation of H3K36 in active gene bodies. This methyl mark is catalyzed by a different SET-domain TrxG protein, Ash1. Ash1 also forms a complex with the protein dCBP/p300. dCBP performs yet another catalytic activity, namely the acetylation of H3K27 and possibly other histone and non-histone residues. The H3K27ac modification is most closely associated with active enhancers, and is mutually exclusive with the PcG-deposited H3K27 methylation mark.

Additional TrxG complexes are capable of modulating the structure of chromatin by remodeling nucleosomes. These ATP-dependent processes slide nucleosomes to create nucleosome-depleted promoters and may help to regularly space the first few nucleosomes of the transcriptional unit. Other ATP-dependent activities include the exchange of histone variants associated with transcriptional activation and labile nucleosomes. The Swi/Snf, Iswi, and CHD family of chromatin remodelers are all members of the TrxG and each have unique functions and differ in the means of their recruitment to target loci (Kingston and Tamkun, 2014, Schuettengruber et al., 2011).

My work on the TrxG focuses on Female sterile (1) homeotic (Fs(1)h). As discussed above, we identified Fs(1)h as a possible interactor with the PcG complex PRC1. Unlike many of the other TrxG proteins, Fs(1)h has not been identified as a constitutive member of a multiprotein complex, and functions as a co-activator contacting numerous transcription-related complexes near transcriptional start sites. Our data support the validity of an Fs(1)h-PRC1 interaction and suggest that this interaction may be important during gene activity transitions during the course of development (see Chapter 3).

Constitutive Heterochromatin and HP1

Constitutive heterochromatin is a prominent feature of all genomes. This type of heterochromatin is characteristically gene poor and repetitive DNA rich (Smith et al., 2007). Such chromatin is found predominantly around the centromere and pericentromeric regions of the

chromosome (Figure 1-5). Heterochromatin components are also found at the telomeres in Drosophila. Heterochromatin is generally important for chromosome stability and the proper segregation of chromosomes (Morris and Moazed, 2007). These functions have implications for the maintenance of genome stability.

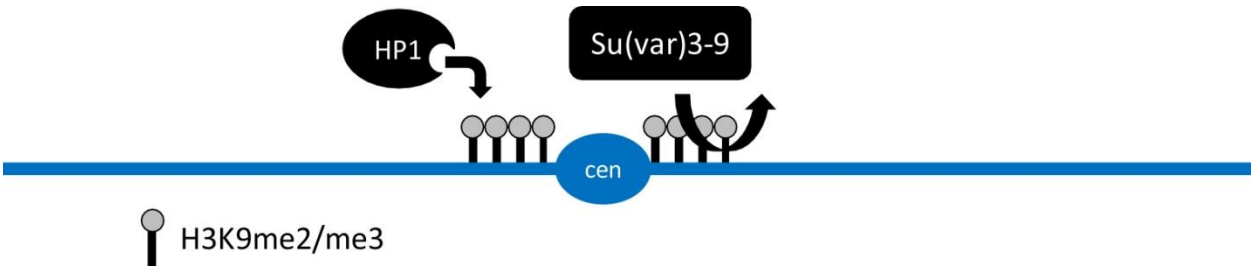
Constitutive heterochromatin is maintained by a very different set of machinery than facultative heterochromatin, though recent evidence suggests that there may be more interplay between these two groups than previously recognized (Cabrera et al., 2015). The foremost of the constitutive heterochromatin effectors is Heterochromatin protein 1 (HP1, also known as Su(var)2-5), which binds throughout constitutive heterochromatin (Grewal and Moazed, 2003). The HP1 protein binds to heterochromatin through its chromodomain. The HP1 chromodomain has affinity for di- and tri-methylated histone H3 at lysine 9 (H3K9me2/me3). These histone modifications are created by the enzyme Su(var)3-9, which is a SET domain protein with methyltransferase activity (Shankaranarayana et al., 2003).

The role of heterochromatin in gene silencing is evident in a process called position effect variegation (PEV). This occurs when chromosome rearrangements result in the juxtaposition of euchromatic genes near heterochromatin, often near the centromere (Elgin and Reuter, 2013). This typically results in variegated silencing of the locus. Several alleles that display this phenomenon exist, but perhaps most well-known is the *white*^{mottled4} (*w*^{m4}) allele. An X-chromosome inversion results in the *white*$^{+}$ gene, whose wild-type function results in red pigmentation of the adult eye, being placed adjacent to heterochromatin. Animals carrying this allele typically have reduced eye pigmentation. If, however, the heterochromatin system is compromised due to genetic perturbations, increased pigmentation is observed. Several additional genes involved in heterochromatin have been identified in screens for suppressors of variegation (that is, mutations resulting in more eye pigmentation).

Figure 1-5: Heterochromatin formation is mediated by Su(var)3-9 and HP1

Heterochromatin plays an important structural role for chromosomes. Found at the centromere (CEN) and pericentromeric regions, the DNA content of constitutive heterochromatin is often highly repetitive in nature. The histone methyltransferase Su(var)3-9 catalyzes di- and tri-methylation of histone H3 on lysine 9. This mark is recognized by the chromodomain of Heterochromatin Protein 1 (HP1).

Figure 1-5 (Continued)



Despite being competent for the repression of euchromatin genes when chromosome structure is altered, the classical view of constitutive heterochromatin paints a picture of a highly repressed state with little cross-talk with euchromatic factors. Experiments in our lab identified additional factors involved in heterochromatin, but generally conformed to this model (Alekseyenko et al., 2014b). It was rather unexpected, then, that my work on the MSL complex led me to examine how an MSL interactor, UpSET, may help to sustain a critical balance between euchromatin and heterochromatin (see Chapter 2).

Summary and Overview of the Dissertation

The different groups of proteins above play a variety of critical roles in the proper regulation of the genome. With the TrxG and MSL complex functioning to achieve proper levels of activation, the PcG and heterochromatin work to prevent aberrant expression. Striking the proper balance between these two states, and finely tuning the exact level of expression, may therefore require more crosstalk between all of these chromatin groups than has previously been recognized.

In this dissertation, I explore several proteins which suggest interrelatedness between these groups that had not previously been characterized. In seeking to understand the contribution of the protein UpSET to MSL-mediated dosage compensation, I found that this protein may have a more prominent role in the maintenance of heterochromatin. Using both genomic and genetic techniques, I provide evidence for UpSET functioning to limit the extent of activation of dosage compensated X-linked genes and to prevent improper expression levels of heterochromatin genes. In a fruitful collaboration with Hyuckjoon Kang, a post-doctoral researcher in the lab, we characterized the Scm protein as a physical link between PRC1 and PRC2. We also explored the extent to which the Fs(1)h protein functionally and physically interacts with the PcG PRC1 complex. Our data are provocative, supporting a

model for the determination of chromatin state based on direct, but mutually exclusive, interactions between members of the PcG and the TrxG.

References:

Akhtar, A., and Becker, P.B. 2000. Activation of Transcription through Histone H4 Acetylation by MOF, an Acetyltransferase Essential for Dosage Compensation in Drosophila. *Mol Cell* 5: 367-375.

Alekseyenko, A.A., Gorchakov, A.A., Kharchenko, P.V., and Kuroda, M.I. 2014a. Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc Natl Acad Sci U S A* 111: 2488-2493.

Alekseyenko, A.A., Gorchakov, A.A., Zee, B.M., Fuchs, S.M., Kharchenko, P.V., and Kuroda, M.I. 2014b. Heterochromatin-associated interactions of Drosophila HP1a with dADD1, HIPP1, and repetitive RNAs. *Genes Dev* 28: 1445-1460.

Alekseyenko, A.A., McElroy, K.A., Kang, H., Zee, B.M., Kharchenko, P.V., and Kuroda, M.I. 2015. BioTAP-XL: Cross-linking/Tandem Affinity Purification to Study DNA Targets, RNA, and Protein Components of Chromatin-Associated Complexes. *Curr Protoc Mol Biol* 109: 21 30 21-21 30 32.

Alfieri, C., Gambetta, M.C., Matos, R., Glatt, S., Sehr, P., Fraterman, S., Wilm, M., Muller, J., and Muller, C.W. 2013. Structural basis for targeting the chromatin repressor Sfmbt to Polycomb response elements. *Genes Dev* 27: 2367-2379.

Ansari, S.A., and Morse, R.H. 2013. Mechanisms of Mediator complex action in transcriptional activation. *Cell Mol Life Sci* 70: 2743-2756.

Brown, J.L., Fritsch, C., Muller, J., and Kassis, J.A. 2003. The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. *Development* 130: 285-294.

Cabrera, J.R., Olcese, U., and Horabin, J.I. 2015. A balancing act: heterochromatin protein 1a and the Polycomb group coordinate their levels to silence chromatin in Drosophila. *Epigenetics Chromatin* 8: 17.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298: 1039-1043.

Chen, Z.-X., Golovnina, K., Sultana, H., Kumar, S., and Oliver, B. 2014. Transcriptional effects of gene dose reduction. *Biology of Sex Differences* 5.

Conrad, T., Cavalli, F.M., Vaquerizas, J.M., Luscombe, N.M., and Akhtar, A. 2012. Drosophila dosage compensation involves enhanced Pol II recruitment to male X-linked promoters. *Science* 337: 742-746.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. 2002. *Drosophila* Enhancer of Zeste/ESC Complexes have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. *Cell* 111: 185-196.

Delcuve, G.P., He, S., and Davie, J.R. 2008. Mitotic partitioning of transcription factors. *J Cell Biochem* 105: 1-8.

Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. 2007. Dynamics of replication-independent histone turnover in budding yeast. *Science* 315: 1405-1408.

Egloff, S., Dienstbier, M., and Murphy, S. 2012. Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet* 28: 333-341.

Elgin, S.C., and Reuter, G. 2013. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harb Perspect Biol* 5: a017780.

Fauth, T., Muller-Planitz, F., Konig, C., Straub, T., and Becker, P.B. 2010. The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. *Nucleic Acids Res* 38: 3209-3221.

Ferrari, F., Alekseyenko, A.A., Park, P.J., and Kuroda, M.I. 2014. Transcriptional control of a whole chromosome: emerging models for dosage compensation. *Nat Struct Mol Biol* 21: 118-125.

Ferrari, F., Jung, Y.L., Kharchenko, P.V., Plachetka, A., Alekseyenko, A.A., Kuroda, M.I., and Park, P.J. 2013a. Comment on "Drosophila dosage compensation involves enhanced Pol II recruitment to male X-linked promoters". *Science* 340: 273.

Ferrari, F., Plachetka, A., Alekseyenko, A.A., Jung, Y.L., Ozsolak, F., Kharchenko, P.V., Park, P.J., and Kuroda, M.I. 2013b. "Jump start and gain" model for dosage compensation in Drosophila based on direct sequencing of nascent transcripts. *Cell Rep* 5: 629-636.

Follmer, N.E., Wani, A.H., and Francis, N.J. 2012. A polycomb group protein is retained at specific sites on chromatin in mitosis. *PLoS Genet* 8: e1003135.

Francis, N.J., Follmer, N.E., Simon, M.D., Aghia, G., and Butler, J.D. 2009. Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell* 137: 110-122.

Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. 2001. Reconstitution of a Functional Core Polycomb Repressive Complex. *Mol Cell* 8: 545-556.

Gelbart, M.E., and Kuroda, M.I. 2009. Drosophila dosage compensation: a complex voyage to the X chromosome. *Development* 136: 1399-1410.

Gelbart, M.E., Larschan, E., Peng, S., Park, P.J., and Kuroda, M.I. 2009. Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nat Struct Mol Biol* 16: 825-832.

Gilchrist, D.A., and Adelman, K. 2012. Coupling polymerase pausing and chromatin landscapes for precise regulation of transcription. *Biochim Biophys Acta* 1819: 700-706.

Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K. 2010. Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* 143: 540-551.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K., Collins, J.B., Li, L., Gilmour, D.S., and Adelman, K. 2008. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev* 22: 1921-1933.

Grewal, S.I., and Moazed, D. 2003. Heterochromatin and epigenetic control of gene expression. *Science* 301: 798-802.

Guo, J., and Price, D.H. 2013. RNA polymerase II transcription elongation control. *Chem Rev* 113: 8583-8603.

Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A., and Lucchesi, J.C. 1997. *mof*, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *EMBO J* 16: 2054-2060.

Ho, J.W., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.A., Minoda, A., Tolstorukov, M.Y., Appert, A., Parker, S.C., Gu, T., Kundaje, A., Riddle, N.C., Bishop, E., Egelhofer, T.A., Hu, S.S., Alekseyenko, A.A., Rechtsteiner, A., Asker, D., Belsky, J.A., Bowman, S.K., Chen, Q.B., Chen, R.A., Day, D.S., Dong, Y., Dose, A.C., Duan, X., Epstein, C.B., Ercan, S., Feingold, E.A., Ferrari, F., Garrigues, J.M., Gehlenborg, N., Good, P.J., Haseley, P., He, D., Herrmann, M., Hoffman, M.M., Jeffers, T.E., Kharchenko, P.V., Kolasinska-Zwierz, P., Kotwaliwale, C.V., Kumar, N., Langley, S.A., Larschan, E.N., Latorre, I., Libbrecht, M.W., Lin, X., Park, R., Pazin, M.J., Pham, H.N., Plachetka, A., Qin, B., Schwartz, Y.B., Shoresh, N., Stempor, P., Vielle, A., Wang, C., Whittle, C.M., Xue, H., Kingston, R.E., Kim, J.H., Bernstein, B.E., Dernburg, A.F., Pirrotta, V., Kuroda, M.I., Noble, W.S., Tullius, T.D., Kellis, M., MacAlpine, D.M., Strome, S., Elgin, S.C., Liu, X.S., Lieb, J.D., Ahringer, J., Karpen, G.H., and Park, P.J. 2014. Comparative analysis of metazoan chromatin organization. *Nature* 512: 449-452.

Izzo, A., Regnard, C., Morales, V., Kremmer, E., and Becker, P.B. 2008. Structure-function analysis of the RNA helicase maleless. *Nucleic Acids Res* 36: 950-962.

Jenuwein, T., and Allis, C.D. 2001. Translating the histone code. *Science* 293: 1074-1080.

Jung, Y.L., Kang, H., Park, P.J., and Kuroda, M.I. 2016. Correspondence of Drosophila Polycomb Group proteins with broad H3K27me3 silent domains. *Fly (Austin)*: 0.

Jurgens, G. 1985. A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature* 316.

Kalashnikova, A.A., Porter-Goff, M.E., Muthurajan, U.M., Luger, K., and Hansen, J.C. 2013. The role of the nucleosome acidic patch in modulating higher order chromatin structure. *J R Soc Interface* 10: 20121022.

Kassis, J.A. 2002. Pairing-sensitive silencing, Polycomb Group Response Elements, and Transposon Homing in Drosophila. *Advances in Genetics* 46.

Kassis, J.A., and Brown, J.L. 2013. Polycomb group response elements in Drosophila and vertebrates. *Adv Genet* 81: 83-118.

Kelley, R.L., Meller, V.H., Gordadze, P.R., Roman, G., Davis, R.L., and Kuroda, M.I. 1999. Epigenetic Spreading of the *Drosophila* Dosage Compensation Complex from *roX* RNA Genes into Flanking Chromatin. *Cell* 98: 513-522.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., Linder-Basso, D., Plachetka, A., Shanower, G., Tolstorukov, M.Y., Luquette, L.J., Xi, R., Jung, Y.L., Park, R.W., Bishop, E.P., Canfield, T.K., Sandstrom, R., Thurman, R.E., MacAlpine, D.M., Stamatoyannopoulos, J.A., Kellis, M., Elgin, S.C., Kuroda, M.I., Pirrotta, V., Karpen, G.H., and Park, P.J. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* 471: 480-485.

Kim, D., Blus, B.J., Chandra, V., Huang, P., Rastinejad, F., and Khorasanizadeh, S. 2010. Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nat Struct Mol Biol* 17: 1027-1029.

Kingston, R.E., and Tamkun, J.W. 2014. Transcriptional regulation by trithorax-group proteins. *Cold Spring Harb Perspect Biol* 6: a019349.

Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Muller, J. 2006. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev* 20: 1110-1122.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* 16: 2893-2905.

Kwak, H., and Lis, J.T. 2013. Control of transcriptional elongation. *Annu Rev Genet* 47: 483-508.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A., and Verrijzer, C.P. 2008. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev* 22: 2799-2810.

Laitem, C., Zaborowska, J., Isa, N.F., Kufs, J., Dienstbier, M., and Murphy, S. 2015. CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. *Nat Struct Mol Biol* 22: 396-403.

Lariviere, L., Seizl, M., and Cramer, P. 2012. A structural perspective on Mediator function. *Curr Opin Cell Biol* 24: 305-313.

Larschan, E., Alekseyenko, A.A., Gortchakov, A.A., Peng, S., Li, B., Yang, P., Workman, J.L., Park, P.J., and Kuroda, M.I. 2007. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* 28: 121-133.

Larschan, E., Bishop, E.P., Kharchenko, P.V., Core, L.J., Lis, J.T., Park, P.J., and Kuroda, M.I. 2011. X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. *Nature* 471: 115-118.

Larschan, E., Soruco, M.M., Lee, O.K., Peng, S., Bishop, E., Chery, J., Goebel, K., Feng, J., Park, P.J., and Kuroda, M.I. 2012. Identification of chromatin-associated regulators of MSL complex targeting in Drosophila dosage compensation. *PLoS Genet* 8: e1002830.

Lee, C., Chang, K.A., Kuroda, M.I., and Hurwitz, J. 1997. The NTPase/helicase activities of *Drosophila* maleless, an essential factor in dosage compensation. *EMBO J* 16: 2671-2681.

Lengsfeld, B.M., Berry, K.N., Ghosh, S., Takahashi, M., and Francis, N.J. 2012. A Polycomb complex remains bound through DNA replication in the absence of other eukaryotic proteins. *Sci Rep* 2: 661.

Lewis, E.B. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* 276.

Lewis, R.A., Wakimoto, B.T., Denell, R.E., and Kaufman, T.C. 1980. Genetic Analysis of the Antennapedia Gene Complex (Ant-C) and Adjacent Chromosomal Regions of DROSOPHILA MELANOGASTER. II. Polytene Chromosome Segments 84A-84B1,2. *Genetics* 95: 383-397.

Lo, S.M., Ahuja, N.K., and Francis, N.J. 2009. Polycomb group protein Suppressor 2 of zeste is a functional homolog of Posterior Sex Combs. *Mol Cell Biol* 29: 515-525.

Lucchesi, J.C., and Kuroda, M.I. 2015. Dosage compensation in Drosophila. *Cold Spring Harb Perspect Biol* 7.

Maenner, S., Muller, M., Frohlich, J., Langer, D., and Becker, P.B. 2013. ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol Cell* 51: 174-184.

Moore, S.A., Ferhatoglu, Y., Jia, Y., Al-Jiab, R.A., and Scott, M.J. 2010. Structural and biochemical studies on the chromo-barrel domain of male specific lethal 3 (MSL3) reveal a binding preference for mono- or dimethyllysine 20 on histone H4. *J Biol Chem* 285: 40879-40890.

Morra, R., Smith, E.R., Yokoyama, R., and Lucchesi, J.C. 2008. The MLE subunit of the Drosophila MSL complex uses its ATPase activity for dosage compensation and its helicase activity for targeting. *Mol Cell Biol* 28: 958-966.

Morris, C.A., and Moazed, D. 2007. Centromere assembly and propagation. *Cell* 128: 647-650.

Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. 2002. Histone Methyltransferase Activity of a *Drosophila* Polycomb Group Repressor Complex. *Cell* 111: 197-208.

Negre, N., Hennetin, J., Sun, L.V., Lavrov, S., Bellis, M., White, K.P., and Cavalli, G. 2006. Chromosomal Distribution of PcG Proteins during *Drosophila* Development. *PLoS Biol* 4.

Pattatucci, A.M., and Kaufman, T.C. 1991. The Homeotic Gene *Sex combs reduced* of *Drosophila melanogaster* is differentially regulated in the Embryonic and Imaginal Stages of Development. *Genetics* 129: 443-461.

Phatnani, H.P., and Greenleaf, A.L. 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* 20: 2922-2936.

Plaschka, C., Nozawa, K., and Cramer, P. 2016. Mediator Architecture and RNA Polymerase II Interaction. *J Mol Biol*.

Price, D.H. 2008. Poised polymerases: on your mark...get set...go! *Mol Cell* 30: 7-10.

Ranjan, A., Wang, F., Mizuguchi, G., Wei, D., Huang, Y., and Wu, C. 2015. H2A histone-fold and DNA elements in nucleosome activate SWR1-mediated H2A.Z replacement in budding yeast. *eLife* 4: e06845.

Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Muller, J. 2010. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* 465: 243-247.

Schuettengruber, B., Martinez, A.M., Iovino, N., and Cavalli, G. 2011. Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* 12: 799-814.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.Y., Bourgon, R., Biggin, M., and Pirrotta, V. 2006. Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nat Genet* 38: 700-705.

Shankaranarayana, G.D., Motamedi, M.R., Moazed, D., and Grewal, S.I. 2003. Sir2 regulates histone H3 lysine 9 methylation and heterochromatin assembly in fission yeast. *Curr Biol* 13: 1240-1246.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J., Wu, C., Bender, W., and Kingston, R. 1999. Stabilization of Chromatin Structure by PRC1, a Polycomb Complex. *Cell* 98: 37-46.

Shilatifard, A. 2012. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu Rev Biochem* 81: 65-95.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311: 844-847.

Simon, J.A., and Kingston, R.E. 2013. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* 49: 808-824.

Smith, C.D., Shu, S., Mungall, C.J., and Karpen, G.H. 2007. The Release 5.1 annotation of Drosophila melanogaster heterochromatin. *Science* 316: 1586-1591.

Soruco, M.M., Chery, J., Bishop, E.P., Siggers, T., Tolstorukov, M.Y., Leydon, A.R., Sugden, A.U., Goebel, K., Feng, J., Xia, P., Vedenko, A., Bulyk, M.L., Park, P.J., and Larschan, E. 2013. The CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage compensation. *Genes Dev* 27: 1551-1556.

Steffen, P.A., and Ringrose, L. 2014. What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat Rev Mol Cell Biol* 15: 340-356.

Straub, T., and Becker, P.B. 2013. Comment on "Drosophila dosage compensation involves enhanced Pol II recruitment to male X-linked promoters". *Science* 340: 273.

Struhl, G. 1981. A gene product required for correct initiation of segmental determination in *Drosophila*. *Nature* 293.

Struhl, G. 1982. Genes controlling segmental specification in the Drosophila thorax. *Proc Natl Acad Sci U S A* 79: 7380-7384.

Sural, T.H., Peng, S., Li, B., Workman, J.L., Park, P.J., and Kuroda, M.I. 2008. The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila melanogaster X chromosome. *Nat Struct Mol Biol* 15: 1318-1325.

Tolhuis, B., de Wit, E., Muijrers, I., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. 2006. Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. *Nat Genet* 38: 694-699.

Vannini, A., and Cramer, P. 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* 45: 439-446.

Vaquerizas, J.M., Cavalli, F.M., Conrad, T., Akhtar, A., and Luscombe, N.M. 2013. Response to Comments on "Drosophila Dosage Compensation Involves Enhanced Pol II Recruitment to Male X-Linked Promoters". *Science* 340: 273.

Villa, R., Forne, I., Muller, M., Imhof, A., Straub, T., and Becker, P.B. 2012. MSL2 combines sensor and effector functions in homeostatic control of the Drosophila dosage compensation machinery. *Mol Cell* 48: 647-654.

Wang, C.I., Alekseyenko, A.A., LeRoy, G., Elia, A.E., Gorchakov, A.A., Britton, L.M., Elledge, S.J., Kharchenko, P.V., Garcia, B.A., and Kuroda, M.I. 2013. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. *Nat Struct Mol Biol* 20: 202-209.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. 2004. Hierarchical recruitment of polycomb group silencing complexes. *Mol Cell* 14: 637-646.

Wang, L., Jahren, N., Miller, E.L., Ketel, C.S., Mallin, D.R., and Simon, J.A. 2010. Comparative analysis of chromatin binding by Sex Comb on Midleg (SCM) and other polycomb group repressors at a Drosophila Hox gene. *Mol Cell Biol* 30: 2584-2593.

Weber, C.M., Ramachandran, S., and Henikoff, S. 2014. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol Cell* 53: 819-830.

Wu, L., Zee, B.M., Wang, Y., Garcia, B.A., and Dou, Y. 2011. The RING finger protein MSL2 in the MOF complex is an E3 ubiquitin ligase for H2B K34 and is involved in crosstalk with H3 K4 and K79 methylation. *Mol Cell* 43: 132-144.

Yamaguchi, Y., Shibata, H., and Handa, H. 2013. Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochim Biophys Acta* 1829: 98-104.

Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. 2011. A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* 332: 977-980.

Zhao, Y., and Garcia, B.A. 2015. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol* 7: a025064.

Zheng, S., Villa, R., Wang, J., Feng, Y., Wang, J., Becker, P.B., and Ye, K. 2014. Structural basis of X chromosome DNA recognition by the MSL2 CXC domain during Drosophila dosage compensation. *Genes Dev* 28: 2652-2662.

Chapter 2
Differential effects of the loss of UpSET in heterochromatin and the dosage-compensated X chromosome in Drosophila

Contributions to this chapter: Charlotte Wang cloned the UpSET-BioTAP transgene and generated transgenic flies. LC-MS/MS was performed at the Taplin Mass Spectrometry Facility at Harvard Medical School. Barry Zee provided LC-MS/MS analysis. Lucy Jung performed the bioinformatics analysis of all ChIP and nascent-RNA experiments. Kyle McElroy performed all other experiments.

Abstract

Chromatin plays a critical role in faithful implementation of gene expression programs. Different post-translational modifications of histone proteins reflect the underlying state of gene activity, and many chromatin proteins write, erase, bind, or are repelled by these histone marks. One such protein is UpSET, the Drosophila homolog of yeast Set3 and mammalian MLL5. Here we show that UpSET is necessary for the proper balance between active and repressed states. Using CRISPR/Cas-9 editing, we generated flies and S2 cells which are mutant for *upSET*. We find that loss of UpSET is lethal in both sexes in flies, but tolerated in S2 cells. Misregulated heterochromatin is apparent by suppressed position effect variegation of the w^{m4} allele in flies, and by ChIP-seq for the H3K9me2 heterochromatin mark and quantification of bulk histone post-translational modifications in S2 cells. Consistent with participation in a histone deacetylase complex, we observe increased spreading of H4K16-acetylation specifically on the dosage compensated male X chromosome. Finally, we demonstrate that gene expression changes at the nascent-transcript level are consistent with a disruption in the balance of active and silent histone modifications. Our findings support a role for UpSET in maintaining heterochromatin, and delimiting activation of X-linked genes by the Male-Specific Lethal dosage compensation complex.

Introduction:

One family of chromatin-associated proteins which post-translationally modify histones is the SET domain-containing proteins. The SET domain catalyzes methylation of histone tails. Different SET domain-containing proteins create unique histone modification signatures, which are associated with different forms of active and repressed chromatin environment. Of the family of SET domain proteins, one paralog is a notable exception to this paradigm. The Drosophila protein UpSET and its homologs Set3 in yeast and MLL5 in mammals are not known to have histone modifying activity. In fact, conserved catalytically important residues in the SET domain are mutated in UpSET/Set3/MLL5 suggesting that protein function must not rely on the competence of the SET domain. Rather than catalyzing histone methylation, these proteins have been characterized in yeast and flies to form a complex with histone deacetylase (HDAC) activity (provided by a different catalytic protein) (Pijnappel et al., 2001, Rincon-Arano et al., 2012).

Phenotypically, Set3 is a non-essential gene in yeast (Pijnappel et al., 2001), with a phenotype of defective transcription kinetics observed only when cells are metabolically challenged (Kim et al., 2012, Wang et al., 2002). Set3 complex (Set3C) was also recently tied to the DNA damage response operating under a model of altered histone acetylation dynamics (Torres-Machorro et al., 2015). MLL5 in mammals has been tied to several different cellular processes including haematopoesis (Heuser et al., 2009, Madan et al., 2009, Zhang et al., 2009), cell cycle progression (Cheng et al., 2008, Deng et al., 2004), cancer (Emerling et al., 2002), and DNA methylation (Yun et al., 2014), though its exact mechanistic role or binding partners in these diverse functions have not been fully resolved. In Drosophila, *upSET*$^{-/-}$ flies were described to be viable, but with a female fertility defect due to derepression of transposable elements in the ovary (Rincon-Arano et al., 2012).

We previously identified the UpSET protein (CG9007) as a top interactor with the MSL3 protein by a crosslinked affinity purification technique (Wang et al., 2013). MSL3 is a chromodomain protein that is a core constituent of the Male-Specific Lethal (MSL) dosage compensation complex in Drosophila. The MSL complex of five genetically-defined proteins and two redundant noncoding RNAs localize specifically to the male X to create a unique chromatin environment and boost expression of active genes (for review see (Lucchesi and Kuroda, 2015)). The chromatin environment that is created by the MSL complex is catalyzed by the Males-absent-on-first (MOF) protein, which acetylates histone 4 on lysine 16 (H4K16ac). This modification has a very stereotypical pattern near transcriptional start sites on autosomes and across the female genome. In contrast, on the male X, H4K16ac is enriched on gene bodies, reflecting the localization of the MSL complex and its putative function in transcriptional elongation.

The interaction between UpSET, a member of an HDAC complex, and MSL3, a member of the MSL complex that makes an acetyl mark, is striking. One question immediately raised is whether UpSET might contribute to dosage compensation. To investigate this, we created an *upSET* mutant Drosophila fly line using the CRISPR/Cas-9 system for genome engineering and a homologous recombination donor. In contrast to previous reports, we observed *upSET*$^{-/-}$ to be lethal in both males and females. This lethality was rescued by a tagged *upSET* transgene. Thus, we were able to analyze the epitope-tagged UpSET protein in the *upSET*-mutant background to identify its genome-wide binding profile. To further assess the molecular defects induced by loss of UpSET, in parallel we created *upSET* mutant S2 cell lines. We surveyed bulk histone post-translational modification levels in these cells, and mapped two specific marks by ChIP-seq. We observed an increase in H4K16ac regions specifically on the male X chromosome and a decrease in the heterochromatin-associated histone 3 lysine 9 dimethyl (H3K9me2) mark on all chromosomes, along with concomitant aberrant transcription of heterochromatic and X-linked genes as measured by nascent-RNA-sequencing. We conclude that heterochromatin and genes on the male X

chromosome are particularly sensitive to a balance of acetylation and deacetylation activities, regulated in part by the UpSET protein.

Results:

***upSET* is an essential gene in Drosophila**

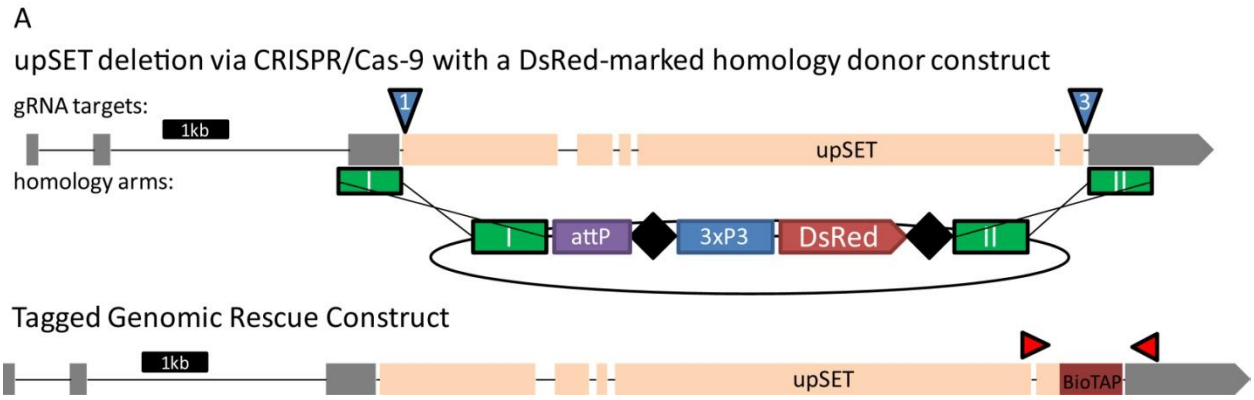
The previous characterization of *upSET* mutant flies utilized the only two then-available lines which carry a P-element and Minos insertion in the *upSET* gene, respectively. However, the insertions carried by both of these lines leave the coding sequence largely intact. While Western blotting suggested no residual protein, given the possibility that those alleles could in fact be hypomorphic instead of complete loss-of-function, we sought to create an *upSET* deletion allele with the coding sequence removed from the genome. To accomplish this, we turned to the versatile CRISPR/Cas-9 genome engineering system. We co-injected *w; w-{nos-cas9}/CyO* embryos, which express Cas-9 in the germ line, with two guide RNA constructs and a homologous recombination donor marked with 3xP3-DsRed (Figure 2-1A). We isolated balanced flies that carried the DsRed marker and confirmed the loss of *upSET* coding sequences by PCR. We found that homozygosity for the *upSET* deletion is lethal, with a low escape rate. The few escapers that were recovered were sickly and did not reproduce with *yw* mates. We were able to rescue this lethality with an UpSET-BioTAP transgene (Figure 2-1B). PCR from rescued individuals confirmed that no wild-type *upSET* DNA remained (Figure 2-1C). Taken together, this suggests that contrary to previous reports, *upSET* is an essential gene in Drosophila, and that the previous alleles are hypomorphs, rather than bona fide loss-of-function.

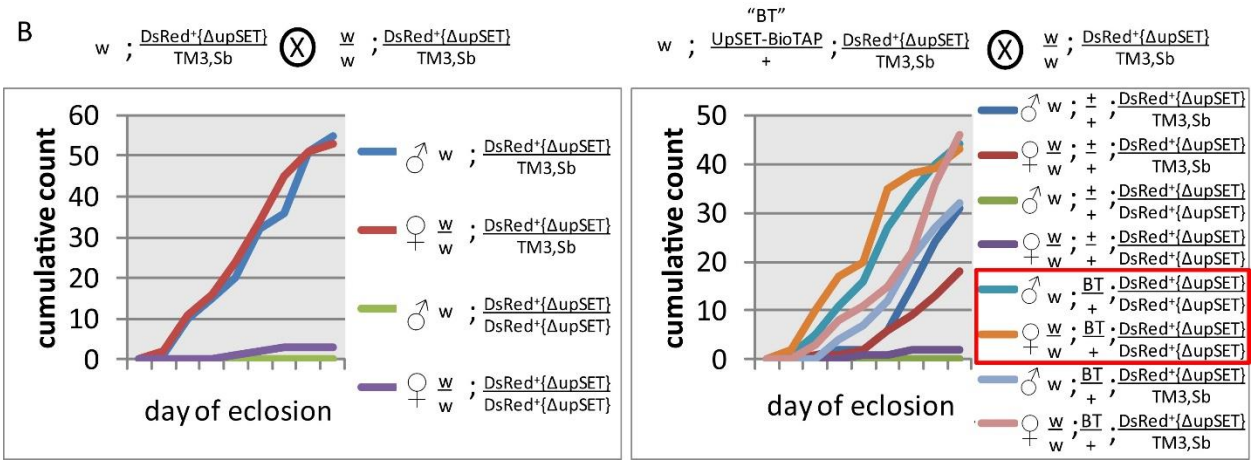
UpSET-BioTAP localizes to TSS of active genes by ChIP-seq

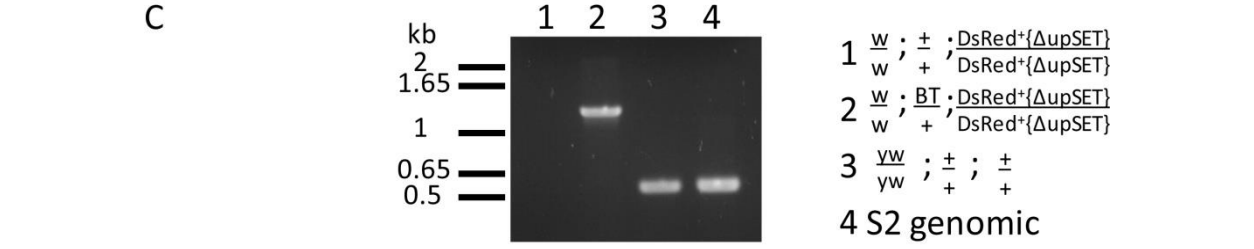
Figure 2-1: *upSET* is an essential gene in Drosophila

A) *upSET* gene span with the location of guide RNA target constructs (blue triangles), homologous recombination donor arms (green rectangles), the BioTAP-tag insertion in the rescue construct (crimson), and PCR primers (red, not to scale) to assess deletion and rescue.
B) *upSET* is homozygous lethal, but is rescued by the UpSET-BioTAP transgene
C) PCR analysis reveals that homozygous escapers lack the 544bp WT product. The rescued flies produce only the band corresponding to the tagged construct.

Figure 2-1 (Continued)







Our previous attempts to determine the localization of UpSET using the BioTAP-tagged transgene were unsuccessful in cell culture due to poor stability of the tagged protein in chromatin preparations (data not shown). We reasoned that more of the tagged protein would be incorporated into chromatin in the *upSET*-deleted background, and so we prepared chromatin for ChIP from 0.1g of mixed 12-24hr embryos carrying the UpSET-BioTAP transgene in the homozygous *DsRed*$^{+}$*{ΔupSET}* background. We immunoprecipated UpSET-BioTAP using the protein-A moiety, and sequenced the resulting material.

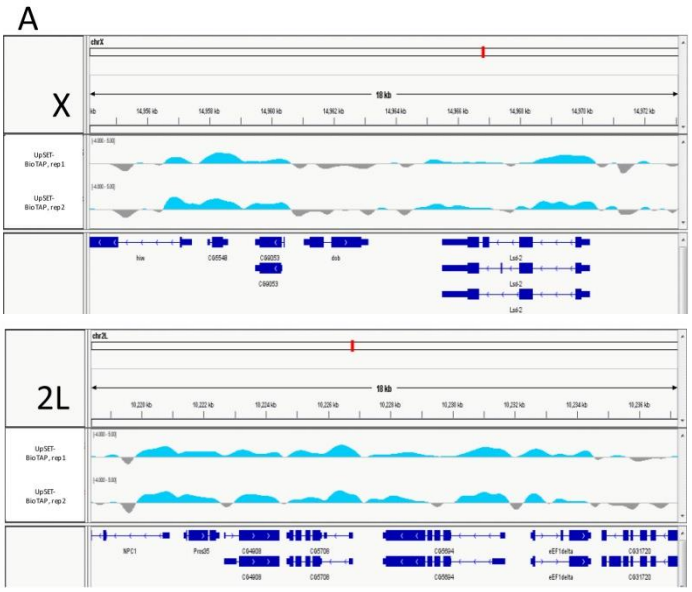
In agreement with previous localization data generated by UpSET Dam-ID in Kc cells (Rincon-Arano et al., 2012), we observed UpSET-BioTAP to localize to active genes by ChIP-seq (Figure 2-2A-B). More specifically, in the BioTAP data there is enrichment for UpSET-BioTAP ChIP peaks in regions carrying the chromatin signatures of transcription-start site proximal and active elongation states (Ho et al., 2014) (Figure 2-3). This is further supported by comparing the overlap of UpSET-BioTAP ChIP peaks with different genome feature annotations. UpSET-BioTAP peak regions display a greater than 2-fold enrichment over the genomic background for promoter and 5' UTR regions and are also enriched for coding exons (Figure 2-2C). Conversely, intron and intergenic regions are depleted from UpSET-BioTAP peaks when compared to the whole genome. When comparing the UpSET-BioTAP dataset to those publically available in the modEncode project, we observe the highest correlation with Pol II-datasets (Figure 2-4), consistent with enrichment at the TSS of genes in the active chromatin context.

We sought to examine whether we could detect an X-specific localization pattern for UpSET-BioTAP as compared to the autosomal pattern, although in mixed-sex embryos, any apparent X-specific localization signal in males would be dampened. We did not detect any difference between UpSET-BioTAP localization at the TSS or throughout the gene body between X-linked and autosomal genes (Figure 2-5A). However, when comparing the intensity of UpSET-BioTAP peaks across the X

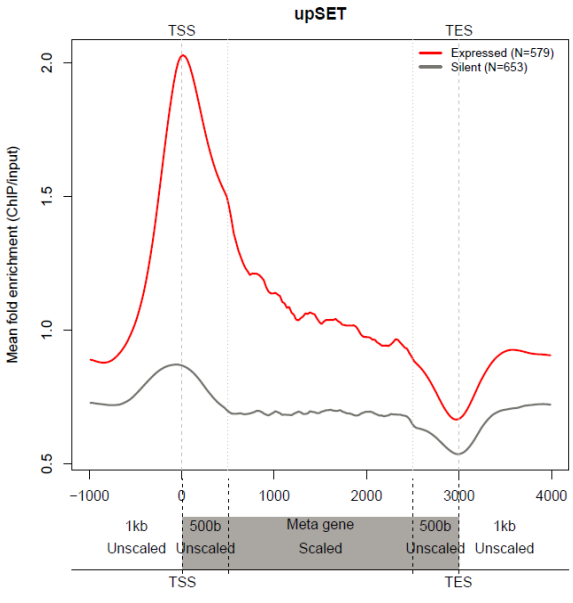
Figure 2-2: UpSET-BioTAP localizes to active TSS in embryos

A) Two representative genome browser views of 18kb regions of the X and 2L chromosomes. There are few differences between UpSET-BioTAP on the X chromosome and binding on the autosomes.
B) Metagene profile for UpSET-BioTAP binding. Gene bodies from TSS+500bp to TES-500bp are scaled. The 1kb up and downstream, and 500bp into the gene on either end are unscaled. UpSET-BioTAP strongly enriches around the TSS. UpSET-BioTAP shows a strong preference for binding at the TSS of active genes. Please note the y-axis is a linear scale.
C) UpSET-BioTAP peaks are enriched for genome regions annotated to be promoters, 5'UTR, and coding exons. Intron and intergenic regions are depleted compared to the whole genome background.

Figure 2-2 (Continued)



B



C

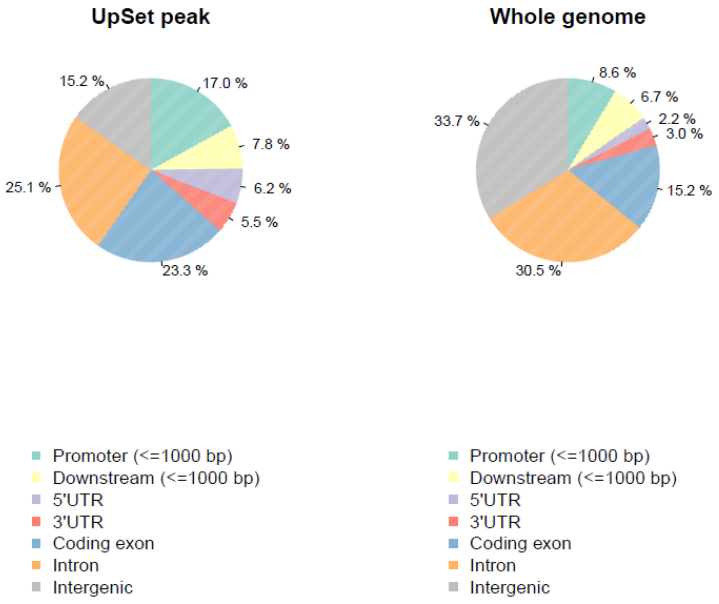
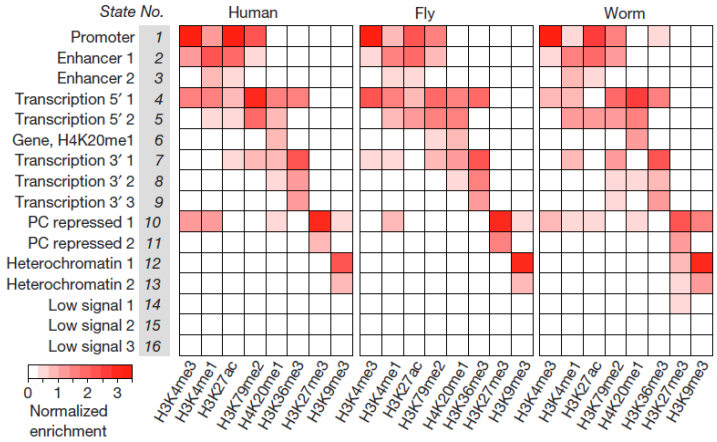


Figure 2-3: UpSET-BioTAP ChIP peaks are enriched for active TSS and transcriptional elongation chromatin signatures

Ho et al (2014) used eight histone post-translational modifications across three species and machine learning to identify 16 different classes of chromatin signatures (top panel). UpSET-BioTAP peaks largely fall into the active signatures, with the highest enrichment of promoter regions (state 1). UpSET-BioTAP peaks also enrich for enhancer and transcription signatures (states 3, 4, 5, 7, and 9).

Figure 2-3 (Continued)



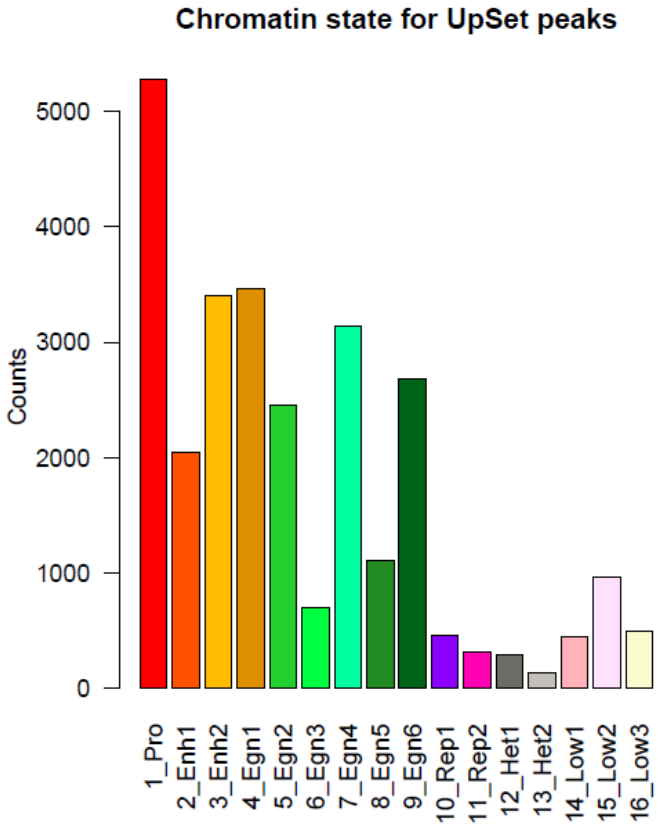


Figure 2-4: UpSET-BioTAP ChIP-seq data correlates most strongly with modEncode ChIP-chip datasets for PolII

Pearson correlation coefficients were calculated for the pairwise comparison of the UpSET-BioTAP ChIP-seq with a number of ChIP-chip profiles for chromosomal proteins generated by the modEncode project using 1kb windows. The highest degrees of correlation were found with active associated proteins and histone PTMs, such as H3K4me3, H3K4me2, and Pol II. It should be noted that the difference in platform (UpSET-BioTAP Chip-seq vs ChIP-chip for modEncode factors) may artificially depress the Pearson correlation coefficient values.

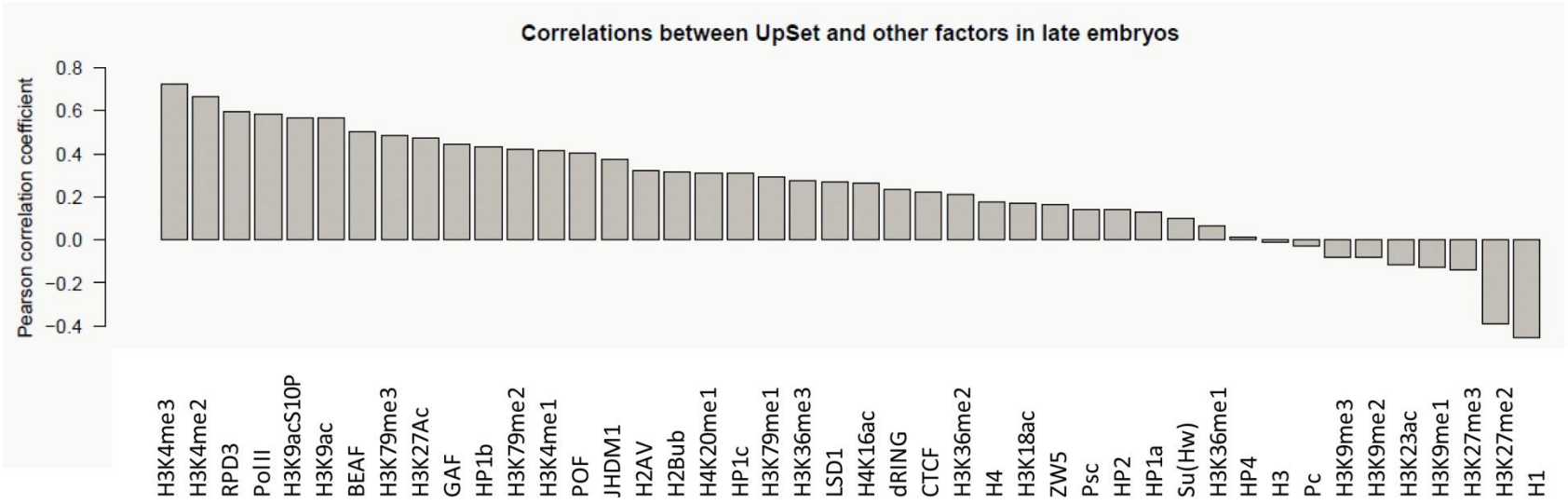


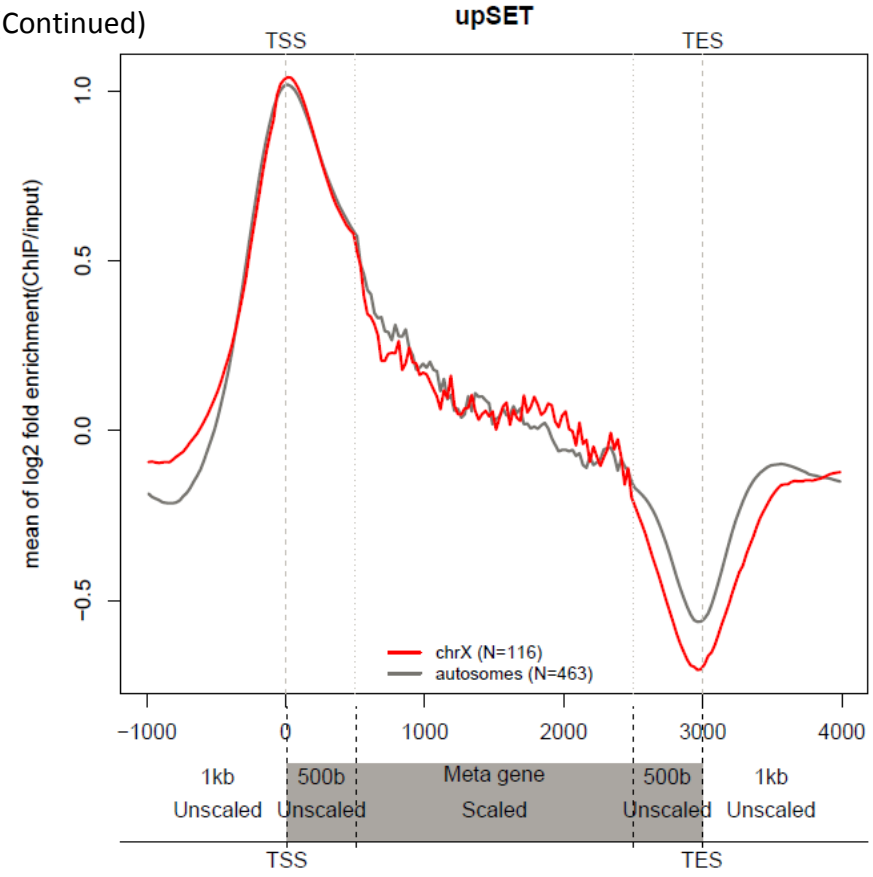
Figure 2-4 (Continued)

Figure 2-5: UpSET-BioTAP shows differential enrichment by chromosome arm

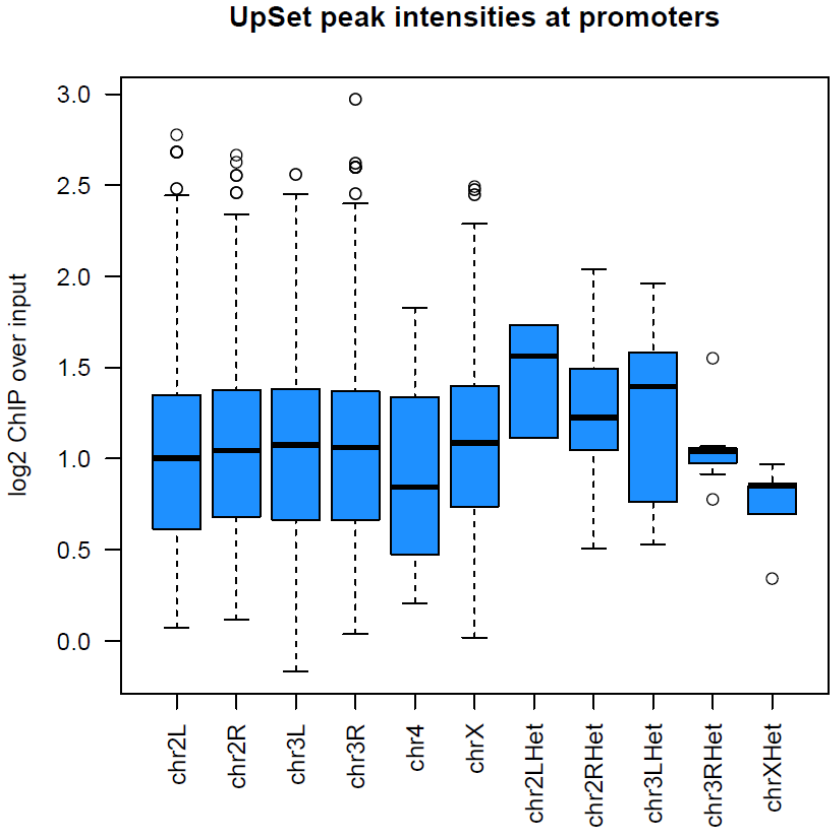
A) Metagene profile comparing the pattern of UpSET-BioTAP across the gene bodies (scaling as in Figure 2-2B) of genes on the X versus the Autosomes. No major differential pattern is detected from the UpSET-BioTAP ChIP-seq from mixed embryos.
B) UpSET-BioTAP ChIP-seq peak intensities at promoters were calculated for genes and grouped by chromosome arm. There is no major enrichment across the euchromatic arms, but higher intensity binding is observed in some, but not all, heterochromatin regions.

Figure 2-5 (Continued)

A



B



chromosome, autosome arms, and heterochromatin, we saw higher values for several of the heterochromatin regions (Figure 2-5B).

CRISPR-engineered S2 cell lines tolerate inactivating UpSET mutations

In order to assess the molecular effects of the loss of UpSET, specifically in male cells, we attempted to generate S2 cell lines stably carrying *upSET* mutations. S2 cells are a male Drosophila cell line which is highly polyploid. We reasoned that these cells may tolerate perturbed chromatin states better than the whole organism, given their tolerance of their non-diploid genome. The ploidy of the genome introduces its own challenges for genome engineering, yet we saw that mutations typically went to fixation when we used the CRISPR/Cas-9 system (Housden et al., 2015). We co-transfected S2 cells with an RFP-expressing marker plasmid and a plasmid co-expressing both the Cas-9 protein and one or more of several different guide RNAs directed toward the *upSET* gene (Figure 2-6A). To isolate single clones, RFP-positive transfected cells were sorted into 96-well plates by FACS. Following regrowth from these single cells, we identified lines that have putative mutations by high-resolution melt assays (HRMA) (Bassett et al., 2013). Statistically significant hits were further analyzed by Sanger sequencing to identify the nature of the molecular lesion at the gRNA target site. In this way, we were able to isolate 3 *upSET* mutant lines which should be devoid of wild-type UpSET protein (Figure 2-6B,C).

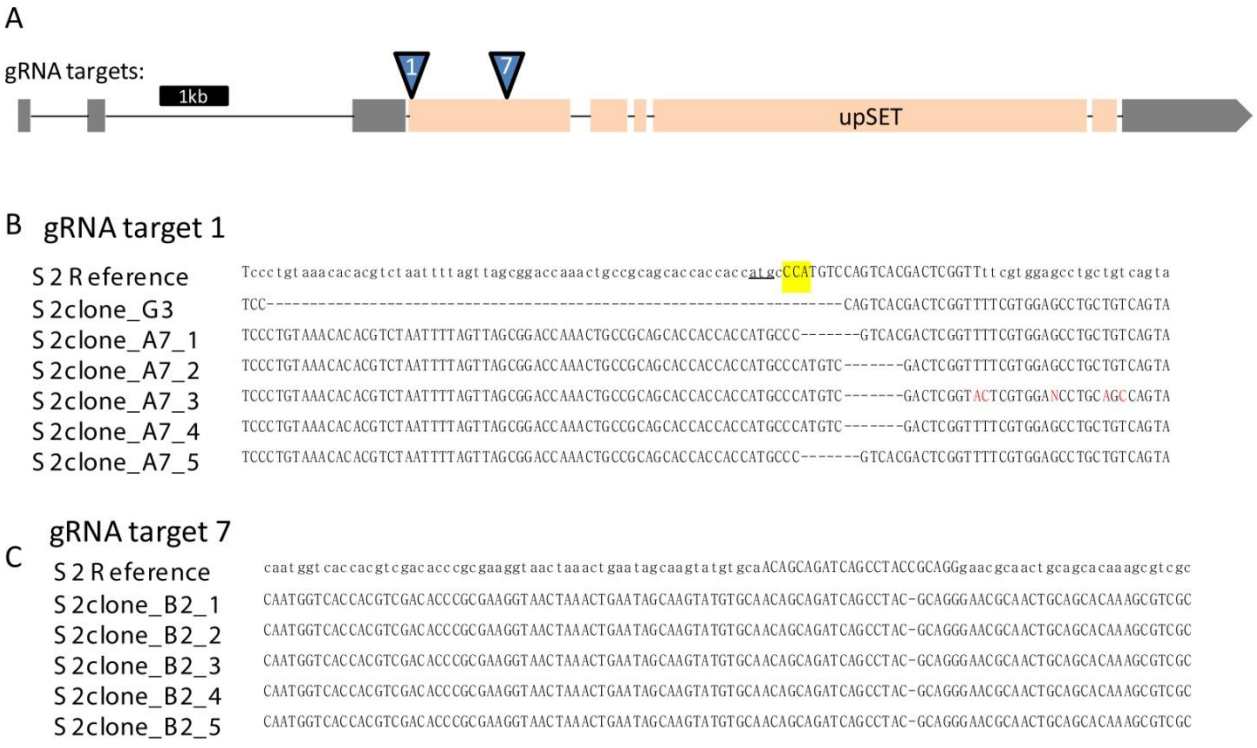
Mutation of *upSET* results in aberrant H4K16ac pattern on the male X chromosome

To determine whether we could detect a molecular phenotype in the *upSET* mutant lines, we performed ChIP-seq using antibodies against H4K16ac, the MOF-catalyzed mark. H4K16ac is known to have two distinct patterns over the bodies of genes in Drosophila (Gelbart et al., 2009). The first, a 5' gene and TSS enrichment, is found on autosomes and the female X chromosome. The second is found only on the dosage compensated (male) X chromosome, and is a much higher enrichment with a 3' bias,

Figure 2-6: Generating *upSET* mutant S2 lines

A) Schematic for the *upSET* gene locus. Non-coding exons are in grey, while coding exons are in light orange. Guide RNA targets for Cas9 are listed.
B) Molecular lesions generated around the *upSET* gRNA #1 target site (capitalized in reference sequence, PAM highlighted), located just downstream of the *upSET* start codon (underlined in reference sequence). Clone G3 has a homozygous 67bp deletion, removing the start codon and 10 additional coding base pairs, as well as 54 bp of the adjacent sequence. Clone A7 has two separate 7bp deletions, both resulting in frameshift mutations and a truncated peptide product.
C) Molecular lesion generated around the *upSET* gRNA #7 target site (capitalized in reference sequence). Clone B2 carries a 1bp deletion, resulting in a frameshift after 367 amino acids of wild-type UpSET protein sequence and a 436 amino acid product.

Figure 2-6 (Continued)



which reflects the localization of the assembled MSL complex toward the 3' end of genes. If UpSET plays a general role in acetylation, one would expect both the X and autosomes to be affected by *upSET* mutations. If, however, UpSET plays a special role in dosage compensation, perhaps a difference in effect on H4K16ac on X compared to the autosomes will be apparent.

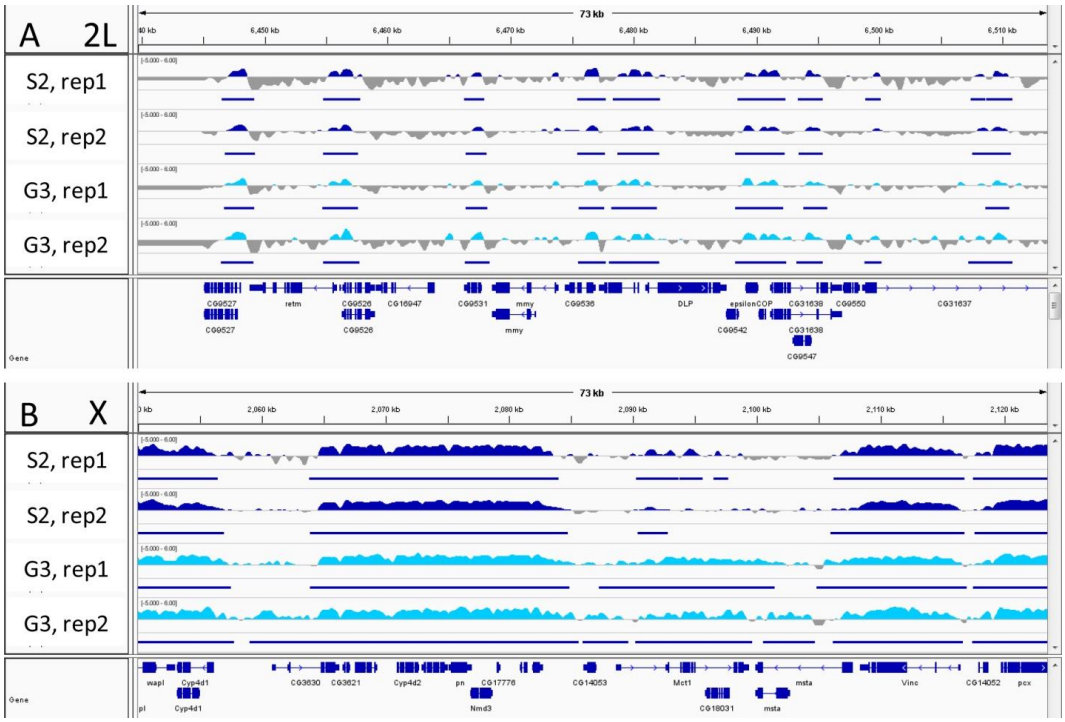
We chose to proceed with ChIP using the (G3) mutant since we expected this largest deletion that removes the start codon and adjacent sequence to be the most deleterious. We observed that the H4K16ac pattern on autosomes stayed remarkably consistent (Figure 2-7A). In comparison, the pattern of H4K16ac on the X chromosome was altered, with additional signal appearing to spread out from preexisting peaks (Figure 2-7B). For genome-wide analysis, we used the SPP package to call peaks after sub-sampling the H4K16ac datasets. We then calculated the genomic coverage of these peaks and compared the result between chromosome arms. In accordance with our visual observation, H4K16ac peaks on chromosome arms covered similar amounts of the genome (in basepairs) in the wild-type and *upSET*-mutant S2 cells, whereas H4K16ac coverage on the X chromosome was increased in *upSET*-mutant cells by ~1Mb (Figure 2-7C). We calculated the fold change (G3 *upSET* mutant vs S2) per H4K16ac peak and observed that a large portion of the peaks on the X chromosome show increased coverage, whereas autosomes remain largely unchanged (Figure 2-7D). That many peaks are changing suggests that the mechanism for increased H4K16ac on the X chromosome in the *upSET* mutant cell line is the consequence of some general, as opposed to site-specific, X-specific process.

Previous work had shown that *upSET*-directed RNAi resulted in spreading of H4K16ac at multiple candidate regions in the female-type Kc cell line by ChIP-qPCR (Rincon-Arano et al., 2012). Our result supports a role for UpSET in regulating dosage compensation specifically, and thus does not seem completely compatible with the previous finding. Given the observed UpSET-MSL3 interaction, it could be that the role UpSET plays in male and female tissues may be different.

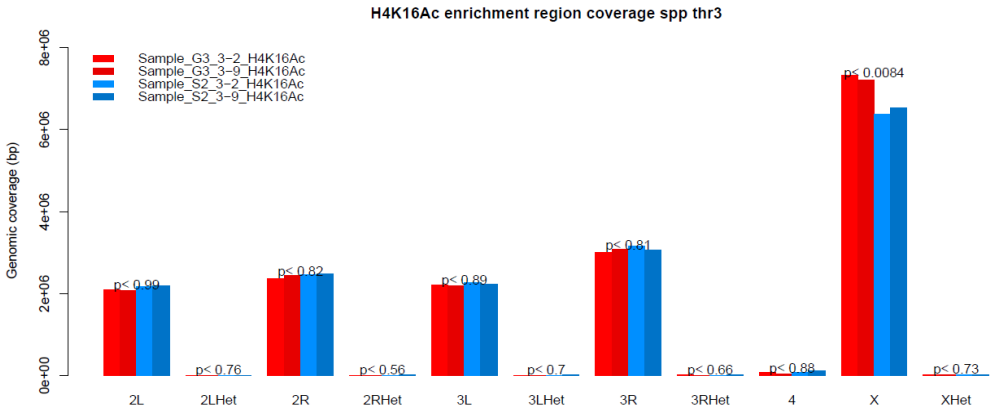
Figure 2-7: *upSET* mutant cells have broader H4K16ac peaks on the X chromosome

A) A representative genome browser view of a 73kb region of the 2L chromosome arm. H4K16ac ChIP-seq tracks show the expected pattern of 5' enrichment for H4K16ac at TSS on autosomes. This pattern remains unchanged in the *upSET* mutant cell line.
B) A representative genome browser view of a 73kb region of the X chromosome. H4K16ac ChIP-seq tracks show the enrichment of H4K16ac on the male X chromosome in S2 cells. The *upSET* mutant cells show expanded spreading of the H4K16ac mark compared to wild type the S2 cells.
C) Quantification of genome by chromosome arm covered by H4K16ac enrichment in wild type S2 and *upSET* mutant G3 cells. The *upSET* mutant cells have an increased coverage (~1Mb) of the X chromosome as compared to autosomes when compared to the parental S2 line. p-values are for a one-sided t-test.
D) Fold change for each peak was calculated as the breadth of the peak in the G3 *upSET* mutant cell line over the breadth of the peak in the S2 parental cell line, when there was any overlap between peaks in both datasets. P-value is for a Wilcoxen test.

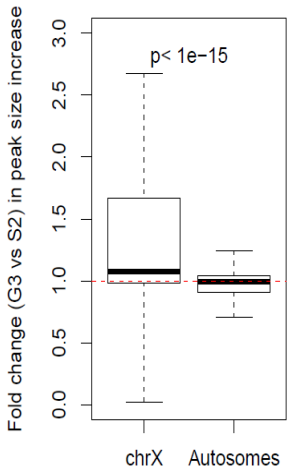
Figure 2-7 (Continued)



C



D



Bulk histone PTM analysis in *upSET* mutant S2 cells reveals perturbed chromatin state

Given that we only tested the dosage compensation mark, H4K16ac, in our ChIP-seq experiment, we sought to obtain a more comprehensive assessment of all histone post-translational modifications to determine whether H4K16ac was among the top modifications altered by the loss of *upSET*. We isolated bulk histones from S2 cells and all three *upSET* mutant cell lines using a salt/acid extraction method (Zee et al., 2016). To quantitatively recover histone peptides with their modification state intact for mass spectrometry analysis, histones were derivatized in solution with a protecting group that allows recovery of histone peptides from reverse phase chromatography. Using the mass difference between the protecting group and various modifications (acetylation, methylation, etc), we were able to quantify the relative abundance within each sample of different modifications in comparison to the unmodified peptide. Since we expected to observe changes in relative amounts of acetylations, we also assessed the histone modifications in S2 cells treated with the general HDAC inhibitor, sodium butyrate (S2but) (Candido et al., 1978).

We observed a slight increase in H4-monoacetylation on the peptide that contains lysines at residues 5, 8, 12, and 16 in *upSET* mutant cells, and a larger increase in the butyrate treated cells (Figure 2-8A). We deconvoluted the source of the acetyl observed on this peptide and saw that there was only a slight trend toward increased H4K16ac in *upSET* mutants (Figure 2-8B). We obtained similar results in replicates (Figure S1).

Unexpectedly, the largest change in modifications on total histones in the *upSET* mutant cells was in H3K9me2 levels, which were greatly diminished compared to wild-type S2 cells (Figure 2-8C). H3K9me2 is a modification known to be enriched in heterochromatin. The HP1 protein, which is critical for heterochromatin formation, interfaces with this mark via its chromodomain. Taken together with the UpSET-BioTAP ChIP observations above, the data suggest an as-yet undetermined role for UpSET in heterochromatin.

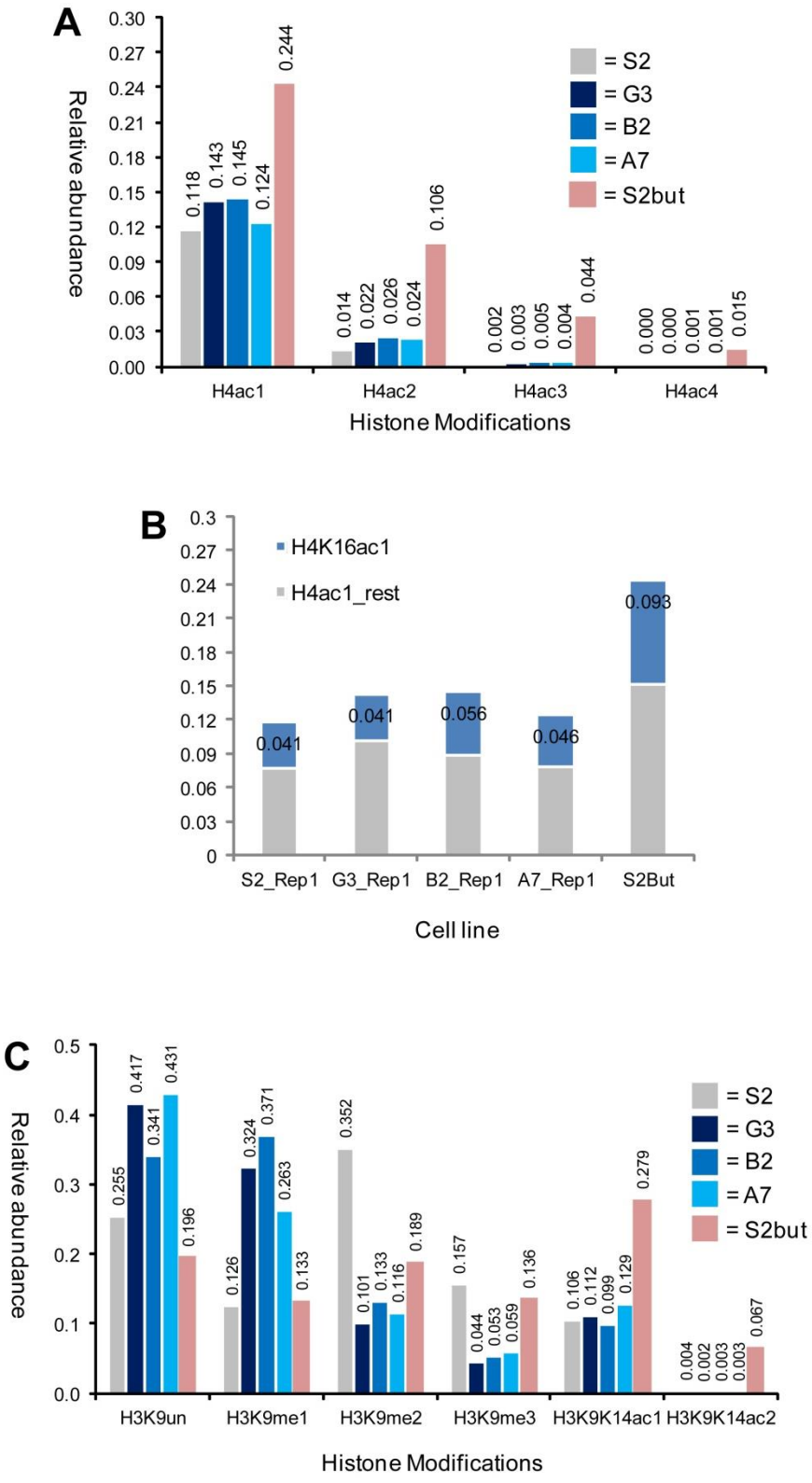
Figure 2-8: Analysis of histone post-translational modifications from bulk histones in *upSET* mutant cells

A) Relative quantification (1.0 = 100%) of H4K5K8K12K16 acetyl patterns in *upSET* mutant cell lines (G3, B2, and A7) or butyrate treated cells (S2but) with respect to the parental S2 line. Butyrate inhibits deacetylases, resulting in high enrichment of acetylation marks. The *upSET* mutant cell lines see a more modest increase in acetylated H4 peptide.

B) Deconvolution of monoacetyl H4 from (A) to identify which residue carries the acetyl mark. There is a slight trend for enrichment of H4K16ac in the *upSET* mutants cells compared to other residues.

C) Relative quantification (1.0 = 100%) of H3K9K14 PTM patterns in *upSET* mutant cell lines (G3, B2, and A7) or butyrate treated cells (S2but) with respect to the parental S2 line. The H3K9me2/me3 marks are depleted from *upSET* mutant S2 cells, suggestive of an effect on heterochromatin.

Figure 2-8 (Continued)



Lower H3K9me2 levels and deficient heterochromatin reflected by ChIP and PEV

To probe whether the UpSET-BioTAP heterochromatin localization and lower bulk levels of H3K9me2 in *upSET* mutant S2 cell lines had functional relevance, we assessed H3K9me2 genomic localization in the G3 *upSET*-mutant cell line. We observed lower overall coverage in heterochromatin in the *upSET* mutant cell line as compared to S2 cells. We also observed loss of H3K9me2 from regions on euchromatic chromosomes arms (Figure 2-9A). Taken together with the bulk histone data from all three *upSET* mutant cell lines and the UpSET-BioTAP localization from embryos, these observations in the G3 *upSET*-mutant cell line support a role for UpSET in heterochromatin, as its loss had clear repercussions on heterochromatin state.

To test this hypothesis in the organism, we tested whether heterozygous loss of *upSET* would have an effect on position effect variegation (PEV) of the *white*^{mottled4} *(w*^{m4}*)* allele seen in the eyes of adult flies. The *w*^{m4} allele carries an inversion on the X chromosome placing the *white* locus adjacent to heterochromatin (Elgin and Reuter, 2013). As such, defects in heterochromatin components result in greater expression of *white*$^{+}$ and a greater extent of red eye pigmentation sectoring (Figure 2-9B). The core components of heterochromatin score strongly in these assays and are collectively called suppressors of variegation (for example, Su(var)3-9, the enzyme responsible for H3K9 dimethylation). We scored the eye sectoring phenotypes of hemizygous *w*^{m4} males and compared heterozygous *upSET* individuals to their balancer carrying siblings and the balancer carrying control (for background effects of the balancers in the genotypes). We observed that loss of *upSET* results in suppression of variegation, that is, a larger number of flies with a greater extent of red pigmentation (Figure 2-9B). This *in vivo* finding, along with our above results, support the conclusion that UpSET plays a role in heterochromatin maintenance.

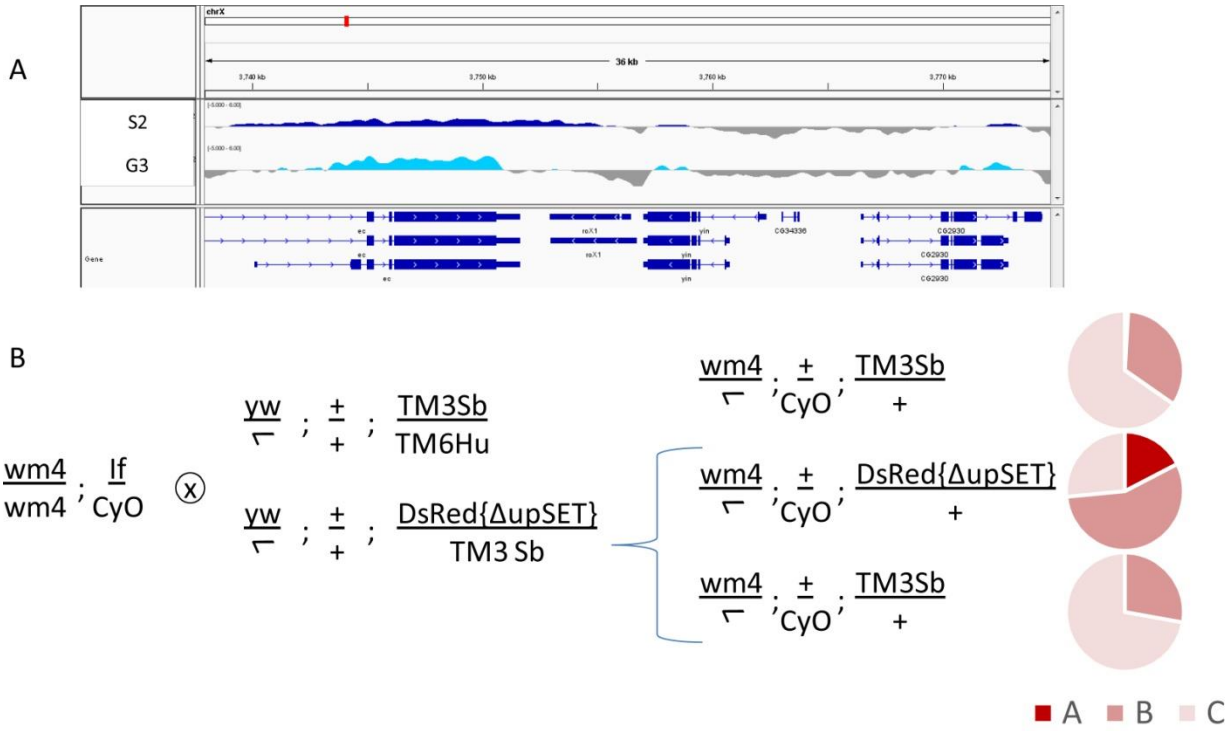
Changes in transcription reflect altered chromatin state

Figure 2-9: H3K9me2 depletion following *upSET* mutation

A) Representative 36kb region of the X chromosome showing decreased H3K9me2 in the G3 *upSET* mutant line.

B) Cross scheme for testing of the effect of heterozygous loss of *upSET* on position effect variegation. The w^{m4} line has a *white* variegating phenotype resulting in white eyes. As heterochromatin becomes less stable, there is loss of repression of *white* resulting in redder eyes. Male flies were sorted at 3 days post-eclosion into three classes: Class A has the highest proportion of red sectors, Class B has intermediate levels, and Class C is largely unpigmented. In comparison to siblings or a control cross, heterozygous loss of *upSET* leads to an increase in red pigmentation, skewing the population toward suppressed variegation.

Figure 2-9 (Continued)



After it became clear that mutation of *upSET* using CRISPR/Cas-9 had an effect on chromatin state, we wondered what the effect on transcription might be. In order to assess whether transcription tracked with the observed chromatin state, we utilized a urea-based method to sequence the nascently transcribed RNA still associated with Pol II. We elected to isolate nascent Pol II-associated RNAs since we hypothesized we might observe any changes in transcription more proximal to chromatin than in the steady state cytosolic pool of mRNA. Indeed, we observed that aberrant transcription occurred in all three *upSET* mutant S2 cell lines in comparison to the parental cell line (Figure 2-10).

Disappointingly, looking for statistically up- or down-regulated genes revealed striking heterogeneity in the transcriptional profiles between the three *upSET*-mutant cell lines. Therefore, we asked whether there might be a unifying principle if we looked instead at trends for groups of genes based on their locations in heterochromatin vs euchromatin or X-linked vs autosomal. To do so, we counted the numbers of genes falling above and below the no-fold-change ($log_2FC=0$) line per grouping (Figure 2-11A-C). The most striking trend, which was consistent for all three *upSET* mutant cell lines, was that the group of genes found in heterochromatin, as determined by the presence of H3K9me2 over the gene in wild-type S2 cells, was upregulated (Figure 2-11D-F, third bar). While the fold change of each gene might not be statistically significant, taken as a whole, the distribution of the heterochromatin genes skewed toward increased expression in a statistically significant manner.

When comparing X-linked versus autosomal genes, there was a weak trend toward upregulation, though this failed to reach the level of statistical significance in all three *upSET* mutant cell lines cell lines (Figure 2-11D-F, second bar). This result is not as striking as the heterochromatin vs euchromatin trend, but in light of the other data presented above, generally supports the same hypothesis. The loss of UpSET has a profound effect on the integrity of heterochromatin, as supported by the PEV of wm4, H3K9me2 ChIP-seq, and bulk histone quantification analyses. Loss of UpSET also has an X-specific effect, clearly observable in the H4K16ac ChIP-seq, yet the bulk histone analysis suggests

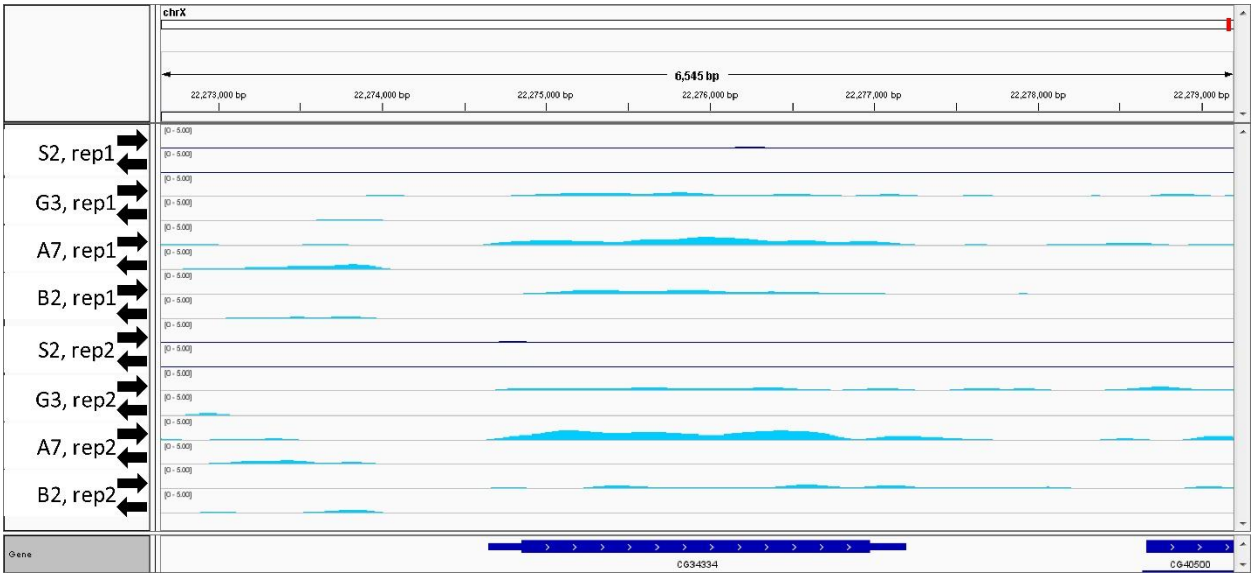
Figure 2-10: Nascent-RNA-seq reveals misregulation of transcription in *upSET* mutants

A) Genome browser view of nascent-RNA sequencing read density over the gene *CG34334* on the X chromosome. *CG34334* is transcribed on the top strand. Few reads are seen in S2 cells (dark blue tracks), but all three *upSET* mutant cell lines transcribe this gene (light blue tracks).

B) Genome browser view as in (A) except of the *CG15578* gene on the X chromosome. *CG15578* is transcribed from the top strand. In S2 cells there are many reads from antisense transcripts in this region, which are largely unchanged in the *upSET* mutant cell lines. All three *upSET* mutant cell lines, however, have increased transcription of the *CG15578* locus and several kb downstream as compared to S2 cells.

Figure 2-10 (Continued)

A



B

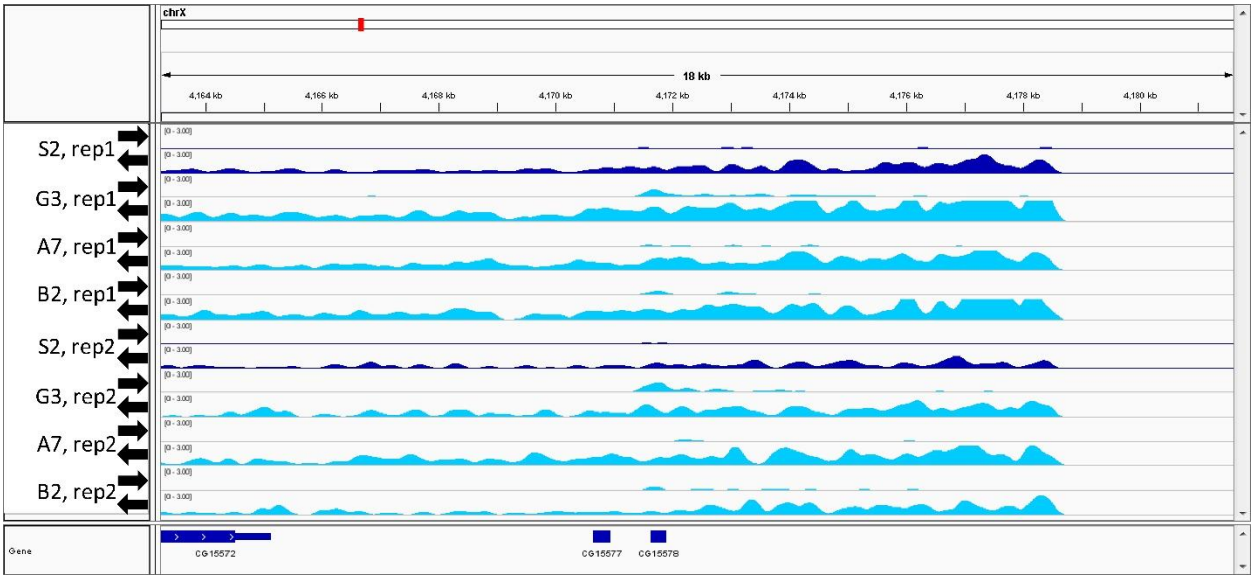


Figure 2-11: Increased transcription of heterochromatin and X-linked genes in *upSET* mutants

A) Scatter plot showing fold-change for individual genes based on their nascent-RNA levels in the G3 *upSET* mutant cell line on the y-axis vs wild type S2 expression level on the x-axis. Autosomal euchromatin genes are highlighted in blue. Euchromatin or heterochromatin annotations were based on H3K9me2 levels in S2 cells.

B) As in (A), except with heterochromatin genes highlighted

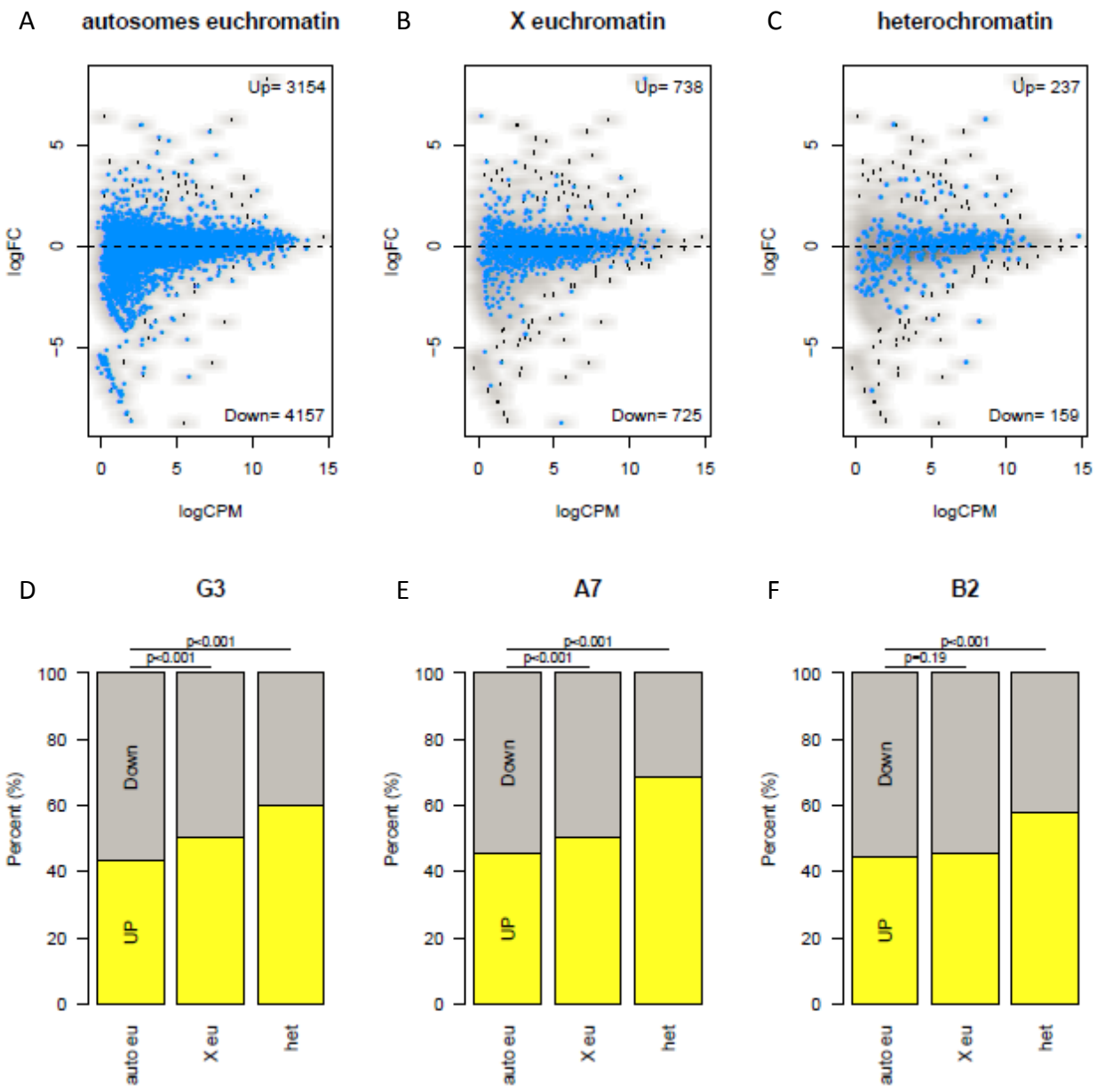
C) As in (A-B), except with X chromosome euchromatin genes highlighted

D) Summary of (A-C) with number of genes falling above and below the zero-fold change line from the G3 *upSET* mutant cell line. The general trends of increased expression on the X chromosome and of heterochromatin genes are statistically significant.

E) As in (D), except using data comparing the A7 *upSET* mutant cell line to S2 cells. The same general trends are observed as in (D).

F) As in (D-E), except using data comparing the B2 *upSET* mutant cell line to S2 cells. The same general trends are observed as in (D-E), however the trend for upregulation of the X chromosome does not reach the threshold for statistical significance in the B2 line.

Figure 2-11 (Continued)



that this effect is not as robust as the effect on heterochromatin. Since it has been proposed that MSL proteins play a role in activation of autosomal heterochromatin genes (Koya and Meller, 2015), it is possible that both the heterochromatin and X-specific effects are due to the UpSET-MSL interaction in male cells. However, we favor a model in which UpSET generally functions to protect heterochromatin by restricting active marks in both males and females, but serves an additional role in males to limit the MSL complex, keeping it in check to properly tune dosage compensation and MSL activity.

Discussion:

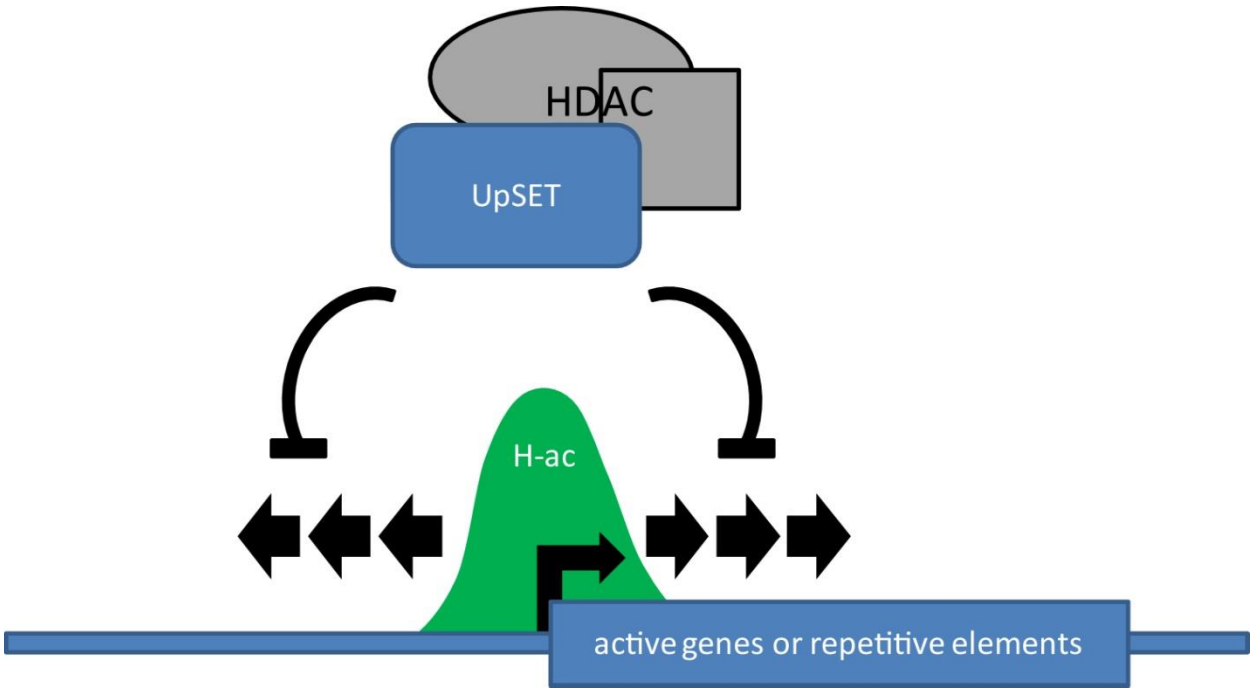
Chromatin and gene expression are intimately linked at the molecular level. Proteins that create and maintain chromatin domains therefore are critical for transcriptional fidelity of gene expression programs. Here we have further investigated one such chromatin protein, the SET-domain containing protein UpSET. Previous characterizations of this protein and its homologs SET3 in yeast and MLL5 in mammals have shown that it assembles into a complex with histone deacetylase activity (Pijnappel et al., 2001). Furthermore, the PHD finger of UpSET has been shown to interact with the histone post-translational modification H3K4me2/3 (Ali et al., 2013, Lemak et al., 2013), which results in recruitment of the HDAC complex to transcription start site proximal locations. Once there, the HDAC complex restricts the spread of these activating marks which prevents the improper activation of neighboring genes (Kim and Buratowski, 2009, Kim et al., 2012, Rincon-Arano et al., 2012). Our results are largely compatible with this model, yet suggest an additional, Drosophila specific, mode of action (Figure 2-12).

The original characterization of UpSET in Drosophila made use of a P-element insertion line, which left the coding region of the gene intact, which was described as homozygous viable with female sterility. Additionally, many of the experiments were performed in female cultured cells and in female tissues. Interestingly, our lab independently discovered UpSET as one of the most enriched proteins in

Figure 2-12: UpSET: a protein with diverse repressive responsibilities

The UpSET protein assembles into a complex with histone deacetylase (HDAC) activity. The exact members of this complex are unclear. Rpd3 and the co-factor Sin3 are candidates that have been shown to interact with UpSET (Rincon-Arano et al., 2012), but are not the direct homologs of the yeast SET3C (Pijnappel et al., 2001). The HDAC activity opposes run-away over activation of genes by limiting the spread of histone acetylation from TSS. Previous reports identified a role for UpSET in silencing of repetitive elements, as well. We have shown that UpSET-mediated repression has a particularly important role on the dosage compensated X chromosome and at heterochromatic genes.

Figure 2-12 (Continued)



crosslinked MSL3 purifications. This led us to seek whether UpSET plays a role in dosage compensation, a male-specific process in Drosophila.

To our surprise, we found that precise deletion of the *upSET* locus was lethal in both males and females. That we were able to rescue this lethality with a tagged UpSET transgene suggests that it is due specifically to the loss of UpSET. Furthermore, it suggests that the previously utilized P-element allele may be hypomorphic and still provide enough UpSET protein for viability, but then result in maternal effect lethality as characterized. Using our tagged allele we determined the genomic localization of UpSET protein in mixed embryos, confirming, while also refining, the previous result for UpSET localization to active genes by Dam-ID. UpSET had been shown to be important for repression of repetitive elements, and we observed UpSET to have enhanced binding at genes in heterochromatin regions as compared to the autosomes.

We then created S2 male cell lines that carried *upSET* mutations that introduce frameshifts to the UpSET open reading frame. The cell culture system proved invaluable for assessing the molecular impact of the loss of UpSET. We observed that loss of UpSET had a profound impact on the state of chromatin, which in broad strokes was consistent across all three cell lines. Consistent with a role related to deacetylation, we saw a modest increase in acetylated histone H4 in bulk histones and spreading of the MSL-deposited H4K16ac mark on the X chromosome in the only line tested by ChIP-seq to date. Perhaps more interestingly, we found that H3K9me2 levels in bulk histones were reduced in all three lines. ChIP-seq for H3K9me2 likewise showed reduced signal globally, with significant losses observed from genic regions of euchromatin. In addition, our nascent-RNA-sequencing showed increases in transcription from heterochromatin regions in all three lines. These molecular findings were further supported by our analysis *upSET*-mutant flies when we tested the impact of the heterozygous *DsRed^{+}{ΔupSET}* deletion on the position effect variegation phenotype of the wm4 allele.

Our findings showed suppressed *white* variegation, suggesting a loss of heterochromatin stability allowing the *white* locus to become expressed more readily.

Interestingly, there has long been an as yet unexplained relationship between heterochromatin and the X chromosome in Drosophila. The X chromosome is observed to be less compact in polytene chromosome preparations and its morphology is particularly sensitive to mutations in heterochromatin components such as Su(var)3-7, Su(var)3-9, and HP1a. Loss of these core heterochromatin factors leads to a swollen X chromosome, whereas the over-expression can lead to enhanced compaction, as compared to changes in autosomes. Furthermore, Jil-1 kinase, which is enriched approximately 2-fold on the male X in an MSL-dependent manner, is thought to prevent the spread of heterochromatin by catalyzing the H3S10ph modification. Jil-1 has a complex interplay with heterochromatin components, as evidence for phosphorylation of Su(var)3-9 (Boeke et al., 2010) and for a composite H3S10phK9me2 epigenetic mark (Wang et al., 2014) has been published. Indeed, using Jil-1 mutant larvae, it has been observed that H3K9me2 spreads from pericentric heterochromatin into the euchromatic gene arm, with a marked increase on the X chromosome in both sexes (modEncode, unpublished observations). Intriguingly this spread appeared to skip over gene bodies, suggesting additional non-Jil-1 mechanisms exist for protecting genes from the spread of heterochromatin.

Differences in experimental methods and in the sex of the cells or tissues used for experiments suggest that UpSET serves slightly different, yet essential, roles in the different sexes. Repression of repetitive elements reported in female cells is likely also a role male cells, yet with the functioning MSL complex positively influencing the expression of heterochromatin genes, UpSET's role may be especially important to keep this influence in check. Given the observed MSL-UpSET interaction, we favor a model in which the MSL complex physically interacts with UpSET in order to limit acetylation to gene bodies and thus fine tune expression levels.

The unique function of UpSET in heterochromatin in Drosophila may make sense in terms of evolutionary history, since the mammalian homolog MLL5 has been implicated in establishing proper DNA methylation, a repressed state that is not found in flies. However, the exact molecular role for MLL5 in this process has not been elucidated (Yun et al., 2014). One can posit that this derived trait in mammals is an extension of the repressive role of SET3 in yeast, with UpSET functioning in an intermediate manner. This expanding alternative repressive role explains how this family of atypical catalytically inactive SET-domain proteins has perdured through evolutionary time.

Materials and Methods:

Generating upSET mutant S2 cells and flies

Mutant S2 cells were generated using the CRISPR/Cas-9 system essentially as described (Housden et al., 2015). Guide RNA sequences were obtained from the Drosophila RNAi Screening Center (DRSC)'s sgRNA design tool (www.flyrnai.org/crispr2). Oligonucleotides were ordered from IDT of the appropriate gRNA sequence with additional bases to allow for ligation into the BbsI site of the pL018 plasmid (a gift from N. Perrimon), which also expresses Cas-9. S2 cells were transfected using Effectene (Qiagen) using 360ng of the appropriate pL018 construct and 40ng of an actin::RFP marker plasmid (a gift from T. Wu). Four days after transfection, single cells in the top 10% of RFP^+ cells (typically ~top3% of total population) were sorted by FACS into conditioned media in 96-well plates. Colony growth from single S2 cells was observed after 2-3 weeks. Independent lines were expanded and tested for mutations by HRMA (high resolution melt assay) using Precision Melt Supermix (Bio-Rad). Lines scoring well in the HRMA had the gRNA target region amplified by PCR using flanking primers, and the resulting product was subcloned into the pCR4Blunt-TOPO vector (Invitrogen). To identify the molecular lesion, 5 bacterial colonies per S2 line were sequenced by Genewiz. Lines with mutations that introduce

frameshifts or deletions of the start codon and adjacent sequence were used for subsequent experiments.

upSET directed gRNA sequences were as follows:

upset1- AACCGAGTCGTGACTGGACATGG.

upset3- AGGCGCGATGCCGTCTGATTAGG.

upset5- TGGCCAGGCGCAGTAGTAATAGG.

upset7- ACAGCAGATCAGCCTACCGCAGG.

Mutant flies were generated by injecting gRNA constructs and a homologous recombination donor into *w; w-{nos-cas9}/CyO* embryos (a gift from N. Perrimon, see also (Housden et al., 2014) for general guidelines for Drosophila CRISPR/Cas-9). The gRNA constructs were designed as above and oligos were ligated into the pU6-BbsI-chiRNA plasmid (addgene#45946) as described (Gratz et al., 2013). The homologous recombination donor was constructed using 1kb of genomic sequence flanking the 5' and 3' gRNA cut sites inserted into the pDsRed-attP plasmid (addgene#51019) (Gratz et al., 2014), which expresses the synthetic marker 3xP3-DsRed in the adult eye. Adults resulting from the injections were outcrossed to *yw* flies, and their progeny were screened for the fluorescent DsRed marker. DsRed-positive progeny were crossed to *yw; TM3/TM6* flies to generate balanced stocks of *yw; +; $DsRed^{+}${ΔupSET}/TM3* or *yw; +; $DsRed^{+}${ΔupSET}/TM6*.

Cloning and Transgenesis for the UpSET-BioTAP allele

The UpSET-BioTAP allele was constructed using the pRedET recombineering system (GeneBridges K002). The genomic region of *upSET* was transferred to the pFly (aka pGS-mw) vector and injected into flies for site-specific integration at 53B2 on the second chromosome (BestGene stock #9736). The resulting flies, *yw; UpSET-BioTAP/UpSET-BioTAP*, were crossed into the *$DsRed^{+}${ΔupSET}*

background. *DsRed*$^{+}$*{ΔupSET}*-homozygous flies carrying one or two copies of the UpSET-BioTAP allele were used for one-step ChIP (see below).

Presence of WT *upSET* or the BioTAP-tagged-*upSET* transgene was assessed by PCR from genomic DNA isolated from 2-3 female flies of the specific genotype. A 544bp product from WT and a 1210bp product from the BioTAP-tagged *upSET* construct are obtained when using the following primer pair:

KAM 201: gctgcacatgtttgatgataagc

KAM202: gtgcaagctcatactttatgcgc

ChIP-seq from S2 cells

S2 cell lines for ChIP-seq were crosslinked in culture media with 1.8% formaldehyde for 10min at room temperature. Formaldehyde was quenched by adding glycine to a final concentration of 0.125M. Fixed cells were pelleted at 1500g at 4C and washed with PBS, ChIP Wash A (10mM HEPES, pH7.6, 10mM EDTA, 0.5mM EGTA, 0.25% Triton) and ChIP Wash B (10mM HEPES, pH7.6, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.01% Triton). Following washing, cells were frozen in liquid nitrogen and stored at -80C. After thawing, cells were resuspended in ChIP Wash B and dounced 5-10 strokes to fully resuspend any aggregates. Cells were then thoroughly resuspended in TE buffer, pH8, and 10% SDS was added to 1% final concentration for lysis. Following lysis, chromatin was washed twice with TE, and finally resuspended in TE+0.5mM PMSF. 10% SDS was added to a final concentration of 0.1%, and 2mL aliquots were made into 15mL polystyrene Falcon conical tubes. Sonication was done using a Bioruptor (Diagenode) on the “high” power setting as 2x10min + 1x4min sessions of 30s on/30s off pulses, with ice bath exchanges between sessions. Following sonication, aliquots were pooled and supplemented sequentially with Triton X100 to a 1% final concentration, Sodium-DOC to 0.1% final, and NaCl to 140mM final. Debris was removed by centrifugation at 14000g at 4C. Sonicated chromatin was

aliquoted as necessary, frozen in liquid nitrogen, and stored at -80C. 500uL of chromatin was used per ChIP, with an additional 100uL chromatin reserved for input. IPs were done overnight using 3uL of α-H4K16ac (Millipore 07-329) (Gelbart et al., 2009) or 3uL α-H3K9me2 (abcam 1220). Immunocomplexes were captured with protein-A-agarose for 3hours at 4C. Bound complexes were washed 5x with RIPA (140mM NaCl, 10mM Tris pH8, 1mM EDTA pH8, 1% Triton, 0.1% SDS, 0.1% sodium deoxycholate), once with LiCl buffer(250mM LiCl, 10mM Tris pH8, 1mM EDTA pH8, 0.5% NP40, 0.5% sodium deoxycholate), twice with TE, and finally resuspended in TE. Input and IPs were treated for 30min with RNase at 37C, then overnight with the addition of proteinaseK and SDS (0.5% final), and crosslinks were reversed for 6hrs at 65C. IP samples were supplemented with NaCl to 140mM final, and both IP and input samples were extracted with an equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol. To maximize recovery in IPs, the organic fraction was extracted with TEN140 (TE+140mM NaCl) and pooled with the initial aqueous phase. All samples were then extracted with an equal volume of 24:1 chloroform:isoamyl alcohol, and precipitated overnight at -80C with sodium acetate and ethanol, in the presence of glycogen. The entirety of the precipitated IP-DNA and ~200ng of input DNA were used to create high-throughput sequencing Illumina libraries using the NEBNext ChIP-seq kit (NEB 6240). Prior to library amplification, size selection was achieved using a 2% agarose gel (Lonza 50111). Sequencing was performed at the Tufts Genomics Core.

Small scale one-step ChIP-seq from BioTAP-tagged embryos

To assess the genomic localization of BioTAP-tagged proteins from a small scale of embryos, immunoprecipitation using only the proteinA moieties of the tag were performed. Using a protocol essentially described elsewhere (Alekseyenko et al., 2008), 0.1 grams of embryos were collected and disrupted using a motorized pestle. Formaldehyde was added to 1% final concentration and incubated for 15min at room temperature. Following quenching of the reaction with glycine and washing, fixed

material was sonicated in RIPA buffer using a Bioruptor, 4 cycles of 30s on/ 30s off on the high setting. Sonicated material was supplemented with TritonX-100 to 1%, Sodium DOC to 0.1%, and NaCl to 140mM, and debris was cleared by centrifugation. Chromatin was aliquoted and stored at -80C until IP. For ChIP, 20-30uL of IgG agarose slurry per IP were washed in RIPA buffer and incubated with chromatin overnight. Bound immunocomplexes were processed for sequencing as described above for ChIP from S2 cells.

ChIP-seq analysis

The adaptor sequences were trimmed with Cutadapt ver. 1.2.1 (Martin, 2011). The reads were aligned to the Drosophila genome (dm3 assembly) using Bowtie ver. 12.0 (Langmead et al., 2009) with a unique mapping option (-m 1). Only uniquely aligned reads were used for the entire analyses. The input normalized fold enrichment profiles were generated using *get.smoothed.enrichment.mle* function of SPP R package (Kharchenko et al., 2008) with a step size of 20 bp and Gaussian kernel bandwidth of 150 bp. The profiles were normalized by the background scaling method. For metagene plots, the regions in the gene body except for 500 bp margins of 5'-end and 3'-end were scaled and averaged after merging two replicates. Genes that are larger than 1.5 kb and its distance from the adjacent gene is larger than 1 kb were only included in the metagene analysis. To estimate gene expression, RNA-seq samples profiled by modENCODE consortium were used for S2 cell and 14-16h embryos (Gerstein et al., 2014). FPKM = 1 was used as a threshold for expressed genes. To detect significantly enriched peaks, *get.broad.enrichment.clusters* function of SPP R package was used with a window size of 1 kb and z-score threshold of 3. The genomic annotation for UpSET and Fs(1)h was performed using CEAS (Shin et al., 2009). For the chromatin annotation, the chromatin segmentations were obtained from the previous studies (S2 from (Kharchenko et al., 2011) and embryos from (Ho et al., 2014)). To compare the H4K16ac genomic coverage between the mutant and control, significantly enriched regions were

detected after subsampling to make the same sequencing depth across samples because the size of significantly enriched regions increases for greater sequencing depth (Jung et al., 2014). To test the significance of the changes in the H4K16ac genomic coverage, a one sided t-test was used. To compare peak breadths between the mutant and control, the fold change in the peak breadth was calculated for each overlapped enrichment region of H4K16ac. To access the significance, a paired one-sided Wilcoxen rank sum test was used.

Bulk histone purification and mass spec

Bulk histones were salt-acid extracted from S2 cell lines for analysis by mass spectrometry as described (Zee et al., 2016).

Position Effect Variegation of the w^{m4} *allele*

Virgin females of the genotype w^{m4}/w^{m4}*; If/CyO* were crossed in parallel to males of the test and control genotypes at 24°C. The test genotype was *yw/Y; +; DsRed*$^{+}${*ΔupSET*}*/TM3,Sb*. The control genotype was *yw/Y; +; TM3,Sb/TM6,Hu,Tb*. Resulting male progeny of these crosses were screened for *CyO* and sorted by the appropriate 3rd chromosome genotypes by balancer chromosome markers. These flies were maintained at 24°C for 3 days, after which variegation of eye pigmentation was assessed.

Nascent-RNA-seq from S2 cells

Nascent-RNA sequencing was done using a urea-based method similar to NET-seq (Churchman and Weissman, 2011, Churchman and Weissman, 2012), and essentially as reported elsewhere (Alekseyenko et al., 2015). In short, 1x10^{7} S2 cells were collected by centrifugation at 300g at 4C, and homogenized in CKS buffer + SUPERase•In RNase inhibitor (Ambion AM2696) + ProteaseArrest (G-Biosciences 786-108) with 3 strokes through a 25G needle. Nuclei were collected by centrifugation and resuspended in CF

buffer + RNasin. NUN buffer was added and samples were vortexed ~30s until a wispy, filamentous precipitate was apparent. This precipitate was spun down and washed 3 times with NUN buffer. Samples were then treated with proteinaseK in CF buffer + 0.5% SDS at 55C for 30 min. Samples were then passed through a 25G needle 5 times to disrupt the chromatin, and incubated an additional 30 min at 55C. Samples were extracted twice with 25:24:1 phenol:chloroform:isoamyl alcohol, once with 24:1 chloroform:isoamyl alcohol, and ethanol precipitated overnight in the presence of glycogen (Ambion AM9510). Resulting nucleic acids were treated with RNase-free TURBO DNase (Ambion AM2238) for 30 min at 37C, with proteinaseK +SDS for an additional 5min at 37C, and then extracted once each with 25:24:1 phenol:chloroform:isoamyl alcohol and 24:1 chloroform:isoamyl alcohol. Nascent-RNA was ethanol precipitated overnight except without the addition of glycogen. Illumina sequencing libraries were constructed using the NEBNext Ultra Directional RNA Library kit (NEB 7420).

Nascent-RNA-seq analysis

The reads were mapped as described above. The tag density profiles were generated using *get.smoothed.tag.density* function of SPP R package with a Gaussian kernel of 100 bp and a step size of 10 bp after library size normalization. To compare the enrichment values between the mutant and control, the reads were separated to sense and anti-sense transcripts. The fold change and significantly changed regions in transcription between conditions were determined using EdgeR (Robinson et al., 2010) after TMM (trimmed mean normalization method) (Ross-Innes et al., 2012). Heterochromatic and euchromatic genes were defined using the heterochromatin and euchromatin boundary information for each chromosome obtained from a previous study based on H3K9me2 enrichment levels (Riddle et al., 2011). To access the significance for the portions of up-regulated genes between groups, a bootstrap method was used (n=1000).

References:

Alekseyenko, A.A., Peng, S., Larschan, E., Gorchakov, A.A., Lee, O.K., Kharchenko, P., McGrath, S.D., Wang, C.I., Mardis, E.R., Park, P.J., and Kuroda, M.I. 2008. A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* 134: 599-609.

Alekseyenko, A.A., Walsh, E.M., Wang, X., Grayson, A.R., Hsi, P.T., Kharchenko, P., Kuroda, M.I., and French, C.A. 2015. The oncogenic BRD4-NUT chromatin regulator drives aberrant transcription within large topological domains. *Genes Dev* 29: 1507-1523.

Ali, M., Rincon-Arano, H., Zhao, W., Rothbart, S.B., Tong, Q., Parkhurst, S.M., Strahl, B.D., Deng, L.W., Groudine, M., and Kutateladze, T.G. 2013. Molecular basis for chromatin binding and regulation of MLL5. *Proc Natl Acad Sci U S A* 110: 11296-11301.

Bassett, A.R., Tibbit, C., Ponting, C.P., and Liu, J.L. 2013. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell Rep* 4: 220-228.

Boeke, J., Regnard, C., Cai, W., Johansen, J., Johansen, K.M., Becker, P.B., and Imhof, A. 2010. Phosphorylation of SU(VAR)3-9 by the chromosomal kinase JIL-1. *PLoS One* 5: e10042.

Candido, E.P.M., Reeves, R., and Davie, J.R. 1978. Sodium Butyrate Inhibits Histone Deacetylation in Cultured Cells. *Cell* 14: 105-113.

Cheng, F., Liu, J., Zhou, S.H., Wang, X.N., Chew, J.F., and Deng, L.W. 2008. RNA interference against mixed lineage leukemia 5 resulted in cell cycle arrest. *Int J Biochem Cell Biol* 40: 2472-2481.

Churchman, L.S., and Weissman, J.S. 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469: 368-373.

Churchman, L.S., and Weissman, J.S. 2012. Native elongating transcript sequencing (NET-seq). *Curr Protoc Mol Biol* Chapter 4: Unit 4 14 11-17.

Deng, L.W., Chiu, I., and Strominger, J.L. 2004. MLL 5 protein forms intranuclear foci, and overexpression inhibits cell cycle progression. *Proc Natl Acad Sci U S A* 101: 757-762.

Elgin, S.C., and Reuter, G. 2013. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harb Perspect Biol* 5: a017780.

Emerling, B.M., Bonifas, J., Kratz, C.P., Donovan, S., Taylor, B.R., Green, E.D., Le Beau, M.M., and Shannon, K.M. 2002. *MLL5*, a homolog of *Drosophila trithorax* located within a segment of chromosome band 7q22 implicated in myeloid leukemia. *Oncogene* 21: 4849-4854.

Gelbart, M.E., Larschan, E., Peng, S., Park, P.J., and Kuroda, M.I. 2009. Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nat Struct Mol Biol* 16: 825-832.

Gerstein, M.B., Rozowsky, J., Yan, K.K., Wang, D., Cheng, C., Brown, J.B., Davis, C.A., Hillier, L., Sisu, C., Li, J.J., Pei, B., Harmanci, A.O., Duff, M.O., Djebali, S., Alexander, R.P., Alver, B.H., Auerbach, R., Bell, K., Bickel, P.J., Boeck, M.E., Boley, N.P., Booth, B.W., Cherbas, L., Cherbas, P., Di, C., Dobin, A., Drenkow, J., Ewing, B., Fang, G., Fastuca, M., Feingold, E.A., Frankish, A., Gao, G., Good, P.J., Guigo, R., Hammonds, A., Harrow, J., Hoskins, R.A., Howald, C., Hu, L., Huang, H., Hubbard, T.J., Huynh, C., Jha, S., Kasper, D., Kato, M., Kaufman, T.C., Kitchen, R.R., Ladewig, E., Lagarde, J., Lai,

E., Leng, J., Lu, Z., MacCoss, M., May, G., McWhirter, R., Merrihew, G., Miller, D.M., Mortazavi, A., Murad, R., Oliver, B., Olson, S., Park, P.J., Pazin, M.J., Perrimon, N., Pervouchine, D., Reinke, V., Reymond, A., Robinson, G., Samsonova, A., Saunders, G.I., Schlesinger, F., Sethi, A., Slack, F.J., Spencer, W.C., Stoiber, M.H., Strasbourger, P., Tanzer, A., Thompson, O.A., Wan, K.H., Wang, G., Wang, H., Watkins, K.L., Wen, J., Wen, K., Xue, C., Yang, L., Yip, K., Zaleski, C., Zhang, Y., Zheng, H., Brenner, S.E., Graveley, B.R., Celniker, S.E., Gingeras, T.R., and Waterston, R. 2014. Comparative analysis of the transcriptome across distant species. *Nature* 512: 445-448.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O'Connor-Giles, K.M. 2013. Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics* 194: 1029-1035.

Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O'Connor-Giles, K.M. 2014. Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in Drosophila. *Genetics* 196: 961-971.

Heuser, M., Yap, D.B., Leung, M., de Algara, T.R., Tafech, A., McKinney, S., Dixon, J., Thresher, R., Colledge, B., Carlton, M., Humphries, R.K., and Aparicio, S.A. 2009. Loss of MLL5 results in pleiotropic hematopoietic defects, reduced neutrophil immune function, and extreme sensitivity to DNA demethylation. *Blood* 113: 1432-1443.

Ho, J.W., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.A., Minoda, A., Tolstorukov, M.Y., Appert, A., Parker, S.C., Gu, T., Kundaje, A., Riddle, N.C., Bishop, E., Egelhofer, T.A., Hu, S.S., Alekseyenko, A.A., Rechtsteiner, A., Asker, D., Belsky, J.A., Bowman, S.K., Chen, Q.B., Chen, R.A., Day, D.S., Dong, Y., Dose, A.C., Duan, X., Epstein, C.B., Ercan, S., Feingold, E.A., Ferrari, F., Garrigues, J.M., Gehlenborg, N., Good, P.J., Haseley, P., He, D., Herrmann, M., Hoffman, M.M., Jeffers, T.E., Kharchenko, P.V., Kolasinska-Zwierz, P., Kotwaliwale, C.V., Kumar, N., Langley, S.A., Larschan, E.N., Latorre, I., Libbrecht, M.W., Lin, X., Park, R., Pazin, M.J., Pham, H.N., Plachetka, A., Qin, B., Schwartz, Y.B., Shoresh, N., Stempor, P., Vielle, A., Wang, C., Whittle, C.M., Xue, H., Kingston, R.E., Kim, J.H., Bernstein, B.E., Dernburg, A.F., Pirrotta, V., Kuroda, M.I., Noble, W.S., Tullius, T.D., Kellis, M., MacAlpine, D.M., Strome, S., Elgin, S.C., Liu, X.S., Lieb, J.D., Ahringer, J., Karpen, G.H., and Park, P.J. 2014. Comparative analysis of metazoan chromatin organization. *Nature* 512: 449-452.

Housden, B.E., Lin, S., and Perrimon, N. 2014. Cas9-based genome editing in Drosophila. *Methods Enzymol* 546: 415-439.

Housden, B.E., Valvezan, A.J., Kelley, C., Sopko, R., Hu, Y., Roesel, C., Lin, S., Buckner, M., Tao, R., Yilmazel, B., Mohr, S.E., Manning, B.D., and Perrimon, N. 2015. Identification of potential drug targets for tuberous sclerosis complex by synthetic screens combining CRISPR-based knockouts with RNAi. *Sci Signal* 8: rs9.

Jung, Y.L., Luquette, L.J., Ho, J.W., Ferrari, F., Tolstorukov, M., Minoda, A., Issner, R., Epstein, C.B., Karpen, G.H., Kuroda, M.I., and Park, P.J. 2014. Impact of sequencing depth in ChIP-seq experiments. *Nucleic Acids Res* 42: e74.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., Linder-Basso, D., Plachetka, A., Shanower, G., Tolstorukov, M.Y., Luquette, L.J., Xi, R., Jung, Y.L., Park, R.W., Bishop, E.P., Canfield, T.K., Sandstrom, R., Thurman, R.E., MacAlpine, D.M., Stamatoyannopoulos, J.A., Kellis, M., Elgin, S.C., Kuroda, M.I.,

Pirrotta, V., Karpen, G.H., and Park, P.J. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* 471: 480-485.

Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* 26: 1351-1359.

Kim, T., and Buratowski, S. 2009. Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* 137: 259-272.

Kim, T., Xu, Z., Clauder-Munster, S., Steinmetz, L.M., and Buratowski, S. 2012. Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. *Cell* 150: 1158-1169.

Koya, S.K., and Meller, V.H. 2015. Modulation of Heterochromatin by Male Specific Lethal Proteins and roX RNA in Drosophila melanogaster Males. *PLoS One* 10: e0140259.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25.

Lemak, A., Yee, A., Wu, H., Yap, D., Zeng, H., Dombrovski, L., Houliston, S., Aparicio, S., and Arrowsmith, C.H. 2013. Solution NMR structure and histone binding of the PHD domain of human MLL5. *PLoS One* 8: e77020.

Lucchesi, J.C., and Kuroda, M.I. 2015. Dosage compensation in Drosophila. *Cold Spring Harb Perspect Biol* 7.

Madan, V., Madan, B., Brykczynska, U., Zilbermann, F., Hogeveen, K., Dohner, K., Dohner, H., Weber, O., Blum, C., Rodewald, H.R., Sassone-Corsi, P., Peters, A.H., and Fehling, H.J. 2009. Impaired function of primitive hematopoietic cells in mice lacking the Mixed-Lineage-Leukemia homolog MLL5. *Blood* 113: 1444-1454.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17: 10-12.

Pijnappel, W.W., Schaft, D., Roguev, A., Shevchenko, A., Tekotte, H., Wilm, M., Rigaut, G., Seraphin, B., Aasland, R., and Stewart, A.F. 2001. The S. cerevisiae SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. *Genes Dev* 15: 2991-3004.

Riddle, N.C., Minoda, A., Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Tolstorukov, M.Y., Gorchakov, A.A., Jaffe, J.D., Kennedy, C., Linder-Basso, D., Peach, S.E., Shanower, G., Zheng, H., Kuroda, M.I., Pirrotta, V., Park, P.J., Elgin, S.C., and Karpen, G.H. 2011. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res* 21: 147-163.

Rincon-Arano, H., Halow, J., Delrow, J.J., Parkhurst, S.M., and Groudine, M. 2012. UpSET recruits HDAC complexes and restricts chromatin accessibility and acetylation at promoter regions. *Cell* 151: 1214-1228.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., Ali, S., Chin, S.F., Palmieri, C., Caldas, C., and Carroll, J.S. 2012.

Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481: 389-393.

Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. 2009. CEAS: cis-regulatory element annotation system. *Bioinformatics* 25: 2605-2606.

Torres-Machorro, A.L., Clark, L.G., Chang, C.S., and Pillus, L. 2015. The Set3 Complex Antagonizes the MYST Acetyltransferase Esa1 in the DNA Damage Response. *Mol Cell Biol* 35: 3714-3725.

Wang, A., Kurdistani, S.K., and Grunstein, M. 2002. Requirement of Hos2 histone deacetylase for gene activity in yeast. *Science* 298: 1412-1414.

Wang, C., Li, Y., Cai, W., Bao, X., Girton, J., Johansen, J., and Johansen, K.M. 2014. Histone H3S10 phosphorylation by the JIL-1 kinase in pericentric heterochromatin and on the fourth chromosome creates a composite H3S10phK9me2 epigenetic mark. *Chromosoma* 123: 273-280.

Wang, C.I., Alekseyenko, A.A., LeRoy, G., Elia, A.E., Gorchakov, A.A., Britton, L.M., Elledge, S.J., Kharchenko, P.V., Garcia, B.A., and Kuroda, M.I. 2013. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. *Nat Struct Mol Biol* 20: 202-209.

Yun, H., Damm, F., Yap, D., Schwarzer, A., Chaturvedi, A., Jyotsana, N., Lubbert, M., Bullinger, L., Dohner, K., Geffers, R., Aparicio, S., Humphries, R.K., Ganser, A., and Heuser, M. 2014. Impact of MLL5 expression on decitabine efficacy and DNA methylation in acute myeloid leukemia. *Haematologica* 99: 1456-1464.

Zee, B.M., Alekseyenko, A.A., McElroy, K.A., and Kuroda, M.I. 2016. Streamlined discovery of cross-linked chromatin complexes and associated histone modifications by mass spectrometry. *Proc Natl Acad Sci U S A*.

Zhang, Y., Wong, J., Klinger, M., Tran, M.T., Shannon, K.M., and Killeen, N. 2009. MLL5 contributes to hematopoietic stem cell fitness and homeostasis. *Blood* 113: 1455-1463.

Chapter 3
Characterization of PRC1 interactors reveals functional link to PRC2 or TrxG proteins

Contributions to this chapter: Hyuckjoon Kang performed the BioTAP-XL analyses of Pc, E(z), and Scm that led to this collaboration, and cloned several critical constructs. Fs(1)h BioTAP-XL in S2 cells was a collaboration between Hyuckjoon Kang and Kyle McElroy. LC-MS/MS was performed at the Taplin Mass Spectrometry Facility at Harvard Medical School. Barry Zee provided LC-MS/MS analysis. Lucy Jung performed the bioinformatics analysis of all ChIP experiments. Kyle McElroy performed all other experiments.

Abstract

The Polycomb Group are key developmental regulators that form several biochemically distinct multiprotein complexes. Investigation of PRC1 and PRC2 interacting proteins on chromatin using BioTAP-XL identified two proteins that may provide insight into the functional crosstalk amongst PcG complexes and between the PcG and the Trithorax group. Subsequent study of Fs(1)h and Scm confirm the physical interactions observed by mass spectrometry on chromatin, and further elucidate the roles these proteins play in establishing gene expression programs. We find that the majority Fs(1)h protein is involved in the well-characterized role it plays in transcriptional activation. The Fs(1)h interaction with the PcG is specific to PRC1, and the timing and location of these interactions may reveal bivalent states on chromatin during the course of development in Drosophila. The bridging of PRC1 and PRC2 by Scm may help to reinforce PcG silencing of key target regions. Taken together, gene activity fate may be dictated by mutually exclusive interaction partners for PRC1.

Introduction:

The Polycomb Group (PcG) are classically defined developmental repressors, first identified in Drosophila for their capacity to maintain an epigenetically silenced state. Subsequent work on these genes has shown that they are required for the proper maintenance of gene expression programs, not only of developmentally important loci, but hundreds of other sites as well (Negre et al., 2006, Schwartz et al., 2006, Tolhuis et al., 2006). While mutants for PcG genes display similar and stereotypical homeotic transformations, the PcG proteins assemble into several distinct multiprotein complexes in order to execute gene repression (Figure 1-3) (Simon and Kingston, 2013).

These complexes have several different biochemical activities and properties. The PRC1 complex contains the subunit Polycomb, whose chromodomain recognizes trimethylated lysine 27 of histone H3. PRC2 contains Enhancer of Zeste, a SET-domain protein, which catalyzes the trimethylation of histone 3 lysine 27 (Cao et al., 2002, Czermin et al., 2002, Muller et al., 2002). The dRing protein is common to both PRC1 and dRAF complexes, yet appears to only have H2A-directed ubiquitylization activity in the context of the latter (Lagarou et al., 2008). The recently characterized complex PR-DUB removes this ubiquitin mark, though the requirement for ubiquitin cycling remains unclear (Schwartz and Pirrotta, 2014). PhoRC has DNA-specific binding capability, but in general is not sufficient in and of itself to recruit the other complexes (Klymenko et al., 2006).

While the biochemical activities each complex possesses is fairly well characterized, how exactly they function together to identify target sites and then the molecular cascade to establish silencing are not clear (McElroy et al., 2014). To try to identify key proteins that may assist the PcG in doing so, Hyuckjoon Kang, a postdoc in the lab, BioTAP-tagged Pc and E(z) subunits as representatives for PRC1 and PRC2, respectively. His proteomic analyses of Pc and E(z) BioTAP-XL experiments in transgenic embryos confirmed that PRC1 and PRC2 are largely distinct complexes at the protein level (Figure 3-1A),

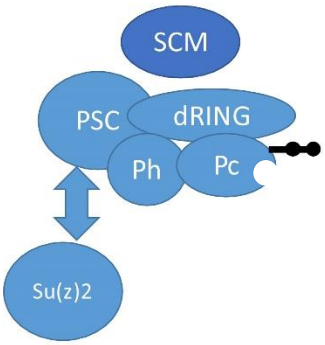
Figure 3-1: Pc and E(z) BioTAP suggest PRC1 and PRC2 are largely distinct and reveal interesting novel interactions

A) Summary of total peptides recovered by BioTAP-XL from Drosophila embryos for PcG proteins in Pc (PRC1) and E(z) (PRC2) experiments. PRC1 and PRC2 are schematized and the BioTAP-tagged subunit is indicated. There is only low level recovery for subunits of the other complex. Figure adapted from Kang, et al, 2015.

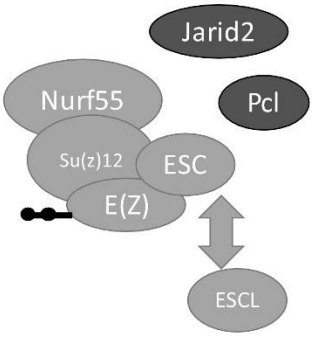
B) Top 10 most enriched proteins from the Pc BioTAP-XL experiments. Total peptide counts for Pc, E(z), MSL3, and in embryo input samples are given. The recovery of Fs(1)h in the Pc pulldown and Scm in the E(z) pulldown were unexpected, and led to our follow-up experiments.

Figure 3-1 (Continued)

A



Complex	Component	Total # of peptides	
		Pc (Embryo)	E(Z) (Embryo)
PRC1	Psc	**110**	7
	Su(z)2	**89**	0
	Sce (dRING)	**67**	2
	ph-p	**39**	2
	Pc	**27**	6
	Scm	**23**	24
	ph-d	**17**	0
PRC2	Su(z)12	4	**113**
	esc	0	**64**
	Nurf55	4	**47**
	Pcl	0	**45**
	E(z)	0	**44**
	Jarid2	2	**44**
	escl	0	**2**



Adapted from Kang et al 2015

B

Complex	Component	Total peptides			
		Pc (Embryo)	Input (Embryo)	E(Z) (Embryo)	MSL3 (Embryo)
PRC1	Psc	110	0	7	0
PRC1*	Su(z)2	89	0	0	0
	Fs(1)h	75	2	1	1
PRC1	Sce (dRING)	67	0	2	0
MOZ/MORF	CG1845	56	1	0	0
MOZ/MORF	enok	45	0	2	0
PRC1	ph-p	39	0	2	0
PRC1	Pc (bait)	27	0	6	0
PRC1*	Scm	23	0	24	0
PRC1*	ph-d	17	0	0	0

even though they colocalize extensively throughout the genome (Jung et al., 2016, Kang et al., 2015). In addition, when examining the proteins that were strongly copurified with Pc and E(z), two stood out in particular (Figure 3-1B).

The first, in the Pc pulldown, was the most strongly co-purified non-PcG protein, and was the product of the *Female sterile (1) homeotic* gene. This gene is the Drosophila version of the well-characterized double bromodomain protein Brd4, a known transcriptional activator and member of the Trithorax group. We were puzzled as to why the members of these seemingly diametrically-opposed chromatin groups would associate, and while we were confident that the interaction was real since it had also been identified by another group using an orthogonal proteomic technique (Strubbe et al., 2011), we undertook experiments to further characterize the nature of this interaction. We utilized BioTAP-tagged Fs(1)h constructs to confirm the PRC1-Fs(1)h reciprocal interaction and identify the breadth of Fs(1)h protein interactions. I also employed recombinant protein expression to provide evidence that Fs(1)h and PRC1 physically interact in a specific manner. This work is ongoing and will be incorporated into a future manuscript co-authored by HJ Kang and L Jung.

The second striking result was that in the E(z) pulldown, the Scm protein, a known member of PRC1, was purified at a high level. In the original characterization of Scm, it was noted that it associates with PRC1 at a substoichiometric ratio, suggesting Scm is not always with PRC1 (Klymenko et al., 2006, Ng et al., 2000, Peterson et al., 2004, Saurin et al., 2001). We reasoned that perhaps Scm might link PRC1 and PRC2, and I performed coimmunoprecipitations using recombinant proteins to confirm that Scm and PRC2 could physically interact. Our work on Scm has been successfully published in (Kang et al., 2015).

The work presented here explores two previously underappreciated facets of the molecular underpinnings of the PcG, which may shed light onto the mechanisms by which the PcG finds its appropriate targets during development and then initiates the maintenance of gene silencing. Both of

these mechanisms appear to require the coordination of many chromatin-associated multiprotein complexes, revealing how complex the interactions within the nucleus must be to properly maintain gene expression programs, and thus, cell identity.

Results:

BioTAP-tagging may allow the isoform-specific study of Fs(1)h

In order to test the validity of the PRC1-Fs(1)h interaction, Hyuckjoon Kang cloned the BioTAP tag into an Fs(1)h genomic transgene. The N-terminal BioTAP tag was placed just upstream of the start codon. The Fs(1)h gene produces several variant transcripts, which lead to two types of protein isoforms: a short form consisting of ~1000 amino acids containing both bromodomains and an extra terminal domain, and a long form consisting of ~2000 amino acids containing the entirety of the short form plus an additional carboxy-terminal motif (CTM). The placement of the N-terminal BioTAP tag should allow the expression of both the long and short isoforms carrying the amino-terminal BioTAP tag. The C-terminal BioTAP tag replaced the stop codon of the long isoform, and is competent only for the tagging of the long form (Figure 3-2A,B). When HJ Kang transfected these constructs into Drosophila S2 cells, and made nuclear and cytoplasmic extracts, it was observed that while the C-terminal tag appropriately tagged only the long form of Fs(1)h, the N-terminal tag appeared to only produce the short form of Fs(1)h (Figure 3-2C). This unexpected result for the Western Blot suggested that our tagging had created a method by which the two isoforms could be studied individually, since it has been reported that the long and the short forms may have unique functions.

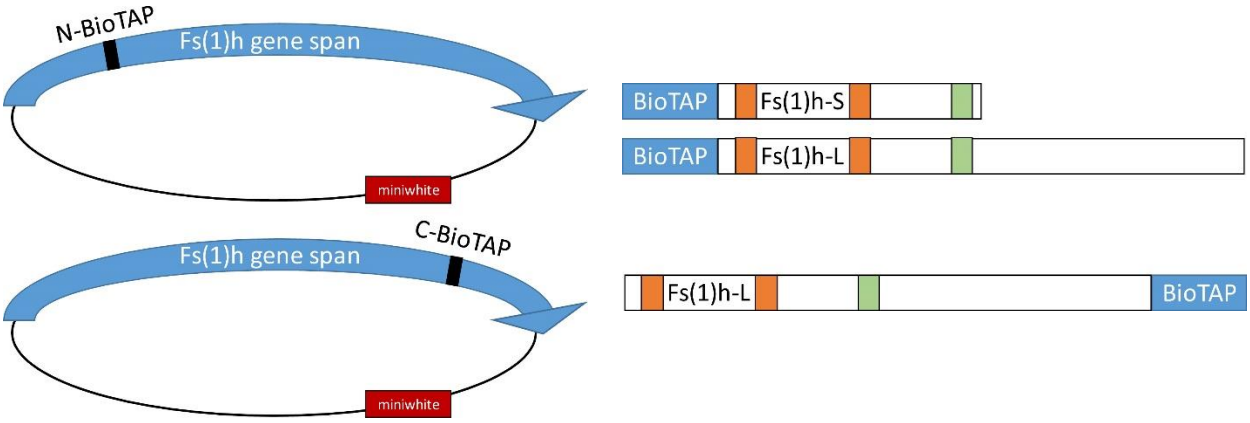
BioTAP-tagged Fs(1)h isoforms form nuclear speckles

Figure 3-2: BioTAP-tagging of Fs(1)h may allow for isoform specific study

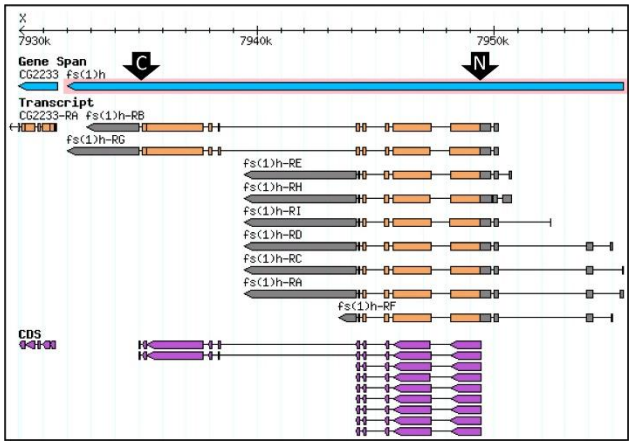
A) Amino- and carboxy-terminal BioTAP tags were cloned into an Fs(1)h genomic transgene. The placement of the tags should allow the N-terminal construct to express both the long and short forms of Fs(1)h, whereas the C-terminal construct can express only the tagged long version. Fs(1)h has two bromodomains (orange) and an extra terminal domain (green), which are common to both proteins.
B) Insertion locations of the N- and C-BioTAP tags in the genomic constructs superimposed over the Fs(1)h genomic locus with annotated transcripts.
C) Western blot for the proteinA moieties of the BioTAP tag from stably transfected S2 cells. Surprisingly, the N-BioTAP transgene appeared to only express the short form of Fs(1)h.

Figure 3-2 (Continued)

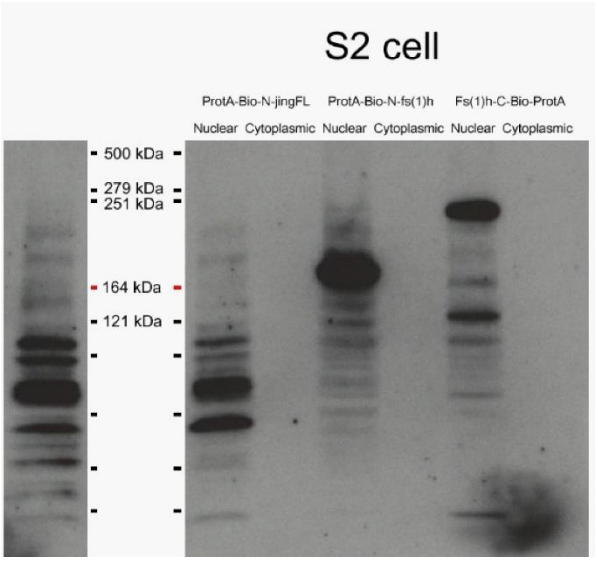
A



B



C



To determine whether the localization of the N- and C-terminal tagged fusion proteins differed, potentially reflecting alternate functions, I employed immunofluorescence using peroxidase-anti-peroxidase (PAP) to detect the tag in S2 cells. Both N- and C-terminal tags resulted in proteins enriched in the nucleus, as observed by Western blot (Figure 3-3). At the resolution of the single nucleus, I was unable to conclude that there were any differences in localization. Both preparations display numerous foci or speckles in the nucleus. In general, it appeared that a higher percentage of cells had robust expression of the N-terminal construct, which could reflect differential success of the transfection and selection steps to generate the stably expressing line. Alternatively, the different locations of the tag could reflect something about the transcription of the *Fs(1)h* locus, though we did not probe this more directly. The numerous foci and granular nuclear appearance of the tagged Fs(1)h staining is reminiscent of the mammalian Brd4 subnuclear localization, which is also observed to have a focal pattern (Dow et al., 2010, Farina et al., 2004).

BioTAP-XL ChIP-seq for Fs(1)h reveals few differences between N- and C-terminal tag

We sought to refine the localization data from the level of immunofluorescence to the level of the genome-wide binding pattern using BioTAP-XL. Using S2 cells stably expressing either the N- or C-terminal BioTAP tag, we prepared chromatin according to the BioTAP-XL procedure (Alekseyenko et al., 2015). We split each sample following tandem affinity purification, and isolated the immunoprecipitated genomic DNA from one quarter of the sample. Illumina libraries were prepared from this material and submitted for next-generation sequencing. The BioTAP-tagged Fs(1)h samples display a high degree of concordance with each other, suggesting that any isoform specific functions might not greatly affect the overall genomic localization (Figure 3-4).

We also compared the BioTAP-tagged Fs(1)h patterns to the genomic localization identified using ant-Fs(1)h antibodies for ChIP-seq experiments by our lab (Hyuckjoon Kang) and by the Paro lab

Figure 3-3: BioTAP-tagged Fs(1)h form nuclear speckles in S2 cells

A) Immunofluorescence resulting from peroxidase-anti-peroxidase (PAP) in untransfected S2 cells. PAP recognizes the proteinA moieties of the BioTAP tag. Some background staining of the cytoplasm is observed. DNA is counterstained with Hoechst.
B) Immunofluorescence in N-BioTAP-Fs(1)h expressing cells. Similarly to mammalian Brd4, Fs(1)h immunofluorescence suggests a granular distribution of the protein within the nucleus.
C) Immunofluorescence in Fs(1)h-C-BioTAP expressing cells. The overall level of expression is somewhat lower than in (B), but the nuclear foci are similar.

Figure 3-3 (Continued)

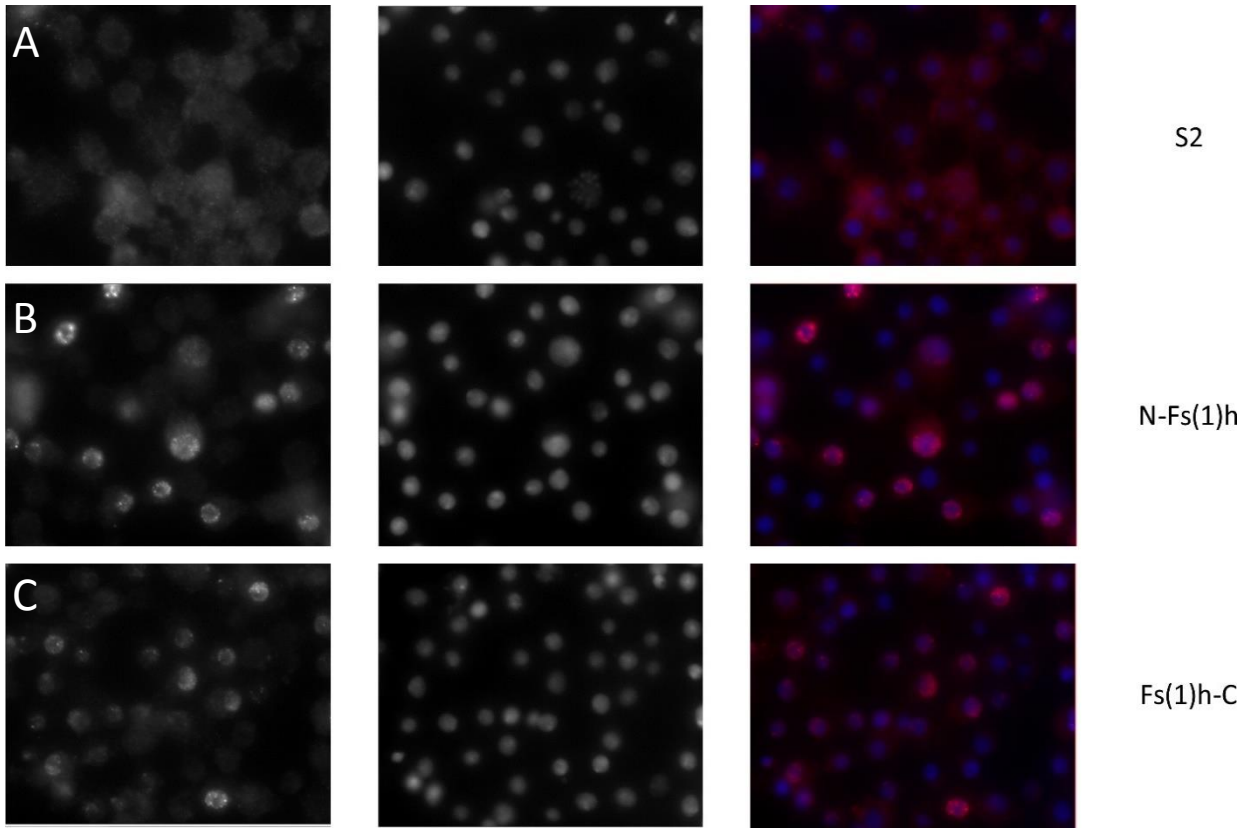


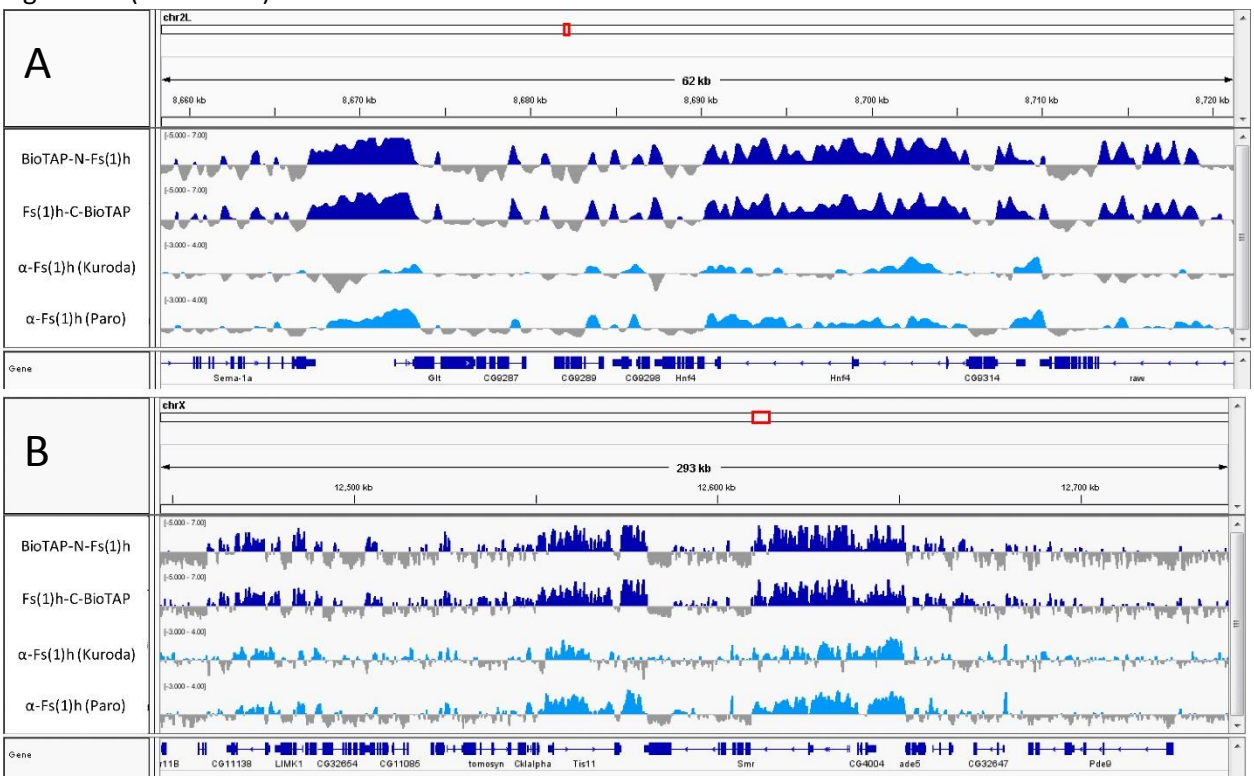
Figure 3-4: BioTAP-tagged Fs(1)h genomic localization is consistent with previous reports

A) A representative 62kb genome browser view of Chr2L. Both N- and C-terminal BioTAP constructs produce a highly concordant profile (dark blue tracks). For comparison, antibody ChIP-seq tracks from our lab and the Paro group are also displayed (light blue). There is strong agreement across all four tracks.

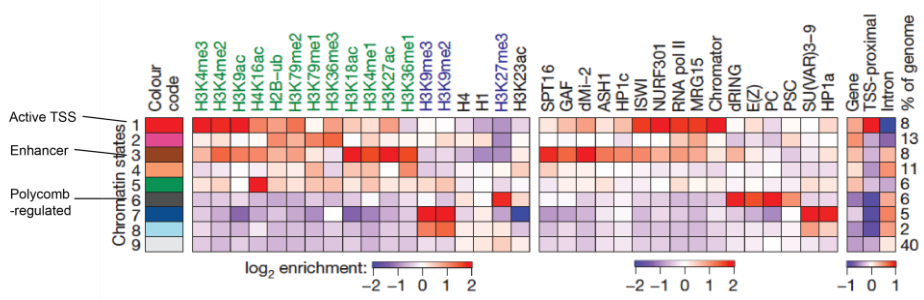
B) A representative 293kb genome browser view of the X chromosome. As in (A), there is strong overlap between the two Fs(1)h BioTAP constructs and antibody ChIP-seq experiments.

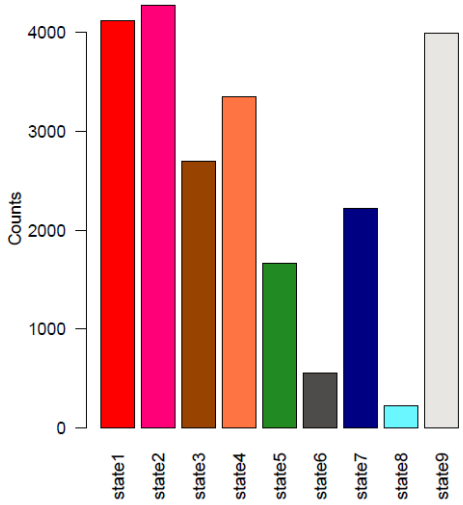
C) Kharchenko et al, (2011) used histone post-translational modifications and machine learning to identify 9 different classes of chromatin signatures (upper left panel of heat map). The upper middle panel shows the enrichment of chromosomal proteins by chromatin signature, and the further right panel shows the enrichment of genome features by chromatin signature. BioTAP-tagged Fs(1)h ChIP-seq peaks largely fall into the active signatures, with the highest enrichment of the transcriptional elongation signature (state 2) and active promoter/TSS-proximal regions (state 1).

Figure 3-4 (Continued)



C





(publically available). The overall patterns corresponded reasonably well, especially in regions of strong Fs(1)h enrichment (Figure 3-4B), and Fs(1)h binds primary to regions enriched for active chromatin marks (Kharchenko et al., 2011) (Figure 3-4C). The BioTAP Fs(1)h experiments appeared to have more signal outside these regions than the antibody ChIP-seq experiments, suggesting that the strength of the association of Fs(1)h at these sites may be weaker and thus only captured by the BioTAP-XL procedure.

It has been reported that the long form of Fs(1)h may interact with insulator proteins, whereas the short form does not (Kellner et al., 2013). In order to see if there was evidence for this in our ChIP-seq data, I visually compared the N- and C-terminally tagged Fs(1)h ChIP-seq patterns to the calculated enriched regions for CTCF, BEAF-32, GAF, Mod(mdg4), and Su(Hw) using both ChIP-chip and ChIP-seq data available from modEncode (datasets 3281, 2638, and 2639 for CTCF; 21 for BEAF-32; 23 for GAF; 24 for Mod(mdg4); and 27 for Su(Hw)). While there are regions where the Fs(1)h-L ChIP-seq track shows some modest enrichment compared to Fs(1)h-S at regions of multiple insulator binding (see * in Figure 3-5), the majority of individual insulator and multiply bound sites either are devoid of Fs(1)h binding or have similar levels in both the N- and C-terminal (S and L forms, respectively) tagged Fs(1)h experiments (see † in Figure 3-5). While the lack of a strong overlap or enrichment over multiply-bound insulator sites does not preclude the possibility of long form Fs(1)h-insulator protein interactions, it does suggest that any such interaction is not likely to be the primary role of Fs(1)h long form.

Fs(1)h BioTAP-XL reciprocally co-purifies PRC1 in addition to expected interactions with co-activators

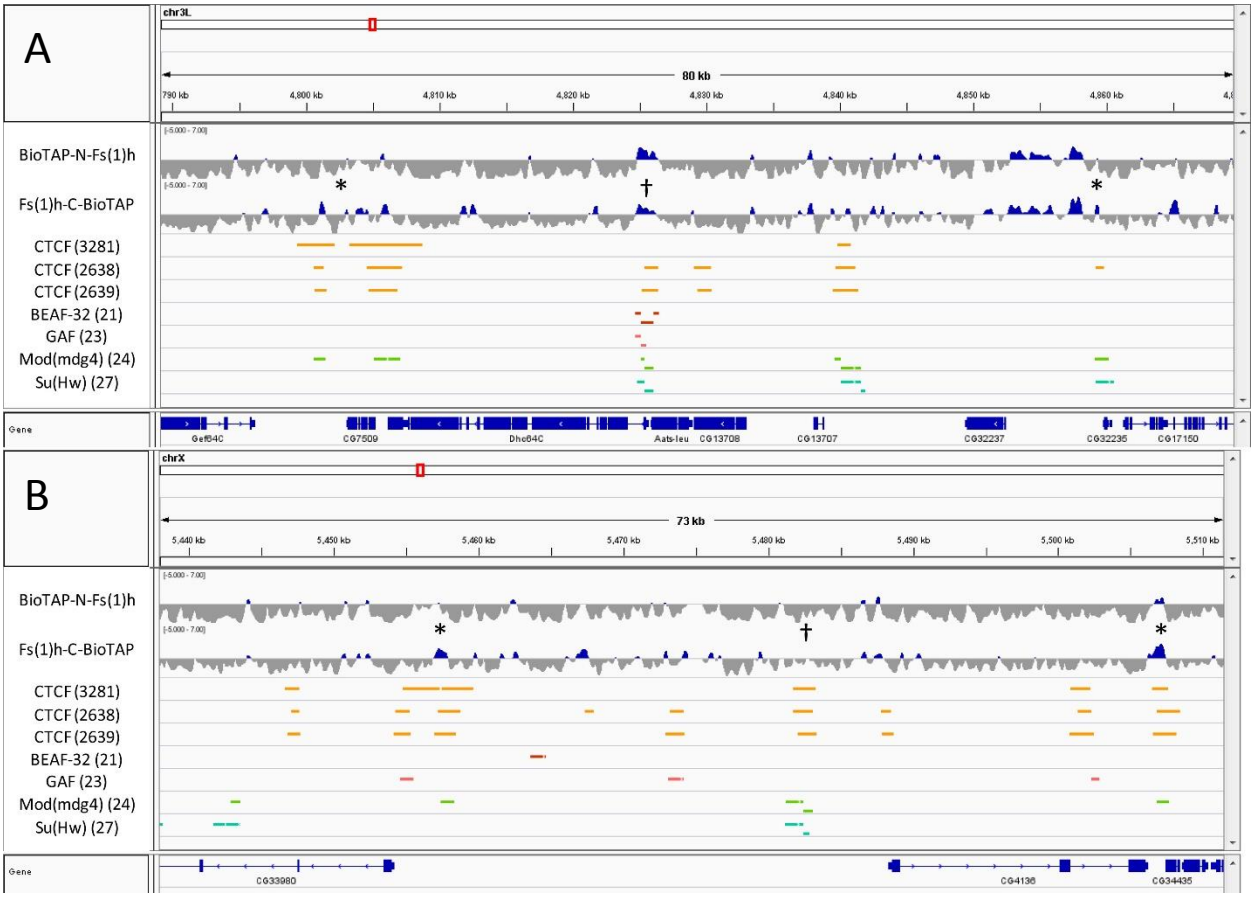
We treated the remainder of the affinity-purified material from the BioTAP-XL experiment with trypsin, purified the resulting peptides, and identified the proteins copurified with N- and C-terminally tagged Fs(1)h by LC-MS/MS. The vast majority of highly enriched proteins have functions related to transcriptional activation and maintenance of the active state (Table S1). This makes a great deal of sense, since Fs(1)h, and TrxG proteins in general, are well known for their ability to maintain gene

Figure 3-5: Weak evidence for a Fs(1)h-L-specific colocalization with insulators

A) An 80kb window of Chr3L. Peak calls were downloaded from modEncode for ChIP-chip experiments for CTCF, BEAF-32, GAF, mod(mdg4), and Su(Hw), and for ChIP-seq experiments for CTCF. These were then visually compared to the N-BioTAP-Fs(1)h and Fs(1)h-C-BioTAP tracks. While there were cases (*) where the Fs(1)h-C-BioTAP transgene showed enrichment over multiply bound insulator sites, many site (†) either lack any enrichment or have similar enrichment between N- and C- BioTAP-tagged Fs(1)h.

B) A 73kb window of ChrX, as in (A).

Figure 3-5 (Continued)



expression. Unlike the majority of TrxG proteins, however, Fs(1)h has not been found to be a constitutive member of any stoichiometric multi-protein complexes. Our proteomics result suggests that Fs(1)h does in fact interact with many other TrxG proteins (Table 3-1), but given the overall levels of enrichment across the different TrxG complexes, I would hypothesize that these interactions exist as a more transient association than a bona fide biochemically stable entity. The molecular model for Fs(1)h function is largely based off of the work of the mammalian field in studying the homolog Brd4. Like Brd4, Fs(1)h has two bromodomains and an extra terminal (ET) domain, with an additional C-terminal motif (CTM) in the long form. The model for Brd4 function is that the two bromodomains mediate the engagement with chromatin by recognition and binding of acetylated lysines on histone N-terminal tails. Acetylation marks are established by other enzymes in the active context of chromatin, and Fs(1)h/Brd4 binding facilitates active transcription by recruitment of several pro-transcription complexes, such as the Mediator complex and P-TEFb. It has also been reported that Brd4 is an atypical kinase which directly targets the PolII CTD (Devaiah et al., 2012), though evidence for this in Drosophila has not been shown.

Our proteomics does indeed lend support for Fs(1)h primarily functioning in transcriptional activation. The Mediator complex is well represented in the data, with 25 subunits identified. The strongest of these is MED1, which is part of the middle module of Mediator; there does not seem to be any overarching preference for one module or another, however, suggesting Fs(1)h may have multiple modes of interaction with Mediator. We also recover both the catalytic subunit of P-TEFb, CDK9, and the regulatory subunit, cyclin T (CycT). Some very elegant biochemistry previously has shown that as much as 50% of cellular P-TEFb is associated with Brd4 in mammalian tissue culture cells via a Brd4/CycT interaction (Yang et al., 2005). The specificity of the Brd4/CycT interaction appears to also be the case in Drosophila, as CycT is one of the highest enriched proteins in the pulldowns and more strongly enriched than CDK9. Interestingly, of the other two complexes that have a role in the transition to productive

Table 3-1: Fs(1)h strongly enriches for proteins related to its co-activator function

Total peptides recovered in the N-BioTAP-Fs(1)h, Fs(1)h-C-BioTAP, or S2 input samples are listed. Selected lists for TrxG members, Mediator complex members, and transcriptional regulatory factors PTEF-b, NELF, and DSIF are shown. The module within Mediator complex is also listed for each subunit and Fs(1)h shows no preference for any one module.

TrxG

Component	N-Fs(1)h	Fs(1)h-C	Input (S2)
fs(1)h	613	374	7
ash1	3	0	0
ash2	0	4	0
brm	26	11	8
bap55	16	11	5
bap60	14	9	5
dalao	10	8	2
dom	18	5	2
e(bx)	22	13	8
osa	19	9	2
gpp	2	3	0
iswi	30	19	11
kis	26	30	2
lid	14	17	5
lolal	2	2	0
mor	45	27	9
nurf38	4	0	2
trx	0	0	0
Trl	5	1	0
trr	24	8	0
Utx	8	3	0
vtd	12	4	0

Mediator complex

Component	N-Fs(1)h	Fs(1)h-C	Input (S2)	Module
MED1	28	31	0	middle
MED13	19	7	0	kinase
MED14	19	11	0	backbone
MED26	18	10	0	unassigned
MED17	16	11	0	head
MED12	12	7	0	kinase
MED8	10	2	0	head
MED23	9	12	0	tail
MED6	9	7	0	head
MED19	8	0	0	middle
MED7	8	11	0	middle
MED10	8	3	0	middle
MED21	6	4	0	middle
MED29	6	4	0	tail
MED30	5	2	1	head
MED11	5	4	0	head
MED16	5	6	0	tail
MED27	4	3	0	tail
MED24	3	4	0	tail
MED25	3	3	0	unassigned
MED4	3	3	0	middle
MED15	2	2	0	tail
MED20	2	0	0	head
MED31	1	1	0	middle
MED28	0	1	0	unassigned

P-TEFb

Component	N-Fs(1)h	Fs(1)h-C	Input (S2)
CycT	33	69	0
CDK9	10	16	0

NELF

Component	N-Fs(1)h	Fs(1)h-C	Input (S2)
Nelf-A	23	34	0
NELF-B	15	15	0
TH1 (NELF-D)	7	14	0
Nelf-E	6	9	0

DSIF

Component	N-Fs(1)h	Fs(1)h-C	Input (S2)
Spt5	6	4	2
Spt4	0	0	1

Table 3-1 (Continued)

elongation, only one is found in the Fs(1)h data. We recovered all four subunits of the NELF complex, which is phosphorylated by P-TEFb to allow for pause release into elongation (Yamaguchi et al., 2013). The DSIF complex (Spt4/Spt5) is also a P-TEFb target, yet we only lowly enrich Spt5 over input and Spt4 is absent. Why Fs(1)h might mediate the targeting of one pausing factor, but not the other, is unclear.

While the majority of highly enriched proteins have to do with the co-activator function of Fs(1)h, we were still able to identify PRC1 components and insulator proteins in the Fs(1)h pulldowns. Thus, by reciprocal copurification our data support that Fs(1)h interacts with the subunits of PRC1 (Table 3-2). The proteomic data suggests the PcG-Fs(1)h interaction is highly specific for the PRC1 complex, since the PRC2 complex components are absent except for the Jarid2 subunit. Since Jarid2 is a substoichiometric subunit of PRC2, it is possible that a Jarid2-Fs(1)h interaction occurs outside of the context of the intact PRC2 complex. Whether this has functional relevance in vivo is being explored further in the lab.

Several insulator proteins are found in the Fs(1)h lists, but do not have a bias toward the C-terminal (long form) list (Table 3-3). The identified insulators in order of enrichment are Mod(mdg4), BEAF-32, CP190, and Su(Hw). CP190 and Su(Hw) were recovered at levels below input and furthermore, CTCF and GAF were not identified. Mod(mdg4) is the strongest candidate for functional interaction given its highest recovery of this set of insulators. Mod(mdg4) was previously shown to co-immunoprecipitate the long form of Fs(1)h (Kellner et al., 2013). However, in the same experiment, GAF, Su(Hw), and CP190 also displayed this property, and our proteomic data do not indicate that these interactions are particularly strong. Taken together with the above ChIP-seq visual analysis, there may indeed be some weak evidence for an Fs(1)h long form-specific interaction with insulators, but again likely does not constitute a primary role for this protein.

Constructs do no provide isoform specific pulldown

Table 3-2: Fs(1)h BioTAP-XL experiments specifically recover PRC1

Total peptides recovered in the N-BioTAP-Fs(1)h, Fs(1)h-C-BioTAP, or S2 input samples are listed for components of PRC1 and PRC2. Validating the recovery of Fs(1)h in the PRC1 BioTAP-XL experiments, PRC1 is observed to be recovered in both Fs(1)h experiments. This interaction appears to be highly specific to PRC1, since PRC2 subunits are largely not recovered.

Table 3-2 (Continued)

PRC1 Complex	BioTAP-N-Fs(1)h	Fs(1)-C-BioTAP	Input (S2)
Psc	12	13	0
Su(z)2	18	4	0
Sce (dRING)	7	4	0
ph-p	5	1	1
Pc	9	2	0
Scm	5	2	0
ph-d	2	1	0
PRC2 Complex			
Su(z)12	0	0	0
esc	0	1	0
CG4236 (Nurf55)	0	0	4
Pcl	0	0	0
E(z)	0	0	0
Jarid2	8	5	0
escl	0	0	0

Table 3-3: Low recovery of insulator proteins by Fs(1)h BioTAP-XL experiments

Total peptides recovered in the N-BioTAP-Fs(1)h, Fs(1)h-C-BioTAP, or S2 input samples are listed for several of the insulator proteins. With the exception of Mod(mdg4) and BEAF-32, insulator proteins are not recovered or observed at counts lower than in input samples. Any insulator-related interaction is a minor function of Fs(1)h in comparison to its role in transcriptional activation (compare counts to Table 3-1, Table S1)

Table 3-3 (Continued)

	N-total	C-total	input
CTCF	0	0	0
GAF	0	0	0
BEAF-32	3	3	1
Su(Hw)	1	0	4
Mod(mdg4)	10	10	2
CP190	9	6	12

Despite the Western blot suggesting that the N-terminal tagged construct solely expressed the short form of Fs(1)h, the proteomic data strongly suggests this is not the case. When we looked at the peptide coverage over the length of the Fs(1)h protein, it became clear that the N-terminal construct purified many peptides that would only be found in the long form of the protein (Figure 3-6). However, the C-terminal tag (long-form only) had relatively more representation of this region than the short form. One possibility is that the N-terminal construct is competent to express tagged long-form at a low level that was not clear on the Western blot. Alternatively, since Brd4 is known to dimerize, perhaps the lower levels of long-form specific peptide observed in the N-terminal construct experiment represent those peptides identified through tagged-short form Fs(1)h to native long form Fs(1)h interactions.

The reason for the seemingly isoform-specific Western blot (Figure 3-2C) remains a mystery, but when I look at the domain architecture of the FlyBase curated transcripts, I would offer the following speculation. Of the 8 annotated transcripts, 6 code for the short form, while 2 code for the long form (Figure 3-2B). Of the 6 short form transcripts, many share the same 3' UTR, but have variable 5' UTRs. In fact, for all short form transcripts, the 5' UTR begins much further upstream than the 5' UTR of the two long form transcripts. By adding in the BioTAP tag at the 5' end, perhaps we've mimicked a more distal transcriptional start site, which might bias the synthesis of short-form transcripts over long-form transcripts.

Confirmation of the Fs(1)h-PRC1 interaction using recombinant protein expression

While we did specifically affinity purify PRC1 components with Fs(1)h (as compared to PRC2 components), the low level provided impetus for us to confirm the interaction using an orthogonal technique. To do so, I turned as before to recombinant protein expression in Sf9 cells using the Baculovirus system. I obtained pFastBac constructs for the four PRC1 components or Baculoviral aliquots for the core four PRC2 components, all generously provided by the Kingston lab (MGH). In

Figure 3-6: Fs(1)h long form is relatively enriched in C-terminal pulldown, but not absent from N-terminal pulldown

Histogram for the peptide counts (y-axis) obtained per unique peptide along the length of the Fs(1)h protein (x-axis). Black vertical bar indicates end of Fs(1)h short form (peptides from the right can be recovered from Fs(1)h long form proteins only). While long form-specific peptides are enriched in the C-terminal BioTAP-XL experiment relative to in the N-terminal BioTAP-XL experiment, they are not absent from the N-terminal experiment. This suggests that the construct may indeed direct the synthesis of N-terminally tagged long form protein, or that through dimerization native long form protein is recovered.

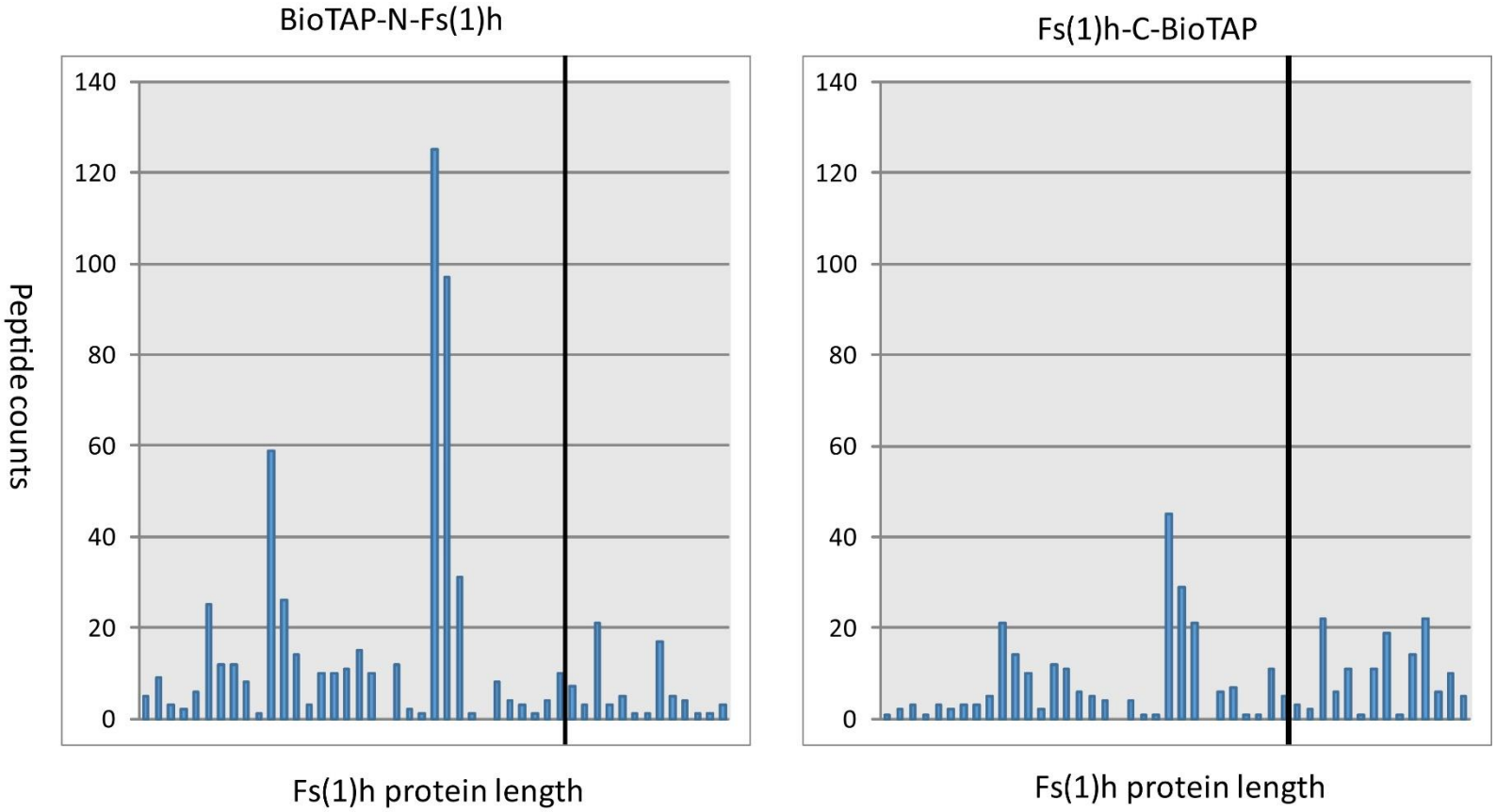


Figure 3-6 (Continued)

order to express Fs(1)h, my collaborator, HJ Kang cloned the cDNA of Fs(1)h-S into the pFastBac backbone. I then made Baculoviruses for PRC1 and Fs(1)h, and amplified all viruses in Sf9 cells.

I then co-infected Sf9 cells with a variety of viruses to express the following combination of proteins: Fs(1)h alone (negative control), Fs(1)h and PRC1, Fs(1)h and PRC2 (non-specific control). The PSC and ESC subunits of PRC1 and PRC2, respectively, were Flag tagged in these experiments to facilitate purification. I made nuclear extracts from the infected cells, and incubated extracts with anti-Flag resin to immunoprecipitate the PcG complexes. I washed the immunocomplexes with a stringent salt series (2M KCl max) and eluted bound complexes using Flag peptide (Figure 3-7). I then assessed protein contents of the elutions by silver staining of SDS-PAGE gels and by Western blotting. If Fs(1)h interacted with the PcG complex, it too would be observed in the eluate. I observed that in Flag elutions from Fs(1)h-only pulldowns, there was indeed some background binding of the Fs(1)h protein evident by a faint ~150 kD band (Figure 3-8A). This band was much more intense in the PRC1+Fs(1)h elutions. By silver stain it was also possible to identify the four recombinant PRC1 proteins (Flag-tagged PSC FPSC], Ph, Pc, and dRING), validating that the complex is assembled. Similarly, in the PRC2+Fs(1)h elutions, the assembly of PRC2 complex (Suz12, E(z), NURF, and Flag-tagged ESC [F-ESC]) was strongly evident (it should be noted that FESC and NURF bands comigrate). In the PRC2+Fs(1)h elutions, the putative Fs(1)h band was not strongly present, similar to the Fs(1)h-only level. There is, however, a more intense band in the similar region, but I contend it is clearly a different band.

To confirm the putative Fs(1)h band is in the Fs(1)h-PRC1 pulldown only, we probed a Western blot with anti-Fs(1)h antibody. Whereas Fs(1)h was expressed at similar levels in the PRC1 and PRC2 coexpression extracts (and more highly expressed in the Fs(1)h alone extract), there was strong signal in the PRC1+Fs(1)h elution only (Figure 3-8B). To verify the efficiency of Flag-peptide-mediated elution, we also boiled the post-elution bead resin. We observed no additional Fs(1)h bound to the resin in the Fs(1)h or PRC2+Fs(1)h, but a weak signal in the PRC1+Fs(1)h lane, suggesting incomplete elution from

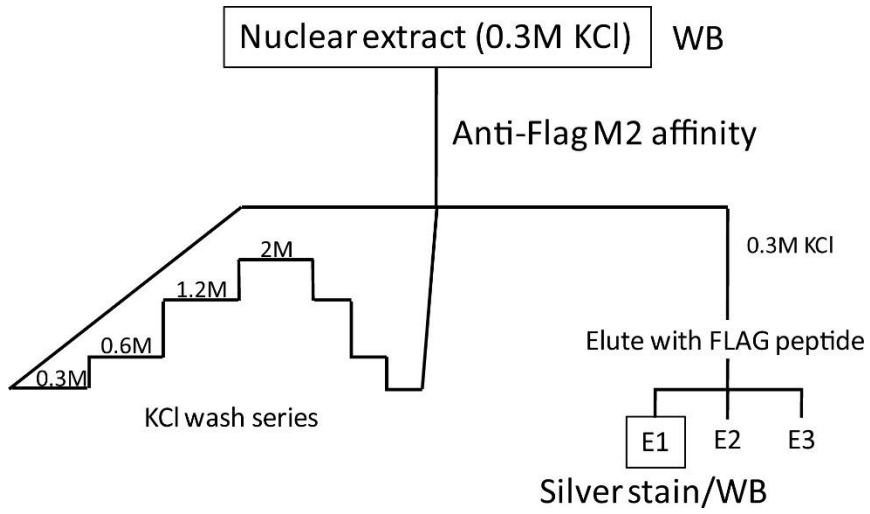
Figure 3-7: Schematic for testing for an interaction between Fs(1)h and PcG proteins

A) Purification scheme. Nuclear extracts were prepared from Sf9 cells co-infected by the appropriate Baculoviruses. Proteins were immunoprecipitated using anti-Flag resin. The bound complexes were washed with a series of buffers with increasing KCl salt. Bound complexes were eluted in 300mM KCl with 0.4mg/mL Flag peptide. Input and elution fractions were analyzed by silver stain and Western blot (WB), as indicated.

B) To test for the interaction between Fs(1)h and PRC1, infected nuclear extracts were prepared for Fs(1)h-only, Fs(1)h coexpressed with PRC1, and Fs(1)h coexpressed with PRC2. If the Fs(1)h-PRC1 interaction is specific, Fs(1)h should only be recovered in the Flag immunoprecipitation when co-expressed with PRC1.

Figure 3-7 (Continued)

A



B

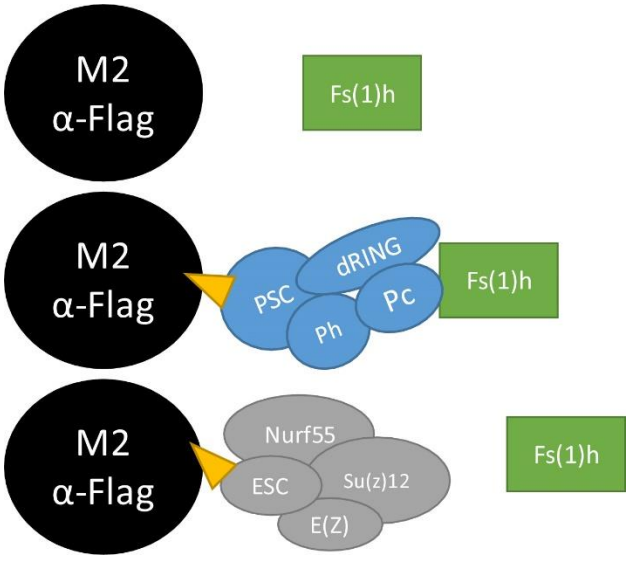
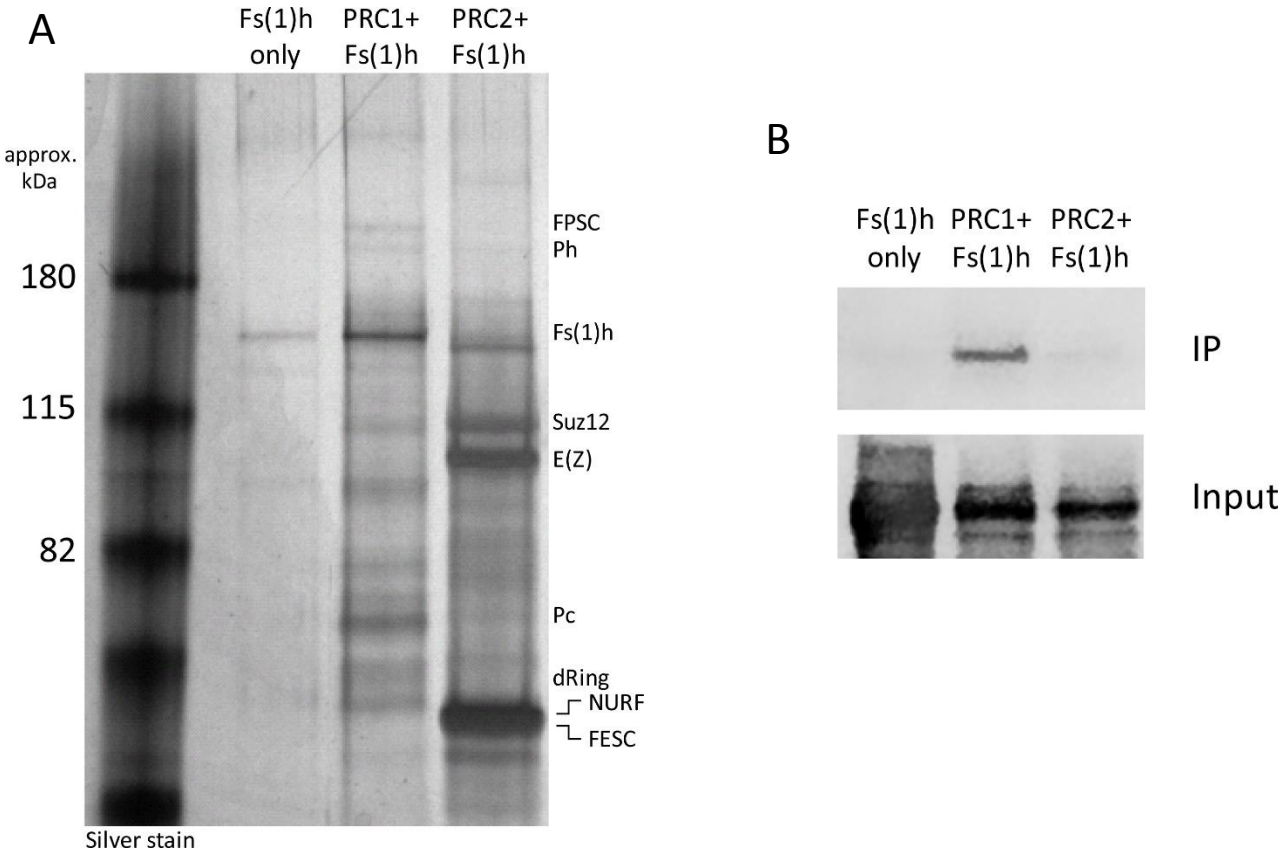
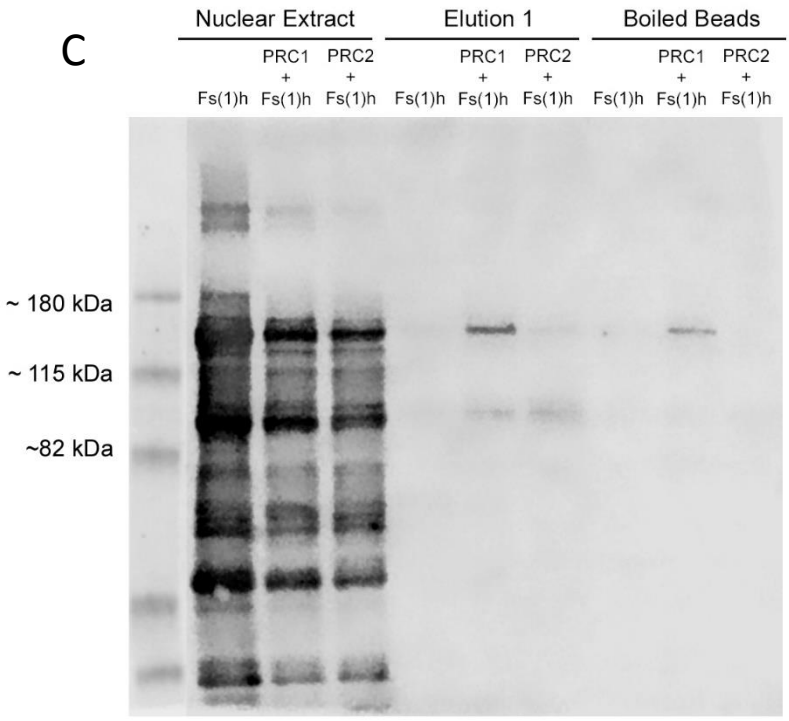


Figure 3-8: Recombinant Fs(1)h does specifically interact with recombinant PRC1 in Sf9 nuclear extracts

A) Assembly of PRC1 and PRC2 were confirmed by silver staining. All four members of each core complex are recovered in the appropriate pulldowns. The putative Fs(1)h band is observed weakly when expressed alone or with PRC2, but is strongly recovered in the PRC1 IP.
B) Anti-Fs(1)h Western blots from input and IP samples show that Fs(1)h is co-immunoprecipitated specifically by PRC1.
C) Full Western blot image from B, along with boiled bead samples to assess extent of elution by Flag peptide. While Flag elution is incomplete, there is no background binding of un-eluted Fs(1)h observed on the beads in the Fs(1)h only or PRC2+Fs(1)h samples.

Figure 3-8 (Continued)





the resin. I conclude from this experiment that the Fs(1)h interaction with PRC1 is both real and highly specific to PRC1 amongst the PcG complexes. These results will be incorporated for publication into a future manuscript co-authored by HJ Kang and L Jung.

Scm is a shared subunit between PRC1 and PRC2

One conundrum surrounding Sex-comb on midleg (Scm) is that it has a strong homeotic phenotype, much akin to mutations of core members of PRC1 and PRC2, yet, Scm was known only to be a substoichiometric member of PRC1. When HJ Kang identified Scm, but not other PRC1 subunits, in the E(z) (PRC2) BioTAP-XL proteomics experiments, it immediately lent itself to follow-up. Perhaps Scm might be a functional bridge between PRC1 and PRC2 to coordinate these complexes on chromatin and maintain silent domains. HJ Kang reciprocally BioTAP-tagged Scm, and observed that both PRC1 and PRC2 subunits were recovered in the BioTAP-XL proteomics, with no apparent preference for either complex. HJ Kang also identified numerous other repressive complexes (NCoR, LINT). HJ Kang proposed a model in which Scm is highly important during the course of development for the coordination of not only PRC1 and PRC2, but also other repressive complexes, to erase active marks and establish silenced domains.

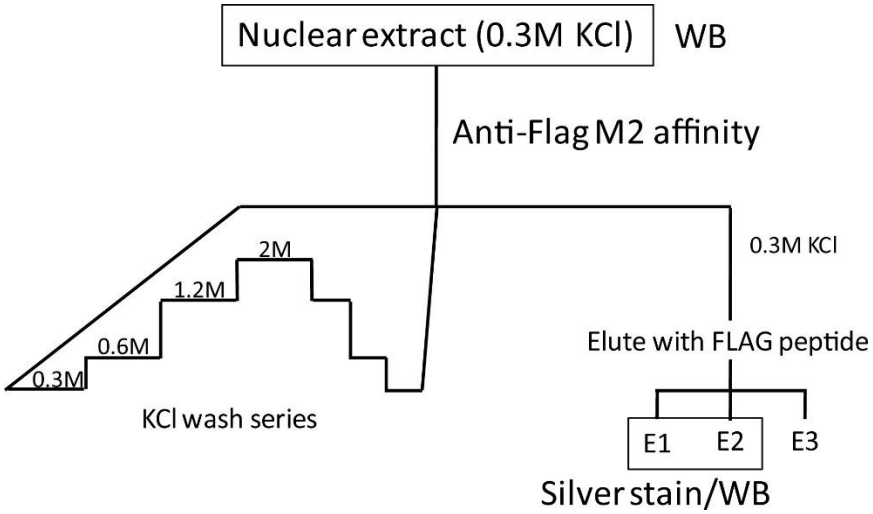
In order to test this model, I sought to show that Scm could physically interact with PRC2 using recombinant proteins. Scm had not previously been identified in biochemical purifications of PRC2, so it was unknown whether this interaction would even occur using recombinant proteins outside of the chromatin context. Joon cloned Scm into the pFastBac vector and I made Scm-expressing Baculovirus. I employed the PRC2 viral stocks as above, and coexpressed Scm with PRC2 or expressed Scm alone (Figure 3-9). I once again confirmed the assembly of PRC2 by silver staining (Figure 3-10A). Unfortunately, the similar molecular weights of Scm and Su(z)12 made it difficult to unambiguously identify each band on the silver stain, though I would argue that the differences in silver stain intensity

Figure 3-9: Schematic for testing for an interaction between Scm and PRC2

A) Purification scheme. Nuclear extracts were prepared from Sf9 cells co-infected by the appropriate Baculoviruses. Proteins were immunoprecipitated using anti-Flag resin. The bound complexes were washed with a series of buffers with increasing KCl salt. Bound complexes were eluted in 300mM KCl with 0.4mg/mL Flag peptide. Input and elution fractions were analyzed by silver stain and Western blot (WB), as indicated.
B) To test for the interaction between Scm and PRC2, infected nuclear extracts were prepared for Scm-only and Scm with PRC2. If Scm interacts with PRC2, it would be recovered when co-expressed.

Figure 3-9 (Continued)

A



B

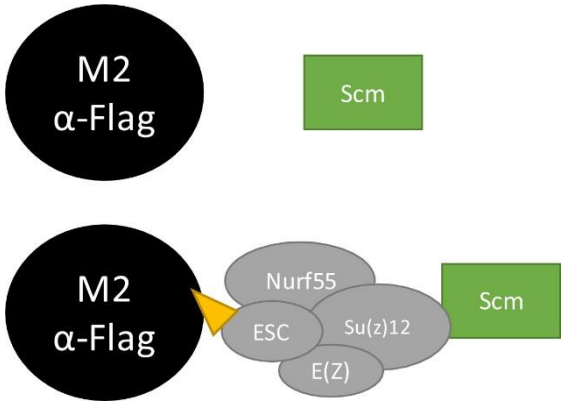


Figure 3-10: Recombinant Scm does interact with recombinant PRC2 in Sf9 nuclear extracts

A) Assembly of PRC2 was monitored by silver staining. Bands corresponding to the four core members of PRC2 were observed (NURF and F-esc comigrate). Unfortunately, the similar molecular weights of Su(z)12 and Scm made it difficult to conclude whether Scm was recovered.
B) Western blotting for Scm shows that Scm is strongly co-purified by PRC2.

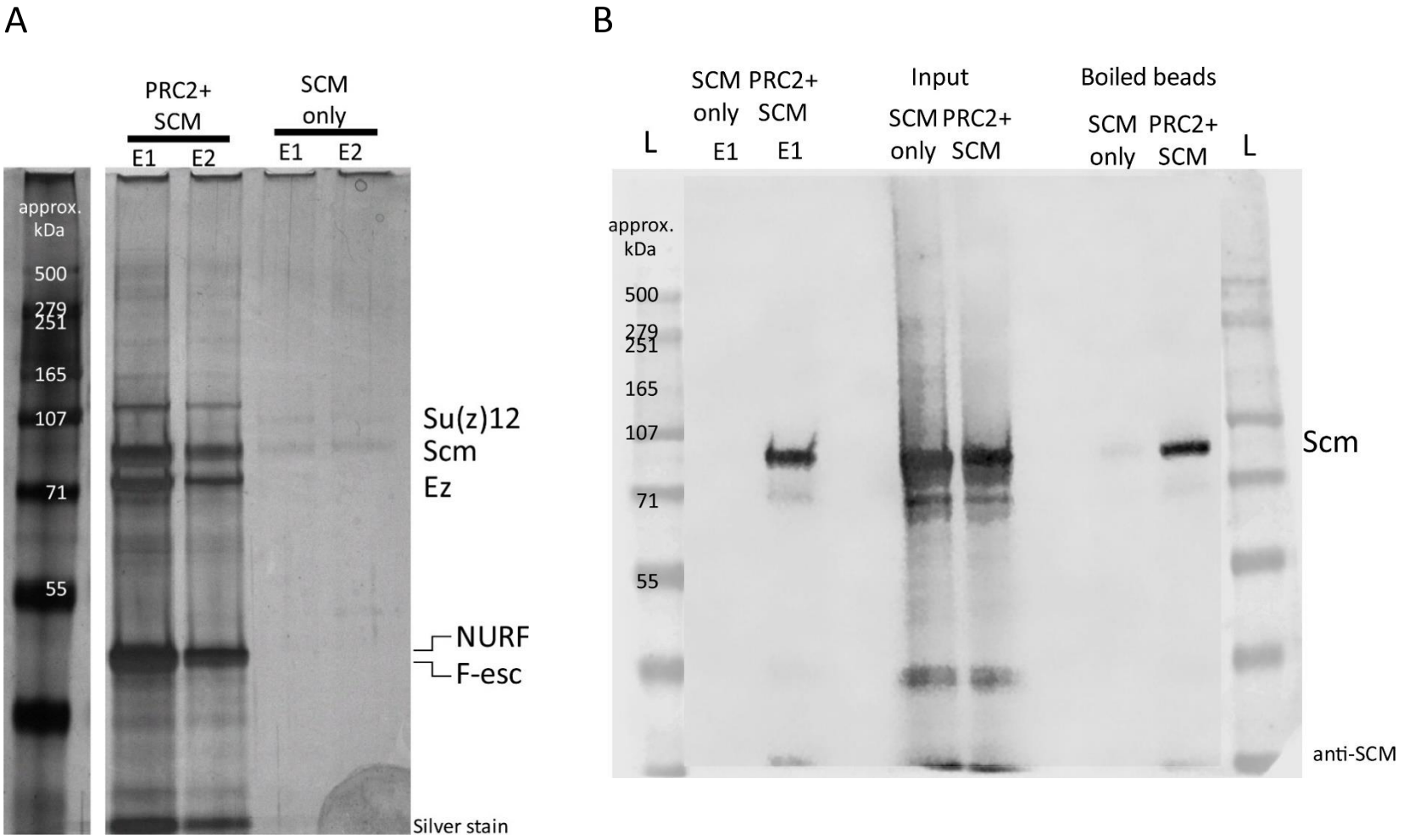


Figure 3-10 (Continued)

between the upper and lower portions of the Scm/Su(z)12 band suggest that at least two proteins are present. For unambiguous identification of the Scm protein, we probed a Western blot using anti-Scm (Figure 3-10B). It was clear from the Western blot that Scm did in fact interact with PRC2. I also boiled the post-Flag eluted beads, and as was the case for Fs(1)h, observed incomplete elution from the beads, in which Scm was only recovered where PRC2 was present.

For the Fs(1)h recombinant protein interaction experiment, PRC2 functioned as a non-specific control. That is, not observing Fs(1)h in the PRC2+Fs(1)h elution supported the specificity of the Fs(1)h-PRC1 interaction, as compared to the argument that Fs(1)h would interact with any protein already stuck to the Flag resin. To perform a similar experiment, I co-expressed Scm with Flag-tagged MCM3 (F-MCM3), another protein I was working with (see Appendix 1). The Baculovirus expressing MCM3 also co-expressed MCM5, which serves as an internal positive control for F-MCM3 interaction competence. I observed that Scm was co-immunoprecipitated only at background levels by F-MCM3, similar to the level seen in Scm-only extracts by silver stain (Figure 3-11A). F-MCM3 was able to pulldown its bona fide partner MCM5, at a roughly stoichiometric level. In this same experiment I co-expressed F-ESC with Scm and observed that Scm was still affinity purified in this case. This was very clear in the silver stain, but perhaps even more striking in the anti-Scm Western blot (Figure 3-11B). This suggests that an intact PRC2 complex is not needed to interact with Scm, and that the PRC2-Scm interaction may be mediated by the Esc subunit. Further testing of each pairwise interaction between Scm and PRC2 complex subunits would be needed to conclude this is the case.

Discussion:

In addition to my focus on the MSL complex (Chapter 1), I enjoyed a productive collaboration with HJ Kang on PcG interactions. The main conclusion from my work in this collaboration is that the

Figure 3-11: The Scm-PRC2 interaction is likely mediated at least in part through ESC

A) In order to test whether Scm would non-specifically interact with any Flag-tagged protein, Scm was co-expressed with Flag-MCM3. The Flag-MCM3 virus also expresses MCM5, which should be recovered. Silver staining of elution fractions suggested that Scm was recovered at similar levels in the Scm-only and Scm+F-MCM3/MCM5 experiment, suggesting the observed interaction in Figure 10 is specific to PRC2. Furthermore, co-expression of Flag-Esc with Scm also recovered Scm protein, suggesting that the Scm-PRC2 interaction is mediated at least in part by Esc.
B) Western blotting confirmed the silver stain result and expression of Scm protein in input fractions.

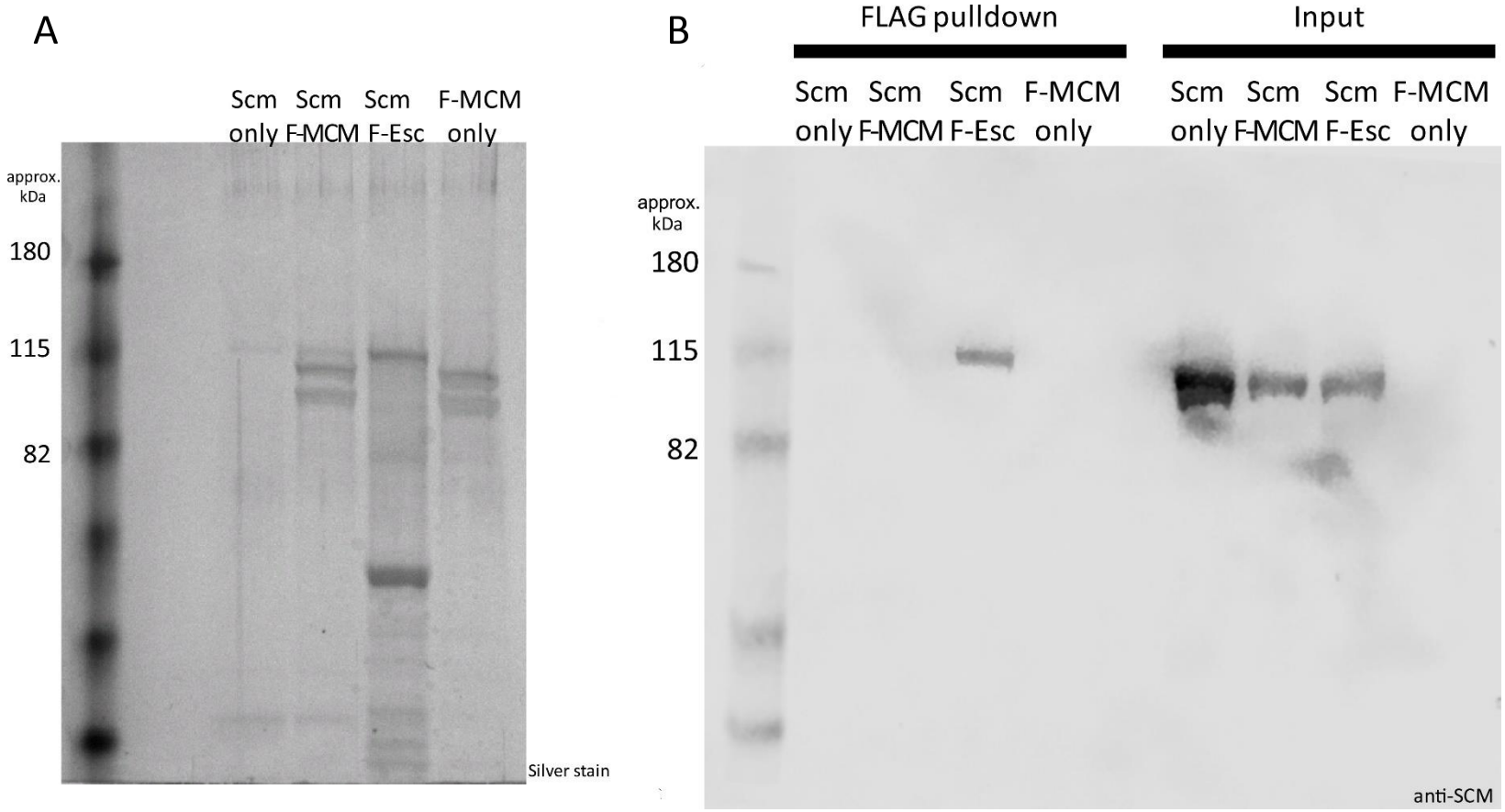


Figure 3-11 (Continued)

interactions that occur between chromatin-bound complexes are more extensive than previously appreciated. The use of the BioTAP-XL technique identified several surprising potential interactions between the two major PcG complexes and other proteins, and I was able to test some of these interactions using recombinant protein expression and co-immunoprecipitation. In additional follow-up studies, I further explored the interactome and genome localization of the TrxG protein, Fs(1)h. Based on the proteomic work, we saw that the majority of Fs(1)h protein is likely to be involved in the expected transcriptional co-activator function, whereas the Fs(1)h-PRC1 interaction is likely to be only at a select subset of loci.

The implications of the Scm-PRC2 interaction may be profound. The recombinant protein interaction experiments suggest that Scm can interact with the core of PRC2 without additional accessory factors, and I observed subsequently that Scm may in fact interact with Esc, which is an obligate cofactor for E(z) methyltransferase enzymatic function. The intimacy of the Scm interaction with the PRC2 catalytic core led us to explore whether Scm gene function in the fly was necessary for the proper patterning of H3K27me3. On polytene chromosomes, we observed redistribution of H3K27me3 in Scm-depleted conditions, suggesting that Scm may have a role in reinforcement of PcG-silenced chromatin regions (Kang et al., 2015).

In the mammalian field, especially in stem cells, genomic regions that carry both active and repressive chromatin proteins and histone PTMs have been identified (Harikumar and Meshorer, 2015, Voigt et al., 2013). These regions, often termed bivalent or poised due to the coexistence of opposing marks, have not been characterized in Drosophila. The co-localization and physical interaction between PRC1 and Fs(1)h suggest that these states may transiently exist during the course of development in Drosophila as well. An intriguing possibility that we are exploring is whether the presence of Scm, which we have shown tracks extremely well with H3K27me3 (Jung et al., 2016, Kang et al., 2015), is indicative of PcG-mediated silencing, whereas PRC1 presence alone does not specify this fate. Taken together,

these data suggest a model in which the mutual exclusivity of Scm vs Fs(1)h and other co-activators for PRC1 interaction may determine whether a region is fated for activation or repression (Figure 3-12). Along with HJ Kang and L Jung, we are delving deeper into the interactions that may help to determine these fate-choice decisions.

Materials and Methods:

Cloning

BioTAP-tagged Fs(1)h was generated using recombineering as in Chapter 2. Both N- and C-terminal constructs were generated. Baculoviral constructs for the recombinant expression of Fs(1)h and Scm were cloned into pFastBac1 using traditional restriction enzyme cloning methods and the appropriate cDNA. Constructs for PRC1 components were a kind gift from the Kingston lab (MGH).

Western Blotting

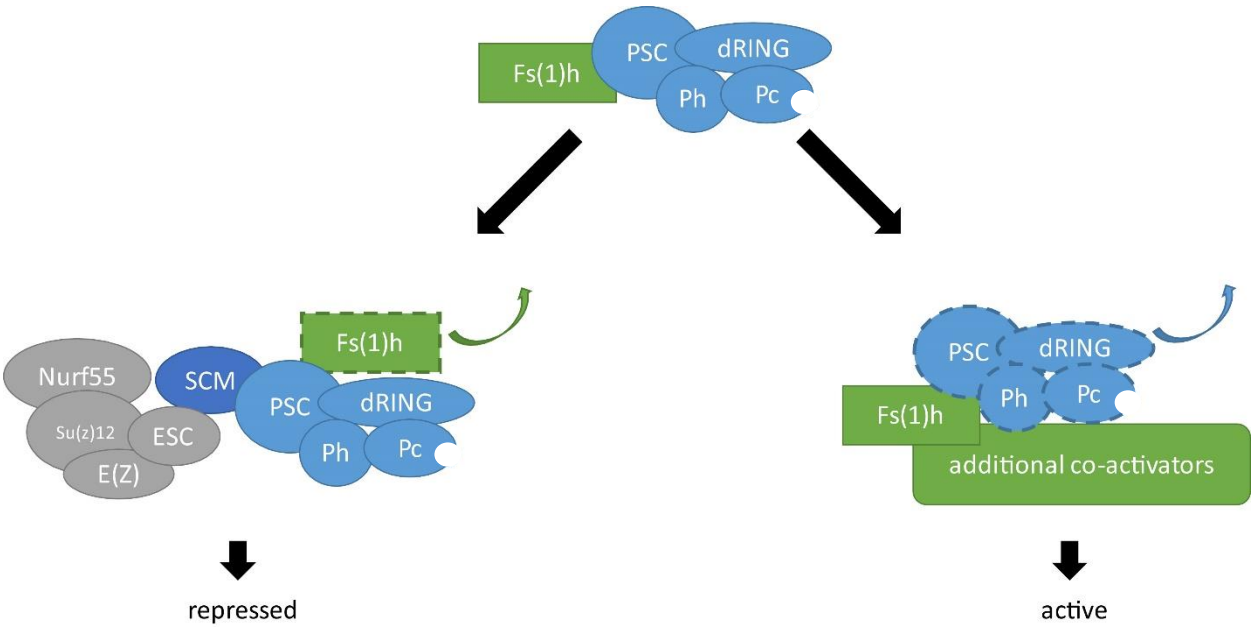
Crude nuclear extracts (as in Figure 3-2) were generated using hypotonic lysis of cells, with the supernatant as the crude cytoplasmic extract. 4-12% Bis-Tris acrylamide gradient gels were used to resolve proteins, and proteins were transferred to nitrocellulose membranes using the Invitrogen Mini Transfer kit. BioTAP-tagged Fs(1)h was detected using peroxidase-anti-peroxidase (PAP), which recognizes the proteinA moieties of the BioTAP tag, used at a dilution of 1:1000. Anti-Fs(1)h (1:1000) was a kind gift from I. Dawid. Anti-Scm (1:500) was a kind gift from J. Simon.

Immunofluorescence

Figure 3-12: Mutually exclusive interactions may dictate gene activity fate

During the course of development, association of Fs(1)h and PRC1 may poise genes for proper induction kinetics. As other interactions occur on chromatin, either Fs(1)h or PRC1 is evicted from sites allowing for either robust repression or activation of genes.

Figure 3-12 (Continued)



Immunofluorescence in S2 was performed essentially as described (Fanti et al., 2008), omitting the flash-freezing of slides. PAP (1:100) was used to detect the BioTAP tag along with anti-Rabbit fluorescent secondary antibody (1:500).

BioTAP-XL for Fs(1)h

BioTAP-XL was performed as described (Alekseyenko et al., 2015). Briefly, BioTAP-tagged Fs(1)h-expressing S2 cells were lysed by Dounce homogenization and the nuclei were crosslinked for 30min at room temperature with 3% formaldehyde. Chromatin was solubilized by sonication, and tandem affinity purification was performed, first with IgG-agarose, then with streptavidin-agarose. Samples were split for DNA and protein analyses. To isolate DNA, proteins were degraded by proteinaseK and crosslinks were reversed with SDS and a 6hr incubation at 65C. Illumina libraries were prepared and sequenced. To isolate peptides for mass spectrometry identification, samples were treated with Trypsin and desalted by C18 column purification.

ChIP-Seq Data analysis

ChIP-seq analysis was the same as in Chapter 2.

Proteomic analysis

Total peptides recovered by mass spectrometry were compared to the appropriate input sample and other BioTAP-XL experiments performed in the lab (as non-specific controls, as applicable). Fold enrichment over input was calculated by adding a pseudocount of 1 to all peptide counts. To calculate coverage of the Fs(1)h bait protein, number of peptides were counted for the available trypsin fragments.

Baculovirus and recombinant protein interaction assays

Baculoviruses expressing proteins of interest were generated using the Invitrogen Bac-to-Bac kit. Briefly, pFastBac1 constructs were transformed into DH10Bac bacteria and subjected to blue-white selection. Genomic DNA was isolated from appropriate white clones and transfected into Sf9 cells using Cellfectin II (Invitrogen). Viruses for PRC2 viruses were a kind gift from the Kingston lab (MGH). Viruses were amplified in Sf9 cells to generate high titer stocks. Co-infection of Sf9 cells with the appropriate viruses was performed for 72hrs, after which infected nuclear extracts were generated using a hypotonic lysis and salt extraction. Protein-protein interactions were assessed by co-immunoprecipitation by anti-Flag (M2) resin (Invitrogen). Assembly of PRC1/2 complexes was monitored by silver staining, and presence of Fs(1)h or Scm were assessed by silver staining, if possible, and by Western blotting.

Silver staining

Gels were fixed overnight in a solution of 40% methanol, 10% acetic acid. Two 20 minute rehydrations with 10% ethanol, 5% acetic acid were followed by a 30 minute incubation in 60uM DTT in water. Gels were soaked in 0.1% silver nitrate ($AgNO_3$) for 30 minutes, then rinsed briefly in water. Gels were developed with frequent changes of 0.0185% formaldehyde, 3% sodium carbonate (Na_2CO_3) until the desired darkness was achieved. Staining was stopped using 5% acetic acid.

References:

Alekseyenko, A.A., McElroy, K.A., Kang, H., Zee, B.M., Kharchenko, P.V., and Kuroda, M.I. 2015. BioTAP-XL: Cross-linking/Tandem Affinity Purification to Study DNA Targets, RNA, and Protein Components of Chromatin-Associated Complexes. *Curr Protoc Mol Biol* 109: 21 30 21-21 30 32.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298: 1039-1043.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. 2002. *Drosophila* Enhancer of Zeste/ESC Complexes have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. *Cell* 111: 185-196.

Devaiah, B.N., Lewis, B.A., Cherman, N., Hewitt, M.C., Albrecht, B.K., Robey, P.G., Ozato, K., Sims, R.J., 3rd, and Singer, D.S. 2012. BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. *Proc Natl Acad Sci U S A* 109: 6927-6932.

Dow, E.C., Liu, H., and Rice, A.P. 2010. T-loop phosphorylated Cdk9 localizes to nuclear speckle domains which may serve as sites of active P-TEFb function and exchange between the Brd4 and 7SK/HEXIM1 regulatory complexes. *J Cell Physiol* 224: 84-93.

Fanti, L., Perrini, B., Piacentini, L., Berloco, M., Marchetti, E., Palumbo, G., and Pimpinelli, S. 2008. The trithorax group and Pc group proteins are differentially involved in heterochromatin formation in Drosophila. *Chromosoma* 117: 25-39.

Farina, A., Hattori, M., Qin, J., Nakatani, Y., Minato, N., and Ozato, K. 2004. Bromodomain protein Brd4 binds to GTPase-activating SPA-1, modulating its activity and subcellular localization. *Mol Cell Biol* 24: 9059-9069.

Harikumar, A., and Meshorer, E. 2015. Chromatin remodeling and bivalent histone modifications in embryonic stem cells. *EMBO Rep* 16: 1609-1619.

Jung, Y.L., Kang, H., Park, P.J., and Kuroda, M.I. 2016. Correspondence of Drosophila Polycomb Group proteins with broad H3K27me3 silent domains. *Fly (Austin)*: 0.

Kang, H., McElroy, K.A., Jung, Y.L., Alekseyenko, A.A., Zee, B.M., Park, P.J., and Kuroda, M.I. 2015. Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in *Drosophila*. *Genes Dev* 29: 1136-1150.

Kellner, W.A., Van Bortle, K., Li, L., Ramos, E., Takenaka, N., and Corces, V.G. 2013. Distinct isoforms of the Drosophila Brd4 homologue are present at enhancers, promoters and insulator sites. *Nucleic Acids Res* 41: 9274-9283.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., Linder-Basso, D., Plachetka, A., Shanower, G., Tolstorukov, M.Y., Luquette, L.J., Xi, R., Jung, Y.L., Park, R.W., Bishop, E.P., Canfield, T.K., Sandstrom, R., Thurman, R.E., MacAlpine, D.M., Stamatoyannopoulos, J.A., Kellis, M., Elgin, S.C., Kuroda, M.I., Pirrotta, V., Karpen, G.H., and Park, P.J. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* 471: 480-485.

Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Muller, J. 2006. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev* 20: 1110-1122.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A., and Verrijzer, C.P. 2008. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev* 22: 2799-2810.

McElroy, K.A., Kang, H., and Kuroda, M.I. 2014. Are we there yet? Initial targeting of the Male-Specific Lethal and Polycomb group chromatin complexes in Drosophila. *Open Biol* 4: 140006.

Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. 2002. Histone Methyltransferase Activity of a *Drosophila* Polycomb Group Repressor Complex. *Cell* 111: 197-208.

Negre, N., Hennetin, J., Sun, L.V., Lavrov, S., Bellis, M., White, K.P., and Cavalli, G. 2006. Chromosomal Distribution of PcG Proteins during *Drosophila* Development. *PLoS Biol* 4.

Ng, J., Hart, C.M., Morgan, K., and Simon, J.A. 2000. A *Drosophila* ESC-E(Z) Protein Complex is Distinct from Other Polycomb Group Complexes and Contains Covalently Modified ESC. *Mol Cell Biol* 20: 3069-3078.

Peterson, A.J., Mallin, D.R., Francis, N.J., Ketel, C.S., Stamm, J., Voeller, R.K., Kingston, R.E., and Simon, J.A. 2004. Requirement for sex comb on midleg protein interactions in Drosophila polycomb group repression. *Genetics* 167: 1225-1239.

Saurin, A.J., Shae, Z., Erdjument-Bromage, H., Tempst, P., and Kingston, R.E. 2001. A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins. *Nature* 412: 655-660.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.Y., Bourgon, R., Biggin, M., and Pirrotta, V. 2006. Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nat Genet* 38: 700-705.

Schwartz, Y.B., and Pirrotta, V. 2014. Ruled by ubiquitylation: a new order for polycomb recruitment. *Cell Rep* 8: 321-325.

Simon, J.A., and Kingston, R.E. 2013. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* 49: 808-824.

Strubbe, G., Popp, C., Schmidt, A., Pauli, A., Ringrose, L., Beisel, C., and Paro, R. 2011. Polycomb purification by in vivo biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners. *Proc Natl Acad Sci U S A* 108: 5572-5577.

Tolhuis, B., de Wit, E., Muijrers, I., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. 2006. Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. *Nat Genet* 38: 694-699.

Voigt, P., Tee, W.W., and Reinberg, D. 2013. A double take on bivalent promoters. *Genes Dev* 27: 1318-1338.

Yamaguchi, Y., Shibata, H., and Handa, H. 2013. Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochim Biophys Acta* 1829: 98-104.

Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. 2005. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* 19: 535-545.

Chapter 4
Summary and Perspectives

Advantages and obstacles to successful analysis of regulatory factors on chromatin

My studies presented in this dissertation suggest that interactions on chromatin may include crosstalk between disparate regulatory complexes which had not previously been recognized. Working with chromatin is a challenge due to the insolubility of factors when bound to DNA. Experimenters must always consider the balance between using soluble complexes removed from chromatin versus introducing crosslinks to stabilize the on-chromatin interactions. Both techniques are subject to their own subset of artifacts and biases. In this work I employed both approaches in an attempt to generate a more complete picture of the interactions as they are found in the cell.

The BioTAP-XL technique is very powerful in identifying new, functionally relevant interactors which may be lost when complexes are solubilized from chromatin. Time and again, our lab has shown that proteins partners identified by BioTAP-XL provide insight for understanding aspects of chromatin biology (Alekseyenko et al., 2014a, Alekseyenko et al., 2014b, Alekseyenko et al., 2015, Kang et al., 2015, Wang et al., 2013). The use of BioTAP-XL as a discovery tool in conjunction with traditional biochemical experiments can be a useful approach for making new discoveries.

However, a shortcoming of the BioTAP-XL technique that we have recognized through my work is that not all proteins are amenable to the technique. The UpSET protein is large (~330kDa), not highly expressed, and has been prone to degradation in chromatin preparations. While UpSET may be an especially difficult case (Rincon-Arano, personal communication), other large proteins in the lab have encountered similar problems due to lack of stability through the BioTAP-XL procedure. There are several approaches that may ameliorate this problem that future experiments could attempt.

Although we prefer to express native levels of tagged factors whenever possible, one option could be to change the expression level of the bait protein. Creating constructs using cDNA and a strong promoter such as actin5C or similar could boost the incorporation of the protein into chromatin. Removal of any potential competition from untagged bait protein is another option which I employed in

the UpSET ChIP-seq experiments. In the absence of wild-type UpSET, I was able to perform one-step ChIP-seq in transgenic UpSET-BioTAP embryos, however, in that case the full BioTAP-XL procedure still failed. I did not determine whether the failure of the two step procedure might be explained by a poor biotinylation efficiency of the tagged UpSET protein. If so, modifying the BioTAP tag to employ a sequence recognized by and requiring an exogenous biotin ligase (such as the BirA system (Beckett et al., 1999, Chapman-Smith and Cronan, 1999)) in conjunction with the actin-driven tagged cDNA could ensure greater chromatin incorporation and biotin ligation efficiency. The drawback of changing the tag is the requirement of the two-component system, which we have consciously avoided in the standard BioTAP tag in order to decrease the number of manipulations in the system.

Another option would be to alter the way we solubilize the chromatin. Our current approach employs sonication, which is a vigorous process and may contribute to protein degradation. Several alternative methods to solubilize chromatin exist, though typically it is impossible to get sufficient yields of chromatin without any sonication. One method would be to use micrococcal nuclease (MNase), which cleaves linker DNA between nucleosomes. MNase is typically used to ascertain the locations of nucleosomes and alone is insufficient to completely solubilize chromatin, however, and much of the starting material would remain insoluble due to the inability of the enzyme to access compact chromatin regions (Wal and Pugh, 2012). Often MNase is paired with sonication, albeit for shorter durations than employed for BioTAP-XL or ChIP techniques. Deoxyribonuclease I (DNaseI) is another enzyme typically used to release DNA from open chromatin regions (Shibata and Crawford, 2009). Much like MNase, DNaseI cannot penetrate densely packed chromatin structures. Alternatives to DNaseI, such as Benzonase and Cyansase (Grontved et al., 2012), also exist, which have slightly different sequence biases in their cleavage profiles. Perhaps a combination of these enzymes used together or in sequence could greatly reduce the amount of sonication required for chromatin solubilization. Unfortunately,

efficient chromatin solubilization without any duration of sonication does not seem, at this time, to be possible.

Future improvements of CRISPR/Cas-9 mutagenesis for the generation of mutant S2 cell lines

In this work I also utilized genome engineering techniques, which over the last few years have exploded in their application across model organisms and beyond (Cong et al., 2013, Housden et al., 2014, Mali et al., 2013). I found the CRISPR/Cas-9 system to be extremely efficient in introducing insertion-deletion (indel) mutations in S2 cells. One concern with any genome engineering technique is whether there are off-target effects from the mutagenic agent. Studies of CRISPR/Cas-9 have shown that there is potential for off-target mutagenesis, but the cleavage rates at alternative sites are typically low (Fu et al., 2013, Hsu et al., 2013). Recently, the Cas-9 enzyme and/or the guide-RNA (gRNA) have been further engineered to diminish the potential for off-target effects (Fu et al., 2014, Ran et al., 2013). Use of these re-engineered enzymes/guides in Drosophila in the future could greatly increase the confidence in on-target mutations being the only changes likely to occur.

The lethal phenotype that I observed after CRISPR/Cas9 mutagenesis of *upSET* in the fly was not due to off-target effects, as I could rescue the lethality with a wild type tagged *upSET* transgene. A rescue experiment is more difficult in S2 cells in culture, where transgene copy number is not easily modulated. An alternative method to address concerns about off-targets is to generate independent mutations using distinct gRNA sequences. If frameshift mutations at different locations in the target gene produce the same phenotype, confidence would be greatly enhanced. In my studies of *upSET* in cell culture, I employed several different guides to try to address any concerns about off-target effects, and saw the same general trends for upregulation of heterochromatin genes.

That the general trend in gene regulation and bulk histone levels were consistent across *upSET* mutant cell lines gives me confidence that these phenotypes are caused by the loss of UpSET. However,

a preliminary finding from follow-up ChIP experiments suggests that it is possible that the selective pressure to grow from a single cell coupled to the long time frame that is necessary to obtain S2 cell colonies may result in a variety of stochastic gene expression and chromatin environment alterations in the cell. Whether the phenotypes described in Chapter 2 are due to these pressures rather than the loss of UpSET is, formally, a concern. An additional concern is the duration for which these *upSET* mutant cells have been cultured subsequent to their isolation, and in between genomic experiments. It is possible that this duration actually causes selection within the mutant population for or against a subpopulation of cells that is further obfuscating our analyses. Therefore, prior to publication, I am working to obtain sufficient amounts of tissue from a condition where this chronic loss of UpSET will not have occurred for as many cellular generations and therefore decrease any selection for/against a specific subset of cells. To do so, I am hoping to collect a sufficient number of *upSET*-deleted homozygous larvae (the progeny of heterozygous parents) in order to assess the genomic localization of H4K16ac and H3K9me2 and/or assess the state of transcription at the nascent-RNA level from imaginal disc or other larval tissues.

Drosophila S2 cells do not readily propagate clonally from a single cell, which has historically presented a problem for the field, but recently the use of conditioned media has made recovery possible. The typical efficiency of colonial growth from even wild-type single S2 cells is 15-20%, which is poor compared to the efficiency observed in many mammalian cell lines (Cieciura et al., 1956, Puck et al., 1956). Presumably there could be a number of gene expression programs that would allow for continued growth from a single cell, which could introduce heterogeneity between replicates even of the same genotype. Whether there is heterogeneity in the expression patterns of single-clone S2 cell lines has not been commented on in the literature, given the recent introduction of the technique.

One way to test whether the heterogeneity in gene expression that I observed is intrinsic to the single cell cloning process would be to isolate cells that have been through the engineering and cloning

process without becoming mutant for UpSET (or other target). In hindsight, these might be better wild type control cells than untreated and non-clonal S2 parental cells. These could be somewhat cumbersome and expensive experiments, however, depending on the number of samples needed to deeply assess heterogeneity. If cells lacking mutations in the target gene also lack the mutant phenotype, though, it is likely that the phenotype is due to loss of the target gene as opposed to any heterogeneity introduced by the single cell cloning procedure.

While the transcriptional profiles at the individual gene level showed heterogeneity in my *upSET* mutant S2 cell lines, one gene in particular was upregulated across all lines, though to varying degrees. The locus is *roX1*, which in wild type S2 cells is lowly expressed for unknown reasons. I have verified the low level in wild-type and the high level of induction of the locus in the *upSET* mutant cell lines both by qPCR and in the nascent-RNA experiment. This locus is particularly interesting because *roX1* is a noncoding RNA component of the MSL complex, and our interest in UpSET was initially based on its recovery in MSL3 BioTAP-XL pulldowns. The re-expression of *roX1* in the *upSET* mutant cells proceeds by an unknown mechanism, but is accompanied by the loss of H3K9me2 ChIP-seq reads over the locus, although this needs to be tested in additional cell lines. If *roX1* becomes expressed, presumably it would be incorporated into intact MSL complexes.

Interestingly, the upregulation of heterochromatin genes could be explained by the re-expression of *roX1*. Recently, several subunits of the MSL complex were proposed to serve a role in activation of genes located in heterochromatin (Koya and Meller, 2015), though whether the mechanism for this is direct is unclear. The noncoding RNA *roX1* had a particularly potent role in this process. Therefore, the activation of *roX1* may render the MSL complex once again competent to boost the transcription of heterochromatin genes, explaining our observations in the *upSET* mutant cells.

This is not the only connection reported between the dosage compensation complex and heterochromatin. Genetic lesions inactivating both *roX* RNAs leads to mislocalization of the MSL

complex to the chromocenter (heterochromatin) and the repeat-rich 4th chromosome (Deng and Meller, 2006). Some have suggested this reflects the ancestral binding capabilities of the MSL complex and its preference for a degenerate GA-repeat motif (Figueiredo et al., 2014). Furthermore, there have been some hints through proteomics for incipient interactions between MLE and HP1-depositing proteins (Cugusi et al., 2015). One could then speculate that the gene expression changes we observe in the *upSET* mutant cells are through an MSL-dependent mechanism. Depleting members of the MSL complex from the *upSET* mutant cells and analyzing the transcriptome could be used to test whether this is the case. Alternatively, due to the concerns about mutant cell culture discussed above, I may test for genetic interactions in the fly using the *upSET* deleted allele and various mutants of interest (*roX*, *Su(var)205,* etc).

Summary

In this dissertation I have presented my findings that the loss of the protein UpSET, which was identified by our lab in previous experiments as a highly ranked MSL complex interactor, has a weak effect on X chromosome gene expression, but a more potent effect on heterochromatic genes. These gene expression changes were accompanied by global changes to histone post-translational modifications, observed by ChIP-seq and by bulk histone analysis. Unfortunately, BioTAP-XL using UpSET as the bait protein failed for unknown reasons, though modifications to the method may be used in future work to successfully identify UpSET-interacting proteins. I have also presented some of the data generated over the course of a successful collaboration with Hyuckjoon Kang on the nature of Polycomb Group interactions on chromatin. Future genetic experiments on PcG proteins in S2 cells using improved genome engineering techniques informed by my work with *upSET* mutant S2 cell lines

may provide us with a system to quickly test our hypotheses with a plurality of genetic lesions unavailable in the whole fly.

References:

Alekseyenko, A.A., Gorchakov, A.A., Kharchenko, P.V., and Kuroda, M.I. 2014a. Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc Natl Acad Sci U S A* 111: 2488-2493.

Alekseyenko, A.A., Gorchakov, A.A., Zee, B.M., Fuchs, S.M., Kharchenko, P.V., and Kuroda, M.I. 2014b. Heterochromatin-associated interactions of Drosophila HP1a with dADD1, HIPP1, and repetitive RNAs. *Genes Dev* 28: 1445-1460.

Alekseyenko, A.A., Walsh, E.M., Wang, X., Grayson, A.R., Hsi, P.T., Kharchenko, P., Kuroda, M.I., and French, C.A. 2015. The oncogenic BRD4-NUT chromatin regulator drives aberrant transcription within large topological domains. *Genes Dev* 29: 1507-1523.

Beckett, D., Kovaleva, E., and Schatz, P.J. 1999. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* 8: 921-929.

Chapman-Smith, A., and Cronan, J.E., Jr. 1999. In vivo enzymatic protein biotinylation. *Biomol Eng* 16: 119-125.

Cieciura, S.J., Marcus, P.I., and Puck, T.T. 1956. Clonal growth in vitro of epithelial cells from normal human tissues. *J Exp Med* 104: 615-628.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819-823.

Cugusi, S., Kallappagoudar, S., Ling, H., and Lucchesi, J.C. 2015. The Drosophila Helicase Maleless (MLE) is Implicated in Functions Distinct From its Role in Dosage Compensation. *Mol Cell Proteomics* 14: 1478-1488.

Deng, X., and Meller, V.H. 2006. roX RNAs are required for increased expression of X-linked genes in Drosophila melanogaster males. *Genetics* 174: 1859-1866.

Figueiredo, M.L., Kim, M., Philip, P., Allgardsson, A., Stenberg, P., and Larsson, J. 2014. Non-coding roX RNAs prevent the binding of the MSL-complex to heterochromatic regions. *PLoS Genet* 10: e1004865.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 31: 822-826.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 32: 279-284.

Grontved, L., Bandle, R., John, S., Baek, S., Chung, H.-J., Liu, Y., Aguilera, G., Oberholtzer, C., Hager, G.L., and Levens, D. 2012. Rapid genome-scale mapping of chromatin accessibility in tissue. *Epigenetics Chromatin* 5.

Housden, B.E., Lin, S., and Perrimon, N. 2014. Cas9-based genome editing in Drosophila. *Methods Enzymol* 546: 415-439.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., and Zhang, F. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31: 827-832.

Kang, H., McElroy, K.A., Jung, Y.L., Alekseyenko, A.A., Zee, B.M., Park, P.J., and Kuroda, M.I. 2015. Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in *Drosophila*. *Genes Dev* 29: 1136-1150.

Koya, S.K., and Meller, V.H. 2015. Modulation of Heterochromatin by Male Specific Lethal Proteins and roX RNA in Drosophila melanogaster Males. *PLoS One* 10: e0140259.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. 2013. RNA-guided human genome engineering via Cas9. *Science* 339: 823-826.

Puck, T.T., Marcus, P.I., and Cieciura, S.J. 1956. Clonal growth of mammalian cells in vitro; growth characteristics of colonies from single HeLa cells with and without a feeder layer. *J Exp Med* 103: 273-283.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154: 1380-1389.

Shibata, Y., and Crawford, G.E. 2009. Mapping regulatory elements by DNaseI hypersensitivity chip (DNase-Chip). *Methods Mol Biol* 556: 177-190.

Wal, M., and Pugh, B.F. 2012. Genome-wide mapping of nucleosome positions in yeast using high-resolution MNase ChIP-Seq. *Methods Enzymol* 513: 233-250.

Wang, C.I., Alekseyenko, A.A., LeRoy, G., Elia, A.E., Gorchakov, A.A., Britton, L.M., Elledge, S.J., Kharchenko, P.V., Garcia, B.A., and Kuroda, M.I. 2013. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. *Nat Struct Mol Biol* 20: 202-209.

Appendix 1
Follow-up Studies on the kinase Jil-1

Contributions to this chapter: Annette Plachetka cloned the Jil-1 BioTAP-tagged construct. Art Alekseyenko oversaw the Jil-1 BioTAP-XL experiments. LC-MS/MS was performed at the Taplin Mass Spectrometry Facility at Harvard Medical School. Lucy Jung provided bioinformatics support. Kyle McElroy performed all other experiments.

Abstract:

The kinase Jil-1 has been extensively studied since its discovery in 1996. Despite this intense research focus, no consensus has been reached as to the critical function of Jil-1 in chromatin. The Jil-1 protein has several interesting properties that effect the morphology and stability of the genome as observed in polytene chromosomes and may play a general role in transcriptional activation. *Jil-1* mutants, however, display a pleiotropic phenotype and are homozygous lethal, confounding investigations into its function. Using a BioTAP-tagged *Jil-1* genomic clone, we expressed BioTAP-Jil-1 in male S2 and female Kc cells. We used this allele to identify the sex-specific localization of Jil-1 throughout the genome and to characterize any sex-specific protein interactors. The BioTAP-XL sequencing localization generally agreed with the ChIP-chip data generated previously. We identified few sex-specific interacting proteins save the MSL complex. We followed up on the interaction between Jil-1 and the Mini-chromosome maintenance (MCM) complex, which functions as the replicative helicase, to see whether Jil-1 might influence the timing of replication via regulatory phosphorylation of the MCM complex. The in vitro kinase assay suggested that, at least in vitro, Jil-1 does not phosphorylate the MCM complex, and we suspended our investigation at that stage.

Introduction:

The Jil-1 protein was first uncovered as an antigen for a monoclonal antibody in a screen for proteins whose nuclear localization was dynamic throughout the cell cycle (Johansen, 1996, Johansen et al., 1996). That work resulted in the partial cloning of the gene, and its identification as a tandem-serine kinase. Given the similar domain structure to the Janus kinase (JAK) family, but the dissimilarity of the residues in the catalytic core, the protein was named JIL-1 and founded a new family of kinases. The same lab further characterized the kinase activity of Jil-1 and found it was capable of phosphorylating Histone H3 (Jin et al., 1999). Intriguingly, they also discovered that Jil-1 has a unique, sex-specific enrichment on the male X chromosome, where it binds about 1.8 fold stronger than on the autosomes (Jin et al., 2000, Jin et al., 1999). This enrichment was shown to be MSL-dependent in nature, and recombinant tagged-Jil-1 (V5 tag or GST) could cross-immunopurify MSL complex subunits from S2 cell lystates (Jin et al., 2000).

The Johansen lab continued their work to characterize Jil-1 and identified the residue that Jil-1 phosphorylates on Histone H3 as the Serine 10 residue (Wang et al., 2001). Histone 3 serine 10 phosphorylation (H3S10ph) is a well-known mark for mitotic chromosomes, yet Jil-1 catalyzes this mark during interphase. Interestingly, the Corces lab had only a few months previously published their work using immunofluorescence to show that H3S10ph was dramatically reorganized to polytene chromosome puffs following heat shock, implicating this mark in transcriptional response (Nowak and Corces, 2000). This juxtaposed with the Johansen lab's finding in (Wang et al., 2001) that *Jil-1* seemed to have a very general role in polytene chromosome morphology using a series of truncated mutants. Despite the wildly deformed polytene chromosomes in *Jil-1* mutants, the MSL complex was found to still bind to the X chromosome, suggesting that the while the enhanced binding of Jil-1 to the X chromosome was MSL-dependent, the deposition of MSL on the X chromosome was Jil-1-independent.

Another function for Jil-1, this time related to heterochromatin, was uncovered when the Supressor of position-effect variegation [Su(var)] mutant *Su(var)3-1* was identified as antimorphic carboxy-terminal truncation alleles of Jil-1 (Ebert et al., 2004). This role in heterochromatin was further bolstered by the close examination of ultrastructure of polytene chromosomes in *Jil-1* mutants. It was observed by transmission electron microscopy of polytene chromosomes that there appeared to be ectopic contacts between chromosomal regions (Deng et al., 2005). One effect of these ectopic contacts was the loss of proper euchromatin/heterochromatin organization resulting in an intermingling of compact and open chromatin domains. Intriguingly, the male X chromosome did not behave in this way. It instead was observed to be highly dispersed, appearing as a shortened blob-like entity with no trace of proper organization remaining.

It was later observed in these deformed polytene chromosome preparations that there was an expansion of H3K9me2 and HP1a immunostaining onto the euchromatin arms from the heterochromatin chromocenter (Zhang et al., 2006). The X chromosome seemed particularly effected, though this time showing strong upregulation of these heterochromatin components throughout both the male and female X chromosomes. This observation piqued the interest of a post-doc in the Kuroda lab, Annette Plachetka, who confirmed this result using the higher-resolution technique of ChIP-on-chip of H3K9me2 from *Jil-1* mutant larvae (unpublished observations). She also assessed the level of expression of a panel of genes in heterozygous and homozygous male larvae, but observed only a small change in expression, leaving the precise role for Jil-1 in dosage compensation unanswered.

This reflects the main debate in the field in general about whether the main role of Jil-1 is primarily to maintain chromosome structure and genome organization, or to play a more proximal to transcription and gene expression. Certainly there are data to support both roles and these two major roles need not be mutually exclusive. The polytene chromosome results discussed above are further bolstered by findings that Jil-1 interacts physically or genetically with other chromosomal structural

proteins such as Chromator (Rath et al., 2006), Su(var)3-9 (Boeke et al., 2010) and Su(var)3-7 (Deng et al., 2010); that tethered Jil-1 induces chromatin structure remodeling without the recruitment of RNA Pol II (Deng et al., 2008); and that different domains of Jil-1 can redundantly rescue polytene chromosome morphology (Bao et al., 2008). Evidence supporting the role in gene expression include data showing that a series of *Jil-1* mutants have differential effects on dosage compensation of the *white* gene (Lerach et al., 2005); that H3S10ph may facilitate release of RNA Pol II from promoter-proximal pausing (Ivaldi et al., 2007) (notably disputed in (Cai et al., 2008)); that chromatin of induced genes in mammals require H3S10ph or H3S28ph (Drobic et al., 2010); and that the 14-3-3 proteins which coordinate transcriptional elongation are reliant on H3S10ph (Karam et al., 2010).

One notable study from the Becker lab looked comprehensively at the relationship between Jil-1 and transcription. They observed no correlation between intensity of Jil-1 binding and level of expression (Regnard et al., 2011). They also reported that the expression of the majority of genes were unaffected in *Jil-1* RNAi depletions, but there was a modest, yet significant, depletion of X-linked expression. The implication of this work is that Jil-1 may reinforce the active state, but alone is insufficient to establish it; this finding is indeed compatible with a role in both chromosome structure and in transcription. A general model for Jil-1 typically proposes that Jil-1 and the H3S10ph mark counteract heterochromatin proteins and heterochromatin-associated marks. Evidence for this dichotomy can be gleaned from all the above work, by a more targeted study of the effect of H3S10ph on expression and H3K9me2 (Cai et al., 2014), and by the fact that different domains of Jil-1 rescue different *Jil-1*-null defects (morphology vs H3K9me2 spread vs viability) (Li et al., 2013).

Interestingly, in the H3K9me2 ChIP-chip analysis by Annette Plachetka, H3K9me2 was observed to spread into euchromatin, as observed by immunofluorescence, but appeared to largely skip active gene bodies, suggesting that redundant mechanisms in addition to Jil-1 must protect genes against the spread of heterochromatin. Furthermore, most recently it was determined that some of the antibodies

used to characterize H3K9me2 and H3S10ph were ineffective at discerning the individual mark from the H3K9me2S10ph composite mark (Wang et al., 2014). While this composite mark is observed only in heterochromatin and on chromosome 4, this result calls into question any model for Jil-1 function invoking H3S10ph as an ironclad barrier to H3K9me2 deposition, and suggests a more dynamic interplay between euchromatin and heterochromatin than previously appreciated.

In order to investigate the role that Jil-1 plays in dosage compensation, members of our lab tagged Jil-1 with an amino-terminal BioTAP bipartite tag. I used this construct to express BioTAP-Jil-1 in both S2 (male) and Kc (female) cells. Using the BioTAP-XL procedure (Alekseyenko et al., 2015) and the transgenic S2 and Kc cells, I was able to identify the genomic localization of BioTAP-Jil-1 and compare it to the known sex-specific differential pattern. From the same input chromatin I was able to characterize Jil-1 interacting proteins. Consistent with the pleiotropic effect observed in *Jil-1* mutants, Jil-1 interacts with a large array of proteins. The only remarkable sex-specific difference in interactors, however, were components of the MSL complex. In comparison to other BioTAP-XL pulldowns by the lab, the mini-chromosome maintenance (MCM) complex was a candidate Jil-1-specific interactor. As Jil-1 is an established kinase, I sought to determine if it might regulate the activity of the MCM complex by phosphorylation, and thereby affect the timing of replication (Bochman and Schwacha, 2009), which is known to be early for the male X chromosome (Chatterjee and Mukherjee, 1977). The results of the kinase assay using recombinant proteins however suggested that Jil-1 was not able to phosphorylate any of the available residues on the MCM complex, leaving the link between replication timing and the X chromosome a mystery.

Results and Discussion:

BioTAP-Jil-1 localization recapitulates known sex-specific Jil-1 binding patterns

Annette Plachetka, a former post-doc in the lab, cloned a BioTAP tag into the amino-terminus of the *Jil-1* genomic locus using the pRedET recombineering system. The tagged *Jil-1* locus was transferred to the pGS ("pFly") vector for transfection and expression. To generate stably expression BioTAP-tagged S2 and Kc cells, I co-transfected pGS-mw-BioTAP-Jil-1 and pHygro, which confers resistance to the antibiotic Hygromycin, using the calcium phosphate transfection kit. Following selection, I verified that BioTAP-Jil-1 was expressed by running crude nuclear extracts and probing for the protein A moieties of the BioTAP tag using peroxidase-anti-peroxidase (PAP) (Figure A1-1b). Since it was reported that Jil-1 can autophosphorylate (Jin et al., 1999) and may be regulated by phosphorylation, I tested whether the inclusion of phosphatase inhibitors when making crude nuclear preparations would have an effect on protein stability. I found that this did not appear to be the case and expression appeared unchanged +/- phosphatase inhibitors. An unknown band (double asterisk), common to tagged extracts, may be a degradation product of BioTAP-Jil-1, but is largely depleted from the nuclear extract of fractionated S2 cells (Figure A1-1b, details below). Furthermore, when BioTAP-Jil-1 was detected in immunofluorescence experiments in S2 (Figure A1-2a) or Kc (data not shown) cells by PAP, the signal is localized to the nucleus. Depending on the level of overall staining in the nucleus, the pattern of BioTAP-Jil-1 staining appears somewhat reminiscent of the X-chromosomal staining of MSL1 by immunofluorescence (Figure A1-2b). These data suggest that the BioTAP-Jil-1 transgene is functional in S2 and Kc cells, though transgenic complementation of Jil-1 mutant flies was not performed.

To analyze protein and DNA interactions of BioTAP-Jil-1, I performed the BioTAP-XL procedure (Alekseyenko et al., 2015) using BioTAP-Jil-1-expressing S2 and Kc cells. Following tandem affinity purification, I split the samples, and from a one quarter aliquot from each of the S2 and Kc cell

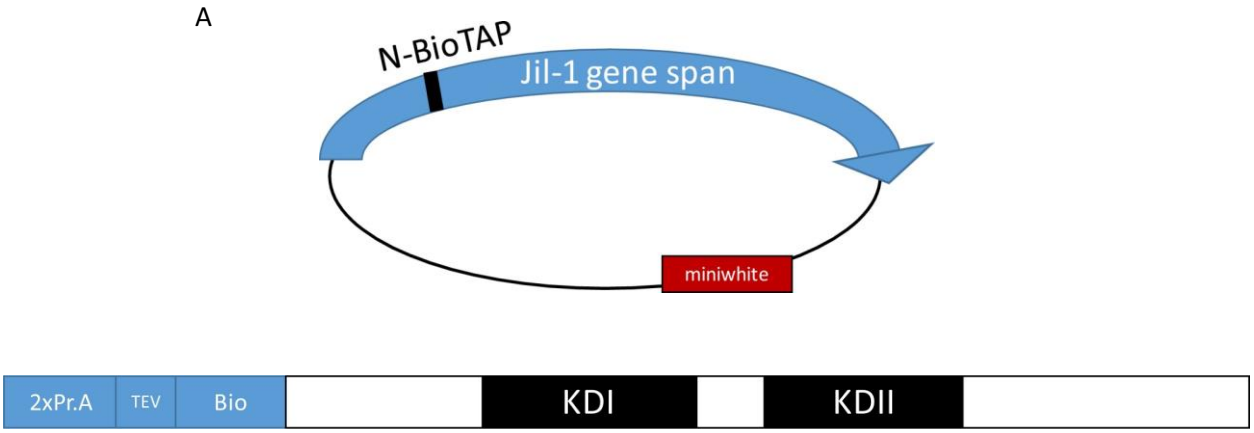
Figure A1-1: BioTAP-Jil-1 expression in S2 and Kc cells

A) An amino-terminal BioTAP-tag was cloned into a *Jil-1* genomic construct. The resulting protein carries the proteinA-biotin acceptor tag, with the addition of a TEV protease recognition site between the two tag halves. Jil-1 has two kinase domains (I and II) and is known to autophosphorylate.

B) S2 and Kc cells were transfected with the BioTAP-Jil-1 construct. To verify expression and stability of the BioTAP-Jil-1 protein with and without phosphatase inhibitors, crude nuclear extracts were prepared and expression was assessed by Western blot against the proteinA portion of the tag using peroxidase-anti-peroxidase (PAP). The inclusion of Phosphatase inhibitors had no effect on BioTAP-Jil-1 stability.

Figure A1-1 (Continued)

A



B

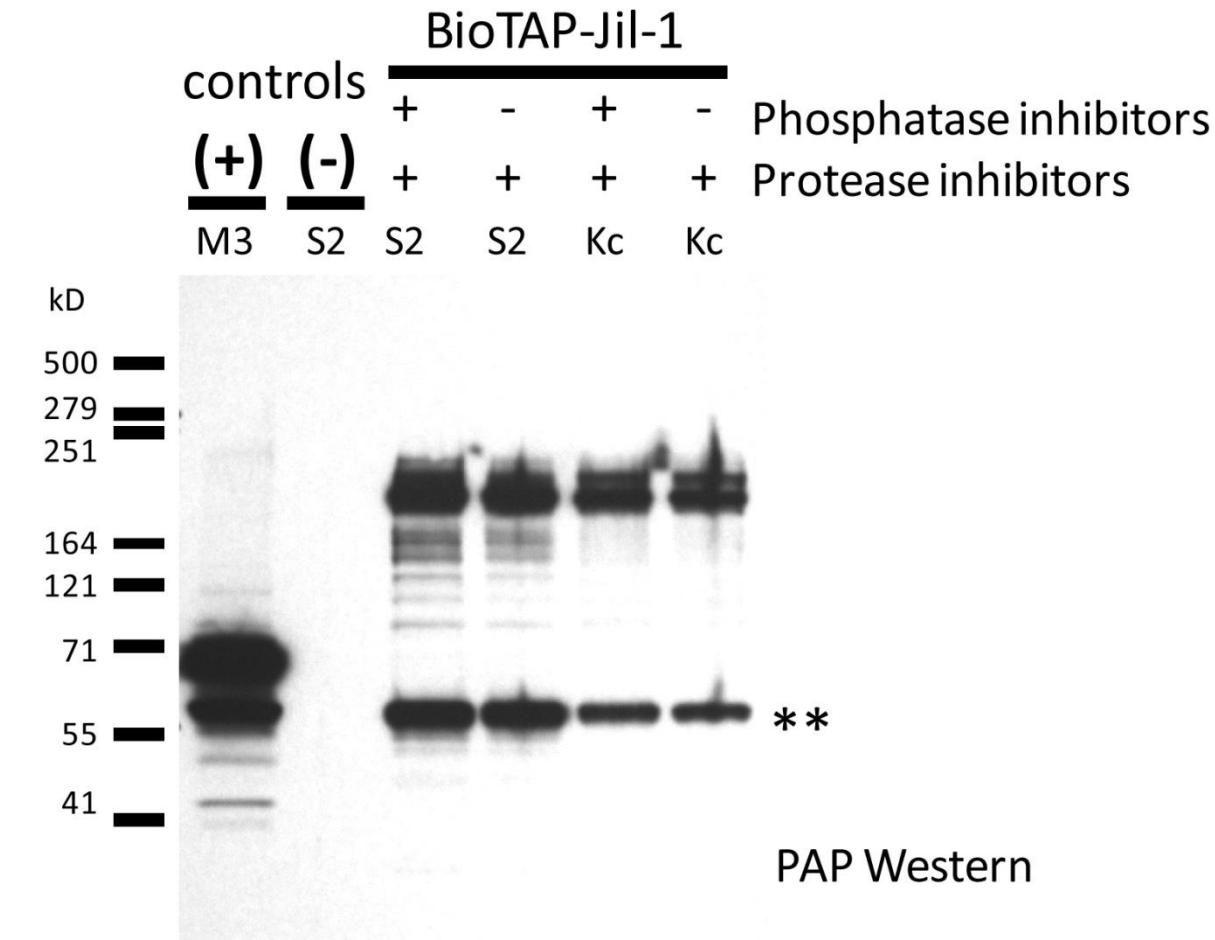
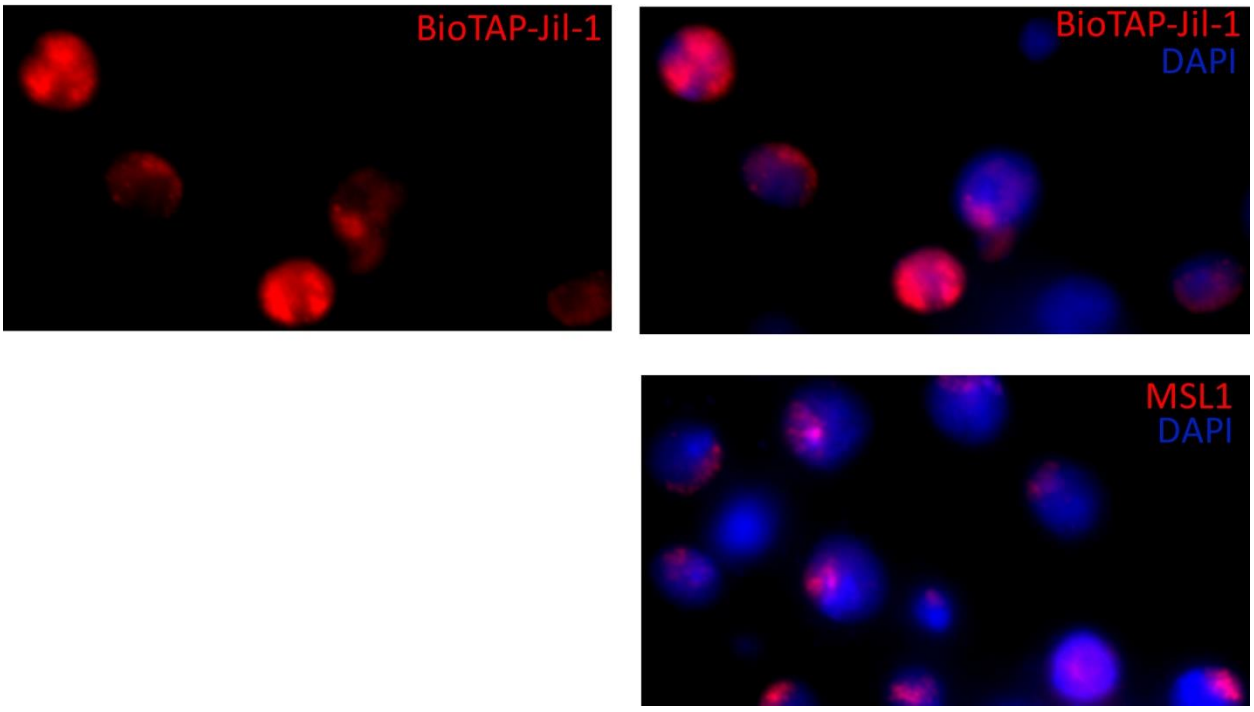


Figure A1-2: Immunofluorescence of BioTAP-Jil-1 in S2 cells is reminiscent of MSL1 staining in S2 cells

A) BioTAP-Jil1-expressing S2 cells were immunostained using PAP. DNA was counter stained with Hoechst. The resulting fluorescent pattern was, in some nuclei, reminiscent of the strong sectoring of MSL staining for the X chromosome. Jil-1 is known to be enriched on the X chromosome in male cells, and this result suggests BioTAP-Jil-1 is active in S2 cells.

B) Anti-MSL1 immunofluorescence in S2 cells, counterstained with Hoechst.

Figure A1-2 (Continued)



purifications, isolated the affinity-purified genomic DNA, prepared Illumina sequencing libraries, and had those libraries sequenced. We compared the resulting data to the ChIP-chip results generated by the modEncode project using anti-Jil-1 antibody and chromatin from S2 and Kc cells (datasets 945 and 3037, respectively). Visual comparison of the tracks in the genome browser show a high degree of concordance between BioTAP-Jil-1 ChIP-seq and anti-Jil-1 ChIP-chip (Figure A1-3). The enrichment of Jil-1 on the male X chromosome is evident, and regions of sex-specific Jil-1 binding are recapitulated by the BioTAP-Jil-1 ChIP-seq (Figure A1-4). Since the ChIP-seq agreed highly with the ChIP-chip data, we did not perform additional analyses on the BioTAP-Jil-1 ChIP-seq datasets.

Few sex-specific BioTAP-Jil-1 interactions identified by mass spectrometry

That we see BioTAP-Jil-1 localizing properly by ChIP lent further support to the functionality of the tagged protein, so I sought to identify any sex-specific interacting partners by liquid chromatography tandem mass spectrometry (LC-MS/MS). The goal of this effort was to understand the targeting or function of Jil-1, and whether additional factors could clarify the dosage compensation, chromatin structural, or transcriptional roles for Jil-1. Using the three-quarter remainders of the tandem affinity purified samples from the BioTAP-XL procedure, I performed on-bead trypsinization to generate peptides for mass spec. The list of proteins and their peptide counts recovered in the two Jil-1 experiments were compared to untagged and input controls (to address specificity and enrichment, respectively), and the pared down list was analyzed for enrichment of known pathways or complexes.

An immediate result of the mass spec was the confirmation by reciprocal pulldown of the Jil-1-MSL complex interaction. All members of the MSL complex were found in the S2 cell IP, but were absent from the KC cell IP (Table A1-1), consistent with their relative instability (MSL1 & MSL3) or lack of expression (MSL2) in female cells. Furthermore, MOF and MLE, which are expressed in both male and female cells, were still recovered in the Kc cell IP. This suggests that the interaction with MOF and MLE

Figure A1-3: BioTAP-Jil-1 ChIP-Seq from S2 and Kc are highly similar to modEncode anti-Jil-1 ChIP-chip

A) A representative 182kb region on Chr3R. BioTAP-Jil-1 ChIP-seq in S2 and Kc cells agree quite well with each other on autosomes. Comparison of the BioTAP-Jil-1 experiments to anti-Jil-1 ChIP-chip experiments from modEncode shows a high degree of correlation.
B) As in (A), a representative 75kb region of Chr2L. Jil-1 is known to have a preference for the bodies of active genes.

Figure A1-3 (Continued)

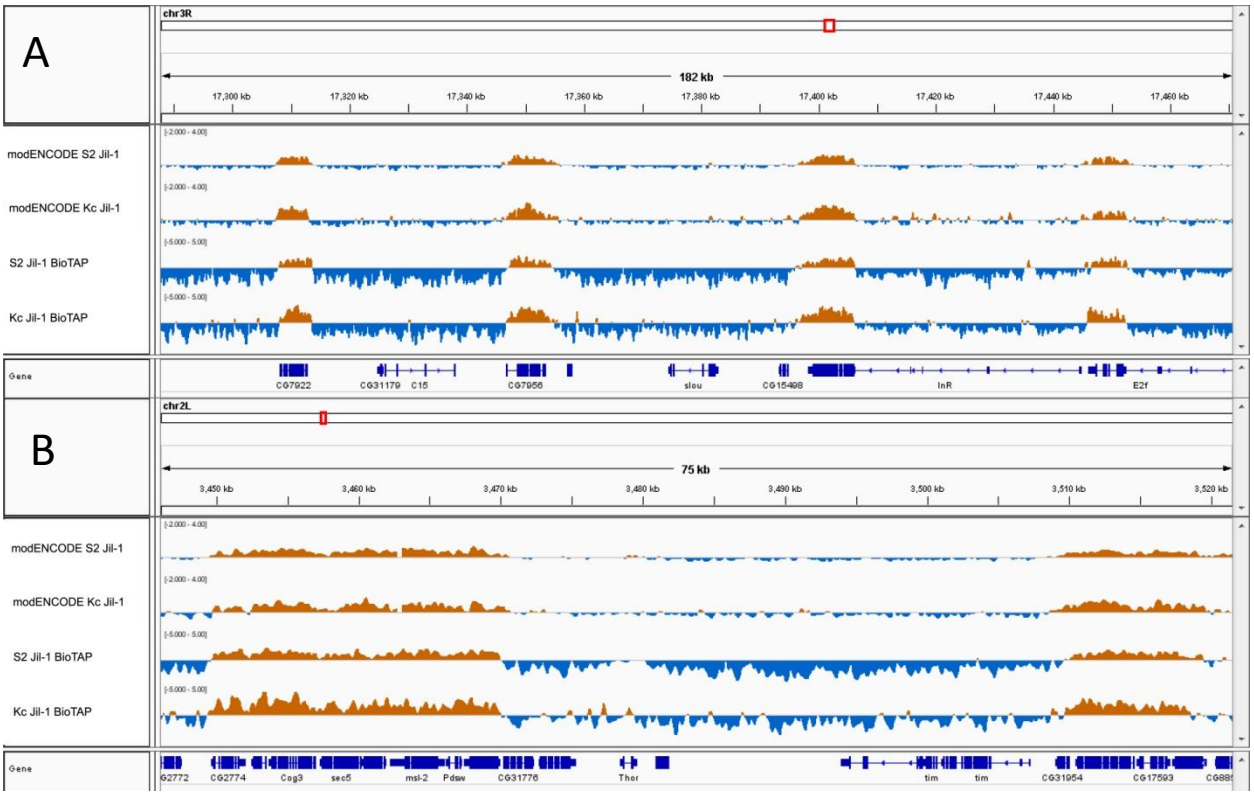


Figure A1-4: BioTAP-Jil-1 recapitulates sex specific differences observed between S2 and Kc cells

A) Jil-1 is known to have an X-specific enrichment in male cells. The BioTAP-Jil-1 samples recapitulate several of these observed sex-specific differences. A 343kb region of ChrX is shown. In S2 cells BioTAP-Jil-1 is more broadly enriched than in Kc cells. These regions agree with the patterns observed in the anti-Jil-1 ChIP-chip modEncode data.

B) As in (A), a representative 73kb region of ChrX. As on autosomes, Jil-1 binding is enriched over gene bodies.

Figure A1-4 (Continued)

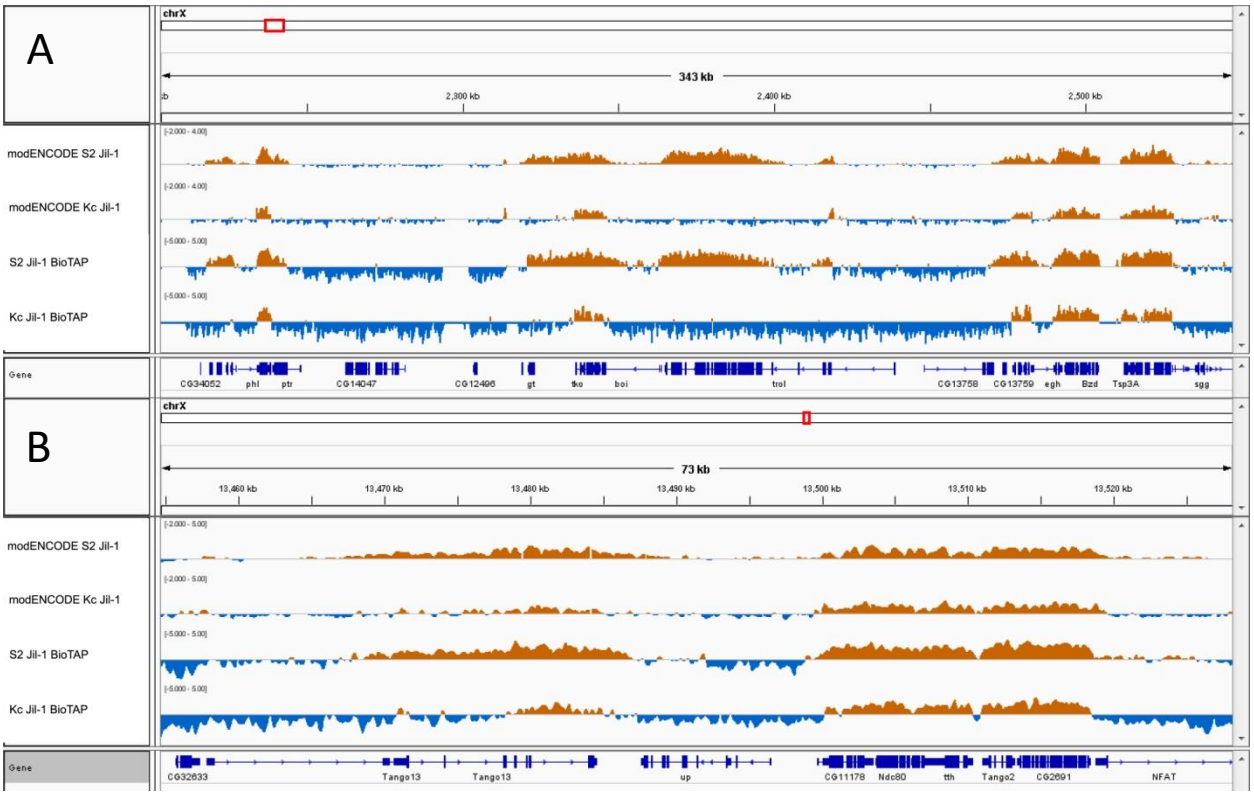


Table A1-1: BioTAP-XL of Jil-1 recovers MSL and MCM complexes

Total peptide counts from BioTAP-XL proteomics experiments for Jil-1 baits in S2 (green) and Kc (pink) cells. BioTAP-Jil-1 is recovered at similar levels from both cell lines. The core components of the MSL complex are recovered in S2 cells but not Kc cells. This results supports the literature (Alekseyenko et al 2012) and shows the interaction occurs by reciprocal IP (Wang et al 2013). Interestingly, the MCM complex is recovered in both S2 and Kc cells, and was therefore selected for follow-up experiments. Other potentially functional interactors mined from the literature are listed, though the enrichments of many of these proteins compared to input and controls (MSL3, HP1, and POF) are somewhat low to strongly support critical functional interactions.

Table A1-1 (Continued)

Gene Name	Jil-1 S2	Jil-1 Kc	untagged	input	MSL3	MSL3	HP1	Pc	E(z)
JIL-1	123	119	0	0	13	15	3	3	0
msl-1	7	0	0	0	35	23	9	0	0
msl-2	1	0	0	0	11	10	2	0	0
msl-3	4	0	0	3	33	21	3	0	0
mof	4	2	0	1	26	24	19	0	0
mle	1	1	0	2	19	13	0	0	0
Mcm2	7	7	0	0	0	0	10	0	0
Mcm3	10	5	0	0	0	0	9	0	0
dpa	12	15	0	0	0	0	11	0	2
Mcm5	7	8	0	0	0	0	10	1	1
Mcm6	6	4	0	0	0	0	6	0	0
Mcm7	3	6	0	0	0	0	7	0	0
14-3-3 epsilon	36	22	0	5	4	4	30	8	7
14-3-3 zeta	17	16	0	4	3	2	9	5	4
RpA-70	18	20	0	0	3	0	2	3	7
HP1	7	5	0	8	5	4	92	1	9
Su(var)3-9	1	0	0	1	0	0	6	0	0

may exist outside of the context of the intact MSL complex, but that the majority of Jil-1 protein interacts with proteins distinct from the dosage compensation machinery. Unfortunately, no other enriched sex-specific differences were identified to suggest a mechanism for differential Jil-1 activity or targeting in the two sexes (Table S2).

As mentioned above, one model for Jil-1 function with respect to transcriptional activation invokes H3S10ph as a requisite mark for the scaffolding protein family 14-3-3, whose recruitment to chromatin facilitates transcriptional elongation. The two Drosophila paralogs of 14-3-3 (epsilon and zeta) are enriched greater than 4-fold in both S2 and Kc pulldowns (Table A1-1), which seems to lend support to the validity of this hypothesis (Karam et al., 2010). Interestingly, recent work has shown that Jil-1 is capable of phosphorylating Su(var)3-9, a heterochromatin component (Boeke et al., 2010). While no sex-specificity for this interaction was presented in vivo, we did observe Su(var)3-9 in the S2 IP, but not in the Kc IP. However, the recovery of Su(var)3-9 in the BioTAP-Jil-1 S2 experiment was not enriched over input recovery levels. So our results regarding a potential Jil-1/Su(var)3-9 interaction should be considered inconclusive or would require additional validation.

BioTAP-Jil-1 mass spec recovers the replication helicase complex, MCM2-7

One intriguing finding in the BioTAP-Jil-1 mass spec data is that the entirety of the MCM2-7 complex is recovered in both S2 and Kc experiments. This ubiquitously expressed complex has been found as a non-specific contaminant in other chromatin-based protein isolation techniques (R. Kingston, personal communication), however in our BioTAP-XL method, these proteins are generally not recovered (Table A1-1). The mini-chromosome maintenance complex (MCM) is a heterohexameric ring which functions as the replicative helicase along with accessory proteins Cdc45 and the GINS complex (Moyer et al., 2006). The complex is recruited to origins through interactions directly with ORC

components, and is loaded onto DNA in a poised, salt-stable double-hexamer configuration in a Cdt1-dependent manner (Samson and Bell, 2013).

Notably, none of the ORC components, Cdt1, Cdc45, nor any of the members of the GINS complex are recovered in the Jil-1 pulldowns, suggesting that the involvement between Jil-1 and the MCM complex may be outside of the context of the replicative helicase. Indeed, several lines of evidence implicate the MCMs in functions outside of replication, including transcription (Snyder et al., 2009), chromatin remodeling, and cell-cycle checkpoint regulation (for review, see (Forsburg, 2004)). A combination of these processes via the MCM complex could explain the role of Jil-1 not only in dosage compensation (X-specific transcription), but also general transcription, genome replication timing (X chromosome known to be early replicating), and polytene chromosome morphology.

The MCM complex is expressed at a level much higher than is needed for replication, leading to the hypothesis that there may be distinct populations of MCM complexes doing a variety of nuclear tasks. The tails of the MCM complex proteins are thought to be integral to the regulation of MCM activity. With respect to the well characterized task of replication initiation and helicase function, the MCM complex subunits need to be phosphorylated by DDK and CDK, two cell-cycle-specific kinases. Interestingly, there are additional phosphorylation marks on the MCM proteins themselves that are not accounted for by DDK or CDK activity, and in fact, some DDK-dependent phosphorylation sites apparently require other pre-existing phosphorylation marks (Figure A1-5) (Randell et al., 2010). The kinases responsible for these other phosphorylation sites are unknown. This raised the intriguing possibility that Jil-1 may in fact have a regulatory interaction with the MCM complex, whereby phosphorylation of the MCM complex by Jil-1 could activate it for transcription or replication tasks.

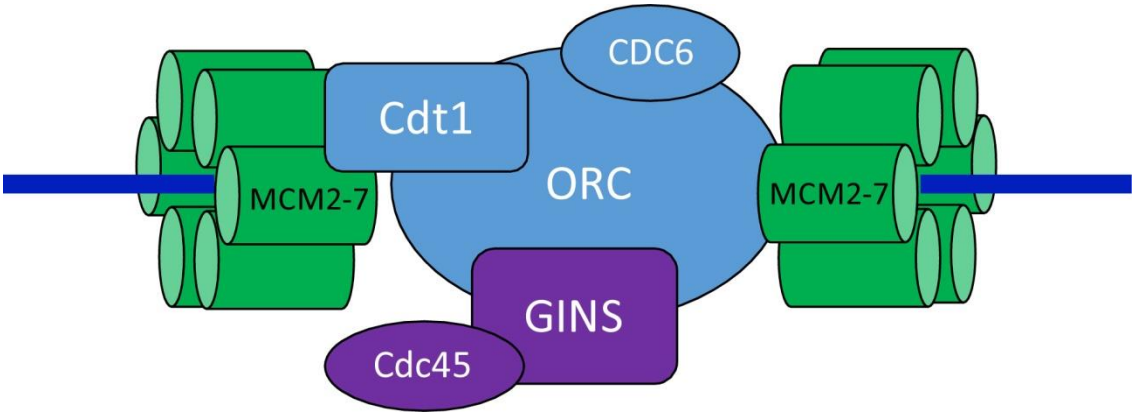
BioTAP-Jil-1 kinase is active in vitro, but does not appear to phosphorylate MCM2-7

Figure A1-5: The MCM complex is the replicative helicase and is regulated by phosphorylation

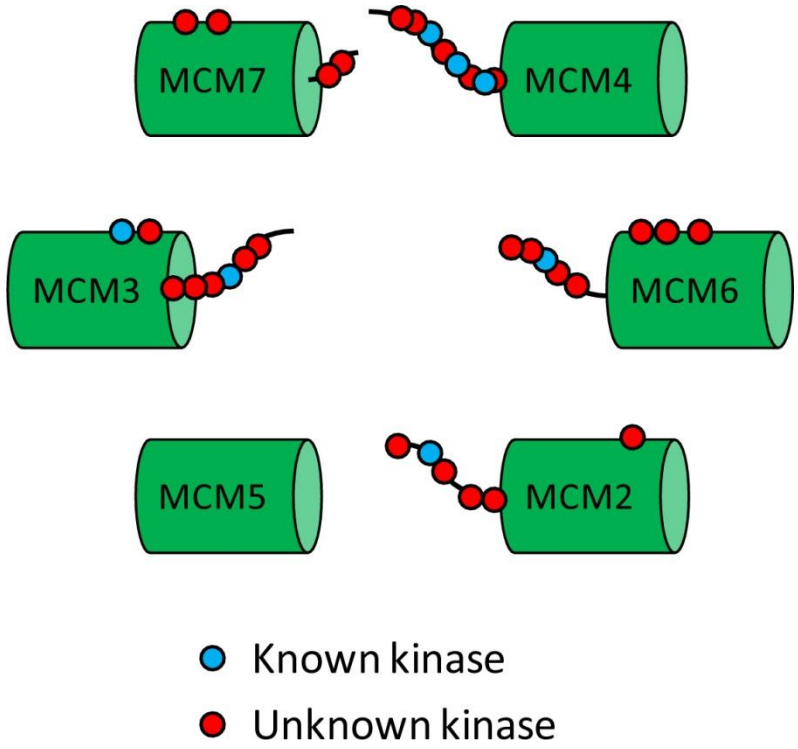
A) The MCM complex is a hexameric ring of AAA+ ATPases which functions as the replicative helicase along with Cdc45 and the GINS complex. MCM is loaded in at origins of replication through interactions with ORC and Cdt1. Only the MCM hexamer is recovered in the BioTAP-Jil-1 proteomics.
B) The MCM complex is densely phosphorylated by numerous kinases. While some are known (schematized in blue), many are catalyzed by unknown kinases (red). These modifications are necessary for MCM activation and are thought to be an integration point for signaling pathways.

Figure A1-5 (Continued)

A



B



In order to test this hypothesis, I endeavored to perform an in vitro kinase assay using recombinant Jil-1 and recombinant MCM complex. To do so, I obtained MCM-expressing baculoviruses generously provided by the Botchan lab (UC-Berkeley). After viral amplification, Sf9 cells were co-infected with these baculoviruses such that all MCM components should be expressed. I then Flag-purified the complex from Sf9 nuclear extracts, and performed ion-exchange column chromatography to obtain stoichiometric hexamer complexes (Figure A1-6A). I then tested for phosphorylation by recombinant Jil-1.

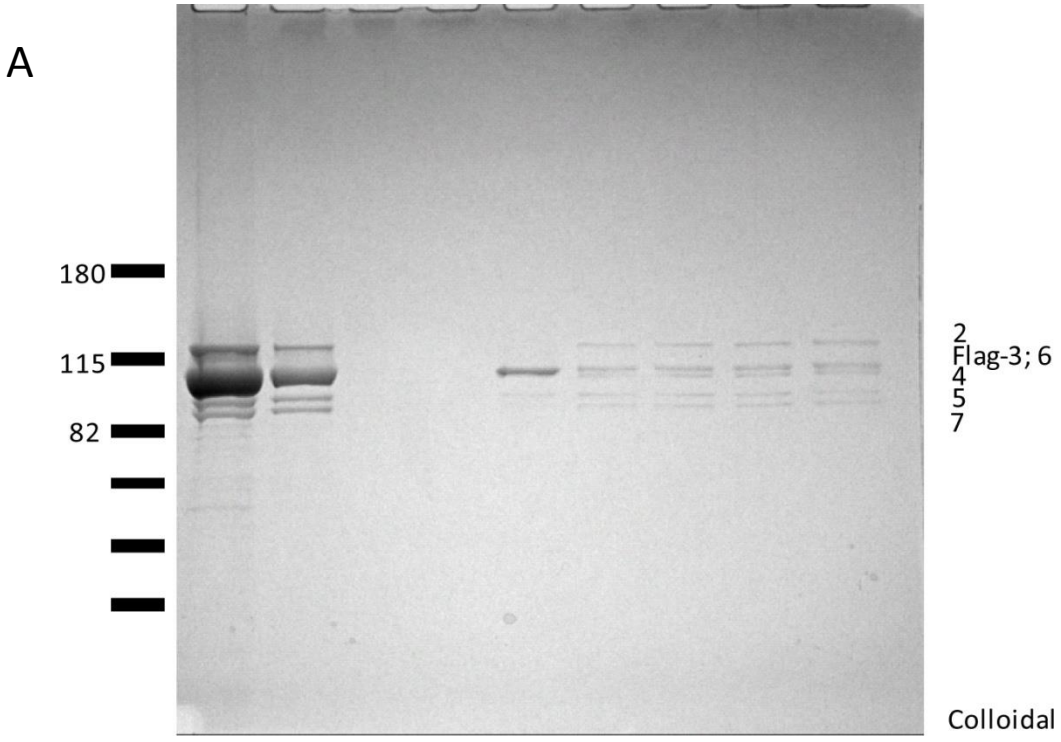
I originally attempted to express a GST-tagged Jil-1 construct in bacteria as the source for recombinant Jil-1 protein. While I was able to efficiently induce GST-Jil-1 in bacteria, the protein was consistently found in the insoluble pellet following cell lysis (Figure A1-7A). Only trace amounts of GST-Jil-1 were recovered using a glutathione resin, limiting the usefulness of this expression method for the kinase assay. I then turned to the BioTAP-Jil-1 construct expressed in S2 cells (Figure A1-7B). I fractionated S2 cells and pulled BioTAP-Jil-1 down using IgG-resin. I then liberated Bio-Jil-1 from the resin using the TEV protease, which cleaves between the Protein A and Biotin-accepting sequences of that early version of the BioTAP tag. As a control, I performed the same fractionation, IgG pulldown, and TEV protease treatment from untransfected S2 cells (Figure A1-7C). I then used the TEV-treated supernatant, which should contain Bio-Jil-1 from the BioTAP-Jil-1 transfected S2 cells and only trace contaminants from the untransfected S2 cells as the source of kinase for the in vitro assay.

To test the in vitro kinase activity of the TEV supernatants (ie, Bio-Jil-1 vs trace contaminants), I used recombinant Xenopus histone H3 (a gift from Bob Kingston, MGH) and salt/acid-extracted S2 histones (Figure A1-6B), which should function as positive controls. When Bio-Jil-1 alone was incubated with P32-labeled ATP, I observed a strong signal, suggesting that autophosphorylation was occurring. In comparison, there was still some signal in the untransfected S2 purification, albeit much less, suggesting that traces of additional kinases may have been pulled down non-specifically. However, when incubated

Figure A1-6: Purification of substrates for in vitro kinase assays

A) To purify the MCM complex, Baculovirus-infected nuclear extracts were prepared as described. To obtain stoichiometric complexes, the Flag-purified material was further purified using ion exchange column chromatography. The resulting fractions have what appear to be stoichiometrically equal MCM hexamers.
B) Jil-1 is known to phosphorylate histone H3 at serine 10 (H3S10ph). Histones therefore were utilized as positive controls to test the activity of the purified recombinant Jil-1. Using a hybrid acid/salt extraction method from S2 cells, the core 4 histones were recovered at high purity. Histone H1 may also be present in these preparations at low levels.

Figure A1-6 (Continued)



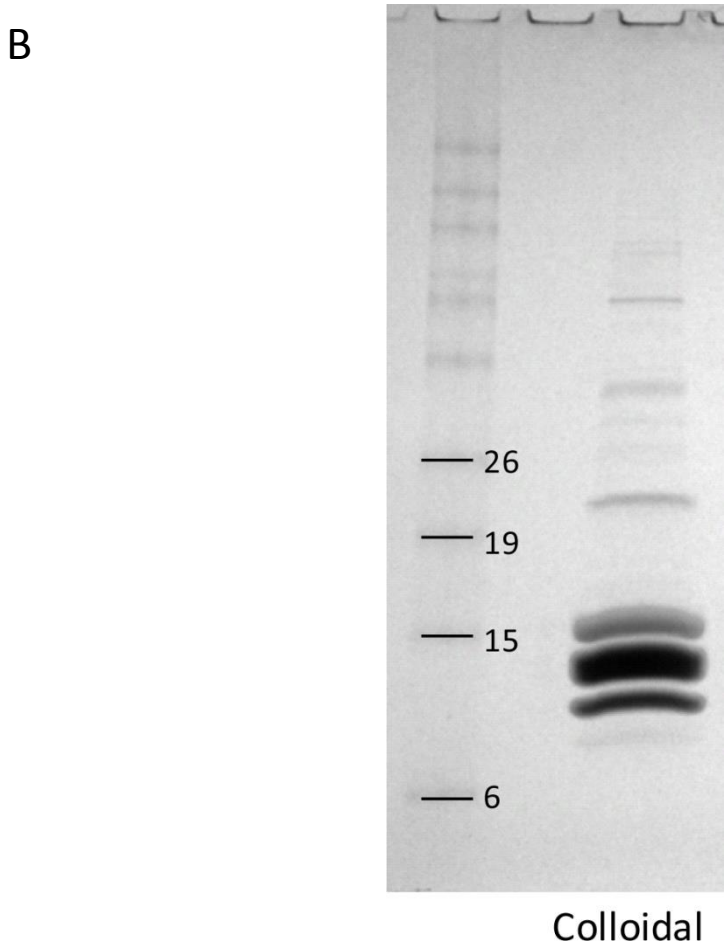


Figure A1-7: Expression of recombinant Jil-1 for in vitro kinase assays

A) Initial efforts to express Jil-1 centered on GST-Jil-1 expression in bacteria. While induction was achieved (* for GST-Jil-1, † for GST-ADD-WT domain control), and the induced band reacted with a GST antibody, the GST-Jil-1 fusion was present at only very low levels in elution fractions.
B) BioTAP-Jil-1 was expressed in stably transfected S2 cells. Nuclear extracts (NE) were enriched for BioTAP-Jil-1 over other potential contaminant bands as compared to whole cell extracts (WCE) or cytoplasmic extracts (CE).
C) For in vitro kinase reactions, BioTAP-Jil-1 was pulled down using IgG-agarose and Bio-Jil-1 was liberated from the resin using TEV protease, which recognizes an amino acid sequence between the ProteinA moieties and biotin accepting sequence in this earlier version of the BioTAP tag. As a negative control, S2 nuclear extracts were treated in the same way.

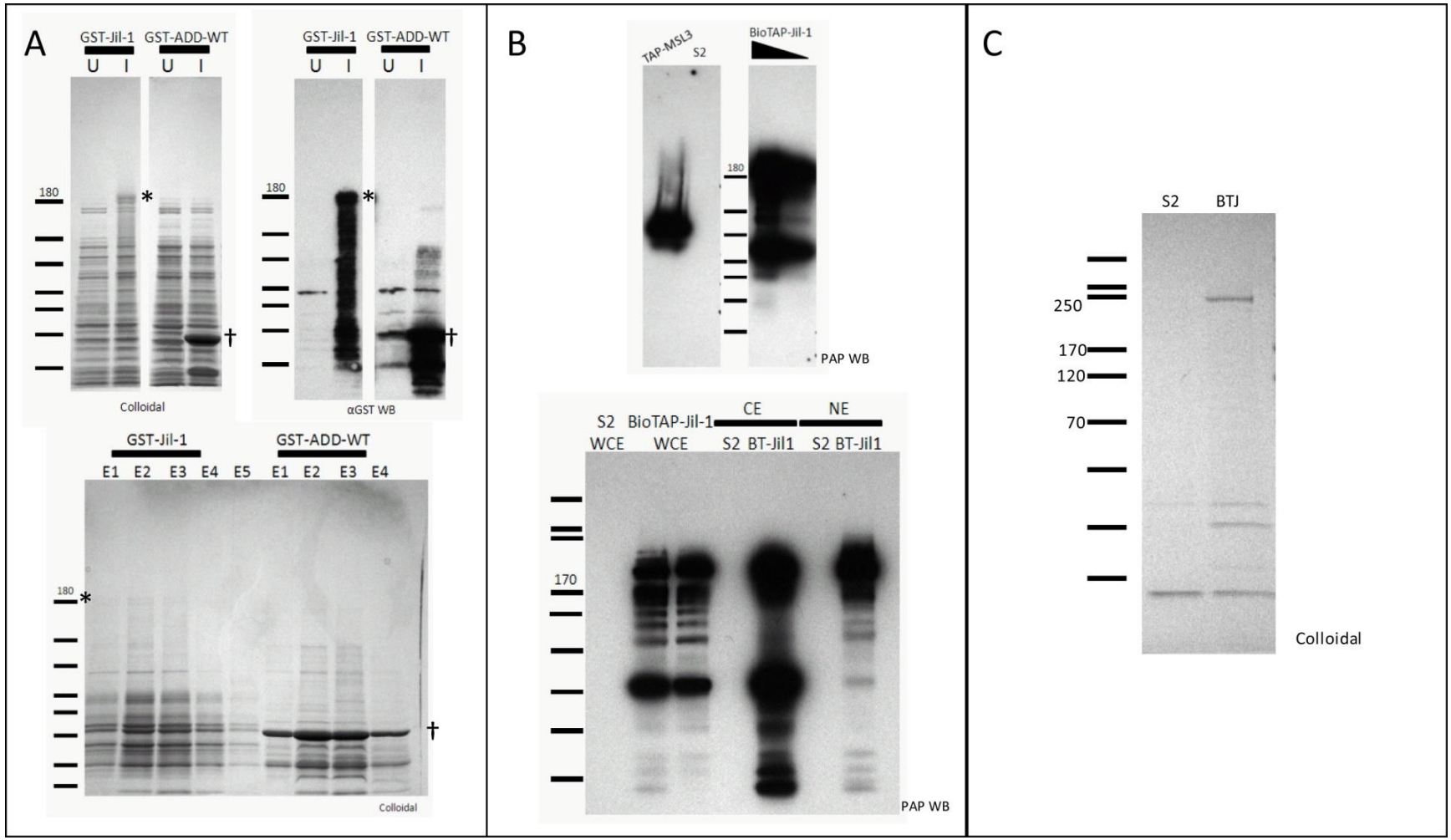


Figure A1-7 (Continued)

with either Xenopus histone H3 or S2 bulk histones, I observed robust phosphorylation at the appropriate size for H3 in the BioTAP-Jil-1 TEV supernatant only (Figure A1-8A).

To test for phosphorylation of the MCM complex by Jil-1, I mixed the stoichiometric MCM complexes with the TEV supernatants from BioTAP-Jil-1 and untransfected S2 cells as above. While there is evidence of weakly phosphorylated bands appearing in the Bio-Jil-1 sample, several of the bands either existed previously in the Bio-Jil-1 alone lane, or appeared to also exist in the S2+MCM control lane (Figure A1-8B). Given the robust signal observed with the positive control H3 and with the autophosphorylation of Bio-Jil-1, we felt that the lack of a strong signal in the region of the MCM complex suggested a negative result for this experiment.

Even though the in vitro kinase assay generated a negative result for the phosphorylation of the MCM complex by Jil-1, the nature of the MCM/Jil-1 interaction is intriguing since it could potentially explain numerous of the Jil-1 phenotypes. If the MCM/Jil-1 interaction is real, it may not rely on regulatory information conferred by phosphorylation. Alternatively, it is possible that future efforts could determine a set of conditions in vitro that allow for phosphorylation to occur, or more careful proteomic study of the MCM complex isolated in BioTAP-Jil-1 experiments may be employed to attempt to identify an in vivo phosphorylation event. Jil-1 clearly plays a role in open chromatin structure, though, and perhaps this role brings it into contact with the MCM complex either as the replication fork is transiting or in any of the other putative functions for the MCM complex.

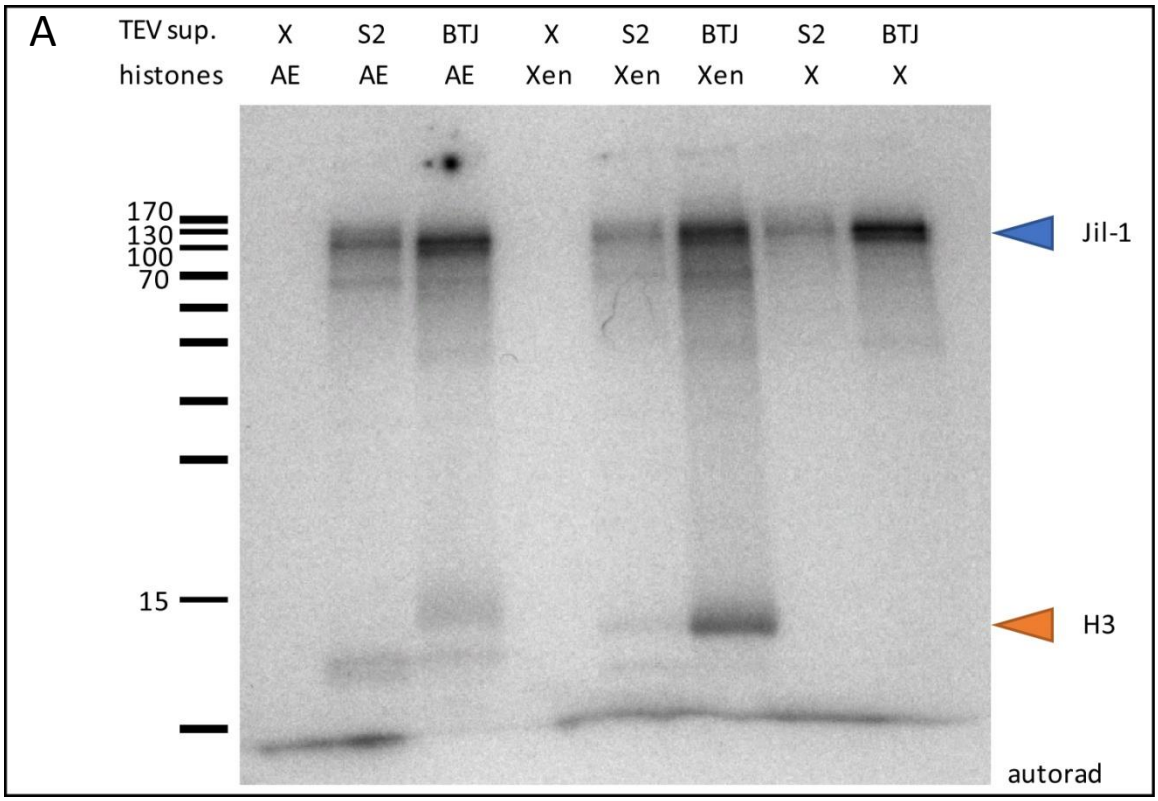
Despite our efforts, Jil-1 remains an intriguing enigma in the field. The wide range of phenotypes in *Jil-1* mutants suggest a diverse array of potential functions. From its clear involvement with the MSL complex to the less clear mechanism of this involvement, and from its potential role in transcription to evidence for a much more general role in genome organization, Jil-1 will undoubtedly be the focus of numerous research groups for years to come.

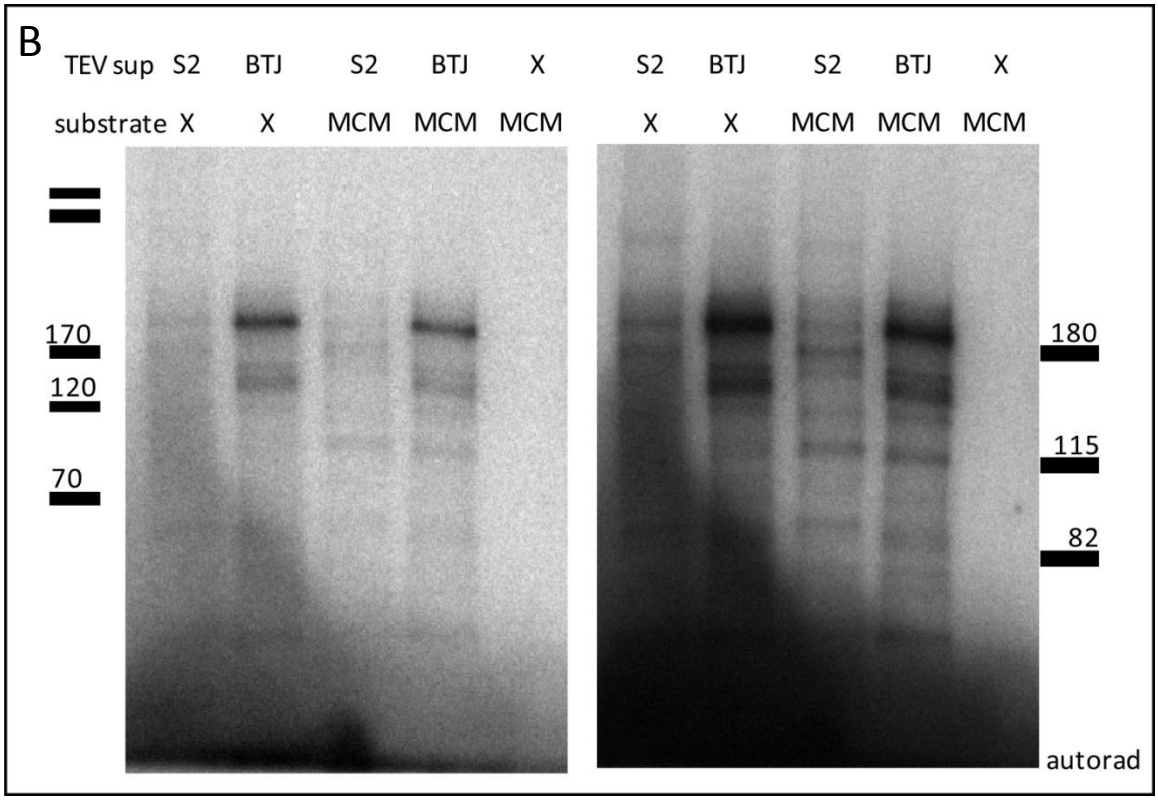
Figure A1-8: Jil-1 does not appear to phosphorylate the MCM complex in vitro

A) In order to confirm the activity of Jil-1 in the TEV supernatant, BioTAP-Jil-1 TEV supernatant (BTJ) or S2 TEV supernatant (S2) were incubated with salt/acid extracted (AE) core Drosophila histones or with recombinant Xenopus histone H3 (Xen) and P32-labeled ATP. We observed strong auto-phosphorylation of the Jil-1 band. The S2 fraction also had some background kinase activity from trace contaminant kinases. Only the BioTAP-Jil-1 TEV supernatant had kinase activity toward the histones, however.

B) As in (A), except TEV supernatant kinase sources were incubated with the stoichiometric MCM complexes. There did not appear to be any strong indication that BioTAP-Jil-1 phosphorylated any of the members of the MCM complex. Two exposures of the same autoradiograph are shown.

Figure A1-8 (Continued)





Materials and Methods:

Cloning

An early version of the BioTAP tag was cloned into a genomic transgene of *Jil-1* by recombineering as in Chapter 2 and 3.

Western Blotting

Western blotting was performed as in Chapter 3. Anti-GST (Sigma G7781) was used at 1:10,000.

Immunofluorescence

Immunofluorescence in S2 and Kc cells was performed as in Chapter 3. Anti-MSL1 was used at a dilution of 1:100.

BioTAP-XL for Jil-1

BioTAP-XL was performed as in Chapter 3 and in (Alekseyenko et al., 2015). BioTAP-Jil-1 expressing S2 and Kc cells were used to identify sex-specific interactors.

ChIP-seq analysis

ChIP-seq analysis was the same as in Chapter 2.

Recombinant Jil-1 expression

GST-Jil-1 and GST-ADD-WT carrying bacterial strains were induced at OD_{600}=~0.5 using 1mM IPTG. GST-Jil-1 was found to be largely insoluble regardless of induction temperature or duration. BioTAP-Jil-1 was transfected into S2 cells to generate a stable line. Extracts were made as from Sf9 cells in Chapter 2.

BioTAP-Jil-1 was immunoprecipitated using IgG-agarose and liberated from the resin using TEV protease, which cleaves an earlier version of the BioTAP tag between modules.

Recombinant MCM expression (Flag/baculo, similar to chapter 3; purification)

Baculovirus stocks for MCM-expressing viruses were a kind gift from the Botchan lab (UC-Berkeley). Viruses were re-amplified and co-infected as in Chapter 3. Purification of stoichiometric MCM2-7 complexes was performed with ion exchange chromatography as described (Ilves et al., 2010). Ion exchange chromatography was performed using the fast protein liquid chromatography (FPLC) set up of the Winston lab (HMS).

Histone extraction (salt/acid)

Histones were extracted using a salt/acid hybrid protocol essentially as described (Zee et al., 2016).

In vitro kinase assay

TEV supernatants containing recombinant Bio-Jil-1 or from a control S2 lysate (trace contaminants only) were incubated with acid-extracted S2 histones, recombinant Xenpous H3 (a gift from the Kingston lab), or stoichiometric MCM complexes in the presence of P^{32}-labeled ATP. After a 20 minute incubation at 30C, the reaction was stopped by the addition of 2x SDS sample buffer. Histone in vitro kinase reactions were resolved on 18% acrylamide gels, while MCM in vitro kinase reactions were resolved on 8% acrylamide gels. Gels were dried and exposed to film for a few hours or overnight.

References:

Alekseyenko, A.A., McElroy, K.A., Kang, H., Zee, B.M., Kharchenko, P.V., and Kuroda, M.I. 2015. BioTAP-XL: Cross-linking/Tandem Affinity Purification to Study DNA Targets, RNA, and Protein Components of Chromatin-Associated Complexes. *Curr Protoc Mol Biol* 109: 21 30 21-21 30 32.

Bao, X., Cai, W., Deng, H., Zhang, W., Krencik, R., Girton, J., Johansen, J., and Johansen, K.M. 2008. The COOH-terminal domain of the JIL-1 histone H3S10 kinase interacts with histone H3 and is required for correct targeting to chromatin. *J Biol Chem* 283: 32741-32750.

Bochman, M.L., and Schwacha, A. 2009. The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol Mol Biol Rev* 73: 652-683.

Boeke, J., Regnard, C., Cai, W., Johansen, J., Johansen, K.M., Becker, P.B., and Imhof, A. 2010. Phosphorylation of SU(VAR)3-9 by the chromosomal kinase JIL-1. *PLoS One* 5: e10042.

Cai, W., Bao, X., Deng, H., Jin, Y., Girton, J., Johansen, J., and Johansen, K.M. 2008. RNA polymerase II-mediated transcription at active loci does not require histone H3S10 phosphorylation in Drosophila. *Development* 135: 2917-2925.

Cai, W., Wang, C., Li, Y., Yao, C., Shen, L., Liu, S., Bao, X., Schnable, P.S., Girton, J., Johansen, J., and Johansen, K.M. 2014. Genome-wide analysis of regulation of gene expression and H3K9me2 distribution by JIL-1 kinase mediated histone H3S10 phosphorylation in Drosophila. *Nucleic Acids Res* 42: 5456-5467.

Chatterjee, R.N., and Mukherjee, A.S. 1977. Chromosomal Basis of Dosage Compensation in Drosophila. IX. Cellular Autonomy of the Faster Replication of the X Chromosome in Haplo-X Cell of *Drosophila melanogaster* and Synchronous Initiation. *The Journal of Cell Biology* 74: 168-180.

Deng, H., Bao, X., Cai, W., Blacketer, M.J., Belmont, A.S., Girton, J., Johansen, J., and Johansen, K.M. 2008. Ectopic histone H3S10 phosphorylation causes chromatin structure remodeling in Drosophila. *Development* 135: 699-705.

Deng, H., Cai, W., Wang, C., Lerach, S., Delattre, M., Girton, J., Johansen, J., and Johansen, K.M. 2010. JIL-1 and Su(var)3-7 interact genetically and counteract each other's effect on position-effect variegation in Drosophila. *Genetics* 185: 1183-1192.

Deng, H., Zhang, W., Bao, X., Martin, J.N., Girton, J., Johansen, J., and Johansen, K.M. 2005. The JIL-1 kinase regulates the structure of Drosophila polytene chromosomes. *Chromosoma* 114: 173-182.

Drobic, B., Perez-Cadahia, B., Yu, J., Kung, S.K., and Davie, J.R. 2010. Promoter chromatin remodeling of immediate-early genes is mediated through H3 phosphorylation at either serine 28 or 10 by the MSK1 multi-protein complex. *Nucleic Acids Res* 38: 3196-3208.

Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., and Reuter, G. 2004. Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. *Genes Dev* 18: 2973-2983.

Forsburg, S.L. 2004. Eukaryotic MCM Proteins: Beyond Replication Initiation. *Microbiology and Molecular Biology Reviews* 68: 109-131.

Ilves, I., Petojevic, T., Pesavento, J.J., and Botchan, M.R. 2010. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell* 37: 247-258.

Ivaldi, M.S., Karam, C.S., and Corces, V.G. 2007. Phosphorylation of histone H3 at Ser10 facilitates RNA polymerase II release from promoter-proximal pausing in Drosophila. *Genes Dev* 21: 2818-2831.

Jin, Y., Wang, Y., Johansen, J., and Johansen, K.M. 2000. JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the Male Specific Lethal (MSL) dosage compensation complex. *The Journal of Cell Biology* 149: 1005-1010.

Jin, Y., Wang, Y., Walker, D.L., Dong, H., Conley, C., Johansen, J., and Johansen, K.M. 1999. JIL-1: A novel chromosomal tandem kinase implicated in transcriptional regulation in *Drosophila*. *Mol Cell* 4: 129-135.

Johansen, K.M. 1996. Dynamic Remodeling of Nuclear Architecture During the Cell Cycle. *Journal of Cellular Biochemistry* 60: 289-296.

Johansen, K.M., Johansen, J., Baek, K.-H., and Jin, Y. 1996. Remodeling of Nuclear Architecture During the Cell Cycle in *Drosophila* embryos. *Journal of Cellular Biochemistry* 63: 268-279.

Karam, C.S., Kellner, W.A., Takenaka, N., Clemmons, A.W., and Corces, V.G. 2010. 14-3-3 mediates histone cross-talk during transcription elongation in Drosophila. *PLoS Genet* 6: e1000975.

Lerach, S., Zhang, W., Deng, H., Bao, X., Girton, J., Johansen, J., and Johansen, K.M. 2005. JIL-1 kinase, a member of the male-specific lethal (MSL) complex, is necessary for proper dosage compensation of eye pigmentation in Drosophila. *Genesis* 43: 213-215.

Li, Y., Cai, W., Wang, C., Yao, C., Bao, X., Deng, H., Girton, J., Johansen, J., and Johansen, K.M. 2013. Domain requirements of the JIL-1 tandem kinase for histone H3 serine 10 phosphorylation and chromatin remodeling in vivo. *J Biol Chem* 288: 19441-19449.

Moyer, S.E., Lewis, P.W., and Botchan, M.R. 2006. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A* 103: 10236-10241.

Nowak, S.J., and Corces, V.G. 2000. Phosphorylation of histone H3 correlates with transcriptionally active loci. *Genes Dev* 14: 3003-3013.

Randell, J.C., Fan, A., Chan, C., Francis, L.I., Heller, R.C., Galani, K., and Bell, S.P. 2010. Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7. *Mol Cell* 40: 353-363.

Rath, U., Ding, Y., Deng, H., Qi, H., Bao, X., Zhang, W., Girton, J., Johansen, J., and Johansen, K.M. 2006. The chromodomain protein, Chromator, interacts with JIL-1 kinase and regulates the structure of Drosophila polytene chromosomes. *J Cell Sci* 119: 2332-2341.

Regnard, C., Straub, T., Mitterweger, A., Dahlsveen, I.K., Fabian, V., and Becker, P.B. 2011. Global analysis of the relationship between JIL-1 kinase and transcription. *PLoS Genet* 7: e1001327.

Samson, R.Y., and Bell, S.D. 2013. MCM loading--an open-and-shut case? *Mol Cell* 50: 457-458.

Snyder, M., Huang, X.Y., and Zhang, J.J. 2009. The minichromosome maintenance proteins 2-7 (MCM2-7) are necessary for RNA polymerase II (Pol II)-mediated transcription. *J Biol Chem* 284: 13466-13472.

Wang, C., Li, Y., Cai, W., Bao, X., Girton, J., Johansen, J., and Johansen, K.M. 2014. Histone H3S10 phosphorylation by the JIL-1 kinase in pericentric heterochromatin and on the fourth chromosome creates a composite H3S10phK9me2 epigenetic mark. *Chromosoma* 123: 273-280.

Wang, Y., Zhang, W., Jin, Y., Johansen, J., and Johansen, K.M. 2001. The JIL-1 Tandem Kinase Mediates Histone H3 Phosphorylation and is required for maintenance of chromatin structure in Drosophila. *Cell* 105: 433-443.

Zee, B.M., Alekseyenko, A.A., McElroy, K.A., and Kuroda, M.I. 2016. Streamlined discovery of cross-linked chromatin complexes and associated histone modifications by mass spectrometry. *Proc Natl Acad Sci U S A*.

Zhang, W., Deng, H., Bao, X., Lerach, S., Girton, J., Johansen, J., and Johansen, K.M. 2006. The JIL-1 histone H3S10 kinase regulates dimethyl H3K9 modifications and heterochromatic spreading in Drosophila. *Development* 133: 229-235.

Supplementary Information

Figure S1: Replicate analysis of histone post-translational modifications from bulk histones in upSET mutant cells

A) Relative quantification (1.0 = 100%) of H4K5K8K12K16 acetyl patterns in *upSET* mutant cell lines (G3 and B2) with respect to the parental S2 line. Trends observed with these replicates are the same as in Figure 2-8.

B) Deconvolution of monoacetyl H4 from (A) to identify which residue carries the acetyl mark. There is a slight trend for enrichment of H4K16ac in the *upSET* mutants cells compared to other residues. Trends observed with these replicates are the same as in Figure 2-8.

C) Relative quantification (1.0 = 100%) of H3K9K14 PTM patterns in *upSET* mutant cell lines (G3 and B2) with respect to the parental S2 line. The H3K9me2/me3 marks are depleted from *upSET* mutant S2 cells, suggestive of an effect on heterochromatin. Trends observed with these replicates are the same as in Figure 2-8.

Figure S1 (Continued)

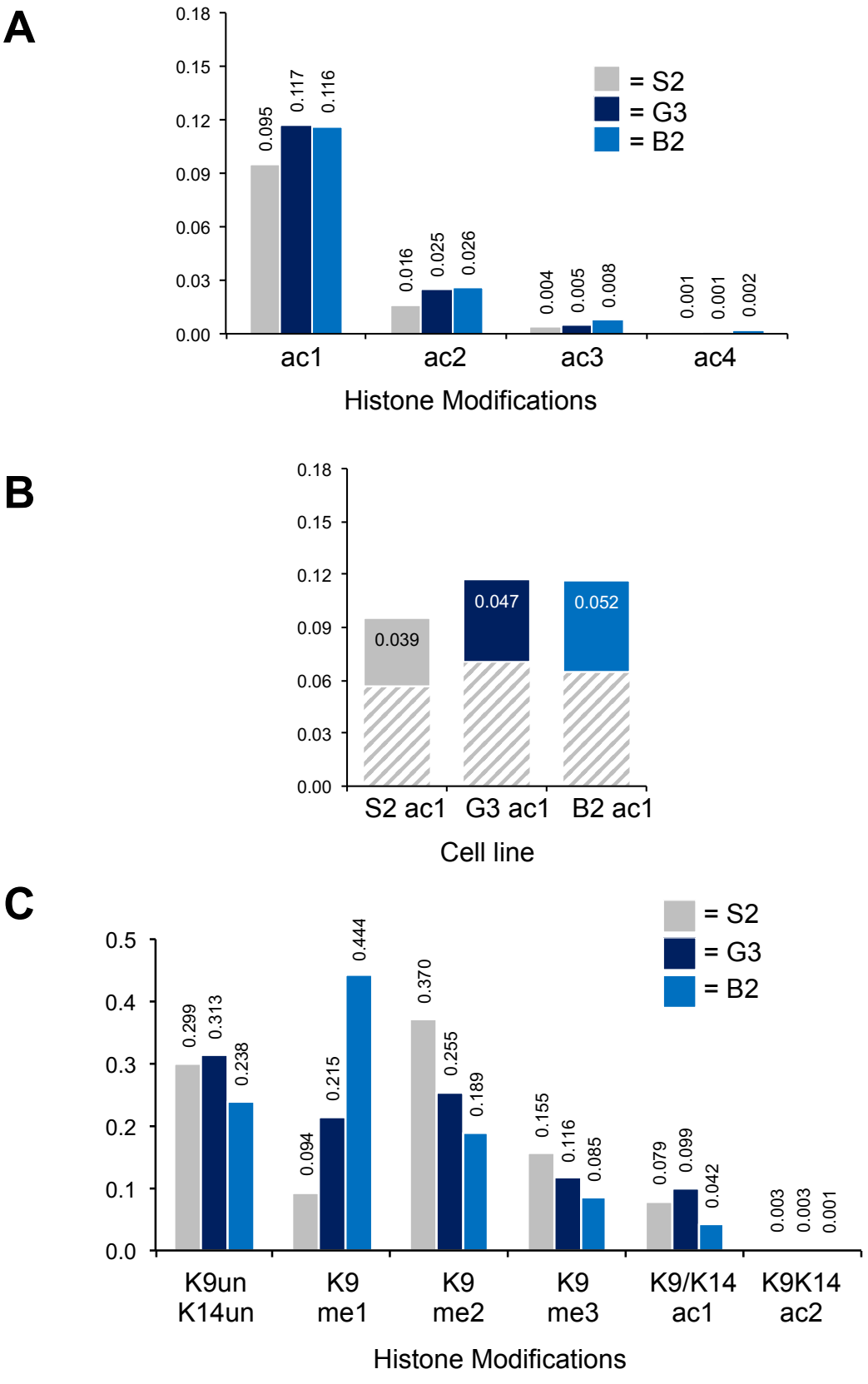


Table S1: BioTAP-tagged Fs(1)h associated proteins in S2 cells

BioTAP-N-Fs(1)h					Fs(1)h-C-BioTAP			
Gene Name	Peptides	Total Peptides	Input (S2) Total peptides		Gene Name	Peptides	Total Peptides	Input (S2) Total peptides
fs(1)h	58	613	7		fs(1)h	57	374	0
Hcf	45	110	6		ACC	44	71	30
Top2	45	85	6		Mtor	43	59	6
MESR4	40	87	0		Hcf	37	73	0
Mtor	34	42	30		CycT	34	69	6
rno	33	60	0		Top2	32	44	0
bip2	32	66	0		MESR4	27	51	0
east	30	49	0		kis	26	30	0
Taf1	28	48	0		Hsc70-4	25	60	1
wapl	27	48	0		CG1815	23	36	0
CG2247	26	51	0		bip2	23	34	0
CG1815	26	40	1		Nelf-A	20	34	0
Mi-2	25	36	2		MED1	20	31	2
kis	25	26	0		Mi-2	19	25	0
SMC1	23	30	0		CG8677	19	21	0
Cap	23	29	0		CG4751	18	23	0
CG8677	23	27	0		glu	18	20	12
mor	22	45	0		CG7946	17	30	0
Ncoa6	22	26	0		SMC2	17	22	0
CG7946	21	50	12		mor	16	27	17
Hsc70-4	21	44	0		Hsp83	16	24	0
Taf2	21	34	0		rno	16	23	0
dre4	21	34	16		wapl	16	20	0
Iswi	21	30	11		Nipped-B	16	18	16
trr	21	24	0		dre4	15	21	0
E(bx)	21	22	8		cher	15	15	41
MED1	20	28	0		Lam	14	20	0
brm	20	26	8		Cap	14	16	0
CG4747	19	33	12		Ncoa6	14	15	12
CycT	19	33	0		CG4747	13	25	0
SA	19	30	0		CG2247	13	25	0
sno	19	24	0		CG4564	13	24	11
Taf6	18	29	2		Iswi	13	19	6
Nipped-B	18	22	0		woc	13	19	0
Lam	17	24	41		east	13	17	5
e(y)3	17	21	0		lid	13	17	0
Sin3A	17	21	1		SMC1	13	15	0
dom	17	18	0		CG1910	12	24	0
rad50	16	21	0		betaTub56D	12	22	13
skd	16	19	0		Act5C	12	22	4
glu	16	18	0		Jafrac1	12	22	0
woc	15	28	6		Ef2b	12	18	0
Nelf-A	15	23	0		rad50	12	17	13
bon	15	19	0		pzg	12	16	0
MED26	15	18	0		simj	12	16	0
ptip	15	17	0		sno	12	15	0
MEP-1	15	17	0		CG12702-RA	12	14	25
Taf5	14	23	0		CG42232	12	13	0
fbl6	14	22	0		emb	12	13	0
Taf7	14	22	0		ATPCL	12	13	17
Gnf1	14	19	16		mod	11	22	7
Su(z)2	14	18	0		14-3-3epsilon	11	20	0
RpII215	14	17	0		nej	11	19	0
CG34422	14	16	0		RpA-70	11	16	12
CG42232	14	16	25		Hsc70-3	11	16	8

Tfb1	13	28	1		pont	11	14	0
CG1910	13	23	0		tou	11	13	12
pzg	13	19	13		Chc	11	13	8
osa	13	19	0		E(bx)	11	13	0
CG1647	13	17	0		RpII215	11	11	0
ACC	13	15	0		Ef1alpha48D	10	33	0
lid	13	14	5		fbl6	10	16	0
Rif1	13	14	0		Psc	10	13	2
MBD-R2	13	13	2		Taf6	10	13	0
CG15439	12	26	0		e(y)3	10	13	0
Act5C	12	26	13		enok	10	13	4
CG4564	12	20	0		rept	10	13	8
pont	12	20	8		RfC4	10	12	4
MED14	12	19	0		Ubqn	10	12	0
Hsc70-3	12	18	12		Taf1	10	12	3
CG5098	12	18	0		RpII140	10	12	0
Ubqn	12	16	4		CG34422	10	11	0
hay	12	16	0		Taf7	10	11	1
SMC2	12	16	0		Sin3A	10	10	2
RnrL	12	15	0		Jra	9	16	0
Bap60	12	14	0		TH1	9	14	0
Smr	12	13	0		Bub3	9	14	8
RpII140	12	12	3		Hrb27C	9	13	0
fu2	11	24	0		MED23	9	12	0
Bub1	11	23	0		ncd	9	11	0
Bub3	11	21	0		MED17	9	11	0
nej	11	15	0		blanks	9	11	5
rept	11	14	4		Bap55	9	11	0
hang	11	13	0		ptip	9	11	5
Kdm2	11	12	0		Aac11	9	11	8
tlk	11	11	0		brm	9	11	1
Nipped-A	11	11	0		MTA1-like	9	11	0
Jra	10	23	2		Smr	9	10	0
Ef1alpha48D	10	21	0		Bub1	9	10	0
Top1	10	19	0		MED26	9	10	0
row	10	19	0		SA	9	10	0
14-3-3zeta	10	16	3		Smc5	9	10	0
Ssrp	10	16	0		Rif1	9	9	0
ebi	10	15	0		TER94	9	9	0
Chro	10	13	7		spen	9	9	0
Psc	10	12	0		Cdk9	8	16	5
kto	10	12	0		Rpd3	8	15	3
Br140	10	12	0		14-3-3zeta	8	13	1
simj	10	11	0		Tfb1	8	12	0
mip120	10	10	0		DNApol-alpha180	8	11	28
CG6905	10	10	4		zip	8	10	0
His2B	9	17	19		polybromo	8	9	0
MED17	9	16	0		Sap130	8	9	0
kay	9	15	0		pds5	8	9	0
Prp8	9	13	13		mre11	8	9	0
polybromo	9	12	0		hang	8	8	2
Sap130	9	11	0		MBD-R2	8	8	0
Hsc70-5	9	11	5		Psi	8	8	20
D19A	9	11	0		His4	7	25	0
BRWD3	9	9	0		NELF-B	7	15	1
pps	9	9	0		Pen	7	13	0
Cul-4	9	9	2		zfh1	7	13	4
His4	8	54	20		sqd	7	12	0
His2Av	8	33	8		Chd64	7	12	0
Rpd3	8	21	5		alphaTub84B	7	11	0
CG6543	8	21	3		kay	7	11	0
CtBP	8	18	0		MED14	7	11	0

Dref	8	17	0		fu2	7	11	1
Bap55	8	16	5		mod(mdg4)	7	10	0
NELF-B	8	15	0		Hel25E	7	10	0
scu	8	13	6		Su(var)2-10	7	9	0
Dek	8	13	0		osa	7	9	0
vtd	8	12	0		Bap60	7	9	0
CG16711	8	12	0		CtBP	7	9	0
14-3-3epsilon	8	12	7		regucalcin	7	9	0
e(y)1	8	12	0		Nelf-E	7	9	0
mrn	8	11	0		Ssrp	7	9	0
mod	8	11	17		IntS9	7	9	5
Bap170	8	10	1		ERp60	7	9	0
pds5	8	10	0		trr	7	8	0
tou	8	9	0		tai	7	8	0
ear	8	9	0		CG16711	7	8	0
Klp3A	8	9	0		Hsc70Cb	7	8	3
dikar	8	9	0		RfC3	7	8	0
Cp190	8	9	12		dikar	7	8	0
Cdk7	8	9	0		Hop	7	8	2
Ef2b	8	9	0		BcDNA.LD23634	7	7	0
nito	8	8	7		skd	7	7	0
spen	8	8	0		ebi	7	7	0
Jarid2	8	8	0		Karybeta3	7	7	16
cg	7	19	0		Gnf1	7	7	0
PNUTS	7	12	0		Mcm3	7	7	0
srp	7	10	0		kto	7	7	8
mod(mdg4)	7	10	1		His2Av	6	16	3
MED8	7	10	0		CG6543	6	15	19
Cdk9	7	10	0		His2B	6	14	1
lola	7	9	2		Droj2	6	11	0
Hmt4-20	7	9	0		Cyp1	6	10	0
Snr1	7	9	3		TfIIS	6	9	1
MED23	7	9	0		su(f)	6	9	0
BEST:LD07122	7	9	0		Rm62	6	8	0
Pen	7	9	1		Klp3A	6	8	0
wah	7	8	0		eIF-4a	6	8	0
lilli	7	8	0		CG5524	6	8	5
CG1737-RA	7	8	0		Hrb87F	6	7	3
zf30C	7	8	0		caz	6	7	0
mip130	7	8	0		CG4785	6	7	1
Utx	7	8	0		Mdh2	6	7	0
gfzf	7	7	2		Dref	6	7	10
CstF-64	7	7	0		pAbp	6	7	6
Sfmbt	7	7	0		scu	6	7	0
tefu	7	7	0		IntS1	6	7	0
bel	7	7	4		Nipped-A	6	7	4
Smc5	7	7	0		pUf68	6	6	0
CG5524	7	7	0		TfIIA-L	6	6	0
Ars2	7	7	14		MEP-1	6	6	23
dalao	6	10	0		sle	6	6	1
Rm62	6	9	0		Rat1	6	6	3
RfC38	6	9	0		vkg	6	6	4
MED6	6	9	0		CG6905	6	6	0
CG7154	6	9	0		RnrL	6	6	0
Lpt	6	8	0		msk	6	6	0
MED19	6	8	0		sxc	6	6	2
RfC4	6	8	8		Cul-4	6	6	0
Chd64	6	8	0		Klp10A	6	6	0
CG10333	6	8	0		IntS6	6	6	0
TfIIS	6	7	0		MED16	6	6	0

BcDNA.LD23634	6	7	2		msps	6	6	0
CG4751	6	7	0		Dek	5	10	0
nop5	6	7	0		Hsp68	5	10	0
Ote	6	7	14		srp	5	9	0
Kap-alpha3	6	7	0		CstF-64	5	9	0
mre11	6	7	0		dalao	5	8	0
stwl	6	7	0		cg	5	8	12
su(f)	6	7	1		Ef1gamma	5	8	3
CoRest	6	7	0		Gp93	5	8	0
enok	6	7	0		Srp54	5	8	7
Top3alpha	6	7	0		smt3	5	8	0
JIL-1	6	7	0		Pp2A-29B	5	8	0
rdx	6	7	0		Taf2	5	7	0
Orc1	6	7	0		MED6	5	7	3
IntS1	6	7	0		SelD	5	7	2
Psi	6	7	0		CG16941	5	7	0
Nelf-E	6	6	0		MRG15	5	6	2
TfIIFalpha	6	6	0		lola	5	6	12
CG4849	6	6	3		Cp190	5	6	0
CG4785	6	6	0		PNUTS	5	6	3
CG6686	6	6	0		Snr1	5	6	10
hyx	6	6	3		Su(var)205	5	6	2
Caf1-180	6	6	0		CG8149	5	6	0
CG2469	6	6	0		mrn	5	6	0
Caf1	5	12	0		bon	5	6	5
wds	5	12	2		l(2)35Df	5	6	0
skpA	5	12	0		Asun	5	6	1
Taf8	5	11	0		dpa	5	5	2
His1	5	10	0		Tctp	5	5	0
Pc	5	9	0		Mcm2	5	5	0
Su(var)2-10	5	8	0		dom	5	5	0
TfIIA-L	5	8	0		su(s)	5	5	0
MED7	5	8	0		Taf5	5	5	0
row	5	8	12		Jarid2	5	5	1
Cyp1	5	8	0		Bap170	5	5	0
Sce	5	7	0		scra	5	5	0
Hrb27C	5	7	8		yip2	5	5	0
yip2	5	7	0		barr	5	5	0
CG7185	5	7	6		Kdm2	5	5	0
Taf4	5	7	0		E(Pc)	5	5	7
Hrb87F	5	6	5		nito	5	5	0
E(Pc)	5	6	0		CG5913	5	5	4
CG12702-RA	5	6	0		mdy	5	5	0
sle	5	6	23		Sfmbt	5	5	0
Spt5	5	6	2		Chd1	5	5	11
RfC3	5	6	3		CG5931	5	5	0
Aac11	5	6	5		Mcm5	5	5	0
CG5931	5	6	11		Sav1	4	31	2
Rcd5	5	6	4		wds	4	7	0
Prp19	5	6	6		Caf1	4	6	0
RpI135	5	6	6		crp	4	6	0
Map205	5	6	8		Top1	4	6	0
Pep	5	6	16		RfC38	4	6	2
CG3542	5	6	0		Gale	4	6	0
blanks	5	6	0		dco	4	5	0
Rrp1	5	5	0		CoRest	4	5	0
l(3)mbt	5	5	0		Lpt	4	5	0
Hsp83	5	5	17		mip120	4	5	0
Synd	5	5	0		BEST:LD07122	4	5	0
atms	5	5	0		CG3523	4	5	0
mus210	5	5	0		JIL-1	4	5	0
lds	5	5	0		Top3alpha	4	5	0

Map60	5	5	0		yki	4	5	0
Rpl1	5	5	4		CG17078-RA	4	5	6
ERp60	5	5	5		CG7185	4	5	0
sxc	5	5	0		CG8839	4	5	15
CG2118	5	5	0		SF2	4	5	0
IntS4	5	5	0		Gdi	4	5	0
ncd	5	5	0		Cdk7	4	5	0
CG14438	5	5	0		mus101	4	5	0
CG17002	5	5	0		shot	4	5	0
smt3	4	15	7		nop5	4	4	0
Mat1	4	9	0		ear	4	4	2
sqd	4	7	4		Spt5	4	4	2
SelD	4	7	3		Sym	4	4	0
Xpd	4	7	0		CG1647	4	4	0
TH1	4	7	0		Usp7	4	4	4
IntS9	4	7	0		bel	4	4	3
caz	4	6	3		tho2	4	4	7
Su(var)205	4	6	10		Chro	4	4	0
alphaTub84B	4	6	0		bin3	4	4	0
CG1622-RA	4	6	1		Larp7	4	4	0
CG9007	4	6	0		CG5098	4	4	0
MED21	4	6	0		CG2807	4	4	0
CG4203	4	6	0		TfIIFbeta	4	4	0
betaTub56D	4	6	0		Mcm6	4	4	0
CG6379	4	6	0		Cpsf73	4	4	0
CG9302	4	5	0		pic	4	4	3
ph-p	4	5	1		IntS11	4	4	0
ttk	4	5	0		shn	4	4	0
Scm	4	5	0		CG32737	4	4	0
MTA1-like	4	5	1		Caf1-180	4	4	5
Ssl1	4	5	1		RpS14a	4	4	0
Brd8	4	5	0		CG5384	4	4	0
Klp10A	4	5	0		Moe	4	4	5
CG10600	4	5	0		Hsc70-5	4	4	0
CG2199	4	5	9		defl	4	4	14
crp	4	5	0		Ars2	4	4	0
CG11076	4	5	1		DNApol-epsilon	4	4	0
CG17471	4	5	0		pho	4	4	0
Rcd1	4	5	3		Cul-3	4	4	8
Chd3	4	5	0		Map205	4	4	0
CG18292-RA	4	4	0		CG2118	4	4	0
Alh	4	4	0		Su(z)2	4	4	2
CG4282	4	4	0		CAP-D2	4	4	0
Adf1	4	4	2		CG6227	4	4	0
CG9797	4	4	3		Ctf4	4	4	0
IntS6	4	4	0		Acf1	4	4	0
zfh1	4	4	0		CG33129	3	6	0
D19B	4	4	0		pnr	3	6	2
Mes-4	4	4	0		tsr	3	6	16
Sgf29	4	4	0		Pep	3	5	0
barr	4	4	0		cg	3	5	0
Caf1-105	4	4	0		Hsc70-1	3	5	0
CG6841	4	4	3		CG30122	3	5	0
Chd1	4	4	0		Mat1	3	5	6
CG17078-RA	4	4	0		Prp19	3	5	0
CG3163	4	4	1		CG1646	3	5	0
ct	4	4	0		CG18292-RA	3	4	0
CG10631	4	4	0		vtd	3	4	0
l(2)37Cb	4	4	1		Zn72D	3	4	0
CG12909	4	4	0		Sce	3	4	0
snama	4	4	0		Jupiter	3	4	0
CG15356-RA	4	4	0		hay	3	4	0

Ref1	4	4	7		CG17233	3	4	0
CAP-D2	4	4	2		Alh	3	4	0
CG16941	4	4	2		MED21	3	4	0
Hsc70-1	3	10	0		Hmt4-20	3	4	1
ix	3	6	0		U2af50	3	4	0
Hrb98DE	3	6	0		dod	3	4	0
Trl	3	5	0		ash2	3	4	0
Eaf6	3	5	0		Nap1	3	4	0
CG3689	3	5	0		IntS4	3	4	0
MED30	3	5	1		MED24	3	4	0
MED11	3	5	0		row	3	4	0
U2af50	3	5	1		IntS12	3	4	0
Jafrac1	3	5	4		CrebB-17A	3	4	0
MED16	3	5	0		Rrp1	3	4	1
CG4951	3	5	2		mof	3	4	0
Srp54	3	5	0		bl	3	4	0
CG8142	3	5	2		lilli	3	4	0
abs	3	5	4		atms	3	4	0
CstF-50	3	5	0		CG8507	3	4	0
MRG15	3	4	0		MED7	3	4	0
Jupiter	3	4	0		CG34422	3	4	0
CG10209	3	4	0		Map60	3	4	0
CG17912	3	4	0		Taf4	3	4	0
SF2	3	4	15		Smu1	3	4	0
mof	3	4	1		msn	3	3	0
Nurf-38	3	4	0		Mes2	3	3	0
DMAP1	3	4	0		CG3689	3	3	1
cactin	3	4	0		CG1316	3	3	0
CG33129	3	4	0		RpII33	3	3	0
RpII33	3	4	0		Akap200	3	3	3
CG3548	3	4	1		CG6841	3	3	0
B52	3	4	0		ran	3	3	0
IntS2	3	4	0		lig	3	3	0
PNUTS	3	4	0		Art1	3	3	0
CG7593	3	4	0		Dip-B	3	3	0
bl	3	4	0		TfIIFalpha	3	3	0
CG8507	3	4	0		Hr96	3	3	4
cl	3	4	0		Fs(2)Ket	3	3	0
vkg	3	4	3		Nurf-38	3	3	10
tth	3	3	0		nonA	3	3	0
Mes2	3	3	0		swm	3	3	0
MED24	3	3	0		Bre1	3	3	0
MED25	3	3	0		Xpd	3	3	0
IntS12	3	3	0		GstT1	3	3	0
Jupiter	3	3	0		CG2469	3	3	0
CG8436	3	3	3		Khc	3	3	0
Gp93	3	3	3		gro	3	3	0
hang	3	3	0		IntS10	3	3	0
Wdr82	3	3	0		Rad23	3	3	0
nbs	3	3	0		p47	3	3	0
CG7379	3	3	0		MED4	3	3	0
wee	3	3	0		Vap-33-1	3	3	2
CG3995	3	3	0		Nup50	3	3	0
CG1888	3	3	0		CG4203	3	3	0
omd	3	3	0		CG6693	3	3	3
mtSSB	3	3	3		CG13185	3	3	0
Asun	3	3	0		ct	3	3	12
regucalcin	3	3	0		Nup358	3	3	0
Hnf4	3	3	0		CG15356-RA	3	3	0
Bj1	3	3	9		cmet	3	3	1
pnr	3	3	0		CG3163	3	3	0
MED4	3	3	0		cdc2	3	3	0
Asx	3	3	0		Utx	3	3	0

psq	3	3	0		MED25	3	3	0
CG32737	3	3	0		ns1	3	3	0
CG12592	3	3	0		CG12909	3	3	0
CG11873	3	3	0		Vinc	3	3	0
Su(var)2-HP2	3	3	0		CG8036	3	3	0
Rad23	3	3	0		Cg25C	3	3	0
Ino80	3	3	1		sip1	3	3	0
BcDNA.LD23876	3	3	0		RecQ4	3	3	3
CG11266	3	3	1		CG10139	3	3	0
CG3680	3	3	1		l(2)04524	3	3	0
Su(var)2-HP2	3	3	4		msi	3	3	3
CG34422	3	3	0		Nlp	3	3	0
Cpsf100	3	3	1		rin	3	3	1
Rbbp5	3	3	0		mus209	3	3	0
ncm	3	3	0		BRWD3	3	3	0
swm	3	3	0		MED27	3	3	2
opa1-like	3	3	0		CG4598	3	3	0
Cbp80	3	3	3		CG10977	3	3	0
Orc5	3	3	0		Rpn3	3	3	3
pic	3	3	0		Vha26	3	3	0
Pitslre	3	3	2		mus210	3	3	4
CG3756	3	3	2		Mad1	3	3	0
eIF-4a	3	3	0		sgg	3	3	1
Sym	3	3	2		Hlc	3	3	2
IntS3	3	3	0		DnaJ-1	3	3	5
CG12288	3	3	0		CtBP	2	6	0
prod	3	3	2		Hmu	2	6	0
Rat1	3	3	1		skpA	2	4	0
CG5787	3	3	22		CG32767	2	4	0
Orc6	3	3	0		Kap-alpha3	2	4	0
ash1	3	3	0		ix	2	4	0
CG4266	3	3	0		PNUTS	2	4	0
Dp	3	3	0		MED11	2	4	0
CG10984	3	3	0		RpL40	2	4	0
Rbp2	3	3	0		La	2	4	1
XRCC1	3	3	0		Ssl1	2	4	0
meso18E	3	3	0		FKBP59	2	4	0
apt	3	3	0		CG2990-RB	2	4	0
msps	3	3	0		MED10	2	3	0
MED10	2	8	0		Wdr82	2	3	0
RpL40	2	8	0		Pa1	2	3	0
ttk	2	4	0		CG2691	2	3	0
crol	2	4	0		CG7593	2	3	0
Hsp68	2	4	0		tlk	2	3	0
lin-52	2	4	0		CG4282	2	3	1
MED27	2	4	0		Set	2	3	0
DnaJ-1	2	4	2		BEAF-32	2	3	3
ran	2	4	0		rump	2	3	0
Dsp1	2	3	0		Tcp-1eta	2	3	1
CG5181	2	3	0		Cpsf100	2	3	0
cg	2	3	0		Ssb-c31a	2	3	0
Tim13	2	3	0		e(y)1	2	3	0
Trap1	2	3	1		bic	2	3	0
hoip	2	3	0		CG15439	2	3	0
CG6227	2	3	0		D1	2	3	3
yki	2	3	0		smid	2	3	0
WRNexo	2	3	0		gpp	2	3	0
Zif	2	3	0		CG5958-RA	2	3	0
CG10139	2	3	3		Etl1	2	3	0
CG12112-RA	2	3	0		CG17233	2	3	0
CG17493	2	3	0		how	2	3	0
Rbp1	2	3	1		mamo	2	2	0

CG13708	2	3	0		kuk	2	2	0
SmB	2	3	4		Cpsf160	2	2	0
Nlp	2	3	3		CG11964	2	2	0
ns1	2	3	0		wah	2	2	0
EloA	2	3	2		ttk	2	2	0
Bre1	2	3	0		ush	2	2	0
CG3909	2	3	0		Pc	2	2	1
CG10977	2	3	0		CG1240	2	2	0
Mcm3	2	3	0		ps	2	2	0
Taf11	2	3	0		Taf8	2	2	3
CG13096	2	3	15		Zn72D	2	2	0
Cg25C	2	3	0		CG10600	2	2	0
Fib	2	3	15		CG11120	2	2	0
msn	2	2	0		mxc	2	2	0
CG18004-RB	2	2	0		CG7288	2	2	0
CG8149	2	2	2		CG13349	2	2	0
sbb	2	2	0		CG2051	2	2	0
CG33097	2	2	2		CG30122	2	2	0
MBD-like	2	2	0		Mes-4	2	2	0
TfIIB	2	2	0		CaBP1	2	2	0
Pa1	2	2	0		CG11164	2	2	0
CG10672	2	2	0		CG17912	2	2	0
CG7987	2	2	0		Rop	2	2	0
Nop56	2	2	0		Spt6	2	2	2
mip40	2	2	0		Rpb5	2	2	0
Nap1	2	2	0		CG2852	2	2	0
CG11120	2	2	0		chic	2	2	0
CG17233	2	2	0		H	2	2	0
Taf12	2	2	0		ncm	2	2	0
dod	2	2	0		lost	2	2	0
Zn72D	2	2	0		CG12592	2	2	0
chif	2	2	0		eIF4G	2	2	0
Smu1	2	2	0		Hdac3	2	2	0
Tango4	2	2	0		MED15	2	2	0
Hpr1	2	2	3		REG	2	2	0
CG7971	2	2	0		pav	2	2	2
BtbVII	2	2	0		CG8142	2	2	0
CG11999	2	2	4		beag	2	2	0
CaBP1	2	2	0		CG10979	2	2	0
CrebB-17A	2	2	0		CG3542	2	2	0
CG2807	2	2	0		Jupiter	2	2	0
Droj2	2	2	1		CG31075	2	2	0
Mdh2	2	2	1		G9a	2	2	0
pUf68	2	2	4		omd	2	2	0
CG4617	2	2	0		ttk	2	2	0
Ef1gamma	2	2	12		Cap	2	2	0
Nop60B	2	2	0		D19A	2	2	0
Acn	2	2	0		DMAP1	2	2	0
ball	2	2	0		CG4813	2	2	0
SF1	2	2	0		Rae1	2	2	0
Su(var)3-3	2	2	0		CG4266	2	2	0
msn	2	2	0		Scm	2	2	0
MED15	2	2	0		ham	2	2	0
tho2	2	2	3		Rbbp5	2	2	0
Taf10	2	2	0		velo	2	2	1
CG12262	2	2	0		Dlc90F	2	2	0
CG31291	2	2	0		chinmo	2	2	3
scaf6	2	2	0		CG4849	2	2	2
D1	2	2	0		Adf1	2	2	3
Prp3	2	2	2		CG14641	2	2	0
cora	2	2	0		Rpt4	2	2	0
glo	2	2	3		bur	2	2	0
gro	2	2	0		Tpr2	2	2	0

gpp	2	2	0		EndoGI	2	2	0
emb	2	2	0		Ald	2	2	1
dik	2	2	0		Rbp1	2	2	0
Lint-1	2	2	0		Taf11	2	2	0
Etl1	2	2	0		hoip	2	2	0
CG5514-RA	2	2	0		Br140	2	2	0
Karybeta3	2	2	0		CG17544-RA	2	2	0
Mad1	2	2	4		CG15784	2	2	4
CG9135	2	2	0		abs	2	2	0
Ada2b	2	2	0		egg	2	2	6
CG2691	2	2	0		CG1371	2	2	2
Vap-33-1	2	2	0		Pabp2	2	2	0
HmgZ	2	2	0		CG1888	2	2	0
CG11107	2	2	10		Gapdh1	2	2	0
su(s)	2	2	0		CG4887	2	2	0
pen	2	2	4		alphaCop	2	2	0
Acf1	2	2	0		CG17002	2	2	0
Tip60	2	2	0		Cap-D3	2	2	0
lola	2	2	0		Eb1	2	2	0
CG7477	2	2	0		Acon	2	2	0
lolal	2	2	0		CycB	2	2	0
REG	2	2	0		fru	2	2	3
CG8289	2	2	3		Cbp80	2	2	0
lost	2	2	0		CG1640	2	2	0
eRF1	2	2	0		MED8	2	2	0
pho	2	2	0		Fancd2	2	2	2
scra	2	2	0		fl(2)d	2	2	0
Fen1	2	2	0		EfTuM	2	2	6
XNP	2	2	0		betaTub97EF	2	2	0
l(2)35Df	2	2	5		Hnf4	2	2	0
IntS11	2	2	3		Rcd1	2	2	0
CG17446	2	2	0		CG7918	2	2	7
Hel25E	2	2	0		Ref1	2	2	0
msk	2	2	0		zf30C	2	2	1
CG7920	2	2	0		Su(var)3-9	2	2	8
Orc4	2	2	0		Pdi	2	2	0
HP1b	2	2	0		cora	2	2	0
CG5913	2	2	0		CycH	2	2	0
CG3884	2	2	0		psq	2	2	31
shn	2	2	0		alpha-Spec	2	2	0
CG13773	2	2	0		CG14438	2	2	0
CG4598	2	2	2		BcDNA.LD23876	2	2	1
CkIIalpha	2	2	0		skap	2	2	0
CG10979	2	2	0		Eaf6	2	2	0
ctrip	2	2	0		Gas41	2	2	0
Tfb4	2	2	0		lolal	2	2	0
AGO2	2	2	3		Spt20	2	2	0
thoc7	2	2	1		CG8858	2	2	0
SmD2	2	2	1		pps	2	2	6
CG9601-RA	2	2	0		RpI135	2	2	0
CG33695	2	2	0		CG3884	2	2	0
CG15514	2	2	0		PHGPx	2	2	0
Set1	2	2	2		CG7427	2	2	0
mad2	2	2	0		Fkbp13	2	2	0
CG18600	2	2	3		apt	2	2	0
G9a	2	2	0		HP1b	2	2	0
Hdac3	2	2	0		row	2	2	3
clu	1	5	0		CG8436	2	2	0
CtBP	1	4	5		Nop56	2	2	0
CG30116	1	4	0		Trip1	2	2	0
BEAF-32	1	3	0		B52	2	2	0
baf	1	3	2		CG5118	2	2	0

htl	1	3	0		hyd	2	2	0
CG9293	1	2	0		CG6904	2	2	0
CG1240	1	2	1		CG7239	2	2	0
CG4360	1	2	0		Ssdp	2	2	10
CG10324	1	2	0		CG11107	2	2	0
CG6272	1	2	0		exba	2	2	2
MED20	1	2	0		coil	2	2	0
SmE	1	2	2		Rpn9	2	2	0
fd68A	1	2	0		CG8929	2	2	0
Sfmbt	1	2	0		CG5516	2	2	0
CG1316	1	2	1		CG30116	1	4	1
CHES-1-like	1	2	0		CG1622-RA	1	3	1
TfIIEalpha	1	2	0		Trap1	1	3	0
CG17233	1	2	0		clu	1	3	0
x16	1	2	3		CG12391	1	2	0
NLaz	1	2	0		crol	1	2	4
Rpb10	1	2	0		SmB	1	2	0
ctp	1	2	0		CG14715	1	2	0
CG3491	1	2	0		CG5181	1	2	0
TotM	1	2	0		CG10543	1	2	1
CG18766	1	1	0		Tim8	1	2	1
E2f2	1	1	0		MED30	1	2	0
Su(var)2-10	1	1	0		lin19	1	2	0
Rpb5	1	1	2		WRNexo	1	2	1
CG4788-RA	1	1	0		RanGap	1	2	0
Rtf1	1	1	0		CG10543	1	2	0
CG10555	1	1	0		CG17385	1	2	0
Nacalpha	1	1	1		msn	1	2	0
Saf-B	1	1	2		SC35	1	2	0
CG12391	1	1	0		CG13992-RA	1	2	0
CG14715	1	1	0		htl	1	2	0
CG7288	1	1	0		sec23	1	2	0
Or7a	1	1	0		Cct1	1	2	0
CycA	1	1	0		CkIIalpha	1	2	1
ph-d	1	1	0		SmD3	1	2	0
chm	1	1	0		CG5554	1	2	0
lola	1	1	0		CstF-50	1	2	0
Usp7	1	1	0		CG9987	1	2	0
JHDM2	1	1	0		ebd1	1	2	4
U2af38	1	1	1		lark	1	2	0
CG5118	1	1	0		RpL3	1	2	0
CG13992-RA	1	1	0		CG3918	1	2	0
Max	1	1	0		CG11058	1	2	0
Cap	1	1	0		Trl	1	1	0
smid	1	1	3		tsh	1	1	0
tai	1	1	0		CG12104	1	1	1
CG15784	1	1	0		ph-p	1	1	0
CG16838	1	1	0		mip130	1	1	0
su(Hw)	1	1	4		CG9293	1	1	10
CG4221	1	1	0		koi	1	1	0
defl	1	1	0		CG5590	1	1	0
nonA	1	1	10		mad2	1	1	0
CG5694	1	1	0		Cap-G	1	1	1
Pof	1	1	0		Nacalpha	1	1	4
mrt	1	1	0		CG5641	1	1	0
Octbeta3R	1	1	0		CG5514-RA	1	1	0
noi	1	1	2		Tango4	1	1	0
mdy	1	1	4		ATbp	1	1	2
HmgD	1	1	0		HP1c	1	1	0
CG10543	1	1	0		CG7154	1	1	0
HP5	1	1	0		pnt	1	1	0
kin17	1	1	0		CG8435	1	1	0
asf1	1	1	0		Cctgamma	1	1	0

Xrp1	1	1	0		CG7477	1	1	0
ush	1	1	0		Su(var)2-10	1	1	0
CG5180	1	1	0		NTPase	1	1	0
ph-d	1	1	0		raps	1	1	0
ACC	1	1	0		CG42668	1	1	0
CG8097	1	1	0		RpL10Ab	1	1	0
Arp8	1	1	0		CG6418	1	1	0
lat	1	1	0		wal	1	1	0
mus209	1	1	1		Bx42	1	1	0
alpha-Adaptin	1	1	6		Tim13	1	1	1
exba	1	1	0		CG11266	1	1	0
Spt6	1	1	0		Phf7	1	1	0
CG14641	1	1	3		CG5525	1	1	0
SC35	1	1	0		srpk79D	1	1	0
YL-1	1	1	0		Sfmbt	1	1	0
tsr	1	1	2		CG3011	1	1	0
BtbVII	1	1	0		CG2926	1	1	0
Rpb4	1	1	0		Tm1	1	1	14
lin19	1	1	0		RpS8	1	1	0
Su(var)3-9	1	1	1		CG3198	1	1	0
Myb	1	1	1		esc	1	1	0
CG6674	1	1	0		kst	1	1	10
ebd1	1	1	0		RpLP0	1	1	0
Cpsf73	1	1	0		mle	1	1	0
beag	1	1	0		Rpn5	1	1	0
CG16972	1	1	0		Fer1HCH	1	1	0
RpL11	1	1	1		CG9135	1	1	0
da	1	1	0		Cypl	1	1	0
Fech	1	1	0		Rb97D	1	1	0
Cdk8	1	1	0		fd68A	1	1	0
chic	1	1	0		l(3)mbt	1	1	0
d4	1	1	0		meso18E	1	1	1
NHP2	1	1	7		FANCI	1	1	0
CG6700	1	1	0		lola	1	1	0
CG10907	1	1	0		eIF-4E	1	1	0
dgt1	1	1	0		CG12262	1	1	2
sec6	1	1	0		SmE	1	1	0
ssx	1	1	0		ACC	1	1	0
CG9915	1	1	0		CG3760	1	1	3
Cpsf160	1	1	0		CG3605	1	1	0
CycB	1	1	0		jumu	1	1	0
CG7637	1	1	1		CG4045	1	1	0
Atu	1	1	0		dUTPase	1	1	0
Zn72D	1	1	3		geminin	1	1	7
LKR	1	1	0		CG11092	1	1	0
PSR	1	1	0		Cct5	1	1	0
RpS19a	1	1	9		Hrb98DE	1	1	0
cathD	1	1	1		Clic	1	1	1
pit	1	1	5		U2af38	1	1	0
CG4360	1	1	0		da	1	1	0
CG10462	1	1	0		mip40	1	1	0
Cypl	1	1	0		CG10932	1	1	0
exo70	1	1	0		nxf2	1	1	2
CG9866	1	1	0		gfzf	1	1	1
l(1)G0148	1	1	0		RpL11	1	1	7
TFAM	1	1	0		CG7194	1	1	0
Ef1beta	1	1	6		CG13142	1	1	0
Dlc90F	1	1	1		Su(dx)	1	1	0
CG15107	1	1	0		CG2862	1	1	0
CG4291-RA	1	1	0		CG1218	1	1	0
CG17118	1	1	0		lds	1	1	0
Hmu	1	1	0		Gapdh2	1	1	0

CG5727	1	1	0		CG3224	1	1	0
CG12608	1	1	0		CG5198	1	1	0
Arf79F	1	1	2		cindr	1	1	0
fru	1	1	0		CG8258	1	1	3
CG7692	1	1	0		Nup214	1	1	0
Atac1	1	1	0		ade5	1	1	0
CG3363	1	1	0		CG11208	1	1	0
CG15477	1	1	0		tum	1	1	0
Cyp12d1-d	1	1	0		CG10324	1	1	0
Deaf1	1	1	0		Rpn2	1	1	0
shn	1	1	0		gammaCop	1	1	0
CG13624	1	1	0		CG12258	1	1	0
CG2051	1	1	0		CG3226	1	1	0
Tctp	1	1	2		Aats-asp	1	1	0
CG9987	1	1	0		CG6388	1	1	0
Kdm4B	1	1	0		Nop60B	1	1	0
Clic	1	1	0		CG13865	1	1	0
Rbf2	1	1	0		HmgZ	1	1	1
futsch	1	1	0		Rab1	1	1	0
sgg	1	1	0		CG7945	1	1	0
FKBP59	1	1	0		CG6179	1	1	0
Nup50	1	1	2		zip	1	1	9
MrgBP	1	1	0		CG2199	1	1	0
crol	1	1	0		MED28	1	1	0
CG4400	1	1	0		lat	1	1	0
Arf102F	1	1	0		BEST:LD13441	1	1	0
Nup358	1	1	12		CG8223	1	1	0
TfIIFbeta	1	1	0		Asx	1	1	0
Rab2	1	1	0		awd	1	1	0
put	1	1	0		Lasp	1	1	0
schlank	1	1	3		tefu	1	1	0
Ets97D	1	1	0		polo	1	1	0
CG16865	1	1	0		Mnn1	1	1	0
Rrp6	1	1	0		SF1	1	1	0
su(w[a])	1	1	0		mus309	1	1	0
CG14442	1	1	0		Msh6	1	1	0
parvin	1	1	0		Mcm7	1	1	0
Rpb7	1	1	1		Slob	1	1	0
CG14868	1	1	0		Parp	1	1	0
msl-3	1	1	3		CG6272	1	1	0
CG3847	1	1	0		ss	1	1	1
CG10274	1	1	0		CG9911	1	1	4
sdk	1	1	0		pen	1	1	0
Cf2	1	1	0		px	1	1	3
Gasp	1	1	0		sqh	1	1	0
CG8119	1	1	0		eIF3-S4-1	1	1	4
sog	1	1	0		Rcd5	1	1	0
cher	1	1	0		CG3339	1	1	0
CG5941	1	1	0		PSR	1	1	0
Msh6	1	1	0		RpS3	1	1	3
CG12163	1	1	1		x16	1	1	0
CG4707	1	1	0		EG:25E8.3	1	1	0
lark	1	1	4		CG9281	1	1	0
CG9331	1	1	0		CG8478	1	1	0
Gdi	1	1	0		Syp	1	1	0
CG5641	1	1	4		dgt5	1	1	0
CG10993	1	1	0		CG5004-RA	1	1	0
Klp67A	1	1	0		HmgD	1	1	0
CG14712	1	1	0		CG7033	1	1	0
yemalpha	1	1	0		sec13	1	1	0
Hira	1	1	0		Fen1	1	1	0
l(1)G0020	1	1	9		RPA2	1	1	6
msl-2	1	1	0		Ef1beta	1	1	0

CG7564	1	1	2		BtbVII	1	1	0
His2A	1	1	7		Tango1	1	1	0
TfIIEbeta	1	1	0		ort	1	1	0
koi	1	1	10		betaTub60D	1	1	0
mxc	1	1	0		CG10631	1	1	2
CG14073	1	1	0		noi	1	1	0
Klp61F	1	1	0		r-l	1	1	0
BCL7-like	1	1	0		CG33695	1	1	5
CG10754	1	1	0		Mlc-c	1	1	0
Cdk12	1	1	0		CG14544	1	1	2
eIF-2alpha	1	1	0		CG9018	1	1	2
dco	1	1	0		Prp3	1	1	0
Nopp140	1	1	12		Anp32a	1	1	0
Spn27A	1	1	0		CG7987	1	1	0
Prp3	1	1	0		Hel89B	1	1	0
CG32708	1	1	0		Msp-300	1	1	0
Mhc	1	1	0		MrgBP	1	1	0
Doa	1	1	0		CG6621	1	1	0
Cas	1	1	0		CG7504	1	1	0
Tpr2	1	1	0		CG6700	1	1	0
CG42668	1	1	0		Fmr1	1	1	1
CG11123	1	1	3		CG4164	1	1	0
RpS14a	1	1	5		Cklalpha	1	1	0
Cyp28a5	1	1	0		CG12065	1	1	0
CG13223	1	1	0		Mms19	1	1	2
cmet	1	1	0		baf	1	1	0
CG6961	1	1	0		BtbVII	1	1	0
CG5071	1	1	0		eIF-1A	1	1	0
CG18815	1	1	0		eco	1	1	0
CG6015	1	1	0		CG17209	1	1	0
cic	1	1	0		Spt3	1	1	0
CG8840	1	1	0		CG10565	1	1	0
SmD3	1	1	1		CG12608	1	1	0
CG7376	1	1	0		D19B	1	1	0
bys	1	1	3		HP5	1	1	9
Tfb5	1	1	0		l(1)G0020	1	1	0
CG3702	1	1	0		CG3995	1	1	0
Gale	1	1	2		mrt	1	1	3
CG31955	1	1	0		CG8289	1	1	7
CG2233	1	1	0		CG13097	1	1	0
CG11058	1	1	0		CG14442	1	1	0
CG8635	1	1	0		msb1l	1	1	1
Tbp	1	1	0		Cpr	1	1	0
Fancd2	1	1	0		CG7920	1	1	2
gp210	1	1	17		Arf79F	1	1	0
MED31	1	1	0		MTA1-like	1	1	2
comt	1	1	0		Doa	1	1	3
chn	1	1	0		glo	1	1	17
CG5789-RA	1	1	0		gp210	1	1	0
Lap1	1	1	0		CG9213	1	1	0
Mtr3	1	1	0		Crc	1	1	3
CG42550	1	1	0		CG9797	1	1	0
CG15207	1	1	0		CG11788	1	1	0
OstDelta	1	1	8		BCL7-like	1	1	0
YT521-B	1	1	0		CG13887	1	1	0
Mcm5	1	1	0		CG17493	1	1	0
Uvrag	1	1	0		ytr	1	1	0
CG4164	1	1	1		RpL30	1	1	0
Hop	1	1	0		CG9004	1	1	1
cdc2	1	1	0		RpS12	1	1	0
RpS28b	1	1	2		Rrp4	1	1	0
NAT1	1	1	0		Pros45	1	1	0
					Actr13E	1	1	0

					wee	1	1	9
					Bj1	1	1	8
					OstDelta	1	1	0
					AnnIX	1	1	0
					CG9302	1	1	0
					Cpr64Ac	1	1	0
					CG1703	1	1	0
					tweek	1	1	0
					GstE3	1	1	0
					ZAP3	1	1	0
					RpL23	1	1	0
					CG10754	1	1	0
					Cnx99A	1	1	0
					ctrip	1	1	3
					Hpr1	1	1	2
					CG3756	1	1	0
					CG1109	1	1	2
					CG33097	1	1	0
					ps	1	1	0
					CG31368	1	1	0
					gem	1	1	1
					CG7637	1	1	0
					pan	1	1	0
					CG3511	1	1	0
					CG13690	1	1	0
					lmp	1	1	0
					Rpn12	1	1	0
					Tina-1	1	1	0
					ras	1	1	0
					beta'Cop	1	1	0
					MED31	1	1	0
					TfIIEalpha	1	1	0
					Uch-L3	1	1	0
					CG5044	1	1	0
					RanBPM	1	1	0
					Dsp1	1	1	0
					rempA	1	1	0
					l(1)G0156	1	1	0
					Dpy-30L1	1	1	0
					CG7338	1	1	0
					ph-d	1	1	0
					CG17737	1	1	0
					exo84	1	1	14
					Ote	1	1	0
					CG3815	1	1	12
					RpL7	1	1	0
					CG4936	1	1	0
					CG5174-RA	1	1	0
					Arp8	1	1	0
					CG10132	1	1	0
					IntS2	1	1	0
					CG7770	1	1	0
					ver	1	1	0
					CG13923	1	1	1
					Nup205	1	1	0
					Nsf2	1	1	0
					Cat	1	1	0
					eIF2B-delta	1	1	0
					CG13850	1	1	0
					pnt	1	1	0
					Lint-1	1	1	0
					Trn-SR	1	1	0
					tex	1	1	0

					Rrp6	1	1	0
					IntS8	1	1	0
					Jheh2	1	1	0
					Ahcy13	1	1	0
					MAPk-Ak2	1	1	0
					scf	1	1	0
					Brd8	1	1	0
					Eig71Ec	1	1	0
					Orc5	1	1	2
					prod	1	1	0
					CG4617	1	1	0
					eIF-2alpha	1	1	0
					CG1979	1	1	0
					kay	1	1	0
					CG12168	1	1	0
					abo	1	1	0
					CG5933	1	1	0
					CG14005	1	1	0
					cactin	1	1	0
					CG9727	1	1	0
					CG7834	1	1	3
					eIF3-S10	1	1	0
					shi	1	1	0
					robl	1	1	0
					Es2	1	1	0
					Hsp60	1	1	0
					CG1965	1	1	0
					Arf102F	1	1	0
					Sec16	1	1	0
					alien	1	1	0
					Prx6005	1	1	0
					MAGE	1	1	0
					snama	1	1	0
					Klp67A	1	1	0
					CG9667	1	1	0
					l(1)G0230	1	1	3
					pea	1	1	0
					Adk3	1	1	0
					Atx2	1	1	0
					Pax	1	1	0
					Acn	1	1	0
					Cyt-c-p	1	1	0
					l(1)G0007	1	1	0
					Myo31DF	1	1	0
					Rab2	1	1	0
					CLIP-190	1	1	0
					Irbp	1	1	0
					CG15514	1	1	0
					DNApol-alpha73	1	1	0
					CHES-1-like	1	1	0
					GstO2	1	1	0

Table S2: BioTAP-Jil-1 associated proteins in S2 and Kc cells

S2 BioTAP-Jil-1				Kc BioTAP-Jil-1			
Gene Name	Peptides	Total Peptides	Input S2 (Total peptides)	Gene Name	Peptides	Total Peptides	Input S2 (Total peptides)
Hsp83	29	68	1	Top2	29	44	57
Hsc70-4-RA	27	81	1	Hsc70-4-RA	28	70	1
JIL-1-RB	26	114	0	JIL-1-RB	27	106	28
CG7946	25	72	1	Hsp83	27	40	0
CG11198	24	27	0	CG11198	27	33	16
Top2	21	35	1	CG4747	25	52	7
CG4747	18	46	1	CG7946	25	85	9
Lam	16	20	0	RpA-70	16	20	44
Hop	16	21	0	betaTub56D	15	21	18
regucalcin-RD	14	30	0	eIF-4a-RD	14	21	5
betaTub56D	14	20	1	lid	13	14	31
Act5C	14	33	1	Act5C	12	18	7
eIF-4a-RD	14	29	1	dpa	12	15	9
stwl	14	22	0	14-3-3epsilon	11	22	10
14-3-3epsilon	13	36	1	rept	11	12	12
CG2118	13	24	0	pont	11	12	0
RpA-70	12	18	0	CG16972	11	12	0
CG10630	11	25	1	Cap	11	11	2
Jafrac1	11	26	1	stwl	10	11	0
ATPCL	10	12	0	CG10630	10	18	12
zip	10	10	1	TER94-RC	10	13	8
dpa	10	12	0	Mi-2	10	10	5
lid	10	10	1	Hcf	10	13	6
Hsc70Cb	10	15	0	Bap55	10	15	44
CG1815	10	10	0	Ef2b	10	12	13
mor	9	10	1	dre4	9	10	0
cher	9	10	1	scra	9	10	0
Iswi	9	10	1	CG2118	9	11	2
TER94-RC	9	15	1	Psa-RA	9	9	4
Ef2b	9	13	1	14-3-3zeta	8	16	25
Chd64	8	13	0	mod	8	9	1
Chc	8	8	1	Chd64	8	8	4
14-3-3zeta	8	17	1	Sin3A	8	10	24
Clic	8	20	0	His2B	8	29	50
CG12030	8	14	1	Lam	8	10	13
Cyp1	8	18	0	Cyp1	8	12	4
CG8149	8	12	1	Hop	8	9	15
cher	8	12	1	ERp60	8	10	42
Tcp-1zeta	8	10	0	His4	8	48	3
Tcp-1eta	8	10	0	Ef1alpha100E	8	17	8
Mcm3	8	10	0	E(bx)	8	10	4
CtBP	8	11	1	Ahcy13	8	11	28
Ef1alpha100E	8	20	1	alphaTub84D	7	12	12
dre4	8	10	1	Iswi	7	10	7
alpha-Spec	8	8	1	CtBP	7	11	3
alphaTub84D	7	14	1	ATPCL	7	8	0
mod	7	8	1	CG5524	7	7	6
Ef1gamma	7	14	1	Mcm2	7	7	5
Dek	7	9	1	CG6084	7	9	7
Rpd3	7	13	1	Droj2	6	6	2
Mcm2	7	7	0	Dek	6	7	14
ERp60	7	8	1	pzg	6	7	15
EndoGI	7	9	0	ran	6	8	7
Khc	7	7	0	Jafrac1	6	8	6
His4	7	40	1	Top2	6	8	14
Mcm5	7	7	0	Tcp-1eta	6	8	0
Mtor	7	7	1	CG8149	6	7	6

Hcf	7	8	1	Hsc70Cb	6	8	3
l(2)35Df	7	7	1	CG10417	6	6	12
rad50-RB	7	8	1	Mcm5	6	8	0
sqd	6	15	1	CG1815	6	6	12
Rm62-RE	6	7	1	CG8258	6	6	35
Mi-2	6	6	1	alpha-Spec	6	6	14
PyK	6	8	0	sqd	5	12	16
pont	6	6	1	Ef1alpha48D-RA	5	10	0
CG8036	6	8	0	JIL-1	5	13	12
dUTPase	6	10	0	Su(var)205	5	5	4
Bap55	6	7	1	Dref	5	6	6
Vinc	6	6	0	Rpd3	5	7	0
kst	6	7	1	EndoGI	5	5	10
anon1A3	5	6	1	mor	5	5	8
JIL-1	5	9	0	Hrb98DE	5	8	5
CG7033	5	5	0	Idh	5	6	8
Su(var)205	5	7	1	Uba1	5	5	9
Jupiter-RD	5	5	0	Gapdh1	5	7	2
CG4699-RA	5	5	1	CG42232	5	5	1
Droj2	5	9	0	Eip55E	5	6	29
CG4236	5	11	1	Rm62-RE	5	6	3
Hrb98DE	5	10	1	rad50-RB	5	5	3
Su(var)2-HP2	5	6	1	CG9273	5	5	5
Mcm6-RA	5	6	0	CG4236	5	9	6
His2B	5	19	1	CG2982	5	5	9
eIF4G	5	5	1	row	5	5	0
CaBP1	5	5	0	CG34422	5	6	0
rept	5	6	1	CG11006	5	5	16
ran	5	8	1	CG8036	5	8	0
pic	5	5	0	MBD-R2	5	6	0
CG8258	5	9	0	tum	5	5	5
Map205	5	6	1	pic	5	6	3
Ars2	5	8	1	Mcm7	5	6	0
T-cp1	5	5	0	anon1A3	4	4	13
Ef1alpha48D-RA	4	8	1	Ef1gamma	4	4	4
Hrb87F	4	4	1	wds	4	4	4
Ald	4	4	0	Hrb87F	4	8	0
CG5384	4	5	0	CG4699-RA	4	5	5
Sin3A	4	4	1	smt3	4	8	1
SelD	4	6	1	CG3226	4	4	0
Gapdh1	4	7	0	Clic	4	5	37
FKBP59	4	5	0	Mtor	4	4	15
Ef1beta	4	6	1	CG7033	4	4	2
ncd	4	6	1	CG6554	4	4	0
mof	4	4	0	Msh6	4	4	0
RfC4	4	5	1	geminin	4	5	0
eEF1delta	4	6	1	pav	4	4	8
CG3609	4	4	0	Tctp	4	5	1
CG42232	4	4	1	Nlp	4	8	2
RfC38	4	5	1	CG2051	4	4	0
Klp3A	4	4	0	dom	4	4	0
Eno	4	4	0	CoRest-RF	4	4	6
Hel25E	4	8	1	His2Av	4	7	5
La	4	6	0	RfC38	4	5	9
glu	4	4	0	Gnf1	4	4	13
granny-smith	4	5	0	Ssrp	4	4	16
Cct5	4	4	0	His1:CG33852-RA	4	6	6
MBD-R2	4	5	1	Hel25E	4	5	2
L(2)04524	4	6	0	dUTPase	4	5	5
Karybeta3	4	4	0	Mcm3	4	5	2

ebi-RA	4	4	1	CG17337	4	4	13
Gnf1	4	5	1	PyK	4	5	3
Pep	4	6	1	CaBP1	4	4	2
E(bx)	4	5	1	caz	3	4	3
gro	4	4	0	CG1240	3	5	3
brm	4	4	1	Jupiter-RD	3	3	0
pds5	4	5	0	CG9894-RB	3	3	4
msl-3	4	4	1	Mi-2	3	3	15
CG16972	4	4	0	Chc	3	3	2
CG9894-RB	3	3	0	Gapdh2	3	3	3
Nurf-38	3	6	1	TfIIS	3	3	0
Snr1	3	3	1	CG10630	3	3	6
TfIIS	3	3	0	Nurf-38	3	4	2
Ubi-p63E	3	6	1	Snr1	3	3	0
CG1910	3	3	0	CG2827-RA	3	3	18
CG17337	3	6	0	Bj1	3	3	2
wds	3	4	1	Klp10A	3	4	2
CG3226	3	3	0	osa-RC	3	3	7
shrb	3	4	1	emb	3	3	1
Nlp	3	10	1	CG16817	3	3	3
Ote	3	3	1	Top1-RB	3	3	12
CG1640	3	3	0	Karybeta3	3	3	3
kis	3	3	1	La	3	4	5
REG	3	4	0	mod(mdg4)-RD	3	3	2
emb	3	3	0	Mcm6-RA	3	4	5
pzg	3	3	1	Usp7	3	3	5
CG9135	3	4	0	Bre1	3	3	0
Dis3	3	3	1	r-l	3	3	6
CG17544	3	3	0	Ubi-p63E	3	8	6
Eb1	3	3	0	Ef1beta	3	3	10
smt3	3	11	1	Pdi	3	3	1
CG5362	3	5	0	Prosalpha1	3	3	6
Ssrp	3	3	1	l(2)35Df	3	3	5
CG2051	3	3	0	CG9135	3	3	4
ade5	3	6	0	Ssb-c31a	3	4	12
cher	3	5	1	128up	3	3	0
zip-RC	3	3	1	MEP-1	3	3	4
His1:CG33852-RA	3	4	1	RfC4	3	3	0
bur	3	3	0	glu	3	3	8
tsr	3	4	1	Ars2	3	4	2
Gapdh2	3	4	0	Cdc37-RA	3	3	5
CG9987	3	3	0	cul-4	3	4	0
Idh	3	4	0	Klp3A	3	3	2
msk	3	3	0	ncd	3	3	2
mus101	3	3	0	CG6227	3	3	6
Map60	3	3	1	Pp2A-29B	3	3	2
CG6673	3	4	0	crp	3	4	0
His2Av	3	5	1	mre11	3	3	3
woc	3	3	1	ade5	3	3	0
zip-RC	3	3	1	spel1	3	3	2
CG6554	3	6	0	MTA1-like	3	3	8
CG6388	3	6	0	Rack1	3	3	0
CG5363	3	3	0	DppIII	3	3	0
Rrp1	3	3	1	CG10722	3	3	0
Moe-RB	3	7	1	Smc5	3	3	6
CG5931	3	3	1	Chro	3	4	5
exba	3	3	0	CG13185	3	3	5
AnnIX	3	3	1	brm	3	3	10
mod(mdg4)-RD	3	3	1	Ca-P60A	3	3	0
Cat	3	3	0	gro	3	3	4
CG11107	3	6	1	DNApol-delta	3	3	0
Usp7	3	3	0	MRG15	2	2	1

shot	3	3	1	CG7911	2	2	2
Lasp	3	3	0	emb	2	2	14
CG12171	3	3	0	Nap1	2	3	12
Hsc70-3	3	3	1	Hrb27C	2	2	1
CG9273	3	3	0	glu	2	3	1
BcDNA:GH10614	3	3	0	Anp32a	2	3	0
Pgk	3	3	0	shrb	2	2	0
Pp2A-29B	3	3	0	CG2446	2	2	2
CG34422	3	3	0	zip	2	2	15
Bre1	3	3	0	growl	2	2	3
CG6783	3	4	0	Taf6	2	2	7
DppIII	3	5	0	Acf1	2	2	9
CG1516	3	3	0	mdy	2	2	9
CG2767	3	3	0	Tcp-1zeta	2	2	5
CG3714	3	3	0	CG5384	2	2	0
CG8290	3	3	0	CG15141	2	3	3
lark	3	3	1	D1	2	2	3
GstD1	3	4	0	BcDNA:LD22118	2	2	2
caz	2	2	1	REG	2	2	0
CG7911	2	2	0	Klp3A	2	2	3
betaTub97EF	2	2	1	Set	2	3	0
CG14715	2	3	0	CG3523	2	2	2
SMC4	2	2	0	glu	2	2	2
CG1316	2	2	0	dom-RD	2	2	0
Mcm7	2	3	0	CG3609	2	2	5
chic	2	3	0	FKBP59	2	2	0
Nap1	2	2	0	Spt5	2	2	0
Uba1	2	2	0	HmgZ	2	2	5
sle	2	2	1	bl	2	2	0
growl	2	3	1	CG4045	2	2	2
Tctp	2	2	1	tsr	2	2	1
mus209	2	2	0	Mcm3	2	2	11
RfC3	2	2	1	Rrp1	2	2	2
xl6	2	2	1	SelD	2	2	1
osa-RC	2	2	1	Mlc-c	2	2	0
Spt6-RA	2	2	0	cg	2	2	0
Pros28.1	2	2	0	CG13350	2	3	3
Mlc-c	2	3	1	CG9797	2	2	4
CoRest-RF	2	2	1	Cas	2	2	0
Dref	2	3	1	Rad23	2	2	0
Ssb-c31a	2	3	0	alien	2	2	4
CG4045	2	2	0	CG3689	2	2	1
CG2947	2	3	0	Caf1-180-RB	2	2	4
CG5899	2	2	1	CG6783	2	3	1
CG5524	2	2	0	Dsp1	2	2	0
Mi-2	2	3	0	mof	2	2	4
nudC	2	2	0	Map60	2	2	2
D1	2	3	1	Khc	2	2	5
Klp10A	2	3	0	Psi	2	2	0
cg	2	2	0	Caf1-105	2	2	0
scra	2	2	0	barr	2	2	2
CG3523	2	2	0	Kap-alpha3	2	2	4
CG3011	2	2	0	Bap60	2	2	0
SMC1	2	2	0	CG3523-RA	2	2	2
sec13	2	2	0	Rcd5	2	2	0
RnrL	2	3	0	CG6133	2	2	0
Vha26	2	2	1	Asx	2	2	1
CG6905	2	2	1	Pp4-19C	2	3	0
CG11006	2	3	1	ImpL3-RA	2	2	0
eIF-5A	2	3	1	sds22	2	2	8
CG7194	2	2	1	His2A	2	6	2

RhoGAP92B-RA	2	2	0	CG1737	2	2	1
sqh	2	2	1	mus209	2	2	4
Cam-RB	2	2	1	CkIalpha	2	3	0
Pros26.4	2	2	0	RnrS	2	2	8
dom	2	2	0	sop	2	2	3
DnaJ-1	2	2	1	ebi-RA	2	2	2
Aac11	2	2	1	fs(1)h	2	2	0
Pen	2	2	0	Aats-ala	2	2	0
cora	2	2	0	pps-RA	2	2	4
CG1972	2	2	1	CG16941	2	2	0
awd	2	2	0	CG10600	2	2	0
mts	2	3	0	pds5	2	2	0
CG8223	2	2	0	GstE6	2	2	28
CG6523	2	2	0	LamC	2	2	5
Dhc64C	2	2	1	Pgk	2	2	0
swm	2	2	1	Cat	2	2	2
CG16817	2	2	0	SMC1	2	2	0
CG17665	2	2	0	CG1972	2	3	2
Elf	2	2	0	chic	2	2	0
Prp19	2	3	1	L(2)04524	2	2	9
MTA1-like	2	2	0	Cp190	2	2	1
CG13887	2	2	0	CG6418	2	3	0
CG2469	2	2	0	CG3680-RA	2	2	16
128up	2	2	1	CG11107	2	2	0
Vha55	2	4	0	CG12276	2	2	8
CG13350	2	2	0	Cctgamma	2	2	0
Prx2540-1	2	2	0	CG11980	2	2	0
CG2852	2	2	0	enok	2	2	0
Dsp1	2	2	0	CG6693	2	2	15
smid	2	2	1	CG13096	2	2	10
row	2	2	1	CG34422	2	2	2
Prp8	2	2	1	AnnIX	2	2	0
Nop60B-RC	2	2	1	Pros29	2	2	0
CG2827-RA	2	2	0	Uch	2	2	0
Rnp4F	2	2	0	EG:115C2.8	2	2	0
CG10527	2	2	0	GstD1	2	2	2
His2A	2	15	1	Rpn7	2	2	7
Dip-B	2	3	0	cul-2	2	2	3
CG5854	2	2	0	T-cp1	2	2	0
GstE6	2	2	0	CG7338	2	2	2
RhoGDI	2	2	0	DMAP1	2	2	3
Cortactin	2	2	0	mts	2	2	0
Cdc37-RA	2	2	0	CG6061	2	2	3
srp	2	2	0	lark	2	2	9
ScpX	2	2	0	RnrL	2	2	0
cul-4	2	2	1	l(2)35Bg	1	1	9
CG3909	2	3	0	CG30122	1	1	4
Tudor-SN	2	2	1	Hsp70Ba-RA	1	1	4
Top3alpha	2	2	0	Trap1	1	1	1
mle	2	2	1	CG30122-RB	1	1	3
Tm1-RA	2	2	0	CG5899	1	1	1
CG12259	2	2	0	CG17950	1	1	3
Gdi	2	2	0	betaTub97EF	1	1	0
aru	2	2	0	Mnf	1	1	2
Rcd1	2	2	1	Ubqn	1	1	0
DNApol-alpha60	2	2	0	kay-RA	1	1	0
CG5014	2	2	1	CG1910	1	1	2
CG10333	2	2	1	Crc	1	1	7
CG17493	2	2	0	Hsp27	1	1	1
Ahcy13	2	2	0	dre4	1	1	2
Past1	2	2	0	Pros28.1	1	1	0
CG3983-RB	2	2	1	CG5355	1	1	3

CG6418	2	2	0	CG5641	1	1	0
RpS18-RB	2	2	1	Map60	1	1	2
east-RB	2	2	0	bic	1	1	3
CG10080	2	2	0	Gp93	1	1	4
CG40045	1	1	0	RfC3	1	1	3
Hsp27	1	1	0	CG2446	1	1	0
Rpb5	1	1	1	CG34422	1	1	1
kay-RA	1	1	0	baf	1	1	1
CG30122	1	1	0	skpA	1	1	4
CG6995	1	1	1	Cbp80	1	1	5
Hsp70Ba-RA	1	1	0	eEF1delta	1	1	1
Trap1	1	1	0	CG6084	1	1	3
Anp32a	1	1	0	Pros45	1	1	5
mor	1	1	0	CG5525	1	1	0
Hrb27C	1	1	1	CG17912	1	1	1
CG6554	1	1	0	CG5355	1	1	11
CG7477	1	1	0	Fs(2)Ket	1	1	0
Gp93	1	1	1	vtd	1	1	0
baf	1	1	1	CG3760	1	1	7
msl-2	1	1	0	RpL24	1	1	1
NELF-A	1	1	0	mor	1	1	0
Aats-glupro	1	1	0	14-3-3zeta-RI	1	1	6
Bap60	1	1	1	Hsp26	1	1	3
CG4802	1	1	0	dalao	1	1	12
CG15141	1	1	0	nito	1	1	4
CG17273	1	1	0	Rpn9	1	1	0
CG5355	1	1	0	CG32626-RD	1	1	10
CG2862	1	1	0	RpS8	1	1	2
MED21	1	1	0	U2af50	1	1	2
Karybeta3	1	1	0	RpL11	1	1	2
Crc	1	1	0	Aats-thr	1	1	0
Df31	1	1	0	Rbp1-like	1	1	0
Klp3A	1	1	0	lig	1	1	0
dod-RA	1	1	0	Su(var)2-HP2	1	1	0
Su(var)3-9	1	1	0	M(2)21AB	1	1	3
alt	1	1	1	CG2469	1	1	3
Pcd-RA	1	1	0	smid	1	1	2
HmgZ	1	1	0	RpS12	1	1	0
DNApol-delta	1	1	0	Dox-A2	1	1	0
RpS7	1	1	0	U2af38	1	1	0
CG5355	1	1	0	CG6673	1	1	2
CG3760	1	1	0	CkIIbeta	1	1	1
alt	1	1	1	Ald	1	1	5
Fdh	1	1	0	Nipped-A	1	1	8
RpL10Ab	1	1	1	eIF-5A	1	1	1
Papss-RD	1	1	0	Sep1	1	1	15
c12.1	1	1	0	Pen	1	1	2
HP1c	1	1	1	CG1354-RB	1	1	8
Pdi	1	1	1	alphaCop	1	1	0
CG10139	1	1	1	tlk	1	1	1
wal	1	1	0	CG11089	1	1	1
dalao	1	1	1	mip130	1	1	0
glu	1	1	0	Su(var)2-HP2	1	1	0
RpL24	1	1	1	CG10990	1	1	6
skpA	1	1	0	hts	1	1	6
da-RA	1	1	0	Tm1-RA	1	1	0
U2af38	1	1	0	PNUTS	1	1	0
Set	1	1	0	CG6388	1	1	0
RpLP1	1	1	1	Tango4	1	1	1
Dp	1	1	0	CG12547	1	1	1
ALiX	1	1	0	ytr	1	1	1
Cpsf160	1	1	0	BcDNA:GH10614	1	1	0

emb	1	1	0	ftz-f1	1	1	6
CG6197	1	1	0	eIF-2alpha	1	1	1
Msh6	1	1	0	Vha44	1	1	2
sip2-RD	1	1	0	TER94	1	1	1
Kap-alpha3	1	1	1	Trx-2	1	1	0
mle	1	1	0	CG9987	1	1	1
Pros35	1	1	0	Cpsf160	1	1	1
BcDNA:LD22118	1	1	0	CG17255	1	1	1
CG3353	1	1	0	Clc	1	1	5
Pgd	1	1	0	CG7497-RB	1	1	2
Mms19	1	1	0	exba	1	1	30
smc1	1	1	1	Prp8	1	1	5
CHES-1-like-RB	1	1	0	gammaCop	1	1	0
TER94	1	1	0	CG9330	1	1	0
Top2	1	1	1	nbs	1	1	0
Prosbeta7	1	1	0	CG7375	1	1	0
CG5077	1	1	1	Jra	1	1	0
betaggt-II	1	1	0	mle	1	1	0
CG9617	1	1	0	barr	1	1	0
p38b	1	1	0	Rnp4F	1	1	0
14-3-3zeta-RI	1	1	0	Rab1	1	1	12
CG9232	1	1	0	beta-Spec	1	1	0
CG6693	1	1	0	Df31	1	1	0
Rpn1	1	1	0	swm	1	1	0
CG6089	1	1	1	CG6776	1	1	1
CG13185	1	1	1	Mcm7	1	1	15
#VALUE!	1	1	0	SF2	1	1	3
fs(1)h	1	1	1	Pabp2	1	1	4
CG9641	1	1	0	Rpt4-RA	1	1	4
Uba2	1	1	0	CG9156	1	1	1
RpS28b	1	1	1	Cam-RB	1	1	1
ash2	1	1	0	Chd3	1	1	2
CG6891-RA	1	1	0	Uba2	1	1	0
Lag1	1	1	1	CG32654	1	1	7
Argk	1	1	0	Prp19	1	1	0
LBR	1	1	1	Pros54	1	1	3
Pp4-19C	1	1	0	CG5363	1	1	0
Rpn9	1	1	0	CG4400	1	1	0
cher	1	1	0	HP1b	1	1	0
CG32955	1	1	0	CG10979	1	1	0
Rab11-RB	1	1	0	Bub3	1	1	5
Aats-asn	1	1	0	Arf79F	1	1	8
CG5602	1	1	0	RpL5	1	1	4
l(2)35Bg	1	1	0	Sym	1	1	0
CG8142	1	1	1	SMC4	1	1	2
r-l	1	1	0	CG5602	1	1	0
CstF-64	1	1	1	CG2118	1	1	0
Chd3	1	1	0	dom-RE	1	1	0
CG5525	1	1	0	CG10565	1	1	1
Nacalpha	1	1	0	CG4203	1	1	7
CG7497-RB	1	1	1	Nopp140	1	1	0
CG17259	1	1	0	CG6617	1	1	3
CG1354-RB	1	1	0	su(f)	1	1	2
eIF3-S4-1	1	1	0	Vha68-2	1	1	2
CG13349	1	1	0	R	1	1	2
M(2)21AB	1	1	0	Vinc	1	1	3
CG1240	1	1	0	CG8801-RA	1	1	0
cyp33	1	1	0	CG3415	1	1	0
dlg1	1	1	0	CG2909	1	1	22
Snx6	1	1	0	Hsc70-3	1	1	2
shn	1	1	0	CG18591	1	1	0
tum	1	1	0	MSBP	1	1	0

Taf6	1	1	1	CG7593	1	1	0
vib	1	1	0	CG5482	1	1	0
CG7375	1	1	0	dlg1	1	1	0
bic	1	1	0	CG3162	1	1	0
Klc	1	1	0	da-RA	1	1	0
Jra	1	1	1	CG13350	1	1	0
Jhd1	1	1	0	Prx2540-1	1	1	0
BRWD3	1	1	1	XNP	1	1	0
Trx-2	1	1	0	Gas41	1	1	5
CG13690	1	1	0	RpL23a	1	1	9
pelo	1	1	0	RpS18-RB	1	1	9
Pros45	1	1	0	CG13900	1	1	2
CG5222	1	1	0	Prosalpha5	1	1	0
CG2118	1	1	0	cher	1	1	1
CG9471	1	1	0	awd	1	1	1
Asx	1	1	0	Su(var)2-10	1	1	5
wapl	1	1	0	BcDNA.GH07921	1	1	0
hyx	1	1	1	Cortactin	1	1	0
Top1-RB	1	1	1	CG7878	1	1	0
Drp1	1	1	0	Rrp42	1	1	3
CG6776	1	1	0	CG12909	1	1	2
Bub3	1	1	0	Akap200	1	1	0
CG6379	1	1	0	VhaSFD	1	1	1
CG8677-RA	1	1	0	CG32528	1	1	10
CG1115	1	1	0	Eno	1	1	0
Cct1	1	1	0	Tps1	1	1	0
spel1	1	1	0	CG5174	1	1	0
Rack1	1	1	1	CG6876	1	1	0
mip130	1	1	0	e(y)1	1	1	0
Vha68-2	1	1	1	CHES-1-like-RB	1	1	0
sta-RD	1	1	1	CG14618	1	1	0
Clc	1	1	0	hyd	1	1	1
CG40801	1	1	0	pnut	1	1	0
CG7770	1	1	0	Parp	1	1	4
fs(1)N	1	1	0	eIF3-S4-1	1	1	5
poe	1	1	0	SmB	1	1	1
CG7182	1	1	0	CG7288	1	1	2
CG11164	1	1	0	CG7246	1	1	6
CG6061	1	1	1	xl6	1	1	5
CG9934	1	1	0	CG11844-RB	1	1	0
cactin	1	1	0	CG9934	1	1	9
Hel89B	1	1	0	CG2691-RA	1	1	0
CG1910	1	1	0	tum	1	1	5
RpS26	1	1	1	RpL6	1	1	0
eIF-2alpha	1	1	0	Uev1A	1	1	0
CG9156	1	1	0	Rpn11	1	1	6
Uev1A	1	1	0	r	1	1	0
Rpn11	1	1	0	CG7324	1	1	12
r	1	1	0	CG6543	1	1	15
CG7324	1	1	0	CG13096	1	1	0
CG6543	1	1	1	CG9286	1	1	0
CG13096	1	1	1	CG9331-RA	1	1	0
CG9286	1	1	0	Fcp1	1	1	0
CG9331-RA	1	1	0	CG11334	1	1	8
Fcp1	1	1	0	RpS14a	1	1	3
CG11334	1	1	0	betaCop	1	1	0
RpS14a	1	1	1	RpS8	1	1	0
betaCop	1	1	0	Sod	1	1	11
RpS8	1	1	1	Fs(2)Ket	1	1	0
Sod	1	1	0	Tfb1	1	1	0
Fs(2)Ket	1	1	1	CG11138	1	1	0
Tfb1	1	1	0	Nedd8	1	1	0

CG11138	1	1	0	S6k	1	1	0
Nedd8	1	1	0	CG1354	1	1	0
S6k	1	1	0	CstF-64	1	1	0
CG1354	1	1	0	CG12065	1	1	4
CstF-64	1	1	0	Cklalpha	1	1	0
CG12065	1	1	0	CG7154	1	1	0
Cklalpha	1	1	0	TfIIA-L	1	1	0
CG7154	1	1	0				
TfIIA-L	1	1	0				